Application of Generalized Discrete Choquet Integral in Multivariate Statistical Analysis

Davor RADENOVIĆ and Goran DIMIĆ and Dragan RADOJEVIĆ

Mihajlo Pupin Institute, Volgina 15
11000 Beograd, Yugoslavia
Phone: (+381)11–772720, Fax: (+381)11–774265
Email: {davor, pedja, draganr}@L100.imp.bg.ac.yu

Abstract

The Choquet integral is a promising tool for multivariate statistical analysis (MVSA) in the presence of interaction between variables. The logical representation of Choquet integral enables using of the existing techniques of MVSA more efficiently. In this paper this is illustrated on the problem of discriminant analysis. Extending the list of variables by combining the variables using a parameterised $T-$norm can give us excellent classification results.

Keywords: logical representation; Choquet integral; discriminant analysis; interaction between variables, parameterised $T-$norm.

1 Introduction

The multivariate (discriminant) analysis can be simply defined as the application of methods that deal with reasonably large number of measurements (i.e. variables) made on each object in one or more samples simultaneously. The measurements relate to characteristics or attributes of the objects that are being recorded. We will call these the variables. The classical approach does not take into consideration the interactions among the variables. But, the presence of interactions between the variables is usual in real problems. The fuzzy measures and fuzzy integrals offer great potentials as tools for aggregating variables' values in the presence of interactions between the variables. The publication of some papers [3], [4] and [5] has intensified the interest in this approach.

In this paper we present the usage of the discrete Choquet integral in discriminant analysis in the presence of interactions among the variables. Since, in general case, there is no clear interpretation when dealing with fuzzy measures, by using an equivalent logical representation the problem is reduced on extension of the list of variables. All logical expressions of the variables are connected using AND operator that is defined as a $T-$norm.

The well known definitions of discrete fuzzy measure, and discrete Choquet integral are given in Section 2. The logical representation of discrete fuzzy Choquet integral is given in Section 3. The discriminant analysis is described in Section 4. The usage of logical representation in discriminant analysis is described in Section 5.

2 Fuzzy measures and Choquet integral

In this paper only discrete spaces and the finite universe Ω of p elements (attributes, features,...), $\Omega = \{X_1, ..., X_p\}$ is considered. $\mathcal{P}(\Omega)$ is the power set of Ω, while $|\Omega|$ denotes the cardinal of a set Ω, and $A \setminus B$ denotes the set difference. $\wedge$, $\vee$ denote min and max respectively.

The additivity property for (probability) measures is usually a hard constraint for real problems. Sugeno, [11], introduced fuzzy measure and integrals, as a generalization of the usual definition of a measure by relaxation of additive property. The concept of fuzzy measure and fuzzy integral is closely related to the twenty years older concept of capacity, proposed by Choquet, [1]. Fuzzy measures include as particular cases probability measures, possibility and necessity measures, belief and plausibility functions, etc. [4].

Definition 1 *A fuzzy measure* μ *is a mapping* $\mu : \mathcal{P}(\Omega) \rightarrow [0,1]$ *such that, for every* A *and* B *in* $\mathcal{P}(\Omega)$:

1. $\mu(\emptyset) = 0$,
2. *if* $B \subseteq A$, *then* $\mu(B) \leq \mu(A)$,

where: Ω *is any set of elements,* $\mathcal{P}(\Omega)$ *is the set of fuzzy subsets of* Ω, *and* $A, B, \ldots$ *are subsets of* Ω.

Mirofushi and Sugeno, [7], defined fuzzy Choquet integral using a concept introduced by Choquet in capacity theory.

Definition 2 *Let* μ *be a discrete fuzzy measure on* Ω, *whose elements are denoted by* $X_1, \ldots, X_p$. *The discrete Choquet integral with respect to* μ *is defined by*

$$C_\mu(X_1, \ldots, X_p) := \sum_{i=1}^{p} \left(X_{(i)} - X_{(i-1)}\right) \mu\left(A_{(i)}\right)$$

where $\cdot_{(i)}$ *indicates that the indices have been permuted so that* $0 \leq X_{(1)} \leq \ldots \leq X_{(p)}$, *and* $A_{(i)} := \left\{X_{(i)}, \ldots, X_{(p)}\right\}$, *and* $X_{(0)} = 0$.

