EVALUATING THE CREDIT RISK OF BANK CUSTOMERS BY COMBINED METHOD OF SUPPORT VECTOR MACHINE AND ALGORITHM OF PARTICLES SWARM OPTIMIZATION

MOJTABA RASULI

Tarbiat Modares University, Tehran, Iran
Alzahra University, Tehran, Iran

Abstract - In the problems of risk identification, the imbalance of bank customer's information should be considered and a remedy should be adopted for them; because, it is very effective on the model efficiency. Also, the known algorithms with supervisor like neural networks were used for improvement of results. The time of learning and constructing the model in the neural networks is longer, but, the time of testing it is short. Imbalance of data is very effective on learning in neural networks and accuracy of its results, and the rate of learning the data sample with less size is lower. In this research, a combinative method including support vector machine of particles swarm optimization was presented so that through it, we can discern more number of risks with more accuracy and less cost. The results of this research show that by using this combination, the accuracy of this method has been increased more than the neural network method and basic support vector machine and also, the discernment rate has been optimized in comparison with two main methods.

Keywords - Particles Swarm Optimization, Credit Risk, Customers and Support Vector Machine

I. INTRODUCTION

Nowadays, the combinative background in banking industry has been changed significantly. The cause of this affair in fact includes some factors like new laws, globalization and technology development, conversion of bank services to the product and significant increase of customers' demand. Transformation in the bank activities and increase of banks complexity have created new discussions like fraud in the bank domain. Development of new technologies has opened many ways for fraudulent and criminal persons to be able to commit fraud [2-1]. Creation of a new information system, techniques of fraud identification, in addition to identify and analyze the frauds and scams accomplished in an organization, somehow by cognizing the behavior of users or customers, has tried in predicting their future behavior and reduces the risk of doing the frauds [4]. In the real world, the data are usually imbalanced, in a manner that some classes have more number of samples than other classes. Of course, the imbalance degree (the ratio of the number of majority class samples to the number of minority class samples) may be low or high [5].Most of the classification algorithms assume that distribution of classes is alike and in the event that distribution of classes is imbalanced, these algorithms will have problem in their discernment and practically, they will be tended toward majority class and the minority class in which the number of data is very low, is mostly ignored [6]. Sampling is one of the methods of overcoming on learning problems in imbalanced data [7]. In this method, with the aim of balancing the data set, a series of mechanisms is used for changing the imbalanced data set. The studies show that most of the basic classifications have better act on balanced data and this is a satisfying cause for using the sampling technique [8-9-10-11]. Due to the very direct and indirect costs of fraud, the banks and financial and monetary institutes severely seek to accelerate and increase the speed of act in cognizing the activities of swindlers and defrauders. This affair due to its direct effect on serving the customers of banks and institutes is to reduce the operational costs and be remained as a presenter of valid and reliable financial services [13]. Therefore, applying the techniques of identifying the fraud in order to prevent from deceitfully acts in the banking system is unavoidable. With regard to the issues, the classifier and its structure, each one, has its special difficulties and advantages and no classifier can be found that doesn't have any fault and difficulty. These difficulties can be reduced only by combination of them and the weakness points of one of them can be covered by the strength points of another one. Some reasons which cause lack of cancellation of credit risk of banks, are as follows:

1. Lack of considering all specifications and elements of classification. There are also some specifications which have occurred during special economic situation that prediction of this situation is an impossible affair.

2. Existence of obtrusive specifications and elements that only have caused the increase of calculation volume and have no effect on classification. These elements in addition to increase the calculation volume, have had bad effect on classification and deviate the path of classifiers.

In this paper, a method according to the support vector machine and algorithm of particles swarm optimization will be presented for evaluation of credit risk of customers.

