Travel Time Prediction with Support Vector Regression

Chun-Hsin Wu, Chia-Chen Wei, Da-Chun Su, Ming-Hua Chang and Jan-Ming Ho

***Abstract*—Travel time prediction is essential to the development of advanced traveler information systems. In this paper, we apply support vector regression (SVR) for travel-time predictions and compare its results to other baseline travel-time prediction methods using real highway traffic data. Since support vector machines have greater generalization ability and guarantee global minima for given training data, it is believed that support vector regression will perform well for time series analysis. Compared to other baseline predictors, our results show that the SVR predictor can reduce significantly both relative mean errors and root mean squared errors of predicted travel times. We demonstrate the feasibility of applying SVR in travel-time prediction and prove that SVR is applicable and performs well for traffic data analysis.**

***Index Terms*—support vector machines, support vector regression, time series analysis, travel time prediction, intelligent transportation systems**

I. Introduction

Advanced Traveler Information Systems (ATIS) is a major application essential to intelligent transportation systems (ITS). As well, to various ITS applications, such as route guidance systems and ramp metering systems, accurate estimation of roadway-traffic conditions, especially travel times, are even more critical to the traffic flow management. With precise travel-time predictions, route guidance systems and ramp metering systems can assist travelers and traffic-control centers to better adjust traveler schedules and control traffic flow.

Travel-time calculation depends on vehicle speed, traffic flow and occupancy, which are highly sensitive to weather conditions and traffic incidents. These features make travel-time predictions very complex and difficult to reach optimal accuracy. Nonetheless, daily, weekly and seasonal patterns can still be observed at a large scale. For instance, daily patterns distinguish rush hour and late night traffic, weekly patterns distinguish weekday and weekend traffic, while seasonal patterns distinguish winter and summer traffic. The time-varying feature germane to traffic behavior is the key to travel-time modeling.

Since the creation of the SVM theory by V.Vapnik in 1995 at the AT&T Bell Laboratories [1], the application of SVM to time-series forecasting has shown many breakthroughs and plausible performance. Moreover, the rapid development of support vector machines (SVM) in statistical learning theory encourages researchers actively focus on applying SVM to various research fields like document classifications and pattern recognitions. On the other hand, applications of Support Vector Regression (SVR) [2], such as forecasting of financial market [3], estimation of power consumption [4], reconstruction of chaotic systems [5] and prediction of highway traffic flow [6], are also under development. The time-varying properties of SVR applications resemble the time-dependency of traffic forecasting, combined with many successful results of SVR predictions encourage our research in using SVR for travel-time modeling.

SVM possess great potential and superior performance as is appeared in many previous researches [4][7]. This is largely due to the structural risk minimization (SRM) principle in SVM that has greater generalization ability and is superior to the empirical risk minimization (ERM) principle as adopted in neural networks [8]. In SVM, the results guarantee global minima whereas ERM can only locate local minima. For example, the training process in neural networks, the results give out any number of local minima that are not promised to include global minima. Furthermore, SVM is adaptive to complex systems and robust in dealing with corrupted data. This feature offers SVM a greater generalization ability that is the bottleneck of its predecessor, the neural network approach [2].

The main idea of the traffic forecasting is based on the fact that traffic behaviors possess both partially deterministic and partially chaotic properties. Forecasting results can be obtained by reconstructing the deterministic traffic motion and predicting the random behaviors caused by unanticipated factors. Suppose that currently it is time t. Given the historical data $f(t-1), f(t-2), \ldots,$ and $f(t-n)$ at time $t-1, t-2, \ldots, t-n$, we can predict the future value of $f(t+1), f(t+2), \ldots$ by analyzing historical data set. Hence, future values can be forecasted based on the correlation between the time-variant historical data set and its outcomes.

