International Journal of Soft Computing 13 (4): 129-133, 2018
ISSN: 1816-9503

Comparative Study of Kernel Function for Support Vector Machine on Financial Dataset

Noviyanti Santoso and Wahyu Wibowo
Department of Business Statictics, Institut Teknologi Sepuluh Nopember, Kampus ITS Sukolilo, 60111 Surabaya, Indonesia

Abstract: Due to the increasing number of business failures effect from economic crisis, it is challenging to develop a financial distress prediction model. The prediction model is the early warning system that has any advantages for companies, consumer, creditors investors and the economy of country in general. We develop SVM Model with different kernel function such as linear, polynomial and radial basis function. We purposed tuning method with 10-fold cross validation to find the best pair of parameters for each kernel function. The result shows that SVM model using radial basis kernel with optimal parameter C = 5 and γ = 1 is obtain appropriate accuracy, the AUC value is 0.72.

Key words: Financial distress, kernel function, support vector machine, polynomial, SVM Model, accuracy

INTRODUCTION

During the recent economic crisis the importance of the prediction of financial distress become obvious due to the increasing number of business failures. The prediction of financial distress is especially, important for creditors but there are several other stakeholders who can profit from an efficient early warning system that is able to predict the financial distress of companies. For example, consumer investor and regulators that use early warning systems to supervise the "financial health" of the companies. The cost of corporate bankruptcy is definitely high but not only for bankers. Business failure can cause losses in other sectors as well. When a company goes bankrupt, the market where the bankrupt company operated on becomes less competitive which is obviously costly for the whole economy. This is the main reason why researchers try to increase the predictive power of the early warning systems (Virag and Nyitrai, 2013).

The increasing availability of economic and financial data has led to the need for reliable tools for modelling and analysis. Data mining and applied statistical methods are the appropriate tools to extract knowledge from such data. Data mining can be defined as the process of selection, exploration and modelling of large databases in order to discover models and patterns that are unknown a priori. It differs from applied statistics mainly in terms of its scope whereas applied statistics concerns the application of statistical method to the data at hand. Data mining is a whole process of data extraction and analysis aimed at the production of decision rules for specified business goals. In other words, data mining is a business intelligence process (Giudici, 2003).

Support Vector Machine (SVM) is technique of prediction (Vapnik, 1995) used for classification or regression analysis. SVM is in a class with ANN in terms of functionality and problem conditions that can be solved. Both are in the supervised learning. Both researchers and practitioners have applied this technique a lot in solving real problems financial distress prediction, bankruptcy and profit forecasting. Proven in many implementations, SVM gives better results than ANN, especially in terms of solutions achieved. ANN finds local optimal as its solution while SVM finds a global optimal.

Most techniques of data mining or machine learning are developed by linearity assumptions. So, the resulting of algorithm is limited to linear cases. Therefore, if a classification case shows nonlinearity, algorithm such as percptron cannot handle it. Scholkopf and Smola (2002) introduce kernel method to solve that problem. It uses kernel method to transform data from input space into a high dimensional feature space in which it searches for a separating hyperplane. Therefore by mapping the data to feature space making classification object easier. Kernel functions that usually used in SVM literature are linear, polynomial, radial basis and sigmoid (Haykin, 1999).

Choice of different kernel functions will result in different SVMs and different performances (Williamson *et al*., 1999). On selection of the kernel

Corresponding Author: Noviyanti Santoso, Department of Business Statictics, Institut Teknologi Sepuluh Nopember, Kampus ITS Sukolilo, 60111 Surabaya, Indonesia

function, the different parameters have to be varied in order to obtain higher classification accuracy. Choosing the best kernel function for a specific problem is still an ongoing research issue. Tsang *et al*. (2005) proposed an idea for identifying a suitable kernel for the given data set. Maji *et al*. (2008) proposed a method for the correct estimation of intersection kernel SVMs which is almost simple and has appropriate classification accuracy but resulted in increase in runtime compared to RBF and polynomial kernels due to large number of support vectors for each classifier (Hasdoonk, 2005). Regardless of the ability of classification, problems still remain, particularly in choosing the efficient kernels of support vector machines for a specific application (Ektefa *et al*., 2011).

The aim of this research is to compare the performance among SVM kernel functions namely linear, polynomial and radial basis for financial distress prediction.

