Using Scaling Methods to Improve Support Vector Regression's Performance for Travel Time and Volume Predictions

Citation for the original published paper (version of record):

Lin, A., Zhang, M., Selpi, S. (2018). Using Scaling Methods to Improve Support Vector Regression's Performance for Travel Time and
Volume Predictions. Time Series Analysis and Forecasting, Contributions to Statistics: 115-127.
http://dx.doi.org/10.1007/978-3-319-96944-2_8

Using Scaling Methods to Improve Support Vector Regression's Performance for Travel Time and Traffic Volume Predictions

Amanda Yan Lin[1,2], Mengcheng Zhang[1,2], and Selpi[1]

Department of Mechanics and Maritime Sciences[1] and
Department of Computer Science and Engineering[2],
Chalmers University of Technology, SE-412 96 Göteborg, Sweden
l.amanda0821@gmail.com
mengchengz@outlook.com
selpi@chalmers.se

Abstract. Long queues often happen on toll roads, especially at the tollgates. These create many problems, including having an impact on the regular roads nearby. If travel time and traffic volume at the tollgates can be predicted accurately in advance, this would allow traffic authorities to take appropriate measures to improve traffic flow and the safety of road users. This paper describes a novel combination of scaling methods with Support Vector Machines for Regression (SVR) for travel time and tollgate volume prediction tasks, as part of the Knowledge Discovery and Data Mining (KDD) Cup 2017. A new method is introduced to handle missing data by utilising the structure of the road network. Moreover, experiments with reduced data were conducted to evaluate whether conclusions from combining scaling methods with SVR could be generalised.

Keywords: Travel time prediction; traffic volume prediction; tollgate; SVR; time series analysis; SVR with scaling; support vector regression

1 Introduction

Traffic jams are common scenes in most roads, including toll roads or controlled-access roads. The tollgates, in particular, are well known as bottleneck, especially during rush hours and holidays. Reliable methods to predict future traffic flow and demands are important for traffic management authorities and road users. With precise predictions, the traffic regulators can decide how to deal with the problems (e.g., to open more tollgates or divert traffic at upstream intersections) and road users can plan their routes better. This paper is an extended version of [1], where we combined Support Vector Machine for Regression (SVR) with scaling methods for predicting travel time and traffic volume (as part of a competition in Knowledge Discovery and Data Mining (KDD) Cup 2017 [2]) for a given road and tollgate during rush hours, knowing the previous two-hour data and some days before. While the base of the work here is the same as in [1], a new piece of work (i.e., testing the generalisation of our methods) is added here. Furthermore, we elaborate on our own method to fill in the missing data by utilising the road network topology. For completeness purpose, we include the relevant parts of the base work here.

Travel time is the time taken from a designated start point to a designated end point. Traffic volume is a record of the number of vehicles at a designated point. Travel time and volume calculations depend on many stochastic factors, such as weather condition, holidays, time of day, and season, making the tasks of predicting travel time and traffic volume are still challenging todate.

SVR is a version of Support Vector Machine (SVM) for regression that was proposed in 1996 by Vladimir N. Vapnik, Harris Drucker, Christopher J. C. Burges, Linda Kaufman and Alexander J. Smola [3]. SVR is chosen here due to past researches that have shown good performances using SVR in different areas, including financial time series forecasting [4], stock market price forecasting [5], real-time flood stage forecasting [6], and also travel time prediction [7].

The rest of this paper is arranged as follows. Section 2 describes the data and the objectives of this work. Section 3 describes the related work. In Section 4, the methods used are introduced. We describe and discuss the results of our experiments for travel time prediction and traffic volume prediction in Sections 5 and 6, respectively. We test if the conclusions from Sections 5 and 6 could be generalised in Section 7. The final conclusions are presented in Section 8.

