8-1-2005

Nonlinear Multiregressions Based on Choquet Integral for Data with both Numerical and Categorical Attributes.

Jin Hui

Nonlinear Multiregressions Based on Choquet Integral for Data with both Numerical and Categorical Attributes

A Thesis

Presented to the

Department of Mathematics

and the

Faculty of the Graduate College

University of Nebraska

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

University of Nebraska at Omaha

by

Jin Hui

August 2005

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

THESIS ACCEPTANCE

Acceptance for the faculty of the Graduate College,
University of Nebraska, in partial fulfillment of the
requirements for the degree Master of Arts,
University of Nebraska at Omaha.

Committee

Dr. Zhenyuan Wang _____

Dr. Vyacheslav Rykov _____

Dr. Haifeng Guo _____

Chairperson _____

Date 4/29/2005

Nonlinear Multiregressions Based on Choquet Integral for Data with both Numerical and Categorical Attributes

Jin Hui (MA)

University of Nebraska, 2005

Advisor: Zhengyuan Wang

Abstract: Based on generalized Choquet integrals with respect to signed fuzzy measures, a model of nonlinear multiregression that can catch the interaction among predictive attributes toward the objective attribute can be established. In this model, some predictive attributes are numerical while the others are categorical. A numericalization technique is adopted to project each state of a categorical attribute that has more than two states to a multi-dimensional space optimally through a genetic algorithm, in which some regression coefficients are determined from data. To reduce the complexity of the genetic algorithm, the other regression coefficients such as the values of the signed fuzzy measure are determined by an algebraic method.

In conclusion, this paper improves the previous relative work in several aspects:

(1) Using a signed fuzzy measure to replace the generalized fuzzy measure such that the regression can more appropriately describe the relation among the objective attribute and the predictive attributes.

(2) To reduce the complexity of the genetic algorithm that is used to search the optimal estimation of the regression coefficients, taking a part of the unknown

regression coefficients, the values of the signed fuzzy measure, out from the chromosome involved in the genetic algorithm.

(3) Optimally projecting the states of the categorical attribute(s) into a partial ordering space instead of a total ordering space as done in the previous work, to "numericalize" the categorical attribute(s) when there are more than two states for a predictive attribute.

CONTENTS

ACKNOWLEDGEMENTS

Mathematics has always captured my attention and imagination. I truly believe that an advanced education in math can pave a successful way for my career in the competitive world. This was the primary reason for me to apply to the master program in the Department of Mathematics.

In 2003, I came to America to pursue my master's degree in the Department of Mathematics at the University of Nebraska at Omaha. Here, I met those who shared my curiosity and penchant for truth seeking in the field of mathematics and statistics. I will always remember and cherish this two sweet years in my life.

It is my pleasure to thank the many people who have made this thesis possible. Dr. Zhenyuan Wang for his phenomenal direction and support on this work. Dr. Vyacheslav Rykov and Dr. Haifeng Guo for being two wonderful advisors with their door are always open. Dr. Jack Heidel for offering me the nice TA job and uncountable help the whole time. I am grateful to my parents for their long-distance and incredible moral support. I also wish to thank my awesome husband for all he is and does. At last, I want to say hi to my coming baby, she will arrive this October, I can't wait to see her.

CHAPTER 1

INTRODUCTION

A new type of nonlinear multiregression based on Choquet integrals with respect to generalized fuzzy measures has been used to discover dependency pattern among attributes recently[25]. The specialty of the new technique is numericalizing categorical attributes to build a nonlinear multiregression when the predictive attributes have mixed type data, numerical and categorical attributes. However, in some cases, when using the generalized fuzzy measure, due to its nonnegativity, the regression can hardly appropriately describe the relation among the objective attribute and the predictive attributes. When all the regression coefficients are determined via a genetic algorithm, each chromosome is so long that the complexity of the algorithm is rather high. In this thesis, the values of the signed fuzzy measure, which is defined on the power set of all predictive attributes, are determined via an algebraic method, then only a few parameters are arranged as genes to form chromosomes in the genetic algorithm. Therefore, the running time of the program is shortened and the accuracy of the result is increased. Furthermore, we develop an evolutionary numericalization technique to convert each categorical attribute to be numerical by optimally projecting its states into a partial ordering multi-dimensional space

CHAPTER 2

Set Theory

Set Theory is the mathematical science of the infinite. It studies properties of sets, abstract the objects that pervade the whole modern mathematics. The language of set theory, in its simplicity, is sufficiently universal to formalize all mathematical concepts and thus set theory, along with Predicate Calculus, constitutes the true Foundations of Mathematics. As a mathematical theory, Set Theory possesses a rich internal structure, and its methods serve as a powerful tool for applications in many other fields of Mathematics.

Set Theory, with its emphasis on consistency and independence proofs, provides a gauge for measuring the consistency strength of various mathematical statements. There are four main directions of current research in set theory, all intertwined and all aiming at the ultimate goal of the theory: to describe the structure of the mathematical universe. They are: inner models, independence proofs, large cardinals, and descriptive set theory.

What's the Essence of set theory? As the name says, the objects of study of set theory are sets. As sets are fundamental objects that can be used to define all other concepts in mathematics, they are not defined in terms of more fundamental concepts. Rather, sets are introduced either informally, and are understood as something self-evident, or, as is now standard in modern mathematics, axiomatically, and their properties are postulated by the appropriate formal axioms.

2.1 Set inclusion and characteristic function

The concept of set inclusion has remained insufficiently developed in the fuzzy set literature to be of much use to social scientists. However, a fully fledged concept of fuzzy set inclusion, along with appropriate statistical methods for evaluating it, could be very useful.

A set consists of some points. In other words, a set is just a collection of things. We can talk about the set of even numbers, the set of people who are taking algebra or the set of books about large white whales. The things in the set are called the elements of the set.

Usually, sets are denoted by capital English letters such as A, B, E, F, U, X; while points (elements) are denoted by low case English letters such as a, b, x, y. As some special sets, the set of all real numbers is denoted by R, and the set of all positive integers is denoted by N.

Let X be a nonempty set. Unless otherwise stated, all sets that we consider are subsets of X. X is called the universe of discourse. The elements of X are called points. X may contain finite, countably infinite, or uncountably infinite number of points. A set that consists of a finite number of points $x_1, x_2, \ldots\ldots, x_n$ (or, a countably infinite number of points $x_1, x_2, \ldots$) may be denoted by $\{x_1, x_2, \ldots, x_n\}$ ($\{x_1, x_2, \ldots\}$, respectively). A set containing on point is called the empty set, and is denoted by $\varnothing$.

If x is a point of X and E is a subset of X, the notation

$$x \in E$$

means that x belongs to E, i.e., x is an element of E; and the statement that x does not belong to E is denoted by

$$x \notin E.$$

Thus, for every point x in X, we have

$$x \in X$$

and

$$x \notin \varnothing$$

A set of sets is called a class. If E is a set and $\mathscr{C}$ is a class, then

$$E \in \mathscr{C}$$

means that the set E belongs to the class $\mathscr{C}$.

If, for each x, $\pi(x)$ is a proposition concerning x, then the symbol

$$\{x \mid \pi(x)\}$$

denotes the set of all those points x for which $\pi(x)$ is true, that is,

$$x_0 \in \{x \mid \pi(x)\} \Leftrightarrow \pi(x_0) \text{ is true.}$$

By replacing point x with set E, such a symbol may be used to indicate a class. For example,

$$\{E \mid x \in E\}$$

denotes the class of those sets that contain the point x.

Example 2.1. Let X=$\{1,2,\cdots\}$. The set $\{x \mid x$ is odd and less than 5$\}$ is $\{1,3\}$.

Example 2.2. Let X be the set of all real numbers, which is often referred to as the real line or one-dimensional Euclidean space. The set $\{(a,b) \mid -\infty < a < \infty\}$ is the set consisting of all real numbers.

If E and F are sets, the notation

$$E \subseteq F \text{ or } F \supseteq E$$

means that E is a subset of F, i.e., every point of E belongs to F. In this case, we say that$F$ includes E, or that E is included by F. For every set E, we have

$$\varnothing \subseteq E \subseteq X$$

Two sets E and F are called equal iff

$$E \subseteq F \text{ and } F \subseteq E;$$

that is, they contain exactly the same points. This is denoted by

$$E = F$$

The symbols $\subseteq$ or $\supseteq$ also may be used for classes. If $\mathscr{C}$ and $\mathscr{F}$ are classes, then

$$\mathscr{C} \subseteq \mathscr{F}$$

means that every set of $\mathscr{C}$ belongs to $\mathscr{F}$, that is, $\mathscr{C}$ is a subclass of $\mathscr{F}$.

If $E_1, E_2, \ldots, E_n$ are nonempty sets, then

$$E = \{(x_1, x_2, \ldots, x_n) \mid x_i \in E_i, i=1,2, \ldots, n\}$$

is called an n-dimensional product set, it is denoted by

$$E = E_1 \times, E_2 \times \ldots \times E_n$$

Similarly , if $\{E_t \mid t \in T\}$ is a family of nonempty sets, where T is an infinite index set, then

$$E = \{x_t, t \in T \mid x_t \text{ for each } t \in T\}$$

is called an infinite-dimensional product set.

Example 2.3. Let X be the set of all real numbers, i.e., $X = R$. Interval [0, 3] is a subset of interval (0, 5). We have

$$\chi_{[0,3]}(x) = \begin{cases} 1, & \text{if } 0 \le x \le 3 \\ 0, & \text{otherwise} \end{cases},$$

$$\chi_{(0,5)}(x) = \begin{cases} 1, & \text{if } 0 < x < 5 \\ 0, & \text{otherwise} \end{cases},$$

and $\chi_{[0,3]} \le \chi_{(0,5)}$.

