



Intelligent System Based Assessments of Academic Journals

Pan Su

Supervisors: Prof. Qiang Shen
Dr. Changjing Shang

Ph.D. Thesis
Department of Computer Science
Institute of Mathematics, Physics and Computer Science
Aberystwyth University

August 17, 2015

Declaration and Statement

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed (candidate)

Date

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where **correction services**¹ have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed (candidate)

Date

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed (candidate)

Date

¹This refers to the extent to which the text has been corrected by others.

Abstract

The assessment of academic journals is becoming more and more critical to many scientific activities. Such assessment can have a significant influence upon many issues ranging from publishing articles in academic journals to determining which universities are qualified to undertake major research projects in a country. Recently, journal ranking has been introduced as an official research assessment tool in many countries. Both peer review and indicator based methods have been exploited for this task in an effort to provide quantitative tools for ranking, evaluating and comparing academic journals. A number of data-driven evaluation measures have been developed based on citations, downloads and other statistical aspects of journals, which have the advantages of being more objective, while consuming less time and finance compared with assessments which based on peer review by human experts.

In this thesis, several intelligent system based methods for journal ranking are presented. The proposed approaches mainly utilise data-driven techniques including: 1) clustering algorithms, which are able to detect groups of academic journals that have similar indicator scores; 2) fuzzy aggregations, which provide more flexible and reliable aggregation of impact indicators than the use of Euclidean and Manhattan distances in journal ranking; and 3) clustering ensembles, by which linguistic variables are introduced to support interpretive clustering of journals. In addition, the mathematical properties of Ordered Weighed Averaging (OWA) aggregation of fuzzy relations are exploited to enhance their application in clustering. Also, link-based fuzzy clustering ensembles are proposed to improve the accuracy and robustness of fuzzy clustering. A method for expediting fuzzy clustering ensembles is introduced to reduce the effort in dealing with the growth of data volumes. Systematic experimental results demonstrated that these methods are not only flexible and interpretable, but also accurate in capturing and reflecting the impact of academic journals.

The proposed approaches lead to a number of further developments. These include: selecting and grouping journal impact indicators; learning the weighting vector of OWA aggregation of fuzzy relations; applying fuzzy similarities to explore the boundary region of rough sets; and using the link-based fuzzy consensus function to support re-sampling based clustering ensemble. The proposed approaches for journal ranking can also be employed to solve clustering or ranking problems in applications other than journal assessment.

Acknowledgements

I would like to express my uppermost gratitude to my supervisors: Prof. Qiang Shen and Dr. Changjing Shang, for their motivation, encouragement, and guidance, which have been essential at all stages of my research.

I am grateful to Dr. Neil S. Mac Parthaláin, Dr. Richard Jensen, Dr. Ren Diao, Dr. Chaw Seng Woo, and Prof. Xizhao Wang for their constant inspiration and for contributing to the original impetus for this research. I am also very thankful to Dr. Chengyuan Chen, Dr. Nitin Kumar Naik, Dr. Shangzhu Jin, and Dr. Yitian Zhao for their suggestions and support.

I would like to express my appreciation to my external examiner: Prof. Mahdi Mahfouf, and internal examiner: Prof. Chris Price, for their valuable suggestions and patience.

I am grateful to all of my colleagues in the Department of Computer Science. I am especially grateful to Jingping Song, Liang Shen, Ling Zheng, Peter Matthew Scully, Yongfeng Zhang, and Zhenpeng Li. I have enjoyed many useful and entertaining discussions with them. My sincere gratitude goes to Tianhua Chen for the stimulating discussions, insightful advices, and collaborative efforts.

My sincere gratitude goes to my entire family: my parents Jingjun Su and Xiaohong Zhou, my dear wife Zhenyu Wang, and her parents Huxiao Wang and Shucui Yan. I would like to specially thank my wife for her charming smile, quiet patience, and endless support. Her tolerance of my occasional negative moods is a testament in itself of her unyielding devotion and love. The completion of my Ph.D. would not have been possible without her kind encouragement.

I would like to thank all my friends both in the United Kingdom and in China. I would also like to thank Pink Floyd for the inspiring music which used to keep company with me in many restless nights.

Contents

Contents	i
List of Figures	v
List of Tables	vii
List of Algorithms	ix
1 Introduction	1
1.1 Intelligent Data-driven Learning	2
1.1.1 Supervised Learning	3
1.1.2 Unsupervised Learning	3
1.2 Research Quality Assessment	4
1.2.1 National Projects of Research Quality Assessment	5
1.2.2 Journal Ranking	6
1.3 Thesis Structure	7
2 Background	13
2.1 Academic Publication Databases (APDs)	15
2.1.1 IEEE Xplore	16
2.1.2 AMiner	16
2.1.3 Web of Science	17
2.1.4 Scopus	19
2.1.5 Google Scholar	19
2.2 Indicators for Assessment of Journal Quality	20
2.2.1 Statistical-based Indicators	20
2.2.2 Recent Developments of Learning-based Evaluation	25
2.3 Link-based Framework of Bibliometric Data	31
2.3.1 Link-based Framework	32

2.3.2	Categorisation of Journal Indicators	35
2.4	Summary	38
3	Fused-link: Distance-based Aggregation of Journal Indicators	39
3.1	Fused-links	40
3.2	Applications of Fused-links	43
3.2.1	Classification	43
3.2.2	Clustering	45
3.3	Experimental Setup	46
3.3.1	Datasets	46
3.3.2	Evaluation Criteria	48
3.4	Results and Discussion	48
3.5	Summary	55
4	OWA-based Aggregation of Fuzzy Relations for Journal Ranking	56
4.1	Preliminaries	57
4.1.1	Fuzzy Relation	58
4.1.2	Ordered Weighted Averaging (OWA) Aggregation	60
4.1.3	Induced OWA	62
4.1.4	OWA Aggregation with Dependent Weights	63
4.2	OWA Aggregation of Fuzzy Relations for Journal Ranking	65
4.2.1	OWA Aggregation of Fuzzy Relations	65
4.2.2	Experimentation	70
4.3	Nearest Neighbour Guided Induced OWA and Journal Ranking	73
4.3.1	k Nearest Neighbour Induced OWA	73
4.3.2	Experimentation	76
4.4	Summary	81
5	T-transitivity on Ordered Weighted Aggregation of Fuzzy Relations	83
5.1	Preliminaries	85
5.2	Ordered Weighted Aggregation of Fuzzy Relations	87
5.3	Weighting Vectors	92
5.4	Application to Cluster UCI Datasets	96
5.4.1	Hierarchical Clustering	97
5.4.2	Aggregated Similarity based Hierarchical Clustering	98
5.4.3	Empirical Evaluation	99
5.4.4	Results and Discussion	101

5.5	Summary	107
6	Ensemble of Fuzzy Clusters for Journal Ranking	108
6.1	Preliminaries	109
6.1.1	Clustering Ensemble	110
6.1.2	Fuzzy Clustering Ensemble	111
6.2	Fuzzy Aggregation and Clustering Ensemble based Journal Ranking .	114
6.2.1	Indicator-based Generation of Fuzzy Base-clusters	114
6.2.2	Base-cluster Grouping and Aggregation of Memberships . . .	117
6.2.3	Refinement for Ranking	120
6.3	Experiment and Evaluation	121
6.3.1	Experiment Setup	122
6.3.2	Results and Discussion	124
6.4	Summary	127
7	Link-based Fuzzy Clustering Ensembles	128
7.1	Preliminaries	129
7.2	New Pairwise Similarity Matrices for Fuzzy Clustering Ensemble . . .	131
7.2.1	<i>FCO</i> : Co-association Matrix for Fuzzy <i>c</i> -means	131
7.2.2	<i>FLink</i> : Link-based Pairwise Similarity Matrix for Fuzzy <i>c</i> -means	132
7.2.3	<i>FCTS</i> : Connected-triple-based Pairwise Similarity Matrix for Fuzzy <i>c</i> -means	134
7.2.4	Fuzzy <i>c</i> -means Ensemble based on Similarity Matrix	136
7.3	Experimentation	137
7.3.1	Experimental Setup	137
7.3.2	Results and Discussion	138
7.4	Summary	141
8	Computational Considerations of Fuzzy Clustering Ensemble	142
8.1	Background	143
8.2	Feature-based Hierarchical Fuzzy Clustering Ensemble	144
8.2.1	Feature-based Generation of Ensemble Members	145
8.2.2	Similarity between Fuzzy Base-clusters	146
8.2.3	Base-cluster Grouping via Hierarchical Clustering	147
8.2.4	Final Assignment of Data Points	149
8.3	Initial Application to Big Data	149
8.4	Experimentation and Evaluation	152

8.4.1	Experimental Setup	152
8.4.2	Clustering Quality	153
8.4.3	Time Efficiency	156
8.5	Summary	158
9	Conclusion	160
9.1	Summary of Thesis	160
9.2	Future Work	163
9.2.1	On Assessments Methods	163
9.2.2	On OWA Aggregation	164
9.2.3	On Fuzzy Clustering Ensemble	164
	Appendix A Publications Arising from the Thesis	166
A.1	Journal Articles	166
A.2	Book Chapter	167
A.3	Conference Papers	167
	Appendix B Datasets Employed in the Thesis	168
	Appendix C List of Acronyms	172
	Bibliography	174

List of Figures

1.1	Relationships between Thesis Chapters	8
2.1	Components of a Digital Bibliographic System	15
2.2	Example of Citation Potential	23
2.3	Link-based Framework of Bibliographic Data	34
2.4	Frequency vs. Structure and Author vs. Reader Impact Indicators	37
3.1	Example of Fused-links between Journals	43
3.2	Screenshot of JCR	47
3.3	Trend of Accuracy against an Increase in k	50
3.4	Trend of Within-1 Accuracy against an Increase in k	50
4.1	Examples of Similarity Relation Function	60
4.2	OWA-based Journal Clustering Procedure	69
4.3	Two Linear Stress Function Used in Deriving W_{and} and W_{or}	71
4.4	Linear Stress Functions for Weighting Vector Generation	78
4.5	Trend of Accuracy against Reliability	79
5.1	Change of R^{min} Aggregated Relation against Weighting Vector	96
5.2	Change of R^{L} Aggregated Relation against Weighting Vector	97
5.3	Fuzzification of Feature Value	101
5.4	Trend of Accuracy against Normalised Attitudinal Character	102
6.1	Clustering Ensemble	111
6.2	Example of Clustering Ensemble	111
6.3	Fuzzy Aggregation and Clustering Ensemble based Journal Ranking	115
6.4	Filtering the Fuzzy c -means Result	117
6.5	Spearman's Correlation to RJL Results	126
7.1	Example of Pairwise Similarity Matrices for Clustering Ensemble	130

7.2	Trend of Accuracy against Diversity	139
8.1	Time Cost vs. Increase of Features	156
8.2	Time Cost vs. Increase of Instances	157
8.3	Time Cost vs. Increase of both Features and Instances	157

List of Tables

2.1	Link-based Categorisation of Indicators	37
3.1	Examples of Journals: Fused-link	42
3.2	Summary of Datasets: Fused-link	48
3.3	Comparison of Classification Accuracy (%): Fused-link	51
3.4	Comparison of Classification Within-1 Accuracy (%): Fused-link	51
3.5	Comparison of Clustering Accuracy (%): Fused-link	53
3.6	Comparison of Clustering Within-1 Accuracy (%): Fused-link	53
3.7	Example of Ranks by JIF, RJL and Fused-link	54
4.1	Examples of Journals: OWA	66
4.2	Comparison of Accuracy (%): OWA	72
4.3	Comparison of Within-1 Accuracy (%): OWA	72
4.4	Examples of Journals: k NN-IOWA	76
4.5	Summary of Datasets: k NN-IOWA	77
4.6	Comparison of Best Achieved Results: k NN-IOWA	80
5.1	Example of Stress Function	94
5.2	Aggregated Result of Examples	94
5.3	Summary of Datasets: T -transitivity	100
5.4	Comparison of Accuracy (%): T -transitivity	103
5.5	Comparison of Accuracy: Paired t -test	104
5.6	Comparison of Accuracy (%) of $N-R^{\min}$: T^{\min} -transitive vs. Not T^{\min} -transitive	104
5.7	Comparison of Accuracy (%) of $F-R^{\min}$: T^{\min} -transitive vs. Not T^{\min} -transitive	105
5.8	Comparison of Accuracy (%) of $N-R^L$: T^L -transitive vs. Not T^L -transitive	105
5.9	Comparison of Accuracy (%) of $F-R^L$: T^L -transitive vs. Not T^L -transitive	105

6.1	Label-assignment Matrix	112
6.2	Binary Cluster-association Matrix	112
6.3	Fuzzy Cluster-association Matrix	113
6.4	Summary of Datasets: FACE	122
6.5	r_s between Scores of Individual Indicator and RJI Rank	123
6.6	Averaging Weight of Indicator in OWA with Andness Weighting Vector	124
6.7	Averaging Weight of Indicator in DOWA	125
6.8	Averaging Weight of Indicator in 3NN-DOWA	125
7.1	Summary of Datasets: Link-based Fuzzy Clustering Ensemble	137
7.2	Comparison of Accuracy (%): Fixed Cluster Number	140
7.3	Comparison of Accuracy (%): Random Cluster Number	140
8.1	Summary of Datasets: HFCE	152
8.2	Comparison of Averaged Silhouette Index	154
8.3	Comparison of Accuracy (%): HFCE	155

List of Algorithms

3.2.1	Fused-link-based k -means for Journal Clustering	45
4.2.1	OWA-based k -means for Journal Clustering	68
6.2.1	Fuzzy Aggregation and Clustering Ensemble based Journal Ranking (FACE)	119
8.2.1	Hierarchical Fuzzy Clustering Ensemble (HFCE)	150

Chapter 1

Introduction

WITH the fast development of Information Technology (IT) and wide application of data acquisition equipment, a huge volume of data in various formats can be collected in daily business and scientific activities. Nowadays, more and more people have recognised that important and valuable information which can assist humans in decision-making is embedded in such data. Therefore, how to make full and effective use of the huge amount of data is becoming a critical challenge in every fields of profession. Data mining and knowledge discovery, together with other database technologies, provide an effective way to solve this problem. Data mining is a computational process which is mainly based on artificial intelligence, machine learning, and statistical techniques. It is capable of providing highly automated analysis of the original data, making inductive reasoning, and extracting abstract knowledge. In business activities for instance, data mining can help decision makers to predict customer behaviors, adjust marketing strategies, reduce their risks, and make the right decisions [20].

Since data mining can bring competitive advantages and economic benefits to businesses, a large number of educational organisations and IT companies worldwide have launched relevant research. Many classic algorithms and softwares have been developed, and such techniques can be categorised according to different criteria. For example, according to the representation of data, data mining techniques may be categorised to: relational, transactional, object-oriented, object-relationship, and data warehouse. According to the specific formats of the processed data, they can be categorised to: text, multimedia, time series, Internet, etc. Based on different

types of targeting knowledge, they can be categorised to: association, classification, clustering, feature analysis, etc. Also, data mining techniques can be divided into two categories based on their range of application: domain-specific techniques and common techniques. Domain-specific data mining can only provide solutions to a specific area and is usually based on expert knowledge from that area.

As the volume of data grows rapidly, the traditional, manual knowledge discovery process becomes increasingly time-consuming and expensive in many tasks [64]. This is especially the case for problem domains where classic approaches rely heavily on the opinions of domain experts, who must have an intricate understanding of the problem at hand [50], such as the task of assessing academic journals that this thesis is focussed on. When assessing the quality of an academic journal, expert opinions are often subjective and/or inconsistent between different individuals [164]. More importantly, with the development of online academic databases, the bibliographic data in its present form may contain a large number of objects (scholars, publications, etc.) and links (co-authors, citations, etc.), which are impractical (if not impossible) in most cases for human beings to analyse.

Classification and clustering techniques present possible approaches that attempt to reduce the overall burden of human experts in assessing academic journals. Classification methods typically work by transforming the underlying human knowledge and criteria for assessing journal qualities into rules, decision trees, and many other forms of classifiers through learning from the ranking results of experts [11]. Clustering approaches, being the main focus of this thesis, can search for and identify groups of journals so that the qualities of academic journals in each group are similar to each other.

1.1 Intelligent Data-driven Learning

Both classification and clustering algorithms can automatically build models for making predictions or decisions by learning useful information from data, and the effectiveness of both classification and clustering analysis is strongly dependent on the quality of data to which they are applied. For example, noisy data may affect the accuracy of the resulting data models [139]. To conduct classification analysis, it is required that the available data contains a certain amount of labelled examples, while in clustering analysis, only unlabelled examples are required. With most (if not all)

of the existing data-driven methods, it is true that the training examples will be more useful if they reflect the real problem. In terms of machine learning, classification and clustering are within the frameworks of supervised and unsupervised learning, respectively.

1.1.1 Supervised Learning

In supervised learning [117], examples with inputs and outputs are presented as the training set, and an algorithm is required to return a mapping from the inputs to outputs, which fits the observed training set. Formally, given a training set of N examples, each of which is in the form $(x_a \in X, y_a \in Y), a = 1, \dots, N$, a function $g : X \rightarrow Y$ is to be learnt, where X represents the input space and Y is the output space. In classification analysis, Y usually represents a set of discrete, non-ordered labels of examples. The function g is an element of a set of possible functions G that can be learnt by an algorithm, usually called the hypothesis space. The learnt function is evaluated based upon how well it predicts the labels for a particular set of examples, which is named as the testing set. The testing set can be a separate set of labelled examples or can be generated from all available labelled examples by using the cross-validation [15].

The training model described above is the discriminative training. The other model in the supervised learning framework is the generative training. The discriminative model merely performs classification or regression based upon labelled examples, while the generative model also explains how the data were generated. In this thesis, only the discriminative model of supervised learning is discussed. Examples of the discriminative model include decision trees, neural networks, and support vector machines.

1.1.2 Unsupervised Learning

Unlike in the supervised learning, algorithms of unsupervised learning are given only unlabelled examples, and they can show the hidden structure of data in the sense of how the examples are organised. Many algorithms developed in the unsupervised learning framework are based on data mining methods used to preprocess data such as the clustering analysis and principal component analysis. Clustering analysis is the task of grouping or segmenting a collection of examples into subsets (called clusters) such that those within each cluster are more similar to each other than examples

assigned to other clusters. An example can be described by a set of features, or by its similarities to other examples in the dataset [91].

Central to all the factors of clustering analysis is the notion of the degree of similarity (or dissimilarity) between the examples. A clustering algorithm attempts to group the examples into the same cluster when the degree of defined similarity between them is high. However, the notion of a “cluster” cannot be precisely defined [60]. Different clustering algorithms usually form different models of clusters even when they work on the same set of examples. Typical models of clustering analysis include: connectivity model, centroid model, density model, and so on. Normally, each example can only belong to one cluster in a clustering result. However, in fuzzy clustering, each example can belong to each cluster to a certain degree (called membership).

By contrast to classification, there are no explicit target outputs associated with each input in clustering. However, it is possible to develop from the formal framework of unsupervised learning that the goal of clustering is to build representations of the input that can be used for decision making, predicting future inputs, etc. In other words, the resulting groups of examples are usually the matter of interest in clustering analysis, while in terms of classification, the resulting discriminative power of grouping examples is of interest.

1.2 Research Quality Assessment

Bringing influential and significant research output to society is one of the key objectives of many universities and institutes, and also the life-long pursuit of scholars. Additionally, the assessment of research quality is becoming more and more critical to many scientific activities. Such assessment can have impact upon many issues ranging from publishing articles in academic journals/conferences to determining which universities are qualified to undertake major research projects in a country. According to [31, 32], if in the words of the old song “money makes the world go round”, it is research quality assessment which, in the higher education context at least, makes the money go round.

1.2.1 National Projects of Research Quality Assessment

Recently, many countries have implemented their national projects for academic outputs evaluation and quantitative assessment, such as the Research Excellence Framework (REF) in UK and the Excellence in Research for Australia (ERA). Generally speaking, the projects organised nation-wide are abundant in funding and human resource. In such cases, inviting expert panels to accomplish a comprehensive peer review is probably the most reliable methodology to make an unbiased result. Although the results from national projects could be more sophisticated than other assessment projects, the huge time and financial costs make this approach difficult to carry out frequently.

Research Excellence Framework in UK

The Research Excellence Framework (REF) is the new system for assessing the quality of research in UK Higher Education Institutions (HEIs) [160]. Its original project is the Research Assessment Exercise (RAE). The primary purpose of REF is to produce assessment outcomes for each submission made by HEIs. The final outcomes of REF can have a profound influence on many aspects of research activities in the UK. For example, the funding bodies intend to use the assessment outcomes to inform the selective allocation of their research funding to HEIs. The assessment can provide accountability for public investment in research and produces evidence of the benefits of this investment.

Technically, REF is a process of expert review. HEIs are invited to make submissions which will be assessed by expert sub-panels, working under the guidance of main panels. Sub-panels can apply a set of generic assessment criteria and level definitions, to produce an overall quality profile for each submission. The primary outcome of the assessment will be an overall quality profile awarded to each submission, showing the proportion of the submission that meets each of the starred levels ranging from: “four-star” which indicates “quality that is world-leading in terms of originality, significance and rigour” to “Unclassified” which indicates “quality that falls below the standard of nationally recognised work. Or work which does not meet the published definition of research for the purposes of this assessment”.

Excellence in Research for Australia

The Excellence in Research for Australia (ERA) is conducted by the Australian Research Council (ARC), with objectives which include: establish an evaluation framework that gives government, industry, business and the wider community assurance of the excellence of research conducted in Australian higher education institutions; provide a national stocktake of discipline level areas of research strength and areas where there is opportunity for development in Australian higher education institutions; and so on. ERA evaluates research undertaken in higher education institutions using measures of research volume and indicators of research quality, application and recognition. It outlines performance in each of the disciplines evaluated at each institution.

One of the main outcomes of ERA is the Ranked Journal List [90, 186]. The ARC spent several years on preparation and consultations related to journal ranking, and released its full ranked lists of more than 20,000 unique peer-reviewed journals in ERA 2010 and ERA 2012, respectively. Over 700 experts are involved to assist the ARC in developing this journal ranking list in which each journal has a single quality rating and is assigned to one or more disciplines defined by its research topic [11].

1.2.2 Journal Ranking

Generally, the assessment of research quality is mainly implemented as the evaluation of research outputs in various forms. Taking computer science as an example, the research outputs can be presented as prototype, software, research paper, etc. It can be seen from those national research assessments that no matter in which research fields, three common entities are critical to the evaluation of research outputs. They are academic journals, researchers (organised by universities or departments) and their publications. The evaluating tasks on the three entities are interlinked and sometimes circular. For example, the quality of a research publication potentially reflects, and is affected by, the prestige of that journal and the authors' research experience. Therefore, the existing assessment projects and methods usually have certain aims and scales.

Journal ranking is widely used in a number of research fields to evaluate an academic journal's impact and quality. The aim of journal rankings is to reflect the place of a journal within the research field, which has impact on the prestige of the

journal and the relative difficulty of getting a paper published in it. Recently, journal ranking has been introduced as an official research evaluation tool in some countries such as Norway, Australia and France [154].

Assessment of journal quality offers various benefits to the academic community. The assessment outcomes help researchers to demonstrate their accomplishments to colleagues, administrators and tenure and promotion committee members, especially those not familiar with the research domain. By knowing the ranks of journals, researchers may be more willing to submit their stronger works to the journals with a higher rank. And also, libraries may consider the assessment outcomes of journals to decide which journals should be future subscription. Students may consult the ranking lists to identify the most credible sources for theories, ideas, and research methods [169]. However, every coin has two sides. Some universities pay too much attention to journal rankings so that their research members are required to publish papers only in a very limited number of top journals [177]. In some newly emerging research areas, there are few journals can be the top-listed ones amongst journals whose topics are in mainstreams, which makes researchers from those areas feel difficult to get their works published. Furthermore, By only concentrating on publishing papers in a small set of journals, some researchers will dedicate their effort to meeting the reviewers' requirements rather than making truly contribution to a research field. Therefore, some researchers just reject journal rankings and choose to submit their papers to the best journals from their own perspective.

In conclusion, the essential scientific value is embedded in the content rather than the location of publications. Top-ranked journals may still contain works of low quality and contribution. In this thesis, several intelligent data-driven methods are developed and applied to the assessment of journals, which allow more objective, interpretable, and intuitive ranking of journals.

1.3 Thesis Structure

This section outlines the structure of the remainder of this thesis. The relationships between the technical chapters are illustrated in Figure 1.1. The dependencies between these chapters are denoted as solid arrows. The chapters which mainly concern the intelligent methods for journal assessments are shown in blue boxes, while the other chapters discuss general problems of intelligent methods extended

from the journal assessments, which are shown in pink boxes. A comprehensive list of publications arising from the work of the thesis is provided in Appendix A.

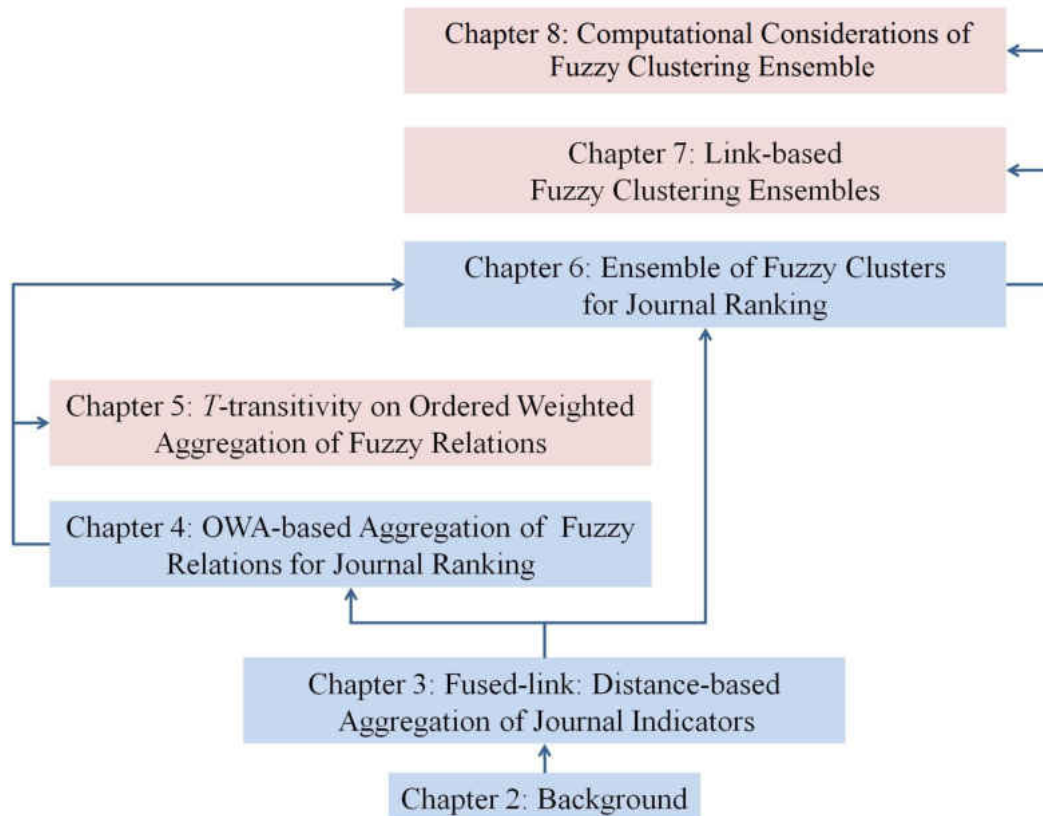


Figure 1.1: Relationships between Thesis Chapters

Chapter 2: Background

This chapter provides a background introduction to modern techniques involved in the assessment of academic journals, which is organised into two core parts: academic publication database and journal impact indicators. This chapter has firstly introduced the concept and development of academic publication database. Secondly, several well-known representatives of these such as: IEEE Xplore, AMiner, and Web of Science have been explained in detail, as they are the main examples adopted to demonstrate the publisher-oriented, researcher-oriented, and index-oriented databases, respectively. More importantly, a selection of popular journal impact indicators are reviewed. Their underlying respective inspirations span a wide range of techniques, including bibliometrics, machine learning, and data mining.

The methods which investigate the correlation and aggregation of existing indicators are reviewed, on which the establishment of journal clustering methods proposed in this thesis is based.

In addition, a link-based framework of bibliographic data is identified. Those reviewed journal impact indicators are uniformly categorised by the sets of links based on which they are calculated.

Chapter 3: Fused-link: Distance-based Aggregation of Journal Indicators

In this chapter, an approach based on links between journals is proposed for the capturing and fusion of various journal impact indicators. In particular, a number of popular indicators are combined and transformed to fused-links between academic journals, and two distance metrics: Euclidean and Manhattan distances are utilised to support the development and analysis of the fused-links. The approach is applied to both supervised and unsupervised learning, in an effort to estimate the impact and therefore the assessment of journals.

This chapter also presents the details of experimental design for demonstrating the performance of fused-link. The construction of datasets and the evaluation criteria are also used in other relative chapters. Results of systematic experimental evaluation demonstrate that by exploiting the fused-links, simple algorithms such as k Nearest Neighbours (k NN), and especially the clustering methods such as k -means, can perform as well as advanced techniques like support vector machines, in terms of accuracy and within-1 accuracy, while exhibiting the advantage of being more intuitive and interpretable.

Chapter 4: OWA-based Aggregation of Fuzzy Relations for Journal Ranking

Fuzzy relations form the basis for many developments and applications of fuzzy systems. Measures of fuzzy similarity have been proposed in the literature for comparing objects. In this chapter, aggregated fuzzy relations are generated between academic journals to compare their performance with respect to different journal impact indicators. In particular, various indicators may be employed to construct several distinctive fuzzy similarity relations, which may be subsequently combined

via the use of the Ordered Weighted Average (OWA) operator. This OWA-aggregated measure preserves reflexivity and symmetry, with T -transitivity conditionally preserved if appropriate weighting vectors are selected. Different similarity measures and weighting vectors are compared for the task of journal clustering, in an effort to estimate the ranking of academic journals.

Besides the original OWA aggregation operator, a nearest neighbour guided induced OWA operator (abbreviated as k NN-IOWA), is also proposed in this chapter. k NN-IOWA is a special case of the generic induced OWA where the input arguments are ordered by the average distances to their k nearest neighbours. The weighting vectors in k NN-IOWA are defined, which are used to interpret the overall behaviour of the operator's reliability. k NN-IOWA is also applied for building aggregated fuzzy relations between academic journals, based on their indicator scores. Its result is compared against different types of aggregation operator and tested on six bibliometric datasets.

The results of experimental evaluation demonstrate that by using OWA-based aggregation of fuzzy similarity relations, simple techniques such as k -means can perform well in terms of standard accuracy and within-1 accuracy.

Chapter 5: T -transitivity on Ordered Weighted Aggregation of Fuzzy Relations

In the application of OWA-based aggregation of fuzzy similarity to support journal ranking and other decision making processes, a key question relates to what underlying mathematical properties of the model can be preserved in the process of constructing or aggregating similarity relations. In this chapter, the properties of OWA aggregated fuzzy similarity have been studied from a mathematical point of view. In particular, two types of aggregation are investigated: 1) for component relations defined by minimum T -norm, the min-max operators are employed to aggregate them; and 2) for component relations defined by Łukasiewicz T -norm, the sum-product operators are employed. The condition of when the proposed aggregated relations preserve T -transitivity is examined, and its impact upon clustering procedures is experimentally investigated.

Fuzzy similarity relations generated from different feature patterns are aggregated via the use of the respective proposed aggregators. The aggregated fuzzy relations are

employed as measures of similarity amongst feature patterns to perform hierarchical clustering. Results on UCI benchmarks demonstrate that the aggregated similarities following the proposed approach produce better hierarchical clusters than those of classic aggregators (e.g., min, max, and average).

Chapter 6: Ensemble of Fuzzy Clusters for Journal Ranking

Fuzzy techniques, such as fuzzy aggregations and fuzzy clustering, have been proven effective for many applications of decision making and multi-criteria evaluation. In this chapter, fuzzy aggregation techniques, the OWA operators in particular, are further applied to aggregate the scores of academic journals under different impact indicators. A fuzzy aggregation based fuzzy clustering ensemble method is proposed for ranking academic journals. Specifically, several distinctive fuzzy clusters of journals are constructed based on their performance with respect to different journal impact indicators, which may be subsequently combined via the use of various OWA operators. A rank can therefore be generated based on the memberships of journals to those combined fuzzy clusters. The nearest neighbour guided aggregation operators can enhance the reliability while the fuzzy clustering enhances interpretability of the ranking procedure.

The ranking results of academic journals from six subjects are compared with the journal list published in Excellence in Research for Australia (ERA). The fuzzy aggregation and clustering ensemble based approach shows its advantage in providing ranking results that are generally more accurate with a linguistic interpretation.

Chapter 7: Link-based Fuzzy Clustering Ensembles

This chapter further studies the clustering ensemble techniques for aggregating fuzzy clusters. Besides its effectiveness in journal ranking, clustering ensemble offers a general approach for aggregating multiple clustering results in order to improve the overall clustering robustness and stability. It also helps improve accuracy by combining clustering results from component methods that utilise different parameters (e.g., number of clusters), avoiding the need for carefully pre-setting the values of such parameters in a single clustering process. Link-based consensus methods for the ensemble of fuzzy c -means are proposed in this chapter. They employ a fuzzy graph to represent the relationships between component clusters upon which to derive the final ensemble clustering results.

The proposed methods are tested against typical traditional methods on various benchmark datasets. The experimental results demonstrate that the proposed fuzzy link-based clustering ensemble approach generally outperforms the others in terms of accuracy.

Chapter 8: Computational Considerations of Fuzzy Clustering Ensemble

To reinforce such link-based fuzzy clustering ensemble, this chapter presents another ensemble approach for fuzzy clustering, with an aim to be applied for clustering of a larger scale of data. The proposed algorithm first generates fuzzy base-clusters with respect to each data feature and then, employs a fuzzy hierarchical graph to represent the relationships between the resulting base-clusters. Whilst the work employs fuzzy *c*-means and hierarchical clustering in generating base-cluster and implementing consensus function respectively, when applied to large datasets it has lower time complexity than the original fuzzy *c*-means and hierarchical clustering.

The resultant ensemble clustering mechanism is tested against traditional clustering methods on various benchmark datasets. Experimental results demonstrate that it generally outperforms its crisp counterparts and the single linkage agglomerative clustering, in terms of accuracy in conjunction with time efficiency, thereby showing that it has the potential for application in clustering big data.

Chapter 9: Conclusion

This chapter summarises the key contributions made by the thesis as well as a discussion of topics which form the basis for future research.

Appendices

Appendix A lists the publications arising from the work presented in this thesis, containing both published papers, and those currently under review for journal publication. Appendix B provides information regarding the UCI benchmark datasets employed in the thesis. Appendix C summaries the acronyms employed throughout this thesis.

Chapter 2

Background

THE evolution of the electronic age has led to the development of numerous academic publication data on the Internet, offering search facilities on a particular subject and the ability to perform research quality analysis. Scientific journals may differ with respect to their importance, status and prestige in a specific academic area. With the development of online resources of academic publications, journal impact indicators are designed to assess significance and performance of individual journals, their role and position in the international formal communication network, their quality or prestige as perceived by scholars, by analysing citations, downloads and comments of individual journals.

Concerning the common approaches for journal quality assessment, there are two general methodologies: 1) peer review, which is a judgment based on expert opinion; and 2) indicator measure, which is a judgment that is based on quantitative data. Each of these approaches expresses its inherent and specific strengths and weaknesses. Although peer review can integrate more complexity of human opinions into evaluation assessments, it also creates a potential for subjectivity of peer panels [36]. The indicator-based assessment of journals relies on the development in computer and information technologies. Many academic organisations and commercial publishing institutions maintain large publication/indexing databases, enabling the calculation of citations and other possible journal impact indicators. These indicator-based assessments are not only more objective, but also use less time and finance compared with those assessments based on peer review by human experts.

Generally speaking, a digital bibliographic system involves several components, and each component has its specific functions. An illustrative diagram of these components and the dependence between them are shown in Figure 2.1 [85]. Similar to many other information systems, the core in Figure 2.1 is the data source. In the digital bibliographic system, most of the electronic versions of academic articles are provided by publishing companies themselves. Abstracting and indexing system summarises the essential contents of documents. The main function of an indexing system is to facilitate the speedy location of specific items of publications. The indexing system is able to rearrange the abstracted materials, bring together distributed articles to a topic and arrange them in classified order. A good index sometimes is a valuable aid for scholars to select the publishing place of their scientific works. The outside layer of the diagram shown in Figure 2.1 is the application layer. Typical applications of a digital bibliographic system include: 1) search engine, which allows users to search the database with keywords and find relevant academic publications very quickly; 2) quality report, which provides indicator-based analysis about citations of a single publication or a set of publications (such as a journal or a proceeding). Note that not all the three layers are necessarily included in modern digital bibliographic systems. For example, Google Scholar [88] uses its powerful search technique to collect useful indexing information from the Internet rather than providing data source by itself. Meanwhile, a number of publishing companies also provide highly integrated indexing and searching services.

The implementation of such an bibliographic system specifically for assessment of journals is also practical. For example, BibNetMiner is an system designed for sophisticated information network mining for bibliographic data. It also contains a multi-level architecture [180]. The bottom level contains the information extraction and analysis engine which provide the data source for the system. The middle layer is the functional module layer, which implements the major assessment algorithms based on the clustering and ranking information derived from the information network analysis. The top layer contains a user-friendly and visualization-enhanced interface, which interacts with users and responds to their requests. For ease of explanation and organisation, in this thesis, the term Academic Publication Database (APD) is adopted to describe all the relevant services of collecting, indexing, searching, analysing, and other manipulating of academic publication data.

The remainder of this chapter is structured as follows. The detailed introduction of APD and several representative ones of APDs are presented in Section 2.1. The

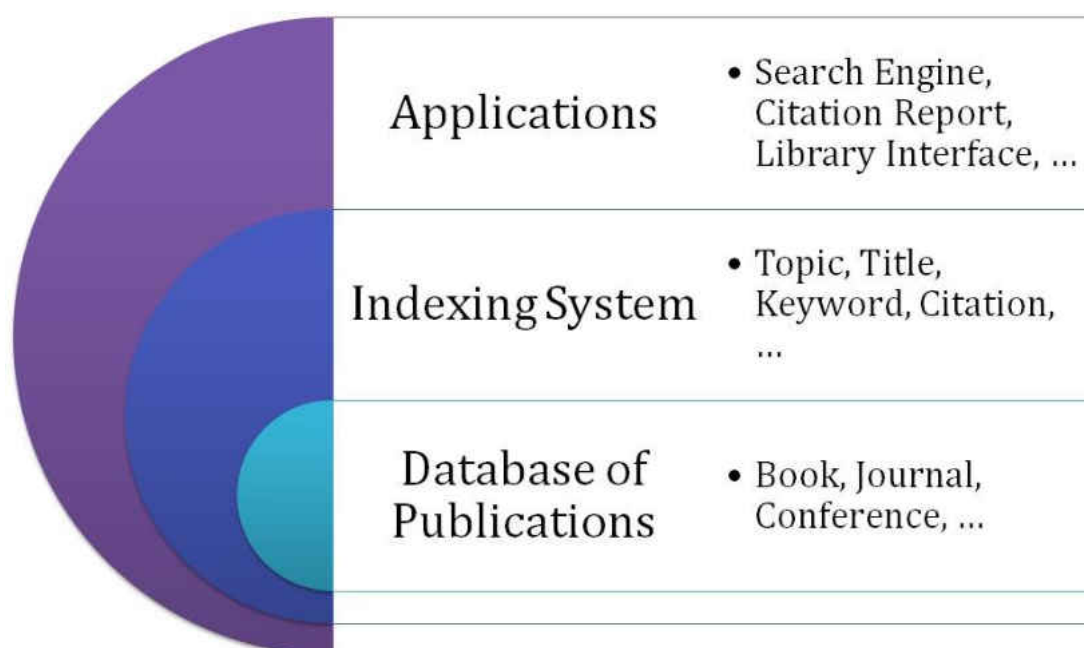


Figure 2.1: Components of a Digital Bibliographic System

main focus of this thesis lies in the use of indicator-based journal quality evaluation. Therefore, the statistical-based and learning-based assessments of academic journal quality are of significant importance to the development of the present work. They are covered in Section 2.2 in detail. A comprehensive description of a link-based framework of bibliometric data is given in Section 2.3, based on which the indicators can be re-defined and categorised. Finally, Section 2.4 summarises this chapter.

2.1 Academic Publication Databases (APDs)

An APD is a database of bibliographic records, an organised digital collection of references to published literature, including journal and newspaper articles, conference proceedings, reports, government and legal publications, patents, books, etc. In contrast to library catalogue entries, a large proportion of the bibliographic records in APD contain very rich subject descriptions in the form of keywords, subject classification terms, or abstracts [65]. APDs are mainly used to find journal articles, while some databases will also index other types of material, such as book chapters, theses, conferences, patents, etc. An APD may be general in scope or cover a specific academic discipline, and a significant number of bibliographic databases are still

proprietary, available by licensing agreement from vendors, or directly from the indexing and abstracting services that create them.

A number of APDs are established by publishing companies and evolve into digital libraries, providing the full-text of the indexed contents. Others may cover citations, scholars, and research topics to create more complete disciplinary search engine systems. Nowadays, the distinction between a database and a search engine is unclear for many online APDs as they are providing integrated services to users. Several typical APDs that cover the research area of computer science are outlined below, in which IEEE Xplore represents publisher-oriented, AMiner represents researcher-oriented, and Web of Science, Scopus, Google Scholar represent index-oriented APDs, respectively.

2.1.1 IEEE Xplore

IEEE Xplore (<http://ieeexplore.ieee.org/>) is a typical publisher-oriented APD. It indexes, abstracts, and provides full-text for articles and papers mainly from the Institute of Electrical and Electronics Engineers (IEEE) and the Institution of Engineering and Technology (IET). Many of the materials covered by IEEE Xplore are on computer science, electronics and relevant engineering. The IEEE Xplore digital library provides online access to more than three million full-text documents from some of the world's most highly cited publications in electrical engineering, computer science and electronics. The IEEE Xplore provides access to many highly cited journals in computer science such as: IEEE Transactions on Fuzzy Systems, IEEE Transactions on Pattern Analysis and Machine Intelligence and IEEE Transactions on Neural Networks and Learning Systems. Guests can only have access to basic abstracts of academic articles and conduct free keyword searches based on content of abstracts. Full-text access rights and advanced search options are based on institution's subscription.

Similar to IEEE Xplore, ScienceDirect (<http://www.sciencedirect.com/>) is also a full-text APD owned by another main international publisher of scientific publications, the Elsevier. It covers over 2,500 journals and more than 33,000 book titles published by Elsevier.

2.1.2 AMiner

AMiner (<http://aminer.org/>), previously named as ArnetMiner, is a free online service used to index and search academic social networks. It was firstly initiated and created

by Professor Jie Tang from Tsinghua University, China, and its relevant research was funded by the Chinese National High-tech R&D Program and the National Science Foundation of China. AMiner is designed to search and perform data mining operations against academic publications on the Internet. In addition to this, it also uses social network analysis to identify connections between researchers through their publications [182]. This allows it to provide a series of researcher-oriented services such as expert finding, reviewer recommendation, academic performance evaluation, and so on.

AMiner was created as a research project in social influence analysis, social network ranking, and social network extraction. AMiner is famous in academia for its ability of identifying relationships between and drawing statistical correlations about research and researchers. It solves the problem of how to extract researcher profiles from the Internet and integrate the extracted information from different sources. A powerful search service based on the integrated information is also provided for users. So far, the statistics from the website of AMiner indicate that about 40 million researchers and about 80 million publications are indexed.

ResearchGate (<http://www.researchgate.net/>) is another social networking site, which is designed for scientists and engineers to share papers, ask and answer questions, and find collaborators. Similar to AMiner, it is also a researcher-oriented APD. However, ResearchGate has many of the features that are typical amongst social network sites, such as updating user profiles, finding other users with similar interests and sharing/reviewing academic papers like blogs. It differs from other social networks in that it is designed for researchers and scientists. From 2009 to 2011, the site grew from 25,000 users to more than 1 million [47].

2.1.3 Web of Science

Web of Science (WoS), previously known as Web of Knowledge, is an online subscription-based scientific citation indexing service maintained by Thomson Reuters that provides a comprehensive citation search. Its most noteworthy and enduring output is its indexing-oriented products covering the disciplines of arts and humanities, social sciences and science and its in-depth exploration of specialised subfields within these disciplines [4].

Science Citation Index Expanded

WoS consists of several core sub-APDs, one of which is the famous Science Citation Index Expanded (SCI-E) known by many scholars in the research area in science and engineering. SCI-E is a larger version of the Science Citation Index (SCI) which is a citation index originally produced by the Institute for Scientific Information (ISI) in 1964. SCI-E covers more than 6,500 journals ranged from 1900 to the present, across 150 disciplines [92]. Due to the reputation of WoS and its rigorous selection process of journals, those journals indexed by SCI-E are usually described as the world's leading ones of science and technology. More importantly, getting papers published in these journals is becoming one of the requirements for many universities to offer tenure to their research staff.

Journal Citation Reports

Journal Citation Reports (JCR) is an annual publication of bibliometric analysis provided by WoS. It reports valuable information about academic journal citations in the sciences and social sciences, including indicator-based analysis of journals such as the Journal Impact Factor (JIF). The JCR was originally published based on Science Citation Index only. Currently, JCR has become a distinct service of WoS, which is based on citations compiled from both the SCI-E and the Social Science Citation Index (SSCI) [78].

As with other bibliometric approaches, the JCR has its own limitations. Recently, much debate has surrounded the overuse of JCR pointing out certain deficiencies of the calculation processes of JIF and other indicators reported in WoS. These criticisms were mainly around that several indicators, especially the JIF, are field-specific and can be easily manipulated by editors, self-citations or by changing the editorial policies, which makes the entire process essentially nontransparent [164]. In spite of the criticisms, the general contribution of the WoS to the development of indexing services and to the establishment of bibliometrics and scientometrics is considerable. JCR also has a significant influence on the behaviour of both the scholars who need to get their research findings published (for example, in research performance or promotion reviews, many universities make direct reference to staff's publication record in journals reported in JCR) and the librarians who seek for sources of publications.

2.1.4 Scopus

Besides the full-text APD, Elsevier also provides an abstract and citation indexing APD which is named as Scopus (<http://www.scopus.com/>). Instead of depending only on Elsevier's publications, Scopus established a content selection and advisory board to maintain an open and transparent content coverage policy regardless of publisher. It covers over 55 million records from more than 5,000 international publishers and also evaluates journals by using indicators.

Compared with WoS, Scopus is easy to navigate, even for the novice user. Its ability to search both forward and backward from a particular citation can be found to be helpful to users. Besides, the multidisciplinary aspect allows users to easily search outside of their discipline. One advantage of WoS over Scopus is the depth of coverage, with the full WoS database going back to 1945 and Scopus going back to 1966. However, Scopus and WOS complement each other as neither resource is all inclusive. Libraries which are able to afford them usually will subscribe to both tools [34].

2.1.5 Google Scholar

Google Scholar (<http://scholar.google.com/>) is a freely accessible web search engine that indexes the metadata of scholarly literature across disciplines produced by Google. The Google Scholar index includes most peer-reviewed journals around the world, plus scholarly books and other non-peer online documents. While Google does not publish the size of Google Scholar's database, it is estimated in [146] that Google Scholar contains roughly 160 million documents as of May 2014. Google Scholar resembles the subscription-based APDs including Scopus, WoS, and etc. Free users of the relevant APDs will be able to access only an abstract and the citation details of an article.

Besides its competitive indexing and searching services, Google Scholar also provides convenient personalised services for users who have Google accounts. It allows users to save search results into the "Google Scholar library", a personal collection which the user can search separately and organise by tags. It also allows individual scholars to create personal "Scholar Citations profiles", public scholar profiles that are editable by scholars themselves. Google Scholar can automatically calculate and display the scholar's total citation count, h-index [6], and i10-index [104].

In spite of the WoS, Scopus, and Google Scholar which generally have coverage on multiple research areas, there are a good number of index-oriented APDs which mainly cover a specific research area. In the field of computer science, such well-known APDs include DBLP (<http://dblp.uni-trier.de/>) and the CiteSeerX (<http://citeseer.ist.psu.edu/>).

2.2 Indicators for Assessment of Journal Quality

Journal impact indicators provide quantitative tools for ranking, evaluating and comparing academic journals. Formally, such an indicator is a mapping from an academic journal to a real number which usually represents the score of the journal. Journals with higher indicator scores are usually deemed to be more important than those with lower ones. However, each indicator has its strengths and limitations, and their ranking results can be quite different [164, 173]. The following section reviews several well-known statistical journal indicators.

2.2.1 Statistical-based Indicators

As pointed out in Section 2.1, with the development in computer and information technologies, many organisations and commercial institutions are able to maintain very large publication databases, upon which the calculation of citations and many complex journal impact statistics are available. Indeed, various indicators of journal impact have been designed for different purposes via the use of different databases. For examples, WoS has several measures of journals included in its annual citation report, and Scopus provides the Source Normalized Impact per Paper (SNIP) [124] based on its own database. Amongst these, the most well-known and perhaps, the most utilised is the journal impact factor provided by WoS.

2.2.1.1 Journal Impact Factor

The Journal Impact Factor (often abbreviated as JIF) of a journal is a non-negative number which indicates the average count of citations to the source items published in that journal. It was devised by Eugene Garfield who is also the founder of SCI and JCR. Formally, the JIF of a journal is calculated by dividing the number of current year citations to the articles published in that journal during the previous two years. For example, the 1992 JIF score of a journal is calculated as:

2.2. Indicators for Assessment of Journal Quality

- A = total cites in 1992
- B = 1992 cites to articles published in that journal during 1990-91 (this is a subset of A)
- C = number of articles published in that journal during 1990-91
- $D = B/C = 1992$ JIF score

The standard JIF relates to a specific two-year time period; it is possible to calculate it for any desired period. The JCR also includes a 5-year JIF which is the average number of times articles from the journal published in the past five years have been cited in the current year. It is calculated by dividing the number of citations in the JCR year by the total number of articles published in the five previous years. For example, the 1992 5-year JIF score of a journal is calculated as:

- A = total cites in 1992
- B = 1992 cites to articles published in that journal during 1987-91 (this is a subset of A)
- C = number of articles published in that journal during 1987-91
- $D = B/C = 1992$ 5-year JIF score

The 5-year JIF complements very well the standard 2-year JIF for indicating the prestige, reputation and influence of the journals through the ratio of the citation counts of articles published in the journals to average productivity of journals for a longer time span [103]. A detailed comparison between 2-year and 5-year JIFs, and an example of how they change the rank order journals can be found in [77].

JIF and 5-year JIF eliminate some of the bias of such counts which favor large journals over small ones, or frequently issued journals over less frequently issued ones [76]. For a long time, JIF is usually recognised as a metric for the relative importance of a journal within a research field. However, much debate has surrounded the overuse of the JIF, which include that citation behavior varies among fields of science and JIF cannot tell this systematic differences, self-citations are not excluded, and several other unreflected matters can affect the value of JIF heavily.

2.2.1.2 Source Normalised Impact per Paper

Recently, Scopus introduced an indicator named Source Normalised Impact per Paper (SNIP). It tries to solve the problem that citation frequencies in some research fields are on average significantly lower than in others, so that its scores can be explained without a field classification system in which the boundaries of fields are explicitly defined [124]. SNIP employs a source normalised approach to correct for differences in citation practices between scientific fields.

A key concept in SNIP is called Raw Impact per Paper (RIP) published in the journal, which is not essentially different from a three-year JIF, but the citing and cited publications are included only if they are indexed by the Scopus rather than WoS. Another key concept in SNIP is called citation potential which indicates how frequently papers in a subject field cite other papers. It is defined as the average number of cited references per paper in a set of papers. For example, there is a set of four citing papers, and their citation relationships with three cited articles of a journal are shown in Figure 2.2. The numbers in the upper squares indicate the number of cited references contained in each source paper. The citation potential in a set of source articles is defined as the average number of cited references per source article, which amounts in this example $(3 + 2 + 1 + 2)/4 = 2$.