2 Data and Work Objectives

The data used here is from the KDD Cup 2017. It consists of four types, i.e., road network topology, time-stamped records of actual vehicles driving from intersections to tollgates (called vehicle trajectories data), traffic volume at tollgates, and weather data. The road network is represented as a sequence of road links and implemented as a directed graph (Fig. 1). The network includes three intersections (A, B, C) and three tollgates (1, 2, 3). These make up ten routes. Only data from vehicles using Amap navigation software was included in the vehicle trajectories data [2]. Therefore, there was quite a lot of missing data in the provided data set.

The objectives of this work are to address the following tasks:

- Task 1 - Travel time prediction: Given training data described above for the period of 19th July to 24th October, predict the average travel time for each route during rush hours (08:00-10:00 and 17:00-19:00), per 20-minute interval, for the period of 25th October to 31st October.
- Task 2 - Traffic volume prediction: Given training data described above for the period of 19th September to 24th October, estimate the volume for each of the five tollgate-direction pairs (Tollgate 1-entry, Tollgate 1-exit, Tollgate 2-entry, Tollgate 3-entry, and Tollgate 3-exit) during rush hours, per 20-minute interval, for the period of 25th October to 31st October.
- Task 3 - Test the generalisation of the methods used, i.e., testing if the conclusions from Tasks 1 and 2 still hold if we use different data as input.

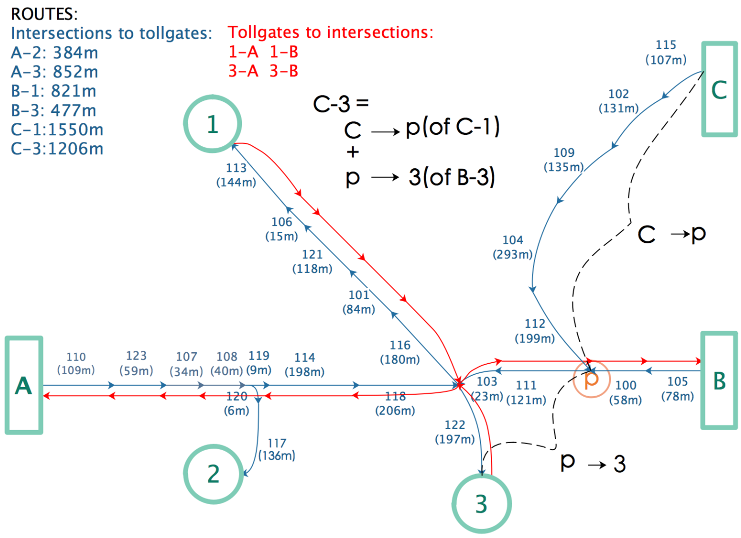

Fig. 1. The link-representation of road network. Each route is composed by a sequence of links, each link is represented by an arrow. The value without parentheses over a link represents the unique id of the link and the value in parentheses represents the length of the link.

3 Related Work

Traffic flow prediction, a well-known problem in traffic network, has been studied by many researchers. Both statistical (data-driven) and analytical approach (model-based) had been tried for such predictions (see a recent review in [8]). The statistical approach uses time series data consisting variables such as travel times, speeds, and volumes as input and predict the near future travel time based on historical traffic patterns. This approach assumes that the current or near future travel time will have similar pattern as historical travel time. While the analytical approach deduces the travel time from traffic conditions. The traffic conditions in turn are predicted from traffic propagation on the network by using traffic simulators. The statistical approach is suitable to be used when there are good amount of historical data while the analytical approach can be applied to the situation with changes in input factors, for example, adding additional networks [8]. Compared with analytical approach, an obvious advantage of statistical approach is that there are lots of ready-to-use software packages, the approach does not need much expertise about traffic flow modelling [9].

SVR belongs to the statistical approach and is a data-driven method. An application of SVR for highway travel time prediction has been studied by Wu et al. in [7]. There exist two main differences between data-sets in Wu et al.'s paper and in our project, one is that they collected the data from different highways while our data were collected between different intersections and tollgates, the other is in our data we have special holidays and lots of missing data, but they avoided special holidays and set the data loss rate within some threshold values. In addition, we use feature scaling as a data pre-processing step which was not included in Wu et al.'s work.

