A MULTIPLE CRITERIA SORTING APPROACH BASED ON DISTANCE FUNCTIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

BİLGE ÇELİK

IN PARTIAL FULFILLMENT OF THE REQUIREMENT
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

MAY 2011

Approval of the thesis:

A MULTIPLE CRITERIA SORTING APPROACH BASED ON DISTANCE FUNCTIONS

submitted by **BİLGE ÇELİK** in partial fulfillment of the requirements for the degree of **Master of Science in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen ________________
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Sinan Kayalıgil ________________
Head of Department, **Industrial Engineering**

Assoc. Prof. Dr. Esra Karasakal ________________
Supervisor, **Industrial Engineering Dept., METU**

Assist. Prof. Dr. Cem İyigun ________________
Co-Supervisor, **Industrial Engineering Dept., METU**

Examining Committee Members:

Prof. Dr. Meral Azizoğlu ________________
Industrial Engineering Dept., METU

Assoc. Prof. Dr. Esra Karasakal ________________
Industrial Engineering Dept., METU

Assist. Prof. Dr. Cem İyigün ________________
Industrial Engineering Dept., METU

Assist. Prof. Dr. Sinan Gürel ________________
Industrial Engineering Dept., METU

Assist. Prof. Dr. Sevgi Özkan ________________
Information Systems Dept., METU

Date: 26.5.2011

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: BİLGE ÇELİK

Signature :

ABSTRACT

A MULTIPLE CRITERIA SORTING APPROACH BASED ON DISTANCE FUNCTIONS

Çelik, Bilge

M.Sc., Department of Industrial Engineering

Supervisor: Assoc. Prof. Dr. Esra Karasakal

Co-supervisor: Assist. Prof. Dr. Cem İyigun

May 2011, 111 pages

Sorting is the problem of assignment of alternatives into predefined ordinal classes according to multiple criteria. A new distance function based solution approach is developed for sorting problems in this study. The distance to the ideal point is used as the criteria disaggregation function to determine the values of alternatives. These values are used to sort them into the predefined classes. The distance function is provided in general distance norm. The criteria disaggregation function is determined according to the sample preference set provided by decision maker. Two mathematical models are used in order to determine the optimal values and assign classes. The method also proposes an approach for handling alternative optimal solutions, which are widely seen in sorting problems. Probabilities of belonging to each class for an alternative are calculated using the alternative optimal solutions and provided as the outputs of the model. Decision maker assigns the alternatives into classes according to these probabilities. The method is applied to five data sets and results are provided for different performance measures. Different distance norms are tried for each data set and their performances are evaluated for each data set. The probabilistic approach is also applied to UTADIS. The performance of the distance based model and modified UTADIS are compared with the previous sorting methods such as UTADIS and classification tree. The developed method has new aspects such as using distances to ideal point for sorting purpose and providing probabilities of belonging to classes. The handling of alternative optimal solutions within the method instead of a post-optimality analysis is another new and critical aspect of the study.

Keywords: Multi-criteria sorting, distance based sorting, distance functions, probabilistic sorting.

ÖZ

UZAKLIKLIK FONKSİYONLARINA BAĞLI ÇOK KRİTERLİ SIRALAMA YÖNTEMİ

Çelik, Bilge

Yüksek Lisans, Endüstri Mühendisliği Bölümü

Tez Yöneticisi: Doç. Dr. Esra Karasakal

Ortak Tez Yöneticisi: Yrd. Doç. Dr. Cem İyigun

Mayıs 2011, 111 sayfa

Sıralama problemi, alternatiflerin birden fazla kriterdeki değerlerine göre, önceden belirlenmiş sıralı sınıflara atanmasını içerir. Bu çalışmada, uzaklık fonksiyonuna dayalı bir sınıflandırma yöntemi geliştirilmiştir. Alternatiflerin, ideal noktaya olan uzaklıkları kriter birleştirme fonksiyonu olarak kullanılarak alternatiflerin değerleri belirlenir. Bu değerler, alternatiflerin sınıflara atanması için kullanılır. Uzaklık fonksiyonu, yöntem içinde genel uzaklık normunda kullanılır. Kriter birleştirme fonksiyonu, karar vericinin hazırladığı örnek bir tercih listesine göre belirlenir. Alternatiflerin optimal değerlerini ve atanacakları sınıfları belirlemek için iki matematiksel model kullanılır. Sınıflandırma yöntemi, sıralama problemlerinde sıklıkla görülen alternatif optimal çözümler için de bir çözüm önerisi getirir. Alternatif optimal çözümlere göre, sınıflara ait olma olasılıkları belirlenir ve yöntemin çıktıları olarak sunulur. Karar verici bu olasılıklara göre alternatifleri sınıflara atar. Çözüm yöntemi beş farklı veriye uygulanmış ve performans ölçütlerinin sonuçları sunulmuştur. Her veri kümesi için farklı uzaklık normları uygulanmış ve performansları karşılaştırılmıştır. Olasılıksal yaklaşım UTADIS yöntemine de uygulanmıştır. Uzaklık fonksiyonuna dayalı yöntem ve değiştirilmiş UTADIS'in sonuçları, klasik UTADIS ve sınıflandırma ağacı gibi varılan yöntemlerle karşılaştırılmıştır. Geliştirilen yöntemin, varılan çözüm yöntemlerinden farkı, ideal noktaya olan uzaklıklara göre sınıflandırma yapması ve alternatifleri sadece bir sınıfa

atamak yerine, alternatiflerin farklı sınıflara ait olma olasılıklarını hesaplamasıdır. Yöntemin bir başka yeni ve önemli özelliği, alternatif optimal çözümleri optimal sonrası ele almak yerine, yöntem içerisinde kullanmasıdır.

Anahtar Kelimeler: Çok kriterli sınıflandırma, uzaklığa dayalı sınıflandırma, uzaklık fonksiyonları, olasılıksal sınıflandırma.

To My Mother and Father...

ACKNOWLEDGEMENTS

First of all, I would like to thank profoundly my mother and father for their endless love and faithful support.

Then, I would like to express my special thanks to my dear supervisor, Assoc. Prof. Dr. Esra Karasakal for her never-ending support, continuous understanding, invaluable patience, and guidance throughout this study. Her guidance made me clarify and realize my goals which I will remember forever.

I also would like to show my endless gratitude to my co-supervisor Assist. Prof. Dr. Cem İyigün for his encouragements, guidance, and valuable comments during this study.

I offer my regards and blessings to all my friends Volkan Gümüşkaya, Aykut Bulut, Ayşegül Demirtaş, Kerem Demirtaş and my other dear colleagues who supported me in any respect during the completion of this thesis. Most of all, my belowed room mate Erdem Çolak for his endless support and patience during this study. His positive attitude encouraged me to stand the most difficult moments.

Finally, I would like to express my special thanks and loves to my beloved fiancé Boran Ekin Aydın and my dear sister Neşen Çelik for their endless support and love during this study.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

The aim of multiple criteria decision aid methodology is to assist a decision maker for analysis of a set of alternatives. The structure of the analysis may be in 3 different forms (Roy, 1996):

1. *Choice Problems:* Identification of best alternative or a limited set of the best alternatives.
2. *Ranking Problems;* Ranking of alternatives from best to worst one.
3. *Classification/Sorting Problems:* Assignment of alternatives into predefined homogenous classes which can be ordinal or nominal.

While choice and ranking problems include judgments based on relative comparison of alternatives which changes according to the set of alternatives chosen, classification/sorting decisions require absolute judgments to assign alternatives into groups which are defined independent of the set of alternatives (Zopounidis and Doumpos,2002).The classification and sorting problems differ in the type of the classes that the alternatives are grouped. In the classification problems, classes are nominal whereas sorting problems include ordinal classes which are ranked from the most preferred class to the least one.

Multiple criteria sorting problem is the assignment of alternatives into predefined ordered classes according to their values on several attributes. The sorting problem is a very common problem which is encountered in many different areas of application such as biostatistics, resource allocation, energy policy evaluation, financial management and so on. Being involved as a problem in a wide range of areas, it has been studied by researchers from several disciplines in the last forty years.

The aim of this study is to develop a method that assists decision maker to sort alternatives with highest accuracy. Decision maker provides a set of assigned alternatives and according to this set, the new alternatives are assigned to the classes by the developed method. The method involves mathematical models that minimize the classification error of known alternatives. Although similar models are used in the previous studies with the same objective, the selection of secondary objectives to identify the alternative optimal solutions has been a problem which is focused in some studies and ignored in many others. Despite the main objective chosen as the classification error in the previously defined set, the main focus in the problem is the accuracy of the classification of unknown alternatives. So, the alternative optimal solutions may result in different accuracy levels for those unknown alternatives. Instead of choosing an alternative optimal solution randomly or with a secondary objective, this study proposes an approach that identifies and utilizes the alternative optimal solutions to define assignment probabilities to each class. There are studies that provide possible classes that an alternative can belong according to the alternative solutions identified by secondary objectives, yet to our knowledge there is no study providing probabilities to the decision maker about the assignment of alternatives to the classes.

The structure of the model used is similar to the linear discriminant functions in which the values on each criterion are added after multiplied by the weights of the criterion to find the value of an alternative. In this study, instead of adding the actual values of criteria for an alternative, we use the distance of each alternative to the ideal point. This approach enables evaluation of different distance norms to find the best suited one to the structure of the data and also eliminates the necessity to modify the data according to its type as higher-the-better or lower-the-better since we consider the distances.

The rest of the thesis is organized as follows. In Chapter 2, the related literature on sorting problems is provided. In Chapter 3, we provide a theoretical background on sorting problems in detail and the terminology used. The distance norms used in our models are also defined in this section. In Chapter 4, the solution approach is defined by presenting the mathematical models and probability calculation process as well as the interpretation of the model's outputs. In Chapter 5, computational experiments of the proposed method are presented. The data sets used are explored and the results on these data sets are evaluated in this chapter. The comparison of these results with the previous studies is also presented in this chapter. Finally, we conclude in Chapter 6 with some future research directions identified.

CHAPTER 2

LITERATURE REVIEW ON SORTING PROBLEMS

For the classification/sorting problems, the developed approaches can be grouped as parametric and non-parametric methods. The parametric methods constitute the first studies in this area which are statistical approaches such as the Linear Discriminant Method (LDF) developed by Fisher (1936) and its extension as Smith's (1947) Quadratic Discriminant Method (QDF), and the econometric approaches such as logit (Bliss, 1934) and probit analysis (Berkson,1944). Yet, these statistical approaches have several drawbacks such as their parametric structure and statistical assumptions. The non-parametric approaches other than Multiple Criteria Decision Analysis (MCDA) methods are neural networks, machine learning, fuzzy set theory and rough sets (Zopounidis and Doumpos, 2004). One of the machine learning approaches is classification tree method. The C4.5 algorithm (Quinlan, 1993), an algorithm developed for classification trees, has several advantages such as handling of qualitative attributes and missing information over ID3 algorithm (Quinlan, 1983) which is the previously developed algorithm (Zopounidis and Doumpos, 2004).

The MCDA methods developed for classification/sorting problems can be grouped into two main categories as the techniques based on the direct interrogation of the decision maker and Preference Disaggregation Analysis (PDA) methods. The first category requires the decision maker to specify the preferential information directly to construct the model. The second group minimizes the effort of the decision maker by providing a proper basis to identify the preferences of the decision maker. So, in PDA methods, rather than giving the information on how the decisions are made, the decision maker actually makes the decisions (Zopounidis and Doumpos, 2004). In classification/sorting problems, these decisions include the classification of a limited set of alternatives which is known as preference set. Then, PDA methods try to construct a criteria aggregation model that can represent the decision maker's preferences best compared to the information gained from the preference set. Some of the most commonly used PDA methods are UTA methods, outranking relation methods and discriminant analysis. In this study, we will provide a detailed review of the PDA methods since the method developed in this study also falls into this category.

Outranking relation methods are employed for both classification and sorting problems. For those problems, the outranking degree of an alternative is determined by comparing the alternative to the reference values that distinguishes the classes. The comparison should be done for each criterion to determine whether an alternative outranks a reference value. An alternative outranks a reference value if the degree of outranking is greater than a pre-determined threshold value. The most popular method that is based on the outranking relations is ELECTRE TRI methods (Yu, 1992; Roy and Bouyssou, 1993).

A second PDA approach to the sorting problems is the utility function based approach among which the most widely used ones are UTA methods, most significantly its variant UTADIS (Devaud et al, 1980; Jacquet-Lagrèze and Siskos, 1982) and MHDIS method (Zopounidis and Doumpos, 1999). Both methods use an estimate of the decision maker's utility function to correctly sort the alternatives into preference ordered classes. While UTADIS uses a single utility function which classifies all the alternatives, MHDIS uses more than one utility function to sort the alternatives in a step-by-step manner. In classical UTADIS method, an additive utility function of decision maker is estimated by adding the marginal utility functions of each criterion. The marginal utility functions transform the value of the alternative on a criterion to its correspondence in utility scale between 0 and 1. The global utility of an alternative is used to assign the alternative to a class according to the thresholds of each class determined in the model based on the preference set as shown in Figure 2.1.There are many extensions of UTADIS with additive utility function studied by Doumpos and Zopounidis (1998) and extensions with multiplicative utility function (Keeney and Raiffa, 1993).

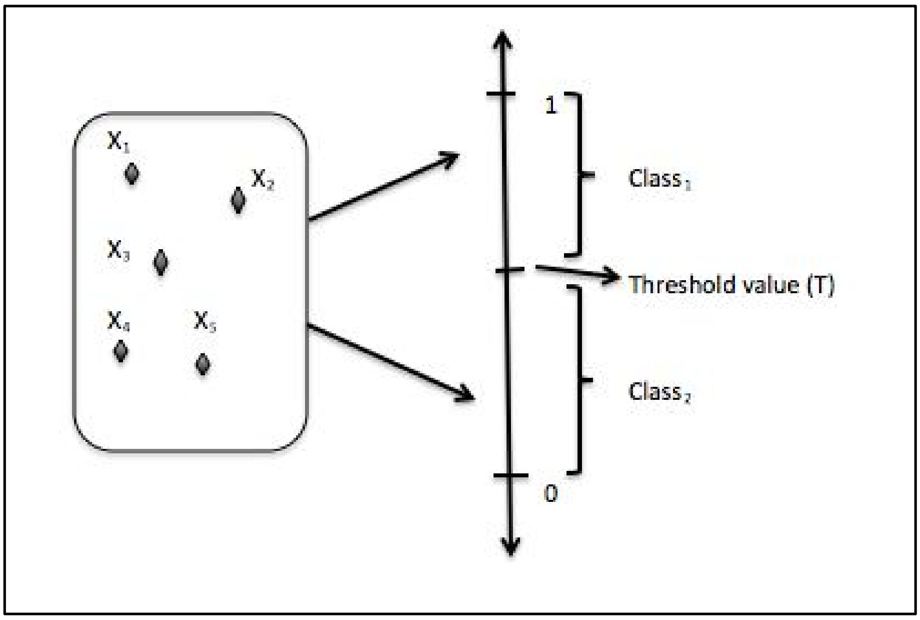


Figure 2-1 Classification of alternatives in UTADIS method in 2-class case[1]

Another PDA approach to the classification/sorting problems is discriminant analysis. It is studied in statistical methods, search methods and mathematical programming (MP) methods. We already mentioned the statistical approaches in the parametric methods. Genetic algorithm approach of Koehler (1989) is an example of search methods. The mathematical programming methods for linear discriminant analysis try to determine a hyperplane $w'x = c$, where w is the weight vector of criteria, c is scalar denoting the threshold value between

[1]Adapted from *Country Risk Evaluation*, K.Kosmidou, M. Doumpos and C. Zopounidis, Springer Optimization and Its Applications, Vol. 15, 2008.

classes and x as the value vector of alternatives, that partitions the p-dimensional Euclidean space into a closed half space $w'x \leq c$ and an open half space $w'x > \mathrm{c}$ (Erengüç and Koehler, 1990). The number of hyperplanes increases as the number of groups increase.

The general form of the MP methods can be shown as (Erengüç and Koehler, 1990):

$$
\begin{aligned}
&minimize\ f(w,c) \\
&subject\ to\ xw \leq c \quad x \in C_1 \\
&\qquad yw > c \quad y \in C_2 \\
&\qquad w \geq 0, c\ u.r.s
\end{aligned}
$$

where w is the weight vector, c is the threshold and x and y denote the alternatives belonging to class 1 and 2 respectively. This general form and the approaches that will be mentioned are constructed for 2-group case where there is no preferential difference between the groups (classification). The 2-group restriction is relaxed by either introducing new thresholds (Freed and Glover, 1981) or by utilizing more than one discriminant function (Bennet and Mangasarian, 1994). Although discriminant analysis is used mostly for classification problems, it can be applied to sorting problems by adding boundary sequencing constraints (Freed and Glover, 1979).

The MP approaches differ according to the objective function. The weights of the criteria are chosen to minimize the classification error. The objective includes a function of the exterior or interior errors. Exterior error denotes the deviation of an incorrectly classified alternative from the cut off value of the group that it is assigned to, while interior error is deviation of a correctly classified alternative from the cut off value. The most common MP approaches are MMD models (minimize maximum exterior deviation), MSD models (minimize sum of exterior deviations), MSID models (minimize sum of exterior deviations and maximize sum of interior deviations), Hybrid models, MIP and NLP models. The MMD and MSD models have linear objectives which are in L-1 norm and L-∞ norms respectively. The NLP model has a similar constraint set whereas its objective is the general L-p norm of the exterior errors. The model is nonlinear for other than p=1 and p=∞ norms. A study of Stam and Joachimsthaler (1989) examines different L-p norm objectives where $1 \leq p \leq \infty$ and concludes that best performing values of p are $1 \leq p \leq 3$ and $p = \infty$. Other than evaluation of different L-p norms for the objective function, there is no research on the utilization of different L-p norms to find the value of an alternative.

A similar PDA approach to discriminant analysis which is developed for sorting problems is a distance based mathematical model of Hipel et al (2005). The model aggregates the values of alternatives on each criteria by calculating the weighted squared Euclidean distance of alternatives the centroid and sorts them according to this distance. This model is important since it is the only study that utilizes distances for criteria aggregation. Yet, it only considers a distance norm similar to the squared weighted Euclidean distance but it does not take square of the weights which are decision variables. It is not a regular distance norm but it results in a linear constraint set. The authors propose the handling of alternative optimal solutions as a future research direction. We mentioned this model since it is the most similar study in literature to the proposed method in this research. Yet, we consider distances to the ideal point and we propose an approach in general L-p norm with a empirical evaluation of different p values applied on several data sets. A method that uses the L-p norm distances of alternatives to the ideal point is VIKOR (Opricovic, 1998) which is an approach developed for ranking problems. It uses an aggregation function that adds the distances to the ideal point by multiplying them with the associated weights of the criteria which is similar to our aggregation function. Still, our approach is unique for the sorting problems not only because of the aggregation function but also its handling approach of the alternative solutions to define probabilities of class assignment.