The calculation of citation potential of a journal starts by delineating the journal's subject field. The subject field of a journal is defined as the set of all papers in the year of analysis with at least one reference to the journal. Since the SNIP is calculated by Scopus and each APD has its own coverage, the calculation of the citation potential does not count the total number of cited references in a paper, but the number of cited references published in journals processed for the Scopus database. The result of a database-dependent citation potential is renamed as Database Citation Potential (DCP), which is further normalised to Relative Database Citation Potential (RDCP) as:

$$\text{RDCP} = \frac{\text{DCP}}{\text{median}(\text{DCP})}. \quad (2.1)$$

where $\text{median}(\text{DCP})$ denotes the median DCP value of all journals in the Scopus database. For example, the Journal of Electronic Materials gained the median DCP in the Scopus database of the year 2007, which is 6.87. Therefore, when calculating the RDCP of 2007, all journals' DCP values are divided by the same number 6.87. As

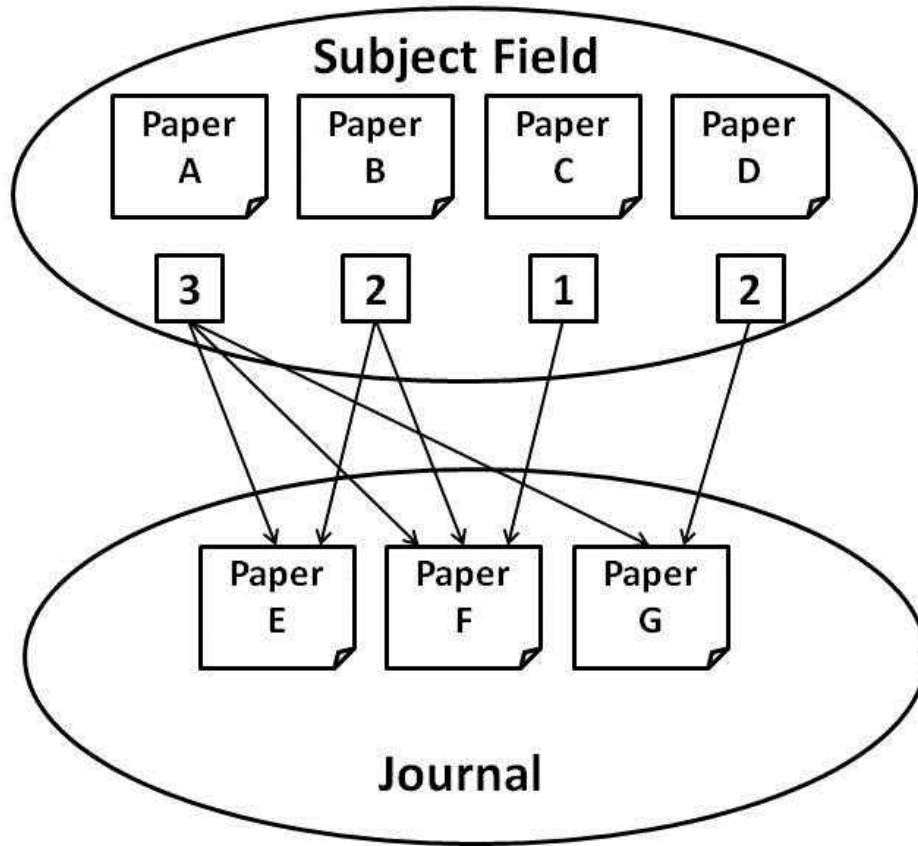


Figure 2.2: Example of Citation Potential

a consequence, half of the journals in the Scopus database have a RDCP value that is higher than 1.0 and half of them have a RDCP value that is lower than 1.0. Based on RIP and RDCP, the SNIP is defined as:

$$\text{SNIP} = \frac{\text{RIP}}{\text{RDCP}}. \quad (2.2)$$

Since there is no normalisation for field differences, the RIP indicator can only reflect the average citation impact of the publications of a journal, without correcting for differences in citation behaviours between scientific fields. By using the ratio of a journal's RIP value to its RDCP value, the SNIP indicator is able to provide a measure of citation impact that allows for meaningful comparisons between journals from different research fields [193].

After the original definition and implementation of the SNIP by Scopus, a number of modifications were recently made to the indicator. Although some systematic

differences can be observed between the revised and original version, the differences of ranking journals between them turn out to be relatively small [193]. The details of these modifications and discussions can be found in [124, 159, 193].

2.2.1.3 Researcher Behaviour based Indicators

Besides the citations amongst papers and journals, the behaviour of researchers is also very informative in reflecting the impact and quality of academia journals. In [95], an approach which examines the publishing behaviour of full-time, tenured faculty members from leading universities is proposed with an aim to rating journals. It is assumed that those Selected Influential Scholars (SIS) can have a positive impact on their collective publications, and hence the data extracted from them could be indicative of the reputation of academic journals. Based on this assumption, four behaviour measurements have been proposed:

- Publishing Breadth (PB) which reflects that if a high percentage of SIS have authored articles in a journal, then that journal is better than other ones in which smaller percentages of SIS have authored:

$$PB = \frac{\text{Number of SIS who have published in this journal}}{\text{Number of SIS}}; \quad (2.3)$$

- Publishing Mode (PM) which reflects that if a journal is the most frequent publication outlet for a higher percentage of SIS than that of other journals, its PM score is higher than others' PM scores accordingly:

$$PM = \frac{\text{Number of SIS whose most frequently publishing place is this journal}}{\text{Number of SIS}}; \quad (2.4)$$

- Publishing Intensity (PI) which reflects that the most important journals are those with the highest average number of articles authored by SIS:

$$PI = \frac{\text{Number of papers authored by SIS in this journal}}{\text{Number of SIS}}; \quad (2.5)$$

- Publishing Weight (PW) defined as the average number of papers authored by those SIS who have published at least one paper in this journal. A high PW score for a journal suggests that those who have published there find it to be an appropriate publishing place:

$$PW = \frac{\text{Number of papers authored by SIS in this journal}}{\text{Number of SIS who have published in this journal}}. \quad (2.6)$$

These behaviour-based indicators opened an avenue of evaluating the impact of a journal based on the contribution influential scholars have made to that journal. However, this approach may contain subjective factors in the selection of those investigating scholars. In the research of [95], the implementation of the behaviour based approach is based on a project which covers 106 scholars from 31 leading universities in USA. To avoid possible bias in the selection of influential scholars, it is crucial to have an independent entity to identify those leading universities and scholars, which makes its implementation much more complicated than the citation-based indicators.

It is worth noticing that these behaviour-based indicators rely on the assumption that scholars have significant impacts on journals. However, its reversed way which uses the citations between journals and papers to assess scholars is more popular in the literature, such as the h-index [6], g-index [59] and i10-index [104]. The h-index was originally proposed to evaluate scholars according to their citation distributions and was extended to evaluate journals in [30]. It is defined as: a journal has index h if h of its papers have at least h citations each. Considering the number and recognition degrees of these statistical-based indicators, it can be concluded that the analysis of citations is the mainstream for the assessments of both journals and scholars.

2.2.2 Recent Developments of Learning-based Evaluation

The statistical-based indicators usually have intuitive and simple definitions, and some of them have achieved dominant positions with the popularity of APDs by which those indicators are provided. However, with the developments of APDs, many new types of informative data for journal assessment other than total number of citations are also available now. Accordingly, it is necessary to revise and improve those traditional statistical-based indicators, and many achievements in machine learning and data mining have becoming more and more popular in measuring the impact of academic publications.

To better capture the impact of academic journals in the huge publishing data, a variety of assessments have been proposed based on more complex approaches. Here, the term “complex approaches” refers to those computational methods borrowed from the algorithms in machine learning or data mining which usually contain a procedure of learning weight, importance and correlation amongst instances from

sizeable data. Those computational methods include PageRank [123], Social Network Analysis (SNA) [200], Principal Component Analysis (PCA) [204], and so on.

2.2.2.1 Eigenfactor

Similar to JIF and SNIP, the Eigenfactor is another citation-based impact indicator for scientific journals. Borrowing methods from Google's PageRank algorithm which ranks the influence of web pages [33], the Eigenfactor rates journals according to the number of their incoming citations, with citations from highly ranked journals weighted to make a larger contribution to the Eigenfactor than those from poorly ranked journals [18]. In other words, for two journals which received the same number of citations, the one which has citations from more significant journals will result in a higher Eigenfactor score than the other one.

The calculation of Eigenfactor is based on a citation matrix in which each entry e_{ij} indicates the number of citations from the j -th journal in the year of interest to articles published in the i -th journal during the preceding five years. When constructing the citation matrix for Eigenfactor, all self-citations are omitted. It helps the Eigenfactor to avoid the criticism which is commonly received by JIF that "it is easily operated by self-citations". The constructed citation matrix is then normalised by the total number of outgoing citations from each journal to create a matrix Z where each entry z_{ij} is defined as:

$$z_{ij} = \frac{e_{ij}}{\sum_k e_{kj}}. \quad (2.7)$$

The citation matrix Z is then normalised by dividing each entry in a column by the sum of that column to form a new matrix H . Similar to the Google's PageRank approach, the journal influence vector V is defined as the leading eigenvector of a modified matrix based on H . The i -th element in V is employed to weight citations from the i -th journal, i.e., the importance of journals is learned from the citations between them. Formally, the vector containing all journals' Eigenfactor scores is calculated as:

$$\text{Eigenfactor} = 100 \frac{H \cdot V}{\sum_i [H \cdot V]_i}, \quad (2.8)$$

where $[H \cdot V]_i$ indicates the i -th element in resultant vector of $H \cdot V$, and the Eigenfactor score of the i -th journal (denoted as Ef_i) is the i -th element in the resultant vector of Equation (2.8).

The Eigenfactor score provides a measure of the influence of total articles in a journal, rather than the influence per article in that journal. To make the results of Eigenfactor comparable to the JIF, an indicator named Article Influence is also proposed. The score of Article Influence for the i -th journal is calculated as:

$$AI_i = 0.01 \frac{Ef_i}{a_i}, \quad (2.9)$$

where a_i is the number of articles published by the i -th journal during the preceding five years divided by the total number of articles published by all journals of interest during the same time window. In both Eigenfactor and Article Influence, larger scores are gained by journals generating higher impact on their research fields.

Although the Eigenfactor and Article Influence are not directly maintained by WoS, they are already included by JCR as two measures of journal importance together with those indicators maintained by WoS. The Scopus has an indicator which is similar to the Eigenfactor, named as SCImago Journal Rank (SJR) [63]. The calculation of SJR is very similar to that of Eigenfactor, with the former being based on the Scopus database while the latter on the WoS. It also worth noticing that the idea of using the eigenvector to represent the importance of nodes in a network was firstly suggested in 1976 for ranking scientific journals [152], before it is widely known as the PageRank algorithm with Google. Ramifications of this idea for journal ranking can be found in [58] and [190].

2.2.2.2 Social Network Analysis

Social network analysis (SNA) is a type of techniques which are originally designed for investigating social structures through the use of network and graph theories [200]. It characterises structures of networks and graphs with a set of nodes (actors) and a set of links (relationships) that connect them. Since it has been founded, SNA is extensively applied to a wide range of disciplines including marketing and anti-terrorism. As the Eigenfactor is derived from web mining, recently many methods derived from SNA are employed to analysis the bibliographic data.

In [25], a set of connections amongst journals which named as the Reader Generated Network (RGN) is extracted from a digital library download log, and the journal impact rankings are calculated based on several social networking centrality metrics. The inter-connections in RGN are defined on the basis of the user download sequence

of publications in the library. The main assumptions underlying the generation of the RGN are: 1) if publications from a pair of journals are downloaded within a short time sequence by one user, a close relation between the pair of journals is expected; 2) the more frequently a certain pair of publications are downloaded by users, the greater degree of connection is assumed between the two journals in which this pair of publications are published. The indicators of journal impact are then calculated from the resulting networks using various social networking centrality metrics which include:

- Degree centrality: the sum of the degrees of edges connecting to a journal, normalised by the total degree of connections in the RGN;
- Closeness centrality: the average distance of shortest paths from a journal to all other journals in the RGN;
- Betweenness centrality: the frequency by which a journal is part of the shortest path between any pair of journals in the RGN.

An empirical study in [25] showed that its final ranking list deviates strongly from that of JIF, which indicates that the indicators derived from the RGN reflect different views from conventional journal impact. Another concept which is similar to RGN is the “Co-readership” [119]. This co-readership between two documents is established when at least one user has added the two documents to one user cart.

In machine learning and data mining, clustering and ranking are often regarded as two different techniques, each of which is usually individually employed to solve different problems. However, the RankClus framework can generate clusters as well as ranks for both journals and their authors based on a heterogeneous citation/publication network [179]. The RankClus algorithm starts from a random initialisation of clusters, and then it goes into an iteration with three steps: 1) rank journals/authors within each cluster; 2) estimate the current result based on objective functions; 3) adjust journals/authors to form new clusters. Not only the technical facet of RankClus is sophisticated, the underlying heuristics of this framework is also very enlightening and can be used in many other types of social network. These heuristics are [180]:

- A conference/journal is highly ranked if it publishes many papers from a good number of highly ranked scholars;
- A scholar is highly ranked if he/she has many publications in highly ranked conferences/journals;
- Scholars who have publications in the same conferences/journals are very likely in the same research field;
- A conference/journal belongs to one research field if it publishes papers of scholars who are in that research field;
- A group of conferences/journals belong to the same field if they mainly publish papers in that research field.

The quality of clustering and ranking in RankClus are mutually enhanced, and the clusters are getting more accurate and the ranking is getting more meaningful in each iteration. Moreover, its rank with clustering results can provide more informative views of bibliographic data.

2.2.2.3 Principal Component Analysis

Since many indicators for ranking, evaluation and comparison of academic journals have been proposed in the literatures, it is interesting to investigate which ones of them are most suitable for the measurement of journal impact. Besides the researches which directly compare the coverage of APDs [102, 142, 192] and compare the scores of a set of journals under different indicators [88, 89], a Principal Component Analysis (PCA) of existing journal impact indicators is reported in [24].

PCA is an orthogonal transformation which converts possibly correlated variables into linearly uncorrelated variables. These uncorrelated variables are called principal components. After the transformation, the first principal component has the largest possible variance, so that the original data points are separated as much as possible on that dimension; the second principal component in turn has the highest variance possible under the constraint that it is orthogonal to the first component, and so on [108].

About forty indicators that were calculated on the basis of both citation and user-log data are investigated in [24]. The Spearman rank correlations [93] are

calculated between journal ranking results provided by each pair of indicators. When these indicators are ranked by their average correlation to all other indicators, the JIF and SJR are very lowly-ranked (34 and 38 respectively), which indicates that they are isolated from other indicators. The resultant correlation matrix of all indicators is then subjected to a PCA process in which an eigenvalue decomposition is employed to identify the principal components [108]. By the analysis of PCA results, the authors of [24] concluded that:

- The group of user-log-based indicators are more strongly correlated to each other than the group of citation-based indicators;
- The indicators based on user-log represent more about a journal's prestige (long-term impact) than the citation-based indicators, while the JIF and SJR represent more about a journal's popularity (short-term impact);
- Some citation-based indicators are more closely related to their user-log-based counterparts than they are to other citation-based indicators.

Besides these interesting conclusions, the authors also revealed two critical challenges in the research of interaction between indicators: 1) an indicator is usually closely related to an APD, so it is difficult to distinguish the properties of an indicator from properties of the APD from which that indicator score was calculated; 2) there is no such universally accepted, gold standard of impact measure to calibrate new indicators to.

2.2.2.4 Aggregation of Existing APDs/Indicators

Apart from the research of correlations between those existing journal impact indicators, another promising direction is the integration of information provided by various indicators and APDs. In [118], an integration of multiple citation sources which include Google Scholar, Google Books, Google Blogs, PowerPoint presentations and course reading lists are investigated for two selected journals. The citation count of proposed integration (named as Integrated Online Impact) was nearly twice as high as those of both WoS and Scopus, which confirms that online citations are sufficiently numerous to be useful for the impact assessment of research.

The integration of citation sources is more like a technique of information acquisition. In order to make full use of the existing rankings of journals and compensate

for limitations of any individual ranking, a promising direction is combining several rankings into a single ranking [96, 161]. Following this appeal, one distinguishing approach which is based on the fuzzy measures and fuzzy integrals [135] has been proposed recently for the problem of combining existing indicators [11].

The importance of fuzzy integrals toward applications is its capability to express the possible interaction amongst members of the universe. This phenomenon cannot be captured by the standard Lebesgue integral, though it always plays a prominent role [135]. When solving the problem of aggregating journal impact indicators, the fuzzy integral can provide a model through which the citation statistics can be integrated in an understandable way, which in turn might contribute to the judicial use of such indicators in decision-making processes for experts in general. In [11], different indicator scores which are reported in the JCR are treated as features of journals while journal ranks published by the ERA are employed as labels to conduct supervised learning in the form of fuzzy integral. Their findings show that it is difficult to model the interactions between journal impact indicators due to inconsistencies and lack of monotonicity, but that the Choquet integral still performs well as a classifier.

Whilst such an initial approach of combining the fuzzy integral with journal assessment is very promising, so far there have only been a limited number of application papers concerning this direction. Much remains to be done in enabling machine learning techniques, especially the fuzzy-based learning models, to become robust and generic techniques in order to support journal quality assessments.

2.3 Link-based Framework of Bibliometric Data

Mining and analysis of link-based data are emerging research areas in machine learning and knowledge discovery. Compared with traditional data mining tasks, link mining focusses on applications where the datasets have rich structures, patterns, and linkages between objects [79]. Such a type of datasets are usually represented as networks or graphs. Also, those networks and graphs can be heterogeneous, that is, the nodes and links in such data structure can be of different types and hence have different attributes or features. Clearly, this is a very common issue in the bibliometric data where journals, papers, and scholars are involved. As indicated by the Bibliometric Information Retrieval System (BIRS), APD users like both the

graphical nature of information organisation and multi-level browsing systems [53]. However, existing researches in this direction often focus on particular datasets such as the news of Reuters [38], encyclopaedia articles [143], and blogs [23]. In this section, a link-based data structure for the assessments of academic journals is introduced, with demonstrations of its abilities in categorising journal impact indicators.

2.3.1 Link-based Framework

Generally, three categories of real-world entities are involved when referring to APDs: the journal, the publication (i.e., the paper) and the author(s). Amongst these, there are usually two types of direct relationships:

- 1) A paper can only be submitted to one journal, whilst a journal issue includes several accepted papers. This implies that the publication-journal relation is many-to-one.
- 2) A publication can be co-authored by more than one scholar and a scholar can have a number of publications. Thus, the publication-scholar relation is many-to-many.

Other than the above two types of straightforward relation, indirect and non-obvious relationships underlying the publication domain may also exist. For example, these may include the situation where well-known scholars are selected as reviewers or editors for certain journals. However, such links are not as many as the publication-journal links and publication-scholars in an APD, hence they are not considered in the present definition for simplicity. To describe both the three categories of entities and the two types of relationships between such entities in a formal representation the following link-based framework is introduced:

Definition 1. Let J , P , and S denote three sets of nodes which represent the three categories of entities typically involved in an APD, respectively: $J = \{j_x | x = 1, \dots, N^J\}$, j_x is a journal and N^J is the number of journals covered by the APD; $P = \{p_y | y = 1, \dots, N^P\}$, p_y is a paper and N^P is the number of papers covered by the APD; $R = \{s_z | z = 1, \dots, N^S\}$, s_z is a scholar and N^S is the number of scholars covered by the APD. The relationships between these entities, which are binary relations, can be readily

grouped into three subsets of links such that $L^{PJ} = \{(p_x, j_y) | (p_x, j_y) \in P \times J\}$, if p_x is published in j_y ; $L^{PS} = \{(p_x, s_y) | (p_x, s_y) \in P \times R\}$, if p_x is authored by s_y ; $L^{PP} = \{(p_x, p_y) | (p_x, p_y) \in P \times P\}$, if p_x is cited by p_y . The bibliographic framework of the APD is herein defined as a network: $\langle J \cup P \cup S, L^{PJ} \cup L^{PS} \cup L^{PP} \rangle$.

Such a link-based framework can be illustrated in Figure 2.3. In particular, a set of nodes of the same type and the links between these nodes are called a *layer*. Links between nodes that belong to different layers are termed *inter-layer* links and those within a given layer are termed *intra-layer* links.

Given a data source of journal articles, the implementation of such a framework is feasible. Most of the links involved can be established by observations such as reference lists and author lists in papers, which are accessible from APDs. For example, Digital Bibliographic Library Browser (DBL-Browser) is one of the bibliographical data browsers which implement the link-based framework. It provides a user-friendly interface which is able to search authors and publications by analysing social networks on the basis of L^{PJ} , L^{PP} and L^{PS} [113]. Note that the dissemination of academic research can be achieved in different ways. However, most of these are in the form of articles published in journals, proceedings, books, and theses. In particular, academic journals typically present the archived version of a certain research outcome, normally safeguarded by peer-reviews. It is because of this observation that journals are herein considered as the primary medium for academic publications. Other forms of publication can be investigated similarly with the link-based framework.

The nodes in different subsets can have different attributes. For example, $j_x \in J$ indicates a journal, thus it can have attributes such as publisher, ISSN number, age and so on; $p_y \in P$ can have attributes such as accepted time, published time, title, etc. Similarly, the links can also have different interpretations. However, certain attributes and links can be derived from a given framework. In this case, they can and should be omitted from the definition. For example, a citation count between two given journals can be derived as the aggregation of the defined links on the publication layer:

$$\text{Citation}(j_x, j_y) = |\{(p_i, p_j) | (p_i, p_j) \in L^{PP}, (p_i, j_x) \in L^{PJ}, (p_j, j_y) \in L^{PJ}\}| \quad (2.10)$$

where $j_x, j_y \in J$, $\text{Citation}(j_x, j_y) \in \mathbb{N}$ indicates the number of citations from journal j_y to j_x . Note that citations between journals are non-symmetric. If $j_x \neq j_y$, it is

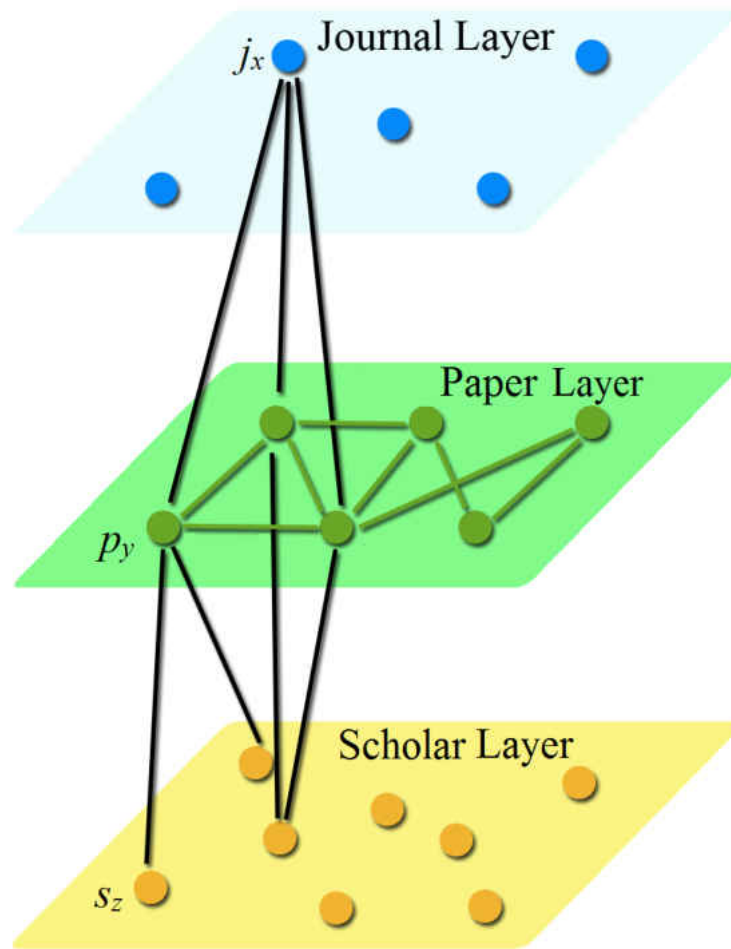


Figure 2.3: Link-based Framework of Bibliographic Data

not necessary that $\text{Citation}(j_x, j_y) = \text{Citation}(j_y, j_x)$. Note also that $\text{Citation}(j_x, j_x)$ indicates the number of self-citations for j_x .

The link-based framework has an intuitive appeal in capturing potentially useful knowledge for the evaluation of journal impacts. Although self-citation may be a controversial issue in research output assessment, the frequency of citations is generally a useful metric for demonstrating journal impact. Most of the conventional journal impact indicators are based on citations between journal articles. Therefore, it is interesting to note that existing journal impact indicators can normally be re-defined by manipulating the intra-layer and inter-layer links. For example, the JIF is defined as the ratio of the number of current year citations to the number of all articles published in that journal during the previous two years. Under the link-based

framework, the JIF value of a journal $j_x \in J$ in year t can be reformulated as:

$$\text{JIF}_t(j_x) = \frac{|\{(p_i, p_j) | (p_i, p_j) \in L^{PP}, (p_i, j_x) \in L^{PJ}\}|}{|\{p_i | (p_i, j_x) \in L^{PJ}\}|} \quad (2.11)$$

where p_i is defined on years $t - 2$ to $t - 1$; p_j is defined on year t . Thus, if p_i is an article published in 2010-2011, and p_j is an article published in 2012, then the above denotes the 2012 JIF value of a journal j_x , $\text{JIF}_{2012}(j_x)$. The alternative of 5-year JIF can also be calculated similarly. Indeed, the above reformulation of JIF is able to represent the 5-year JIF only with a minor modification by extending the time constraint on the citation source to 5 years, that is, p_i is defined on years $t - 5$ to $t - 1$.

Apart from the paper layer, the scholar layer also contains useful information of journal impact such as the researcher behaviour based indicators [95]. The provided four indicators described in Equations (2.3)-(2.6) can be redefined on the scholar layer. For simplicity, only the reformulation of the PB indicator is shown here, though the other three can easily be done in the same way. The publishing breadth of a journal indicates that: if a high percentage of selected authors have published articles in a certain journal, then that journal is rated higher than the others in which smaller percentages of those selected authors have authored articles. Given a journal $j_x \in J$, this concept is captured in the framework as:

$$\text{PB}(j_x) = \frac{|\{s_z | (p_y, s_z) \in L^{PR}, (p_y, j_x) \in L^{PJ}\}|}{|S|} \quad (2.12)$$

where S is a set of selected influential scholars.

In addition to the typical indicators illustrated above, many other forms of knowledge can be captured using the link-based framework. They may be designed from different perspectives and hence, different links may be involved and aggregated in describing the knowledge. Taking the CiteSpace [40] as an example, the link-based framework is employed to create visualised knowledge for characterising and interpreting the structure and dynamics of co-citation clusters.

2.3.2 Categorisation of Journal Indicators

Since many indicators have been proposed in the literature, their categorisation is very helpful for studying their different characters and usages. However, most of

the researching works in the literature did not cover this problem, except a two-dimensional space is proposed to categorise the existing indicators in [25] (as shown in Figure 2.4).

In this two-dimensional space, one axis indicates whether the raw data is defined by readers or by authors. The other axis corresponds to whether the calculation is based on frequency or structural metrics. Following this taxonomy, the JIF is categorised to the author-frequency quadrant while the RGN should be deemed as a reader-structural one. However, since the calculations of the indicators are becoming more complex and integrated, making the decision of frequency or structural based is sometimes difficult. If one considers SNIP and Eigenfactor as examples, they could be accepted as frequency-based indicators since it counts the number of citations from other journals to the considered journal; but they also could be accepted as structural ones since their weights of subject field and citation source are defined in a more structural way.

Another issue of the two-dimensional space is its inability in distinguishing the citation-based methods with the author-defined methods. Although the references of a publication are listed by its author(s), the citations from one citing paper to the cited papers should be treated as linkages between publications which are different from linkages between scholars and publications. The concept of “author-defined” can be more used to describe those indicators which are based on the reputations of authors, such as the PB, PM, PI, and PW. Therefore, the citation-based methods should not be mixed with the author defined ones when categorising journal impact indicators.

Besides distinguishing the statistical-based and learning-based indicators, by which the indicators reviewed in this chapter are organised, an additional link-based categorisation is employed here to further differentiate journal impact indicators. From such a link-based perspective, those existing indicators are classified by sets of links from which they are calculated. For examples, both the JIF and PB are author-oriented and frequency-based in the two-dimensional categorisation. However in the link-based categorisation, the JIF is based on the intra-layer links L^{PP} and inter-layer links L^{PJ} , while the PB is defined on L^{PS} and L^{PJ} . The link-based categorisation of the indicators described in this chapter is given in Table 2.1, where L^{JJ} denotes a set of intra-links amongst journals. In the Eigenfactor and SJR, L^{JJ} is the citation

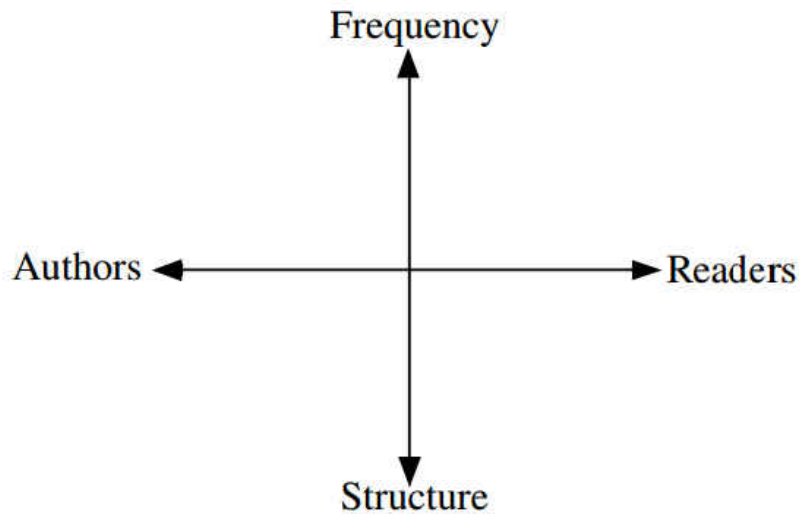


Figure 2.4: Frequency vs. Structure and Author vs. Reader Impact Indicators

Table 2.1: Link-based Categorisation of Indicators

Indicator	L^{PP}	L^{JJ}	L^{PS}	L^{PJ}
JIF, 5-year JIF	✓			✓
RIP, RDCP, SNIP	✓			✓
h-index for journal	✓			✓
PB, PM, PI, PW			✓	✓
RankClus	✓		✓	✓
Eigenfactor, SJR		✓		✓
RGN		✓		✓

count between journals which is defined by Equation (2.10); while in the RGN, L^{JJ} is defined by the user logs.

By aggregating various indicators which are defined on different sets of links, the resultant assessment can produce an overall estimate of journal impact which considers a wider range of journal influence than that produced by each individual indicator.

2.4 Summary

This chapter has firstly introduced the concept and development of APD for the purpose of supporting the assessment of academic journals. The components and their functions of an APD have been described. Several well-known representatives of APDs: IEEE Xplore, ArnetMiner, WoS, Scopus, and Google Scholar have been explained in detail.

More importantly, this chapter has presented a review of several famous journal impact indicators. Their underlying respective inspirations span a wide range of techniques, including bibliometrics, machine learning, and data mining. Furthermore, to facilitate the establishment of a fusion of existing indicators for journal clustering in the next chapter, methods which investigate the correlation and aggregation of existing indicators have been briefly reviewed.

In addition, to illustrate the information contained by APDs in a uniform manner, a link-based framework of bibliographic data has also been identified. Existing journal impact indicators, irrespective of whether they are statistical-based or learning-based, are uniformly categorised by the sets of links based on which they are calculated. If an indicator is already introduced in this chapter, its name will be directly used without any interpretation in the remainder of this thesis. Otherwise, a brief introduction will be given for any newly encountered indicator.

Chapter 3

Fused-link: Distance-based Aggregation of Journal Indicators

EACH indicator has its strengths and limitations and their ranking results can be different. A promising direction which can compensate the shortcomings of an individual indicator, is the integration of existing indicators. This chapter presents an approach for link-based indicator fusion as an alternative for analysing the quality of academic journals. Journals are evaluated in a multi-dimension space which is constructed by various impact indicators. The strength of a fused-link between two given journals is defined as the inverse of distance between those two journals. Two distance metrics: Euclidean and Manhattan distances are utilised to support the development and analysis of the fused-links.

Both classification and clustering algorithms built upon the basis of fused-links are tested against five datasets of journals from the area of computer science and informatics. Through comparisons with the use of advanced learning mechanisms such as support vector machines and decision trees, the systematic experimental results demonstrate that the proposed fused-link based approach helps to capture and reflect the impact of academic journals and consequently, their qualities effectively.

The structure of this chapter is organised as follows. Section 3.1 describes the concept of the fused-links between journals. Section 3.2 presents their application to both classification and clustering for journal ranking. Sections 3.3 and 3.4 present experimental evaluation of the proposed approach, combined with a discussion of the results. Finally, Section 3.5 summarises the chapter.

3.1 Fused-links

With the aid of on-line APDs such as WoS and Scopus, the calculation of individual journal impact indicator may be carried out efficiently and updated with the time. A number of indicators are indeed widely accepted and applied by scholars. Unfortunately, none of them is powerful enough to characterise all aspects of journal impact in the real-world. To compensate the potential bias of using single indicators, thereby enriching links in the journal layer, the following fused-link approach is proposed that integrates the results of various indicators.

The calculation of journal impact indicators depends mainly on the links between publications L^{PP} and the links between publications and scholars L^{PS} (see Table 2.1 and Figure 2.3). This is supported by the argument that citations between publications is often the single most significant link between journals, in assessing the impact of a certain journal. However, since journal ranking is in essence to compare the quality of journals, the links on the journal layer are important for discerning an impact indicator. An indicator that employs journal links typically aims to evaluate a single journal and usually focusses on one particular aspect of journal impact, such as the Eigenfactor and RGN. When human experts assess the quality of journals, peer-comparison is commonly and sensibly used to support their judgement.

Although a journal is a collection of articles and its impact is heavily affected by the quality of its articles, journals also possess properties that cannot be simply extracted from their collected publications or their authors/readers. As shown in Section 2.3, most of the existing journal impact indicators can be reformulated as calculations defined over links. These indicators describe different features of the links regarding a certain journal. Journals which have similar impact indicator scores can be expected to exhibit similar features of their links.

Note that different indicators may have different ranges of values. For example, most of the journals have JIF scores between 0 to 5, while their Eigenfactor scores are from 0 to 0.1. Thus, to ease comparisons between different journals with multiple features, a data-driven normalisation is adopted to unify the representation of the individual underlying indicators. That is, given a set of journals J , for each indicator I' , the score of a journal $I'(j_x) \in \mathbb{R}$ is normalised to $I(j_x) \in [0, 1]$ as:

$$I(j_x) = \frac{I'(j_x) - \min_{j_i \in J}(I'(j_i))}{\max_{j_i \in J}(I'(j_i)) - \min_{j_i \in J}(I'(j_i))}. \quad (3.1)$$

To assess the quality of journals with respect to multiple indicators, a concept termed *fused-link* is introduced here, which measures how similar two journals are in terms of multiple indicators collectively. The more similar the stronger the corresponding fused-link. Formally, the strength of a fused-link between two journals is defined as follows:

Definition 2. Given ‘ m ’ normalised journal impact indicators I_1, \dots, I_m , a journal j_x is evaluated by them: $\vec{I}_{j_x} = (I_1(j_x), \dots, I_m(j_x))$. The strength of a fused-link between two journals $j_x, j_y \in J, x \neq y$ is defined as:

$$Fused(j_x, j_y) = \begin{cases} \frac{m}{d(\vec{I}_{j_x}, \vec{I}_{j_y})} & , \quad \text{if } d(\vec{I}_{j_x}, \vec{I}_{j_y}) \neq 0 \\ \infty & , \quad \text{otherwise.} \end{cases} \quad (3.2)$$

where $d(\vec{I}_{j_x}, \vec{I}_{j_y})$ is the distance between \vec{I}_{j_x} and \vec{I}_{j_y} with respect to a distance metric d .

The employment of a distance metric in defining the fused-link strengths helps to capture the intuition that two journals linked by a strong connection should have similar scores with respect to the same given indicators. That is, the impact of such linked journals on the relevant academic research should be similar. Conventional methods evaluate journals with regard to individual features. The fused-links allow not only for the assessment of journal impact by considering multiple indicators together, but also for the use of the neighbourhood information on the link strengths to determine the impact of a journal if it is less known to a certain researcher or community.

To illustrate the concept of fused-links, a set of five journals are individually evaluated using two separate indicators of JIF and Eigenfactor, as listed in Table 3.1. The fused-links between these journals are evaluated by the use of the Euclidean distance metric, that is, a fused-link between two journals is defined as:

$$E\text{-Fused}(j_x, j_y) = \begin{cases} \frac{m}{\left(\sum_{i=1}^m (I_i(j_x) - I_i(j_y))^2 \right)^{\frac{1}{2}}} & , \quad \text{if } \sum_{i=1}^m (I_i(j_x) - I_i(j_y))^2 \neq 0 \\ \infty & , \quad \text{otherwise.} \end{cases} \quad (3.3)$$

Table 3.1: Examples of Journals: Fused-link

J	I'_1 (JIF)	I_1	I'_2 (Eigenfactor)	I_2
j_1	7.806	1.000	0.00571	0.074
j_2	5.027	0.554	0.05002	1.000
j_3	2.683	0.178	0.00895	0.141
j_4	2.674	0.177	0.01409	0.249
j_5	1.574	0.000	0.00219	0.000

The resulting fused-links and their strengths are shown in Figure 3.1. Each point in Figure 3.1 has two coordinates, representing the normalised Eigenfactor and JIF scores. The distance between any two points indicates the dissimilarity of the pair of journals with respect to the given two impact indicators. For instance, j_4 has a stronger fused-link to j_3 as compared to j_2 and j_1 , since the position of j_4 is closer to that of j_3 than to either the position of j_2 or that of j_1 , within the feature space constructed by the two impact indicators. This situation can be easily generalised to situations of a higher dimensionality. Although the Euclidean metric is employed in this definition, any other distance metric may be used as alternative.

To build the fused-links, both Euclidean and Manhattan distance metrics are employed in order to facilitate comparisons as different distances measures may have significant influence of ranking. The Manhattan distance is chosen because it is well used in performing path-finding and many other tasks [150, 176], and also because it is computationally simple. It is defined as the sum of absolute differences of two objects on each dimension. Formally, given two journals $j_x, j_y \in J, x \neq y$, the strength of their fused-link using Manhattan distance is:

$$\text{M-Fused}(j_x, j_y) = \begin{cases} \frac{m}{\sum_{i=1}^m |I_i(j_x) - I_i(j_y)|} & , \text{if } \sum_{i=1}^m |I_i(j_x) - I_i(j_y)| \neq 0 \\ \infty & , \text{otherwise.} \end{cases} \quad (3.4)$$

Note that in the following presentation, the ‘‘M-’’ prefix represents the cases where Manhattan distance metric is used to calculate fused-links while ‘‘E-’’ prefix represents the cases where Euclidean distance is employed.

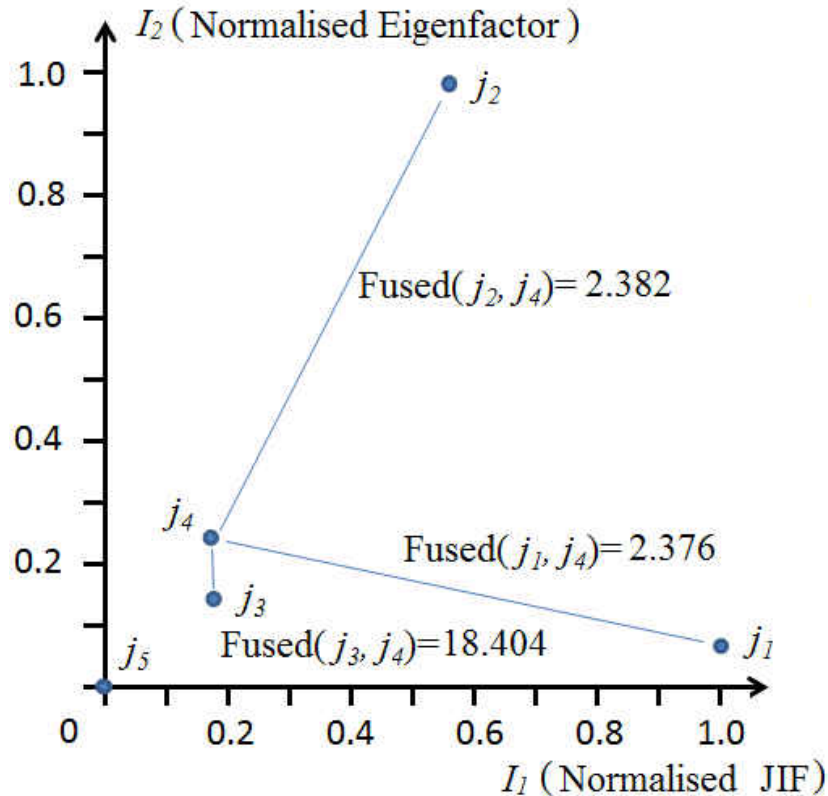


Figure 3.1: Example of Fused-links between Journals

3.2 Applications of Fused-links

Fused-links are designed to aid in assessing the ranking of different journals with respect to different performance indicators. If a-priori knowledge of the impact of those journals of interest is acknowledged (for example, with a small set of journals having been labelled as “A*-class”, “A-class” or “B-class”), fused-links can be used to classify the rank of other unlabelled journals. If however, the background of journal ranking is not available, fused-links are useful to detect groups of journals which have impacts of a similar strength. These two types of potential application correspond to the classical problems of classification and clustering, as presented below.

3.2.1 Classification

This is to identify which class a new or unseen observation belongs to, on the basis of a training set of data containing instances whose class labels are known. For the journal ranking problem where a ranking list is available, the aim is to determine the quality

position of journals which have not yet been ranked. Unfortunately, sometimes the number of journals on a given ranking list is very small and incomplete. For example, only a “top-10” list is provided. In this case classification results can be rather poor. However, fused-links between journals indicate the similarity of the impact of individual journals, the ranks of two journals with strong links are likely to be ‘close’ to each other. This provides an intuitive means to estimate the rank of one journal using those of its neighbours. From this observation, the k Nearest Neighbours (k NN) classifier [46] is herein combined with fused-links to classify journal ranks.

To classify an unlabelled object (or journal in the present problem), the k nearest objects of the unlabelled are selected to vote for which class it belongs. By “nearest” it refers to those objects within a feature space whose distance to the unlabelled object is measured to be the shortest. The fused-link strength between any two journals is defined in a feature space constructed by various impact indicators, so that it is readily applicable to k NN as a distance metric. Intuitively, journals with a close distance between them are more likely to have a similar impact. Thus, the ranking of those nearest neighbours offer more useful information than the ranks of their far neighbours. Inspired by this observation, the Distance-weighted k NN (D- k NN) method [56] is modified here to classify the ranks of journals, where the votes of the nearest neighbours are weighted by their corresponding fused-link strengths to the journal under consideration. This leads to the following fused-link based method for journal ranking classification.

Suppose that a set of journal ranks C is given, and $\mu'(c, j_x) \in \{1, 0\}, c \in C, j_x \in J$ is a binary membership function which satisfies that if j_x is of rank c then $\mu'(c, j_x) = 1$, else $\mu'(c, j_x) = 0$. The rank $\mu(j_x) \in C$ of an originally unranked journal $j_x \in J$ is determined by:

$$\mu(j_x) = \arg \max_{c \in C} \sum_{i=1}^k \text{Fused}(j_x, j_i^x) \cdot \mu'(c, j_i^x) \quad (3.5)$$

where $j_i^x \in J, i = 1, \dots, k$ is the i -th nearest neighbour of j_x .

The implementation of this classification method is straightforward. The simplest, brute-force way to find an object’s nearest neighbours is to compute its distance to all the training examples. Such an implementation has a test time complexity of $O(N^J)$ [201], where N^J is the size of dataset.

3.2.2 Clustering

In general, the task of clustering is to assign a set of objects into groups (namely clusters) such that the objects in the same group are similar to each other, and dissimilar to those in the other clusters [105]. In this chapter, journal clustering attempts to seek a partition of a collection of academic journals J , using fused-link strengths as the similarity measures between any two given journals. There are a number of generic clustering methods that may be employed to implement the approach. Among them, the k -means algorithm is popular due to its simplicity [105] and success in solving real-world problems [42, 83, 167]. Having noticed this, this algorithm is herein integrated with fused-links to demonstrate the proposed ideas. The pseudo-code for the implementation of fused-link-based k -means algorithm is showed in Algorithm 3.2.1.

k : the number of clusters;
 $J = \{j_x | x = 1, \dots, N^J\}$: a set of journals represented by their indicator scores;
 $C = \{c_i | i = 1, \dots, k\}$: k centroids of the k clusters;
 $\mu(j_x) \in \{1, \dots, k\}$: the cluster label of $j_x, j_x \in J$.

- 1: **for** $i = 1, \dots, k$ **do**
- 2: randomly pick c_i from J
- 3: $\mu(c_i) = i$
- 4: **end for**
- 5: **repeat**
- 6: **for** $x = 1, \dots, N^J$ **do**
- 7: $\mu(j_x) = \arg \max_{i \in \{1, \dots, k\}} \text{Fused}(j_x, c_i)$
- 8: **end for**
- 9: **for** $i = 1, \dots, k$ **do**
- 10: $c_i = \text{average of } \{j_x | \mu(j_x) = i\}$
- 11: **end for**
- 12: **until** none of $c_i \in C$ changed
- 13: **return**

Algorithm 3.2.1: Fused-link-based k -means for Journal Clustering

The time complexity of k -means is $O(N^J kl)$, where N^J is the number of journals in the journal layer, k is the number of clusters, and l is the number of iterations taken by the algorithm to converge. Usually, k and l are fixed in advance and so the algorithm has linear time complexity in the size of the data: $O(N^J)$ [105].

3.3 Experimental Setup

This section reports on the work carried out for evaluation of proposed fused-link to journal ranking with classification and clustering. The design of the experiments is first presented, including the construction of datasets and the performance criteria employed.

3.3.1 Datasets

A credible demonstration of the ranking ability of the link-based methods is to compare their ranking results with human expert opinion. The Ranked Journal List (RJL) provided by the ERA 2010 has involved a large group of scholars to rank a large number of academic journals. Although many debates surrounded the end result of RJL [44], it has been employed by researchers as a benchmark to compare journal ranking outcomes produced by automated mechanisms versus human experts [11]. In the present experiments, to evaluate the ranking results of clustering and classification based on the proposed fused-link based approach, the result of RJL is also assumed to be the ground truth in comparing the “accuracy” of different methods. Each journal in RJL has a rank in the domain $Ranks = \{A^*, A, B, C\}$, where rank A^* indicates top journals in a certain research area, and the significance and popularity of journals are decreasing, while the percentage of amount increases from rank A^* down to rank C. Following this, each journal studied in the experiments below is assigned a label taken from the domain *Ranks*.

Seven indicators that are reported in the JCR (2010) are selected as base-indicators to compute fused-link strengths between journals. These are: Total Cites (number of times the journal being cited in 2010); JIF; 5-year JIF; Immediacy Index (ratio of cites to the current articles over the number of those articles); Cited Half-Life (median age of the articles cited); Eigenfactor; and Article Influence (ratio of the Eigenfactor score to the total number of articles considered). JCR has a long history of applications for researchers and librarians in choosing their reading lists. All indicator score calculations in JCR are based on the same range of journals, i.e., journals which are indexed by WoS. A screenshot of the JCR that showing journals from one of the subject categories: Computer Science–Artificial Intelligence of year 2010 is given in Figure 3.2.

Mark	Rank	Abbreviated Journal Title <i>(linked to journal information)</i>	ISSN	JCR Data ^(j)						Eigenfactor [®] Metrics ^(j)	
				Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Articles	Cited Half-life	Eigenfactor [®] Score	Article Influence [®] Score
<input type="checkbox"/>	1	IEEE T. PATTERN ANAL.	0162-8828	25060	5.308	7.534	0.625	176	9.5	0.04969	2.802
<input type="checkbox"/>	2	INT. J. COMPUT. VISION.	0920-5691	9898	5.151	6.986	0.808	99	6.9	0.02168	2.648
<input type="checkbox"/>	3	SIAM J. IMAGING SCI.	1936-4954	322	4.500	4.500	0.273	44	1.9	0.00351	2.661
<input type="checkbox"/>	4	IEEE T. EVOLUT. COMPUT.	1089-778X	4681	4.403	5.432	0.333	54	8.0	0.00702	1.371
<input type="checkbox"/>	5	MED. IMAGE ANAL.	1361-8415	2901	4.364	4.630	0.705	61	6.5	0.00720	1.308
<input type="checkbox"/>	6	INT. J. NEURAL SYST.	0129-0657	816	4.237	2.581	0.421	38	3.6	0.00122	0.318
<input type="checkbox"/>	7	INT. J. INF. TECH. DECIS.	0219-6220	534	3.139	2.191	0.356	45	3.0	0.00078	0.198
<input type="checkbox"/>	8	J. MACH. LEARN. RES.	1532-4435	4766	2.974	4.967	0.456	114	6.1	0.02175	2.448

Figure 3.2: Screenshot of JCR

Note that in this work, it is the method that integrates individual impact indicators together with fused-links that is investigated, rather than the selection of base indicators themselves. In fact, experts from different research areas may select different base-indicators or journal-sets reflecting their own preferences to compare journals. This is different from classifier/clustering ensemble, where diversity in base-classifiers/clustering members is an important factor to consider in building a robust ensemble [170].

In collecting sets of journals for the experiments, five research branches of computer science in the JCR are selected: Artificial Intelligence (AI), Information Systems (IS), Interdisciplinary Applications (IA), Software Engineering (SE), and Theories & Methods (TM). Amongst them, only those journals that are ranked both in RJL and indexed by the JCR are considered as a valid data object (in order to have the ground truth to entail comparison). If a journal contains missing values in the JCR, then its whole data record is removed from the experimental data. The resulting datasets are then grouped with respect to the above five subdisciplines. A summary of the datasets is shown in Table 3.2. In RJL, journals are allocated a rank with respect to a

discipline group, so that the A* rank accounts for the top 5%, the A rank accounts for the next 15%, the B rank accounts for the next 30%, and the bottom 50% is rank C. Since ranges of journals and divisions of research branches are different in RJL and JCR, unlike in the use of JCR, journals ranked from A* down to C-graded are no longer evenly distributed, but nonlinearly.

Table 3.2: Summary of Datasets: Fused-link

Dataset	A*	A	B	C	Total
Artificial Intelligence (AI)	11	24	31	21	87
Information Systems (IS)	15	19	32	20	86
Interdisciplinary Applications (IA)	7	27	25	24	83
Software Engineering (SE)	10	19	32	8	69
Theories & Methods (TM)	13	31	16	12	72

3.3.2 Evaluation Criteria

Accuracy is used to indicate the consistency between the proposed link-based methods and RJL, which is defined as the ratio of correctly classified or clustered objects to the total number of objects in the dataset. As the outcomes of applying link-based approach are compared with the ranking results of RJL, the “correctly classified/clustered objects” are deemed to be the journals whose assigned ranks are consistent with their ranks in expert-devised RJL. The *within-1* accuracy [11] is another criterion used, which is often adopted in classification or clustering problems where the classes/clusters of the objects are ordered (as they are for the present journal ranking task). Following this criterion, an A*-ranked journal is a “correctly classified object” if it is classified into either A* or A; an A-ranked journal is also deemed as correctly classified if the result is A*, A, or B; etc.