4 Methods

Experiments using SVR with and without scaling methods were conducted. The scaling methods investigated include Standard-scaling, Min-Max-scaling, and Robust-scaling (Section 4.2). A combination of our own method (called Complementary) and linear interpolation was used to fill in the missing data. The use of different combinations of features was tested. Cross-validation was used to measure the predictive performance of each model built using different scaling method and feature set. Generalisation of the methods were tested.

4.1 Support Vector Regression

The Support Vector Regression (SVR) uses the same principles as the support vector machine for classification (SVC). The goal of SVR is to find a function, with at most ϵ deviation from the actual target y. The problem can be written as a convex optimization problem

$$\text{minimize } \frac{1}{2}\left\|w\right\|^2 \quad \text{subject to } \begin{aligned} y_i - \langle w, x_i \rangle - b &\leq \epsilon \\ \langle w, x \rangle + b - y_i &\leq \epsilon \end{aligned}$$

If the problem is not feasible, slack variables ξ_i, ξ_i^* are introduced. The formulation becomes

$$\text{minimise } \frac{1}{2}\left\|w\right\|^2 + C\sum_{i=1}(\xi_i + \xi_i^*) \quad \text{subject to } \begin{aligned} y_i - \langle w, x_i \rangle - b &\leq \epsilon + \xi_i \\ \langle w, x \rangle + b - y_i &\leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned}$$

where the constant $C > 0$ is penalty parameter. More about SVR can be found in [10][11]. In this project, we used the SVR implementation from Scikit-learn library in Python [12].

4.2 Scaling Methods

Scaling is a way to systematically alter all the values in a data set. The simplest method, Min-Max-scaling, is rescaling the data to a fixed range, usually $[0, 1]$ or $[-1, 1]$. For a given data set X, a Min-Max-scaling is typically done via the following equation:

$$lb + \frac{X - min(X)}{max(X) - min(X)}(ub - lb),$$

where lb is a lower bound of the range, ub is an upper bound [13].

One common and widely used scaling method is Standard-scaling. The idea of Standard-scaling is to make the values of each feature in the data have zero-mean and unit-variance, according to

$$\frac{X - mean(X)}{standard\ deviation(X)}.$$

Another scaling method is Robust-scaling, which is based on the median and the interquartile range. If the data set X contains many outliers, Robust-scaling often gives better results [14]. Robust-scaling is defined as

$$\frac{X - median(X)}{IQR},$$

where IQR is interquartile range [14].

4.3 Error Measurements and Validation Method

Mean Absolute Percentage Error (MAPE) has been chosen by KDD cup team to evaluate the predictions. The MAPE is defined

$$MAPE = \frac{1}{R}\sum_{r=1}^{R}(\frac{1}{T}\sum_{t=1}^{T}|\frac{d_{rt} - p_{rt}}{d_{rt}}|). \quad (1)$$

For task 1 (travel time prediction), d_{rt} and p_{rt} are the actual and predicted average travel time for route r during time window t. For Task 2 (volume prediction), R is the number of tollgate-direction pairs (1-entry, 1-exit, 2-entry, 3-entry and 3-exit), T is the number of time windows in the testing period, and d_{rt} and p_{rt} are the actual and predicted traffic volume for a specific tollgate-direction pair r during time window t.

Cross validation was used to assess the predictive performance of our models.

5 Travel Time Prediction

To build a good model for Task 1, we addressed the sub-task to estimate the average travel time, per 20 minutes interval, from designated intersections to tollgates, for the hours 08:00-10:00 and 17:00-19:00 during 18th to 24th October, with training data from 19th July to 17th October. In order to test our models, the previous two-hours data of the period to be predicted were used as test data.