The Choquet integral is a generalization of the Lebesgue integral, and it coincides with the Lebesgue integral when the measure is additive. The discrete fuzzy Choquet integral enables modeling of positive interaction and redundancy between attributes.

In the next section the logical representation of the discrete fuzzy Choquet integral is proposed as the solution to this problem.

3 The logical representation of Choquet integral

The logical representation of Choquet integral is based on the next properties:

(a) linearity of discrete Choquet integral by measures, [9], [2]. If some fuzzy measure can be represented as linear convex combination of some other fuzzy measures then Choquet integral for this measure is equivalent to linear convex combination of Choquet integrals for that fuzzy measures,

(b) Choquet integral for logical (0,1) fuzzy measure is equivalent to logical expression of the variables [8], and

(c) any fuzzy measure can be represented as a convex combination of logical fuzzy measure. Components of the fuzzy measure vector of a variables set are equal to fuzzy measures of elements of the variables power set.

In [10] is given a new logical representation of Choquet integral that is based only on AND operator.

Proposition 1 *Discrete Choquet integral of* $X_i, (i = 1, \ldots, p)$, *for any fuzzy measure* μ, *can be represented as:*

$$C_\mu(X_1, \ldots, X_p) = \sum_{A_i \subset \Omega} \left(\mu(A_i) \sum_{B_k \subset \Omega \setminus A_i} \left((-1)^{|B_k|} \bigwedge_{X_p \in A_i \cup B_k} X_p \right) \right).$$

Proof is given in [10].

Example 1. The logical fuzzy measure for the tree attribute case and the logical expressions corresponding to logical discrete fuzzy Choquet integral are defined as:

$$\begin{aligned} C_\mu(X_1, X_2, X_3) &= \mu_1 (X_1 - X_1 \wedge X_2 - X_1 \wedge X_3 + X_1 \wedge X_2 \wedge X_3) \\ &+\mu_2 (X_2 - X_1 \wedge X_2 - X_2 \wedge X_3 + X_1 \wedge X_2 \wedge X_3) \\ &+\mu_3 (X_3 - X_1 \wedge X_3 - X_2 \wedge X_3 + X_1 \wedge X_2 \wedge X_3) \\ &+\mu_{12} (X_1 \wedge X_2 - X_1 \wedge X_2 \wedge X_3) \\ &+\mu_{13} (X_1 \wedge X_3 - X_1 \wedge X_2 \wedge X_3) \\ &+\mu_{23} (X_2 \wedge X_3 - X_1 \wedge X_2 \wedge X_3) \\ &+\mu_{123} (X_1 \wedge X_2 \wedge X_3). \end{aligned}$$

From previous results follows next proposition:

Proposition 2 *Discrete Choquet integral of* X_i, $(i = 1, ..., p)$, *for any fuzzy measure* μ, *can be represented as:*

$$C_m(X_1, ..., X_p) = C_\mu(X_1, ..., X_p) = \sum_{A_i \subset \Omega} \left(m(A_i) \bigwedge_{X_p \in A_i} X_p \right),$$

where m *is Möbius transformation of* μ *defined as:* $m(A) := \sum_{B \subset A} (-1)^{|A \setminus B|} \mu(B), \forall A \subset \Omega.$

Proof is given in [10].

Example 2. Choquet integral of three attributes can be represented as, [10]:

$$\begin{aligned} C_\mu(X_1, X_2, X_3) &= \mu_1 X_1 + \mu_2 X_2 + \mu_3 X_3 + (\mu_{12} - \mu_1 - \mu_2)(X_1 \wedge X_2) + (\mu_{13} - \mu_1 - \mu_3)(X_1 \wedge X_3) \\ &\quad + (\mu_{23} - \mu_2 - \mu_3)(X_2 \wedge X_3) + (\mu_{123} - \mu_{12} - \mu_{13} - \mu_{23} + \mu_1 + \mu_2 + \mu_3)(X_1 \wedge X_2 \wedge X_3) \\ &= m_1 X_1 + m_2 X_2 + m_3 X_3 + m_{12}(X_1 \wedge X_2) + m_{13}(X_1 \wedge X_3) + m_{23}(X_2 \wedge X_3) \\ &\quad + m_{123}(X_1 \wedge X_2 \wedge X_3). \end{aligned}$$

So, the Choquet integral, in general case, can be represented as a linear convex combination of the variables and logical expressions with AND operators on all combinations on the variables.

