UNIVERSITI PUTRA MALAYSIA

FUZZY CLUSTERING METHOD AND EVALUATION BASED ON MULTI CRITERIA DECISION MAKING TECHNIQUE

FADHAA OTHMAN SAMEER

FS 2018 28

FUZZY CLUSTERING METHOD AND EVALUATION BASED ON MULTI CRITERIA DECISION MAKING TECHNIQUE

By

FADHAA OTHMAN SAMEER

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

January 2018

DEDICATIONS

- To my Prophet Mohamed who taught the humanity. To respectful father and a lovely mother who taught me the meaning of courage and always had confidence in me.
- To my brothers Safhaa, Deyhaa, Muhmad, Ahmed and Alaa also my sister Nuras and for all their charge and support of me a lot and made it all possible for me.
- To all my friends who accompanied me through the different parts of my study.

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Doctor of Philosophy

FUZZY CLUSTERING METHOD AND EVALUATION BASED ON MULTI CRITERIA DECISION MAKING TECHNIQUE

By

FADHAA OTHMAN SAMEER

January 2018

Chairman : Associate Professor Mohd. Rizam Abu Bakar, PhD
Faculty : Science

In the financial sector, credit scoring is one of the most successful operational research techniques. Credit scoring is an evaluation of the risk connected with lending to clients (consumers) or an organization. In actual credit scoring-related problems, generally inaccurate parameters or input data are used due to incomplete or inaccessible information being provided. Thus, designing a successful credit scoring model is then becoming more complex. Furthermore, the fuzzy approach is more efficient than the others to handle imprecisions and uncertainties. Hence, fuzzy clustering analysis such as the Gustafson-Kessel (GK) algorithm is seen to be a very important tool in the field of credit scoring. In a credit scoring problem with cluster analysis, finding a subset of features from large data sets is a very important issue. In addition, two other important problems are the requiring predefined number of clusters and selecting initial centres of clusters. Thus in this study we intend to overcome these problems by determining a feature subset and the number of the cluster problems after developing an algorithm which simultaneously solved these issues. This proposed algorithm is developed based on heuristic method named modified binary particle swarm optimization (MBPSO) with kernel fuzzy clustering method as a fitness function. The proposed algorithm is used as a pre-processing method for data followed by Gustafson-Kessel (GK) algorithm to classify credit scoring data. For the third problem a modified of Kohonen Network (MKN) algorithm was proposed to select the initial centres of clusters. A similar degree between points was utilized to get similarity density, and then by means of maximum density points selecting them as weights of the Kohonen algorithm. After the optimization of the weights by modified version of the Kohonen Network method these weights will be set as the initial centres of the Gustafson-Kessel (GK) algorithm. Hence, we proposed a complete method by combining MBPSO, MKN and GK (MBPSO+MKN+GK). The new proposed method (MBPSO+MKN+GK) Gustafson-Kessel algorithm (GK)integrated with modified of Kohonen Network algorithm (MKN) and modified binary particle swarm optimization (MBPSO) was used to classify the credit scoring data. Multi-criteria decision making was used for measuring the overall preference values of these methods and considered all the criteria at the same time. The technique for order preference by similarity to ideal solution (TOPSIS) was used for ranking the fuzzy clustering processes having multiple criteria. Furthermore, the weights of the criteria were determined by using the modified fuzzy analytic hierarchy process (MFAHP) with ranking function. Simulation experiments were carried out to investigate the performance of methods with different number of samples and different number of features. Also these methods were applied on two credit scoring datasets of German and Australian. For a real problem application, we consider the data from Gulf Commercial Bank in Iraq. This study revealed that the GK along with the MBPSO algorithm showed a better performance as compared to the GK algorithm alone. Also, the GK and MKN algorithms together were better than GK alone. But the best performance of all will be the MBPSO+MKN+GK.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

KAEDAH PENGELOMPOKAN KABUR DAN PENILAIAN BERASASKAN TEKNIK MEMBUAT-KEPUTUSAN PELBAGAI-KRITERIA

Oleh

FADHAA OTHMAN SAMEER

Januari 2018

Pengerusi : Associate Professor Mohd. Rizam Abu Bakar, PhD
Fakulti : Science

Dalam sektor kewangan, pemarkahan kredit adalah salah satu daripada teknik-teknik penyelidikan operasi yang paling berjaya. Pemarkahan kredit ialah penilaian risiko berhubung pemberian pinjaman kepada klien (pengguna) atau suatu organisasi. Dalam permasalahan sebenar berkaitan pemarkahan kredit, secara umumnya parameter atau data input yang tidak tepat digunakan oleh kerana maklumat tidak lengkap atau tidak tercapai yang diberikan. Maka mereka bentuk model pemarkahan kredit yang berjaya menjadi bertambah kompleks. Tambahan pula, pendekatan kabur lebih cekap daripada yang lain dalam menangani ketidaktepatan dan ketakpastian. Oleh itu analisis berkelompok kabur seperti algoritma Gustafson-Kessel (GK) dilihat sebagai satu kaedah yang penting dalam bidang pemarkahan kredit. Dalam permasalahan pemarkahan kredit dengan analisis berkelompok, mencari subset ciri daripada set data yang besar adalah isu sangat penting. Tambahan, dua permasalahan penting yang lain adalah keperluan pratertakrif bilangan kelompok dan mencari pusat kelompok permulaan. Maka dalam kajian ini kami mencadangkan satu kaedah untuk mengatasi masalah ini dengan penentuan subset ciri dan masalah bilangan kelompok selepas membangunkan satu algoritma yang mampu menyelesaikan masalah ini secara serentak. Algoritma yang dicadang ini dibangun berdasarkan kepada kaedah heuristik yang dipanggil pengoptimuman kerumunan zarah biner terubahsuai (MBPSO) dengan kaedah pengelompokan intipati kabur sebagai fungsi kebugaran. Algoritma yang dicadang digunakan dalam kaedah praprosesan untuk data dan diikuti dengan algoritma Gustafson-Kessel (GK) untuk mengelaskan data pemarkahan kredit. Untuk masalah ketiga algoritma rangkaian Kohonen (MKN) dicadang untuk memilih pusat permulaan kelompok. Darjah setara antara titik digunakan untuk mendapatkan kepadatan keserupaan dan seterusnya menggunakan min kepadatan titik pemilihan maksimum sebagai pemberat algoritma Kohonen. Selepas pengoptimuman pemberat menggunakan kaedah rangkaian Kohonen terubahsuai,

