Development of New Models Using Machine Learning Methods Combined with Different Time Lags for Network Traffic Forecasting

Derman Akgöl

Mehmet Fatih Akay (advisor)

DISSERTATION.COM

Irvine, California
Boca Raton, Florida

Development of New Models Using Machine Learning Methods Combined with Different Time Lags for Network Traffic Forecasting

Dissertation.com
Boca Raton, Florida • Irvine, California
USA • 2017

ISBN-10: 1-61233-460-1 (ebk.)
ISBN-13: 978-1-61233-460-8 (ebk.)

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS

Derman AKGÖL

DEVELOPMENT OF NEW MODELS USING MACHINE LEARNING METHODS COMBINED WITH DIFFERENT TIME LAGS FOR NETWORK TRAFFIC FORECASTING

DEPARTMENT OF COMPUTER ENGINEERING

ADANA, 2016

ABSTRACT

The purpose of this thesis is to forecast the amount of network traffic in Transmission Control Protocol/Internet Protocol (TCP/IP) -based networks by using different time lags and various machine learning methods including Support Vector Machines (SVM), Multilayer Perceptron (MLP), Radial Basis Function (RBF) Neural Network, M5P (a decision tree with linear regression functions at the nodes), Random Forest (RF), Random Tree (RT), and Reduced Error Prunning Error (REPTree), and statistical regression methods including Multiple Linear Regression (MLR) and Holt-Winters and compare the performance of statistical and machine learning methods. Two different Internet Service Providers (ISPs) as a traffic data (bits) have been utilized to build traffic forecasting models. The datasets have been split into training and test sets. The first 66% of the datasets have been utiliced as training sets and the reamining have been used as test sets. The performance of the forecasting models for the datasets has been assessed using Mean Absulote Percentage Error (MAPE). The results show that generally, SVM based and M5P based models perform better than the ones obtained by the other methods.

Key Words: Machine learning, time series, traffic engineering, time lags.

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Assoc. Prof. Dr. M. Fatih AKAY, for his supervision guidance, encouragements, patience, motivation, useful suggestions and his valuable time for this work.

I would like to thank members of the MSc thesis jury, Asst. Prof. Dr. Zehan KESİLMİŞ and Asst. Prof. Dr. Buse Melis ÖZYILDIRIM, for their suggestions and corrections.

I would also like to thank Cukurova University Scientific Research Projects Center for supporting this work (Project no: FYL-2015-5265)**.**

Last but not the least, I would like to thank my family for their endless support and encouragements for my life and career.

CONTENTS **PAGE**

1. INTRODUCTION

1.1. Internet Traffic Prediction

Internet traffic is the transfer of data across the Internet. Amount of Internet traffic is rising, and several and large number of packets are sent thorough all over the world (Hasegawa et al., 2001). Internet/network traffic analysis and modeling is an effective structure to characterize network performance; therefore, it has been a crucial point in many studies (Chen et. al., 2012).

Lately, since communication and network technologies have developed quickly, traffic characteristic is altering extremely. The research about Internet traffic analysis and modeling has varied from the large time scale to the small time scale. The researches have indicated that the traffic characteristics of the small time scale reacted differently from the large time scale's one (Chen et. al., 2012). Since the characteristic of Internet traffic is getting more and more sophisticated, this creates new difficulties to the management of the network. In order to overcome these problems, development of Internet/network traffic forecasting models have become one of the most active research areas. A traffic forecasting model needs to be able to express the characteristics of the network in the past and also, needs to be forecast the development of this network in the near future (Wang et al., 2008).

The predictability of Internet traffic is substantial point in many fields such as adaptive application, admission control, wireless and network management (Rutka and Lauk, 2015). To predict Internet traffic is very important to understand communication networks, optimize resources and have a better quality of service. As well, by comparing the real traffic with the forecast, anomalies (such as security attacks, viruses, etc.) can be detected with the help of traffic forecasting (Cortez et. al.,2007). Also, the predicted results can be used as a significant reference for the bandwidth allocation Internet traffic control and error control in management (Bai et. al.,2009).

The field of Time Series Forecasting (TSF) considers the prediction of chronologically sorted predictors, where the aim is to identify a complicated system as a black-box and forecasting its behavior based on historical data. The TSF approaches can be split into two parts as univariate and multivariate according to the number of variables used, i.e. one or more variables. Multivariate methods are expected to generate better results when the variables are correlated have been used (Cortez et al., 2007).

1.2. Previous Work

In (Hasegewa et. al., 2001), Local Linear Approximation (LLA), Radial Basis Function Neural Network (RBF), Support Vector Machines (SVM) were applied to develop Internet traffic models. First, these three methods were utilized to predict chaotic time series, and then a simple version of the LLA method was chosen because of easiness to apply and have high predictability. As a result, it has been observed that the nonlinear time series prediction method could be active for prediction of Internet traffic.

In (Sang and Li, 2002), the predictability of network traffic was assessed by using two metrics, namely the bounded error and minimum prediction error. The Auto-Regressive Moving Average (ARMA) and Markov-Modulated Poison Process (MMPP) methods have been used for assessment of the accuracy of the prediction models. The results have shown that proper traffic measurement and multiplexing improved accuracy for both ARMA and MMPP based prediction models.