Example 2.4. Let $X = \{u, v\}$, $A = \{u\}$, and $B = \{v\}$. Then, neither $A \subseteq B$ nor $B \subseteq A$. In fact, we have

$$\chi_A(x) = \begin{cases} 1, & x = u \\ 0, & x \ne v \end{cases},$$

$$\chi_B(x) = \begin{cases} 1, & x = u \\ 0, & x \ne v \end{cases},$$

and neither $\chi_A \le \chi_B$ nor $\chi_B \le \chi_A$.

Example 2.5. Let X_1 and X_2 be one dimensional Euclidean spaces. Then $X = X_1 \times X_2 = \{(x_1, x_2) \,|\, x_1 \in (-\infty, \infty),\ x_2 \in (-\infty, \infty)\}$ is the two-dimensional Euclidean space. The set $\{(x_1, x_2) \,|\, x_1 < x_2\}$ is a half (open) plane over the line $x_2 \equiv x_1$, while the set $\{(x_1, x_2) \,|\, x_1^2 + x_2^2 \le r^2\}$ is the close circle centering at the origin with a radius r, where $r>0$.

If E is a set, the function x_E, defined for all $x \in X$ by

$$x_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases}$$

is called the characteristic function of the set E. The correspondence between sets and their characteristic functions is one to one, that is

$$E = F \Leftrightarrow X_E(x) = x_F(x), \quad \forall x \in X.$$

It is easy to see that

$$E \subset F \Leftrightarrow X_E(x) \leq x_F(x), \quad \forall x \in X.$$

and that

$$x_x \equiv 1, \quad x_\phi \equiv 0$$

2.2 Operations on sets

Sets can be combined in a number of different ways to produce another set. Here four basic operations are introduced and their properties are discussed.

Let $\mathscr{C}$ be any class of subsets of X. The set of all those points of X that belong to at least one set of the class $\mathscr{C}$ is called the union of the sets of $\mathscr{C}$. This is denoted by

$$\bigcup \mathscr{C}$$

If to every t of a certain index set T there corresponds a set E_t, then the union of the sets of class

$$\{E_t | t \in T\}$$

may be also denoted by

$$\bigcup_{t \in T} E_t \text{ or } \bigcup_t E_t ,$$

Especially, when

$$\mathscr{C}=\{E_1,E_2\}.$$

Then $\bigcup \mathscr{C}$ is denoted by

$$E_1 \cup E_2;$$

and if

$$\mathscr{C}=\{E_1,E_2,\ldots,E_n\} \quad (\mathscr{C}=\{E_1,E_2,\ldots,\}),$$

then $\bigcup \mathscr{C}$ is denoted by

$$E_1 \cup E_2 \cup \ldots \cup E_n \text{ or } \bigcup_{i=1}^{n} E_i \qquad (\bigcup_{i=1}^{\infty} E_i \text{ respectively}).$$

The set of all those points of X which belong to every set of the class $\mathscr{C}$ is called the intersection of the sets of $\mathscr{C}$. This is denoted by $\bigcap \mathscr{C}$. Symbols similar to those used for unions are available, such as $\bigcap_{t\in T} E_t$ (or $\bigcap_{t} E_t$), $E_1 \cap E_2, E_1 \cap E_2 \cap \ldots \cap E_n$ (or $\bigcap_{i=1}^{n} E_i$) and $\bigcap_{i=1}^{\infty} E_i$. If F is a set, the class $\{E \cap F \mid E \in \mathscr{C}\}$ is denoted by $\mathscr{C} \cap F$.

Example 2.6. Let $X=\{a,b,c,d\}$, $\mathscr{C}=\{\{a\},\{c,d\}\}$, and $F=\{a,b\}$. Then

$\mathscr{C} \cap F=\{\{a\},\varnothing\}$.

Example 2.7. Let $X=(-\infty,\infty)$, $\mathscr{C}=\{[a,b] \mid -\infty < a \le b < \infty\}$, and F=[0,10].

Then, $\mathscr{C} \cap F=\{a,b \mid 0 \le a \le b \le 10\}$, that is, the class of all closed subintervals of the unit closed interval.

It is convenient to adopt the conventions that

$$\bigcup_{t\in T} E_t = \varnothing$$

and

$$\bigcap_{t\in T} E_t = X$$

When T is empty.

Proposition 2.1. The following statements are equivalent:

(1) $E \subseteq F$

(2) $E \cup F = F$

(3) $E \cap F = E$

Two sets E and F are called disjoint iff

$$E \cap F = \varnothing$$

A class $\mathscr{C}$ is called disjoint iff every two distinct sets of $\mathscr{C}$ are disjoint; in this case we refer to the union of the sets of $\mathscr{C}$ as a disjoint union.

If E is a set, the set of all those points of X that do not belong to E is called the complement of E. This is denoted by $\overline{E}$.

In general, we have the following principle of duality: Any valid identity among sets, obtained by unions, intersections, and complements, remains valid, if the symbols

$$\cap, \subseteq, \text{ and } \varnothing$$

are interchanged with $\cap$, $\subseteq$, and X respectively (and if the equality and complementation are left unchanged).

If E and F are sets, the set of all those points of E that do not belong to F is called the difference of E and F. This is denoted by

$$E - F$$

If $E \supseteq F$, the difference $E - F$ is called proper.

$E - F$ can be expressed by the intersection and the complement, that is,

$$E - F = E \cap \overline{F}$$

The symmetric difference of E and F, in symbols

$$E \Delta F$$

is defined by

$$E \Delta F = (E - F) \cup (F - E)$$

The difference can be expressed in terms of the intersection and the complement, that is, $A - B = A \cap \overline{B}$. By using De Morgan's law $\overline{A \cup B} = \overline{A} \cap \overline{B}$, we can express the union in terms of the intersection and the complement as follows:

$$A \cup B = \overline{\overline{A} \cap \overline{B}}.$$

Similarly, by using De Morgan's law $\overline{A \cap B} = \overline{A} \cup \overline{B}$, we can express the intersection in terms of the union and the complement as well:

$$A \cap B = \overline{\overline{A} \cup \overline{B}}.$$

So, we have only two basic set operators: either the intersection and the complement, or the union and the complement.

Chapter 3

Signed Fuzzy Measures

Measure theory initially was created to provide a detailed analysis of the notion of length of subsets of the real line and more generally area and volume of subsets of Euclidean spaces. The terms "measure", "measurable", etc. have very appropriate technical definitions (usually involving sigma-algebras) that can make them appear difficult to understand. However, the technical nature of the definitions is extremely important, since it gives a firm footing to concepts that are the basis for much of analysis (including some of the slippery underpinnings of calculus).

3.1 Measures

μ is called additive on $\mathscr{C}$ iff $\mu(E \cup F) = \mu(E) + \mu(F)$ whenever $E \in \mathscr{C}$, $F \in \mathscr{C}$, $E \cup F \in \mathscr{C}$, and $E \cap F = \varnothing$.

μ is called σ-additive (countably additive) on $\mathscr{C}$ iff $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ whenever $\{A_i\}$ is a disjoint sequence of sets in $\mathscr{C}$ and $\bigcup_{i=1}^{\infty} A_i$ is also in $\mathscr{C}$.

μ is called a measure on $\mathscr{C}$ iff μ is σ-additive and there exists $E \in \mathscr{C}$ such that $\mu(E) < \infty$.

If μ is a measure on $\mathscr{C}$ and $\varnothing \in \mathscr{C}$, then $\mu(\varnothing) = 0$.

Let $(X, \mathscr{R}_\sigma)$ be a measurable space and μ be a measure on $\mathscr{R}_\sigma$ (we also say that μ is a measure on $(X, \mathscr{R}_\sigma)$). Then triple $(X, \mathscr{R}_\sigma, \mu)$ is called a measure space.

Example 3.1. Let $X = \{a, b, c\}$, $\mathscr{C} = \mathscr{P}(X)$, and $\mu(E) = |E|$ for $E \in \mathscr{C}$. Then μ is a finite measure on $\mathscr{C}$.

Example 3.2. Let $X = \{x_1, x_2, \ldots\}$ and $\mu(E) = \sum_{x_i \in E} 3^{-i}$. Then μ is a regular measure on $\mathscr{P}(X)$.

Let $(X, \mathscr{R}_\sigma, \mu)$ be a measure space.

Property (M1) μ is monotonic, that is, $\mu(E) \le \mu(F)$ whenever $E \in \mathscr{R}_\sigma$, $F \in \mathscr{R}_\sigma$, and $E \subseteq F$.

Property (M2) μ is continuous from below, that is $\lim_{i\to\infty} \mu(E_i) = \mu(\bigcup_{i=1}^{\infty} E_i)$ whenever $E_i \in \mathscr{R}_\sigma$, $i = 1, 2, \ldots$, and $\{E_i\}$ is nondecreasing.

Property (M3) μ is continuous from above, that is $\lim_{i\to\infty} \mu(E_i) = \mu(\bigcap_{i=1}^{\infty} E_i)$ whenever $E_i \in \mathscr{R}_\sigma$, $i = 1, 2, \ldots$, $\{E_i\}$ is nonincreasing , and there exists i_0 such that $\mu(E_{i_0}) < \infty$.

3.2 Fuzzy Measures

Fuzzy measures and integrals were introduced by Sugeno in 1974. The motivation is the following: suppose we pick an element ω from a set Ω, but we don't know which one. For a subset A of Ω, you are asked to guess whether $\omega \in A$. It is argued that humans can subjectively express their degrees of confidence with values in [0,1]. Thus, to each $A \subseteq \Omega$ the value $\upsilon(A)$ is assigned expressing a belief

that $\omega \in A$. Therefore, it is sound to require that $\upsilon(\varnothing) = 0$, $\upsilon(\Omega) = 1$, and υ is monotone increasing.