3.4 Results and Discussion

In this set of experiments, each dataset is split into subsets for 3-fold cross validation [16]. This helps to make both the training and the testing sets contain enough amount of data [115]. The reported results are based on an average of 50 times of the 3-fold cross validation.

For comparison, the standard (non distance weighted) k NN and D- k NN are implemented and run, with each having an Euclidean distance-based version and a Manhattan distance-based version. That is, four different methods are considered. In addition, the number of nearest-neighbours k for each method is set from 3 to 7 over different runs. The difference between k NN and D- k NN is that the standard k NN gives all nearest-neighbours equal weights when assigning the rank of an unlabelled object. In this case, the $\text{Fused}(j_x, j_i^x)$ in equation (3.5) is replaced by 1. In response to the different values of k , all of the results show a consistent trend for the four methods on every dataset. Thus, the accuracies and within-1 accuracies measured on these datasets are averaged respectively to represent their overall performances. The trends of their accuracies and within-1 accuracies are shown in Figures 3.3 and 3.4, respectively.

Generally, all these four methods increase their accuracy and with-1 accuracy along with the increase of k . This indicates that when ranking a journal by the ranks of its neighbours, considering more references will generate a more accurate rank. It is shown that for both k NN and D- k NN, the use of Manhattan distance metric leads to a better result than using the popular Euclidean distance, in terms of both accuracy and with-1 accuracy. One possible reason for having achieved such results is the use of limited journal datasets, each of which has less than one hundred journals (instances), whilst consisting of seven base-indicators (dimensions). This also conforms to the finding obtained in [3] that the Manhattan distance metric is preferable for high dimensional data mining applications.

Importantly, the results also show that for both Manhattan and Euclidean based methods, D- k NN has a better performance than k NN. In particular, the Manhattan distance based D- k NN performed the best amongst all four methods (while the largest number of nearest neighbours employed, $k = 7$, leads to the highest performance). This reflects the advantage of utilising fused-link strengths as a weighting parameter for journal rank classification.

To further analyse the results achievable by the link-based methods, advanced classification techniques such as the Support Vector Machines (SVM) [181] and decision trees [157] are also tested. The Sequential Minimal Optimization (SMO) [153] and J48¹ [158] algorithms released with the Weka software [203] are used to

¹J48 is a Java implementation of the C4.5 decision tree in the Weka software.

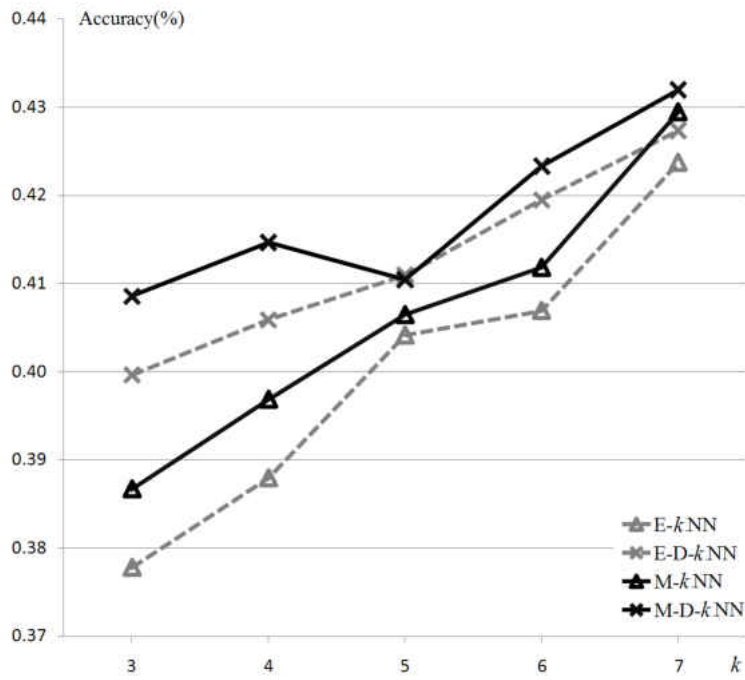


Figure 3.3: Trend of Accuracy against an Increase in k

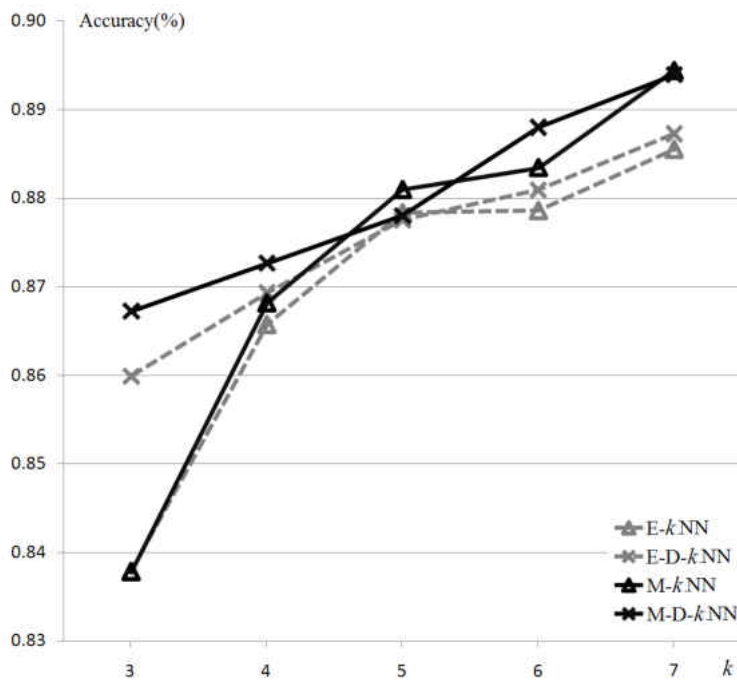


Figure 3.4: Trend of Within-1 Accuracy against an Increase in k

represent these two classical classification approaches, with the polynomial kernel selected to implement SMO. The resultant accuracy (%) and the within-1 accuracy (%) are shown in Tables 3.3 and 3.4, respectively. To validate the significance of the experiment results, the paired-t tests are carried out between M-D- k NN and the rest on each dataset.

Table 3.3: Comparison of Classification Accuracy (%): Fused-link. The best two performances on each dataset are highlighted in boldface, the sign “(*)/(v)” indicates that the corresponding result is significantly ($p < 0.05$) worse/better than that achieved by M-D- k NN

Dataset	J48	SMO	E-D- k NN	M-D- k NN
AI	41.47±4.73(*)	44.02±2.29(*)	44.64±3.54	45.20±5.11
IS	38.37±4.98(*)	39.72±1.65(*)	43.14±3.93	43.70±3.91
IA	33.37±4.12(*)	37.54±3.70	39.33±3.46(v)	38.41±3.73
SE	47.01±4.25(v)	53.16±1.36(v)	43.30±4.89	43.77±4.38
TM	43.78±5.69	44.86±1.82	43.28±4.11(*)	44.92±4.37
Avg.	40.80	43.86	42.74	43.20

Table 3.4: Comparison of Classification Within-1 Accuracy (%): Fused-link. The best two performances on each dataset are highlighted in boldface, the sign “(*)/(v)” indicates that the corresponding result is significantly ($p < 0.05$) worse/better than that achieved by M-D- k NN

Dataset	J48	SMO	E-D- k NN	M-D- k NN
AI	89.86±3.27(*)	97.45±0.99(v)	96.02±2.00	96.14±1.70
IS	83.72±3.18(*)	87.05±1.80(*)	88.70±3.00(*)	89.37±2.61
IA	80.00±3.63(*)	83.37±3.53	81.57±2.75(*)	82.55±2.93
SE	90.09±2.35(*)	92.12±1.53(v)	89.94±2.44(*)	91.36±2.23
TM	84.08±3.85(*)	83.36±0.20(*)	87.42±3.01	87.56±2.90
Avg.	85.55	88.67	88.73	89.40

In terms of accuracy, the results show that M-D- k NN outperforms J48 and SMO on all the datasets investigated except SE. Moreover, M-D- k NN is able to achieve a higher within-1 accuracy than all other methods. This demonstrates that fused-link

based approach is capable of classifying journals effectively by combining the existing impact indicators in an intuitive way.

More particularly, J48 has a lower accuracy/within-1 accuracy compared with other methods and it is significantly worse than M-D- k NN on most tested dataset. This is likely due to its being overfitted to the observed journals [11]. One possible explanation for the less ideal performance of link-based methods on the dataset SE is that B-ranked journals are dominant in the number in dataset and the number of journals belonging to each rank is rather unbalanced. In this case, a classifier biased to the dominant class may happen to have a higher accuracy than un-biased ones. SVM has shown its robust classification ability in many problems [55, 75]. However, the main disadvantage of SVM is that it works more like a “black-box”; its training procedure is less intuitively interpretable [17]. In situations where an interpretable result is required, an intuitive way such as direct integration of journal impact factors for ranking (which is what fused-links approach basically does) will be more welcomed by scholars, than function-based methods like SVM. As such, the link-based approach clearly offers more promising application to the journal ranking problem, given their generally high accuracy and easy interpretability.

Similar to the classification application, Euclidean-distance-based and Manhattan-distance-based k -means are compared for clustering. The number of clusters is set to the number of ranks ($k = 4$). The label of majority journals in each cluster is deemed as the rank of the journals within it. 50 times random centroid initialisation are tested for both methods [149]. The overall averaged and best-10 averaged results are reported respectively in Tables 3.5 and 3.6 for the k -means approach.

The first result to note is that in terms of accuracy, when applied to the k -means algorithm, the difference due to the use of two different distance metrics is not so obvious as with the case of D- k NN. Overall, the accuracies of the proposed clustering methods are better than that achievable by J48 and are only slightly lower than those attainable with fused-link-based classification. However, the clustering results in terms of within-1 accuracy are not so good as those obtained by the classification methods. A likely reason is that the k -means algorithm is sensitive to initial partitioning. Nevertheless, the best-10 averaged results outperform the overall averaged ones significantly. This demonstrates with the initial centroids appropriately selected, k -means may result in very good performance. The results

Table 3.5: Comparison of Clustering Accuracy (%): Fused-link. The results which are better than the best achievable by their classification counterparts are highlighted in boldface, the sign “(*)/(v)” indicates the corresponding result is significantly ($p < 0.05$) worse/better than that achieved by M- k -means

Dataset	E- k -means	M- k -means	E-best-10	M-best-10
AI	44.28±3.03(*)	49.26±2.06	47.82±1.23	51.95±0.48
IS	43.15±2.79(v)	41.19±3.06	45.58±0.74	45.23±1.16
IA	39.47±3.30	39.40±1.78	42.77±1.02	42.17±1.39
SE	42.67±2.27	42.23±4.39	45.80±1.40	48.12±3.79
TM	42.83±2.69	43.21±2.81	45.56±1.28	46.53±2.95
Avg.	42.48	43.06	45.50	46.80

Table 3.6: Comparison of Clustering Within-1 Accuracy (%): Fused-link. The results which are better than the best achievable by their classification counterparts are highlighted in boldface, the sign “(*)/(v)” indicates the corresponding result is significantly ($p < 0.05$) worse/better than that achieved by M- k -means

Dataset	E- k -means	M- k -means	E-best-10	M-best-10
AI	89.15±3.55(*)	92.01±2.06	93.10±0.00	94.02±0.48
IS	80.84±4.32	81.82±2.23	83.72±3.10	86.16±1.69
IA	79.25±2.72(*)	80.41±2.55	83.61±2.06	83.49±0.58
SE	88.37±3.88(v)	83.86±4.39	93.62±1.56	91.01±2.54
TM	82.13±3.31	82.81±4.78	85.83±1.71	87.50±1.60
Avg.	83.95	84.18	87.98	88.44

also show that the clusters formed by following the fused-link based approach reflect the distribution of journal ranks in RJL.

To further appreciate the appropriate determination of journal ranks using the proposed method, a particular example of generated ranks is provided in Table 3.7. Journals which have top-10 JIF values in dataset AI are selected and their ranks with fused-link are given by voting, from the 50 times the results obtained using Manhattan-distance-based clustering. The table shows that even the journals which have a very high JIF value can have a corresponding B or C-rank in RJL. This indicates that the result of journal ranking by JIF itself may significantly deviate from

that returned by human experts. All the journals listed in Table 3.7 are ranked at least to the level of B by fused-link clustering and 6 of the 10 ranks are identical to those given by RJL. It is also interesting to note that two journals that are not in the top-10 JIF list, but usually highly commended by the research community: Artificial Intelligence (JIF: 2.533, JIF-rank: 20) and Machine Learning (JIF: 1.967, JIF-rank: 33) are ranked to A* by both the RJL and the proposed method.

Table 3.7: Example of Ranks by JIF, RJL and Fused-link

JIF-rank	Title	JIF	RJL	Fused-link
1	IEEE T PATTERN ANAL	5.308	A*	A*
2	INT J COMPUT VISION	5.151	A	A*
3	SIAM J IMAGING SCI	4.500	not given	–
4	IEEE T EVOLUT COMPUT	4.403	A*	A*
5	MED IMAGE ANAL	4.364	A*	A*
6	INT J NEURAL SYST	4.237	B	B
7	INT J INF TECH DECIS	3.139	C	B
8	J MACH LEARN RES	2.974	A	A*
9	COMPUT LINGUIST	2.971	A*	A*
10	IEEE T IMAGE PROCESS	2.918	A*	A*
11	IEEE COMPUT INTELL M	2.905	C	B

Although the link-based methods are better than the SMO and J48 in terms of accuracy, the general results for accuracy regarding the journal ranks is low. This may be caused by the (no) selection of indicators. If selected indicators are not sufficiently different amongst themselves or they do not offer sufficient discriminative power over the quality of journals, the journals of different ranks may be heavily overlapped. This will obviously lead to poor performance, especially when the cluster number k is set to a small number. Another reason may be that most of the journals are not obviously better or worse than others, thus their ranks are more likely to be affected by the subjective bias of the experts. For example, the journal ranking list produced by one country is likely to involve higher ranks for the journals published in that country. Despite these observations, the experiments conducted have generally demonstrated that the fused-link approach works well overall in performing the task of ranking academic journals.

3.5 Summary

This chapter has presented a link-based framework for classification and clustering, with a focussed application to the problem of journal ranking. It has proposed a fused-link based representation between journals, which allows for fusion of different journal impact indicators to support the assessment of academic journal quality. Both classification and clustering algorithms built upon the basis of fused-links are tested against five datasets of journals from the area of computer science and informatics. Experimental results have shown that the ranking results of the fused-link methods are consistent with RJI, which is produced by a large group of journal-ranking specialists. Through comparisons with the use of advanced learning mechanism such as support vector machines and decision trees, the systematic experimental results demonstrate that the proposed fused-link based approach helps to capture and reflect the impact of academic journals while being more interpretable.

The proposed work is very promising. However, it also opens up an avenue for significant further investigation. For instance, it may be useful to establish a more flexible way to control the aggregation of indicator scores other than the employment of distance metrics in an effort to further strengthen the ability of creating interpretable outcomes. It is also interesting to use linguistic terms rather than precise scores to evaluate a journal's performance on each individual impact indicator as they may be more intuitive and closer to experts' cognitive concept of journal quality. Chapters 4 and 6 will focus on these issues, respectively.

Chapter 4

OWA-based Aggregation of Fuzzy Relations for Journal Ranking

AGGREGATION of several input values into a single output value is an indispensable tool in a wide range of applications such as human resource management [37], group decision making [207], industrial problems [223], etc. Different types of aggregation operator have been proposed in the literature. A popular aggregation method is the Ordered Weighted Averaging (OWA) operator originally introduced in [209]. It provides a parameterised family of aggregation operators, including as special cases the maximum, the minimum and the average calculus [35].

Academic journal ranking is a specific application problem addressed here in which OWA may also play a significant role. Instead of direct aggregation of the individual scores, Chapter 3 introduced another direction which employs the distance metrics over journals that are placed in a multi-dimensional space with each dimension representing a certain impact indicator. In this chapter, OWA operators are employed to aggregate fuzzy similarities between journals in terms of their impact indicators, thereby generating clusters of journals that reflect their individual indicator scores. For each indicator, a fuzzy similarity relation amongst journals is generated, the OWA operator is then employed to aggregate the resulting similarity relations. The OWA-aggregated relation is applied to support the classic k -means clustering algorithm in order to generate clusters of journals according to their indicator scores.

As pointed out in [136]: “the solution (of journal assessments) appears to be in a combination of peer review and objective indicators. These indicators should be

assessed for relevance and reliability”. It is therefore useful to enhance the reliability in performing aggregation of publication impact indicators for the task of academic journal ranking. Inspired by this observation, the present chapter also proposes a nearest neighbour guided induced OWA operator, denoted as k NN-IOWA hereafter, for developing aggregated fuzzy relations between journals, based on their impact indicator scores. The proposed operator is a special case of the Induced OWA (IOWA) [12, 134, 213], with two characters that distinguish it from other IOWA operators: 1) the elements of the order inducing vector represent the relative reliabilities of the associated arguments, and 2) the value of the reliability measure depends on the distribution of the arguments. That is, for each individual argument, its average distance to the other k nearest arguments is calculated and transformed into its corresponding element in the order inducing vector.

The remainder of this chapter is structured as follows. Section 4.1 introduces the concepts of fuzzy similarity relations and the basics of the OWA and IOWA aggregation operators. Section 4.2 describes the definition of OWA-aggregated fuzzy relations among journals and its application to k -means clustering for journal ranking. The experimental evaluation of the proposed approach, along with a discussion of the results is also provided in this section. Sections 4.3 defines k NN-IOWA and presents the experimental analysis of its application to k -means clustering for journal ranking. Finally, Section 4.4 summarises this chapter.

4.1 Preliminaries

The modelling of imprecise and qualitative knowledge, as well as the handling of uncertainty at various stages are possible through the use of fuzzy sets [216]. Traditional crisp sets allow only full membership or no membership at all. As an extension of the classic set theory, fuzzy sets allow partial membership. In other words, membership or non-membership of an element to a crisp set is described by a characteristic function in the binary pair $\{0, 1\}$, while that to a fuzzy set is characterised by an membership function that takes values in the interval $[0, 1]$. In this case, a given element can be a member of more than one fuzzy set. Fuzzy logic, which is based on the theory of fuzzy sets is capable of supporting humans in a wide range of applications [114]. This section introduces several key notions from the fields of fuzzy sets and fuzzy logic.

4.1.1 Fuzzy Relation

The concept of similarity is a basic concept in human cognition. Similarity plays an essential role in taxonomy, recognition, case-based reasoning and many other fields. Particularly, fuzzy relations [217] and their properties play an important role in both theoretical development and practical application of constructing similarity metrics.

Definition 3. Let X be a nonempty universe. A fuzzy relation $R: X \times X \rightarrow [0, 1]$ is

- reflexive iff $\forall a \in X, R(a, a) = 1$;
- symmetric iff $\forall a, b \in X, R(a, b) = R(b, a)$;
- T -transitive iff $\forall a, b, c \in X, R(a, b) \geq T(R(a, c), R(c, b))$.

Here, T is a T -norm [168] e.g., a mapping $T(x, y) : [0, 1] \times [0, 1] \rightarrow [0, 1]$ which satisfies

- commutativity: $T(x, y) = T(y, x)$;
- monotonicity: $T(x, y) \leq T(x', y')$, if $x \leq x'$ and $y \leq y'$;
- associativity: $T(x, T(y, z)) = T(T(x, y), z)$; and
- the boundary condition $T(x, 1) = x$.

A number of T -norms are proposed in the literature, the common ones include:

- the minimum T -norm: $T_{\min}(x, y) = \min(x, y)$,
- the product T -norm: $T_p(x, y) = x \cdot y$, and
- the Łukasiewicz's T -norm: $T_L(x, y) = \max(x + y - 1, 0)$.

If T is the min operator, the above definition coincides with the definition of similarity relations in [217]. There exist many different definitions of similarity metrics which have been employed with success for different purposes such as cluster analysis, classification, recognition and diagnostics and most of the similarity metrics are defined by using T -norms. However, formulation of a valid, general-purpose

definition of similarity is a challenging problem [7]. In this chapter, the proposed aggregation methods focus on the use of fuzzy relations as similarity metrics in clustering academic journals according to their indicator scores.

Given a set of journals $J = \{j_1, \dots, j_{N'}\}$ and a journal impact indicator $I : J \rightarrow \mathbb{R}$, many functions proposed in the literature can be employed to perform pairwise comparison of journal indicator scores into fuzzy similarity relations. For example:

$$R_I(j_x, j_y) = 1 - \frac{|I(j_x) - I(j_y)|}{\max\{I(j_i) | j_i \in J\} - \min\{I(j_i) | j_i \in J\}}, \quad (4.1)$$

$$R_I(j_x, j_y) = \exp\left(-\frac{(I(j_x) - I(j_y))^2}{2\delta^2}\right), \quad (4.2)$$

$$R_I(j_x, j_y) = \max\left(\min\left(\frac{I(j_y) - (I(j_x) - \delta)}{I(j_x) - (I(j_x) - \delta)}, \frac{(I(j_x) + \delta) - I(j_y)}{(I(j_x) + \delta) - I(j_x)}\right), 0\right). \quad (4.3)$$

where $I(j_x)$ and $I(j_y)$ denote the scores of journal $j_x, j_y \in J$ respectively, and δ^2 is the variance of the scores $\{I(j_x) | j_x \in J\}$. Other definitions for implementing the similarity between objects can be found in [166, 195].

The matrix representation of a fuzzy similarity relation R is called a fuzzy pairwise similarity matrix, denoted as $M(R)$ with elements in $[1, 0]$ and “1” is on the diagonal if such fuzzy relation functions satisfy reflexivity. A 3-D representation of the three examples of fuzzy similarity relations is shown in Figure 4.1. For example, given a set of four journals $\{j_1, j_2, j_3, j_4\}$ whose scores under an indicator I are $\{1.0, 2.0, 3.0, 4.0\}$, and the fuzzy similarity relation R_I between them is evaluated by Equation (4.1), then the resulting fuzzy similarity matrix is:

$$M(R_I) = \begin{pmatrix} 1.0 & 0.66 & 0.33 & 0.0 \\ 0.66 & 1.0 & 0.66 & 0.33 \\ 0.33 & 0.66 & 1.0 & 0.66 \\ 0.0 & 0.33 & 0.66 & 1.0 \end{pmatrix}$$

where each entry m_{xy} in $M(R_I)$ indicates the membership of a journal pair (j_x, j_y) to the fuzzy relation R_I , for instance, m_{12} represents that journals j_1 and j_2 are deemed to be of a similarity degree of 0.66.

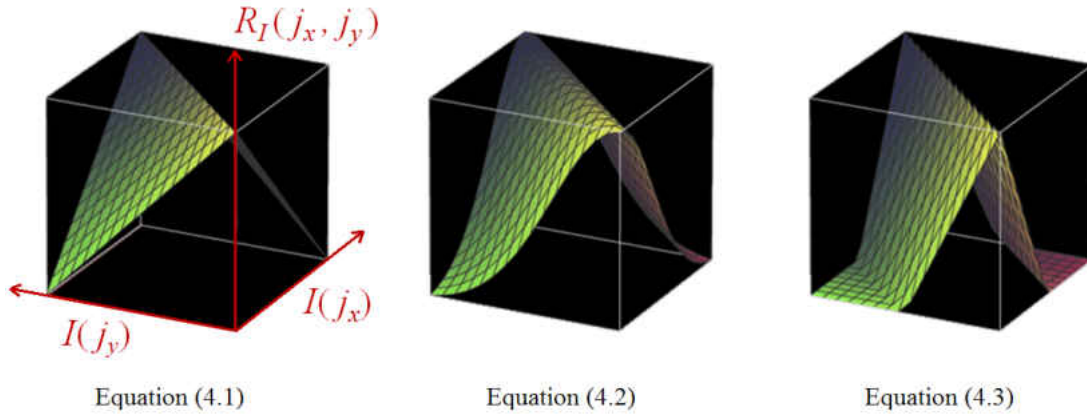


Figure 4.1: Examples of Similarity Relation Function

4.1.2 Ordered Weighted Averaging (OWA) Aggregation

When dealing with real-world problems, the opinions of different experts are usually aggregated in order to provide more accurate and reliable solutions. Similarly, numeric measures of certain properties are also typically aggregated when addressing a given problem, with the weighted average being popularly used. Apart from the classical aggregation operators (such as average, maximum and minimum), another interesting and more general type of aggregation operator is the Ordered Weighted Averaging (OWA) operator [131, 163, 209]. OWA is a family of aggregation operators which are parameterised based on the ordering of the inputs. The fundamental aspect of this family of operators is the reordering step in which the inputs are rearranged in descending order and then integrated into a single aggregated value. Formally, a mapping $A_{\text{owa}} : \mathbb{R}^m \rightarrow \mathbb{R}$ is called an OWA operator if:

$$A_{\text{owa}}(a_1, \dots, a_m) = \sum_{i=1}^m w_i a_{\pi(i)} \quad (4.4)$$

where $a_{\pi(i)}$ is a permutation of $a_i \in \mathbb{R}, i = 1, \dots, m$, which satisfies $a_{\pi(i)}$ is the i -th largest of the a_1, \dots, a_m , and $w_i \in [0, 1], i = 1, \dots, m$ is a collection of weights that satisfies $\sum_{i=1}^m w_i = 1$.

For presentational simplicity, the weights of an OWA operator are hereafter denoted as a weighting vector $W = (w_1, \dots, w_m)$, in which the i -th component is w_i . Different choices of the weighting vector W can lead to different aggregation results. The ordering of inputs gives OWA the normally nonlinear feature. Three special

cases of the OWA operator are the classical *mean*, *max* and *min*. The *mean* operator results by setting $w_i = 1/m$, the *max* by $w_1 = 1$ and $w_i = 0$ for $i \neq 1$, and the *min* by $w_m = 1$ and $w_i = 0$ for $i \neq m$. These weighting vectors are denoted as W_{mean} , W_{max} and W_{min} respectively in the remainder of the thesis. Obviously, an important feature of the OWA operator is that it is a mean operator which satisfies:

$$\min\{a_1, \dots, a_m\} \leq \sum_{i=1}^m w_i a_{\pi(i)} \leq \max\{a_1, \dots, a_m\}. \quad (4.5)$$

Such an operator provides aggregation between the maximum and the minimum of the arguments. This boundedness implies that it is idempotent; that is, if all $a_i = a, a \in \mathbb{R}$, then $A_{\text{owa}}(a_1, \dots, a_m) = a$.

Different weighting vectors can be used to express the different aggregation behaviours of OWA operator. A measure which is used to interpret the overall behaviour of an OWA operator is the Attitudinal Character (A-C) [211] (usually called as the measure of “orness” [57]). It gives an idea of whether an aggregation operator behaves similarly to conjunction (influenced by smaller inputs) or disjunction (influenced by larger inputs). In particular, the orness measure (A-C measure) of an OWA operator with the weighting vector W is defined by:

$$\text{orness}(W) = \text{A-C}(W) = \frac{1}{m-1} \sum_{i=1}^m ((m-i)w_i). \quad (4.6)$$

A higher orness value indicates the aggregation is more similar to disjunction. It can be calculated that $\text{orness}(W_{\text{mean}}) = 0.5$, $\text{orness}(W_{\text{max}}) = 1$ and $\text{orness}(W_{\text{min}}) = 0$. It has been proven that if an OWA weighting vector satisfies the buoyancy property [13], $w_i \geq w_j$ for $i < j$, ($i, j = 1, \dots, m$), the related OWA operator displays the properties of a norm, and hence it can be used to provide distance measures [212]. The generated distance measure has been applied to solve group decision making [208] and semi-supervised clustering [13].

A useful method for generating the OWA weights is by the use of a so-called stress function [211], enabling formal characterisation of the resulting OWA aggregation operator. This can be accomplished using a function $h : [0, 1] \rightarrow \mathbb{R}^+$ to stress the places where to obtain significant values for the weighting vector. Formally, a weighting vector of OWA is defined by a stress function h as follows.

Definition 4. [211] Let $h : [0, 1] \rightarrow \mathbb{R}^+$ be a non-negative function on the unit interval. The OWA weights $W = (w_1, \dots, w_i, \dots, w_m)$ can be defined as:

$$w_i = \frac{h(\frac{i}{m})}{\sum_{j=1}^m h(\frac{j}{m})}. \quad (4.7)$$

This method of obtaining the OWA weighting vector has a number of useful features. For instance, the $h(x)$ values associated with the lower portion of the left side of $[0, 1]$ reflect those weights associated with the larger argument values, while the values associated with the right side of the unit interval reflect the weights associated with the smaller values in the aggregation. Other properties are omitted here but can be found in [211].

4.1.3 Induced OWA

A key step of OWA aggregation is the ordering of the arguments which transforms the original argument vector $(a_1, \dots, a_i, \dots, a_m)$ into an ordered argument vector $(a_{\pi(1)}, \dots, a_{\pi(i)}, \dots, a_{\pi(m)})$. The ordering used in OWA depends upon the actual value of the arguments as $a_{\pi(i)}$ is the i -th largest of the arguments. A more general strategy towards the ordering of the arguments has been proposed in [213]. This has led to the development of a generalised approach to OWA aggregation, termed the Induced OWA (IOWA). In IOWA, each of the input values is represented as a two-tuple $\langle u_i, a_i \rangle$ that is referred to as an OWA pair. The input arguments $(a_1, \dots, a_i, \dots, a_m)$ are ordered on the basis of the values u_i . In particular, the procedure for calculating the IOWA aggregation over these OWA pairs is defined by:

$$A_{\text{iowa}}(\langle u_1, a_1 \rangle, \dots, \langle u_m, a_m \rangle) = \sum_{i=1}^m w_i a_{\pi'(i)}$$

where $a_{\pi'(i)}$ is from the permutation of $\langle u_i, a_i \rangle$ which satisfies that $\langle u_{\pi'(i)}, a_{\pi'(i)} \rangle$ has the i -th largest amongst all u_i , and $w_i \in [0, 1], i = 1, \dots, m$ is a collection of weights which satisfies that $\sum_{i=1}^m w_i = 1$. $U = (u_1, \dots, u_m)$ is called the order inducing vector. The bounding property exhibited by IOWA aggregation is similar to that by OWA: $\min\{a_1, \dots, a_m\} \leq A_{\text{iowa}}(\langle u_1, a_1 \rangle, \dots, \langle u_m, a_m \rangle) \leq \max\{a_1, \dots, a_m\}$. Idempotency also holds in IOWA: If all $a_i = a, a \in \mathbb{R}$, then $A_{\text{iowa}}(\langle u_1, a_1 \rangle, \dots, \langle u_m, a_m \rangle) = a$, no matter which order inducing vector U and weighting vector W are used. Note

that if two or more OWA pairs have identical values of u_i , their argument values are averaged firstly before aggregation.

The introduction of inducing vector helps improve the flexibility of the ordering process in OWA aggregation. OWA operators can be rephrased as special cases of IOWA operators where $u_i = a_i$ for all $i = 1, \dots, m$. In IOWA, different order inducing vectors can lead to different results of aggregation. Hence, the interpretation of orness of the weighing vectors is also dependent on the choice of a given order-inducing vector.

4.1.4 OWA Aggregation with Dependent Weights

When combining multiple arguments using pre-defined weighting vectors in OWA and IOWA, the weights in aggregation are normally argument-independent as they are not necessarily related to the inputs they are applied to. In such cases, unduly high or low weights might be given by false or biased judgments and hence, a typical OWA operator would suffer drastically from giving the highest priority to either the highest or the lowest value [26]. To achieve more reliable outcomes, a type of OWA operators with dependent weights have been introduced in the literature, in which the normal-distribution of argument values is used to determine the weight vector. This type of OWA operators considers a strong intuitive appeal for deriving the weighting vectors of aggregation to the concept of data reliability, aiming to decrease the effect of potential outliers in input arguments.

In particular, the Dependent OWA (DOWA) operators [206] utilise weighting vectors that are derived in accordance with the average of arguments. Let (a_1, \dots, a_m) be the argument vector, and μ be the average value of this argument set, where $\mu = \frac{1}{m} \sum_{i=1}^m a_i$. The similarity between any argument a_i and the average value μ can be calculated as follows:

$$s(a_i, \mu) = 1 - \frac{|a_i - \mu|}{\sum_{j=1}^m |a_j - \mu|}. \quad (4.8)$$

From this, a weighing vector can be generated by applying the following:

$$w_i = \frac{s(a_i, \mu)}{\sum_{j=1}^m s(a_j, \mu)} \quad (4.9)$$

$$A_{\text{dowa}}(a_1, \dots, a_m) = \sum_{i=1}^m w_i a_i \quad (4.10)$$

Besides measuring reliability of arguments by their distances to the average value, there are other alternatives. In k NN-DOWA [27] for example, the reliability of an argument is based on its nearest neighbours. This type of reliability helps differentiate amongst a collection of arguments such that an argument whose value is similar to its k neighbours [10] is deemed reliable and can be assigned with a higher weight. In contrast, an argument that is largely different from its neighbours is discriminated as an unreliable member. Formally, the reliability measure of an argument a_i , $i = 1, \dots, m$ in k NN-DOWA is defined as:

$$r^k(a_i) = 1 - \frac{\sum_{t=1}^k |a_i - n_t^{a_i}|}{\max_{j, j' \in \{1, \dots, m\}} |a_j - a_{j'}|} \quad (4.11)$$

where $n_t^{a_i}$, $t = 1, \dots, k$ is the t -th nearest neighbour of the argument a_i , and the distance measure d used to perform neighbour-searching is $d(a_j, a_{j'}) = |a_j - a_{j'}|$, $j, j' = 1, \dots, m$. Note that other distance metrics may be used for this. However, for computational simplicity, the absolute distance metric is used here.

Having obtained the reliability values of all arguments concerned, they are normalised to form the weighing vectors in k NN-DOWA. Given the reliability value $r^k(a_i)$ of each argument $a_i, i = 1, \dots, m$, the corresponding k NN-DOWA operator $A_{\text{dowa}}^k : \mathbb{R}^m \rightarrow \mathbb{R}$ can be specified by:

$$A_{\text{dowa}}^k(a_1, \dots, a_m) = \sum_{i=1}^m w_i^k a_i \quad (4.12)$$

where $w_i^k = r^k(a_i) / \sum_{j=1}^m r^k(a_j)$. k NN-DOWA and DOWA are order independent (termed *neat* in the literature) [210], as they generate the same outcome regardless of the order of argument values. k NN-DOWA has been applied to the task of alien detection, where different similarity measures of textual entities are combined. Similar to k NN-DOWA, the Cluster-DOWA forms cluster of arguments to detect outliers [26]. One crucial assumption in all these methods is that arguments which have high reliability values should be highly weighted.

4.2 OWA Aggregation of Fuzzy Relations for Journal Ranking

Whilst the application of fused-link to cluster journals is promising, there is much further research to be carried out in making these techniques more flexible in order to support activities in journal quality assessment. Direct use of distance metrics makes it difficult for human experts to choose appropriate similarity measures in order to compare journals. It may be impossible to decide the degree of “orness” in the aggregation. To compensate for the potential bias of using fixed and equaled weights on indicators, as well as enriching the flexibility of measuring similarity relations amongst journals, an OWA-aggregated fuzzy relation is proposed to integrate fuzzy similarity measures for the assessment of journals.

4.2.1 OWA Aggregation of Fuzzy Relations

Formally, given a set of journals $J = \{j_1, \dots, j_N\}$, and a set of journal impact indicators $I = \{I_1, \dots, I_m\}$, the fuzzy similarity between two journals $j_x, j_y \in J$ with respect to the indicator $I_i, i = 1, \dots, m$ is represented by $R_{I_i}(j_x, j_y)$, and the OWA-aggregated fuzzy relation between j_x and j_y is defined by:

$$R_{\text{owa}}(j_x, j_y) = A_{\text{owa}}(R_{I_1}(j_x, j_y), \dots, R_{I_m}(j_x, j_y)) = \sum_{i=1}^m w_i R_{I_i}(j_x, j_y) \quad (4.13)$$

where w_i is the weighting vector which satisfies $\sum_{i=1}^m w_i = 1$, and $R_i(j_x, j_y)$ is the i -th largest in $\{R_{I_1}(j_x, j_y), \dots, R_{I_m}(j_x, j_y)\}$. The transformation from single indicator scores $I_i(j_x)$ and $I_i(j_y)$ to the similarity relation $R_{I_i}(j_x, j_y)$ can be achieved using either of Equations (4.1)–(4.3).

In so doing, the OWA-aggregated relation R_{owa} implements a mapping from multiple similarity relations to one relation: $R^m \rightarrow R$. Obviously, the aggregated relation also preserves the reflexivity and symmetry if all R_{I_1}, \dots, R_{I_m} are fuzzy similarities. The aggregated relation does not always display T -transitivity however. One special case is when the weight vector $W_{\min} = (0, \dots, 0, 1)$, where the aggregation operator can be formed as $R_{\text{owa}}(j_x, j_y) = \min\{R_{I_1}(j_x, j_y), \dots, R_{I_m}(j_x, j_y)\}$, which has been shown to preserve T -transitivity [191]. For problems where it is important to preserve T -transitivity, fuzzy transitive closure must be computed [48].

Table 4.1: Examples of Journals: OWA

	I_1	I_2	I_3	I_4
j_1	7.806	10.716	0.00571	0.867
j_2	5.027	7.228	0.05002	0.591
j_3	2.683	3.752	0.00895	0.387
j_4	2.674	3.255	0.01409	0.299
j_5	1.574	1.454	0.00219	0.287
δ	2.45342	3.68844	0.02028	0.24523

To illustrate the concept of OWA-aggregated fuzzy relations, assume that five journals are individually evaluated using four separate indicators: JIF (I_1), 5-year JIF (I_2), Eigenfactor (I_3) and Immediacy Index (I_4), as listed in Table 4.1. The fuzzy similarity relation with respect to each indicator is evaluated by the use of Equation (4.2). The resulting fuzzy similarity relations are shown in Equations (4.14)–(4.17). Suppose that the weighting vector in R_{owa} is $W = (0.1, 0.2, 0.3, 0.4)$, the aggregated fuzzy relation amongst these journals is shown in Equation (4.18). Each entry in the relation indicates the degree of aggregated relation between the corresponding pair of journals, with respect to the four given impact indicators in a joint manner. For instance, the impact of j_4 is more similar to that of j_3 as compared to those of j_2 and j_1 .

$$M(R_{I_1}) = \begin{pmatrix} 1.0 & 0.526 & 0.113 & 0.112 & 0.048 \\ 0.526 & 1.0 & 0.634 & 0.631 & 0.411 \\ 0.113 & 0.634 & 1.0 & 1.0 & 0.931 \\ 0.112 & 0.631 & 1.0 & 1.0 & 0.932 \\ 0.048 & 0.411 & 0.931 & 0.932 & 1.0 \end{pmatrix} \quad (4.14)$$

$$M(R_{I_2}) = \begin{pmatrix} 1.0 & 0.639 & 0.168 & 0.129 & 0.043 \\ 0.639 & 1.0 & 0.641 & 0.560 & 0.294 \\ 0.168 & 0.641 & 1.0 & 0.991 & 0.824 \\ 0.129 & 0.560 & 0.991 & 1.0 & 0.888 \\ 0.043 & 0.294 & 0.824 & 0.888 & 1.0 \end{pmatrix} \quad (4.15)$$

$$M(R_{I_3}) = \begin{pmatrix} 1.0 & 0.074 & 0.987 & 0.918 & 0.985 \\ 0.074 & 1.0 & 0.105 & 0.174 & 0.049 \\ 0.987 & 0.105 & 1.0 & 0.968 & 0.946 \\ 0.918 & 0.174 & 0.968 & 1.0 & 0.842 \\ 0.985 & 0.049 & 0.946 & 0.842 & 1.0 \end{pmatrix} \quad (4.16)$$

$$M(R_{I_4}) = \begin{pmatrix} 1.0 & 0.531 & 0.147 & 0.068 & 0.061 \\ 0.531 & 1.0 & 0.708 & 0.492 & 0.464 \\ 0.147 & 0.708 & 1.0 & 0.936 & 0.920 \\ 0.068 & 0.492 & 0.936 & 1.0 & 0.999 \\ 0.061 & 0.464 & 0.920 & 0.999 & 1.0 \end{pmatrix} \quad (4.17)$$

$$M(R_{owa}) = \begin{pmatrix} 1.0 & 0.528 & 0.486 & 0.435 & 0.426 \\ 0.528 & 1.0 & 0.613 & 0.536 & 0.372 \\ 0.486 & 0.613 & 1.0 & 0.985 & 0.924 \\ 0.435 & 0.536 & 0.985 & 1.0 & 0.941 \\ 0.426 & 0.372 & 0.924 & 0.941 & 1.0 \end{pmatrix} \quad (4.18)$$

Similar to the fused-link, the OWA-aggregated fuzzy relation can also be applied to the task of journal ranking by combining with existing distance/similarity based algorithms. If a-priori ranks of certain journals that are of interest are acknowledged, supervised learning techniques can be used to model the relation between indicators and labels or to classify the rank of other unlabelled journals. However, in many applications, expert assessors may prefer to rank journals in an objective way where only the impact indicators are considered, rather than any a-priori published ranks. In this case, unsupervised learning techniques are very useful for detecting groups of journals which may have similar impact measures. In this section, the aggregated fuzzy relations are employed to aid in assessing the clustering of different journals with respect to different indicators, with an aim to partition a collection of academic journals. Particularly, the k -mean clustering algorithm is employed, in which the similarity measures between any two given journals are evaluated by the proposed OWA-aggregated fuzzy relations.

The integrated approach is shown in Figure 4.2. The indicator scores of journals are firstly collected from APDs such as JCR, and then transformed into entries in the

fuzzy similarity relations between journals. As indicated previously, any of Equations (4.1)–(4.3) may be used in order to perform such calculation. The OWA weighting vector may be given by expert assessors or derived from stress functions [211]. Note that for k -means (and other clustering algorithms where only the similarity to clustering centroids is considered), the complete specification of the similarity relations for each pair of objects may be not necessary.

The pseudo-code for the implementation of the integrated k -means algorithm is shown in Algorithm 4.2.1. The time complexity of the classical k -means is $O(N^J kl)$ [105], where N^J is the number of journals to be clustered, k is the number of clusters, and l is the number of iterations taken by the algorithm before termination. For the calculation of OWA-aggregated fuzzy relation between two journals, an additional sorting process is required. Since the number of indicators is usually far smaller than the number of journals, a basic sorting algorithm is acceptable in terms of its complexity, and the resulting overall time complexity is $O(N^J m^2 kl)$, where m is number of indicators of interest. Usually, k , m and l are fixed in advance and so the algorithm has linear $O(N^J)$ time complexity with respect to dataset size.

k : the number of clusters;
 $J = \{j_x | x = 1, \dots, N^J\}$: a set of journals evaluated by m indicators;
 $C = \{c_i | i = 1, \dots, k\}$: k centroids of the k clusters;
 $\mu(j_x) \in \{1, \dots, k\}$: the cluster label of $j_x, j_x \in J$.

- 1: choose a similarity measure for calculation of R_1, \dots, R_m
- 2: decide the weighting vector $W = (w_1, \dots, w_m)$ for R_{owa}
- 3: **for** $i = 1, \dots, k$ **do**
- 4: randomly pick c_i from J
- 5: $\mu(c_i) = i$
- 6: **end for**
- 7: **repeat**
- 8: **for** $x = 1, \dots, N^J$ **do**
- 9: $\mu(j_x) = \arg \max_{i \in \{1, \dots, k\}} R_{\text{owa}}(j_x, c_i)$
- 10: **end for**
- 11: **for** $i = 1, \dots, k$ **do**
- 12: $c_i = \text{average of } \{j_x | \mu(j_x) = i\}$
- 13: **end for**
- 14: **until** none of $c_i \in C$ changed
- 15: **return**

Algorithm 4.2.1: OWA-based k -means for Journal Clustering

4.2. OWA Aggregation of Fuzzy Relations for Journal Ranking

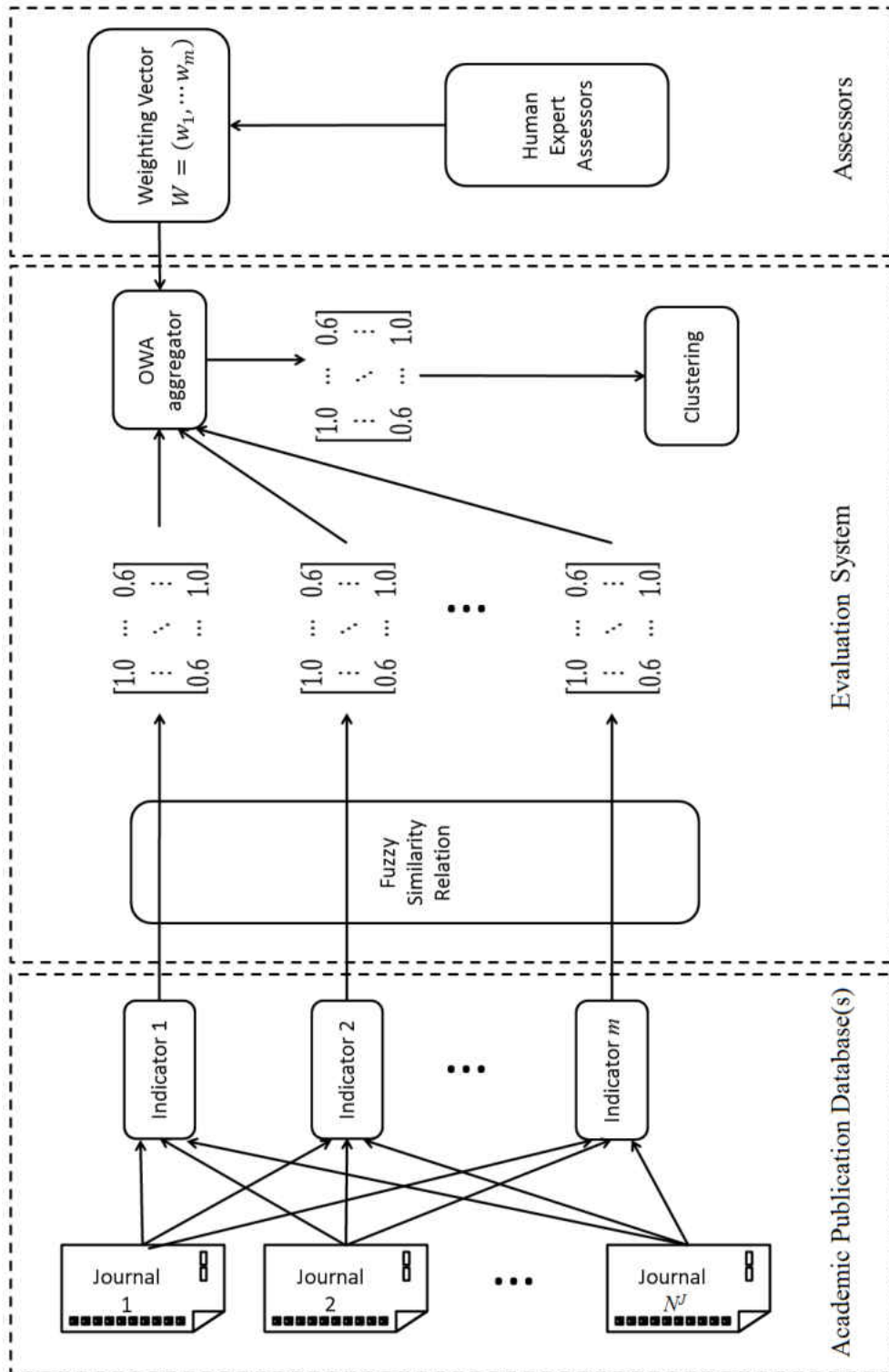


Figure 4.2: OWA-based Journal Clustering Procedure

Finally, it is worth indicating that the degree of “orness” in aggregating the journal impact indicators are easier to interpret than distance metrics, owing to the use of OWA aggregator. This is of particular significance to performing journal ranking and assessment tasks, where human experts and users tend to prefer interpretable descriptions. Also, the use of OWA aggregation of fuzzy similarity relations between journals allows the results to reflect better intuition behind common practice in journal ranking, where multiple indicators are necessary whilst only one overall impact value is ultimately employed when judging a journal’s standing.

4.2.2 Experimentation

This section presents an experimental evaluation of the proposed work. It shows the setup of the experiments carried out and also discusses the results obtained. In order to demonstrate the journal ranking results generated using the proposed methods in a credible fashion, they are compared with human expert opinion. The Ranked Journal List (RJL) provided by the ERA 2010 has invited a group of scholars to rank a very large number of academic journals. Similar to Chapter 3, to evaluate the ranking results of proposed approach, RJL is assumed to be the ground truth when comparing different similarity measures and weighing vectors in the present experiments. The “accuracy” and “within-1 accuracy” are adopted in order to analyse the consistency between the proposed approach and RJL. The details of the construction of datasets and the evaluation criteria which are used in this experiment are described in Section 3.3. A summary of these datasets is given in Table 3.2.

All three equations listed in Equations (4.1)–(4.3) are employed to carry out clustering. For each similarity measure, five different weighting vectors for aggregating the similarities relations are tested:

$$\begin{aligned}
 W_{\min} &= (0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0), \\
 W_{\text{and}} &= (0.036, 0.071, 0.107, 0.143, 0.179, 0.214, 0.250), \\
 W_{\text{mean}} &= (1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7), \\
 W_{\text{or}} &= (0.250, 0.214, 0.179, 0.143, 0.107, 0.071, 0.036), \\
 W_{\max} &= (1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0).
 \end{aligned}$$

W_{and} and W_{or} are derived from the linear stress function and linearly decreasing stress function, respectively (as shown in Figure 4.3) [211]. Besides the aggregated fuzzy

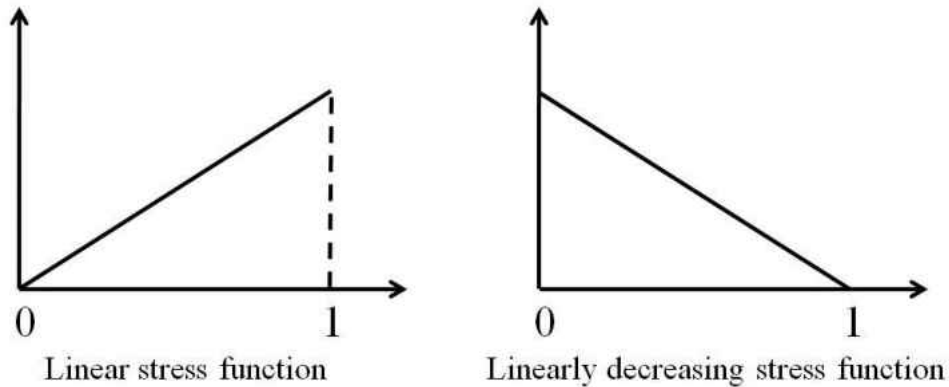


Figure 4.3: Two Linear Stress Function Used in Deriving W_{and} and W_{or}

relations, the Euclidean distance based k -means is also implemented and compared. The number of clusters is set to the number of ranks ($k = 4$). 50 times random centroid initialisation are tested. The overall averaged and best-10 averaged results [149] are reported in Tables 4.2 and 4.3 respectively.

The results show that for all three similarity measures, the weighting vector W_{or} achieved the best in terms of both accuracy and within-1 accuracy when applied to the k -means algorithm. This indicates that the weighting vector which shows an *orness* behaviour (namely, inputs with high values dominate the aggregation result) is more preferable when OWA operators are used. However, the performance of the extreme *orness* case as of W_{max} is slightly worse compared with the remainder. This implies that the general impact scores of journals are similar if they have high similarities for *more than one* indicator.

