Fuzzy Utility Based Decision Analysis in the Credit Scoring Problem

Ersin Kuset Bodur

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the Degree of

Doctor of Philosophy
in
Applied Mathematics and Computer Science

Eastern Mediterranean University
June 2012
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Elvan Yılmaz
Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Doctor of Philosophy in Applied Mathematics and Computer Science.

Prof. Dr. Nazım Mahmudov
Chair, Department of Mathematics

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Doctor of Philosophy in Applied Mathematics and Computer Science.

Assoc. Prof. Dr. Rashad Aliyev
Supervisor

Examining Committee

1. Prof. Dr. İsmail Burhan Türkşen ____________________
2. Assoc. Prof. Dr. Rashad Aliyev ____________________
3. Asst. Prof. Dr. Mustafa Menekay ____________________

ABSTRACT

A method that uses fuzzy c-means (FCM) is proposed for credit scoring based on unsupervised learning of a set training data. Data vectors are composed of significant applicant attributes and corresponding expert decisions. Two new statistical cost functions J_m and J_σ are introduced to evaluate the candidate models by k-fold cross validation based on the mean and the standard deviation of the decision attributes. A linguistic approach based on the fuzzy-valued Choquet integral is suggested to rank the consumer loan applicants. The lower and upper imprecise probabilities are used as a capacity measure in Choquet integral to determine the utility ranking of the consumer loan applicants.

This thesis proposes an algorithm to calculate the applicant's non-expected utility by using imprecise probabilities of accepted cases over the Fuzzy C-Means clusters for fuzzy Choquet integral. The method is applied on consumer loan evaluations for a financial institution to verify expert decisions in parallel to extracting linguistic rules of decision making. In the suggested approach linguistic fuzzy valued Choquet integral is used as measure of fuzzy utility. The results indicate that the proposed method is successful in ranking the consumer loan applications with only six fails in total of 135 applications.

Keywords: Fuzzy c-means, Fuzzy clustering, Sugeno integral, Fuzzy valued Choquet integral, imprecise probability

ÖZ

Bulanık-c-ortalaması (FCM) kullanan ve anlamlı başvuru nitelikleri ve karşılığı uzman kararından oluşan bir modelleme veri kümesiyle yönlendirilmemiş öğrenmeye dayanan kredi puanlama metodu önerilmektedir. Değerlendirmeye aday modelleri k-kat çapraz sağlamayla karar niteliğinin ortalama ve standard sapmasına dayanan J_m ve J_σ adında iki yeni istatistiksel maliyet fonksiyonu tanımlanmaktadır. Tüketici kredisi için başvuran müşterileri sıralamak üzere bulanık değerli Choquet integrale dayanan sözel bir yaklaşım önerilmektedir. Choquet integralde kapasite ölçüsü olarak tüketici kredisi başvurularının fayda sıralamasını belirlemek üzere alt ve üst belirsizlik olasılıkları kullanılmaktadır.

Bu tez başvuranların umulmadık faydasını hesaplamak üzere Choquet integralin kapasite ölçüsü olarak bulanık-c-ortalaması kümelerindeki uzmanlarca onaylananların belirsiz olabilirliliği kullanan bir algoritma önermektedir. Önerilen metod bir finans kuruluşunun tüketim kredisi değerlendirmelerinde bir yandan karar vermenin sözel kurallarını bulurken, diğer taraftan uzman kararlarını sınamak üzere uygulanmıştır. Önerilen yaklaşımda bulanık-sayı değerli ölçüt kullanan Choquet integrali kullanılmıştır. Sonuçlar ileri sürülen metodun tüketici kredisi başvurularını sıralamada toplam 135 başvurudan yalnızca altı yanılmayla başarılı olduğunu göstermektedir.

Anahtar Kelimeler: Bulanık-c-ortalaması, bulanık sınıflandırma, Sugeno integral, bulanık değerli Choquet integral, kesin olmayan ihtimal

To My Family

ACKNOWLEDGMENT

It is my duty and pleasure to express my deep appreciation to my supervisor Assoc. Prof. Dr. Rashad Aliyev for his continuous help, valuable suggestions and encouragement during preparation of this dissertation.

Thanks to all my teachers for their helpful instructions.

I wish to thank my husband who supported me in many ways during the preparations of this thesis. Finally, special thanks to my mom and my son. I would like to dedicate this thesis to my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Utility is a wide-ranging concept which carries various deep meaning in different fields such as in economics, decision theory, game theory, etc. Utility is used in economics to measure "the relative satisfaction of goods and services" by Jeremy Bentham and John Stuart Mill [1]. For an investor it is the profit of an investment, and for a player it is gain or loss at the end of the game. Once measure of utility is developed, it is possible to compare utilities of substantial goods and services. Consistent comparison of the decisions by their expected utilities provides reliability of decisions, and forms the foundation of the Decision Theory based on the expected utility function, and utility theory [2].

Utility is a reward associated with an outcome of the action for each of the possible states of the world influencing the outcome of the action. The utility of an action scores the decision maker's attitudes toward possible risk and reward values [3]. Utility of an action is also called payoff, and it reflects the desirability of the outcomes of that action to the player, for any reason. When the outcomes are random, weighted probabilities of payoffs reflect the player's attitude towards risk [4].

For a given set of alternatives X, utility function $u: X \rightarrow R$ ranks the preferences relative to each other in increasing or decreasing order by a preference relation $\leq$ or

$\geq$. Function $u(x)$ rationalizes $\leq$ on X if for every $x, y \in X$, $u(x) \leq u(y)$ if and only if $x \leq y$. If $u(x)$ rationalizes $\leq$, then it implies that $\leq$ is complete, transitive and rational [1].

A utility function does not need to be a scoring function for preferences, but the legitimate use of calculus of mathematical expectations is possible only with numerical utility functions [5]. In applications, there may be representations of utility functions in tabular, graphical, and mathematical formats [1].

The first known expected utility function was defined by D. Bernoulli in 1738, while attempting to solve St. Petersburg Paradox, where the expected monetary payoff alone was inadequate for reasoning about the choices of the players in tossing coins. His idea to use the probabilities of outcomes in finding the optimum decision built up into the final formalism and axiomatic foundation of utility theory by major contribution of many researchers including [5], [6] and [7]. The topological existence conditions of utility functions as a representation of preference ordering were stated by G. Depreu [8].

Other quantitative decision-making practices use other forms of utility theory. Operations Research (OR) is emerged during World War II [9], aiming the scientific analysis of decision making. OR has evolved into management science, keeping many principles of decision analysis (DA) in its area. Mathematical models of decision analysis incorporate the preferences and probability assumptions of the decision maker along with the structure of the decision problem. Decision is considered to be an irrevocable allocation of resources, and the decision maker is an individual who has the power to commit the resources of the organization [10].

DA deals with organizational decisions, and concerning with the appropriateness of the decision-making process rather than the individuality of the decision maker or the relations of power holders within an organization. Later forms of DA did attempt to include in their thoughts the result of prospect theory. Namely, they tried to integrate biases of human judgment into their model-building processes [10].

Utility theory states that consistent, reliable, and rational comparisons of decisions optimize expected utilities of outcomes. Expected Utility Theory is a general set of assumptions and axioms that outline rational decision making for decisions with random outcomes [11]. It recommends to weight for the utility of each outcome by the probability of occurrence and then decides for the outcome that results in the greatest weighted sum.

Utility theory formed by von Neumann and Morgenstern is an axiomatically stated single objective optimization of expected utilities. Von Neumann and Morgenstern established the rational framework of expected utility functions in game theory by the following three properties: (i) the higher utility require a desirable outcome. In order to explore desirable outcome, in other words the best decision, one need the largest expected utility. (ii) For three possibilities; if choice $'a'$ is better than $'b'$ and $'b'$ is better than $'c'$, then necessarily $'a'$ is better than $'c'$, which is called the transitivity axiom. (iii) If players are indifferent between two outcomes or choices, then necessarily the expected utilities will be the same. These three assumptions underlie the rational framework for decision making under uncertainty that expected utility theory provides. Moreover, for application in economic actions, transferability of one players' utility to another is necessary [4].

A rational decision maker would exhibit certain characteristics (usually expressed in the form of axioms), and then the solution of certain problems (expressed in a formal way) may be solved based on sound mathematical principles.

The theory of Bounded Rationality was proposed by H. Simon for real organizations, where decisions are not fully rational. Limitations of information, cognitive capacity, and attention may impose restrictions on decision maker, and the optimum decision cannot be obtained [12].

In many cases the decision shall be taken in a limited time or deciding to a solution that satisfies the objectives to a certain extent is sufficient rather than searching for an optimal solution. Satisficing is the term used for searching a sufficiently satisfactory decision instead of the exactly optimal one [12].

The chapters of this thesis are organized as follows. The introduction is given in Chapter 1. Chapter 2 reviews basic definitions and principles that are fundamental parts of the structure of the next chapters. Chapter 3 describes main variables of the data set and fuzzy c-means (FCM) algorithm. Chapter 4 contains the description of the fuzzy utility functions, fuzzy valued Choquet integral and the developed ranking method introducing model of credit scoring. The experimental results are discussed in Chapter 5. Finally, conclusions are given in Chapter 6.

Chapter 2

PRELIMINARY DEVELOPMENTS and DEFINITIONS

2.1 General Review

Credit scoring is a procedure of separating specific subgroups in a population of objects. In [13], a general approach for classifying objects using mathematical programming algorithms is investigated. The approach is based on optimizing a utility function, which is quadratic in indicator parameters and is linear in control parameters. The power and usefulness of fuzzy classification rules for data mining purposes are studied in [14]. For this purpose, an evolution strategy and a genetic algorithm are recommended as evolutionary fuzzy rule learners. Their performances are compared against Nefclass, neuro-fuzzy classifier, and selection of other well-known classification algorithms on number of publicly available data sets and two real life Benelux financial credit scoring data sets. An approach in order to develop a TOPSIS classifier to show its usage in credit scoring, providing a way to deal with large sets of data using machine learning is suggested in [15].

In paper [16], two real world credit data sets in the University of California Irvine Machine Learning Repository are chosen. Support-Vector-Machine (SVM) and clustering-launched classification (CLC) are given to discuss the advantages of CLC to predict credit scoring. A hybrid mining approach in the design of an effective credit scoring model based on clustering and neural network techniques are introduced in [17] using the clustering techniques to preprocess the input samples with the objective of indicating unrepresentative samples into isolated and inconsistent clusters, and then he used neural networks to construct the credit scoring model. Most credit assessment models are based on simple credit scoring functions estimated by discriminate analysis.

Utility theory focuses on methods making decision under risk aversion starting by Bernoulli [18]. Preference relations are used in order to discuss the best alternative, thus, the theory of decision making is used in many disciplines such as Economy, Operational Research, Management, Artificial Intelligence, etc. Utility has an important role in decision making and investigation of researcher's showed us how this area can be used to examine and to solve decision problems. The proof of existence of fuzzy utility function is proposed in [19]. The utility function is the measure of preferences. Generalization of classical utility theory is projected and basic preferences are defined by means of rational fuzzy preference relations in [20]. A generic procedure for construction of multi-dimensional utility in case of mutual utility independence of base vector attributes is offered in [21]. A general approach in order to classify objects using mathematical programming algorithms based on optimizing utility function is presented in [13], with a utility function that is quadratic in indicator parameters and linear in control parameters.

A fuzzy multipurpose decision making problem is studied in [22] to establish a general model that covers all possible representations by means of preference orderings, utility functions and preference relations. The process to verify the assumption and the evaluation of the resulting utility function including sufficient conditions for multi-attribute utility function is considered in [23]. Classical weighted arithmetic mean is common in analysis of the problems in decision making as presented in [24], where the Choquet integral is used as a weighted arithmetic mean aggregation tool after a classification process on the data set. Fuzzification of Choquet integral as a fuzzy number is discussed in [25]. Fuzzy measures and nonlinear integrals in data mining are proposed in set function identification, nonlinear multi-regression, nonlinear classification, networks, and fuzzy data analysis in [26].

In multi-criteria decision making problems, the theory of fuzzy measures has been used by many authors; firstly, fuzzy measure as a generalization of the classical probability measure was given by Sugeno. The classical probability measure theory has an important role in decision theory; consequently, fuzzy measure has been investigated by numerous researchers to determine the best decision for their problems. For example, Modave and Grabisch examined the associations between additive representation in decision making and measurement theory in order to propose a Choquet representation theorem in multi-criteria decision making [27]. Graphical explanations of the Choquet integral, viewed as an aggregation operator in the case of two elements are given by Grabisch in [28]. Numerous methods have been presented to construct Choquet integral-based utility function representing decision maker's preferences.

A methodology for building a non-additive utility function in terms of Choquet integral for multi-criteria problems is defined in [29]. An effective decision theory under uncertainty when the environment of fuzzy events and fuzzy states are characterized by imprecise probabilities is intended by authors in [30]. The theory is based on a non-expected fuzzy utility function represented by a fuzzy-valued Choquet integral with a fuzzy number valued fuzzy measure constructed from imprecise probabilities. In [31], a synthesis within the application is offered on the application of fuzzy integral as an innovative tool for criteria aggregation in decision problems.

A complexity-based method in order to construct fuzzy measures by the discrete Choquet integral is suggested in [32] to evaluate the student's performance based on a basic Competence test. Credit rating for commercial loans is an important issue for loan officers of a bank. In [33], a fuzzy credit-rating approach is proposed to deal with the problem arisen from the credit rating table used in Taiwan. The credit-rating criteria are modeled as hierarchical decision structures.

2.2 Statement of the Problem

The process which is carried out for a business or an individual application of credit to determine eligibility of the applicant for a loan is loan credit evaluation and approval. The loan may also be restricted to pay for goods and services over an extended period. An important factor in evaluation process of business loan credit applications is the credit worthiness, which score intends to measure history of trustworthiness, moral character, and expectations of continued performance of the financial credit applicant. However, for the consumer credit applications, it is difficult to collect sufficient information about the applicants to evaluate their credit worthiness score. This work attacks to the problem of the credit evaluation and approvals by modeling the finalized expertise decision based on a set of applicant attributes, such as income, age, credit history, requested loan amount, etc, which are considered by the financial institutions important in evaluating credit applications. The aim of the model is to obtain rules in terms of the applicant attributes to predict the expert decision for a new credit applicant and, to score the credit eligibility of the applicants for the applied loan. Note that scoring the credit eligibility is also called the credit-scoring, and it is valid only for the applied loan, whereas the credit worthiness scores intent to measure the general credit eligibility of the applicants.

This thesis focuses mainly on a fuzzy-valued Choquet integral based consumer credit evaluation model to explore the evaluation criteria. The expert prevision is explored from a set of consumer applications for which experts provided an evaluation decision using fuzzy clustering and transferred its results into a fuzzy linguistic rule base. The fuzzy linguistic rule base is used to determine imprecise prevision of the experts in the form of fuzzy upper and lower imprecise probabilistic measure of rule prototypes. It extracts the experts prevision using fuzzy clustering by describing the expert prevision through a set of flexible fuzzy linguistic rules. The financial institutions mostly request a ranking of the applicants rather than only a final decision of denied or accepted for each applicant. However, experts typically classify the applicants denied or accepted.

A utility score provides flexibility in finalizing the decision such as which applicants might be accepted when the financial resources are increased; as well as it provides a verification of experts decision. If an applicant's utility score is low but experts decided for it accepted, or vice versa, such inconsistent cases may be detected and re-evaluated by the experts to prevent any material mistakes in the decision.

2.3 Preliminaries

Various fuzzy methods including Fuzzy Clustering have been developed for decision making by following the introduction of Fuzzy Set Theory by L. A. Zadeh [34]. Clustering is one of the important tools in decision making as well as in pattern recognition, data mining, and data modeling. Clustering partitions a data set according to a similarity measure for the objects in the data set.

Clustering a data set into a number of partitions may explore the general characteristic relations between the arguments of the data set which contains sufficiently rich samples of a decision process. The main purpose of clustering is to divide the given data into homogeneous clusters according to the similarity [35].

Unsupervised learning is achieved by natural grouping or meaningful partition of similar data items in a data set. The extraction of knowledge from a data set by clustering is called unsupervised learning if clustering is based on a similarity measure rather than corrective actions supervised by the known relations [36]. Similarity is fundamental to the definition of clustering. The clustering of input-output data provides unsupervised learning by collecting similar data vectors into the same cluster [37].

In clustering data sets with scalar type attributes commonly a distance measure is used as a similarity measure. A similarity measure of the vectors may be established by various different representations of distance. Different distance measures are commonly used in fuzzy clustering algorithms [38], [39]. One of the types of similarity measures in the fuzzy clustering is the distance measure; different representations of distance are used to establish the similarity in the clustering. The following list of distance measures has been used in different sources.

i. Euclidean distance, where $d_2(a, b) = (\sum_{i=1}^{s} (a_i - b_i)^2)^{1/2}$.
ii. Minkowski distance, where $d_p(a, b) = d_p(a, b) = [\sum_{i=1}^{s} |a_i - b_i|^p]^{1/p}$.

iii. Mahalanobis distance, where $d_A(a,b) = \sqrt{(a-b)^T A^{-1}(a-b)}$ and A^{-1} represents the inverse of the covariance matrix of each cluster.

iv. Hamming distance, $d_1(a,b) = \sum_{i=1}^{s} |a_i - b_i|$.

v. Maximum distance, $d_\infty(a,b) = max_{i=1,\dots,x} |a_i - b_i|$.