Numerous studies have focused on the accurate prediction of travel time of highways: time series analysis, Bayesian classification, Kalman filtering, ARIMA model, linear model, tree method, neural networks, and simulation models [8][9][10][11][12][13][14][15]. The simulation models, such as METANET, SIMRES, STM or Paramics, predict travel time using microscopic or macroscopic simulators. Most of the other models are data-driven models based on statistical analysis. A

All of the authors are with the Institute of Information Science, Academia Sinica, Taipei, Taiwan 115 (phone: +886-2-2788-3799; fax: +886-2-2782-4814; e-mail: {wuch, ccwei, starking, justin, hoho}@ iis.sinica.edu.tw)

travel-time prediction dataset can be characterized as: historical data, current data, or both historical and current data [16]. Typically, the input data for these methods are vehicle speed, travel time, traffic flow, density or occupancy.

The way input and output data are manipulated before and after a chosen prediction algorithm differentiate the indirect and direct travel-time prediction methods [17]. Indirect methods input travel-time dependent variable, namely, speed, from the outcome of the prediction algorithm the variable is then converted to travel time. On the other hand, direct methods input travel time by preprocessing raw travel-time dependent variables.

In this paper, we use support vector regression to predict the travel time of the highway and show that SVR is applicable to travel-time prediction and outperforms many previous methods. In Section II we introduce support vector regression briefly. In Section III we explain our experimental procedure. Then we present the methods and results of different travel time predictors in Section IV and Section V, respectively. Section VI concludes the paper.

II. Support Vector Regression

Considering a set of training data $\{(x_1, y_1), \ldots, (x_l, y_l)\}$, where each $x_i \subset R^n$ denotes the input space of the sample and has a corresponding target value $y_i \subset R$ for $i=1,\ldots, l$ where l corresponds to the size of the training data [17][18]. The idea of the regression problem is to determine a function that can approximate future values accurately.

The generic SVR estimating function takes the form:

$$f(x) = (w \cdot \Phi(x)) + b \tag{1}$$

where $w \subset R^n, b \subset R$ and Φ denotes a non-linear transformation from R^n to high dimensional space. Our goal is to find the value of w and b such that values of x can be determined by minimizing the regression risk:

$$R_{reg}(f) = C\sum_{i=0}^{\ell}\Gamma(f(x_i) - y_i) + \frac{1}{2}\|w\|^2 \tag{2}$$

where $\Gamma(\cdot)$ is a cost function, C is a constant and vector w can be written in terms of data points as:

$$w = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)\Phi(x_i) \tag{3}$$

By substituting equation (3) into equation (1), the generic equation can be rewritten as:

$$\begin{aligned} f(x) &= \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)(\Phi(x_i)\cdot\Phi(x)) + b \\ &= \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)k(x_i, x) + b \end{aligned} \tag{4}$$

In equation (4) the dot product can be replaced with function $k(x_i, x)$, known as the kernel function. Kernel functions enable dot product to be performed in high-dimensional feature space using low dimensional space data input without knowing the transformation Φ. All kernel functions must satisfy Mercer's condition that corresponds to the inner product of some feature space. The radial basis function (RBF) is commonly used as the kernel for regression:

$$k(x_i, x) = \exp\left\{-\gamma|x - x_i|^2\right\} \tag{5}$$

Some common kernels are shown in Table 1. In our studies we have experimented with these three kernels.

Kernels	**Functions**
Linear	$x \cdot y$
Polynomial	$[(x * x_i) + 1]^d$
RBF	$\exp\left\{-\gamma\|x - x_i\|^2\right\}$

Table 1.Common kernel functions

The ε -insensitive loss function is the most widely used cost function [18]. The function is in the form:

$$\Gamma(f(x) - y) = \begin{cases} |f(x) - y| - \varepsilon, & for \quad |f(x) - y| \geq \varepsilon \\ 0 & otherwise \end{cases} \tag{6}$$

By solving the quadratic optimization problem in (7), the regression risk in equation (2) and the ε -insensitive loss function (6) can be minimized:

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

subject to

$$\sum_{i=1}^{\ell}\alpha_i - \alpha_i^* = 0, \quad \alpha_i, \alpha_i^* \in [0, C] \tag{7}$$

The Lagrange multipliers, α_i and α_i^*, represent solutions to the above quadratic problem that act as forces pushing predictions towards target value y_i. Only the non-zero values of the Lagrange multipliers in equation (7) are useful in forecasting the regression line and are known as support vectors. For all points inside the ε -tube, the Lagrange multipliers equal to zero do not contribute to the regression function. Only if the requirement $|f(x) - y| \geq \varepsilon$ (See Figure 1) is fulfilled, Lagrange multipliers may be non-zero values and used as support vectors.

