Combining Support Vector Regression with Scaling Methods for Highway Tollgates Travel Time and 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
`{yanlin,mezhang}@student.chalmers.se`
`selpi@chalmers.se`

Abstract. Toll roads or controlled-access roads are very commonly used, e.g. in Asia. Drivers expect to drive smoother and faster on the toll roads compared to on regular roads. However, long queues on toll roads, particularly at the tollgates, often happen and create many problems. Being able to accurately predict travel time and volume of the tollgates would allow appropriate measures to improve traffic flow and safety to be taken. This paper describes a novel investigation on the use of scaling methods with Support Vector Regression (SVR) for highway tollgates travel time and volume prediction tasks as well as an investigation of the most important features for these tasks. Experiments were done as part of the Knowledge Discovery and Data Mining (KDD) Cup 2017. Suitability of certain scaling methods for different types of time series and reasoning why certain features are important for these tasks are also discussed.

Keywords: Traffic flow prediction; traffic volume prediction; highway tollgates; time series analysis; SVR with scaling; robust scaling; SVR;

1 Introduction

Transportation problems are increasing along with urbanisation and motorisation. Traffic jams are common scenes in most roads. Toll roads or controlled-access roads are no exception. Highway tollgates, in particular, have been well known as bottleneck in traffic networks, especially during rush hours and holidays. Reliable methods to predict future traffic flow are important for traffic management authorities as well as the road users. With precise predictions, the traffic regulators can decide how to deal with traffic jam or some other problems of highway tollgates (e.g.,to deploy additional toll collectors and/or divert traffic at upstream intersections). Such accurate predictions can also help road users to plan their journey.

In this paper, we address two prediction tasks, travel time prediction and traffic volume prediction, as part of a competition in Knowledge Discovery and Data Mining (KDD) Cup 2017 [7]. The tasks are to predict travel time and

volume for a given road and tollgate during rush hours, knowing the previous two-hour data and some days before. Support Vector Machine for Regression (SVR) with different scaling methods are applied for these prediction tasks.

Travel time is a measurement of time from a designated start point to a designated end point, which is the raw element for a number of performance measures in different transportation analyses [3]. Traffic volume is a record of the number of vehicles at a designated point. Both travel-time and volume calculations depend on many stochastic factors, such as weather conditions, holidays, time of the day, season, etc, making the tasks of predicting travel time and traffic volume become challenging.

SVR is a version of SVM for regression that was proposed in 1996 by Vladimir N. Vapnik, Harris Drucker, Christopher J. C. Burges, Linda Kaufman and Alexander J. Smola [10]. It is an application of SVM (Support Vector Machine) for time-series forecasting [3]. SVR has shown some good performances in different areas, such as financial time series forecasting [4], stock market price forecasting [5] and real-time flood stage forecasting [6]. It was applied for travel-time prediction and achieved good result as well [3]. Those successful results motivate us to use SVR for travel time and volume predictions. However, compared to the previous work, we introduce a novel approach by combining scaling methods with SVR. Furthermore, we also introduce a special approach to fill in the missing data for these tasks.

The rest of this paper is arranged as follows. Section 2 describes the raw data and the two prediction tasks. Section 3 introduces the methods used. We describe and discuss the results of our experiments for travel time prediction and traffic volume prediction in Section 4 and Section 5, respectively. The conclusions are presented in Section 6.

2   Data and tasks description

The data used here was provided by organizers of the KDD Cup 2017. Four different types of data set were provided: road network topology of the area (Fig. 1), vehicle trajectories, traffic volume at tollgates, and weather data for the area.

The road network is represented as a sequence of road links and implemented as a directed graph (Fig. 2). The network includes three intersections (A, B, C) and three tollgates (1, 2, 3). These make up six routes: routes from Intersection A to Tollgates 2 and 3, routes from Intersection B to Tollgates 1 and 3, and routes from Intersection C to Tollgates 1 and 3 (Fig. 1).

