Credit Scoring with an Improved Fuzzy Support Vector Machine Based on Grey Incidence Analysis

Baiheng Yi
The School of Economics and Management
Nanjing University of Aeronautics and Astronautics (NUAA)
Nanjing, China
E-mail: ysb900818@126.com

Jianjun Zhu
The School of Economics and Management
Nanjing University of Aeronautics and Astronautics (NUAA)
Nanjing, China
E-mail: zhujianjun@nuaa.edu.cn

***Abstract*—Credit scoring has become increasingly important as the economy recovers, and thus a huge amount of customer credit data is collected by commercial banks and finance corporations. With the rise of machine learning, credit risk can be assessed more easily according to historic data, and support vector machine (SVM) is considered to be an "off-the-shelf" supervised learning algorithm to solve the classification problem successfully. In this paper, an improved fuzzy support vector machine (FSVM) is proposed to overcome the classification problem caused by noise and outliers. First, the notion of mean grey incidence degree is defined to describe the relevance among the training samples. Then, homogeneous and heterogeneous class centers are selected as two reference points in order to discriminate noise and outliers from the valid data. Finally, a fuzzy membership function is given for the purpose of FSVM training. As an empirical study, two credit data set are chosen to demonstrate the feasibility of the model.**

***Keywords*—credit scoring; fuzzy support vector machine; grey incidence analysis**

I. Introduction

In recent years, credit scoring has played an important role in the capital market[1][2], which can be viewed as an approach of assessing the potential risk from the credit applicants. Credit managers are faced with the problem of minimizing credit risk as well as trying their best to increase the amount of credit lending[3]. To solve this, a predictable model for estimating the probability of default is developed, and each applicant is given a credit score to describe his/her credit level. A cutoff value is generally set to avoid the default risk, and any applicant with a credit score lower than this will be rejected. Compared with the past, new credit data collected by bank and other financial institution are growing substantially as the demand for credit increases exponentially. Particularly, economic data from The People's Bank of China shows that the new RMB lending in 2014 was 97,832 billion, almost doubled the data in 2008 with only 49,036 billion. On the other hand, the burgeoning online credit industry stormed in the financial market, more accurately, in 2014, over 0.63 billion investors borrowed and 1.16 billion lent in China[4]. Therefore, credit scoring and risk management are urgently needed.

Many of the traditional manual approaches by professional bank managers are gradually out of date on account of the inefficiency, and the emerging data mining and machine learning technologies[5] are taking its place with obvious advantages like saving time and cost as well as being consistent and objective. A variety of classification methods are implemented in credit scoring cases such as statistical analysis[6], logistic regression[7], neural network[8], and support vector machine (SVM)[9][10].

Among all state-of-the-art algorithms, SVM is believed by many researchers to be the best. The kernel based support vector machine attracts a lot of attention due to its inexpensive computational complexity. Unfortunately, one of the main shortcomings in the application of classic SVM is the sensitivity to noise and outliers, which would cause an over-fitting problem. For example, if the noise is in the training set and is learned by the SVM model, then the classification rule would be followed when predicting the new test sample and as a result that would lead to a poor classifier performance or a lower classification accuracy. To overcome this drawback, Lin[11] proposed Fuzzy SVM method with assigning unique fuzzy membership value to each point according to their distance from the homogeneous class center. Due to the increase in classification accuracy, FSVM is widely discussed and studied. Nevertheless, there still exists room for improvement on reducing the effect of noise and achieving high classification results[12][13].

In this paper, we propose an improved fuzzy support vector machine based on grey incidence analysis[15]. We indicate the disadvantage of adopting Euclidean distance to measure the differences among the training samples, and instead, grey incidence analysis is used to improve this. Moreover, two class centers are defined as the criterion to distinguish noise from valid data, and fuzzy membership value are set based on Ha's method[14] – fuzzy membership value increases from the class center and when the distance to its homogeneous class center reaches a threshold, this sample is treated as a noise.

II. Study on a Novel FSVM for Credit Scoring

For the sake of completeness, this section first gives the basic concept of absolute degree of grey incidence and FSVM, and then introduces the improved FSVM based on grey incidence analysis and its application on credit scoring.

