On the uniqueness of the expression for the Choquet integral with linear core in classification

Weiwei Zhang

Abstract

The Choquet integral has been applied in data mining, such as nonlinear multiregressions and nonlinear classifications. Adopting signed efficiency measure in the Choquet integral makes the models more powerful. Another idea for generalizing the above-mentioned models is to use a linear core in the Choquet integral. This has been successfully used in nonlinear mulregression. However, there is a uniqueness problem for presenting the Choquet integral in classification models such that it is difficult to explain the exact contribution from each individual attributes, as well as their combinations, towards the target. In this work, an additional restriction on the parameters is given to guarantee the uniqueness of the expression.

1. Introduction

The Choquet integral of a weighted integrand with respect to a monotone measure has been used by Kebin Xu [6] for classification. As a new type of classification methods which attempt to use nonlinear integral as an aggregation tool, it achieves good performance on classification problems though it requires relatively long training time on identifying the monotone measure and the weights when the number of feature attributes is large. Moreover, this classifier is very informative and flexible for dealing with data sets involving strong interactions among the feature attributes toward the classification. However, the nonnegativity of the weights and the monotonicity of the set function restrict the classification action in the first quadrant and the vertex angle being obtuse.

In this paper, the weight function and the monotone measure are replaced by a linear
core and a signed efficiency measure respectively. They extend the applicability and make the classification much more powerful. The first one makes the data being in all four quadrants. This means any real number can be an observation. The second one makes the vertex angle being any angle, but not only obtuse. The improvement has geometric interpretation and uses in practical application. However, the idea of using linear core in Choquet integral in classification models has a uniqueness problem of expression of the Choquet integral. We discuss the problem in the paper and enhance the restriction on the parameters.

The paper is organized as follows. In Section 2, the relevant mathematical concepts are introduced. In Section 3, we show the information fusion and the connection between observations of feature attributes and the objective attribute. Section 4 presents a classification problems. In Section 5 we discuss the importance factor of feature attributes. There is a uniqueness problem in nonlinear classification with linear core. This is discussed in Section 6. Section 7 presents an idea to overcome the uniqueness problem. Finally, brief conclusions and a few ideas for future research are presented in Section 8.

2. Relevant mathematical concepts

Let $X$ be a nonempty set, called the universal set. We also let $\mathcal{F}$ be a $\sigma$-algebra, consisting of subset of $X$. If $X$ is finite, usually we take $\mathcal{P}(X)$ as $\sigma$-algebra $\mathcal{F}$. $(X, \mathcal{F})$ is called a measurable space.

**Definition 1** [5]. Set function $\mu : \mathcal{F}(X) \rightarrow (-\infty, \infty)$ is called a monotone measure (is also called fuzzy measure) if

1. $\mu(\emptyset) = 0$;
2. $\mu(E) \geq 0 \quad \forall E \in \mathcal{F}$;
(3) \( \mu(E) \leq \mu(F) \) if \( E \subseteq F \) whenever \( E, F \in \mathcal{F} \), and \( E \subseteq F \).

If \( \mu(X) = 1 \), monotone measure \( \mu \) is called normalized.

**Definition 2** [5]. Set function \( \mu : \mathcal{F}(X) \to (-\infty, \infty) \) is called an efficiency measure (is also called generalized measure) if it satisfies the condition (1) and (2) given in Definition 1.

Any monotone measure is a special case of an efficiency measure.

**Definition 3** [5]. Set function \( \mu : \mathcal{F}(X) \to (-\infty, \infty) \) is called a signed efficiency measure (is also called signed generalized measure) if it satisfies only condition (1) given in Definition 1.

When a signed efficiency measure (as well as the monotone measure or efficiency measure) is used in classification, \( \mu(X) \) assumes value 1 and the values at the other sets are denoted as \( \mu_1, \mu_2, \ldots, \mu_{2^n-2} \) respectively according to a binary code for the subscript, that is, \( \mu_1 = \mu(\{x_1\}) \), \( \mu_2 = \mu(\{x_2\}) \), \( \mu_{2^n-3} = \mu(\{x_1, x_3, \ldots, x_n\}) \), \( \mu_{2^n-2} = \mu(\{x_2, \ldots, x_n\}) \).