4 Discriminant analysis. The Two–Group Problem

Discriminant analysis involves deriving linear combination of the variables that will discriminate between the a priori defined groups in such a way that the misclassification error rates are minimized.. This can be done by maximizing the between-group variance relative to the within-group variance [12]. In this paper we consider the two-group problem.

4.1 Fisher's Approach

Suppose a population G is made up of two groups G_1 and G_2. A measurement $\mathbf{X}$ consisting of p characteristics is observed from G. Our task is to develop an assignment rule for $\mathbf{X}$ that will allocate this observation to G_1 or G_2. To assist in defining a rule, the researcher has access to n observations, of which n_1 are form G_1 and n_2 are from G_2.

Under the assumption that the true mean vector for G_i is $\mu_i, i = 1, 2$, and that the variance-covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ have a common value $\mathbf{\Sigma}$, Fisher suggested finding a linear combination of $\mathbf{X}$ so that the ratio of the difference in the means of the linear combination in G_1 and G_2 to its variance is maximized. In other words, denoting the linear combinations by $\mathbf{Y} = \mathbf{b}'\mathbf{X}$, we wish to find a vector of weights $\mathbf{b}$ so that we maximize the criterion

$$\triangle = \frac{\mathbf{b}'\boldsymbol{\mu}_1 - \mathbf{b}'\boldsymbol{\mu}_2}{\mathbf{b}'\mathbf{\Sigma}\mathbf{b}}.$$

It is not difficult to show that $\mathbf{b}$ is proportional to $\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. The linear combination, therefore, is not unique; only ratios of the coefficients are. Thus, any set of coefficients can be multiplied by any constant.

In applications the parameters will usually not be known. Hence the samples of n_i observations from each G_i are used to define a sample-based rule by replacing μ_i with $\bar{\mathbf{x}}_i$, the estimated mean vector in G_i, and $\mathbf{\Sigma}$ with $\mathbf{S}$, the pooled sample variance-covariance matrix. These estimates are given by

$$\bar{\mathbf{x}}_i = (\bar{x}_{i1}, \bar{x}_{i2}, \ldots, \bar{x}_{ip})$$

$i = 1, 2$, and

$$\mathbf{S} = \frac{1}{n_1 + n_2 - 2}(\mathbf{x}_1'\mathbf{x}_1 + \mathbf{x}_2'\mathbf{x}_2)$$

where $\bar{x}_{ij} = \sum_{l=1}^{n_i} X_{jl}/n_i$, $i = 1, 2$, $j = 1, 2, ..., p$, and where $\mathbf{x}_1'$ si the $(p \times n_1)$ matrix of observations in deviation form taken from G_1, and $\mathbf{x}_2'$ si the $(p \times n_2)$ matrix of observations in deviation form taken from G_2. Replacing parameters with their respective sample-based estimates shows that

$$\hat{\mathbf{b}} = \mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

where $\mathbf{S}^{-1}$ is the inverse of the pooled sample variance-covariance matrix.

The mean value of the discriminant function is commonly referred to as the group centroid. The group centroids, denoted by $\bar{Y}_i$, where i is used to identify the group being considered, are obtained by applying the vector of discriminant coefficients to the mean score vector for each group: that is, $\bar{Y}_i = \hat{\mathbf{b}}\bar{\mathbf{x}}_i$.

The point of separation is obtained by

$$Y_c^* = \frac{n_2\bar{Y}_1 + n_1\bar{Y}_2}{n_1 + n_2}.$$

5 Using the logical representation of Choquet integral in discriminant analysis

The discriminant analysis described in the previous section can be applied in the case of presence of the interactions among the variables. The list of the variables is extended by new "generalized" variables that are logical expressions on all combinations of the original variables with AND operators defined as $T-$norm functions.