Vehicle trajectories data lists time-stamped records of actual vehicles driving from intersections to tollgates. Specifically the data about vehicle trajectories consists of intersection ID, tollgate ID, vehicle ID, date time when the vehicle enters the route, trajectory (sequence of link traces with each trace consists of a link ID, time entering the link, and total travel time (in seconds) passing the link), and total travel time (in seconds) from the intersection to the tollgate. Only data about vehicles using Amap navigation software was included in the
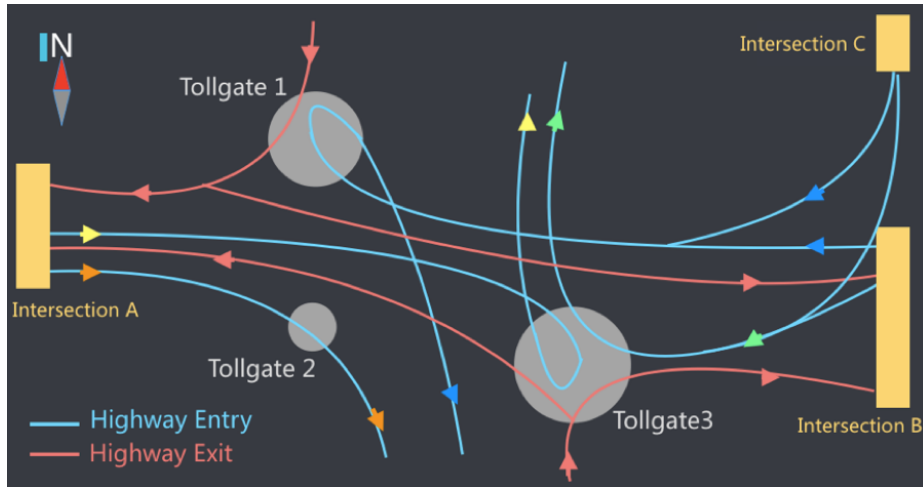
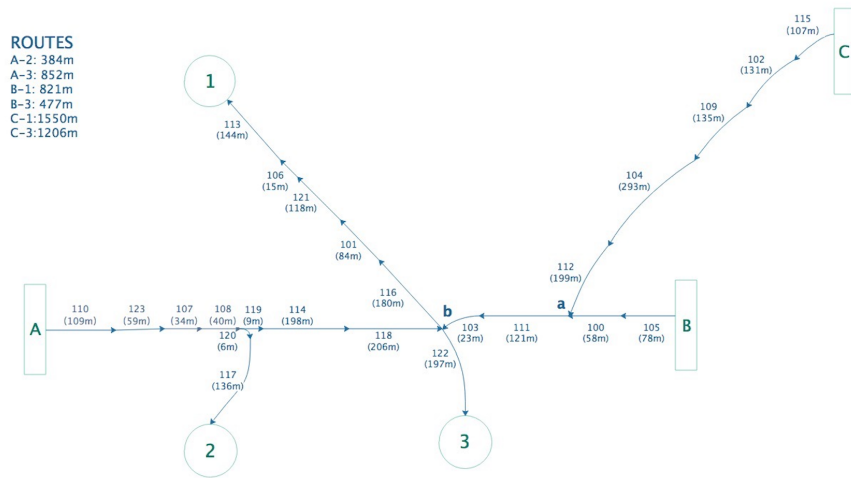**Fig. 1.** An overview of road network. This figure is taken from the description of KDD CUP 2017 [7].

Fig. 2. 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. The total length of each route is presented at the upper left corner.

vehicle trajectories data [7]. Therefore, there was quite a lot of missing data in the provided data set.

The data about traffic volume at tollgates consists of date time when a vehicle passes the tollgate, tollgate ID, direction (0 for entry, 1 for exit), vehicle model (integer 0 to 7 to indicate the capacity of the vehicle), boolean values indicating if the vehicle uses electronic toll collection (ETC) or not, and vehicle type (0 for passenger vehicle, 1 for cargo vehicle).

The weather data consists of weather related measurements collected every three hours in the target area. Specifically the data consists of date, hour, air pressure (in hundred Pa), sea level pressure (in hundred Pa), wind direction (in degrees), wind speed (in m/s), temperature (in Celsius degrees), relative humidity, and precipitation (in mm).

