Classification Model based on Choquet Integral Discriminant

Yaqin Qiao*, Xuezhen Dai, Zaibin Chang
Public Department, Xi’an Traffic Engineering Institute, Xi’an 71000, China
947394171@qq.com
aleibing@ultrawise.com.cn

Abstract. At present, the classification method based on nonlinear integral has been widely used. In this paper, Choquet integral value is used as a nonlinear integral discriminant, and a new discriminant classification model is proposed. The classification model maps the data sets which cannot be linearly separable in low dimensional space into high-dimensional space by nonlinear integral discriminant, and then classifies them by Choquet integral. Because there are too many parameters of classification model in the process of classification, particle swarm optimization is used to optimize the parameters, and then combined with the idea of support vector machine to classify the model. Finally, through the MATLAB programming experiment, the classification model proposed in this paper is compared with the SVM classification model, and the experimental results show the feasibility and effectiveness of the model.

1. Introduction
In machine learning, statistics and pattern recognition, classification is an important research content. Given a training set \( T = (X_{\text{new}}, Y_{\text{new}}) \), \( X_{\text{new}} \) is an example set, including \( m \) samples and \( n \) attributes. \( Y_{\text{new}} \) is the category set to which the sample belongs. Classification is to determine which category a new sample belongs to under the training set training model. In essence, this classification is an optimization problem and has been widely used in knowledge discovery and data mining. At present, many methods have been proposed to solve the classification problem, such as decision tree [1,2], Bayesian network [3], neural network [4], support vector machine [5].

At present, the classification method based on nonlinear integral has been widely used and achieved some good results. Reference [7] used fuzzy integral in the field of image processing. Grabisch and Sugeno proposed to use fuzzy t-conorm integral for classification [8]. Based on the possibility theory, the Choquet integral is used to classify statistical patterns with non additive measures. Then, an optimal classification model based on non additive measure is proposed [10]. A 2-additive classification with feature selection is proposed in [11]. Literature [12] gives a classification method by Choquet integral and logistic regression. In [13], Choquet integral is used for supervised classification. Classification can also be achieved by using weighted Choquet integral projection [14] or directly classifying data using generalized signed fuzzy measures [15]. We can take Choquet integral [16] as a special nonlinear integral and Choquet integral value as nonlinear integral discriminant to construct a new classifier. Wang and his collaborators made outstanding contributions in this field [13,15,17,20].

Based on Choquet integral value as a nonlinear integral discriminant, a new discriminant classification model is proposed in this paper. The idea of this classification model is that the low-dimensional data sets which can not be linearly separable are mapped into high-dimensional space by nonlinear integral discriminant, and then classified by Choquet integral. In the process of classification,
because the parameters of classification model are too many, the optimization of the model is carried out by combining the idea of support vector machine. This model not only solves the data set which cannot be linearly separable in low dimension, but also greatly simplifies the complexity of the model.

The main research contents of this paper are as follows: firstly, the concept of assignment measure and Choquet integral is given. Secondly, Choquet integral value and discriminant of taking Choquet integral value as nonlinear integral are given. Then a classification model based on Choquet integral discriminant is proposed. Finally, through MATLAB programming experiment, the classification model and support vector machine \cite{6} classification model are calculated. The experimental results show that the model is feasible and effective.

2. Related works

Given a training set \( T = (X_{m \times n}, Y_{m \times 1}) \), where the attribute set is \( X = \{x_1, x_2, \ldots, x_n\} \), there are a total of \( m \) samples. Each sample contains the values of all attributes. The value of an attribute is numerical and described by \( n \)-dimensional vector \( f = (f(x_1), f(x_2), \ldots, f(x_n)) \) which is called characteristic attribute. The set of all possible values for a class \( \{Y_1, Y_2, \ldots, Y_m\} \) is represented by \( Y \), where each \( Y_i \) \((k = 1, 2, \ldots, m)\) represents a specified class. Therefore, the \( j \)-th sample record consists of the \( j \)-th observation value for all characteristic attributes and classification attributes, and there are \((f_j(x_1), f_j(x_2), \ldots, f_j(x_n), Y_j) \), there into \( j = 1, 2, \ldots, n; k_j \in \{1, 2, \ldots, m\} \).

The purpose of classification is to build a classifier from the training set. When the feature attributes have new samples, we can use the model to determine the class of the new samples. Based on this, we propose a classification model based on Choquet integral discriminant (CMCID).

**Definition 1.** Let \( X = \{x_1, x_2, \ldots, x_n\} \) be a nonempty finite set of attributes and \( P(X) \) be the power set of \( X \). A signed efficiency measure \( \mu \) on \( P(X) \) is a set function \( \mu : P(X) \rightarrow (\infty, -\infty) \), with \( \mu(\phi) = 0 \).