FCM clustering algorithm uses Euclidean distance to cluster data set. Mainly, the algorithm, the distance measure, and the character of application affect the result of clustering. Three main types of clustering methods are proposed: partitioning, hierarchical, and fuzzy methods. In partitioning clustering methods, the data set is divided into a preset number of clusters. K-Means [40] is a standard partitioning clustering method based on K centroids of a random initial partition. Hierarchical clustering methods generate a hierarchy between clusters: small clusters include very similar items to combine the larger clusters which contain relatively dissimilar items, and produce a clustering space represented by a dendogram to show the hierarchical cluster structure.

Fuzzy clustering is an algorithmic fuzzy data analysis approach, generally obtained by fuzzification of classical algorithms using the fuzzy set theory [41]. In contrast to the classical set theory, where an object either belongs to a set or not, in the fuzzy set theory an object belongs to a set partially with the degree of membership between 0 and 1, [34]. There are two fundamental methods of fuzzy clustering; a Fuzzy c-Means Clustering method based on fuzzy c-partitions and Fuzzy Equivalence Relation method based on hierarchical clustering method [42].

Ruspini established an algorithm for hard c-mean partitioning which divides data to *c* number of clusters by assigning each data item to exactly one cluster to illustrate the cluster structure of a given data set [43]. He also recommended an algorithm for fuzzy partition. Dunn generalized this clustering process to a fuzzy ISODATA clustering technique, and Bezdek used Dunn's process to develop FCM algorithm where a data object belongs to all fuzzy clusters with different degrees of membership. FCM is a method to find out the fuzzy representation of a data set by partitioning each item into *c* fuzzy clusters [44].

Fuzzy measure and Choquet integral are the generalizations of the classical probability measure and Lebesque integral, respectively. The concept of fuzzy measure and the theory of fuzzy integral based on fuzzy measure have been established by Sugeno using min and max operators. Later on, fuzzy measures and fuzzy integrals were discussed in different sources, such as [45], [46], [47] and [48]. In this section, the definition of fuzzy function, fuzzy measure and fuzzy utility functions are represented by the Choquet integral.

The notations that will be used throughout this section are: R is the set of real numbers, $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ denotes the universal set, and $P(\mathcal{X})$ is a non-empty family of subsets of $\mathcal{X}$ including $\emptyset$ and $\mathcal{X}$. $(\mathcal{X}, P(\mathcal{X}))$ is a measurable space and $P(\mathcal{X})$ is a σ -algebra which is defined on the non-empty set $\mathcal{X}$.

Suppose $\prec$ is a preorder then 0 is a minimal element in $\mathcal{X}$ and 1 is a maximal element in $\mathcal{X}$. And, also suppose E^n is the space of all fuzzy subsets of R^n satisfying the conditions of normality, convexity, and upper continuous with compact support; i.e., $E^1_{[a,b]}$ denotes the space of fuzzy sets of $[a,b] \in R$.

Definition 2.3.1 Let $\mathcal{X}$ be non empty set and let $P(\mathcal{X})$ be non-empty family of subsets of $\mathcal{X}$. $P(\mathcal{X})$ is called σ –algebra of $\mathcal{X}$ if it satisfies the following three properties:

i. $\emptyset, \mathcal{X} \in P(\mathcal{X})$.
ii. Let $A \in \mathcal{X}$. If $A \in P(\mathcal{X})$ then the complement of A is also in $P(\mathcal{X})$.
iii. If $A_1, A_2, \ldots \in P(\mathcal{X})$ then $\cup_{n=1}^{\infty} A_n \in P(\mathcal{X})$.

Definition 2.3.2 Let $\tilde{A}$ be fuzzy set defined in $\mathcal{X}$ with membership function $\mu_{\tilde{A}}(x) : \mathcal{X} \to [0, 1]$ for every x in $\mathcal{X}$. The fuzzy complement $\tilde{A}^C$ of $\tilde{A}$ is defined by membership function $\mu_{\tilde{A}^c}(x) = 1 - \mu_{\tilde{A}}(x)$ for every x in $\mathcal{X}$. $\mu_{\tilde{A}}(x)$ is defined as the degree such that x belongs to $\tilde{A}$, therefore, $\mu_{\tilde{A}^c}(x)$ is defined as the degree such that x does not belong to $\tilde{A}$. In classical set theory, $A \cap A^C = \emptyset$ and $A \cup A^C = U$ for crisp sets, however, these properties do not satisfy for fuzzy sets [49].

Definition 2.3.3 A set function $\mu : P(\mathcal{X}) \longrightarrow [0,1]$ is called a fuzzy measure on measurable space if it satisfies the following statements [31]:

i. $\mu(\emptyset) = 0, \mu(\mathcal{X}) = 1$
ii. If $A \subseteq B$ then $\mu(A) \leq \mu(B)$ for all $A, B \in P(\mathcal{X})$.

μ is a normalized monotonic set function. For $A,\ B \subseteq P(\mathcal{X})$, a fuzzy measure μ is additive if $\mu(A \cup B) = \mu(A) + \mu(B)$, where $A \cap B = \emptyset$.

Definition 2.3.4 Let $\mathcal{X}$ be a nonempty fuzzy set and let $\tilde{F}(\mathcal{X}) = \{\, \tilde{B} \mid \mu_{\tilde{B}} : \mathcal{X} \longrightarrow [0,1] \,\}$ be the class of all fuzzy subsets of $\mathcal{X}$. And, let $\tilde{F}_1(\mathcal{X})$ be subclass of $\tilde{F}(\mathcal{X})$. $\tilde{F}_1(\mathcal{X})$ is a fuzzy σ −algebra if the following properties are satisfied, [50] :

i. $\emptyset, \mathcal{X} \in \tilde{F}_1(\mathcal{X})$, where $\emptyset(x) = 0$ and $\mathcal{X}(x) = 1$ for every x in $\mathcal{X}$.

ii. if $\tilde{B} \in \tilde{F}_1(x)$, then the complement of $\tilde{B}$ is also in $\tilde{F}_1(x)$, i.e. $\tilde{B}^C \in \tilde{F}_1(\mathcal{X})$

iii. if $\{\tilde{B}_n\} \subset \tilde{F}_1(\mathcal{X})$, then $\cup_{n=1}^{\infty} \tilde{B}_n \in \tilde{F}_1(\mathcal{X})$.

A *signed fuzzy measure* μ is a set function $\mu : \tilde{F}_1(\mathcal{X}) \longrightarrow (-\infty, \infty)$ that satisfies $\mu(\emptyset) = 0$.

Definition 2.3.5 A fuzzy number is a fuzzy set $\tilde{a} : R \longrightarrow [0,1]$ satisfying the properties [49]:

i. $\tilde{a}$ is normal, that is, there exists and x in R such that $\tilde{a}(x) = 1$.

ii. $a_r = \{\, x \mid \tilde{a}(x) \geq r \,\}$ is closed interval $[a_r^-, a_r^+]$, $r \in (0,1]$.

A fuzzy infinity denoted by $\tilde{\infty}$ is a fuzzy number satisfying the condition that for every positive real number M, there exists $r_0 \in [0,1]$ such that $a_{r_0}^- < -M$ or $M < a_{r_0}^+$

Definition 2.3.6 Let $\tilde{A},\ \tilde{B}$ be two fuzzy sets in E^n. The Hausdorff distance of $\tilde{A}$ and $\tilde{B}$ is given by

$$\tilde{d}_{fH}(\tilde{A}, \tilde{B}) = \int_{r\in[0,1]} r/[d_H(A^{r=1}, B^{r=1}),\ sup_{r\leq\bar{r}\leq 1} d_H(A^{\bar{r}}, B^{\bar{r}})].$$

Definition 2.3.7 Let F_1 be a set of fuzzy numbers. A fuzzy number-valued fuzzy measure (also called z −fuzzy-measure) on $\tilde{F}_1(\mathcal{X})$ is a fuzzy number-valued fuzzy set function $\tilde{\mu} : \tilde{F}_1(\mathcal{X}) \longrightarrow E^1$, where E^1 denotes the space of fuzzy set of $[a, b] \subset R$ with the following properties [50]:

i. $\tilde{\mu}(\emptyset) = 0$;

ii. if $\tilde{B} \subseteq \tilde{C}$, then $\tilde{\mu}(\tilde{B}) \leq \tilde{\mu}(\tilde{C})$

iii. if $\tilde{B}_1 \subseteq \tilde{B}_2 \subseteq \cdots$, and $\tilde{B}_n \in \tilde{F}_1(\mathcal{X})$, then $\tilde{\mu}\left(\bigcup_{n=1}^{\infty} \tilde{B}_n\right) = \lim_{n\to\infty} \tilde{\mu}(\tilde{B})$

iv. $\tilde{B}_1 \supseteq \tilde{B}_2 \supseteq \cdots$, $\tilde{B}_n \in \tilde{F}_1(\mathcal{X})$ and there exists n_0 such that $\tilde{\mu}\left(\tilde{B}_{n_0}\right) \neq \tilde{\infty}$ then $\tilde{\mu}\left(\bigcap_{n=1}^{\propto} \tilde{B}_n\right) = \lim_{n\to\infty} \mu(\tilde{B}_n)$.

In the definition above the limits are defined according to the fuzzy Hausdorff distance, [51].

Other notations that will be used throughout this section are: let $\tilde{A}$ and $\tilde{B}$ be fuzzy sets in E^n, where E^n is the space of all fuzzy subsets of R^n. Meanwhile, $(\mathcal{X}, \tilde{F}_1(\mathcal{X}))$ is called a fuzzy measurable space, and $(\mathcal{X}, \tilde{F}_1(\mathcal{X}),\ \tilde{\mu})$ is called a *z-fuzzy-measure* space.

Definition 2.3.8 Let $(\mathcal{X}, \tilde{F}_1(\mathcal{X}), \tilde{\mu})$ be a z-fuzzy-measure space. $f: \mathcal{X} \longrightarrow (-\infty, \infty)$ is called a fuzzy measurable function if $\mathcal{X}_{F_\beta} \in \tilde{F}_1(\mathcal{X})$ where $F_\beta = \{x \in \mathcal{X} \mid f(x) > \beta\}$ and $\mathcal{X}_{F_\beta}(x)$ =(1 if and only if $x \in F_\beta$); and 0 if and only if $x \notin F_\beta$) with $b \in (-\infty, \infty)$. M' denotes a set of all fuzzy measurable functions, and M'_+ denotes a set of non-negative fuzzy measurable functions.

Let N_R denote the set of all closed intervals of the real line. $\bar{f}: \mathcal{X} \longrightarrow N_R$ is fuzzy measurable if both $f_1(x) = [\bar{f}(x)]_1$, the left end point of interval (x), and $f_2(x) = [f(x)]_2$, the right end point of interval (x) are fuzzy measurable functions of x.

Fuzzy integral is an operator on [0,1] which is used to solve the multi-criteria decision problems and also it is used in many applications, such as [52], [53]. While there are two well-known types of fuzzy integrals for utility evaluation: Sugeno fuzzy integral and Choquet integral; we only focus on the Choquet integral for a positive and measurable function.

Definition 2.3.9 Let $\mathcal{X}$ be a non empty set. And, let $P(\mathcal{X})$ be an σ-algebra defined on $\mathcal{X}$. $\mu(\mathcal{X}) : P(\mathcal{X}) \longrightarrow [0,1]$ is a Sugeno fuzzy measure if the following conditions hold [26]:

i. $\mu(\emptyset) = 0, \mu(\mathcal{X}) = 1$
ii. If $A_1 \subseteq A_2$ then $\mu(A_1) \leq \mu(A_2)$ for all $A_1, A_2 \in P(\mathcal{X})$.
iii. If $A_1 \subseteq A_2 \subseteq \cdots \in P(\mathcal{X})$ then $\lim_{n\to\infty} \mu(A_n) = \mu(\lim_{n\to\infty} A_n)$.

Definition 2.3.10 Let μ be the fuzzy measure satisfying given properties. The Sugeno integral $f: P(\mathcal{X}) \rightarrow (0, \infty)$ is defines as, [26],

$$\int f \, d\mu = \sup_{\alpha \in [0,\infty)} [\alpha \wedge \mu(F_a)] \text{ such that } F_a = \{ x \mid f(x) \geq \alpha \}$$

For an α-cut set of a function of a non-negative function f in $[0, \infty)$.

Definition 2.3.11 Choquet integral of a nonnegative function $h: P(\mathcal{X}) \longrightarrow R_0^+$ with respect to a fuzzy measure g on $\mathcal{X}$ is defined by

$$E_g(x) = \int_0^x g(H_\alpha) d\alpha,$$

where $H_\alpha = \{ x \in X \mid h(x) \geq \alpha \}$.

Choquet integral $E_g(h)$ for a finite X always exists, and it is a generalization of mathematical expectation if g is a probability measure. Proofs of the following important properties of $E_g()$ are available in [54]:

i. If $\alpha \in R_0^+$, then $E_g(\alpha) = \alpha$.

ii. If $h(x) < h'(x) \;\; \forall x \in X,$ then $E_g(h) \leq E_g(h')$.

iii. If $g(A) < g'(A)$ $, \forall A \subseteq X,$ then $E_g(h) \leq E_{g'}(h)$ for every h such that $h : \mathcal{X} \longrightarrow R_0^+$.

iv. $E_g(I_A) = g(A)$.

v. If $b, \, c \in R_0^+,$ then $E_g(c + bh) = c + bE_g(h)$ for every h, such that $h : X \longrightarrow R_0^+$.

$E_g(h)$ is an additive functional for probability measures.

Denote h_i a value of a function h at point $x_i \in X$. Then, the Choquet integral of h denoted by $E_g(h)$ is expressed as $E_g(h) = \sum_{i=1}^{n}\big((h_i - h_{(i+1)})g(A_{(i)})\big)$, where subscript i shows that the indices are permuted in order to have $h_{(1)} \geq h_{(2)} \geq \cdots \geq h_{(n)}$ and $h_{(n+1)} = 0$, $A_{(i)} = \{x_{(1)}, \dots, x_{(i)}\}$.

Two functions h and h' are called equi-ordered, and denoted by $h \simeq h'$, if and only if either h is a constant function or for each pair $x_i,\ x_j \in X$ such that $h_i \leq h_j$, it follows necessarily that $h_i{}' \leq h_j{}'$. For a fuzzy measure g on X, the proof of ordered additivity of the Choquet integral, $E_g(h) \leq E_g(h + h') = E_g(h) + E_g(h')$ if function h is equi-ordered with h', is presented in [55]. Consequently, although $E_g()$ is in general non-additive, it is additive for equi-ordered functions. Finally, Choquet integral $E_P()$ with the probability measure P, instead of an arbitrary fuzzy measure, corresponds to the mathematical expectation with respect to P, and is simply

$$E_p(h) = \sum_{i=1}^{n} p_i h_i\,, \qquad (2.1)$$

where $p_i = P(\{x_i\})$, where $i = 1, \dots, n$.

Let f be a classical function from $\mathcal{X}$ into $\mathcal{Y}$ and let $\mathcal{X}$ and $\mathcal{Y}$ be the domain and the range of f , respectively. In [41], Zimmermann stated that there are three categories of the fuzzy function as generalizations of the classical function $f : \mathcal{X} \longrightarrow \mathcal{Y}$.

Defining a fuzzy function we mean a function whose values are fuzzy numbers. Let $\mu_{\tilde{f}(x)}$ represent the membership function of the fuzzy number $f(x)$.

Definition 2.3.12 Fuzzy function $\tilde{f}$ is defined from $\mathcal{X}$ into the power sets $\wp(\mathcal{Y})$ in $\mathcal{Y}$, if and only if $\mu_{\tilde{f}(x)}(y) = \mu_{\tilde{R}}(x,y)$ for every (x,y) in $\mathcal{X} \times \mathcal{Y}$. For $0 < r \leq 1$, $\tilde{f}_1^r$ and $\tilde{f}_2^r$ are the level functions of $\tilde{f}$ so that $\tilde{f}_2^r$ denotes $sup\{ z \in dom(\mu_{\tilde{f}(x)}) : \mu_{\tilde{f}(x)}(z) \geq r\}$ and $\tilde{f}_1^r$ denotes $inf\{ z \in dom(\mu_{\tilde{f}(x)}) : \mu_{\tilde{f}(x)}(z) \geq r\}$, respectively.

Fuzzy Utility Function Utility of a decision is introduced by Bernoulli to measure of the risk connected to the decision, concluding that the future value of the decision is expected to be the sum of the products of probabilities of the consequences by their expected losses and gains [18]. Mathieu-Nicot introduced the concept of fuzzy expected utility [56], and Billot introduced a set of theorems to extend Ponsard`s result to a convex fuzzy utility proving that the preferences may be sorted in a convex fuzzy utility function [19].