The objectives of this project are to address the following tasks as best we can:

− Task 1 : Travel time prediction. Given training data described above for the period of 19th July to 24th October, estimate the average travel time of vehicles 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 4-exit) during rush hours, per 20-minute interval, for the period of 25th October to 31st October.

3   Methods

Experiments using SVR with and without scaling methods were conducted. The scaling methods investigated include Standard-scaling, Mix-Max-scaling, and Robust-scaling (Section 3.2). The use of different combinations of features was also tested. Cross-validation was used to measure the predictive performance of each model built using different scaling method and feature set.

3.1   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

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

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

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

where the constant $C > 0$ is penalty parameter. More about SVR can be found in [2][13].

In this project, we used the SVR implementation from Scikit-learn library in Python [12].

3.2   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 [15].

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 [11]. Robust-scaling is defined as

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

where IQR is interquartile range [11].

3.3   Error measurements and validation method

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

For Task 1 (travel-time prediction), the MAPE is defined

$$MAPE_{travel-time} = \frac{1}{R}\sum_{r=1}^{R}(\frac{1}{T}\sum_{t=1}^{T}|\frac{d_{rt} - p_{rt}}{d_{rt}}|) \tag{1}$$

In the Eq. 1 above, 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), the MAPE is defined:

$$MAPE_{volume} = \frac{1}{M}\sum_{m=1}^{M}(\frac{1}{T}\sum_{t=1}^{T}|\frac{f_{mt} - p_{mt}}{f_{mt}}|). \tag{2}$$

In the Eq. 2, M 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 f_{mt} and p_{mt} are the actual and predicted traffic volume for a specific tollgate-direction pair m during time window t.

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

4   Travel time prediction

In order to build a model that can make good estimations for Task 1, we were given training data for the period of 19th July to 17th October, and was asked to estimate the average travel time, per 20 minutes interval, from designated intersections to tollgates during rush hours (08:00-10:00 and 17:00-19:00) for the period of 18th October to 24th October. For the purpose of testing our models, data from the previous two-hours of the rush hours period to be predicted were used as test-data.

Since there are many missing data, particularly in routes B-1, B-3, C-1 and C-3, the missing data were filled in before experiments were run. In these four routes, the missing data were filled in by applying "complementary" method and then linear interpolation. "Complementary" method means that if there is no data in a 20-minute time window of a route, this missing part will be filled in by the relevant part of other route(s) data. For instance, if the data for a specific time window is missing in route C-3, we gather part of data for that specific time window from route C-1 to get Intersection C to point p (C → p) and part of data from route B-3 to get point p to Tollgate 3 (p → 3) to fill the missing part in C-3 (see Fig. 2). Similar ways were done for the routes B-1, B-3 and C-1. For the routes A-2 and A-3, the missing data were only filled in by linear interpolation.

As we assumed the travel time of a given route in the morning and afternoon are independent of each other, the same prediction procedure was applied for every route in the morning and afternoon respectively. SVR was used as the main prediction method. After testing with several experiments (with different values chosen randomly), radial basis function (RBF) was chosen as the kernel function, with $\gamma = 0.005$ and $\epsilon = 0.5$. Parameter C was chosen according to

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

where $\bar{y}$ and σ_y are the mean and the standard deviation of the y values of training data [9]. SVR with RBF has been found less sensitive to preprocessing of data such as scaling [15].

Many cross-validation experiments were conducted: using different scaling methods, different amount of training data, and different features sets. Two basic features were always included: time window position and the previous two-hour travel time. *Time window position*: The prediction is for every 20-minute time window of the rush hours (rush hours are defined as 08:00-10:00 and 17:00-19:00), therefore the rush hours are split into six 20-minute time windows. For example, for the rush hours in the morning, 08:00-08:20 is the first position, 08:20-08:40 is the second, and so on. *Previous Two-hour travel time*: They are the two-hour travel time data before the rush hours. For instance, the previous two-hour travel time for the rush hours in the morning are the data from 06:00-08:00. They are also split into six 20-minute time windows.

