Prediction of Profitability of Industries using Weighted SVR

Divya Tomar

Student
Indian Institute of Information Technology
Allahabad, India

Ruchi Arya

Student
Indian Institute of Information Technology
Allahabad, India

Sonali Agarwal

Lecturer
Indian Institute of Information Technology
Allahabad, India

Abstract-In order to measure the profitability of an industry by predicting Pre-Tax Operating Margin by applying regression technique on Price/Sales Ratio and Net Margin of various industries. Prediction of Pre-Tax Operating Margin is done using Support vector Regression (SVR). We present a model in this paper in order to solve the problem of over-fitting which is due to noise and outliers in dataset. For this a weighted coefficient based approach is proposed that reduces the prediction error and provides the higher accuracy than simple support vector regression. At last, the comparison of SVR using different kernel functions with weight is done and results of experiments shows that LS-SVR with RBF kernel function using weighted coefficient have better accuracy.
Keywords-Pre-Tax Operating Margin, Price/Sales, Support Vector Regression, Weighted Coefficient

I. INTRODUCTION

An industry is very keen in knowing the profits obtained by other. But the information regarding this is not available easily. Therefore to compare the profitability among industries, some metrics are needed to be taken into consideration. Pre-Tax Operating Margin is one of those terms. It is an important term for any industry to measure the profitability of that industry. By using the accounting records of an industry, Pre-tax Operating Margin is used for measuring the operating efficiency. When we have to compare among different industries, it can be a good metric if it is predicted correctly on the basis of Price/Sales and Net Margin. For finding the value of a stock comparative to its past performance price/sales metric is used. For calculating this we have to divide the company's market cap by its revenue earned in the most recent year. Another measure of profitability is Net Margin. PSRs and Net Margin vary greatly when we compare between industries. They are mainly used for internal comparison. So they are not considered a very effective metrics to compare among different industries. Therefore an approach is proposed in this paper which uses the PSRs and Net margin to predict the Pre Tax Operating Margin which is used to compare among different industries. Since different industries have different number of firms, so Pre-Tax Operating Margin is also depend upon the number of firms an industry have. So we have introduced a weighted coefficient that is varying according to number of firms. We have assigned more weight to those industries that have more number of firms and less weight to those that have comparatively small number of firms. By assigning weights, we used to optimize the predicted value of Pre Tax Operating margin.

Data mining is a technique to find out useful and undiscovered information from large amount of data. For example find out the relationships among various variables in order to identify the dependency among variables on each other. The regression model refers to construct the mathematical model based on the known data, and it's mainly aims at forecasting for the unknown data. Support vector Machine is a modern machine learning approach developed on statistical learning theory. In 1990s, Vapnik V et.al put forward the concept of SVM. It works on structural risk minimization principle and has good generalization ability [1].SVM is excellent in handling non-linear, small size and high dimensional problems [2]. It uses non linear map from original data input space to feature space. For mapping the non-linearly separable data samples into feature space require kernel functions. There are number of kernel functions used for this purpose but polynomial and RBF kernel functions are used very frequently. Different kernel functions have different ability to handle non-linear problems [3]. SVR has good performance for the predictions of non-linear problems [4, 5].There are three most potently used methods of Support vector regression are: ε –SVR, υ –SVR and LS-SVR (Least square Support vector regression). Vapnik introduced a loss function that is known as ε -insensitive function. Precision of the regression estimation is controlled by ε that makes ε -SVR more robust and sparse. This sparseness is used for the estimation of dependence relation for vast data in the high feature space. Schlökopf [6] and others proposed υ –SVR that controls the number of support vectors. If number of support vector increases the model become more complex .It also adjusts ε automatically according to the model complexity and the slack variables. LS-SVR (1east square-SVR) [7] proposed by Suykens and others to reduce the computational complexity. In LS-SVR slack term is used as square norm in objective function. It transforms inequality restrictions into equality restrictions to find the solution of a Karush Kuhn Tucker (KKT) system.

The noise and outliers present in the dataset leads to over-fitting problem in SVR that affects the performance of regression. The better observation values do not mean the better values on the other points. So to solve over-fitting problem Scholars put forward the concept of the SVR using weighted coefficient. The content of paper is as follows; we discuss about the related research work in section 2. Section 3 deals with the different SVR techniques to predict the pre-tax operating margin. In Section 4, experimental details are elaborated and also explain the usefulness of the approach on the basis of result obtained. Section 5 concludes the paper.