**Definition 2.** Let \( f = \{f(x_1), f(x_2), \ldots, f(x_n)\} \), \( \mu \) be a signed efficiency measure on \( P(X) \). The Choquet integral of \( f \) with respect to a signed efficiency measure \( \mu \) on \( X \), it is defined as follows

\[
(C) \int f d\mu = \sum_{i=1}^{n} [f(x^*_i) - f(x_{i-1}^*)] \mu([x^*_i, x^*_{i+1}, \ldots, x^*_n])
\]

Where \( f(x^*_0) = 0 \), \( x^*_1, x^*_2, \ldots, x^*_n \) is a permutation of \( x_1, x_2, \ldots, x_n \) such that \( f(x^*_1) \leq f(x^*_2) \leq \cdots \leq f(x^*_n) \). Choquet integral can be calculated by:

\[
(C) \int f d\mu = \sum_{j=1}^{2^m-1} z_j\mu_j
\]

Where \( m_j = m(j_1, \ldots, j_m) \) if \( j \) is expressed in binary digits as \( j = j_1 2^m + j_2 2^{m-1} + \cdots + j_m 2^0 \) and if \( z_j > 0 \) or \( j = 2^m - 1 \), \( z_j = \min_{(x^*_j) \in \{f(x^*_j)\}} \max_{(x^*_j) \in \{f(x^*_j)\}} (f(x^*_j)) \); otherwise \( z_j = 0 \) for \( j = 1, 2, \ldots, 2^m - 1 \).

Here, \( \text{frc}(j / 2^m) \) equals the fractional part of \( j / 2^m \), so it can be rewritten by the following simple one:

\[
|i| \{\text{frc}(j / 2^m)\} \in [1, 2^m - 1] = \{i | j = 1\} \{i | \text{frc}(j / 2^m) \in [0, 2^m / 2^m - 1]\} = \{i | j = 0\}
\]

The meaning of this new substitutive formula shows that we can apply the values of in linear function to depict the value of \( m \) the Choquet integral.

3. Classification Model based on Choquet Integral Discriminant

Here is an example of a two-dimensional space. Explain the geometric meaning of points where Choquet integral is equal. Let \( X = \{x_1, x_2\} \), assign measure \( m(\{x_1\}) = 0.2, m(\{x_2\}) = 0.5, m(X) = 1 \), \( f(x_1) = y \), and definition 2 to obtain Choquet integral of \( f \). If \( x \) and \( y \) change in two-dimensional space, then the Choquet integral of \( f \) will change respectively. What is the graph of all points
where the Choquet integral value is equal to the constant value $c$?

From above, we can get it

$$
c = (C)\hat{f}dm = \int (m(X) - m(x_1))x + m(x_2)y, \quad x ? y
\int (m(X) - m(x_1))y + m(x_2)x, \quad x > y.
$$

(3)

By introducing the assignment measure into the above formula, we get the following results:

$$
0.5x + 0.5y = c, \quad x ? y
0.8y + 0.2x = c, \quad x > y.
$$

(4)

If $c = 1$, then

$$
0.5x + 0.5y = 1, \quad x ? y
0.8y + 0.2x = 1, \quad x > y.
$$

(5)

It shows that points with equal Choquet values form a broken line $L$ in 2-dimensional space, as shown in Fig.1. In other words, all points with equal Choquet values can be projected to a point on $y = x$. When the two kinds of points are classified, the classification results can be obtained by the projection points on $y = x$. This method is like a linear discriminant, but the classification boundary is a broken line rather than a straight line. The line $y = x$ also known as the projection axis. Generally speaking, the key is to select the appropriate projection axis and assign the measure. Of course, the projection axis of $y = x$ should not be fixed, but another line can be selected. Therefore, the following concepts are introduced.

![Fig.1 Points with equal Choquet integral values form a broken line in the plane](image)

**Definition 3.** Let be $f = \{f(x_1), f(x_2), L, f(x_n)\}$ be a function defined on $X = \{x_1, x_2, L, x_n\}$, $a = (a_1, a_2, L, a_n), b = (b_1, b_2, L, b_n), m$ the assigned measure on $P(X)$. Then $(C)\hat{f}(a + bf) dm$ is called weighted Choquet integral of $f$ about $a, b$, where, $(a + bf)(x_i) = a_i + bf(x_i)$. The line $a_1 + bx_1 = a_2 + bx_2 = L = a_n + bx_n$ is called the weighted axis of $n$-dimensional space about $a, b$.

**Notes:**

1. If $a = (0, 0, L, 0), b = (1, 1, L, 1)$, then the weighted Choquet is same as in the definition 1, and the weighted axis is $x_1 = x_2 = L = x_n$.