CHAPTER 3

THEORETICAL BACKGROUND

3.1 Sorting Problems

The PDA methods offer a basis to the decision maker that makes identification of her preferences easier. In sorting problems, this basis is usually a *preference set* (or training set) which is the set of alternatives that the decision maker has already classified to the predefined groups. The preference set is used to extract information on the relative importance of each criterion and the cut off values between classes. The model then classifies the alternatives in the testing set in accordance with the classification in preference set.

		Criteria				Class
Alternatives		A_1	A_2	....	A_q	
	X_1	A_{11}	A_{12}	....	A_{1q}	C_i
	X_2	A_{21}	A_{22}	....	A_{2q}	C_j
	⋮	⋮	⋮	⋮	⋮	⋮
	X_m	A_{m1}	A_{m2}	....	A_{mq}	C_k

Figure 3-1 Preference set in sorting problems

In sorting problems, the aim is to assign a finite set of alternatives X = $\{X_1, X_2, \dots, X_m\}$ to previously defined k groups $\{C_1, C_2, \dots, C_k\}$.Alternative X_i is a vector of values on each criterion, $X_i = (A_{i1}, A_{i2}, \dots, A_{iq})$ where A_{ij} is the score of alternative i on criterion j. Each criterion in sorting problems is ordinal and they can be either “*higher the better*” or “*lower the better*”. If a criterion is higher the better, then the higher scores on that criterion is preferred to the lower scores and it is vice versa for lower the better criteria. The alternatives’ preference order is determined according to these performances on each criterion. An attribute can be categorical ({Small, Medium, Large}) or linear ({1,2,3..}). Linear attributes can be continuous or discrete. The trade-off between the criteria is described by the *weights* of criteria. In criteria aggregation methods, these weights are used as coefficients for scores on each criterion to determine the value of an alternative. This process transforms $R^q \rightarrow R^1$ so the alternatives are ordered in one dimensional space and a threshold value can be calculated for each class. A *threshold* is the value that separates two classes on R^1. The thresholds and the weights are determined by the sorting model in order to classify the alternatives in preference set with minimum error. Then, the developed method with established parameters can be used to sort the alternatives in testing set. A typical model construction is shown in Figure 3.2 (Doumpos and Zopounidis, 2004).

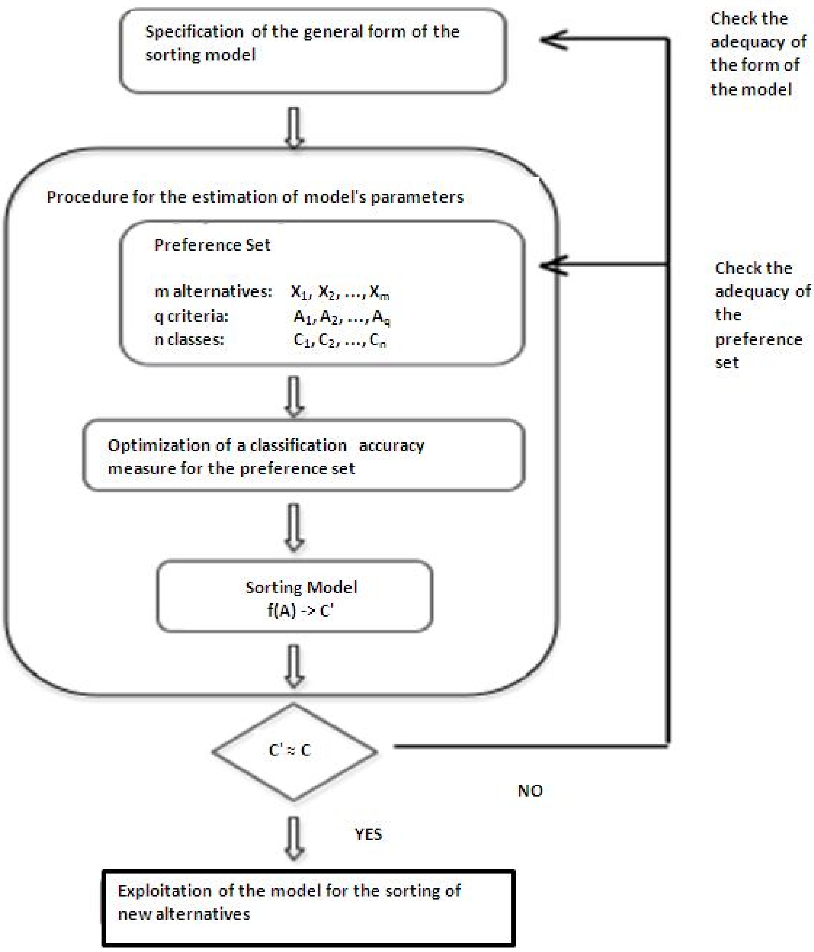


Figure 3-2 Model construction for the sorting problems

3.2 Distance Norms

In this study, we employ different L-p distance norms for the aggregation of criteria to determine the value of an alternative. The general weighted distance of p-norm between two points (X_1, X_2, … , X_n) and (Y_1, Y_2,…,Y_n) is as shown below:

$$\left(\sum_{i=1}^{n}\left(w_i\left|x_i - y_i\right|\right)^{p}\right)^{1/p} \tag{1}$$

When $p = 1$, the distance norm becomes *Manhattan distance* which is the rectilinear distance between two points. When $p = 2$, the distance norm becomes a more familiar and more widely used norm which is *Euclidean distance*. The square of Euclidean distance is known as *Squared Euclidean distance.* As p order gets higher, the larger valued dimensions get more dominating and when $p = \infty$, the distance is equal to max $(X_1, X_2 \dots X_n)$, which is known as *Tchebycheff distance.* Each distance norm in the objective function constitutes a different form of contours (Figure 3.3).

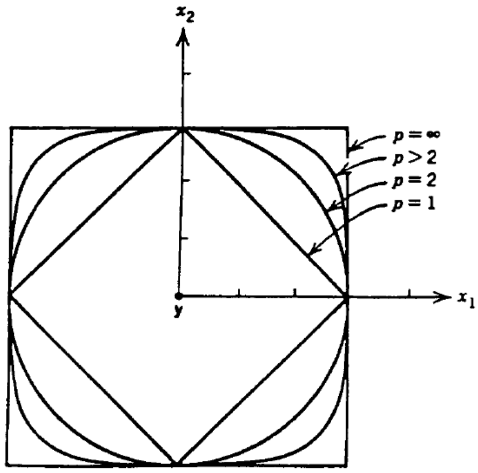


Figure 3-3 Contours of L-p norm distances from y[2]

The distance norms are utilized in the distance calculation of alternatives to the ideal point. *Ideal point* is the point compromising the best values in each criterion which can be maximum or minimum depending on the type of the criteria. It is calculated as shown below:

$$\mathrm{I}_\mathrm{i}^* = max\left(\mathrm{A}_\mathrm{ij} \middle| \mathrm{j} = 1,2, \dots, \mathrm{m}\right) \quad \forall \mathrm{i} \in H \tag{2}$$

$$I_i^* = min\left(A_{ij} \middle| j = 1,2, \dots, m\right) \quad \forall i \in L \tag{3}$$

$$I^* = (I_1^*, I_2^*, \dots, I_q^*) \tag{4}$$

[2] Adapted from the article *Solving the Classification Problem in Discriminant Analysis Via Linear and Nonlinear Programming Methods,* A.Stam and E.A.Joachimsthaler, Decision Sciences, Vol.20, 285-293,1989.

where H is the set of criteria which are higher the better, A_{ij} is the score of alternative i on criterion j, L is the set of criteria which are lower the better and there are m alternatives. I_i^* is the best value in criterion i and I^* is the ideal point in R^q which combines the best values in each criteria.

3.3 UTADIS

UTADIS method is an extension of UTA methods for sorting of alternatives. The method estimates the utility function of the decision maker based on the preference set which is used as a criteria aggregation function. The global utility function of the decision maker is estimated by adding the marginal utility functions for each criterion. The global utility function is:

$$U(a) = \sum_{j=1}^{q} w_j \, u_j(a_j) \tag{5}$$

where a = $(a_1, a_2, \ldots, a_q)$ is the vector of criteria, w_j is the weight factor of criteria j and $w_1 + w_2 + \cdots + w_q = 1$. $u_j(a_j)$ is the marginal utility function that shows the value of score a_j on criteria j for the decision maker. The criteria aggregation function of UTADIS is different than the discriminant function by this property that takes marginal utility value of a score rather than the score itself. The method tries to estimate these marginal utility functions with minimum classification error of the preference set.

The marginal utility functions are monotone and increasing from 0 to 1 in the range of the criteria.

$$u_j(a_j^*) = 1 \tag{6}$$

$$u_j(a_j') = 0 \tag{7}$$

where a_j^* and a_j' are best and worst values for criteria j respectively. In order to avoid nonlinearity due to writing global utility function as a product of two unknowns, weights of the criteria and marginal utility functions, the global utility function is transformed as;

$$U(a) = \sum_{j=1}^{q} u_j{}'(a_j) \tag{8}$$

where $u_j'(a_j) = w_j\, u_j(a_j)$, $u_j'(a_j^*) = w_j$ and $u_j'(a_j') = 0$. In the transformed structure, the global utility function is still in the range of (0,1) but the marginal utility functions vary between $(0, w_j)$ which is the weight of the associated criteria. To estimate the marginal utility function, the interval $[a_j', a_j^*]$ on each criterion is divided into $p_j - 1$ subintervals with p_j break points $a_j^1\, (= a_j'), a_j^2, \dots, a_j^{p_j} (= a_j^*)$. The marginal utility function is estimated by estimating the utility values at breakpoints as shown below:

$$w_{js} = u_j'(a_j^s) - u_j'(a_j^{s-1}) \tag{9}$$

$$u_j'(a_j^h) = \sum_{s=1}^{h-1} w_{js} \tag{10}$$

where w_{js} is the utility value corresponding to the interval s on criterion j. The marginal utility value of any score a_{jk} of alternative k on criterion j is calculated by linear interpolation using the w_{js} values.

$$u_j'(a_{jk}) = \sum_{s=1}^{r_{jk}-1} w_{js} + w_{j,r_{ji}} \frac{a_{jk} - a_j^{r_{jk}}}{a_j^{r_{jk}-1} - a_j^{r_{jk}}} \tag{11}$$

where r_{jk} is the subinterval that alternative k belongs on criterion j. The global utility function of alternative i is then calculated by summing up these marginal utility values.

The objective of UTADIS is to minimize classification error of alternatives in preference set. The misclassified alternatives are the ones that are assigned to a group different than i although they belong to class i (assigned by decision maker in preference set). The classification error for the misclassified alternatives is calculated as:

$$\varepsilon_k^- = max\{\, 0, U(a_k) - u_{i-1}\} \tag{12}$$

$$\varepsilon_k^+ = max\{\, 0, u_i - U(a_k)\} \tag{13}$$

for every $x_k \in C_i$ where u_i denotes the threshold value of class i and ε_k^- and ε_k^+ are the errors of assignment to a higher (better) class and a lower (worse) class respectively. The linear model that finds the optimal w_{jt} values and thresholds of classes with the objective of minimizing the total classification error is shown below.

Indexes

$K =$ number of alternatives in preference set

$q =$ number of criteria

$n =$ number of classes

$p_j =$number of breakpoints on each criterion

$k \in \{1,2, \dots, K\}$ for alternatives in preference set

$j \in \{1,2, \dots, q\}$ for criteria

$i \in \{1,2, \dots, n\}$ for classes

$s \in \{1,2, \dots, p_j - 1\}$ for intervals on each criterion

Parameters

$r_{jk} =$ subinterval that alternative k belongs on criterion j

$C_i =$ set of alternatives in preference set which belongs in class i

$x_k =$ alternative k in preference set

$m_i =$ number of alternatives in class i

$a_{jk} =$ score of alternative k on criterion j

$a_j^t =$ breakpoint t on criterion j

$p, \delta_1, \delta_2 =$ small positive constants (0.001; 0.0001; 0.0001)

Decision Variables

w_{js} = utility value corresponding to interval s on criterion j

u_i = threshold value between class i and i+1

ε_k^- = error of assignment of alternative k to a higher class

ε_k^+ = error of assignment of alternative k to a lower class

UTADIS Model

$$\min \sum_{i=1}^{n} \left[\frac{\sum_{\forall x_k \in C_i} (\varepsilon_k^+ + \varepsilon_k^-)}{m_i} \right] \tag{14}$$

s.t.

$$\sum_{j=1}^{q} \left[\sum_{s=1}^{r_{jk}-1} w_{js} + w_{j,r_{jk}} \frac{a_{jk} - a_j^{r_{jk}}}{a_j^{r_{jk}-1} - a_j^{r_{jk}}} \right] - u_1 + \varepsilon_k^+ \geq \delta_1 \quad \forall x_k \in C_1 \tag{15}$$

$$\sum_{j=1}^{q} \left[\sum_{s=1}^{r_{jk}-1} w_{js} + w_{j,r_{jk}} \frac{a_{jk} - a_j^{r_{jk}}}{a_j^{r_{jk}-1} - a_j^{r_{jk}}} \right] - u_i + \varepsilon_k^+ \geq \delta_1$$

$$\forall x_k \in C_i \ (1,2,3,\dots,n-1) \tag{16}$$

$$\sum_{j=1}^{q} \left[\sum_{s=1}^{r_{jk}-1} w_{js} + w_{j,r_{jk}} \frac{a_{jk} - a_j^{r_{jk}}}{a_j^{r_{jk}-1} - a_j^{r_{jk}}} \right] - u_{i-1} - \varepsilon_k^- \leq -\delta_2$$

$$\forall x_k \in C_i \ (1,2,3,\dots,n-1) \tag{17}$$

$$\sum_{j=1}^{q} \left[\sum_{s=1}^{r_{jk}-1} w_{js} + w_{j,r_{jk}} \frac{a_{jk} - a_j^{r_{jk}}}{a_j^{r_{jk}-1} - a_j^{r_{jk}}} \right] - u_{i-1} - \varepsilon_k^- \geq -\delta_2 \quad \forall x_k \in C_n \tag{18}$$

$$\sum_{j=1}^{q} \sum_{s=1}^{p_j-1} w_{js} = 1 \tag{19}$$

$$u_{i-1} - u_i \geq p \tag{20}$$

$$w_{js}, \varepsilon_k^+, \varepsilon_k^- \geq 0 \tag{21}$$

The optimal w_{js} and u_k values are used to determine the utility values of alternatives in test set and their assigned class according to their position between thresholds.

3.4 Classification Tree

Classification (decision) tree is a data mining technique for classification problems based on some decision rules used to partition the alternatives. The decision rules are implemented in IF…THEN conditions. C4.5 algorithm (Quinlan, 1993) uses these rules to construct a tree as shown in Figure 3.4.

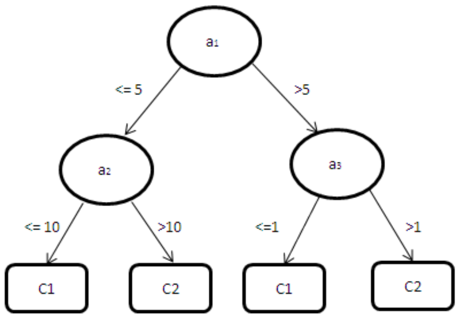


Figure 3-4 Classification decision tree

Each decision node in the tree represents an attribute. The branches coming out of a node separate the alternatives according to the decision rule indicated. The branching process continues until the separated alternatives belong to one class. The tree is formed by use of alternatives in preference set. The algorithm proceeds in iterative steps such as:

1. Let S_t be the set of alternatives that reach a node t.
2. If S_t contains alternatives that belong the same class C_i, then t is a leaf node labelled as C_i .
3. If S_t contains alternatives that belong to more than one class, a decision rule on an attribute is used to split the alternatives into smaller subsets.

Apply the procedure to each subset recursively starting from the root node.

The decision rules are about how to split to attributes in each attribute node. There are binary splits which create 2 branches and multi-way splits which create to more than 2 branches, e.g. as many parts as the values on that attribute for categorical attributes. For continuous attributes, another decision is to determine the best split to branch. There are several performance measures such as GINI index or entropy to measure the performance of split decision.

The resulting tree can be pruned (post-pruning) or the algorithm can be stopped before a full tree is formed (pre-pruning) in order to avoid over fitting which is the case where the tree classifies the alternatives in preference set with a high accuracy but fails to succeed in the test set. After the final tree is conducted, it is used to determine the classes of test variables by using the decision rules for each alternative to reach one of the leaf nodes.

CHAPTER 4

SOLUTION APPROACH

The common property of previously mentioned existing models for sorting problems is that they form the criteria aggregation function by estimating weights of the criteria and define thresholds once with the preference set. The values of the weights and thresholds are chosen such that the classification of alternatives in preference set is done with minimum error. Yet, classification error may be minimized by more than one set of values which result in alternative optimal solutions. Although these are alternative optimals resulting in the same classification error for the preference set, they may not result in the same classification accuracy for the testing set. So, randomly choosing one of these solutions may not be the best choice for higher accuracy of the model. The alternative optimal solutions are handled by further analysis with secondary objectives in many studies (Köksalan and Özpeynirci, 2008). This requires evaluation of many possible objectives and the objectives that give better performance in some data sets may fail to do so in other data sets. Also, using secondary objectives do not consider all possible alternative optimal solutions; therefore do not guarantee to obtain the best alternative solution that gives highest accuracy. None of the previous methods include identification and handling of alternatives in the main structure of the method but offer as a further analysis up to our knowledge. Our solution approach in this study proposes a solution by calculating the minimum and maximum values of these decision variables in order to determine the probabilities of belonging to a class for an alternative in the test set.

The proposed solution approach in this study includes two optimization models and a probability calculation algorithm as shown in Figure 4.1. The first model calculates the minimum and maximum values of thresholds for each class that gives the minimum classification error for preference set. The second model takes this error as a binding constraint and calculates the minimum and maximum values of alternatives in the test set which occur as a result of the alternative weights that gives the minimum classification error. Then, these values are used by probability calculation algorithm to define the probabilities. The details of each of these steps are given in the following subsections.