A. Absolute degree of grey incidence[15]

Definition 1. Assume that $x_i = (x_i(1), x_i(2), \cdots, x_i(n))$ is an n - dimension vector, if there exists $x_i^0 = (x_i(1) - x_i(1),\ x_i(2) - x_i(1), \cdots, x_i(n) - x_i(1))$, then x_i^0 is called zero starting point of x_i, denoted $x_i^0 = (x_i^0(1), x_i^0(2), \cdots, x_i^0(n))$.

Definition 2. Assume that two vectors $x_i = (x_i(1), x_i(2), \cdots, x_i(n))$ and $x_j = (x_j(1), x_j(2), \cdots, x_j(n))$ are of the same length, let $a_i = \int_1^n x_i^0 dt$ and $a_i - a_j = \int_1^n x_i^0 - x_j^0 dt$, then

$$\varepsilon_{ij} = \frac{1 + |a_i| + |a_j|}{1 + |a_i| + |a_j| + |a_i - a_j|} \tag{1}$$

is called the absolute degree of grey incidence of x_i and x_j, or absolute degree of incidence for short.

B. FSVM for credit scoring

Consider a given training set $T = \{(x_1, y_1, s_1), (x_2, y_2, s_2), \cdots, (x_l, y_l, s_l)\}$, each training point $x_i \in R^n$ has its corresponding label $y_i = \{+1, -1\}$ and a fuzzy membership $s_i\ (0 < s_i \le 1)$ In credit scoring models, x_i denotes the attributes of customers while y_i is the scoring result for each person. The credit scoring problem is trying to find an optimal hyper-plane $g(x) = (w^T \cdot x + b)$ with largest margin $\|w\|$ which separates the positive (good credit) from the negative (bad credit). The FSVM primal optimization problem can be written as[11]:

$$\begin{aligned} &\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{R} s_i \xi_i \\ &\text{s.t.}\ \ y_i(w^T \cdot x_i + b) \ge 1 - \xi_i, \quad i = 1, 2, \cdots, l \\ &\qquad \xi_i \ge 0, \qquad\qquad\qquad\quad i = 1, 2, \cdots, l \end{aligned} \tag{2}$$

where $C > 0$ is a penalty parameter to errors, since credit data is hardly linear separable in real world, and C makes the algorithm work for the non-linearly separable case as well as making ite less sensitive to outliers. s_i represents the unique contribution of each sample to the classifier hyper-plane. ξ_i is a slack variable to measure the violation of constraints. A smaller s_i means less effect of $\xi_i C$ to the classifier and less importance of the corresponding training point x_i.

By using Lagrange multipliers, the primal problem can be transformed into its dual form:

$$\begin{aligned} &\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j \\ &\text{s.t.}\ \ \sum_{i=1}^{n} \alpha_i y_i = 0, \qquad i = 1, 2, \cdots, l \\ &\qquad 0 \le \alpha_i \le s_i C, \qquad i = 1, 2, \cdots, l \end{aligned} \tag{3}$$

Note that we can solve the dual problem in lieu of the primal. When α_i are found, plug them back into the Karush-Kuhn-Tucker conditions, we can derive the following classifier function:

$$f(x) = \text{sign}(w^T \cdot x + b) = \text{sign}\left(\sum_{i=1}^{l} \alpha_i y_i \langle x_i \cdot x_j \rangle + b\right) \tag{4}$$

Finally, for each new sample with given x_i, its label can be predicted by (4).

C. Limitation of classic fuzzy membership function

Fuzzy membership is a parameter that measures how meaningful the training point is in the classification problem. It is obvious that the importance of some points such as noise is less than those support vectors, so the former ones should be either ignored or misclassified.

In classic fuzzy SVM algorithms, the assigning of a fuzzy membership value based on Euclidean distance can be futile or ineffective in many cases. Intuitively, to measure the differences among samples under the same dimension, this requires that samples are spherically distributed. However, in real-world datasets, not all of them that satisfy spherical distribution, therefore, the use of Euclidean distance is not reasonable. Take a real-world Australian credit data set[16] for example, we extract two attributes from the data for the convenience of showing the problem so that the x and y axis represents continuous numerical attribute 2 and attribute 3, respectively (see Fig. 1). Notice that, for all samples, there is no specific regularity of distribution; points uniformly scatter from a center to the space around. Hence, using Euclidean distance as the distance measure would surely fail in the example above.

