Australian Journal of Basic and Applied Sciences, 7(7): 639-644, 2013
ISSN 1991-8178

Biased Gaussian Kernel Support Vector Machines –A Forecasting Method

[1]V. Anandhi and [2]Dr. R. Manicka Chezian

[1]A*ssistan*t Professor(Computer Science), Department of Forest Resource Management, Forest College and Research Institute, Mettupalayam 641 301,Tamil Nadu, India.
[2]Associate Professor, Department of Computer Science,NGM College, Pollachi-642001, Tamil Nadu, India.

Abstract: There is an annual increase in the annual consumption of paper per person, due to which a forecast on demand is necessary to improve India's socio economic development.Demand and supply data from Tamilnadu Newsprint and Papers Limited (TNPL) is used for forecasting. In this paper, the demand and supply pattern has been assessed for pulpwood and the same has been forecasted using modified Support Vector Machines (SVM).When parameters C and r (in the case of a Gaussian kernel) are chosen correctly, SVMs provide a good out-of-sample generalization.The Gaussian Kernel of the SVM is modified to improve the accuracy of the predictions of demand and supply of pulpwood.

Key words: Forecast, Pulpwood, Support Vector Machines (SVM), Mean Magnitude Relative Error (MMRE) and Median Magnitude Relative Error (MdMRE).

INTRODUCTION

India has of 600 paper mills of which between 30 to 40% use wood (Forest Survey of India, 2009) as raw material. Current raw material required is slightly more than 5 million metric cube as against a domestic supply of 2.6 million metric cube leading to 45% short fall in supply. There is an alarming increase in demand without a commensurate increase in supply. Demand and supply data from Tamilnadu Newsprint and Papers Limited (TNPL) is used for forecasting. Demand and supply assessed for pulpwood is used for forecasting by using modified Support Vector Machines (SVM).

Literature Review:

Support vector machines (SVM) are used as they reduce the time and expertise needed to construct/train price forecasting models. Also SVM has lower tune-able parameters with parameter values choice being less critical for good forecasting results. SVM can optimize its structure (tune its parameter settings) on input training data provided. SVM training includes solving quadratic optimization as it has only a unique solution and does not involve weights random initialization as training NN does. So an SVM with the similar parameter settings and trained on identical data provides identical results. This increases SVM forecast repeatability while reducing training runs number needed to locate optimum SVM parameter settings (Sansom, D.C., 2003).

Data non-regularity enables SVMs to be used for regression analysis, for example when data is not distributed regularly or has a known distribution (Zhang, L., 2001). Information to be transformed is evaluated prior to entering classification techniques score. SVM techniques' advantages are given below:

1. SVMs gain flexibility through kernel introduction in the choice of threshold form separating instances which do not need to be linear or have similar functional form for all data. This is because its function is non-parametric and operates locally.

2. No assumptions are necessary as kernel contains a non-linear transformation and its functional transformation ensures that data is linearly separable. Transformation is implicit on a robust theoretical basis without the need for human judgment.

3. When parameters C and r (in the case of a Gaussian kernel) are chosen correctly, SVMs provide a good out-of-sample generalization ensuring that selecting an appropriate generalization grade ensures that SVMs are robust, even with a biased training sample.

4. As optimality problem is convex, SVMs deliver unique solutions which are advantageous when compared to Neural Networks which have local minima linked multiple solutions and so might not be robust over samples.

In this paper, the demand and supply pattern has been assessed for pulpwood and the same has been forecasted using modified Support Vector Machines (SVM). The Gaussian Kernel of the SVM is modified to improve the accuracy of the predictions of demand and supply of pulpwood. The following sections discuss the methodology used for forecasting, followed by the experimental results and discussion.

Corresponding Author: V. Anandhi, Assistant Professor(Computer Science), Department of Forest Resource Management, Forest College and Research Institute, Mettupalayam 641 301,Tamil Nadu, India.
E-mail: anandhivenugopal@gmail.com

Methodology:

A novel SVM kernel is proposed in this work and is compared with M5, a standard linear regression algorithm.

M5 Algorithm:

Various tree-building algorithms like C4.5 determine the attributes which best classifies data remaining. Tree construction is iterative. Decision trees' advantage is the interpretation of immediate version to rules by decision-makers., Regression/model trees (Breiman, L., 1984) are used for numeric prediction in data mining. Both build decision trees where each leaf ensures local regression for a specific input space, the difference being that decision trees generate constant output values for input data subsets, model trees generate linear (first -order) models for subsets.

