Proceedings of the World Congress on Engineering 2015 Vol I
WCE 2015, July 1 - 3, 2015, London, U.K.

A modified ABC to Optimize the Parameters of Support Vector Machine for Predicting Bankruptcy

Yassine Abbazi, Khalid Jebari, Aziz Ettouhami

***Abstract*—This paper investigated a modified Artificial Bee Colony (ABC) to find the optimal values of Support Vector Machine (SVM) parameters C and σ for predicting bankruptcy. Based on key financial ratios, (ABC-SVM) model is used to find a suitable classification by which non bankrupt firms are separated from bankrupt ones. The results obtained by using our model are benchmarked against the performance of Discriminant Analysis(DA) and logistic regression(LR) on the same data.**

***Index Terms*—Bankruptcy Prediction, Support Vector Machine, Artificial Bee Colony.**

I. Introduction

BANKRUPTCY prediction represented an important research topic in accounting and finance for the last decades. Due its vital interest, many researchers investigated their efforts by analysing and exploring two ways [1]. The first way is based on statistical methods. The first pioneer who inaugurated and enlightened this way is Beaver [2], he designed, by financial ratios, experimental models for examining corporate failures. Followed by Altman [3] who applied multi-discriminant analysis. This study inspired other researchers to utilize multivariate techniques such as logistic regression analysis by Ohlson [4] and probit approach by Zmijewski [5]. The second way explores artificial intelligence namely decision tree [6], [7], fuzzy set theory [8], rough sets theory [9], [10], [11],case based reasoning [12], [13], [14], evolutionary algorithms [15], [16], support vector machine [17], [18] and artificial neural networks [19], [20], [21], [22].

Support vector machines (SVMs), were developed by Boster, Guyon, and Vapnik [23] to provide better solutions to decision boundary than could be obtained using the traditional neural network. Since the new model was proposed Cortes and Vapnik [24]. SVM has been successfully applied to numerous financial applications, including time series forecasting [25], [26], [27] and bankruptcy prediction [28].

Min and Lee [17] stated that the optimal parameter search on SVM plays a crucial role to build a bankruptcy prediction model with high prediction accuracy and stability. To make an efficient SVM model, two extra parameters: C and σ^2 have to be carefully chosen. The first parameter, C, determines the trade-offs between the minimization of the fitting error and the minimization of the model complexity. The second parameter, σ^2, is the bandwidth of the radial basis function (RBF) kernel. Hence, the goal of this study is to propose a model that can determine the optimal parameters (C and σ^2) of SVMs to yield the highest predictive accuracy. Our model (ABC-SVM) is based on SVM and a modified Artificial bee colony is used to optimize the value of the parameters C and σ^2. As the selection by biased roulette wheel may result a premature convergence of the ABC algorithm, we present in this study a selection technique based on the concept of the fitness sharing [29]. The model was tested on the bankruptcy to compare its accuracy with Discriminant Analysis and logistic regression. The remainder of this paper is organized as follows. The basic ideas of support vector machine is reviewed in Section 2. Artificial Bee Colony is presented in section 3. Research design for our model modified artificial bee colony based SVM is proposed in Section 4 to describe its enhanced ideas. An example of experimental study for predicting bankruptcy is illustrated in Section 4 and conclusions are in the last section.

Laboratoire Conception et Systemes (Microlectronique et Informatique); Faculty of Sciences Mohammed V-Agdal;
University UM5A Rabat Morocco
Multidisciplinary faculty El Jadida;
University Chouaib Doukkali, El Jadida Morocco
e-mail: khalid.jebari@gmail.com