In (Tong et. al., 2004), authors used boosting for network traffic prediction by considering it as a classical regression problem. The recommended algorithm took over Feed-Forward Neural Network (FFNN) as the basic learner, to acquire the non-linear characteristic of network traffic and increase using boosting to constrain overfitting. The results have shown that boosting by using FFNN is an effective method.

In (Alarcon-Aquiro and Barria, 2006), authors used multiresolution Finite-Impulse-Response Neural Network (FIRNN) based learning algorithm by applying the analysis into wavelet theory to predict network traffic. The Maximal Overlap Discrete Wavelet Transform (MODWT) and Multiresolution Multi-Layer

Perceptrons (MMLP) have been used in this study. The results have indicated that the network traffic prediction was improved by using FIRNN learning algorithm and MODWT.

In (Cortez et. al., 2006), authors used Neural Network Ensemble (NNE), Naïve-Benchmark, Holt-Winters, Auto-Regressive Integrated Moving Average (ARIMA) to develop Transmission Control Protocol/Internet Protocol (TCP/IP) network prediction models. The results have shown that NNE performed much better than other TSF methods.

In (Cortez et. al., 2007), Neural Network (NN), Holt-Winters, and Naïve-Benchmark have been utilized to build TCP/IP network forecasting models. It has been shown that NN-based models gave lower error rates than the ones obtained by the other methods. Also, it has been reported that Naïve-Benchmark based models gave the highest error rates TCP/IP network forecasting.

In (Wang et. al., 2008), authors have considered Back Propagation (BP) neural networks models to develop Internet traffic forecasting models by using Radial Basis Function optimized by Genetic Algorithm (GA-RBF) and BP Neural Network. According to the results obtained, GA-RBF based models performed much better than BP Neural Networks.

In (Zhang and Liu, 2009), Least Squares Support Vector Machine (LS-SVM) has been utilized to forecast Internet traffic. Also, authors have used five more machine learning methods including Kalman Filtering (KF), ARMA, Historical Mean (HM), RBF, and Support Vector Regression (SVR) for comparison. It was concluded that LS-SVM based models performed much better than the ones obtained by the other methods in different time scales such as a week, a day, and peak periods.

In (Jiang et. al., 2009), ARMA and FARIMA have been used for network traffic prediction. In this paper, the relation between time scales and time series models has not been examined. The results have shown that the performance of Auto-Regressive Fractionally Moving Average (FARIMA) based models did not give any advantage over the other method.

In (Huang and Sadek, 2009), the SPN has been utilized to forecast Internet traffic. Two more machine learning methods including BP and Nearest

Neighborhood (NNB) algorithm have been used for comparison. The results have shown that the accuracy of Spinning Network (SPN) based models showed superior performance.

In (Chabaa et. al., 2009), authors have applied the Adaptive Neuro-Fuzzy Inference System (ANFIS) to forecast TCP/IP network. The authors concluded that ANFIS based model has been very effective to forecast Internet traffic.

In (Castro-Neto et. al., 2009), authors have used Online Support Vector Regression (OL-SVR), which is an application of statistical regression method and three more machine learning methods including Gausian Maximum Likelihood (GML), Holt-Exponential Smoothing (HES), and Artificial Neural Network (ANN) for comparison of Internet traffic forecasting models. The results have shown that even though GML performed better than other methods in literature, in this study, OL-SVR has performed slightly better than GML.

In (Bermelon and Rossi, 2009), SVR has been used to forecast link load forecasting. The results have indicated that SVR based models have better result than the ones obtained by the other methods. That is, Moving Average (MA) and Auto-Regressive (AR) were not good enough methods for link load forecasting.

Detailed information of the studies and methods used in each study between the years 2001 and 2009 are given in Table 1.1.

In (Syed et. al., 2010), authors used the wavelet filters based Seasonal Autoregressive Moving Average (SARIMA) method to develop Internet traffic forecasting models. Authors also used simple SARIMA method to compare results of wavelet based SARIMA. The results showed that wavelet based SARIMA markedly improved the performance of forecasting models.

In (Chabaa et. al., 2010), Multi-Layer Perceptrons (MLP) based models have been developed for Internet traffic forecasting. The Levenberg-Marquardt (LM) and Resilient Back Propagation (RP) algorithms also have been applied for comparison purposes. It was concluded that LM and RP were very effective algorithms to prove accuracy of Internet traffic forecasting.