M5 algorithm constructed trees have leaves linked to multivariate linear models with tree nodes being chosen over an attribute that maximizes expected error reduction as output parameter function (Quinlan, J.R., 1992) of standard deviation. M5 algorithm builds decision trees which divide attribute space into orthohedric clusters having an axis paralleling border. The advantage of Model trees is its quick conversion into rules; each tree branch has the following condition: attribute $\leq$ value or attribute > value.

M5 algorithm (Quinlan, J.R., 1992) uses the "divide-and conquer" principle. Set N is linked with either a leaf or a test that splits N into subsets which correspond to the test outcomes is chosen with the same process being applied to subsets recursively. M5 algorithm splitting criterion is based on treating the class values standard deviation that reach a node as a measure of its error. It calculates anticipated error reduction due to each attribute being tested at that node. The following is the formula to compute SDR (standard deviation reduction) (Witten, I.H. and E. Frank, 2000):

$$SDR = sd\left(N\right) - \sum_{i} \left|\frac{N_i}{N}\right| sd\left(N_i\right)$$

where N -set of examples that reaches the node;
N_i- subset ofexamples that have the ith outcome of the potential set;
sd - standard deviation.

M5 after examining possible splits selects one that maximizes anticipated the expected error reduction. Splitting in M5 ceases when class values of all instances that reach a node vary slightly, or only few instances are left. Relentless division leads to over elaborate structures which require pruning, for instance substituting a sub tree with a leaf., A smoothing process compensates sharp discontinuities - in the final stage - occurring between adjacent linear models in the pruned tree's leaves specially for models made from smaller training examples. Smoothing updates adjacent linear equations so that predicted neighbouring input vectors outputs which correspond to different equations have a closer value.

Biased Gaussian Kernel Support Vector Machines(BGK-SVM):

SVMs (support vector machines) are supervised learning methods meant for classification and regression analysis (Vapnik, V., 2000). An SVM training algorithm builds a classification model which reveals whether a new sample falls into a category for which the system is trained. An SVM model is represented by drawing a hyperplane/set of hyperplanes dividing a category into two sets. SVM ensures achievement of a good separation through a hyperplane having the longest distance to the nearest class' training data points. The larger the margin, lower the classifier's generalization error (Vapnik, V., 1997). SVM algorithm is as follows:

For a set of training data $\left(x_i, y_i\right)$, $i = 1,..,n$, where $x_i \in \Re^d$ is a feature vector and $y_i \in \{+1,-1\}$ indicates the class value of x_i, the optimization problem is solved asfollows:

$$\min_{w,b,\xi} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$$

subject to

$$y_i\left(w^T\Phi\left(x_i\right)+b\right) \geq 1-\xi_i \text{ for i=1...n}$$

$$\xi_i \geq 0$$

where $\Phi : \Re \rightarrow \mathrm{H}$, H being high dimensional space $w \in \mathrm{H}$, and $\mathrm{b} \in \Re$. $C \geq 0$ is a parameter controlling margin errors minimization and margins maximization. Φ is chosen to ensure the existence of an efficient kernel function K., The above optimization problem, in practice, is solved through using Lagrange Multiplier method. Sequential Minimal Optimization (SMO) (Chang, C.C. and C.J. Lin, 2001) is an algorithm to solve SVM QP problem. SMO's advantage its ability to solve Lagrange multipliers without recourse to numerical QP optimization. The Lagrangian form is as follows:

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K\left(x_i, x_j\right) - \sum_{i=1}^{n}\alpha_i$$

subject to

$0 \le \alpha_i \le C$ for i=1...n

$$\sum_{i=1}^{n}\alpha_i y_i = 0$$

w is computed as followson solving the optimization problem:

$$w = \sum_{i=1}^{n}\alpha_i y_i \Phi\left(x_i\right)$$

x_i is a support vector if $\alpha_i \ne 0$. New instance x is computed by the following function:

$$f\left(x\right) = \sum_{i=1}^{n_S}\alpha_i y_i K\left(s_i, x\right) + b$$

Where s_i are support vectors and n_S is the number of vectors.

