Title
Forecasting Electric Vehicle Arrival & Departure Time On UCSD Campus using Support Vector Machines

Author
Xu, Zhuo

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Forecasting Electric Vehicle Arrival & Departure Time on UCSD Campus using Support Vector Machines

A thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Engineering Sciences (Applied Ocean Science)

by

Zhuo Xu

Committee in charge:

Professor Jan Kleissl, Chair
Professor Renkun Chen
Professor Lynn Russell

2017

The Thesis of Zhuo Xu is approved and acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2017

TABLE OF CONTENTS

SIGNATURE PAGE iii
TABLE OF CONTENTS iv
LIST OF FIGURES v
LIST OF TABLES vi
ACKNOWLEDGMENTS vii
ABSTRACT OF THE THESIS viii
1. INTRODUCTION 1
2. THEORY 2
2.1 The Forecasting Model 2
2.2 Methodology 5
2.3 Case Study 8
3. RESULTS AND DISCUSSION 12
4. CONCLUSIONS 16
5. FUTURE WORK 17
REFERENCES 18

LIST OF FIGURES

Figure 1: Operation Flow Chart of the SVM Model 4

Figure 2: Arrival & Departure Time of All EVs on UCSD Campus of 2012.. 6

Figure 3: Arrival and Departure Time of Designated EVs on UCSD Campus of 2012, 2013 and 2014 7

Figure 4: Observed & Predicted EV Arrival Time Comparison of 2014 on UCSD Campus (Averaged value of 5^{th}, 15^{th} and 45^{th} week) 13

Figure 5: Observed & Predicted EV Departure Time Comparison of 2014 on UCSD Campus (Averaged value of 5^{th}, 15^{th} and 45^{th} week) 13

LIST OF TABLES

Table 1: Sample of raw EV charging data on UCSD campus from 2012 5

Table 2: Input Parameters for the SVM Model (One Week) 10

Table 3: MAPE, RMSE & MBE Performance for Year 2012, 2013 & 2014 by Using Vector Support Machine method 14

Table 4: MAPE, RMSE & MBE Performance for Year 2012, 2013 & 2014 by Using Reference Forecast.. 14

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my academic advisor, Professor Jan Kleissl, for giving me this wonderful opportunity to conduct my research under his supervision. I really appreciate his patience, encouragement for this thesis. It is a great honor to be a part of his research team. This work would not have been finished without the assistance and support from him.

I would also like to thank Guang Wang, whose valuable advice always helped me come up with new ideas and directions for the research. I want to thank you for guiding me through all the difficult moments and helping me explore the right direction.

Lastly and most importantly I would like to thank my parents who have been and will always be my greatest inspiration in my life. Thank you for providing me unconditional support and bearing the loneliness of not having me around for the last five years.

ABSTRACT OF THE THESIS

Forecasting Electric Vehicle Arrival & Departure Time on UCSD Campus Using Support Vector Machines

by

Zhuo Xu

Master of Science in Engineering Sciences (Applied Ocean Science)

University of California, San Diego, 2017

Professor Jan Kleissl, Chair

Nowadays, with the development of technologies, road transportation in US has a tendency of shifting from traditional oil consumption method to more electric demand oriented vehicles. Therefore, with the increasing number of electric vehicles in California [1], we need to develop an algorithm to distribute electric power more appropriately. Large penetration of electric vehicles (EV) will increase the electricity demand and load forecasting plays a central role in the operation and electric power. This thesis presented a short – term arrival and departure time forecasting by using Support Vector Machine, which is an artificial intelligence technique that can be used for data prediction. Forecast results were compared for each year and it was found that larger training model has relatively smaller percentage of error in terms of accuracy.

1. INTRODUCTION

The fleet of plug in electric vehicles in the United States is larger than any other countries in the world. The statistic indicates that there are approximately 506,450 [1] highway plug in electric vehicles from 2008 to 2016 [2]. As a matter of fact, California is one of the states in US that provided relatively large number of electric vehicles. There is a statistic showing that approximately 40% [3] of US electric cars sales are in California.

San Diego is one of the most important cities in Southern California and it has many plug in electric vehicles customers. University of California, San Diego (UCSD) has a lot of commuters that utilizing electric vehicles near every day, this also include the transportation vehicles on campus. In this paper, the Electric Vehicles that particularly are being used by those commuters of UCSD were studied.

The purpose of this thesis is to predict the arrival and departure time of electric vehicles on UCSD campus to better distribute electric power more effectively on the grid. Therefore, EV's arrival and departure dataset in 2012, 2013 and 2014 were selected for analysis. In this thesis, support vector machine (SVM) method was used to analyze the data in those 3 years. 45 weeks of training and 5 weeks of testing dataset were also selected for each year. 1 week validation dataset was used as observed dataset to compare with the predicted dataset forecasted by training and testing data. Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Mean Bias Error (MBE) were computed and compared as results [4,5].

2. THEORY

2.1 The Forecasting Model

Load forecasting can be separated into three major categories based on the timeline: short term, medium term and long term forecasting. For short term forecasting (STF), it is primarily used to predict the arrival and departure time from one hour to one week. It plays a significant role for the next day unit commitment and grid operation optimization. The accuracy is essential for the power system stability and energy cost minimization. For medium term forecasting (MTF), it is primarily used for the prediction of arrival and departure time from one week to one year in advance. Medium term forecasting is vital for power system planning. Operation as maintenance and load switching are determined by accurate medium term. Long term forecasting (LTF) is primarily used for predicting the electricity load from one year to many years in advance. Therefore, it has a very strong impact on the decisions for investments on infrastructure and new generation units.