Table 1.1 Summary of studies between the years 2001 and 2009

Study	Methods
Hasegawa et. al. (2001)	LLA, RBF, SVM
Sang et. al. (2002)	ARMA, MMPP
Tong et. al. (2004)	Boosting with FFNN
Papagiannaki et. al. (2005)	WMRA, ARIMA
Alarcon-Aquino et. al. (2006)	FIRNN, MMLP, MODWT
Cortez et. al. (2006)	NNE, Naïve-Benchmark, Holt-Winters, ARIMA
Cortez et. al. (2007)	NN, Holt-Winters, Naïve-Benchmark
Wang et. al. (2008)	GA-RBF, BP
Zhang et. al. (2009)	LS-SVM, KF, ARMA, HM, RBF, SVR
Jiang et. al. (2009)	ARMA, FARIMA
Huang et. al. (2009)	SPN, BP, NNB
Chabaa et. al. (2009)	ANFIS
Chang et. al. (2009)	GM, ARMA, RBF, GARCH, ANFIS, ASVR-ANFIS/NGARCH
Castro-Neto et. al. (2009)	OL-SVR, GML, HES, ANN
Bermelon et. al. (2009)	SVR, MA, AR

LLA, Local Linear Approximation; **RBF**, Radial Basis Function Neural Network; **SVM**, Support Vector Machine; **ARMA**, Auto-Regressive Moving Average; **MMPP**, Markov-Modulated Poisson Process; **FFNN**, Feed-Forward Neural Network; **WMRA**, Wavelet Multiresolution Analysis; **ARIMA**, Autoregressive Integrated Moving Average; **FIRNN**, Finite-Impulse-Response Neural Network; **MMLP**, Multiresolution Multilayer Perceptron; **MODWT**, Maximal Overlap Discrete Wavelet Transform; **NNE**, Neural Network Ensemble; **NN**, Neural Networks; **GA-RBF**, Radial Basis Function optimized by Genetic Algorithm; **BP**, Back Propagation; **KF**, Kalman Filtering; **HM**, Historical Mean; **SVR**, Support Vector Regression; **FARIMA**, Autoregressive Fractionally Moving Average; **SPN**, Spinning Network; **NNB**, Nearest Neighborhood; **ANFIS**, Adaptive Neuro-Fuzzy Inference System; **GM**, Grey Model; **GARCH**, Generalized Autoregressive Conditional Heteroscedasticity; **ASVR-ANFIS/NGARCH**, ANFIS with nonlinear GARCH by adaptive SVR; **OL-SVR**, online support vector regression; **GML**, Gaussian maximum Likelihood; **HES**, Holt-Exponential Smoothing; **MA**, Moving Average; **AR**, Auto-Regressive.

In (Tan et. al., 2010), the effectiveness of using ARMA based models on Internet traffic forecasting has been analyzed. The results of the study were positive and it has shown that the most accurate forecasting models have been obtained by step size of 30 seconds.

In (Hong et. al., 2011), SVR with Continuous Ant Colony Optimization (SVRCACO) based models have been developed for inter-urban traffic forecasting since SVR has been widely used in literature. The results have indicated that

SVRCACO based models performed better than the ones obtained by SARIMA, which has been used for comparison.

In (Kim, 2011), authors used the Auto-Regressive - Generalized Auto-Regressive Conditional Heteroscedasticity (AR-GARCH) to develop Internet traffic forecasting models. The performance of AR-GARCH based models has been compared with the ones developed by ARIMA. The results have shown that AR-GARCH based models were more accurate than ARIMA based models.

In (Hong, 2011), authors used the Seasonal SVR with Chaotic Annealing Algorithm (SSVRCSA) method to develop inter-urban traffic forecasting models. Also, SARIMA, Back Propagation Neural Network (BPNN), and Seasonal Holt-Winters (SHW) based models have been built for comparison purposes. The authors concluded that SSVRCSA based models yielded more accurate results than the ones obtained by the other methods.

In (Chen et. al., 2012), the FNT based models have been built for prediction of network traffic. Flexible Neural Tree (FNT) methods have been developed by using genetic programming. The results have indicated that FNT with genetic programming model was accurate to develop network traffic forecasting models and performed better than the FNT with the FFNN.

In (Miguel et. al., 2012), -Lagged Feed Forward Networks (TLFN) based models have been developed to predict long-term Internet traffic. Also, the authors used Holt-Winters to compare with TLFN. The authors concluded that TLFN based models gave more accurate results than the ones obtained by Holt-Winters.

In (Cortez et. al., 2012), the authors used Naïve-Benchmark, Holt-Winters, ARIMA, and ANN to forecast Internet traffic. It has been concluded that while ANN based models gave the best result for 5-minute and hourly data sets, Holt-Winters based models were more accurate than the ones obtained by the other method for daily data set.

In (Oliveira et. al., 2014), MLP and Stacked Autoencoder (SAE) have been utilized to build Internet traffic forecasting models. The results of MLP and SAE based models have been compared. The results have shown that even though SAE was a complex method, MLP based models performed much better than SAE based models.

In (Ratrout and Gazder, 2014), authors have used LR and ANN methods including MLP and RBF to develop Internet traffic forecasting models for daily data sets. Accuracy of the models has been examined in this paper. Authors concluded that ANN based models performed much better than Linear Regression (LR) based models. However; when ANN methods were examined separately, the performance of MLP and RBF based models were very close to each other.

In (Kamińska-Chuchmala, 2014), geostatistical estimation method called Ordinary Kriging (OK) has been used for spatial Internet traffic load forecasting. Authors concluded that giving estimated values on whole considered area was an advantage of the OK method. On the other hand; the accuracy of OK based models was not good enough.