Note that W_{or} is selected for all three fuzzy similarity measures and with its use, Equation (4.3) has achieved the best average accuracy and within-1 accuracy. Compared with the other two measures, it contains a threshold to ignore insignificant similarity values between two journals, while emphasising high similarity values. Generally, for the proposed approach, the best accuracy is achieved if both the similarity measure and the weighting vector are focused on high similarity values.

In classical clustering, measures implemented with the Euclidean and Manhattan distance metrics are more commonly used. For the present journal ranking problem, the OWA-aggregated fuzzy relation has shown higher accuracy and within-1 accuracy. It is worth noting that if the fuzzy similarity for each indicator is generated using

Table 4.2: Comparison of Accuracy (%): OWA. The best performance on each dataset is highlighted in boldface

Eqn.	Equation (4.1)					Equation (4.2)					Equation (4.3)					Euclidean Distance
	W_{\min}	W_{and}	W_{mean}	W_{or}	W_{\max}	W_{\min}	W_{and}	W_{mean}	W_{or}	W_{\max}	W_{\min}	W_{and}	W_{mean}	W_{or}	W_{\max}	
AI	42.87	50.80	51.95	52.53	43.45	43.45	51.72	50.57	52.53	47.47	42.53	51.03	52.30	55.06	47.36	47.82
IS	43.26	44.42	45.23	50.12	47.67	43.02	43.72	45.47	46.51	46.63	39.53	46.40	47.67	48.95	48.95	45.58
IA	45.06	41.20	42.17	44.58	39.64	42.53	43.61	42.05	45.78	39.16	40.96	44.22	43.98	45.42	39.16	42.77
SE	44.06	47.25	48.12	46.09	41.30	42.90	45.36	44.93	47.39	40.43	42.17	46.23	47.83	46.81	40.43	45.80
TM	42.22	42.64	46.53	47.64	39.31	44.44	47.50	49.31	48.61	39.86	46.67	49.17	48.47	49.72	38.75	45.56
Avg.	43.49	45.26	46.80	48.19	42.27	43.27	46.38	46.46	48.17	42.71	42.37	47.41	48.05	49.19	42.93	45.50

Table 4.3: Comparison of Within-1 Accuracy (%): OWA. The best performance on each dataset is highlighted in boldface

Eqn.	Equation (4.1)					Equation (4.2)					Equation (4.3)					Euclidean Distance
	W_{\min}	W_{and}	W_{mean}	W_{or}	W_{\max}	W_{\min}	W_{and}	W_{mean}	W_{or}	W_{\max}	W_{\min}	W_{and}	W_{mean}	W_{or}	W_{\max}	
AI	88.51	93.68	94.02	94.14	82.53	90.57	92.76	92.76	93.56	86.09	87.13	92.64	93.45	93.68	85.63	93.10
IS	81.86	86.51	86.16	87.33	85.81	80.93	85.93	88.02	85.35	85.58	82.79	86.40	85.70	88.60	85.00	83.72
IA	81.93	81.81	83.49	83.25	79.88	82.77	84.22	85.18	84.58	79.16	83.61	84.22	84.22	84.34	78.80	83.61
SE	92.46	91.74	91.01	89.71	81.88	85.22	89.28	85.80	88.41	80.00	84.49	89.42	89.71	88.84	80.58	93.62
TM	86.67	85.83	87.50	88.89	83.89	88.19	89.44	90.42	90.83	83.61	90.97	86.11	88.61	89.17	84.31	85.83
Avg.	86.28	87.91	88.44	88.66	82.80	85.54	88.33	88.44	88.55	82.89	85.80	87.76	88.34	88.93	82.86	87.98

Equation (4.1) and the OWA weighing vector is set to W_{mean} , the resulting k -means is just identical to the Manhattan distance based k -means.

Note that journals of a certain rank may often be heavily overlapped with journals of other ranks. Therefore, the low accuracy of journal ranks using classification/clustering is not unexpected. After all, most of the journals are not obviously better or worse than others, although their ranks are more likely to be affected by the preference of the human assessors. Besides, the assumed ground truth is itself not accurate. In light of this, it may be interesting to develop an reliable means for determining the relative ranking positions of academic journals using OWA-based fuzzy relations.

4.3 Nearest Neighbour Guided Induced OWA and Journal Ranking

The weighting vectors in OWA and IOWA are normally argument-independent as they are not necessarily related to the inputs they are applied to. However, with the argument-dependent OWA, weights are indeed determined on the basis of the input arguments. Particularly, in k NN-DOWA [27] for example, the *reliability* of an argument is defined as the appropriateness of using that argument as the aggregated outcome, aiming to decrease the effect of potential outliers in input arguments. k NN-DOWA has been applied to the task of alien detection, where different similarity measures of textual entities are combined [27]. One crucial assumption in k NN-DOWA is that arguments which have high reliability values should be highly weighted. However, empirical results have shown that in certain situations, dependent weights do not always perform as expected. Besides, retaining more diversity of base members in the aggregated output is sometimes preferable [122, 51]. Inspired by these observations, and in order to generalise the dependent determination of the weighting vectors in k NN-DOWA, the k Nearest Neighbour Induced OWA (k NN-IOWA) is herein proposed.

4.3.1 k Nearest Neighbour Induced OWA

k NN-IOWA is designed to be a special case of IOWA, where each input two-tuple is $\langle r^k(a_i), a_i \rangle$ with $r^k(a_i)$ representing the reliability measure of a_i as with k NN-DOWA defined in Equation (4.11), where k is a predefined number of nearest neighbours

to be considered. Particularly, the input arguments $(a_1, \dots, a_i, \dots, a_m)$ are ordered with respect to their induced values $(r^k(a_1), \dots, r^k(a_i), \dots, r^k(a_m))$. Formally, kNN -IOWA is a mapping $A_{iowa}^k : \mathbb{R}^m \rightarrow \mathbb{R}$ and the kNN -IOWA aggregation over the given arguments is calculated as follows:

$$\begin{aligned} A_{iowa}^k(a_1, \dots, a_m) &= A_{iowa}(\langle r^k(a_1), a_1 \rangle, \dots, \langle r^k(a_m), a_m \rangle) \\ &= \sum_{i=1}^m w_i a_{\pi^k(i)} \end{aligned} \quad (4.19)$$

where $a_{\pi^k(i)}$ is from the permutation of OWA pairs $\langle r_{\pi^k(i)}, a_{\pi^k(i)} \rangle$ which satisfies that $r_{\pi^k(i)}$ has the i -th largest amongst $r^k(a_i), i = 1, \dots, m$, and $w_i \in [0, 1]$ is a collection of weights that satisfies $\sum_{i=1}^m w_i = 1$.

As a special case of IOWA, the bounding property of kNN -IOWA is similar to that of the IOWA operators: $\min\{a_1, \dots, a_m\} \leq A_{iowa}^k(a_1, \dots, a_m) \leq \max\{a_1, \dots, a_m\}$. The idempotency also holds: If all $a_i = a$ then $A_{iowa}^k(a_1, \dots, a_m) = a$. Note that if two or more arguments have an identical value of the reliability measure, their argument values are averaged before being aggregated.

Interestingly, the weights in kNN -IOWA are independent of the argument values. Any weights that satisfy $\sum_{i=1}^m w_i = 1$ can be employed in the process of aggregating the sorted arguments. This flexibility in weight determination offers a degree of freedom to control the behaviour of the resulting kNN -IOWA aggregation operator. The stress function which is designed for obtaining weights in OWA can be employed in kNN -IOWA in a similar way as with the existing work, in implementing the control of the reliability of kNN -IOWA. This work has an intuitive appeal in that high weights are associated with large reliability values. The reverse holds also; if high weights are associated with small reliability values the aggregated outcome will then be not reliable or trustworthy. In the situation where users do not have a-priori knowledge of weight settings, both the weights of high reliability (such as kNN -DOWA) and those of low reliability can be tested, then the one with better performance can be selected. To compensate for the potential bias of using single indicators, thereby enriching the reliability of fuzzy similarity relations amongst journals, kNN -IOWA is herein employed to integrate fuzzy similarity measures. This also offers a useful testbed to examine the utility of the above-proposed kNN -IOWA aggregation.

Given a set of academic journals $J = \{j_1, \dots, j_{N^j}\}$ and a journal impact indicator $I : J \rightarrow \mathbb{R}$, fuzzy similarity measures defined in Equations (4.1)–(4.3) can be employed

to perform pairwise comparison of journal indicator scores into a similarity relation $R_I : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$. More generally, given J and a set of journal impact indicators $I = \{I_1, \dots, I_m\}$, the fuzzy similarity between two journals $j_x, j_y \in J$ with respect to the indicator $I_i \in I$ is represented by $R_{I_i}(j_x, j_y)$, and the k NN-IOWA aggregation of these similarities between j_x and j_y can be computed by:

$$R_{\text{iowa}}^k(j_x, j_y) = A_{\text{iowa}}^k(R_{I_1}(j_x, j_y), \dots, R_{I_m}(j_x, j_y)) \quad (4.20)$$

where the weighting vector may be defined offline (say, by the user) or learned from historical data, and $R_{I_i}(j_x, j_y)$ are ordered with respect to their reliability values which are subsequently based on their k nearest neighbours.

To illustrate the computation process of $R_{\text{iowa}}^k(j_x, j_y)$, suppose that three journals are individually evaluated using four separate indicators: JIF (I_1), 5-year JIF (I_2), Eigenfactor (I_3) and Immediacy Index (I_4), as listed in Table 4.4. Also, without losing generality, suppose that the fuzzy similarity relation with respect to each indicator is evaluated by the use of Equation (4.1). This leads to the following similarities between journal j_1 and j_2 , which are each assigned with respect to one of the four individual indicators: $R_{I_1}(j_1, j_2) = 0.46$, $R_{I_2}(j_1, j_2) = 0.50$, $R_{I_3}(j_1, j_2) = 0$, and $R_{I_4}(j_1, j_2) = 0.43$. Suppose that 2NN-IOWA (i.e., $k = 2$) is adopted to perform aggregation. Two nearest neighbours are therefore considered in calculating the reliability of arguments. This results in the ordered argument vector of $(0.46, 0.43, 0.50, 0)$, with the corresponding order inducing vector $(0.92, 0.87, 0, 0.88)$. Given that the weighting vector in A_{iowa}^2 is $W = (0.40, 0.30, 0.20, 0.10)$, the aggregation result of the four fuzzy similarities between j_1 and j_2 is $R^2(j_1, j_2) = 0.410$. Comparatively, with the same W , the aggregated similarity between j_1 and j_2 using the original OWA operator is 0.422, which is closer to the argument of largest value (0.50, given by the 5-year JIF) rather than the argument which has the largest reliability (0.46, given by the JIF). Intuitively, in the tasks such as journal ranking, a reliable aggregated output is preferable to the aggregated output that is simply close to a single extreme argument.

It is worth indicating that the computational results from applying the proposed aggregation operator are easier to interpret than the concept of orness, owing to the use of the reliability measure for order inducing. This is of importance when performing journal ranking and assessment, as it mirrors the way that human experts

Table 4.4: Examples of Journals: k NN-IOWA

	I_1	I_2	I_3	I_4
j_1	7.806	10.716	0.00571	0.867
j_2	5.027	7.228	0.05002	0.591
j_3	2.683	3.752	0.00895	0.387

make such decisions, where multiple indicators are necessary whilst only one overall impact value (that takes into consideration of the multiple indicators) is ultimately employed when judging a journal's standing.

4.3.2 Experimentation

Generally speaking, the pairwise relations obtained by the application of k NN-IOWA can be utilised in a variety of similarity/distance-based learning algorithms. To keep consistency with Section 3.3 and Section 4.2.2, the k -means is herein integrated with k NN-IOWA to perform academic journal clustering.

In order to evaluate the performance of different aggregation operators for journal ranking, their clustering results are compared with human expert decisions as reflected in the Ranked Journal List (RJL) that is provided by the ERA 2010 [11]. Following the setup of previous experimentations, the result of RJL (2010) is assumed to be the ground truth in comparing the “accuracy” of different methods in this section. Each journal in RJL has a rank in the domain $Ranks = \{A^*, A, B, C\}$, where the label A^* indicates top journals in a certain research area, and the significance of journals decreases from it down to the label C. Each journal studied in the experiments below is therefore assigned a label also taken from this domain.

The collection of datasets is similar to that described in Section 3.3. However, the number of journals in each dataset are increased where journals from six main areas in the JCR Science Edition 2010 are selected:

- Agriculture (Agricultural Economics & Policy, Agricultural Engineering, Dairy & Animal Science, Multidisciplinary);
- Chemistry (Analytical, Applied, Inorganic & Nuclear, Medicinal, Multidisciplinary, Organic, Physical);

4.3. Nearest Neighbour Guided Induced OWA and Journal Ranking

- Computer Science (Artificial Intelligence, Cybernetics, Hardware & Architecture, Information Systems, Interdisciplinary Applications, Software Engineering, Theories & Methods);
- Materials Science (Biomaterials, Ceramics, Characterization & Testing, Coatings & Films, Composites, Multidisciplinary, Paper & Wood, Textiles);
- Medicine (General & Internal, Legal, Research & Experimental, Medical Ethics, Medical Informatics, Medical Laboratory Technology);
- Physics (Applied, Atomic, Molecular & Chemical, Condensed Matter, Fluids & Plasmas, Mathematical, Multidisciplinary, Nuclear, Particles & Fields).

Amongst them, only those journals that are both ranked in RJL and indexed by the JCR are considered as valid data objects (in order to have the ground truth to entail comparison). If a journal is missed from the JCR, then it is removed from the experimental data. A summary of the resulting datasets is shown in Table 4.5. Scores for seven indicators as reported in the JCR Science Edition 2010 are selected to generate fuzzy similarities amongst journals. These indicators are: Total Cites, JIF, 5-year JIF, Immediacy Index, Cited Half-Life; Eigenfactor and Article Influence. All these indicators are normalised to $[0, 1]$ by using Equation (3.1) before they are employed to generate similarity relations between journals. Two criteria, “accuracy” and “within-1 accuracy” are adopted in order to analyse the consistency between the proposed k NN-IOWA based clustering and RJL.

Table 4.5: Summary of Datasets: k NN-IOWA

Number of Instances	A*	A	B	C	Total
Agriculture	3	35	39	31	108
Chemistry	37	70	95	143	345
Computer Science	44	101	108	67	320
Material Science	26	61	80	61	228
Medicine	20	39	73	107	239
Physics	30	50	73	56	209

To examine the relationship between journal clustering accuracy and the reliability of the weighting vectors for k NN-IOWA (that is equivalent to the orness of OWA),

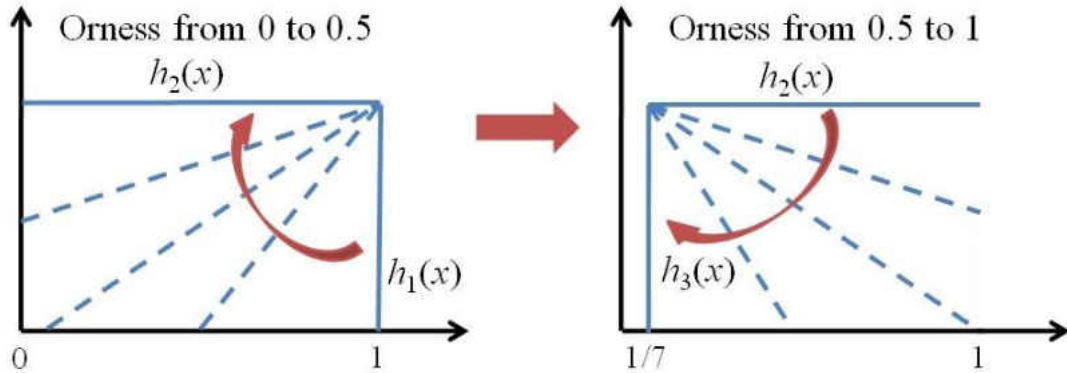


Figure 4.4: Linear Stress Functions for Weighting Vector Generation. $h_1(x)$, $h_2(x)$, and $h_3(x)$ (solid lines) are the stress functions whose corresponding weighting vectors have orness equal to 0, 0.5, and 1, respectively; other eighteen functions (dot lines) are selected to generate the weighting vectors whose orness are in the intervals (0,0.5) and (0.5,1)

twenty-one weighting vectors are generated using linear stress functions with the orness values approximately uniformly distributed from zero to one (see Figure 4.4). Figure 4.5 shows the change of accuracy (Y-axis) with respect to the orness(W) of the weighing vectors (X-axis) in both k NN-IOWA and OWA. Each point in Figure 4.5 is an averaged value of 50 random centroid initialisation, and Equation (4.1) is employed to generate the similarity between journals regarding each indicator.

To facilitate comparison, DOWA [206] and k NN-DOWA are also implemented, with their results shown in Figure 4.5 as straight dot-lines. For the alternatives of similarity measure, Equations (4.2) and (4.3) are also employed to carry out clustering to enrich the comparison. The best achieved results on the selected datasets are reported in Table 4.6.

For five of the six datasets, the accuracies achieved by the use of non-dependent aggregation operators (k NN-IOWA and OWA) generally increase with the increase of the reliability/orness of the weighting vectors. The performance of k NN-IOWA in relation to the weighting vectors of extreme reliability values is more stable than that of OWA. This indicates that the use of nearest neighbours as guidance for ordering arguments entails more reliable output in aggregation operators, which in turn allows the generation of better results in journal ranking. Figure 4.5 also shows that the

4.3. Nearest Neighbour Guided Induced OWA and Journal Ranking

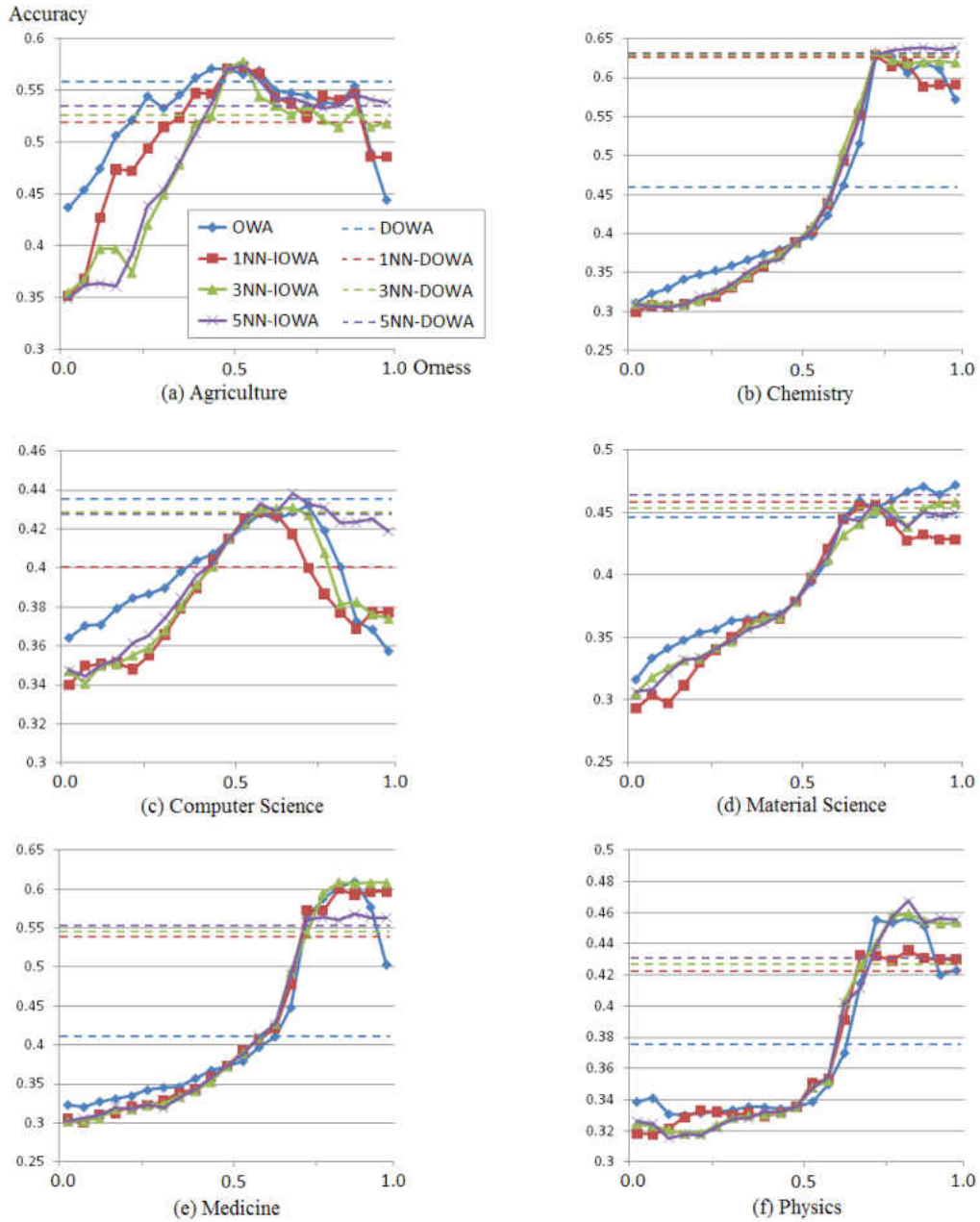


Figure 4.5: Trend of Accuracy against Reliability

Table 4.6: Comparison of Best Achieved Results: kNN-IOWA. The results are given in the format of “aggregation operator-accuracy(%)-reliability/orness”, where kNN indicates kNN-IOWA

	Eqn. (4.1)		Eqn. (4.2)		Eqn. (4.3)	
	Accuracy (%)	Within-1 (%)	Accuracy (%)	Within-1 (%)	Accuracy (%)	Within-1 (%)
Agriculture	3NN-57.93-0.527	5NN-96.22-0.527	3NN-58.12-0.516	3NN-96.88-0.833	5NN-57.60-1.000	5NN-96.78-1.000
Chemistry	5NN-63.85-0.889	5NN-98.46-0.944	5NN-64.38-0.944	5NN-98.13-0.944	1NN-64.13-0.944	5NN-97.15-0.622
Computer Science	5NN-43.84-0.667	OWA-86.89-0.667	5NN-44.47-0.423	OWA-89.74-0.833	3NN-44.41-1.000	OWA-86.84-0.333
Material Science	OWA-47.19-1.000	5NN-91.41-0.622	1NN-47.20-0.722	1NN-90.93-0.667	OWA-47.89-0.722	5NN-91.18-0.503
Medicine	OWA-60.95-0.889	OWA-92.01-0.889	OWA-61.72-0.889	OWA-91.49-0.944	1NN-58.12-0.889	OWA-88.75-0.944
Physics	5NN-46.75-0.833	3NN-91.00-0.722	5NN-45.35-0.889	1NN-94.01-0.833	1NN-47.85-0.577	1NN-93.88-0.722

performance of the k NN-IOWA is not very sensitive to the selection of k in the most of the tested dataset. Except on the Agriculture dataset, the results of $k = 1, 3, 5$ start to show difference when $\text{orness}(W) \approx 0.75$.

Note that the outcomes of using dependent weighting vectors in DOWA are not so good as those of using dependent weighting vectors in k NN-DOWA. This may be due to the fact that the k NN-based operators, including both k NN-DOWA and k NN-IOWA, are able to assign high weights to arguments which are close to the other relevant arguments, while DOWA only emphasises on the arguments close to their means. Thus, if individual journal indicators focus on rather different aspects, say the calculation of JIF and 5-year JIF includes self-citations while that of Eigenfactor and Article Influence excludes self-citations, then k NN-based methods can achieve better results than DOWA. However, the accuracy reachable by using dependent weighting vectors is not so high as that achievable by the use of carefully selected weighting vectors. This shows that although dependent methods can help aggregation operators to learn weights from arguments, human intervention for carefully choosing the weights is still necessary in situations where higher accuracies are required.

Generally, the weighting vectors which have $\text{orness}(W) > 0.5$ achieved the best results in terms of both the standard accuracy and within-1 accuracy. This indicates that the weighting vectors that exhibit a high reliability are preferable when k NN-IOWA/OWA operators are used for journal clustering. The results also show that if journals have high similarities for *more than one* indicator, the aggregated impact scores of journals may also be similar. This may be expected as there are only seven individual indicators considered.

It is interesting to note that the aggregated fuzzy relation has shown a higher accuracy and within-1 accuracy when compared with the use of Manhattan distance (which is commonly adopted in classical clustering algorithms). In fact, if the fuzzy similarity for each indicator is generated using Equation (4.1) with $\text{orness}(W) = 0.5$, the accuracies of all non-dependent aggregation operators are identical to those obtained using Manhattan distance-based k -means, which is clearly reflected by the intersections on $\text{orness}(W) = 0.5$ in Figure 4.5.

4.4 Summary

This chapter has presented two techniques for computing OWA-aggregated fuzzy relations amongst academic journals. Firstly, the OWA-aggregated fuzzy relation is

proposed and applied to the conventional k -means algorithm for clustering journals according to their impact indicators. Secondly, a nearest neighbour guided induced OWA operator: k NN-IOWA is proposed. The resultant aggregation operators have the strength of controlling the degree of orness and reliability of the aggregated output, respectively. They also have been applied for building aggregated fuzzy relations between academic journals, on the basis of the individual indicator scores of the journals concerned. Experimental results indicate that the proposed methods are more consistent with the trend in the RJL than their counterparts which employ the Euclidean and Manhattan distances.

Chapter 5

T-transitivity on Ordered Weighted Aggregation of Fuzzy Relations

AGGREGATION of several input values into a single output value is an indispensable tool not only for theoretical researches such as mathematics or physics, but for many real-world applications in engineering, economical, social, and other fields. The problem of aggregation is indeed very broad and heterogeneous, with a good number of aggregation operators having been developed, ranging from the simple arithmetic mean to more complicated fuzzy methods, including minimum/maximum, uninorm, and other alternative *T*-norm/*T*-conorm [14]. In particular, a parameterised mean-like aggregation operator, namely the ordered weighted averaging (OWA), has been introduced [209] and successfully applied in different areas [41, 133, 208]. Essentially, by selecting an appropriate weighting vector, an OWA operator helps to capture and reflect the uncertain nature of human judgments in problem-solving, generating an aggregated result that lies between the (conventional) two extremes of minimum or maximum combination of multi-featured data objects [212].

In general, relations holding amongst data points form the basis for many developments and applications of fuzzy systems. The construction of fuzzy relations and the mathematical treatment of fuzzy similarity have been studied from different points of view, with many papers stressing the close relationship between the concept of similarities and that of distance measures [22, 120]. In fact, similarity between fuzzy values can be directly measured using distance or pseudo distance functions

[28, 43]. However, in their applications to supporting multicriteria decision making [202], which forms a major challenge for practical fuzzy systems, a key question is what underlying properties of the data can be preserved in the process of constructing or aggregating similarity relations. For certain applications such as prototype-based reasoning where clusters of objects that are similar to certain prototypical samples are sought [151], properties such as reflexivity and transitivity [183] may not be necessary. Yet, there are many other situations in which it is desirable to maintain the symmetry and a degree of transitivity over the homogeneous similarity classes or granules whose members possess these properties. Also, from a practical point of view, symmetric and transitive classes or granules are easy to interpret, which generally facilitates the clustering process [166].

To enhance the aggregation of fuzzy relations with such desired properties entailed, this chapter discusses the properties of two OWA-based aggregations of fuzzy relations. It is theoretically proven that these ordered weighted aggregations of fuzzy relations allow the aggregated results to retain the T_{\min} -transitive and T_{\perp} -transitive properties respectively, if their weights are ascendant ordered. To demonstrate the effectiveness of such ordered weighted aggregation of fuzzy relations, they are applied to cluster data patterns by using hierarchical clustering. The similarities between data points are measured using the ordered weighted aggregation of component fuzzy relations held over individual features. Experimental results on several UCI datasets demonstrate that the aggregated similarities following the proposed approach produce better hierarchical clusters than the application of classic aggregators (e.g., min, max and average).

The rest of this chapter is organised as follows. Section 5.1 introduces the basic concepts of the aggregation of fuzzy relations Section 5.2 presents the two modified types of ordered weighted aggregation of similarity measures, including a discussion of their properties. The decision on the weighting vectors for the proposed OWA operators using stress functions is investigated in Section 5.3. Section 5.4 describes the application of these aggregated fuzzy relations as similarity measures in clustering problems, supported by experimental results. The chapter is summarised in Section 5.5.

5.1 Preliminaries

Let X denote a finite set, $R_i : X \times X \rightarrow [0, 1]$, $i = 1, \dots, m$ denotes m fuzzy relations (named as component relations) on X , and $w_1, \dots, w_m \in [0, 1]$ denote weights, associated to these relations. The aggregation process aims at providing a relation R summarising the data R_1, \dots, R_m together with the weights w_1, \dots, w_m . Here, the aggregated degree of fuzzy relation $R(a, b) \in [0, 1]$, $a, b \in X$ depends on the local comparisons $R_1(a, b), \dots, R_m(a, b)$ between the two patterns. The component relations usually represent the similarities of patterns from different perspectives such as opinions from different experts, multiple criteria of evaluation and different features of describing data.

Definition 5. [70] *The aggregation of the component relations R_1, \dots, R_m , with weights w_1, \dots, w_m is a relation R over X such that:*

$$R(a, b) = A(R_1(a, b), \dots, R_m(a, b), w_1, \dots, w_m) \quad (5.1)$$

where A is a mapping $[0, 1]^{2m} \rightarrow [0, 1]$, non-decreasing in the first m places, with $A(0, \dots, 0, w_1, \dots, w_m) = 0$ and $A(1, \dots, 1, w_1, \dots, w_m) = 1$.

Both the weighted and non-weighted aggregation procedures have been studied in the literature, with several of them focus on mathematical properties while others focus on the effectiveness of applications to real applications [14]. For the purposes of aggregating fuzzy relations, both the norm-conorm and sum-product operators are investigated in existing methods. Usually, the norm-conorm operators are employed to aggregate a more general type of fuzzy relations while the sum-product operators are usually applied to fuzzy relations which preserve T_L -transitivity [70, 178]. In the following, several definitions and properties about aggregations of fuzzy relations are introduced.

Definition 6. [70] *The optimistic aggregated fuzzy relation is:*

$$R_{opt}(a, b) = S_{i=1, \dots, m} T(w_i, R_i(a, b)). \quad (5.2)$$

The pessimistic aggregated fuzzy relation is:

$$R_{pess}(a, b) = T_{i=1, \dots, m} S(N(w_i), R_i(a, b)). \quad (5.3)$$

Here, S is a T -conorm and N is a strong negation. T -conorms are dual to T -norms. Given a T -norm, the complementary conorm is defined by:

$$S(x, y) = 1 - T(1 - x, 1 - y). \quad (5.4)$$

In order to have the intuitive explanation of the two aggregators, all the m weights can be assumed to be either 0 or 1. Hence, Equations (5.2) and (5.3) can be rewritten as $R_{\text{opt}}(a, b) = S_{\{i|w_i=1\}}R_i(a, b)$ and $R_{\text{pess}}(a, b) = T_{\{i|w_i=1\}}R_i(a, b)$ respectively, by which $R_{\text{opt}}(a, b)$ can be viewed as the degree of truth of the statement “there exists at least one significant criterion for which a is in relation with b ” and $R_{\text{pess}}(a, b)$ as the degree of truth of the statement “ a is in relation with b for all significant criteria” [70]. It has been proved that $R_{\text{opt}} \geq R_{\text{pess}}$. If the minimum, maximum and the standard negation $N(x) = 1 - x$ are selected as the T -norm, T -conorm and negation in definition 6 respectively, then

$$R_{\text{opt}}(a, b) = \max_{i=1, \dots, m} \min(w_i, R_i(a, b)) \quad (5.5)$$

and

$$R_{\text{pess}}(a, b) = \min_{i=1, \dots, m} \max(1 - w_i, R_i(a, b)) \quad (5.6)$$

Theorem 1. [70] *If R_1, \dots, R_m are T_{\min} -transitive fuzzy relations, and f_1, \dots, f_m are non-decreasing mappings from $[0, 1]$ into $[0, 1]$, then $R = \min_{i=1, \dots, m} f_i(R_i)$ is T_{\min} -transitive.*

It is easy to conclude from theorem 1 that if R_1, \dots, R_m are T_{\min} -transitivity, then Equation (5.6) preserves T_{\min} -transitivity.

Theorem 2. [178] *The weighted average R of m T_L -transitive fuzzy relations $R(a, b) = \sum_{i=1}^m w_i R_i(a, b)$ with $w_i \geq 0$ and $\sum_{i=1}^m w_i = 1$ is also T_L -transitive.*

Theorem 3. [165] *Let R_1, \dots, R_m be m T_L -transitive fuzzy relations then, $R = A(R_1, \dots, R_m)$ is T_L -transitive iff the De Morgan’s dual of $R = A(a)$, defined as $A'(a) = 1 - A(1 - a_1, \dots, 1 - a_m)$, satisfies the following condition: $\forall a, b, c \in [0, 1]^m | c = a + b; A'(c) \leq A'(a) + A'(b)$.*

Based on the pessimistic aggregated fuzzy relation using min-max and the sum-product weighted aggregation of T_L -transitive relations, this chapter investigates two types of ordered weighted aggregation of fuzzy relations accordingly. The definitions and properties of the two types of aggregation are introduced in the following section.

5.2 Ordered Weighted Aggregation of Fuzzy Relations

As mentioned earlier, the relationships between fuzzy relations and similarity/distance metrics have been widely studied. According to [8], aggregating operations can also be interpreted in terms of distance. The utility of these operators and their weighted counterparts for modelling compensatory aggregation situations have been proven in fuzzy decision making. More recently, [13] and [212] show that when appropriate weights are selected, an OWA operator can also perform as a distance metric, e.g., a positive mapping which satisfies identity, symmetry, and triangle inequality. In essence, OWA is a family of aggregation operators which is a special weighted average based on the ordering of the inputs. The fundamental aspect of this family of operators is the reordering step in which the inputs are rearranged in descending order and then integrated into a single aggregated value. Inspired by the OWA aggregation, two types of ordered weighted aggregation of fuzzy relations are studied in this chapter. The two types of aggregation are based on min-max operator and sum-product operators, respectively.

Definition 7. *The ordered weighted aggregation of the local relations R_1, \dots, R_m , with the weighting vector (w_1, \dots, w_m) based on min-max operator is a relation R^{\min} over X such that:*

$$R^{\min}(a, b) = \min_{i=1, \dots, m} \max(1 - w_i, R_{\pi(i)}(a, b)) \quad (5.7)$$

where $R_{\pi(i)}(a, b)$ is a permutation of $R_i(a, b)$, $i = 1, \dots, m$, which satisfies $R_{\pi(i)}(a, b)$ being the i -th largest of the $R_i(a, b)$, and $w_i \in [0, 1]$, $i = 1, \dots, m$ is a collection of weights that satisfies $\max_{i=1, \dots, m} w_i = 1$.

Definition 8. *The ordered weighted aggregation of the local relations R_1, \dots, R_m , with the weighting vector (w_1, \dots, w_m) based on sum-product operator is a relation R^{\sum} over X such that:*

$$R^{\sum}(a, b) = \sum_{i=1}^m w_i R_{\pi(i)}(a, b) \quad (5.8)$$

where $R_{\pi(i)}(a, b)$ is a permutation of $R_i(a, b)$, $i = 1, \dots, m$, which satisfies $R_{\pi(i)}(a, b)$ is the i -th largest of the $R_i(a, b)$, and $w_i \in [0, 1]$, $i = 1, \dots, m$ is a collection of weights that satisfies $\sum_{i=1}^m w_i = 1$.

In so doing, the two OWA-like aggregated relations R^{\min} and R^{L} implement two mappings from multiple similarity relations to one relation: $R^m \rightarrow R$ respectively. It worth noticing that the constraint on the weights in R^{\min} is different from R^{L} . This is due to the requirement of Definition 5 (Section 5.1) that when all the component relations R_1, \dots, R_m are zero, the constraint $\max_{i=1, \dots, m} w_i = 1$ can assure $R^{\min} = 0$. Obviously, both R^{\min} and R^{L} are “mean” operators which satisfy the boundedness:

$$\min_{i=1, \dots, m} R_i(a, b) \leq R^{\min}(a, b), R^{\text{L}}(a, b) \leq \max_{i=1, \dots, m} R_i(a, b).$$

Since these two aggregators are designed for combining fuzzy relations, the reflexivity, symmetry, and the T -transitivity are also considered here. It is easy to prove that the aggregated relations R^{\min} and R^{L} preserve the reflexivity and symmetry if R_1, \dots, R_m are reflexive and symmetric. However, the aggregated relation does not always display T -transitivity.

The discussion of transitivity and symmetry on fuzzy relations has drawn much attention, especially its effectiveness and interpretation in real applications. When a reasoning process is prototype-based, such as generating clusters of patterns that are similar to certain prototypical samples, it seems that the transitivity and symmetry properties do not have to be a necessary requirement. However, there may be other situations that require homogeneous similarities and clusters/granules whose members satisfy symmetry and transitive property. Such clusters are easy to distinguish from a practical point of view, and the knowledge extracted from one of these clusters can also be applied in the same fashion to all the others of these clusters. Therefore, symmetry and transitivity in similarity relations can be considered to be useful properties for knowledge extraction in many scenarios. Following such a desirable direction, the T_{\min} -transitivity and T_{L} -transitivity of the proposed R^{\min} and R^{L} are investigated, respectively. Although transitivity is not always preserved in the above two forms of aggregation in nature, it is proven that transitive aggregated results can also be obtained by them if their weighting vectors satisfy certain constraints.

Theorem 4. *Let R_1, \dots, R_m be T_{\min} -transitive relations, (w_1, \dots, w_m) be the weighting vector in R^{\min} such that $w_i \leq w_j$ for $i < j$, then R^{\min} is T_{\min} -transitive.*

Proof. Assume that $\pi^1(j)$, $\pi^2(j)$ and $\pi^3(j)$ are three permutations of $j = 1, \dots, m$ which satisfies $R_{\pi^1(j)}(a, b)$, $R_{\pi^2(j)}(a, c)$ and $R_{\pi^3(j)}(c, b)$ are the j -th largest value in

5.2. Ordered Weighted Aggregation of Fuzzy Relations

$R_{1,\dots,m}(a, b)$, $R_{1,\dots,m}(a, c)$ and $R_{1,\dots,m}(c, b)$, respectively. $w'_j = 1 - w_j$. For all $a, b, c \in X$, since R_1, \dots, R_m are T_{\min} -transitive, then

$$R^{\min}(a, b) = \min_{j=1,\dots,m} \max(w'_j, R_{\pi^1(j)}(a, b)) \geq \min_{j=1,\dots,m} \max(w'_j, \min(R_{\pi^1(j)}(a, c), R_{\pi^1(j)}(c, b))). \quad (5.9)$$

Since the operator \max distributes over the \min operator and the associativity of \min , the right side of Equation (5.9) equals to:

$$\min_{j=1,\dots,m} \left(\max(w'_j, R_{\pi^1(j)}(a, c)), \max(w'_j, R_{\pi^1(j)}(c, b)) \right) = \min \left(\min_{j=1,\dots,m} \max(w'_j, R_{\pi^1(j)}(a, c)), \min_{j=1,\dots,m} \max(w'_j, R_{\pi^1(j)}(c, b)) \right). \quad (5.10)$$

Given that $w_i \leq w_j \Rightarrow w'_i \geq w'_j$ and $R_{\pi^2(i)}(a, c) \geq R_{\pi^2(j)}(a, c)$ for $i < j$, then $R^{\min}(a, c) = \min_{j=1,\dots,m} \max(w'_j, R_{\pi^2(j)}(a, c))$ is equal to $\max(w'_m, R_{\pi^2(m)}(a, c))$, which is the minimum value amongst all the permutations of $R_{1,\dots,m}(a, c)$ combined with $w'_{1,\dots,j}$, $w'_i \geq w'_j$ for $i < j$. Then,

$$R^{\min}(a, c) = \min_{j=1,\dots,m} \max(w'_j, R_{\pi^2(j)}(a, c)) \leq \min_{j=1,\dots,m} \max(w'_j, R_{\pi^1(j)}(a, c)),$$

and similarly,

$$R^{\min}(c, b) = \min_{j=1,\dots,m} \max(w'_j, R_{\pi^3(j)}(c, b)) \leq \min_{j=1,\dots,m} \max(w'_j, R_{\pi^1(j)}(c, b)).$$

It can be implied that Equation (5.10) is greater or equals to

$$\min \left(\min_{j=1,\dots,m} \max(w'_j, R_{\pi^2(j)}(a, c)), \min_{j=1,\dots,m} \max(w'_j, R_{\pi^3(j)}(c, b)) \right) = \min(R^{\min}(a, c), R^{\min}(c, b)).$$

□

In fact, with the assist of Theorem 3, it can be proven that when the component relations are T_L -transitive in $R^L(a, b)$, the aggregated relation is T_L -transitive if and only if the weighting vector satisfy the additional condition that $w_i \leq w_j$ for $i < j$.

Theorem 5. *Let R_1, \dots, R_m be T_L -transitive relations, (w_1, \dots, w_m) be the weighting vector in T_L , R^L is T_L -transitive iff when $w_i \leq w_j$ for $i < j$.*

Proof. It can be concluded from Theorem 3 that R_L is T_L -transitive \iff its De Morgan's dual $N(a)$ satisfies $\forall a, b, c \in [0, 1]^m | a = b + c; N(c) \leq N(a) + N(b)$. Assume that $a' = 1 - a$, then $N(a) = 1 - \sum_{i=1}^m w_i a'_{\pi'(i)}$ where $a'_{\pi'(i)}$ is a permutation of $a'_i \in [0, 1], i = 1, \dots, m$, which satisfies $a'_{\pi'(i)}$ is the i -th largest of the a'_i . Since $a' = 1 - a$, the descent permutation of a' : $(1 - a)_{\pi'(i)}$ can be replaced by an ascent permutation of a : $a_{\pi(i)}$ where $a_{\pi(i)} \leq a_{\pi(j)}$ for $i < j$, so that $(1 - a)_{\pi'(i)} = 1 - a_{\pi(i)}$. Then:

$$\begin{aligned}
 N(a) &= 1 - \sum_{i=1}^m w_i a'_{\pi'(i)} = 1 - \sum_{i=1}^m w_i (1 - a)_{\pi'(i)} \\
 &= 1 - \sum_{i=1}^m w_i (1 - a_{\pi(i)}) \\
 &= 1 - \sum_{i=1}^m (w_i - w_i a_{\pi(i)}) \\
 &= 1 - \sum_{i=1}^m w_i + \sum_{i=1}^m w_i a_{\pi(i)} \\
 &= \sum_{i=1}^m w_i a_{\pi(i)}.
 \end{aligned}$$

Hence, the De Morgan's dual of R_L , $N(a)$ can be seen as an OWA aggregation of $a \in [0, 1]^m$ with a increasingly ordered. According to [212], an OWA aggregation is a norm (which satisfies the triangle inequality $f(a) + f(b) = f(a + b)$) if and only if the OWA weighting vector satisfy the additional condition that $w_i \geq w_j$ for $i < j$. In the case of $N(a)$, both the input arguments and the weighting vector are reversely ordered as they were in the original OWA operator, so that it satisfies $\forall a, b, c \in [0, 1]^m | c = a + b; N(c) \leq N(a) + N(b) \iff$ the weighting vector satisfy the additional condition that $w_i \leq w_j$ for $i < j$. \square

Theorem 6. Let R_1, \dots, R_m be T_L -transitive relations, (w_1, \dots, w_m) be the weighting vector in T_L such that $w_i \leq w_j$ for $i < j$, then R^L is T_L -transitive.

Proof. Assume that $\pi^1(j), \pi^2(j)$ and $\pi^3(j)$ are three permutations of $j = 1, \dots, m$ which satisfies $R_{\pi^1(j)}(a, b)$, $R_{\pi^2(j)}(a, c)$ and $R_{\pi^3(j)}(c, b)$ are the j -th largest value in $R_{1, \dots, m}(a, b)$, $R_{1, \dots, m}(a, c)$ and $R_{1, \dots, m}(c, b)$ respectively. For all $a, b, c \in X$, since

R_1, \dots, R_m are T_L -transitive, then

$$\begin{aligned}
 R^L(a, b) &= \sum_{j=1}^m w_j R_{\pi^1(j)}(a, b) \geq \\
 &\quad \sum_{j=1}^m w_j \cdot \max(R_{\pi^1(j)}(a, c) + R_{\pi^1(j)}(c, b) - 1, 0) \geq \\
 &\quad \sum_{j=1}^m \max(w_j R_{\pi^1(j)}(a, c) + w_j R_{\pi^1(j)}(c, b) - w_j, 0) \geq \\
 &\quad \max\left(\sum_{j=1}^m (w_j R_{\pi^1(j)}(a, c) + w_j R_{\pi^1(j)}(c, b) - w_j), 0\right) \geq \\
 &\quad \max\left(\sum_{j=1}^m w_j R_{\pi^1(j)}(a, c) + \sum_{j=1}^m w_j R_{\pi^1(j)}(c, b) - 1, 0\right)
 \end{aligned}$$

e.g.,

$$R^L(a, b) \geq T_L\left(\sum_{j=1}^m w_j R_{\pi^1(j)}(a, c), \sum_{j=1}^m w_j R_{\pi^1(j)}(c, b)\right) \quad (5.11)$$

Since the sum-product $\sum_{j=1}^m w_j R_{\pi(j)}(a, c)$ with $w_i \leq w_j$ for $i < j$ attains its minimal value amongst all the permutations of $R_{1, \dots, m}(a, c)$ when $\pi(j) = \pi^2(j)$ e.g., the $R_{1, \dots, m}(a, c)$ are descent ordered and $w_{j=1, \dots, m}$ are ascend ordered. Therefore, $\sum_{j=1}^m w_j R_{\pi^1(j)}(a, c) \geq \sum_{j=1}^m w_j R_{\pi^2(j)}(a, c)$ and similarly,

$$\sum_{j=1}^m w_j R_{\pi^1(j)}(c, b) \geq \sum_{j=1}^m w_j R_{\pi^3(j)}(c, b).$$

Then, the right side of Equation (5.11) is greater than or equal to

$$T_L\left(\sum_{j=1}^m w_j R_{\pi^2(j)}(a, c), \sum_{j=1}^m w_j R_{\pi^3(j)}(c, b)\right) = T_L(R^L(a, c), R^L(c, a)).$$

□

Theorem 7. Let R_1, \dots, R_m be m component fuzzy relations, (w_1, \dots, w_m) be the weighting vector in R^{\min} such that $w_i \leq w_j$ for $i < j$, then $R^{\min}(a, b) = \min_{i=1, \dots, m} R_i(a, b)$.

Proof. From the definition of R^{\min} , it can be obtained that $R_{\pi(i)}(a, b) \geq R_{\pi(j)}(a, b)$ for $i < j$, and $R_{\pi(m)}(a, b) = \min_{j=1, \dots, m} R_j(a, b)$. Assume $w'_j = 1 - w_j$ for $j = 1, \dots, m$,

then $w'_i \geq w'_j$ for $i < j$. Since $R_{\pi(i)}(a, b) \geq R_{\pi(j)}(a, b)$ and $w'_i \geq w'_j$, for $i < j$, then $\max(w'_1, R_{\pi(1)}(a, b)) \geq \dots \geq \max(w'_m, R_{\pi(m)}(a, b))$. Therefore,

$$\begin{aligned} R^{\min}(a, b) &= \min_{i=1, \dots, m} \max(1 - w_i, R_{\pi(i)}(a, b)) \\ &= \max(w'_m, R_{\pi(m)}(a, b)) \\ &= \max(1 - w_m, R_{\pi(m)}(a, b)) \end{aligned}$$

According to the definition of R^{\min} , $\max_{i=1, \dots, m} w_i = 1$ and since $w_i \leq w_j$ for $i < j$ then $w_m = 1$.

$$\begin{aligned} R^{\min}(a, b) &= \max(1 - w_m, R_{\pi(m)}(a, b)) \\ &= \max(0, R_{\pi(m)}(a, b)) \\ &= R_{\pi(m)}(a, b) = \min_{i=1, \dots, m} R_i(a, b) \end{aligned}$$

□

It is worth noticing that the purpose of adding the constraint $\max_{i=1, \dots, m} w_i = 1$ on R^{\min} is to make it satisfy the requirement of $\text{Agg}(0, \dots, 0, w_1, \dots, w_m) = 0$ in Definition 5. If the constraint is removed from the definition of R^{\min} , Theorem 4 still holds while the result of $R^{\min}(a, b)$ will not equal to $\min_{i=1, \dots, m} R_i(a, b)$, but only equals to $\max(1 - w_m, R_{\pi(m)}(a, b))$.

Note that the proposed two aggregations can also be employed to define similarity measures between two fuzzy sets. If a and b are two fuzzy sets defined on a non-empty universe X , the component fuzzy relation $R_i(a, b)$ is deemed as a component similarity of the two fuzzy sets measured by the observation of element $x_i \in X$, then the similarity between a and b can be expressed as the aggregation of component similarities over all elements in X [195] and hence, can be calculated by Definitions 7 or 8.

5.3 Weighting Vectors

A common pitfall with existing aggregation operators is the inability to provide an explanatory means by which a user can utilise to enhance the individual perception of arguments' importance. To resolve this shortcoming, the stress function has been introduced [211] as a simple mechanism for attaining interpretability in OWA

aggregations, which formalises characterisation of the resulting OWA aggregation operator. This can be accomplished using a function $h : [0, 1] \rightarrow \mathbb{R}^+$ to stress the places where to obtain significant values for the weighting vector (see Equation (4.7)). This method of obtaining the OWA weighting vector has a number of useful features. For instance, the $h(x)$ values associated with the lower portion of the left side of $[0, 1]$ reflect those weights associated with the larger argument values, while the values associated with the right side of the unit interval reflect the weights associated with the smaller values in the aggregation. Different types of stress function can be used to express the distribution of weights and hence, different aggregation behaviours. A measure which is used to interpret the overall behaviour of an aggregation operator is the Attitudinal Character (see Equation (4.6)). It gives an idea of whether an aggregation operator behaves similarly to conjunction/andness (influenced by smaller inputs) or disjunction/orness (influenced by larger inputs).

The concepts of attitudinal character and stress function can also be applied to the proposed two aggregations of fuzzy relation. Similar to other argument-independent methods, this approach is practical under the circumstances where human experience is relevant. However, since the constraint on weighting vector in $R^{\min}(a, b)$ is different to that of the original OWA aggregation, a modification of the normalisation in Equation (4.7) is needed. The resultant weighting vector of $R^{\min}(a, b)$ is defined as:

$$w_i = \frac{h(\frac{i}{m})}{\max_{j=1, \dots, m} h(\frac{j}{m})}, \quad (5.12)$$

where h is a stress function $h : [0, 1] \rightarrow \mathbb{R}^+$. Accordingly, the measure of attitudinal character for $R^{\min}(a, b)$ is normalised as:

$$A-C'(W) = \frac{A-C(W)}{\sum_{i=1}^m w_i}. \quad (5.13)$$

Since the weighting vector of R^{\min} satisfy that $\sum_{i=1}^m w_i = 1$, then $A-C'(W) = A-C(W)$ for R^{\min} .