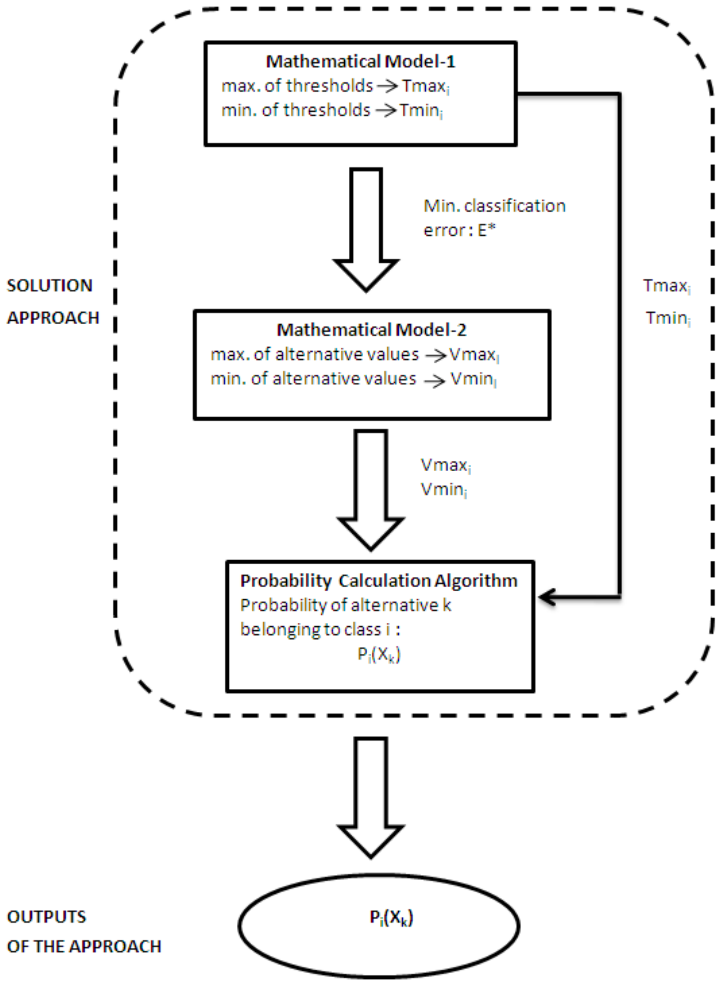


Figure 4-1 Steps of the solution approach

4.1 Criteria Aggregation Function

Before description of the criteria aggregation function and the details of the models, the problem should be defined in a formal way with the notations that will be used in the following sections. The problem is to sort the alternatives in preference set $P = \{X_1, X_2, \dots, X_K\}$ into n predefined classes $\{C_1, C_2, \dots, C_n\}$ with minimum classification error. The classes are ordered such that C_1 is the most preferred and C_n is the least preferred class. An alternative X_k is defined on a set of criteria $A = \{a_1, a_2, \dots, a_q\}$ such as $X_k = (a_{k1}, a_{k2}, \dots, a_{kq})$ where the a_{kj} values are the scores of alternative k on criterion j. The sorting process is done by defining a criteria aggregation function that maps the alternatives on R^1 by defining proper weights of the criteria $(w_1, w_2, \dots, w_q)$ and thresholds that separate the classes $(T_1, T_2, \dots, T_{n-1})$. The established criteria aggregation function is used to sort the alternatives in test set $T = \{X_1, X_2, \dots, X_L\}$ which will be quite different in our model as it will be explained later.

The general structure of the model is similar to the MP formulations of discriminant analysis yet the criteria aggregation function is quite different. Instead of adding the scores of alternatives on each criterion, the distance of an alternative to the ideal point is calculated and taken as the value of that alternative. The L-p norm distance of an alternative x_k to ideal point I^* is:

$$D(x_k, I^*) = \left[\sum_{j=1}^{q} \left(w_j \left|a_{kj} - I_j^*\right|\right)^p\right]^{1/p} = V_k \qquad (22)$$

where a_{kj} is the score of alternative k on criterion j, w_j is the weight of criterion j, I_j^* is the best value on criterion j where the alternatives are defined on q criteria. V_k is the value of alternative x_k which is the mapped value from $R^q \rightarrow R^1$. Different p norms are applied in the model and evaluated in following sections.

On R^1 space, the V_k values are separated into different classes by establishing thresholds in the range $(0, \infty)$ since distances can not be negative. The lower the V_k value of alternative x_k, the more preferred the alternative is since as an alternative gets closer to the ideal point it is more preferred.

$$V_i < V_j \;\rightarrow\; X_i \gg X_j \tag{23}$$

So, class-1 is in the range of $(0, T_1)$, class-2 is in the range of (T_1, T_2), and so on as shown in Figure 4.2.

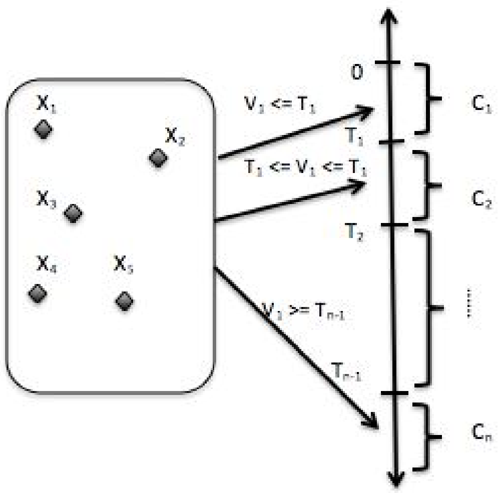


Figure 4-2 Sorting of alternatives to classes

The alternatives are assigned to classes according to the following rule:

$$V_k \leq T_1 \;\rightarrow X_k \in C_1 \tag{24}$$

$$T_{i-1} < V_k \leq T_i \;\rightarrow X_k \in C_i \quad \forall i = 2,3, \ldots, n-1 \tag{25}$$

$$V_k > T_{n-1} \;\rightarrow X_k \in C_n \tag{26}$$

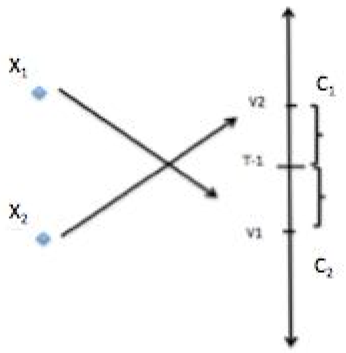
This distance approach eliminates the need for pre-transformation of scores on each criterion to increasing (higher-the-better) or decreasing (lower-the-better) structure.

The main objective of the model is to construct several criteria aggregation functions that minimize the classification error of alternatives in preference set. The classification error for the misclassified alternatives is calculated as:

$$e_k^- = max\{ 0, T_{i-1} - V_k \} \tag{27}$$

$$e_k^+ = max\{ 0, V_k - T_i\} \tag{28}$$

for every $x_k \in C_i$ where T_i denotes the threshold value of class i, T_{i-1} is the threshold value for the lower class, and e_k^+ and e_k^- are the errors of assignment to a lower (worse) class and a higher (better) class respectively.



$$e_2^- = V_2 - T_1$$

$$e_1^+ = T_1 - V_1$$

Figure 4-3 Errors of alternatives $X_1 \in C_1$ and $X_2 \in C_2$

4.2 Model-1

The first model is similar to the previously mentioned sorting methods except that only thresholds and target classification error are permanently determined. The criteria aggregation function is estimated in the first phase by stabilizing weights and thresholds in the previous studies (Erengüç and Koehler, 1990; Freed and Glover, 1979). In our model, different weight vectors are used to find maximum and minimum values of thresholds but these weights are not kept for a permanent criteria aggregation function since in the second phase, the objective changes as to find the possible maximum and minimum values of alternatives in the test set. So, in our method the criteria aggregation function changes with different weights since the objective in the first model is to find the alternative values of thresholds with minimum classification error. In the second model, within the alternative values that give the minimum classification error found in model-1, the maximum and minimum distances to the ideal point for each test alternative are determined.

The decision variables and parameters of the model are defined as follows:

Indexes

$n =$ number of classes

$q =$ number of criteria

$K =$ number of alternatives in preference set

$$t = \begin{cases} 1 \text{ for maximum threshold values} \\ 2 \text{ for minimum threshold values} \end{cases}$$

$i \in \{1,2,\dots,n-1\}$ for thresholds

$j \in \{1,2,\dots,q\}$ for criteria

$k \in \{1,2,\dots,K\}$ for alternatives in preference set

Parameters

a_{kj} = value of alternative k on criterion j

I_j^* = value of ideal point on criterion j

∂ = a small constant (0.005)

C_i = set of alternatives in preference set which belongs to class i

Decision Variables

w_{tj} = weight of criterion j

e_{kt}^{+} = error of assignment of alternative k to a lower class

e_{kt}^{-} = error of assignment of alternative k to a higher class

$Tmax_i$ = maximum value of threshold i separating class i and i+1

$Tmin_i$ = minimum value of threshold i separating class i and i+1

Model-1

$$\min \ \sum_{t=1}^{2}\sum_{k=1}^{K}(e_{kt}^{+}+e_{kt}^{-})+\partial\sum_{i=1}^{n-1}(Tmin_i - Tmax_i) \tag{29}$$

s. t.

$$\left[\sum_{j=1}^{q}(w_{1j}\left|a_{kj}-I_j^*\right|)^p\right]^{1/p} - e_{k1}^{+} \leq Tmin_1 \quad \forall k \in C_1 \tag{30}$$

$$\left[\sum_{j=1}^{q}(w_{1j}\left|a_{kj}-I_j^*\right|)^p\right]^{1/p} + e_{k1}^{-} \geq Tmin_{i-1} \quad \forall k \in C_i \ (i=2,3,\dots n) \tag{31}$$

$$\left[\sum_{j=1}^{q}(w_{1j}\left|a_{kj}-I_j^*\right|)^p\right]^{1/p} - e_{k1}^{+} \leq Tmin_i \quad \forall k \in C_i \ (i=2,3,\dots n-1) \tag{32}$$

$$\left[\sum_{j=1}^{q}(w_{2j}\left|a_{kj}-I_j^*\right|)^p\right]^{1/p} - e_{k2}^{+} \leq Tmax_1 \quad \forall k \in C_1 \tag{33}$$

$$\left[\sum_{j=1}^{q}(w_{2j}\left|a_{kj}-I_j^*\right|)^p\right]^{1/p} + e_{k2}^{-} \geq Tmax_{i-1} \quad \forall k \in C_i \ (i=2,3,\dots n) \tag{34}$$

$$\left[\sum_{j=1}^{q}(w_{2j}\left|a_{kj}-I_j^*\right|)^p\right]^{1/p} - e_{k2}^{+} \leq Tmax_i \quad \forall k \in C_i \ (i=2,\dots n-1) \tag{35}$$

$$Tmax_i \geq Tmax_{i-1} \qquad \forall i \in \{1,2,\dots,n-1\} \tag{36}$$

$$Tmin_i \geq Tmin_{i-1} \qquad \forall i \in \{1,2,\dots,n-1\} \tag{37}$$

$$Tmax_i \geq Tmin_i \qquad \forall i \in \{1,2,\dots,n-1\} \tag{38}$$

$$\sum_{j=1}^{C} w_{tj} = 1 \qquad \forall t \in \{1,2\} \tag{39}$$

$$w_{tj} \geq 0, e_{kt}^{+} \geq 0, e_{kt}^{-} \geq 0 \qquad \forall t \in \{1,2\}, \forall j \in \{1,2,\dots,q\}, \forall k \in \{1,2,\dots,K\} \tag{40}$$

The index t in the model is used to create two different criteria aggregation functions for each value of t. For t=1, a set of weights (w_{1j}) is found that minimizes the threshold values ($Tmin$) and for t=2, another set of weights (w_{2j}) is found that maximizes the threshold values ($Tmax$) among the alternative threshold values that result in same minimum classification error. Each of these weight vectors sum up to 1 for all criteria. (e_{1t}^{+}, e_{1t}^{-}) and (e_{2t}^{+}, e_{2t}^{-}) are the errors resulting in two criteria aggregation functions. The idea is to be able to explore each set of values of alternative solutions that results in minimum threshold and maximum threshold. If one criteria aggregation function was used, only one set of values would be found which would maximize the range ($Tmax - Tmin$). We can explain this idea by the following example. Assume that one criteria aggregation function was used instead of two. For two class case, two alternative optimal solutions shown in Figure 4.4 would result in the same ($Tmax - Tmin$) value and same classification error (0) so the model would give any of the two as the optimal solution. Yet, assuming that these two are the only alternative optimal solutions, we want to identify $Tmax$ value of the second solution and $Tmin$ value of the first solution. In order to do that, we need to identify two different configurations of alternatives which is not possible without using two different weight spaces.

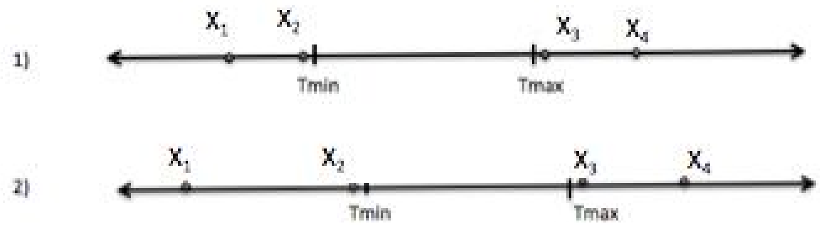


Figure 4-4 Two alternative optimal solutions in 2-class case where $\{X_1, X_2\} \in C_1$ and $\{X_3, X_4\} \in C_2$

The first objective is to minimize total error and the secondary objectives are to maximize the maximum threshold value and minimize the minimum threshold value. When there is only one solution (only one set of weights and thresholds) that gives the minimum classification error, then there is only one value for each threshold that minimizes classification error. So, the maximum and minimum threshold variables take the same value. The constraints between Eq.30 and Eq.32 ensure that the alternatives assigned to class-i in preference set take values out of the range of the correct class. Otherwise the error variable takes a value equal to the distance out of the correct class's threshold. The constraints between Eq.33 and Eq.35 are for the same purpose but this time for the decision variables defined for maximum thresholds. Eq.36 and Eq.37 are defined to ensure that thresholds of better classes are smaller than thresholds of worse classes. Eq.39 equates the sum of weights for each criterion to 1 for each weight set.

The total classification error of model-1 is shown below which is given to the second model as a constraint:

$$\sum_{t=1}^{2} \sum_{k=1}^{K} (e_{kt}^{+} + e_{kt}^{-}) = E^{*} \tag{41}$$

4.3 Model-2

The second model is quite similar to the first one yet the alternatives in the test set are also included in this model. The objective is to find the maximum and minimum values of each alternative in the test set among the solutions that give the minimum classification error. In order to do that, both the classification of alternatives in the preference set and establishing the values of alternatives in the test set are done simultaneously. The minimum classification error found in model-1 (E^{*}) is taken as an upper bound on the classification error of alternatives in the preference set.

$$\sum_{\forall k \in P} e_k \leq E^* \quad (42)$$

Again, different criteria aggregation functions are used to find the $Vmax$ and $Vmin$ values of alternatives on set T. For each $Vmax_l$ or $Vmin_l$ of alternative l, a different set of weights and thresholds is found if there exist alternative optimal solutions with E^* classification error. So, there are $2L$ sets of weights and thresholds which result in $2Lq$ weight variables and $2L(n-1)$ thresholds which were only $2q$ and $2(n-1)$ respectively in the first model. This increases the computational effort especially when the L_P norm is different than L_1 and L_∞, which makes both Model-1 and Model-2 nonlinear and becomes a problem especially with massive data sets. Since the classification error is taken as a constraint, the objective is to maximize $Vmax$ and minimize $Vmin$.

$$\min \quad \sum_{\forall l \in T}(Vmin_l - Vmax_l) \quad (43)$$

The total weight of all criteria for each weight set is again equated to 1. The threshold values for each class are limited between the minimum and maximum values found in Model-1.

$$Tmin_i \leq T_i \leq Tmax_i \quad \forall i \quad (44)$$

<u>*Additional Indexes*</u>

L = number of alternatives in test set

$l \in \{1,2,\dots,L\}$ for alternatives in test set

<u>*Additional Parameters*</u>

a_{lj} = score of alternative l on criterion j

E^* = total classification error

$Tmin_i$ = minimum value of threshold separating class i and i+1

$Tmax_i$ = maximum value of threshold separating class i and i+1

Decision Variables

w_{ltj} = weight of criterion j

e_{klt}^{+} = error of assignment of alternative k to a lower class

e_{klt}^{-} = error of assignment of alternative k to a higher class

T_{lti} = value of threshold i

$Vmax_l$ = maximum value of alternative l in test set

$Vmin_l$ = minimum value of alternative l in test set

Model-2

$$\min \quad \sum_{l=1}^{L}(Vmin_l - Vmax_l) \tag{45}$$

s. t.

$$\left[\sum_{j=1}^{q}(w_{ltj}\left|a_{kj} - I_j^*\right|)^p\right]^{1/p} - e_{klt}^{+} \leq T_{lt1}$$

$$\forall k \in C_1, \forall l \in \{1,2,\dots,L\}, \forall t \in \{1,2\} \tag{46}$$

$$\left[\sum_{j=1}^{q}(w_{ltj}\left|a_{kj} - I_j^*\right|)^p\right]^{1/p} + e_{klt}^{-} \geq T_{lti-1}$$

$$\forall k \in C_i \ (i = 2,3,\dots n), \forall l \in \{1,2,\dots,L\}, \forall t \in \{1,2\} \tag{47}$$

$$\left[\sum_{j=1}^{q}(w_{ltj}\left|a_{kj} - I_j^*\right|)^p\right]^{1/p} - e_{klt}^{+} \leq T_{lti}$$

$$\forall k \in C_i \ (i = 2,\dots n-1), \forall l \in \{1,2,\dots,L\}, \forall t \in \{1,2\} \tag{48}$$

$$\left[\sum_{j=1}^{q}(w_{l1j}\left|a_{lj} - I_j^*\right|)^p\right]^{1/p} \geq Vmax_l \quad \forall l \in \{1,2,\dots,L\} \tag{49}$$

$$\left[\sum_{j=1}^{q}(w_{l1j}\left|a_{lj} - I_j^*\right|)^p\right]^{1/p} \leq Vmin_l \quad \forall l \in \{1,2,\dots,L\} \tag{50}$$

$$T_{lti} \geq T_{lti-1} \quad \forall l \in \{1,2,\dots,L\}, \forall t \in \{1,2\}, \forall i \in \{2,3,\dots,n-1\} \tag{51}$$

$$\sum_{j=1}^{q} w_{ltj} = 1 \qquad \forall l \in \{1,2,\dots,L\}, \forall t \in \{1,2\} \tag{52}$$

$$\sum_{t=1}^{2} \sum_{k=1}^{K} \sum_{l=1}^{L} (e_{klt}^{+} + e_{klt}^{-}) \le LE^{*} \tag{53}$$

$$Tmax_i \ge T_{lti} \qquad \forall l \in \{1,2,\dots,L\}, \forall t \in \{1,2\}, \forall i \in \{1,2,3,\dots,n-1\} \tag{54}$$

$$Tmin_i \le T_{lti} \qquad \forall l \in \{1,2,\dots,L\}, \forall t \in \{1,2\}, \forall i \in \{1,2,3,\dots,n-1\} \tag{55}$$

$$w_{ltj} \ge 0, e_{klt}^{+} \ge 0, e_{klt}^{-} \ge 0 \qquad \forall t \in \{1,2\}, \forall j \in \{1,2,\dots,q\}, \forall k \in \{1,2,\dots,K\},$$

$$\forall l \in \{1,2,\dots,L\} \tag{56}$$

The constraints of Model-2 are equivalent to Model-1 except that dimension l is included in the decision variables.

