PSO-based Linear SVM Classifier Selection for Credit Risk Evaluation Modeling Process

Paulius Danenas and Gintautas Garsva
Department of Informatics, Kaunas Faculty, Vilnius University, Muitines St. 8, LT- 44280 Kaunas, Lithuania

Keywords: Support Vector Machines, Linear SVM, Particle Swarm Optimization, Credit Risk, Evaluation, Bankruptcy, Machine Learning

Abstract: A research on credit risk evaluation modelling using linear Support Vector Machines (SVM) classifiers is proposed in this paper. The classifier selection is automated using Particle Swarm Optimization technique. Sliding window approach is applied for testing classifier performance, together with other techniques such as discriminant analysis based scoring for evaluation of financial instances and correlation-based feature selection. The developed classifier is applied and tested on real bankruptcy data showing promising results.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

Credit risk evaluation is defined as one of the most important domains in financial sector as it shows the ability to regenerate income by lending money; yet, calculation of the possibility to get back the money invested is the most critical problem. Machine learning and artificial intelligence techniques are novel and state-of-the-art methods which help to develop tools for this problem by overcoming the drawbacks of statistical tools and deriving more robust and accurate solutions.

Discriminant analysis was one of the first techniques applied in credit evaluation (Altman, 1968). Support Vector Machines (SVM) classifiers gained a lot of attention as they showed abilities to get classification results comparable to Neural Networks but avoiding their main difficulties such as local minimas. Selection of hyperparameters is a sophisticated task thus various metaheuristic and evolutionary techniques have been adopted for solving this task including swarm intelligence techniques such as Ant colony Optimization (Zhou et al, 2007). Particle Swarm Optimization (abbr. PSO) has previously been applied for SVM optimization in credit risk domain – personal credit scoring (Xuchuan et. al, 2007), financial distress prediction (Chen et al., 2010; Wang, 2010), consumer credit scoring analysis (Yun et al., 2011). Linear SVM (LIBLINEAR) has also been tested to show competitive results to original C-SVC classifier (Danenas et. al, 2010; Danenas et al, 2011), which proved that they can be a good alternative in terms of both complexity and speed. According to these aspects, linear SVM and PSO are selected for model development. The research presented in this paper proposes a hybrid method based on linear Support Vector Machines classification and Particle Swarm Optimization. The proposed method is also tested in "sliding window" approach manner, which means that it can be useful to identify more general trends. Moreover, proposed approach might be useful while trying to improve the performance of these methods by identifying the most relevant financial attributes and developing a new classifier based on that particular technique.

2 USED METHODS

Support Vector Machines (SVM). SVM solves following quadratic minimization problem:

$$\min -\sum_{i=1}^{\ell}\alpha_i + \frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)$$

$$\textit{subject to} \sum_{i=1}^{\ell} y_i \alpha_i = 0,\ \forall i : 0 \le \alpha_i \le C$$

where the number of training examples is denoted by l, training vectors $X_i \in R, i = 1,..,l$ and a vector $y \in R^l$ such as $y_i \in [-1;1]$. α is avector of l values where each component α_i corresponds to a training example (x_i, y_i). If training vectors x_i are not linearly

Danenas P. and Garsva G..
PSO-based Linear SVM Classifier Selection for Credit Risk Evaluation Modeling Process.
DOI: 10.5220/0004006403380341
In *Proceedings of the 14th International Conference on Enterprise Information Systems* (ICEIS-2012), pages 338-341
ISBN: 978-989-8565-10-5

Copyright © 2012 SCITEPRESS (Science and Technology Publications, Lda.)

separable, they are mapped into a higher (maybe infinite) dimensional space by the kernel function $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$.