For instance, considering $a, b, c \in X$ are three patterns, $A_{1, \dots, 4}$ are four fuzzy sets based on which the component similarity relations between the three patterns are built, with $\mu_{1, \dots, 4}(a) = (0.63, 0.94, 0.97, 0.62)$, $\mu_{1, \dots, 4}(b) = (0.01, 0.49, 0.25, 0.97)$ and $\mu_{1, \dots, 4}(c) = (0.68, 0.91, 0.62, 0.68)$, respectively. Given that for $\forall a, b \in X, a \neq b$

the T_{\min} -transitive and T_L -transitive fuzzy relations are defined by the T_{\min} and T_L norms as: $R_i^{\min}(a, b) = \min(\mu_i(a), \mu_i(b))$ and $R_i^L(a, b) = 1 - |\mu_i(a) - \mu_i(b)|$, $i = 1, \dots, 4$, respectively. The corresponding T_{\min} -transitive similarity relations are then: $R_{1,\dots,4}^{\min}(a, b) = (0.01, 0.49, 0.25, 0.62)$, $R_{1,\dots,4}^{\min}(a, c) = (0.63, 0.91, 0.62, 0.62)$, $R_{1,\dots,4}^{\min}(c, b) = (0.01, 0.49, 0.25, 0.68)$, and the T_L -transitive ones over the three examples are: $R_{1,\dots,4}^L(a, b) = (0.38, 0.55, 0.28, 0.65)$, $R_{1,\dots,4}^L(a, c) = (0.95, 0.97, 0.65, 0.94)$, $R_{1,\dots,4}^L(c, b) = (0.33, 0.58, 0.63, 0.71)$. Table 5.1 presents the resulting weights obtained from the application of stress functions, while Table 5.2 shows the aggregated results of $R_{1,\dots,4}^{\min}$ and $R_{1,\dots,4}^L$ by using Equation (5.7) and (5.8) respectively.

Table 5.1: Example of Stress Function

	$h(x) =$	Weighting Vector	A-C'(W)
W_1	1, for $x = 0$; 0, otherwise.	$R^{\min} : (1.00, 0.00, 0.00, 0.00)$ $R^L : (1.00, 0.00, 0.00, 0.00)$	1.00
W_2	$1.25 - x$	$R^{\min} : (1.00, 0.75, 0.50, 0.25)$ $R^L : (0.40, 0.30, 0.20, 0.10)$	0.63
W_3	$c, c \in (0, 1]$	$R^{\min} : (1.00, 1.00, 1.00, 1.00)$ $R^L : (0.25, 0.25, 0.25, 0.25)$	0.50
W_4	x	$R^{\min} : (0.25, 0.50, 0.75, 1.00)$ $R^L : (0.10, 0.20, 0.30, 0.40)$	0.33
W_5	0, for $x = 0$; 1, otherwise.	$R^{\min} : (0.00, 0.00, 0.00, 1.00)$ $R^L : (0.00, 0.00, 0.00, 1.00)$	0.00

Table 5.2: Aggregated Result of Examples

W	W_1	W_2	W_3	W_4	W_5
$R^{\min}(a, b)$	0.6200	0.4900	0.0100	0.0100	0.0100
$R^{\min}(a, c)$	0.9100	0.6200	0.6200	0.6200	0.6200
$R^{\min}(c, b)$	0.6800	0.4900	0.0100	0.0100	0.0100
$R^L(a, b)$	0.6500	0.5290	0.4650	0.4010	0.2800
$R^L(a, c)$	0.9700	0.9260	0.8775	0.8290	0.6500
$R^L(c, b)$	0.7100	0.6220	0.5625	0.5030	0.3300

It can be seen from the given example that generally, the results of aggregated relations are impacted by the stress functions. Higher attitudinal character values will result in higher-valued aggregated similarities. However, due to the aggregations based on min-max operators are more easily to produce discrete results than the ones based on sum-product operators, R^{\min} is not as sensitive as R^{\perp} to the change of weights. It also worth noticing that the example also demonstrates that R^{\min} dose not pre-serve T_{\min} -transitive under W_1 by validating $R^{\min}(a, b) < \min(R^{\min}(a, c), R^{\min}(c, b))$. Similarly, under W_1 and W_2 , $R^{\perp}(a, b)$ is less than $T_{\perp}(R^{\perp}(a, c), R^{\perp}(c, b))$, which means that R^{\perp} dose not preserve T_{\perp} -transitive neither.

To further demonstrate how the changing of weighting vectors has impact on the behaviours of R^{\min} and R^{\perp} aggregations respectively, an example of a two dimensional dataset is employed. The two dimensions represent the membership values of patterns to the two fuzzy sets A_1 and A_2 , respectively. The Z-axis in Figure 5.1 and 5.2 indicates the aggregated similarities of all the patterns in the space to the pattern p , with $\mu_1(p) = 0.5$ and $\mu_2(p) = 0.5$ being the memberships of p to A_1 and A_2 , respectively. It can be seen from Figure 5.1 that the Z-axis values of $p \in [0.5, 1] \times [0.5, 1]$ are equal to 0.5, which forms a flat area in that subspace. When $W = (1, 1)$ or $A-C'(W) < 0.5$, two slopes are adjacent to that flat area. When non-transitive weighting vectors ($A-C'(W) > 0.5$) are employed, two “stairs” are raised up from the bottom of the slopes, and the “heights” of those stairs are influenced by the weighting vectors. It is worth noticing that the “stairs” project non-convex contour lines in Figure 5.1, which reflects the non-transitivity of the aggregated similarity.

Since $R_i^{\perp}(a, b) = 1 - |\mu_i(a) - \mu_i(b)|$ is a distance-like component relation, the contour lines are symmetric about $(0.5, 0.5)$ in R^{\perp} aggregation (as shown in Figure 5.2). When $W = (0, 1)$, a pyramid-like surface is formed with $(0.5, 0.5)$ reaches the peak value 1 and the contour lines are squares. With the value of $A-C'(W)$ moving from 0 to 1, ridges are raising up from the four slopes of the pyramid-like surface, and the straight boundaries of those square contours reaching out accordingly. Similar to the situation of using R^{\min} aggregation, when non-transitive weighting vectors ($A-C'(W) > 0.5$) are employed, the contour lines in Figure 5.2 are non-convex (see the “star” shape in Figure 5.2 when $W = (0.8, 0.2)$).

It can be seen from the above examples that not only the T -transitivity but also the behaviours of the R^{\min} and R^{\perp} aggregation can be controlled by tuning the stress

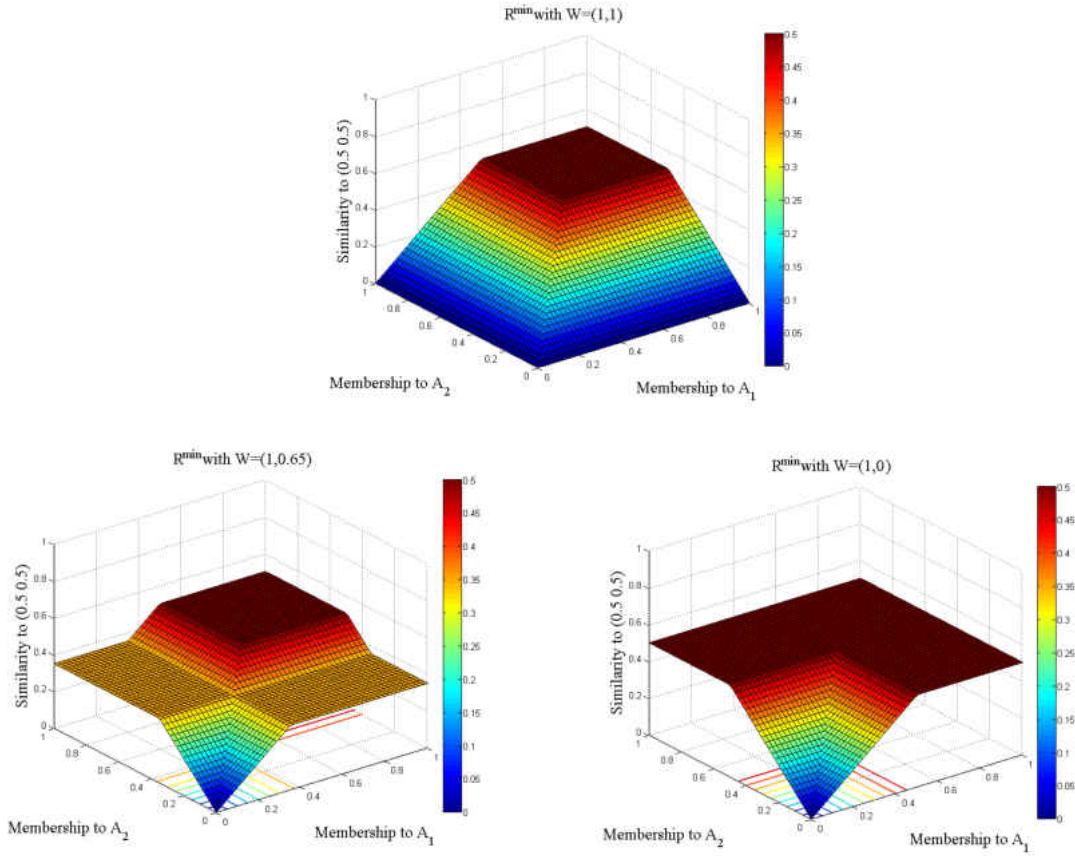


Figure 5.1: Change of R^{\min} Aggregated Relation against Weighting Vector

function and in turn, tuning the weights in the weighting vector. Inspired by the success with OWA aggregation in real applications, the proposed ordered weighted aggregations of fuzzy relations are herein applied to the problem of hierarchical clustering where different weighting vectors are tested in the proposed two aggregations and the resultant fuzzy relations are employed as similarity metrics.

5.4 Application to Cluster UCI Datasets

In order to investigate the effectiveness of proposed ordered weighted aggregation of fuzzy relations in clustering patterns as well as demonstrate their potential in solving real problems, an application of the aggregated relation to the hierarchical clustering is presented. Clustering is useful in several exploratory decision making and machine learning situations, including data mining, document retrieval, image segmentation, and pattern classification. Generally, the task of clustering is to assign

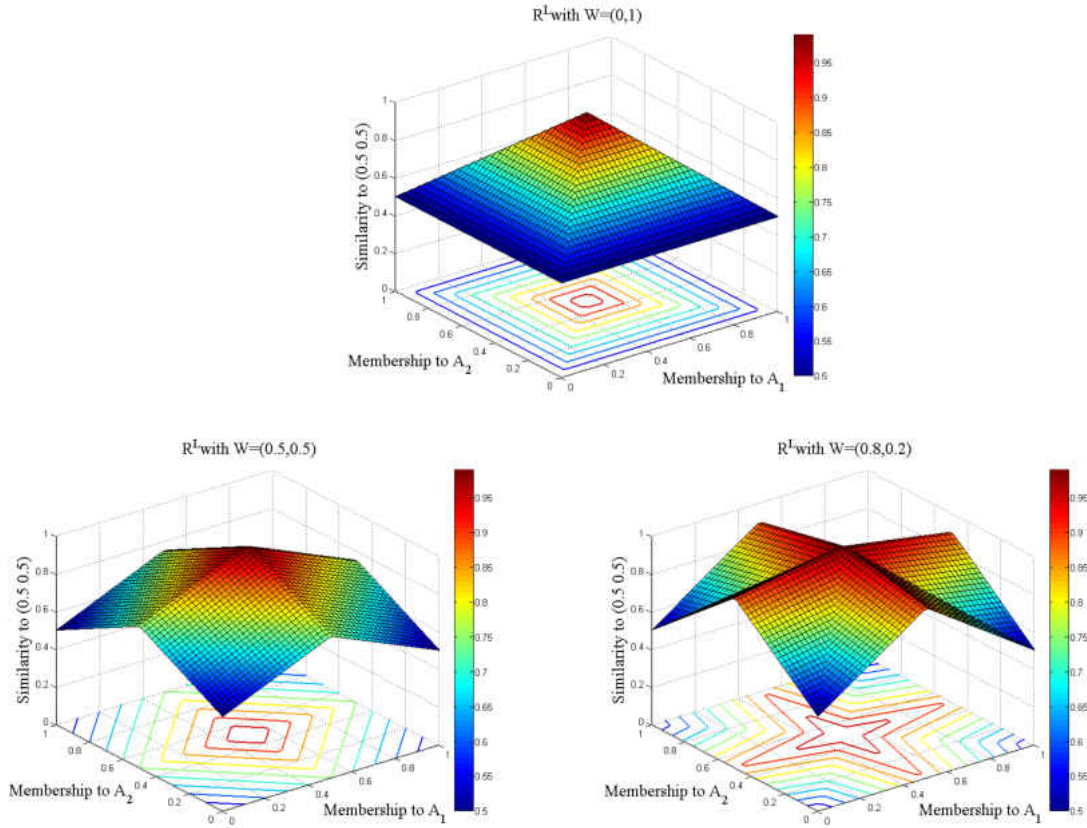


Figure 5.2: Change of R^L Aggregated Relation against Weighting Vector

a set of patterns into groups (namely clusters) such that the patterns in the same group are similar to each other, and dissimilar to those in the other clusters. Since the likeness amongst members is fundamental to the definition of a cluster, a measure of the similarity between two patterns is essential to most clustering procedures. This application uses the proposed aggregated fuzzy relations to justify the similarity of each pair of patterns and grouping similar patterns into the same cluster by using the hierarchical clustering algorithm. Its performance is assessed over a number of benchmark datasets from the UCI Machine Learning Repository [71].

5.4.1 Hierarchical Clustering

Hierarchical clustering is one of the most significant developments in clustering algorithms. In particular, hierarchical clustering builds a cluster hierarchy or a tree/dendrogram of clusters. Every cluster node contains child clusters; sibling clusters partition the points covered by their common parent. Such an approach

allows exploring data on different levels of granularity. Hierarchical clustering methods are categorised into agglomerative (bottom-up) and divisive (top-down) [19, 205]. An agglomerative clustering starts with one-point (singleton) clusters and recursively merges two or more most appropriate clusters. A divisive clustering starts with one cluster of all patterns and recursively splits the most appropriate cluster. The process continues until a stopping criterion (frequently, the requested number k of clusters) is achieved.

The reason for using the hierarchical clustering to test the proposed aggregated fuzzy relations is that any forms of similarity or distance can be applied to the hierarchical clustering directly. Consequently, the clustering results are mainly dependent on the weights employed in the proposed aggregations of similarities. Many hierarchical clustering algorithms are variants of the single-link, complete-link, and minimum-variance algorithms. Amongst these, the single-link and complete-link algorithms are commonly seen. These two algorithms differ in the way they characterise the similarity between a pair of clusters. In the single-link method, the distance between two clusters is the minimum of the distances between all pairs of patterns drawn from the two clusters (one pattern from the first cluster, the other from the second). In the complete-link algorithm, the distance between two clusters is the maximum of all pairwise distances between patterns in the two clusters. In either case, two clusters are merged to form a larger cluster based on minimum distance (maximum similarity) criteria. Both the single-link and complete-link algorithms do not contain the step of random initialisation, which makes their outputs more stable than the classic k -means. However, the complete-link algorithm produces tightly bound or compact clusters while the single-link algorithm suffers from a chaining effect [105]. Due to these reasons, the complete-link hierarchical clustering is selected herein to test the performance of the proposed aggregation of fuzzy relations in clustering analysis.

5.4.2 Aggregated Similarity based Hierarchical Clustering

A component fuzzy relation can be regarded as the similarity/dissimilarity metric of patterns based on a certain feature. The resulting aggregated relation reflects the overall estimation of similarity between patterns based on all considered features. In many clustering and classification problems, a more general type of fuzzy data other than the example in Section 5.3 is often considered. In such a kind

of fuzzy data, each feature does not represent a set of memberships to a single fuzzy set, but multiple sets of memberships with regard to several linguistic labels. These labels provide qualitative descriptions of a feature such as “very large”, “normal”, “very small”, etc. In so doing, each feature is described by a fuzzy partition rather than a single fuzzy set. Formally, given a data set of N patterns $\{p_1, \dots, p_N\}$, each pattern $p_a, a = 1, \dots, N$ is described by m features, and the i -th feature is expressed by a set of memberships $\mu^{il}, l = 1, \dots, L_i$ to L_i grades of linguistic labels (for simplicity in the notation, the number of linguistic labels L_i for all features is assumed to be L), each pattern p_a can be characterised by the following values: $p_a = ((\mu_a^{11}, \dots, \mu_a^{1L}), \dots, (\mu_a^{i1}, \dots, \mu_a^{iL}), \dots, (\mu_a^{m1}, \dots, \mu_a^{mL})), l = 1, \dots, L$. In this case, the fuzzy relation between each pair of patterns are firstly built “within” each individual feature, and then the proposed ordered weighted aggregation is employed to aggregate the similarities evaluated by different features. This is formally summarised as follows:

- Step 1. Acquire the fuzzy similarity relations $R_i(p_a, p_b)$ based on the i -th feature. According to [166], the T_{\min} -transitive and T_L -transitive similarity relations based on L linguistic labels can be obtained by the following two equations, respectively:

$$R_i^{\min}(p_a, p_b) = \begin{cases} \inf_{l \in L_{ab}^i} (\min(\mu_a^{il}, \mu_b^{il})), & \text{if } L_{ab}^i \neq \emptyset \\ 1, & \text{otherwise} \end{cases} \quad (5.14)$$

where $L_{ab}^i = \{l | \mu_a^{il} \neq \mu_b^{il}\}$.

$$R_i^L(p_a, p_b) = \inf_{l=1, \dots, L} (1 - |\mu_a^{il} - \mu_b^{il}|) \quad (5.15)$$

- Step 2. Aggregate R_i^{\min} or $R_i^L, i = 1, \dots, m$ by using R^{\min} or R^L , i.e., Equation (5.7) or Equation (5.8), respectively.
- Step 3. Apply complete-link hierarchical clustering to the dataset based on the aggregated fuzzy relation.

5.4.3 Empirical Evaluation

To evaluate the performance of proposed methods, they are experimentally tested over six datasets obtained from UCI benchmark repository [71], where true labels

of patterns are known but are not explicitly used in the clustering process. In order to easily convert the feature values to the fuzzy memberships belongs to linguistic terms, all the feature values in the selected datasets are treated as numeric values. The details of these datasets are summarised in Table 5.3.

Table 5.3: Summary of Datasets: T -transitivity

Datasets	Instances	Attributes	Classes
Iris	150	4	3
Sonar	208	60	2
Glass	214	9	6
Wine	178	13	3
Ecoli	336	7	8
Ionosphere	351	34	2

Both the normalisation and fuzzification of original datasets are implemented to entail a comparison. For each pattern $p_a, a = 1, \dots, N$, its i -th feature value $F_i(p_a) \in \mathbb{R}, i = 1, \dots, m$ is normalised to $F'_i(p_a) \in [0, 1]$ by:

$$F'_i(p_a) = \frac{F_i(p_a) - \min_{j=1, \dots, N} (F_i(p_j))}{\max_{j=1, \dots, N} (F_i(p_j)) - \min_{j=1, \dots, N} (F_i(p_j))}. \quad (5.16)$$

For the fuzzification, each normalised feature value is further transformed to five membership degrees with regard to the linguistic terms defined in Figure 5.3 [166]. It is worth noticing that the normalisation employed here can be viewed as a special case of the fuzzification in which only one linguistic term “High” is defined as $F'_i(p_a)$.

For the hierarchical clustering algorithm, the complete-link is selected to estimate the similarity between two clusters, and the number of clusters is set to the number of known classes of each dataset. The resulting clusters are evaluated in terms of accuracy since the group truth for each dataset is known. In order to examine the relationship between clustering accuracy and the attitudinal character of weighting vectors in the proposed aggregations, 21 weighting vectors are generated using linear stress functions with their attitudinal character values distributed from zero to one. There are ten weighting vectors ($A-C'(W) \in [0, 0.5]$) preserve transitivity while the

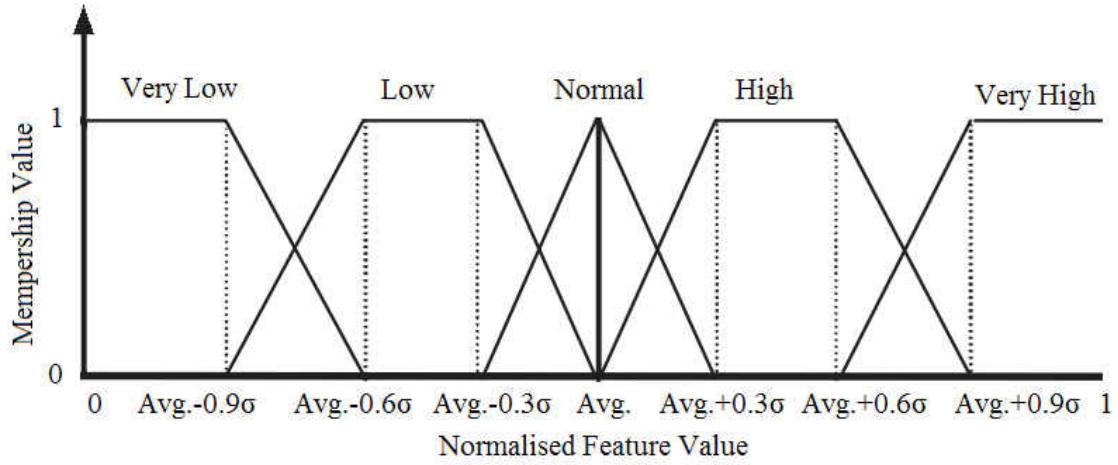


Figure 5.3: Fuzzification of Feature Value

other eleven weighting vectors ($A-C'(W) \in (0.5, 1]$) do not. The clustering process runs only one time for each weighting vector on each dataset, since the complete-link hierarchical clustering does not contain random parameters and multiple runs only produce identical results. Figure 5.4 shows the change of accuracy (Y-axis) with respect to the attitudinal character value of weighting vector (X-axis). In Figure 5.4 and the following tables, “N/F” represents normalisation/fuzzification of datasets, and “Rmin/RL” represents R^{\min}/R^L aggregations, respectively. The resultant clustering accuracy rates are also summarised in Table 5.4, where the average, standard deviation, and highest/lowest value achieved by each method are reported. In order to validate the significance of the experimental results, paired-t tests are carried out amongst the four methods, in which the weighting vectors are the same in each pair. The winners are listed in cells of Table 5.5, with “-” indicating that the difference between the compared methods is “not significant” ($p > 0.05$). Since on each dataset there are only 21 pairs of weighting vectors tested (the number of observations is usually beyond 30 in the paired t-test), the results from all datasets are collected together to form a relatively large set of observations for the paired-t test, and the according paired t-test results are indicated as “overall” in Table 5.5.

5.4.4 Results and Discussion

Firstly, the results achieved by the T -transitive aggregations and those by the not T -transitive aggregations are compared. Since it is difficult to define a non-transitive

5.4. Application to Cluster UCI Datasets

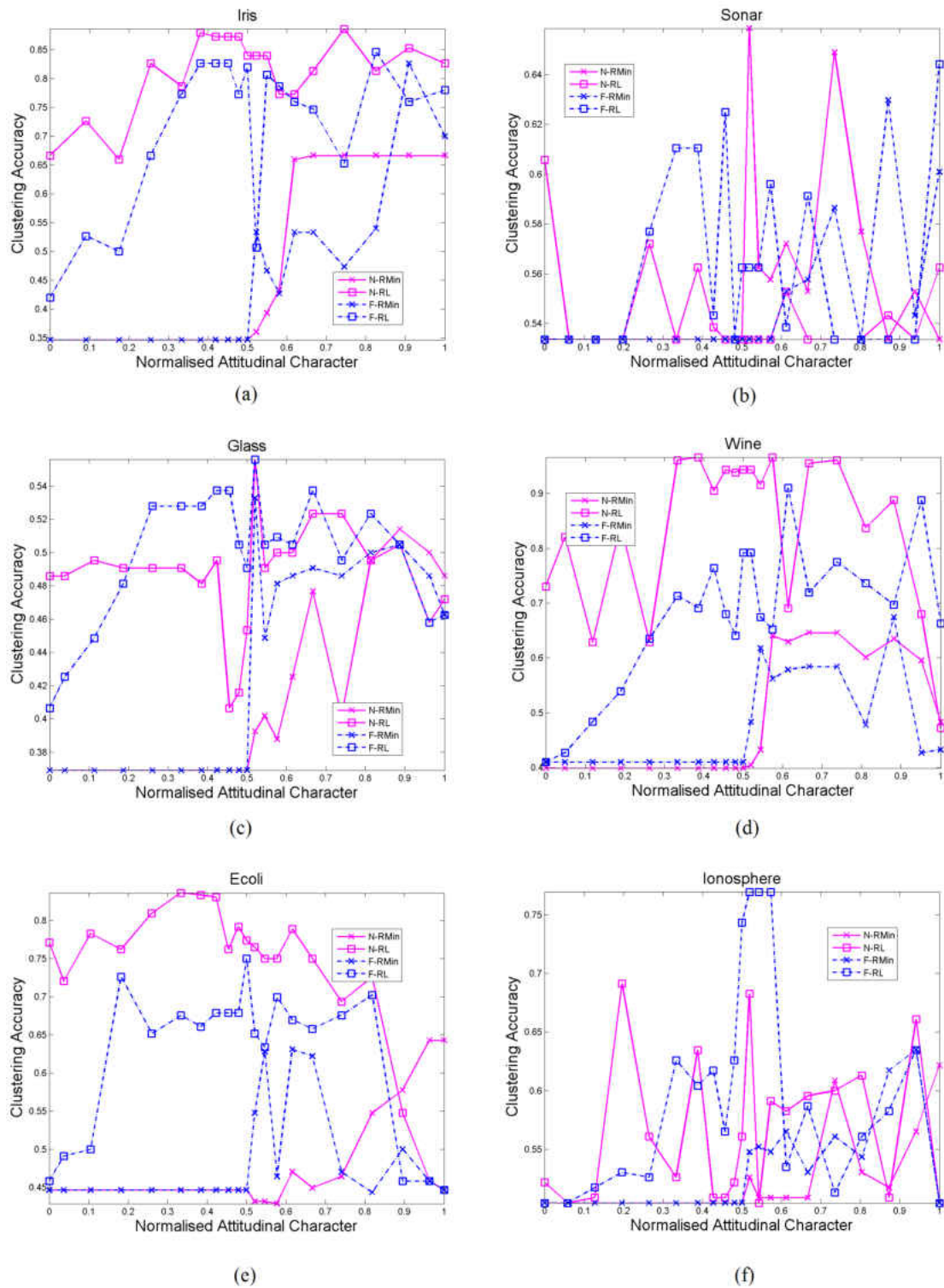


Figure 5.4: Trend of Accuracy against Normalised Attitudinal Character

Table 5.4: Comparison of Accuracy (%): T -transitivity

	Average \pm Standard Deviation			Maximum Value-Minimum Value				
	$N-R^{\min}$	$N-R^L$	$F-R^{\min}$	$F-R^L$	$N-R^{\min}$	$F-R^L$		
Iris	46.00 \pm 15.05	80.57 \pm 7.13	45.46 \pm 14.39	70.51 \pm 14.19	66.67-34.67	88.67-66.00	82.67-34.67	84.67-42.00
Sonar	55.34 \pm 3.62	54.33 \pm 0.85	54.65 \pm 2.66	56.32 \pm 3.59	65.87-53.37	60.58-53.37	62.98-53.37	64.42-53.37
Glass	40.68 \pm 5.28	48.64 \pm 3.32	42.57 \pm 6.26	49.87 \pm 3.94	51.40-36.92	55.61-40.65	53.27-36.92	55.61-40.65
Wine	48.10 \pm 10.85	83.90 \pm 14.37	47.30 \pm 8.74	68.00 \pm 13.11	64.61-39.89	96.63-47.19	67.42-41.01	91.01-41.01
Ecoli	47.60 \pm 6.63	73.09 \pm 11.08	48.20 \pm 6.54	61.92 \pm 10.13	64.29-42.86	83.63-44.64	63.10-44.35	75.00-44.64
Ionosphere	52.15 \pm 3.43	56.63 \pm 6.26	53.11 \pm 3.87	59.96 \pm 9.20	62.17-50.44	69.13-50.44	63.48-50.44	76.96-50.44

Table 5.5: Comparison of Accuracy: Paired t -test

	$N-R^{\min}$ vs. $N-R^L$	$F-R^{\min}$ vs. $F-R^L$	$N-R^{\min}$ vs. $F-R^{\min}$	$N-R^L$ vs. $F-R^L$
Iris	$N-R^L$	$F-R^L$	-	$N-R^L$
Sonar	-	-	-	$N-R^L$
Glass	$N-R^L$	$F-R^L$	-	-
Wine	$N-R^L$	$F-R^L$	-	$N-R^L$
Ecoli	$N-R^L$	$F-R^L$	-	$N-R^L$
Ionosphere	$N-R^L$	$F-R^L$	-	-
Overall	$N-R^L$	$F-R^L$	-	$N-R^L$

Table 5.6: Comparison of Accuracy (%) of $N-R^{\min}$: T^{\min} -transitive vs. Not T^{\min} -transitive

	Average±Standard Deviation		Best-accuracy (A-C'(W))	
	T_{\min} -transitive	Not T_{\min} -transitive	T_{\min} -transitive	Not T_{\min} -transitive
Iris	34.67±0.00	58.47±13.17	34.67 ([0.00, 0.50])	66.67 ([0.67, 1.00])
Sonar	53.37±0.00	57.50±4.39	53.37 ([0.00, 0.50])	65.87 (0.52)
Glass	36.92±0.00	44.81±5.06	36.92 ([0.00, 0.50])	51.40 (0.89)
Wine	39.89±0.00	57.13±9.41	39.89 ([0.00, 0.50])	64.61 (0.67, 0.74)
Ecoli	44.64±0.00	50.86±8.66	44.64 ([0.00, 0.50])	64.29 (0.96, 1.00)
Ionosphere	50.44±0.00	54.04±4.31	50.44 ([0.00, 0.50])	62.17 (1.00)

counterpart for a transitive aggregation, the paired t -test is not available in this comparison. In this case, the average accuracy (with standard deviation) and the highest accuracy (with the attitudinal character value(s) where it is achieved) are reported in Tables 5.6-5.9. If the highest accuracy is obtained by more than two weighting vectors, their attitudinal character values are given by intervals. It is worth noticing that not all the values in the interval are tested in this experiments, but only those discrete points which are shown in Figure 5.4.

The results show that for the application of R^{\min} aggregation to the complete-link hierarchical clustering, the use of non-transitive weighting vectors leads to the better average and highest accuracy on most of the datasets then those of transitive

Table 5.7: Comparison of Accuracy (%) of $F-R^{\min}$: T^{\min} -transitive vs. Not T^{\min} -transitive

	Average±Standard Deviation		Best-accuracy (A-C'(W))	
	T_{\min} -transitive	Not T_{\min} -transitive	T_{\min} -transitive	Not T_{\min} -transitive
Iris	34.67±0.00	57.33±12.70	34.67 ([0.00, 0.50])	82.67 (0.91)
Sonar	53.37±0.00	56.06±3.39	53.37 ([0.00, 0.50])	62.98 (0.87)
Glass	36.92±0.00	48.79±2.28	36.92 ([0.00, 0.50])	53.27 (0.52)
Wine	41.01±0.00	54.21±8.26	41.01 ([0.00, 0.50])	67.42 (0.88)
Ecoli	44.64±0.00	52.11±7.90	44.64 ([0.00, 0.50])	63.10 (0.62)
Ionosphere	50.44±0.00	56.04±3.87	50.44 ([0.00, 0.50])	63.48 (0.94)

Table 5.8: Comparison of Accuracy (%) of $N-R^L$: T^L -transitive vs. Not T^L -transitive

	Average±Standard Deviation		Best-accuracy (A-C'(W))	
	T_L -transitive	Not T_L -transitive	T_L -transitive	Not T_L -transitive
Iris	78.85±9.17	82.47±3.44	88.00 (0.38)	88.67 (0.75)
Sonar	54.68±2.37	53.95±1.03	60.58 (0.00)	56.25 (1.00)
Glass	47.20±3.22	50.23±2.76	49.53 (0.11, 0.42)	55.61 (0.52)
Wine	84.63±12.91	83.09±16.50	96.63 (0.39)	96.63 (0.57)
Ecoli	78.84±3.63	66.76±13.15	83.63 (0.33)	78.87 (0.62)
Ionosphere	54.98±6.06	58.43±6.26	69.13 (0.20)	68.26 (0.52)

Table 5.9: Comparison of Accuracy (%) of $F-R^L$: T^L -transitive vs. Not T^L -transitive

	Average±Standard Deviation		Best-accuracy (A-C'(W))	
	T_L -transitive	Not T_L -transitive	T_L -transitive	Not T_L -transitive
Iris	67.09±17.09	74.27±9.66	82.67 ([0.38,0.45])	84.67 (0.83)
Sonar	56.34±3.64	56.30±3.73	62.50 (0.46)	64.42 (1.00)
Glass	49.24±4.68	50.56±3.00	53.74 (0.42,0.46)	55.61 (0.52)
Wine	61.59±13.22	75.06±9.07	79.21 (0.50)	91.01 (0.61)
Ecoli	63.18±9.99	60.53±10.62	75.00 (0.50)	70.24 (0.82)
Ionosphere	57.87±7.34	62.26±10.82	74.35 (0.50)	76.96 ([0.52, 0.57])

weighting vectors. The main reason is that when the weighting vectors employed in the R^{\min} aggregation satisfy the constraint in Theorem 4 to retain transitivity, the R^{\min} aggregation will behave as the classic min operator, and its aggregated results will be $\min_{i=1,\dots,m} R_i(a, b)$ (see Theorem 7). The extreme pessimistic behaviour of min operator undermines the quality of clustering on these tested datasets. The results also imply that preserving T_{\min} -transitivity for R^{\min} aggregation does not necessarily provide a good performance of it in hierarchical clustering effectively. Generally, both $N-R^{\min}$ and $F-R^{\min}$ achieve their highest accuracies when the attitudinal character values of weighting vectors are closed to 1, suggesting that the weighting vectors which show orness (namely, inputs with high values dominate the aggregation result) are more preferable when R^{\min} operators are employed for conducting hierarchical clustering. However, the performance of the extreme orness case, i.e., W_{\max} is slightly worse than the best one achievable. This implies that the R^{\min} aggregator can produce better result than the classic max and min operators in the aggregation of T_{\min} -based fuzzy similarities on different features.

The performance of non- T^{\perp} -transitive aggregations is not obviously better than that of T^{\perp} -transitive aggregations in general. In fact, the component similarity relation used in $N-R^{\perp}$ (Equation (5.15) with only one linguistic term defined by Equation (5.16)) is just the negation to the classic Euclidean/Manhattan distance metric on $[0,1]$. Particularly, when W_{mean} is applied to $N-R^{\perp}$ aggregation, its resultant hierarchical clustering is identical to the Manhattan (“city block”) distance based hierarchical clustering (see the points on the line $N-R^{\perp}$ with $A-C'(W) = 0.5$ in Figure 5.4). By comparing the position of where the $N-R^{\perp}$ aggregated similarity relation obtain its highest accuracy, it can be concluded that the proposed R^{\perp} aggregated similarity has the potential of providing better clusters than the classic Manhattan distance in terms of hierarchical clustering. It also can be seen from Tables 5.7 and 5.9 that the R^{\perp} aggregator can outperform the classic max and min operators in the aggregation of fuzzy similarities with appropriate weighting vectors.

Secondly, a comparison is made between the quality of the clusters generated by R^{\min} and that of R^{\perp} . For both the normalisation and fuzzification of datasets, the accuracies of the R^{\perp} aggregations are significantly better than those achieved by the R^{\min} aggregations in five of the six datasets (see the second and third columns in Table 5.5). This indicates that the application of R^{\perp} aggregated fuzzy relations to hierarchical clustering entails more effective information for comparing patterns.

This is mainly because the component similarities employed in R^{L} is symmetric about a fixed pattern on each dimension, while the component similarities employed in R^{min} tend to assign small similarity values to patterns with low memberships. As a consequence in the hierarchical clustering, patterns which have low memberships are difficult to merge with other patterns and cause “singleton” clusters. Another possible reason is that the R^{L} aggregation produces continuous valued results while R^{min} aggregation leads to the result that many patterns may have the same degree of similarity to a certain pattern), which in turn allows R^{L} generates better results than R^{min} .

Finally, this is to compare the results on normalised datasets and those on fuzzified datasets of using the proposed two aggregations in the hierarchical clustering. It can be seen from Table 5.5 that when R^{L} aggregation is applied, the accuracies achieved on three normalised datasets are significantly better than those achieved on their fuzzified counterparts. One possible explanation is that the component similarity relation defined in Equation (5.15) is too “pessimistic”, which means that many patterns have $R_i^{\text{L}}(p_a, p_b) = 0$ after the fuzzification. Intuitively, fuzzification can help the min-based similarity metric to eliminate the effect of always assigning small similarity values to patterns with low memberships and hence, improve the performance of $F-R^{\text{min}}$. However, the paired t-test in the experimental results failed to show that. One possible explanation is that the membership functions showed in Figure 5.3 are not suitable for defining the linguistic terms on some of the tested datasets.

5.5 Summary

This chapter has studied the T -transitivity of ordered weighted aggregation of fuzzy relations and their applications as pairwise similarity matrices for hierarchical clustering. The proposed aggregated similarities take the advantage of OWA-like aggregators in that the behaviour of the aggregations can be controlled by stress functions. Furthermore, the conditions regarding when the aggregated similarity may preserve T -transitivity are also investigated. Experimental results on six UCI datasets indicate that the proposed ordered weighted aggregations generally outperform the conventional aggregators/distances in hierarchical clustering.

Chapter 6

Ensemble of Fuzzy Clusters for Journal Ranking

TO avoid the bias which is caused by using a single impact indicator, ensemble of multiple indicators is a promising method to produce a more robust ranking result. Besides fuzzy aggregation, other fuzzy techniques such as fuzzy clustering, have proven to be effective for many applications of decision making and multi-criteria evaluation [5, 39, 81]. In this chapter, a fuzzy aggregation based fuzzy clustering ensemble method is proposed for ranking academic journals. In particular, the OWA operators with dependent weights, are applied to aggregate the scores of academic journals.

Distinctive fuzzy clusters of journals are constructed based on their performance with respect to different journal impact indicators. These may be subsequently combined via the use of various OWA operators. For each individual indicator, fuzzy clusters of journals which are labelled with linguistic terms are generated. The OWA operators are then employed to aggregate various fuzzy clusters according to their linguistic labels, and the final fuzzy clusters are formed. Two refinement methods are also introduced to support the evaluation in order to generate a rank of journals according to their memberships to those fuzzy clusters. The overall ranking process is not only more reliable and interpretable than ranking by the original indicator scores but also very intuitive. The k NN-DOWA can enhance the reliability while the fuzzy clustering improves the interpretability of the ranking procedure. The ranking

results of academic journals from six subjects are compared with the RJL published in ERA (2010).

The remainder of this chapter is organised as follows. Section 6.1 introduces the basics of clustering ensemble and fuzzy clustering ensemble. Section 6.2 describes the details of the proposed fuzzy aggregation and clustering ensemble based method for journal ranking. Section 6.3 presents the experimental evaluation of the proposed approach, along with a discussion of the results. Finally, Section 6.4 concludes the chapter.

6.1 Preliminaries

Clustering is one of the important approaches within the framework of unsupervised learning, which is helpful for finding the hidden structure of unlabelled datasets. In general, the task of clustering is to assign objects to groups (namely clusters) such that objects in the same group are similar to each other, and dissimilar to those in the other clusters [105]. A good number of clustering algorithms have been proposed in the literature and have been successfully applied to many datasets [121]. For a given problem, different algorithms, and indeed even the same algorithm with different parameter settings (e.g., the number of clusters assumed), typically lead to different solutions [100]. Hence, an inexperienced user runs the risk of fetching an inappropriate clustering method. Also, in unsupervised learning, there is usually no ground truth against which the result can be evaluated. Therefore, it is extremely difficult for users to decide which algorithm is appropriate for a certain problem domain [87].

To overcome the aforementioned limitations, improving the accuracy as well as robustness of individual clustering methods, clustering ensembles have emerged as effective solutions. Similar to the classifier ensemble [51] and feature selection ensemble [170], the clustering ensemble combines results of various clustering algorithms and may do so in different ways. One of the main objectives of the combination is to achieve accuracy superior to those of individual clustering [175]. By combining multiple partitions of a set of data points into a single consolidated clustering, the performance of clustering ensembles generally depends on both the quality of ensemble components and the methods of aggregating those components. This has been empirically verified in [87, 121].

6.1.1 Clustering Ensemble

Formally, the clustering ensemble problem can be described as follows. Let $X = \{x_1, \dots, x_N\}$ be a set of N data points and $\Pi = \{\pi_1, \dots, \pi_i, \dots, \pi_m\}$ be m base-clustering members. Each base-clustering member returns a set of clusters $\pi_i = \{C_1^i, \dots, C_{K_i}^i\}$ such that $\bigcup_{k=1}^{K_i} C_k^i = X$, where K_i is the number of clusters constructed by that member. For each $x \in X$ and each base-clustering member π_i , $C^i(x) \in \pi_m$ denotes the cluster label to which the data point x belongs in π_i . The task of clustering ensemble is to find a new clustering result π^* for the given dataset X , which summarises the information embedded in the whole clustering ensemble Π [100].

Two key procedures are involved in the development of a clustering ensemble technique. First, base-clustering members are generated, typically by artificially diversifying methods for parameter settings and data re-sampling. Second, a consensus function is then applied to those base-clustering members to generate the final clustering result. The procedure of clustering ensemble is illustrated in Figure 6.1.

A number of methods have been proposed that have helped to address these issues. For example, in order to ensure diversity of component clustering members, different parameter configurations of a given clustering algorithm have been tested [73, 80]; re-sampling techniques have also been applied to diverse base-clusters [54, 112]. Regarding the techniques for the issue of consensus, existing methods include: feature-based approaches where each base-clustering member provides cluster labels as a feature describing data points, the resultant new dataset is then utilised to formulate the final solution [141, 184]; pairwise similarity based approaches which create a matrix, containing the pairwise similarity measures amongst data points, then any similarity-based clustering algorithm (say, hierarchical clustering) can be applied [73]; graph-based approaches which manipulate data partitions by exploiting the graph representation [68, 175, 215].

A consensus function can be generally viewed as a map from a set of base-clustering members to one final partition of the original data $f : \Pi \rightarrow \pi$. Once the base-clusters are generated from the data, a variety of consensus functions that are readily available may be applied to derive the final data partition. Most of the consensus functions utilise an ensemble-information matrix which aggregates the base-clustering members. Given the ensemble of Figure 6.2, two types of such a

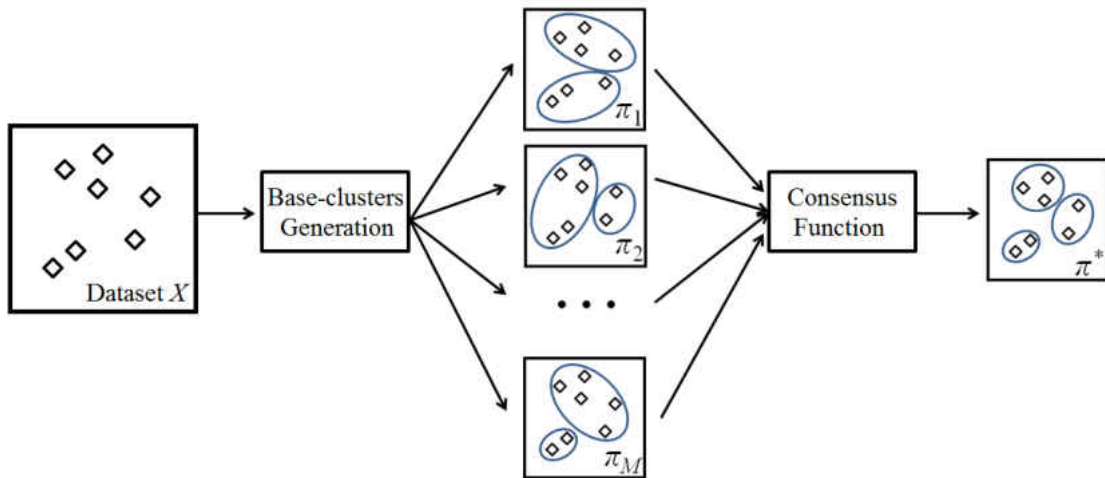


Figure 6.1: Clustering Ensemble

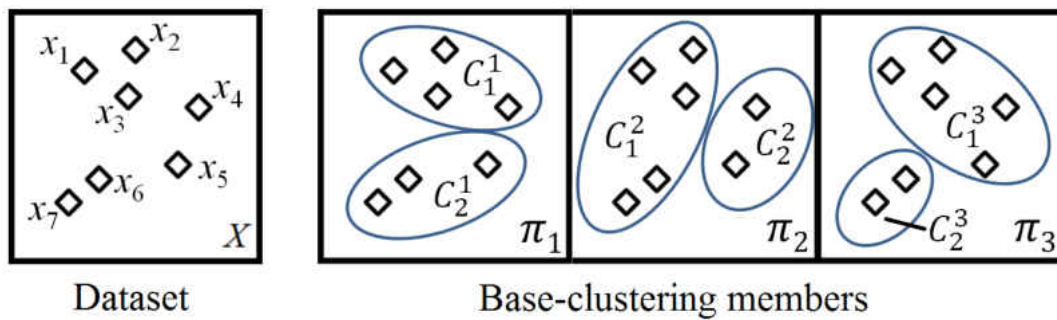


Figure 6.2: Example of Clustering Ensemble

matrix: the label-assignment matrix and the binary cluster-association matrix are illustrated in Tables 6.1 and 6.2, respectively.

Usually, a categorical data clustering algorithm is further applied to this type of ensemble-information matrix to achieve the final partition of the original data. Alternatively, an ensemble may be represented as a graph, where the nodes are base-clusters or data points and links between them define the relationships holding amongst the clusters and data points [189].

6.1.2 Fuzzy Clustering Ensemble

If a crisp clustering algorithm such as k -means [128] is used in the generation of clusters, the association degree of a data point belonging to a specific cluster is either

Table 6.1: Label-assignment Matrix

	π_1	π_2	π_3
x_1	C_1^1	C_1^2	C_1^3
x_2	C_1^1	C_1^2	C_1^3
x_3	C_1^1	C_1^2	C_1^3
x_4	C_1^1	C_2^2	C_1^3
x_5	C_2^1	C_2^2	C_1^3
x_6	C_2^1	C_1^2	C_2^3
x_7	C_2^1	C_1^2	C_2^3

Table 6.2: Binary Cluster-association Matrix

	C_1^1	C_2^1	C_1^2	C_2^2	C_1^3	C_2^3
x_1	1	0	1	0	1	0
x_2	1	0	1	0	1	0
x_3	1	0	1	0	1	0
x_4	1	0	0	1	1	0
x_5	0	1	0	1	1	0
x_6	0	1	1	0	0	1
x_7	0	1	1	0	0	1

1 or 0. However, there are other popular clustering algorithms such as fuzzy c -means [21] that naturally produce clusters of data with uncertain boundaries. In other words, fuzzy c -means can generate fuzzy partitions for a given dataset. It has been successfully applied to image segmentation [126], chemical structures clustering [94], and many scenarios in medicine and healthcare [2, 130]. Each cluster in a fuzzy partition $\tilde{\pi}$ is a fuzzy set $\tilde{C}_k, k = 1, \dots, K$ where $\tilde{C}_k(x) \in [0, 1]$ represents the degree of a data point $x \in X$ belonging to the corresponding fuzzy cluster. Usually, this degree is normalised with all the clusters in a partition to satisfy that $\sum_{k=1}^K \tilde{C}_k(x) = 1$.

Formally, a fuzzy (or soft) clustering ensemble can be described as follows [156]. Let $X = \{x_1, \dots, x_N\}$ be a set of N objects and $\Pi = \{\tilde{\pi}_1, \dots, \tilde{\pi}_i, \dots, \tilde{\pi}_m\}$ be m fuzzy ensemble members. Each ensemble member returns a set of fuzzy clusters $\tilde{\pi}_i = \{\tilde{C}_1^i, \dots, \tilde{C}_{K_i}^i\}$ such that for each $x \in X$, $\sum_{k=1}^{K_i} \tilde{C}_k^i(x) = 1$, where K_i is the number

of fuzzy clusters constructed by that member, and $\tilde{C}_k^i(x) \in [0, 1]$ denotes the degree of which the data point x belongs to the fuzzy cluster \tilde{C}_k^i .

The fuzzy clusters generated by all ensemble members together form a set of fuzzy base-clusters for the ensemble: $\{\tilde{C}_1, \dots, \tilde{C}_M\} = \bigcup_{i=1}^m \tilde{\pi}_i$, where $M = \sum_{i=1}^m K_i$. An example of the so-called instance-cluster matrix of a fuzzy clustering ensemble is shown in Table 6.3. The task of a fuzzy clustering ensemble is: for a given dataset X , find a new fuzzy partition $\tilde{\pi}^*$ (or a crisp partition π^*) which summarises the information embedded in the whole clustering ensemble Π .

Table 6.3: Fuzzy Cluster-association Matrix

	C_1^1	C_2^1	C_1^2	C_2^2	C_1^3	C_2^3
x_1	0.6	0.4	0.6	0.4	0.6	0.4
x_2	0.8	0.2	0.8	0.2	0.8	0.2
x_3	0.5	0.5	0.9	0.1	0.8	0.2
x_4	0.7	0.3	0.2	0.8	0.8	0.2
x_5	0.2	0.8	0.4	0.6	0.6	0.4
x_6	0.4	0.6	0.6	0.4	0.1	0.9
x_7	0.0	1.0	0.7	0.3	0.1	0.9

Note that a key difference between crisp and fuzzy clustering ensemble is that the latter works on fuzzy clusters. If the fuzzy base-clusters are defuzzified into crisp clusters, many of the consensus functions designed for crisp clustering ensemble can be borrowed for building clustering ensembles of fuzzy base-clusters. However, valuable information may be lost in defuzzification and hence, the quality of the ensemble may be adversely affected [156]. For example, in the crisp clustering ensemble, the instance-cluster matrix is sparse and its contained information for a clustering ensemble is incomplete [100]. Thus, although fuzzy base-clusters contain more information, conventional consensus functions for crisp clustering ensemble cannot directly make use of such information. It is due to this observation that fuzzy or soft clustering ensembles have been introduced in the literature [156, 214] and followed on herein. Inspired by these observations, and in order to utilise the interpretability of fuzzy linguistic terms, a fuzzy aggregation based fuzzy clustering ensemble method is proposed in this chapter for ranking academic journals.

6.2 Fuzzy Aggregation and Clustering Ensemble based Journal Ranking

A number of indicators have been widely accepted and applied by scholars, which typically focus on one particular aspect of academic journals such as citations. Scores of journals gained under various indicators can be directly integrated by using the aggregation methods introduced in Section 4.1. However, when human experts assess the quality of academic journals, linguistic terms are commonly and sensibly used to support their judgement. Therefore, the interpretable estimation of journals' quality with respect to labelled fuzzy clusters (rather than the pure scores) under each indicator is utilised in this chapter in order to conduct journal ranking. In particular, a linguistic term based consensus method is proposed to regroup the fuzzy clusters generated by different indicators and OWA operators with dependent weights are applied to integrate fuzzy memberships.

The proposed journal ranking method is nicknamed FACE to reflect the fact that it is based on fuzzy aggregation and cluster ensemble. Briefly, its working process starts by creating ensemble members using fuzzy *c*-means on each of the journal impact indicators which are available (and selected) from on-line databases of academic publications. The resultant (fixed number of) fuzzy clusters, termed base clusters for easy reference, are associated with predefined linguistic labels. The preference relation amongst linguistic terms is then employed to group the base clusters. The OWA operators are used to group base clusters belonging to different cluster ensemble members, forming the final fuzzy clusters. The method may also involve the following two optional steps: 1) defuzzifying the resultant fuzzy clusters such that each data point (i.e., journal) belongs to just one final crisp cluster (which may still be associated with a linguistic label) and hence, introducing a weak ranking amongst all journals; and 2) combining the memberships of a given journal to all fuzzy clusters into a single index of rank, thereby giving an absolute rank amongst all journals. An illustrative flowchart of the FACE algorithm is showed in Figure 6.3 and the following subsections detail its key operations.