Quite a lot of data points were missing in the data set. Before running experiments, the missing data were filled in by applying our own "Complementary" method (Fig. 1) and then linear interpolation. "Complementary" is a method to fill in missing data in a route with the relevant part of other route(s) data. For instance, if there is missing data for a specific time window in route C-3, we gather part of that specific time window data from C-1 to get the data from

Intersection C to point p (C $\rightarrow$ p) and part of data in route B-3 to get the data from point p to Tollgate 3 (p $\rightarrow$ 3) to fill in the missing part in C-3 (see Fig. 1). "Complementary" was applied to the data in routes B-1, B-3 and C-1 in the same way. Since there are not that many missing data in routes A-2 and A-3, the missing parts were only filled in by linear interpolation.

We assumed that, in the morning and afternoon, the travel time of every given route are independent of each other, as a result, we applied the same prediction procedure on each route in the morning and afternoon separately. After iterative trial and error experiments with different parameter values chosen randomly, radial basis function (RBF) was selected as our kernel function, with parameters $\gamma = 0.005$ and $\epsilon = 0.5$. Furthermore, parameter C was chosen based on

$$max(|\bar{y} + 3\sigma_y|, |\bar{y} - 3\sigma_y|) \tag{2}$$

where $\bar{y}$ and σ_y are the mean and standard deviation of the y values from training data [15]. It has been found that SVR with RBF is less sensitive to data preprocessing methods, such as scaling [13].

Several cross-validation experiments, for example, testing different scaling methods, different number of training data, and different feature sets, were conducted. Time window position and the previous two-hour travel time were treated as two basic features during the process. *Time window position*: since the prediction is for the rush hours (the definitions of rush hours are 08:00-10:00 and 17:00-19:00), every 20 minutes interval, the rush hours are divided into six 20-minute time window. For instance, for the rush hour in the afternoon, 17:00-17:20 is the first time window position, 17:20-17:40 is the second, and so on. *Previous two-hour travel time*: it means the two-hour travel time data before the rush hours. For example, the previous two-hour travel time for the rush hours in the afternoon are the data from 15:00 to 17:00. They are divided into six 20-minute time window as well.

Obviously, the travel time is a product of dynamic interplay of traffic demand and traffic supply[16]. High traffic flow denotes high traffic demand. The factors, including temporal effects, such as daily pattern, weekly pattern and holiday, have influences on the traffic demand[7]. The factors, for example, crashes, road works, weather and so on, have influences on the traffic supply. For this reason, we added some extra features into the prediction one by one. The predictive performance of every resulting model was evaluated by comparing the results from validation and prediction phases. We show extra features that can capture the traffic demand as follows. *Special days*: holidays, weekends or working days. *Tollgate volume*: it means the traffic volume at the tollgate in the target route. For instance, when predicting the travel time in route A-2, the tollgate volume means the volume at Tollgate 2 (shown in Fig. 1). *Adjacent tollgate volume*: it means the traffic volume at the tollgate adjacent to route to be predicted. When two routes are from the same intersection and then go to different tollgates, while one of them is the route to be predicted, another one will become the adjacent route. For instance, route A-2's adjacent tollgate volume is the volume at Tollgate 3.

Table 1. Average MAPE from cross-validation experiments with basic features, using two sets of training data (all means 19/7 to 17/10 and part means 19/9 to 17/10). Test set is from 18/10 to 24/10.

Scaling method	validation using all data	prediction using all data	validation using part of data	prediction using part of data
Robust-scaling	0.2302	0.1886	0.1901	0.2073
Standard-scaling	0.2296	0.1902	0.1888	0.2083
Min-Max-scaling, [0,1]	0.2276	0.1935	0.1811	0.1928
No scaling	0.2464	0.2081	0.1977	0.2001

The predictive performances of using SVR combined with different scaling methods are presented in Table 1. This table also shows the results of the experiments using two different amount of training data (”all” means training data from 19/7 to 17/10 and ”part” means training data from 19/9 to 17/10).