pemberat ini jadikan sebagai pusat algoritma Gustafson-Kessel (GK). Seterusnya kami mencadangkan kaedah lengkap dengan mengabungkan MBPSO, MKN dan GK (MBPSO+MKN+GK). Pembuatan keputusan multi-kriteria digunakan untuk mengukur keutamaan menyeluruh kaedah ini dan mengambil kira kesemua kriteria secara serentak. Kaedah keutamaan susunan melalui kesamaan dengan penyelesaian ideal (TOPSIS) digunakan dalam proses pemeringkatan pengklasteran kabur untuk kriteria berbilang. Tambahan lagi, pemberat kriteria ditentukan menggunakan proses hierarki analisis kabur terubahsuai (MFAHP) bersama fungsi keutamaan. Kajian simulasi dilakukan untuk menilai prestasi setiap kaedah menggunakan saiz sampel berlainan dan bilangan ciri berlainan. Juga, kaedah-kaedah ini digunakan ke atas dua set data pemarkahan Jerman dan Australia. Untuk aplikasi permasalahan sebenar, kami menggunakan data Gulf Commercial Bank di Iraq. Kajian ini menunjukkan algoritma GK dan MBPSO menunjukkan prestasi yang lebih baik berbanding jika algoritma GK bersendiri. Juga algoritma GK dan MKN bersama lebih baik dari algoritma GK bersendiri. Tetapi prestasi terbaik dikalangan kesemua kaedah adalah MBPSO+MKN+GK.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank GOD for the bounties who has granted me. I thank him for giving me the ability to deal with my challenges during my research and letting me accomplishing this thesis.

I would like to thank Assoc. Prof. Dr. Mohd Rizam Abu Bakar, my supervisor, for the professional, scientific, and personal guidance he has generously given me through this long journey.

My gratitude also goes to the members of my supervisory committee, Prof. Dr. Iden Hassan Hussain Al-Kanani and Assoc. Prof. Dr. Lee Lai Soon and Assoc. Prof. Dr. Leong Wah June for their expertise and important contributions, useful suggestions helpful comments and personal mentor.

Also importantly, my special thanks to my family, my parents, my sister and brothers who encouraged me not to miss my hope in doing my research and supported me a lot mentally.

I would like to thank the Ministry of Higher Education and Scientific Research of Iraq for the financial supporting of the scholarship. Further thanks to Gulf Commercial Bank for supporting my research by providing of data. Much gratitude is also due to the Universiti Putra Malaysia members who created an environment in which Ph.D. students can flourish. My acknowledgement would be incomplete without mentioning of my friends who made wonderful memories for me. Thanks you all.

I certify that a Thesis Examination Committee has met on 23 January 2018 to conduct the final examination of Fadhaa Othman Sameer on her thesis entitled "Fuzzy Clustering Method and Evaluation Based on Multi Criteria Decision Making Technique" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Zarina Bibi binti Ibrahim, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Habshah binti Midi, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Mohd Bakri bin Adam, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Anthony Bellotti, PhD
Senior Lecturer
Imperial College London
United Kingdom
(External Examiner)

NOR AINI AB. SHUKOR, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 28 March 2018

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Mohd. Rizam Abu Bakar, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairperson)

Lee Lai Soon, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Leong Wah June, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Iden Hassan Hussain Al-Kanani, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ______________________ Date: ______________________

Name and Matric No: Fadhaa Othman Sameer, GS41271

Declaration by Members of Supervisory Committee

This is to confirm that:
- the research conducted and the writing of the thesis was under supervision;
- supervision responsibilities as stated in Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature : ______________________________
Name of Chairman of Supervisory Comittee : Associate Professor Dr. Mohd Rizam Abu Bakar

Signature : ______________________________
Name of Member of Supervisory Comittee : Associate Professor Dr. Lee Lai Soon

Signature : ______________________________
Name of Member of Supervisory Comittee : Associate Professor Dr. Leong Wah June

Signature : ______________________________
Name of Member of Supervisory Comittee : Professor Dr. Iden Hassan Hussain Al-Kanani

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

$\widetilde{A}$	Fuzzy set
BPSO	Binary particle swarm optimization
C	Clusters centres matrix
d	Distance function
F	Covariance matrix
FCM	Fuzzy C-Means clustering algorithm
FS	Features selection
GK	Gustafson-Kessel algorithm
J	Objective function
KFCM	Kernel fuzzy clustering method
KN	Kohonen network
MBPSO	Modified binary particle swarm optimization
MFAHP	Modified fuzzy analytic hierarchy process
MKN	Modified Kohonen network
$\mu_{\widetilde{A}}(x)$	Membership of x in $\widetilde{A}$
n	The sample size
PSO	Particle swarm optimization
p	Features of data
SOM	Self-organizing map
TOPSIS	Technique for order preference by similarity to ideal solution
U	Partition matrix
X	Data set

CHAPTER 1

INTRODUCTION

1.1 Scope of analysis

In recent years, finance officers and bankers, all over the world, have faced many challenges. These challenges primarily relate to the complexity of financial markets, as a result of growing demand for financing. This complication in financial environments forces the financiers to find appropriate tools for measuring losses which may arise in the banking sector (Choudhary et al., 2009). This can be done through measuring the major risks named market risk, operational risk and credit risk. As the major operation of banks and financial institutions include lending activities, credit risk is therefore a major risk observed in the banking and financial system. To safeguard against losses which may arise when faced with non-performing loans, measuring the risk and making adequate provisions, based on those calculations, is a must for every financial institution. In this regard, during the recent decades, risk management and especially credit risk management has attracted many academics as well as researchers from the financial sector (Shojai et al., 2010).

1.2 General Review

The credit risk can be described as the risk of the principal loss or the loss if the financial returns which occur due to the borrower's failure to either fulfil the agreed upon obligations or repay the loan. Credit risk arises and can be expected when a borrower does not meet their obligation in relation to future cash flows. Therefore, there is uncertainty over the borrower's financial performance in the future. As a result, in recent years, financiers have sought tools and means to enable them to calculate the borrower's credit worthiness. Credit scoring is a type of risk evaluation process which is seen to be a very important decision made by the financial institutions for preventing any wrong decisions which could potentially lead to huge losses. The various classification models belong to the data mining approach used by the decision makers to decrease the credit risk of the customers. As accuracy is an essential criterion for selecting the appropriate credit scoring model, many studies have been carried out for improving the efficiency of the credit scoring models (Khashei et al., 2013). A majority of the studies have used data mining techniques for investigating problems related to credit scoring. Many data mining models were proposed and applied. Also, in the early 1930-40s, the mail-order companies used the numerical scorecards for credit scoring (Thomas et al., 2005). It is an established fact that the data mining techniques could prove to be helpful to the decision makers and the financial managers to find out the hidden knowledge from the input data regarding their clients and their different requests. Also, this knowledge helps the financial institutions decrease the risk of clients (Khashei et al., 2013). Since the beginning, credit is always has always been for enhancing consumption. Many forms of credit like borrowing, lending, instalments and payments before or after the delivery of service or goods, consumption credit etc. already exist and help in enabling smoother transactions and increase economic growth. Even today, these consumer credit alternatives facilitate consumption and improve the economy. Many recent credit scoring models have been proposed based on the operational research, statistical techniques, and artificial intelligence (AI) (Thomas et al., 2005). However,the statistical techniques can be applied when determining a linear relationship between the dependent and the independent variables. Also artificial intelligence techniques have a major limitation which lies in their poor understandability because of the black box nature. It is very complicated for artificial intelligence techniques to make knowledge representation and these techniques require a vast number of training samples and long learning time (Thomas et al., 2005).