The credit-scores of applicants target to sort the applicants according to their future contribution to the profit of financial institution. The unsupervised learning ability of the FCM provides the applicants to be partitioned into c fuzzy clusters, and assigns membership functions to the applicants for each cluster. This thesis proposes a method to construct fuzzy utility function of the credit applicants by using the FCM membership degrees $\mu_{i,k}$ of the applicants, and the accepted-rates $R_{a,i}$ of the partitions

$$R_{a,i} = N_{a,i}/N_i$$

where N_i is the total number of applicants those belong to partition i with the highest membership degree compared to the other partitions. And, $N_{a,i}$ is the number of accepted applicants those belong to partition i with the highest membership degree compared to the other partitions.

Each FCM partition corresponds to a characteristic class of risk factor with different expected success rate. Under the assumption of having expected success rate $E_{a,i}$ of partition i equal to $R_{a,i}$, and assuming that the membership values $\mu_{i,k}$ indicates the probability ending up in the partition, a non-additive fuzzy expected utility value is obtained by expression

$$U_{n,k} = \max_{i=1,\ldots,c} (\mu_{a,i,k}, R_{a,i}) \tag{2.2}$$

which means that an applicant cannot accumulate utility-scores from multiple partitions. Non-additive fuzzy expected utility is known as not a rational utility since the utility shall measure the sum of all benefits of the choice. However, in some cases the utilities may not be additive because of the future alternates are not independent, i.e., if one occurs the other cannot occur.

The axiomatic definition of a probability Let a space of all events $\mathcal{X}$ be given that (a) $P(A) \geq 0$ $f$or all A in $\mathcal{X}$. (b) $P(\mathcal{X}) = 1$. (c) $P(A \cup B) = P(A) + P(B)$ if A and B are disjoint sets in $\mathcal{X}$. The number $P(A)$ corresponds to the probability of an event A in $\mathcal{X}$. Then, the probability of an event A has the properties; (a) $0 \leq P(A) \leq 1$, (b) $P(A) = 1 - P(A^c)$, (c) $P(\emptyset) = 0$, (d) $P(\mathcal{X}) = 1$. The set A^c is $\{ \omega \in \chi : \omega \notin A \}$, $P(A) + P(A^c) = 1$. Let us suppose $\underline{P}(A)$ and $\overline{P}(A)$ are the lower and upper probabilities for an event A, respectively. Then, we have $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$, for every event A in $\mathcal{X}$. $\underline{P}(\emptyset) = \overline{P}(\emptyset) = 0$, and $\underline{P}(\mathcal{X}) = \overline{P}(\mathcal{X}) = 1$. If $A_1 \subseteq A_2$, this implies that $\underline{P}(A_1) \leq \underline{P}(A_2)$ and $\overline{P}(A_1) \leq \overline{P}(A_2)$. Also for all events A in $\mathcal{X}$, $\underline{P}(A) \leq P(A) \leq \overline{P}(A)$, $\underline{P}(A^c) = 1 - \overline{P}(A)$, A^c is complement of A.

Definition 2.3.13 If the sample space $\mathcal{X}$ is finite and the following three properties are satisfied

i. $\underline{P}(\emptyset) = \overline{P}(\emptyset) = 0$, and $\underline{P}(\mathcal{X}) = \overline{P}(\mathcal{X}) = 1$
ii. $A_1 \subseteq A_2$, implies that $\underline{P}(A_1) \leq \underline{P}(A_2)$ and $\overline{P}(A_1) \leq \overline{P}(A_2)$
iii. $\underline{P}(A) \leq P(A) \leq \overline{P}(A)$ for all events $A \in \mathcal{X}$

then the functions $\underline{P}(A) = \inf(P(A))$ and $\overline{P}(A) = sup\,(P(A))$ are called lower probability, upper probability measures, respectively, [57].

Application of probability theory has drawbacks in the problems involving in evaluation of expert decisions since the expert's perception has nonlinear character which is not suitable for probabilistic utility calculations. Perceptions are both imprecise, and fuzzy in character. Fuzzy Choquet integral of imprecise probabilities is a successful tool for determination of non-expected utility [58].

Let $X = \{\, x_k \mid k = 1, \dots, n \,\}$ be a set of input vectors, with the corresponding set of output $Y = \{\, y_k \in \{0,1\} \mid k = 1, \dots, n \,\}$. Let β_i denote the membership value of $x_i \in X$ in fuzzy set B. The fuzzy mean of the output based on B is given by

$$P_{mean(y=1)} = \left(\sum_{i=1}^{n_D} \beta_{i,k} y_k \right) / \left(\sum_{i=1}^{n_D} \beta_{i,k} \right) \tag{2.3}$$

The lower probabilities of $(y = 1)$ can be evaluated for the available α −cut points of fuzzy set,

$$P_{lower(y=1)}(\alpha = \beta_k) = \left(\sum_{\beta_{i,k} \le \alpha} \beta_{i,k} y_k \right) / \left(\sum_{\beta_{i,k} \le \alpha} \beta_{i,k} \right) \tag{2.4}$$

The lower edge of the fuzzy imprecise probability set may be easily obtained by curve fitting on the evaluated points $P_{lower(y=1)}(\alpha = \beta_k),\ k = 1, \dots, n.$

For the upper edge of the fuzzy probability of $(y = 1)$, we may use a curve-fit on the points calculated by the $(y = 1)$ cases above the α –cut

$$P_{upper(y=1)}(\alpha = \beta_k) = \left(\sum_{\beta_{i,k} \ge (1-\alpha)} \beta_{i,k} y_k \right) / \left(\sum_{\beta_{i,k} \ge (1-\alpha)} \beta_{i,k} \right) \tag{2.5}$$

Definition 2.3.14 Let $\overline{f} : X \longrightarrow N_R$ be a fuzzy measurable interval-valued function on X and $\overline{\mu}$ be a fuzzy number-valued fuzzy measure on $\overline{f}$. The Choquet integral of $\overline{f}$ with respect to $\overline{\mu}$ is defined by

$$\overline{E}_g = \{ \int f d\overline{\mu} \,|\, f(x) \in \overline{f}(x),\ \forall x \in X,\ f : X \longrightarrow R \text{ is measurable} \}$$

$$= \int \overline{f}\, d\overline{\mu} \tag{2.6}$$

A fuzzy valued function $\overline{f} : X \longrightarrow E^n$ is fuzzy measurable if the r-cut $\overline{f}^r(x) = \{ y \mid \overline{f}(y) \geq r \}$ that belongs to the fuzzy measure is a fuzzy measurable interval-valued function for every $r \in (0,1]$, where $\overline{f}(y)$ is the membership value of $\overline{f}$ at y.

One of the advantages of Choquet integral is a possibility to illustrate decision making under uncertainty. A Choquet integral may be taken at only individual level which is equivalent to ordinary weighted average, or considering proper weights for the couplets of permutations, triplets of permutations, etc.

Definition 2.3.15 The fuzzy ranking is the method of comparing fuzzy numbers. One way to compare fuzzy numbers is to convert a fuzzy number to a crisp number by concerning a mapping function, i.e. if A is a fuzzy number, then $F(A) = a$, where a is a crisp number.

The aim of such a fuzzy ranking is to express the best scores of decision making problems by crisp preferences of alternatives, such that in many cases final scores of alternatives are represented in terms of fuzzy numbers. There are several methods to sort fuzzy numbers by ranking crisp numbers; each one has advantages besides disadvantages. For example in [42], fuzzy numbers are compared by defining the Hamming distance, and by determining α-cut and also through the extension principle.

Chapter 3

DATA SET AND DATA SET ANALYSIS

3.1 Main Variables and the Data Set

In this thesis, our plan is to estimate a non-expected (non-additive) utility of each consumer loan credit applicant using an available dataset that contains the expert decision attributes as the input features, and the binary expert decision result (*accepted or denied*) as output of the application.

Consumer Loan Data Set The analyzed data set is obtained from a finance institution which provides credit to appropriate applicants. The available data set contains totally $n_D = 135$ cases of complete input features of the credit loan applicants, and corresponding expert decisions. Each case composes a data vector (x_k, y_k) with the following 10 attributes:

1. Net income (USD), scalar
2. Age (Years), scalar
3. Last employment period (Years), scalar
4. Credit history (Negative, Positive)
5. Purpose of loan (General purpose, Flat refurbishment, Car purchase, Flat purchase)
6. Requested loan amount (USD), scalar
7. Loan-maturity (Years), scalar

8. Proposed number of guarantors (Number), scalar
9. Collateral (None, Not applicable, Car, Flat)
10. Expert decision (Denied, Accepted).

From these ten attributes, the first nine are input attributes $x_k = (x_{k,1}, \dots, x_{k,nk})$, $n_X = 9$, where k stands for the k^{th} applicant, and the last attribute states the experts decision for the case: either accepted ($y_k = 1$, total of 103 applicants) or denied ($y_k = 0$, total of 32 applicants). Some statistical properties of the input and output such as minimum, maximum, mean and standard deviation are calculated for the consumer loan data set. Then, we converted the nominal attributes into numerals and normalized all attributes into the $[0,1]$ range to avoid anomalies of large difference between the ranges of each attribute before applying fuzzy c-means (FCM) to partition the data set into n_C fuzzy clusters.

3.2 Fuzzy Clustering Algorithm

Fuzzy c-means (FCM) is a well-known fuzzy clustering algorithm to cluster a numerical data set into c clusters [59]. FCM clusters a finite set of data vectors $X = \{x_1, x_2, \dots, x_k, \dots, x_n\} \subset \mathcal{R}^p$, where the dimension of the vector space is P. A fuzzy c partition $\wp = \{\mu_1, \mu_2, \dots, \mu_k, \dots, \mu_n\}$ is a family of subsets of X. $\wp$ is the class of fuzzy sets such that μ_{rk} denotes the value of the degree of membership of object $x_k = \{x_{k,1}, x_{k,2}, \dots, x_{k,n}\} \in \mathcal{R}^p$ in the i-the partition, for all $k = 1, 2, \dots, n$ and $i = 1, \dots, c$. A fuzzy c-partition $\wp$ of the given set of data X satisfies the constraints:

$$\sum_{i=1} \mu_i(x_k) = 1 \text{ for every } x \in X$$

and

$$0 < \sum_{x \in X}^{c} \mu_i(x_k) < 1 \quad \text{for } i = 1,\ldots,c. \tag{3.1}$$

The vectors $v_1, v_2, \ldots, v_k, \ldots, v_c$ are the cluster centers corresponding to each cluster c and $d(x, v_i)$ is the distance function.

The matrix $U = [\mu_{i,k}]_{cxn}$ is the cxn fuzzy partition matrix of X. In other words, this is the matrix of degrees of memberships of data objects x_k, $k = 1, \ldots, n$ in each cluster i, where $i = 1, \ldots, c$. The aim of the FCM algorithm is to find the best possible fuzzy partitions that minimize the objective function J such that the objective function is defined by

$$J(U,V) = \sum_{i=1}^{c} \sum_{k=1}^{n} \left(\mu_{ik}(x)\right)^m \|x_k - v_i\|^2 , \tag{3.2}$$

where $m \in (1, \infty)$ is the fuzzification power, and $\|x_k - v_i\|^2$ is the distance function between x_k and v_i, [42], [44]. In order to minimize the objective function J, the c fuzzy cluster centers $v_1^{(t)}, v_2^{(t)}, \ldots, v_c^{(t)}$ using the fuzzy $c-$partition matrix $U^{(t)}$ are calculated by

$$v_i^{(t)} = \frac{\sum_{k=1}^{n} \left[\mu_{j,k}^{(t)}\right]^m x_k}{\sum_{k=1}^{n} \left[\mu_{j,k}^{(t)}\right]^m}, \quad i = 1,\ldots,c \tag{3.3}$$

and also the fuzzy c −partition matrix $U^{(t+1)}$ is revised using the c fuzzy cluster centers from

$$\mu_{i,k}^{(t+1)} = [\sum_{j=1}^{c} (\frac{d(x_k, v_i^{(t)})}{d(x_k, v_j^{(t)})})^{\frac{2}{m-1}}]^{-1}. \tag{3.4}$$

The number of clusters c and the fuzzification power m of FCM have an important role in unsupervised learning of the relations in a data set. Fuzzification power m close to 1 makes the clustering to approach nearly too hard clustering, and higher m values makes the clusters more and more fuzzily mixed to each other. Typically m values near 2 are satisfactory for proper generalization in unsupervised learning. If c is decided appropriately each cluster center corresponds to a prototype in the data set.

In the literature there are many proposals for fuzzy validity functions to test the validity of the partitions generated by FCM [59]. However, recently researchers avoid cluster validity indices by searching the best performing c and m value in a set of specific cases [60], [61] and [62].

The following steps can be applied for Fuzzy C-Means (FCM) algorithm, [41], [42]. Suppose m and e are real, small positive numbers, respectively. At the beginning the real number m and a small positive number ϵ are selected to terminate this algorithm.

Step 1: Compute the fuzzy c −partition matrix $U^{(0)}$, i.e. compute $\mu_1^{(k)}, \mu_2^{(k)}, \cdots, \mu_c^{(k)}$ for $k = 0$ according to the given c.

Step 2: Then, compute the c fuzzy cluster centers $v_1^{(k)}, v_2^{(k)}, \cdots, v_c^{(k)}$ using the fuzzy $\mu_1^{(k)}, \mu_2^{(k)}, \cdots, \mu_c^{(k)}$, c- partition matrix $U^{(k)}$ obtained by

$$v_i^{(t)} = \frac{\sum_{k=1}^{n} (\mu_{j,k}^{(t)})^m x_k}{\sum_{k=1}^{n} (\mu_{j,k}^{(t)})^m} \qquad \text{for } i = 1, \ldots, c.$$

Step 3: revise the fuzzy c-partition matrix $U^{(k+1)}$ using the c fuzzy cluster centres from

$$\mu_i^{(k+1)}(x) = \left[\sum_{i=1}^{c} \left(\frac{d\left(x, v_i^{(k)}\right)}{d\left(x, v_j^{(k)}\right)} \right)^{\frac{2}{m-1}} \right]^{-1}$$

for each $i = 1, \ldots, c$, if $d(x, v_j^{(x)}) \neq 0$, $i = 1, \ldots, c$.

But if $d\left(x, v_j^{(x)}\right) = 0$ then $\mu_{jk} = 1$ when $j = i$ or $\mu_{jk} = 0$ when $j \neq i$.

Step 4: If $\left| v_j^{(k+1)} - v_j^k \right| \leq \varepsilon$ then terminate, otherwise repeat step 2.

Chapter 4

CONTINUOUS MODEL OF CREDIT SCORING

4.1 Validity Test for the Best Model

A fuzzy credit-scoring model requires trimming of structural and non-structural parameters to a training data set. For FCM based modeling cluster validity indices may be a remedy to decide on structural parameters such as the number of clusters c, and the fuzzification power m of the model. But, the cluster validity indices only validates that the vectors are concentrated in the neighborhood of the cluster centers rather than validating the accuracy of predicting the unknown target attribute.

Mosteller introduced k-sample method of validating for the significance test [63]. Pickard and Cool introduced the cross validation method based on splitting the data set to training and verification partitions [64]. The major drawback of Pickard's cross validation is a significant reduction of the training data, which is not tolerable for small data sets. In most application the evaluation of validation tests of the models are carried statistically by k-Fold cross-validation (CV) method. k-Fold CV is based on construction of k models using randomly partitioned training and verification data sets in a moving window pattern as described by Mosteller.

A special case of the k-Fold CV is called leave-one-out-cross-validation (LOOCV), where the validation is based on total n, each case reserving a single training vector for validation purpose and using all others for training. LOOCV is known statistically most sound cross verification method. But, it requires highest computational effort compared to k-Fold CV when $k < n$.

The evaluation of k models which are obtained by k-Fold CV method is carried using accuracy measures on candidate models. The primary cost function is based on the fail rate in estimating the positive and negative expert decisions. However, many models give exactly same fail-rate due to exactly same fail counts. This thesis introduces two measures of modeling accuracy. The first measure J_m is based on the maximum of mean of centers

$$\overline{v}_i = \frac{1}{k}\sum_{j=1}^{k} v_{j,i} \tag{4.1}$$

of k-Fold ensemble of FCM cluster-center vectors $v_{j,i}$ where $j = 1, \dots, k$ points the k-Fold models, $i = 1, \dots, c$ is the FCM cluster number.

$$J_m = \max_{i=1,\dots,c} (\overline{v}_{i,d}(1-\overline{v}_{i,d})) \tag{4.2}$$

where $\overline{v}_{i,d}$ is the expert-decision component of $\overline{v}_i$. For both negative decision ($\bar{v}_{i,d} = 0$) and positive decision ($\bar{v}_{i,d} = 1$), the product $\bar{v}_{i,d}(1 - \bar{v}_{i,d})$ is zero, and smaller J_m indicates corresponding clusters that are more significant. Our second cost measure J_σ is the maximum of mean of standard deviation of decision attribute of the cluster centers, $v_{j,i,d}$, for all k-Fold CV models.

$$\sigma_{f,i}^2 = \frac{1}{k}\sum_{j=1}^{k} (v_{j,i,d} - \overline{v}_{i,d})^2 \tag{4.3}$$

and

$$J_\sigma = \max_{i=1,\dots,c} (\sigma_{f,j,k}) \qquad (4.4)$$

The fail-rate is the most significant cost measures among these three cost measures, J_m is based on the mean of fuzzy means of the decision attributes in each cluster, and expected to have higher significance than J_σ, which is based on the variance of the fuzzy means of decision attributes.

