IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Generalized discriminant classification model in view of Choquet integral

Generalized discriminant classification model in view of Choquet integral

Xuezhen Dai[1], Zhenni Hu[2] and Zaibin Chang[2,*]

Xi'an Traffic Engineering Institute, Shannxi xi’an, China

*Corresponding author e-mail: daixuezhenqq@126.com

Abstract. Fisher discriminant analysis is a common technique for data processing. The paper proposes a new generalized discriminant classification model in view of Choquet integral. Firstly, through the deep analysis of data classification with regard to Choquet integral, the weighted axis in n-dimensional space is introduced, and each data point in n-dimensional space is corresponding to a point in (2n-1)-dimensional space according to the given weighted axis. Secondly, using the Fisher criterion, a new classification model in view of Choquet integral is presented. Moreover, the algorithms for solving the classification model are built. Finally, some experimental outcomes illustrate the feasibility of the algorithm for solving classification problems.

1. Introduction

Classification plays a vital role in areas of machine learning, statistics, and pattern recognition. Given a data set, which includes *n* attributes, *m* classes, *l* records of each attributes, and the corresponding classes, a classification is an approach to distinguishing which class includes a new record of attributes.

Recently, the classification methods with regard to nonlinear integrals are studied and some encouraging results have been obtained. In these methods, using a nonlinear integral to fuse the values of a new record and to determine which class it belongs to. Paper [1] uses the fuzzy integral in the field of image processing. Grabisch and Sugeno propose to apply fuzzy *t*-conorm integrals to classification[2]. Paper [3] adopts the possibility theory on the Choquet integral. Paper [4] proposes the 2-additive classification with feature selections. Paper [5] gives a classification method by using Choquet integrals and Logistic regression. Paper [6] uses Choquet integral in supervised classification. We can use the Choquet integral[9] as a special nonlinear integral and make it to be a key role in classification. In particular, Wang and his collaborators make an outstanding contribution in this field [6-8,10].

In this paper, based on the Fisher linear discriminant and the Choquet integral, a novel discriminant analysis model for classification is proposed. The idea is using fisher criterion to get the best classification boundary in a new high dimensional space, which is obtained by the ascending dimension mapping. Moreover, being the dimension of new space is very high, the solving of classification boundary is transformed to solve the combination of the each record in the record set according to the reproducing kernel theory. By taking this approach, the amount of variable parameters could be minimized from 2^n+n-1 to 3*n*, which greatly simplifies the complexity of the problem.

The paper is outlined in the following way. In Section 2, coming after the introduction part, a signed efficiency measure and Choquet Integrals are described. Section 3 proposes new discriminant analysis

IOP Conf. Series: Materials Science and Engineering **490** (2019) 042042 doi:10.1088/1757-899X/490/4/042042

models and algorithms based on Choquet integral. In section 4, experimental results illustrate the feasibility of the algorithm. Eventually, the last section is to conclusion part.

2. Related works

Let $X=\{x_1,x_2,\cdots,x_n\}$ be the attributes set. In classification, a data set consisting of l example records, called training set, is given. Each record contains the values of feature attributes and the corresponding class. Positive integer l is called the data size. The values of attributes are numerical, and described by an n-dimensional vector, $f=(f(x_1),f(x_2),\cdots,f(x_n))$. The range of attributes is called the feature space. The set of all possible values of classes, $\{C_1,C_2,\cdots,C_m\}$ is denoted by C, where each $C_k, k=1,2,\cdots,m$, refers to a specified class. Thus, the j-th sample record consists of the j-th observation for all feature attributes and the classifying attribute, and is denoted by $(f_j,C_{k_j})=(f_j(x_1),f_j(x_2),\cdots,f_j(x_n),C_{k_j}),\ j=1,2,\cdots,l,$ where k_j belongs to {1,2,…, m}.

The goal of classification is to construct a classification model expressed by the characteristic attributes. A discriminant analysis model is demonstrated in view of Choquet integral. Some related definitions are introduced as following.

Definition 1. Let $X=\{x_1,x_2,\cdots,x_n\}$ be a nonempty finite set of attributes and P(X) be the power set of X. A signed efficiency measure μ on P(X) is a set function $\mu: P(X)\to(-\infty,+\infty)$, with $\mu(\varnothing)=0$.