4.4 Model-1 and Model-2 for L_∞

The models given in sections 4.2 and 4.3 are valid for all L_p norms except $p = \infty$ since its distance function is quite different than the others. The distances in each dimension are not added as in the other norms but only the maximum of them is taken as the distance of two points. In this case, the L_∞ distance (*Tchebycheff distance*) is calculated as follows;

$$D(x_k, I^*) = \max_j \left(w_j \left| a_{kj} - I_j^* \right| \right) = V_k \tag{57}$$

The function is not included in the model like this but instead it is inserted in the objective in order to construct a linear model. The first objective is to define these distances correctly so its coefficient is 1. The secondary objective is to minimize the total classification error as in the other models. Among the alternative optimal solutions that give the minimum classification error, the ones that give the minimum and maximum values of thresholds and testing alternatives are chosen in model-1 and model-2 respectively.

All the variables and parameters are same except the decision variable Dp and Dt, which are the global values (distances to ideal point) of alternatives in preference set and test set respectively. Model-1 and model-2 for L_∞ are shown below.

Model-1:

Additional Decision Variables

Dp_{kt} = weighted L_∞ distance of alternative k ∈ P to ideal point

Additional Parameters

∂ = a small positive constant (0.005)

β = a small positive constant greater than ∂ (0.01)

Model-1

$$\min \quad \beta \sum_{k=1}^{K} \sum_{t=1}^{2} (e_{kt}^{+} + e_{kt}^{-}) + \partial \sum_{i=1}^{n-1} (Tmin_i - Tmax_i) + \sum_{k=1}^{K} \sum_{t=1}^{2} Dp_{kt} \quad (58)$$

s.t.

$$Dp_{kt} \geq w_{tj}(a_{kt} - I_j^*) \quad \forall k \in \{1,2,\dots,K\}, \forall t \in \{1,2\}, \forall j \in \{1,2,\dots,q\} \quad (59)$$

$$Dp_{k1} - e_{k1}^{+} \leq Tmin_1 \quad \forall k \in C_1 \quad (60)$$

$$Dp_{k1} + e_{k1}^{-} \geq Tmin_{i-1} \quad \forall k \in C_i (i = 2,3,\dots n) \quad (61)$$

$$Dp_{k1} - e_{k1}^{+} \leq Tmin_i \quad \forall k \in C_i (i = 2,\dots n-1) \quad (62)$$

$$Dp_{k2} - e_{k2}^{+} \leq Tmax_2 \quad \forall k \in C_1 \quad (63)$$

$$Dp_{k2} + e_{k2}^{-} \geq Tmax_{i-1} \quad \forall k \in C_i (i = 2,3,\dots n) \quad (64)$$

$$Dp_{k2} - e_{k2}^{+} \leq Tmax_i \quad \forall k \in C_i (i = 2,\dots n-1) \quad (65)$$

$$Tmax_i \geq Tmax_{i-1} \quad \forall i \in \{1,2,\dots,n-1\} \quad (66)$$

$$Tmin_i \geq Tmin_{i-1} \quad \forall i \in \{1,2,\dots,n-1\} \quad (67)$$

$$Tmax_i \geq Tmin_i \quad \forall i \in \{1,2,\dots,n-1\} \quad (68)$$

$\sum_{j=1}^{C} w_{tj} = 1 \qquad \forall t \in \{1,2\}$ (69)

$w_{tj} \geq 0, e_{kt}^{+} \geq 0, e_{kt}^{-} \geq 0 \qquad \forall t \in \{1,2\}, \forall j \in \{1,2,\ldots,q\}, \forall k \in \{1,2,\ldots,K\}$ (70)

The constraints of the model are equivalent to the constraints of the general distance norm approach. The only difference is the calculation of criteria aggregation function which is refered as Dp_{kt}.

Model-2:

Additional Decision Variables

Dp_{klt} = weighted L_∞ distance of alternative k ∈ P to ideal point
Dt_{lt} = weighted L_∞ distance of alternative l ∈ T to ideal point

Additional Parameters

∂ = a small positive constant (0.005)

Model-2

min $\quad \partial \sum_{l=1}^{L}(Vmin_l - Vmax_l) + \sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{t=1}^{2} Dp_{klt} + \sum_{l=1}^{L}\sum_{t=1}^{2} Dt_{lt}$ (71)

s.t.

$Dp_{klt} \geq w_{ltj}\left(a_{kt} - I_j^*\right) \quad \forall k \in \{1,2,\ldots,K\}, \forall t \in \{1,2\}, \forall l \in \{1,2,\ldots,L\},$

$\forall j \in \{1,2,\ldots,q\}$ (72)

$Dt_{lt} \geq w_{ltj}\left(a_{kt} - I_j^*\right) \quad \forall l \in \{1,2,\ldots,L\}, \forall t \in \{1,2\}, \forall l \in \{1,2,\ldots,L\},$

$\forall j \in \{1,2,\ldots,q\}$ (73)

$Dp_{klt} - e_{klt}^{+} \leq T_{lt1} \qquad \forall k \in C_1, \forall t \in \{1,2\}, \forall l \in \{1,2,\ldots,L\}$ (74)

$Dp_{klt} + e_{klt}^{-} \geq T_{lti-1} \quad \forall k \in C_i (i = 2,3,\ldots n), \forall t \in \{1,2\}, \forall l \in \{1,2,\ldots,L\}$ (75)

$Dp_{klt} - e_{klt}^{+} \leq T_{lti} \qquad \forall k \in C_i (i = 2,\ldots n-1), \forall t \in \{1,2\}, \forall l \in \{1,2,\ldots,L\}$ (76)

$Dt_{l1} \leq Vmin_l \qquad \forall l \in \{1,2,\ldots,L\}$ (77)

$$Dt_{l2} \geq Vmax_l \qquad \forall l \in \{1,2,\dots,L\} \tag{78}$$

$$\sum_{j=1}^{q} w_{ltj} = 1 \qquad \forall l \in \{1,2,\dots,L\}, \forall t \in \{1,2\} \tag{79}$$

$$\sum_{t=1}^{2} \sum_{k=1}^{K} \sum_{l=1}^{L} (e_{klt}^{+} + e_{klt}^{-}) \leq LE^{*} \tag{80}$$

$$Tmax_i \geq T_{lti} \qquad \forall l \in \{1,2,\dots,L\}, \forall t \in \{1,2\}, \forall i \in \{1,2,3,\dots,n-1\} \tag{81}$$

$$Tmin_i \leq T_{lti} \qquad \forall l \in \{1,2,\dots,L\}, \forall t \in \{1,2\}, \forall i \in \{1,2,3,\dots,n-1\} \tag{82}$$

$$w_{ltj} \geq 0, e_{klt}^{+} \geq 0, e_{klt}^{-} \geq 0 \qquad \forall t \in \{1,2\}, \forall j \in \{1,2,\dots,q\}, \forall k \in \{1,2,\dots,K\},$$

$$\forall l \in \{1,2,\dots,L\} \tag{83}$$

The constraints of Model-2 are equivalent to Model-1 with additional dimension l for each decision variable. The constant ∂ is set to 0.005 by empirical evaluation of different values to ensure that the minimum and maximum values of alternatives are set among the solutions with the weighted distances of alternatives assigned correctly.

4.5 Probability Calculation

After the two models are solved, the optimal $Vmax_l$, $Vmin_l$, $Tmin_i$ and $Tmax_i$ values are used in order to define the probability of belonging to each class for the alternatives in test set. Since values of the alternatives are defined as intervals rather than points and so are the thresholds, the sorting of alternatives in test set into classes is not straightforward. If the thresholds were taken as points, the interval of an alternative could lie on only one class (Figure 4.5-a), or more than one classes (Figure 4.5-b), and the probability could be calculated by taking ratio of the part of an alternative lying on one class to the whole with the assumption of uniform distribution. In the first case, the probability of the class that the interval of alternative lies on is assigned as 1 and the other classes 0 since all the possible values of an alternative are between its minimum and maximum values. In the second case, all the classes that the interval of alternative covers take positive probabilities which are calculated according to the ratios lying on each class which sum up to 1 for all classes for an alternative.

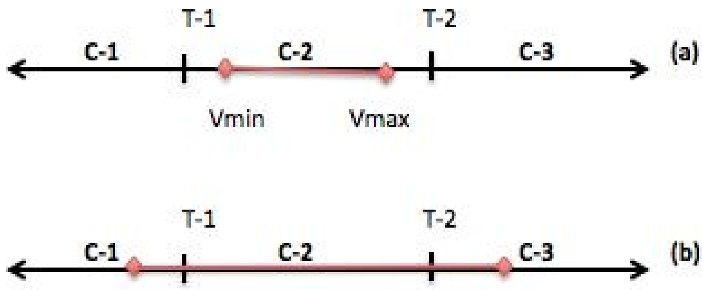


Figure 4-5 (a) Alternative lies on only one class. (b) Alternative lies on more than one classes

Yet, in our model, the thresholds also have intervals. This makes the identification of classes between thresholds more complicated. When the maximum value of a threshold i is less than the minimum value of the following threshold $i+1$, it is rather simple. Between the minimum and maximum values of a threshold i, class i and $i+1$ are possible. Between the minimum value of a threshold i and the maximum value of threshold $i-1$, only possible class is i. Between the maximum value of threshold i and minimum value of threshold $i+1$, only possible class is class i+1. The situation where intervals of thresholds do not coincidence is shown in Figure 4.6.

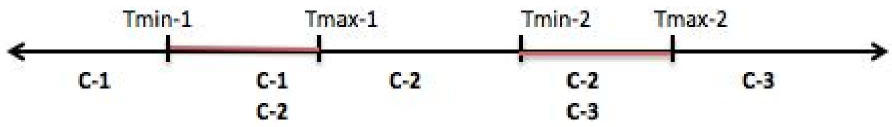


Figure 4-6 Classes between intervals when there is no intersection of threshold intervals

When the intervals of thresholds intersect, the classes are assigned in a similar way. Yet, in that case, the number of classes that can exist in an interval can be more than 2 depending on the number of intervals intersecting. The class assignments in the case of 2 and 3 intervals intersecting are shown in Figure 4.7.

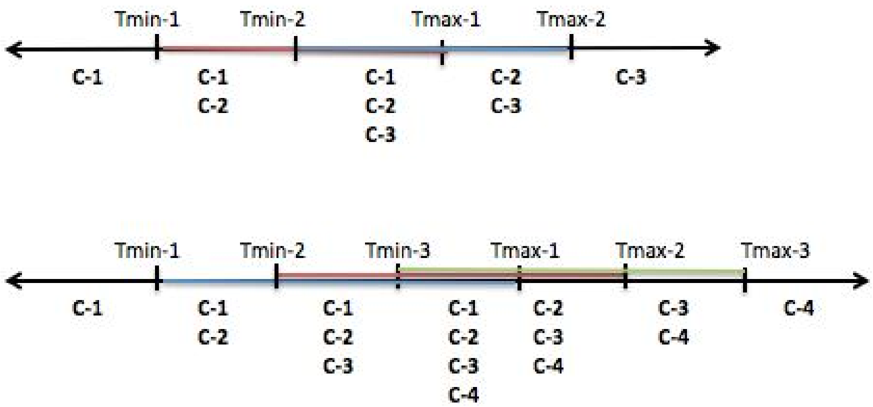


Figure 4-7 Classes between intervals when there are 2 and 3 intersecting intervals of thresholds

The algorithm to identify the possible classes between intervals is shown below:

Let $\{S_1, S_2, \ldots, S_n\}$ and $\{E_1, E_2, \ldots, E_n\}$ denote the order of intervals which each class start and end. For instance, $S_1 = 1$ and $E_1 = 3$ means class-1 starts in interval 1 and continues until interval 4. So, a class is open for the intervals between S_i and E_i. Let R denote the number of intervals and there are n classes. Let T = $\{T_1, T_2, \dots, T_{2(n-1)}\}$ be the ordered set of thresholds and C = $\{C_1, C_2, \dots, C_n\}$ be the set of classes.

Step.0. Set $S_1 = 1$ and $E_n = R$. Set the iteration counter i=0.

Step.1. Increase i by 1. For $\forall j \in \{1,2,\dots,n\}$, if $T_i = Tmin_j$, set $S_{j+1} = 1$ and if $T_i = Tmax_j$, set $S_j = 1$.

Step.2. If i=R, stop. Otherwise, go to step-1.

The algorithm follows the rule that $Tmin_j$ opens class $(j + 1)$ and $Tmax_j$ closes class j. Opening a class means that class can exist after that interval and closing means that class can not exist after that interval.

Before proceeding to the probability calculation phase, we need to introduce two concepts, optimistic case and pessimistic case. As it is shown, rather than assigning an interval between thresholds to a class permanently, in these methods the classes are defined on the intervals that they can exist. This means a class may or may not exist on its possible interval, depending on the choice of the thresholds. Since there are infinite choices to define the boundaries of the classes, we consider only two cases which are more meaningful. The first one is the case where the boundaries of a class are as wide as possible, which is the optimistic case for that class. In this case, all the possible intervals that a class can exist are taken as the intervals of that class. The second one is the case where the boundaries of a class are as narrow as possible, which is the pessimistic case. In this case, only the intervals that a class exists for 100% of the time are taken as the boundaries of that class. These intervals are the ones that only one class can exist. If a class is not defined alone in any interval, then in the pessimistic case, that class has no interval. The optimistic boundaries of classes with the maximum and minimum thresholds given are shown in Figure 4.8.

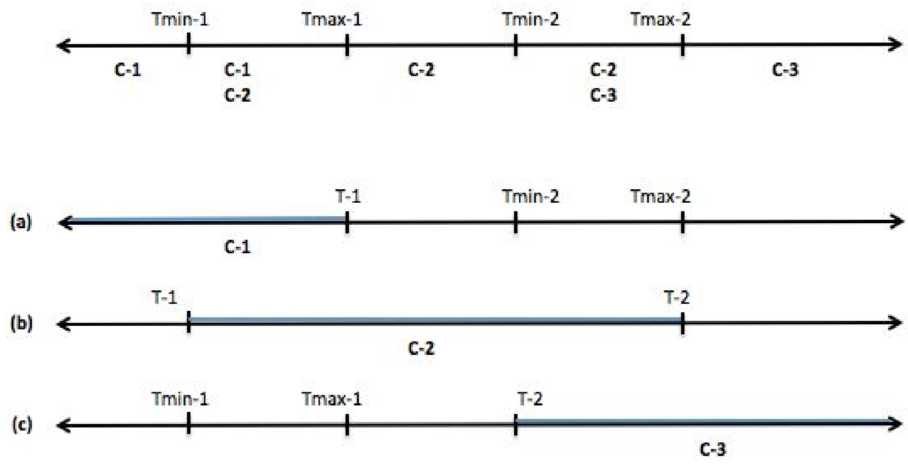


Figure 4-8 The optimistic intervals for (a) class-1, (b) class-2, (c) class-3

As it can be seen in Figure 4.8, to find the optimistic interval for a class, the threshold before the class is taken as its minimum value and the threshold after the class is taken as its maximum value. The pessimistic boundaries for each class for the same case are shown in Figure 4.9.

When the optimistic or pessimistic intervals are found for a class, the other classes and their thresholds are not considered. So the boundaries of only one class are established at a time. When there are only two classes, the optimistic case of one class is the pessimistic case of the other class.

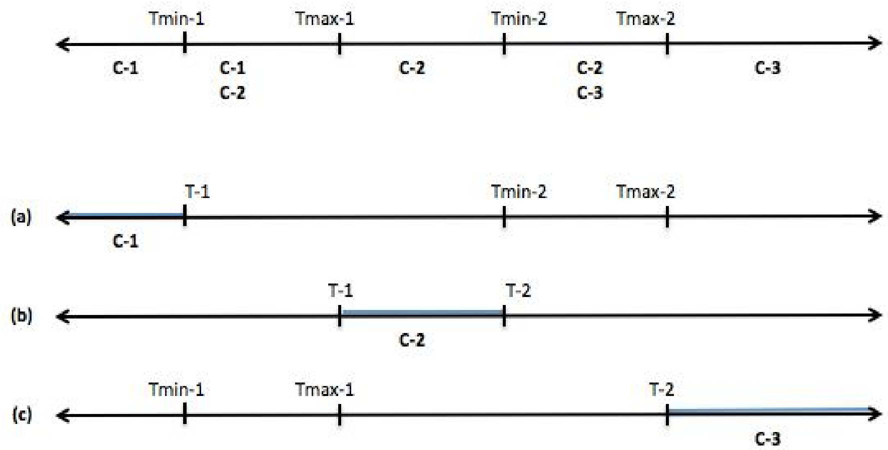


Figure 4-9 The pessimistic intervals for (a) class-1, (b) class-2, (c) class-3

The probabilities of belonging to a class are also calculated according to the optimistic and pessimistic cases. The optimistic probabilities for each class are defined for the optimistic cases of those classes and vice versa for the pessimistic probabilities. When calculating the probabilities, two distributions, which are uniform distribution and triangular distribution, are considered. The cumulative distribution function for the uniform distribution when the lower and upper points are a and b respectively is:

$$F(X) = \begin{cases} 0 & for\ x < a \\ \frac{x-a}{b-a} & for\ a < x < b \\ 1 & for\ \ x > b \end{cases} \tag{84}$$

If the alternatives are uniformly distributed between their minimum and maximum values, the probabilities of belonging to each class for an alternative are calculated as the ratios of the proportion of the alternative intersecting with the interval of the class. Let A_{ki} be the length of the part of alternative X_k that intersects with class i. Then, the probability of belonging to class i for X_k is:

$$P_i\,(X_k) = \frac{A_{ki}}{\sum_{i=1}^{n} A_{ki}} \tag{85}$$

A_{ki} values are determined for the optimistic and pessimistic cases of the classes and the probabilities are given for each case. The probability calculation for uniform distribution is shown in Figure 4.10. Optimistic probability for class-1 is $P_1\,(X_k) = \frac{Tmax_1 - Vmin}{Vmax - Vmin}$ for the case shown in Figure 4.10 and the pessimistic probability is 0 since the interval of alternative does not coincidence with the pessimistic interval of class-1.