2. The weighted Choquet integral can be calculated by $(C)\hat{f}(a + bf) dm = \sum_{j=1}^{n} z_j m_j$, where $z_j$ as follows, if $z_j > 0$ or $j = 2^s - 1$, $z_j = \min_{j \in \{0/2^s \mid 1, 2^s\}} (a_i + bf(x_i))$, otherwise $z_j = 0$, for $j = 1, 2, L, 2^n - 1$. 

3
In the following, the weighted Choquet integral is called Choquet integral for short.

3. In \( n \)-dimensional space, the graphic of all points whose Choquet integral equals to a constant value \( c \) is called Choquet hyperolane in the following. That is, the Choquet integral of each point on Choquet hyperolane is equal. Each Choquet hyperolane points the weighted axis \( a_1 + b_1 x_1 + a_2 x_2 = L = a_n + b_n x_n \) on one point, so the weighted axis is called projection axis.

\[ \text{Definition 4.} \quad \text{From } (C) \Rightarrow (a + b f)dm = \sum_{j=1}^{2^{n-1}} z_j m_j, \text{ giving the following map: } F : R^n \rightarrow R^{2^n - 1} \]

\[ F(f) = (z(1), z(2), \ldots, z(2^n - 1)), \text{ where } z(j) \text{ is the coefficient under } m_j, j = 1, 2, \ldots, 2^n - 1. \text{ The map } F \text{ is called the ascending dimension mapping based on Choquet integral. It is called a raised dimensional mapping space based on Choquet integral. Under the given weighted axis, the points in the } n \text{-dimensional space are mapped to the } 2^n - 1 \text{-dimensional space through the ascending dimension mapping space.} \]

So the classification model based on Choquet integral discriminant (CMCID) is given as follows:

**Input:** \( X \), training set

1. (1) Given training set

\[ T = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_m, Y_m)\} \in (R^n \times R), \text{ where, } X_i \in R^n, Y_i \in \{-1, 1\}, i = 1, 2, \ldots, m \]

2. Let \( p \) is the maximum number of iterations. Randomly generate \( M \) individuals, \( a, b \) as the initial population. Where \( a \) and \( b \) the coefficient representing the weighted axis.

3. (3) In the given initial \( a, b \), the new training set \( T' \) is generated by dimension increasing mapping \( F \).

\[ T' = \{(\Phi(X_1), Y_1), (\Phi(X_2), Y_2), \ldots, (\Phi(X_m), Y_m)\} \in (R^{2^{n-1}} \times R). \]

4. (4) Constructing and solving convex quadratic programming problems

\[ \min \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} Y_i Y_j (\Phi(X_i), \Phi(X_j)) \alpha_i \alpha_j - \sum_{i=1}^{m} \alpha_i \]

\[ \text{s.t. } \sum_{i=1}^{m} Y_i \alpha_i = 0, \alpha_i \geq 0, i = 1, 2, \ldots, m \]

Get a solution \( \alpha^* = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_m^*)^T \).

5. (5) Solving, \( \mu = \sum_{i=1}^{m} \alpha_i^* Y_i \Phi(X_i) \).

6. (6) The Choquet integral value at each point is calculated. The distance between the two types of data on the projection axis is calculated. \( d \) is taken as the individual fitness.

7. (7) Check whether to meet the termination condition (the maximum number of iterations), if holds, output \( a, b \) and \( \mu \) the maximum fitness value of the individual, and then turn to step (9). Otherwise, proceed to step (8).

8. (8) Through mutation, crossover, select operator to get a new population, return(2).

9. (9) Input the test data, according to \( (C) \int f d\mu = \sum_{j=1}^{2^{n-1}} z_j \mu_j \) and Choquet integral value \( c \) determine its classification.

**Output:** the class of test set

4. Applications

Example 1. Two dimensional data records are shown in Fig.2 and are required to be divided into two categories. We use four algorithms to classify them (solved by programming in MATLAB): Fig.3 shows the classification results of support vector machine (linear kernel). Fig.4 shows the classification result of support vector machine (Gaussian kernel). Fig.5 is the classification result of amid algorithm proposed in this paper (Choquet integral). As can be seen from Fig.4, the boundary of support vector machine Gaussian kernel function has certain fitting property. Fig.5 shows that the given classification algorithm can not only separate nonlinear data, but also has good classification.
effect for this type of data.