Let X be a nonempty set, $\mathscr{C}$ be a nonempty class of subsets of X, and $\mu : \mathscr{C} \to [0, \infty)$ be a nonnegative, extended real-valued set function defined on $\mathscr{C}$. Generally, we use the following conventions:

$$\sup_{x \in \phi} \{x | x \in [0, \infty]\} = 0,$$

$$\sum_{i \in \phi} a_i = 0,$$

where $\{a_i\}$ is a real number sequence.

Now, we give the following four definitions for fuzzy measures.

If μ satisfies (FM1) $\mu(\varnothing) = 0$ when $\varnothing \in \mathscr{C}$ (vanishing at $\varnothing$)

and (FM2) $E \in \mathscr{C}, F \in \mathscr{C}$, and $E \subseteq F$ imply $\mu(E) \leq \mu(F)$ (monotonicity),

then μ is called a fuzzy measure on $(X, \mathscr{C})$.

In addition, if fuzzy measure μ satisfies

(FM3) $\{E_n\} \subseteq \mathscr{C}$, $E_1 \subseteq E_2 \subseteq \cdots$, and $\bigcup_{n=1}^{\infty} E_n \in \mathscr{C}$ imply $\lim_n \mu(E_n) = \mu(\bigcup_{n=1}^{\infty} E_n)$,

then μ is called a lower semi-continuous fuzzy measure.

Similarly, if fuzzy measure μ satisfies

(FM4) $\{E_n\} \subseteq \mathscr{C}$, $E_1 \subseteq E_2 \subseteq \cdots$, $\mu(E_1) < \infty$, and $\bigcap_{n=1}^{\infty} E_n \in \mathscr{C}$ imply

$$\lim_n \mu(E_n) = \mu(\bigcap_{n=1}^{\infty} E_n),$$

then μ is called a upper semi-continuous fuzzy measure.

At last, if fuzzy measure μ satisfies both

(FM3) $\{E_n\} \subseteq \mathscr{C}$, $E_1 \subseteq E_2 \subseteq \cdots$, and $\bigcup_{n=1}^{\infty} E_n \in \mathscr{C}$ imply $\lim_n \mu(E_n) = \mu(\bigcup_{n=1}^{\infty} E_n)$

and (FM4) $\{E_n\} \subseteq \mathscr{C}$, $E_1 \subseteq E_2 \subseteq \cdots$, $\mu(E_1) < \infty$, and $\bigcap_{n=1}^{\infty} E_n \in \mathscr{C}$ imply

$$\lim_n \mu(E_n) = \mu(\bigcap_{n=1}^{\infty} E_n),$$

then μ is called a continuous fuzzy measure.

Now, let's define μ', the dual of μ:

$$\mu'(A) = \mu(X) - \mu(\overline{A})$$

3.3 Signed fuzzy measures

The nonadditivity of fuzzy measure can effectively describe the interaction among the contributions from each attribute toward some target. Some works have shown successful applications of the Choquet integral in nonlinear multiregressions, classifications, and decision-makings, where the values of fuzzy measure are usually regarded as unknown parameter to be elicited from training data sets. However, fuzzy measure restricts its values to be nonnegative, which may arise limitations in practical

applications. It is not reasonable to assume that all values of fuzzy measure are nonnegative before we elicit them from real data. Sometimes, negative values may exist, such as in pharmacological, financial, and sociological applications. Thus, a more generalized fuzzy measure, the signed fuzzy measure, should be considered and investigated. The signed fuzzy measure may assume some negative values and, therefore, may be nonmonotonic.

Let X be the factor space which is composed of n predictive attributes $X = \{x_1, x_2, ..., x_n\}$. A signed fuzzy measure is a set function μ defined on $(X, \mathcal{P}(X))$ satisfying $\mu(\varnothing) = 0$, where $\mathcal{P}(X)$ represents the power set of X. Usually, we can assume that $\mu(X) \geq 0$, other wise it won't get the unique value.

Set function $\mu : \mathcal{P}(X) \rightarrow (-\infty, \infty)$ is called a signed fuzzy measure if $\mu(\phi) = 0$. μ is simply called a fuzzy measure if it is nonnegative and monotonic[1,2]. μ is said to be regular if

$$|\mu(A)| \leq 1, \forall A \subseteq X, \text{ and } \max_{A \subseteq X} \mu(A) - \min_{A \subseteq X} \mu(A) = 1$$

A signed fuzzy measure is not necessary additive. Its nonadditivity reflects the interaction among predictive attributes toward the objective attribute. Example 3.3 shows why a signed fuzzy measure is needed in case the interaction among predictive attributes cannot be ignored and how it describes the interaction.

Example 3.3. Three workers x_1, x_2, and x_3 are hired for manufacturing the same kind of products. Their respective efficiencies are 6,5, and 9 products per day. However, x_1 and x_2 can manufacture 14 products per day if they work together; x_1 and x_3 can only manufacture 7 products per day; x_2 and x_3 together can manufacture 25 products per day. These efficiencies can be expressed by a signed fuzzy measure

μ on the power set of $\{x_1, x_2, x_3\}$. The values of μ are listed in Table 3.1. Here inequality $\mu(\{x_1, x_2\}) > \mu(\{x_1\}) + \mu(\{x_2\})$ means that x_1 and x_2 have a good cooperation, while inequality $\mu(\{x_1, x_3\}) < \mu(\{x_1\}) + \mu(\{x_3\})$, and even $\mu(\{x_1, x_3\}) < \mu(\{x_3\})$, means that x_1 and x_3 have a very bad relationship and they are not suitable for working together. The nonadditivity of μ represents the interaction among three workers toward the number of total products. In general, a signed fuzzy measure may take negative values, though μ is nonnegative in this example actually.

In case these three workers work always separately, then their joint efficiencies should satisfy the additivity. Their respective efficiencies and the joint efficiencies for all possible combinations of workers are also listed in Table I as the values of additive measure λ. When they work together, we have to consider their joint efficiency, listed in Table 3.1 as the values of set function μ.

Table 3.1. Measure λ and nonadditive measure μ used in this example

Set	Value of λ	Value of μ
$\varnothing$	0	0
$\{x_1\}$	5	6
$\{x_2\}$	6	5
$\{x_1, x_2\}$	11	14
$\{x_3\}$	8	9
$\{x_1, x_3\}$	13	7
$\{x_2, x_3\}$	14	16
$\{x_1, x_2, x_3\}$	19	25

If these three workers x_1, x_2 ,and x_3 are hired for 10, 15, and 7 days respectively to manufacture the products. We use function f to denote the number of working

days for them, then the total number of products they manufactured during such working periods would be the Lebesgue-like integral of f with respect to additive measure λ:

$$\int f d\lambda = 10 \times 6 + 15 \times 5 + 7 \times 9 = 198$$

However, they work together actually. The interaction among them must be considered. Suppose that they begin to work from the same day and continue until the end of their respective working period. In such a manner, during the first 7 days, all workers work together with efficiency $\mu(\{x_1, x_2, x_3\})$, and the number of products is $f(x_3) \cdot \mu(\{x_1, x_2, x_3\}) = 7 \times 25 = 175$; during the next $f(x_1) - f(x_3)$ days, workers x_1 and x_2 work together with efficiency $\mu(\{x_1, x_2\})$, and the number of products is $[f(x_1) - f(x_3)] \cdot \mu(\{x_1, x_2\}) = 3 \times 14 = 42$; during the last $f(x_2) - f(x_1)$ days, only x_2 works alone with efficiency $\mu(\{x_2\})$, and the number of products is $[f(x_{2)}) - f(x_1)] \cdot \mu(\{x_2\}) = 5 \times 5 = 25$. Thus, the total number of products manufactured by these work during the periods mentioned above should be 175+42+25=242. This is just the Choquet integral $\int f d\mu$.

The interaction among workers appearing in this Example is today different from the concept of correlation in statistics. The latter is used to describe the relation between the observations of two predictive attributes and dose not concern the objective attribute.

CHAPTER 4

CHOQUET INTEGRALS

The Choquet integral with respect to fuzzy measure is often used in information fusion and data mining as a nonlinear aggregation tool.

4.1 Choquet integrals

Let X be a nonempty finite set, μ is a signed fuzzy measure defined on $\mathcal{P}(X)$.

Choquet integral was introduced by Choquet in potential theory with the concept of capacity. Then, it has been used for utility theory in the field of economic theory, and has been used for image processing and recognition. The Choquet integral of a nonnegative function defined on $A \subseteq X$ with respect to a fuzzy measure was proposed by Murofushi and Sugeno. It is defined as

$$(c)\int_A f d\mu = \int_0^\infty \mu(F_\alpha \cap A) d\alpha,$$

where $F_\alpha = \{x \in X | f(x) \geq \alpha\}$.

The Choquet integral of a real-valued function f on domain A ($A \subseteq X$) with respect to μ, $\int_A f d\mu$, is defined by the formula

$$\int_A f d\mu = \int_{-\infty}^0 \mu(F_\alpha \cap A) - \mu(A)] d\alpha + \int_0^\infty \mu(F_\alpha \cap A) d\alpha,$$

when not both terms in the right side are infinite.

If $A=X$, then the formula is reduced to be:

$$\int f d\mu = \int_{-\infty}^0 \mu(F_\alpha) - \mu(X)] d\alpha + \int_0^\infty \mu(F_\alpha) d\alpha$$

if at least one term on the right-hand side is finite. Function f is called integrable on A if $\int_A f d\mu < \infty$.