Both monotone measures and efficiency measures are special cases of signed efficiency measures.

Let \( f : X \to (-\infty, \infty) \) be a measurable function with respect to \( \mathcal{F} \). When \( \mu \) is a monotone measure, the Choquet integral of \( f \) with respect to \( \mu \) can be defined as follows.
Definition 4 [5]. Let $\mu$ be a monotone measure on $(X, \mathcal{F})$. The Choquet integral of $\mathcal{F}$ with respect to $\mu$, denoted by $(C)\int f \, d\mu$, is defined as

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

if not both Riemann integrals in (1) are infinite, where $F_{\alpha} = \{ x \mid f(x) \geq \alpha \}$ for $\alpha \in (-\infty, \infty)$ and is called $\alpha$-level set of $f$.

The $\alpha$-level set $F_{\alpha}$ is nonincreasing with respect to $\alpha$. Since $\mu$ is nondecreasing, $\mu(F_{\alpha})$ is nonincreasing function of $\alpha$, such that both Riemann’s integral in Definition 4 are well defined.

As a special case, when $f$ is nonnegative, formula (1) is reduced to be

$$(C)\int f \, d\mu = \int_{0}^{\infty} \mu(F_{\alpha}) \, d\alpha$$

In most cases, formula (1) is also feasible for the Choquet integral with respect to efficient measures.

In any database, the number of attributes is always finite. Let $X = \{x_1, x_2, \ldots, x_n\}$ denote a finite set of attributes. Then $(X, \mathcal{P}(X))$ is a measurable space. Each record (or, observation) of $x_1, x_2, \ldots, x_n$, denoted by $f(x_1), f(x_2), \ldots, f(x_n)$ respectively, is just a real-valued function on $X$. Since the power set of $X$ is taken as the $\sigma$-algebra, any real-valued function on $X$ is measurable. In this case, the Choquet integral can be generalized to be taken with respect to efficiency measures. Though $\mu(F_{\alpha})$ is no longer monotonic function of $\alpha$, it is, in fact, a function of $\alpha$ with bounded variation, such that Riemann integrals in expression (1) are still well defined. Up to now, the Choquet integral of real-valued function $f$ can be taken with respect to any efficiency measure.

If $\mu$ is a signed efficiency measure, then it can be expressed as a difference of two
efficiency measures:

\[ \mu = \mu_1 - \mu_2, \]  

(3)

where both \( \mu_1 \) and \( \mu_2 \) are efficiency measure satisfying \( \mu_1(E) \cdot \mu_2(E) = 0 \) for every \( E \in \mathcal{F} \). Such a decomposition is called the least decomposition of \( \mu \).

**Definition 5** [5]. Let \( \mu \) be a signed efficiency measure and \( f \) be a real-valued measurable function on \((X, \mathcal{F})\).

The Choquet integral of \( f \) with respect to \( \mu \) is defined as

\[
(C)\int f \, d\mu = (C)\int f \, d\mu_1 - (C)\int f \, d\mu_2
\]

if not both Choquet integral in formula (4) are infinite, where \( \mu_1 \) and \( \mu_2 \) are efficiency measures and form the least decomposition of \( \mu \).

As mentioned before, in any database the number of attributes is always finite, that is, \( X \) is a finite set. In this case, there is a simple formula for calculating the value of \( (C)\int f \, d\mu \) once \( f \) and \( \mu \) are given. First, the values of function \( f \), \( \{f(x_1), f(x_2), ..., f(x_n)\} \), are rearranged into a nondecreasing order as,

\[ f(x_1^*) \leq f(x_2^*) \leq ... \leq f(x_n^*) \]

where \( (x_1^*, x_2^*, ..., x_n^*) \) is a permutation of \( (x_1, x_2, ..., x_n) \). Then the Choquet integral of \( f \) with respect to \( \mu \) can be calculated by

\[
(C)\int f \, d\mu = \sum_{i=1}^{n} [f(x_i^*) - f(x_{i-1}^*)] \cdot \mu(\{x_1^*, x_2^*, ..., x_n^*\})
\]

with a convention \( f(x_0^*) = 0 \).