2- The suggested method

As it was mentioned, types of problems can be created in the act of customer's classification and

their classification can make a mistake. To solve the first problem, we need to add the useful and effective specifications in the act of classification that needs the investigative work, brief study of banking system and bank's customers socially. But, two other difficulties can be reduced by computer and calculation algorithms. In this paper, also, the aim is to reduce these two problems. The main problem with support vector machine is the number of those parameters which should be regulated before any type of test. With regard to this issue that there is no specified law for regulation of these parameters, but, correct determination of it causes the success of education. In this type of cases, selection of expert person can be wrong or inappropriate. Using the optimization algorithms like the particles swarm algorithm can be very useful for selection of optimum structure. In this event, each generation of the society will be a unique support vector machine in this form that the particles swarm algorithm should find the appropriate structure of the support vector machine, in this event, the support vector machine itself should be the fitness function. The tests done in this paper consist of:

1. In the first test, at first, the effect of reducing the dimension has been studied by using PCA and we will do the test for reduced dimensions of 5, 7, 10, 12 and 15.

2. The support vector machine optimized by the algorithm of the particles swarm optimization of model. In the continuation, we will study each one of the tests.

3- The suggested method

As we mentioned above, the suggested method has several stages that in the continuation, we explain each one of them.

3-1- Reading the data set

In this stage, the bank's customers with a specificity which has been defined in the data base of EDGAR, are read as input. It can be mentioned that the customers of different banks can be used as input and also, the desirable specificities can be added to it or diminished from it.

3-2- Preparation of data

Since, in the intended data set, the data don't have an alike format, in this stage, they should become uniform and homogenous and after it, the clearance and amalgamation should be done which is practically a difficult and time-consuming work. Elimination of some records and fields from accomplished works is done in this stage and we eliminated many records which have been unnecessary and didn't help in construction of a good classification and identification of bad customers and we reduce the data volume. Sometimes, in the initial data set, there were blank amounts in the fields of some records or in some cases, the amounts allocated to the variables were wrong which were corrected or eliminated. In this stage, despite of clearance and pre-processing of data, a number of customers that had insufficient information, was eliminated from the data set.

3-3- Reduction of dimensions by PCA

PCA is a method without control. This method reduces the dimension in a manner that separability of classes is protected as far as possible. PCA technique is the best method to reduce the dimensions of data in linear form. Namely, by eliminating the inconsequential coefficients gained from this conversion, the lost information is less than the other methods. Of course, application of PCA isn't limited to reduce the dimensions of data and is used in other fields like identification of pattern and discernment of face too [15]. In this method, new coordinates axes have been defined for data and the data are expressed according to these new coordinates axes. The first axis should be placed in a manner that the variance of data is maximized. In this same order, the next axes perpendicular on all previous axes are placed in a manner that the data have the most distribution in that direction. In the following figure, this issue has been shown for two-dimensional data [16].

3-4- Incarnation of data

Since, this set of data includes disconnected information (like marriage status specificity), so, this type of nominal arrays should be converted to the logical ones. The logical array is placed in the path of specificities in the form of zero and one. In the figure (1), for incarnation of this conversion, we showed the data in the form of two distinct classes so that we could observe their distribution.

As we know, there have been 16 specificities in the data set and for drawing them, we reduced them up to two dimensions. Our outcome is this issue that whether the customer is the member of deposit with due date or not. And this issue has been shown in the binary form of yes and no in this figure.

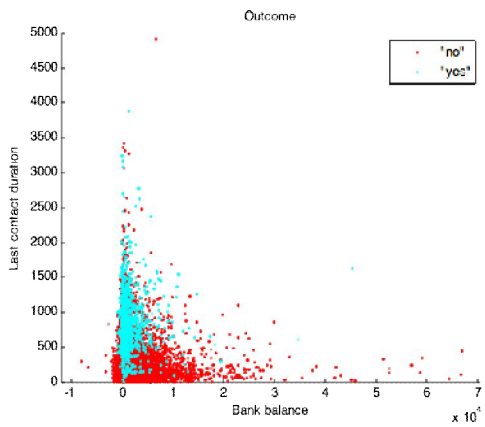

Figure 1: Information and data distribution

3-5- The particles swarm optimization

The particles swarm optimization is an algorithm for strong and efficient evolutionary calculations and it was developed by Kennedy and Eberhart in 1995. This is an optimization method based on population in which population is named group. A group formed