Definition 2. Let $f=\{f(x_1),f(x_2),\cdots,f(x_n)\}$. μ be a signed efficiency measure on P(X). The Choquet integral of f with respect to a signed efficiency measure μ on X, denoted by $(C)\int f d\mu$, is defined as follows

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

$$=\sum_{i=1}^{n}[\mu(\{x_i^*,x_{i+1}^*,\cdots,x_n^*\})-\mu(\{x_{i+1}^*,x_{i+2}^*,\cdots,x_n^*\})]f(x_i^*)$$

Where $f(x_0^*)=0$ $(x_1^*,x_2^*,\cdots,x_n^*)$ is a permutation of $(x_1,x_2,\cdots,x_n)$ such that $f(x_1^*)\le f(x_2^*)\le\cdots\le f(x_n)$.

Choquet integral can be calculated by[10]:

$$(C)\int f d\mu=\sum_{j=1}^{2^n-1}z_j\mu_j, \qquad (1)$$

in which $\mu_j=\mu(\cup_{j_i=1}\{x_i\})$ if j is expressed in binary digits as $j_nj_{n-1}\cdots j_1$ for every $j=1,2,\cdots,2^n-1$ and if $z_j>0$ or $j=2^n-1$, $z_j=\min_{i|frc(j/2^i)\in[1/2,1)}(f(x_i))-\max_{i|frc(j/2^i)\in[0,1/2)}(f(x_i))$; otherwise $z_j=0$ for $j=1,2,\cdots,2^n-1$.

Here, $frc(j/2^i)$ equals the fractional part of $j/2^i$, so it can be rewritten by the following simple one:

$\{i\mid frc(j/2^i)\in[1/2,1)\}=\{i\mid j=1\}$, $\{i\mid frc(j/2^i)\in[0,1/2)\}=\{i\mid j=0\}$.

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

Without loss of generality, the number of states of the classifying attribute can be defined as m. Th us, its classification can be labelled as m-classification. Clearly, we can dissemble any m-classification problem into (m-1) 2-classification ones. Therefore, we mainly concentrate on the 2-classification models.

IOP Conf. Series: Materials Science and Engineering **490** (2019) 042042 doi:10.1088/1757-899X/490/4/042042

3. Generalized discriminant models based on Choquet integral

Definition 3. Let $f=\{f(x_1),f(x_2),\cdots,f(x_n)\}$ be a function defined on $X=\{x_1,x_2,\cdots,x_n\}$, $a=(a_1,a_2,\cdots,a_n)$, $b=(b_1,b_2,\cdots,b_n)$, μ is the assigned measure on *P*(*X*).

Then $(C)\int(a+bf)d\mu$ is called weighted Choquet integral of *f* about *a*, *b*. where, $(a+bf)(x_i)=a_i+b_if(x_i)$. The line $a_1+b_1x_1=a_2+b_2x_2=\cdots=a_n+b_nx_n$ is called the weighted axis of n-dimensional space about *a*, *b*.

Notes:

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

2. The weighted Choquet integral can be calculated by

$$(C)\int(a+bf)d\mu=\sum_{j=1}^{2^n-1}z_j\mu_j \qquad (2)$$

where z_j as follows, if $z_j>0$ or $j=2^n-1$,

$z_j=\min_{i|frc(j/2^i)\in[1/2,1)}(a_i+b_if(x_i))-\max_{i|frc(j/2^i)\in[0,1/2)}(a_i+b_if(x_i))$, otherwise $z_j=0$ for $j=1,2,\cdots,2^n-1$.

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 broken-hyperplane in the following. That is , the Choquet integral of each point on Choquet broken-hyperplane is equal. Each Choquet broken-hyperplane joints the weighted axis $a_1+b_1x_1=a_2+b_2x_2=\cdots=a_n+b_nx_n$ on one point, so the weighted axis is called projection axis.