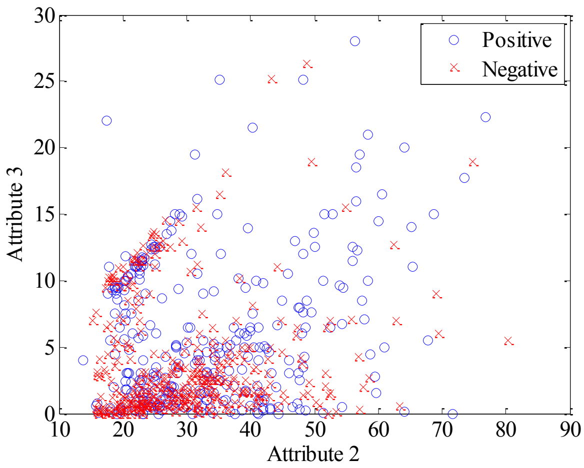

Fig. 1 Original Australian credit data

Although previous work proved that FSVM is more robust to class data set with noise and outliers than many other algorithm, a problem still occurs when the fuzzy membership values are set in a descending order from the class center to the periphery, in which case the importance of both support vectors and outliers would be reduced simultaneously. Since the classifier hyper-plane is merely related to a small amount of support vectors, this would surely lead more errors on classification. Moreover, for the imbalanced data, the noise and outliers could not be effectively split out from the valid data set.

To solve the problem indicated above, we propose a new method of assigning the samples' fuzzy membership value, on one hand the contribution of support vectors to the dual problem (3) can be strongly maintained, and the impact of noise and outliers can be diminished on the other. The total classification accuracy of the credit scoring model can be improved.

D. Mean degree of grey incidence

Given a training set $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_l,y_l)\}\in(R^n\times Y)^l$, where x_i denotes an n-dimension vector $x_i=(x_i(1),x_i(2),\cdots,x_i(n))$, namely l customers and n attributes for each of them in credit scoring problem. $y_i=\{+1,-1\}$ is the corresponding scoring result for each customer. As a general rule, data sets need to be preprocessed before using SVM to establish the model and grey incidence analysis requires the same dimension to compare the relevance of two sequences, we first normalize the training set by the following equation:

$$X_i(k)=\frac{x_i(k)-\min_i x_i(k)}{\max_i x_i(k)-\min_i x_i(k)} \tag{5}$$

Here, $k=1,2,\cdots,n$ and $i=1,2,\cdots,l$. We rewrite the original training set T into a non-dimensional form $S=\{(X_1,y_1),\ (X_2,y_2),\cdots,(X_l,y_l)\}\in(R^n\times Y)^l$. Unlike calculating the differences among samples by Euclidean distance, we use (6)to measure how a sample is related to the homogeneous samples, and the large r_i is, the higher our degree of "confidence" that it is not a noise or outlier.

Definition 3. Assume a non-dimensional training set $S=\{(X_1,y_1),(X_2,y_2),\cdots,(X_l,y_l)\}\in(R^n\times Y)^l$, the number of positive and negative samples are l^+ and l^-, respectively, and $l^++l^-=l$. Define

$$r_i=\begin{cases}\dfrac{1}{l^+-1}\times\sum\limits_{j=1\cap j\neq i\cap y_j=+1}^{j=l}\varepsilon_{ij}, & \text{if } y_i=+1\\[2ex] \dfrac{1}{l^- -1}\sum\limits_{j=1\cap j\neq i\cap y_j=-1}^{j=l}\varepsilon_{ij}, & \text{if } y_i=-1\end{cases} \tag{6}$$

is sample x_i's in-class mean absolute degree of grey incidence, or mean degree of grey incidence for short.