Moreover; the drawbacks of the application of mathematical programming techniques for risk assessment of the credit are mathematical programming techniques need a computation of effort that is rarely known by the business and financial analysts (Marques et al., 2013). Though intelligent and statistical techniques are classification modelsthey are crispy models that use the classic logic in their procedures of modelling. However; these two techniques can not be able to get effective models for the uncertainties that exist in the relationships and data (Khashei et al., 2009). Therefore; to overcome these drawbacks, a soft models or fuzzy models have been proposed to the financial sector which is a second category of credit scoring models. The probable future performance of the customers is changing over time which is humanity behaviour. In the case of credit applications it is difficult to gather relevant information about individuals for determining their credit worthiness score. Also the final decision of the experts based on the applicant characteristics like credit history, age, income, loan required etc. However, in actual credit scoring problems, the input parameters and data are often incorrect as the information could be unattainable or incomplete (Lin, 2010). Therefore, there are many complex difficulties which appear while tackling the uncertainties during the designing of an appropriate credit scoring model. Furthermore, the fuzzy approach is more efficient than the others to handle imprecisions and uncertainties (Safaei et al., 2008).

The data collected provides a further basis for analysing, reasoning, making decisions and finally, for understanding different phenomena and objects. One important data analysis activity involves classifying or grouping the data into sets of clusters or categories. Objects placed in one group showed similar properties depending on a particular criterion (Zakrzewska, 2007). Cluster analysis or clustering is defined as the identification of the subsets of similar objects. This subset generally and intuitively is seen to correspond to points which are similar to one another as compared to points from other clusters. These points in the same cluster display a similar label. Clustering can be conducted in an unsupervised manner by determining the similar subset of points, without using a predefined idea of a cluster (Lim and Sohn, 2007).

Thus, the determination of the clusters having labels is based on data instead of using a specific model or a perspective of this data. There is a huge need for developing reliable models which can help in predicting the defaults (Bezdek et al., 1984).

One of the statistical method that be non-parametric which can be used for credit scoring is clustering analysis. The essential advantages of this type of analysis is that a specification of distribution does not assume for data, so it is a suitable method where sufficient priorly hypotheses does not exist. It is therefore exploratory and identifies the most likely solution which makes it suitable for use in credit risk analysis (Huang and Dun, 2008). The clustering is seen to be a substantial unsupervised learning problem; and like other similar problems, it also deals with classifying and categorising the collection of the unlabelled data (Kaufman and Rousseeuw, 2009). Furthermore, a fuzzy clustering is more flexibly than hard clustering where each object has memberships in all clusters instead of in a single cluster. The membership matrix provides more information to help the users to decide the core and boundary objects of clusters (Bai et al., 2013).

Many of the recently published studies have stated that combining many classifiers (or classifier ensembles) is better than using a single classifier (Tsai and Wu, 2008). Hence, hybrid models which have combined the advantages of many models are a very hot topic of research. The simple hybrid model consists of a process of developing a credit evolution model which is further classified into three steps: feature subset selection (FSS), determining the parameters of model and classification. This approach helps in selecting different methods for the above-mentioned three steps. The feature selection step is very important as it restricted the input feature number to improve the prediction accuracy and further decrease the computational complexity.

Furthermore, credit scoring databases in particular are often large and contain a lot of redundant and irrelevant features, so it is more demanding in terms of computation cost to classify such data. However, this difficulty can be overcome by using a feature selection method. Thus the selection of features is one of the important and most challenging issues in credit scoring.

1.3 Statement of the problem

Loan credit approval or evaluation is defined as the process which is carried out for any individual or business application for credit, which determines the eligibility of the customer for the loan (Louzada et al., 2012). The evaluation systems of credit risk is an important role in the financial decision-making. These systems diminish possible risks to enable faster decisions of credit and to reduce the cost of credit analysis. The need to control and effectively to manage credit risk has led financial institutions to strive to improve the designed techniques for credit. As a result the development of various quantitative models was promulgated by financial institutions and consulting companies. Thus, the numbers of academic studies about credit scoring were developed to show a diversity of classification methods applied to distinguish good and bad customers (Louzada et al., 2016). A majority of the studies which developed computation-based credit scoring models had the objective of obtaining better classification accuracy than other models. One of the statistical methods that is non-parametric which can be used for credit scoring is clustering analysis. The essential advantages of this type of analysis is that a specification of distribution is not assumed for data, so it is a suitable method where sufficient prior hypotheses does not exist (Lim and Sohn, 2007).

The researchers seek to solve the uncertainty and imprecisions in data and the future performance of the customers based on fuzzy clustering methods. Credit scoring databases in particular are often large and contain a lot of redundant and irrelevant features, so it is more demanding in terms of computation cost to classify such data. However, this difficulty can be overcome by using a feature selection method. Thus the selection of features is one of the important issue and most challenging in credit scoring. The models of the following researchers (Hoffmann et al. (2007), Lahsasna et al. (2010), Zhou (2012) and Gholamian et al. (2013)) adopted fuzzy clustering methods for handling uncertainty to evaluate credit risk problem without using feature subset selection techniques.

As result, many researchers have used fuzzy clustering with features subset selection in the credit scoring problem (Mehdizadeh (2009), Sadatrasoul et al. (2015) and Zhou and Li (2016)). In addition, there are two main problems in clustering analysis which include the number of clusters and the initial centres of clusters. In clustering analysis, one of the most challenging hard problems to solve is the number of clusters. Usually this number is fixed before clustering data by researchers. However, having a predetermined number of clusters is not realistic for a lot of data analysis in the real world. Also, the above researches used only features selection without integrating the two problems of feature selection and the number of clusters as pre-processing before clustering the data. After having gained the most relevant features and the optimal number of clusters, the initial centres of clusters were selected by the optimization method. Therefore, this study proposes a general and comprehensive algorithm to classify the credit scoring problem based on clustering analysis given fuzzy environment features selection and the two problems of clustering algorithms.