II. RELATED WORK

The research work [9] of Chang-An Wu and Hong-Bing Liu an improved approach for support vector regression (SVR) is given which is better than and has smaller errors as compare to traditional SVR .In this work all training data are divided into class based on signs of error i.e. positive class and negative class and support vector is used for the construction of regression model which is used for the forecasting of unknown data .The traditional method for learning such as neural network uses ERM that minimize the sample total error during training process due to this the problem of over fitting come into existence and the generalization ability of model is limited.

Weimin Huang and Leping Shen [10] tried to overcome the problem of over fitting that come into existence with traditional regression caused by noises and outliers. They introduce a weight factor which is calculated for each data sample based on its distance to the center to the smallest enclosing hyper sphere in the feature space. Result of experiment indicates that the proposed weighted method reduces the error of regression as compare to SVR and has better accuracy also. In research paper [11] an extension of recently introduce Least Square Support Vector machine (LSSVM) is presented Experiment result shows that extended Least Square Support Vector Machine has better runtime, more number of support vectors and less Mean Square error.

In recent work [12] author proposed an effective weighted support vector regression method based on sample simplification. This paper First step is the sample simplification and the method can select useful samples effectively; and then weighted support vector regression is used to train the selected samples. Result of this approach is to compare the time used by three method ε _SVR, ε _SVR (outlier), Weighted_SVR (outlier), SWSVR (outlier).In another new approach [13] for classification model based on SVR-RBF algorithm is given. the result of research shows that SVR-RBF algorithm has a better accuracy.

In this research work [14] a new method for optimizing Support Vector Machines for regression problems is presented. This algorithm searches for efficient feasible directions. Within these selected directions, the best one is chosen, i.e. the one, coupled with an optimal step analytical evaluation that ensures a maximum increase of the objective function. Paper [15] introduced an improved algorithm for LSSVM which have better accuracy and faster running time.

In another recent paper work [16] hybrid support vector regression is used for time series prediction. Nonlinear regression and times series problems are successfully solved by this. In this article author used a hybrid SVM model. In research work [17] a novel approach for prediction based on genetic algorithm is proposed. Result of experiment is to compare the error between traditional neural network and proposed method and error analysis shows that proposed method represents better predictability.

III. SUPPORT VECTOR REGRESSION

Support Vector Machine is a new approach in the field of machine learning. It has many attractive feature and good empirical performance. Initially SVM is developed for classification purpose in order to separates the different class labels in multi dimensional space by constructing hyper-planes. But newly it have been extended to the regression problems.SVM solve both regression and classification tasks.SVM provides optimum, unique solution that is absent from the problem of local minima. We have presented in this research work a weighted Support vector regression with hybrid kernel function. It can be used for both linearly separable and non-linearly separable training data points. In non-linearly separable case it transforms the data samples into a higher dimensional space that is known as feature space, from low dimension input space.SVM then constructs an optimal hyper-plane in this feature space to separates the data points.

Suppose sample data is given as:

$$D= \{(x^1, y^1), \ldots, (x^n, y^n)\},\ x^i \epsilon R^n \text{ and } y^i \epsilon R.$$

Where, x^i=input vector.

Y^i=output data.

Estimates the unknown function using following equation:

$$y = f(x) = w.x + b \qquad (1)$$

SVR performs minimization of the parameters $||w||^2$ and -insensitive loss function to solve linear regression problems .Slack variables ξ_i, ξ_i^*, i=1,…n is introduced to perform minimization of -insensitive loss function .Slack variable is used to measure the deviation of training data points outside -insensitive zone. Perform minimization of following function:

$$minimize\ \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi i + \xi i^*) \qquad (2)$$

$$y_i - w.x_i - b \leq \varepsilon + \xi i$$

$$\text{Subject to } w.x_i + b - y_i \leq \varepsilon + \xi i^* \qquad (3)$$

$$\xi_i, \xi_i^* \geq 0, i = 1, \ldots n$$

The constant C >0 is a pre-specified value that are used to measures the trade-off between the model complexity which is associated to deviations larger than ε and smoothness of ‘f’. ξ, ξ^* are the usual slack variables that represent bounds on upper and lower boundary on the outputs of the system. We used positive Lagrange multipliers α, α^*, η, η^* to solve the problem of optimization (2) by first converting it into dual form. We can write the Lagrangian equation as follows:

$$L = \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi i + \xi i^*) - \sum_{i=1}^{n}(\eta_i \xi i + \eta_i^* \xi i^*) - \sum_{i=1}^{n}\alpha_i(\varepsilon + \xi i + y_i - (\langle w, x_i\rangle + b)) -$$

$$\sum_{i=1}^{n}\alpha_i^*\ (\varepsilon + \xi i^* - y_i + (\langle w, x_i\rangle + b)) \qquad (3)$$

