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 $\sigma$ 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 represents 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 Boser, 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 $\sigma$ 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, $\sigma$, 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 $\sigma$) 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 $\sigma$.

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 $h(x) > 0$ 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:

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

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:

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

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:
\( \sigma \) denotes the variance of the Gaussian Kernel. According to the Lagrangian the optimal hyperplane is reformulated into the following problem:

\[
\max \{ L(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \} \tag{3}
\]

s.t. \( \sum_{i=1}^{l} \alpha_i y_i = 0 \leq \alpha_i \leq C, i=1, 2, ..., l \)

III. ARTIFICIAL BEE COLONY

Bees are social insects that live in highly organized colonies [32]. Three groups 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 \( C = 1, 2, ..., C_{\text{max}} \)

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:

\[
v_{ij} = x_{ij} + \phi_{ij} (x_{ij} - x_{kj}) \tag{4}
\]

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 \( \phi_{ij} \) 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:

\[
P_i = \frac{fit_i}{\sum_{j=1}^{N} fit_j} \tag{5}
\]

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

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

\[
x_i = x_{\text{min}} + \text{rand}(0, 1)(x_{\text{min}} - x_{\text{max}}) \tag{6}
\]

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 \( \sigma^2 \), are directly coded with real-valued data. The proposed model was developed and implemented in C++ language.
B. Modified ABC

In ABC algorithm, position of nectar source is presented by the solution vector \( x = (X_1, \ldots, X_N)^T \), 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:

\[
P_i = \frac{f_i t_i}{\sum_{j=1}^{N} f_j t_j}
\]

(7)

Where \( f_i t_i \) is the fitness of the solution \( x \) and is formulated as follows:

\[
f_i t_i = \frac{1}{1 + f(x_i)}
\]

(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:

\[
f_{sh}(i) = \frac{f_i t_i}{\sum_{j=1}^{N} \text{sim}(i,j)}
\]

(9)

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

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

(10)

Where \( \sigma_s \) denotes a threshold of dissimilarity and \( \alpha \) is a constant which regulates the shape of the sharing function. So, the onlooker selects the nectar with the probabilistic selection becomes:

\[
P_i = \frac{f_{sh}(i)}{\sum_{j=1}^{N} f_{sh}(j)}
\]

(11)

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 works make conclusions based on 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/ml/mlearn/MLRepository.html
\(^4\)http://mill.ucsd.edu/index.php?page=Datasetssubpage=Classification
acuracy 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 $\sigma$ 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 $\sigma$, 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 generic 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