from particles N which are moving around D-dimensional search space. In this state in algorithm, the situation of ith particle in each group is defined in the form of $x_i = (x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{ik})$ and the speed for ith particle of algorithm can be written in the form of $v_i = (v_{i1}, v_{i2}, \dots, v_{ij}, \dots, v_{ik})$ [8]. The outset of doing the particles swarm optimization is in this form that a group of particles (solutions) is created randomly and by updating the generations, they try to find the optimum solution. In the initial stage of giving the amounts, the initial population of particles should be created. In the particles optimization algorithm, each particle in fact shows a solution of problem and a solution, here, is the same path of vehicle that in each step, each path is updated by using two best amounts. The first case is the best situation that the particle already has succeeded for reaching to it. The mentioned situation is known with the name of pbest. If we define the situation of particles in the form of $p_i = (p_{i1}, p_{i2}, \dots, p_{iD})$, the best situation of ith particle will be named P^{best}. If we define the amount of these situations which have been already gained by $\mathbf{pbest}_i$ in the form of $g_i = (g_1, g_2, \dots, g_D)$, we will name the best value of this set, $\mathbf{gbest}_i$.

The initial amounts are given to PSO by random particles population. Then, the algorithm with search is executed for optimum solutions and consequently, it does the generation updating. In each generation, the situation and speed of ith particle are shown by $\mathbf{pbest}_i$ and $\mathbf{gbest}_i$ and are updated by using the relations (1) and (2) [8].

$$v_{id}^{new} = w \times v_{id}^{old} + c_1 \times r_1 \times (pbest_{id} - x_{id}^{old}) + c_2 + r_2 \times (gbest_d - x_{id}^{old}) \quad (1)$$

$$x_{id}^{new} = x_{id}^{old} + v_{id}^{new} \quad (2)$$

In the above relations, r_1 and r_2 are the random numbers in the range of (0,1). c_1 and c_2 are the fixed amounts of acceleration that undertake the control and manner of the round of a movement particle in a generation. The speeds of v_{id}^{old} and v_{id}^{new} in order show the speed of old and new particles. x_{id}^{old} and x_{id}^{new} in order show the current and updated situation of particle. W is the inertia weight which has been defined for controlling the effect of previous speed of particles in a flow. The general performance of particles optimization algorithm can be shown in the form of figure (2).

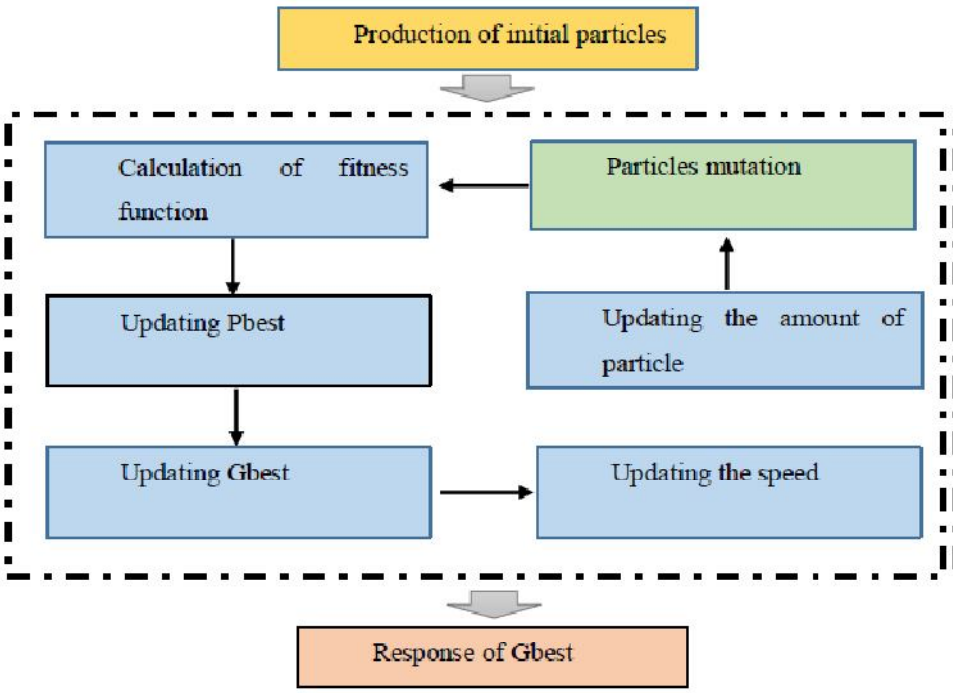

Figure 2- The general performance of PSO algorithm

3-6- The support vector machine

In this stage, the prepared data are educated by the algorithm of non-linear two-class support vector machine. The support vector machine is a binary classification which acts by using the data mapping from the main entrance space to a space with higher dimension for separation of them. This model searches a super-page that its distance with data of two classes is maximum. In this method, it is tried that for gaining the border of classes, a system with minimum capacity or in better words, a system with minimum complexity to be implemented.