The Choquet integral has been successfully applied in data mining. Xu [6] adopts the
Choquet integral having weighted integrand with respect to monotone measure in classification. Xu’s classification model is over restricted. The data in that model must lie in the first quadrant. The project line must go through origin and lie in the first quadrant. Furthermore, the angle between the projection directions described by the two branches of the classifying boundary must be obtuse. To improve the classification model we allow the data possessing real value, that is the data can appear in any quadrant. Moreover, the set function $\mu$ is generalized from monotone measure to be signed efficient measure. At the same time, the weighted integrand is replaced by linear core $a + bf$, where $a$ and $b$ are real-valued functions defined on $X$, $f$ is an observation of $x_1, x_2, ..., x_n$, $\mu$ is a signed efficiency measure. Functions $a$ and $b$ can be expressed as n-vectors, i.e., $a = (a_1, a_2, ..., a_n)$ and $b = (b_1, b_2, ..., b_n)$. They should satisfy the following constraints:

1. $a_i \geq 0$ for $i = 1, 2, ..., n$, with $\min_{1 \leq i \leq n} a_i = 0$;
2. $-1 \leq b_i \leq 1$ for $i = 1, 2, ..., n$, with $\max_{1 \leq i \leq n} |b_i| = 1$;
3. $\mu(X) = 1$.

Such a classification model is more flexible and powerful than weighted model.

3. Information fusion and identification of set functions

Suppose the data in the data base is complete and have $l$ observations with a form as:

\[
\begin{array}{cccccc}
  x_1 & x_2 & ... & x_n & y \\
  f_{11} & f_{12} & ... & f_{1n} & y_1 \\
  f_{21} & f_{22} & ... & f_{2n} & y_2 \\
  & & ... & & \\
  f_{l1} & f_{l2} & ... & f_{ln} & y_l \\
\end{array}
\]
where $y$ is the objective attribute, which has categorical value, and positive integer $l$ denotes the size of the data it should be much larger than $n$, usually, not less than 5 times of $2^n$. The row
\[ f_{f_1} \ f_{f_2} \ \ldots \ f_{f_n} \ y_j \]
is the $j$ th observation of attributes $x_1, x_2, \ldots, x_n$ and $y$, $j=1,2,\ldots,l$. 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 $f^{(j)}$, and we write $f_{ji} = f^{(j)}(x_i), i=1,2,\ldots,n$ for $j=1,2,\ldots,l$.

4. **Nonlinear classifications based on the Choquet integral with linear core**

Any classification problem with $n$ classes (called $n$-classification problem) can be divided into $n-1$ successive 2-classification problems, in which $y$ has only 2 categorical values, say, I and II. Thus, in this work only 2 classes are discussed.

Given the data shown in Section 3, we want to determine the values of unknown parameters $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ and $\mu_1, \mu_2, \ldots, \mu_{2^n-2}$ as well as the classifying critical value $c$ through a learning procedure optimally. The optimization criterion can be chosen as minimization misclassification rate for the given learning data. The process can be implemented by using a soft computing technique, such as the genetic algorithms and pseudo gradient search [6]. Once a new observation $f$ (as a new sample) is available, calculate the value of $\hat{c} = (C) \int (a + bf) d\mu$. Comparing $\hat{c}$ with the critical value $c$, if $\hat{c} \leq c$, then we classify the new sample into one class, say, class I; otherwise, the new sample classified into class II.