Obviously, the travel-time are a result of dynamic interplay of traffic demand and traffic supply [14]. High traffic flow indicates high traffic demand. Factors influencing traffic demand include temporal effects like daily and weekly pattern, as well as holiday [3]. Factors influencing the traffic supply includes crashes, road works, weather, etc. For this reason, extra features were added one by one and the predictive performance of each resulting model was evaluated by comparing the validation and the prediction result. Additional features that can capture the traffic demand are as follows. *Special days*: working days, weekends, or holidays. *Tollgate volume*: this feature is the volume of the tollgate of the target route. For example, when predicting the travel time of route A-2, the tollgate volume is the volume at tollgate 2 (shown in Fig. 2). *Adjacent tollgate volume*: this feature is the volume of the predicting route's adjacent tollgate. If two routes come from the same intersection and go to different tollgates, one of the two is the predicting route, as a result, another is the adjacent route. The tollgate of the adjacent route is called adjacent tollgate. For example, for route A-2, the adjacent tollgate volume is the volume of tollgate 3.

Table 1. Average MAPE from 13-fold cross-validation experiments with features: time window position and two-hour travel time; Data used for training are from 19/7 to 17/10; Data used to test are from 18/10 to 24/10;

Scaling method	validation result	prediction of test data
Robust-scaling	0.2302	0.1886
Standard-scaling	0.2296	0.1902
Min-Max-scaling, [0,1]	0.2276	0.1935
No scaling	0.2464	0.2081

The predictive performances of using SVR combined with different scaling-methods are presented in Table 1 and Table 2. The results of the experiments using two different amount of training data sets are shown in Table 1 (training

Table 2. Average MAPE from 4-fold cross-validation experiments with features: time window position and two-hour travel time; Data used for training are from 19/9 to 17/10; Data used to test are from 18/10 to 24/10;

Scaling method	validation result	prediction of test data
Robust-scaling	0.1901	0.2073
Standard-scaling	0.1888	0.2083
Min-Max-scaling, [0,1]	0.1811	0.1928
No scaling	0.1977	0.2001

Table 3. Average MAPE from different cross-validation experiments with Min-Max-scaling in range [0,1] and features: time window position and two-hour travel time; Data used for training are from 19/9 to 17/10; Data used to test are from 18/10 to 24/10;

Extra Feature(s)	validation result	prediction of test data
None	0.1811	0.1928
Special days	0.1795	0.1920
Tollgate volume (vol)	0.1770	0.1931
Tollgate volume & special days	0.1773	0.1938
Tollgate vol. & adjacent tollgate vol.	0.1771	0.1900

data from 19/7 to 17/10) and in Table 2 (training data from 19/9 to 17/10). The results of the experiments using different sets of features are shown in Table 3.

Comparing Table 1 and Table 2, 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 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. Based on the experiments with the same amount of training data (data from 19th September to 17th October), adding more features (tollgate volume and adjacent tollgate volume) provides better validation and prediction results (Table 3).

The best experimental result from the travel-time prediction task appears in Table 1 by applying Robust-scaling with the two basic features (the previous two-hour travel time and time window position). From Table 1 and Table 2, 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.

5   Traffic volume prediction

Similarly to Task 1, in order to build a good model for Task 2, we addressed the following sub-task: given training data for the period of 19th September to 17th October, estimate the average volume for each of the tollgate-direction pairs, per 20 minutes interval, during rush hours (08:00-10:00 and 17:00-19:00) for the period of 18th October to 24th October.

As we assumed the volume of a given tollgate direction pair in the morning and in the afternoon are independent of each other, the same prediction procedure was applied for all tollgate direction pairs in the morning and afternoon respectively. The average error of all tollgate direction pairs was calculated using MAPE defined in Eq. 2. SVR was applied for the volume prediction too. After testing with several experiments (with different values chosen randomly), radial basis function (RBF) was chosen as the kernel function with $\gamma = 0.01$ and $\epsilon = 0.01$. Parameter C was chosen according to Eq. 3.

The feature selection strategy for volume prediction was similar as for travel time prediction (Section 4). The two basic features here were time window position and the previous two-hour volume. The previous two-hour volume means the two hours volume before the rush hours to be predicted and time window position is similar as in Section 4.

