Global Journal of Science Frontier Research
Environment & Earth Science
Volume 13 Issue 3 Version 1.0 Year 2013
Type : Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

Regression Analysis for Predicting Wood Pulp Demand by PSO Optimization

By V. Anandhi & Dr. R. Manicka Chezian

Tamil Nadu Agricultural University, India

Abstract - In today's world, consumption of paper and paperbased products is increasing in all the fields. Wood pulp which is extracted from the wood chips is the most commonly used raw material to manufacture the papers. Demand and supply of the wood pulp determines the social-economical development of a country. Many forecasting methods are used to predict the future demands of the wood pulp so that the supply chain management can be planned. In this paper, support vector regression analysis methods are used to predict the demands of wood pulp and Particle Swarm Optimization (PSO) algorithm is proposed to optimize the parameters of kernel functions. Regression models were created by using the data collected from TNPL. The parameters such as Mean Magnitude Relative Error (MMRE) and Median Magnitude Relative Error (MdMRE) are used for evaluating the results. Evaluated result shows that proposed SVM regression with PSO approach gave improved accuracy with significant decrease in MMRE and MdMRE.

Keywords : wood pulp, demand supply management, support vector regression analysis, particle swarm optimization, RBF.

GJSFR-H Classification : FOR Code: 860399

REGRESSION ANALYSISFOR PREDICTING WOOD PULP DEMAND BY PSO OPTIMIZATION

Strictly as per the compliance and regulations of :

Regression Analysis for Predicting Wood Pulp Demand by PSO Optimization

V. Anandhi [α] & Dr. R. Manicka Chezian [σ]

Abstract - In today's world, consumption of paper and paper-based products is increasing in all the fields. Wood pulp which is extracted from the wood chips is the most commonly used raw material to manufacture the papers. Demand and supply of the wood pulp determines the social-economical development of a country. Many forecasting methods are used to predict the future demands of the wood pulp so that the supply chain management can be planned. In this paper, support vector regression analysis methods are used to predict the demands of wood pulp and Particle Swarm Optimization (PSO) algorithm is proposed to optimize the parameters of kernel functions. Regression models were created by using the data collected from TNPL. The parameters such as Mean Magnitude Relative Error (MMRE) and Median Magnitude Relative Error (MdMRE) are used for evaluating the results. Evaluated result shows that proposed SVM regression with PSO approach gave improved accuracy with significant decrease in MMRE and MdMRE.

Keywords : wood pulp, demand supply management, support vector regression analysis, particle swarm optimization, RBF.

I. Introduction

Pulp is a processed material from wood fibers and it can be used for the production of paper and paperboard. Different types of woods such as soft woods and hard woods are available to make the wood pulp in the pulp mills. Usually grounded wood is processed in pulp mills to produce the wood pulp and this wood pulp is used to produce papers in the paper industries. Sometimes waste papers are recycled in the paper mills and combined with wood pulp in the paper industries to produce papers [1]. Some mechanical processes, chemical processes and combined processes are used to create the pulp from the wood fibers. In the chemical process, small pieces of wood are heated with some standard so lutions in the required temperature and pressure. This heating dissolves the glue which combines the fibers and leaves the fibers of the type of cellular and gives paper with high quality. Pulp is retrieved up to 60% of the load used by the pieces of wood taken for heating.

Pulping by mechanical process separates cellulose fibers either with wet grind stones or using refiners. Wet grind stones are used to press the logs and refiners use metal disks to pass the wood chips in the revolving disks. Large quantity of pulp is yielded in mechanical pulping and sometimes the quality of pulp is low because some wastes cannot be detached. So mechanical pulping is used to produce lower cost grades of paper (used for newsprint). To produce varying quality of paper different combinations of chemical and mechanical processes can be used. The usage of paper and its related products are increasing in every year in exponential way. From the year 1981 to 2000, the consumption of paper in india was increased about 0.83 to 0.88 percentage in every year. But from the year 2000 to 2010 consumption was increased by more than 2%. The increase in the conumption of is caused by the factors such as changing lifestyle, growth in economics, raised rate of literacy, growth of media and the need of paper with high quality [2]. To meet the exponential growing need of the paper pulp, implantation of the trees is the only solution. Many studies about forests and agriculture plans for future suggested the implantation. Therefore, during the year 2004 to 2007 around 91 % of plantations were done in the farmlands of tamilnadu [3]. The papers can be classified into three types. They are cultural paper, industry paper and low cost paper. Cultural papers are used for the purpose of writing, industrial papers are used for packing purposes and low quality paper is used for printing newspapers. Some of the issues addressed in national forest policies [4] are,

1) Competition for land is increased among foresting, residential and other activities of development.
2) Investment by the private bodies to maintain the forest is very low.
3) Most of the industry developed products depend on the forest materials.
4) Some forests are degraded by nature.

In India, totally around 600 paper mills are running and most of the industries use wood as a most significant raw material [5]. Domestic supply of wood per year is about 2.6 million metric cubes and it is around 45 % of the industry consumption. Regression analysis is one of the statistical analysis method used to find the relationship between one dependent variable and many independent variables. There are two types of regression. They are linear regression and polynomial regressions. Linear regression uses one independent

Author α : Assistant Professor(Computer Science), Department of Forest Resource Management, Tamil Nadu Agricultural University, Forest College and Research Institute, Mettupalayam 641 301, Tamil Nadu, India. E-mail : anandhivenugopal@gmail.com
Author σ : Associate Professor, Department of Computer Science, NGM College, Pollachi-642001, Tamil Nadu, India.

variable to predict the output variable. But multiple regressions use a series of independent variable to predict one dependent variable [6]. Classification and regression are two important statistics used for prediction. Classification maps an independent set of input values into an output value in a predetermined set of values. But prediction maps an independent set of values into a numeric value based on a predetermined relationship [7]. Support vector machine (SVM) is a classification algorithm used for classifying data into two predetermined classes. The aim of SVM is to maximize the margins between two classes so that the errors in the classification can be reduced. The data instance lies in the margins are called as support vectors. Accuracy achieved in classification inspires, to use this algorithm for regression analysis. Support vector regression (SVR) is based on the development of margin kernel methods. This SVR is used to find a relationship between input independent variables and output dependent variable. This relationship maps the values of variables in the input domain into the real output value [8]. Training sample was selected with possible input and output combinations [9]. Genetic Algorithms (GAs) are stochastic search algorithms used for problems that need optimization. GA takes initial population as a set of possible candidate solutions. Using the steps of selection, cross over and mutation of individual solutions in each iteration GA finds an optimal solution. Every solution is evaluated by the fitness function [10]. To improve the results of any regression model genetic algorithms can be used to tune the parameters used to represent the relationship between the input and output variables. But genetic algorithms do not use past search history and does not guarantee for optimized solutions. So swarm intelligence approaches inspired from the behaviour of some animals or insects can be used instead of GA algorithms for optimization. In this work, support vector regressions are used to forecast the demand supply of wood pulp and particle swarm optimization (PSO) is used for optimizing the parameters used for forming kernel function of SVM.

II. Literature Review

V. Anandhi, Dr. R. Manicka Chezian and Dr. K.T. Parthiban presented Forecast of Demand and Supply of Pulpwood using Artificial Neural Network [11]. To predict the future demand and supply of the wood pulp, a prediction model was developed by using artificial neural networks (ANN). The consumption of paper over last 10 years was collected from the tamilnadu newsprint and papers limited and used for training the ANN. Back propagation network was used for training the data. Every time validation error was checked and whenever the validation error was greater than a limit for 6 iterations, the training was terminated. This made the tuning of ANNs towards the forecasting problem of demand supply of wood pulps. This study created awareness in the need of wood pulp in future similar to other products of agriculture. V. Anandhi and R. Manicka Chezian presented Support Vector Regression to Forecast the Demand and Supply of Pulpwood [12]. To forecast the demand supply of wood pulp, SVM regression analysis was used with kernel functions. SVM was a machine learning method used to project the given data into a higher dimensional space. It maximized the margins between two different classes and minimized the regression error margin. Training data was collected from the demand and supply of paper mills over many years and support vector regression model is created. After training any data irrespective of training data could be given for testing. Training and testing using this method gave improved accuracy. Smruti Rekha Das et al presented an Improving RBF Kernel Function of Support Vector Machine using Particle Swarm Optimization [13]. SVM had been used as the most important and accurate tool for the classification, regression and any detection based algorithms. The main goal of SVM was to find the best margin used to separate the data and the outcome of interest was optional. SVM was used in linearly separable problems. To improve the accuracy of classification and regression using SVM optimization was introduced. In SVM-RBF kernel design Particle swarm optimization (PSO) was used to optimize the selection of parameters. After selection the number of features was reduced by principle Component analysis (PCA). Using the same data set algorithms using linear, polynomial and RBF kernel functions were tested and compared. Results showed that PSO optimization for RBF kernel gave improved accuracy.

Dimitri P. Solomatine and And Michael Baskara L. A. Siek presented Flexible and Optimal M5 Model Trees with Applications to Flow Predictions [14]. Numerical prediction was otherwise called as regression. M5-modelled decision tree was one of the classification methods and the tree was created by portioning the large size input space using the entropy measures and gave unique labels in the leaf nodes. The actual values of input parameters were applied from the root node down to the leaf node and the leaf node gave the classified label. For regression, leaf nodes were assigned the average of the output variable. If the input domain has complex regression relationships among them, the leaf nodes were designed to give the relationship functions rather than average value of the output variable. M5 tree was created progressively by dividing the input domain into many subsets. The division was stopped when the regression error was lesser then some acceptable error level from the training data. Reza Shah Hosseini, Saied Homayouni and Reza Safari presented a Modified algorithm based on Support Vector Machines for classification of hyper spectral

images in a similarity space [15]. SVM-based classifier is a more accurate classification method when the size of the sample space is very small. The most important aspect of SVM classifier was the kernel function. The main purpose of kernel function was to minimize the distance between instances inside the cluster and maximize the distance between instances of different clusters. The design of kernel functions was done at the time of training process and then tested with test data. Finding best kernel was the time consuming process.

III. Methodology

The SVM-based regression [16] is a supervised strategy used to tune many number of parameters internally, whose values influence the SVM-generated approximation function and should be priory set by the trainer. SVM uses many number of trial-and-error searches to get the lowest regression error. PSO algorithms use the particle swarm intelligence and have more advantages than GA based algorithms. In GA, each iteration deletes the worst solutions by saving only the good solutions but in PSO, all the particles utilize the good solutions obtained by other particles.

To evaluate the accuracy of the regression methods several measures based on the calculation of error is used. Error is calculated by the difference between the actual and predicted values. Absolute percentage of error is represented by Magnitude Relative Error (*MRE*). MRE is computed by the following formula.

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

Mean Magnitude Relative Error (*MMRE*) is calculated by taking the average of *MREs* over all the reference projects. For some input domains the *MMRE* is sensitive to outliers if the number of observations is too large [17]. To reduce this sensitivity Median MREs can be used for n number of projects. Formulas to calculate MMRE and MdMRE are given as follows.

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

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

IV. Results and Discussion

Sample data is collected from the company Tamilnadu News Print Ltd. Supply demand combinations of wood pulp measured in metric tones are calculated for the past 10 years. Sample data collected is shown in table 1.

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

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

Source: TNPL Management Plan

Figure 1 shows the actual and predicted values using algorithms such as M5, SVM kernel, SVM with RBF kernel, and SVM-kernal and PSO optimization method. Result shows that PSO optimization for parameters of kernel function gives minimum variations between the actual and predicted values.

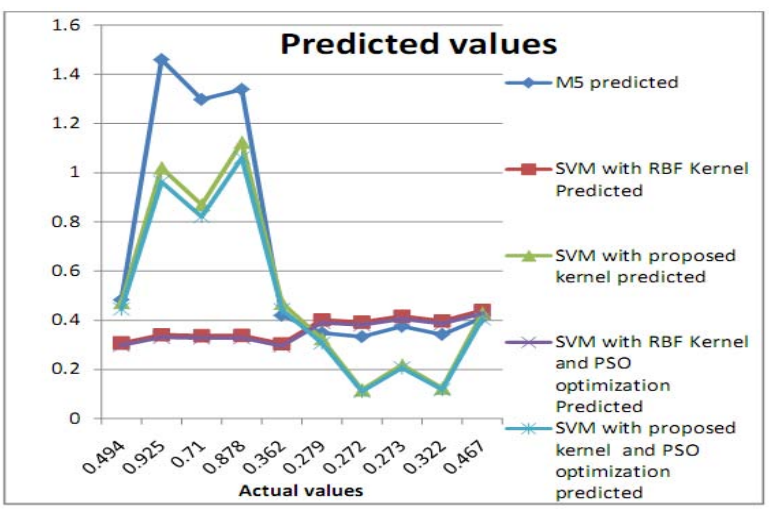

Figure 1 : Actual Vs Predicted Values

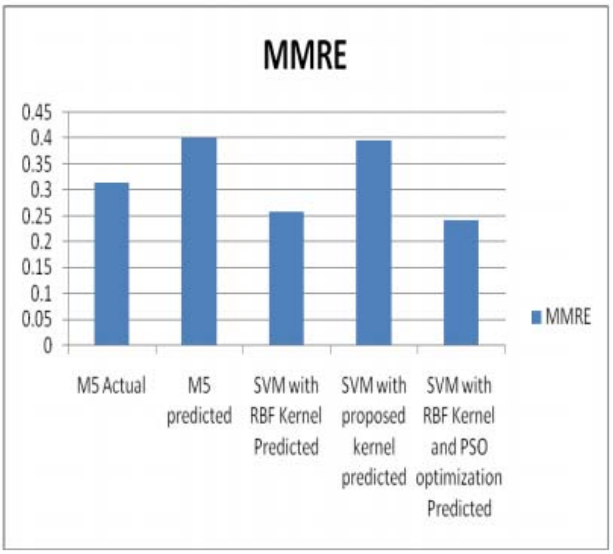

Figure 2 : MMRE

Figure 2 shows the Mean of Magnitude Relative error (MMRE) for the different regression methods using the experimental data. Graphical representation result shows that proposed PSO optimized algorithm with SVM kernel function for regression gives lowest MMRE.

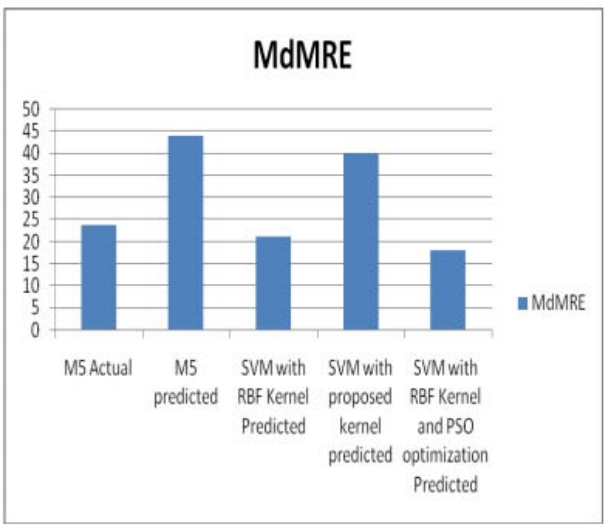

Figure 3 : MdMRE

Figure 3 shows the error measure using Mean of Magnitude Relative error (MMRE) for the different regression methods M5, SVM-kernal, SVM-RBF kernel with PSO optimization using the experimental data. Graphical representation result shows that proposed PSO optimized algorithm with SVM kernel function for regression gives lowest MdMRE.

V. Conclusion

Wood pulp demand and supply chain is an important factor in social and economical development of a country. As the demand is increasing in exponent way, forecasting the need of pulp is useful for the prediction of the market demand, planning alternative sources and planning implantation. In this paper, support vector regression analysis methods are used to predict the demands of wood pulp and Particle Swarm Optimization (PSO) algorithm is used to optimize the parameters of kernel functions to increase the accuracy of prediction. The parameters such as mean magnitude relative error (MMRE) and Median Magnitude Relative Error (MdMRE) are used for evaluating the results. Data is collected from TNPL and various algorithms such as M5, SVM, SVM with RBF kernel and SVM kernel functions with PSO optimization were used. Evaluated results show that proposed SVM regression with PSO approach gave improved accuracy with significant decrease in MMRE and MdMRE.

VI. Acknowledgements

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 Références Referencias

1. Ernest Orlando Lawrence. (1999). India's Pulp and Paper Industry: Productivity and Energy Efficiency, Berkeley National Laboratory.
2. Sandeep Kumar Kujur. (1980-81). Globalisation, Energy efficiency and Material Consumption in a Resource based Industry: A Case of India's Pulp and Paper Industry.
3. Parthiban K T and Govinda Rao M. (2008). Pulp wood based Industrial Agro forestry in Tamil Nadu – Case Study. Indian Forester".
4. *"National Forest Policy".* (1988). Government of India Ministry of Environment and Forests, New Delhi.
5. *Ministry of Environment and Forests, Govt. of India.* Forest Survey of India. (2009).
6. http://www.investopedia.com/terms/r/regression.asp.
7. Ruggeri, Kenett and Faltin, (2008). "Classification and Regression Tree Methods".
8. K. P. Soman, R. Loganathan, and V. Ajay, "Support vector machines and other kernel methods".
9. H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines". *9*, 155, 1997.
10. Belavendram, N, (2010). "Application of Genetic Algorithms for Robust Parameter Optimization". *International Journal of Automotive and Mechanical Engineering (IJAME),* 211-220.
11. V. Anandhi and R. Manicka Chezian, (2012). "Forecast of Demand and Supply of Pulpwood using Artificial Neural Network". *International Journal of Computer Science and Telecom-munications, 3.*

12. V. Anandhi and R. Manicka Chezian, (2013). "Support Vector Regression to Forecast the Demand and Supply of Pulpwood". *International Journal of Future Computer and Communication, 2*.
13. Smruti Rekha Das, Pradeepta Kumar Panigrahi, Kaberi Das, (2013). "Improving RBF Kernel Function of Support Vector Machine using Particle Swarm Optimization". *International Journal of Advanced Computer Research (IJACR), 2*.
14. Dimitri P. Solomatine and Michael Baskara L. A. (2010). Siek presented Flexible and Optimal M5 Model Trees with Applications to Flow Predictions.
15. Reza Shah Hosseini, Saied Homayouni and Reza Safari, (2012). "Modified algorithm based on Support Vector Machines for classification of hyper spectrual images in a similarity space". *Journal of Applied Remote Sensors, 6(1)*.
16. Gabriele Moser and Sebastiano B. Serpico, (2009). "Automatic Parameter Optimization for Support Vector Regression for Land and Sea Surface Temperature Estimation from Remote Sensing Data". *47*.
17. Yamille del Valle, G. K.-C. (2008). "Particle Swarm Optimization: Basic Concepts, Variants and Applications in Power Systems". *12*.
18. Bhatnagar, R., Bhattacharjee, V., & Ghose, M. K, (2010.). "Software Development Effort Estimation–Neural Network Vs. Regression Modeling Approach". *International Journal of Engineering Science and Technology, 2(7)*, 2950-2956.