The constant C introduced in equation (2) determines penalties to estimation errors. A large 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; this allows the minimization of margin with errors, thus higher generalization ability. If C goes to infinitely large, SVR would not allow the occurrence of any error and result in a complex model, whereas when C goes to zero, the result would tolerate a large amount of errors and the model would be less complex.

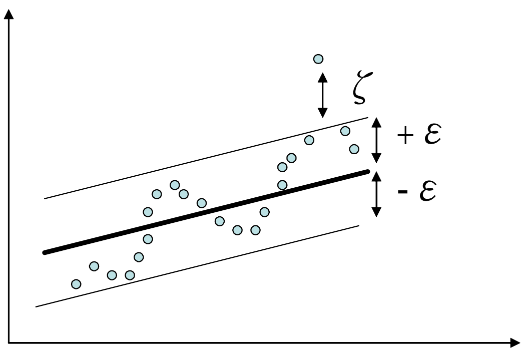

Figure 1. Support vector regression to fit a tube with radius ε to the data and positive slack variables ζ_i measuring the points lying outside of the tube.

Now, we have solved the value of w in terms of the Lagrange multipliers. For the variable b, it can be computed by applying Karush-Kuhn-Tucker (KKT) conditions which, in this case, implies that the product of the Lagrange multipliers and constrains has to equal zero:

$$\begin{aligned} &\alpha_i(\varepsilon+\zeta_i-y_i+(w,x_i)+b)=0 \\ &\alpha_i^*(\varepsilon+\zeta_i^*+y_i-(w,x_i)-b)=0 \end{aligned} \quad (8)$$

and

$$\begin{aligned} &(\mathrm{C}-\alpha_i)\zeta_i=0 \\ &(\mathrm{C}-\alpha_i^*)\zeta_i^*=0 \end{aligned} \quad (9)$$

where ζ_i and ζ_i^* are slack variables used to measure errors outside the ε -tube. Since $\alpha_i,\alpha_i^*=0$ and $\zeta_i^*=0$ for $\alpha_i^*\in(0,\mathrm{C})$, b can be computed as follows:

$$\begin{aligned} &b=y_i-(w,x_i)-\varepsilon \quad for\ \alpha_i\in(0,C) \\ &b=y_i-(w,x_i)+\varepsilon \quad for\ \alpha_i^*\in(0,C) \end{aligned} \quad (10)$$

Putting it all together, we can use SVM and SVR without knowing the transformation.

III. Experimental Procedure

A. Data Preparation

The traffic data is provided by the Intelligent Transportation Web Service Project (ITWS) [19][20] at Academia Sinica, a governmental research center based in Taipei, Taiwan. The Taiwan Area National Freeway Bureau (TANFB) constantly collects vehicle speed information from loop detectors that are deployed at 1 km intervals along the Sun Yet-Sen Highway. The TANFB Web site provides the raw traffic information source, which is updated once every 3 minutes. The loop detector data is employed to derive travel time indirectly: the travel time information is computed from the variable speed and the known distance between detectors.

Since traffic data may be missed or corrupted, we select a better portion of the dataset of the highway between February 15 and March 21, 2003. During this five-week period there are no special holidays and the data loss rate is not over some threshold value; which could bias our results if not properly managed. We use data from the first 28 days as the training set and use the last 7 days as our testing set. We examine the travel times over three different distances: from Taipei to Chungli, Taichung and Kaohsiung, which cover 45-km, 178-km and 350-km stretches, respectively. In addition, we examine the travel times of 45-km distance between 7am and 10am further since travel time of short distance in rush hour changes more dynamically. Figure 2 shows the travel-time distribution of the short distance on a daily and weekly basis, respectively. We can find the daily similarities and the instant dynamics from the daily and weekly patterns.