We know that Fisher linear discriminant model is to find a suitable straight line, such that the data points can be separated after projecting to the straight line, see Fig. 1. The data in Fig. 2 is not linear separable, so Fisher linear discriminant method is invalid, but they can be classified by their Choquet integral. That is to say, we can find the suitable integral value *c*, so the Choquet integral of each point in first class is greater than *c*, and the Choquet integral of each point in second class is less than *c*. The geometric meaning is the classification of the projection points, which are the data points project onto the weighted axis. The broken line *L* is the classification boundary in Fig. 2. The method is same as the Fisher linear discriminant method. The difference is that the projection direction is not orthogonal to the projection axis, and the dividing boundary is no longer a hyperplane in given method. If the signed efficiency measure is taken as the additive measure, and the weighted axis is taken as orthogonal to the dividing line, then the classification model in view of Choquet integral is degraded to the Fisher linear classification model. Therefore, the goal of data classification in this way is to find an appropriate a projection axis and a signed efficiency measure, such that the projection point of the two kinds of points can be divided. So the following definitions are introduced.

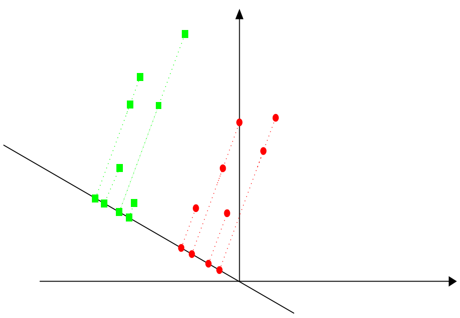

Figure 1. Geometric representation of two-dimensional fisher linear discriminant analysis

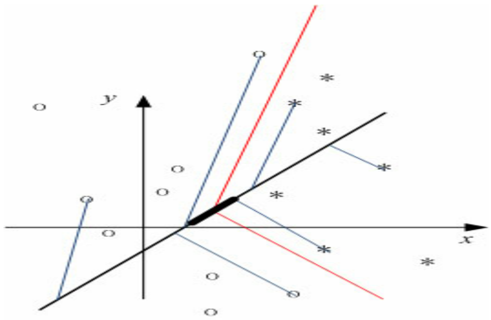

Figure 2. Geometric representation of a two dimensional classification model based on Choquet integral

Definition 4. From (2), giving the following map: $\Phi : R^n \to R^{2^n-1} : \Phi(f) = (z(1), z(2), \cdots, z(2^n - 1),$, where *z*(*j*) is the coefficient under $\mu_j, j = 1,2,\cdots,2^n - 1$ in (2). The map Φ is called the ascending dimension mapping based on Choquet integral.

Under given weighted axis, by ascending dimension mapping Φ, the points in *n*-dimensional space are mapped to 2^n-1 dimensional space. From figure 2, We can see that the original points are not linearly separable in n-dimensional space, but they may become linearly separable in 2^n-1-dimensional space. Let D is the training set of *n*-dimensional space, D1 for class 1 and D2 for class2. l_1 is the number of points in the class 1, l_2 is the number of points in the class 2, $l = l_1 + l_2$. Through the ascending dimension mapping Φ, for given weighted axis, each point *f* in D can be transferred Φ (*f*), and the Fisher criterion function is:

$$J_F(\mu) = \frac{\mu^T S_b \mu}{\mu^T S_\mu \mu} \qquad (3)$$

Where S_b, S_μ is the corresponding class scatter matrix and intra class scatter matrix in $2^n - 1$ space, that is,

$$S_b = (A_1 - A_2)(A_1 - A_2)^T, A_k = \frac{1}{l_k}\sum_{x_i \in X_k} (\Phi(f_i)), k = 1,2$$

$$S_\mu = \sum_{x_i \in X_k} (\Phi(f_i) - A_k)(\Phi(f_i) - A_k)^T, k = 1,2$$

Using fisher method, we can get $\mu = S_\mu^{-1}(A_1 - A_2)$ is the best solution for the given *a*,*b*.

Then the key problem is how to get the best weighted axis, the genetic algorithm is used to obtain it. So the classification algorithm of the generalized discriminant analysis model based on Choquet integral (GDAMBC) is given as follows:

Input: X ,training set.

Out: the class of test set.

(1)Given training set

$$T = \{(f_1, Y_1), (f_2, Y_2), \cdots, (f_l, Y_l)\} \in (R^n \times Y), \text{where, } f_i \in R^n, Y_i \in Y = \{1, -1\}, i = 1,2,\cdots,l.$$

(2)Let *m* is the largest number of evolution. Randomly generated *M* individuals, *a*, *b* as the initial population. a and b indicate the coefficient of the weighted axis.