In Table 1, one can see that using fewer weeks data for training gives better validation results, but worse prediction results. This also means that our experiments did not show anything conclusive about the influence of season on the travel time prediction (note that the period 19/7 to 18/9 is a summer season). Similarly, our experiments (not shown here due to space) suggest that most of the weather-related features did not increase predictive performance of our models. If any, only temperature was worth adding.

The best experimental result from the travel time prediction task is achieved by applying Robust-scaling with the two basic features (the previous two-hour travel time and time window position). Table 1 also shows that using scaling method gives better predictive performance compared to no scaling. Robust-scaling seems to be particularly good for time series with more varying patterns (that include summer season), while Min-Max-scaling seems to be particularly good for time series with more similar patterns.

6 Traffic Volume Prediction

In order to build a good model for Task 2, we addressed this sub-task: estimate the average volume for every tollgate-direction pair, per 20 minutes interval, during rush hours (08:00-10:00 and 17:00-19:00) from 18th October to 24th October using training data from 19th September to 17th October.

We assumed that, in the morning and afternoon, the volume at a given tollgate direction pair are independent of each other, as a result, we applied the same prediction procedure on each tollgate direction pair in the morning and afternoon separately. MAPE defined in Eq. 1 was used to calculate the average error for every tollgate direction pair. We applied SVR for the volume prediction as well. After iterative trial and error experiments with different parameter values chosen randomly, radial basis function (RBF) was selected as our kernel function, with parameters $\gamma = 0.01$ and $\epsilon = 0.01$. Furthermore, parameter C was selected by Eq. 2.

Similar feature selection strategy, as in Section 5, is used for this task. Time window position and the previous two-hour volume were treated as two basic features during the process. Time window position is the same as in Section 5 and the previous two-hour volume means the two-hour volume data before the rush hours to be predicted.

The performances of combining different scaling methods with SVR are presented in Table 2. For the volume prediction, applying SVR combined with a scaling method gives a huge improvement to the result compared with only using SVR. And again, it appears that Robust-scaling is particularly good for time series with more varying patterns. Note that the period of 1st October to 7th October is a big holiday period in China and it is widely known that the traffic volume is unusual during that period.

Table 2. Average MAPE from cross-validation experiments with features: time window position and previous two-hour volume. Training data are from 19/9 to 17/10. Test data are from 18/10 to 24/10.

Scaling method	validation result	prediction of test data
Robust-scaling	0.2710	0.1472
Standard-scaling	0.2717	0.1502
Min-Max-scaling, [0,1]	0.3467	0.1526
No scaling	1.0374	0.3128

Traffic volume may depend on many factors, including time of day, day of week, holiday, weather, etc. For this reason, an additional feature called special days to capture the holidays and weekends effect was added. Other features, extracted from the provided volume data, including the number of vehicles with ETC and the number of vehicles have vehicle model n ($n \in [0, 7]$), were also tested in our experiments (not shown here due to space). The predictive performance increases when we add special days to the feature set. The best performance (with average MAPE 0.2691 in validation phase and 0.1436 in prediction phase) is from an experiment where special days is included in the feature set, suggesting the feature special days is a very important feature for traffic volume prediction.

7 Generalisation

Based on the experimental results from the previous sections, we conclude that: (i) SVR with a scaling method performs better compared to without scaling; (ii) robust-scaling is especially good for time series with varying patterns; and (iii) Min-max-scaling is especially good for time series with similar patterns. These conclusions could depend on the provided input data. Here, we want to test if the conclusions could be generalised if different traffic data set is used. Due to the lack of other traffic data set, we analyse the following question instead:

Do these conclusions still hold if some of the data had been missing?