Consider that the mean grey incidence degree might be only in a narrow range, we map its value into $[0,1]$ by using (7), denoted r_i^{norm}. This will in fact come in handy later.

$$r_i^{norm}=\begin{cases}\dfrac{r_i-r_{\min}^+}{r_{\max}^+-r_{\min}^+}, & \text{if } y_i=+1\\[2ex] \dfrac{r_i-r_{\min}^-}{r_{\max}^--r_{\min}^-}, & \text{if } y_i=-1\end{cases} \tag{7}$$

Here, $r_{\min}^+=\min\{r_i\mid y_i=+1\}$, $r_{\max}^+=\max\{r_i\mid y_i=+1\}$, $r_{\min}^-=\min\{r_i\mid y_i=-1\}$, $r_{\max}^-=\max\{r_i\mid y_i=-1\}$.

Based on the concept of mean degree of grey incidence, we will next define the positive and negative class center, and then give the fuzzy membership function.

E. Fuzzy membership function based on two class centers

For the purpose of distinguishing noise and outliers more effectively, we choose positive and negative class centers as two reference points to judge whether the sample is misclassified or not. Obviously, a sample is more likely to be correctly classified when it is close to homogeneous class center and far from the heterogeneous one.

Definition 4. Assume a non-dimensional training set $S=\{(X_1,y_1),(X_2,y_2),\cdots,(X_l,y_l)\}\in(R^n\times Y)^l$, the mean degree of grey incidence of each sample X_i is r_i, if there exists

$$\begin{aligned}X_{cen}^+&=\{X_i\mid r_i=r_{\max}^+\cap y_i=+1\}\\ X_{cen}^-&=\{X_i\mid r_i=r_{\max}^-\cap y_i=-1\}\end{aligned} \tag{8}$$

then X_{cen}^+ is called the positive class center, and X_{cen}^- the negative class center.

X_{cen}^{+} and X_{cen}^{-} are regarded as two reference points, thus, we can calculate the absolute degree of grey incidence between each sample and two reference points, denoted $\varepsilon_{i\_cen}^{+}$ and $\varepsilon_{i\_cen}^{-}$. Assume X_i belongs to +1 class, if $\varepsilon_{i\_cen}^{+} > \varepsilon_{i\_cen}^{-}$, i.e., X_i is more related to its homogeneous class center, then X_i should be considered as valid sample (classified correctly), and vice versa. According to the theory in [14], the fuzzy membership value should be set in ascending order from the class center to the periphery, and when the value is greater than a threshold, the sample might be noise or outliers so that a small s_i should be given. Finally, we give the fuzzy membership function as follows:

$$s_i = \begin{cases} 1 - r_i^{norm} + \sigma, & \varepsilon_{i\_cen}^{+} \ge \varepsilon_{i\_cen}^{-} \cap y_i = +1 \\ 1 - r_i^{norm} + \sigma, & \varepsilon_{i\_cen}^{-} \ge \varepsilon_{i\_cen}^{+} \cap y_i = -1 \\ r_i^{norm} + \sigma, & \varepsilon_{i\_cen}^{+} < \varepsilon_{i\_cen}^{-} \cap y_i = +1 \\ r_i^{norm} + \sigma, & \varepsilon_{i\_cen}^{-} < \varepsilon_{i\_cen}^{+} \cap y_i = -1 \end{cases} \tag{9}$$

Here, $\sigma > 0$ is used to avoid the case $s_i = 0$. For the samples close to their homogeneous class center, r_i^{norm} is comparatively large, so a conversely small $1 - r_i^{norm} + \sigma$ is given. If a sample is far from its homogeneous class center, then r_i^{norm} is generally small, such that $\varepsilon_{i\_cen}^{+}$ and $\varepsilon_{i\_cen}^{-}$ would be compared to identify whether it is a valid sample or noise. Note that in (9), line3 and line4 indicates the sample is noise and the corresponding r_i^{norm} is inherently small, whereas line1 and line2 shows the sample is valid and a conversely large $1 - r_i^{norm} + \sigma$ is given.

To summarize, the overall framework of implementing our proposed methodology on the credit scoring problem can be listed in the following steps:

Step 1: Data preprocessing. Normalize original credit data set T, obtain new training set S.