(3)In the given *a*, *b*, the new training set T' is constructed by Φ.

$$T' = \{(\Phi(f_1), Y_1), (\Phi(f_2), Y_2), \cdots, (\Phi(f_l), Y_l)\} \in (R^{2^n-1} \times Y)$$

(4) Solving, $A_k = \frac{1}{l_k}\sum_{x_i \in X_k} (\Phi(f_i)), k = 1,2$ and $S_\mu = \sum_{x_i \in X_k} (\Phi(f_i) - A_k)(\Phi(f_i) - A_k)^T, k = 1,2$, then we can get the best μ from $\mu = S_\mu^{-1}(A_1 - A_2)$.

(5)Let $Y_1 = \frac{1}{l_1}\sum_{j=1}^{l_1} y(f_j)$, $Y_2 = \frac{1}{l_2}\sum_{j=1}^{l_2} y(f_{j+l_1})$,where $y(f) = \sum_{j=1}^{2^n-1} z_j \mu_j$. If $Y_1 < Y_2$, let $c = (c_1 + c_2)/2$, where $c_1 = \max_{i=1,2,\ldots l1}(y(f_i))$, $c_2 = \min_{i=1,2,\ldots l2}(y(f_i))$. If $Y_1 > Y_2$, let $c = (c_1 + c_2)/2$, where $c_1 = \min_{i=1,2,\ldots l1}(y(f_i))$, $c_2 = \max_{i=1,2,\ldots l2}(y(f_i))$.

(6) Using c to classify each point in D. let *e*=error classification rate, if $e \neq 0$, $d = e$, else if $e = 0$, let $d = -\frac{\mu^T S_b \mu}{\mu^T S_\mu \mu}$. Let *d* as fitness value to optimize the parameters of the weighted axis by genetic algorithm.

(7) Check whether to meet the termination conditions (maximum number of iterations), if holds, get μ and boundary value c, go to 9. If not holds, go to 8.

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

(9)Enter the test data, according to $y(f) = \sum_{j=1}^{2^n-1} z_j \mu_j$ and c to determine its classification.

From the above algorithm, the classification of test data can be obtained. When the number of attributes is large, the dimension of ascending dimension data is more large. In order to solve the

IOP Conf. Series: Materials Science and Engineering **490** (2019) 042042 doi:10.1088/1757-899X/490/4/042042

problem, the reproducing kernel method is used to decrease the number of variable parameters. From reproducing kernel theory, the best projection can be expressed by $\mu = \sum_{i=1}^{l} \alpha_i \Phi(f_i)$, which means the solving of a singed efficiency measure can be transferred by solving of α. In the method, the number of variable parameters is reduced greatly.

Bring $\mu = \sum_{i=1}^{l} \alpha_i \Phi(f_i)$ into (3),the expression of the Fisher criterion function is expressed as:

$$J_F(\alpha) = \frac{\alpha^T S_B \alpha}{\alpha^T S_W \alpha} \quad \text{where} \quad \alpha = (\alpha_1, \alpha_2, \cdots, \alpha_l),\ S_B = (q_1 - q_2)(q_1 - q_2)^T,$$

$$q_k = (\tfrac{1}{l_k}\sum_{x_i \in X_k} (\Phi(f_1), \Phi(f_i)), \tfrac{1}{l_k}\sum_{x_i \in X_k} (\Phi(f_2), \Phi(f_j)), \cdots, \tfrac{1}{l_k}\sum_{x_i \in X_k} (\Phi(f_l), \Phi(f_j)), k = 1, 2.$$

$$S_W = \tfrac{1}{l_1}\sum_{j=1}^{l_1} (p_j - q_1)(p_j - q_1)^T + \tfrac{1}{l_2}\sum_{j=1}^{l_2} (p_j - q_2)(p_j - q_2)^T,\ p_j = ((\Phi(f_1), \Phi(f_j), (\Phi(f_2), \Phi(f_j), \cdots, (\Phi(f_l), \Phi(f_j))^T .$$

We can get the discriminant function is $y(f) = \sum_{i=1}^{l} \alpha_i (\Phi(f), \Phi(f_i)), \alpha = (\alpha_1, \alpha_2, \cdots, \alpha_l)$,which is the weighted Choquet integral value of each point. Similarly, by using genetic algorithm to optimize the weighted axis, and the classification algorithm of a new generalized discriminant analysis model based on Choquet integral is obtained as follows:

Input: X ,training set.