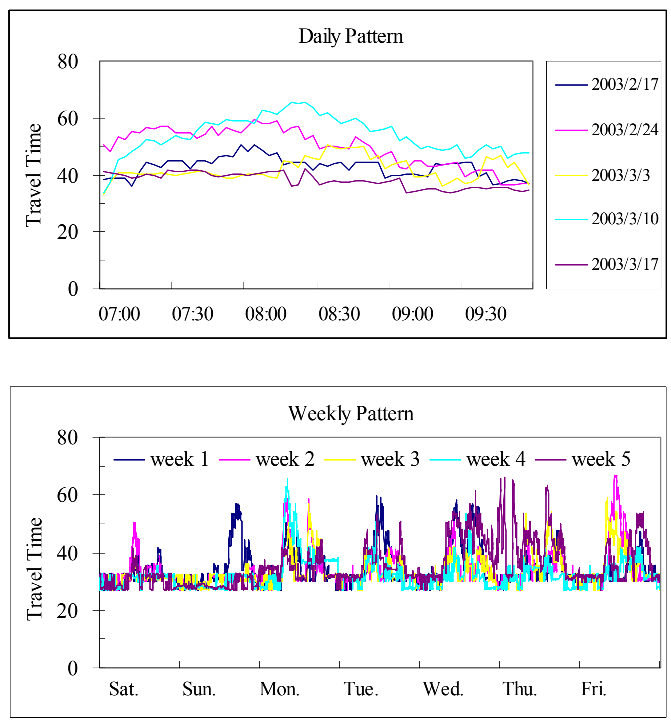

Figure 2. Daily and weekly travel-time distributions traveling from Taipei to Chungli, 45-km stretch, during 7am to 10am for five Wednesdays and five weeks between February 15 and March 21, 2003.

B. Prediction Methodology and Error Measurements

Suppose the current time is t, we want to predict $y(t+l)$ for the future time $t+l$ with the knowledge of the value $y(t-n)$, $y(t-n+1)$,…, $y(t)$ for past time $t-n$, $t-n+1$, …, t, respectively.

The prediction function is expressed as:

$$y(t+l) = f(t, l, y(t), y(t-1), \ldots, y(t-n))$$

We examine the travel times of different prediction methods for departing from 7am to 10am during the last week between March 15 and March 21, 2003. Relative Mean Errors (RME) and Root Mean Squared Errors (RMSE) are applied as performance indices.

$$RME = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - Y_i^*}{Y_i}\right|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - Y_i^*}{Y_i}\right|^2}$$

where Y_i is the observation value and Y_i^* is the predicted value.

IV. Travel Time Predicting Methods

To evaluate the applicability of travel-time prediction with support vector regression, some common baseline travel-time prediction methods are exploited for performance comparison.

A. Support Vector Regression Prediction Method

As discussed previously, there are many parameters that must be set for travel-time prediction with support vector regression. We have tried several combinations, and finally chose a linear function as the kernel for performance comparison with ε=0.01 and C=1000. In our experiences, however, Radial Basis Function (RBF) kernel also performed as well as linear kernel in many cases. The SVR experiments were done by running mySVM software kit with training window size equal to five [21].

B. Current Travel Time Prediction Method

This method computes travel time from the data available at the instant when prediction is performed [13]. The travel time is defined by:

$$T(t,\Delta) = \sum_{i=0}^{L-1}\frac{x_{i+1} - x_i}{v(x_i, t-\Delta)}$$

where Δ is the data delay, L is the number of sections, $(x_{i+1}$-$x)$ denotes the distance of a section of a highway, and $v(x_i,\ t$-$\Delta)$ is the speed at the start of the highway section.

C. Historical Mean Prediction Method

It is the travel time obtained from the average travel time of the historical traffic data at the same time of day and day of week:

$$\overline{T}(t) = \frac{1}{w}\sum_{i=1}^{w} T(i,t)$$

where w is the number of weeks trained and $T(i,\ t)$ is the past travel time at time t of historical week i.

V. Results

The experiment results of travel time prediction over short distance in rush hour are shown in Figure 3. As expected, the historical-mean predictor cannot reflect the traffic patterns that are quite different from the past average, and the current-time predictor is usually slow to reflect the changes of traffic patterns. Since SVR can converge rapidly and avoid local minimum, the SVR predictor performs very well in our experiments.