Consequently, the support vector machine by using less educational data than competitor methods can estimate the borders of system with an appropriate accuracy, without defacing the system generalizability. In fact, the main aim of the classifier of support vector machine is to achieve a function of $f(x)$ which is the super-page determinative. This super-page separates two classes from each other optimally. Consider the following linear sample set [3]:

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\}, \quad i = 1,2,\dots,l \ , x_i \in R^n, y_i \in \{-1, +1\}$$
(3)

In the above relation, x_1 is the input of problem and y_1 is the outcome corresponding to it and they are considered as the class labels. The general form of separator linear function is shown in the form of relation (4).

$$Specificity = \frac{TN}{TN+FN}$$

$$f(x) = (w.x) + b$$
(4)

In this relation, $w.x$ shows the internal multiplication of vectors w and x. An optimum classification is resulted when the following relation is established.

$$(w.x) + b = 0$$
(5)

The optimum classification under the linear conditions can be gained from the relation (6).

$$f(x) = sgn\left(\sum_{i=1}^{l} \alpha_i^* y_i (x_i.x) + b^*\right) where \ b^* = y_i - \sum_{i=1}^{l} \alpha_i^* y_i (x_i.x_j)$$
(6)

In the suggested method that non-linear classification is used, the non-linear mapping from a low-dimensional sample to a higher dimensional specificity space is utilized. Then, by using Kernel's function of $k(x_i, x_j)$ the linear classification is converted to the non-linear classification in this space. Therefore, the non-linear classification space is gained by using the relation (7) [1]:

$$f(x) = sgn\left(\sum_{i=1}^{l} \alpha_i^* y_i \, k(x_i.x) + b^*\right)$$
(7)

1-The results of tests

To compare the classifiers, some standards and criteria should be introduced. There are different criteria for evaluation and comparison of classifiers with each other. The criteria that we use in evaluation, are used when the classification has been exerted on two classes and the desirable and undesirable classes are discerned. We define these criteria as follows:

The True Positive: This criterion which is also known as TP, shows this issue that what number of those data that belong to the positive class, is discerned true. In other words, when we assume one of the classes positive and another one is assumed negative, this criterion occurs in a time that the classifier discerns the positive classes truly.

The False Positive: This criterion which is also said FP, shows the percent of the number of those data which belong to the negative class and the classifier classifies them positive falsely.

The True Negative: This criterion which is also said TN, shows the percent of those data which have belonged to the negative class and the classifier classifies them truly and belonging to the negative class.

The False Negative: This criterion which is also said FN, expresses the percent of those data which belong to the positive class, but, their classifier attributes to the negative class.

Specificity: It shows the ratio of the true negative to the sum of negative elements:

$$Specificity = \frac{TN}{TN+FN}$$
(8)

Accuracy: It shows the sensitivity which expresses the percent of the ratio of true discernment to the sum of discernments:

$$Accucary = \frac{TP+TN}{TP+TN+FP+FN}$$
(9)

3-7- The results gained from the first test

To evaluate the results gained from this test, we have used according to those criteria that we mentioned above. In the event of reduction of data dimensions to 5 numbers, the results of table 1 have been gained. As it is clear from the table, the accuracy of classification in the combination of type 2 is better than type 1 and an accuracy about 77.22 percent can be considered for quintuple specificity space.

1- The results of data classification with quintuple specificity space

	TN	FN	FP	TP	Spec	Accu
Combination of type 1	57.6763	42.3237	4.1409	95.8591	0.4232	0.7677
Combination of type 2	62.1564	37.8436	6.1672	93.8328	0.4144	0.7722

We have brought the results of data classification with septet specificity space in the table 2. By reduction of specificity space to 7 numbers, the accuracy in the combination of type 2 becomes better than type 1 and it hasn't been changed remarkably in comparison with quintuple specificity space. In this specificity space, we consider the accuracy of combination of type 2 which is equal to 77.99 percent as the accuracy of reduction of dimensions to septet specificity space.