6.2.1 Indicator-based Generation of Fuzzy Base-clusters

When translating a set of real-valued scores into a linguistic term which is closer to the use of natural language, it is a common practice to employ fuzzification

6.2. Fuzzy Aggregation and Clustering Ensemble based Journal Ranking

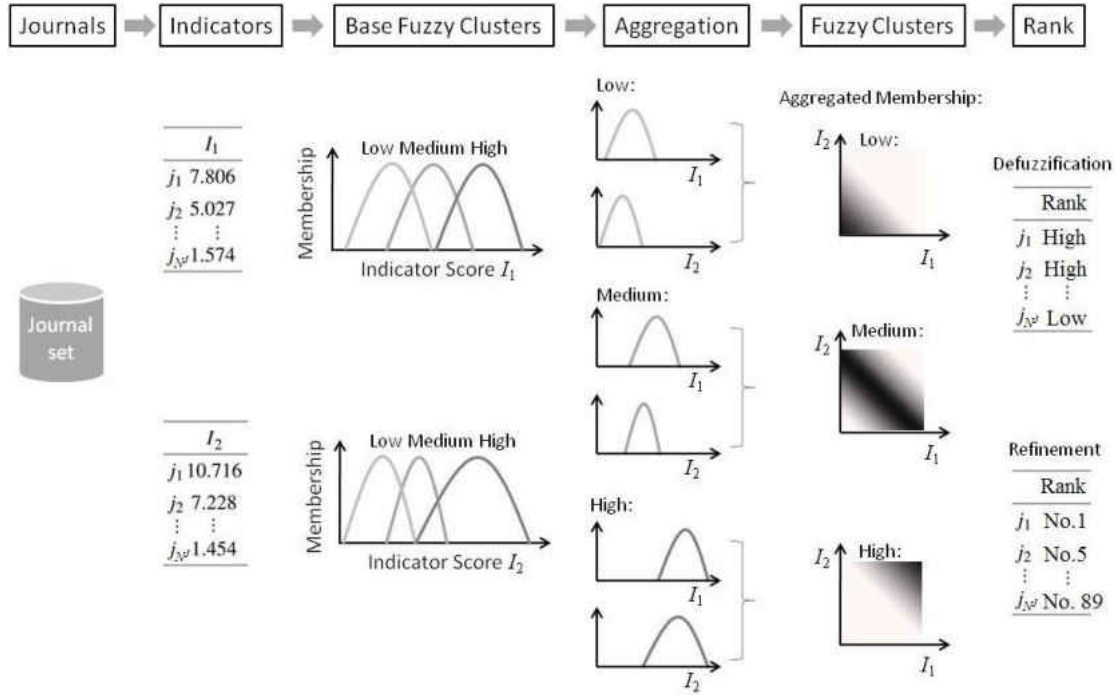


Figure 6.3: Fuzzy Aggregation and Clustering Ensemble based Journal Ranking

techniques. For the present work, fuzzy c -means, which is able to retain the non-binary memberships of each data point to all clusters, is adopted to translate the numerical indicator scores into predefined linguistic terms. Without losing generality, suppose that a set of journals J is evaluated with regard to m impact indicators I_1, \dots, I_m , and that each indicator $I_i, i = 1, \dots, m$ is a mapping $I_i : J \rightarrow \mathbb{R}$. Also, it is naturally presumed that the higher impact indicator score is assigned to a journal the higher impact that journal is believed to possess. For each indicator I_i , fuzzy c -means is then utilised to form clusters for J with respect to $\{I_i(j) | j \in J\}$ and a pre-specified number K (which indicates the number of fuzzy subsets of J that are required to be constructed). From this, K fuzzy sets are formed with $\tilde{C}_1^i(j), \dots, \tilde{C}_K^i(j)$ representing the memberships of a journal $j \in J$ belonging to the resulting individual fuzzy clusters, respectively.

When linguistic terms are employed to describe a variable, a preference ordering relation is usually defined on the set of linguistic terms such as *Bad* \prec *Acceptable* \prec *Good* or *Low* \prec *Medium* \prec *High*. In general application of fuzzy clustering, such an ordered labelling scheme over the clusters is not necessary. However, in FACE, labelling the clusters is not only helpful to understand the relative quality of journals in a cluster, but also important to organise base clusters in the subsequent aggregation

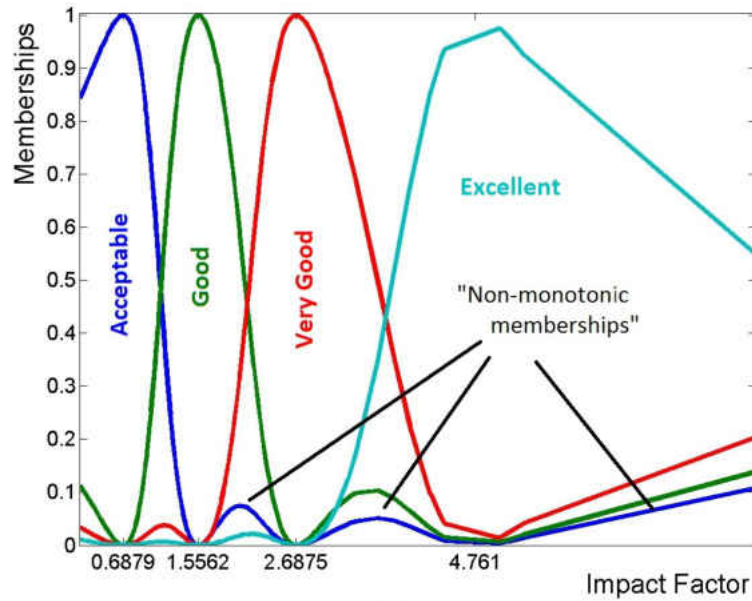
process. The required labelling may be accomplished by consulting human experts in the field. Yet, since the fuzzy clusters are herein generated according to an individual impact indicator whose values are totally ordered, the value of each cluster centre can be employed to signify the overall relative quality of that cluster. Thus, given a set of K pre-defined set of linguistic terms $L = \{L_1, \dots, L_K\}$ which satisfy that $L_1 \prec \dots \prec L_K$, the fuzzy clusters $\tilde{C}_1^i, \dots, \tilde{C}_K^i$ can be readily sorted in ascending order with regard to their cluster centres and then, are labelled with L_1, \dots, L_K respectively.

Note that a possible drawback of employing fuzzy c -means to implement fuzzification is that a data point's membership to a cluster is not monotonically decreasing with its distance to the cluster centre. This is caused by the mechanism of normalisation which is inherent in the fuzzy c -means algorithm. If the fuzzy clusters are defuzzified into crisp clusters by assigning each object to the cluster with which it obtains the maximum membership, the non-maximum (and non-monotonic) memberships will have no impact on the final crisp result and hence, will be ignored. However, in FACE, memberships of a journal to all those linguistically labelled clusters are useful in the subsequent aggregation. Therefore, a filtering process is applied to the resultant fuzzy memberships to ensure that the membership of a journal to a cluster is monotonically decreasing with its distance to the cluster centre. Such a filtering process can be implemented using the following two steps:

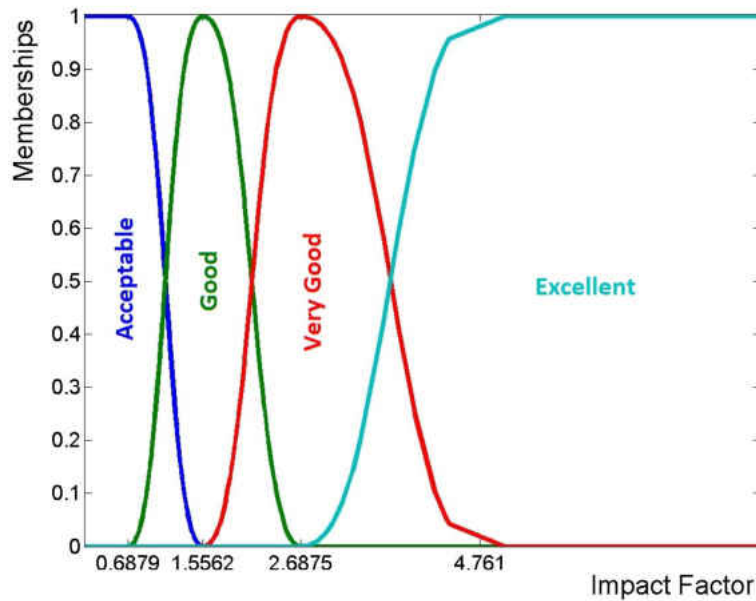
1. For each labelled fuzzy cluster $\tilde{C}_{L_k}^i, k = 2, \dots, K$, set membership $\tilde{C}_{L_k}^i(j) = 0$ for each $j \in J$ where $I_i(j)$ is smaller than the centre of $\tilde{C}_{L_{k-1}}^i$; and for each labelled fuzzy cluster $\tilde{C}_{L_k}^i, k = 1, \dots, K - 1$, set membership $\tilde{C}_{L_k}^i(j) = 0$ for each $j \in J$ where $I_i(j)$ is greater than the centre of $\tilde{C}_{L_{k+1}}^i$;
2. For each journal $j \in J$, update its memberships to all the clusters by normalisation:

$$\tilde{C}_{L_k}^i(j) = \frac{\tilde{C}_{L_k}^i(j)}{\sum_{t=1}^K \tilde{C}_{L_k}^t(j)}. \quad (6.1)$$

Figure 6.4 shows an example of such a filtering process. Figure 6.4(a) is the fuzzy c -means result on a selective set of journals in Computer Science which are evaluated by the Impact Factor of 2010. Figure 6.4(b) is the filtered result using the above method.



(a)



(b)

Figure 6.4: Filtering the Fuzzy *c*-means Result

6.2.2 Base-cluster Grouping and Aggregation of Memberships

Having gone through the fuzzification process as described in the preceding subsection, $m \times K$ fuzzy clusters are generated and labelled. In this step, the $m \times K$ fuzzy base clusters are grouped into K final clusters which are again labelled by the pre-defined set of linguistic terms L . Generally, cluster ensemble algorithms involve

unsupervised grouping of base clusters, with many working methods available in the literature to implement such grouping, including: feature-based, graph-based, voting-based, and so on [98]. However, since the fuzzy clusters of journals generated in FACE are automatically labelled (though using predefined linguistic terms), an intuitive “supervised” grouping of them becomes feasible (which is less challenge than using the unsupervised approaches). This is described below.

Given the $m \times K$ labelled base clusters $\tilde{C}_{L_1}^1, \dots, \tilde{C}_{L_K}^1, \tilde{C}_{L_1}^2, \dots, \tilde{C}_{L_K}^{m-1}, \tilde{C}_{L_1}^m, \dots, \tilde{C}_{L_K}^m$, owing to their inherent ordering, they can be (re-)categorised into K groups $C_1 = \{\tilde{C}_{L_k}^i | k = 1, i = 1, \dots, m\}, \dots, C_K = \{\tilde{C}_{L_k}^i | k = K, i = 1, \dots, m\}$. Each $C_k, k = 1, \dots, K$ contains all the clusters with label L_k and $\tilde{C}_{L_k}^i$ is the fuzzy cluster which is generated by impact factor I_i .

To illustrate the construction of C_k , for simplicity, a crisp counterpart of C_K is addressed first. Consider a voting system in which each indicator votes for the top-rated “excellent” journals, for example. Those in C_K^i (i.e., the crisp counterpart of $\tilde{C}_{L_K}^i$) are the journals voted by I_i and hence, C_K contains all the journals that are each regarded as an “excellent journal” by at least one of I_1, \dots, I_m . Similarly, in general, C_k contains all those journals in the vote which are deemed to be of the quality level expressed by L_k . In such a crisp voting system, the votes can be summed for each journal and the winners can be ranked by how many ballots they have attracted. In FACE, however, each journal is not necessarily voted to having just one single quality level in a boolean way, but can have multiple explicit partial memberships assigned, indicating that it may be of different quality levels (though to various degrees). To make the best use of such information contained within such a voting system, more advanced aggregation operators rather than the simple sum/average are employed here to summarise the (both full and partial) votes, thereby deriving the final membership of a journal to a certain labelled fuzzy cluster C_k .

From this, given the K groups $C_1 = \{\tilde{C}_{L_k}^i | k = 1, i = 1, \dots, m\}, \dots, C_K = \{\tilde{C}_{L_k}^i | k = K, i = 1, \dots, m\}$, the memberships of a journal $j \in J$ to the final labelled fuzzy clusters $\tilde{C}_{L_k}^*, k = 1, \dots, K$ can be computed by:

$$\tilde{C}_{L_k}^*(j) = A(\tilde{C}_{L_k}^1, \dots, \tilde{C}_{L_k}^m). \quad (6.2)$$

where A is an aggregation operator. Then, the memberships of j to all the final fuzzy clusters are normalised by $\tilde{C}_{L_k}^*(j) = \tilde{C}_{L_k}^*(j) / \sum_{k=1}^K \tilde{C}_{L_k}^*(j)$. The full algorithm of FACE is shown in Algorithm 6.2.1.

Inputs: $J = \{j_1, \dots, j_x, \dots, j_{N^J}\}$: a dataset of N^J journals, where $j_x = (I_1(j_x), \dots, I_i(j_x), \dots, I_m(j_x)) \in \mathbb{R}^m$ and $I_i(j_x)$ is the score of journal j_x evaluated by impact indicator I_i ;

$L = \{L_1 \prec \dots \prec L_k \prec \dots \prec L_K\}$: a set of K linguistic terms with a preference relation;

Outputs: $\tilde{\pi}^* = \{\tilde{C}_1^*, \dots, \tilde{C}_K^*\}$: a fuzzy partition of J with linguistic label L .

- 1: **for** $i = 1 : m$ **do**
- 2: create sub-dataset $J_i = \{I_i(j_1), \dots, I_i(j_{N^J})\}$
- 3: create ensemble member $\tilde{\pi}_i = \{\tilde{C}_1^i, \dots, \tilde{C}_K^i\}$ using fuzzy c -means on J_i
- 4: sort $\tilde{C}_1^i, \dots, \tilde{C}_K^i$ to $\tilde{C}_{\pi(1)}^i, \dots, \tilde{C}_{\pi(K)}^i$ so that the cluster center of $\tilde{C}_{\pi(k)}^i$ is smaller than the cluster center of $\tilde{C}_{\pi(k')}^i$, for $k < k'$ ($k, k' = 1, \dots, K$)
- 5: label $\tilde{C}_1^i, \dots, \tilde{C}_K^i$ with L_1, \dots, L_K , respectively
- 6: **end for**
- 7: regroup all the fuzzy clusters $\bigcup_{i=1}^m \tilde{\pi}_i$ to create K groups of fuzzy base clusters $C_1 = \{\tilde{C}_{\pi(1)}^1, \dots, \tilde{C}_{\pi(1)}^m\}, \dots, C_K = \{\tilde{C}_{\pi(K)}^1, \dots, \tilde{C}_{\pi(K)}^m\}$
- 8: **for** $k = 1 : K$ **do**
- 9: **for** $x = 1 : N^J$ **do**
- 10: $\tilde{C}'_k(j_x) = A(\tilde{C}_{\pi(k)}^1(j_x), \dots, \tilde{C}_{\pi(k)}^m(j_x))$ where A is an aggregation operator
- 11: **end for**
- 12: **end for**
- 13: **for** $k = 1 : K$ **do**
- 14: **for** $x = 1 : N^J$ **do**
- 15: normalise $\tilde{C}'_k(j_x)$ by $\tilde{C}^*_k(j_x) = \tilde{C}'_k(j_x) / \sum_{k=1}^K \tilde{C}'_k(j_x)$, such that $\sum_{k=1}^K \tilde{C}^*_k(j_x) = 1$
- 16: **end for**
- 17: **end for**
- 18: label $\tilde{C}_1^*, \dots, \tilde{C}_K^*$ with L_1, \dots, L_K , respectively

Algorithm 6.2.1: Fuzzy Aggregation and Clustering Ensemble based Journal Ranking (FACE)

What is required now is the choice of a method to implement the aggregation operator A . As one of the possible mechanisms to perform the task of information aggregation, the concept of data reliability has been introduced [26], with successful extended applications for classification and feature selection. It works by exploiting the proximity to clusters of arguments and hence, can be rather inefficient. Recently, an enhanced version, termed k NN-DOWA, has been proposed in [27], where a hierarchical clustering process required by the original approach is replaced by a search of nearest neighbours. Although a number of aggregation operators are available in the literature and many of them have been applied to decision making [14, 35], they typically require subjective specification of the aggregation weights.

Here, k NN-DOWA is adopted to aggregate the memberships of journals voted by different impact indicators. This is feasible because: 1) the weights used in the aggregation are learned from the arguments automatically; and 2) the weights assigned to the arguments represent their reliability, which can be collected as useful “by-products” to further analyse and interpret the reliability of the underlying impact indicators.

For a dataset with n points and m features, the time complexity of the original fuzzy c -means is $O(nmK)$, where K is the number of clusters [116]. Since FACE employs fuzzy c -means on one dimension dataset for m (the number of impact indicators) times, the time complexity of FACE in generating the base clusters is also $O(N^J mK)$. The time complexity of consensus step depends on the aggregation operator A . Suppose that the complexity of consensus is $O(A)$, then the overall time complexity of FACE is $O(N^J mK) + O(A)$. Take k NN-DOWA as an example, the time complexity of k NN-DOWA is $O(m^2)$ [27]. Therefore, if it is adopted to aggregate the memberships of journals, the complexity of the consensus step is $O(N^J m^2 K)$, and the overall time complexity of FACE is $O(N^J mK) + O(N^J m^2 K) = O(N^J m^2 K)$.

6.2.3 Refinement for Ranking

Consider an example where the pre-defined set of linguistic terms is $\{Acceptable, Good, VeryGood, Excellent\}$ and the preference ordering relation is $Acceptable \prec Good \prec VeryGood \prec Excellent$. Suppose that the evaluation result of a journal using FACE is represented as a vector such as $(0.1, 0.1, 0.3, 0.5)$, whose elements denote the degrees of the journal belonging to the four (quality level) clusters, respectively. This form of result gives a “soft” evaluation of the quality of journals and is generally more informative than simply assigning journals to just one crisp cluster. Nevertheless, in many practical research quality assessment scenarios, it is not the absolute classification of journal qualities that is sought after, but the relative ranking amongst possible competitors. In order to decide on a rank over journals, using the information contained within the evaluation result vectors, two methods of transforming soft partition to ranks are provided here.

The first is to assign a journal to the cluster(s) in which it has the maximum membership. That is, taking the strategy of winners taking all. In so doing, the linguistic label associated with the final fuzzy cluster that possesses the maximum

membership degree becomes the quality level of that journal, i.e.,

$$\text{rank of } j = \arg \max_{L_k \in L} \tilde{C}_{L_k}^*(j). \quad (6.3)$$

Noted that L_K is the highest rank available for all journals while L_1 represents the lowest rank. Obviously, this method can only provide a fixed number of (i.e., K) ranks amongst the journals.

The alternative method is to assign a significance score to each of the linguistic terms and then, to sort the journals with respect to the weighted sum of the scores and journal (quality level) cluster memberships. For example, the scores can be set to $L_k = k$, reflecting the order of these quality levels. Then, the ranking over a set of journals can be obtained by sorting the journals in a descending order, according to:

$$\text{rank index of } j = \sum_{k=1}^K k \tilde{C}_{L_k}^*(j). \quad (6.4)$$

Compared against the first method, this second approach can provide a more detailed rank over journals. Note that the final ranks produced by the two methods are however, not necessarily in the same order. That is, journal j may be ranked higher than j' using the first method, but it may be ranked lower than j' if the second method is applied. The actual ranking outcomes depend on which method is used which in turn, depends on the results of the cluster ensemble. For example, suppose that the fuzzy evaluation of j is $(0.4, 0, 0, 0, 6)$ and that of j' is $(0, 0, 0.6, 0, 4)$, then j is ranked higher than j' using Eqn. (6.3) and lower using Eqn. (6.4). This is not a surprise, as these methods reflect different focus of attention, similar to the use of conventional defuzzification techniques, where a different defuzzification method may well result in a different defuzzified inference outcome. In real application, so long as an approach is consistently utilised throughout, the ranking results will be consistent.

6.3 Experiment and Evaluation

This section presents an experimental evaluation of the proposed FACE algorithm for journal ranking. It shows the setup of the experiments carried out and then discusses the results obtained.

6.3.1 Experiment Setup

In order to demonstrate the journal ranking results generated using the proposed FACE algorithm in a credible fashion, six indicators that are reported in the JCR (2010) are selected as base-indicators, and the RJL (2010) is assumed to be the ground truth. The employed six indicators include: Total Cites (TC), Journal Impact Factor (JIF), 5-year JIF (5-JIF), Immediacy Index (II), Eigenfactor (Ei), and Article Influence (AI). A summary of the resultant datasets is shown in Table 6.4. Each of these datasets contains over two hundred journals. The details of these indicators and the construction of datasets can be found in Section 3.3.

It is worth noticing that all these six indicators assign higher scores to journals with more citations to varying degrees. Such measurement generally indicates higher scores imply finer ranks for each of the indicators mentioned. Besides the indicators included in JCR, many other indicators are available from various academic publication databases. Note that in this work, it is the methods which aggregate individual impact indicators together with data reliability that are investigated, rather than the selection of base indicators themselves. Despite that more various indicators could be carefully selected for different disciplines by experts, only the indicators reported in JCR are tested here without losing generality.

Table 6.4: Summary of Datasets: FACE

Number of Instances	A*	A	B	C	Total
Chemistry	37	70	95	143	345
Computer Science	44	101	108	67	320
Material Science	26	61	80	61	228
Mathematics	52	84	127	69	332
Medicine	20	39	73	107	239
Physics	30	50	73	56	209

In statistics, Spearman's rank correlation coefficient r_s is a nonparametric measure of statistical dependence between two given variables [45]. It assesses how well the relationship between the two variables can be described using a monotonic function. If there are no identical data points, a perfect Spearman correlation of +1 or -1 occurs when each of the variables is a perfect monotonic function of the other. The sign of

r_s indicates the direction of association between one variable, say x (calling it the independent variable) and the other, say y (the dependent variable). If y tends to increase when x increases, r_s is positive, and if y tends to decrease when x increases, r_s is negative. $r_s = 0$ indicates that there is no tendency for y to either increase or decrease as x increases. The r_s between scores of each individual indicator and the RJI ranks are listed in Table 6.5.

Table 6.5: r_s between Scores of Individual Indicator and RJI Rank

Indicators	TC	JIF	5-JIF	II	Ei	AI
Chemistry	0.6665	0.7962	0.8152	0.7557	0.7019	0.8125
Computer Science	0.4914	0.4603	0.5023	0.3188	0.4130	0.5480
Material Science	0.6463	0.6153	0.6413	0.6045	0.6634	0.7185
Mathematics	0.5923	0.5610	0.5884	0.5262	0.6427	0.7287
Medicine	0.5401	0.4961	0.5010	0.5083	0.5368	0.5375
Physics	0.4501	0.6659	0.7299	0.5586	0.5095	0.7614
Average	0.5645	0.5991	0.6297	0.5454	0.5779	0.6844

Table 6.5 shows that all these indicators have a positive r_s value with respect to the RJI scores. This indicates that generally, if the scores of a journal on these indicators tend to increase, then their ranks in RJI increase also. However, for each indicator, its correlation levels to RJI are different from dataset to dataset. From their average performance on these datasets it can be seen that AI is the most correlated indicator to the rank of RJI, while II is the least relevant indicator. JIF and 5-JIF, which are commonly used in many real-world quality assessment scenarios, are more highly relevant to the results of RJI, as compared to TC and Ei.

To support systematic comparison, the quality levels of the journals that are awarded with respect to each of the individual indicators are aggregated using five different operators, namely: DOWA, kNN-DOWA and OWA with W_{mean} , $W_{andness}$ and W_{orness} . Scores of each indicator are (separately) normalised to $[0, 1]$ before clustering and aggregation. The weighting vectors in the OWA operators are not weight-dependent, thus a pre-definition of them are required. Instead of using the simple W_{max} and W_{min} , W_{orness} and $W_{andness}$ are employed (which are derived from linear stress functions in Figure 4.3). In particular, $W_{orness} = (0.29, 0.24, 0.19, 0.14, 0.09, 0.05)$,

$W_{andness} = (0.05, 0.09, 0.14, 0.19, 0.24, 0.29)$ and $W_{mean} = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$, given that there are six indicators to be aggregated in each of the experiments carried out. Note that $W_{andness}$ is directly implemented as the reverse of W_{orness} .

6.3.2 Results and Discussion

Both DOWA and kNN-DOWA use dependent weighting vectors, the resulting weights represent the reliability of the corresponding arguments. In DOWA, the reliability is measured by the similarity of an argument to the average of all arguments, while in kNN-DOWA, the reliability is done by the similarity of an argument to its k nearest neighbours. Since there are six indicators to be aggregated, the k in kNN-DOWA is set to 3, indicating that the majority of all 5 neighbours are considered. Tables 6.6-6.8 show the average weights that are computed for each impact indicator in OWA with $W_{andness}$, DOWA and 3NN-DOWA, respectively. Every entry in these tables is subtracted by $1/6$ (the average weight of each indicator) from its real value, so that a positive number means that the indicator is more highly weighted than the average, and that a negative one means that it is weighted lower than the average.

Table 6.6: Averaging Weight of Indicator in OWA with Andness Weighting Vector

Indicators	TC	JIF	5-JIF	II	Ei	AI
Chemistry	0.0002	-0.0045	-0.0478	-0.0194	0.0349	0.0366
Computer Science	0.0470	-0.0045	-0.0333	0.0043	0.0061	-0.0196
Material Science	-0.0152	-0.0073	-0.0081	-0.0119	0.0080	0.0345
Mathematics	-0.0011	-0.0023	0.0010	-0.0009	0.0072	-0.0039
Medicine	0.0021	0.0174	-0.0362	-0.0162	0.0214	0.0115
Physics	-0.0083	-0.0015	-0.0183	0.0156	0.0161	-0.0036
Average	0.0041	-0.0005	-0.0238	-0.0048	0.0156	0.0093

It can be seen from Table 6.6 that Ei, AI and TC have positive weights, while 5-JIF, II and JIF have negative weights when a conjunctive aggregation is run. These results indicate that on most journals, Ei, AI and TC tend to give lower scores compared with the other three indicators. Both DOWA and 3NN-DOWA weighted JIF, 5-JIF and AI higher than the other three indicators, which shows that they are considered more

Table 6.7: Averaging Weight of Indicator in DOWA

Indicators	TC	JIF	5-JIF	II	Ei	AI
Chemistry	-0.0062	-0.0055	-0.0018	0.0096	-0.0090	0.0129
Computer Science	-0.0098	0.0099	0.0140	-0.0173	-0.0063	0.0095
Material Science	-0.0097	0.0048	0.0003	0.0030	-0.0058	0.0074
Mathematics	-0.0076	0.0078	0.0059	0.0029	-0.0036	-0.0054
Medicine	-0.0080	0.0053	0.0033	-0.0045	-0.0117	0.0156
Physics	-0.0072	0.0159	0.0110	-0.0210	-0.0054	0.0067
Average	-0.0081	0.0064	0.0055	-0.0046	-0.0070	0.0078

Table 6.8: Averaging Weight of Indicator in 3NN-DOWA

Indicators	TC	JIF	5-JIF	II	Ei	AI
Chemistry	-0.0040	-0.0015	0.0029	0.0075	-0.0063	0.0014
Computer Science	-0.0046	0.0114	0.0162	-0.0312	-0.0011	0.0093
Material Science	-0.0172	0.0092	0.0062	0.0050	-0.0111	0.0079
Mathematics	-0.0093	0.0120	0.0094	0.0060	-0.0050	-0.0131
Medicine	-0.0100	0.0119	0.0120	-0.0087	-0.0152	0.0100
Physics	-0.0069	0.0174	0.0157	-0.0345	-0.0017	0.0100
Average	-0.0087	0.0101	0.0104	-0.0093	-0.0067	0.0043

“reliable” when used with these two aggregation operators. Note that individually, each of these three indicators also gains a relatively high r_s coefficient to the RJL result (see Table 6.5).

The r_s coefficients between the aggregated scores and the RJL results are depicted as the dot-lines in Fig. 6.5. On five out of the six datasets, 3NN-DOWA achieves the best or second best r_s results amongst all the five aggregation operators. However, its performance on the Mathematics dataset is not so good as those obtained using other aggregation operators. A possible reason is that the most RJL-relevant indicators are Ei, AI and TC on the Mathematics dataset while 3NN-DOWA puts more weight on JIF, 5-JIF and II. Similar to 3NN-DOWA, OWA with $W_{andness}$ also shows good results on these datasets, which indicates that the ranks produced by RJL are more like conjunctive of the impact indicators rather than the disjunctive of them.

6.3. Experiment and Evaluation

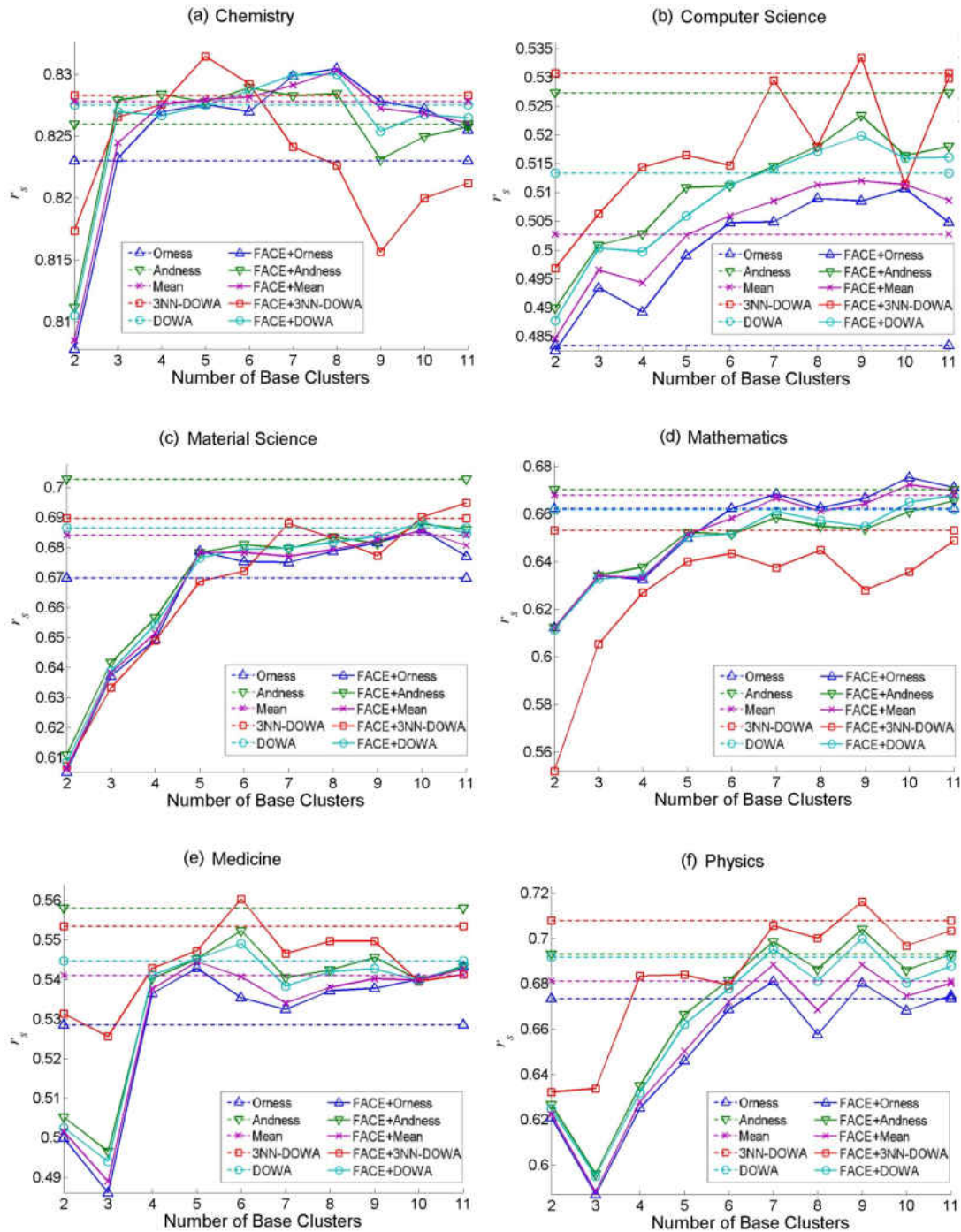


Figure 6.5: Spearman's Correlation to RJL Results

The solid lines in Fig. 6.5 show the r_s coefficients between journal ranks obtained with FACE and those by RJL. The number of base clusters on each impact indicator is consecutively set from 2 to 11 (to support a wide range of tests). Since the

direct aggregation of pure scores can provide a detailed rank, to entail an unbiased comparison, Eqn. (6.4) is employed to produce a ranking of the journals based on the final fuzzy clusters returned by FACE. As the fuzzy c -means algorithm starts with a random initialisation, each point on the solid lines is the average of 30 independent runs. However, as the impact of the initialisation of fuzzy c -means to any one dimensional dataset is small, the standard deviation of the results is very small. Therefore, standard deviations are not presented in Fig. 6.5.

In general, the r_s coefficient between the result of FACE and RJI is increased when the number of base-clusters is increased. The first result to notice is that on five out of the six datasets, the solid lines can reach above the highest dot lines. This indicates that using an appropriately selected number of base clusters, FACE can outperform the direct aggregation of individual indicator scores. These results also show that when FACE is employed, the highest r_s values on five out of the six datasets are achieved by the use of 3NN-DOWA. Overall, the results of 3NN-DOWA are better than those achievable using other aggregation operators on the following datasets: Computer Science, Medicine and Physics. Unfortunately, similar to the situation when 3NN-DOWA is applied to directly aggregate indicator scores, its performance on the Mathematics dataset is not so good as those obtained using other aggregation operators. Nevertheless, 3NN-DOWA generally achieves better results than DOWA both in direct aggregation and in FACE.

6.4 Summary

This chapter has presented a fuzzy aggregation and clustering ensemble based method for academic journal ranking, with a focussed application to aggregating indicators from the JCR provided by Web of Science. The proposed method works by exploiting data-reliability based aggregation of fuzzy clusters that are generated from scores returned by individual impact indicators. It helps strengthen the interpretability of the assessment outcomes for academic journals, thanks to the use of quality level terms with inherent linguistic meaning. Experimental results on real-world journals from six subject areas have shown that the ranking results of the proposed method are consistent with those by RJI, which are produced by a large group of journal-ranking specialists. Compared with the direct aggregation of individual indicator scores, the present work has an advantage in producing ranking results that are closer to the ground truth.

Chapter 7

Link-based Fuzzy Clustering Ensembles

ALTHOUGH much effort has been made in the development of clustering ensembles, modelling a mechanism that is effective for integrating multiple data partitions in a clustering ensemble is far from trivial. The development and application of clustering ensembles are still at an early stage [98]. Most of the existing clustering ensemble methods are based on crisp base-clusterings. However, interesting departures from the traditional work have recently been reported, such as that reported in [155, 214] where the problem of aggregating “soft” base-clustering members is defined.

Following this desirable trend, in this chapter, link-based consensus approaches for building ensembles of fuzzy c -means are proposed. Different from ensembles of crisp clusters, the proposed methods are able to handle fuzzy components. The work also differs from the link-based crisp clustering ensemble [99, 100], since it employs a graph with fuzzy links to represent the relationships between base-clusters and to refine the pairwise similarity matrix for generating the ensembles. With a number of UCI benchmark datasets [71], the proposed methods are tested against their crisp counterparts and those that utilise a fuzzy co-association matrix without link-based refinement. The experimental results demonstrate that the fuzzy link-based clustering ensemble methods developed herein perform better than their counterparts in terms of accuracy.

The remainder of this chapter is organised as follows. Section 7.1 introduces the basics of pairwise similarity matrices as a type of consensus function for clustering ensemble. Section 7.2 defines fuzzy co-association matrix and link-based pairwise similarity matrices, and presents their applications to agglomerative clustering in an attempt to create ensembles of fuzzy clusters. Section 7.3 reports on the experimental evaluation of the proposed approach and discuss the results. Finally, Section 7.4 concludes this chapter.

7.1 Preliminaries

Apart from the consensus functions described in Section 6.1, pairwise similarity matrices form another type of consensus method which creates a matrix, containing the pairwise similarity among data points. Any similarity-based clustering algorithm (e.g., hierarchical clustering) can be applied on such a pairwise similarity matrix to form the final clusters. There have been various approaches for this [73, 99]. Using the co-association (CO) matrix [72, 73] as an example: given a set X of N data points and let $\Pi = \{\pi_1, \dots, \pi_m\}$ be m base-clustering members, the functionality of each base-clustering member $\pi_i \in \Pi, i = 1, \dots, m$ is equivalent to transferring the data into an $N \times N$ similarity matrix, using Equation (7.1):

$$S_i(x_a, x_b) = \begin{cases} 1, & \text{if } C^i(x_a) = C^i(x_b) \\ 0, & \text{otherwise.} \end{cases} \quad (7.1)$$

where $C^i(x)$ denotes the cluster label to which the data point x belongs in π_i . Having obtained all the m similarity matrices regarding the base-clustering members, they are aggregated to form the so-called co-association matrix using Equation (7.2). An illustrative example of such a matrix is shown in Figure 7.1.

$$CO(x_a, x_b) = \frac{1}{m} \sum_{i=1}^m S_i(x_a, x_b). \quad (7.2)$$

The entries in a CO matrix therefore capture the similarities between data points x_a and x_b ($x_a, x_b \in X$).

With the co-association matrix used as the similarity measure between data points, the consensus partition can be obtained by applying similarity based clustering

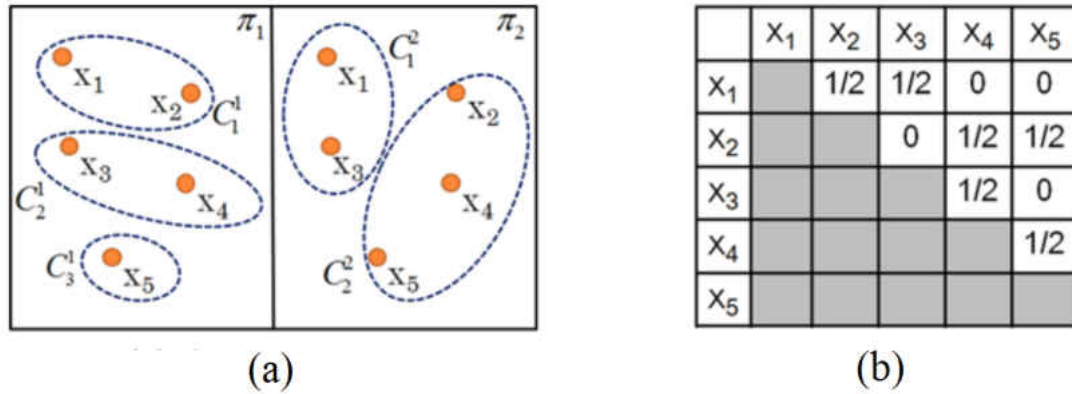


Figure 7.1: Example of Pairwise Similarity Matrices for Clustering Ensemble

algorithms. In [73], an algorithm is applied to find a minimum spanning tree after obtaining the co-association matrix. The co-association matrix is deemed as an adjacency matrix of a graph, a tree that contains all the nodes of the graph and the minimum weights in their edges is searched. Then, the weak links between nodes are cut with respect to a given threshold. This is equivalent to cutting the dendrogram produced by the single link agglomerative hierarchical clustering [105] using the threshold.

Based on the original co-association matrix, several modifications have been proposed to refine the similarity between data points for clustering ensemble. In [125], a new hierarchical clustering algorithm is applied to the co-association matrix to improve the quality of the consensus partition. This algorithm is based on the development of the concept of normalised edges to measure similarity between clusters.

In the definition of co-association matrix, Equation (7.1) takes only the values 0 or 1. The resultant new similarity between data points is computed only by taking into account whether the two objects belong to the same cluster or not [189]. The main drawback of using such a CO matrix is that many entries of it are zeros, which implies that two corresponding data points are assigned to different clusters by all base-clustering members. Investigations revealed that the zero-similarity values can be as much as 75% in some UCI datasets [99]. Unfortunately, this characteristic is commonly encountered with the crisp clustering ensembles, thereby significantly limiting the quality of the final data partition that is to be generated by any given consensus function [100].

To solve this problem, two similarity matrixes: Connected-Triple Based Similarity (CTS) and SimRank Based Similarity (SRS) are proposed in [99]. The CTS works on the basis that if two objects share a link with a third object, then this is indicative of similarity between those two data points. The SRS reflects the underlying assumption that neighbors are similar if their neighbors are similar as well. Also, [188] presented a weighted co-association matrix, which considers the similarity using the sizes of the clusters, the number of clusters in each partition and the original similarity values between the data points. Furthermore, [196] introduced the concept of probability accumulation matrix. These matrices take into account more information than the traditional co-association and can measure the pair-wise correlation between objects in higher resolution [189].

7.2 New Pairwise Similarity Matrices for Fuzzy Clustering Ensemble

The link-based refining methods proposed in [99] are herein extended for the ensemble of fuzzy clustering with an aim to refine the underlying sparse-information ensemble matrices. In particular, fuzzy c -means are employed to generate base-clustering members. This leads to the following CO matrix-based method which is named as FCO , and two link-based methods, named as $FLink$ and $FCTS$ respectively, for fuzzy clustering ensemble.

7.2.1 FCO : Co-association Matrix for Fuzzy c -means

Fuzzy c -means is an effective method to generate a fuzzy partition of a given data set. Each cluster in a partition $\tilde{\pi}_i$ is a fuzzy set $\tilde{C}_k^i, k = 1, \dots, K_i$ where $\tilde{C}_k^i(x_a) \in [0, 1]$ represents the degree of a data point $x_a \in X$ belongs to the corresponding fuzzy cluster. Usually, this degree is normalised with all the clusters in a partition to satisfy that $\sum_{k=1}^{K_i} \tilde{C}_k^i(x_a) = 1$.

Following the representational form used in crisp clustering ensemble (for notational consistency), the similarity measure of two objects $x_a, x_b \in X$ with respect to each base-clustering member, $S_{\tilde{i}}(x_a, x_b)$ and subsequently, the FCO matrix are defined in Equation (7.3) and Equation (7.4) respectively:

$$S_{\tilde{i}}(x_a, x_b) = \sum_{k=1}^{K_i} \min(\tilde{C}_k^i(x_a), \tilde{C}_k^i(x_b)) \quad (7.3)$$

$$FCO(x_a, x_b) = \frac{1}{m} \sum_{i=1}^m S_i(x_a, x_b). \quad (7.4)$$

Since $\sum_{k=1}^{K_i} \tilde{C}_k^i(x_a)$ is normalised to 1, it follows that $S_i(x_a, x_b) \in [0, 1]$ and $FCO(x_a, x_b) \in [0, 1]$. Note that Equation (7.3) is a generalised version of Equation (7.1). If the degree of a data point belongs to a crisp cluster is represented as $\tilde{C}_k^i(x_a) \in \{0, 1\}$, then Equation (7.3) can also be applied to crisp clustering ensemble equivalently.

One of the properties of fuzzy c -means is that most of the data points have non-zero memberships to many or even all clusters. This feature is very helpful for clustering ensemble helping to retain more details of the base-clustering members in the pairwise similarity matrix. Even two data points which are not assigned in the same cluster in crisp clustering can also have non-zero values in the FCO matrix with regard to the definition Equation (7.4). This gives potentially finer discrimination of the data points.

7.2.2 *FLink*: Link-based Pairwise Similarity Matrix for Fuzzy c -means

In clustering ensemble, base-clustering members are usually generated from the same dataset. Hence, the resulting base-clusters in a clustering ensemble may share common data points. These shared data points create the linkage amongst base-clusters and therefore, it is possible to estimate the similarity of any base-cluster pair by exploring the underlying link information. Note that the concept of a graph formulated from a set of base-clusters and a set of weighted links between them has been introduced previously, as of [99]. Given a clustering ensemble as defined in Section 6.1, a graph $\langle V, L \rangle$ can be constructed where $V = \bigcup_{i=1}^m \pi_i = \{C_1, \dots, C_M\}$, $M = \sum_{i=1}^m K_i$ is the set of vertices each representing a base-cluster, and L is a set of weighted links between the clusters. The weighted links between base-clusters C_a and C_b , $a, b = 1, \dots, M$ is defined as:

$$w(C_a, C_b) = \frac{|C_a \cap C_b|}{|C_a \cup C_b|} \quad (7.5)$$

where $|X|$ stands for the cardinality of a set X .

In crisp clustering ensemble, however, base-clusters within the same base-clustering member do not have common data points with each other, that is, $\forall C_k^i, C_{k'}^i \in \pi^i$, if

$k \neq k'$, then $C_k^i \cap C_{k'}^i = \emptyset$ for $k, k' = 1, \dots, K_i$. The weights of those links between the clusters within the same base-clustering member are of a value of zero. Further refinement will therefore be necessary before they can be used in the emerging ensemble. In order to retain more information from base-clustering members and refine the *FCO* matrix for fuzzy *c*-means ensembles, a fuzzy graph of fuzzy *c*-means ensemble is proposed.

Formally, given a set of fuzzy base-clusters $C = \{\tilde{C}_1, \dots, \tilde{C}_M\}$ on a dataset $\{x_1, \dots, x_N\}$, a fuzzy graph $\langle C, \tilde{L} \rangle$ is defined on the set of the fuzzy base-clusters where \tilde{L} is a fuzzy set of links defined on $C \times C$. The membership of a link between \tilde{C}_a and \tilde{C}_b , $a, b = 1, \dots, M$ to the fuzzy set \tilde{L} is computed by:

$$\tilde{L}(\tilde{C}_a, \tilde{C}_b) = \frac{\sum_{t=1}^N \min(\tilde{C}_a(x_t), \tilde{C}_b(x_t))}{\sum_{t=1}^N \max(\tilde{C}_a(x_t), \tilde{C}_b(x_t))} \quad (7.6)$$

where $\tilde{C}_a(x_t)$ indicates the degree of a data point x_t belonging to a fuzzy cluster \tilde{C}_a . Obviously, $\tilde{L}(\tilde{C}_a, \tilde{C}_b) \in [0, 1]$, $\tilde{L}(\tilde{C}_a, \tilde{C}_a) = 1$ and $\tilde{L}(\tilde{C}_a, \tilde{C}_b) = \tilde{L}(\tilde{C}_b, \tilde{C}_a)$. The degree assigned to the link connecting fuzzy clusters \tilde{C}_a and \tilde{C}_b is thus defined in accordance with the proportion of their overlapping degree on all data points in X . In so doing, even for two fuzzy base-clusters within the same base-clustering member, the weight of the link between them is possible to be of a non-zero value. As such, in general, each base-cluster may have a link to all the other base-clusters, and the fuzzy degree of a given link represents the similarity between the corresponding two base-clusters.

Given a fuzzy graph, link-based pairwise similarity matrix of data points can be introduced using the fuzzy weights associated with the links. In particular, for a clustering member $\tilde{\pi}_i$, the link-based similarity of data points x_a and x_b , $a, b = 1, \dots, N$ can be estimated by:

$$LS_{\tilde{\pi}_i}(x_a, x_b) = \begin{cases} 1, & \text{if } a = b \\ \tilde{L}(\arg \tilde{C}_{\max}^i(x_a), \arg \tilde{C}_{\max}^i(x_b)) \times \min(\tilde{C}_{\max}^i(x_a), \tilde{C}_{\max}^i(x_b)), & \text{otherwise} \end{cases} \quad (7.7)$$

where $\tilde{C}_{\max}^i(x_a) = \max_{k=1, \dots, K_i} \tilde{C}_k^i(x_a)$ and $\arg \tilde{C}_{\max}^i(x_a) \in \pi_i$ representing the fuzzy cluster in which x_a has the maximum membership. In case of a draw, a random pick is made amongst those even clusters. From this, it has a natural appeal to define the similarity of two data points in the overall fuzzy *c*-means clustering ensemble as:

$$FLink(x_a, x_b) = \frac{\sum_{i=1}^m LS_{\tilde{\pi}_i}(x_a, x_b)}{m}. \quad (7.8)$$

Different from *FCO*, the link-based similarity defined in Equation (7.7) only associates a data point x_a to the cluster of which x_a has the maximum membership degree. If two data points happen to have the maximum degrees in the same cluster, then their similarity values assigned by $LS_{\tilde{L}}$ is deemed to be the smaller degree value of the two, since $\tilde{L}(\tilde{C}_a, \tilde{C}_a) = 1, a = 1, \dots, M$. Otherwise, the link-based similarity of two data points x_a and x_b is defined as the smaller value of their respective maximum degrees times the weight of the link between those two base-clusters where x_a and x_b have the maximum degree values.

Note that non-zero weighted links may exist not only between base-clusters within a single base-clustering member, e.g., $\exists \tilde{L}(\tilde{C}_k^i, \tilde{C}_{k'}^i) > 0$, but also between base-clusters cross base-clustering members, e.g., $\exists \tilde{L}(\tilde{C}_k^i, \tilde{C}_{k'}^j) > 0$ for $i \neq j$ and $i, j = 1, \dots, m$. As $LS_{\tilde{L}}$ does not employ links across base-clustering members, it can be computed efficiently in terms of both time and memory space required. However, in crisp clustering ensemble, links across base-clustering members are employed to estimate the similarity within base-clustering members using means such as the connected-triple [113], thereby improving the quality of the final ensemble result. Inspired by this observation, and to test whether the across links may indeed help refine *FLink* further while allowing for consistent comparison with link-based crisp clustering ensemble, the connected-triple is also applied to \tilde{L} in the present work as will be described next.

7.2.3 *FCTS*: Connected-triple-based Pairwise Similarity Matrix for Fuzzy *c*-means

The connected-triple approach has been used in a bibliographic dataset to assess the similarity amongst author names and identify possible duplicates [113]. It assumes that if two nodes are both connected to a third node then it is indicative of similarity between those two nodes. The connected-triple is also applied to the weighted crisp clustering ensemble graph $\langle V, L \rangle$ of Equation (7.5) to generate the similarity of nodes within clustering members [99]. Specifically, the weighted connected-triple deems the similarity of two base-clusters C_a and C_b , $a, b = 1, \dots, M$ as the sum of the minimum weights to every common neighbour of theirs:

$$w'(C_a, C_b) = \sum_{t=1}^M \min(w(C_a, C_t), w(C_b, C_t)) \quad (7.9)$$

where $M = \sum_{i=1}^m K_i$ denotes the total number of base-clusters of all base-clustering members. $w'(C_a, C_b)$ may also be normalised such that:

$$n_w'(C_a, C_b) = \frac{w'(C_a, C_b)}{w'_{\max}} \quad (7.10)$$

where w'_{\max} is the maximum $w'(C_a, C_b)$ value of any two base-clusters C_a and C_b . Having obtained this, the similarity of two data points x_a and x_b , $a, b = 1, \dots, N$ with base-clustering member C^i , $i = 1, \dots, m$ can be defined by:

$$S'_i(x_a, x_b) = \begin{cases} 1, & \text{if } C^i(x_a) = C^i(x_b) \\ n_w'(C^i(x_a), C^i(x_b)) \times DC, & \text{otherwise} \end{cases} \quad (7.11)$$

where $DC \in [0, 1]$ is a constant decay factor. The connected-triple-based similarity matrix for base-clusters is defined the same as Equation (7.2):

$$CTS(x_a, x_b) = \frac{\sum_{i=1}^m S'_i(x_a, x_b)}{m}. \quad (7.12)$$

In a similar way, the fuzzy version of CTS can be introduced, where $\tilde{L}(\tilde{C}_a, \tilde{C}_b)$, $a, b = 1, \dots, M$ is refined using the connected-triple to become:

$$L'(\tilde{C}_a, \tilde{C}_b) = \sum_{t=1}^M \min(\tilde{L}(\tilde{C}_a, \tilde{C}_t), \tilde{L}(\tilde{C}_b, \tilde{C}_t)) \quad (7.13)$$

and then normalised to:

$$n_L'(\tilde{C}_a, \tilde{C}_b) = \frac{L'(\tilde{C}_a, \tilde{C}_b)}{L'_{\max}} \quad (7.14)$$

where L'_{\max} is the maximum $L'(\tilde{C}_a, \tilde{C}_b)$ value of any two fuzzy base-clusters \tilde{C}_a and \tilde{C}_b . Therefore, the similarity of two data points x_a and x_b , $a, b = 1, \dots, N$ with base-clustering member \tilde{C}^i can be modified to:

$$LS'_i(x_a, x_b) = \begin{cases} 1, & \text{if } a = b \\ L'(\arg \tilde{C}_{\max}^i(x_a), \arg \tilde{C}_{\max}^i(x_b)) \times \min(\tilde{C}_{\max}^i(x_a), \tilde{C}_{\max}^i(x_b)), & \text{otherwise} \end{cases} \quad (7.15)$$

where $\tilde{C}_{\max}^m(x_a) = \max_{k=1, \dots, K_i} \tilde{C}_k^i(x_a)$ and $\arg \tilde{C}_{\max}^i(x_a) \in \pi_i$ represents the fuzzy cluster of which x_a has the maximum membership. As before, if a draw incurs, one of those even clusters is randomly taken. The similarity of two data points in the overall fuzzy c -means clustering ensemble is computed by:

$$FCTS(x_a, x_b) = \frac{\sum_{i=1}^m LS'_i(x_a, x_b)}{m}. \quad (7.16)$$

Besides the CTS , other link-based methods such as the SimRank-based similarity matrix [99] can also be modified to ensemble fuzzy c -means clustering members. However, in spite of their effectiveness, the implementation of link-based similarity methods (even the $FCTS$) similarly suffer from high computational time requirements. This drawback originates within the algorithms, whose simplified variation may not be able to maintain the original performance [101]. Hence, the $FCTS$, which requires less computational time compared with other link-based methods is presented in this chapter.

7.2.4 Fuzzy c -means Ensemble based on Similarity Matrix

The overall process of using the proposed matrices in building clustering ensembles is similar to that of the existing work that uses pairwise similarity matrices (e.g., [80]). To save space only the two main steps are outlined below:

1. Fuzzy c -means are used on the dataset X for M times to generate fuzzy base-clusters. The diversity of base-clustering members is ensured by a combination of re-sampling the original datasets, different numbers of learned clusters, and different initial centroids for fuzzy c -means. Note that in theory, many other methods used in crisp clustering ensemble can also be used in place of fuzzy c -means ensemble, though the current work only uses the latter for simplicity.
2. Any of the three proposed methods (FCO , $FLink$, $FCTS$) can be used to generate a pairwise similarity matrix of data points, exploiting the information embedded in base-clustering members. From this, a pairwise similarity based clustering algorithm, such as hierarchical clustering, can then be employed to generate the final partition of the dataset as the output of clustering ensemble.

7.3 Experimentation

This section presents an experimental evaluation of the proposed work. It first outlines the setup of the experiments carried out and then discusses the results obtained. One experiment is designed to test the trend of accuracy when the diversity of base-clustering members is changed, and the other to compare the performances of different methods.

7.3.1 Experimental Setup

To evaluate the performance of proposed methods, they are experimentally tested over six datasets obtained from UCI benchmark repository [71], where true labels of instances are known but are not explicitly used in the clustering ensemble learning process. The details of these datasets are summarised in Table 7.1. The final results of the resulting clustering ensembles are evaluated in terms of accuracy as the group truth for each dataset is known.

Table 7.1: Summary of Datasets: Link-based Fuzzy Clustering Ensemble

Datasets	Instances	Attributes	Classes
Iris	150	4	3
Wine	178	13	3
Parkinsons	195	22	2
Glass (Identification)	214	9	6
Ecoli	336	7	8
Ionosphere	351	34	2

The fuzzy c -means clustering algorithm is used to generate the base-clustering members. Thirty clustering-members are created ($m = 30$) and the cluster centroids are randomly initialised in each run. Two agglomerative clustering approaches (complete-linkage and average-linkage) [80] are selected to implement the consensus function. These consensus functions divide data points into clusters using the underlying similarity matrix $FLink$, FCO , $FCTS$, or CTS . For a fair comparison, the number of final clusters on each dataset is set to that of its true classes and the decay factor (DC) of CTS is commonly set to 0.5 [99], and the base-clustering results used in CTS are defuzzified from the base fuzzy c -means used in the other three fuzzy methods.

7.3.2 Results and Discussion

Sensitivity of proposed methods

This is to check the robustness of the approach against the diversity of base-clustering members. To vary the base-clustering members, the maximum number of base-clusters K_{\max} in each test is set from 3 to 30 with an increment step of 3, and the number of base-clusters in each clustering-member K_i is randomly chosen from $[3, K_{\max}]$. Figure 7.2 shows the change of accuracy with respect to the increase of diversity in base-clustering members where agglomerative clustering with average-linkage is used as the consensus function. Each point in Figure 7.2 is an averaged value of 50 runs.

For five of the six datasets, the accuracies of the three proposed methods (*FLink*, *FCTS* and *CTS*) generally increase along with the increase of diversity. This indicates that the use of link-based pairwise similarity matrices in fuzzy *c*-means ensemble entails more differences in base-clustering members, which in turn allows the generation of better results. The outcome of using *FCO* seems to be more stable as compared with link-based methods. This indicates that *FCO* is not sensitive to the number of clusters in each base-clustering member. An intuitive explanation is that in fuzzy *c*-means, each data point has gained a certain membership to all the clusters. Thus, the base-clustering members which have a smaller number of clusters can retain as much information as the ones of a larger cluster number. However, the accuracy of *FCO* is not so high as that achievable by the link-based methods in general. This shows that although fuzzy *c*-means can help *FCO* to keep more information for building ensembles, the link-based refinements are helpful in generating more effective pairwise similarity matrices.

Accuracy comparison between link-based methods

This is to further analyse the results achievable by the link-based methods, using a fixed number ($K_i = \lceil \sqrt{N} \rceil$) or a random number ($K_i \in [3, \lceil \sqrt{N} \rceil]$) of clusters in each base-clustering member. The resultant accuracies are shown in Tables 3 and 4 respectively, where each number in these tables is an averaged value based on 50 runs. To validate the significance of the experimental results, the paired-t tests are carried out between *FLink* and the rest on each dataset. In each “pair” of results,

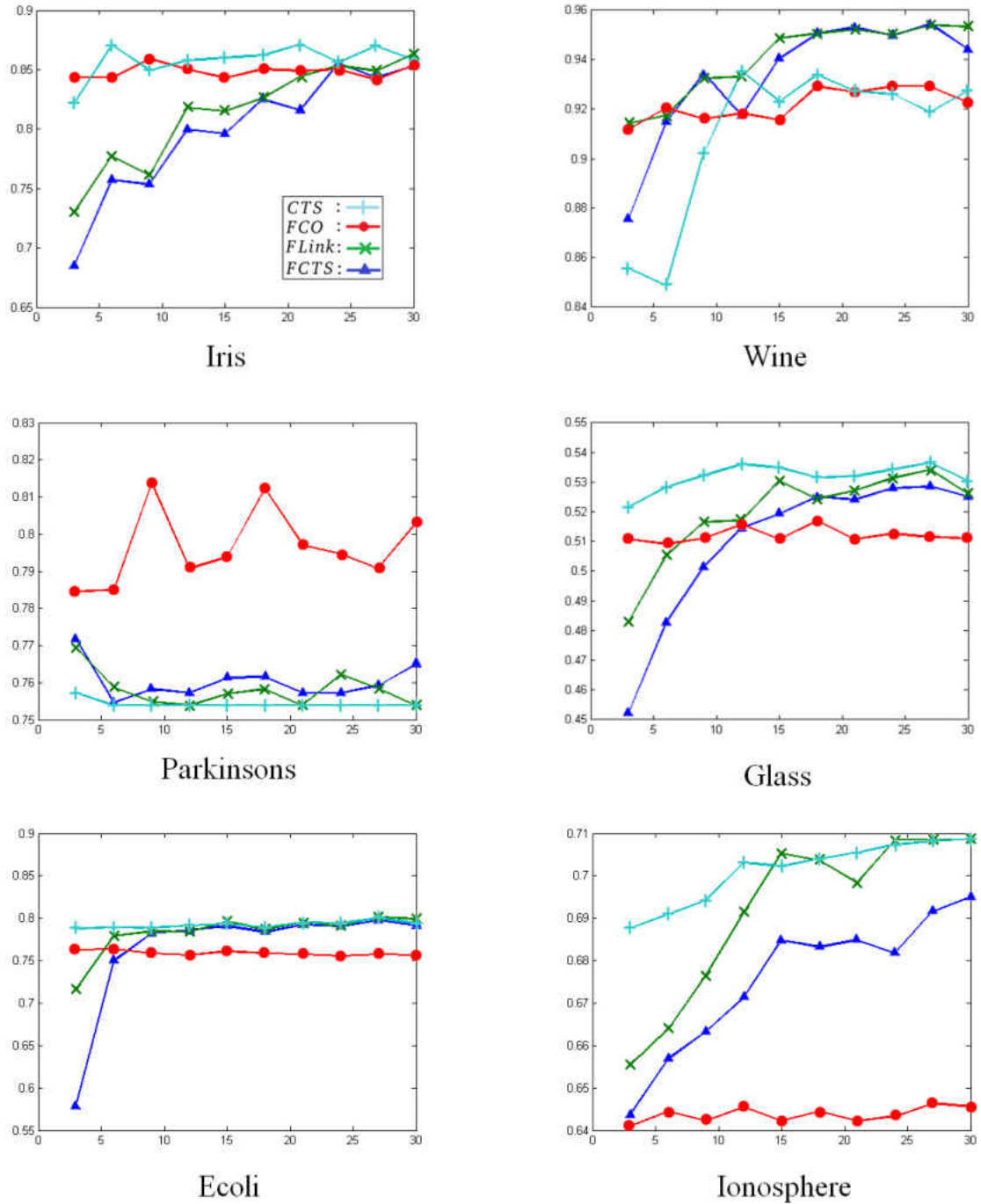


Figure 7.2: Trend of Accuracy against Diversity

Table 7.2: Comparison of Accuracy (%): Fixed Cluster Number. The best-2 results on each dataset are highlighted in boldface, the sign “(*)/(v)” indicates that the corresponding result is significantly ($p < 0.05$) worse/better than that of *FLink*

	Complete-link			Average-link				
	<i>FLink</i>	<i>FCO</i>	<i>FCTS</i>	<i>CTS</i>	<i>FLink</i>	<i>FCO</i>	<i>FCTS</i>	<i>CTS</i>
Iris	86.36	87.60 (v)	80.97(*)	71.35(*)	77.53	80.91 (v)	67.20(*)	76.80
Wine	94.51	91.58(*)	94.45	80.75(*)	94.45	90.31(*)	94.44	71.99(*)
Parkinsons	81.92	81.54(*)	81.92	76.18(*)	81.92	75.38(*)	82.05	80.58(*)
Glass	48.25	45.37(*)	48.31	52.60 (v)	51.31	47.79(*)	49.28(*)	57.34 (v)
Ecoli	79.53	76.15(*)	79.90 (v)	75.29(*)	82.86	64.99(*)	83.58 (v)	77.29(*)
Ionosphere	64.10	64.10	64.10	64.10	64.10	64.10	64.10	64.10
Means	75.7783	74.3900	74.9417	70.0450	75.3617	70.5800	73.4417	71.3500

Table 7.3: Comparison of Accuracy (%): Random Cluster Number. The best-2 results on each dataset are highlighted in boldface, the sign “(*)/(v)” indicates that the corresponding result is significantly ($p < 0.05$) worse/better than that of *FLink*

	Complete-link			Average-link				
	<i>FLink</i>	<i>FCO</i>	<i>FCTS</i>	<i>CTS</i>	<i>FLink</i>	<i>FCO</i>	<i>FCTS</i>	<i>CTS</i>
Iris	86.21	85.52	85.73	85.51	85.12	85.03	84.81	86.15
Wine	95.16	91.40(*)	95.13	86.58(*)	95.33	93.02(*)	95.38	91.72(*)
Parkinsons	75.84	76.27	76.28 (v)	75.38(*)	75.84	79.57 (v)	76.45	75.38(*)
Glass	52.83	51.20(*)	52.53	52.74	52.98	51.33(*)	52.84	53.53 (v)
Ecoli	78.98	77.11(*)	79.33	76.92(*)	79.24	75.89(*)	78.95	78.86
Ionosphere	68.43	68.17	67.35	70.51 (v)	70.83	64.39(*)	69.38(*)	70.79
Means	76.2417	74.9450	76.0583	74.6067	76.5567	74.8717	76.3017	76.0717

the generation of base-clustering members is based on the same number of clusters and same initialisation centroids.

The results show that for both fixed and random K_i , the use of link-based pairwise similarity matrix *FLink* leads to the best average accuracy over the six datasets, in building fuzzy c -means ensembles. However, the performance of *FCTS* is not significantly better than *FLink* in general. This implies that the connected-tripe method does not necessarily further refine *FLink* effectively. Note that both *FLink* and *FCTS* achieve a better accuracy than *CTS* on most of the datasets. Although the *CTS* which employs the connected-triple to infer the similarities amongst clusters within each base-clustering member, it seems that the inferred similarities are not as effective as those generated by the fuzzy links \tilde{L} in *FLink* and *FCTS*. Particularly, *FLink* can use the fuzzy links $\tilde{L}(\tilde{C}_k^i, \tilde{C}_{k'}^i)$ where $k, k' = 1, \dots, K_i$ directly without inferring them from $\tilde{L}(\tilde{C}_k^i, \tilde{C}_{k'}^j), i \neq j$, the time for running the connected-triple method (or the other similar refinement) is saved. In conclusion, *FLink* entails higher accuracy but lower time-consumption than *CTS*.

7.4 Summary

This chapter has presented the notion of link-based pairwise similarity matrices for fuzzy c -means ensemble. The proposed methods are based on the co-association matrix and its link-based refinements for crisp clustering ensemble. The proposed matrices take the advantage of fuzzy c -means in that each data point can have a membership to all clusters. A set of fuzzy links between base-clusters is defined and a fuzzy graph is employed to generate the link-based similarity matrices. Experimental results on six UCI datasets indicate that the proposed approach generally outperforms the conventional *CTS*. Furthermore, the link-based methods also help to build better pairwise similarity matrices as compared to the non-link based matrix *FCO*.

Chapter 8

Computational Considerations of Fuzzy Clustering Ensemble

CLUSTERING ensembles organically integrate individual component methods which may utilise different parameter settings and features, and which may themselves be generated on the basis of different representations and learning mechanisms. Such a technique offers an effective means for aggregating multiple clustering results in order to improve the overall clustering accuracy and robustness. To reinforce such development, this chapter presents another clustering ensemble approach for fuzzy clustering, with an aim to be applied for clustering of big data. The proposed algorithm first generates fuzzy base-clusters with respect to each data feature and then, employs a fuzzy hierarchical graph to represent the relationships between the resulting base-clusters. Whilst the work employs fuzzy *c*-means and hierarchical clustering in generating base-cluster and implementing consensus function respectively, when applied to large datasets it has lower time complexity than the original fuzzy *c*-means and hierarchical clustering. The resultant ensemble clustering mechanism is tested against traditional clustering methods on various benchmark datasets. Experimental results demonstrate that it generally outperforms crisp clustering ensembles and single linkage agglomerative clustering, in terms of accuracy in conjunction with time efficiency, thereby showing that it has the potential for application in clustering big data.

The rest of this chapter is organised as follows. Section 8.1 introduces the background of the clustering algorithms designed for big data. Section 8.2 presents

the proposed Hierarchical Fuzzy Clustering Ensemble (HFCE) approach in detail, including a discussion of its advantages over the existing clustering ensemble methods. In addition to the theoretical development of this fuzzy clustering ensemble approach, its potential application to clustering big data is also introduced. In Section 8.3, the potential to exploit this method as a clustering tool for big data through parallel implementation is described. Section 8.4 reports on the experimental set up and analyses the results. The chapter is summarised in Section 8.5.

8.1 Background

Dealing with big data has become inevitable in many real-world problems. Recently, a new trend of and indeed challenge for data mining has arisen with the exponential growth and also availability of large amount of complex data. Applying conventional data mining techniques directly to big data is difficult or even impossible due to its intolerable computational time. Besides, the high dimensional and multi-model features may degrade the performance of conventional learning algorithms [194]. A number of research directions have been proposed in the literature to overcome such difficulties, including re-sampling data and distributing or parallelising conventional algorithms [132].

Clustering algorithms have emerged as an alternative powerful meta-learning tool to accurately analyse the massive volume of data generated by modern applications [62]. In general, the task of clustering is to assign objects to groups (namely clusters) such that data points in the same group are similar to each other, and dissimilar to those in the other clusters [105]. A good number of clustering algorithms have been proposed in the literature, and successfully applied to a range of problems [1, 29, 129, 174]. However, clustering big data is more challenging than dealing with traditional data modelling and analysis problems. Many of the existing clustering methods such as the k -means and fuzzy c -means are NP-hard, and hence they are very time consuming for handling big data.

To tackle the aforementioned problem feasible techniques have been proposed. Most of which work by extending the existing approaches (that have been developed for non big data) through analysing a selected, manageable amount of samples of the original data [197] and then, exploiting the sample-based modelling results to derive a partition for the overall data. These methods differ usually only in terms of how the

sample-based analysis is carried out, including the CLARA algorithm [111] and the CLARANS algorithm [140]. Clustering big data has also led to distributed and parallel implementations [145]. One such approach is to directly extend existing clustering methods by taking advantage of distributed network environments in which the overall computation effort is shared by collaborative computing facilities [225]. Whilst these promising results have been reported, much remains to be done in order to have a more efficient and effective fuzzy clustering approach that is suitable for big data. Inspired by this observation, a hierarchical fuzzy clustering ensemble (HFCE) method (which is applicable for distributed and parallel application) is proposed in this chapter.

8.2 Feature-based Hierarchical Fuzzy Clustering Ensemble

Apart from conventional clustering algorithms whose outputs are hard partitions of data, there are alternative approaches such as EM [49] and fuzzy c -means [21], which generate soft or fuzzy partitions of data with a natural appeal. In order to take advantage of the aforementioned ensemble techniques over fuzzy clustering, an additional “hardening” process would be required for the fuzzy cluster assignments. This process may result in loss of information that is conveyed by the uncertainty measures of the relevant cluster assignments. This is particularly true for application settings where the underlying clustering algorithms access only a partial view of the data, such as in distributed data mining [156]. Yet, most of the existing clustering ensemble methods are based on crisp clusters. However, interesting departures from such work have recently been reported, including sCSPA, sMCLA and sHBGF (which are the fuzzy versions of the graph/hypergraph-based algorithms CSPA, MCLA and HBGF, respectively) [156].

Following such recent development, this section presents a link-based hierarchical consensus-based approach for building ensembles of fuzzy c -means. The proposed hierarchical clustering ensemble algorithm starts by creating ensemble members using fuzzy c -means on each feature in dataset. The resulting fuzzy base-clusters and the links between them are represented in a fuzzy graph. The idea of hierarchical clustering is then employed to iteratively group the nodes based on the fuzzy links, in order to create the hierarchical structure that leads to the final clusters. It also

yields instance-wise fuzzy cluster membership estimation, which may be defuzzified such that each data point belongs to just one final cluster if required. The following details the key operations of this algorithm.

8.2.1 Feature-based Generation of Ensemble Members

In general, no constraints are necessarily imposed over the generation procedure through which the clustering partitions are obtained. In fact, different (component) clustering algorithms or the same algorithm with different parameter settings can be applied. If so desired, even different representations for data points, different subsets of data points or their projections on different subspaces may be used [67, 215]. To perform a conventional clustering ensemble task, a moderate-sized dataset can be clustered several times in order to obtain ensemble members.

In dealing with big data, however, the computational overheads of running a single clustering procedure on a complete dataset may already be intolerable, multiple executions of clustering on the whole dataset are impractical. Reducing the complexity of dataset for each clustering member offers a reasonable way of solving this problem. In this work, an m -dimensional dataset is divided into M one-dimensional subsets, and m times of fuzzy c -means are carried out on those one-dimensional subsets.

In practice, projecting data onto different subspaces or choosing different subsets of features may lose information (e.g., correlations between features) which can be important to detect the underlying patterns of the data [106]. Unfortunately, this is also true in the proposed method where the qualities of the resulting individual ensemble members are generally not so good as those created from the direct use of all accessible features. Despite this observation, as demonstrated in the existing ensemble learning frameworks [199, 221], relatively weak component results are still commonly used. Whilst individual ensemble members may be simple, if jointly utilised in conjunction with an appropriate consensus function, weak base-clusters are capable of producing high quality ensemble results.

Such an individual feature-based partition strategy however, has further limitations in dealing with datasets that have redundant or interactive features. If the features are redundant, HFCE may produce redundant base-clusters accordingly. Also, if certain individual features are interactive with each other, the useful information embedded in the interactions will be lost. A possible approach to solving these

problems is to use feature selection or grouping techniques [52, 222, 224] in guiding the partition of the original data. However, the assessment of any redundancy and interaction amongst features incurs additional computational cost in the generation of base-clusters and therefore, improvement over these issues can be rather time-consuming when dealing with datasets with a high dimensionality. Thus, in the current design and implementation of HFCE, advanced feature partition strategies are sacrificed to compensate for its execution speed. Finding a rapid feature partition algorithm to support HFCE remains active as further research.

It may be difficult to know a-priori which base-clustering algorithm(s) will be appropriate for a given clustering problem. It is generally advisable and also, a common practice to employ those clustering algorithms that are known to be able to reflect and make use of most information embedded in the data. This is obvious as the more information each clustering member holds, the more information there is for the consensus function to work on. Based on this understanding, fuzzy c -means, which is able to retain the non-binary memberships of each data point to all clusters is adopted as the base algorithm for the generation of ensemble members in this work.

8.2.2 Similarity between Fuzzy Base-clusters

In the above proposed strategy for ensemble member generation the base-clusters are created by partitioning the dataset with respect to different individual features. However, all data points used come from the same original dataset. As such, the resulting base-clusters may share certain points. These shared data points naturally create linkages amongst base-clusters and therefore, it is possible to estimate the similarity of any base-cluster pair by exploring the underlying link information [224].

Note that the concept of a graph formulated from a set of base-clusters and a set of weighted links between them has been introduced previously, as in Section 7.2.2. In order to retain more information from fuzzy clustering components and reflect the interactions between different features which are embedded in the original dataset, a fuzzy graph of fuzzy c -means ensemble is employed here. Given a clustering ensemble, a graph $\langle C, \tilde{L} \rangle$ can be constructed where $C = \bigcup_{i=1}^m \pi_i = \{C_1, \dots, C_M\}$, $M = \sum_{i=1}^m K_i$ is the set of vertices each representing a base-cluster, and \tilde{L} is a fuzzy set of links between the clusters. The membership of a

link $(\tilde{C}_a, \tilde{C}_b)$, $a, b = 1, \dots, M$ to the fuzzy set that represents the fuzzy relation \tilde{L} is defined as given in Equation (7.6):

$$\mu_{\tilde{L}}(\tilde{C}_a, \tilde{C}_b) = \frac{\sum_{t=1}^N \min(\tilde{C}_a(x_t), \tilde{C}_b(x_t))}{\sum_{t=1}^N \max(\tilde{C}_a(x_t), \tilde{C}_b(x_t))}$$

Therefore, each fuzzy base-cluster can be represented as a node and the membership of a link to \tilde{L} can be represented as the strength or weight of that link. Obviously, $\mu_{\tilde{L}}(\tilde{C}_a, \tilde{C}_b) \in [0, 1]$, $\mu_{\tilde{L}}(\tilde{C}_a, \tilde{C}_a) = 1$ and $\mu_{\tilde{L}}(\tilde{C}_a, \tilde{C}_b) = \mu_{\tilde{L}}(\tilde{C}_b, \tilde{C}_a)$. The degree assigned to the link connecting fuzzy clusters \tilde{C}_a and \tilde{C}_b is thus defined in accordance with the proportion of their overlapping degree on all data points in X . In so doing, for two fuzzy base-clusters within the same ensemble member, the weight on the link between them is possible to be of a non-zero value. As such, in general, each fuzzy base-cluster may have a link to all the other fuzzy base-clusters.

The defined fuzzy degree of each given link intuitively captures the underlying similarity between the corresponding two fuzzy base-clusters. These fuzzy links are of particular significance in this work. Since the ensemble members are generated from one-dimensional subsets, information on the interactions or correlations between features is lost in compromise with the gain of computing time. However, by employing a link-based consensus function that makes use of similarities between base-clusters, such information can be (re-)captured.

The similarities between base-clusters carry the information of how close they are to one another, and this information is useful to merge redundant base-clusters. In crisp clustering ensembles, further refinement will have to be carried out in an effort to estimate the similarities between base-clusters within the same ensemble member. As indicated in Section 7.2.3, connected-triple links cross ensemble members are computed. Fortunately, in fuzzy c -means ensemble, non-zero weighted links exist not only between those base-clusters within a single base-clustering member, but also between base-clusters cross different base-clustering members. Since no additional refinement is needed (as otherwise needed for the crisp case), the similarity measures can be readily computed, making significant savings in time and memory space.

8.2.3 Base-cluster Grouping via Hierarchical Clustering

In this step, fuzzy base-clusters are grouped into a certain number of final clusters to become the output of the ensemble. Fuzzy base-clusters are artificially treated as

data instances and those original data points given in the dataset that belong to a base-cluster are regarded as a feature for the artificial data instance. In other words, the “instance-cluster matrix” in Table 6.3 is transposed to a “cluster-instance matrix”, which generally speaking, has a large number of features but a relatively smaller number of instances involved than the real instance-cluster matrix.

In the existing work on fuzzy clustering ensembles, each base-cluster maintains the non-binary membership values of all those data points belonging to it. This makes fuzzy base-clusters more informative but more storage-consuming than their crisp counterparts. Fortunately, link-based clustering approaches such as single-linkage clustering do not need to re-access the original memberships of the data points once the similarity matrix is obtained, making them less sensitive to high-dimensional data. This is important in an effort to deal with the grouping of base-clusters, since otherwise iteratively visiting fuzzy base-clusters directly can be very time-consuming, if not prohibitive.

By using single-linkage hierarchical clustering algorithm, grouping fuzzy base-clusters can be achieved without the need of updating the cluster centroids. A matrix $L = [l(a, b)]_{M \times M}$ can be constructed with the indices of its rows and columns representing the indices of base-clusters, and each entry $l(a, b), a, b = 1, \dots, M$, of the matrix representing the similarity value of the corresponding base-clusters, e.g., $\mu_{\tilde{L}}(\tilde{C}_a, \tilde{C}_b)$. From this, the grouping of the base-clusters depends upon L only, rather than upon the memberships of the original data points belonging to these base-clusters. The subsequent steps of single-linkage clustering are quite simple after the similarities amongst base-clusters are known: Applying a simple sorting procedure over those similarity values and then using a threshold or a given number of total clusters required to merge the clusters. Therefore, both the time and memory resources required for iteratively visiting instance-cluster matrix or the pair-wised similarity matrix are saved. Note that single-linkage clustering obtains exactly the same results by agglomerating small clusters into larger ones (bottom up) as by dividing larger clusters into smaller ones (top down) [84]. In the following, this method is referred to as single-linkage agglomerative clustering (SLAC).

Apart from SLAC, a multilevel scheme for partitioning irregular graphs, METIS [110] has been used to group base-clusters in sMCLA. However, METIS can only produce a given number of balanced groups of clusters. Since each group is required to contain the same number of base-clusters, this method is not suitable for use in

the present approach where one-dimensional individual feature-based partition of the dataset is assumed. For example, in the situation where the dataset contains certain outlier values in one feature, a base-cluster consists of only the outliers can be generated. Such a cluster should be grouped with itself rather than with any other normal ones. Therefore, algorithms which return absolutely balanced groups of clusters may damage the overall quality of the resultant cluster groups. In light of this observation and considering both the time complexity and the quality of clustering, SLAC is selected to group base-clusters in HFCE.

Recall the basic idea of hierarchical clustering, that is to build a tree of data clusters that are successively merged into similar groups, with each level of the resulting tree being a segmentation of the original data [107]. By applying SLAC to group fuzzy base-clusters rather than data points or crisp clusters, each level of the resulting tree comprises groups of base fuzzy clusters. In order to obtain the overall data partition by the ensemble, an additional step which transforms the groups of fuzzy base-clusters into the final fuzzy clusters of the original data points is needed.

8.2.4 Final Assignment of Data Points

At each level of the resultant hierarchical tree, all those base-clusters contained within a certain cluster-group are collapsed to form one single fuzzy cluster. At the leaf level of the tree each fuzzy base-cluster contains a membership value for every data point that is deemed to belong to the cluster. Such a membership for a given final fuzzy cluster is computed as the normalised mean of its memberships to all those base-clusters that are grouped together. To produce a crisp final partition of the original data, each point is assigned to the cluster group to which it has the highest membership. Note that if so desired, other aggregation operators rather than the average may also be employed to implement this of course.

Summarising the above development, the proposed hierarchical fuzzy clustering ensemble (HFCE) learning algorithm is given in Algorithm 8.2.1.

8.3 Initial Application to Big Data

Although much effort has been made in the development of clustering ensembles, the application of clustering ensemble techniques is still at an early stage [98]. Little

Inputs: $X = \{x_1, \dots, x_t, \dots, x_N\}, x_t = (a_1^t, \dots, a_m^t) \in \mathbb{R}^m$: a dataset of N instances and m features;
 $K_1, \dots, K_i, \dots, K_m$: number of base-clusters in each ensemble member;
 K : final number of clusters.
Outputs: $\tilde{\pi}^* = \{\tilde{C}_1^*, \dots, \tilde{C}_K^*\}$: a fuzzy partition of X .

- 1: **for** $i = 1 : m$ **do**
- 2: create sub-dataset $X_i = \{a_i^1, \dots, a_i^N\}$
- 3: create ensemble member $\tilde{\pi}_i = \{\tilde{C}_1^i, \dots, \tilde{C}_{K_i}^i\}$ using fuzzy c -means on X_i
- 4: **end for**
- 5: merge the ensemble members to create a set of fuzzy base-clusters
 $C = \{\tilde{C}_1, \dots, \tilde{C}_M\} = \bigcup_{i=1}^m \pi_i$,
where $M = \sum_{i=1}^m K_i$
- 6: **for** $a = 1 : n - 1$ **do**
- 7: **for** $b = i + 1 : n$ **do**
- 8: $\mu_{\tilde{L}}(\tilde{C}_a, \tilde{C}_b) = \frac{\sum_{t=1}^N \min(\tilde{C}_a(x_t), \tilde{C}_b(x_t))}{\sum_{t=1}^N \max(\tilde{C}_a(x_t), \tilde{C}_b(x_t))}$
- 9: **end for**
- 10: **end for**
- 11: create a partition $\pi^C = \{C_1^C, \dots, C_k^C, \dots, C_K^C\}$ on C based on \tilde{L} using hierarchical clustering
- 12: **for** $k = 1 : K$ **do**
- 13: $\mu'_{\tilde{C}_k^*}(x_t) = \text{average of } \{\tilde{C}_i(x_t) | \tilde{C}_i \in C_k^C\}$
- 14: **end for**
- 15: normalise $\mu'_{\tilde{C}_k^*}(x_t)$ to $\mu_{\tilde{C}_k^*}(x_t)$,
such that $\sum_{k=1}^K \mu_{\tilde{C}_k^*}(x_t) = 1$

Algorithm 8.2.1: Hierarchical Fuzzy Clustering Ensemble (HFCE)

has been successfully done for big data. This section proposes an initial idea as to how HFCE may be potentially utilised to handle big data clustering, based on an investigation into its time complexity.

For a dataset with N data points and m features, the time complexity of the original fuzzy c -means is $O(mNK)$ where K is the number of clusters [116]. The hierarchical agglomerative clustering has a time complexity of $O(N^2 \log N)$ [105]. Since the agglomerative clustering is employed only for grouping base-clusters in HFCE, its use leads to a complexity of $O(M^2 \log M)$, where $M = \sum_{i=1}^m K_i$ is the total number of fuzzy base-clusters generated. To calculate the final fuzzy partition the algorithm also involves an additional time complexity of $O(MNK)$.

For big data it may be difficult to expect that Algorithm 8.2.1 can be implemented on a single computer of moderate computational power. However, HFCE can be

implemented in a parallel way. Suppose that there are $P + 1$ computers within a certain parallel computer network. One computer acts as the host node, which is in charge of assigning tasks and collecting results [218], and the other P computers work as the real computing nodes. Then, such a parallel computation network can be applied to implement the three main components of HFCE, reducing its overall time complexity. The parallelisation is outlined below:

1) Generation of ensemble members: Each ensemble member is denoted by π_i , where $i = 1, \dots, m$ and each π_i corresponds to a certain feature of the dataset. If $P < m$, then the host node first assigns π_1, \dots, π_p to the P computing nodes. Once any of the P ensemble member is generated, the host node assigns the next component of $\pi_{p+1}, \pi_{p+2}, \dots$ in order to the free computing nodes. This process iterates until π_m is reached. If $P \geq m$, the assignment is straightforward. Since fuzzy c -means only takes a few iterations to converge on one-dimensional data, the burden of each computing node is very low.

2) Single-linkage-based grouping of fuzzy base-clusters: A number of parallel versions of SLAC have been proposed in the literature (e.g., [145, 162]). For simplicity, an intuitive parallel method (not the best time-saving one) is introduced here. As any pair of fuzzy base-clusters have in general, a fuzzy link between them, the membership values of all the links $\mu_{\tilde{C}_a, \tilde{C}_b}$ can be collectively represented as a pair-wise similarity matrix $L_{M \times M}$. Since this matrix is symmetric and the similarity of one element to itself is not considered in SLAC, $\frac{M(M-1)}{2}$ similarity values are needed to be computed. The host node decomposes the task of calculating these similarity values into the P computing nodes. Note that only a limited number of fuzzy base-clusters are generated for each feature, and that this number is generally much smaller than the number of data points. Thus, after $L_{M \times M}$ is obtained, the parallelisation of following steps in SLAC for base-cluster grouping may not be necessary, but optional for a powerful computing node.

3) Computation of the memberships of data points to the final clusters: similar to the working of step 1), the final clustering result can be represented as a matrix $A_{N \times K}$. Given that $N > P, p = \lfloor N/P \rfloor$, the host node can decompose $A_{N \times K}$ into P disjointed sub-matrices $A_{[1, \dots, p] \times K}, A_{[p+1, \dots, 2p] \times K}, \dots, A_{[p \times (P-1), \dots, N] \times K}$. The required membership calculations regarding the resulting P sub-matrices can then be assigned to the P computing nodes, respectively.

8.4 Experimentation and Evaluation

This section presents an experimental evaluation of the proposed work. It first outlines the setup of the experiments carried out and then discusses the results obtained. One experimentation is designed to test the quality of clusters which are generated using HFCE, in comparison to those produced by the use of alternative approaches, and another to show the time efficiency of HFCE in contrast with the original fuzzy c -means and single-linkage agglomerative clustering.

8.4.1 Experimental Setup

To evaluate the performance of the proposed approach, the algorithm is tested over nine datasets with continuous attributes [198] obtained from the UCI benchmark repository [71], where the underlying true labels of the data points are known (which are not explicitly used in the clustering ensemble learning process but in the computation of clustering accuracy). The details of these datasets are summarised in Table 8.1.

Table 8.1: Summary of Datasets: HFCE

Datasets	Instances	Attributes	Classes
Iris	150	4	3
Sonar	208	60	2
Statlog Heart	270	13	2
Parkinsons	195	22	2
Ionosphere	351	34	2
Pima Indians Diabetes	768	8	2
Yeast	1484	8	10
Statlog Landsat Satellite	4435	36	6
Spambase	4601	57	2

In HFCE, fuzzy c -means is used to implement fuzzy ensemble members. In each ensemble member, the number of base-clusters $K_i, i = 1, 2, \dots, m$ is set to the number of given classes of the dataset. The cluster centroids are randomly initialised in each run. The single-linkage clustering technique is selected to implement base-cluster

grouping. It organises base-clusters into a hierarchical tree using the underlying similarity matrix. For comparison, an ensemble of crisp clusters (HCCE) with a similar underlying mechanism to that of HFCE is also implemented. To have a common ground for this comparative study, the base-clusters used in HCCE are those defuzzified from the base fuzzy c -means used in HFCE and the number of final clusters on each dataset is again set to that of its true classes. The output of HFCE is defuzzified by assigning a data point to the cluster to which it reaches the maximum membership.

8.4.2 Clustering Quality

In order to gauge clustering quality, two types of criterion are usually employed, measuring how well a clustering partitions the given data into the underlying groupings, namely, the internal and external criteria [98]. In particular, the goodness of a clustering ensemble is estimated using the averaged Silhouette index which measures the compactness of resultant clusters without referring to the ground truth (internal). If however, the class labels are available for all the data involved in the experiments, the final clustering results can be evaluated using the accuracy which measures how well the clusters match the given true labels of the data points (external). In this experiment, the quality of the final clustering outcomes is also assessed using these two criteria.

The resultant averaged Silhouette index and clustering accuracy rates are shown in Tables 8.2 and 8.3 respectively, where each number in these tables is an averaged value based on 50 runs. In order to compare the ensemble-based clustering methods with the conventional clustering methods, the results of fuzzy c -means (FCMC) and single-linkage agglomerative clustering (SLAC) are also included. To validate the significance of the experimental results, paired-t tests are carried out. The baseline for comparison is the result of running HFCE.

Experimental results show that HFCE achieves better compactness and accuracy than HCCE on most of the datasets. This indicates that the information embedded in fuzzy base-clusters are more effective to generate final ensemble partitions than that embedded in crisp base-clusters. Also, the final clusters generated by HFCE lead to better accuracies than SLAC over seven datasets. However, the performance of HFCE is not significantly better than FCMC in general. A likely reason is that

Table 8.2: Comparison of Averaged Silhouette Index ($[-1, 1]$). The best-two results on each dataset are highlighted in boldface, the sign “(*)/(v)” indicates that the corresponding result is significantly ($p < 0.05$) worse/better than that of HFCE

Dataset	HFCE	HCCE	SLAC	FCMC
Iris	0.678±0.003	0.267±0.243(*)	0.700±0.000(v)	0.685±0.000(v)
Sonar	0.136±0.020	0.175±0.093(v)	0.446±0.000(v)	0.267±0.009(v)
Statlog Heart	0.184±0.000	0.204±0.093	0.261±0.000(v)	0.342±0.000(v)
Parkinsons	0.349±0.000	0.354±0.223	0.792±0.000(v)	0.417±0.000(v)
Ionosphere	0.230±0.000	0.219±0.086	0.496±0.000(v)	0.331±0.000(v)
Pima Indians Diabetes	0.469±0.000	0.283±0.081(*)	0.648±0.000(v)	0.381±0.012(*)
Yeast	0.032±0.036	-0.089±0.042(*)	0.640±0.000(v)	-0.048±0.034(*)
Statlog Landsat Satellite	0.289±0.054	-0.154±0.067(*)	-0.015±0.000(*)	0.471±0.000(v)
Spambase	0.644±0.000	0.052±0.028(*)	0.875±0.000(v)	0.140±0.007(*)

Table 8.3: Comparison of Accuracy (%): HFCE. The best-two results on each dataset are highlighted in boldface, the sign “(*)/(v)” indicates that the corresponding result is significantly ($p < 0.05$) worse/better than that of HFCE

Dataset	HFCE	HCCE	SLAC	FCMC
Iris	89.89±0.25	65.24±9.48(*)	66.67±0.00(*)	89.33±0.00(*)
Sonar	54.33±0.00	59.54±6.05(v)	53.37±0.00(*)	55.22±0.48(v)
Statlog Heart	67.41±0.00	60.93±6.33(*)	55.93±0.00(*)	79.26±0.00(v)
Parkinsons	75.38±0.00	75.38±0.00	75.38±0.00	75.38±0.00
Ionosphere	62.61±0.00	55.48±3.14(*)	50.87±0.00(*)	74.78±0.00(v)
Pima Indians Diabetes	65.10±0.00	68.64±3.21(v)	65.23±0.00(v)	66.67±0.00(v)
Yeast	41.96±3.59	33.66±1.28(*)	32.35±0.00(*)	43.01±0.13
Statlog Landsat Satellite	61.98±0.76	40.35±3.95(*)	24.28±0.00(*)	72.51±0.00(v)
Spambase	60.75±0.00	60.60±0.00(*)	60.62±0.00(*)	76.71±0.00(v)

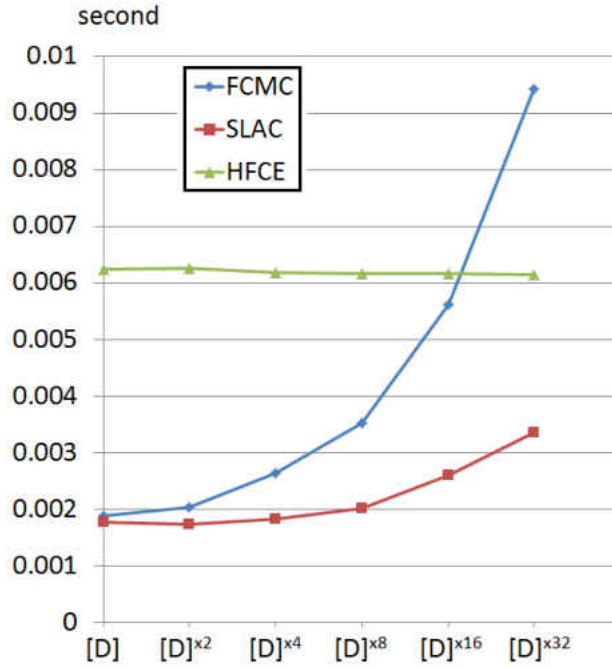


Figure 8.1: Time Cost vs. Increase of Features

the tested datasets are sensitive to the interaction of features, while the similarities amongst fuzzy base-clusters cannot completely capture the interaction of features in those datasets. Importantly, HFCE only employs ensemble members which generate base-clusters each involving just one feature.

8.4.3 Time Efficiency

This set of experiments is to empirically check the time efficiency of HFCE, as compared to that of the original SLAC and FCMC. To ensure that the results are easy to analyse, the simple iris dataset is used. However, to vary the scale of each experiment, the original dataset is artificially enlarged either horizontally (by duplicating features), or vertically instances (by duplicating data points), or both. In the following presentation, the original dataset is denoted as $[D]$, its horizontally double-sized dataset $[D \ D]$ is denoted as $[D]^{\times 2}$, and its vertically double-sized dataset $\begin{bmatrix} D \\ D \end{bmatrix}$ is denoted as $[D]_{\times 2}$, etc. The experiments are carried out on a computer with Intel(R) Core(TM)2 Duo 3.00 GHz \times 2 CPU, 4 GB RAM, and Windows 7 (64-bit) Operating

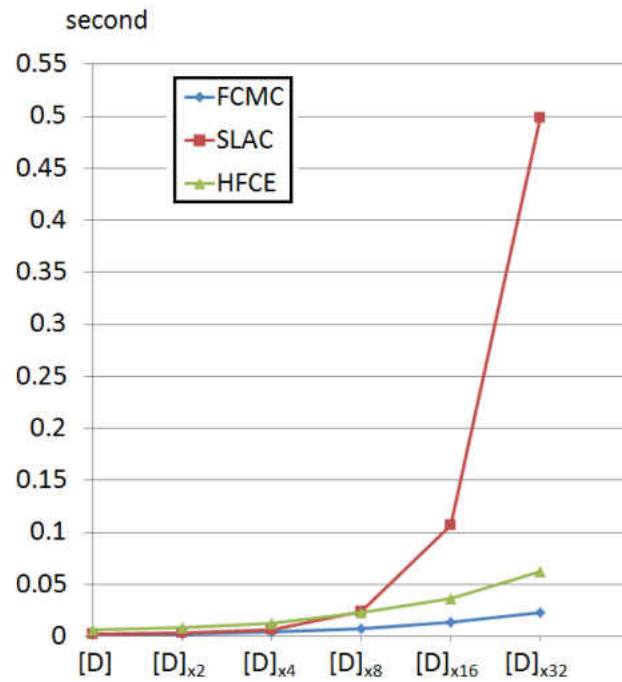


Figure 8.2: Time Cost vs. Increase of Instances

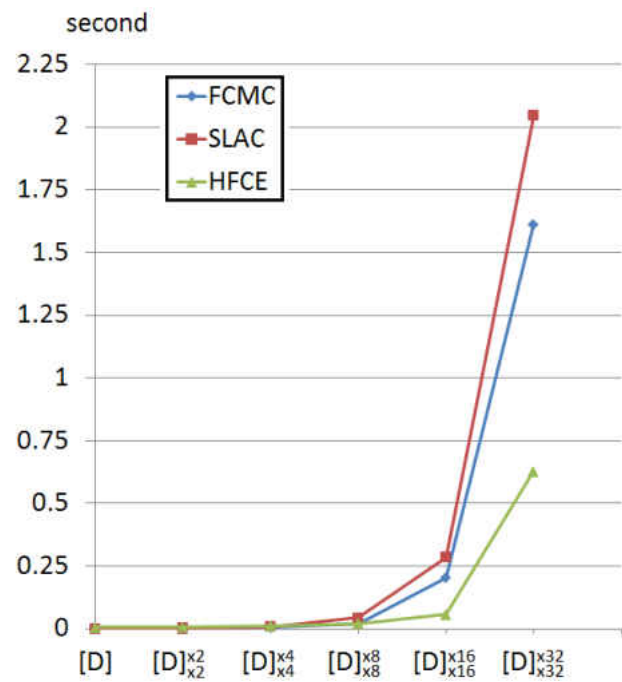


Figure 8.3: Time Cost vs. Increase of both Features and Instances

system. All three methods under comparison are implemented in series with Matlab 7.11 win64 version. Each point in these figures is an averaged value of 50 runs.

Figures 8.1-8.3 present the time cost of running each of the aforementioned three methods (HFCE, SLAC and FCMC) in response to the increase of the number of features, the number of data points and the number of both factors, respectively. It is clear that the execution time of these methods generally increases along with the increase of data size. However, HFCE shows a more stable performance than its counterparts when both the number of features and instances are increased (see Figure 8.3). This shows that the use of feature-based partition of dataset and the pairwise similarity matrices entails more time efficiency in the proposed hierarchical fuzzy c -means ensemble, which in turn indicates the potential of HFCE in dealing with big data.

Note that the outcome of using HFCE on datasets with increased features seems to be more stable as compared with FCMC (Figure 8.1). An intuitive explanation is that HFCE only computes the one-dimension distance between data points, which makes it far less sensitive to the “curse of dimensionality”. However, HFCE still suffers from the increase of data points to a certain extent. Nevertheless, it is not so drastic as the algorithms which have a time complexity of $O(N^2)$ or above, as reflected in Figure 8.2. Thus, the proposed HFCE takes the advantage of SLAC when dealing with an increasing number of features and that of fuzzy c -means when dealing with an increasing number of data points. The downside is that although reasonable, the accuracy of HFCE is not so high as that achievable by the original fuzzy c -means in general. Nevertheless, HFCE allows for higher time efficiency than fuzzy c -means without a drastic loss of clustering quality.

8.5 Summary

This chapter has presented an approach for feature-based ensemble member generation and for link-based hierarchical base-clusters grouping, in building hierarchical fuzzy (c -means) clustering ensembles. The proposed work takes the advantage of the fast speed of generating fuzzy base-clusters on one dimension data by using fuzzy c -means, and it consider the links between base-clusters to (re-)capture the interactions between features. It also takes the advantage of hierarchical clustering in that the iterative access of data points is replaced by the computation of pair-wised similarity measures. Experimental results on nine popular benchmark datasets indicate that the proposed approach generally outperforms its crisp counterparts (HCCE and single-linkage agglomerative clustering). Furthermore, it also has the potential

to process big data as the approach entails a higher time efficiency compared to the original fuzzy c -means and hierarchical clustering. It would be therefore useful to examine the effects of the suggested parallel implementation in a real problem setting which involves big data.

Chapter 9

Conclusion

THIS chapter presents a summary of the proposed intelligent assessments of academic journals and their extensions as detailed in the preceding chapters. Having reviewed the existing information technologies applied to journal rankings in the literature, the thesis has demonstrated that the developed aggregations of similarities and the clustering ensemble methods have effectively improved the accuracy, interpretability and reliability in the aggregation of impact indicators for the assessments of academic journals. The extended study of the T -transitivity of the OWA-aggregated fuzzy relations enhances the theoretical aspect of the thesis, providing both the proofs and applications of the T -transitive OWA-aggregated fuzzy relations. Several modifications of the existing clustering ensembles have also been proposed, which exploits the link-based and speeding-up of fuzzy clustering ensemble. The capabilities and potential of the developed applications have been experimentally validated, and compared with either the original approaches, or relevant techniques in the literature.

9.1 Summary of Thesis

A number of famous APDs and journal impact indicators have been reviewed in Chapter 2. Their underlying respective inspirations span a wide range of research areas, including bibliometrics, machine learning, and data mining. Furthermore, to facilitate the aggregation of existing indicators for intelligent assessments of journals, methods which investigate the correlation and aggregation of existing indicators have

been briefly reviewed. A link-based framework of bibliographic data has also been identified, by which the existing journal impact indicators are uniformly categorised by the sets of links based on which they are calculated.

The fused-link and its application to classification and clustering of journals are described in Chapter 3. It allows for fusion of different journal impact indicators to support the assessment of academic journal quality by using distance metrics. Both classification and clustering algorithms built upon the basis of fused-links are tested against datasets of academic journals. Through comparisons with the use of advanced learning mechanism such as support vector machines and decision trees, the systematic experimental results demonstrate that the proposed fused-link based approaches help to capture and reflect the impact of academic journals while being more interpretable.

To establish a more flexible way to control the aggregation of indicators, the OWA-aggregated fuzzy relations amongst academic journals are proposed in Chapter 4. Firstly, the classic OWA based aggregation of fuzzy relations is proposed and applied to the conventional k -means algorithm for clustering journals. Secondly, a nearest neighbour guided induced OWA operator: k NN-IOWA is proposed. The proposed aggregation operators have been applied to build aggregated fuzzy relations between academic journals on the basis of the individual indicator scores. The proposed methods have the strength of controlling the degree of orness and reliability of the aggregated output.

Chapter 5 has studied the mathematical properties of ordered weighted aggregation of fuzzy relations, and its application as pairwise similarity matrices to hierarchical clustering. The conditions of when the aggregated similarities preserve T -transitivity are investigated by mathematical proofs. The proposed aggregated similarities take the advantages of OWA aggregators. Not only the degree of orness but also the T -transitivity of the aggregated fuzzy relations can be controlled by stress functions. Furthermore, experimental results indicate that the proposed ordered weighted aggregations generally outperform the conventional aggregators/distances in hierarchical clustering.

In order to enhance the assessments of academic journals with the interpretability of linguistic terms, Chapter 6 has presented a fuzzy aggregation and clustering ensemble based method for journal ranking, with its application to aggregate indicators

from the JCR provide by WoS. The proposed method is developed on the basis of data-reliability based aggregation of fuzzy clusters. Compared with direct aggregation of indicator scores, the proposed approach shows its advantage in providing ranking results that are generally more reliable with data-driven weights as well as more interpretable with linguistic terms.

Chapter 7 has discussed the fuzzy clustering ensemble in a more general scenario and provided the consensus of fuzzy link-based pairwise similarity matrices. The proposed matrices take the advantage of fuzzy c -means in which each data point can have a membership to all clusters. A set of fuzzy links between base-clusters is defined, and a fuzzy graph is employed to generate the link-based similarity matrices. Experimental results indicate that the proposed approach generally outperforms the conventional CTS which is the link-based refinements of co-association matrix designed for crisp clustering ensemble. Furthermore, the link-based methods also help to build better pairwise similarity matrices as compared to the non-link based matrix FCO .