1.4 Motivation

The systems of credit risk evaluation have a main role in decision-making of the financial sector. These systems diminish possible risks to enable faster decisions of credit and to reduce the cost of credit analysis. The need to control and effectively to manage credit risk has led financial institutions to strive to improve the designed techniques for credit (Louzada et al., 2016). The discovery or extracting knowledge from datasets is one among the most desirable tasks in credit scoring. The data mining can deal with this problem in the form of data clustering which is one of the most common data mining tasks which can be used for credit scoring (Lim and Sohn, 2007). The difficulties of handling uncertainties while designing a successful credit scoring model is becoming more complex. Furthermore, the fuzzy approach is more efficient than the others to handle imprecisions and uncertainties (Safaei et al., 2008). The credit scoring databases in particular are often large and contain a lot of redundant and irrelevant features so it is more demanding in terms of computation cost to classify such data. However, this difficulty can be overcome by using a features selection method. Moreover, the particle swarm optimization algorithm (PSO) is a type of global optimisation algorithm that is used for solving the problems wherein the optimal solution lies as a point in the parameters of multi-dimensional space. The PSO shows a better-structured neighbourhood which helps it display a better recombination method as compared to the GA. Also, it contains a velocity term which helps it attain a faster convergence to a good solution. The binary and the continuous PSOs are used for the filter and the wrapper methods and for the single and multiple objective feature selection (Poli et al., 2007). One main problem in traditional BPSO is that the new position of a particle is solely decided by its velocity while the particles current position hardly has any in influence in updating its next position (Zhang and Liu, 2008). So a modification of the updating function of the (BPSO) is needed by using velocity and position to update the next position for every particle in the swarm in contrast to the original binary particle swarm optimization which used only velocity. In addition, there are two main problems in clustering analysis which include the number of clusters and the initial centres of clusters. This thesis intends to develop a new BPSO to solve the feature subset selection and the number of clusters simultaneously.

Also the currently used techniques for initialising the cluster centres are classified into random sampling techniques density estimation techniques and distance optimisation processes. Out of these, the random sampling is very popular owing to its simplicity. The artificial neural network (ANN) has been widely utilized for as long as three decades for clustering and data classification. Also, the competitive or the winner takes all types of artificial neural network (ANN) are used for clustering the input data. In the competitive learning, the similar data patterns are often grouped by a network and can be represented by one unit neuron. In the comparative information designs are frequently gathered by a system and can be spoken to by one unit neuron. This type of collection is carried out consequently utilizing information relationship. Some well known illustrations where the (ANN) technique was used for clustering is the Kohonen algorithm self organising maps (SOM) (Khan et al., 2001). Since the Kohonen algorithm (KA) used random selecting technique to gain initial weights we seek to modify this method by using density selecting technique. After this the optimisation weights were selected as initial centres of the clustering model to overcome the drawbacks of the existing method. This new method has higher performance than the existing method in accuracy of classification of the credit scoring problem. The multi-criteria decision-making (MCDM) models were utilized to assess a finite set of alternatives (fuzzy clustering methods) with respect to multiple criteria (external and internal) at the same time.

1.5 Preliminaries

Many fuzzy methods, like the Fuzzy Clustering methods, were developed to help in making decisions, based on the fuzzy set theory (Zadeh, 1965). Clustering plays an important role in decision making and also helps in data mining, pattern recognition, and data modelling. The clustering technique classifies the data set based on the measure of similarity of the objects present in the data. The clustering of the data into many partitions helps in exploring the characteristic relationships between the objects in the data which already contains many samples of any decision procedure. The principle objective of the clustering process is classifying the data into more similarity-based symmetric clusters.

Unsupervised learning refers to extracting knowledge from any data set using the clustering technique if this technique used the similarity measure instead of the corrective acts that are supervised by known relations. The similarity between the objects in the data set is vital for clustering. One type of similarity measure involved in fuzzy clustering includes the distance measurement and many types of distance measures are used in the existing fuzzy clustering algorithms. The clustering algorithms classify the data objects (including entities, patterns, observances, instances, units) into a particular cluster (categories, groups or subsets). The cluster is defined as an aggregate of many points within a test space and the distance between any two points within the cluster is seen to be lesser than the distance present between a point inside and that present outside the cluster (Kaufman and Rousseeuw, 2009). Also a cluster is defined as a continuous region of space (D-dimensional feature space) which contains a higher density of points. Two such high-density regions are separated from each other by a space which contains lower density points. In the above definitions, the cluster has been defined based on the internal homogeneity and an external separation i.e., the objects within one cluster are similar to one another, while the objects in the different clusters are dissimilar from each other.

1.6 The objectives of research

This research aims to propose a model under the fuzzy clustering method for credit scoring with the following objectives:

(1) To modify the binary particle swarm optimization algorithm to identify the number of clusters and to choose the most relevant features subset.

(2) To modify the Kohonen algorithm to select the initial centres of clusters for the Gustafson-Kessel algorithm.

(3) To develop a mathematical model based on fuzzy clustering analysis (Gustafson-Kessel method) for credit risk assessment.

(4) To develop the fuzzy analytic hierarchy process (MFAHP) with ranking function to compute the weights for criteria (external and internal).

(5) To extend the multi criteria decision-making (MCDM) model (TOPSIS method) by hybridizing with the modified fuzzy analytic hierarchy process (MFAHP) to evaluate a finite set of fuzzy clustering methods.

(6) To validate and verify the improvement of Gustafson -Kessel method on simulation data, standard real data and data from Gulf Commercial Bank in Iraq.

1.7 Scope of the present work

The scope of this study involves a proposed modified binary particle swarm optimization (MBPSO) which is a hybrid model between the fuzzy clustering method (kernel fuzzy c-means method) and the binary particle swarm optimization method to predefine the number of clusters and to choose the most relevant features subset. Also proposed a modified Kohonen network algorithm (MK)for the selection of the initial centres for the Gustafson-Kessel algorithm (GK). After that the two modification methods (MBPSO) and (MK) with (GK) used as a new classification method. Multi-criteria decision making (MCDM) models were used to evaluate a finite set of alternatives (fuzzy clustering methods) with respect to multiple criteria (external and internal). Simulation study was conducted in this study. In addition, the credit scoring problem was studied. Here two datasets Australian and German were investigated to test the methods comprehensively. A set of real data from Gulf Commercial Bank in Iraq was also used for analysis purposes.