5. **The Shapley value and important index**
Once the values of parameters, \( a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n, c, \) and \( \mu_1, \mu_2, \ldots, \mu_{n-2}, \)
in the classification model are determined through a learning procedure based on given
data, it is expected to use them to illustrate the importance of various attributes to the
target, making classification.

Each feature attribute plays its own respective role in the classification. When signed
efficient measure \( \mu \) is additive, that is, \( \mu \) is a classical measure on \( \mathcal{P}(X) \), the importance of
an attribute \( x_i \) to the classification can be quantified by \( \mu(\{x_i\}) \) directly since there is no
interaction among feature attributes toward the target. However, in most real problems,
the involved signed efficient measure \( \mu \) is not additive. For example, \( \mu(\{x_i, x_j\}) \) may be
different from \( \mu(\{x_i\}) + \mu(\{x_j\}) \). Thus, the importance of attribute \( x_i \) is reflected not only
by \( \mu(\{x_i\}) \), but also by \( \mu(\{x_i, x_j\}) - \mu(\{x_j\}) \) for every \( j \neq i \). The following concept of
Shapley value can be used as an important index of each feature attribute towards target.

**Definition 6** [6]. The Shapley value of feature attribute \( x_i \) with respect to \( \mu \) is defined by

\[
v_i = \sum_{k=0}^{n-1} \gamma_k \sum_{K \subseteq X \setminus \{x_i\}, |K|=k} (\mu_K - \mu_K), \quad i = 1, 2, \ldots, n
\]

(5)

with

\[
\gamma_k = \frac{(n-k-1)!k!}{n!} = \frac{1}{C_{n-1}^k}, \quad k = 0, 1, \ldots, n-1
\]

(6)

where \( \mu_K = \mu(\{x_i\} \cup K), \mu_K = \mu(K), |K| \) indicates the cardinality of subset \( K \) of \( X \), and
\( 0! = 1 \) as usual.

This definition is proposed by Shapley [1] on an axiomatic basis in cooperative game
theory and discussed by Grabisch [2], as well as some scholars else. Value \( v_i \) is a
hierarchical average of the additional effects of adding attribute \( x_i \) to every set in \( \mathcal{P}(X) \).
(X-\{x_i\}). It can be shown that \(\sum_{i=1}^{n} v_i = 1\). In case signed efficient measure \(\mu\) is additive, the above formula is reduced as \(v_i = \mu_i, \quad i = 1, 2, \ldots, n\).

6. Artificial data used to show the uniqueness problem

In Xu’s classification model [6], a weighted integrand is used in the Choquet integral, which is an aggregation to fuse the data of feature attributes. In this work, the weighted integrand is generalized by using a linear transformation of \(f\), called the linear core of the Choquet integral. Simultaneously, the observation of feature attributes, \(f\), is allowed to be any linear value and the nonadditive set function \(\mu\) can take any real value as well, that is, \(\mu\) is a signed efficient measure. Thus, the new classification model based on the Choquet integral becomes

\[ c = (C)\int (a + bf)d\mu. \]

Recently, people find that, given a data set for classification problem, the estimated value of parameters may be various when this model is used, though the corresponding classifying boundary is the same. This causes a difficulty for explaining the importance of feature attributes towards the classification based on the value of parameters.

An artificial data set now is used to show the problem on the uniqueness of the Choquet integral expression with linear core in the classification model.

There are two feature attributes, two classes, and 26 records in the data set, that is, \(X = \{x_1, x_2\}, \quad C = \{C_1, C_2\} \quad \text{and} \quad l = 20\).
The data are shown in Figure 1 by ◊ for class I and ⋄ for class II respectively, where \( f_1 \) and \( f_2 \) are two coordinates on the plane. Geometrically, the classification by the Choquet integral with linear core can be described as follows. There is a projection axis \( L \) located by equation \( a_1 + b_1 f_1 = a_2 + b_2 f_2 \). For each sample point \((f_1, f_2)\), around with the directions presented by the two branches of the contours of the Choquet integral, is projected onto axis \( L \) as a point with value \( c = (C) \int (a+bf) d\mu \). Thus, the 2-dimensional classification problem is converted to be a one dimensional classification problem on \( L \), which can be solved by using only one critical value as the boundary of two classes.