In (Rutka and Lauks, 2015), authors have used FFNN for prediction of network traffic. Also, the accuracy of the predicted traffic and estimated prediction interval has been examined. As a result, authors reported that numerical studies of real traffic traces to verify the prediction of real network traffic was not so easy.

In (Katris and Daskalaki, 2015), authors used FARIMA, two ANNs method including MLP and RBF, Holt-Winters, ARIMA/GARCH, FARIMA/GARCH, hybrid FARIMA+RBF, and hybrid FARIMA+MLP. It was concluded that the hybrid models performed much better than the other methods. On the other hand; FARIMA/GARCH was more effective method than ARIMA/GARCH and Holt-Winters for Internet traffic forecasting.

In (Akgol et. al., 2015), SVM, MLP, RBF, and Random Tree (RT) have been employed to forecast Internet traffic. The data sets in the study of (Cortez et. al., 2012) have been used in this study. Two different time lags for each time scale (5 minute, hourly, and daily) have been utilized to build Internet traffic models. The authors concluded that small time scale (5 minute) gave better results than the large time scale (daily) for all methods except RBF. Also, it was concluded that SVM and MLP based models performed better than the ones developed by the other methods while RBF based models performed worse than the ones developed by the other methods.

In (Akgol and Akay, 2016), authors have used different machine learning methods including SVM, MLP, RBF, Random Forest (RF), RT, and Reduced Error Pruning (REPTree) to forecast Internet traffic. The data sets in the study of (Cortez et. al., 2012) have been utilized in this study. Also, three different time lags have been used for each data set (5 minute and hourly) to develop Internet traffic forecasting models. Mean Absulote Percentage Error (*MAPE)* has been used as a performance metric. The results have shown that SVM based models yielded lower *MAPE* values than the ones obtained by the other methods.

Detailed information of the studies and methods used in each study between the years 2010 and 2016 are given in Table 1.2.

Table 1.2 Summary of studies between the years 2010 and 2016.

Study	Methods
(Chen et. al., 2010)	BP-ANN
(Syed et. al., 2010)	Wavelet Filter based SARIMA, SARIMA
(Chabaa et. al., 2010)	MLP, LM, RP
(Tan et. al., 2010)	ARMA
(Hong et. al., 2011)	SVRCACO, SARIMA
(Liu et. al., 2011)	CTSA, LSVM
(Kim, 2011)	AR-GARCH, ARIMA
(Hong, 2011)	SSVRCSA, SARIMA, BPNN. SHW
(Chen et. al., 2012)	FNT-GP, FNT-FFNN
(Miguel et. al., 2012)	TLFN, Holt-Winters
(Cortez et. al., 2012)	Naïve-Benchmark, Holt-Winters, ARIMA, ANN
(Maurya et. al., 2012)	FIS
(Oliveira et. al., 2014)	MLP, SAE
(Ratrout et. el., 2014)	LR, MLP, RBF
(Kamińska-Chuchmała, 2014)	OK
(Rutka et. al., 2015)	FFNN
(Katris et. al., 2015)	FARIMA, MLP, RBF, Holt-Winters, ARIMA-GARCH, FARIMA-GARCH, hybrid FARIMA-RBF and FARIMA-MLP
(Akgol et. al., 2015)	SVM, MLP, RBF, RT
(Akgol et. al., 2016)	SVM, MLP, RBF, RF, RT, REPTree

BP-ANN, Back propagation Artificial Network; **SARIMA**, Seasonal Auto Regressive Moving Average; **MLP**, Multilayer Perceptron; **LM**, Levenberg-Marquard; **RP**, Resilient Back Propagation; **ARMA**, Auto Regressive Moving Average; **SVRCACO**, Support Vector Regression with Continuous Ant Colony; **CTSA**, Chaotic Time Series Analysis; **LSVM**, Local Support Vector Machine; **AR-GARCH**, Autoregressive-generalized conditional heteroscedascity; **ARIMA**, Autoregressive Integrated Moving Average; **SSVRCSA**, Seasonal Support Vector Regression with Chaotic Simulated Annealing; **BPNN**, Back-Propagation Neural Network; **SHW**, Seasonal Holt-Winters; **FNT-GP**, Flexible Neural Tree with Genetic Programming; **FNT-FFNN**, FNT with Feed Forward Neural Network; **TLFN**, Time-Lagged Feed Forward Networks; **ANN**, Artificial Neural Network; **FIS**, Fuzzy Inference System; **SAE**, Stacked Autoencoder; **LR**, Linear Regression; **RBF**, Radial Basis Function Neural Network; **OK**, Ordinary Kriging; **FARIMA**, Autoregressive Fractionally Integrated Moving Average; **RT**, Random Tree; **RF**, Random Forest; **REPTree**, Reduced Error Pruning.

1.3. Motivation, Purpose and Contributions of This Thesis

In literature, to the best of our knowledge, although there exists several studies which predict the network traffic with the help of statistical as well as machine learning regression methods, there is no comprehensive study that compares the performance of different machine learning methods for prediction of network traffic on different data sets using several time lags.