Fan et al. (Fan et al., 2008) proposed a family of linear SVM and logistic regression classifiers for large-scale SVM classification which do not use kernel functions for transformation into other dimensional space; although with less flexibility, it can perform effectively especially using large amounts of data. The formulations of the algorithms by are given in the paper of Fan et al.; four of them (L2-regularized L1-loss SVC, L2-regularized L2-loss SVC, L2-regularized logistic regression, L1-regularized L2-loss SVC) are used in the experiment. These classifiers are formulated as minimization problems, but they all share the concept of cost parameter C and bias usage. This proposes a possibility for heuristic selection of classifiers themselves.

Particle Swarm Optimization (PSO). The PSO algorithm, introduced by Kennedy and Eberhart, is based on behavior of flock of birds which search for food randomly in some area, knowing only the distance from the food. Thus all the particles have one fitness PSO is expressed in terms of particles (birds) and searched target described by fitness value; the location of each particle is determined by velocity describing its flying direction and distance. Two extreme values are tracked by each particle - the optimal solution found by the particle itself (*pbest*), and the optimal solution found by the whole swarm (*gbest*). Unmodified PSO algorithm is used in this research, thus its details are not presented in this paper, but can be found in other sources, such as (Kennedy et al., 2001).

2.1 PSO Approach for Linear SVM Optimization

A classification technique based on Particle Swarm Optimization and linear SVM combination, namely PSO-LinSVM is proposed in this paper. Each particle P = <p$_1$;p$_2$;p$_3$> is represented as follows:

p_1 – integer value, that represents the algorithm used for classification:

0 - L2-regularized logistic regression
1 – L2-regularized L2-loss SVC
2 – L1-regularized L2-loss SVC
3 - L2-regularized L1-loss SVC

p_2 – real value, cost parameter C

p_3 –real value, which represents bias term

The fitness function is defined as maximization of sum of TPR values:

$$f(fitness) = \sum_{N_C}^{1} TPR_i ,$$

where N_C is the number of classes. Most of the authors (Wang, Chin et al.) choose accuracy for fitness evaluation; however, in case of imbalanced learning, accuracy is not the best option, so sum of TP rate values is selected for this case, which allows selection of classifier that balances between identification of both “majority” and “minority” classes. These evaluations are obtained by performing k-fold cross-validation training; k is considered to be quite small (k = 5 is used for the experiment), considering the amount of data used in research. The optimal solution can be obtained only in case of perfect classification; as this happens very rarely, the main goal is to find best satisfactory solution.

2.2 Sliding Window Testing Approach

This research adopts techniques used earlier by Danenas et al. (Danenas et al., 2010; Danenas et al., 2011), extending it with PSO application for classifier optimization step. Thus the modified algorithm is defined as follows:

1. Evaluate each financial entry manually or by using expert techniques to compute bankruptcy classes (discriminant models used in banking are sued in this research).
2. Apply data preprocessing steps – elimination of unevaluated instances, data imputation and standardization.
3. Perform the following steps for each $m \in [1, n-k]$, where *n* is the total number of periods, *k* is the number of periods are used for forecasting:
   a. Feature selection;
   b. Classifier and parameter selection, using Particle Swarm Optimization;
   c. Train classifier using data from first *m* periods.
   d. Apply hold-out testing using data from period $p,\ p \in [m+1, m+k]; p \in \mathrm{N}$.

Note, that feature selection step is important for 2 reasons:

1. Quality and complexity - data dimensionality reduction;

2. Ratio importance - a new classifier based on other evaluator but using a set of statistically significant attributes obtained from the data is developed.

The output of each iteration in experimental stage is the trained classifier and the list of selected attributes for each period.

3 EXPERIMENT RESULTS

3.1 Data used in the Experiment

The dataset that was applied for the experiment consists of entries from 785 USA Transportation, Communications, Electric, Gas, And Sanitary Services companies with their 1999-2008 yearly financial records (balance and income statement) from financial EDGAR database.

Each instance has 51 financial attributes (indices used in financial analysis). “Risky” and “Non-Risky” classes were formed using Zmijewski’s scoring technique widely used in banking.