The two classes in the given data can be actually well separated (with misclassification

<table>
<thead>
<tr>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>classification</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/5</td>
<td>6</td>
<td>I</td>
<td>21/20</td>
<td>7</td>
<td>II</td>
</tr>
<tr>
<td>1/4</td>
<td>3</td>
<td>I</td>
<td>11/10</td>
<td>17/3</td>
<td>II</td>
</tr>
<tr>
<td>1/3</td>
<td>5</td>
<td>I</td>
<td>6/5</td>
<td>21/5</td>
<td>II</td>
</tr>
<tr>
<td>1/2</td>
<td>2</td>
<td>I</td>
<td>6/5</td>
<td>9</td>
<td>II</td>
</tr>
<tr>
<td>1/2</td>
<td>6</td>
<td>I</td>
<td>3/2</td>
<td>1</td>
<td>II</td>
</tr>
<tr>
<td>3/5</td>
<td>7</td>
<td>I</td>
<td>2</td>
<td>16</td>
<td>II</td>
</tr>
<tr>
<td>2/3</td>
<td>2</td>
<td>I</td>
<td>2</td>
<td>20</td>
<td>II</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>I</td>
<td>3</td>
<td>9</td>
<td>II</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>I</td>
<td>3</td>
<td>10</td>
<td>II</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>I</td>
<td>3</td>
<td>11</td>
<td>II</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>II</td>
<td>4</td>
<td>4</td>
<td>II</td>
</tr>
<tr>
<td>1/5</td>
<td>9</td>
<td>II</td>
<td>4</td>
<td>8</td>
<td>II</td>
</tr>
<tr>
<td>4/5</td>
<td>8</td>
<td>II</td>
<td>4</td>
<td>16</td>
<td>II</td>
</tr>
</tbody>
</table>
rate 0) by a contour of the Choquet integral with linear core. The contour is a broken line indicated by $f_2 = \frac{20}{9} f_1 + \frac{80}{9}$ when $a_1 + b_1 f_1 \leq a_2 + b_2 f_2$ and $f_2 = \frac{40}{3} f_1 + 20$ when $a_1 + b_1 f_1 > a_2 + b_2 f_2$. The vertex of the broken line is $(1, \frac{20}{3})$, which is on axis $L_1$ possessing equation $f_2 = \frac{10}{3} f_1 + \frac{10}{3}$. These are shown in Figure 1. The corresponding values of parameters in the Choquet integral are $a=(1,0)$, $b=(1,\frac{3}{10})$, $\mu(\emptyset) = 0$, $\mu(\{f_1\}) = \frac{4}{5}$, $\mu(\{f_2\}) = \frac{3}{5}$, and $c = 2$.

However, for the same boundary it is not difficult to find other values of parameters in the expression of Choquet integral’s contour. For example, $a=(0,0)$, $b=(1,\frac{3}{20})$, $\mu(\emptyset) = 0$, $\mu(\{f_1\}) = \frac{2}{3}$, $\mu(\{f_2\}) = \frac{3}{4}$, and $c = 1$. The corresponding contour of the Choquet integral with these parameters’ values coincides with the previous one, though the projection axis $L_2$ is different from $L_1$. These are also illustrated in Figure 1. The detailed calculation for this example is given in Appendix.