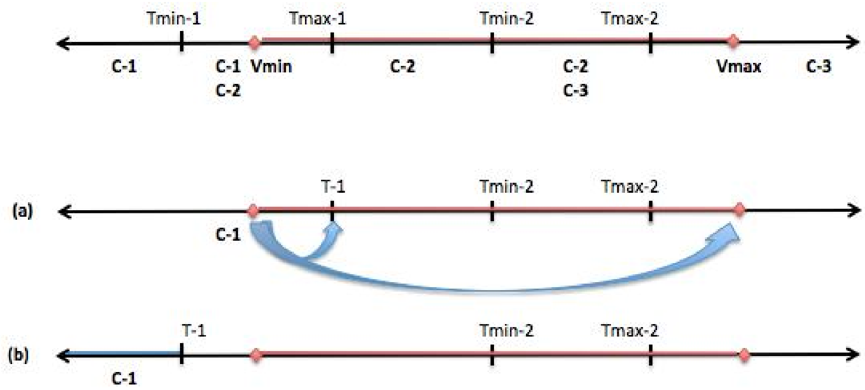


Figure 4-10 The probability calculation for uniform distribution in (a) optimistic and (b) pessimistic case for class-1

The cumulative distribution function for triangular distribution when a, b and c are lower, upper and mode values respectively is:

$$F(X) = \begin{cases} \frac{(x-a)^2}{(b-a)(c-a)} & for\ a \leq x \leq c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & for\ c \leq x \leq b \end{cases} \tag{86}$$

The mode is taken as the middle point between $Vmin$ and $Vmax$ values in probability calculations. The distribution of an alternative is as shown in the Figure 4.11.

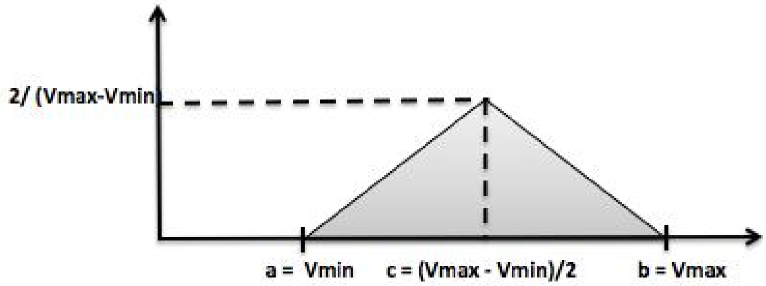


Figure 4-11 Triangular distribution of an alternative between Vmin and Vmax values

Let A_{ki} be the length of the part of alternative X_k that intersects with class i. Depending on the position of the intersection on the interval of the class, the calculation of the probabilities are different. The possible configurations are shown in the Figure 4.12.

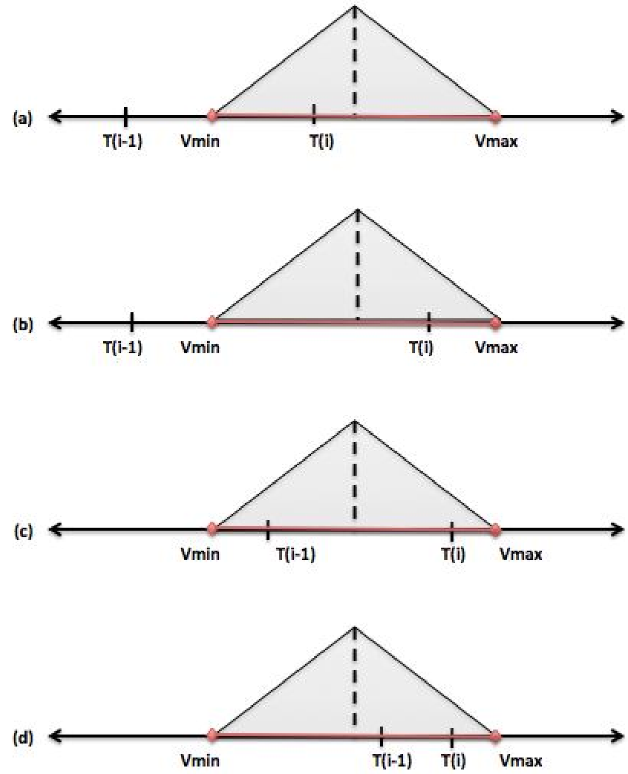


Figure 4-12 Different positions of intersections between classes and alternatives

The corresponding probability calculations for each of the configurations shown in Figure 4.12 are:

(a) $$P_i(X) = \frac{4(T_i - Vmin)}{(Vmax - Vmin)^2} \tag{87}$$

(b) $$P_i(X) = \frac{2((Vmax - Vmin) - 2T_i)}{(Vmax - Vmin)^2} \tag{88}$$

(c) $$P_i(X) = \frac{2((Vmax - Vmin) - 2T_i)}{(Vmax - Vmin)^2} - \frac{4(T_{i-1} - Vmin)}{(Vmax - Vmin)^2} \tag{89}$$

(d) $$P_i(X) = \frac{2((Vmax - Vmin) - 2T_i)}{(Vmax - Vmin)^2} - \frac{2((Vmax - Vmin) - 2T_{i-1})}{(Vmax - Vmin)^2} \tag{90}$$

Each of these probabilities assuming triangular and uniform distributions is calculated for both optimistic and pessimistic positions of classes. The output of the method for each alternative has the form shown in Figure 4.13.

		Uniform	Triangular
Optimistic	Class-1	$P_1(X_k)$	$P_1(X_k)$
	Class-2	$P_2(X_k)$	$P_2(X_k)$
	⋮	⋮	⋮
	Class-n	$P_n(X_k)$	$P_n(X_k)$
Pessimistic	Class-1	$P_1(X_k)$	$P_1(X_k)$
	Class-2	$P_2(X_k)$	$P_2(X_k)$
	⋮	⋮	⋮
	Class-n	$P_n(X_k)$	$P_n(X_k)$

Figure 4-13 Outputs of the method

4.6 Modified UTADIS

The developed solution approach is also adapted to classical UTADIS[3] method. Instead of using the criteria aggregation function that uses distances to the ideal point to determine the values of alternatives, the utility function is used in the method. The classical UTADIS method includes only one model that determines the utility thresholds and weights that give the minimum classification error. Then, these established values are used to determine the utility values of alternatives in test set. It is modified such that the alternative optimal solutions that give the minimum classification error are searched and the ones that give minimum and maximum threshold ($Tmax, Tmin$) and utility values of alternatives ($Umax,\ Umin$) are identified in the first and second

[3] For more information on UTADIS, see Chapter-3, Section-3.

model respectively. The probability calculation and evaluation phase is same as the original model so we will focus only on the mathematical models.

4.6.1 UTADIS Model-1

UTADIS estimates marginal utility functions for each criterion which are added to form global utility function of the decision maker. The global utility values of the alternatives are used to sort them into classes, which are separated by thresholds estimated in the model. In classical UTADIS, there is only one model which estimates thresholds and weights of criteria at the same time in order to minimize classification error of alternatives in preference set. In the modified UTADIS, weights are not established once and then used for prediction, but instead different weight sets are used to identify the maximum and minimum values of thresholds and alternatives. In the first model, the maximum and minimum values of thresholds are determined among the alternative optimal solutions that give minimum classification error. The linear model is shown below:

Additional Indexes

p_j = number of breakpoints on each criterion

$s \in \{1,2,\dots,p_j-1\}$ for intervals on each criterion

Additional Parameters

r_{jk} = subinterval that alternative k belongs on criterion j

a_j^t = breakpoint t on criterion j

$s, \delta_1, \delta_2, \partial$ = small positive constants (0.001; 0.0001; 0.0001;0.005)

Decision Variables

w_{jst} = utility value corresponding to interval s on criterion j

$umax_i$ = maximum value of threshold value separating class i and class i+1

$umin_i$ = minimum value of threshold value separating class i and class i+1

$\mathrm{\varepsilon}_{\mathrm{kt}}^{-}$ = error of assignment of alternative k to a higher class

$\mathrm{\varepsilon}_{\mathrm{kt}}^{+}$ = error of assignment of alternative k to a lower class

UTADIS Model-1

$$\min \quad \sum_{t=1}^{2}\sum_{i=1}^{n}\left[\frac{\sum_{\forall x_k \in C_i}(\varepsilon_{kt}^{+}+\varepsilon_{kt}^{-})}{m_i}\right] + \partial\sum_{i=1}^{n-1}(umin_i - umax_i) \tag{91}$$

s.t.

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jk}-1} w_{js1} + w_{j,r_{jk},1}\frac{a_{jk}-a_j^{r_{jk}}}{a_j^{r_{jk}-1}-a_j^{r_{jk}}}\right] - umax_1 + \varepsilon_{k1}^{+} \geq \delta_1 \qquad \forall x_k \in C_1 \tag{92}$$

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jk}-1} w_{js1} + w_{j,r_{jk},1}\frac{a_{jk}-a_j^{r_{jk}}}{a_j^{r_{jk}-1}-a_j^{r_{jk}}}\right] - umax_i + \varepsilon_{k1}^{+} \geq \delta_1$$

$$\forall x_k \in C_i \ (i = 2,3,\ldots,n-1) \tag{93}$$

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jk}-1} w_{js1} + w_{j,r_{jk},1}\frac{a_{jk}-a_j^{r_{jk}}}{a_j^{r_{jk}-1}-a_j^{r_{jk}}}\right] - umax_{i-1} - \varepsilon_{k1}^{-} \leq -\delta_2$$

$$\forall x_k \in C_i \ (i = 2,3,\ldots,n-1) \tag{94}$$

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jk}-1} w_{js1} + w_{j,r_{jk},1}\frac{a_{jk}-a_j^{r_{jk}}}{a_j^{r_{jk}-1}-a_j^{r_{jk}}}\right] - umax_{i-1} - \varepsilon_{k1}^{-} \geq -\delta_2 \qquad \forall x_k \in C_n \tag{95}$$

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jk}-1} w_{js2} + w_{j,r_{jk},2}\frac{a_{jk}-a_j^{r_{jk}}}{a_j^{r_{jk}-1}-a_j^{r_{jk}}}\right] - umin_1 + \varepsilon_{k2}^{+} \geq \delta_1 \qquad \forall x_k \in C_1 \tag{96}$$

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jk}-1} w_{js2} + w_{j,r_{jk},2}\frac{a_{jk}-a_j^{r_{jk}}}{a_j^{r_{jk}-1}-a_j^{r_{jk}}}\right] - umin_i + \varepsilon_{k2}^{+} \geq \delta_1$$

$$\forall x_k \in C_i \ (i = 2,3,\ldots,n-1) \tag{97}$$

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jk}-1} w_{js2} + w_{j,r_{jk},2}\frac{a_{jk}-a_j^{r_{jk}}}{a_j^{r_{jk}-1}-a_j^{r_{jk}}}\right] - umin_{i-1} - \varepsilon_{k2}^{-} \leq -\delta_2$$

$\forall\, x_k \in C_i \; (i = 2,3,\dots,n-1)$ (98)

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jk}-1} w_{js2} + w_{j,r_{jk},2}\frac{a_{jk}-a_j^{r_{jk}}}{a_j^{r_{jk}-1}-a_j^{r_{jk}}}\right] - umin_{i-1} - \varepsilon_{k2}^{-} \geq -\delta_2 \quad \forall\, x_k \in C_n \tag{99}$$

$$\sum_{j=1}^{q}\sum_{s=1}^{p_j-1} w_{jst} = 1 \qquad \forall t \in \{1,2\} \tag{100}$$

$$umax_{i-1} - umax_i \geq s \qquad \forall i \in \{2,3,\dots,n-1\} \tag{101}$$

$$umin_{i-1} - umin_i \geq s \qquad \forall i \in \{2,3,\dots,n-1\} \tag{102}$$

$$umax_i \geq umin_i \qquad \forall i \in \{1,2,3,\dots,n-1\} \tag{103}$$

$w_{jst} \geq 0\,, \varepsilon_{kt}^{+} \geq 0\,, \varepsilon_{kt}^{-} \geq 0 \qquad \forall j \in \{1,2,\dots,q\}, \forall s \in \{1,2,\dots,p_j-1\},$

$\forall t \in \{1,2\}, \forall k \in \{1,2,\dots,K\}$ (104)

4.6.2 UTADIS Model-2

In the second model, among the alternatives that give the minimum classification error found in model-1, the ones that result in maximum and minimum values of alternatives in test set are identified. The original UTADIS does not need a second model to find values of alternatives in the test set since the optimization is done once to establish the utility function and then, if no postoptimality analysis is done, this utility function is used to find values of alternatives of test set. In the modified version, we find these values again in an optimization model because alternative optimal solutions are searched with the secondary objective of minimizing $Umin$ and maximizing $Umax$ values. The second model of modified UTADIS is shown below:

Additional Indexes

L = number of alternatives in test set

$l \in \{1,2,\dots,L\}$ for alternatives in test set

Additional Parameters

a_{jl} = score of alternative l on criterion j

E^* = total classification error from model-1

$umin_i$ = minimum value of threshold i seperating class i and i+1

$umax_i$ = maximum value of threshold i seperating class i and i+1

$s,\ \delta_1, \delta_2$ = small positive constants (0.001; 0.0001; 0.0001)

Decision Variables

w_{jstl} = utility value corresponding to interval s on criterion j

$Vmax_l$ = maximum global utility value of alternative l

$Vmin_l$ = minimum global utility value of alternative l

ε_{ktl}^{-} = error of assignment of alternative i to a higher class

ε_{ktl}^{+} = error of assignment of alternative i to a higher class

u_{itl} = threshold value seperating class i and i+1

UTADIS Model-2

$$\min \quad \sum_{l=1}^{L} (Umin_l - Umax_l) \tag{105}$$

s.t.

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jk}-1} w_{js1l} + w_{j,r_{jk},l}\frac{a_{jk}-a_j^{r_{jk}}}{a_j^{r_{jk}-1}-a_j^{r_{jk}}}\right] - u_{1tl} + \varepsilon_{ktl}^{+} \geq \delta_1$$

$$\forall\, x_k \in C_1, \forall\, l \in \{1,2, \dots L\}, \forall t \in \{1,2\} \tag{106}$$

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jk}-1} w_{js1l} + w_{j,r_{jk},1,l}\frac{a_{jk}-a_j^{r_{jk}}}{a_j^{r_{jk}-1}-a_j^{r_{jk}}}\right] - u_{itl} + \varepsilon_{ktl}^{+} \geq \delta_1$$

$$\forall\, x_k \in C_i\ (i = 2,3, \dots, n-1), \forall\, l \in \{1,2, \dots L\}, \forall t \in \{1,2\} \tag{107}$$

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jk}-1} w_{js1l} + w_{j,r_{jk},1,l}\frac{a_{jk}-a_j^{r_{jk}}}{a_j^{r_{jk}-1}-a_j^{r_{jk}}}\right] - u_{i-1,tl} - \varepsilon_{ktl}^{-} \leq -\delta_2$$

$$\forall\, x_k \in C_i\ (i = 2,3,\dots,n-1), \forall\, l \in \{1,2,\dots L\}, \forall t \in \{1,2\} \tag{108}$$

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jk}-1} w_{js1l} + w_{j,r_{jk},1,l}\frac{a_{jk}-a_j^{r_{jk}}}{a_j^{r_{jk}-1}-a_j^{r_{jk}}}\right] - u_{i-1,tl} - \varepsilon_{ktl}^{-} \geq -\delta_2$$

$$\forall\, x_k \in C_n, \forall\, l \in \{1,2,\dots L\}, \forall t \in \{1,2\} \tag{109}$$

$$\sum_{j=1}^{q}\sum_{s=1}^{p_j-1} w_{jstl} = 1 \qquad \forall t \in \{1,2\},\ \forall\, l \in \{1,2,\dots L\} \tag{110}$$

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jl}-1} w_{js1l} + w_{j,r_{jl},1,l}\frac{a_{jl}-a_j^{r_{jl}}}{a_j^{r_{jl}-1}-a_j^{r_{jl}}}\right] \geq Umax_l \qquad \forall l \in \{1,2,3,\dots,L\} \tag{111}$$

$$\sum_{j=1}^{q}\left[\sum_{s=1}^{r_{jl}-1} w_{js2l} + w_{j,r_{jl},2,l}\frac{a_{jl}-a_j^{r_{jl}}}{a_j^{r_{jl}-1}-a_j^{r_{jl}}}\right] \leq Umin_l \qquad \forall l \in \{1,2,3,\dots,L\} \tag{112}$$

$$u_{itl} - ult_{i-1,tl} \geq s \qquad \forall i \in \{2,3,\dots,n-1\}, \forall\, l \in \{1,2,\dots L\}, \forall t \in \{1,2\} \tag{113}$$

$$umax_i \geq u_{itl} \qquad \forall i \in \{1,2,3,\dots,n-1\}, \forall\, l \in \{1,2,\dots L\}, \forall t \in \{1,2\} \tag{114}$$

$$umin_i \leq u_{itl} \qquad \forall i \in \{1,2,3,\dots,n-1\}, \forall\, l \in \{1,2,\dots L\}, \forall t \in \{1,2\} \tag{115}$$

$$\sum_{t=1}^{2}\sum_{l=1}^{L}\sum_{k=1}^{K}(\varepsilon_{ktl}^{+} + \varepsilon_{ktl}^{-}) \leq LE^{*} \tag{116}$$

$$w_{jstl} \geq 0, \varepsilon_{ktl}^{+} \geq 0, \varepsilon_{ktl}^{-} \geq 0 \qquad \forall j \in \{1,2,\dots,q\}, \forall s \in \{1,2,\dots,p_j-1\},$$

$$\forall t \in \{1,2\}, \forall k \in \{1,2,\dots,K\}, \forall\, l \in \{1,2,\dots L\} \tag{117}$$

4.6.3 Heuristics for Determining Subintervals on Each Criterion

When the criterion is categorical, the breakpoints in the criterion range are determined as these categories. For instance, if there are 3 categories on a criterion such as low (1), medium (2) and high (3); the breakpoints on that criterion are taken as 1,2 and 3. But, if the criterion is continuous, determining the subintervals is not that straightforward. There are 2 heuristics proposed for this problem by Doumpos and Zopounidis (2004). The empirical study of Doumpos and Zopounidis (2001) shows that the second heuristic performs better than the first heuristic in terms of increasing stability of the model and classification performance. So, we implement second heuristic (HEUR2) for determining the breakpoints in this study. HEUR2 considers the distribution of alternatives, which belongs to different groups, on each criterion scale. HEUR2 has 5 steps to determine the breakpoints as shown below:

Step 1. Rank the alternatives in preference set for each criterion, according to their score a_{kj} on that criterion from the least preferred one to the most. Set the minimum acceptable number of alternatives (β) belonging to a subinterval equal to zero.