Table 1: Main characteristics of datasets used in experiments.

Year	Entries labeled as		Total entries	No of selected attributes	Bankrupt 1 years after	Bankrupt >1 year after
	Risky (R)	Not risky (NR)				
1999	376	166	542	11	-	-
2000	423	192	615	8	0	0
2001	383	226	609	13	2	1
2002	376	239	615	11	1	0
2003	417	220	637	9	0	0
2004	460	194	654	9	1	1
2005	478	173	651	8	1	4
2006	375	118	493	8	0	1
2007	367	112	479	11	0	6
2008	38	12	50	8	-	-
Total	**3693**	**1652**	**5345**		**5**	**13**

Note that ratios in original Zmijewski were not used in order to avoid linear dependence between variables. Main characteristics of the datasets formed for the experiment are presented in Table 1. It also shows financial ratios which were considered relevant by feature selection procedure; the number of such features is larger than the ones which are considered in original evaluator.

3.2 Computational Results

Correlation-based feature subset selection (Hall, 2001) algorithm with Tabu search for search in attribute subsets was applied for feature selection.

The search space for PSO was set to $C \in [0;50]$, $bias \in [0;1]$, as well as the number of run iterations was set to 10. PSO was configured to run with 20 particles and inertia rate of 0.8. Velocity for p_2 was set to 3, for p_3 was set to 0.2.

Table 2 presents the results obtained by PSO-LinSVM classifier: classifier parameters, obtained by PSO, classification accuracy together with True Positive and F-Measure rates for each class. It is clear that classification accuracy did not show stable increase while providing the classifier with more data each year. While performing testing procedure with first year data, accuracy decreased to 80% in 2004 although next year it returned to 83.8% was relatively stable, and later in fell to 82%; similar trends might be identified while analyzing testing results obtained with Year 2 and Year 3 data. It is important to note that instances marked as “risky” were identified better.

Table 2: Experimental classification results.

Training period			2000	2001	2002	2003	2004	2005	2006	2007
Linear classifier			L1-SVM (dual)	L2-SVM (dual)	L2-SVM (dual)	L2-RLR	L2-SVM (primal)	L2-SVM (dual)	L2-SVM (dual)	L2-SVM (primal)
C			15,3157	47,8343	24,7346	29,0490	22,3727	38,0860	6,5322	48,0734
Bias			1,000	0,196	0,749	0,797	0,873	0,838	0,436	0,508
Year 1	Accuracy		77,941	78,409	80,220	83,689	80,640	83,806	82,887	82,000
	TP	R	0,969	0,952	0,981	0,987	0,952	0,957	0,970	0,974
		NR	0,461	0,521	0,464	0,482	0,412	0,462	0,385	0,333
	F-Measure	R	0,846	0,843	0,867	0,895	0,878	0,900	0,896	0,892
		NR	0,609	0,653	0,618	0,637	0,535	0,579	0,520	0,471
Year 2	Accuracy		80,032	77,080	84,146	83,232	83,806	84,742	82,000	-
	TP	R	0,979	0,947	0,985	0,990	0,957	0,959	0,974	-
		NR	0,521	0,436	0,503	0,407	0,462	0,496	0,333	-
	F-Measure	R	0,857	0,844	0,897	0,896	0,900	0,905	0,892	-
		NR	0,670	0,568	0,653	0,567	0,579	0,611	0,471	-
Year 3	Accuracy		77,237	80,488	83,384	86,032	84,124	84,000	-	-
	TP	R	0,966	0,952	0,987	0,987	0,967	0,974	-	-
		NR	0,405	0,456	0,418	0,462	0,444	0,417	-	-
	F-Measure	R	0,848	0,873	0,897	0,915	0,902	0,902	-	-
		NR	0,551	0,582	0,576	0,615	0,575	0,556	-	-
Average testing accuracy			**78,403**	**78,660**	**82,583**	**84,318**	**82,857**	**84,183**	**82,444**	**82**