The aim of this thesis is to extend the work of Cortez et. al. (2012) and forecast/predict the amount of traffic in TCP/IP-based networks by using various machine learning methods. The machine learning methods that have been employed are SVM, MLP, RBF, M5P (Decision Tree with Linear Regression Functions at the Nodes), RF, RT, and REPTree and the statistical regression methods employed are MLR and Holt-Winters. The performance of statistical and machine learning regression methods has been compared in this thesis.

The main differences between this research proposal and the studies from related literature can be summarized as follows:

- This is the first study ever used trees as machine learning methods such as M5P, RF, RT, and REP tree to forecast Internet traffic.
- To the best our knowledge, there is no study that used different time lags in the same data set to develop new forecasting models. By using different time lags along with two data sets, this will be comprehensive study regarding the number of models to be developed for forecasting network traffic.
- In addition, the SVM method has not been examined according to the kernel type in any of the studies in literature. This is the first study that compares the performance of SVM with different kernel.

1.4. Overview of Data Sets

Two different Internet Service Providers (ISPs) traffic data (bits) have been utilized to build traffic forecasting models. Related with the time scales, the forecasting types can be explained as below (Cortez et. al., 2012):

- Real time; which includes data that not going beyond a few minutes and supposes an online forecasting system;
- Short term; from one to several hours, important to detect anomalies or control optimality;
- Middle term; from one to several days, employed to plan facilities;
- Long term; generally carried several months/years and used for strategic decisions.

Because of the characteristics of the Internet traffic collected, the data sets used in this thesis have been categorized as the first three types. Thus, three time series were formed for each data set by collecting all inputs in a given period of time. Each data set consists of different time scales including 5-minutes, 1-hour, and 1-day.

1st data set: The first data set (referred to as DS1) has been provided from a private ISP with centers in eleven European cities. The data set was saved between 07.06.2005, at 6:57 a.m. and 29.07.2005, at 11:17 a.m. The DS1 data was recorded every 30 seconds (Cortez et al., 2012).

2nd data set: The second data set (referred to as DS2) has been provided from United Kingdom education research networking association (UKERNA). The data set was saved between 19.11.2004, at 9:30 a.m. and 27.01.2005, at 11:11 a.m. The DS2 data was saved every 5 minutes (Cortez et al., 2012).

The graphic of the selected time scales are given in Figure 1.1, Figure 1.2, Figure 1.3, Figure 1.4, Figure 1.5, Figure 1.6, respectively.