Step 2. Form all non-overlapping subintervals $\left[a_j^s, a_j^{s+1}\right]$ where the alternative with score equal to a_j^s and the alternative with score equal to a_j^{s+1} belong to different groups.

Step 3. Check the number of alternatives that lie on each subinterval formed after step-2. If the number of alternatives in a subinterval is less than β, then merge this interval with the precedent one. (Skip this when β=0)

Step 4. Compare the number of subintervals in each criterion to the number of constraints in the LP model. If the number of subintervals leads to specification of more than $m_1 + 2\sum_{i=2}^{n-1} m_i + m_n$ variables (w), where m_i is the number of constraints for class i, then set $\beta = \beta + 1$ and go to step-3. Otherwise, stop.

4.7 Evaluation of Results

The probabilities provided to the decision maker can be interpreted differently by different decision makers with various world views. Since the probabilities are given for optimistic and pessimistic cases, there can be several combinations of these probabilities and the evaluation of them depends on the decision maker. For instance, three possible probabilities of belonging to class i and j for an alternative in a two class case are shown in Figure 4.14.

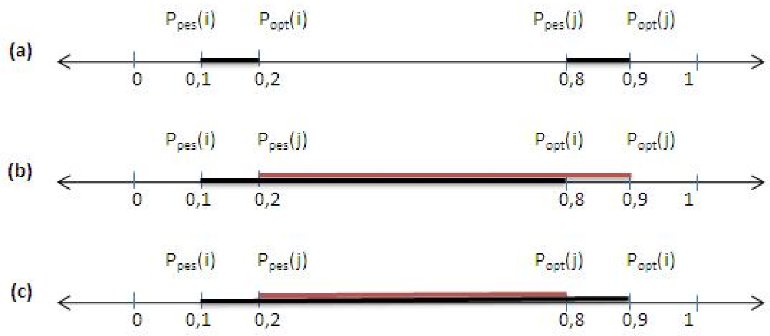


Figure 4-14 Three different combinations of pessimistic and optimistic probabilities of an alternative for class i and j

The situation in Figure 4.14 (a) can be evaluated easily since both the pessimistic and optimistic probabilities for class-j are higher than class-i. So, assigning that alternative to class-j is obvious. In situation (b), the difference is less clear since the probabilities are quite similar. Still, in each of optimistic and pessimistic cases, class-j is higher. The evaluation of last situation (c) depends heavily on the decision maker. Although the optimistic probability of class-i is higher, class-j is more probable if we look at the pessimistic probabilities. So, the decision changes for a risk-averse and risk-seeking person. Risk-averse decision makers would rely on the pessimistic probabilities more and prefer maximizing the minimum probability and choose class-j, while risk-seeking decision makers would rely on optimistic probabilities and prefer maximizing the maximum probability by choosing class-j. The accuracy calculations for this study are done by the classes that would be chosen by a risk-seeking decision maker considering mostly the optimistic probabilities.

CHAPTER 5

COMPUTATIONAL EXPERIMENTS

5.1 Data Sets

The developed method is implemented on 5 data sets which are retrieved from UCI Machine Learning Repository, a study of Hipel and Kilgour (2005) and a study of Fernandez et al (2009). The data sets are chosen such that all the criteria and classes are ordinal so the problem is a sorting problem. Some of the criteria are categorical and some are continuous. The categorical data is transformed into quantitative by assigning numbers to each category. Each of the data set is separated into two groups as training (preference) and test sets. 65% of the data sets are taken as training and 35% as the test data. The general information about the data sets is given in Table 5.1.

Table 5-1 Data sets used in the computational experiments

Data set	Number of data points	Number of sttributes	Classes	Number of alternatives in each class
Water Supplies	19	7	1	10
			2	9
R&D Projects	81	4	1	6
			2	28
			3	27
			4	4
			5	10
			6	3
			7	1
			8	2

Table 5-1 Continued

Assistant	151	3	1	49
			2	50
			3	52
Cars	220	6	1	14
			2	10
			3	44
			4	152
Credit	150	20	1	83
			2	67

5.1.1 Water Supplies Data Set

The water usage data set is retrieved from the study of Hipel and Kilgour (2005). It is a relatively small data set with 19 data points. The data set includes alternatives for best water resources. The alternatives are sorted into 2 classes such as:

- Acceptable – Class-1
- Unacceptable – Class-2

There are 7 criteria to evaluate the alternatives, which are shown in Table 5.2.

Table 5-2 Criteria for the water supplies data set with their ranges and types

Criteria	**Range**	**Type**
Project investment cost	Millions of dollars	Lower the better
Project operating cost	Millions of dollars	Lower the better
Project negative infrastructure impact	0-100	Lower the better
Project negative environmental impact	0-100	Lower the better
Project implementation risk	0-100	Lower the better
Project supply capability	Million imperial gallons per day	Higher the better
Quality of water the project will deliver	0-100	Higher the better

All of the criteria are continuous and the value ranges are not compatible as seen in the table. So, normalization is done first to make the ranges of each criterion equivalent and prevent the dominance of one criterion. The normalization is applied by dividing each score on a criterion by the range on that criterion and therefore equating the ranges to (0,1) for each criterion. The function is as shown below:

$$a'_{kj} = \frac{a_{kj}}{\max_k(a_{kj}) - \min_k(a_{kj})}$$

where a_{kj} and a'_{kj} are the original and standardized values of alternative k on criterion j.

5.1.2 R&D Projects Data Set

The R&D projects data set is retrieved from the study of Fernandez et al (2009). The data includes 81 alternatives, which are the R&D projects to be evaluated on 4 criteria. The alternatives are classified into 8 classes:

- Exceptional – Class-1
- Very high – Class-2
- High – Class-3
- Above average – Class-4
- Average – Class-5
- Below average – Class-6
- Low – Class-7
- Very low – Class-8

The criteria are shown in Table 5.3.

Table 5-3 Criteria for the R&D projects data set with their ranges and types

Criteria	Range	Type
Economic outcomes	1-7	Higher the better
Social outcomes	1-7	Higher the better
Scientific outcomes	1-7	Higher the better
Improvement of research competence	1-7	Higher the better

Since all the criteria are categorical and defined on equal ranges, no standardization is applied to this data set. The interesting point of the data set is that some of the alternatives are classified in 2 classes such as "exceptional or very high" which is well handled by the proposed method.

5.1.3 Assistant Data Set

This data set is retrieved from the UCI Machine Learning Repository. There are 151 data points in the data set. The alternatives are assistants to be evaluated and sorted into 3 classes:

- High – Class-1
- Medium – Class-2
- Low – Class-3

There are 3 criteria that are shown in Table 5.4.

Table 5-4 Criteria for the assistant data set with their ranges and types

Criteria	Range	Type
Native English speaker	Native (1), Non-Native (2)	Lower the better
Semester of teaching	Regular (1) , Summer (2)	Lower the better
Class size	Number of students registered	Higher the better

The last criterion "class size" is standardized into the range (0,1) by dividing the values to the range of that criterion similar to the previous data sets.

5.1.4 Car Data Set

This data set is retrieved from the UCI Machine Learning Repository. There are 220 data points in the data set. The alternatives are cars to be evaluated and sorted into 4 classes such as:

- Very good – Class-1
- Good – Class-2
- Acceptable – Class-3
- Unacceptable – Class-4

The alternatives are sorted according to 6 criteria as shown in Table 5.5.

Table 5-5 Criteria for the car data set with their ranges and types

Criteria	Range	Type
Price	Very high (1), High (2), Medium (3), Low (4)	Higher the better
Maintenance cost	Very high (1), High (2), Medium (3), Low (4)	Higher the better
Number of doors	2 doors (1), 3 doors (2), 4 doors (3), More than 4 doors (4)	Higher the better

Table 5-5 Continued

Number of person that can be carried	2 persons (1), 4 persons (2), More than 4 persons (3)	Higher the better
Luggage boot capacity	Small (1), Medium (2), Big (3)	Higher the better
Safety	Low (1), Medium (2), High (3)	Higher the better

Since the criteria are categorical, they are quantified by assigning numbers to each category as shown in Table 5.4.

5.1.5 Credit Data Set

The credit data set is retrieved from UCI Machine Learning Repository. The data set includes 150 data points. The alternatives are credit applicants which are sorted into 2 classes such as:

- Approved – Class-1
- Not approved – Class-2

The alternatives are sorted according to 20 criteria as shown in Table 5.6.

Table 5-6 Criteria for the credit data set with their ranges and types

Criteria	Range	Type
Status of existing check account	No check account (1) Account with no money (2) Account with less than $200 (3) Account with more than $200 (4)	Higher the better

Table 5-6 Continued

Duration of the credit application	Months	Higher the better
Credit history of the applicant	Critical account (1), Delay in paying off in the past (2), Paid back existing credits duly till now (3), Paid back all credits at this bank duly (4) Paid back all credits duly or taken no credits till now (5)	Higher the better
Purpose of the credit application	New car (1) Used car (2) Furniture or equipment (3) Radio or television (4) Domestic appliances (5) Repair (6) Education (7) Vacation (8) Retraining (9) Business (10)	Higher the better
Amount of the credit application	Dollars	Higher the better
Amount of saving accounts of the applicant	No information or no account (1) Account is less than \$100 (2) Account is between \$100 and \$500 (3) Account is between \$500 and \$1000 (4) Account is above \$1000 (5)	Higher the better

Table 5-6 Continued

Employment status of the applicant	Unemployed (1) Employed less than 1 year (2) Employed more than 1, less than 4 years (3) Employed more than 4, less than 7 years (4) Employed more than 7 years (5)	Higher the better
Installment rate in percentage of disposable income	0-100	Higher the better
Marital status and sex of the applicant	Male and divorced (1) Female and divorced (2) Male and single (3) Male and married or widowed (4) Female and single (5)	Higher the better
Whether or nor there exists other debtors and guarantors	None (1) Co-applicant (2) Guarantor (3)	Higher the better
Duration of the residence of the applicant	Years	Higher the better
Properties belong to the applicant	No information or no property (1) Car (2) Building society savings agreement or life insurance (3) Real estate (4)	Higher the better
Age of the applicant	Years	Higher the better

Table 5-6 Continued

Whether or not there exists other installment plans of the applicant	Installment plans to bank (1) Installment plans to stores (2) No installment plans (3)	Higher the better
Housing information of the applicant	Rent (1) Owns the house (2) House is for free (3)	Higher the better
Number of existing credits of the applicant on this bank	0 - ~	Higher the better
Job of the applicant	Unemployed or unskilled-non-resident (1) Unskilled-resident (2) Skilled employee or official (3) Manager or self-employed (4)	Higher the better
Number of people liable to provide maintenance for the applicant	1 person (1) 2 persons (2)	Higher the better
Whether or not the applicant has telephone	Does not own telephone (1) Owns a telephone (2)	Higher the better
Whether or not the applicant is a foreign worker	Foreign worker (1) Not a foreign worker (2)	Higher the better

All the criteria which are not categorical are standardized as shown before into the range (0,1).

5.2 Performance Measures

The most commonly used performance measure for classification/sorting methods is the *accuracy* of the prediction of the method for the alternatives in test set. Accuracy is the proportion of the correctly classified alternatives in the test set. Table 5.7 shows the number of correctly and incorrectly classified alternatives for 2-class problem.

Table 5-7 Number of incorrectly classified alternatives for 2-class case

	Predicted Class		
Actual Class		**Class=1**	**Class=2**
	Class=1	**X**	**Y**
	Class=2	**Z**	**T**

The accuracy of the case shown in the table is calculated as:

$$Accuracy = \frac{X + T}{X + Y + Z + T}$$

Accuracy is one of the performance measures used in this study. Yet, since the output of the proposed method is not similar to the usual classification/sorting methods such as a single class for one alternative, accuracy alone is not enough to see the actual performance of the method. The other performance measures used in this study are *covering* and $accuracy^2$. Covering is used to identify whether the correct class has a positive probability for an alternative although it is not the class with highest probability. It is useful when the output for an alternative consists of similar probabilities for different classes such as:

$$P_1(x_k) = 0.55$$
$$P_2(x_k) = 0.50$$

where the correct class of alternative x_k is class-2. The accuracy measure does not consider the 0,45 probability of belonging to correct class yet the existence of that probability is important since a decision maker may decide on that class based on that probability. So, covering counts the alternatives similar to this one and takes the proportion of these alternatives as a performance measure. If there are K alternatives which has positive probabilities in their correct classes (highest or not) among N alternatives, then covering is calculated as:

$$Covering = \frac{K}{N}$$

Accuracy² calculates the accuracy considering not only the class with highest probability but also the class with second highest probability where the difference between these probabilities is less than 0.2 as shown below:

$$\textbf{(a)} \quad \left.\begin{array}{l} P_1(x_k) = 0.90 \\ P_2(x_k) = 0.85 \end{array}\right\} \text{Consider class-2 for accuracy}^2$$

$$\textbf{(b)} \quad \left.\begin{array}{l} P_1(x_k) = 0.90 \\ P_2(x_k) = 0.40 \end{array}\right\} \text{Do not consider class-2 for accuracy}^2$$

Table 5.8 shows the number of alternatives that are correctly classified and among the incorrectly classified alternatives, the ones with positive probabilities.

Table 5-8 Number of alternatives according to the classes with first and second highest probability for 2-class case

	Predicted Class				
Actual Class		**Class¹=1**		**Class¹=2**	
		Class²=None	**Class²=2**	**Class²=None**	**Class²=1**
	Class=1	X_1	X_2	Y_1	Y_2
	Class=2	Z_1	Z_2	T_1	T_2

Where $Class^1$ and $Class^2$ are the classes with highest and second highest positive probabilities where the difference between probabilities is less than 0.2. In this case, the accuracy and $accuracy^2$ calculations are done as follows:

$$Accuracy = \frac{(X_1 + X_2) + (T_1 + T_2)}{\sum_{i=1}^{2}(X_i + Y_i + T_i + Z_i)}$$

$$Accuracy^2 = \frac{(X_1 + X_2) + (T_1 + T_2) + Y_2 + Z_2}{\sum_{i=1}^{2}(X_i + Y_i + T_i + Z_i)}$$

Other than these 3 measures that consider the correctness of the predicted classes, there is another performance measure considering the ability of the method to predict all classes. This means the thresholds determined by the model result in intervals for each class. This is a necessary performance measure because the method can result in missing classes where the thresholds below and above the class are equal, which makes the interval of that class disappear. This may result when the number of alternatives in one class is significantly less than the other classes so the relative importance of that class for that data set is very small. So increasing the interval of other classes result in smaller classification error and the class with fewer alternatives is closed. This is not a desired situation because the method should be able to identify the ranges for all classes. The performance measure that considers this situation is *missing # of classes*. Missing number of classes counts the classes that do not have an interval and consequently, there are no positive probabilities for that class in the outputs of the model. The first 3 performance measures are desired to be as high as possible whereas in this performance measure, the less is the better.

5.3 Results

The method is applied to each of the five data sets for different distance norms and the results are compared in order to identify the best-suited distance norm.

For the L_p norms different than L_1 and L_∞, the mathematical models are nonlinear. Also, for computational concerns, the squared Euclidean distance is also used to compare its performance with Euclidean distance. GAMS compiler is used for the execution of the models. The solver is specified as MINOSD for nonlinear models and CPLEX for the linear models. Since the distance function for each norm is convex, optimality is guaranteed for nonlinear models. The results of proposed method and modified UTADIS are given for each data set with the results of performance measures previously defined, computation times and classification error. In each case, the uniform and triangular distribution assumptions resulted in the same probabilities so the results are not shown for each distribution separately.

5.3.1 Results of Distance-Based Method

In this section, the results of the proposed method with the criteria disaggregation function based on distances are shown. For water supplies data set, the results are provided in Table 5.9.

Table 5-9 Results of water supplies set

Water Supplies Data Set						
Distance norm	**Accuracy**	**Accuracy^2**	**Covering**	**# of missing classes**	**Classification Error**	**Computation Time**
L_1	**0.67**	**1.00**	**1.00**	**-**	**0**	**0.42 sec**
L_2	0.5	0.5	0.33	1	0	0.42 sec
$squared\ L_2$	0.58	0.7	1.00	-	0	0.45 sec
L_3	0.5	0.5	0.5	1	0	0.74 sec
L_4	0.5	0.5	0.5	1	0	0.46 sec
L_5	0.29	0.29	0.29	1	0	0.87 sec
L_{10}	0.5	0.5	0.5	1	0	0.43 sec
L_{20}	0.5	0.5	0.5	-	0	0.84 sec
L_{30}	0.33	1	1	-	0	2.67 sec
L_∞	0.67	0.67	0.67	-	0	0.50 sec

The best performing distance norm for water supplies data set is L_1 norm with highest accuracy. The L_∞ norm also results in same level accuracy. but L_1 norm has higher values on covering and accuracy^2 performance measures. The classification error is zero for all distance norms and the computation times are very low due to the small size of the data set.

The outputs of the method with L_1 norm for water supplies data set are provided in Appendix-A.

The results of R&D projects data set are provided in Table 5.10.

Table 5-10 Results of R&D projects data set

R&D Projects Data Set						
Distance norm	**Accuracy**	**Accuracy^2**	**Covering**	**# of missing classes**	**Classification Error**	**Computation Time**
L_1	**0.83**	**0.83**	**0.83**	**-**	**0.22**	**1.35 sec**
L_2	0.69	0.69	0.69	1	0.33	18.32 sec
$squared\ L_2$	0.69	0.69	0.69	1	6.31	1.42 sec
L_3	0.55	0.55	0.55	1	1.99	22.67 sec
L_4	0.55	0.55	0.55	1	0.4	17.92 sec
L_5	0.5	1	1	-	0.44	39.14 sec
L_{10}	0.57	0.57	0.57	1	0.54	48.59 sec
L_{20}	0.36	0.36	0.36	2	0.57	37.34 sec
L_{30}	0.31	0.31	0.31	2	0.57	30.53 sec
L_∞	0.26	0.26	0.26	4	0.88	8.95 sec

The best compromise of results is seen in L_1 distance norm for R&D projects data set with 0.83 accuracy, no missing classes and a small computation time due to linear model. The classification error is also at minimum for L_1 distance norm.

The outputs of the method with L_1 norm for R&D projects data set are provided in Appendix-B.

The results of assistant data set are given in Table 5.11.