The above example shows the uniqueness for the expression of classifying boundary based on the Choquet integral with linear core is violated. Table 2 lists all classification results by using above-mentioned two classifiers. Both of them classify the given data perfectly.
Unfortunately, the corresponding Shapely values of attributes of $x_1$ and $x_2$ obtained from these two classifiers are different. In fact, in the first classifier, $\mu(\emptyset) = 0, \mu(\{x_1\}) = \frac{4}{5}, \mu(\{x_2\}) = \frac{3}{5}, \mu(X) = 1$, we get

$$
\gamma_0 = \frac{(2-0-1)!0!}{2!} = \frac{1}{C_0^{2-1}2} = \frac{1}{2}
$$

$$
\gamma_1 = \frac{(2-1-1)!!}{2!} = \frac{1}{C_1^{2-1}2} = \frac{1}{2}
$$

$$
v_1 = \frac{1}{2} \left( \frac{4}{5} - 0 \right) + \frac{1}{2} \left( 1 - \frac{3}{5} \right) = \frac{3}{5}
$$

$$
v_2 = \frac{1}{2} \left( \frac{3}{5} - 0 \right) + \frac{1}{2} \left( 1 - \frac{4}{5} \right) = \frac{2}{5}
$$

While in the second classifier, $\mu(\emptyset) = 0, \mu(\{x_1\}) = \frac{2}{3}, \mu(\{x_2\}) = \frac{3}{4}, \mu(X) = 1$, we get

Figure 1
\[ \gamma_0 = \frac{(2-0-1)!0!}{2!} = \frac{1}{C_0^{2-1}2} = \frac{1}{2} \]
\[ \gamma_1 = \frac{(2-1-1)!1!}{2!} = \frac{1}{C_1^{2-1}2} = \frac{1}{2} \]
\[ v_1 = \frac{1}{2} \left( \frac{2}{3} - 0 \right) + \frac{1}{2} \left( 1 - \frac{3}{4} \right) = \frac{11}{24} \]
\[ v_2 = \frac{1}{2} \left( \frac{3}{4} - 0 \right) + \frac{1}{2} \left( 1 - \frac{2}{3} \right) = \frac{13}{24} \]

We see that the Shapley values in these two classifier are different though they share the same classify boundary.

**Table 2**

<table>
<thead>
<tr>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>classification</th>
<th>( a_1 + b_1 f_{j_1} )</th>
<th>( a_2 + b_2 f_{j_2} )</th>
<th>( C=2 )</th>
<th>( f_1 )</th>
<th>( a_1 + b_1 f_{j_1} )</th>
<th>( a_2 + b_2 f_{j_2} )</th>
<th>( C=1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/5</td>
<td>6</td>
<td>I</td>
<td>6/5</td>
<td>9/5</td>
<td>24/25</td>
<td>1/5</td>
<td>9/10</td>
<td>29/40</td>
<td></td>
</tr>
<tr>
<td>1/4</td>
<td>3</td>
<td>I</td>
<td>5/4</td>
<td>9/10</td>
<td>59/50</td>
<td>1/4</td>
<td>9/20</td>
<td>2/5</td>
<td></td>
</tr>
<tr>
<td>1/3</td>
<td>5</td>
<td>I</td>
<td>4/3</td>
<td>3/2</td>
<td>43/30</td>
<td>1/3</td>
<td>3/4</td>
<td>31/48</td>
<td></td>
</tr>
<tr>
<td>1/2</td>
<td>2</td>
<td>I</td>
<td>3/2</td>
<td>6/10</td>
<td>33/25</td>
<td>1/2</td>
<td>3/10</td>
<td>13/30</td>
<td></td>
</tr>
<tr>
<td>1/2</td>
<td>6</td>
<td>I</td>
<td>3/2</td>
<td>9/5</td>
<td>42/25</td>
<td>1/2</td>
<td>9/10</td>
<td>8/10</td>
<td></td>
</tr>
<tr>
<td>3/5</td>
<td>7</td>
<td>I</td>
<td>5/8</td>
<td>21/10</td>
<td>19/10</td>
<td>3/5</td>
<td>21/20</td>
<td>15/16</td>
<td></td>
</tr>
<tr>
<td>2/3</td>
<td>2</td>
<td>I</td>
<td>5/3</td>
<td>3/5</td>
<td>109/75</td>
<td>2/3</td>
<td>3/10</td>
<td>49/90</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>I</td>
<td>2</td>
<td>2/10</td>
<td>83/55</td>
<td>1</td>
<td>3/10</td>
<td>23/30</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>I</td>
<td>2</td>
<td>9/10</td>
<td>89/50</td>
<td>1</td>
<td>9/20</td>
<td>49/60</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>I</td>
<td>2</td>
<td>6/5</td>
<td>46/25</td>
<td>1</td>
<td>3/5</td>
<td>13/15</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>II</td>
<td>2</td>
<td>3</td>
<td>13/5</td>
<td>1</td>
<td>3/2</td>
<td>11/8</td>
<td></td>
</tr>
</tbody>
</table>
7. An improving idea