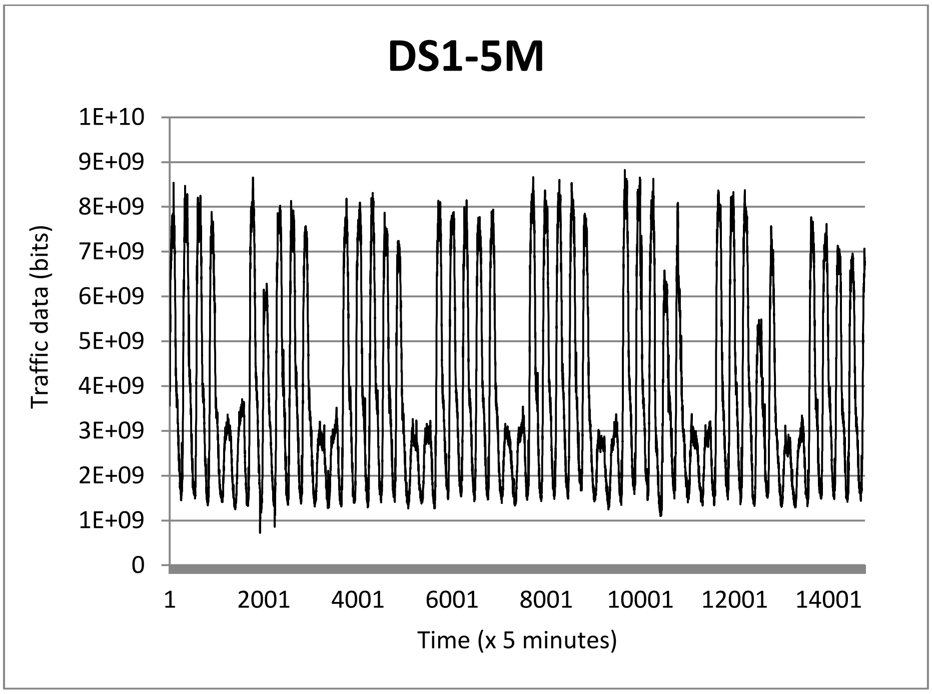

Figure 1.1. The Internet traffic time series for DS1-5M

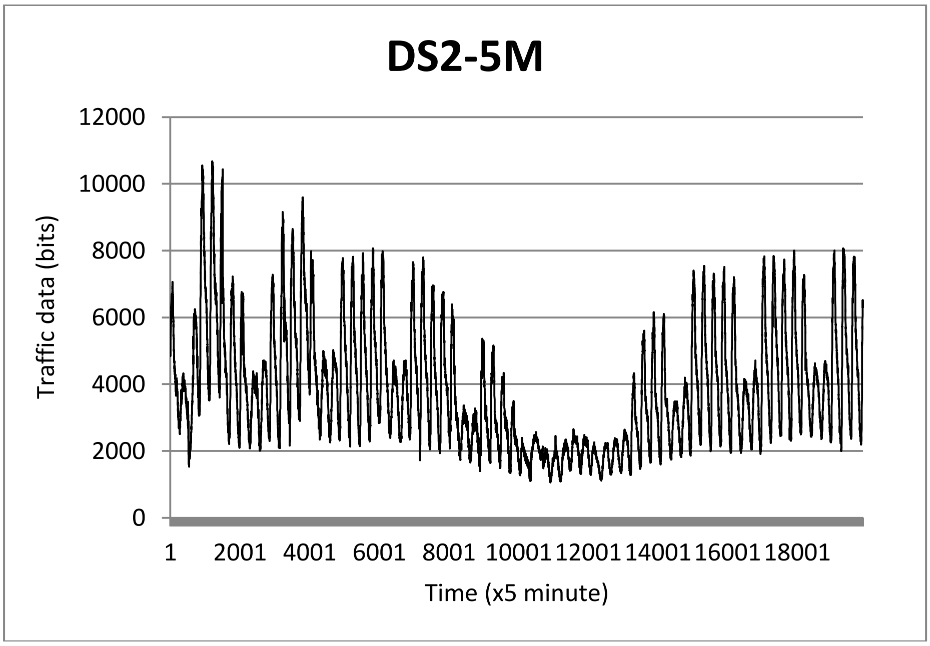

Figure 1.2. The Internet traffic time series for DS2-5M

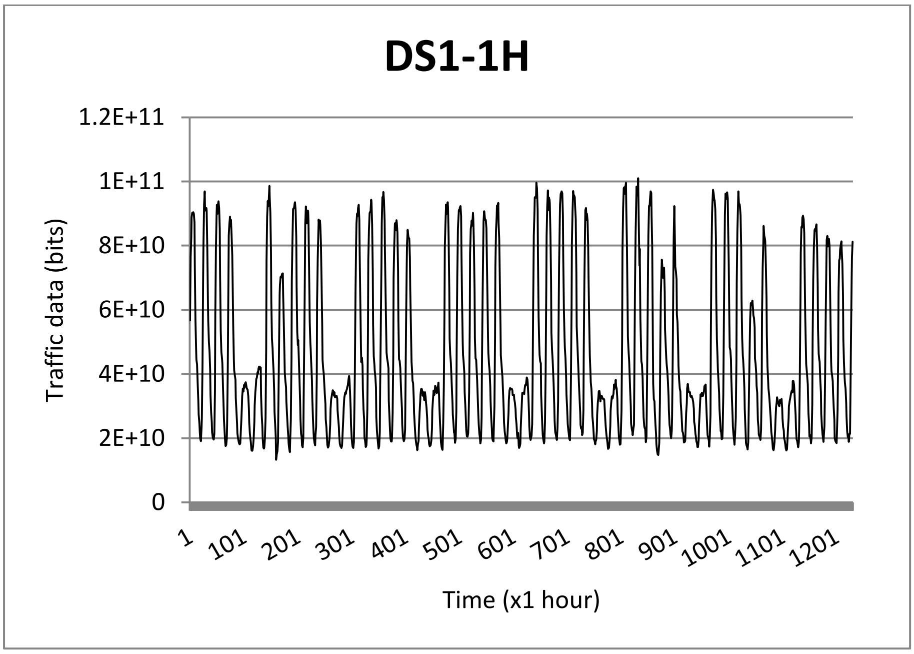

Figure 1.3. The Internet traffic time series for DS1-1H

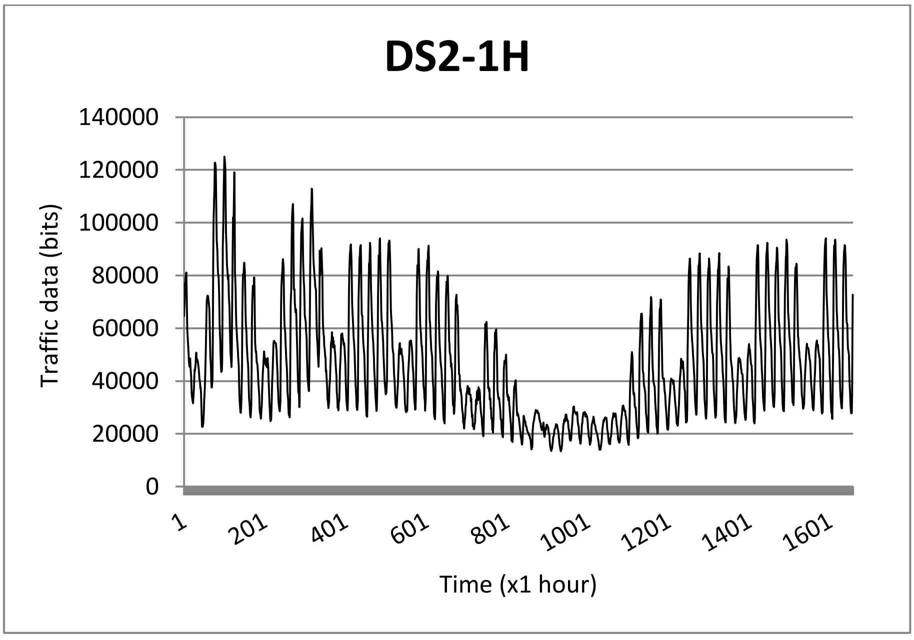

Figure 1.4. The Internet traffic time series for DS2-1H