In this thesis, we are mainly concerned with determining a numerical solution for classifying the credit score data with the help of fuzzy clustering analysis. This thesis has been organised as per the general structure of the dissertations and thesis submitted to the Universiti Putra, Malaysia. This thesis consists of 5 chapters.

Chapter 1: provides a general introduction about credit scoring, problem statement, objectives and contributions of this thesis.

Chapter 2: provides a literature review and it investigates the studies that are related to the research topic. These studies include some which have been published earlier and some which are still in progress with regards to the credit scoring processes including the various clustering methods.

Chapter 3: this chapter contains the proposed methods, including the modified binary particle swarm optimisation for predefining the cluster number and selecting the most relevant feature subset simultaneously. And the modified Kohonen network algorithm which helps in selecting the initial centres of the Gustafson-Kessel algorithm. Also, the modified form of TOPSIS method was proposed for ranking the fuzzy clustering techniques using multiple criteria.

Chapter 4: describes the results and the discussion when the real and the predicted data was studied using various inference methods. All the results obtained were evaluated, tabulated and discussed in further detail. In this chapter, the major findings have been highlighted and these findings were then compared to the objectives performance metrics that were used in this study for method evaluation.

Chapter 5: presents the conclusions of this study. Also future work and recommendations for some further research have been described in the chapter. This chapter also includes some remarks regarding the fulfilment of the objectives of this study. Also some suggestions are given for improving this work in future.

BIBLIOGRAPHY

Abdou, H. A. and Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management*, 18(2-3):59–88.

Abrahams, C. R. and Zhang, M. (2008). *Fair lending compliance: Intelligence and implications for credit risk management*. John Wiley & Sons.

Akko, S. (2012). An empirical comparison of conventional techniques, neural networks and the three stage hybrid adaptive neuro fuzzy inference system (anfis) model for credit scoring analysis: The case of turkish credit card data. *European Journal of Operational Research*, 222(1):168–178.

Alam, P., Booth, D., Lee, K., and Thordarson, T. (2000). The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: an experimental study. *Expert Systems with Applications*, 18(3):185–199.

Aliev, R. A., Pedrycz, W., Guirimov, B. G., Aliev, R. R., Ilhan, U., Babagil, M., and Mammadli, S. (2011). Type-2 fuzzy neural networks with fuzzy clustering and differential evolution optimization. *Information Sciences*, 181(9):1591–1608.

Altman, E. I., Marco, G., and Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks. *Journal of banking & finance*, 18(3):505–529.

Amani, F. A. and Fadlalla, A. M. (2017). Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems*, 24:32–58.

Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on neural networks*, 12(4):929–935.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., and Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the operational research society*, 54(6):627–635.

Bai, C., Shi, B., Liu, F., and Sarkis, J. (2018). Banking credit worthiness: Evaluating the complex relationships. *Omega*.

Bai, L., Liang, J., Dang, C., and Cao, F. (2013). A novel fuzzy clustering algorithm with between-cluster information for categorical data. *Fuzzy Sets and Systems*, 215:55–73.

Bandyopadhyay, S., Saha, S., and Pedrycz, W. (2011). Use of a fuzzy granulation–degranulation criterion for assessing cluster validity. *Fuzzy Sets and Systems*, 170(1):22–42.

Baraldi, A. and Blonda, P. (1999). A survey of fuzzy clustering algorithms for pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29(6):778–785.

Basirzadeh, H. and Abbasi, R. (2008). A new approach for ranking fuzzy numbers based on α-cuts. *Journal of Applied Mathematics & Informatics*, 26(3-4):767–778.

Bensaid, A. M., Hall, L. O., Bezdek, J. C., Clarke, L. P., Silbiger, M. L., Arrington, J. A., and Murtagh, R. F. (1996). Validity-guided re-clustering with applications to image segmentation. *IEEE Transactions on Fuzzy Systems*, 4(2):112–123.

Bezdek, J. C., Ehrlich, R., and Full, W. (1984). Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3):191–203.

Bittmann, R. M. and Gelbard, R. M. (2007). Decision-making method using a visual approach for cluster analysis problems; indicative classification algorithms and grouping scope. *Expert Systems*, 24(3):171–187.

Boeringer, D. W. and Werner, D. H. (2004). Particle swarm optimization versus genetic algorithms for phased array synthesis. *IEEE Transactions on Antennas and Propagation*, 52(3):771–779.

Bradley, P. S., Mangasarian, O. L., and Street, W. N. (1998). Feature selection via mathematical programming. *INFORMS Journal on Computing*, 10(2):209–217.

Campello, R. J. (2010). Generalized external indexes for comparing data partitions with overlapping categories. *Pattern Recognition Letters*, 31(9):966–975.

Chan, F. T. and Kumar, N. (2007). Global supplier development considering risk factors using fuzzy extended ahp-based approach. *Omega*, 35(4):417–431.

Chatterjee, S., Corbae, D., Nakajima, M., and Ríos, R. (2007). A quantitative theory of unsecured consumer credit with risk of default. *Econometrica*, 75(6):1525–1589.

Chen, C. (2000). Extensions of the topsis for group decision-making under fuzzy environment. *Fuzzy sets and systems*, 114(1):1–9.

Chen, Q., Chen, Y., and Jiang, W. (2016). Genetic particle swarm optimization-based feature selection for very-high-resolution remotely sensed imagery object change detection. *Sensors*, 16(8):1204.

Choudhary, A. K., Harding, J. A., and Tiwari, M. K. (2009). Data mining in manufacturing: a review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, 20(5):501–521.

Danenas, P., Garsva, G., and Gudas, S. (2011). Credit risk evaluation model development using support vector based classifiers. *Procedia Computer Science*, 4:1699–1707.

Das, S., Abraham, A., and Konar, A. (2008). Automatic kernel clustering with a multi-elitist particle swarm optimization algorithm. *Pattern recognition letters*, 29(5):688–699.

De Andrés, J., Lorca, P., de Cos Juez, F. J., and Sánchez, Lasheras, F. (2011). Bankruptcy forecasting: A hybrid approach using fuzzy c-means clustering and multivariate adaptive regression splines (mars). *Expert Systems with Applications*, 38(3):1866–1875.

Desai, V. S., Crook, J. N., and Overstreet, G. A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, 95(1):24–37.

Ding, Y. and Fu, X. (2016). Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm. *Neurocomputing*, 188:233–238.

Fichthorn, K. A. and Weinberg, W. H. (1991). Theoretical foundations of dynamical monte carlo simulations. *The Journal of chemical physics*, 95(2):1090–1096.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188.