II. Support vector machine

SVM is a binary classification method by supervised learning [30]. This method relies on the existence of a linear classifier in a suitable space and the use of kernel functions that allow optimum separation of the data. the aim is to find a linear classifier (hyperplane) that will separate the data and maximize the distances between these two classes. The simplest designing is a linear classification problem obtained by linear combination of the input vector $x = (X_1, ..., X_N)^T$ with w is the normal vector to the hyperplane:
$h(x) = w^T x + w_0$
if $h(x) \geq 0$ than x is decided in class 1 else x in class -1
The main objective is to find a hyperplane which minimizes margin error and is described as a set of support vectors. Finding these vectors from training data is formulated as quadratic optimization problem:

$$Min\frac{1}{2}\|w\|^2 + C\Sigma_{i=1}^{l}\xi_i \tag{1}$$

subject to $y_i\ (w^T\phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0$
where C is a regularization parameter that determines the trade-off between the maximum margin and the minimum classification error. The decision function is defined as $sgn(\phi(x).w$
If training vectors are not linearly separable, they can be represented in a larger (probably infinite) dimensional space by using kernel function:
$K(x_i, x_j \equiv \phi(x_i)^T\phi(x_j)$
Popular kernel functions [31] in machine learning theories are Gaussian (RBF), Polynomial, Linear and Multi layer perceptron. In this paper we have used the Gaussian Kernel:

$$K(x_i, x_j) = exp(-\frac{\|}{x_i} - x_j\|^2 2\sigma^2 \tag{2}$$

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

σ denotes the variance of the Gaussian Kernel. According to the Lagrangian the optimal hyperplane is reformulated into the following problem:

$$max\{L(\alpha) = \Sigma_{i=1}^{l}\alpha_i - \frac{1}{2}\Sigma_{ij}\alpha_i\alpha_j y_i y_j K(x_i, x_j)\} \quad (3)$$

s.t. $\Sigma_{i=1}^{l}\alpha_i y_i$, $0 \leq \alpha_i \leq C$, i=1. 2, ..., l

III. Artificial bee colony

Bees are social insects that live in highly organized colonies [32]. Three croups consist this colony, where each has a specific job: employed bees are associated specific food source, onlooker bees choose a food source according to the dance of employed bees, and scout bees search randomly the food source.

In ABC algorithm, the location of the food source is the possible solution to the problem, and the amount of that nectar source is an objective value called fitness. Employed bees are allocated to different food sources so as to maximize the total contribution of nectar. The colony must optimize the overall efficiency of collection. The distribution of bees is based on many factors such as the amount of nectar and the distance between the food source and the hive. The number of active or inactive employees represents the number of candidate solutions. In the first step, the algorithm generates an initial population of N solutions randomly distributed. Each vector $x_i (i = 1, 2, ..., N)$ represent a solution for optimization problem. After initialization, the population of solutions is subjected to repeated cycles $C = 1, 2, ..., C_{max}$ these cycles represent process searches made by foragers active, inactive and scouts.

The active foragers looking in the neighbourhood of the previous source $x_i$new sources v_i more nectar, they then calculate their fitness. To produce a new food source from the former, we use the following differential expression against:

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \quad (4)$$

with $k \in \{1, 2, ..., N_F\}$, N_F is the number of active foragers, $j \in \{1, 2, ..., N_F\}$, N are randomly selected indices, $k \neq i$ is randomly determined and ϕ_{ij} is a random number belonging to the interval $[-1, 1]$ it controls the production of a food source in the neighbourhood of x_{ij}

After the discovery of each new source of food v_{ij}, a selection mechanism is adopted, i.e. this source of foods is evaluated by artificial bee, its performance is compared with x_{ij}. if the nectar from this source is equal to or better than the previous source, the latter is replaced by the new. In the contrary case the former is preserved. At the end of the research process, the active foragers share information on nectar food sources and their locations with other bees through dance. In recent assess these information from all active foragers and choose food sources based on the probability value $P_i$associated to this source, and calculated by the following formula:

$$P_i = \frac{fit_i}{\Sigma_{j=1}^{N} fit_j} \quad (5)$$

Where fit_i is the fitness of the solution, which is proportional to the amount of nectar food source at position i.

```
Data: N: Number of the solutions
Result: Best solution
Initialization ;
N random solutions x_i;
c ← 1 Evaluate solutions ;
while c < C_max do
    foreach Employee do
        Produce new solutions v_i by using the formula
        4 ;
        Evaluate new solutions ;
        Apply greedy selection ;
    end
    Calculate the probability P_i ;
    foreach Onlooker do
        Select solution depending on P_i;
        Produce new solutions v_i from the selected
        ones by using the formula 5 ;
        Evaluate new solutions ;
        Apply greedy selection ;
    end
    if abandoned solution for the scout then
        Replace with new solution by using equation 6
    end
    c ← c + 1;
end
```