Step 2: Calculate sample X_i's mean degree of grey incidence r_i in its homogeneous class and map it into interval $[0,1]$, denoted r_i^{norm}.

Step 3: Search the sample X_i with max r_i in two classes, denoted X_{cen}^{+} and X_{cen}^{-} respectively, namely positive and negative class center.

Step 4: Calculate the absolute degree of grey incidence between each sample X_i and two reference points X_{cen}^{+} and X_{cen}^{-}, denoted $\varepsilon_{i\_cen}^{+}$ and $\varepsilon_{i\_cen}^{-}$.

Step 5: Discriminate noise from the valid sample by comparing $\varepsilon_{i\_cen}^{+}$ and $\varepsilon_{i\_cen}^{-}$, then endow fuzzy membership value s_i to each sample.

Step 6: Plug s_i into training set S, obtain new training set with fuzzy membership $S' = \{(X_1, y_1, s_1), (X_2, y_2, s_2), \cdots, (X_l, y_l, s_l)\}$.

Step 7: Select appropriate parameters, train S' with FSVM.

III. EXPERIMENTS

A. Data

In this section, we compare our proposed method GFSVM with classic SVM[5] and Lin's FSVM[11] to verify the effectiveness of GFSVM. Two real-world data sets from UCI Machine Learning Repository[16] are chosen. One is Australian credit approval database, the other is German credit data.

In the Australian set, there are 690 instances with 307 (44.5%) being positive and 383 (55.5%) being negative. Despite the fact that there is no specific attributes description (only the symbolic letters of the former attributes name are used for the sake of trade secret) for Australian credit approval database, it is still commonly used in the credit scoring problem. Details of numerical Australian credit approval database are shown in TABLE 1.

B. The original German set consists of 700 (70%) positive and 300 (30%) negative instances with 20 attributes recording the personal information and financial history of applicants. Part of these attributes are qualitative and cannot be used in machine learning. Hence, researchers reformed it into a numerical version with four more attributes, which is used in our paper. We list the original description of German database in *Training parameters selection*

The selection of SVMs (includes all modified FSVMs) training parameters is still an imperfectly solved problem. Many believe that cross-validation grid search is a feasible approach to pick up C and $gamma$. Besides, kernel is widely used in SVM training, for the reason that, training sample can be mapped into high dimension feature space by only doing inner product. Linear, Polynomial (Poly), and Radial Basis Function (RBF) kernels are selected in our test. We first run the 5-folds cross-validation to get a general range of C and $gamma$, under which value the classifiers cross-validation accuracy can reach the peak. Then, a localized search starts within a narrower range and smaller step. As a result, in German database, we search C in $[1,30]$ with a step of 1 and $gamma$ in $[0.02, 0.2]$ with a step of 0.02. For Australian database, C is in $[1,40]$ with a step of 1 and $gamma$ in $[0.05,1]$ with a step of 0.05.

TABLE 2 .

TABLE 1. NUMERICAL VERSION OF AUSTRALIAN DATABASE

Number	Type	Value
Attribute 1	Nominal	0, 1 (formerly: a, b)
Attribute 2	Continuous	13.75-80.25
Attribute 3	Continuous	0-28
Attribute 4	Nominal	1, 2, 3 (formerly: p, g, gg)

Attribute 5	Nominal	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 (formerly: ff, d, i, k, j, aa, m, c, w, e, q, r, cc, x)
Attribute 6	Nominal	1, 2, 3, 4, 5, 6, 7, 8, 9 (formerly: ff, dd, j, bb, v, n, o, h, z)
Attribute 7	Continuous	0-28.5
Attribute 8	Nominal	1, 0 (formerly: t, f)
Attribute 9	Nominal	1, 0 (formerly: t, f)
Attribute 10	Continuous	0-67
Attribute 11	Nominal	1, 0 (formerly: t, f)
Attribute 12	Nominal	1, 2, 3 (formerly: s, g, p)
Attribute 13	Continuous	0-2000
Attribute 14	Continuous	0-100,000