Out: the class of test set

(1) Given training set

$T = \{(f_1, Y_1), (f_2, Y_2), \cdots, (f_l, Y_l)\} \in (R^n \times Y)$ Where, $f_i \in R^n, Y_i \in Y = \{1, -1\}, i = 1, 2, \cdots, l$.

(2)Let *m* is the largest number of evolution. Randomly generated *M* individuals, *a*, *b* as the initial population. a, b indicates the coefficient of the weighted axis.

(3)In the generation of *a*, *b*, the new training set is constructed by Φ .

$$T' = \{(\Phi(f_1), Y_1), (\Phi(f_2), Y_2), \cdots, (\Phi(f_l), Y_l)\} \in (R^{2^n - 1} \times Y)$$

(4)Under the new data T' , solving, $q_k = (\tfrac{1}{l_k}\sum_{x_i \in X_k} (\Phi(f_1), \Phi(f_i)), \tfrac{1}{l_k}\sum_{x_i \in X_k} (\Phi(f_2), \Phi(f_j)), \cdots, \tfrac{1}{l_k}\sum_{x_i \in X_k} (\Phi(f_l), \Phi(f_j)), k = 1, 2$

$$S_W = \tfrac{1}{l_1}\sum_{j=1}^{l_1} (p_j - q_1)(p_j - q_1)^T + \tfrac{1}{l_2}\sum_{j=1}^{l_2} (p_j - q_2)(p_j - q_2)^T,$$

then we can get the best $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_l)$ from $\alpha = S_W^{-1}(q_1 - q_2)$.

In general, in order to ensure the non singularity of K_W, $K_{W} + \gamma I$ is used instead of K_W, where γ is a very small number and *I* is the unit matrix.

(5) Let Y1= $\frac{1}{l_1}\sum_{j=1}^{l_1} y(f_j)$, Y2= $\frac{1}{l_2}\sum_{j=1}^{l_2} y(f_{j+l_1})$, where $y(g) = \sum_{i=1}^{l} \alpha_i (\Phi(g), \Phi(f_i)),\ \alpha = (\alpha_1, \alpha_2, \cdots, \alpha_l)$. If $Y_1 < Y_2$, let $c=(c_1+c_2)/2$, where $c_1=\max_{i=1, 2, \ldots l1},(y(f_i))$, $c_2=\min_{i=1, 2, \ldots l2}(y(f_i))$. If $Y_1 > Y_2$, let $c=(c_1+c_2)/2$, where $c_1=\min_{i=1, 2, \ldots l1},(y(f_i))$, $c_2=\max_{i=1, 2, \ldots l2}(y(f_i))$.

(6)Using *c* to classify each point in D. let *e*=error classification rate, if $e \neq 0$, $d=e$, else if $e=0$, let $d=-\frac{\alpha^T S_B \alpha}{\alpha^T S_W \alpha}$. Let *d* as fitness value to optimize the parameters of the weighted axis by genetic algorithm.

(7) Check whether to meet the termination conditions (maximum number of iterations), if holds, get $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_l)$ and boundary value c, go to 9. If not holds, go to 8.

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

(9)Enter the test data *f*, according to the discriminant function

$y(f) = \sum_{i=1}^{r} \alpha_i (\Phi(f), \Phi(f_i)), \alpha = (\alpha_1, \alpha_2, \cdots, \alpha_r)$ and c to determine its category.

IOP Conf. Series: Materials Science and Engineering **490** (2019) 042042 doi:10.1088/1757-899X/490/4/042042

4.Applications

Example 1. The 2-dimensional data records are given in Fig. 3， ○ for class1 and + for class2. Three algorithms are used to classify them. Fig. 4 shows the results of the application of Fisher linear discriminant function classification, Fig. 5 shows the classification results of the Fisher Gauss kernel function(sigma=0.1925), Figure 5 is the classification result obtained by algorithm 1, *a*=[0.1869, -0.0643], *b*=[0.2029, 1.0000]. From the figures, the boundary of Gauss kernel function has a certain over fitting. Fig. 6 shows that the given classification algorithm can not only separate the nonlinear data, but also has better classification and prediction ability than the Fisher method by Gauss kernel function.