4.2 Fuzzy Utility Function Construction

FCM cannot be applied directly on the data set since many of the attributes are nominal. In the preprocessing phase of the process the nominal attributes were converted to numerals by assigning an integer number to each nominal symbol starting from zero. Thus, preprocessing converts the "negative" credit history to "0", and "positive" to "1". Similarly, "General purpose" in purpose of loan is replaced by "0", "Flat refurbishment" by "1", "Car purchase" by "2", and "Flat purchase" by "3". Similar replacements are applied to collateral and credit decision attributes as well. Finally, preprocessing splits the data set randomly to training and verification sets, to concern k −fold cross validation with $k = n$, which is called leave-one-out-cross-validation (LOOCV).

Figure 1 shows the process diagram to obtain FCM based decision model from the training data set.

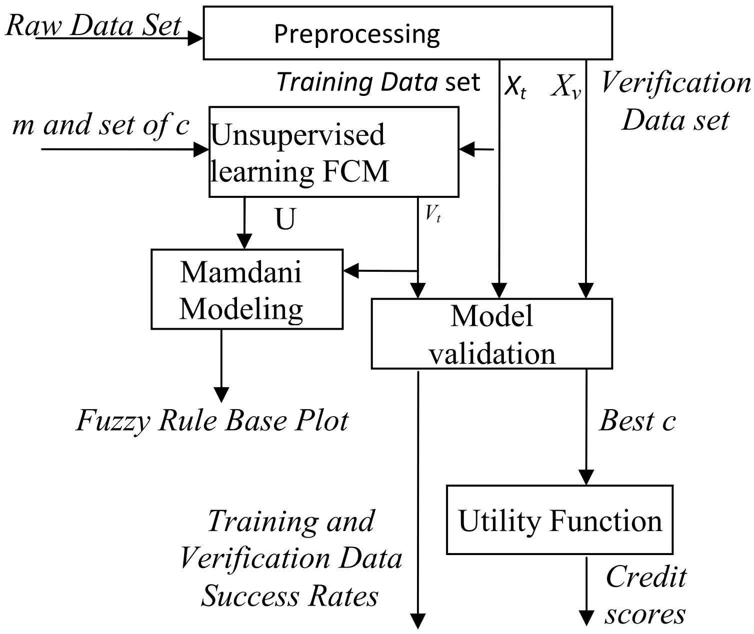

Figure 1. Process diagram of decision making

FCM is applied to the processed data set for $m = 2$ and all values of $c \in \{2, 3, \ldots, 15\}$ for 120 times with random initialization of cluster-centers to reduce the effects of ill FCM initialization which has been described by Hathaway et.al. [65]. The FCM result with the smallest FCM cost, equation 3.2, is used for the modeling. In parallel to Mamdani modeling, FCM membership expression, equation 3.4, is applied for decision making to decide on $x_{k,p}$ using the success rates of the dominant cluster, where x_k belongs with highest membership value. However, since the decision attribute $x_{k,p}$ is unknown for the verification data set, an equal- located-parameter $\alpha = 0.5$ (in between 0 and 1) is used in place of $x_{k,p}$, as proposed by Mosteller [63].

The training and verification fail-rates for positive and negative credit decisions are shown in Table 1, where % fail is the sum of percent fail of denied and percent fail of accepted objects.

$$\% fail = 100\left(N_{ap}/N_a + N_{dp}/N_d\right), \tag{4.6}$$

N_{ap} and N_{dp} are predicted numbers of accepted and denied objects. Also, N_a and N_d are actual number of accepted and denied objects of all training vectors or verification vectors. J_m and J_σ are the secondary cost measures for the estimation accuracy.

At the same time, Table 1 indicates the lowest percent verification fail-rate and lowest secondary costs obtained with $c = 7$. The model with $c = 7$ appears to be the best performing model among all models according to the primary and secondary cost functions J_m and J_σ .

A conventional visual fuzzy rule base of the training data is obtained using the Mamdani modeling of the data set [66], which is used and described by [67] and [60]. The fuzzy rule base of Mamdani type fuzzy model of the training data is determined by the cluster centers $\{v_1, v_2, \dots, v_c\}$, training data vectors X, and FCM partition matrix U. Table 2 demonstrates the results of FCM cluster-centers with $c=7$. The visual Mamdani Fuzzy-Rule-Base for the best performing model ($c=7$) is shown in Figure 2.

Table 1. The fail counts for denied and accepted applicants

C	Fail events in training			Fail events with verification				
	for deny	for accept	% Fail	for deny	For accept	% fail	J_m	J_σ
2	46	1	53.2%	46	1	53.2%	0.2454	0.0231
3	11	3	22.9%	12	5	31.4%	0.1213	0.0043
4	7	6	29.7%	7	9	40.9%	0.0776	0.0779
5	0	16	59.3%	0	18	66.7%	0.1710	0.1131
6	10	2	18.2%	11	3	22.9%	0.0749	0.0031
7	8	2	16.0%	10	2	18.2%	0.0577	0.0022
8	8	3	19.7%	10	4	25.6%	0.0633	0.0109
9	4	3	15.4%	8	4	23.4%	0.0659	0.0206
10	4	3	15.4%	5	5	23.9%	0.1250	0.2327
11	4	3	15.4%	6	4	21.3%	0.2496	0.4376
12	5	3	16.5%	10	7	36.7%	0.1113	0.2261
13	3	9	36.6%	8	11	49.3%	0.2390	0.4354
14	5	7	31.3%	4	12	48.7%	0.1650	0.3673
15	6	3	17.6%	8	10	45.6%	0.0962	0.2696

The cross sectional plot of FCM membership values along each attribute of the input vectors at each cluster-center is shown in Figure 3. It forms a clear and direct visual representation of the membership expression of FCM for a particular cluster-center matrix V_t.

The sharp peaks on the fuzzy sets of the 3, 4 and 5th rules are placed to display the cluster-centers. These three rules are weak in membership values, and also there are fewer objects with maximum membership values in them.

Table 2. FCM cluster-centers with $c = 7$

i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
1	0.18	0.25	0.33	0.10	0.41	0.13	0.83	0.84	0.43	0.03
2	0.07	0.14	0.13	0.96	0.50	0.09	0.83	0.85	0.51	0.06
3	0.31	0.39	0.53	0.98	0.50	0.24	0.59	0.87	0.76	0.94
4	0.32	0.39	0.53	0.98	0.50	0.24	0.59	0.87	0.76	0.94
5	0.32	0.39	0.53	0.98	0.50	0.24	0.59	0.87	0.76	0.94
6	0.18	0.22	0.28	0.99	0.58	0.18	0.83	0.95	0.71	0.96
7	0.48	0.64	0.70	0.99	0.95	0.51	0.85	0.93	0.97	0.98

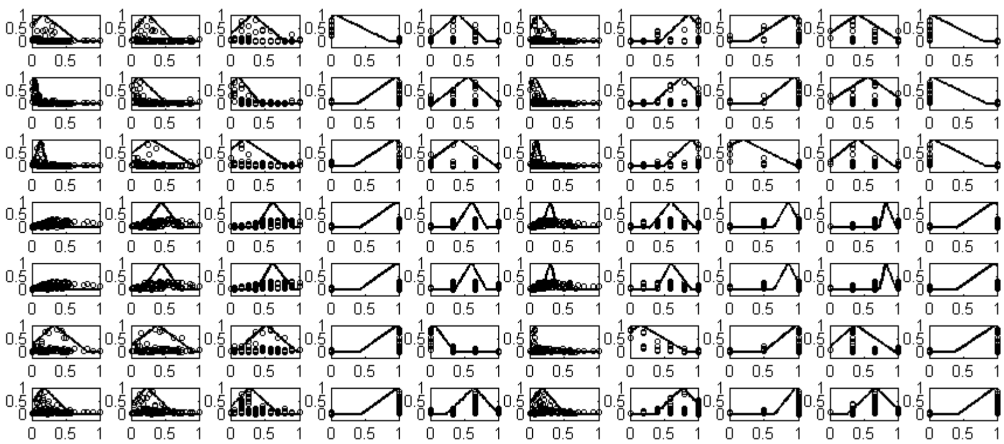

Figure 2. Mamdani rule base of decision model for $c=7$

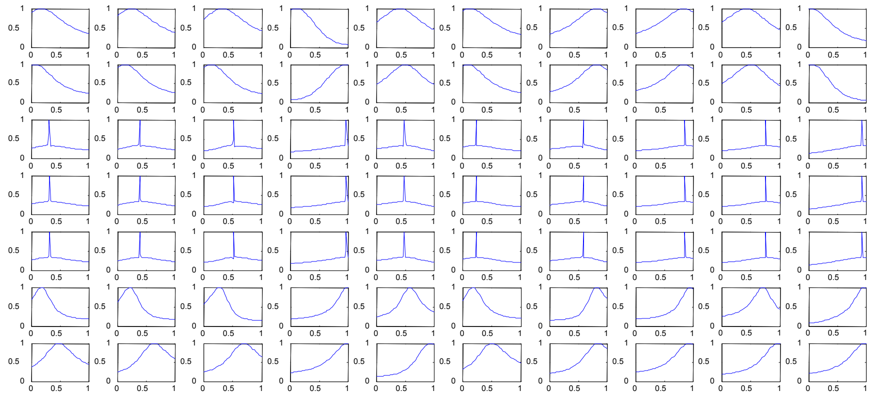

Figure 3. Cross sectional plots of FCM membership expression for $c=7$

The additive and non-additive utility-scores of the applicants shown in Figure 4 are obtained by equations 4.1 and 4.2. In Figure 4, squares indicate predicted, circles indicate expert decision such that higher position means accepted position, lower position means denied position.

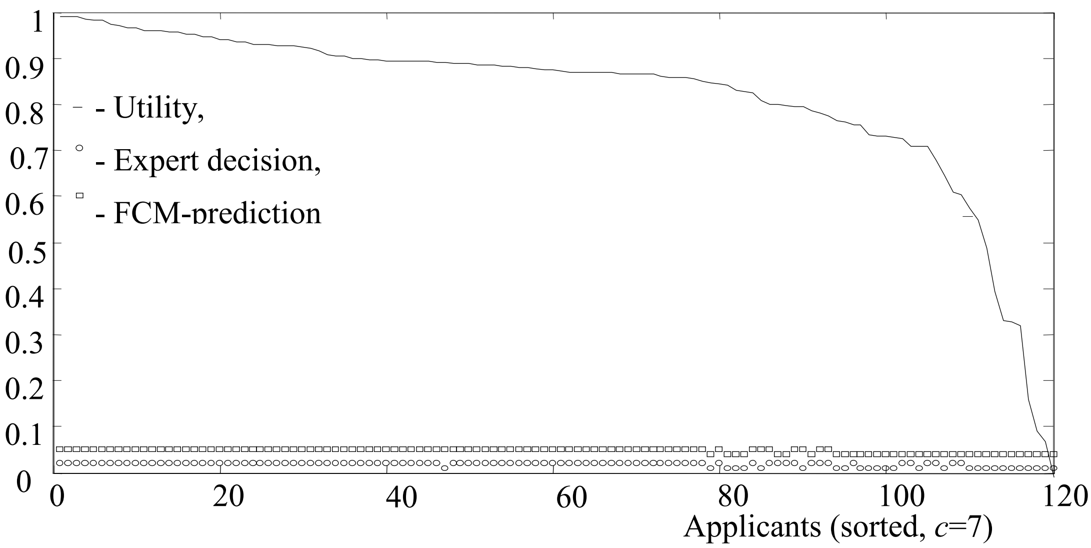

Figure 4. Non-additive utility predicted for the applicants

4.3 The Methodology of the Credit Scoring

4.3.1 Fuzzy Modeling

Fuzzy modeling is a method in order to explain the feature of a system using fuzzy rules, [68]. In this thesis, the fuzzy rule base that represents the input-output relation of an available data set $XY = \{ (x_k, y_k) \mid k = 1, \cdots, n_D\}$ is obtained from the fuzzy clusters which are assigned by the FCM algorithm [44]. FCM has unsupervised learning ability by connecting the similar data items in the same cluster and consequently discovering the input-output relation of the modeled system [37]. FCM algorithm partitions the data set into clusters by assigning a membership value $u_{xy,i,k}(xy_k)$ to each data vector $xy_k = (x_k, y_k)$ to indicate its membership value in cluster i based on the similarity between the fuzzy cluster centers v_i, $i = 1, \cdots, n_c$ and the data vector xy_k.

The Euclidean distance between the normalized data vectors usually form the similarity measure of FCM clustering algorithm. Each of n_C fuzzy clusters contains similar input-output cases with membership values closer to 1, and is processed to extract multi-input fuzzy rules [61], [62]. The i^{th} rule of the conventional type of fuzzy rule base is obtained from the projections of the FCM-membership values $u_{xy,i,k}(xy_k)$ on $x - u$ Cartesian space of each input feature using convex-points [34], [67] and [60].

The statistical properties of the consumer loan data set for the input and output are listed in Table 3. And, also consumer loan approval expert decision of the data set can be seen in Table 3.

Table 3. Some statistical properties of the consumer loan data set

Attributes	**Min.**	**Max**	**Mean**	**St. Dev.**
Income : $\boldsymbol{x_1}$	104	3400	1029	633
Age : $\boldsymbol{x_2}$	20	60	35	9.5
Employment $\boldsymbol{x_3}$	0.5	10	4.5	2.4
Cr. History: $\boldsymbol{x_4}$	0	1	0.9	0.30
Purpose: $\boldsymbol{x_5}$	0	3	1.7	0.94
Amount: $\boldsymbol{x_6}$	1000	25000	6900	4468
Maturity: $\boldsymbol{x_7}$	6	36	27	8.6
Guar.: $\boldsymbol{x_8}$	0	2	1.7	0.5
Collateral $\boldsymbol{x_9}$	0	3	2.2	0.8
Exp. Decision: $\boldsymbol{y}$	0	1	0.8	0.4

Meanwhile, Figure 5 shows the process diagram of Choquet expected utility scoring to obtain FCM-based decision model from the training data set. The representation of a triangular membership function is obtained by three parameters $\{C, D, E\}$, which are well defined in the range $[0,1]$ for a range of $x \in [0,1]$. The left, top and right corners of the triangle are represented by x_L, x_C and x_R, respectively. C is the x value of the top corner of the triangle. D is the measure of the width w of the base of the triangle, where $D = (x_R - x_L)^{1/2}/2 = \sqrt{w}/2$.

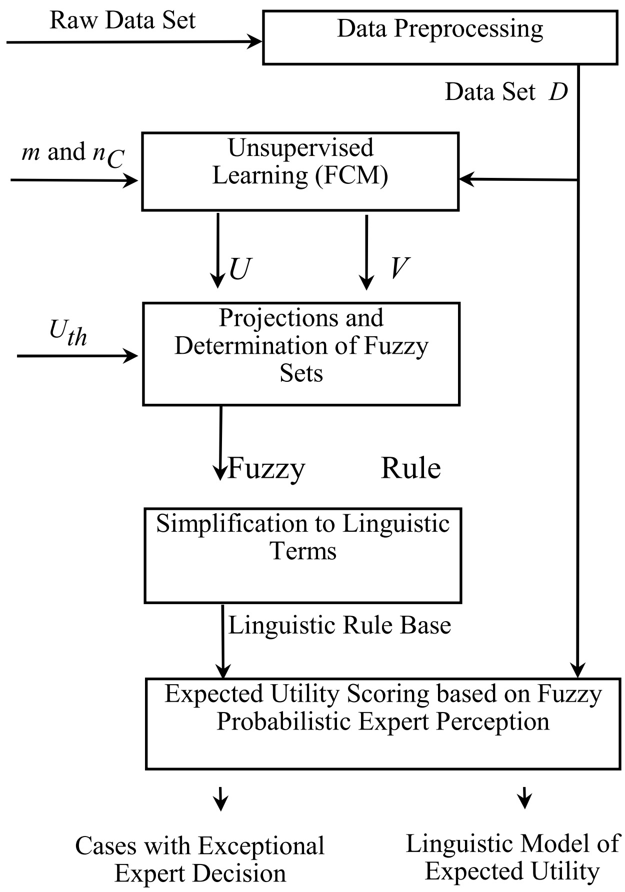

Figure 5. Process diagram of Choquet expected utility scoring

And, E is the tilt of the triangle, where $E = (x_R - x_C)/(x_R - x_L)$. For equation E, the distance from top to right corners is divided by the distance from left to right corners. The corner points of the triangle are defined as $x_C = C$, $x_R = wE + x_C$ and $x_L = x_R - w$, where $w = 4D^2$. The triangular membership function is described by three parameters x_C, x_L and x_R, where $x_C \in [0,1]$ and x_L and x_R may be out of the range.