Table 5-11 Results of assistant data set

Assistant Data Set						
Distance norm	**Accuracy**	**Accuracy^2**	**Covering**	**# of missing classes**	**Classification Error**	**Computation Time**
L_1	0.32	0.32	0.32	2	3	0.95 sec
L_2	0.44	0.44	0.44	-	2.24	2 min 11.10 sec
$squared\ L_2$	0.38	0.38	0.38	2	3	1.18 sec
L_3	**0.48**	**0.48**	**0.48**	**-**	**2.01**	**2 min 1.31 sec**
L_4	0.46	0.46	0.46	-	1.89	1 min 54.89 sec
L_5	0.46	0.46	0.46	-	1.84	2 min 22.90 sec
L_{10}	0.46	0.46	0.46	-	1.74	1 min 7.17 sec
L_{20}	0.46	0.46	0.46	-	1.7	1 min 28.57 sec
L_{30}	0.46	0.46	0.46	-	1.77	1 min 23.90 sec
L_∞	0.32	0.32	0.32	-	2.5	1.54 sec

For assistant data set, all distance norms perform poorly compared to the other data sets. Still, the best results of performance measures are seen in L_3 distance norm. The classification errors are positive and higher compared to other data sets for all distance norms.

The outputs of the method with L_3 norm for assistant data set are provided in Appendix-C.

The results of the cars data set are shown in Table 5.12.

Table 5-12 Results of the cars data set

Cars Data Set						
Distance norm	**Accuracy**	**Accuracy²**	**Covering**	**# of missing classes**	**Classification Error**	**Computation Time**
L_1	0.84	0.84	0.84	-	2.4	34.28 sec
L_2	**0.88**	**0.88**	**0.88**	**-**	**0.47**	**9 min 9.21 sec**
$squared\ L_2$	0.88	0.88	0.88	1	1.73	34 sec
L_3	0.87	0.87	0.87	-	0.28	14 min 34.60 sec
L_4	0.88	0.88	0.88	1	0.25	24 min 9.42 sec
L_5	0.87	0.87	0.87	-	0.24	11 min 8.10 sec
L_{10}	0.83	0.83	0.83	-	0.4	29 min 43.59 sec
L_{20}	0.83	0.83	0.83	1	0.51	16 hrs 57 min
L_{30}	0.83	0.83	0.83	1	0.57	20 hrs 32 min
L_∞	0.85	0.85	0.85	1	0.78	53.70 sec

For cars data set, all distance norms perform quite well with high accuracy levels and less missing classes. Still, the best results are seen in L_2 distance norm. The computational times are higher compared to other data sets since the size of the data is higher. For L_{20} and L_{30}, the computational times are higher compared to other distance norms and the nonlinear models did not result in optimal solutions for this data set.

The outputs of the method with L_2 norm for cars data set are provided in Appendix-D.

The results of credit data set are given in Table 5.13.

Table 5-13 Results of credit data set

Credit Data Set						
Distance norm	**Accuracy**	**Accuracy^2**	**Covering**	**# of missing classes**	**Classification Error**	**Computation Time**
L_1	0.4	0.4	0.4	-	0	1.34 sec
L_2	0.6	0.6	0.6	-	0	6 min 19.32 sec
$squared\ L_2$	0.36	0.36	0.36	1	0	1.43 sec
L_3	**0.72**	**0.74**	**0.84**	**-**	**0**	**21 min 17.26 sec**
L_4	0.58	0.58	0.58	-	0	11 min 43.90 sec
L_5	0.58	0.58	0.58	-	0	17 min 23.06 sec
L_{10}	0.42	0.42	0.42	-	0	7 min 21.53 sec
L_{20}	0.52	0.52	0.52	-	0	6 min 42.23 sec
L_{30}	0.44	0.44	0.44	-	0	57 min 59.06 sec
L_∞	0.36	0.36	0.36	1	0.05	10.06 sec

The best performing distance norm for credit data set is L_3 with 0.72 accuracy, 0.84 covering and no missing classes. The computational time is quite greater than other norms yet the difference in other performance measures is more significant than this drawback. The classification errors are 0 or slightly different than 0 for all distance norms.

The outputs of the method with L_3 norm for credit data set are provided in Appendix-E.

The results of the distance based method for different distance norms show that although there isn't a pattern related to the increase or decrease in p-norm for any data set, the best results occur in smaller values of p. The p values greater than 10 require a high computational effort for non-linear models and the results may not be the optimal solutions. Still, we see that it is not right to fix the distance-based method to only one distance norm, since different distance norms may fit a data set better and result in better performance as in the case of five data sets.

The highest probability classes for each data set and their actual classes are given in Appendix-F for the best performing L_p norms.

5.3.2 Results of Modified UTADIS

Modified UTADIS is applied to each of the five data sets and the results are provided in this section. The HEUR2 defined in Chapter-4 is applied first in order to determine the breakpoints on each criterion. The results for all data sets are given in Table 5.14.

Table 5-14 Results of modified UTADIS for each data set

Modified UTADIS Results						
Data Set	**Accuracy**	**Accuracy^2**	**Covering**	**# of missing classes**	**Classification Error**	**Computation Time**
Water Supplies	0.5	0.5	0.5	1	0	0.73 sec
R&D Projects	0.9	0.92	1	-	0	0.82 sec
Assistant	0.46	0.46	0.46	2	0.17	0.78 sec
Cars	0.91	1	1	-	7.27	0.89 sec
Credit	0.64	0.64	0.64	1	0	0.87 sec

It is seen that modified UTADIS performs quite well for the R&D projects and cars data set with accuracies greater than 0.90. Also, the computation time required is very low (less than a second) since the model is linear. The missing classes occur in this method too. For assistant, water supplies and credit data sets, there is only one class open. The comparison of modified UTADIS with the proposed method and the previous methods is provided in the next section.

The highest probability classes for each data set and their actual classes are given in Appendix-G for the modified UTADIS method.

5.4 Comparison with Previous Methods

The classification tree and the classical UTADIS are applied to the data sets and the results are compared with the proposed method and modified UTADIS. The results of the best performing L_p norms are selected for the proposed distance based method in each data set and compared with the other methods. The results of modified UTADIS are compared with UTADIS to see whether the modification provides any improvement. For the classification tree, XLMiner program is used to obtain the results. The classification error for classification trees are given as percentage of misclassified alternatives rather than total misclassified distance as in the other methods. The breakpoints of classical UTADIS are same with the modified UTADIS, which are found by HEUR2 previously mentioned. The accuracy^2 and covering performance measures do not take any value for UTADIS and classification tree methods since they do not provide outcomes as probabilities.

Table 5.15 shows the results of each method for water supplies data set.

Table 5-15 Results of each method for water supplies data set

Water Supplies Data Set						
Method	**Accuracy**	**Accuracy^2**	**Covering**	**# of missing classes**	**Classification Error**	**Computation Time**
Proposed method (L_1)	**0.67**	**1.00**	**1.00**	**-**	**0**	**0.42 sec**
Classification Tree	0.5	-	-	1	0	1.18 sec
UTADIS	0.5	-	-	1	0	0.32 sec
Modified UTADIS	0.5	0.5	0.5	1	0	0.73 sec

For the water supplies data set, it is seen that the proposed method performs best among all the other sorting methods. There are no missing classes and the covering and accuracy^2 levels are at 100%. The computation time is insignificant for each method since the data set is relatively small. Also, it is seen that modified UTADIS performs same as UTADIS for this data set.

The results of R&D projects for each method are provided in Table 5.16.

Table 5-16 Results of each method for R&D projects data set

R&D Projects Data Set						
Method	**Accuracy**	**Accuracy^2**	**Covering**	**# of missing classes**	**Classification Error**	**Computation Time**
Proposed method (L_1)	0.83	0.83	0.83	-	0.22	1.35 sec
Classification Tree	0.42	-	-	-	0.40	2.03 sec
UTADIS	0.6	-	-	-	12.1	0.281 sec
Modified UTADIS	**0.9**	**0.92**	**1**	**-**	**0**	**0.828 sec**

The modified UTADIS method performs better than other methods for the R&D projects data set. It is seen that modification improved the performance of UTADIS for this data set. The proposed method also performs well with 0.83 accuracy. The computation times are again low but the classification error is positive for all methods except modified UTADIS. Yet, although it is not seen in the table, modified UTADIS has a drawback for this data set. The case where highest probabilities are equal for more than one class occurs more frequently in modified UTADIS compared to other methods. In that case, if the correct class is one of the highest probabilities, it is taken as a correct classification and increase the accuracy level. Yet, these cases are not as informative as the other results for the decision maker.

Table 5.17 shows the results of each method for assistant data set.

Table 5-17 Results of each method for assistant data set

Assistant Data Set						
Method	**Accuracy**	**Accuracy^2**	**Covering**	**# of missing classes**	**Classification Error**	**Computation Time**
Proposed method (L_3)	**0.48**	**0.48**	**0.48**	**-**	**2.01**	**2 min 1.31 sec**
Classification Tree	0.42	-	-	-	0.44	2.3 sec
UTADIS	0.46	-	-	1	0.24	0.34 sec
Modified UTADIS	0.46	0.46	0.46	2	0.17	0.78 sec

For the assistant data set the proposed method performs better than the other approaches. Still, it is seen that the performance of each sorting method is lower compared to other data sets. The computation times are relatively low and the classification errors are positive for each method. The modified UTADIS performs similarly as the UTADIS with same accuracy but more missing classes.

Table 5.18 shows the results of each method for cars data set.

Table 5-18 Results of each method for cars data set

Cars Data Set						
Method	**Accuracy**	**Accuracy^2**	**Covering**	**# of missing classes**	**Classification Error**	**Computation Time**
Proposed method (L_2)	0.88	0.88	0.88	-	0.47	9 min 9.21 sec
Classification Tree	0.77	-	-	1	0.21	2.71 sec
UTADIS	0.67	-	-	1	7.27	0.28 sec
Modified UTADIS	**0.91**	**1**	**1**	**-**	**7.27**	**0.89 sec**

Modified UTADIS method performs better than other methods for cars data set as seen in the Table 5.18. Yet, the drawback mentioned before is valid for this data set too with a relatively high number of equal probability cases. So, the proposed method is also a good solution approach for this data set.

Table 5.19 shows the results of each solution approach for credit data set.

Table 5-19 Results of each method for credit data set

Credit Data Set						
Method	**Accuracy**	**Accuracy^2**	**Covering**	**# of missing classes**	**Classification Error**	**Computation Time**
Proposed method (L_3)	**0.72**	**0.74**	**0.84**	**-**	**0**	**21 min 17.26 sec**
Classification Tree	0.64	-	-	-	0.23	3.14 sec
UTADIS	0.64	-	-	1	0	0.51 sec
Modified UTADIS	0.64	0.64	0.64	1	0	0.87 sec

For the credit data set, the best performing solution approach is proposed method as seen in Table 5.19. There are no missing classes and the classification error is zero. The computation time is relatively high since L_3 model is nonlinear and the data set is not as small as the other data sets. Again, it is seen that modified UTADIS performs same as the UTADIS.

It is seen that proposed distance-based method performs better than other methods in most of the 5 data sets in terms of the defined performance measures. Modified UTADIS results in slightly better values for some data sets yet its results include a relatively high percentage of equal probability cases. Still, it performs at least as good as the classical UTADIS for each data set which shows that it is a promising modification for UTADIS. Classification tree performs worse than the other methods in all cases yet it is useful to see its results since it provides an idea about the data set and how informative it is.

For the proposed method, the computation times are lower for the L_p norms that perform best. Yet, the nonlinearity of the models for distance norms other than L_1 and L_∞ is a drawback. The computational effort would increase exponentially for larger data sets and greater L_p norms.

CHAPTER 6

CONCLUSION

A new solution method for sorting problems is developed in this study. It is a PDA approach including a distance function based method as criteria aggregation function and instead of identifying only one class for each alternative, it provides probabilities of belonging to each class for an alternative. The decision maker can evaluate these probabilities and decide on the assigned class. So, the method provides a second opportunity to include decision maker's preferences in the sorting process. The method is given in general distance norm and in the computational experiments, it is seen that different distance norms may fit different data sets better so defining the method with a general distance norm is more promising. The probabilistic approach is also applied to UTADIS in order to handle the alternative optimal solutions.

The results of the distance function based method are compared with the results of modified UTADIS and previous methods such as UTADIS and classification trees. Distance function based method performs better than the previous methods for the five data sets. It is seen that modified UTADIS always performs at least as good as classical UTADIS for each of the five data set and for some data sets, it performs better than distance function based method. Computational effort is a challenging issue especially for the massive data sets since the model becomes nonlinear for certain distance norms. It is also seen that as the distance norm gets larger, the computation time required increases and usually performance of the method decreases for larger distance norms such as L_{10}, L_{20} and so on.

A possible research direction on this subject can be application of combined distance norms to the distance based criteria aggregation function. The combined distance norms consist of distance norms for both continuous and categorical data, which eliminates the need for quantifying the categorical data. Another possible research topic may be application of probabilistic approach to mathematical programming based discriminant analysis for classification problems. Since the alternative optimal solutions issue is also valid for classification problems, calculating probabilities by defining maximum and minimum values may improve the performance of discriminant analysis as in the UTADIS case. . Also, an additional study to one of these proposed research topics may be calculation of the coefficient of secondary objectives by a theoretical approach instead of empirical evaluation. The comparisons of the developed approach with the previous sorting methods are based on the five data sets. A theoretical comparison of the methods in order to identify the superior properties of each method may be another future research direction

REFERENCES

Bennet, K.P., Mangasarian, O.L. (1994). Multicategory discrimination via linear programming. *Optimization Methods and Software*, 3, pp. 27–39.

Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American Statistical Association*, 39, pp. 357–365.

Błaszczynski, J., Greco, S., Słowinski, R. (2007). Multi-criteria classification – A new scheme for application of dominance-based decision rules European Journal of Operational Research 181 (2007) 1030–1044.

Bliss, C.I. (1934). The method of probits. *Science* 79, pp. 38–39.

Chen, Y., Hipel, K.H. and Kilgour, D.M. (2005). Case-based distance models with and application in water resources management. IEEE International Conference, 1, pp. 215-220.

Chen, Y., Kilgour, D.M. and Hipel, K.W. (2008). A case-based distance method for screening in multiple-criteria decision aid. *Omega,* 36, pp. 373–383.

Devaud, J.M., Groussaud, G. and Jacquet-Lagrèze, E. (1980). UTADIS: Une méthode de construction de fonctions d'utilité additives rendant compte de jugements globaux. *European Working Group on Multicriteria Decision Aid*, Bochum.

Doumpos, M. and Zopounidis, C. (1998). The use of the preference disaggregation analysis in the assessment of financial risks. *Fuzzy Economic Review*, 3/1, pp. 39-57.

Doumpos, M. and Zopounidis, C. (2001). Developing sorting models using preference disag- gregation analysis: An experimental investigation. *Fuzzy Sets in Management, Economy and Marketing*, *World Scientific*, Singapore, pp. 51-67.

Doumpos, M., Zopounidis, C. (2004). Developing sorting models using preference disaggregation analysis: An experimental investigation. *European Journal of Operational Research*, 154 (3), pp. 585–598.

Fernandez, E., Navarro, J. and Bernal, S. (2009). Multi-criteria sorting using a valued indifference relation under a preference disaggregation paradigm. *European Journal of Operational Research*, 198, pp. 602–609.

Fisher, R.A. (1936). The use of multiple measurements in taxonomy problems. *Annals of Eugenics,* 7, pp. 179–188.

Fielding, A.H., Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation

Freed, N. and Glover, F. (1986). Evaluating alternative linear programming models to solve the two-group discriminant problem. *Decision Sciences,* 17, pp. 151-62.

Freed, N., and Glover, F., (198l). Simple but powerful goal programming models for discriminant problems. *European Journal of Operational Research*, 7, pp. 44-60.

Jacquet-Lagrèze, E., Siskos, Y. (1982). Assessing a set of additive utility functions for multicriteria decision making: The UTA method. *European Journal of Operational Research*, 10, pp. 151–164.

Joachimsthaler, E. A. and Stam, A. (1988). Four approaches to the classification problem in discriminant analysis: an experimental study. *Decision Sciences*, 19, pp. 322-33.

Keeney, R., Raiffa, H. (1993). Decisions with Multiple Objectives: Preferences and Value Trade-offs. *Cambridge University Press*, Cambridge.

Koehler, G. J. (1989). Discriminant functions using a genetic algorithm. Working paper, Department of Decision and Information Sciences, University of Florida.

Koehler, G.J., Erenguc, S.S. (1990). Minimizing misclassifications in linear discriminant analysis. *Decision Sciences*, 21, pp. 63–85.

Koksalan, M., Ozpeynirci, S.B. (2009). An interactive sorting method for additive utility functions. *Computers & Operations Research*, 36, 2565—2572.

Larnttand, K.F., Moy, J.M. (1997). An experimental comparison of some recently developed linear programming approaches to the discriminant problem. *Computers Ops Res.,* 24 (7), pp. 593—599.

Markowski, C.A., Markowski, E.P. (1987). An experimental comparison of several approaches to the discriminant problem with both qualitative and quantitative variables. *European Journal of Operational Research,* 28, pp. 74–78.

Opricovic, S. (1998). Multicriteria Optimization of Civil Engineering Systems. *Faculty of Civil Engineering*, Belgrade.

Quinlan, J. (1993). C4.5: Programs for Machine Learning. *Morgan Kaufmann Publishers*, Los Altos, CA.

Quinlan, J.R. (1983). Learning efficient classification procedures and their application to chess end games. Machine Learning: An Artificial Intelligence Approach, *Tioga Publishing Company*, Palo Alto, CA.

Roy, B. (1996). Multicriteria Methodology for Decision Aiding. *Kluwer Academic Publishers*, Dordrecht.

Roy, B., Bouyssou, D. (1993). Aide Multicritère à la Décision: Méthodes et Cas. *Economica*, Paris.

Smith, C.A.B. (1947). Some examples of discrimination. *Annals of Eugenics,* 13, pp. 272–282.

Stam, A. (1997). Nontraditional approaches to statistical classification: Some perspectives on Lp-norm methods. *Annals of Operations Research,* 74, pp. 1–36.

Stam, A., Joachimsthaler, E.A. (1989). Solving the classification problem via linear and nonlinear programming methods. *Decision Sciences*, 20, pp. 285–293.

Wilson, D.R., Martinez, T.R. (1997). Improved Heterogeneous Distance Functions. *Journal of Artificial Intelligence Research,* 6, pp. 1-34.

Yu, W. (1992). ELECTRE TRI: Aspects methodologiques et manuel d'utilisation. *Document du Lamsade*, 74, Universite de Paris-Dauphine.

Zopounidis, C., Doumpos, M. (2000). Building additive utilities for multi-group hierarchical discrimination: The MHDIS method. *Optimization Methods and Software*, 14 (3), pp. 219–240.

Zopounidis, C., Doumpos, M. (2002). Multicriteria classification and sorting methods: A literature review. *European Journal of Operational Research*, 138 (2), pp. 229–246.

APPENDIX A

THE OUTPUTS OF THE DISTANCE BASED METHOD WITH L_1 NORM FOR WATER SUPPLIES DATA SET

```
Data-1

Optimistic:

 Triangular  Uniform
1 0.997984   0.968254
2 0.939027   0.825397

Pessimistic:

 Triangular  Uniform
1 0.060973   0.174603
2 0.002016   0.031746

.....................
Data-2

Optimistic:

 Triangular  Uniform
1 0.951994   0.845070
2 0.960325   0.859155

Pessimistic:

 Triangular  Uniform
1 0.039675   0.140845
2 0.048006   0.154930

.....................
Data-3

Optimistic:

 Triangular  Uniform
1 1.000000   1.000000
2 0.758034   0.652174

Pessimistic:

 Triangular  Uniform
1 0.241966   0.347826
2 0.000000   0.000000

.....................
Data-4

Optimistic:

 Triangular  Uniform
1 1.000000   1.000000
2 0.908163   0.785714

Pessimistic:

 Triangular  Uniform
1 0.091837   0.214286
2 0.000000   0.000000

.....................
Data-5

Optimistic:

 Triangular  Uniform
1 0.768800   0.660000
2 0.948800   0.840000

Pessimistic:

 Triangular  Uniform
1 0.051200   0.160000
2 0.231200   0.340000

.....................
Data-6

Optimistic:

 Triangular  Uniform
1 0.946446   0.836364
2 1.000000   1.000000

Pessimistic:

 Triangular  Uniform
1 0.000000   0.000000
2 0.053554   0.163636

.....................
Data-7

Optimistic:

 Triangular  Uniform
1 0.992871   0.940298
2 0.924705   0.805970

Pessimistic:

 Triangular  Uniform
1 0.075295   0.194030
2 0.007129   0.059701

.....................
Data-8

Optimistic:

 Triangular  Uniform
1 0.931413   0.814815
2 0.921963   0.802469

Pessimistic:

 Triangular  Uniform
1 0.078037   0.197531
2 0.068587   0.185185

.....................
Data-9

Optimistic:

 Triangular  Uniform
1 1.000000   1.000000
2 0.942545   0.830508

Pessimistic:

 Triangular  Uniform
1 0.057455   0.169492
2 0.000000   0.000000

.....................
Data-10

Optimistic:

 Triangular  Uniform
1 0.997919   0.967742
2 0.947971   0.838710

Pessimistic:

 Triangular  Uniform
1 0.052029   0.161290
2 0.002081   0.032258

.....................
Data-11

Optimistic:

 Triangular  Uniform
1 0.989796   0.928571
2 0.908163   0.785714
```

Figure A - Probabilities of water supplies data set

```
Pessimistic:

  Triangular   Uniform
1 0.091837     0.214286
2 0.010204     0.071429

......................
Data-12

Optimistic:

  Triangular   Uniform
1 0.895195     0.771084
2 0.943098     0.831325

Pessimistic:

  Triangular   Uniform
1 0.056902     0.168675
2 0.104805     0.228916

......................
```

Figure A -continued

APPENDIX B

THE OUTPUTS OF THE DISTANCE BASED METHOD WITH L_1 NORM FOR R&D PROJECTS DATA SET

Data-1

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-2

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-3

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-4

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-5

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-6

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Figure B - Probabilities of R&D projects data set

Data-7

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-8

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-9

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-10

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-11

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-12

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Figure B -continued

Data-13

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-14

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-15

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-16

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-17

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-18

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Figure B -continued

```
Data-19

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000
4 0.000000    0.000000
5 0.000000    0.000000
6 0.000000    0.000000
7 0.000000    0.000000
8 0.000000    0.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000
4 0.000000    0.000000
5 0.000000    0.000000
6 0.000000    0.000000
7 0.000000    0.000000
8 0.000000    0.000000

........................
Data-20

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 1.000000    1.000000
3 0.000000    0.000000
4 0.000000    0.000000
5 0.000000    0.000000
6 0.000000    0.000000
7 0.000000    0.000000
8 0.000000    0.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 1.000000    1.000000
3 0.000000    0.000000
4 0.000000    0.000000
5 0.000000    0.000000
6 0.000000    0.000000
7 0.000000    0.000000
8 0.000000    0.000000

........................
```

```
Data-21

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000
4 0.000000    0.000000
5 0.000000    0.000000
6 0.000000    0.000000
7 0.000000    0.000000
8 0.000000    0.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000
4 0.000000    0.000000
5 0.000000    0.000000
6 0.000000    0.000000
7 0.000000    0.000000
8 0.000000    0.000000

........................
Data-22

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000
4 0.000000    0.000000
5 0.000000    0.000000
6 0.000000    0.000000
7 0.000000    0.000000
8 0.000000    0.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000
4 0.000000    0.000000
5 0.000000    0.000000
6 0.000000    0.000000
7 0.000000    0.000000
8 0.000000    0.000000

........................
```

```
Data-23

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000
4 0.000000    0.000000
5 0.000000    0.000000
6 0.000000    0.000000
7 0.000000    0.000000
8 0.000000    0.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000
4 0.000000    0.000000
5 0.000000    0.000000
6 0.000000    0.000000
7 0.000000    0.000000
8 0.000000    0.000000

........................
Data-24

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 1.000000    1.000000
3 0.000000    0.000000
4 0.000000    0.000000
5 0.000000    0.000000
6 0.000000    0.000000
7 0.000000    0.000000
8 0.000000    0.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 1.000000    1.000000
3 0.000000    0.000000
4 0.000000    0.000000
5 0.000000    0.000000
6 0.000000    0.000000
7 0.000000    0.000000
8 0.000000    0.000000

........................
```

Figure B -continued

Data-25

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Data-26

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Data-27

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Data-28

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Data-29

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Data-30

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Figure B -continued

Data-31

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-32

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	1.000000	1.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	1.000000	1.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-33

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	1.000000	1.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	1.000000	1.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-34

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-35

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	1.000000	1.000000
7	1.000000	1.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-36

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	1.000000	1.000000
7	1.000000	1.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Figure B -continued

Data-37

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-38

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-39

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-40

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	1.000000	1.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	1.000000	1.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-41

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	1.000000	1.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Data-42

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	1.000000	1.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000
5	1.000000	1.000000
6	0.000000	0.000000
7	0.000000	0.000000
8	0.000000	0.000000

....

Figure B -continued

APPENDIX C

THE OUTPUTS OF THE DISTANCE BASED METHOD WITH L_3 NORM FOR ASSISTANT DATA SET

```
Data-1

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 0.000000     0.000000
3 1.000000     1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 0.000000     0.000000
3 1.000000     1.000000

......................
Data-2

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000
3 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000
3 0.000000     0.000000

......................
Data-3

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000
3 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000
3 0.000000     0.000000

......................
Data-4

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000
3 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000
3 0.000000     0.000000

......................
Data-5

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000
3 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000
3 0.000000     0.000000

......................
Data-6

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000
3 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000
3 0.000000     0.000000

......................
Data-7

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 0.000000     0.000000
3 1.000000     1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 0.000000     0.000000
3 1.000000     1.000000

......................
Data-8

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000
3 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000
3 0.000000     0.000000

......................
Data-9

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 0.000000     0.000000
3 1.000000     1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 0.000000     0.000000
3 1.000000     1.000000

......................
```

Figure C - Probabilities of assistant data set

```
Data-10

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000

......................
Data-11

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000

......................
Data-12

Optimistic:

  Triangular   Uniform
1 1.000000    1.000000
2 0.000000    0.000000
3 0.000000    0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000    1.000000
2 0.000000    0.000000
3 0.000000    0.000000

......................
```

```
Data-13

Optimistic:

  Triangular   Uniform
1 1.000000    1.000000
2 0.000000    0.000000
3 0.000000    0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000    1.000000
2 0.000000    0.000000
3 0.000000    0.000000

......................
Data-14

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000

......................
Data-15

Optimistic:

  Triangular   Uniform
1 1.000000    1.000000
2 0.000000    0.000000
3 0.000000    0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000    1.000000
2 0.000000    0.000000
3 0.000000    0.000000

......................
```

```
Data-16

Optimistic:

  Triangular   Uniform
1 1.000000    1.000000
2 0.000000    0.000000
3 0.000000    0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000    1.000000
2 0.000000    0.000000
3 0.000000    0.000000

......................
Data-17

Optimistic:

  Triangular   Uniform
1 1.000000    1.000000
2 0.000000    0.000000
3 0.000000    0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000    1.000000
2 0.000000    0.000000
3 0.000000    0.000000

......................
Data-18

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 1.000000    1.000000

......................
```

Figure C -continued

Data-19

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000

....

Data-20

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-21

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

....

Data-22

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-23

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-24

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-25

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-26

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

....

Data-27

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Figure C -continued

Data-28

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-29

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000

....

Data-30

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

....

Data-31

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000

....

Data-32

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000

....

Data-33

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

....

Data-34

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-35

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-36

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

....

Figure C -continued

Data-37

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

....

Data-38

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-39

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-40

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-41

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-42

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-43

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-44

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-45

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

....

Figure C -continued

Data-46

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-47

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-48

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-49

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000

....

Data-50

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000

....

Figure C -continued

APPENDIX D

THE OUTPUTS OF THE DISTANCE BASED METHOD WITH L_2 NORM FOR CARS DATA SET

Data-1

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

....

Data-2

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000

....

Data-3

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000

....

Data-4

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000

....

Data-5

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	0.000000	0.000000
4	0.000000	0.000000

....

Data-6

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

....

Data-7

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

....

Data-8

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

....

Data-9

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

....

Figure D - Probabilities of cars data set

Data-10

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

....

Data-11

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000

....

Data-12

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

....

Data-13

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

....

Data-14

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-15

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-16

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

....

Data-17

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-18

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Figure D - continued

Data-19

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-20

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-21

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-22

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-23

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-24

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-25

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-26

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-27

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Figure D - continued

Data-28

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-29

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-30

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-31

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-32

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-33

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-34

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-35

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-36

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.000000	0.000000

....

Figure D - continued

Data-37

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-38

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-39

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-40

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-41

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-42

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-43

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

....

Data-44

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-45

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Figure D - continued

Data-46

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-47

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-48

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-49

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-50

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-51

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-52

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-53

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-54

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Figure D - continued

Data-55

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-56

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-57

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-58

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-59

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-60

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-61

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	1.000000	1.000000
4	0.000000	0.000000

....

Data-62

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Data-63

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	1.000000	1.000000

....

Figure D - continued

```
Data-64

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

----------------------
Data-65

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

----------------------
Data-66

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

----------------------
Data-67

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

----------------------
Data-68

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

----------------------
Data-69

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

----------------------
Data-70

Optimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000    0.000000
2 0.000000    0.000000
3 0.000000    0.000000
4 1.000000    1.000000

----------------------
```

Figure D - continued

APPENDIX E

THE OUTPUTS OF THE DISTANCE BASED METHOD WITH L_3 NORM FOR CREDIT DATA SET

Data-1

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-2

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-3

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000

....

Data-4

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-5

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000

....

Data-6

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-7

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-8

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-9

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-10

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-11

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-12

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Figure E – Probabilities of credit data set

```
Data-13

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-14

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-15

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

......................
Data-16

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-17

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-18

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-19

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

......................
Data-20

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-21

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-22

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-23

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

......................
Data-24

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

......................
```

Figure E - continued

Data-25

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-26

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-27

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-28

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-29

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-30

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-31

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-32

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-33

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-34

Optimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000

Pessimistic:

	Triangular	Uniform
1	0.000000	0.000000
2	1.000000	1.000000

....

Data-35

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Data-36

Optimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

Pessimistic:

	Triangular	Uniform
1	1.000000	1.000000
2	0.000000	0.000000

....

Figure E - continued

```
Data-37

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-38

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-39

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-40

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

......................
Data-41

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

......................
Data-42

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

......................
Data-43

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-44

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

......................
Data-45

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-46

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-47

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
Data-48

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
```

Figure E - continued

```
Data-49

Optimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

Pessimistic:

  Triangular   Uniform
1 0.000000     0.000000
2 1.000000     1.000000

......................
Data-50

Optimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

Pessimistic:

  Triangular   Uniform
1 1.000000     1.000000
2 0.000000     0.000000

......................
```

Figure E - continued

APPENDIX F

THE HIGHEST PROBABILITY CLASSES OF EACH DATA SET FOR THE BEST PERFORMING L_p NORM DISTANCE BASED METHODS

Table F – Results of distane-based method for each data set

Data Point	Water Supplies		R&D Projects		Assistant		Cars		Credit	
	Real	Pred.	Real	Pred.	Real	Pred.	Real	Pred.	Real	Pred.
1	1	1	1	1	1	3	1	3	1	1
2	1	2	2	2	1	1	1	1	1	1
3	1	1	2	2	1	2	1	2	1	1
4	2	1	2	2	1	1	2	1	1	1
5	2	2	2	2	1	1	2	2	1	1
6	2	2	2	2	1	1	3	3	1	1
7	2	1	2	2	1	1	3	3	1	1
8	2	1	2	2	1	1	3	3	1	1
9	1	1	2	2	1	3	3	3	1	1
10	1	1	2	3	1	3	3	3	1	1
11	1	1	2	2	1	3	3	3	1	1
12	2	2	2	2	1	1	3	3	1	1
13			2	3	1	1	3	3	1	1
14			2	2	1	3	3	4	1	1
15			2	2	2	1	3	4	1	1
16			3	3	2	1	3	3	1	1
17			3	3	2	1	4	4	1	1
18			3	3	2	3	4	4	1	1
19			3	3	2	2	4	4	1	1
20			3	2	2	3	4	4	1	1
21			3	3	2	1	4	4	1	1
22			3	3	2	3	4	4	1	1
23			3	3	2	3	4	4	1	1
24			3	2	2	3	4	4	1	1
25			3	3	2	3	4	4	1	1
26			3	3	2	1	4	4	1	1
27			3	3	2	3	4	4	1	1
28			3	3	2	3	4	4	1	1
29			3	3	2	2	4	4	1	1
30			3	3	2	1	4	4	1	1
31			3	3	2	2	4	4	1	1
32			5	5	2	2	4	4	1	1
33			5	5	2	1	4	4	2	1
34			5	3	2	3	4	4	2	2
35			6-7	6-7	3	3	4	4	2	1
36			6	6	3	1	4	1	2	1

Table F – continued

37			2	2	3	1	4	4	2	1
38			3	3	3	3	4	4	2	1
39			4	4	3	3	4	4	2	2
40			4	5	3	3	4	4	2	2
41			4	4	3	3	4	4	2	1
42			4	5	3	3	4	4	2	1
43					3	3	4	3	2	1
44					3	3	4	4	2	1
45					3	1	4	4	2	1
46					3	3	4	4	2	1
47					3	3	4	4	2	1
48					3	3	4	4	2	1
49					3	3	4	4	2	2
50					3	1	4	4	2	1
51							4	4		
52							4	4		
53							4	4		
54							4	4		
55							4	4		
56							4	4		
57							4	4		
58							4	4		
59							4	4		
60							4	4		
61							4	3		
62							4	4		
63							4	4		
64							4	4		
65							4	4		
66							4	4		
67							4	4		
68							4	4		
69							4	4		
70							4	4		

APPENDIX G

THE HIGHEST PROBABILITY CLASSES OF EACH DATA SET FOR MODIFIED UTADIS METHOD

Table G – Results of modified UTADIS method for each data set

Data Point	Water Supplies		R&D Projects		Assistant		Cars		Credit	
	Real	Pred.	Real	Pred.	Real	Pred.	Real	Pred.	Real	Pred.
1	1	1	**1**	1	1	2	1	**1..2**	1	1
2	1	1	**2**	1..2	1	2	1	**1..2**	1	1
3	1	1	**2**	2	1	2	1	**1..2**	1	1
4	2	1	**2**	2	1	2	2	**1**	1	1
5	2	1	**2**	2	1	2	2	**1..2**	1	1
6	2	1	**2**	1..2	1	2	3	**3..4**	1	1
7	2	1	**2**	1..2	1	2	3	**3**	1	1
8	2	1	**2**	1..2	1	2	3	**3**	1	1
9	1	1	**2**	2	1	2	3	**3**	1	1
10	1	1	**2**	2..3	1	2	3	**3**	1	1
11	1	1	**2**	2	1	2	3	**3**	1	1
12	2	1	**2**	1..2..3	1	2	3	**3**	1	1
13			**2**	2	1	2	3	**3**	1	1
14			**2**	2	1	2	3	**4**	1	1
15			**2**	2	2	2	3	**4**	1	1
16			**3**	3	2	2	3	**3**	1	1
17			**3**	3	2	2	4	**3..4**	1	1
18			**3**	3	2	2	4	**4**	1	1
19			**3**	3	2	2	4	**4**	1	1
20			**3**	2	2	2	4	**4**	1	1
21			**3**	3	2	2	4	**4**	1	1
22			**3**	3	2	2	4	**4**	1	1
23			**3**	3	2	2	4	**4**	1	1
24			**3**	2	2	2	4	**4**	1	1
25			**3**	3	2	2	4	**4**	1	1
26			**3**	3	2	2	4	**4**	1	1
27			**3**	3	2	2	4	**4**	1	1
28			**3**	3	2	2	4	**3..4**	1	1
29			**3**	3	2	2	4	**4**	1	1
30			**3**	3	2	2	4	**3..4**	1	1
31			**3**	3	2	2	4	**4**	1	1
32			**5**	4..5	2	2	4	**4**	1	1
33			**5**	4..5	2	2	4	**4**	2	1
34			**5**	3	2	2	4	**4**	2	1

Table G – continued

35			**6**	6..7	3	2	4	**4**	2	1
36			**6**	4..5..6..7	3	2	4	**3**	2	1
37			**2**	1..2	3	2	4	**3**	2	1
38			**3**	3	3	2	4	**4**	2	1
39			**4**	4..5	3	2	4	**4**	2	1
40			**4**	4..5	3	2	4	**4**	2	1
41			**4**	4..5	3	2	4	**4**	2	1
42			**4**	4..5	3	2	4	**4**	2	1
43					3	2	4	**4**	2	1
44					3	2	4	**4**	2	1
45					3	2	4	**4**	2	1
46					3	2	4	**4**	2	1
47					3	2	4	**4**	2	1
48					3	2	4	**4**	2	1
49					3	2	4	**4**	2	1
50					3	2	4	**4**	2	1
51							4	**4**		
52							4	**4**		
53							4	**4**		
54							4	**4**		
55							4	**3..4**		
56							4	**4**		
57							4	**4**		
58							4	**4**		
59							4	**4**		
60							4	**4**		
61							4	**3**		
62							4	**4**		
63							4	**4**		
64							4	**4**		
65							4	**4**		
66							4	**4**		
67							4	**4**		
68							4	**4**		
69							4	**4**		
70							4	**4**		