If some of the data had been missing, we start from slightly different data. The basic idea to address this question is: randomly delete some values from the original data (pretend those values were missing) and repeat the same experiment with reduced input data. The procedure can be summarised as follows:

1. delete $p\%$ of the original data randomly.
2. fill in the originally missing data and the deleted data using Complementary and linear interpolation.
3. take the data after Step 2, for each of the three scaling methods, run the experiment with a fixed feature set and a fixed SVR-setting. For simplicity reason, we use the basic feature set (time window position and the previous two-hour travel time), RBF-kernel, and SVR parameters $\epsilon = 0.5, \gamma = 0.005$. The output from this step is a table similar to Table 1, but with changed values.
4. repeat Step 1 to Step 3 100 times.

Task 1 was investigated with five levels of deletion $(10\%, 20\%, 30\%, 40\%, 50\%)$. For brevity, only the results from 10% and 50% are reported here (Fig. 2).

The results from Task 1 (Fig. 2) show that the performance of no scaling (black) is noticeably worst compared to that of the other methods for both validation (solid lines) and prediction (dashed lines). For validation, the performances of Robust-scaling (solid red) and Standard-scaling (solid blue) are very similar, while Min-max-scaling (solid green) is slightly better than the other scaling methods. For prediction, the performances of Robust-scaling (dashed red) and Standard-scaling (dashed blue) deteriorate more than Min-max-scaling (dashed green) as more data are deleted. A possible explanation of this is that as more data are deleted, the more outliers disappear and are replaced with smoother values (since we use complementary and linear interpolation to fill in the deleted data). This suggests that Min-max-scaling is especially good for time series with similar patterns.

A modified generalisation procedure was applied for Task 2, with the same five levels of deletion. The modifications include: In Step 2, only linear interpolation is used to fill in the originally missing data and the deleted data; In Step 3, use time window position and the previous two-hour volume as feature set and set SVR parameters $\epsilon = 0.01, \gamma = 0.01$.

The results from Task 2 (Fig. 3) show that in all cases the performance of no scaling (black) is worst compared to that of the other methods for both validation (solid lines) and prediction (dashed lines). The performances of Robust (red) and Standard-scaling (blue) are very similar in both validation and prediction. Unlike the results of generalisation from Task 1, the validation performance of Min-max-scaling (solid green) deteriorates more than the other scaling methods

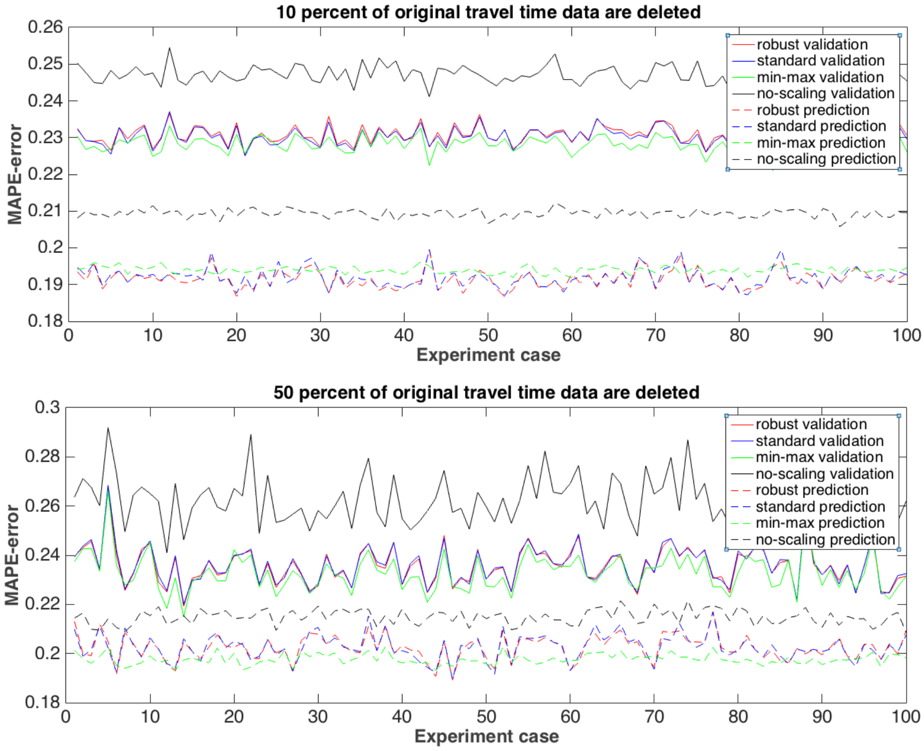

Fig. 2. Validation and prediction error for 100 experiments with 10% (top) and 50% (bottom) deleted data for the generalisation of Task 1 (travel time prediction).