The presented format of C, D and E provides advantage in tuning the model by evolutionary optimization methods since they are valid all through [0,1] and they are independent measure of top corner position, width of base, and tilt of triangle [60]. After we determined the convex points $(u_{i,k}, x_{k,j})$ such that $k = k_1, k_2, \cdots, k_p$, at the right-side and at the left-side of the cluster center $v_{i,j}$, the right and left corner of the triangle were determined using the least squares estimation (LSE) to find the best fitting line passing through the cluster center and near the convex points.

The line parameters (c_0, c_1) satisfy $c_0 + c_1 x + \varepsilon = u$, where ε denotes the deviation of u due to parameters c_0 and c_1. The line passing through $(x_0, u_0) = (v_{i,j}, 1)$ satisfies $c_0 = u_0 - c_1 x_0$ and it reduces the line equation $\varepsilon = u - u_0 - c_1(u - x_0)$. Thus, LSE with this constraint is reduced to minimize $\sum \varepsilon_k^2 = \sum\big((u_k - u_0) - c_0(x_k - x_0)\big)^2$. Solution is simplified using matrices $U = \big[(u_{k_1} - u_0), \ldots, (u_{k_p} - u_0)\big]$ and $X = \big[(x_{k_1} - x_0), \ldots, (x_{k_p} - x_0)\big]$ to write the estimation error in the form $(U - c_1 X)^T (U - c_1 X) = (\varepsilon_{k_1}, \ldots, \varepsilon_{k_p})^T = 0$. The solution $c_1 = (X^T X)^{-1}(X^T U)$ is reduced to

$$c_1 = \frac{\sum_k (x_k - x_0)(u_k - u_0)}{\sum_k (x_k - x_0)^2} \tag{4.6}$$

where $c_0 = u_0 - c_1 x_0$.

The left and right corners (x_p, u_p) of the triangular membership function have $u_p = 0$. Accordingly, we obtained x_p of these points from $0 = u_p - u_0 - c_1(x_p - x_0)$ and $u_p - u_0 = c_1(x_p - x_0)$ through

$$x_{p,j} = \frac{v_{i,j} + \sum_k (x_k - v_{i,j})^2}{\sum_k (x_k - v_{i,j})(u_k - 1)} \tag{4.7}$$

The credit scoring problem is studied to evaluate the utility of applicants by non-additive Sugeno integral of the FCM generated probabilistic partition matrix as a measure in [69]. The study determined the optimum $n_C = 7$ of FCM at constant $m = 2$ for minimum fail rate and secondary cost functions [69]. The optimum number of rules $n_C = 7$ provides sufficiently high interpretability and comprehensibility for linguistic representation [70].

In this thesis, we searched a decision making model for the same problem using fuzzy-valued Choquet integral instead of the Sugeno integral to calculate the fuzzy utility functions based on expert probabilities using FCM generated fuzzy linguistic models. In the following parts of this section, we describe the developed methodology of linguistic decision-making modeling algorithm for $n_C = 6$, which gives the lowest fail rates among all searched n_C values.

Table 4 lists the cluster-centers obtained by FCM for $n_C = 6$ and m= 1.7. Figure 6 shows the graphical representation of the fuzzy rule base which is obtained from FCM by determination of the left and right convex points of the projected membership values on each attribute. The projection and linear regression procedure to obtain the fuzzy rule base is described in [60], [62] and [69].

Table 4. FCM cluster centers of data set for $m = 1.7, n_C = 6$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
v_1	0.15	0.22	0.31	0.05	0.44	0.13	0.82	0.87	0.42	0.024
v_2	0.06	0.12	0.13	0.97	0.5	0.08	0.76	0.94	0.53	0.084
v_3	0.13	0.25	0.2	0.92	0.47	0.12	0.87	0.36	0.51	0.112
v_4	0.34	0.43	0.63	0.99	0.41	0.24	0.46	0.91	0.81	0.989
v_5	0.2	0.24	0.31	0.99	0.56	0.19	0.82	0.95	0.74	0.991
v_6	0.48	0.62	0.69	0.99	0.93	0.48	0.81	0.9	0.96	0.993

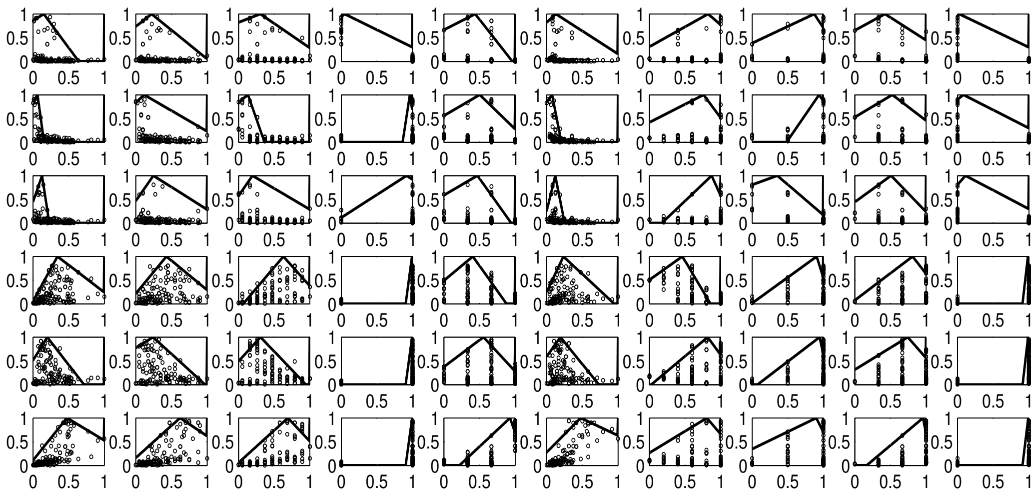

Figure 6. Graphical representation of fuzzy rule base

The FCM generated fuzzy rule base with n_C clusters is expected to contain n_C different fuzzy sets for each variable. Each of these fuzzy sets corresponds to possible fuzzy linguistic terms of the variable.

However, inspecting Figure 6 carefully we realize that many of the terms resembles each other, and the difference between some of the sets are very small to represent them by linguistic terms. To progress the evaluation of the Choquet integrals through less number of linguistic terms, we represented all similar fuzzy sets by a single fuzzy set with rounded average of corner points of all similar fuzzy sets.

Table 5 states the linguistic terms for each variable, the corner points of their triangular membership functions, and the list of rules which are used in the rule base.

The linguistic terms and their fuzzy membership functions that are described in Table 5 were applied to the rule base in Table 6 to illustrate the relations in fuzzy linguistic terms shown in Figure 7. The graphical representation of the fuzzy rule base is displayed in Figure 8. The obtained linguistic fuzzy rule base is a simplified approximation of the FCM generated fuzzy rule base resulting in a loss of information. In spite of loss in prediction accuracy the obtained rule base has less linguistic terms per variable thus, it is a simpler linguistic expression to approximate the multi-input fuzzy locations of the FCM rule-base.

Looking at Table 6, it is possible to think of that only x_9 is sufficient to distinguish denied and accepted cases. But, the fuzzy sets of the linguistic terms weak, good and strong for x_9 are quite similar to each other as shown in Figure 7.

A decision based on only x_9 will have higher probability to fail compared to a decision based on all inputs because the data set contains high uncertainty and fuzzy sets corresponding to linguistic terms of x_9 are closed to each other.

Table 5. Observed linguistic terms of input variables

Attribute	term	x_L	x_C	x_R	in rules
x_1 Net Income (USD)	Low	-0.2	0.1	0.2	2, 3
	medium	-0.6	0.2	0.7	1, 5
	high	0	0.4	1.4	4, 6
x_2 Age (years)	young	-0.6	0.2	1.1	1-3, 5
	Old	-0.2	0.5	1.4	4, 6
x_3 Last Emp. period/year	short	-0.7	0.2	0.9	1-3, 5
	long	-0.1	0.6	1.2	4, 6
x_4 Credit History	negative	-0.7	0.1	1.4	1
	positive	0.9	1.0	1.1	2-6
x_5 Purpose of loan	low loans	-0.5	0.5	1.1	1-5
	high loans	0.2	0.9	1.1	6
x_6 Requested Loan	Low	-0.2	0.1	0.3	2, 3
	medium	-0.2	0.3	1.2	1, 4-6
x_7 Loan maturity, years	short	-0.4	0.5	0.9	1-3, 5, 6
	long	-0.2	0.8	1.3	4
Prop. # of Guarantors	one or two	-1.4	0.4	1.2	3
	two or three	-0.1	0.9	1.2	1, 2, 4-6
	three	0.5	0.9	1.4	6
x_9 Collateral	weak	-0.6	0.5	1.3	1-3
	good	-0.1	0.8	1.4	4, 5
	strong	0.2	1.0	1.1	6

Table 6. Linguistic fuzzy rule base

	Input attributes									Output
i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
1	Med	Young	short	negative	low	large	long	2 or 3	Weak	Denied
2	Low	young	short	positive	low	small	long	2 or 3	Weak	Denied
3	Low	young	short	positive	low	small	long	1 or 2	Weak	Denied
4	High	old	long	positive	low	large	short	2 or 3	Good	Accepted
5	Med	young	short	positive	low	large	long	2 or 3	Good	Accepted
6	High	old	long	positive	high	large	long	3	Strong	Accepted

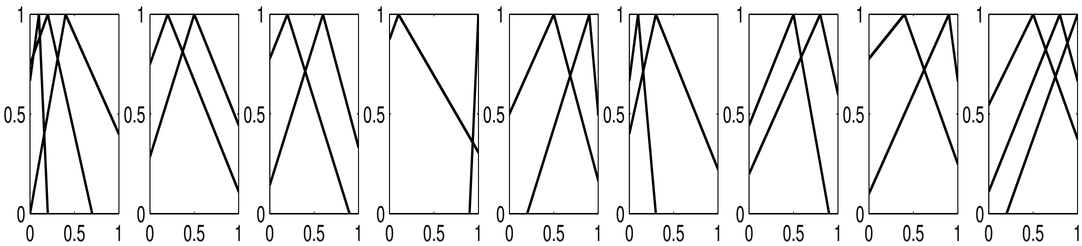

Figure 7. Fuzzy linguistic terms of each input variable

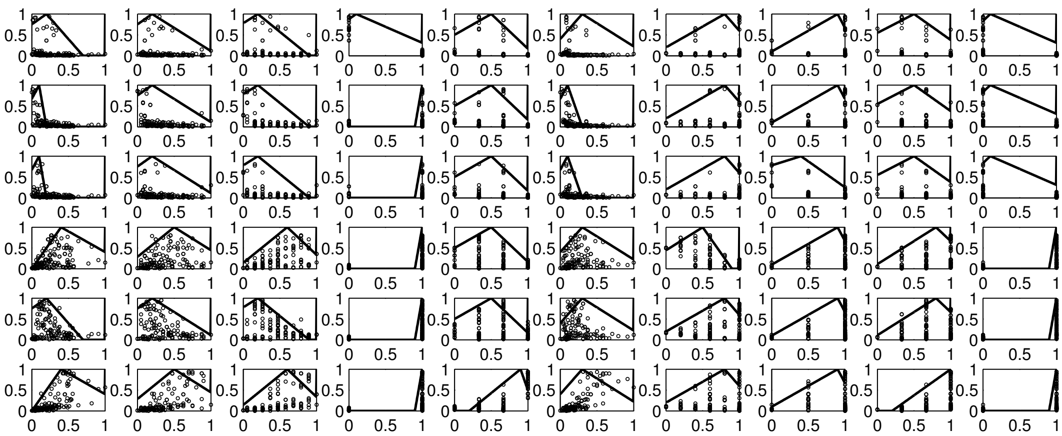

Figure 8. Linguistic fuzzy rule base for utility ranking

4.3.2 Fuzzy Set of a Rule

The antecedent part of each fuzzy rule $i = 1, \ldots, n_C$ describes a multi-dimensional fuzzy set $A_i(x)$ such that it expresses a fuzzy location in the input space. The membership value of x_k in $A_i(x)$ is called the degree of fulfillment of x_k in rule-i and also it is denoted by $\mu_{A,i,k}$. It is calculated by fuzzy logical operations

$$\mu_{A,i,k} = A_i(x_k) = \wedge_{j=1}^{nx} A_{i,j}(x_{k,j}) \tag{4.8}$$

where the logical intersection $\wedge$ denotes a t-norm operation [34], [71] and [41]. In our application, we preferred *min* and *max* functions for t-norm and t-conorm, respectively. The normalized membership values $\mu_{A,i,k}$ of the rule-i through the training data set $\{ x_k | \, k = 1, \ldots, n_D\}$ are called the degree of fulfillment of the rule, and are denoted by $\beta_{i,k}$. The normalized fuzzy sets are denoted by $\bar{f} = \{ f_1, f_2, \ldots, f_{n_C}\}$.

The fuzzy-valued Choquet integral with a fuzzy probabilistic prevision over the rules of the linguistic fuzzy rule base forms a non-additive utility for each case. In the application, we assume that most of the applicants are similar to one of n_C prototypes $\{ f_1, f_2, \ldots, f_{n_C}\}$. We expect that each training observation belongs dominantly to one of n_C rules. The evaluation of fuzzy probabilistic perceptions related to each rule may be calculated using the fuzzy expectation of the expert decisions by a $\alpha -$ cut, with $\alpha \in [0,1]$.

4.3.3 Evaluation of Fuzzy Imprecise Probabilistic Perceptions

The fuzzy measurable interval valued functions $\bar{f} = \{ f_1, f_2, \ldots, f_{n_C}\}$ denote the fuzzy regions that are described by the rules; $i = 1, \ldots, n_C$ and the Choquet integral with expert decision based fuzzy measures $\{P_1, \ldots, P_{n_C}\}$ over $\bar{f}$ shall rank the applicants. A fuzzy-valued fuzzy measure corresponding to f_i is obtained as fuzzy probabilistic imprecise prevision $P_i(x)$. We obtain the fuzzy measures P_i using probability theory by the fuzzy counts of experts accepted decisions in each fuzzy rule, [72].

The overall probability P_i of the rule-i is obtained using the fuzzy expectation of the experts accepted decisions for fulfillment degree $\beta_{i,k}$ of the observation $xy_k = (x_i, y_k)$ in the training dataset $XY = \{yx_k \mid k = 1, \ldots, n_D\}$.

$$p_i = \left(\sum_{k=1}^{n_D} \beta_{i,k} y_k \right) \Bigg/ \left(\sum_{k=1}^{n_D} \beta_{i,k} \right) \tag{4.9}$$

where p_i is the probability of being accepted by experts if the rule-i is fully satisfied i.e. $\beta_{i.k} = 1$. If the probability of "*accepted*" cases were decreasing proportionally to the fulfillment degree of the rule-i it could be the only necessary parameter to estimate the overall probability of the cases. However, we know that the rule is a nonlinear function and the probability for $\alpha < 1$ may highly deviate from its linear estimate αp_i. In this region, an estimate of a lower boundary and an upper boundary of probability may surely help to improve the estimate.

The lower bound of the probability $P_{L,i}(\alpha)$ for $\alpha \in [0,1]$ is obtained by restricting the observations to an α-cut by

$$p_{L,i}(\alpha) = \left(\sum_{\beta_{i,j \le \alpha}} \beta_{i,j} y_k \right) \Bigg/ \left(\sum_{\beta_{i,j \le \alpha}} \beta_{i,j} \right), \quad \alpha \in [0,1] \tag{4.10}$$

where y_k denotes the expert decision and it is 1 for "*accepted*", 0 for "*denied*". The upper bound of the probability of $y=1$ in rule-i is denoted by $P_{U,i}(\alpha)$ and estimated using the points above the α-cut

$$p_{U,i}(\alpha)=\left(\sum_{\beta_{i,j}\geq 1-\alpha}\beta_{i,k}y_k\right)\Big/\left(\sum_{\beta_{i,j}\geq 1-\alpha}\beta_{i,k}\right),\quad \alpha\in[0,1]. \tag{4.11}$$

The plots of the probability points for each observation case (x_k,y_k) for each rule-$i\in\{1,...,n_C\}$ are shown in Figure 9, where x-axis is probability, and y-axis is $\alpha\in[0,1]$.

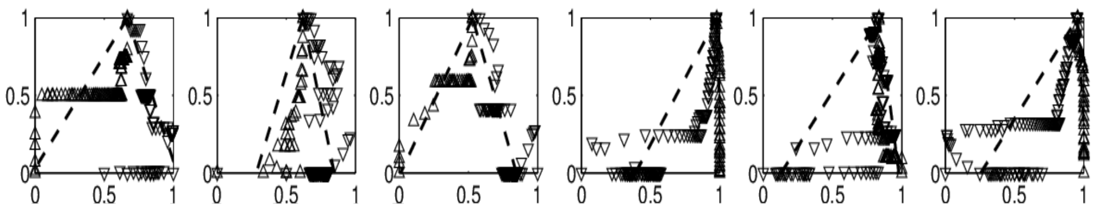

Figure 9. Imprecise probability functions P_i of the linguistic fuzzy rules

The formulas in equations 4.6 and 4.7 have been used in linear regression for the parameters of the linguistic terms. A linear regression with the point p_i set at $\alpha=1$ is carried as explained by equations 4.6 and 4.7.