Then minimizing (3) with respect to the primal variables (w, b, ξi, ξ^*_i), the derivative has to vanish. The dual variables η_i and η^*_i are eliminated, so we obtain the dual optimization problem:

$$\max_{\alpha,\alpha^*} W(\alpha^{(*)}) = -\frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) - \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N}y_i\ (\alpha_i - \alpha_i^*)$$

$$subject\ to \begin{cases}\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i\alpha_i^* \epsilon\ [0, C]\end{cases} \qquad (4)$$

There are only few training samples that have non-zero value of Lagrangian coefficient, those samples are called support vectors (SVs).Usually number of SVs is smaller than the training data points. For non-linear mapping of input data samples to feature space require kernel function K (x_i,x_j). We have performed experiment in this research work using RBF and polynomial kernel function that includes:

(a) Polynomial kernel function:

$$k(x_i, x_j) = (x_i.x_j + 1)^d$$

(b) Radial basis kernel function:

$$k(x_i , x_j)=\exp\left(-\gamma \left\|x_i - x_j\right\|^2\right) \quad for\ \gamma > 0.$$

The estimates of regression function at any given point of time x is then:

$$f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\, k(x_i, x_j) + b \quad (5)$$

By exploiting the Karush-Kuhn-Tucker (KKT) conditions we computed the value of b.

In our experiment we use least square Support vector regression because of its effectiveness and completeness and estimate the model performance of our model with cross validation .Another advantage of LSSVM is that model optimization and calculation can be performed relatively fast. The optimization problem is replaced by the following equations:

$$\min_{w,b,e} J(w, e) = \frac{1}{2} w^T w + \frac{1}{2}\gamma \sum_{i=1}^{n} e_i^2 \qquad (6)$$

Where $y_i = w^T.x + b + e_i, i = 1, \ldots N$

The resulting LSSVM model for function regression becomes:

$$f(x) = \textstyle\sum_{i=1}^{N} \alpha_i\ k(x_i , x_j)\text{+b} \qquad (7)$$

In the traditional LS-SVR, all the training datasets are treated uniformly. However, in many real world problems, different data points contribute differently in the learning of predictive function. We introduce a Weight factor to reduce the prediction error and that is depend on the number of firm .We assign more weight to large number of firm and less weight to when number of firm is small. Assignment of weight is based on the concept that when number of firm is large they contribute to more tax rate. So, we modify the 6 as:

$$\min_{w,b,e} J(w, e) = \frac{1}{2} w^T w + \frac{1}{2}\gamma \sum_{i=1}^{n} q_i.e_i^2 \quad (8)$$

It is clear that a smaller weight factor q_i reduces the effect of the parameters e , such that the corresponding data points is treated as less important and for which weight factor is high is treated as more important.

IV.DATASET DESCRIPTION

In this research work, for the regression purpose we choose a real time marketing dataset of India. This data set is available in excel format, which contains following details:-

TABLE I. DATA SET DETAILS

S.No	Attribute Name	Attribute Description
1.	Industry Group	List of different industry groups
2.	No. of Firms	Number of firms within industry group
3.	Price/Sales ratio	Price to sales is calculated by dividing a stock's current price by its revenue per share for the trailing 12 months: PSR=$\frac{Share\ Price}{Revenue\ Per\ Share}$
4.	Net Margin	Estimated by dividing the net income by the total revenues: $Net\ Margin = \frac{Net\ Income}{Sales}$
5.	Pre Tax Operating Margin	Estimated by dividing the Pre Tax Operating income by the total revenues: $Operating\ Margin = \frac{EBIT(1-t)}{Sales}$ To estimates the value for sector, we use the cumulated values of EBIT(1-t) and sales.