MATERIAL AND METHODS

Support vector machine: SVM was introduced by Vapnik (1995) for classification or regression problem which is very popular these days. SVM include in the supervised learning algorithms in which the learning machine is given with the set of input values with their associated labels. In this technique, we try to find an optimal classifier hyperplane which can separate two sets of data from two different classes. This technique attract researchers because of its convincing perfomance in terms of predict a new data class. We are going to begin with classification case that can be separated linearly. In this case, the classifier function, we are looking for is linear function. This function can be defined as:

$$f(x) = w \times \varphi(x) + b \quad (1)$$

which, x, w$\in\Re^n$ and b$\in\Re$. This classification problem can be formulated as well: we want to find a parameter set (w, b) so, $f(x_i) = \langle w, x\rangle + b = y_i$ for all i. In this technique, we try to find the best classifier function between unlimited function to separate two objects. The best hyperplane is located in the middle between two sets of objects from two classes. Looking for the best hyperplane is equivalent with maximizing margin or distance between two sets of objects from different classes. If $wx_1+b = +1$ is supporting-hyperplane of $+1(wx_1+b = +1)$ class and $(wx_2+b = -1)$ is supporting-hyperplane of $-1(wx_2+b = -1)$, the margin between two classes can be calculated by finding the distance between two supportinghyperplanes of both classes. Specifically, the margin is calculated in

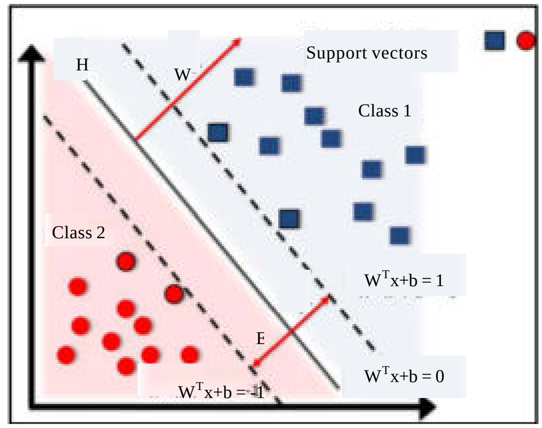

Fig. 1: SVM with maximum margin

this following mathematics, $(wx_1+b = +1-(wx_2+b = -1) = w(x_1-x_1 = 2 = >(w/||w||(x_1-x_2)) = 2/||w||$. Fig. 1 shows how SVM works to find a classifier function with maximum margin. SVM classification is performed with different kernel functions such as liniear function, polynomial function and radial basis function.

Kernel method: Most techniques of data mining or machine learning are developed by linearity assumptions. So, the resulting of algorithm is limited to linear cases. Therefore, if a classification case shows nonlinearity, algorithm such as perceptron cannot handle it. Kernel method (Scholkopf and Smola, 2002) is one to solve it. By kernel method a data x in input space is mapping to feature space F with higher dimension through map φ as well $\varphi: x \rightarrow \varphi(x)$. Therefore, data x as input space become φ(x) in feature space.

Mostly function φ(x) is unavailable or uncountable. But dot product from two vectors can be calculated both in the input space and in the feature space. In other word while φ(x) may unknown, dot product $\langle\varphi(x_1), \varphi(x_2)$ still can be calculated in the feature space. To be able to use kernel method, the constraints need to be expressed into dot product from vector data xi As a consequence, the constraints which describing the problem in classification have to be reformulated, so, it become a dot product. In this feature space do product <•> become < φ(x), φ(x)'>. A kernel function, K (x, x') is able to replace the dot product < (p(x), ip(x)'>. Then in feature space, we can make a linear classifier function which representing nonlinear function in input space. Figure 2 describing an example of feature mapping of two dimension spaces into two dimension feature space. In input space, the data can be separate linearly but we are able to separate it in feature space. Therefore, by adjusting the kernel parameters, the best kernel function can be determined.

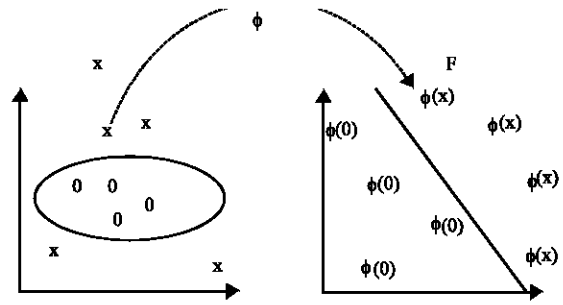

Fig. 2: Mapping kernel from input space to feature space

Liniear kernel: The linear kernel is the simplest kernel function. Kernel algorithms using a linear kernel are often equivalent to their non-kernel counterpart. It is given by:

$$K(x,x')=(x,x')+C \tag{2}$$

Polynomial kernel: A polynomial mapping is a popular method for non-linear modelling. The polynomial kernel is usually preferable as it avoids problems with the hessian becoming zero.

$$K(x,x')=(x,x')+1)^{d} \tag{3}$$

Radial Basis Kernel (RBF): RBF is a popular kernel function used commonly with SVM classification (Fig. 3). This functions most commonly with a Gaussian form as: (Haykin, 1999)

$$K(x,x')=\exp^{f}\left(-\frac{\|x-x'\|^{2}}{2\sigma^{2}}\right) \tag{4}$$

Dataset: In this comparative analysis, we used 693 dataset with 99 manufacture companies that listed on the Indonesia stock exchange in period 2008-2014. The input variables that used refer to the previous research, adopting 20 financial ratio indexes shown in Table 1.