The Choquet integral possesses the following basic properties:

(1) if $\mu(A) = 0$, then $\int_A f d\mu = 0$; $(f(x) \in (-\infty, \infty), \forall x \in X; \mu(A) \in (-\infty, \infty), \forall A \subseteq X)$

(2) $\mu\{x : f(x) \neq 0\} \cap A = \varnothing \Rightarrow \int_A f du = 0$;

$(f(x) \in [0, \infty), \forall x \in X; \mu(A) \in (-\infty, \infty), \forall A \subseteq X)$,

(3) $\int_A c d\mu = c \cdot \mu(A)$; $(f(x) \in (-\infty, \infty), \forall x \in X; \mu(A) \in (-\infty, \infty), \forall A \subseteq X)$

(4) if $f \leq g$, then $\int_A f d\mu \leq \int_A g d\mu$; $(f(x) \in (-\infty, \infty), \forall x \in X; \mu(A) \in [0, \infty), \forall A \subseteq X)$

(5) $\int_A f d\mu = \int f \cdot x_A d\mu$; $(f(x) \in (-\infty, \infty), \forall x \in X; \mu(A) \in (-\infty, \infty), \forall A \subseteq X)$

(6) $\int_A (f + c) d\mu = \int_A f d\mu + c \cdot \mu(A)$;

$(f(x) \in (-\infty, \infty), \forall x \in X; \mu(A) \in (-\infty, \infty), \forall A \subseteq X)$

(7) $\int_A c \cdot f d\mu = c \cdot \int_A f d\mu$ for any $c \geq 0$;

$(f(x) \in (-\infty, \infty), \forall x \in X; \mu(A) \in (-\infty, \infty), \forall A \subseteq X)$

(8) $\int (-f) d\mu = -\int f d\mu'$; $(f(x) \in (-\infty, \infty), \forall x \in X; \mu(A) \in (-\infty, \infty), \forall A \subseteq X)$

where c is a constant, f and g are measurable functions on $(X, \mathscr{F})$, and μ' is the dual of μ (see Section 3.2).

A notable property of the Choquet integral is that $\int (f + g) d\mu \neq \int_A f d\mu + \int_A g d\mu$ in general. This means that the Choquet integral is a nonlinear functional.

To calculate the value of the Choquet integral of a given function f, the values of f, $\{f(x_1), f(x_2), \cdots, f(x_n)\}$, should be first rearranged into a nondecreasing order, that is,

$$f(X_1^*) \le f(x_2^*) \le \cdots \le f(x_n^*),$$

where $(x_1^*, x_2^*, \cdots x_n^*)$ is a permutation of $(x_1, x_2, \cdots, x_n)$. Then, the value of the Choquet integral can be obtained via expression

$$\int f d\mu = \sum [f(x_i^*) - f(x_{i-1}^*)] \cdot \mu(\{x_i^*, x_{i+1}^*, \cdots, x_n^*\})$$

where $f(x_0^*) = 0$ [7].

In case μ is additive, the Choquet integral coincides with the Lebesgue-like integral. So the Choquet integral is a generalization of the Lebesgue-like integral, and we may use the same notation for them. The definition of the Lebesgue-like integral fails when the signed measure is replaced by a signed fuzzy measure. Instead, we should use the Choquet integral or some other appropriate nonlinear integrals[3].

Generally, due to the nonadditivity of μ , the Choquet integral is not linear, that is,

$$\int (f+g) d\mu \ne \int f d\mu + \int g d\mu$$

for some function f and g. However, we have

$$\int af d\mu = a \int f d\mu$$

and

$$\int (f+c) d\mu = \int f d\mu + c \cdot \mu(X)$$

for any nonnegative constant α, constant c, and function f.

This means that the Choquet integral is "marginally linear" though it is nonlinear globally. More properties of the Choquet integral with respect to a nonadditive measure can be found in[4-6].

4.2. Wang's Formula

To be convenient, Wang[8] has proposed a new scheme to calculate the value of a Choquet integral with real-valued integrand by =the inner product of two (2^n-1)-dimensional vectors as

$$\int f d\mu = \sum_{j=1}^{2^n-1} z_j \mu_j , \tag{1}$$

in which

$$z_j = \begin{cases} \min\limits_{i:frc(\frac{j}{2^i})\in[\frac{1}{2},1)} f(x_i) - \max\limits_{i:frc(\frac{j}{2^i})\in[0,\frac{1}{2})} f(x_i), & \text{if it is} > 0 \text{ or } j = 2^n - 1 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

for $j = 1,2,\cdots,2^n - 1$

with a convention that the maximum on the empty set is zero, where $\mathrm{frc}(\frac{j}{2^i})$ is the fractional part of $\frac{j}{2^i}$. In the above formula, if we express j in the binary form $j_n j_{n-1}\cdots j_1$, then $\{ i \mid \mathrm{frc}(\frac{j}{2^i}) \in [\frac{1}{2},1)\} = \{i \mid j_i = 1 \}$ and $\{ i \mid \mathrm{frc}(\frac{j}{2^i}) \in [0,\frac{1}{2})\} = \{i \mid j_i = 0 \}$.

This formula is equivalent to the following scheme[9]:

Let T: $\mathcal{P}(X) \to N$ with $T(A) = \sum_{x_j \in A} 2^{j-1}$. Then

$$C_\mu(f) = \sum_{i=0}^{2^n-1} z_i \mu_i , \tag{3}$$

where

$$z_j = \begin{cases} f(x_j') - f(x_{j-1}') & \text{if } i = T(\{x_j', \cdots, x_n'\}) \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

and $f(x_0') = 0$.

Obviously, the latter calculation scheme is faster than the one provided by Wang if the computation is conducted by hand. However, using a vector multiplication in Wang's formula is more convenient for computation by computer. Furthermore, the purpose of this paper is to provide a model of nonlinear multiregressions based on Choquet integral for data with both numerical and categorical attributes, where the values of singed fuzzy measure μ are a part of unknown parameters. The Wang's formula can directly show the values of z_j's that are entries of the coefficient matrix of a system of linear equations with unknown variable μ_j's. So, the Wang's formula is very convenient for using an algebraic method to estimate as many regression coefficients as possible, and therefore, to reduce the complicity of a genetic algorithm (or an alternative search method) that we have to use for estimating the other regression or classifier coefficients[8,10,11].

CHAPTER 5

MULTIREGRESSION MODELS

5.1 An overview of classical Linear Multiregressions

The model of classical linear multiregressions is used to describe the relation between the objective attribute and several predictive attributes in database. It requires a basic assumption of a group of predictive attributes to the objective attribute is just the sum of the respective contributions of each predictive attributes cannot be ignored in many real-world problems. Such an interaction is an inherent joint potential from the predictive attributes toward the objective attribute and is totally different from the statistical concept of correlation that describes the relation among the appearance of the observed values from the productive attributes. The linear multiregression model fails due to the existence o a strong interaction among predictive attributes.

To overcome the limitation of the model of linear multiregression, people introduced some nonlinear models such as the quadratic model[12] and the projection pursuit[13]. The former uses a quadratic function to replace the linear combination in the classical linear multiregression, and it can be only applied to describe the nonlinear of a very special case. While the latter takes a sum of one-dimensional nonlinear transformations on several linear combinations of the predictive attributes, then determine the unknown parameters via a neural network, and it can only describe (with a high complexity) some kinds of nonlinearity indirectly. Both of these models can describe only limited kinds of the nonlinear contribution from predictive attributes toward the objective attribute in a database.

5.2 Linear multiregressions expressed by the lebesgue like integral

Let X=$\{ x_1, x_2, ..., x_n \}$ be the set of predictive attributes (variables) and Y be the objective attribute (variable). A linear multiregression model has a form of

$$Y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n + N(0, \sigma^2),$$

where $N(0, \sigma^2)$ is a normally distributed variable with mean 0 and variance and $a_0, a_1, a_2, \cdots, a_n$ are regression coefficients that are not necessarily nonnegative. Denote the class of the empty set $\varnothing$ and all singletons $\{\varnothing, \{x_1\}, \{x_2\}, \ldots \{x_n\}\}$ by $\mathscr{S}$. Class $\mathscr{S}$ is a semi-ring. If we define a set function $\lambda : \mathscr{S} \to (-\infty, \infty)$ by $\lambda(\{x_i\}) = a_i$, $i = 1,2,\cdots,n$, and $\lambda(\varnothing)$=0 on semi-ring $\mathscr{S}$, then λ can be uniquely extended from $\mathscr{S}$ to be an additive signed measure on $\mathscr{P}(X)$, the power set of X[4,9]. The additivity of λ means that

$$\lambda(A \cup B) = \lambda(A) + \lambda(B),$$

$\forall A, B \in \mathscr{P}(X)$ with $A \cap B = \varnothing$.

Let the observation of predictive attributes $x_1, x_2, ..., x_n$ be $f(x_1), f(x_2), \ldots, f(x_n)$ respectively. The observation can be regarded as a measurable function f defined on measurable space (X, $\mathscr{P}(X)$). Thus, the linear multiregression model can be expressed as

$$Y = a_0 + \int f d\lambda + N(0, \sigma^2).$$

where $\int f d\lambda$ is the Lebesgue-like integral of f with respect to λ [14]. In such a linear model, for any given set $A \subseteq X$, $\lambda(A)$ can be regarded as the global contribution rate from all predictive attributes in A toward the objective attribute. The Lebesgue-like integral is an aggregation tool that summarizes the contribution from all predictive

attributes toward the objective attribute according to $f = (f(x_1), f(x_2), \ldots, f(x_n))$ and λ linearly. However, to use this model, we need a basic assumption that there is no interaction among predictive attributes toward the objective attribute. This means that the global contribution from a set of the predictive attributes toward the objective attribute is just the sum of individual contributes from each predictive attribute in the set alone. Here, the interaction is totally different from the concept of correlationship in statistics. The latter does not concern the objective attribute, and can be estimated by only using the observations of the predictive attributes.