Fodor, J. C. and Roubens, M. (2013). *Fuzzy preference modelling and multicriteria decision support*. Springer Science & Business Media.

Fraley, C. and Raftery, A. E. (1998). How many clusters? which clustering method? answers via model-based cluster analysis. *The computer journal*, 41(8):578–588.

Gadat, S. and Younes, L. (2007). A stochastic algorithm for feature selection in pattern recognition. *Journal of Machine Learning Research*, 8(5):509–547.

Gajawada, S. and Toshniwal, D. (2012). Hybrid cluster validation techniques. *Advances in Computer Science, Engineering & Applications*, pages 267–273.

Gelbard, R., Goldman, O., and Spiegler, I. (2007). Investigating diversity of clustering methods: An empirical comparison. *Data & Knowledge Engineering*, 63(1):155–166.

Gheyas, I. A. and Smith, L. S. (2010). Feature subset selection in large dimensionality domains. *Pattern recognition*, 43(1):5–13.

Gholamian, M., Jahanpour, S., and Sadatrasoul, S. (2013). A new method for clustering in credit scoring problems. *Journal of Mathematics and Computer Science*, 6:97–106.

Gómez, Skarmeta, A. F., Delgado, M., and Vila, M. A. (1999). About the use of fuzzy clustering techniques for fuzzy model identification. *Fuzzy sets and systems*, 106(2):179–188.

Graves, D. and Pedrycz, W. (2010). Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study. *Fuzzy sets and systems*, 161(4):522–543.

Gustafson, D. E. and Kessel, W. C. (1979). Fuzzy clustering with a fuzzy covariance matrix. *Proceeding of the IEEE Conference*, pages 761–766.

Hand, D. J. and Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3):523–541.

Harris, T. (2015). Credit scoring using the clustered support vector machine. *Expert Systems with Applications*, 42(2):741–750.

He, J., Lan, M., Tan, C., Sung, S., and Low, H. (2004). Initialization of cluster refinement algorithms: A review and comparative study. *Proceeding of International Conference of the Neural Networks, IEEE*, 1:297–302.

Heeren, T. and D'Agostino, R. (1987). Robustness of the two independent samples t-test when applied to ordinal scaled data. *Statistics in medicine*, 6(1):79–90.

Henley, W. and Hand, D. J. (1996). A k-nearest-neighbour classifier for assessing consumer credit risk. *The Statistician*, pages 77–95.

Ho, W., Xu, X., and Dey, P. K. (2010). Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. *European Journal of Operational Research*, 202(1):16–24.

Hoffmann, F., Baesens, B., Mues, C., Van Gestel, T., and Vanthienen, J. (2007). Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms. *European Journal of Operational Research*, 177(1):540–555.

Hsieh, T., Lu, S., and Tzeng, G. (2004). Fuzzy mcdm approach for planning and design tenders selection in public office buildings. *International journal of project management*, 22(7):573–584.

Hu, Q., Yu, D., Liu, J., and Wu, C. (2008). Neighborhood rough set based heterogeneous feature subset selection. *Information Sciences*, 178(18):3577–3594.

Huang, C. and Dun, J. (2008). A distributed pso-svm hybrid system with feature selection and parameter optimization. *Applied Soft Computing*, 8(4):1381–1391.

Huang, C. and Wang, C. (2006). A ga-based feature selection and parameters optimizationfor support vector machines. *Expert Systems with applications*, 31(2):231–240.

Huang, Z., Chen, H., Hsu, C., Chen, W., and Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision support systems*, 37(4):543–558.

Hullermeier, E., Rifqi, M., Henzgen, S., and Senge, R. (2012). Comparing fuzzy partitions: A generalization of the rand index and related measures. *IEEE Transactions on Fuzzy Systems*, 20(3):546–556.

Hung, C. and Chen, Jing, H. (2009). A selective ensemble based on expected probabilities for bankruptcy prediction. *Expert systems with applications*, 36(3):5297–5303.

Hwang, C., Yang, M., Hung, W., and Lee, M. (2012). A similarity measure of intuitionistic fuzzy sets based on the sugeno integral with its application to pattern recognition. *Information Sciences*, 189:93–109.

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323.

Jain, R. (1977). A procedure for multiple-aspect decision making using fuzzy sets. *International Journal of systems science*, 8(1):1–7.

Jiang, M., Ji, F., and Li, R. (2011). The applied research of credit scoring combination model based on sa-ga algorithm. *Proceeding of Fourth International Conference of Business Intelligence and Financial Engineering (BIFE),IEEE*, pages 491–494.

Kabir, M. M., Shahjahan, M., and Murase, K. (2012). A new hybrid ant colony optimization algorithm for feature selection. *Expert Systems with Applications*, 39(3):3747–3763.

Karels, G. V. and Prakash, A. J. (1987). Multivariate normality and forecasting of business bankruptcy. *Journal of Business Finance & Accounting*, 14(4):573–593.

Kaufman, L. and Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

Khan, J., Wei, J. S., Ringner, M., Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C., et al. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature medicine*, 7(6):673–679.

Khashei, M., Bijari, M., and Ardali, G. A. R. (2009). Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (anns). *Neurocomputing*, 72(4):956–967.

Khashei, M., Rezvan, M. T., Hamadani, A. Z., and Bijari, M. (2013). A bi-level neural-based fuzzy classification approach for credit scoring problems. *Complexity*, 18(6):46–57.

Kleinberg, J. (2003). An impossibility theorem for clustering. *Advances in neural information processing systems*, pages 463–470.

Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, 97(1):273–324.

Kou, G., Peng, Y., and Wang, G. (2014). Evaluation of clustering algorithms for financial risk analysis using mcdm methods. *Information Sciences*, 275:1–12.

Lahsasna, A., Ainon, R. N., and Wah, T. Y. (2010). Enhancement of transparency and accuracy of credit scoring models through genetic fuzzy classifier. *Maejo International Journal of Science and Technology*, 4(1):136–158.

Lane, M. C., Xue, B., Liu, I., and Zhang, M. (2013). Particle swarm optimisation and statistical clustering for feature selection. *Proceeding of Australasian Joint Conference of Artificial Intelligence,Springer*, pages 214–220.

Lane, M. C., Xue, B., Liu, I., and Zhang, M. (2014). Gaussian based particle swarm optimisation and statistical clustering for feature selection. *Proceeding of European conference of evolutionary computation in combinatorial optimization,Springer*, pages 133–144.

Lee, K. S. and Geem, Z. W. (2005). A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice. *Computer methods in applied mechanics and engineering*, 194(36):3902–3933.