The results presented in Table 4 and 5 was obtained by The results of performances by using different scaling-methods combined with SVR are presented in Table 4. In addition, the comparisons of performances for different features are presented in Table 5.

Traffic volume depends on many factors, including time of day, day of week, holiday, weather, etc. For this reason, an additional feature called special days (explained in Section 4) to capture the holidays and weekends effect was added. Moreover, other features (basically 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 (see Table 5).

Table 4. Average MAPE from cross-validation experiments with features: time window position and two-hour volume; Data used for training are from 19/9 to 17/10; Data used to test 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

For the volume prediction, applying SVR combined with a scaling method gives a huge improvement to the result compared with only using SVR, see Table 4. And again, it appears that Robust-scaling is particularly good for time series

Table 5. Average MAPE from cross-validation experiments with Robust-scaling and features: time window position & two-hour volume; Data used for training are from 19/9 to 17/10; Data used to test are from 18/10 to 24/10;

Extra Feature	validation result	prediction of test data
None	0.2710	0.1472
special days	0.2647	0.1470
use ETC	0.3605	0.1705
vehicle model (veh. mod.) 1	0.2854	0.1472
veh. model 2	0.2759	0.1621
veh. model 3	0.3240	0.1531
veh. model 4	0.3138	0.1476
veh. model 5	0.3107	0.1504
veh. model 6	0.2708	0.1476
veh. model 7	0.2738	0.1447
veh. model 7 & special days	0.2682	0.1440
veh. mod. 6 & veh. mod. 7 & special days	0.2691	0.1436

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 very different during that period compared to usual days.

The best performance shows up in Table 5, with features: two-hour volume, time window position, vehicle model 6, vehicle model 7, and special days. Table 5 suggests that the feature special days is a very important feature for traffic volume prediction.

6   Conclusion

In this experiment, we demonstrated the application of SVR with scaling methods for travel-time prediction over a very short distance in rush hours and tollgate traffic volume prediction in rush hours. The performances of SVR-predictor combined with three scaling methods (Robust-scaling, Standard-scaling, and Min-Max-scaling) were compared. Our results suggested that SVR with a scaling method performs better compared to without scaling, Robust-scaling is particularly good for time series with varying patterns, and Min-Max-scaling is particularly good for time series with more similar patterns.

Features that capture different travel-time/volume influencing factors were analyzed in the experiments. In general, SVR combined with scaling provides a more accurate prediction than without scaling, especially for volume prediction. Adding extra features (travel-time/volume influencing factors) did not give significant improvement.

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, for the training data containing many outliers (like holiday data) and without deep analysis of the data (no data pruning), SVR combined with a scaling method can still provide reasonable prediction results.

References

1. Ding, A., Zhao, X., Jiao, L.: Traffic flow time series prediction based on statistics learning theory. In: Proceedings of the IEEE 5th international conference on intelligent transportation systems, 727-730 (2002)
2. Smola, A. J., Schlkopf, B.: A tutorial on support vector regression. Statistics and computing, 14(3), 199-222 (2004)
3. 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)
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. KDD2017, `https://tianchi.aliyun.com/competition/information.htm?spm=5176.100067.5678.2.ru0ea4&raceId=231597`, accessed on 15 March 2017
8. Müller, K. R., Smola, A. J., Rtsch, G., Schlkopf, B., Kohlmorgen, J., Vapnik, V.: Predicting time series with support vector machines. In: International Conference on Artificial Neural Networks (pp. 999-1004). Springer Berlin Heidelberg (1997, October)
9. Cherkassky, V., Ma, Y.: Practical selection of SVM parameters and noise estimation for SVM regression. Neural networks, 17(1), 113-126 (2004)
10. 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)
11. RobustScaler, `http://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html`, accessed on 15 March 2017
12. SVR, `http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html`, accessed on 15 March 2017
13. Basak, D., Pal, S., Patranabis, D. C.: Support vector regression. Neural Information Processing-Letters and Reviews, 11(10), 203-224, (2007)
14. van Lint., J. W. C.: Reliable travel time prediction for freeways: bridging artificial neural networks and traffic flow theory. TRAIL Research School, (2004)
15. 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)