2-The results of data classification with septet specificity space

	TN	FN	FP	TP	Spec	Accu
Combination of type 1	55.1797	44.8203	3.9197	96.0803	0.4482	0.7569
Combination of type 2	62.5792	37.4207	6.1998	93.8002	0.3784	0.7799

In the table 3, we showed the results of data classification with decuple specificity space for two types of combination. In this table, the combination of types 1 and 2 hasn't had much effect on classification accuracy than two previous dimensions. And for decuple specificity space, our accuracy is about 78.18 percent.

3-The results of data classification with decuple specificity space

	TN	FN	FP	TP	Spec	Accu
Combination of type 1	54.5455	45.4545	3.5722	96.4278	0.4545	0.7549
Combination of type 2	62.5793	37.4207	6.2106	0.3742	0.3742	0.7818

We showed that the results of data classification with a space including 12 specificities in the table 4. In this table, both types of combination have a good accuracy than the previous specificity spaces in a manner that this accuracy has been gained for combination of types 1 and 2 in order equal to 77.67 and 78.28 percent.

4-The results of data classification with a space including 12 specificities

	TN	FN	FP	TP	Spec	Accu
Combination of type 1	59.4080	40.5920	4.0717	95.9283	0.4059	0.7767
Combination of type 2	62.7907	37.2093	6.2215	93.7785	0.3721	0.7828

Similarly, the results of data classification with a space including fifteen specificities have been mentioned in the table 5. In this space, also, both combinations have a good accuracy, but, it has been reduced in comparison with a space including 12 specificities and for combination of types 1 and 2, it has been gained in order equal to 78.38 and 87.18 percent.

5-The results of data classification with a space including fifteen specificities

	TN	FN	TP	FP	Spec	Accu
Combination of type 1	60.2537	39.7463	96.5038	3.4962	0.3975	0.7838
Combination of type 2	62.5793	37.4207	0.3742	6.2106	0.3742	.7818

With regard to the results gained from above tables, it can be concluded that reduction of dimension to 12 specificities has the best conditions for suggested methods of this paper. So, hereafter, we change the specificity space to 12 and we also change other variables. So, in the first test, the best amount of accuracy has been gained equal to 78.28 percent.

3-8- The results gained from the second test

The results of changing the structure of this network for the second test have been shown in the table 6. With regard to the results gained from changing the structure of support vector machine to CMC2 state, it doesn't have better performance than other five suggested states and gives an accuracy equal to 50.16 percent to us.

This issue that we use a combination which is the sigmoid activation function of outcome, will give us better results, in a manner that in the table 4, for changing the CMC1 structure, our accuracy has reached to 77.77 percent.

6- The simulation results of support vector machine for combination of type 1 in the first test

	TN	FN	FP	TP	Spec	Accu
CMC1	59.6195	40.3805	4.0825	95.9175	0.4038	0.7777
CMC2	99.7886	0.2114	99.4680	0.5320	0.0021	0.5016
CMC3	58.9852	41.0148	4.0174	95.9826	0.4101	0.7748
CMC4	57.5053	42.4947	4.0065	95.9935	0.4249	0.7675
CMC5	58.5624	41.4376	4.1260	95.8740	0.4144	0.7722
CMC6	58.1395	41.8605	4.0391	95.9609	0.4186	0.7705

So, it can be said that change in the structure of support vector machine has an important role in classification of information and increase of accuracy.

The results of the second test with change in the size of support vector machine for the first test are in the form of table 7. With regard to the results of this table, combination (CMC21) has a higher accuracy than other combinations and the combination of CMC26 gives an accuracy equal to 77.22 percent to us.