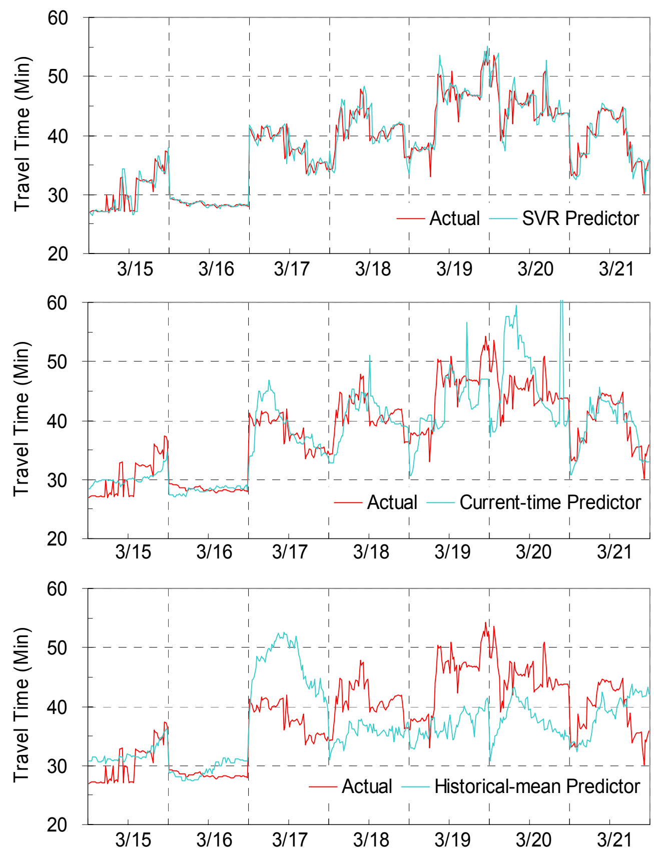

Figure 3. Comparisons of predicted travel times using different predicting methods.

The results in Table 2 show the relative mean errors (RME) and root mean squared errors (RMSE) of different predictors for different travel distances over all the data points of the testing set. They show that the SVR predictor reduces both RME and RMSE to less than half of those achieved by the current-time predictor and historical-mean predictor for all different distances.

In our experiments, as the traveling distance increases, the number of free sections increases more than the number of busy sections, such that the travel time of long distance is dominated by the time to travel free sections more than the time to travel busy sections. So it is not surprising that all the three predictors predict well for long distance, 350 km, but this makes it difficult to compare the performances of the three predictors. For this reason, we specifically examine the testing data points where the predicted error of any predictor is larger than or equal to 5%. As shown in Table 3, the SVR predictor improves not only the overall performance, but also significantly reduces the prediction errors for the cases where there are worse prediction errors in any one of the predictors.

RME	Current-time Predictor	Historical-mean Predictor	SVR Predictor
45 km	9.29%	12.52%	3.91%
161 km	3.88%	5.01%	1.71%
350 km	2.85%	2.56%	0.96%

RMSE	Current-time Predictor	Historical-mean Predictor	SVR Predictor
45 km	28.75%	16.20%	6.79%
161 km	9.98%	6.66%	2.57%
350 km	5.49%	3.42%	1.33%

Table 2. Prediction results in RME and RMSE of different predictors for traveling different distances (all testing data points).

RME	Current-time Predictor	Historical-mean Predictor	SVR Predictor
45 km	10.53%	14.31%	4.42%
161 km	5.85%	7.81%	2.38%
350 km	6.13%	4.90%	1.21%

RMSE	Current-time Predictor	Historical-mean Predictor	SVR Predictor
45 km	31.19%	17.55%	7.35%
161 km	13.81%	9.00%	3.26%
350 km	10.29%	5.66%	1.63%

Table 3. Prediction results for the testing data points that have greater prediction errors (>= 5%) in any one of the predictors.

VI. Conclusion

Support vector machine and support vector regression have demonstrated their success in time-series analysis and statistical learning. However, little work has been done for traffic data analysis. In this paper we examine the feasibility of applying support vector regression in travel-time prediction. After numerous experiments, we propose a set of SVR parameters that can predict travel times very well. The results show that the SVR predictor significantly outperforms the other baseline predictors. This evidences the applicability of support vector regression in traffic data analysis.