Methodology: In order to obtain the model and predict the data, we partition the data into two parts which are training set and testing set. Afterward we applied greedy stepwise to select significant variables. There are 8 financial ratio indexes which used for modelling, i.e., Sales/Total asset, Earning/DEBT, Retained earning/ Total asset, gross profit ratio, operating profit ratio, net profit ratio, ROI and working Capital/Long term DEBT. In this study, we construct experiments with different kernel function such as linear, polynomial and radial basis. The linear kernel has no parameters, except C. We adjust the C as 0.1, 1, 5, 10, 100 and 1000. Additionally, the degree of polynomial kernel as 1-5 were determined. The radial basis has more parameters to tune, namely gamma (γ). There is no guidance that mentions the best value of gamma. Therefore, we choose 0.1, 1, 5, 10 and 100 to evaluate the radial kernel. From this combination, we will obtain the best parameters for each kernel function.

In the last, we compare performance of kernel function using area under ROC curve (AUC) that appropriate for unbalance data (Wang and Yao, 2013) as show as Fig. 4. Value of AUC is calculated using this following equation. The overall procedure of modelling SVM is illustrated in Fig. 3.

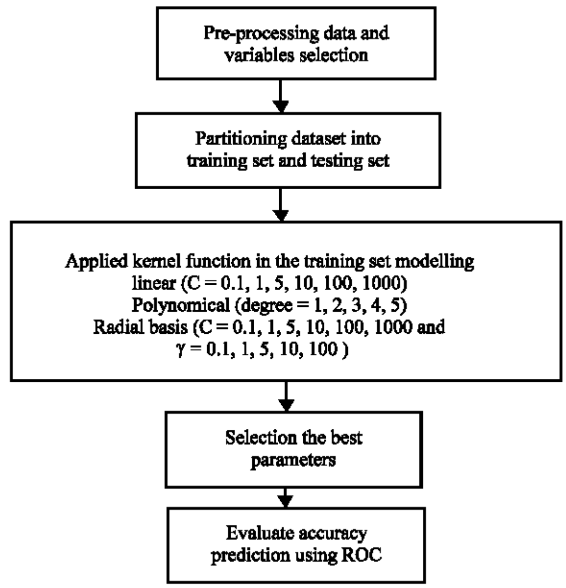

Fig. 3: Procedure of SVM modelling

Table 1: Financial ratio indexes as input variables

Variables	Definition
X1	EBIT/Total asset
X2	Sales/Total asset
X3	Sales/Fixed asset
X4	Earning/DEBT
X5	Current ratio
X6	Working capital/Total asset
X7	ROE
X8	Retained earning/Total asset
X9	Gross profit ratio
X10	Operating profit ratio
X11	Net profit ratio
X12	EBIT/Sales
X13	ROI
X14	Working capital/Long term DEBT
X15	DEBT to equity
X16	Book equity/Total capital
X17	Market value equity/Total capital
X18	Market value equity/Total liabilitas
X19	PER
X20	PBV

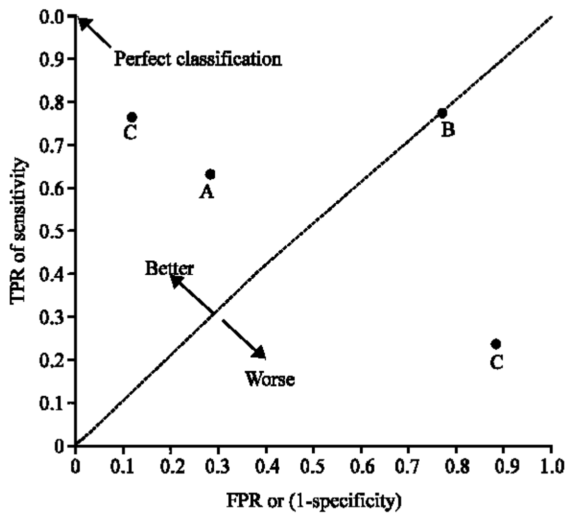

Fig. 4: Area under ROC curve (Random guess)

RESULT AND DISCUSSION

This study constructs experiments with different kernel function such as linear, polynomial and radial basis function. We apply a tuning method to find the best combination of parameters C and γ using 10-fold cross validation. According to Table 2, the optimal pair of parameters of each kernel is different. The best performance of testing data is radial kernel (0.72), followed by polynomial kernel with number of degree is 2 (0.63) and the last is linear kernel (0.51). Additionally, Table 2 shown that as degree increases, the AUC of training data increases unsteady while the AUC of testing data increases to the second degree (highest) then decreases constantly. It means that overfitting and underfitting problems occurred in the polynomial kernel performance. Therefore we need more simulation to find the best number of degree in the polynomial kernel.