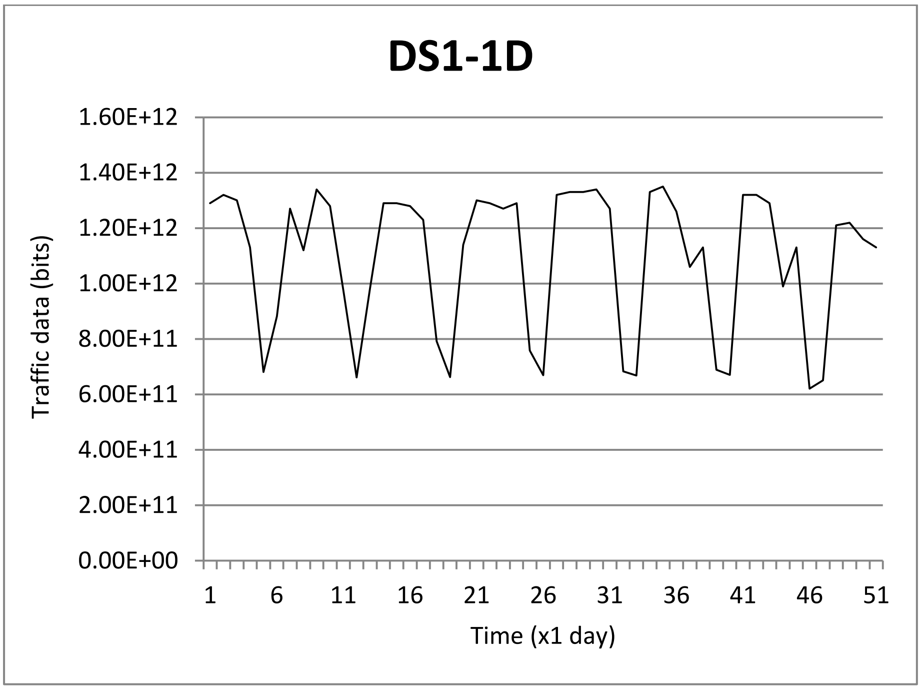

Figure 1.5. The Internet traffic time series for DS1-1D

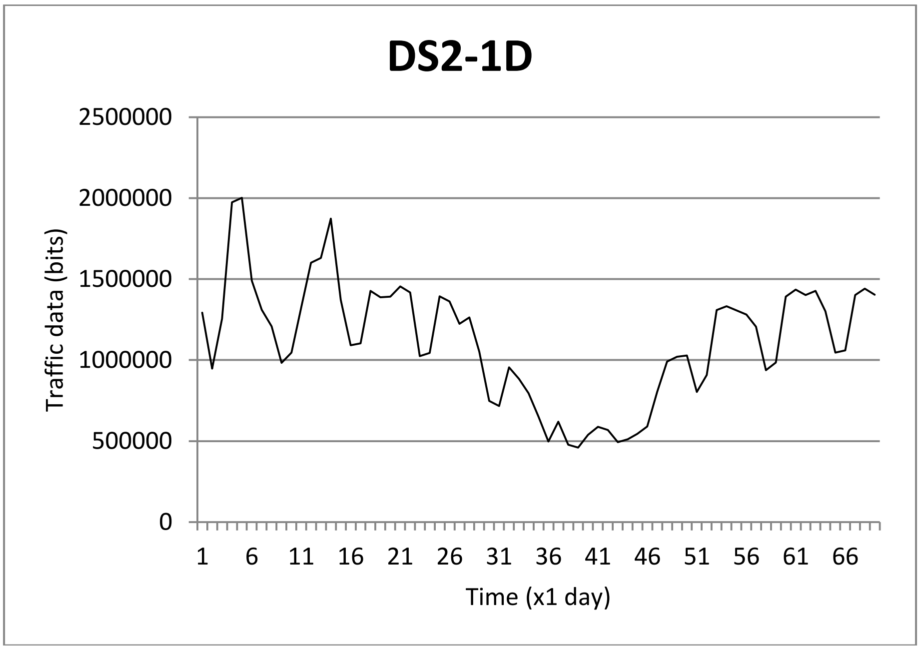

Figure 1.6. The Internet traffic time series for DS2-1D

2. OVERVIEW OF METHODS

2.1. Support Vector Machines

SVM is related to statistical learning theory (Vapnik, 1999), which was introduced in 1992 (Boser, 1992).