Fig. 1. Pseudo code of the ABC algorithm

The food source whose nectar is abandoned by the Scouts replace it with a new source according the formula:

$$x_i = x_{min} + rand(0, 1)(x_{min} - x_{max}) \quad (6)$$

If during a predetermined cycle number called "limit" a position can't be improved, so this source of food is assumed to be abandoned. All these steps are summarized in the algorithm:

IV. Methodology

In this section, we describe the design of our proposed model ABC-SVM, for predicting bankruptcy. The approach of combining a modified artificial bee colony algorithm with SVM is introduced for modelling and optimizing the parameter settings of SVM.

A. *SVM parameters optimization*

In the proposed model, the SVM parameters are optimized by the evolutionary process of the modified ABC. SVM model then performs the bankruptcy prediction using these optimal values. The approach was illustrated in figure 2. The optimal values of SVMs parameters are searching by modified ABC. The values of the two parameters, C and σ^2, are directly coded with real-valued data. The proposed model was developed and implemented in C++ language.

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

B. *Modified ABC*

in ABC algorithm, position of nectar source is presented by the solution vector $x = (X_1, ..., X_N)^T$, and the quality of nectar source is presented by the objective function of this problem [33]. The onlooker selects the nectar with the probabilistic selection:

$$P_i = \frac{fit_i}{\Sigma_{j=1}^{N}} \tag{7}$$

Where fit_i is the fitness of the solution x and is formulated as follows:

$$fit_i = \frac{1}{1 + f(x_i)} \tag{8}$$

For the objective function f(x), several measurement have been developed to evaluate the predictive accuracy of models. They include the mean absolute error as well as: MAPE, RMSE; the hit ratio and assessment ratio. The hit ratio is used as an indicator of our model performance by comparing it results with discriminant analysis and logistic regression.
The probabilistic selection depends on the fitness values of the solutions in the population. A fitness-based selection scheme might be a biased roulette wheel, ranking based, stochastic universal sampling, stochastic remainder selection, tournament selection or another selection scheme [34]. the biased roulette well, adopted by ABC, promotes replication of the best solutions and the elimination of the worst. Accordingly, if the size of the population of candidate solutions N is insufficient, it is likely that a solution with high adaptation is selected more than once. This solution becomes then dominant and causes premature convergence of the algorithm. It is also possible that a good solution is never selected on the N draws. In this case, it is called genetic drift. another phenomenon that may occur if N is small corresponds to the case where the n selected solutions have very similar degrees of adaptation. In this case of the population of solutions loses genetic diversity and the algorithm behaves more like a random process. The selection method proposed in this article is based on the method fitness sharing. This method consists in modulating the adaptation of each solution according to the density of the population in its neighbourhood. The essential purpose of this technique is to maintain a diversity within the population. Characteristically, the shared function $f_{sh(i)}$ of a solution x_i is:

$$f_{sh}(i) = \frac{fit(i)}{\Sigma_{j=1}^{n} sim(i,j)} \tag{9}$$

Where sim (i, j) is a measure of similarity between the two solutions i and j, which is generally expressed in the form:

$$sh(d(i,j)) = \begin{cases} 1 - \left(\frac{d(i,j)}{\sigma_s}\right)^{\alpha} & \text{if d(i,j)} < \sigma_s \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Where σ_s denotes a threshold of dissimilarity and α is a constant which regulates the shape of the sharing function.
So, the onlooker selects the nectar with the probabilistic selection becomes:

$$P_i = \frac{f_{sh}(i)}{\Sigma_{j=1}^{N}} \tag{11}$$

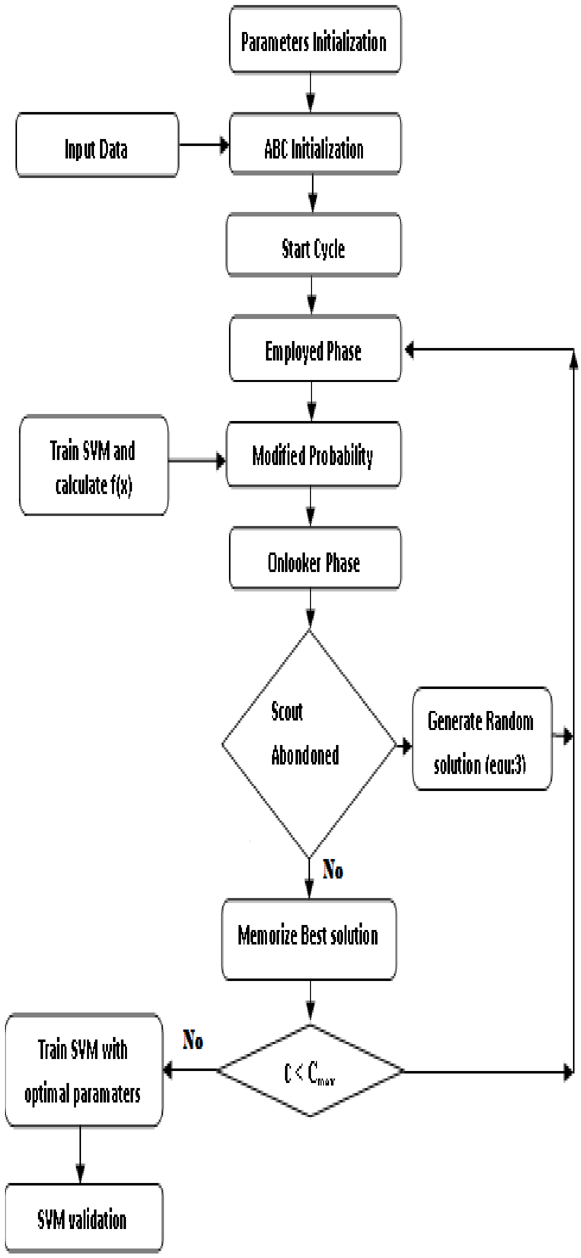

Fig. 2. The proposed approach ABC-SVM

V. Experimental results

Table V shows the data of the four data sets used in this study namely the Australian data[1], German data[2] and Japanese data[3] . In particular, the first three are the information of the credit card application to predict the UCSD[4] data-set's default probability . In practice, financial institutions assess credit risk and loan portfolios and analyze the probability of insolvency or bankruptcy before granting a loan. Note that very few related studies considering applying different data sets in their experiments. In addition, many related work makes conclusions based on sets of private data, which are difficult to assess. Therefore, our model, Discriminant Analysis and logistic regression are examined and compared. Experiments were performed to examine the

[1] http://www.liaad.up.pt/old/statlog/datasets/australian/australian.doc.html
[2] http://www.liaad.up.pt/ML/statlog/datasets/german/german.doc.html
[3] http://www.ics.uci.edu/mlearn/MLRepository.html
[4] http://mill.ucsd.edu/index.php?page=Datasetssubpage=Classification

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
INFORMATION OF THE DATASETS

	Australian	German	Japanese	USCD
No. of variables	14	20	15	40
Non bankruptcy	307	700	296	53 304
Bankruptcy	383	300	357	53 473
Total no. of samples	690	1 000	653	106 777

TABLE II
PREDICTION ACCURACY RATE IN PERCENTAGE

	DA	LR	SVM	ABC-SVM
Australian	83.75	86.10	87.00	88.15
German	84	83.15	82.84	85.16
UCSD competition	82.27	82.65	82.74	85.34
Average	83.34	83.96	84.19	86.21