To deal with the scenarios where the volume of data may be relatively large, a method for expediting the link-based fuzzy clustering ensemble has been devised in Chapter 8. The proposed approach builds hierarchical fuzzy (c -means) clustering ensembles based on the feature-based ensemble member generation and fuzzy link-based consensus. It takes the advantage of the fast speed of using fuzzy c -means to generate fuzzy base-clusters on one dimensional data, and considers the links between base-clusters to (re-)capture the interactions between features. It also takes the advantage of hierarchical clustering that the iterative access of data points is replaced by the computation of pair-wised similarity measures (which can be paralleled). Experimental results indicate that the proposed approach generally outperforms its crisp counterparts. Furthermore, it also has the potential to process big data as the approach entails a higher time efficiency compared to the original fuzzy c -means and hierarchical clustering.

While the proposed work is promising, much can be done to strength it for further investigation. The remainder points out several interesting issues which will be helpful to improve the current research.

9.2 Future Work

As indicated previously, although the work presented in this thesis has introduced a number of techniques potentially useful for academic journal assessment, much remains to be further developed. This section lists several important issues that are worth investigation.

9.2.1 On Assessments Methods

Although several methods for aggregation of impact indicators have been studied in this thesis, allowing both the distance/similarity-based classification and clustering algorithms to conduct the intelligent assessments of academic journals. Also, there are many other similarity measures and clustering methods available in the literature than what have been utilised in the present work. These may be employed as alternatives. It would also be interesting to see this approach applied to some other journal datasets as well, either real-world or synthetic ones to further demonstrate its ability in assessing academic journals. Such work remains active research.

It would also be useful to develop criteria for the selection of the indicators. As the evaluation of journal impacts is becoming popular amongst both publishing companies and higher education institutions, a large number of newly proposed journal impact indicators are available now. It is useful to identify and extract patterns of the interactions amongst these indicators, with an aim to select a subset of indicators which can effectively reflect the quality of journals. Recent advances in feature selection techniques (e.g., [86, 224]) may offer a tool to implement such indicator selection.

Note that a group of journals of a certain rank may often be heavily overlapped with journals of other ranks. Therefore, the low accuracy of journal ranks using clustering is not unexpected. After all, most of the journals are not obviously better or worse than others, although their ranks are more likely to be affected by the preference of the human assessors. Besides, the assumed ground truth is itself not necessarily accurate. It has been pointed out in [108] that there is no such a universally accepted, golden standard of impact measure to calibrate new indicators to. In light of this, it may be interesting to develop a technique that would allow the integration of knowledge on human preference in ranking academic journals. Preference-based reasoning [109, 220] is a popular research area in fuzzy systems.

An investigation into how such techniques may be utilised to support the present work would be beneficial.

9.2.2 On OWA Aggregation

Several approaches have been developed to determine weights for aggregation of a given set of attributes. Some of them are based on the assumption that there exists an expert that can supply crucial information that will be used later to extract the parameters needed to perform aggregation [144]. However, other methods do not require the presence of an expert but the existence of a set of training examples. From such examples, parameters can be inferred through the use of a certain learning mechanism [9, 185]. Following this direction, it would be very interesting to investigate how to learn the weights of the proposed aggregation from datasets. In particular, a study would be necessary to reveal how such learning may be carried out subject to the conditions introduced in Theorem 6 as constraints to control the T_L -transitivity of the learned weighting vectors.

Recently, a rough set feature selection technique [147] which uses the information gathered from both the lower approximation dependency value and a distance metric which considers the number of objects in the boundary region and that of those objects from the lower approximation has been proposed. The use of this measure can result in smaller subset sizes of selected features than those obtained using the dependency function alone. This demonstrates that there is much valuable information to be extracted from the boundary region. That method is a good starting point for further work based on the distance metric for exploring the boundary region of rough sets [147]. Following this thought, by replacing the distance metric with the more flexible T -transitive fuzzy relation, an interesting and promising direction of demonstrating the effectiveness of the OWA aggregated fuzzy relations may be foreseen.

9.2.3 On Fuzzy Clustering Ensemble

From the literature, it is clear that when the volume of data is far beyond the capability of a parallel computing system in problem solving, under sampling the original dataset is a possible way to solve that problem with acceptable time cost [62, 132]. Sampling techniques can speed up the mining process by more than an order of magnitude through the reduction of I/O costs and that of the number of transactions to be considered. They may also be able to make the sampled

database resident in main-memory. Furthermore, sampling may allow for accurate representation of data patterns in a given database with high confidence [219]. It has been demonstrated in Chapter 8 that the proposed feature-based fuzzy clustering ensemble can recapture the interactions between features by the links between base-clusters. It would be interesting to use the links between base-clusters to recapture the patterns or structures embedded in the original dataset, even when the clustering members are based on different subsets of the original data. Under such a further development, the capability of the proposed method for speeding-up fuzzy clustering ensemble may be further improved.

Appendix A

Publications Arising from the Thesis

A number of publications have been generated from the research carried out within the PhD project. Below lists the resultant publications that are in close relevance to the thesis, including both papers already published and articles submitted for review.

A.1 Journal Articles

1. P. Su, C. Shang, and Q. Shen, Link-based approach for bibliometric journal ranking, *Soft Computing*, vol. 17, no. 12, pp. 2399–2410, 2013.
2. P. Su, C. Shang, and Q. Shen, A hierarchical fuzzy cluster ensemble approach and its application to big data clustering, *Journal of Intelligent and Fuzzy Systems*, vol. Preprint, no. Preprint, 2015.
3. T. Chen, Q. Shen, P. Su, and C. Shang, Fuzzy rule weight modification with particle swarm optimisation, under review.
4. P. Su, Q. Shen, T. Chen, and C. Shang, Journal ranking with data reliability and fuzzy cluster ensemble, under review.
5. P. Su, C. Shang, and Q. Shen, On the ordered weighted aggregation of fuzzy similarity relations, under review.

A.2 Book Chapter

6. Q. Shen, R. Diao, and P. Su, Feature Selection Ensemble, Turing Centenary, pp. 289–306, 2012.

A.3 Conference Papers

7. P. Su and Q. Shen, Link-based methods for bibliometric journal ranking, Computational Intelligence (UKCI), 12th UK Workshop on, pp. 1–8, 2012.
8. N. Naik, P. Su, and Q. Shen, Integration of interpolation and inference, Computational Intelligence (UKCI), 2012 12th UK Workshop on, pp. 1–7, 2012.
9. P. Su, C. Shang, and Q. Shen, OWA aggregation of fuzzy similarity relations for journal ranking, Fuzzy Systems (FUZZ-IEEE), 2013 IEEE International Conference on, pp. 1–7, 2013.
10. P. Su, T. Chen, C. Shang, and Q. Shen, Nearest neighbour-guided induced OWA and its application to journal ranking, Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on, pp. 1794–1800, 2014.
11. P. Su, C. Shang, and Q. Shen, Link-based pairwise similarity matrix approach for fuzzy c-means clustering ensemble, Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on, pp. 1538–1544, 2014.
12. T. Chen, Q. Shen, P. Su, and C. Shang, Refinement of fuzzy rule weights with particle swarm optimisation, Computational Intelligence (UKCI), 2014 14th UK Workshop on, pp. 1–7, 2014.
13. T. Chen, Q. Shen, P. Su, and C. Shang, Induction of quantified fuzzy rules with particle swarm optimisation, Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on, accepted.

Appendix B

Datasets Employed in the Thesis

The UCI datasets employed in the thesis are public available benchmark data, available through the UCI machine learning repository [71] which have been drawn from real-world problem scenarios. Their underlying problem domains are described in detail below, where the URL of the respective data sets are also given in order to facilitate easy access.

- Ecoli

<http://archive.ics.uci.edu/ml/datasets/Ecoli>

“The localization site of a protein within a cell is primarily determined by its amino acid sequence. Rule-based expert system for classifying proteins into their various cellular localization sites, using their amino acid sequences, in gram-negative bacteria and in eukaryotic cells.” [97]

- Glass Identification

<http://archive.ics.uci.edu/ml/datasets/Glass+Identification>

This dataset contains 10 attributes which describes the chemical contents of glass. “The study of classification of types of glass (in determining whether the glass was a type of “float” glass or not) was motivated by criminological investigation. At the scene of the crime, the glass left can be used as evidence if it is correctly identified.” [61]

- Ionosphere

<http://archive.ics.uci.edu/ml/datasets/Ionosphere>

“This radar data was collected by a system in Goose Bay, Labrador. This system consists of a phased array of 16 high-frequency antennas with a total transmitted power on the order of 6.4 kilowatts. The targets were free electrons in the ionosphere. "Good" radar returns are those showing evidence of some type of structure in the ionosphere. "Bad" returns are those that do not; their signals pass through the ionosphere. Received signals were processed using an autocorrelation function whose arguments are the time of a pulse and the pulse number.” [171]

- Iris

<https://archive.ics.uci.edu/ml/datasets/Iris>

“This is perhaps the best known database to be found in the pattern recognition literature. Fisher’s paper [69] is a classic in the field and is referenced frequently to this day. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other.”

- Parkinsons

<https://archive.ics.uci.edu/ml/datasets/Parkinsons>

”This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson’s disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD. ” [127]

- Pima Indians Diabetes

<https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>

This dataset is from National Institute of Diabetes and Digestive and Kidney Diseases. “Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage. The diagnostic, binary-valued variable investigated is whether the patient shows signs of diabetes according to World Health Organization criteria.” [172]

- Sonar

[http://archive.ics.uci.edu/ml/datasets/Connectionist+Bench+\(Sonar,+Mines+vs.+Rocks\)](http://archive.ics.uci.edu/ml/datasets/Connectionist+Bench+(Sonar,+Mines+vs.+Rocks))

“The data set contains 111 patterns obtained by bouncing sonar signals off a metal cylinder at various angles and under various conditions, and 97 patterns obtained from rocks under similar conditions. The transmitted sonar signal is a frequency-modulated chirp, rising in frequency. The data set contains signals obtained from a variety of different aspect angles, spanning 90 degrees for the cylinder and 180 degrees for the rock. Each pattern is a set of 60 numbers in the range 0.0 to 1.0. Each number represents the energy within a particular frequency band, integrated over a certain period of time.” [82]

- Spambase

<https://archive.ics.uci.edu/ml/datasets/Spambase>

This dataset is used for classifying email as spam or non-spam. “The ‘spam’ concept is diverse: advertisements for products/web sites, make money fast schemes, chain letters, pornography... Our collection of spam e-mails came from our postmaster and individuals who had filed spam. Our collection of non-spam e-mails came from filed work and personal e-mails, and hence the word ‘george’ and the area code ‘650’ are indicators of non-spam. These are useful when constructing a personalized spam filter. One would either have to blind such non-spam indicators or get a very wide collection of non-spam to generate a general purpose spam filter. ”

- Statlog (Heart)

<http://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29>

“This data set is a heart disease database, with 6 real-valued attributes: 1, 4, 5, 8, 10, 12; 1 ordered attribute:11; 3 binary attributes: 2, 6, 9; and 3 nominal features:7, 3, 13. The class label to be predicted: absence (1) or presence (2) of heart disease.” [66]

- Statlog (Landsat Satellite)

<https://archive.ics.uci.edu/ml/datasets/Statlog+%28Landsat+Satellite%29>

“The database consists of the multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and the classification associated with the central

pixel in each neighbourhood. The aim is to predict this classification, given the multi-spectral values. In the sample database, the class of a pixel is coded as a number.” [66]

- Wine

<http://archive.ics.uci.edu/ml/datasets/Wine>

“These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.” [187]

- Yeast

<https://archive.ics.uci.edu/ml/datasets/Yeast>

This dataset is used for predicting the cellular localization sites of proteins [137, 138].

Appendix C

List of Acronyms

APD	Academic publication database
BIRS	Bibliometric information retrieval system
CTS	Connected-triple based similarity
DCP	Database citation potential
DOWA	Dependent ordered weighted averaging
ERA	Excellence in research for Australia
IEEE	Institute of electrical and electronics engineers
IET	Institution of engineering and technology
IOWA	Induced ordered weighted averaging
ISI	Institute for scientific information
JCR	Journal citation reports
JIF	Journal impact factor
<i>k</i>NN	<i>k</i> Nearest neighbour
OWA	Ordered weighted averaging
RDCP	Relative database citation potential
REF	Research excellence framework
RGN	Reader generated network

RIP	Raw impact per paper
SCI	Science citation index
SCI-E	Science citation index expanded
SJR	SCImago journal rank
SMO	Sequential minimal optimisation
SNIP	Source normalized impact per paper
SRS	SimRank based similarity
SVM	Support vector machines
WoS	Web of science

Bibliography

- [1] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer communications*, vol. 30, no. 14, pp. 2826–2841, 2007.
- [2] M. F. Abbod, D. G. von Keyserlingk, D. A. Linkens, and M. Mahfouf, "Survey of utilisation of fuzzy technology in medicine and healthcare," *Fuzzy Sets and Systems*, vol. 120, no. 2, pp. 331–349, 2001.
- [3] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space," in *Database Theory-ICDT 2001: 8th International Conference London, UK, January 4-6, 2001 Proceedings*, no. 1973. Springer Science & Business Media, 2001, p. 420.
- [4] K. Allen, H. Lancour, J. Daily, A. Kent, H. Lancour, and J. Daily, *Encyclopedia of library and information science*. Taylor & Francis, 1973.
- [5] A. Azadeh, S. Ghaderi, M. Anvari, M. Saberi, and H. Izadbakhsh, "An integrated artificial neural network and fuzzy clustering algorithm for performance assessment of decision making units," *Applied Mathematics and Computation*, vol. 187, no. 2, pp. 584–599, 2007.
- [6] P. Ball, "Index aims for fair ranking of scientists," *Nature*, vol. 436, no. 7053, pp. 900–900, 2005.
- [7] I. Beg and S. Ashraf, "Similarity measures for fuzzy sets," *Appl. and Comput. math*, vol. 8, no. 2, pp. 192–202, 2009.
- [8] G. Beliakov, "Definition of general aggregation operators through similarity relations," *Fuzzy Sets and Systems*, vol. 114, no. 3, pp. 437–453, 2000.
- [9] —, "Learning weights in the generalized owa operators," *Fuzzy Optimization and Decision Making*, vol. 4, no. 2, pp. 119–130, 2005.
- [10] G. Beliakov and S. James, "Using choquet integrals for knn approximation and classification," in *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on*. IEEE, 2008, pp. 1311–1317.
- [11] —, "Citation-based journal ranks: the use of fuzzy measures," *Fuzzy Sets and Systems*, vol. 167, no. 1, pp. 101–119, 2011.

- [12] ———, “Induced ordered weighted averaging operators,” in *Recent Developments in the Ordered Weighted Averaging Operators: Theory and Practice*. Springer, 2011, pp. 29–47.
- [13] G. Beliakov, S. James, and G. Li, “Learning choquet-integral-based metrics for semisupervised clustering,” *Fuzzy Systems, IEEE Transactions on*, vol. 19, no. 3, pp. 562–574, 2011.
- [14] G. Beliakov, A. Pradera, and T. Calvo, *Aggregation functions: A guide for practitioners*. Springer, 2007, vol. 361.
- [15] Y. Bengio and Y. Grandvalet, “No unbiased estimator of the variance of k-fold cross-validation,” *The Journal of Machine Learning Research*, vol. 5, pp. 1089–1105, 2004.
- [16] ———, “Bias in estimating the variance of k-fold cross-validation,” in *Statistical modeling and analysis for complex data problems*. Springer, 2005, pp. 75–95.
- [17] K. P. Bennett and C. Campbell, “Support vector machines: hype or hallelujah?” *ACM SIGKDD Explorations Newsletter*, vol. 2, no. 2, pp. 1–13, 2000.
- [18] C. Bergstrom, “Measuring the value and prestige of scholarly journals,” *College & Research Libraries News*, vol. 68, no. 5, pp. 314–316, 2007.
- [19] P. Berkhin, “A survey of clustering data mining techniques,” in *Grouping multidimensional data*. Springer, 2006, pp. 25–71.
- [20] M. J. Berry and G. Linoff, *Data mining techniques: for marketing, sales, and customer support*. John Wiley & Sons, Inc., 1997.
- [21] J. C. Bezdek, R. Ehrlich, and W. Full, “Fcm: The fuzzy c-means clustering algorithm,” *Computers & Geosciences*, vol. 10, no. 2, pp. 191–203, 1984.
- [22] J. C. Bezdek and J. D. Harris, “Fuzzy partitions and relations; an axiomatic basis for clustering,” *Fuzzy Sets and Systems*, vol. 1, no. 2, pp. 111–127, 1978.
- [23] S. Bhagat, I. Rozenbaum, and G. Cormode, “Applying link-based classification to label blogs,” in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. ACM, 2007, pp. 92–101.
- [24] J. Bollen, H. Van de Sompel, A. Hagberg, and R. Chute, “A principal component analysis of 39 scientific impact measures,” *PloS one*, vol. 4, no. 6, p. e6022, 2009.
- [25] J. Bollen, H. Van de Sompel, J. A. Smith, and R. Luce, “Toward alternative metrics of journal impact: A comparison of download and citation data,” *Information Processing & Management*, vol. 41, no. 6, pp. 1419–1440, 2005.
- [26] T. Boongoen and Q. Shen, “Clus-dowa: A new dependent owa operator,” in *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on*, June 2008, pp. 1057–1063.

- [27] ———, “Nearest-neighbor guided evaluation of data reliability and its applications,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 40, no. 6, pp. 1622–1633, 2010.
- [28] B. Bouchon-Meunier, M. Rifqi, and S. Bothorel, “Towards general measures of comparison of objects,” *Fuzzy Sets and Systems*, vol. 84, no. 2, pp. 143–153, 1996.
- [29] N. S. Boutonnet, M. J. Rooman, M.-E. Ochagavia, J. Richelle, and S. J. Wodak, “Optimal protein structure alignments by multiple linkage clustering: application to distantly related proteins,” *Protein Engineering*, vol. 8, no. 7, pp. 647–662, 1995.
- [30] T. Braun, W. Glänzel, and A. Schubert, “A hirsch-type index for journals,” *Scientometrics*, vol. 69, no. 1, pp. 169–173, 2006.
- [31] D. Bridges, “Assessing the quality of educational research in higher education in the uk,” in *Paper for EERQI Workshop, Leuven 19th to 21st June*, 2008.
- [32] ———, “Research quality assessment in education: impossible science, possible art?” *British Educational Research Journal*, vol. 35, no. 4, pp. 497–517, 2009.
- [33] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” *Computer networks and ISDN systems*, vol. 30, no. 1, pp. 107–117, 1998.
- [34] J. F. Burnham, “Scopus database: a review,” *Biomedical digital libraries*, vol. 3, no. 1, p. 1, 2006.
- [35] T. Calvo, G. Mayor, and R. Mesiar, *Aggregation operators: new trends and applications*. Springer Science & Business Media, 2002, vol. 97.
- [36] D. F. Campbell, “Evaluation universitärer forschung,” *Entwicklungstrends und neue Strategiemuster für wissenschaftsbasierte Gesellschaften. SWS-Rundschau*, vol. 39, no. 4, pp. 363–383, 1999.
- [37] L. Canós and V. Liern, “Soft computing-based aggregation methods for human resource management,” *European Journal of Operational Research*, vol. 189, no. 3, pp. 669–681, 2008.
- [38] S. Chakrabarti, B. Dom, and P. Indyk, “Enhanced hypertext categorization using hyperlinks,” in *ACM SIGMOD Record*, vol. 27, no. 2. ACM, 1998, pp. 307–318.
- [39] C. Chakraborty and D. Chakraborty, “A fuzzy clustering methodology for linguistic opinions in group decision making,” *Applied Soft Computing*, vol. 7, no. 3, pp. 858–869, 2007.
- [40] C. Chen, F. Ibekwe-SanJuan, and J. Hou, “The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis,” *Journal of the American Society for Information Science and Technology*, vol. 61, no. 7, pp. 1386–1409, 2010.
- [41] H. Chen and L. Zhou, “An approach to group decision making with interval fuzzy preference relations based on induced generalized continuous ordered weighted averaging operator,” *Expert Systems with Applications*, vol. 38, no. 10, pp. 13 432–13 440, 2011.

- [42] J. Chen, M. Mahfouf, C. Bingham, Y. Zhang, Z. Yang, and M. Gallimore, "An evolutionary based clustering algorithm applied to data compression for industrial systems," in *Advances in Intelligent Data Analysis XI*. Springer, 2012, pp. 103–113.
- [43] S.-M. Chen, M.-S. Yeh, and P.-Y. Hsiao, "A comparison of similarity measures of fuzzy values," *Fuzzy Sets and Systems*, vol. 72, no. 1, pp. 79–89, 1995.
- [44] S. Cooper and A. Poletti, "The new era of journal ranking: the consequences of australia's fraught encounter with "quality"." *Australian Universities' Review*, vol. 53, no. 1, pp. 57–65, 2011.
- [45] G. W. Corder and D. I. Foreman, *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons, 2014.
- [46] T. Cover and P. Hart, "Nearest neighbor pattern classification," *Information Theory, IEEE Transactions on*, vol. 13, no. 1, pp. 21–27, 1967.
- [47] M. Crawford, "Biologists using social-networking sites to boost collaboration," *BioScience*, vol. 61, no. 9, pp. 736–736, 2011.
- [48] M. De Cock, C. Cornelis, and E. E. Kerre, "Fuzzy rough sets: the forgotten step," *Fuzzy Systems, IEEE Transactions on*, vol. 15, no. 1, pp. 121–130, 2007.
- [49] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.
- [50] R. Diao, "Feature selection with harmony search and its applications," Ph.D. dissertation, Aberystwyth University, 2014.
- [51] R. Diao, F. Chao, T. Peng, N. Snooke, and Q. Shen, "Feature selection inspired classifier ensemble reduction," *Cybernetics, IEEE Transactions on*, vol. 44, no. 8, pp. 1259–1268, 2014.
- [52] R. Diao and Q. Shen, "Nature inspired feature selection meta-heuristics," *Artificial Intelligence Review*, pp. 1–30, 2015. [Online]. Available: <http://dx.doi.org/10.1007/s10462-015-9428-8>
- [53] Y. Ding, G. G. Chowdhury, S. Foo, and W. Qian, "Bibliometric information retrieval system (birs): A web search interface utilizing bibliometric research results," *Journal of the American Society for Information Science*, vol. 51, no. 13, pp. 1190–1204, 2000.
- [54] C. Domeniconi and M. Al-Razgan, "Weighted cluster ensembles: Methods and analysis," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 2, no. 4, p. 17, 2009.
- [55] H. Drucker, S. Wu, and V. N. Vapnik, "Support vector machines for spam categorization," *Neural Networks, IEEE Transactions on*, vol. 10, no. 5, pp. 1048–1054, 1999.
- [56] S. A. Dudani, "The distance-weighted k-nearest-neighbor rule," *Systems, Man and Cybernetics, IEEE Transactions on*, no. 4, pp. 325–327, 1976.

- [57] J. J. Dujmović, “Properties of local andness/oriness,” in *Theoretical Advances and Applications of Fuzzy Logic and Soft Computing*. Springer, 2007, pp. 54–63.
- [58] M. Dunaiski and W. Visser, “Comparing paper ranking algorithms,” in *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*. ACM, 2012, pp. 21–30.
- [59] L. Egghe, “Theory and practise of the g-index,” *Scientometrics*, vol. 69, no. 1, pp. 131–152, 2006.
- [60] V. Estivill-Castro, “Why so many clustering algorithms: a position paper,” *ACM SIGKDD Explorations Newsletter*, vol. 4, no. 1, pp. 65–75, 2002.
- [61] I. W. Evett and E. J. Spiehler, “Rule induction in forensic science,” Central Research Establishment, Home Office Forensic Science Service, Tech. Rep., 1987.
- [62] A. Fahad, N. Alshatri, Z. Tari, A. Alamri, I. Khalil, A. Zomaya, S. Foufou, and A. Bouras, “A survey of clustering algorithms for big data: taxonomy and empirical analysis,” *Emerging Topics in Computing, IEEE Transactions on*, vol. 2, no. 3, pp. 267–279, Sept 2014.
- [63] M. E. Falagas, V. D. Kouranos, R. Arencibia-Jorge, and D. E. Karageorgopoulos, “Comparison of scimago journal rank indicator with journal impact factor,” *The FASEB journal*, vol. 22, no. 8, pp. 2623–2628, 2008.
- [64] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” *AI Magazine*, vol. 17, no. 3, p. 37, 1996.
- [65] J. Feather and P. Sturges, *Routledge International Encyclopaedia of Information and Library Science*. Routledge, 2003.
- [66] C. Feng, A. Sutherland, R. King, S. Muggleton, and R. Henery, “Comparison of machine learning classifiers to statistics and neural networks,” *AI&Statistics-93*, vol. 6, p. 41, 1993.
- [67] X. Z. Fern and C. E. Brodley, “Random projection for high dimensional data clustering: A cluster ensemble approach,” in *ICML*, vol. 3, 2003, pp. 186–193.
- [68] —, “Solving cluster ensemble problems by bipartite graph partitioning,” in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 36.
- [69] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [70] P. Fonck, J. Fodor, and M. Roubens, “An application of aggregation procedures to the definition of measures of similarity between fuzzy sets,” *Fuzzy Sets and Systems*, vol. 97, no. 1, pp. 67–74, 1998.
- [71] A. Frank and A. Asuncion, “UCI machine learning repository,” 2010.

- [72] A. Fred, "Finding consistent clusters in data partitions," in *Multiple classifier systems*. Springer, 2001, pp. 309–318.
- [73] A. L. Fred and A. K. Jain, "Combining multiple clusterings using evidence accumulation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 6, pp. 835–850, 2005.
- [74] X. Fu and Q. Shen, "Fuzzy compositional modeling," *Fuzzy Systems, IEEE Transactions on*, vol. 18, no. 4, pp. 823–840, 2010.
- [75] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000.
- [76] E. Garfield, "Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies," *Science*, vol. 178, no. 4060, pp. 471–479, 1972. [Online]. Available: <http://www.sciencemag.org/content/178/4060/471.short>
- [77] —, "How ISI selects journals for coverage: quantitative and qualitative considerations," *Essays of an Information Scientist*, vol. 13, no. 22, p. 186, 1990.
- [78] —, "The evolution of the science citation index," *International Microbiology*, vol. 10, no. 1, pp. 65–69, 2010.
- [79] L. Getoor and C. P. Diehl, "Link mining: a survey," *ACM SIGKDD Explorations Newsletter*, vol. 7, no. 2, pp. 3–12, 2005.
- [80] A. Gionis, H. Mannila, and P. Tsaparas, "Clustering aggregation," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, p. 4, 2007.
- [81] A. Gomez-Skarmeta, M. Delgado, and M. Vila, "About the use of fuzzy clustering techniques for fuzzy model identification," *Fuzzy Sets and Systems*, vol. 106, no. 2, pp. 179–188, 1999.
- [82] R. P. Gorman and T. J. Sejnowski, "Analysis of hidden units in a layered network trained to classify sonar targets," *Neural Networks*, vol. 1, p. 75, 1988.
- [83] J. M. Górriz, J. Ramírez, E. W. Lang, and C. G. Puntonet, "Hard c-means clustering for voice activity detection," *Speech communication*, vol. 48, no. 12, pp. 1638–1649, 2006.
- [84] J. C. Gower and G. Ross, "Minimum spanning trees and single linkage cluster analysis," *Applied Statistics*, pp. 54–64, 1969.
- [85] R. Grossman, Y. Qin, W. Xu, H. Hulen, and T. Tyler, "An architecture for a scalable, high-performance digital library," in *Mass Storage Systems, 1995. 'Storage-At the Forefront of Information Infrastructures', Proceedings of the Fourteenth IEEE Symposium on*. IEEE, 1995, pp. 89–98.
- [86] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

- [87] S. T. Hadjitodorov, L. I. Kuncheva, and L. P. Todorova, "Moderate diversity for better cluster ensembles," *Information Fusion*, vol. 7, no. 3, pp. 264–275, 2006.
- [88] A.-W. Harzing and R. Van Der Wal, "Google scholar as a new source for citation analysis," *Ethics in Science and Environmental Politics*, vol. 8, pp. 61–73, 2008.
- [89] —, "A google scholar h-index for journals: An alternative metric to measure journal impact in economics and business," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 1, pp. 41–46, 2009.
- [90] N. Haslam and P. Koval, "Possible research area bias in the excellence in research for australia (era) draft journal rankings," *Australian Journal of Psychology*, vol. 62, no. 2, pp. 112–114, 2010.
- [91] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani, *The elements of statistical learning*. Springer, 2009, vol. 2, no. 1.
- [92] Y.-S. Ho, "The top-cited research works in the science citation index expanded," *Scientometrics*, vol. 94, no. 3, pp. 1297–1312, 2013.
- [93] R. V. Hogg, J. McKean, and A. T. Craig, *Introduction to mathematical statistics*. Pearson Education, 2005.
- [94] J. D. Holliday, S. L. Rodgers, P. Willett, M.-Y. Chen, M. Mahfouf, K. Lawson, and G. Mullier, "Clustering files of chemical structures using the fuzzy k-means clustering method," *Journal of chemical information and computer sciences*, vol. 44, no. 3, pp. 894–902, 2004.
- [95] C. W. Holsapple, "A new map for knowledge dissemination channels," *Communications of the ACM*, vol. 52, no. 3, pp. 117–125, 2009.
- [96] C. W. Holsapple, L. E. Johnson, H. Manakyan, and J. Tanner, "A citation analysis of business computing research journals," *Information & Management*, vol. 25, no. 5, pp. 231–244, 1993.
- [97] P. Horton and K. Nakai, "A probabilistic classification system for predicting the cellular localization sites of proteins." in *Proceedings of the 4th International Conference on Intelligent Systems for Molecular Biology*, vol. 4, 1996, pp. 109–115.
- [98] N. Iam-On and T. Boongoen, "Comparative study of matrix refinement approaches for ensemble clustering," *Machine Learning*, pp. 1–32, 2013.
- [99] N. Iam-On, T. Boongoen, and S. Garrett, "Refining pairwise similarity matrix for cluster ensemble problem with cluster relations," in *Discovery Science*. Springer, 2008, pp. 222–233.
- [100] N. Iam-On, T. Boongoen, S. Garrett, and C. Price, "A link-based approach to the cluster ensemble problem," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 12, pp. 2396–2409, 2011.
- [101] N. Iam-on, S. Garrett *et al.*, "Linkclue: A matlab package for link-based cluster ensembles," *Journal of Statistical Software*, vol. 36, no. i09, 2010.

- [102] P. Jacsó, “Comparison and analysis of the citedness scores in web of science and google scholar,” in *Digital libraries: Implementing strategies and sharing experiences*. Springer, 2005, pp. 360–369.
- [103] ———, “Five-year impact factor data in the journal citation reports,” *Online Information Review*, vol. 33, no. 3, pp. 603–614, 2009.
- [104] ———, “Google scholar author citation tracker: is it too little, too late?” *Online Information Review*, vol. 36, no. 1, pp. 126–141, 2012.
- [105] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: a review,” *ACM Computing Surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.
- [106] S. Jin, D. S. Yeung, and X. Wang, “Network intrusion detection in covariance feature space,” *Pattern Recognition*, vol. 40, no. 8, pp. 2185–2197, 2007.
- [107] S. C. Johnson, “Hierarchical clustering schemes,” *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.
- [108] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.
- [109] S. Kaci and L. van der Torre, “Preference-based argumentation: Arguments supporting multiple values,” *International Journal of Approximate Reasoning*, vol. 48, no. 3, pp. 730–751, 2008.
- [110] G. Karypis, R. Aggarwal, V. Kumar, and S. Shekhar, “Multilevel hypergraph partitioning: applications in vlsi domain,” *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 7, no. 1, pp. 69–79, 1999.
- [111] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009, vol. 344.
- [112] S. Y. Kim and J. W. Lee, “Ensemble clustering method based on the resampling similarity measure for gene expression data,” *Statistical methods in medical research*, vol. 16, no. 6, pp. 539–564, 2007.
- [113] S. Klink, P. Reuther, A. Weber, B. Walter, and M. Ley, “Analysing social networks within bibliographical data,” in *Database and Expert Systems Applications*. Springer, 2006, pp. 234–243.
- [114] G. Klir, *Fuzzy sets and fuzzy logic*. Prentice Hall, 1995, vol. 4.
- [115] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proceedings of the 14th international joint conference on Artificial intelligence-Volume 2*. Morgan Kaufmann Publishers Inc., 1995, pp. 1137–1143.
- [116] J. F. Kolen and T. Hutcheson, “Reducing the time complexity of the fuzzy c-means algorithm,” *Fuzzy Systems, IEEE Transactions on*, vol. 10, no. 2, pp. 263–267, 2002.
- [117] S. Kotsiantis, “Supervised machine learning: A review of classification techniques,” *Informatica*, vol. 31, pp. 249–268, 2007.

- [118] K. Kousha, M. Thelwall, and S. Rezaie, "Using the web for research evaluation: the integrated online impact indicator," *Journal of Informetrics*, vol. 4, no. 1, pp. 124–135, 2010.
- [119] P. Kraker, C. Schlögl, K. Jack, and S. Lindstaedt, "Visualization of co-readership patterns from an online reference management system," *Journal of Informetrics*, vol. 9, no. 1, pp. 169–182, 2015.
- [120] C. L. Krumhansl, "Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density." 1978.
- [121] L. I. Kuncheva and S. T. Hadjitodorov, "Using diversity in cluster ensembles," in *Systems, man and cybernetics, 2004 IEEE international conference on*, vol. 2. IEEE, 2004, pp. 1214–1219.
- [122] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Machine Learning*, vol. 51, no. 2, pp. 181–207, 2003.
- [123] A. N. Langville and C. D. Meyer, *Google's PageRank and beyond: the science of search engine rankings*. Princeton University Press, 2011.
- [124] L. Leydesdorff and T. Opthof, "Scopus's source normalized impact per paper (snip) versus a journal impact factor based on fractional counting of citations," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 11, pp. 2365–2369, 2010.
- [125] Y. Li, J. Yu, P. Hao, and Z. Li, "Clustering ensembles based on normalized edges," in *Advances in Knowledge Discovery and Data Mining*. Springer, 2007, pp. 664–671.
- [126] Y. W. Lim and S. U. Lee, "On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques," *Pattern Recognition*, vol. 23, no. 9, pp. 935–952, 1990.
- [127] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello, I. M. Moroz *et al.*, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMedical Engineering OnLine*, vol. 6, no. 1, p. 23, 2007.
- [128] S. Lloyd, "Least squares quantization in pcm," *Information Theory, IEEE Transactions on*, vol. 28, no. 2, pp. 129–137, 1982.
- [129] J. Ma, D. Tian, M. Gong, and L. Jiao, "Fuzzy clustering with non-local information for image segmentation," *International Journal of Machine Learning and Cybernetics*, vol. 5, no. 6, pp. 845–859, 2014.
- [130] M. Mahfouf, M. F. Abbod, and D. A. Linkens, "A survey of fuzzy logic monitoring and control utilisation in medicine," *Artificial intelligence in medicine*, vol. 21, no. 1, pp. 27–42, 2001.
- [131] J. Malczewski, "Ordered weighted averaging with fuzzy quantifiers: Gis-based multi-criteria evaluation for land-use suitability analysis," *International Journal of Applied Earth Observation and Geoinformation*, vol. 8, no. 4, pp. 270–277, 2006.

- [132] P. Malik, "Governing big data: principles and practices," *IBM Journal of Research and Development*, vol. 57, no. 3/4, pp. 1–1, 2013.
- [133] J. M. Merigó and M. Casanovas, "Induced aggregation operators in the euclidean distance and its application in financial decision making," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7603–7608, 2011.
- [134] J. M. Merigó and A. M. Gil-Lafuente, "The induced generalized owa operator," *Information Sciences*, vol. 179, no. 6, pp. 729–741, 2009.
- [135] R. Mesiar and A. Mesiarová, "Fuzzy integrals—what are they?" *International Journal of Intelligent Systems*, vol. 23, no. 2, pp. 199–212, 2008.
- [136] B. Meyer, C. Choppy, J. Staunstrup, and J. van Leeuwen, "Viewpoint research evaluation for computer science," *Communications of the ACM*, vol. 52, no. 4, pp. 31–34, 2009.
- [137] K. Nakai and M. Kanehisa, "Expert system for predicting protein localization sites in gram-negative bacteria," *Proteins: Structure, Function, and Bioinformatics*, vol. 11, no. 2, pp. 95–110, 1991.
- [138] ———, "A knowledge base for predicting protein localization sites in eukaryotic cells," *Genomics*, vol. 14, no. 4, pp. 897–911, 1992.
- [139] D. F. Nettleton, A. Orriols-Puig, and A. Fornells, "A study of the effect of different types of noise on the precision of supervised learning techniques," *Artificial Intelligence Review*, vol. 33, no. 4, pp. 275–306, 2010.
- [140] R. T. Ng and J. Han, "Clarans: A method for clustering objects for spatial data mining," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 14, no. 5, pp. 1003–1016, 2002.
- [141] N. Nguyen and R. Caruana, "Consensus clusterings," in *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*. IEEE, 2007, pp. 607–612.
- [142] M. Norris and C. Oppenheim, "Comparing alternatives to the web of science for coverage of the social sciences literature," *Journal of Informetrics*, vol. 1, no. 2, pp. 161–169, 2007.
- [143] H.-J. Oh, S. H. Myaeng, and M.-H. Lee, "A practical hypertext categorization method using links and incrementally available class information," in *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2000, pp. 264–271.
- [144] M. O'Hagan, "Aggregating template or rule antecedents in real-time expert systems with fuzzy set logic," in *Signals, Systems and Computers, 1988. Twenty-Second Asilomar Conference on*, vol. 2. IEEE, 1988, pp. 681–689.
- [145] C. F. Olson, "Parallel algorithms for hierarchical clustering," *Parallel Computing*, vol. 21, no. 8, pp. 1313–1325, 1995.

- [146] E. Orduña-Malea, J. M. Ayllón, A. Martín-Martín, and E. D. López-Cózar, “About the size of google scholar: playing the numbers,” *arXiv preprint arXiv:1407.6239*, 2014.
- [147] N. Parthalaín, Q. Shen, and R. Jensen, “A distance measure approach to exploring the rough set boundary region for attribute reduction,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 22, no. 3, pp. 305–317, 2010.
- [148] W. Pedrycz and P. Rai, “Collaborative clustering with the use of fuzzy c-means and its quantification,” *Fuzzy Sets and Systems*, vol. 159, no. 18, pp. 2399–2427, 2008.
- [149] J. M. Pena, J. A. Lozano, and P. Larranaga, “An empirical comparison of four initialization methods for the k-means algorithm,” *Pattern Recognition Letters*, vol. 20, no. 10, pp. 1027–1040, 1999.
- [150] V. Perlibakas, “Distance measures for pca-based face recognition,” *Pattern Recognition Letters*, vol. 25, no. 6, pp. 711–724, 2004.
- [151] P. Perner, “Are case-based reasoning and dissimilarity-based classification two sides of the same coin?” *Engineering Applications of Artificial Intelligence*, vol. 15, no. 2, pp. 193–203, 2002.
- [152] G. Pinski and F. Narin, “Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics,” *Information Processing & Management*, vol. 12, no. 5, pp. 297–312, 1976.
- [153] J. Platt *et al.*, “Fast training of support vector machines using sequential minimal optimization,” *Advances in Kernel Methods Support Vector Learning*, vol. 3, 1999.
- [154] D. Pontille and D. Torny, “The controversial policies of journal ratings: Evaluating social sciences and humanities,” *Research Evaluation*, vol. 19, no. 5, pp. 347–360, 2010.
- [155] K. Punera and J. Ghosh, “Soft cluster ensembles,” *Advances in Fuzzy Clustering and its Applications*, pp. 69–90, 2007.
- [156] ———, “Consensus-based ensembles of soft clusterings,” *Applied Artificial Intelligence*, vol. 22, no. 7-8, pp. 780–810, 2008.
- [157] J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.
- [158] ———, *C4. 5: Programs for Machine Learning*. Morgan Kaufmann, 1993.
- [159] F. Radicchi and C. Castellano, “Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts,” *Journal of Informetrics*, vol. 6, no. 1, pp. 121–130, 2012.
- [160] A. E. RAE, “Research excellence framework,” *Policy*, p. 38, 2009.
- [161] R. K. Rainer Jr and M. D. Miller, “Examining differences across journal rankings,” *Communications of the ACM*, vol. 48, no. 2, pp. 91–94, 2005.

- [162] E. M. Rasmussen and P. WILLETT, "Efficiency of hierarchic agglomerative clustering using the icl distributed array processor," *Journal of Documentation*, vol. 45, no. 1, pp. 1–24, 1989.
- [163] C. Rinner and J. Malczewski, "Web-enabled spatial decision analysis using ordered weighted averaging (owa)," *Journal of Geographical Systems*, vol. 4, no. 4, pp. 385–403, 2002.
- [164] R. Rousseau, "Journal evaluation: Technical and practical issues," *Library trends*, vol. 50, no. 3, pp. 418–439, 2002.
- [165] J. M. F. Salido and S. Murakami, "On β -Precision aggregation," *Fuzzy Sets and Systems*, vol. 139, pp. 547–558, 2003.
- [166] J. F. Salido and S. Murakami, "Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations," *Fuzzy Sets and Systems*, vol. 139, no. 3, pp. 635–660, 2003.
- [167] G. Salton, "Developments in automatic text retrieval," *Science*, vol. 253, no. 5023, pp. 974–980, 1991.
- [168] B. Schweizer and A. Sklar, *Probabilistic metric spaces*. Courier Corporation, 2011.
- [169] A. Serenko and M. Dohan, "Comparing the expert survey and citation impact journal ranking methods: Example from the field of artificial intelligence," *Journal of Informetrics*, vol. 5, no. 4, pp. 629–648, 2011.
- [170] Q. Shen, R. Diao, and P. Su, "Feature selection ensemble." *Turing-100*, vol. 10, pp. 289–306, 2012.
- [171] V. G. Sigillito, S. P. Wing, L. V. Hutton, and K. B. Baker, "Classification of radar returns from the ionosphere using neural networks," *Johns Hopkins APL Tech. Dig*, vol. 10, pp. 262–266, 1989.
- [172] J. W. Smith, J. Everhart, W. Dickson, W. Knowler, and R. Johannes, "Using the adap learning algorithm to forecast the onset of diabetes mellitus," in *Proceedings of the Annual Symposium on Computer Application in Medical Care*. American Medical Informatics Association, 1988, p. 261.
- [173] J. Stegmann, "How to evaluate journal impact factors," *Nature*, vol. 390, no. 6660, p. 550, 1997.
- [174] S. J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. K. Chan, "Cost-based modeling for fraud and intrusion detection: Results from the jam project," in *DARPA Information Survivability Conference and Exposition, 2000. DISCEX'00. Proceedings*, vol. 2. IEEE, 2000, pp. 130–144.
- [175] A. Strehl and J. Ghosh, "Cluster ensembles—a knowledge reuse framework for combining multiple partitions," *The Journal of Machine Learning Research*, vol. 3, pp. 583–617, 2003.

- [176] P. Su, Y. Li, Y. Li, and S. C.-K. Shiu, "An auto-adaptive convex map generating path-finding algorithm: genetic convex a*," *International Journal of Machine Learning and Cybernetics*, vol. 4, no. 5, pp. 551–563, 2013.
- [177] J. Suchan, "How academic organizational systems and culture undermine scholarship and quality research: a response to ron dulek," *Journal of Business Communication*, vol. 45, no. 3, pp. 349–356, 2008.
- [178] T. Sudkamp, "Similarity, interpolation, and fuzzy rule construction," *Fuzzy Sets and Systems*, vol. 58, no. 1, pp. 73–86, 1993.
- [179] Y. Sun, J. Han, P. Zhao, Z. Yin, H. Cheng, and T. Wu, "Rankclus: integrating clustering with ranking for heterogeneous information network analysis," in *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*. ACM, 2009, pp. 565–576.
- [180] Y. Sun, T. Wu, Z. Yin, H. Cheng, J. Han, X. Yin, and P. Zhao, "Bibnetminer: mining bibliographic information networks," in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. ACM, 2008, pp. 1341–1344.
- [181] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [182] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "Arnetminer: extraction and mining of academic social networks," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008, pp. 990–998.
- [183] Y. A. Tolias, S. M. Panas, and L. H. Tsoukalas, "Generalized fuzzy indices for similarity matching," *Fuzzy Sets and Systems*, vol. 120, no. 2, pp. 255–270, 2001.
- [184] A. Topchy, A. K. Jain, and W. Punch, "Clustering ensembles: Models of consensus and weak partitions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 12, pp. 1866–1881, 2005.
- [185] V. Torra, "Learning weights for weighted owa operators," in *Industrial Electronics Society, 2000. IECON 2000. 26th Annual Conference of the IEEE*, vol. 4. IEEE, 2000, pp. 2530–2535.
- [186] J. K. Vanclay and L. Bornmann, "Metrics to evaluate research performance in academic institutions: A critique of era 2010 as applied in forestry and the indirect h 2 index as a possible alternative," *Scientometrics*, vol. 91, no. 3, pp. 751–771, 2012.
- [187] B. Vandeginste, "Parvus: An extendable package of programs for data exploration, classification and correlation," *Journal of Chemometrics*, vol. 4, no. 2, pp. 191–193, 1990.
- [188] S. Vega-Pons and J. Ruiz-Shulcloper, "Clustering ensemble method for heterogeneous partitions," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Springer, 2009, pp. 481–488.

- [189] ———, “A survey of clustering ensemble algorithms,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 25, no. 03, pp. 337–372, 2011.
- [190] D. Walker, H. Xie, K.-K. Yan, and S. Maslov, “Ranking scientific publications using a model of network traffic,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2007, no. 06, p. P06010, 2007.
- [191] M. Wallace, Y. Avrithis, and S. Kollias, “Computationally efficient sup-t transitive closure for sparse fuzzy binary relations,” *Fuzzy Sets and Systems*, vol. 157, no. 3, pp. 341–372, 2006.
- [192] W. H. Walters, “Google scholar coverage of a multidisciplinary field,” *Information Processing & Management*, vol. 43, no. 4, pp. 1121–1132, 2007.
- [193] L. Waltman, N. J. van Eck, T. N. van Leeuwen, and M. S. Visser, “Some modifications to the snip journal impact indicator,” *Journal of Informetrics*, vol. 7, no. 2, pp. 272–285, 2013.
- [194] R. Wang, Y.-L. He, C.-Y. Chow, F.-F. Ou, and J. Zhang, “Learning elm-tree from big data based on uncertainty reduction,” *Fuzzy Sets and Systems*, vol. 258, pp. 79–100, 2015.
- [195] W.-J. Wang, “New similarity measures on fuzzy sets and on elements,” *Fuzzy Sets and Systems*, vol. 85, no. 3, pp. 305–309, 1997.
- [196] X. Wang, C. Yang, and J. Zhou, “Clustering aggregation by probability accumulation,” *Pattern Recognition*, vol. 42, no. 5, pp. 668–675, 2009.
- [197] X.-Z. Wang, L.-C. Dong, and J.-H. Yan, “Maximum ambiguity-based sample selection in fuzzy decision tree induction,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 24, no. 8, pp. 1491–1505, 2012.
- [198] X.-Z. Wang, Y.-L. He, and D. D. Wang, “Non-naive bayesian classifiers for classification problems with continuous attributes,” *Cybernetics, IEEE Transactions on*, vol. 44, no. 1, pp. 21–39, 2014.
- [199] X.-Z. Wang, R. Wang, H.-M. Feng, and H.-C. Wang, “A new approach to classifier fusion based on upper integral,” *Cybernetics, IEEE Transactions on*, vol. 44, no. 5, pp. 620–635, 2014.
- [200] S. Wasserman, *Social network analysis: Methods and applications*. Cambridge university press, 1994, vol. 8.
- [201] K. Q. Weinberger and L. K. Saul, “Distance metric learning for large margin nearest neighbor classification,” *The Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.
- [202] J. Williams and N. Steele, “Difference, distance and similarity as a basis for fuzzy decision support based on prototypical decision classes,” *Fuzzy Sets and Systems*, vol. 131, no. 1, pp. 35–46, 2002.
- [203] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

- [204] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1, pp. 37–52, 1987.
- [205] R. Xu, D. Wunsch *et al.*, "Survey of clustering algorithms," *Neural Networks, IEEE Transactions on*, vol. 16, no. 3, pp. 645–678, 2005.
- [206] Z. Xu, "Dependent owa operators," in *Modeling Decisions for Artificial Intelligence*. Springer, 2006, pp. 172–178.
- [207] —, "Induced uncertain linguistic owa operators applied to group decision making," *Information Fusion*, vol. 7, no. 2, pp. 231–238, 2006.
- [208] Z. Xu and J. Chen, "Ordered weighted distance measure," *Journal of Systems Science and Systems Engineering*, vol. 17, no. 4, pp. 432–445, 2008.
- [209] R. R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 18, no. 1, pp. 183–190, 1988.
- [210] —, "Families of owa operators," *Fuzzy Sets and Systems*, vol. 59, no. 2, pp. 125–148, 1993.
- [211] —, "Using stress functions to obtain owa operators," *Fuzzy Systems, IEEE Transactions on*, vol. 15, no. 6, pp. 1122–1129, 2007.
- [212] —, "Norms induced from owa operators," *Fuzzy Systems, IEEE Transactions on*, vol. 18, no. 1, pp. 57–66, 2010.
- [213] R. R. Yager and D. P. Filev, "Induced ordered weighted averaging operators," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 29, no. 2, pp. 141–150, 1999.
- [214] L. Yang, H. Lv, and W. Wang, "Soft cluster ensemble based on fuzzy similarity measure," in *Computational Engineering in Systems Applications, IMACS Multiconference on*. IEEE, 2006, pp. 1994–1997.
- [215] Z. Yu, H.-S. Wong, and H. Wang, "Graph-based consensus clustering for class discovery from gene expression data," *Bioinformatics*, vol. 23, no. 21, pp. 2888–2896, 2007.
- [216] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [217] —, "Similarity relations and fuzzy orderings," *Information Sciences*, vol. 3, no. 2, pp. 177–200, 1971.
- [218] M. J. Zaki, "Parallel and distributed association mining: A survey," *IEEE Cconcurrency*, vol. 7, no. 4, pp. 14–25, 1999.
- [219] M. J. Zaki, S. Parthasarathy, W. Li, and M. Ogihara, "Evaluation of sampling for data mining of association rules," in *Research Issues in Data Engineering, 1997. Proceedings. Seventh International Workshop on*. IEEE, 1997, pp. 42–50.

- [220] M. Zanker, M. Jessenitschnig, and W. Schmid, "Preference reasoning with soft constraints in constraint-based recommender systems," *Constraints*, vol. 15, no. 4, pp. 574–595, 2010.
- [221] J. Zhai, H. Xu, and Y. Li, "Fusion of extreme learning machine with fuzzy integral," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 21, no. supp02, pp. 23–34, 2013.
- [222] J. Zhai, M. Zhai, and C. Bai, "An improved algorithm for calculating fuzzy attribute reducts," *Journal of Intelligent and Fuzzy Systems*, vol. 25, no. 2, pp. 303–313, 2013.
- [223] Q. Zhang, M. Mahfouf, J. R. Yates, and C. Pinna, "Model fusion using fuzzy aggregation: Special applications to metal properties," *Applied Soft Computing*, vol. 12, no. 6, pp. 1678–1692, 2012.
- [224] L. Zheng, R. Diao, and Q. Shen, "Self-adjusting harmony search-based feature selection," *Soft Computing*, pp. 1–13, 2014. [Online]. Available: <http://dx.doi.org/10.1007/s00500-014-1307-8>
- [225] J. Zhou, C. Philip Chen, L. Chen, and H.-X. Li, "A collaborative fuzzy clustering algorithm in distributed network environments," *Fuzzy Systems, IEEE Transactions on*, vol. 22, no. 6, pp. 1443–1456, 2014.