5.3 Multiregressions Based on the Generalized Choquet Integral

The new non-linear multiregression model now is expressed as

$$Y = c + \int (a + bf)\, d\mu + \varepsilon$$

where c is a constant, both a and b are real-valued functions defined on X, f is an observation of attributes in X, μ is a signed fuzzy measure, and $\varepsilon \sim N(0, \sigma^2)$, that is, ε is a normally distributed random variable with mean 0 and variance σ^2. Functions a, b can be expressed as vectors, i.e., $a = (a_1, a_2, \ldots, a_n)$ and $b = (b_1, b_2, \ldots, b_n)$. They should satisfy the constraints that a_i is nonnegative for $i=1,2,\ldots,n$ with $\min_i a_i = 0$, $b_i \in [-1, 1]$ for $i=1, 2,\ldots,n$ with $\max_i b_i = 1$.

It is a nonlinear multiregression model since the Choquet integral is nonlinear with respect to its integrand. In a special case that the signed fuzzy measure is a classical signed measure, this model coincides with the linear multiregression model. So, the new model is a generalization of the classical one.

In this new model, constants c and q, vectors a and b, and set function μ are regarded as unknown regression coefficients. They should be optimally estimated once a proper dataset of all predictive attributes and the objective attribute is available. Meanwhile, σ^2 is also estimated as the regression error.

Determining the regression coefficients in this model is the inverse problem of calculating the Choquet integral. Unfortunately, there is no classical mathematical method that can be used for this purpose. However, this inverse problem can be induced as a nonlinear optimization problem. Regarded as a large-scale search, it can be realized by soft computing techniques (running special genetic algorithms or training neural networks) when a proper sample dataset is available. The procedure of the search can also be regarded as a nonlinear integral projection pursuit.

The Choquet integral of nonnegative measurable function with respect to a fuzzy measure has a common property that $\int_A f d\mu = 0$ if and only if $\mu(\{x|f(x) > 0\} \cap A) = 0$ [4,15,16]. Unfortunately, the generalized Choquet integral does not have such a property since set function μ and integral f loses the nonnegativity in general.

The most important aggregation tool especially in information fusion is the integral attributes. People use nonadditive set functions to describe the interaction and define some integrals with respect to them as new aggregation tools used for information fusion [2,17-19]. In real problems, various backgrounds require a variety of integrals. The weighted average method is the simplest one and can be regarded as a linear integral (the Lebesgue integral (see Appendix C)). However, due to the nonadditivity of the concerned set functions such as fuzzy measures, which is used to

describe the interaction among attributes, the integral is generally nonlinear. In the last twenty years, the theoretical as well as practical significance of nonadditive set functions and nonlinear integrals has increasingly been recognized. The Choquet integral with respect to nonadditive monotone set functions is one kind of nonlinear functionals defined on a subspace of all real valued measurable functions. Unlike the fuzzy integral, which uses the maximum and minimum operators, the Choquet integral is defined via the common addition and multiplication and, therefore, it is a generalization of the classical Lebesgue integral.

CHAPTER 6

MULTIREGRESSIONS WITH BOTH NUMERICAL AND CATEGORICAL ATTRIBUTES

Based on generalized Choquet integrals with respect to signed fuzzy measures, a model of nonlinear multiregression can be established. A numericalization technique is adopted to project each state of a categorical attribute that has more than two states to a multi-dimensional space optimally through a genetic algorithm with the algebraic method.

6.1 THE MODEL OF REGRESSIONS

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of concerned predictive (explanatory) attributes and Y be the objective (response) attribute in a given database. Suppose that all concerned attributes are numerical, and the available data corresponding to these attributes are complete and have the following form:

x_1	x_2	$\cdots$	x_n	Y
f_{11}	f_{12}	$\cdots$	f_{1n}	Y_1
f_{21}	f_{22}	$\cdots$	f_{2n}	Y_2
$\vdots$				$\vdots$
f_{l1}	f_{l2}		f_{ln}	Y_l

where f_{ji} is the j-th real-valued observation of x_i, Y_j is the j-th real-valued observation of Y. l is the size of the data, and should be much larger than n, usually, at least 5 times of 2^n. Each observation of $x_1, x_2, \ldots, x_n$ can be regarded as a function $f: X \to (-\infty, \infty)$. Thus, the j-th observation of $x_1, x_2, \ldots, x_n$ is denoted by f_j.

The interaction among predictive attributes toward the objective attribute is described by a set function μ defined on the power set of X satisfying the condition of vanishing at the empty set, i.e., $\mu : \mathcal{P}(X) \rightarrow (-\infty, \infty)$ with $\mu(\varnothing) = 0$. To be convenient, we also require that $\mu(X)$ is nonnegative. Set function μ is called a signed fuzzy measure.

The new non-linear multiregression model is the one as we mentioned in Section 5.3. In this multiregression model, the regression coefficients are constant c, all elements of vectors a and b, and $\mu(A)$ for every $A \in \mathcal{P}(X) - \{\varnothing\}$. These regression coefficients may be optimally determined by using a genetic algorithm and the least square method based on the above-mentioned data. Finally, σ^2 may be estimated by the regression residual.

6.2 CATEGORICAL AND NUMERICAL ATTRIBUTES

In databases, there may be two types of attributes. One is both numerical and continuous attribute, whose values are assumed to be in some interval of real numbers. Another is categorical (discrete) attribute, which has finitely many different states.

In addition to explore the relationship among categorical attributes or quantitative attributes respectively, people discovered the pattern embedded in the mixture attributes of both categorical and numerical attributes recently. They developed an evolutionary numericalization technique to convert categorical attributes to be numerical such that a nonlinear multiregression model can accommodate both categorical and numerical data. But in this pattern, a categorical attribute was simply projected to a one dimensional space, say, a line, it may be rude an inappropriate for

some categorical attributes when each of them has more than two possible states. To improve this situation, we develop a new pattern by projecting a categorical attribute to a (K–1)-dimensional space when this attributes has K possible states.

6.3 A NEW MODEL ACCOMMODATING BOTH CATEGORICAL AND NUMERICAL ATTRIBUTES

Let $X = \{x_1, x_2, \ldots, x_m, x_{m+1}, x_{m+2}, \ldots, x_n\}$ be the set of all considered predictive attributes, in which $x_1, \ldots, x_m$ are numerical and $x_{m+1}, \ldots, x_n$ are categorical. The set of all possible states of categorical attribute x_i is denoted by $S_i = \{s_{i1}, s_{i2}, \ldots, s_{iK_i}\}$ and is called the range of x_i, where K_i is the number of possible states of attribute x_i and is called the potential of x_i, $i = m + 1, m + 2, \ldots, n$. In each S_i, $i = m + 1, m + 2, \ldots, n$, each state s_{ik}, $k = 1, 2, \ldots, K_i$. Similar to the old function f, f^* denotes a function defined on X. It has a nonnegative real value at each attributes x_i for $i = 1, 2, \ldots, m$, and has a value in (K_i -1)-dimensional space at attribute x_i for $i = m + 1, m + 2, \ldots, n$. Each f_j^* in the data is an observation of such a function. Now let's numericalize attributes $x_{m+1}, \ldots, x_n$. Our numericalization strategy is, for each $i = m + 1, m + 2, \ldots, n$, projecting x_i to a (K_i –1)-dimensional space, say, ($x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(K_i-1)}$). After the projection, we denote X^* to the new set of predictive attributes then optimally assigning a real value to each state $x_i^{(k)}$, $k = 1, 2, \ldots, K_{i-1}$. The optimization is in the sense that, after replacing these states with corresponding real-valued assignments respectively, the regression

$$Y = c + \int (a + bf^*)\, d\mu + \varepsilon$$

fits the data as well as possible. This optimization procedure takes place with optimizing regression coefficients together in a genetic algorithm. Thus, for each s_{ik},

we use a gene to represent it and align all of them in the chromosome. The corresponding value of this gene is denoted by a_{ik} that is the value we want to assign to s_{ik}. For a fixed i, different values a_{ik} indicate the different influences of s_{ik} to the objective attribute, and they should be regularized when being used as the value of function f^* at attribute x_i to calculate the integral of f^*. The regularization is made as follows:

$$a_{ik}^* = \frac{a_{ik}}{\sum_{k=1}^{N} a_{ik}}$$

for $i = m + 1, m + 2, \ldots, n$. In comparison with the model for pure numerical attributes given in section 3, now the number of unknown parameters is increased. In our current model, each chromosome consists of

$$2^n - 2 + n + \sum_{i=m+1}^{n} N_i$$

genes ($2^n - 2$ genes for the values of the set function, n genes for the weights, and N_i genes for the states of x_i, $i = m+1, m+2, \ldots, n$), and there are

$$(2^n - 2) + 2 + (n-1) + \sum_{i=m+1}^{n} (N_i - 1) = 2^n + m - 1 + \sum_{i=m+1}^{n} N_i$$

unknown independent parameters to be determined from data by minimizing the error e:

$$e = \sqrt{\frac{1}{l}\sum_{j=1}^{l}(Y_j - c + \int(a + bf_j)\, d\mu)^2}\ .$$

6.4. ALGORITHM

With respect to multiregressions, an algebraic approach to finding has been recently introduced [8]. The algebraic approach provides a huge savings in the running time of the algorithm because the size of the chromosome in the genetic algorithm is greatly reduced.

A. Preparations

(1) For given n, express positive integer k in binary digits as $k_n k_{n-1} \ldots k_1$ for every k = 1,2,..., 2^n-1

(2) Use μ_k to denote $\mu(A)$ where $A = \bigcup_{K_i=1}\{x_i\}$, k = 1,2,..., 2^n-1.

B. Part 1

Use the least square method to determine $c, \mu_1, \mu_2, \ldots, \mu_{2^n-1}$ when the values of all elements of a and b are specified in the genetic algorithm given in Part 2.

(1) Construct the $l \times 2^n+1$ augmented matrix $Z=[z_{jk}]$ as follows.

$z_{j0} = 1$,

$$z_{jk} = \begin{cases} \min(a_i + b_i f_{ji}^*) - \max(a_i + b_i f_{ji}^*), \text{ if it is positive or } k = 2^n - 1 \\ 0, \qquad \text{otherwise} \end{cases}$$

$z_{j2^n} = y_j$,

for $k = 1,2,\ldots, 2^n-1$ and $j = 1,2,\ldots,l$.

(2) Find the least square solution of the system of linear equations having above augmented matrix for unknown variables $c, \mu_1, \mu_2, \ldots, \mu_{2^n-1}$.

(3) Calculate the regression residual error for $\hat{\sigma}^2$ by

$$\hat{\sigma}^2 = \frac{1}{l}\sum_{j=1}^{l}\left[y_j - c - \int(a + bf_j^*)d\mu\right]^2 = \frac{1}{l}\sum_{j=1}^{l}(y_j - c - \sum_{k=1}^{2^n-1} z_{jk}\mu_k)^2.$$

C. Part 2 (main algorithm)

Use a genetic algorithm to optimize the values of vectors a and b.

(1) Input integers n, l and the data.

(2) Choose a large prime s as the seed for the random number generator. Set the value for each parameter listed in the following.

λ: The bit length of each gene, i.e., λ bits are used for expression each gene. It depends on the required precision of the results.

p: The population size. It should be a large positive even integer. Its default is 200.

α and β: The probabilities used in a random switch to control the choice of genetic operators for producing offspring from selected parents. They should satisfy the condition that $\alpha \geq 0, \beta \geq 0$, and $\alpha + \beta \leq 1$. Their defaults are 0.2 and 0.5 respectively.

ε and δ: Small positive numbers used in the stopping controller. Their defaults are 10^{-6} and 10^{-10} respectively.

w: The limit number of generations that have not significant progression successively. Its default is 10.

(3) Calculate

$$\hat{\sigma}_y^2 = \frac{1}{l}\sum_{j=1}^{l}(y_j - \bar{y})^2,$$

where

$$\bar{y} = \frac{1}{l}\sum_{j=1}^{l} y_j .$$

(4) Randomly create the initial population that consists of p chromosomes. Each chromosome consists of $2n$ genes, denoted by $g_1, g_2, \cdots, g_n, g_{n+1}, g_{n+2}, \cdots, g_{2n}$. The first n of them represent vector a, while the next n of them represent vector b. Each gene consists of λ bits and represents a number in [0, 1) . Initialize counter GC by p, counter WT by 0, and SE by $\hat{\sigma}^2$.

(5) Decode each chromosome to get vectors a and b by the following formulae.

$$a_i = \frac{g_i - m(g)}{(1-g_i)(1-m(g))},$$

$$b_i = \frac{2g_{n+1} - 1}{M(g)},$$

for $i = 1,2,\ldots,n$, where $m(g) = \min_{1 \le k \le n} g_k$ and

$$M(g) = \max_{1 \le k \le n} |2g_{n+k} - 1|$$

(6) For each chromosome in the population, through algorithm Part 1, use a and b obtained above and the data to determine the corresponding optimal values of c, μ, and find the residual $\hat{\sigma}^2$.

(7) The residual error of the r-th chromosome in the current population is denoted by $\hat{\sigma}^2$. Let $m(\hat{\sigma}^2) = \min_{1 \le k \le n} g_k$ and $R = \{r | \hat{\sigma}^2 = m(\hat{\sigma}^2)\}$. Erasing the record saved for the last generation if any, save and $m(\hat{\sigma}^2)$ and a, b, c, μ of r-th chromosomes for all $r \in R$ in the current population. Display GC, WT, and $m(\hat{\sigma}^2)$.

(8) If $m(\hat{\sigma}^2) < \varepsilon \hat{\sigma}_y^2$, then go to (16); otherwise, take the next step.

(9) If $SE - m(\hat{\sigma}^2) < \delta \hat{\sigma}^2$, then $WT + 1 \Rightarrow WT$ and take

(10) the next step; otherwise $0 \Rightarrow WT$, and go to (11).

(11) If $WT > w$, then go to (16); otherwise, take the next step.

(12) The relative goodness of the r-th chromosome in the current population is defined by $G_r = \frac{m(\hat{\sigma}^2)}{\hat{\sigma}_r^2}, r = 1,2,\ldots,p$,

if $m(\hat{\sigma}^2) > 0$.

(13) Let $p_r = \frac{G_r}{\sum_{r=1}^{p} G_r}, r = 1,2,\ldots,p.$

(14) According to the probability distribution $\{p_r | r = 1,2,\ldots,p\}$ (by using a random switch), select two different chromosomes in the current population as the parents. Randomly select one operator among the three-bit mutation (with probability α), the two-point crossover (with probability), and one of the equally likely 48 two-point realignments (with probability $1-\alpha-\beta$) to produce two new chromosomes as the offspring.

(15) Repeat step (13) for $p/2$ times totally to get p new chromosomes. $GC+p \Rightarrow GC$. Save $m(\hat{\sigma}^2)$ in SE.

(16) For each new chromosome, take step (5) and (6) to find the corresponding values of a, b, c, μ and $\hat{\sigma}^2$. Add these new chromosomes into the current population. According to the magnitude of $\hat{\sigma}^2$ (the smaller the better), select p best chromosomes among these $2p$ chromosomes to form the new population. Then go to (7).

(17) Check the sign of μ_{2^n-1} corresponding the r-th chromosome for all $r \in R$. In case $\mu_{2^n-1} < 0$, replace c by $c+\mu_{2^n-1} \max_{1 \le i \le n} a_i$, then replace a_i by $\max_{1 \le i \le n} a_i - a_i$ and switch the sign of vector b and set function μ such that $\mu_{2^n-1} > 0$. Display $s, p, \lambda, \alpha, \beta, \varepsilon, \delta$, and w. After deleting any duplicates, display a, b, c, and μ of r-th chromosomes for all $r \in R$.

(18) Stop.

6.5 AN EXAMPLE

When buying a pre-owned car, assuming that there are only two predictive attributes: mileage and color. i.e. $X = \{u, v\}$. And the objective attribute is the price. Here, mileage, which is denoted by u, is a numerical attribute while color, which is denoted by v, is a categorical attribute and it has three states: black, white and red. We project this categorical attribute: color, to a two-dimensional space (v_1, v_2).

Now $X^* = \{u, v_1, v_2\}$. Though v_1 and v_2 can be any real values, by adjusting the values of a and b they could always be summarized to 3 expressions: (0,0), (0,1), (1,0). We can denote each expression to a color. Without lost of generality, assume

that (0, 0) means black, (1,0) means white, (0,1) means red. This is shown in Table 6.1.

Table 6.1. the Integrand's values respective to the color

State of Categorical Attributes	Integrand's Value at v_1	Integrand's Value at v_2
black	a_2	a_3
white	$a_2 + b_2$	a_3
red	a_2	$a_3 + b_3$

For this example, n=3, 2^n-1=7, so we express positive integer k in binary digits as $k_3k_2k_1$ for every $k = 1,2,\ldots,7$.

Use μ_k to denote $\mu(A)$ where $A = \bigcup_{k_i=1}\{x_i\}$, $k=1,2,\ldots,7$.

We have successfully used above algorithm for a number of examples. All results have been verified. The following is one of them.

We set the data size $k=100$, ($100>5*2^3$). This is shown in Table 6.2.

Table 6.2. the Integrand's values in this example

No.	Mileage/10000	Color	Price
1	0.2509	black	0.2151
2	0.9136	black	0.5382
3	0.2647	red	0.2565
4	0.3764	black	0.2582
5	0.2674	red	0.2567
6	0.4638	white	0.2964
7	0.5721	black	0.356
8	0.7631	white	0.3263
9	0.6199	red	0.42
10	0.0923	black	0.1992
11	0.1736	white	0.2674
12	0.8681	white	0.3572

13	0.5284	white	0.3028
14	0.2837	red	0.2584
15	0.559	white	0.3059
16	0.5256	red	0.3728
17	0.6921	white	0.3192
18	0.9546	red	0.5873
19	0.2774	black	0.2177
20	0.1513	white	0.2651
21	0.0755	white	0.2576
22	0.0802	white	0.258
23	0.6179	red	0.4189
24	0.188	black	0.2088
25	0.6886	black	0.4143
26	0.4862	red	0.3531
27	0.731	white	0.3231
28	0.3332	black	0.2366
29	0.9394	white	0.3936
30	0.9365	white	0.3919
31	0.0619	white	0.2562
32	0.3083	red	0.2641
33	0.9013	black	0.5308
34	0.7107	black	0.4253
35	0.5938	red	0.4069
36	0.7234	red	0.4717
37	0.4771	red	0.3486
38	0.3169	black	0.2284
39	0.6469	white	0.3147
40	0.3328	white	0.2833
41	0.7852	red	0.5026
42	0.1978	white	0.2698
43	0.1733	white	0.2673
44	0.7755	black	0.4577
45	0.1479	red	0.2448
46	0.6111	black	0.3755
47	0.174	red	0.2474
48	0.9637	black	0.5682
49	0.1822	red	0.2482
50	0.254	red	0.2554
51	0.5364	red	0.3782
52	0.4452	black	0.2926
53	0.781	red	0.5005
54	0.1086	white	0.2609
55	0.267	white	0.2767
56	0.7633	black	0.4517
57	0.2603	red	0.256
58	0.5683	white	0.3068
59	0.4921	white	0.2992
60	0.5463	white	0.3046
61	0.0557	black	0.1956
62	0.8498	black	0.4999
63	0.0767	red	0.2377
64	0.3051	black	0.2225
65	0.488	white	0.2988
66	0.0348	black	0.1935
67	0.5364	white	0.3036
68	0.038	black	0.1938
69	0.2421	black	0.2142
70	0.7452	white	0.3245
71	0.5879	white	0.3088
72	0.5613	red	0.3907
73	0.4697	black	0.3048
74	0.2817	red	0.2582
75	0.4735	red	0.3467
76	0.6018	white	0.3102
77	0.9853	white	0.4212
78	0.759	black	0.4495
79	0.2652	black	0.2165
80	0.3208	red	0.2704

81	0.6231	white	0.3123
82	0.7776	black	0.4588
83	0.8356	red	0.5278
84	0.477	black	0.3085
85	0.9621	black	0.5673
86	0.5868	white	0.3087
87	0.7308	black	0.4354
88	0.3981	black	0.2691
89	0.2778	red	0.2578
90	0.1622	red	0.2462
91	0.5078	red	0.3639
92	0.2804	white	0.278
93	0.2155	black	0.2116
94	0.5551	white	0.3055
95	0.1035	red	0.2403
96	0.2338	red	0.2534
97	0.2314	black	0.2131
98	0.8862	black	0.5217
99	0.8269	white	0.3408
100	0.1447	red	0.2445

After running the program, we obtain the results in table 6.3:

Table 6.3. the obtained values compared to the real values

	Real value	Obtained value
μ_1	0.6	0.5874967700443454
μ_2	0.1	0.0
μ_3	0.05	0.0
μ_4	0.2	0.2015611505170548
μ_5	0.7	0.7085416696095773
μ_6	0.3	0.31478312792546953
μ_7	0.4	0.4150728976884508
a_1	0.0	0.0
a_2	0.3	0.04481574775580864
a_3	0.8	0.06015626529852938
b_1	1.0	1.0
b_2	0.6	0.8692660550458715
b_3	0.2	0.9747706422018348
c	0.0	-0.013157917141883932

It is possible that the solution to a problem is not unique. In fact, we formed the problem by using one of its solutions which does not imply it is the only solution. That's why the values of a and b charged and the values of μ are similar to the original.

We have presented an algebraic method involving a genetic algorithm to find the values of all the regression coefficients. And we have seen that the calculation is not complex. The experimental results have been compared with the given values. The comparisons show that the proposed genetic approach is an effective and applicable way to retrieve the values of regression coefficients.

CHAPTER 7

CONCLUSION

Generalized fuzzy measure restricts its values to be nonnegative, which may raise limitations in practical applications. It is not reasonable to assume that all values of fuzzy measure are nonnegative before we elicit them from real data. Sometimes, negative values may exist, such as in pharmacological, financial, and sociological applications. Thus, a more generalized fuzzy measure, the signed fuzzy measure, should be considered and investigated. The signed fuzzy measure may assume some negative values and, therefore, may be mot monotonic.

To deal with the nonlinear multiregression based on Choquet integral for data with both numerical and categorical attributes, we use our new numericalization technique to optimally project the states of the categorical attributes into a partial ordering space instead of a total ordering space. It is useful in many real problems such as regression, classification and network structure optimization, where both types of data are involved and some of the categorical attributes have more than two possible states.

With respect to multiregressions, an algebraic approach to find the values of regression coefficients has been recently introduced. The algebraic approach provides a huge savings in the running time of the algorithm because the size of the chromosome in the genetic algorithm is greatly reduced. The smaller the chromosome, the quicker the genetic algorithm will converge. Our algorithm is based on this work.

APPENDIX A

Summary of Variables

n	number of attributes
l	number of observations
s	seed number (a large prime)
p	population size
w	generation limiter
λ	number of bits per gene
α, β	probabilities used in determining genetic operator
ε, δ	small numbers used as stopping conditions
$\hat{y}_j$	estimate of y_j
$\bar{y}$	average of objective attributes
$\hat{\sigma}_y^2$	residual of objective attributes
q	vector of objective attributes
GC	number of generated chromosomes
WT	tracks variable w
SE	saved error
a , b	functions used in multiregression (they are vectors)
e_0	initial error
$m(\sigma^2)$	smallest residual error
μ_k	denotes $\mu(A)$ where $A = \bigcup_{k_i=1} \{x_i\}$

$\mathcal{U}_1$

$\mathcal{U}_2$

$\mathcal{U}_{12}$

APPENDIX B

C++ Code for Example 6.1

```
#include <stdio.h>
#include <stdlib.h>
#include <iomanip.h>
#include <time.h>
#include <math.h>
#include <iostream.h>
#include <fstream.h>

#define A1 = 0.0
#define A2 = 0.3
#define A3 = 0.8
#define B1 = 1.0
#define B2 = 0.6
#define B3 = 0.2
#define C = 0
#define MU1 0.6
#define MU2 0.1
#define MU3 0.5
#define MU4 0.2
```

```
#define MU5 0.7
#define MU6 0.3
#define MU7 0.4

int min(double x1, double x2, double x3)
{
        double min = 10;

        if (x1 < min) min = x1;
        if (x2 < min) min = x2;
        if (x3 < min) min = x3;

        if (x1 == min) return 1;
        if (x2 == min) return 2;
        if (x3 == min) return 3;

        return -1;
}

int main() {
        srand(time(NULL));
        ofstream outFile ("test.csv", ios::out);
        ofstream outFile2 ("test_categorical.csv", ios::out);

        if(!outFile)
        {
```

```
        cerr << "File could not be opened" << endl;
        exit(1);
}

double x1=0.0;
double x2=0.0;
double x3=0.0;
double y = 0.0;

double choice = 0.0; // black, white or red

int count = 0;

while(count<100)
{
        x1 = (1.0*rand())/(RAND_MAX + 1.0); // between 0 and 1
        choice = (3.0*rand())/(RAND_MAX + 1.0); // black, while or red

        if (choice < 1) // black
        {
                x2 = 0.3;
                x3 = 0.8;
        }
        else if (choice < 2) // white
        {
                x2 = 0.9;
```

```
        x3 = 0.8;
}
else if (choice < 3) // red
{
        x2 = 0.3;
        x3 = 1.0;
}

// determine y value
if (min(x1, x2, x3) == 1)
{
        if (x2 < x3)
        y = x1 * MU7 + (x2 - x1) * MU6 + (x3 - x2) * MU4;

        else
        y = x1 * MU7 + (x3 - x1) * MU6 + (x2 - x3) * MU2;
}
else if (min(x1, x2, x3) == 2)
{
        if (x1 < x3)'
        y = x2 * MU7 + (x1 - x2) * MU5 + (x3 - x1) * MU4;

        else
        y = x2 * MU7 + (x3 - x2) * MU5 + (x1 - x3) * MU1;
}
else if (min(x1, x2, x3) == 3)
```

```
{
        if (x1 < x2)
        y = x3 * MU7 + (x1 - x3) * MU3 + (x2 - x1) * MU2;

        else
        y = x3 * MU7 + (x2 - x3) * MU3 + (x1 - x2) * MU1;
}
else cout << "ERROR!!!" << endl;


if (choice < 1)
{
        outFile << setprecision(4) << setiosflags(ios::fixed | ios::showpoint)
<< x1 << "," << setprecision(4) << setiosflags(ios::fixed | ios::showpoint) << "0,0," << y <<
endl;
        outFile2 << setprecision(4) << setiosflags(ios::fixed | ios::showpoint)
<< x1 << "," << setprecision(4) << setiosflags(ios::fixed | ios::showpoint) << "black," << y
<< endl;
}
else if (choice < 2)
{
        outFile << setprecision(4) << setiosflags(ios::fixed | ios::showpoint)
<< x1 << "," << setprecision(4) << setiosflags(ios::fixed | ios::showpoint) << "1,0," << y <<
endl;
```

```
                    outFile2 << setprecision(4) << setiosflags(ios::fixed | ios::showpoint)
<< x1 << "," << setprecision(4) << setiosflags(ios::fixed | ios::showpoint) << "white," << y
<< endl;
                }
                else
                {
                    outFile << setprecision(4) << setiosflags(ios::fixed | ios::showpoint)
<< x1 << "," << setprecision(4) << setiosflags(ios::fixed | ios::showpoint) << "0,1," << y <<
endl;
                    outFile2 << setprecision(4) << setiosflags(ios::fixed | ios::showpoint)
<< x1 << "," << setprecision(4) << setiosflags(ios::fixed | ios::showpoint) << "red," << y <<
endl;
                }
                count++;
        }
        return 0;
}
```

Note:

For error, $\hat{\sigma}^2 = 7.54350627949288E-7$

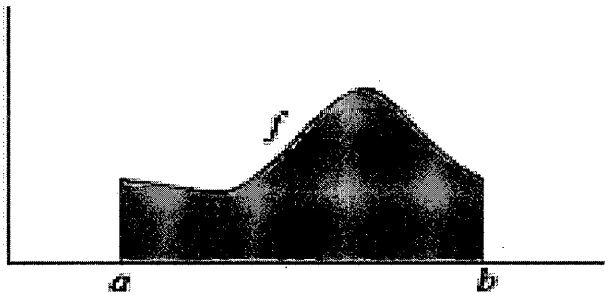
Population size=1000

Total generation=267

running time=5.69min

APPENDIX C

The Lebesgue Integral



The Lebesgue integral can be interpreted as the area under a curve.

The Lebesgue integration has long been understood that for functions with a smooth enough graph (such as continuous functions on closed bounded intervals), the area under the curve could be defined and computed using techniques of approximation of the region by polygons.

The Lebesgue integral is named for Henri Lebesgue (1875-1941). In Lebesgue's theory, integrals are limited to a class of functions called measurable functions. A function f is measurable if the pre-image of any closed interval is in X:

$$f^{-1}([a,b]) \in X$$

It can be shown that this is equivalent to requiring that the pre-image of any Borel subset of R be in X. We will make this assumption from now on. The set of measurable functions are closed under algebraic operations, but more importantly the class is closed under various kinds of pointwise sequential limits:

$$\liminf_{i \in N} f_i, \quad \limsup_{i \in N} f_i$$

are measurable if the original sequence $\{f_i\}_{i \in N}$ consists of measurable functions.

We build up an integral

$$\int_X f d\mu$$

for measurable real-valued functions f defined on E in stages:

Indicator functions: To assign a value to the integral of the indicator function of a measurable set S consistent with the given measure μ, the only reasonable choice is to set:

$$\int_S d\mu = \mu(S)$$

Simple functions: We extend by linearity to the linear span of indicator functions:

$$\int_{S_i} (\sum a_i) d\mu = \sum a_i \mu(S_i)$$

where the sum is finite and the coefficients ak are real numbers. Such a finite linear combination of indicator functions is called a simple function. Note that a simple function can be written in many ways as a linear combination of characteristic functions, but the integral will always be the same.

Non-negative functions: Let f be a non-negative measurable function on E which we allow to attain the value $+\infty$, in other words, f takes values in the extended real number line. We define

$$\int_E f d\mu := \sup\left\{ \int_E s d\mu : s \le f, s \text{ simple } \right\}$$

We need to show this integral coincides with the preceding one, defined on the set of simple functions. There is also the question of whether this corresponds in any way to a Riemann notion of integration. It is not hard to prove that the answer to both questions is yes.

We have defined the integral of f for any non-negative extended real-values measurable function on E. For some functions the integral will be infinite.

Signed functions: To handle signed functions, we need a few more definitions. If f is a function of the measurable set E to the reals (including $\pm\ \infty$), then we can write

$$f(x) = f^+(x) - f^-(x),$$

where

$$f^+(x) = \begin{cases} f(x) \text{ if } f(x) \geq 0 \\ 0 \quad \text{otherwise} \end{cases}$$

$$f^-(x) = \begin{cases} -f(x) \quad \text{if } f(x) < 0 \\ 0 \quad\quad\ \text{otherwise} \end{cases}$$

Note that both $f^+(x)$ and $f^-(x)$ are non-negative functions. Also note that

$$|f(x)| = f^+(x) + f^-(x)\ .$$

If

$$\int |f(x)| d\mu < \infty,$$

then f is called Lebesgue integrable. In this case, both integrals satisfy

$$\int f^+(x) d\mu < \infty,\ \int f^-(x) d\mu < \infty,$$

and it makes sense to define

$$\int f d\mu = \int f^+(x) d\mu - \int f^-(x) d\mu$$

This definition gives the desirable properties of the integral. Complex valued functions can be similarly integrated, by considering the real part and the imaginary part separately.

REFERENCES

1. Z. Wang,"The autocontinuity of set function and the fuzzy integral", Journal of Mathematical Analysis and applications, Vol.99, pp.195-218,1984.
2. Z. Wang and G.J.Klir. Fuzzy Measure Theory Plenum, New York, 1992.
3. Z. Wang, K.S. Leung, M. L. Wong, and J. Fang. "A new type of nonlinear integrals and the computational algorithm", Fuzzy sets and systems, Vol.112, pp.223-231, 2000.
4. L.M.de Campos and M.J. Bolanos. "Characterization and comparison of Sugeno and Choquet integrals", Fuzzy sets and Systems, Vol. 52, pp.61-67, 1992.
5. T. Murofushi, M. Sugeno, and M. Machida. "Non-monotonic fuzzy measures and the Choquet integral", Fuzzy Sets and Systems, Vol.64, pp.73-86, 1994.
6. Z.Wang and G.J.Klir. "Choquet integrals and natural extensions of lower probabilities", International Journal of Approximate Reasoning, Vol.16, pp.137-147, 1997.
7. Z. Wang, K.S.Leung, M.L.Wong, J.Fang, and K.Xu. "Nonlinear nonnegative multi-regressions based on Choquet integrals", International Journal of Approximate Reasoning, Vol. 25, pp. 71-87, 2000.
8. Z. Wang. "A New Genetic Algorithm for Nonlinear Multiregressions Based on Generalized Choquet Integrals", Proc. FUZZ-IEEE, Int. Conf. Fuzzy Syst., pp. 819-921, 2003.
9. M.Grabisch, H.T.Nguyen and E.A.Walker. "Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference, Kluwer Academic,1995

10. K.S.Leung, M.L.Wong, W.Lam, Z.Wang and K.Xu. "Learning Nonlinear Multiregression Networks Based on Evolutionary Computation", IEEE Trans, On Systems, Man and Cybernetics Part B, Vol.32, pp.630-644, 2002

11. K.Xu, Z.Wang and K.S.Leung. Classification by Nonlinear Integral Projections, IEEE Trans. Fuzzy Systems, Vol.11, 187-201

12. W.Mendenhall, and T.Sincich. Statistics for Engineering and the Science, Prentice Hall, Englewood Cliffs, New Jersey, 1996

13. J.H.Friedman and W.Stuetzle. "Projection pursuit regression",Ann.Statist, Vol.76, pp.817-823, 1981

14. P. R. Halmos. Measure Theory, Van Nostrand, New York, 1967.

15. T.Murofushi and M.Sugeno. "An interpretation of fu2zy measures and the Choquet integral as an integral with respect to a fuzzy measure", Fuzzy Sets and Systems 29, pp.201-227,1989.

16. Z.Wang. "Convergence theorems for sequences of Choquet integrals", International Journal of General Systems 26, pp.133-143,1997.

17. J.M. Keller, H.Qiu, and H.Tahani. ''Fumy integral and image segmentation", Proc. NAETPS.86, New Orleans, pp. 324-338, 1986.

18. J. M. Keller and J.Osbom. "Training the fuzzy integral", International Journal of Approximate Reasoning 15, pp.1-24, 1996.

19. M.Sugeno. "Theory of Fuzzy Integral and its Applications", Ph. D. dissertation, Tokyo Institute of Technology, 1974.

20. M.Grabich. "A New algorithm for Identifying Fuzzy Measures and Its Application to Pattern Recognition", Proc. Joint Conf. 4-th IEEE Internat. Conf. On Fuzzy

Systems and 2-nd Internat. Fuzzy Engineering Symposium, Yokohama, Japan, pp.145-150, 1995.

21. J. Han and M. Kambr. Data Mining-Concepts and Techniques, Morgan Kaufmann Publishers, 2000.
22. Z.Wang. "A new model of nonlinear multiregressions by projection pursuit based on generalized Choquet integrals", Proc. FUZZ-IEEE, pp.1240-1244, 2002.
23. Z.Wang, K-S.Leung, and J.Wang. "Determining nonnegative monotone set functions based on Sugeno's integral; an application of genetic algorithms," Fuzzy Sets and Systems, vol.112, pp.155-164, 2002.
24. K.Xu, Y.Ke, and Z.Wang. "A fast algorithm for Choquet-integral-based nonlinear multiregression used in data mining", The Journal of Fuzzy Mathematics, Vol.8, pp.195-201, 2000
25. K.Xu, Z.Wang, M.Wong and K.Leung. "Discover dependency pattern among attributes by using a new type of nonlinear multiregression", International Journal of Intelligent Systems, Vol.161, pp.949-962, 2001.
26. D.Dubois and D.Prade. Fuzzy Number: An overview, Reading in fuzzy sets for intelligent systems, Morgan Kaufman Publishers, San Mateo, pp.112-148,1993.
27. M.Grabisch. "Fuzzy Integral for Classification and Feature Extraction, Fuzzy Measure and Integrals", PhysicaVerlag, pp.415-434,2000.
28. G.J.Klir and B.Yuan. "Fuzzy Sets and Fuzzy Logic-Theory and Applications", Prentice Hall, New Jersey, 1995.
29. Z.Wang and J.Wang. "Determining Fuzzy Measures by Choquet Integral", Proc. ISUMA-NAFIPA'95, IEEE Computer Society Press, pp.274-277, 1995.
30. L.A.Zadeh. "Fuzzy Sets", Information and Control, Vol.8, pp.338-353,1965.

31. D.Ralescu. "The fuzzy integrals", Journal of Mathematical Analysis and applications Vol.75, pp.526-570, 1980.

32. P.Walley. Statistical Reasoning with ImpreciseProbabilities, Chapman and Hall, London, 1991.

(4) 33. Z.Wang, G. J. Klir, and W.Wang. Monotone set, pp. 324-338,1986.

7 → Intro. → 15.

8 → Conference paper → 16

14. mit web → 17

15. resource web → 18

9, 11 → 6, 7

5, 13 → 8, 9.

3 → 10

4 → 11.

6, 10 → 12, 13

12 → 14. distance

9, 10, 1 → 6, 12, 1

7, 11, 14 → 7, 15, 16.

15 → 19

8 → 17.

16 - 20

8 - 18.