Example 3. In the presence of interaction among two variables X_1 and X_2, the new generalized variable is a $T-$norm function $T(X_1, X_2)$. The measurement matrix $\mathbf{X}$ can be shown as

Sample number	X_1	X_2	$T(X_1, X_2)$	class
1	X_{11}	X_{21}	$T(X_{11}, X_{21})$	1
2	X_{12}	X_{22}	$T(X_{12}, X_{22})$	1
...	...	...	...	...
n_1	X_{1n_1}	X_{2n_1}	$T(X_{1n_1}, X_{2n_1})$	1
n_1+1	X_{1n_1+1}	X_{2n_1+1}	$T(X_{1n_1+1}, X_{2n_1+1})$	2
n_1+2	X_{1n_1+2}	X_{2n_1+2}	$T(X_{1n_1+2}, X_{2n_1+2})$	2
...	...	...	...	...
n_1+n_2	$X_{1n_1+n_2}$	$X_{2n_1+n_2}$	$T(X_{1n_1+n_2}, X_{2n_1+n_2})$	2

The next example show how powerful can be the discriminant analysis when applied to the extended list of the variables. We give both the results obtained without and with extended list of variables.

Example 4. In Figure 1 are given 400 two-attribute points that represent two classes A and B. The members of the class A are marked by the symbol '+' and the members of the class B are marked by the symbol '*' (the shaded region). In Figure 2 and 3 are given the classification functions obtained by applying the original and extended list of the variables, respectively. Figure 2 shows that the original discriminant analysis cannot successfully solve such a problem. The number of misclassified points using this approach is 169. But, by adding the new variable that is generated as the Yager's $T-$ norm with parameter equal to 2 we can get the results shown in the Figure 3. The number of missclassified points by this approach is only 19.

6 Conclusion

Choquet integral is promising tool for MVSA in the presence of interaction between variables. The logical representation of Choquet integral enable usage of the existing techniques of MVSA in a much more efficient way. In this paper this is illustrated on the problems of discriminant analysis, which could not be solved on the classical approach by linear discriminant function. The classical linear discriminant analysis is applied in the case of presence of interaction among the variables by extending the list of the variables. The new "generalized" variables are logical expressions on all combinations of original variables with AND operators, defined as $T-$norm functions. Since there is unlimited number of $T-$norm functions and different functions give different classification results, we used a parameterised $T-$norm that we can change and get different result. In our future work we will try to define the procedure for determining the parameter of the $T-$norm that minimizes the misclassification error rates and thus produces the best classification results.

References

[1] G. Choquet, *Theory of capacities*, Ann. Inst. Fourier **5** (1953), 131-295.

[2] D. Denneberg, *Non-Additive Measure and Integral,* Kluwer, Boston, (1994).

Figure 1: Class A points ('+' symbol) and B points ('*' symbol)

Figure 2: Discriminant analysis using original approach

Figure 3: Discriminant analysis using logical representation

[3] M. Grabisch, *Fuzzy integral in multicriteria decision making*, Fuzzy Sets and Systems, **69** (1995), 279-298.

[4] M. Grabisch, *Pattern Classification and Feature Extraction by Fuzzy Integral*, EUFIT'95, 3th European Congress on Intelligent Techniques and Soft Computing, Aachen, (1995), 1465-1469.

[5] M. Grabisch. *The application of fuzzy integral in multicriteria decision making*, European Journal of Operational Research, **89** (1996), 445-456.

[6] M. Grabisch, *k-order additive discrete fuzzy measures and their representation*, Fuzzy Sets and Systems, **92** (1997), 167-189.

[7] T. Mirofushi, M. Sugeno, *An interpretation of fuzzy measure and the Choquet integral as an integral with respect to fuzzy measure*, Fuzzy Sets and Systems **29** (1989), 201-227.

[8] T. Murofushi and M. Sugeno, *Some quantities represented by the Choquet integral*, Fuzzy Sets and Systems, **56** (1993), 229-235.

[9] E.Pap, *Null-Additive Set Functions*, Kluwer, Dordrecht, (1995).

[10] D. Radojević, *Generalization of discrete Choquet Integral*, Submitted for publication to Fuzzy Sets and Systems.

[11] M. Sugeno, *Theory of fuzzy integrals and its applications*, Thesis, Tokyo Institute of Technology, (1974).

[12] W. Dillon, M. Goldstein, *Multivariate Analysis - Methods and Applications*, John Wiley & Sons, New York, (1988).