Lee, T., Chiu, C., Chou, Y., and Lu, C. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, 50(4):1113–1130.

Li, Q., Ren, Y., Li, L., and Liu, W. (2016). Fuzzy based affinity learning for spectral clustering. *Pattern Recognition*, 60:531–542.

Li, X. and Zhong, Y. (2012). An overview of personal credit scoring: Techniques and future work. *International Journal of Intelligence Science*, 2(04):181.

Liang, J. J., Qin, A. K., Suganthan, P. N., and Baskar, S. (2006). Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. *IEEE transactions on evolutionary computation*, 10(3):281–295.

Lim, M. K. and Sohn, S. Y. (2007). Cluster-based dynamic scoring model. *Expert Systems with Applications*, 32(2):427–431.

Lin, H. (2010). Personnel selection using analytic network process and fuzzy data envelopment analysis approaches. *Computers & Industrial Engineering*, 59(4):937–944.

Lin, K., Li, X., Zhang, Z., and Chen, J. (2014). A k-means clustering with optimized initial center based on hadoop platform. *Proceeding of the 9th International Conference of the Computer Science & Education ,IEEE*, pages 263–266.

Lin, S.W., L. Z. C. S. and Tseng, T. (2008). Parameter determination of support vector machine and feature selection using simulated annealing approach. *Applied soft computing*, 8(4):1505–1512.

Lin, S., Shiue, Y., Chen, S., and Cheng, H. (2009). Applying enhanced data mining approaches in predicting bank performance: A case of taiwanese commercial banks. *Expert Systems with Applications*, 36(9):11543–11551.

Liu, H. and Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, 17(4):491–502.

Louzada, F., Ara, A., and Fernandes, G. B. (2016). Classification methods applied to credit scoring: Systematic review and overall comparison. *Surveys in Operations Research and Management Science*, 21(2):117–134.

Louzada, F., Ferreira, S., and Diniz, C. A. (2012). On the impact of disproportional samples in credit scoring models: An application to a brazilian bank data. *Expert Systems with applications*, 39(9):8071–8078.

Lu, Y., Wang, S., Li, S., and Zhou, C. (2011). Particle swarm optimizer for variable weighting in clustering high-dimensional data. *Machine learning*, 82(1):43–70.

Luo, Y., Pang, S., and Qiu, S. (2003). Fuzzy cluster in credit scoring. *Proceeding of the International Conference of Machine Learning and Cyberneticson,IEEE*, 5:2731–2736.

Ma, G., Xu, Z., Zhang, W., and Li, S. (2015). An enriched k-means clustering method for grouping fractures with meliorated initial centers. *Arabian Journal of Geosciences*, 8(4):1881–1893.

Macharis, C., Springael, J., De Brucker, K., and Verbeke, A. (2004). Promethee and ahp: The design of operational synergies in multicriteria analysis.: Strengthening promethee with ideas of ahp. *European Journal of Operational Research*, 153(2):307–317.

Mangasarian, O. L. (1965). Linear and nonlinear separation of patterns by linear programming. *Operations research*, 13(3):444–452.

Marques, A., García, V., and Sánchez, J. S. (2013). A literature review on the application of evolutionary computing to credit scoring. *Journal of the Operational Research Society*, 64(9):1384–1399.

Masoud, H., Jalili, S., and Hasheminejad, S. M. H. (2013). Dynamic clustering using combinatorial particle swarm optimization. *Applied intelligence*, 38(3):289–314.

Mehdizadeh, E. (2009). A fuzzy clustering pso algorithm for supplier base management. *International Journal of Management Science and Engineering Management*, 4(4):311–320.

Moayedikia, A., Jensen, R., Wiil, U. K., and Forsati, R. (2015). Weighted bee colony algorithm for discrete optimization problems with application to feature selection. *Engineering Applications of Artificial Intelligence*, 44:153–167.

Myers, J. H. and Forgy, E. W. (1963). The development of numerical credit evaluation systems. *Journal of the American Statistical association*, 58(303):799–806.

Naik, A., Satapathy, S. C., and Parvathi, K. (2012). Improvement of initial cluster center of c-means using teaching learning based optimization. *Procedia Technology*, 6:428–435.

Nanda, S. J. and Panda, G. (2014). A survey on nature inspired metaheuristic algorithms for partitional clustering. *Swarm and Evolutionary computation*, 16:1–18.

Ngai, E. W., Xiu, L., and Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2):2592–2602.

Ólafsson, S. and Yang, J. (2005). Intelligent partitioning for feature selection. *INFORMS Journal on Computing*, 17(3):339–355.

Olson, D. L. (2004). Comparison of weights in topsis models. *Mathematical and Computer Modelling*, 40(7-8):721–727.

Ozkan, I. and Türkşen, I. B. (2012). Minimax ε-stable cluster validity index for type-2 fuzziness. *Information Sciences*, 184(1):64–74.

Peng, Y., Wang, G., Kou, G., and Shi, Y. (2011). An empirical study of classification algorithm evaluation for financial risk prediction. *Applied Soft Computing*, 11(2):2906–2915.

Peng, Y., Zhang, Y., Kou, G., Li, J., and Shi, Y. (2012). Multicriteria decision making approach for cluster validation. *Procedia Computer Science*, 9:1283–1291.

Poli, R., Kennedy, J., and Blackwell, T. (2007). Particle swarm optimization. *Swarm intelligence*, 1(1):33–57.

Pudil, P., Novovičová, J., and Kittler, J. (1994). Floating search methods in feature selection. *Pattern recognition letters*, 15(11):1119–1125.

Rana, S., Jasola, S., and Kumar, R. (2011). A review on particle swarm optimization algorithms and their applications to data clustering. *Artificial Intelligence Review*, 35(3):211–222.

Reichert, A. K., Cho, C., and Wagner, G. M. (1983). An examination of the conceptual issues involved in developing credit-scoring models. *Journal of Business & Economic Statistics*, 1(2):101–114.

Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1):1–39.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Rubinstein, R. Y. and Kroese, D. P. (2016). *Simulation and the Monte Carlo method*. John Wiley & Sons.

Ruspini, E. H. (1969). A new approach to clustering. *Information and control*, 15(1):22–32.

Saaty, T. L. (2005). *Theory and applications of the analytic network process: decision making with benefits, opportunities, costs, and risks*. RWS publications.

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International journal of services sciences*, 1(1):83–98.

Sadatrasoul, S., Gholamian, M., and Shahanaghi, K. (2015). Combination of feature selection and optimized fuzzy apriori rules: The case of credit scoring. *The International Arab Journal of Information Technology*, 12(2):138–145.

Safaei, N., S.M., M., Tavakkoli, M., R., and Sassani, F. (2008). A fuzzy programming approach for a cell formation problem with dynamic and uncertain conditions. *Fuzzy Sets and Systems*, 159(2):215–236.

Sánchez, J. F. M. and Lechuga, G. P. (2016). Assessment of a credit scoring system for popular bank savings and credit. *Contaduría y Administración*, 61(2):391–417.

Schebesch, K. B. and Stecking, R. (2005). Support vector machines for classifying and describing credit applicants: detecting typical and critical regions. *Journal of the Operational Research Society*, 56(9):1082–1088.

Scitovski, S. and Šarlija, N. (2015). Cluster analysis in retail segmentation for credit scoring. *Croatian Operational Research Review*, 5(2):235–245.

Shang, L., Zhou, Z., and Liu, X. (2016). Particle swarm optimization-based feature selection in sentiment classification. *Soft Computing*, 20(10):3821–3834.

Shojai, S., Feiger, G., et al. (2010). Economists hubris: the case of risk management. *Journal of Financial Transformation*, 28:25–35.

Shunmugapriya, P. and Kanmani, S. (2017). A hybrid algorithm using ant and bee colony optimization for feature selection and classification (ac-abc hybrid). *Swarm and Evolutionary Computation*, 36:27–36.

Siami, M., Hajimohammadi, Z., et al. (2013). Credit scoring in banks and financial institutions via data mining techniques: A literature review. *Journal of AI and Data Mining*, 1(2):119–129.

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437.

Steenackers, A. and Goovaerts, M. J. (1989). A credit scoring model for personal loans. *Insurance: Mathematics and Economics*, 8(1):31–34.

Tao, Li, W. and Maybank (2009). Geometric mean for subspace selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):260–274.

Thistleton, W. J., Marsh, J. A., Nelson, K., and Tsallis, C. (2007). Generalized box–muller method for generating q-gaussian random deviates. *IEEE Transactions on Information Theory*, 53(12):4805–4810.

Thomas, L., Oliver, R., and Hand, D. (2005). A survey of the issues in consumer credit modelling research. *Journal of the Operational Research Society*, 56(9):1006–1015.

Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International journal of forecasting*, 16(2):149–172.

Tsai, C. (2014). Combining cluster analysis with classifier ensembles to predict financial distress. *Information Fusion*, 16:46–58.

Tsai, C. and Chen, M. (2010). Credit rating by hybrid machine learning techniques. *Applied soft computing*, 10(2):374–380.

Tsai, C. and Wu, J. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert systems with applications*, 34(4):2639–2649.

Tseng, F., Filev, D., and Chinnam, R. B. (2017). A mutual information based online evolving clustering approach and its applications. *Evolving Systems*, 8(3):179–191.

Tseng, Fang, M. and Lin, L. (2005). A quadratic interval logit model for forecasting bankruptcy. *Omega*, 33(1):85–91.

Tso, G. K. and Yau, K. K. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9):1761–1768.

Unler, A. and Murat, A. (2010). A discrete particle swarm optimization method for feature selection in binary classification problems. *European Journal of Operational Research*, 206(3):528–539.

Van, B., F. and Engelbrecht, A. P. (2004). A cooperative approach to particle swarm optimization. *IEEE transactions on evolutionary computation*, 8(3):225–239.

Van, D. B., F. and Engelbrecht, A. P. (2006). A study of particle swarm optimization particle trajectories. *Information sciences*, 176(8):937–971.

Wang, J. and Yang, D. (2007). Using a hybrid multi-criteria decision aid method for information systems outsourcing. *Computers & Operations Research*, 34(12):3691–3700.

Weiss, G. M. (2004). Mining with rarity: a unifying framework. *ACM Sigkdd Explorations Newsletter*, 6(1):7–19.

West, D. (2000). Neural network credit scoring models. *Computers & Operations Research*, 27(11):1131–1152.

Wiginton, J. C. (1980). A note on the comparison of logit and discriminant models of consumer credit behavior. *Journal of Financial and Quantitative Analysis*, 15(03):757–770.

Wu, K. and Yang, M. (2005). A cluster validity index for fuzzy clustering. *Pattern Recognition Letters*, 26(9):1275–1291.

Xie, X. L. and Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on pattern analysis and machine intelligence*, 13(8):841–847.

Xu, R. and Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3):645–678.

Xue, B., Zhang, M., and Browne, W. N. (2014). Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. *Applied Soft Computing*, 18:261–276.

Xue, B., Zhang, M., Browne, W. N., and Yao, X. (2016). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, 20(4):606–626.

Yalcin, N., Bayrakdaroglu, A., and Kahraman, C. (2012). Application of fuzzy multi-criteria decision making methods for financial performance evaluation of turkish manufacturing industries. *Expert Systems with Applications*, 39(1):350–364.

Yang, J. and Olafsson, S. (2006). Optimization-based feature selection with adaptive instance sampling. *Computers & Operations Research*, 33(11):3088–3106.

Yeh, I. and Lien, C. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36(2):2473–2480.

Yobas, M. B., Crook, J. N., and Ross, P. (2000). Credit scoring using neural and evolutionary techniques. *IMA Journal of Management Mathematics*, 11(2):111–125.

Yue, Z. (2014). Topsis-based group decision-making methodology in intuitionistic fuzzy setting. *Information Sciences*, 277:141–153.

Yun, L. C., Q. y. and Zhang, H. (2011). Application of the pso-svm model for credit scoring. *Proceeding of the Seventh International Conference of the Computational Intelligence and Security (CIS),IEEE*, pages 47–51.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3):338–353.

Zakrzewska, D. (2007). On integrating unsupervised and supervised classification for credit risk evaluation. *Information technology and control*, 36(1):98–102.

Zhang, W. and Liu, Y. (2008). Multi-objective reactive power and voltage control based on fuzzy optimization strategy and fuzzy adaptive particle swarm. *International Journal of Electrical Power & Energy Systems*, 30(9):525–532.

Zhou, J., L. C. Z. and Li, H. (2016). Fuzzy clustering with the entropy of attribute weights. *Neurocomputing*, 198:125–134.

Zhou, Z. (2012). An improved fuzzy isodata algorithm for credit risk assessment of the eit enterprises. *Modern Economy*, 3(05):686.

Zimmermann, H. (2010). Fuzzy set theory. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3):317–332.

Zorarpacı, E. and Özel, S. A. (2016). A hybrid approach of differential evolution and artificial bee colony for feature selection. *Expert Systems with Applications*, 62:91–103.