Furthermore, based on Table 2, we can show that financial distress of manufacture companies in Indonesia can be better predicted using the radial basis kernel. Unfortunately, the overfitting problems also, occur in radial kernel function.

We simulate the effect of parameters C and γ to evaluate performance of radial basis kernel on SVM model. Figure 5 shows the accuracy prediction by AUC value where γ is fixed at 1 and C increases. The result obtain that AUC of testing data increases until C = 5 and then falling down as increase the C. The results indicate that selection parameter C and γ is important in SVM model.

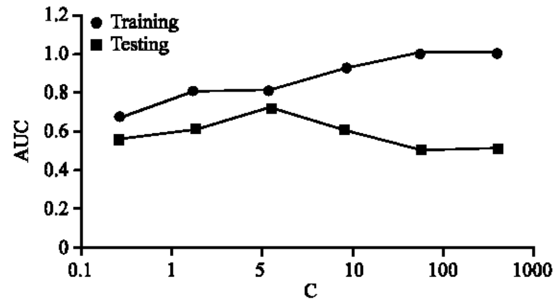

Fig. 5: The accuracy prediction AUC value

Table 2: Performance of SVM kernel on each best parameters (C, γ)

Kernel function	C	γ	Degree	AUC Training	AUC Testing
Linear	1	0.1	-	0.53	0.51
Polynomial	10	5.0	1	0.55	0.52
			2	0.65	0.63
			3	0.73	0.50
			4	0.72	0.50
			5	0.70	0.51
Radial	5	1.0	-	0.82	0.72

The classification of manufacture companies based on financial condition using radial basis is performance appropriately. There are nine distress companies that misclassified to non-distress companies.

CONCLUSION

This research applied SVM model using different kernel function to predict financial distress of manufacture companies in Indonesia. The result shows that radial basis function as kernel on SVM model is outperforms other kernel function. We assign a turning method using 10-fold cross validation to find the best pair of parameters. The highest AUC on testing data occurred when C = 5 and γ = 1 (0.72). However, the prediction accuracy is poor classification. Perhaps it caused by structure of dataset that unbalanced, selected input variables and parameters to conduct the SVM Model. We consider to further research to apply another method of feature selection then using genetic algorithm or particle swarm optimization to get optimal parameters.

REFERENCES

Ektefa, M., F. Sidi, H. Ibrahim, M.A. Jabar and S. Memar, 2011. A comparative study in classification techniques for unsupervised record linkage model. J. Comput. Sci., 7: 341-347.

Giudici, P., 2003. Applied Data Mining: Statistical Methods for Business and Industry. John Wiley & Sons, Hoboken, New Jersey, USA., ISBN:9780470846780, Pages: 376.

Haasdonk, B., 2005. Feature space interpretation of SVMs with indefinite kernels. IEEE. Trans. Pattern Anal. Mach. Intell., 27: 482-492.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. 2nd Edn., Prentice Hall, New Jersey, USA., pp: 696.

Maji, S., A.C. Berg and J. Malik, 2008. Classification using intersection kernel support vector machines is efficient. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK., USA., pp: 1-8.

Scholkopf, B. and A.J. Smola, 2002. Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond. MIT Press, Cambridge, Massachusetts, USA., Pages: 633.

Tsang, I.W., J.T. Kwok and P.M. Cheung, 2005. Core vector machines: Fast SVM training on very large data sets. J. Mach. Learn. Res., 6: 363-392.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. 2nd Edn., Springer, New York, USA., ISBN-13: 9780387945590, Pages: 188.

Virag, M. and T. Nyitrai, 2013. Application of support vector machines on the basis of the first Hungarian bankruptcy model. Soc. Economy, 35: 227-248.

Wang, S. and X. Yao, 2013. Using class imbalance learning for software defect prediction. IEEE. Trans. Reliab., 62: 434-443.

Williamson, R.C., A.J. Smola and B. Scholkopf, 1999. Entropy numbers, operators and support vector kernels. Proceedings of the 4th European Conference on Computational Learning Theory (EuroCOLT'99), March 29-31, 1999, Springer, Nordkirchen, Germany, ISBN:978-3-540-65701-9, pp: 285-299.