Total Instances: 88

V.EXPERIMENT AND RESULTS

We have done our experiment using mat-lab 7.0. We aimed to find a regression curve to approximate the true curve using the given samples. We have taken the dataset of India to predict the Pre-Tax operating Margin of the industries. Weight is given according to number of firms. In this experiment, comparison of different SVR techniques have been performed using polynomial, RBF and hybrid kernels functions and calculate root mean square error and mean absolute error with weight and without weight. Result of experiment shows that LS-SVR using hybrid kernel functions with weighted coefficient function has better accuracy than others.

A. Mean Absolute Error

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n} |e_i|$$

B. Root Mean Squared Error

$$\mathrm{RMSE}=\sqrt{\left(\sum_{i=1}^{n}(f_i - y_i)\right)/n}$$

f_i : Predicted value by model.

y_i: Value actually observed.

n=total number of data points.

TABLE II. COMPARISON OF ROOT MEAN SQUARE ERROR OF DIFFERENT SVR TECHNIQUES

SVR	Kernel functions			
	Polynomial		RBF	
	Without weight	With weight	Without weight	With weight
υ -SVR	0.1460	**0.0148**	0.0845	**0.0158**

ε-SVR	0.0862	**0.0130**	0.0772	**0.0101**
LS-SVR	0.0851	0.0114	0.0596	0.0075

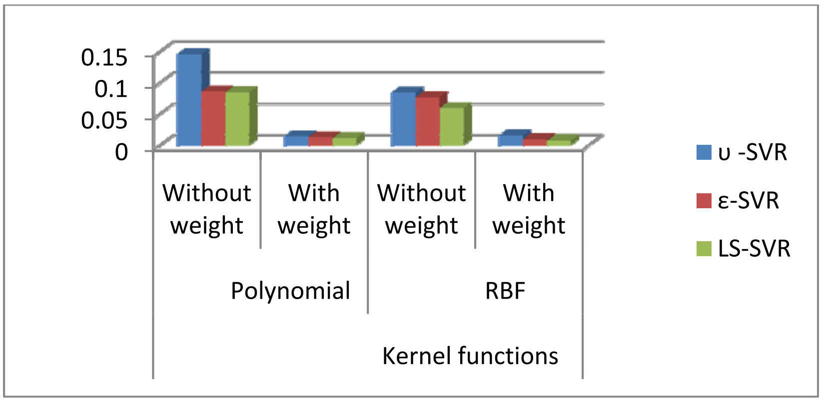

Figure 1. Comparison of root mean squared Error of different SVR techniques

TABLE III. COMPARISON OF MEAN ABSOLUETE ERROR OF DIFFERENT SVR TECHNIQUES

SVR	**Kernel functions**			
	Polynomial		**RBF**	
	Without weight	**With weight**	**Without weight**	**With weight**
υ - SVR	0.0817	**0.0086**	0.0731	**0.0087**
ε-SVR	0.0690	**0.0079**	0.0554	**0.0053**
LS-SVR	0.061	**0.0073**	0.0400	**0.0036**

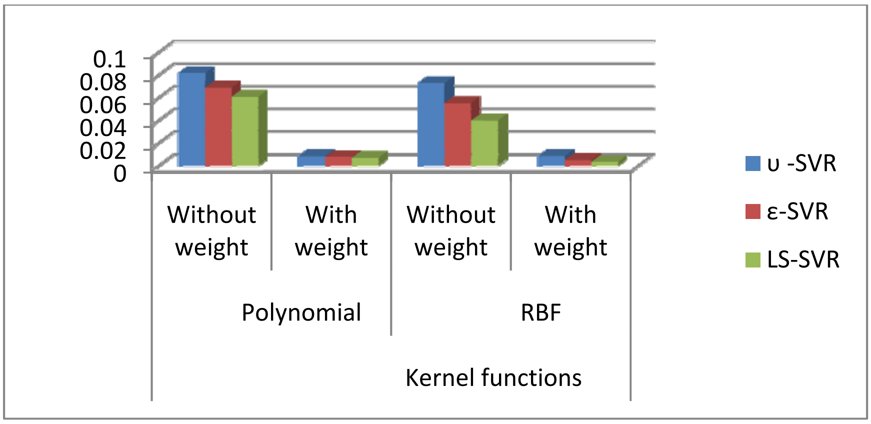

Figure 2. Comparison of mean absolute error of different SVR techniques

VI.CONCLUSION

As there is a great competition among industries regarding their profit obtained. So, special measures need to be taken to compare the profitability among industries. For this purpose we have chosen Pre-Tax Operating

Margin metric. This paper has proposed a novel approach that predicts Pre Tax Operating Margin by applying Support Vector Regression. Normally there are three techniques, namely LS-SVR, υ –SVR, ε –SVR, that are used in SVR. We performed experiment using three kernel functions namely polynomial, RBF kernel function and hybrid kernel functions. Also, weight factor was introduced that is based on number of firms in an industry, to improve the accuracy. Result of experiment shows that LS-SVR using RBF kernels function with weight has highest accuracy as compared to other techniques.