The corner points given in Table 7 are obtained using line fitting. Note that the corner points beyond [0,1] only express the left and right edges of the triangular membership functions shown in Figures 9 and 10, respectively.

Table 7. Corner points of imprecise probability functions P_i

Rule # i :	1	2	3	4	5	6
Mean Prevision (Top Corner) p_i	0.66	0.62	0.53	0.98	0.84	0.96
Lower Prev. (Left Corner) $p_{i,L}(0)$	-0.10	0.33	0.01	0.36	0.12	0.29
Upper Prevision (Right Corner) $p_{i,U}(0)$	1.05	0.88	0.85	1.01	0.98	1.02

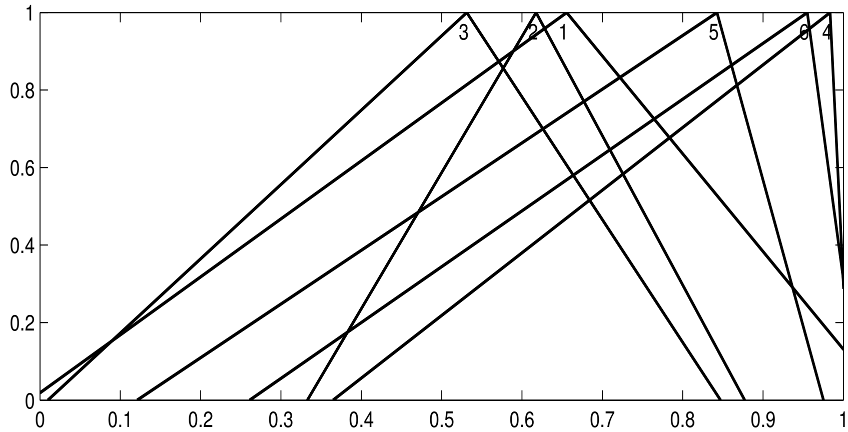

Figure 10. Fuzzy probability functions P_i of the linguistic fuzzy rules

4.3.4 Ranking of the Fuzzy Linguistic Rules

Inspecting Figure 9 gives important clues about the expert's prevision of probabilities. Through the process, the rules were already sorted by the ascending order in their cluster center coordinate as seen in Figure 6. Thus, intuitively we expected that rule-6 has the strongest perceptional probability among the six rules because the fuzzy mean of decision for rule-6 is closest to unity. However, the fuzzy probabilities, equation 4.5, of the rules indicate that the overall probability $p_6 < p_4$. Thus, the capacity measures of the rules are ordered by their center of gravity into $P_3 \prec P_2 \prec P_1 \prec P_5 \prec P_6 \prec P_4$.

4.3.5 Ranking the Utilities of the Applicant-k by Choquet Integral

The fuzzy-valued Choquet utility for the input variables of each observation is obtained by the Choquet integral using max () for t-conorm operation. The fuzzy-valued Choquet utility of case-k were defuzzified to scalar score e_k using the center-of-gravity to rank the cases in their utility scores.

The fuzzy-valued integrals of six cases are shown in Figure 11, where according to their scalar scores they are ranked $E_{109} \prec E_{130} \prec E_{121} \prec E_{129} \prec E_{11} \prec E_{48}$ with the indices indicating the applicant ID number. The fuzzy-valued Choquet utility E_k encloses more details than the score e_k since it also indicates the magnitude of each rule in the utility of case-k.

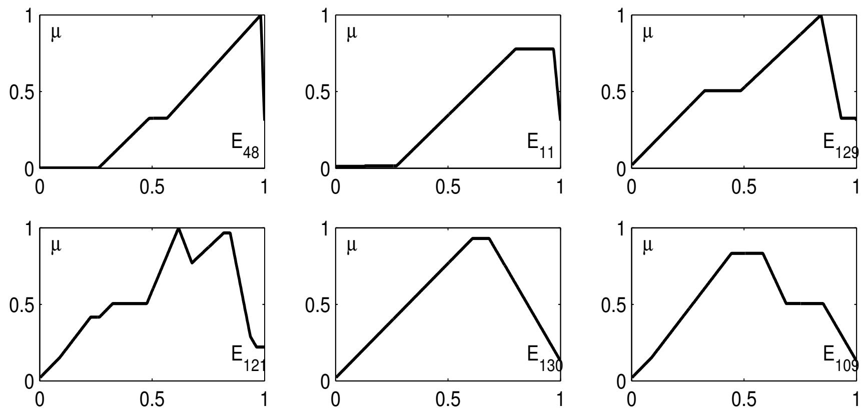

Figure 11. Fuzzy Choquet utility for six cases of applicants

4.4 Decision Analysis

The consumer credit applications are difficult to evaluate due to difficulties in collecting sufficient information about the applicants such as the credit worthiness score of each consumer individually. The loan credit evaluation experts discuss the applications case by case using their experience in their decision.

A decision based on the Decision Theory requires an estimation of the expected utility of each consumer loan credit applicant. The non-additive nonlinearity in expert perception of utility makes the application of the probabilistic expected utility theory unpractical and inaccurate.

The Choquet integral provides a tool to predict the non-expected utility of each applicant using the provisional probabilities over an observation set of cases that contains the expert decisions such as *denied* or *accepted* as output of each case.

In this section, we demonstrate the evaluation of non-expected utility by using the Choquet integral of the measure of capacities obtained by imprecise probabilities over the FCM clusters according to the process diagram shown in Figure 12.

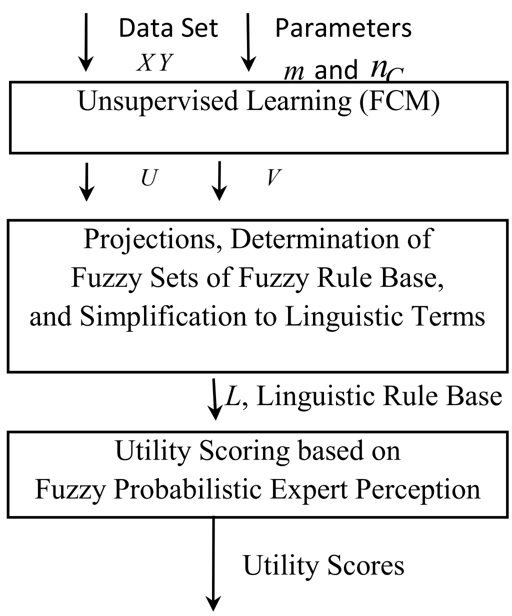

Figure 12. Simplified block diagram of the utility evaluation process

The input and output attributes of the data vectors are specified in Table 8, where ID number is the identification number of the randomly selected cases. The complete consumer loan dataset contains total $n_D = 135$ consumer loan applicants. Table 9 contains the data set D_t that consists of sampled 30 cases and corresponding expert decisions.

Table 8. Input and output attributes of loan applicant case k-set

Symbol	**Type**	**Explanation**
x_{k1}	Scalar	Net income (USD),
x_{k2}	Count	Age (Years)
x_{k3}	Count	Last employment period (Years)
x_{k4}	Nominal	Credit history (0: Negative, 1: Positive)
x_{k5}	Nominal	Purpose of loan (0: General purpose, 1: Flat refurbishment, 2: Car purchase, 3:Flat purchase)
x_{k6}	Scalar	Requested loan amount (USD)
x_{k7}	Scalar	Loan-maturity (Years)
x_{k8}	Scalar	Proposed number of guarantors (Number)
x_{k9}	Nominal	Collateral (0: None, 1: Not applicable, 2: Car, 3: Flat)
y_k	Nominal	Expert decision (0: Denied, 1: Accepted)

Table 9. Data set D_t with 30 loan credit applications

	Income	Age	Emp.	Cre	Pur	Req.	Mat.	Gua.	Col.	Dec.
1	1322	31	8	1	2	6000	24	2	2	1
2	830	30	8	1	1	5000	24	2	3	1
3	381	25	2	1	1	2000	18	2	1	1
4	732	37	4	1	0	2500	12	2	1	1
5	1800	48	7	1	3	8500	30	2	3	1
6	1161	29	4	1	3	7000	30	2	3	1
7	185	23	1	1	0	3000	36	2	1	0
8	710	27	8	1	2	5000	36	2	2	1
9	662	32	4	1	3	6500	36	1	3	1
10	942	26	5	1	0	2000	12	2	1	1
11	1600	45	10	1	2	11000	24	1	2	1
12	586	28	3	1	1	4000	18	2	3	1
13	180	25	2	1	2	2500	36	2	1	0
14	1583	45	10	1	2	11000	24	1	2	1
15	900	26	3	1	2	8500	36	2	2	1
16	1178	28	8	1	1	7500	18	2	3	1
17	250	20	1	1	2	2000	24	2	2	0
18	1543	48	6	1	3	16000	36	2	3	1
19	1296	28	4	1	3	15000	36	1	3	1
20	930	33	5	1	1	7500	36	2	3	1
21	1125	42	5	1	2	6000	24	2	2	1
22	893	32	4	0	1	3000	36	2	1	0
23	2800	40	7	1	1	7000	12	1	3	1
24	1089	28	9	0	2	9500	18	2	2	0
25	987	34	5	1	1	4500	12	2	3	1
26	1351	36	6	1	2	9000	30	1	2	1
27	231	26	2	0	1	3000	36	2	1	0
28	1335	37	7	1	2	4000	12	1	2	1
29	650	30	6	1	1	4500	24	1	3	1
30	1647	50	7	1	0	3000	6	2	1	1

Chapter 5

EXPERIMANTAL RESULTS

5.1 Decision Making by Minimizing Misclassification Rates

The plot of the normalized scalar utility scores $e_{k'}$ against the utility-ordered cases $k' = 1, \ldots, n_D$ is shown in Figure 13, where the expert decision of each case is marked by a ⋄-mark at lower-position for "*denied*" and upper-position for "*accepted*". The upper positioned marks indicate $y = 1$ and expert decision of the case is "accepted" in Figure 13.

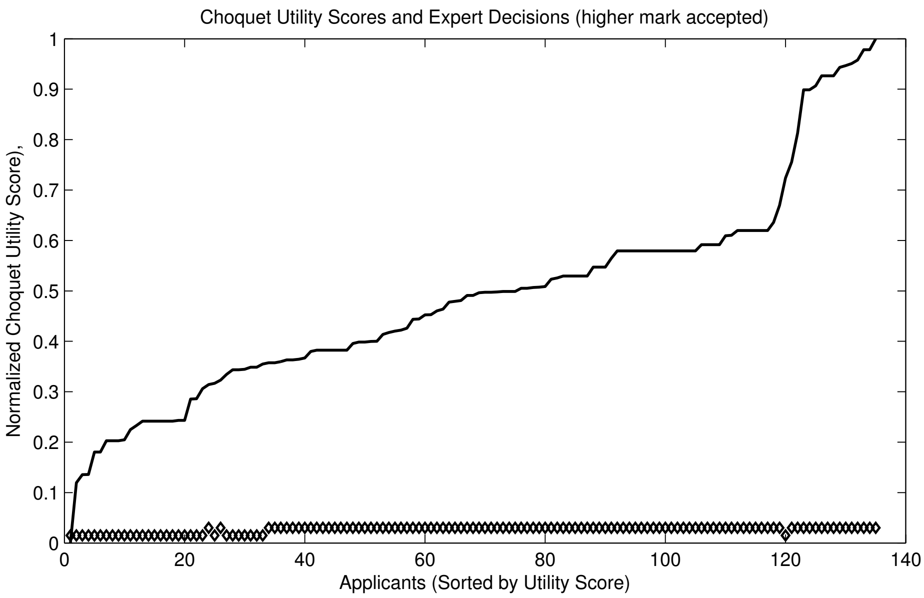

Figure 13. Normalized Choquet utility scores for all applicants sorted in utility score.

The plot indicates the success of the method in ranking the cases in their utility scores with only six exceptional cases. The perfect scoring shall give the first 32 cases with "*denied*" expert-decisions and all remaining with "*accepted*". However, we count 2 fails (*accepted*) in the first 32 cases, and 2 fails (denied) in the remaining 103 cases. Table 10 contains the fail rates of modeling tests carried out for $m = \{ 1.7,\ 1.8,\ 1.9,\ 2.0 \}$ at $n_C = \{ 3, \dots, 12 \}$, and compares them to the fail rates obtained by the Sugeno Integral over the multi-input FCM membership functions as fuzzy measurable fuzzy sets [69]. The Sugeno integral over FCM functions were carried with 120 applicant cases, where total 27 were denied and remaining 93 were accepted by experts. The applied fuzzy utility method in this thesis is superior to the Sugeno-Integral of FCM functions since it gives both less number and also less percent fail. Note that the detected exceptions may be re-examined by the experts to maximize the utility of financial institution by the granted credits.

The secondary fail measure D_D in Table 10 is the sum of the index-distance from sorted-index of denied-failed-case to the accepted region which starts at $k' > 32$. Similarly, D_A is the sum of index-distance from failed-cases in accepted region to the denied region. For $n_C = 6$, $k' \in \{ 24, 29 \}$ are the failed cases in the *denied* region and $k' \in \{ 34, 119 \}$ are the failed cases in the accepted region, giving $D_D = 7 + 9 = 16$, and $D_D = 2 + 87 = 89$. A high D_D means the institution may have a loss of profit due to denied acceptable cases, and a high D_A means that the institution has taken unacceptable risk by accepting unacceptable cases.

The fail rates according to the Sugeno integral are demonstrated in Table 10, [66]. The fail rates indicate that the applied fuzzy-valued Choquet integral in this thesis gives less number of fails although the number of cases is increased from 120 to 135.

Table 10. Number of fails for expected utility prediction

	Linguistic Fuzzy Choquet Integral					Sugeno Integral	
n_C	# Fail	% Fail	D_D	D_A	n_C	#Fail	%Fail
3	36	72	305	641	3	17	31.4
4	8	16	24	115	4	16	40.9
5	30	60	115	506	5	18	66.7
6	**4**	**8**	**13**	**90**	6	14	22.9
7	30	60	171	513	**7**	**12**	**18.2**
8	28	56	106	611	8	14	25.6
9	22	22	123	383	9	12	23.4
10	32	64	166	821	10	10	15.4
11	28	56	224	290	11	10	21.3
12	36	72	196	736	12	17	36.7

Another advantage of the proposed method is the ranking of the rules in their previsional probability of utility score by $P_3 \prec P_2 \prec P_1 \prec P_5 \prec P_6 \prec P_4$. Accordingly, the highest utility section specified by rule-4, and the lowest utility by rule-3 are expressed.

Rule-4: if (net income is_r high) and (age is_r old) and (last employment period is_r long) and (credit history is_r positive) and (purpose of loan is_r low) and (requested loan is_r large) and (loan maturity is_r short) and (proposed number of guarantors is_r 2 or 3) and (collateral is_r good) then (utility is_r highest).

Rule-6: if (net income is_r high) and (age is_r old) and (last employment period is_r long) and (credit history is_r positive) and (purpose of loan is_r high) and (requested loan is_r large) and (loan maturity is_r long) and (proposed number of guarantors is_r 3) and (collateral is_r strong) then (utility is_r higher).

Rule-5: if (net income $is_r\}$ medium) and (age is_r young) and (last employment period is_r short) and (credit history is_r positive) and (purpose of loan is_r low) and (requested loan is_r large) and (loan maturity is_r long) and (proposed number of guarantors is_r 2 or 3) and (collateral is_r good) then (utility is_r high).

Rule-1: if (net income is_r medium) and (age is_r young) and (last employment period is_r short) and (credit history is_r negative) and (purpose of loan is_r low) and (requested loan is_r large) and (loan maturity is_r long) and (proposed number of guarantors is_r 2 or 3) and (collateral is_r weak) then (utility is_r sufficient).

Rule-2: if (net income is_r medium) and (age is_r young) and (last employment period is_r short) and (credit history is_r negative) and (purpose of loan is_r low) and (requested loan is_r large) and (loan maturity is_r long) and (proposed number of guarantors is_r 2 or 3) and (collateral is_r weak) then (utility is_r insufficiently low).

Rule-3: if (net income is_r low) and (age is_r young) and (last employment period is_r short) and (credit history is_r positive) and (purpose of loan is_r low) and (requested loan is_r small) and (loan maturity is_r long) and (proposed number of guarantors is_r 1 or 2) and (collateral is_r good) then (utility is_r lowest).

The described fuzzy Choquet integral may be applied to the new consumer credit applicants to estimate the probable standing of the applicant in the expert's decision range. This feature may be especially useful to satisfy the consumers before they formally apply to the financial institution.

The applied fuzzy-valued Choquet integral to the already evaluated applicants provides the ranking of the applicants in the provisional utility scale according to the expert decisions given for similar cases. The inconsistent cases with an expected utility highly deviated from the decision boundary may be reviewed by the experts to prevent material mistakes in the decision making process.

5.2 Evaluation of Utility Function Based on Choquet Integral

The proposed process for evaluation of utility function by Choquet integral evaluation mainly depends on unsupervised learning ability of FCM by partitioning the similar cases into the same cluster with higher fuzzy membership values. The following algorithm is applied to the data set:

Algorithm

1. Normalize D_t into $[0,1]$ interval to prevent anomalies of FCM algorithm. Normalized data set is denoted by $D_n = \{xy_1, ..., xy_{n_D}\}$ where $xy_k = \{x_{k1}, x_{k2}, ..., x_{k9}, y\}$.