as more data are deleted. For prediction, the performances of all three scaling methods are very similar. A possible explanation for this is that the original pattern (varying pattern) of volume data are mostly preserved after deletion and filling in process.

Based on the results of this section, the conclusions (i) and (iii) still hold. Regarding conclusion (ii), it seems that, for the data with varying pattern, the Robust and Standard-scaling perform very similar and slightly better than Min-max-scaling.

8 Conclusion

This paper demonstrated the application of SVR with scaling methods for travel time and tollgate volume predictions in rush hours. The impact of using three different scaling methods (Robust-scaling, Standard-scaling, and Min-Max-scaling) with SVR-predictor was investigated. Furthermore, experiments to test if the conclusions from Sections 5 and 6 still hold if reduced data is used as input were conducted. Our results suggested that SVR combined with a scaling method

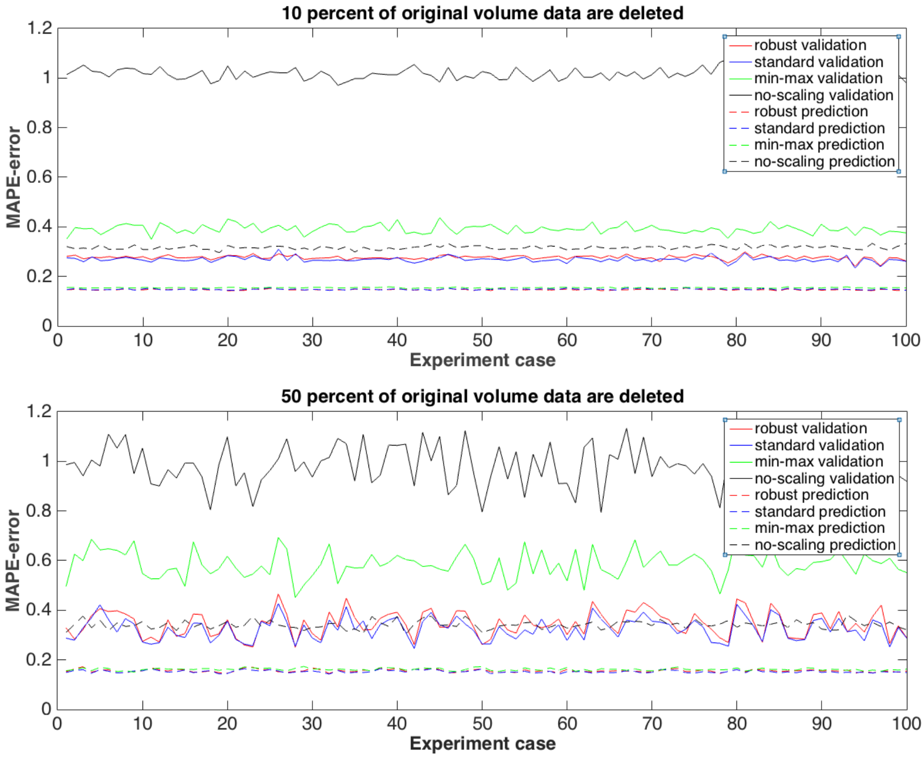

Fig. 3. Validation and prediction error for 100 experiments with 10% (top) and 50% (bottom) deleted data for the generalisation of Task 2 (volume prediction).

provides a more accurate prediction than without scaling, especially for volume prediction task. Min-Max-scaling was found to be particularly good for time series with more similar patterns. The performances of Robust-scaling and Standard-scaling were found to be pretty similar, and they seemed to perform slightly better than Min-max-scaling for time series with varying patterns.