7- The results of the second test with change of support vector machine for the first test

	TN	FN	FP	TP	Spec	Accu
CM C21	62.7907	37.2093	6.2215	93.7785	0.3721	0.7828
CM C22	62.1564	37.8436	6.1998	93.8002	0.3784	0.7798
CM C23	62.5793	37.4207	6.2106	93.7894	0.3742	0.7818
CM C24	62.5793	37.4207	6.1998	93.8002	0.3742	0.7819
CM C25	62.5793	37.4207	6.1998	93.8002	0.3784	0.7799
CM C26	62.1564	37.8436	6.1672	93.8328	0.4144	0.7722

Generally, it can be concluded from the second test that change in the structure of support vector machine has much effect on increase or decrease of accuracy of data classification and the highest amount of accuracy in this test is considered equal to 78.28 percent.

CONCLUSION

In this paper, the stages of classifying the bank customer's information including demonstration of documents and reduction of dimensions, construction and combination of classifier and data classification have been discussed and studied with complete details and with regard to the numerous tests which were done in this paper, the following results can be mentioned:

1.The amount of accuracy by using the suggested combinative method has been approximately improved in order equal to eight, three and two percent in comparison with linear regression, decision tree and support vector machine methods. So, combining the methods of support vector machine and decision tree has better performance than each one of them in classifying the bank customers and this result can be known as the strength point of suggested method.

2. The tests showed that in comparison of classifications by suggested method, the classifier of support vector machine with 79 percent has had the most amount of accuracy in the system. This means that combination of support vector machine with particles swarm is appropriate for classifying the bank customers.

According to the accomplished studies, each one of the classifiers has had its advantages and faults and none of them has superiority over another one absolutely.

REFERENCES

[1] Paulius Danena, Gintautas Garsv: Selection of Support Vector Machines based classifiers for credit risk domain, Expert Systems with Applications 42 (2015) 3194–3204

[2] Yu, L., Yue, W., Wang, S., & Lai, K. K. (2010). Support vector machine based multiagent ensemble learning for credit risk evaluation. Expert Systems with Applications, 37(2), 1351–1360.

[3] Yun, L., Cao, Q.-Y., & Zhang, H. (2011). Application of the PSO–SVM model for credit scoring. Seventh International Conference on Computational Intelligence and Security, 47–51.

[4] Zhang, Z., Gao, G., & Shi, Y. (2014). Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors. European Journal of Operational Research, 237, 335–348.

[5] Horta, I. M., & Camanho, A. S. (2013). Company failure prediction in the construction industry. Expert Systems with Applications, 40(16), 6253–6257.

[6] Hsieh, N.-C., & Hung, L.-P. (2010). A data driven ensemble classifier for credit scoring analysis. Expert Systems with Applications, 37(1), 534–545.

[7] Huang, F. (2008). A particle swarm optimized fuzzy neural network for credit risk evaluation. In Proc. of 2008 second international conference on genetic and evolutionary computing, pp. 153–157.

[8] Kennedy, J., Eberhart, R. C., & Shi, Y. (2001). Swarm intelligence. Morgan Kaufmann Publishers.

[9] Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), 2767–2787.

[10] Kim, K., & Ahn, H. (2012). A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach. Computers & Operations Research, 39(8), 1800–1811.

[11] Kong, W., Cheng, W., Ding, J., & Chai, T. (2010). A reliable and efficient hybrid PSO for parameters optimization of LS-SVM in production rate prediction. In 2010international symposium on computational intelligence and design, pp. 140–143.

[12] Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. Expert Systems with Applications, 40(13), 5125–5131.

[13] Lai, K. K., Yu, L., Zhou, L., & Wang, S. (2006). Credit risk evaluation with least square support vector machine. In G.-Y. Wang, J. F. Peters, A. Skowron, & Y. Yao (Eds.), Rough sets and knowledge technology. Lecture Notes in Computer Science (Vol. 4062, pp. 490–495). Berlin Heidelberg: Springer

[14] le Cessie, S., & van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. Applied Statistics, 41(1), 191–201.

[15] Zhang, D., Zhou, X., Leung, S. C. H., & Zheng, J. (2010). Vertical bagging decision trees model for credit scoring. Expert Systems with Applications, 37(12), 7838–7843.

[16] Kennedy, J., Eberhart, R. (1995). Particle swarm optimization. In IEEE international joint conference on neural network Vol. 4, pp. 1942–1948.

★★★