2. Select the sets of m m and n_C, parameters to search the best m and n_C that minimize the fail rate in expert decisions. For each m and n_C pair:

   i. Apply FCM to D_n in order to get partition matrix U_{FCM} and cluster center matrix F_{FCM}.

ii. Obtain the fuzzy set $A_{i,j}$, of attribute-j for the rule R_i by applying the convex- points method to the $U_{FCM,i} - x_{k_j}$, planes; for all $i = 1, \dots, n_C$.

iii. Use the cluster center $V_{FCM,j,j}$ for the top point of triangular membership function $A_{i,j}$.

iv. Use the least squares regression with fixed top corner on convex points for left and right corners of $A_{i,j}$ to generate a fuzzy rule base $R = \{ R_i | \, i = 1, \dots, n_C \}$.

v. Simplify the rule base R to get a convenient fuzzy linguistic rule base $L = \{ L_i | \, i = 1, \dots, n_C \}$ by reducing fuzzy sets $A_{i,j}$ to linguistic sets $A_{i,j}$.

vi. Calculate the membership degree

$$\mu_{L,k,i} = \cap_{j=1}^{n_C} A_{L,i,j}(x_{k,j}) \tag{5.1}$$

of each xy_k in $L_i \in L$; and normalize $\mu_{L,k,i}$ over $k = 1, \dots, n_D$, to $\beta_{k,i}$.

vii. Calculate the fuzzy mean, upper and lower imprecise accepted probabilities and construct the triangular membership functions P_i of each rule.

$$P_{mean,i} = \frac{\sum_{k=1}^{n_D} \beta_{i,k} y_k}{\sum_{k=1}^{n_D} \beta_{i,k}} \tag{5.2}$$

$$P_{lower,i}(\alpha) = \frac{\sum\limits_{\beta_{i,k}<\alpha} \beta_{i,k} y_k}{\sum\limits_{\beta_{i,k}<\alpha}^{n_D} \beta_{i,k}} \tag{5.3}$$

$$P_{upper,i}(\alpha) = \frac{\sum\limits_{\beta_{i,k}\geq\alpha} \beta_{i,k} y_k}{\sum\limits_{\beta_{i,k}\geq\alpha}^{n_D} \beta_{i,k}} \tag{5.4}$$

viii. For each x_k integrate the fuzzy Choquet expected utility E_k by fuzzy integral of imprecise probabilities P_i over all rules using the membership degree of the case in rule-i

ix. Defuzzify E_k to a scalar expected utility e_k by center-of-gravity, centeroid, or similar methods.

x. Sort the cases in e_k and find the total fail of expert decisions to find the best (n_C, m) pair with the minimum fail.

3. Return the sorted expected utilities e_k calculated for the best (n_C, m) pair.

The cluster centers and the fuzzy membership values obtained by FCM with $n_C = 6$ and m=1.7 on dataset D_t are shown in Table 11 and Table 12. The graphical representation of the fuzzy rule base is displayed in Figure 14 which is obtained using the convex points of projected scatter-plots explained in [60].

Table 11. FCM cluster centers for m = 1.7, $n_C = 6$

$v_{FCM,1}$=(0.15, 0.30,0.27, 0.04, 0.34, 0.08, 0.99, 0.99, 0.02, 0.03)

$v_{FCM,2}$=(0.03, 0.11, 0.10, 1.00, 0.49, 0.05, 0.86, 0.99, 0.19, 0.06)

$v_{FCM,3}$=(0.35, 0.30, 0.86, 0.11, 0.67, 0.53, 0.43, 0.97, 0.53, 0.10)

$v_{FCM,4}$=(0.28, 0.45, 0.42, 1.00, 0.11, 0.05, 0.24, 0.98, 0.07, 1.00)

$v_{FCM,5}$=(0.34, 0.42, 0.57, 1.00, 0.58, 0.34, 0.68, 0.96, 0.84, 1.00)

$v_{FCM,6}$=(0.48, 0.60, 0.70, 1.00, 0.67, 0.49, 0.63, 0.06, 0.69, 1.00)

Table 12. FCM cluster centers of each case $k = 1,\ldots,30$ in each rule

k	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	max	i
1	0.01	0.01	0.91	0.03	0.02	0.02	0.91	3
2	0.05	0.06	0.14	0.45	0.2	0.09	0.45	4
3	0.99	0.01	0	0	0	0	0.99	1
4	0.02	0.02	0.08	0.13	0.51	0.25	0.51	5
5	0.01	0.01	0.87	0.05	0.03	0.03	0.87	3
6	0	0.01	0.03	0.05	0.86	0.05	0.86	5
7	0.31	0.12	0.13	0.16	0.15	0.12	0.31	1
8	0.07	0.1	0.33	0.34	0.1	0.05	0.34	4
9	0.02	0.02	0.1	0.19	0.47	0.2	0.47	5
10	0.01	0.01	0.04	0.07	0.15	0.73	0.73	6
11	0	0.01	0.05	0.06	0.8	0.07	0.8	5
12	0.03	0.89	0.04	0.02	0.01	0.01	0.89	2
13	0.93	0.03	0.01	0.01	0.01	0.01	0.93	1
14	0.03	0.9	0.03	0.02	0.01	0.01	0.9	2
15	0.02	0.04	0.59	0.1	0.18	0.07	0.59	3
16	0	0.01	0.03	0.06	0.86	0.03	0.86	5
17	0.01	0.04	0.26	0.51	0.11	0.07	0.51	4
18	0.5	0.15	0.12	0.08	0.09	0.06	0.5	1
19	0	0.01	0.03	0.03	0.05	0.88	0.88	6
20	0.01	0.01	0.04	0.03	0.05	0.88	0.88	6

Table 12. (continued)

k	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	max	i
21	0.01	0.02	0.45	0.31	0.13	0.09	0.45	3
22	0.98	0.01	0	0	0	0	0.98	1
23	0	0	0.01	0.02	0.95	0.01	0.95	5
24	0	0	0.04	0.9	0.04	0.02	0.9	4
25	0.01	0.01	0.06	0.08	0.81	0.04	0.81	5
26	0.98	0.01	0	0	0	0	0.98	1
27	0.01	0.02	0.15	0.06	0.1	0.65	0.65	6
28	0.02	0.02	0.07	0.09	0.11	0.69	0.69	6
29	0.01	0.01	0.35	0.51	0.08	0.04	0.51	4
30	0.01	0.02	0.07	0.75	0.11	0.05	0.75	4

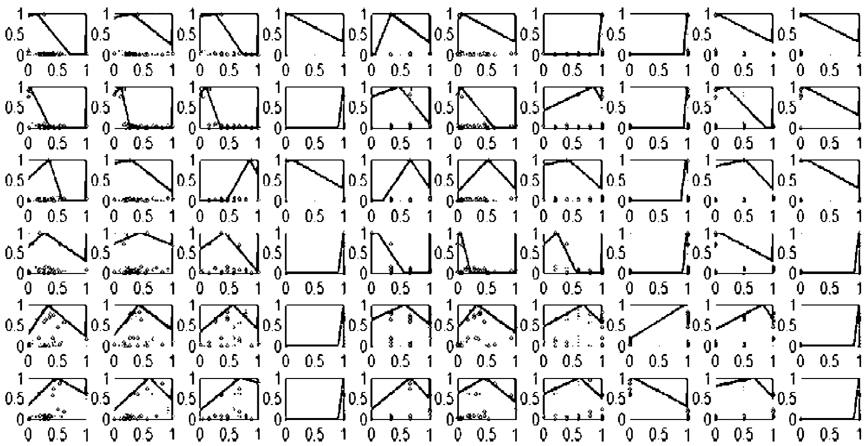

Figure 14. FCM generated fuzzy rule base

The corner points of the similar triangular fuzzy membership functions in the FCM generated rule base are represented by a single fuzzy set to obtain the linguistic terms which are shown in Table 13 and Figure 15.

The graphical representation of the fuzzy linguistic rule base is shown in Figure 16. The linguistic rule base provides the comprehensibility of the rules by simplification and standardization of the fuzzy sets corresponding to each attribute. The approximate membership functions for the imprecise probabilities of $(y=1)$ for each rule is obtained by line-fitting on the evaluated points which are calculated according to equations (5.3) and (5.4).

Table 13. Corner points of the linguistic triangular membership functions

Attributes	**Terms**	**Left**	**Top**	**Right**
x_1	low	-3.20	0.20	0.80
	medium	-0.40	0.20	0.50
	high	-0.30	0.40	1.40
x_2	young	-2.30	0.30	1.20
	old	-0.30	0.50	1.30
x_3	short	-1.10	0.30	0.80
	long	-0.00	0.70	1.70
x_4	negative	-2.40	0.10	1.40
	positive	0.90	1.00	1.00
x_5	high	-2.60	0.20	0.80
	low	-0.20	0.60	1.30
x_6	small	-0.40	0.10	0.90
	large	-0.40	0.40	1.20
x_7	short	-1.00	0.60	1.20
	long	0.90	1.00	1.10
x_8	1-2	-1.00	0.10	1.40
	2-3	0.80	1.00	1.10
x_9	low	-2.00	0.10	1.30
	good	-2.60	0.60	1.30
	high	-0.50	0.80	1.30

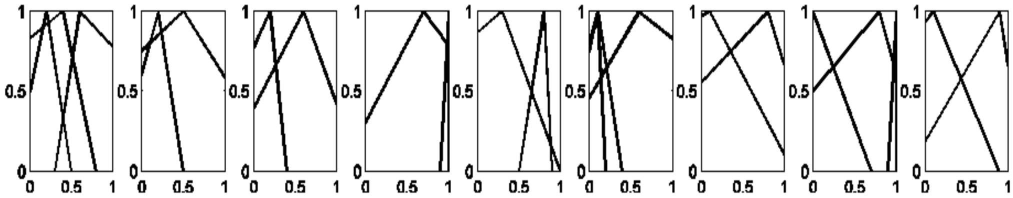

Figure 15. Fuzzy sets of linguistic terms for input attributes

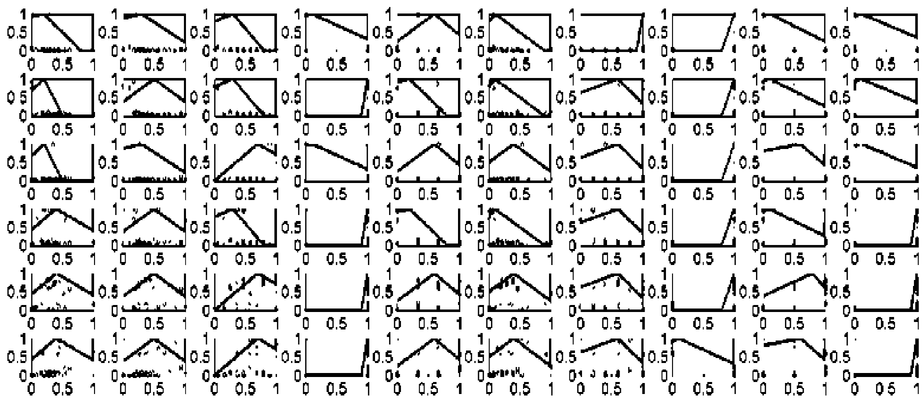

Figure 16. Linguistic fuzzy rule base

In the line-fitting the top point is fixed at the mean value which is obtained by equation (5.2). The degree of fulfillments $\beta_{i,k}$ of the input vectors for each rule, which are used in evaluation of imprecise probabilities, are listed in Table 14. The resulting imprecise probabilities and the fitted fuzzy sets are shown in Figure 17, [72], [73].

The imprecise probabilities for each input vector corresponding to each rule are agglomerated using Choquet integral to a fuzzy utility function $E(x_k)$. Some of the sample fuzzy utilities are displayed in Figure 18.The fuzzy utilities of each input vector is defuzzified by using Center of Gravity method in order to obtain the utility score of the input vector. The scalar utility scores provide the ordering of the loan applicants according to the imprecise probabilities of expert decisions.

Table 14. Degree of fulfillments, defuzzified non-expected utility

k	$\beta_{1,k}$	$\beta_{2,k}$	$\beta_{3,k}$	$\beta_{4,k}$	$\beta_{5,k}$	$\beta_{6,k}$	$e(x_k)$
1	1.00	0.45	0.42	0.03	0.03	0.00	0.6186
2	0.78	0.60	0.45	0.03	0.03	0.00	0.5988
3	0.78	0.74	0.44	0.33	0.29	0.00	0.5286
4	1.00	1.00	0.49	0.37	0.33	0.00	0.5425
5	0.56	0.45	0.00	0.00	0.00	0.00	0.6388
6	0.60	0.56	0.45	0.00	0.00	0.00	0.5792
7	0.46	0.44	0.38	0.15	0.11	0.10	0.494
8	0.60	0.45	0.43	0.03	0.03	0.03	0.5792
9	0.49	0.00	0.00	0.00	0.00	0.00	0.6263
10	0.86	0.83	0.49	0.37	0.33	0.00	0.5365
11	0.86	0.00	0.00	0.00	0.00	0.00	0.6723
12	0.60	0.49	0.45	0.34	0.33	0.00	0.5561
13	0.46	0.44	0.33	0.30	0.30	0.29	0.5077
14	0.86	0.00	0.00	0.00	0.00	0.00	0.6723
15	0.60	0.46	0.45	0.43	0.30	0.30	0.5407
16	0.78	0.60	0.45	0.03	0.03	0.00	0.5988
17	0.30	0.30	0.00	0.00	0.00	0.00	0.4988
18	0.37	0.33	0.00	0.00	0.00	0.00	0.6037
19	0.49	0.00	0.00	0.00	0.00	0.00	0.6263
20	0.60	0.45	0.43	0.38	0.34	0.33	0.5508
21	0.88	0.60	0.45	0.30	0.30	0.00	0.5886

Table 14. (continued)

k	$\beta_{1,k}$	$\beta_{2,k}$	$\beta_{3,k}$	$\beta_{4,k}$	$\beta_{5,k}$	$\beta_{6,k}$	$e(x_k)$
22	1.00	0.65	0.00	0.00	0.00	0.00	0.5064
23	0.59	0.00	0.00	0.00	0.00	0.00	0.642
24	1.00	0.00	0.00	0.00	0.00	0.00	0.6094
25	0.78	0.60	0.45	0.34	0.33	0.00	0.5764
26	0.98	0.00	0.00	0.00	0.00	0.00	0.6772
27	1.00	0.44	0.00	0.00	0.00	0.00	0.4623
28	1.00	0.00	0.00	0.00	0.00	0.00	0.6773
29	0.63	0.00	0.00	0.00	0.00	0.00	0.6481
30	0.37	0.33	0.32	0.00	0.00	0.00	0.6037

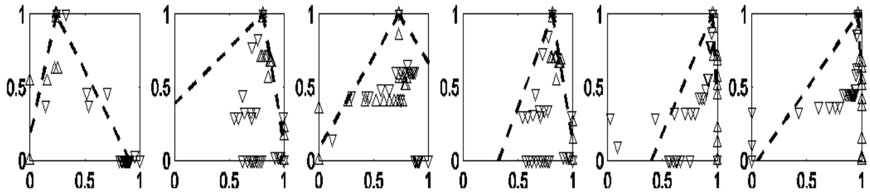

Figure 17. Imprecise probabilities of experts accept decision

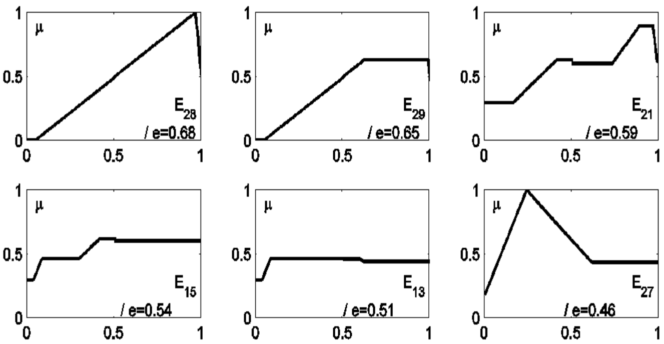

Figure 18. Fuzzy probabilistic Choquet utility of selected cases

The obtained scalar utility values are listed in the last column of Table 14. Figure 19 displays the ordered normalized utility values of the applicants together with the expert decisions.