References

[1] K.-R Nuller, A. J. Smola, G. Ratsch, B. Scholkopf, J. kohlmorgen and V. Vapnik, "Predicting time series with support vector machine," in *Proc. ICANN 1997, Springer LNCS 1327*. pp. 999-1004.

[2] S. R. Gunn, "Support vector machine for classification and regression", Technical Report, U. of Shouthampton, May 1998.

[3] H. Yang, L. Chan and I. King, "Support vector machine regression for volatile stock market prediction," in *Proc. IDEAL 2002, Springer LNCS 24412*, pp. 391-396.

[4] B. J. Chen, M. W. Chang and C. J. Lin, "Load forecasting using support vector machines: a study on EUNITE competition 2001", report for *EUNITE Competition for Smart Adaptive System*. Available: http://www.eunite.org

[5] D. Matterra and S. Haykin, "Support vector machines for dynamic reconstruction of a chaotic system," in *Advances in Kernel Methods*, B. Schölkopf, C.J.C. Burges, and A.J. Smola Eds., pp. 211-241, MIT Press, 1999.

[6] A. Ding, X. Zhao and L Jiao, "Traffic flow time series prediction based on statistics learning theory," in *Proc. IEEE 5th International Conference on Intelligent Transportation Systems 2002*, pp. 727-730.

[7] D. C. Sansom, T. Downs and T. K. Saha, "Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian National Electricity Market Participants", in *Proc. Australasian Universities Power Engineering Conference*, 2002.

[8] J.W.C. van Lint, S.P. Hoogendoorn and H.J. van Zuylen, "Robust and adaptive travel time prediction with neural networks," proceedings of the 6th annual TRAIL Congress (part 2), December 2000.

[9] E. Fraschini and K. Ashausen, "Day on day dependencies in travel time: first result using ARIMA modeling," ETH, IVT Institut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau, February 2001.

[10] X. Zhang, J. Rice and P. Bickel, "Empirical comparison of travel time estimation methods," *Report for MOU 353, UCB-ITS-PRR-99-43*, December 1999.

[11] J. Rice and E. van Zwet, "A simple and effective method for predicting travel times on freeways," in *Proc. IEEE 4th International Conference on Transportation Systems 2001*, pp. 227 -232.

[12] Sheng Li, "Nonlinear combination of travel-time prediction model based on wavelet network," in *Proc. IEEE 5th International Conference on Intelligent Transportation Systems 2002*, pp. 741 -746.

[13] D. Park, and L. R. Ritett, "Forecasting multiple-period freeway link travel times using modular neural networks," *presented at the 77th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 1998.

[14] H. Sun, H. Liu and B. Ran, "Short term traffic forecasting using the local linear regression model," *presented at the 82nd Annual Meeting of the Transportation Research Board*, 2003

[15] X. Zhang and J. A. Rice, "Short-term travel time prediction using A time-varying coefficient linear model", Technical Report, Dept. Statistics, U.C. Berkeley, March 2001.

[16] R. Chrobok, O. Kaumann, J. Wahle and M. Schreckenberg, "Three categories of traffic data: historical, current, and predictive," in *Proc. 9th IFAC Syposium Control in Transportation Sytems 2000*, pp250-25.

[17] V.N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.

[18] K.R. Muller, A. Smola, G. Ratch, B. Scholkopf, J. Kohlmorgen and V.N. Vapnik, "Using support vector support machines for time series prediction," Image Processing Services Research Lab, AT&T Labs.

[19] C.H. Wu, D.C. Su, J. Chang, C.C. Wei, J.M. Ho, K.J. Lin and D.T. Lee, "An advanced traveler information system with emerging network technologies", in *Proc. 6th Asia-Pacific Conference on Intelligent Transportation Systems*, 2003.

[20] C.H. Wu, D.C. Su, J. Chang, C.C. Wei, K.J. Lin and J.M. Ho, "The design and implementation of intelligent transportation web services," in Proc. *IEEE Conference on E-commerce*, 2003.

[21] S. Ruping, "mySVM software," Available: http://www-ai.cs.uni-dortmund.de/SOFTWARE/MYSVM/