REFERENCES

[1] V. N. Vapnik, Statistical Learning Theory, Wiley, 1998.

[2] Deng Nai-Yang, Tian Ying –jie,Support Vector Method: A new Method in Data Mining, Beijing,Science Press , 2004.

[3] O Chapelle, V Vapnik, O Bousquent, S Mukherjee, Choosing multiple parameters for Support Vector Machines, Machine Learning, 46, pp.131-159, 2002.

[4] K R Muller and S Mika, “An Introduction to Kernel-Based Learning algorithms,” IEEE Trans on Neural Networks, 2001, 12(2), pp.199-222.

[5] A.J.Smola and B. Scholkopf, “A tutorial on support vector regression”, Statistics Computer, vol.14, no.3, aug. 2004, pp 199-222.

[6] B.Scholkopf, A.J. Smola, P L Bartlety , “New Support vector algorithms ,” Neural Computation , 2000 , 12 , pp. 1207-1245.

[7] K A J Suyokens, J Vandewalle, “Least squares support vector machine classifiers,” Neural Processing Letters, 1999, 9(3), pp. 293-300.

[8] N.Chistianini and J. Shawe-Taylor, An Introduction to Support Vector Machines, and other kernel-based learning methods, Cambridge University Press, 2000.

[9] Chang-An Wu and Hong-Bing Liu,”An Improved Support Vector Regression Based on Classification”, 2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07).

[10] Weimin Huang, Leping Shen,”Weighted Support Vector Regression Algorithm Based on Data Description”, 2008 ISECS International Colloquium on Computing, Communication, Control, and Management.

[11] Likun Hou, Qingxin Yang，Jinlong An, “An Improved LSSVM Regression Algorithm”, 2009 International Conference on Computational Intelligence and Natural Computing.

[12] Man Tang, Hongbin zhang “An Effective Method for Weighted Support Vector Regression Based on Sample Simplification”, Computing, Communication, Control, and Management, 2009. CCCM 2009.

[13] Yu Chen “Application of a Non-liner Classifier Model based on SVR-RBF Algorithm”, Power and Energy Engineering Conference (APPEEC), 2010 Asia-Pacific.

[14] Mireille Tohmé, Régis Lengellé “F-SVR: A New Learning Algorithm for Support Vector Regression”, Acoustics, Speech and Signal Processing, 2008. ICASSP 2008.

[15] Likun Hou, Qingxin Yang，Jinlong “An Improved LSSVM Regression Algorithm” 2009 International Conference on Computational Intelligence and Natural Computing.

[16] Qiong Li, Yuchen Fu*, Xiaoke Zhou, Yunlong Xu “A hybrid Support Vector Regression for Time Series Prediction” , 2010 Third International Conference on Knowledge Discovery and Data Mining.

[17] Hang Xie, Yuhe Liao, Hao Tang “Reliable prediction system based on support vector regression with genetic algorithms” 2009 Fifth International Conference on Natural Computation.

AUTHORS PROFILE

Divya is currently pursuing M.Tech in Indian Institute of Information Technology; Allahabad with specialization is in Human Computer Interaction. She received her bachelor degree in Computer Science & Engineering in 2009.Her current research interest is in support Vector Machine classification and regression. She has published 2 papers in the area of Support Vector Machine Classification and regression.

Sonali Agarwal is a lecturer in Indian Institute of Information Techology, Allahabad, India. Her research interests include Data Mining, Data Warehousing, E Governance, Knowledge Manage ment and Support Vector Machine. She has published 10 papers in the area of E Goverenance and data mining.

Ruchi Arya is currently pursuing M.Tech in Indian Institute of Information Technology, Allahabad with specialization in Software Engineering. Her current research interest is in Clustering Techniques.