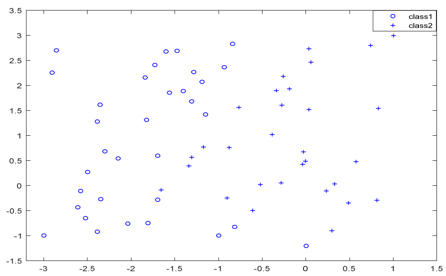

Figure 3. The data records of Example 1

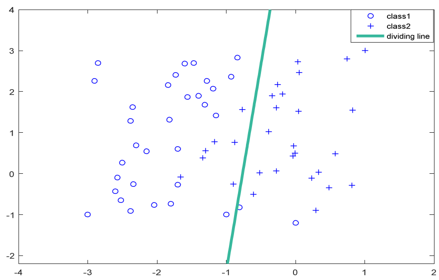

Figure 4. The classification based on Fisher method(linear kernel) of Example 1

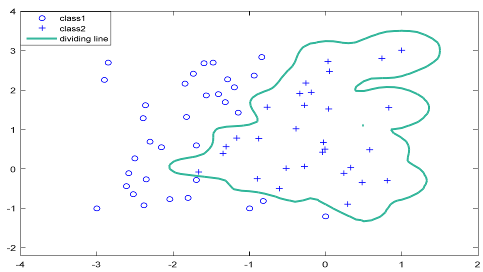

Figure 5. The classification based on Fisher method (Gauss kernel) of Example 1

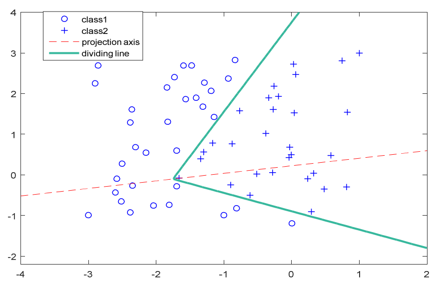

Figure 6. The classification based on GDAMBC of Example 1

Example 2 The 3-dimension data records are given by Fig. 7, ○ for class 1 and + for class 2. Fig. 8 shows the classification results by GDAMBC. Fig.9 shows the fitness value according to the generation. Being the minimum of fitness is negative, so the number of error classification is zero. Fig. 10 shows the classification results by Fisher method(Gauss kernel).