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

4 CONCLUSIONS AND FUTURE DEVELOPMENT

This paper presents an approach for credit risk evaluation using linear SVM classifiers, selected and optimized by Particle Swarm Optimization, combined with sliding window testing technique and feature selection using correlation analysis. Linear SVM classifiers perform well when applied to large scale problems; this is one of the main reasons why they were selected as classification technique. The developed classifiers were applied for real-world dataset, combined with widely applied Zmijewski technique as an evaluator and basis for output formation. Analysis of experimental results shows that the performance still needs to be improved to be more stable and reliable. Particle Swarm Optimization topology has not been investigated in this research, thus further steps will involve more detailed investigation into PSO performance. Imbalanced learning is another field where significant improvements might lead to increase in overall performance; this procedure is especially important if labelling is done automatically (as, in our case, using Zmijewski's model), as this might lead to highly imbalanced datasets. Notably, misidentification of bankrupt company might cost more to the creditor than the misdentification of "healthy" one, thus this problem is especially important if there are much less bankrupt companies or companies with high risk than companies which belong to another classes.

REFERENCES

Altman, E., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. In *The Journal of Finance*, Vol. 23 (4), pp.589–-609. American Finance Association; Blackwell Publishing

Chen, C.-Y., Chen, M.-Y., Hsieh C.-H., 2010. A Financial Distress Prediction System Construction based on Particles Swarm Optimization and Support Vector Machines. In *Proceedings of 2010 International Conference on E-business, Management and Economics* (IPEDR), Vol.3, IACSIT Press, Hong Kong pp. 165-169.

Danenas P., Garsva G., 2010. Credit risk evaluation using SVM-based classifier. In *Lecture notes in Business Information Processing*, Heidelberg: Springer-Verlag Vol. 57, Part 1, 2010, pp. 7-12, Springer.

Danenas P., Garsva G., Gudas S., 2011. Credit Risk Evaluation Model Development Using Support Vector Based Classifiers. In *Proceedings of the International Conference on Computational Science (ICCS 2011)*, Vol. 4, pp. 1699-1707, Procedia Computer Science.

Fan, R., Chang, K., Hsieh, C., Wang, X., Lin, C., 2008. LIBLINEAR: A library for large linear classification, In *The Journal of Machine Learning Research*, Vol. 9, pp.1871–4.

Hall, M.A. *Correlation-based Feature Subset Selection for Machine Learning*. Hamilton, New Zealand (1998)

JSwarm-PSO: Swarm optimization package, http://jswarm-pso.sourceforge.net/

Kennedy, J., Eberhart, R. C., Shi, Y. *Swarm intelligence*. Morgan Kaufmann Publishers, 2001.

Weka 3: Data Mining Software in Java, http://www.cs.waikato.ac.nz/ml/weka/

Wang X., 2010. Corporate Financial Warning Model Based on PSO and SVM. In *2010 2nd International Conference on Information Engineering and Computer Science (ICIECS)*, Wuhan, pp.1-5.

Xuchuan, J.M.Y., 2007. Construction and Application of PSO-SVM Model for Personal Credit Scoring. *Proc. of the 7th international conference on Computational Science, ICCS '07, Part IV*, pp. 158-161.

Yun, L., Cao Q.-Y.; Zhang H., 2011. Application of the PSO-SVM Model for Credit Scoring. Proceedings of *Seventh International Conference on Computational Intelligence and Security (CIS),* pp.47-51.

Zhou J., Zhang A., Bai T., 2008. Client Classification on Credit Risk Using Rough Set Theory and ACO-Based Support Vector Machine. In: *Proceedings of Wireless Communications, Networking and Mobile Computing (WiCOM '08)*, pp.1-4

Zmijewski, M,, 1984. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. In *Journal of Accounting Research*, Vol. 22, pp. 59–82.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS