Link Load Prediction using Support Vector Regression and Optimization

Debashree Priyadarshini
SOA University,
Jaganmohan Nagar,
Bhubaneswar-30

Milu Acharya
Professor
SOA University,
JaganMohan Nagar,
Bhubaneswar-30

Ambika Prasad Mishra
Asst.Professor
SOA University,
Jaganmohan Nagar,
Bhubaneswar-30

ABSTRACT

Prediction techniques are an interesting challenge in many areas like weather, banking, and finance, healthcare and so on. They are also becoming a popular subject in networking domain. This topic explores link load prediction of a network using Support Vector Regression and Optimization techniques. Support Vector Regression(SVR) is robust to outliers and can be used to online and adaptive learning.SVR has been used in other problems of networking like TCP throughput prediction, latency prediction and dynamic bandwidth provisioning in non- stationary traffic.Here,SVR and Optimization technique has been used to assess the bandwidth of a network dynamically.

General Terms
 Network Management

Keywords
Network forecast, Support Vector Regression, Support Vector Machines, Time series, Sequential Minimal Optimization.

1. INTRODUCTION
The heavy usage of Internet applications and other network services have given rise to complex network traffic. The complexity of the network services had led to network congestion [1, 2, 3, 6]. In the complex network services, it has become a very difficult task to allocate bandwidth to all users so that it is judiciously utilized by all [9, 15, 18]. So, dynamic allocation has become a major problem in networking.

Support Vector Regression [5, 10] was used for link load prediction of a network [1].This approach was not suitable for   short-term data and was slow in adapting non-stationary properties of the data. Convolution approach for bandwidth allocation [2] was not suitable in considerations with cost in calculation and storage. This made real-time calculations difficult. Probability-Hop Forecasting Algorithm was not applicable for unexpected turn in traffic parameters [6, 16, and 20].In the current approach, we have been using Support Vector Regression and Optimization techniques for link load prediction.

The remainder of the paper is organized as follows. In Section 2 we briefly explain the concepts of Support Vector Machines and introduced the basic idea of Support Vector Regression. In Section 3, we specify the sensitivity analysis of SVR performance .In Section 4, our proposed model is shown in a logical diagram. In Section 5, the performance metrics which are used to assess the prediction accuracy of SVR are discussed. In Section 6, the experimental conditions and analysis of experiments are discussed. In Section 7, the results of the study are discussed. Finally, concluding remarks and future work are addressed in Section 8.

2. SUPPORT VECTOR MACHINES AND SUPPORT VECTOR REGRESSION (SVM AND SVR)
SVMs were developed to solve the classification problem, but recently they have been extended to the domain of regression [4].

2.1. Support Vector Machine:-

Suppose we are given a training data $\{(x1, y1)… (xl, yl)\}$[5] X * R where 'X' denotes the space of the input patterns [1]. Our goal is to find a function 'f(x)'that has at most 'ε' deviation from the actually obtained targets 'y_i'for all the training data.

For linear case,
   f(x)=<w,x>+b   with w Є X, b Є R ------(1)

$< . , . > \rightarrow$ **dot product in X**

Rewriting (1) in the form of a convex optimization problem, we get
Minimize ½ ||w||²
Subject to
$$yi - < w, xi > -b \leq \varepsilon$$
$$< w, xi > +b - yi \leq \varepsilon \text{ ---- (2)}$$

Flatness in (1) can be ensured by minimizing the norm i.e. $||w||^2 = <w, w>$

Minimize $\frac{1}{2}||w||^2 + C \sum_{i=1}^{l}(\xi i + \xi i *)$--- (3) Subject to

$$yi - < w, xi > -b \leq \varepsilon + \xi i$$
$$< w, xi > +b - yi \leq \varepsilon + \xi i *$$
$$\xi i, \xi i * \leq 0 \text{ ----------------- (4)}$$

A ε-insensitive loss function '$|\xi|_{\epsilon}$' is described as

$$'|\xi|_\epsilon' = \begin{cases} 0 \; if \; |\xi| \leq \varepsilon \\ |\xi| - \in, else \end{cases} \text{ ----- (5)}$$

(3) can be solved more easily in its dual formulation obtained by constructing a Lagrange functions from the objective .The solution of the dual problem yields the function ' f(x)' ,which can be written as a linear combination of training data, Lagrange multipliers α_i , α_i^* and constant term 'b'.

$$f(x) = \sum_{i=1}^{S}(\propto i - \propto i *) < xi, x > +b \text{ --- (6)}$$

$= \sum_{i=1}^{Sv} (\propto i - \propto i *) < xi, x > + b$ ------ (7)

Subject to

$\sum_{i=1}^{S} (\propto i - \propto i *) = 0$

$0 \leq \propto i, \propto i * \leq C$ ---------------- (8)

2.2. Support Vector Regression:-

Support Vector Regression (SVR) [10] attempts to minimize the generalization error bound so as to achieve generalized performance. The idea of SVR is based on the computation of a linear regression function in a high dimensional feature space where the input data are mapped via a nonlinear function. The model produced by Support Vector Regression only depends on a subset of the training data because the cost function for building the model ignores the training data that is close to the model prediction.SVR has been applied in various fields-time series and financial (noisy and risky) prediction, approximation of complex engineering analyses, convex quadratic programming and choices of loss functions [13, 14, 17].

2.3. Kernel Trick:-

The dual formulation provides the key to the nonlinear extensions of SVR.This peculiarity of SVM is called as "Kernel Trick" [11]. The input data is mapped into a higher dimensional space F by a function

$$\Phi: R^d \rightarrow F$$

A linear regression in this new space F is equivalent to a non-linear original space. A function **K (x, x')** is called a kernel if it corresponds to a dot product in some feature space and if it satisfy Mercer's condition [6].

Using Kernel "f(x)" can be written as

$f(x) = \sum_{i=1}^{S} (\propto i - \propto i *) K(xi, x) + b$ ----- (9)

There are three kernel functions usually used [18]:-

1) Polynomial Kernel
   $K(xi, xj) = (xi. xj + 1)^d$
2) Sigmoid Kernel
   $K(xi, xj) = \tanh [c1(xi.xj) + c2]$

3) RBF Kernel
   K (xi, xj) =exp (-||xi-xj||²/σ²) ˆd

3. SENSITIVITY ANALYSIS OF SVR PERFORMANCE

Support Vector Regression is affected by many factors [1, 6] like training size(S), smoothing factor(C), parameter of the kernel function (γ), timescale (τ), and dimension (d) .The timescale (τ) is the observation timescale. The dimension (d) is the minimum number of state variables required to describe the system and their product "τd" is the average system memory length.

In this paper, we firstly reconstruct a set of training samples G={(X$_t$,d$_t$),t= 1,2 ,3,…,n-p } where

X$_t$ =(x$_t$,x$_{t+1}$,… x$_{t+p-1}$) , **d$_t$** = x$_{t+p}$ is the length of time series {**x$_t$**} , and **"p"** is the embedding dimension of the training set G.We then train a SVM using the technique of Sequential Minimal Optimization on a given training set and predict the future value using predictor[7,17,21].

3. PROPOSED PROCESS

Our propose architecture is described schematically in Figure 1 as shown at the bottom of the paper.

4. PERFORMANCE METRICS

To quantitatively assess forecasting performance, the Mean Square Error (MSE), Normalized Mean Square Error (NMSE) and Root Mean Square Error (RMSE) are used[1,6].

1) **MSE:-**

$MSE = \frac{1}{M} \sum_{t=1}^{M} (X_t - \hat{X}_t)^2$-------------- (10)

Where X$_t$ is the observed value of the time series at time t, xˆ is the predicted value of x$_t$ and M is the total number of the predicted value.As the prediction accuracy increases , the MSE becomes smaller.

2) NMSE:-

$NMSE = \frac{1}{\sigma^2} \frac{1}{M} \sum_{t=1}^{M} (X_t - \hat{X}_t)^2$---------- (11)

σ^2 = variance of the time series over the prediction duration.If NMSE > 1 ,the prediction performance is worse than that of a trivial predictor.

3) RMSE:-Root Mean Square assesses the quality of the estimator in terms of its variation and unbiasedness.

RMSE= $\sqrt{(1/n \sum_{i}^{n} (x - x')^2)}$ -------- (12)

5. EXPERIMENTAL CONDITIONS AND ANALYSIS

5.1. Selection of the parameters used in SVM Model

When applying SVM to link load prediction the first to be considered is what kernel function is to be used. The kernel to be used is Multi-linear kernels. The second thing to be considered is the hyper parameters (ε, σ, C and p) of the learning machine. "ε" is the zone of error insensitivity of our cost function. "σ" is the variance. A smaller value of σ implies stronger capacity of the learning machine while a higher value usually leads to minor test error. "C" is the regularization coefficient that determines the trade off between the model and complexity (flatness) and the degree to which deviation larger than "ε" is tolerated in optimization formula (3).A larger C assigns higher penalties to errors so that the regression is trained to minimize error with lower generalization while a small C assigns fewer penalties to errors. If C goes to infinity, SVR would not allow the occurrence of any error and result in a complex model, whereas when C is zero, the result would tolerate a large number of errors and the model would be less complex. "p" is the embedding dimension i.e. the number of past samples we supposed will be useful to model the observed process.

5.2. Experimental Conditions

The dimension'd' and timescale 'τ' are fixed. d is fixed as 5 and τ is 1 second. The dataset is constructed with all possible inputs and outputs.30% of the dataset is selected at random to train in SVM and the remaining 70% is used for prediction as test data. The SVR forecast accuracy can also be estimated in terms of their cost. There are 2 tasks - model training and actual forecasting. For model training, the technique to be used is the Sequential minimal Optimization decomposition technique. In sequential Minimal Optimization (SMO) decomposition technique; the computational complexity is linear in the number of support vectors [21].

6. RESULTS

According to the above mentioned process, we choose 300 traffic data as the experiment data. The first 100 is the training data and the latter 200 is the testing or prediction data. After the model is well trained, we use it to predict. Mean Square Error (MSE) [6], Normalized Mean Square Error (NMSE) [6] and Root Mean Square Error (RMSE) [1] are applied to compare the performance of SVM with Auto-Regressive (AR) [1] and Moving Average Models (MA) [1].

The graph in Figure 2(at the bottom of the paper) shows that the SVM has least error as compared to the other models (Auto-Regressive and Moving-Average).Hence, SVR proves to be a better approach for link load prediction than the other existing models. Support Vector Machines excels previously used models in terms of RMSE; MSE and NMSE.The performance of SVM is quite satisfactory. The theoretical advantages and better performance indicate a very promising implementation of SVM for real-time traffic condition prediction.

7. DISCUSSION AND CONCLUSION

In this paper, SVM Model and Optimization techniques [7] are used to predict link load in a network. This can provide a bright direction in dynamic bandwidth management, admission control and better performance in networks.

SVM is used in link load because of its many positive aspects. Some of these are: SVM models are robust to parameter variation, they can generalize to unseen data, and they are well accepted in continuous and adaptive online learning. This feature is an extremely desirable property in network environments.

This work poses a starting point for further research. The performance of SVM in link load prediction can be enhanced by using some more sophisticated techniques or by considering the statistical properties of the input data.

9. REFERENCES

[1] P.Bermolen, D.Rossi. "Support Vector Regression for Link Load Prediction", Elsevier, Computer Networks, France, 2008

[2] P.Bermolen, D.Rossi. "Network Forecast with Support Vector Machines" TELECOM Paris Tech.France-INFRES Department

[3] B.Krithikaivasan, K.Deka, D.Medhi. "Adaptive Bandwidth Provisioning Envelope Based on Discrete Temporal Network Measurements" School of Computing and Engineering.USA

[4] B.Krithikaivasan, Y.Zeng, K.Deka, D.Medhi. "ARCH-based Traffic Forecasting and Dynamic bandwidth Provisioning for Periodically Measured Non-stationary Traffic" IEEE/ACM Transactions on Networking, vol X, No .Y, 2006

[5] S.R.Gunn. "Support Vector Machines for Classification and Regression" .Technical Report

[6] H.Feng, Y.Shu, S.Wang, M.Ma. "SVM based Models for Prediction WLAN Traffic"

[7] J.C.Platt. "Fast Training of Support Vector Machines using Sequential Minimal Optimization "1Microsoft Way,Redmond,WA 98052,USA

[8] M.Mirza, J.Sommers, P.Bardford, X.Zhu. "A machine learning approach to TCP throughput prediction,Proc.of ACM SIGMETRICS '07,San Diego,USA,2007

[9] W.E.Leland, M.S.Taqqu, W.Willinger, D.V.Wilson, "On the self-similar nature of Ethernet traffic", IEEE Transaction on Networking 2(1994)

[10] A.J.Smola, B.Scholkopf. "A Tutorial on Support Vector Regression" September 30, 2003

[11] N.Cristianini, J.Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods", Cambridge University Press, New York, NY, 1999

[12] S.Haykin, L.Li, "Nonlinear Adaptive prediction of Non-stationary signals", IEEE Transactions on signal Processing, Vol.43, No.2, Feb 1995, pp.526-535

[13] S.Ruping, M.Katharina, "Support Vector Machines and Learning about time" Proc.of IEEE ICASSP'03, Hong Kong, 2003

[14] K.R.Muller, A.Smola, G.Ratsch, B.Scholkopf, J.Kohlmorgen, V.Vapnik. "Predicting time series with Support Vector Machines", Artificial Neural Networks, Lecture Notes in Computer Science,vol.1327,Springer,Berlin,1999

[15] R.Beverly, K.Sollins, A.Berger "SVM Learning of IP Address Structure for Latency Prediction".

[16] G.Wang, "A Survey on training Algorithms for Support Vector Machine Classifiers" Fourth International Conference on Networked Computing and Advanced Information Management

[17] P.J.Brockwell, R.A.Davis "Introduction to Time-Series and Forecasting", Springer, Berlin, 1996

[18] W.Luo, X.Liu, J.Zhang. "SVM based analysis and prediction of network traffic"

[19] I.Mierswa, M.Wurst, R.Klinkenberg, M.Scholz, T.Euler, YALE: rapid prototyping for complex data mining task, Proc. Of ACM SIGKDD'06, P.A.USA, 2006

[20] D.Sun, Q.Wang. "The Development of a Support Vector Machine (SVM) based Traffic Condition Predictor and Comparison with other Methods" 2010 13th International Transportation Systems. Madeira Island, Portugal, Sep 19—22, 2010.

[21] P.H.Chen, R.E.Fan, C.J.Lin. "A Study of SMO-Type Decomposition Methods for Support Vector Machines" .Department of Computer Science, National Taiwan University, Taipei 106, Taiwan

8. AUTHORS PROFILE

Debashree Priyadarshini is pursuing her Master Degree at Institute of Technical Education and Research, SOA University; Bhubaneswar.Her research interests include Soft Computing and Communication Networks.

Milu Acharya obtained her PhD from Utkal University.Currently, she is a Professor in Department of Mathematics, Institute of Technical Education and Research, and Bhubaneswar.Her research interests include Numerical Analysis, Optimization and Complex Analysis.

Ambika Prasad Mishra is an Assistant Professor in Department of Computer Science and Engineering, Institute of Technical education and Research; Bhubaneswar.His research interests include Soft Computing and Algorithm Design.
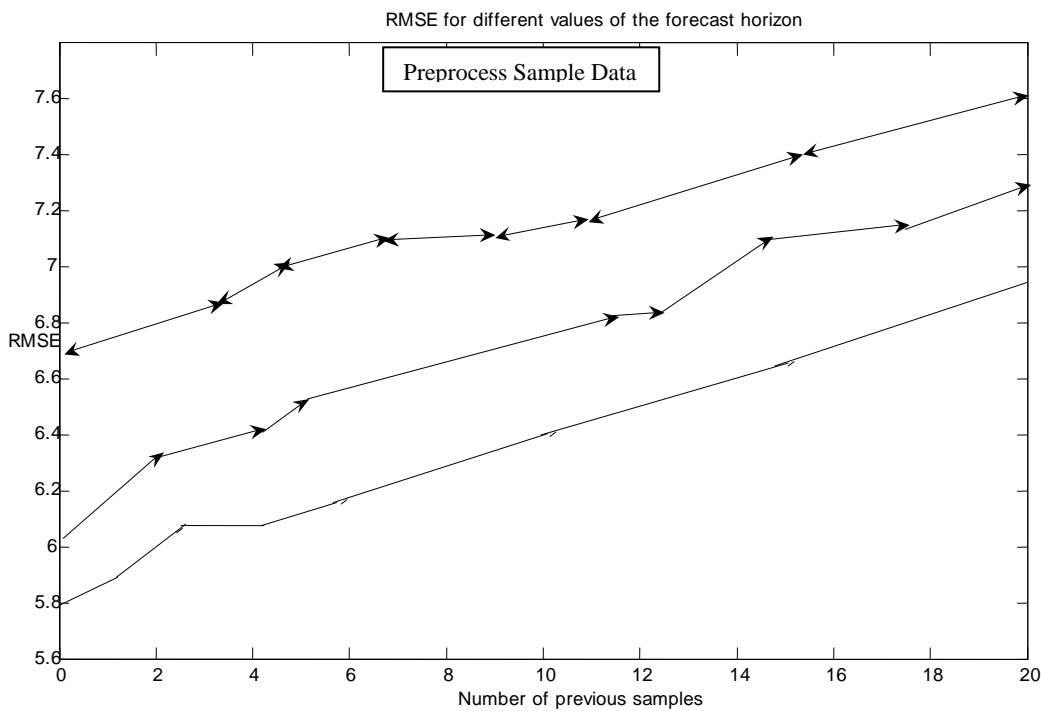
Figure 1:-
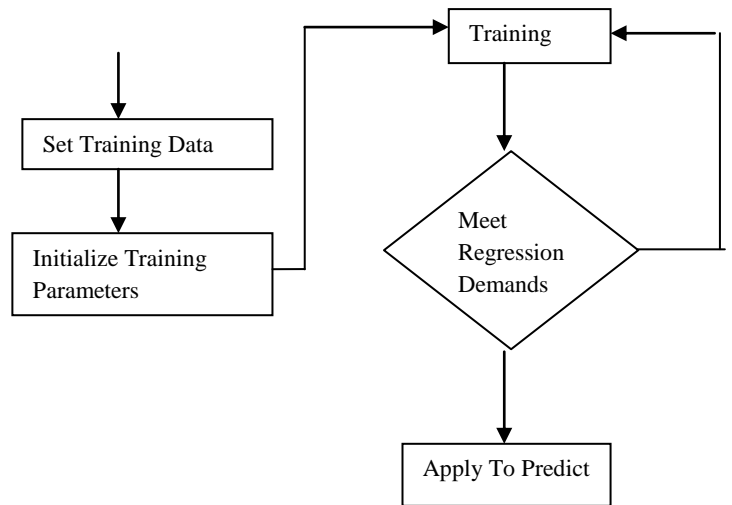
Figure 2:- Results of the proposed Process