TABLE III
TYPE I ERRORS

	DA	LR	SVM	ABC-SVM
Australian	0.06	0.04	0.00	0.00
German	0.13	0.11	0.01	0.01
UCSD competition	0.21	0.18	0.14	0.08
Average	0.13	0.11	0.05	0.03

TABLE IV
TYPE II ERRORS

	DA	LR	SVM	ABC-SVM
hline Australian	0.16	0.14	0.09	0.07
German	0.18	0.22	0.14	0.11
UCSD competition	0.28	0.27	0.20	0.13
Average	0.206	0.21	0.14	0.10

accuracy of the predictions of bankruptcy, Type I and Type II errors. Type I error was defined as the probability that a firm predicted not to fail will in fact fail, while the Type II error was defined as the probability that a firm predicted to bankrupt or not bankrupt.

Table V shows the prediction accuracy rate of each studied models. Our proposed model outperforms the two others. Table V shows the type I and type II errors of the three models. It is interesting to note that our model prediction accuracy rate is the lowest, but it provides the lowest type II error in the Australian and Bankruptcy datasets. Conversely, the others techniques present the highest average type II error.

VI. CONCLUSION

Support vector machine techniques is a widely used bankruptcy prediction model, has shown their applicability in the kind of area research. As the performance of SVM will be weakened if its parameters C and σ are not properly chosen, it is an indispensable task to optimize those parameters for a good performance. In this paper, we introduce a hybrid framework to improve the performance of SVM. Our model is based on Artificial bee colony to determine the optimal value of C and σ, for attaining desired output with an acceptable level of accuracy. Based on related datasets, the experimental results show that the hybrid technique outperforms Discriminant Analysis and logistic regression in terms of prediction accuracy and type I/II errors. For future work, several issues can be further studied. The first one, consists to use a genetic operator like mutation to diversify the ABC solutions. Second, we project to improve support vector machine by using others machine learning and compared them with our proposed model. Finally, it is interesting to use our model in credit scoring or index markets.

REFERENCES

[1] P. R. Kumar and V. Ravi, “Bankruptcy prediction in banks and firms via statistical and intelligent techniques–a review,” *European Journal of Operational Research*, vol. 180, no. 1, pp. 1–28, 2007.

[2] W. H. Beaver, “Financial ratios as predictors of failure,” *Journal of accounting research*, pp. 71–111, 1966.

[3] E. I. Altman, “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy,” *The journal of finance*, vol. 23, no. 4, pp. 589–609, 1968.

[4] J. A. Ohlson, “Financial ratios and the probabilistic prediction of bankruptcy,” *Journal of accounting research*, pp. 109–131, 1980.

[5] M. E. Zmijewski, “Methodological issues related to the estimation of financial distress prediction models,” *Journal of Accounting research*, pp. 59–82, 1984.

[6] M. L. Marais, J. M. Patell, and M. A. Wolfson, “The experimental design of classification models: An application of recursive partitioning and bootstrapping to commercial bank loan classifications,” *Journal of accounting Research*, pp. 87–114, 1984.

[7] M.-Y. Chen, “Comparing traditional statistics, decision tree classification and support vector machine techniques for financial bankruptcy prediction,” *Intelligent Automation & Soft Computing*, vol. 18, no. 1, pp. 65–73, 2012.

[8] Y.-C. Chen, “A study on the quality of credit granting in leasing: Fuzzy set theory approach,” *Soft Computing*, vol. 5, no. 3, pp. 229–236, 2001.

[9] S. Greco, B. Matarazzo, and R. Slowinski, “A new rough set approach to evaluation of bankruptcy risk,” in *Operational tools in the management of financial risks*. Springer, 1998, pp. 121–136.

[10] R. Slowinski and C. Zopounidis, “Application of the rough set approach to evaluation of bankruptcy risk,” *Intelligent Systems in Accounting, Finance and Management*, vol. 4, no. 1, pp. 27–41, 1995.