A. Linear SVM

Assume a training set of N data points, $S = \{x_k, y_k\}$ where $x_k \in R^n$ is an input vector and $y_k \in R$ is an output vector. The SVM problems are related with hyperplanes which separate the data. The equation of the hyperplanes are given using a vector v and a bias c. Decision function for the hyperplanes is $v^t x + c = 0$. The margin of separation ρ can be made maximum by constructing the optimal hyperplane. The support vector method technique aims at building a classifier

$$f(x) = sign\left(v^t.x + c\right). \tag{2.1.}$$

The v and c parameters are limited with

$$\min_i \left| v.x_i + c \right| \geq 1. \tag{2.2.}$$

After the vectors are divided without any problem, and as a result of this dividing process, if the space between the nearest vector and the hyperplane is maximum, it is called to be divided by a hyperplane. Consequently, a dividing hyperplane in standard form has to fulfill the constraints given in (2.3),

$$y_i\left(v.x_i + c\right) \geq 1, \;\; i = 1, 2, ..., n. \tag{2.3.}$$

A point x_i having the distance d from the hyperplane (v,c) is,

$$d\left((v,c),x_i\right)=\frac{y_i\left(x_i.v+c\right)}{\|v\|}\geq\frac{1}{\|v\|} \tag{2.4.}$$

ρ can be calculated as

$$\rho=\frac{2}{\|v\|}. \tag{2.5.}$$

Hence, SVM searches for a separating hyperplane by minimizing

$$\Phi(v)=\frac{1}{2}(v.v). \tag{2.6.}$$

$\Phi(v)$ in (2.6) can be minimized by performing the structural risk minimization principle,

$$\|v\|^2\leq c. \tag{2.7.}$$

h is the series of standard hyperplanes in space that has *n*-dimension and is limited by,

$$h\leq\min\left[\left(R^2c\right),d\right]+1, \tag{2.8.}$$

in which *R* is a hypersphere's radius surrounding all training vectors. As a result of this, minimizing (2.6) is equal to minimization of the upper bound.

The limitations of (2.3) can be reduced by presenting slack numbers $\zeta_i\geq 0, i=1,2,...,n$, therefore (2.3) can be rewritten as

$$y_i\left(vx_i+c\right)\ge 1-\zeta_i,\ i=1,2,\dots,n. \tag{2.9.}$$

Under these circumstances, the problem of optimization becomes

$$\Phi\left(v,\zeta\right)=\frac{1}{2}\left(v.v\right)+C\sum_{i=1}^{n}\zeta_i\ . \tag{2.10.}$$

In (2.10.) C is a user specified positive fixed constant. The saddle point of Lagrangian function is utilized in the solution of the problem given in (2.10).

$$L\left(v,c,\alpha,\zeta,\gamma\right)=\frac{1}{2}\left(v.v\right)C\sum_{i=1}^{n}\zeta_i-\sum_{i=1}^{n}\alpha_i\left|y_i\left(v.x_i+c\right)-1+\zeta_i\right|-\sum_{i=1}^{n}\gamma_i\zeta_i\ . \tag{2.11.}$$

In (2.11), $\alpha_i\ge 0,\zeta_i\ge 0,i=1,2,\dots,n$ are Lagrange multipliers. (2.11) must be solved in terms of v, c, and ζ_i. Classical Lagrangian duality empowers the first issue, turning (2.11) into a dual problem of it and this makes the solution easier. (2.12) shows the dual problem to be solved

$$\max_{\alpha}\left[\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j\left(x_i x_j\right)\right] \tag{2.12.}$$

with constraints

$$\sum_{i=1}^{n}\alpha_i y_i=0,\ 0\le\alpha_i\le B,\ i=1,2,\dots,n \tag{2.13.}$$

There is a unique solution for this classic quadratic optimization problem. In respect of the optimization theory which is called Kuhn-Tucker theorem (2.16.),

$$\alpha_i\left[y_i\left(v.x_i+c\right)-1\right]=0,\ i=1,2,\ldots,n \tag{2.14.}$$

If x_i satisfy (2.15), then (2.14) will have non-zero Lagrange multipliers

$$y_i\left(v.x_i+c\right)=1\,. \tag{2.15.}$$

The subjects in (2.15.) are called support vectors (SV's). The hyperplane is determined with a small subset of the training vectors of the SV's. Therefore, if the optimal solution α_i^* does not take a value of zero, the classifier function is states by

$$f\left(x\right)=\operatorname{sgn}\left\{\sum_{i=1}^{n}\alpha_i^* y_i\left(x_i.x\right)+c^*\right\}\ . \tag{2.16.}$$

In (2.16.) c^* is the answer of (14) for any $\alpha_i^*\neq 0$.

B. Non-linear SVM

The majority of the data sets cannot be decently divided by a linear separating hyperplane. However, they can be linearly divided if mapped into a higher dimensional field by utilizing a nonlinear mapping. Therefore, $z=\phi\left(x\right)$ that converts the input vector x having a dimension d into a vector z having a dimension d is defined and $\phi(\ )$ is selected so that $\left\{\phi\left(x_i,y_i\right)\right\}$ (new training data) is divisible with a hyperplane.

The data points from the input space into some space of higher dimension are mapped by using the function

$$\vartheta\left(.\right):R^n\rightarrow R^{nh}\,. \tag{2.17.}$$

Optimal function (2.9) transforms (2.18) using the same constraints,