C. *Training parameters selection*

The selection of SVMs (includes all modified FSVMs) training parameters is still an imperfectly solved problem. Many believe that cross-validation grid search is a feasible approach to pick up C and $gamma$. Besides, kernel is widely used in SVM training, for the reason that, training sample can be mapped into high dimension feature space by only doing inner product. Linear, Polynomial (Poly), and Radial Basis Function (RBF) kernels are selected in our test. We first run the 5-folds cross-validation to get a general range of C and $gamma$, under which value the classifiers cross-validation accuracy can reach the peak. Then, a localized search starts within a narrower range and smaller step. As a result, in German database, we search C in $[1,30]$ with a step of 1 and $gamma$ in $[0.02,0.2]$ with a step of 0.02. For Australian database, C is in $[1,40]$ with a step of 1 and $gamma$ in $[0.05,1]$ with a step of 0.05.

TABLE 2. ORIGINAL ATTRIBUTES IN GERMAN DATABASE

Number	Type	Description
Attribute 1	Qualitative	Status of existing checking account
Attribute 2	Numerical	Duration in month
Attribute 3	Qualitative	Credit history
Attribute 4	Qualitative	Purpose
Attribute 5	Numerical	Credit amount
Attribute 6	Qualitative	Savings account/bonds
Attribute 7	Qualitative	Present employment since
Attribute 8	Numerical	Instalment rate in percentage of disposable income
Attribute 9	Qualitative	Personal status and sex
Attribute 10	Qualitative	Other debtors/guarantors
Attribute 11	Numerical	Present residence since
Attribute 12	Qualitative	Property
Attribute 13	Numerical	Age in years
Attribute 14	Qualitative	Other instalment plans
Attribute 15	Qualitative	Housing
Attribute 16	Numerical	Number of existing credits at this bank
Attribute 17	Qualitative	Job
Attribute 18	Numerical	Number of people being liable to provide maintenance for
Attribute 19	Qualitative	Telephone
Attribute 20	Qualitative	Foreign worker

D. *Results and analysis*

In our experiments, we used the modified Libsvm program developed by Chang and Lin[17] in National Taiwan University. Results on German and Australian data sets are shown in TABLE 3 and TABLE 4, the best train rate and test rate are labeled in bold.

It is widely known that credit scoring is a difficult task because the credit samples are very often not correctly classified. In most cases the nature of an assessment or scoring method is selecting handful number of criteria from the gathered information, and this, however, is not sufficient to reflect every aspect of a client's life. Therefore, a unanimous standard for evaluating whether a credit scoring method is effective or not is comparing the classification accuracy under the same data set, and the method with lower misclassification rates would normally be considered acceptable. It is necessary for the reader to bear this in mind when the following results are explained.

From the results, both in the German and Australian data sets, the maximal train rate and test rate are obtained by our proposed GFSVM. However, different kernels lead to unique performances in two data sets; in the German set, results of RBF are relatively higher than Linear and Poly, while Poly shows the best performance among the three kernels in the Australian set. The value of parameter C varies with both kernels and classifiers and there is no obvious regularity on the changes. Notice that, under the same kernel and classifier, all train rates are higher than test rate, and this explains that the cross-validation marks the highest accuracy among the n - folds (in this paper $n=5$) data sets.

TABLE 3. RESULTS ON GERMAN DATA SET

		C	*gamma*	*train rate* (%)	*test rate* (%)
Linear	SVM	3	—	75.6	74.2
	FSVM	3	—	75.6	74
	GFSVM	7	—	76.6	75.2
Poly	SVM	13	0.08	75.8	70.6
	FSVM	15	0.06	74.4	69.4
	GFSVM	4	0.12	76	72.4
RBF	SVM	13	0.06	76.4	74
	FSVM	29	0.04	75.4	73
	GFSVM	23	0.02	**77.2**	**76**

TABLE 4. RESULTS ON AUSTRALIAN DATA SET

		C	*gamma*	*train rate* (%)	*test rate* (%)
Linear	SVM	3	—	86.5	84.48
	FSVM	3	—	86	84.48
	GFSVM	23	—	86	83.1
Poly	SVM	1	0.95	88.25	84.48
	FSVM	6	0.25	87.5	83.79