[11] T. E. McKee, “Rough sets bankruptcy prediction models versus auditor signalling rates,” *Journal of Forecasting*, vol. 22, no. 8, pp. 569–586, 2003.

[12] S. M. Bryant, “A case-based reasoning approach to bankruptcy prediction modeling,” *Intelligent Systems in Accounting, Finance and Management*, vol. 6, no. 3, pp. 195–214, 1997.

[13] H. Ahn and K.-j. Kim, “Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach,” *Applied Soft Computing*, vol. 9, no. 2, pp. 599–607, 2009.

[14] H. Li, H. Adeli, J. Sun, and J.-G. Han, “Hybridizing principles of topsis with case-based reasoning for business failure prediction,” *Computers & Operations Research*, vol. 38, no. 2, pp. 409–419, 2011.

[15] K.-S. Shin and Y.-J. Lee, “A genetic algorithm application in bankruptcy prediction modeling,” *Expert Systems with Applications*, vol. 23, no. 3, pp. 321–328, 2002.

[16] H. Etemadi, A. A. A. Rostamy, and H. F. Dehkordi, “A genetic programming model for bankruptcy prediction: Empirical evidence from iran,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 3199–3207, 2009.

[17] J. H. Min and Y.-C. Lee, “Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters,” *Expert systems with applications*, vol. 28, no. 4, pp. 603–614, 2005.

[18] K.-S. Shin, T. S. Lee, and H.-j. Kim, “An application of support vector machines in bankruptcy prediction model,” *Expert Systems with Applications*, vol. 28, no. 1, pp. 127–135, 2005.

[19] G. Zhang, M. Y. Hu, B. E. Patuwo, and D. C. Indro, “Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis,” *European journal of operational research*, vol. 116, no. 1, pp. 16–32, 1999.

[20] C. Charalambous, A. Charitou, and F. Kaourou, “Comparative analysis of artificial neural network models: Application in bankruptcy prediction,” *Annals of operations research*, vol. 99, no. 1-4, pp. 403–425, 2000.

[21] C.-F. Tsai and J.-W. Wu, “Using neural network ensembles for bankruptcy prediction and credit scoring,” *Expert Systems with Applications*, vol. 34, no. 4, pp. 2639–2649, 2008.

[22] G. P. Zhang, *Neural networks in business forecasting*. IGI Global, 2003.

[23] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. New York, NY, USA: ACM, 1992, pp. 144–152.

[24] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[25] L.-J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *Neural Networks, IEEE Transactions on*, vol. 14, no. 6, pp. 1506–1518, 2003.

[26] F. E. Tay and L. Cao, "Application of support vector machines in financial time series forecasting," *Omega*, vol. 29, no. 4, pp. 309–317, 2001.

[27] K.-j. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1, pp. 307–319, 2003.

[28] Y. Ding, X. Song, and Y. Zen, "Forecasting financial condition of chinese listed companies based on support vector machine," *Expert Systems with Applications*, vol. 34, no. 4, pp. 3081–3089, 2008.

[29] B. Sareni and L. Krahenbuhl, "Fitness sharing and niching methods revisited," *Evolutionary Computation, IEEE Transactions on*, vol. 2, no. 3, pp. 97–106, 1998.

[30] I. Steinwart and A. Christmann, *Support vector machines*. Springer Science & Business Media, 2008.

[31] A. J. Smola and B. Schölkopf, *Learning with kernels*. Citeseer, 1998.

[32] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (abc) algorithm," *Applied soft computing*, vol. 8, no. 1, pp. 687–697, 2008.

[33] B. Akay and D. Karaboga, "A modified artificial bee colony algorithm for real-parameter optimization," *Information Sciences*, vol. 192, pp. 120–142, 2012.

[34] K. Jebari, A. Elmoujahid, A. Dik, A. Bouroumi, and A. Ettouhami, "Unsupervised fuzzy tournament selection," *Applied Mathematical Sciences*, vol. 28, no. 1, pp. 2863–2881, 2011.

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)