In order to overcome the uniqueness problem mentioned in Section 6, geometrically, we can strengthen the restriction of the parameters in the model such that the project axis \( L \), passes through the origin, that is, let \( a = (0,0) \). Thus, the classifier becomes

\[
c = (C) \int bf \, d\mu.
\]

In such a new model the expression of the boundary is the contour of the Choquet integral \( (C) \int bf \, d\mu \) is now unique.
8. Conclusions

The paper presents a new model with a linear core based on Choquet integrals for solving data classification problems, it is a significant generalization of the existing weighted Choquet integral model. However, the same boundary may have different values of parameters in the expression of Choquet integral’s contour. It rises a problem about explaining the importance of feature attributes towards the classification based on the value of parameters. Thus, it needs a more strong constraint to make the expression unique. One way we can take is to make the line pass through the origin. Now, any interaction, no matter it is positive or negative, among feature attributes towards the classifying attribute can be described and the data may locate in any quadrant and are projected to any suitable axis passing through the origin.
References


Appendix

Constraints:

1) \( a = (a_1, a_2), a_i \in [0, \infty) \quad i = 1, 2; \)

2) \( b = (b_1, b_2), b_i \in [-1,1] \quad i = 1, 2; \)

3) \( \mu(X) = 1. \)

If \( a_1 + b_1f_1 \leq a_2 + b_2f_2, \)

\[
c = (a_i + b_if_i)\mu(X) + (a_2 + b_2f_2 - a_1 - b_1f_1)\mu(\{f_2\});
\]

If \( a_1 + b_1f_1 > a_2 + b_2f_2, \)

\[
c = (a_2 + b_2f_2)\mu(X) + (a_1 + b_1f_1 - a_2 - b_2f_2)\mu(\{f_1\}).
\]

The projection axis \( L_1 \) given \( a = (1,0), \quad b = (1, \frac{3}{10}), \quad \mu(\emptyset) = 0, \quad \mu(\{f_1\}) = \frac{4}{5}, \)

\( \mu(\{f_2\}) = \frac{3}{5} \) and \( c = 2 \) is the boundary.

If \( 1 + 1 \cdot f_1 \leq 0 + \frac{3}{10}f_2, \)

\[
(1 + 1 \cdot f_1) \cdot 1 + (0 + \frac{3}{10}f_2 - 1 - 1 \cdot f_1) \cdot \frac{3}{5} = 2, \quad \text{simplify it we get}
\]

If \( 1 + f_1 \leq \frac{3}{10}f_2, \)

\[
\frac{1}{4}f_1 + \frac{9}{80}f_2 = 1;
\]

If \( 1 + f_1 > 0 + \frac{3}{10}f_2, \)

\[
(0 + \frac{3}{10}f_2) \cdot 1 + (1 + 1 \cdot f_1 - 0 - \frac{3}{10}f_2) \cdot \frac{4}{5} = 2, \quad \text{simplify it we get}
\]

If \( 1 + f_1 > \frac{3}{10}f_2, \)

\[
\frac{2}{3}f_1 + \frac{1}{20}f_2 = 1.
\]

Because the above two line \( \frac{1}{4}f_1 + \frac{9}{80}f_2 = 1 \) and \( \frac{2}{3}f_1 + \frac{1}{20}f_2 = 1 \) have the same point
of intersection, we get two equations:

\[
\begin{align*}
\frac{1}{4} f_1 + \frac{9}{80} f_2 &= 1, \\
\frac{2}{3} f_1 + \frac{1}{20} f_2 &= 1,
\end{align*}
\]

Then, get the point \( (1, \frac{20}{3}) \).