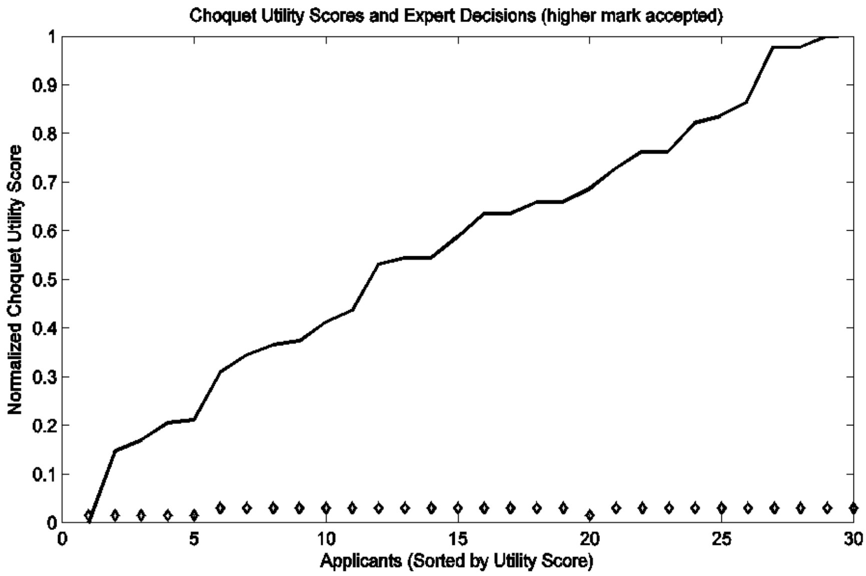

Figure 19. Utility scores defuzzified from fuzzy Choquet utilities

Chapter 6

CONCLUSION

First of all, this thesis is attracted to the problem of credit risk scoring by developing a Fuzzy C-Means clustering based approach. The proposed approach is based on scoring the objects by a non-additive fuzzy utility as described by the process diagram of the decision model. The approach is applied to financial credit data with ten attributes. The computer simulation results show the effectiveness of FCM algorithm for unsupervised learning by fuzzy clustering. The optimal number of cluster-centers c to reach the best performance of the model is obtained by k-Fold cross validation to minimize the fail rates and by two cost functions based on mean and standard deviation of the decision attributes. A non-additive utility measure based on the cluster with maximum fuzzy membership value is used to score the credit risks non-additively. Together with the decision model the Mamdani rule-base of the decision model is recommended to describe the fuzzy model graphically.

In the experimental simulations the fail rate of estimating the expert decisions is dropped down to nearly 18% with the best number of clusters c=7. In accordance with the minimum fail rate, the minimum of the introduced secondary validation costs is obtained by the same best model with c=7. Moreover, on the change of accepted objects the utility scores enable to change the decision attributes of the objects accordingly.

Secondly, ranking the consumer loan credit applicants in the previsional probabilistic utility scores using only the available expert decisions for each case is studied. The Choquet integral over the FCM generates fuzzy linguistic rule base provided a non-additive aggregation based on perception of the experts. The described fuzzy-valued Choquet integral may be applied to the new consumer credit applicants to estimate the probable standing of the applicant in the expert's decision range. This feature may be especially useful to satisfy the consumers before they formally apply to the financial institution.

The described method ranks the applicants in a previsional utility scale according to the expert decisions given for similar cases. The previsional expected utility scale ranking of the applicants make the inconsistent cases with an expected utility highly deviated from the decision boundary detectable, so that these cases may be reviewed by the experts to prevent material mistakes in the decision making process.

Finally, the application of imprecise probabilities as a capacity measure in Choquet integral to determine the utility ranking of the consumer loan applications is discussed. We propose a method to calculate lower and upper imprecise probabilities of a desired output in a fuzzy set, and an algorithm to evaluate the utility scores of the consumer loan applicants. The example application uses a data set of 30 applicants as input, and expert decisions (0 = deny, 1 = accept) as their output. The resulting utility scores are consistent to the output, and these scores provide the ordering of the applicants in the risk of credit. They can be used to classify applicants into accepted and denied classes.

REFERENCES

[1] Binmore, K. (2009). Rational Decisions. *Princeton University Press*, 39-75.

[2] Tapiero, C. (2004). Risk and Financial Management: Mathematical and Computational Methods. *John Wiley and Sons*, 16-50.

[3] Neiger, D. (2008). Diagrammatic Decision-Support Modeling Tools in the Context of Supply Chain Management. Adam, F., Humphreys, P. *Encyclopedia of Decision Making and Decision Support Technologies*. *Information Science Reference, 1*, 245-252.

[4] Ryan, G. (2008). Games of Strategy. *In book, Encyclopedia of Decision Making and Decision Support Technologies edited by Adam, F., Humphreys, P. Information Science Reference*, 1, 402-408.

[5] Von Neumann, J., & Morgenstern, O. (1947). Theory of Games and Economic Behavior. *Princeton NJ, Princeton University Press*, Second edition.

[6] Ramsey, F.P. (1964). Truth and Probability, 1926. In *The Foundations of Mathematics*, Routledge & Kegan Paul, 1931. Also in Kyburg and Smokler, 1964.

[7] Savage, L.J. (1954). The Foundations of Statistics. *John Wiley*, 1954; also (1972), Dover.

[8] Debreu, G. (1959). The Theory of Value. *John Wiley and Sons*, New York.

[9] Howard R.A. (1966). Decision Analysis: Applied Decision Theory. *Proceedings of the Fourth International Conference on Operational Research*, Wiley-Inter Science, 55-71. (1971), Decision Analysis: Applied Decision Theory", Reprinted in Management Decisions for Production Operations, Robert Goodell Brown, *The Dryden Press*, Inc., 513-533.

[10] Howard R.A. (1968). The Foundations of decision Analysis. *IEEE Transactions on Systems Science and Cybernetics SSS*-4(3), 211-219. (1984), readings on The Principles and Applications of Decision Analysis, Edited by Howard, R. A. and Matheson, J. E. Strategic Decisions Group, Menlo Park, California, 97-113.

[11] McElroy, T. (2008). Rational Decision Making, Dual Processes, and Framing: Current Thoughts and Perspectives. In book, *Encyclopedia of Decision Making and Decision Support technologies* edited by Adam, F., Humphreys, P. Information Science Reference, 1, 757-765.

[12] Pomerol, J.-C., & Adam, F. (2008). Understanding the Legacy of Herbert Simon to Decision Support Systems. In book, *Encyclopedia of Decision Making and Decision Support Technologies* edited by Adam, F., Humphreys, P. Information Science Reference, 1, 930-938.

[13] Bugera, V., Konno, H., & Uryasev, S. (2002). Credit cards Scoring with Quadratic Utility Functions. *J. Multi-Crit. Decis. Anal., 11*, 197-211.

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

[15] Desheng, W., & Oslon, D.L. (2006). A Topsis Data Mining Demostration and Application to Credit Scoring. *International Journal of Data Warehousing and Mining.* 2, Issue 3.

[16] Luo, S.-T., Cheng, B.-W., & Hsieh, C.-H. (2009). Prediction model building with clustering-launched classification and support vector machines in credit scoring. *Expert Systems with Applications, 36*, 7562-7566.

[17] Hsieh, N.-C. (2005). Hybrid mining approach in the design of credit scoring models. *Expert systems with Applications 28*, 655-665.

[18] Bernoulli, D. (1954, January). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, 22, No. 1, 23-36.

[19] Billot, A. (1995). An existence theorem for fuzzy utility functions: A new elementary proof. *Fuzzy Sets and Systems*, 74, 271-276.

[20] Cutello, V., & Montero, J. (1999). Fuzzy rationality and Utility Theory Axioms. *Fuzzy information processing society*, NAFIPS, 18th International Conference the North America, 332-336.

[21] Nikolova, N., Ahmed, S., & Tenekedjiev, K. (2008). Generic Procedure for Construction of a Multi-Dimensional Utility Function under Fuzzy rationality. *Int. J. of Computers, Communications and Control, 3*, 422-426.

[22] Chiclana, F., Herrera, F., & Herrera-Viedma, E. (1997). Integrating three representation models in fuzzy multipurpose decision making based on fuzzy preference relations. *Fuzzy sets and Systems*, 33-48.

[23] Keeney, R.L. (1974). Multicative Utility Function. *Operational Research*, 22, No.1, 22-34.

[24] Marichal, J.-L. (2000). An Axiomatic Approach of the Discrete Choquet Integral as a Tool to Aggregate Interacting Criteria. *IEEE Trancations on Fuzzy Systems*, 8, No.6, 800-807.

[25] Yang, R., Wang, Z., Heng, P.-A., & Leung, K.-S. (2005). Fuzzy Numbers and Fuzzification of the Choquet integral. *Fuzzy Sets and Systems, 153*, 95-113.

[26] Wang, Z., Leung, K.-S., & Klir, G.J. (2005). Applying Fuzzy Measures and Nonlinear Integrals in Data Mining. *Fuzzy Sets and Systems, 156*, 371-380.

[27] Modave, F., & Grabisch, M. (1998). Preference Representation by the Choquet Integral: The Commensurability Hypothesis. *Proceeding 7th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* Paris, France.

[28] Grabisch, M. (2000). A Graphical Interpretation of the Choquet Integral. *IEEE Transactions on Fuzzy Systems*. 8, 5, 627-631.

[29] Angilella, S., Greco, S., Lamantia, F., & Matarazzo, B. (2004). Assessing Non-Additive Utility for Multicriteria Decision Aid. *European Journal of Operational Research. 158*, 734-744.

[30] Aliev, R.A., Huseynov, O.H., & Aliev, R.R. (2009). Decision Making with Imprecise Probabilities and its Applications. In Proc. of the *5th International Conference on Soft Computing with Words in System Analysis*, Decision and Control, 282-286.

[31] Grabisch, G. (1996). The Application of Fuzzy integrals in Multicriteria Decision Making. *European Journal of Operational Research, 89*, 445-456.

[32] Shieh, J.-I., Wu, H.-H., & Liu, H.-C. (2009). Applying a complexity-based Choquet integral to evaluate student`s performance. *Expert Systems with Applications, 36*, 5100-5106.

[33] Chen, L.-H., & Chiou, T.-W. (1999). A Fuzzy Credit-rating Approach for Commercial loans: a Taiwan case. *Omega, International Journal of Management Science, 27*, 407-419.

[34] Zadeh, L.A. (1965). Fuzzy Sets. *Information and Control*, 8, 338-353.

[35] Bittmann, R.M., & Gelbard, R.M. (2008). DSS Using Visualization of Multi-Algorithms, Voting, in Adam F., Humphreys P. (Ed's.). *Encyclopedia of Decision Making and Decision Support Technologies*, 297, IGI Global.

[36] Fielding, A.H. (2007). Cluster and Classification for the Biosciences, *Cambridge University press*.

[37] Berenji, H.R., & Khedkar, P.S. (1993). Clustering in Product Space for Fuzzy Inference. 2nd *IEEE International Conference on Fuzzy Systems*, 2, 1402-1407.

[38] Yuan, B., Klir, G.J., & Swan-Stone, J.F. (1995). Evolutionary Fuzzy C-Means Clustering Algorithm. Proceeding of *fourth IEEE International Conference on Fuzzy Systems*, FUZZ-IEEE 95, 4, 2221-2226.

[39] Abonyi, J., & Feil, B. (2007). Cluster Analysis for Data Mining and System Identification, *Birkhauser Verlag* A.G.

[40] Anil, K.J., & Richard, C.D. (1988). Algorithms for Clustering Data. *Prentice-Hall*.

[41] Zimmerman, H.J. (1996). Fuzzy Set Theory and Its Applications. *Kluver Academic Publishers*, 3rd Ed.

[42] Klir, G.J., & Yuan, B. (1995). Fuzzy Sets and Fuzzy Logic Theory and Applications. *Prentice Hall*, 574.

[43] Ruspini, E.H. (1969). A New Approach to Clustering. *Information and Control*, 15, 22-32.

[44] Bezdek, J.C. (1973). Fuzzy Mathematics in Pattern Classification. Ph.D.Thesis, Applied Mathematics Center, *Cornell University*, Ithaca, NY.

[45] Gilboa, I., & Schmeidler, D. (1992). Additive Representations of Non-Additive Measures and the Choquet Integral. *Annals of Operations Research*, 51, 43-65.

[46] Mesiar, R. (2005). Fuzzy Measure and Integrals. Fuzzy Sets and Systems, 156, 365-370.

[47] Yuan, B., & Klir, G.J. (1996). Constructing Fuzzy Measures: A New Method and Its Application to Cluster Analysis. *Proc. NAFIPS`96, Berkley*, CA, 567-571.

[48] Yuan, B., & Klir, G.J. (1994). A Generalization of Sugeno Integrals. In Fuzzy IEEE'94. Proceedings of the *First International Joint Conference of the North American Fuzzy Information Processing Society Biannual Conference*. The Industrial Fuzzy Control and Intelligent Systems Conference, and the NASA Joint Technology, 451-454.

[49] Aliev, R.A., Fazlollahi, B., & Aliev, R.R. (2004). Soft Computing and its Applications in Business and Economics. *Studies in Fuzziness and Soft Computing, Springer*, 157, 26-32.

[50] Zhang, G.-Q. (1992). Fuzzy Number-Valued Fuzzy Measure and Fuzzy Number-Valued Fuzzy Integral on a Fuzzy Set. *Fuzzy sets and systems*, .49, 357-376.

[51] Aliev, R.A., (2008). Modelling and Stability Analysis in Fuzzy Economics. *Applied and Computational Mathematics*, 31-53.

[52] Kandel, A. (1978). Fuzzy Statistics and Forecast Evaluation. *IEEE Trans. Systems Man and Cybernetics*, 8 Issue: 5, 396-401, May.

[53] Sugeno, M., Tsukamoto, Y. & Terano, T. (1974). Subjective Evaluation of Fuzzy Objects. Reprint *IFAC Symposium Stochastic Control*.

[54] Bolanos, M., Campos, L. de, & Gonzalez, A. (1989). Convergence Properties of the Monotone Expectation and Its Application to the Extension of Fuzzy Measures. *Fuzzy Sets and Systems*, 33, 201-212.

[55] Campos, L. de, & Bolanos, M. (1992). Characterization and comparison of Sugeno and Choquet integrals. *Fuzzy Sets and Systems*, 52, 61-67

[56] Mathieu-Nicot, B. (1986). Fuzzy Expected Utility. *Fuzzy Sets and Systems*, 20, 163-173.

[57] Wagner, C.G. (1994). Imprecise Probability and Expert Forecasting. *IEEE Proc. 6th International Conference on Tools with Artificial Intelligence*.

[58] Zadeh, L.A. (2002). Toward a Perception-Based Theory of Probabilistic Reasoning with Imprecise Probabilities. *Journal of Statistical Planning and Inference*, Elsevier, 105, 233-264.

[59] Bezdek, J. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. *Kluver Academic Publishers* Norwell, MA, USA.

[60] Bodur, M., Acan, A., & Akyol, T. (2005). Fuzzy System Modeling with the Genetic and Differential Evolutionary Optimization. *Proceedings of IEEE International Conference on Computational Intelligence for Modeling, Control and Automation*, CIMCA, 100-111.

[61] Türkşen, İ.B. (2008). Fuzzy Functions with LSE. *Journal of Applied Soft Computing*, 1178–1188.

[62] Bodur, M., & Ahmaderaghi, B. (2009). A comparison of Fuzzy Functions with LSE and TS-fuzzy methods in modeling uncertain dataset. *In Proceedings of the 5th International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control* (ICSCCW 2009), 1-4.

[63] Mosteller, F. (1948). A k-Sample Slippage Test for an Extreme Population. *Annals of Mathematical Statistics*, 19, 1, 58–65.

[64] Picard, R. & Cook, D. (1984). Cross-Validation of Regression Models. *Journal of the American Statistical Association* 79 (387), 575–583.

[65] Hathaway, R.J., Bezdek, J.C. & Hu, Y. 2000. Generalized Fuzzy C-Means clustering Strategies Using L_P Norm Distances. *IEEE Transactions on fuzzy systems*. 8 (5), 576-582.

[66] Sugeno, M., & Yasukawa, T. (1993). A fuzzy logic based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems*, 1, 1-24.

[67] Nakanishi, H., Türkşen, İ.B., & Sugeno, M. (1993). A Review and Comparison of Six Reasoning Methods. *Fuzzy Sets and Systems*, 57(3), 257-294.

[68] Aliev, R.A., & Aliev, R.R. (2001). Soft Computing and Its Applications. *World Scientific Publishing Co. Pte. Ltd*. 131-137.

[69] Aliev, R.R., & Kuset Bodur, E. (2010). Fuzzy C-Means Clustering Technique for Credit Scoring Risk Prediction. *Proceedings of the Ninth International Conference on Application of Fuzzy Systems and Soft Computing*, ICAFS-2010, Praque, Czech Republic, 235-244.

[70] Mencar, C., Giovanna Castellano, G., & Fanelli, A.M. (2005). Some Fundamental Interpretability Issues in Fuzzy Modeling. *Proc. EUSFLAT - LFA* 2005, 100-105.

[71] Babuska, R., Verbruggen, H., Hellendroorn, H., & Driankov, D. (1997). Constructing fuzzy models by product space clustering, Fuzzy Model Identification: Selected Approaches. *Springer-Verlag*, 53-90.

[72] Aliev, R.R., & Kuset Bodur, E. (2011). Fuzzy Choquet Integral for Loan Granting based on Imprecise Probability. *Proceedings of the Sixth International Conference on Soft Computing with Words and Perceptions in System Analysis, Decision and Control*, ICSCCW-2011, Antalya, Turkey, 291-299.

[73] Aliev, R.R., & Kuset Bodur, E. (2012). Fuzzy-Valued Choquet Integral Based Utility Ranking in the Credit Scoring Problem. *Applied and Computational Mathematics*. (Accepted).