Features that capture different travel time/volume influencing factors were analysed in the experiments. Although adding the features tollgate volume and adjacent tollgate volume has been found to increase the performance in some of our experiments for travel time prediction, but it is not always the case. The feature special days was found to be useful for volume prediction. Weather-related features were not found to be that useful in our experiments.

When our model was applied to Task 1, the mean absolute percentage error of the travel time prediction is around 0.19, which differs by only 0.02 from the best result obtained by other contestants (this is a competition task, the best prediction result was announced). Similarly, when our model was applied to Task 2, the mean absolute percentage error of the volume prediction is around 0.144, which differs by only 0.03 from the best result. We conclude that SVR combined with a scaling method can still provide a reasonable performance for travel time and traffic volume predictions, even when the training data contains many outliers (like holiday data) and no deep analysis of the data was applied.

Acknowledgments. S. acknowledges strategic funding support from Chalmers Area of Advance Transport while writing this paper.

References

1. Lin, A. Y., Zhang, M. C., & Selpi, .: Combining Support Vector Regression with Scaling Methods for Highway Tollgates Travel Time and Volume Predictions. In: Proceedings of Int. Work-Conference on Time Series Analysis, ITISE 2017, (2017)
2. KDD2017, `https://tianchi.aliyun.com/competition/information.htm?spm=5176.100067.5678.2.ru0ea4&raceId=231597`, accessed on 15 March 2017
3. Drucker, H., Burges, C. J., Kaufman, L., Smola, A., Vapnik, V.: Support vector regression machines. Advances in neural information processing systems, 9, 155-161 (1997)
4. Lu, C. J., Lee, T. S., Chiu, C. C.: Financial time series forecasting using independent component analysis and support vector regression. Decision Support Systems, 47(2), 115-125 (2009)
5. Yeh, C. Y., Huang, C. W., Lee, S. J.: A multiple-kernel support vector regression approach for stock market price forecasting. Expert Systems with Applications, 38(3), 2177-2186 (2011)
6. Yu, P. S., Chen, S. T., Chang, I. F.: Support vector regression for real-time flood stage forecasting. Journal of Hydrology, 328(3), 704-716 (2006)
7. Wu, C. H., Ho, J. M., Lee, D. T.: Travel-time prediction with support vector regression. IEEE transactions on intelligent transportation systems, 5(4), 276-281 (2004)
8. Oh, S., Byon, Y. J., Jang, K., & Yeo, H.: Short-term travel-time prediction on highway: a review of the data-driven approach. Transport Reviews, 35(1), 4-32 (2015)
9. Van Lint, J. W. C., Hoogendoorn, S. P., & van Zuylen, H. J.: Accurate freeway travel time prediction with state-space neural networks under missing data. Transportation Research Part C: Emerging Technologies, 13(5), 347-369 (2005)
10. Smola, A. J., Schlkopf, B.: A tutorial on support vector regression. Statistics and computing, 14(3), 199-222 (2004)
11. Basak, D., Pal, S., Patranabis, D. C.: Support vector regression. Neural Information Processing-Letters and Reviews, 11(10), 203-224, (2007)
12. SVR, `http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html`, accessed on 15 March 2017
13. Crone, S. F., Guajardo, J., & Weber, R.: The impact of preprocessing on support vector regression and neural networks in time series prediction. In: DMIN (pp. 37-44), (2006)
14. RobustScaler, `http://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html`, accessed on 15 March 2017
15. Cherkassky, V., Ma, Y.: Practical selection of SVM parameters and noise estimation for SVM regression. Neural networks, 17(1), 113-126 (2004)
16. van Lint., J. W. C.: Reliable travel time prediction for freeways: bridging artificial neural networks and traffic flow theory. TRAIL Research School, (2004)