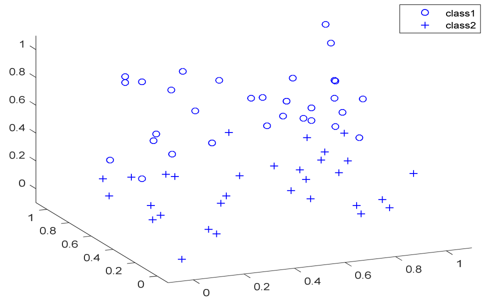

Figure 7. The data records of Example 2

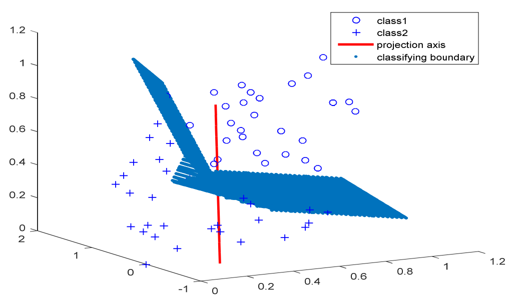

Figure 8. The classification based on GDAMBC of Example 2

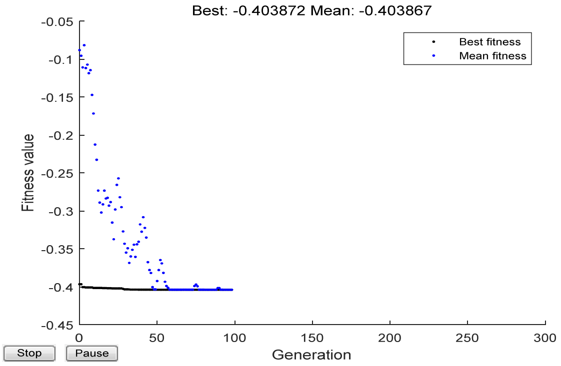

Figure 9. The fitness value according to the generation of Example 2

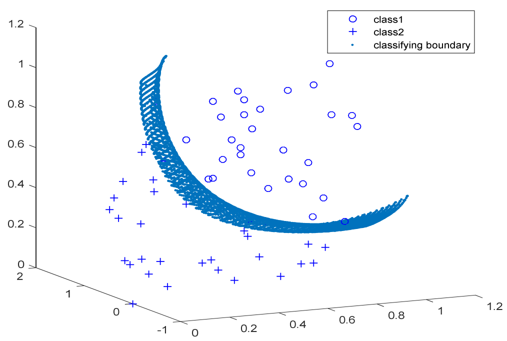

Figure 10. The classification based on Fisher ethod(Gauss kernel) of Example 2

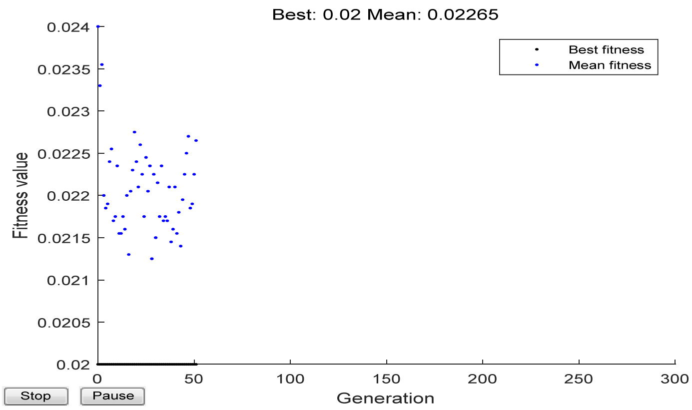

Figure 11. The fitness value according to the generation of Example 3

Example 3. Iris data set. The data set consists of 4 dimensional vector, which indicates the characteristics of 3 different types of petals, containing 150 samples. In this paper, we select the second and the third types of data, including 100 sample points, which are not linearly separable. In paper [6], the error classification number is 3. Using the given optimal models in the paper, From Fig. 11, we can see that the error classification number is decreased by the number of iterations, and the final error classification number is 2. It shows that the given method is effective for data classification.

5.Conclusion

In this paper, using the Choquet integral, new models of data classification are presented. The classification method is like Fisher linear discriminant method. The given models and algorithms are effective and useful from the applications in the paper. Next, we will further study and improve the algorithms for solving complex data classification problems.

Acknowledgments

This work is partially supported by National Natural Science Foundation of China (Grant No. 11501435), Scholarship Fund for Studying Abroad by China Scholarship Council, Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No.12JK0878).

References

[1] 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.
[2] 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.

[3] M. Grabisch, M. Sugeno. Multi-Attribute Classification Using Fuzzy Integral.IEEE International Conferenceon Fuzzy System, San Diego, 8-12 March 1992, pp. 47-54.
[4] 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.
[5] A.Tehrani, W.Cheng, K.Dembczynski, E. Hüllermeier. Learning monotone nonlinear models using the Choquet. Mach Learn, 2012, 89:183-211.
[6] M. Grabisch. Fuzzy Integral for Classification and Feature Extraction in Fuzzy Measures and Integrals: Theory and Applications. Physica Verlag, 2000,415-434.
[7] K. Xu, Z. Wang, P. Heng, K. Leung. Classification by Nonlinear Integral Projections.IEEE Transactions on Fuzzy Systems, 2003, 11(2): 187-201.
[8] J. Chu, Z. Wang, Y. Shi. Analysis to the Contribu-tions 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.
[9] G. Choquet, Theory of Capacities. Annales de l'Institut Fourier, 1954, 5:131-295.
[10] Z. Wang, R. Yang, K. Leung. Nonlinear Integrals and Their Applications in Data Mining. New Jersey: World Scientific Publishing,2010.