Suppose there exists another projection axis \( L_2 \) which has the same contour as the \( L_1 \). Let it have \( a = (0,0) \), \( \mu(X) = Q \), \( \mu(\{ f_1 \}) = s \) and \( \mu(\{ f_2 \}) = t \). According to

\[
0 + b_1 \cdot f_1 = 0 + b_2 \cdot f_2,
\]

we get \( \frac{f_1}{f_2} = \frac{b_2}{b_1} \).

Since we suppose \( L_2 \) has the same contour with \( L_1 \), that means \( L_2 \) has the point \( (1, \frac{20}{3}) \).

Following the constraints about \( b \), we take \( b_1 = 1 \). Now, we have three conditions:

\[
\begin{align*}
\frac{f_1}{f_2} &= \frac{b_2}{b_1} \\
(1, \frac{20}{3}) \\
b_1 &= 1
\end{align*}
\]

and the projection axis \( L_2 \) has \( b = (1, \frac{3}{20}) \).

If \( 0 + f_1 \leq 0 + \frac{3}{20} f_2 \),

\[(0 + 1 \cdot f_1) \cdot Q + (0 + \frac{3}{20} \cdot f_2 - 0 - 1 \cdot f_1) \cdot t = C\]

simplify it we get

If \( f_1 \leq \frac{3}{20} f_2 \),

\[
\frac{(Q-t)}{C} f_1 + \frac{3t}{20C} f_2 = 1;
\]

If \( 0 + f_1 > 0 + \frac{3}{20} f_2 \),
\[(0 + \frac{3}{20} \cdot f_2) \cdot Q + (0 + 1 \cdot f_1 - 0 - \frac{3}{20} \cdot f_2) \cdot s = C\], simplify it we get

If \( f_1 > \frac{3}{20} f_2 \),
\[\frac{s}{C} f_1 + \frac{3(Q-s)}{20C} f_2 = 1.\]

Since it is supposed that both of the projection axis have the same contour, then
\[
\frac{1}{4} f_1 + \frac{9}{80} f_2 = 1 \quad \text{and} \quad \frac{(Q-t)}{C} f_1 + \frac{3t}{20C} f_2 = 1 \quad \text{are the same line},
\]
\[
\frac{2}{3} f_1 + \frac{1}{20} f_2 = 1 \quad \text{and} \quad \frac{s}{C} f_1 + \frac{3(Q-s)}{20C} f_2 = 1 \quad \text{are the same line}.
\]

We get
\[
\frac{1}{4} = \frac{Q-t}{C} \cdot \frac{2}{3} = \frac{s}{C} \cdot \frac{9}{80} = \frac{3t}{20C} \cdot \frac{1}{20} = \frac{3(Q-s)}{20C},
\]

namely, \( Q = C, S = \frac{3}{4} C, t = \frac{2}{3} C \), according to constraints, \( Q = \mu(X) = 1 \), now the projection axis \( L_2 \) has \( a = (0,0), b = (1, \frac{3}{20}), \mu(\emptyset) = 0, \mu(\{f_1\}) = \frac{2}{3}, \mu(\{f_2\}) = \frac{3}{4} \).

Since \( a_1 + b_1 f_1 = a_2 + b_2 f_2 \), when \( a = (1,0), b = (1, \frac{3}{10}) \), the projection axis \( L_1 \) is \( f_2 = \frac{10}{3} f_1 + \frac{10}{3} \), when \( a = (0,0), b = (1, \frac{3}{20}) \), the projection axis \( L_2 \) is \( f_2 = \frac{20}{3} f_1 \).

The contour are \( f_2 = -\frac{20}{9} f_1 + \frac{80}{9} \) and \( f_2 = -\frac{40}{3} f_1 + 20 \).