	GFSVM	9	0.8	**88.5**	**84.83**
	SVM	33	0.25	87.75	84.14
RBF	FSVM	6	0.25	87.5	84.14
	GFSVM	3	0.95	87.75	83.45

IV. CONCLUSION

In this work, we present an improved fuzzy membership function in FSVM classifier with grey incidence analysis which assigns a different membership value according to the samples degree of grey incidence with both homogeneous and heterogeneous class centers. In this way, noise and outliers can be discriminated more accurately. Furthermore, we apply it to the credit scoring problem and make a comparison with classic SVM and FSVM. To evaluate its applicability and performance, German and Australian data sets are used in our experiments and the results verify the fact that GFSVM achieves better performances than SVM and FSVM.

The future work of this author will seek to tackle the data set imbalanced problem, especially in risk controlling cases like credit scoring, bad records is always the minority class and classifier hyper-plane might shift towards it. Therefore, minority class samples, which we are concerned with more, might be misclassified under the circumstances though the classifier can get high total accuracy led by the majority class.

ACKNOWLEDGMENT

The author acknowledge support from “The National Social Science Fund” (14AZD049), “The National Natural Science Fund” (71171112, 71401064).

REFERENCES

[1] Verbraken T, Bravo C, Weber R, and Baesens B. “Development and application of consumer credit scoring models using profit-based classification measures”, European Journal of Operational Research, vol. 238, pp. 505-513, October 2014.

[2] Chaudhuri A. “Modified fuzzy support vector machine for credit approval classification”, AI Communications, vol. 27, pp. 189-211, 2014.

[3] Harris T. “Credit scoring using the clustered support vector machine”, Expert Systems with Applications, vol. 42, pp. 741-750, 2015.

[4] Tsinghua University. “China online lending industry report”, 2014.

[5] Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2007.

[6] Teng G E, He C Z, Xiao J. “Customer credit scoring based on HMM/ GMDH hybrid model”, Knowledge Information System, vol. 36, pp. 731-747, 2013.

[7] Han L, Han L Y, Zhao H W. “Credit scoring model hybridizing artificial intelligence with logistic regression”, Journal of Networks, vol. 8, pp. 253-261, 2013.

[8] Zhao Z Y, Xu S X, Kang B H and et al. “Investigation and improvement of multi-layer perceptron neural networks for credit scoring”, Expert Systems with Applications, vol. 42, pp. 3508-3516, 2015.

[9] Yu L, Yue W Y, Wang S Y, Lai K K. “Support vector machine based multiagent ensemble learning for credit risk evaluation”, Expert Systems with Applications, vol. 37, pp. 1351-1360, 2010.

[10] Niklis D, Doumpos M, Zopounidis C. “Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines”, Applied Mathematics and Computation, vol. 234, pp. 69-81, 2014.

[11] Lin C F, Wang S D. “Fuzzy support vector machines”, IEEE Transactions on Neural Networks, vol. 13, pp. 464-471, 2002.

[12] Wu Z N, Zhang H G, Liu J H. “A fuzzy support vector machine algorithm for classification based on a novel PIM fuzzy clustering method”, Neurocomputing, vol. 125, pp. 119-124, 2014.

[13] Sevakula R K, Verma N K. Clustering based outlier detection in fuzzy SVM[A]. In: 2014 IEEE International Conference on Fuzzy Systems[C]. Beijing: IEEE, pp. 1172-1177, 2014.

[14] Ha M H, Peng G B, Zhao Q H, and Ma L J. “New fuzzy support vector machine”, Computer Engineering and Applications, vol. 45, pp. 151-153, 2009.

[15] Liu S F, Lin Y. Grey information theory and practical applications. London: Springer-Verlag, 2006.

[16] Bache K, Lichman M. UCI Machine Learning Repository. http://archive.ics. uci.edu/ml, Irvine, CA: University of California, 2013.

[17] Chang C C, Lin C J. “LIBSVM: A library for support vector machines”, ACM Transactions on Intelligent Systems and Technology, vol. 2, pp. 1-27, 2011.