SVMs use the inner product as a metric to measure similarity/distance between patterns (Platt, J.C., 1999). The dependent relation between pattern's attribute is mapped as $\left\langle \Phi\left(x\right), \Phi\left(y\right) \right\rangle$ for pattern x and y or is represented as a kernel function as:

$$k\left(x, y\right) = \left\langle \Phi\left(x\right), \Phi\left(y\right) \right\rangle$$

Gaussian RBF kernel is very commonly used and is formulated as

$$k\left(x, y\right) = \exp\left(-\frac{\left\|x - y\right\|^2}{2\sigma^2}\right)$$

To avoid SVM kernel's over and under fitting, its width should adapt to feature space distribution. This work suggests a method to find the ideal width. Width is narrowed and weight assigned is less than 1 in dense areas, whereas in sparse areas width increases with the assigned weight being more than 1. The relationships are as follows:

a. The relation between σ and λ

$$k\left(x, y\right) = \exp\left(-\lambda\left\|x - y\right\|^2\right)$$

b. Relation between similarity and distance

$$d^2 = \left\|\Phi\left(x\right) - \Phi\left(y\right)\right\|^2 = 2 - 2k\left(x, y\right)$$

c. Relation between dense and.sparse in feature space

When Pattern x is in the dense area, the x's adjoining members are computed using weighed nearest neighbor distance formula:

$$sim\_Wnn\left(x\right) = \frac{1}{k}\sum_{i} k\left(x, x_i\right), \quad \mathrm{x}_i \in k - Wnn$$

$sim\_Wnn\left(x\right)$ is the index of density of x's neighborhood.

Regression analysis with MMRE formula (Mean Magnitude Relative Error) and MdMRE (Median Magnitude Relative Error):

It is necessary to measure software estimates accuracy for evaluation and validation. A common evaluation criteria in software engineering (Setiono, R., 2010) is used in this context:

Magnitude Relative Error (*MRE*) computes absolute error percentage between actual and predicted efforts for reference samples.

$$MRE_i = \frac{\left|actual_i - estimated_i\right|}{actual_i}$$

Mean Magnitude Relative Error (*MMRE*) calculates *MREs average* over all reference sampless. As *MMRE* is sensitive to an individual outlying prediction, a median of *MREs* is adopted for *n*samples (*MdMRE*) when there are many observations less sensitive to extreme *MRE values.* Despite the use of *MMRE* for estimation accuracy, there exists much discussion about its efficacy in estimation procedures. *MMRE* has been criticized as being unbalanced in many validation circumstances, resulting often in overestimation (Bhatnagar, R., 2010).

$$MMRE = \frac{1}{n}\sum_{i=1}^{n} MRE_i$$

$$MdMRE = \underset{i}{median}\left(MRE_i\right)$$

RESULTS AND DISCUSSION

The study is based on the data collected at the Tamil naduNewsprint and Papers Limited (TNPL) in Karur District,Tamil Nadu. Over ten years of data are collected for thedemand and the supply patterns of pulp wood. The datacollected is normalized.Predictions are performed by the proposed BGK-SVM based on the observed data. Table 1 shows the sample data of supply and demand of pulpwood in metric tonnes.

Table 1: Sample Data of Supply and Demand of Pulp Wood in MT (Metric Tonne).

Year	Supply (MT)	Demand (MT)
2003	125954	133719
2004	123026	147505
2005	162935	164804
2006	210152	166471
2007	222478	180577
2008	347139	383315

(Source- TNPL Management Plan.)

The Mean Magnitude Relative Error (MMRE) and Median Magnitude Relative Error (MdMRE) are evaluated through various techniques like M5, SVM with RBF kernel and the proposed BGK-SVM.Table 2 provides results of average MMRE and MdMRE for various techniques. Figures 1 and 2 reveal the same.

Table 2: Average MMRE and MdMRE for various techniques.

Technique Used	MMRE	MdMRE
M5	0.314705	23.75804
SVM-RBF	0.400824	43.92262
Proposed BGK-SVM	0.258144	21.34087

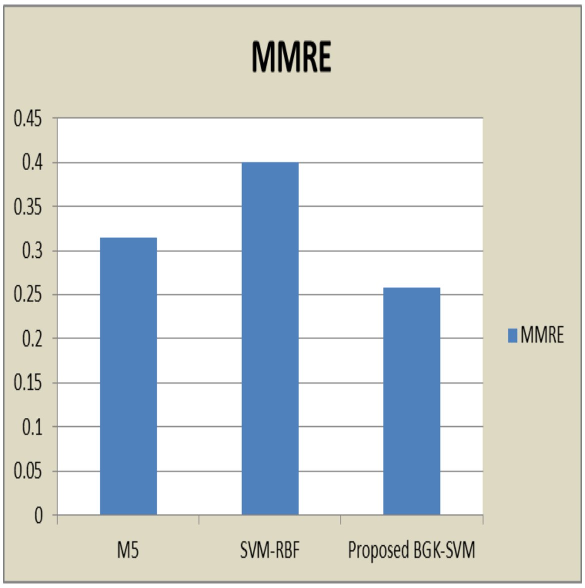

Fig. 1: MMRE for different techniques used.

From figure 1 it can be seen that the proposed technique decreases the MMRE by 64.44 % compared to SVM with RBF kernel and by 18% compared to M5 algorithm.

From figure 2 it is seen that the proposed Gaussian kernel decreases the MdMRE by 51.34% compared to SVM with RBF kernel. Simlarly the MdMRE decreases by 10.17% compared to M5 alogrithm.

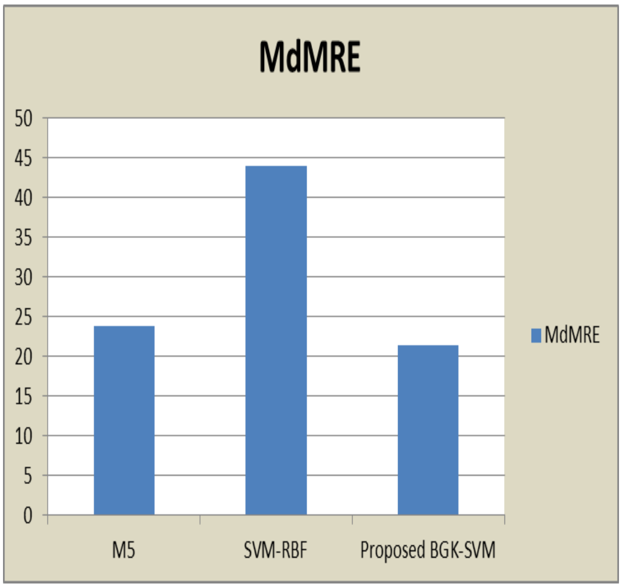

Fig. 2: MdMRE for different techniques used.

Conclusion:

This paper proposes the use of biased Gaussian Kernel Support Vector Machines to forecast demand and supply of pulpwood. It also uses Mean Magnitude Relative Error (MMRE) and Median Magnitude Relative Error (MdMRE) as evaluation criteria. This study also uses demand and supply data collected from TNPL. Simulation results demonstrate that the performance of the proposed method is satisfactory.

ACKNOWLEDGMENTS

Sincere thanks to Mr. R. Sreenivasan, Assistant General Manager, Tamil Nadu News print and Papers Limited (TNPL), karur, Dr. K. T. Parthiban, Professor and Head, Tree Breeding, Th. P Durairasu, Dean, FC&RI, Dr. M. Anjugam, Professor and Head, Forest College and Research Institute for their guidance and support.

REFERENCES

Bhatnagar, R., V. Bhattacharjee, M.K. Ghose, 2010. Software Development Effort Estimation–Neural Network Vs. Regression Modeling Approach. International Journal of Engineering Science and Technology, 2(7): 2950-2956.

Breiman, L., J. Friedman, R. Olshen and C.J. Stone, 1984. Classification and Regression Trees, Chapman and Hall, New York.

Chang, C.C. and C.J. Lin, 2001. *LIBSVM: a library for support vector machines*.

Forest Survey of India, 2009. Ministry of Environment and Forests, Govt. of India.

Platt, J.C., 1999. ”Fast training of support vector machines using sequential minimal optimization”. Advances in kernel methods: Support vector machines, B. Schokopf et al. (ed.), MIT Press.

Quinlan, J.R., 1992. “Learning with continuous classes.” Proc., Australian Joint Conf. on Artificial Intelligence, World Scientific, Singapore, 343-348.

Sansom, D.C., T. Downs, T.K. Saha, 2003. Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants. Journal of Electrical and Electronics Engineering, Australia, 22(3): 227-234.

Setiono, R., K. Dejaeger, W. Verbeke, D. Martens, B. Baesens, 2010. Software effort prediction using regression rule extraction from neural networks. In Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on (2: 45-52). IEEE.

Vapnik, V., 2000. *The nature of statistical learning theory*, Springer, NY.

Vapnik, V., S. Golowich and A. Smola, 1997. “Support vector method forfunction approximation, regression estimation, and signal processing,”*Neural Information Processing Systems*, vol. 9, MIT Press, Cambridge, MA.

Witten, I.H. and E. Frank, 2000. Data mining: Practical machine learning tools and techniques with Java implementations, Morgan Kaufmannp.

Zhang, L., F. Lin, B. Zhang, 2001. Support vector machine learning for image retrieval. In Image Processing, 2001. Proceedings. 2001 International Conference on (2: 721-724). IEEE.