Example 2. The three-dimensional data records are shown in Fig.6, which are required to be divided into two categories. We use four algorithms to classify them (solved by programming in MATLAB): Fig.7 shows the classification results of support vector machine (linear kernel). Fig.8 shows the classification result of support vector machine (Gaussian kernel). Fig.9 is the classification result obtained by amid algorithm proposed in this paper. As can be seen from Fig.8, the boundary of support vector machine Gaussian kernel function has a certain degree of fitting. In Fig.9, the nonlinear data is separated by a fold surface. For this kind of data, the algorithm proposed in this paper has a good classification effect.
Example 3. Iris data set, which is composed of 4-dimensional vectors, represents the characteristics of three different types of petals and contains 150 samples. In this paper, we choose the second and third data, including 100 nonlinear separable sample points. In reference [13], the number of error classifications is 3. Using the optimal model given in this paper, we can see from Fig.10 that the number of error classifications reduces the number of iterations, and the final number of error classifications is 2, which indicates that the given method is effective for data classification. Fig.11 shows the Choquet integral value obtained from the second and third class data of iris data set through the algorithm proposed in this paper. The ordinate is the integral value of Choquet and the abscissa is the number of values.

5. Conclusion
In this paper, a new data classification model is proposed by using Choquet integral discriminant. The given model and algorithm are effective and useful from the application examples in this paper. Next, we will further study and improve the algorithm to solve the problem of complex data classification.

Acknowledgments
This work was financially supported by Project of science and Technology Department of Shanxi Province (2020JQ-923).

References
[1] J. R. Quinlan. Induction of Decision Tree. Machine Learning, 1986(1):81-106.
[2] J. R. Quinlan. C4. 5: Programs for Machine Learning. [S. 1. ]; Morgan Kaufman, 1993.
[3] T. Leonard, JSJ Hsu. Bayesian methods: an analysis for statisticians and interdisciplinary
researchers. Cambridge University Press, 1999.
[4] S. I. Gallant. Perceptron-based learning algorithms. IEEE Transactions on Neural Networks, 1990, 1(2):179-191.
[5] C. Cortes, V. Vapnik. Support-vector networks. Machine Learning, 1995,20, 273-297.
[6] Suykens J A K, Vandewalle J. Least Squares Support Vector Machine Classifiers[J]. Neural Processing Letters, 1999, 9(3):293-300.
[7] J. M. Keller, H. Qiu, H. Tahani. The fuzzy integral and image segmentation. In North Amer. Fuzzy Information Proc. Soc. , pages 324-339, New Orleans, June 1986.
[8] M. Grabisch, M. Sugeno. Fuzzy integrals and dual measures: application to pattern classification. InSino-Japan Joint Meeting on Fuzzy Sets & Systems, Beijing, China, October 1990.
[9] M. Grabisch, M. Sugeno. Multi-Attribute Classification Using Fuzzy Integral. IEEE International Conferenceon Fuzzy System, San Diego, 8-12 March 1992, pp. 47-54.
[10] M. Grabisch and J. M. Nicolas. Classification by Fuzzy Integral: Performance and Tests. Fuzzy Sets System, 1994, 65(2-3): 255-271.
[11] L. Mikenina, H. J. Zimmermann. Improved Feature Selection and Classification by the 2-Additive Fuzzy Measure. Fuzzy Sets and Systems,1999, 107(2): 197-218.
[12] A. Tehrani, W.Cheng, K. Dembczynski, E. Hüllermeier. Learning monotone nonlinear models using the Choquet.Mach Learn, 2012, 89:183-211.
[13] M. Grabisch. Fuzzy Integral for Classification and Feature Extraction in Fuzzy Measures and Integrals: Theory and Applications.Physica Verlag, 2000,415-434.
[14] K. Xu, Z. Wang, P. Heng, K. Leung. Classification by Nonlinear Integral Projections. IEEE Transactions on Fuzzy Systems, 2003, 11(2): 187-201.
[15] J. Chu, Z. Wang, Y. Shi. Analysis to the Contributions from Feature Attributes in Nonlinear Classification Based on the Choquet Integral. 2010 IEEE International Conference on Granular Computing (GrC), San Jose, 14-16 August 2010, pp. 677-682.
[16] G. Choquet, Theory of Capacities. Annales de l’Institut Fourier, 1954, 5:131-295.
[17] Z. Wang, K. S. Leung, G. J. Klir. Applying fuzzy measures and nonlinear integrals in data mining. Fuzzy Sets and Systems, 2005, 156(3): 371-380.
[18] N. Yan, Z. Chen, Y. Shi, Z. Wang, G. Huang. Using Non-Additive Measure for Optimization-Based Nonlinear Classification. American Journal of Operations Research, 2012, 2, 364-373.
[19] R. Yang, Z. Wang, P. A. Heng, K.S. Leung. Classification of heterogeneous fuzzy data by Choquet integral with fuzzy-valued integrand. IEEE Trans. Fuzzy Syst, 2007, 15(5):931-942.
[20] Z. Wang, R. Yang, K. Leung. Nonlinear Integrals and Their Applications in Data Mining. New Jersey: World Scientific Publishing, 2010.