Support Vector Machines (SVM) is a machine learning algorithms that analyze the data used for classification and regression analysis [8,9]. It contains all the main features that characterize maximum margin algorithm. It also can easily manage high dimensional data and are flexible in modeling diverse sources of data. The major modeling framework includes training and testing a set of data. Also, the purpose of SVM is to produce a model based on training data which predicts the target values. For instance, given a training set

$\left(x_{i,Z_i}\right), i = 1, \dots, l$ where $x_i \in R^n$ and $z \in R^n$, the support vector machines (SVM) require the solution of the following optimization problem:

$$min_{w,b,\varphi} = \frac{1}{2} w^T w + C \sum_{i=1}^{l} \varphi_i$$

Corresponding to

$$w^T \emptyset(x_i) + b - z_i \leq \varepsilon + \varphi_i$$

$$\varphi_i \geq 0, i = 1, \dots l.$$

Figure 1 below describes the flowchart of modeling process of this algorithm

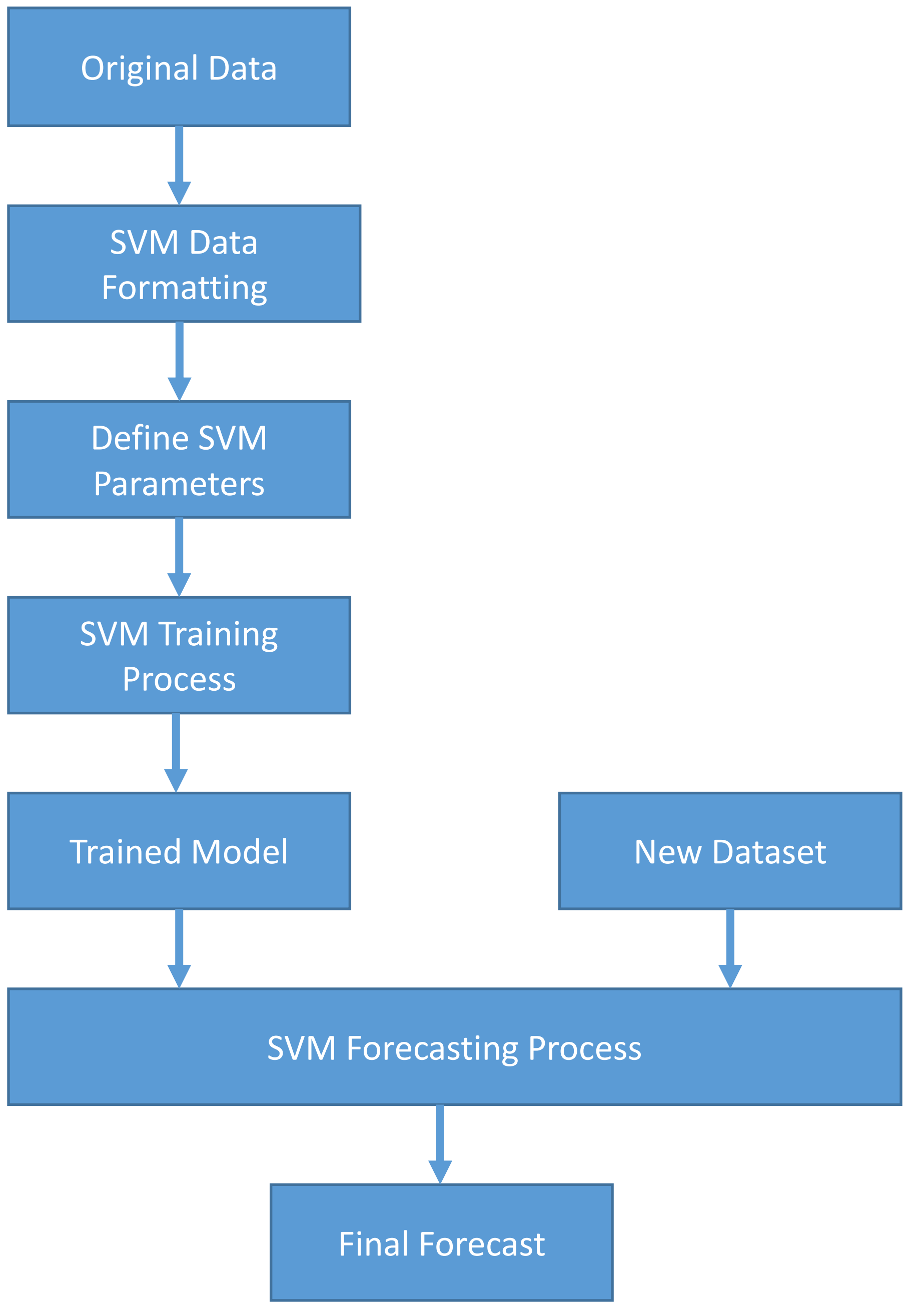

Figure 1: Operation Flow Chart of the SVM Model

2.2 Methodology

In this paper, EV data from 2012 to 2014 on UCSD campus were focused primarily.

Table 1: Sample of raw EV charging data on UCSD campus from 2012

City	State	Postal Code	Serial Numl	Member Guest	Connection Time	Disconnect Time	Cumulative Energy
La Jolla	CA	92037	216136	4116688	1/8/2015 19:31	1/8/2015 21:09	5.06299
La Jolla	CA	92037	216132	4BB304	1/8/2015 19:35	1/8/2015 21:43	9.71313
La Jolla	CA	92037	216130	2075682	1/8/2015 18:35	1/8/2015 21:03	9.22815
La Jolla	CA	92037	211655	5576616	1/8/2015 15:40	1/8/2015 17:19	5.9929185
La Jolla	CA	92037	216137	6319260	1/8/2015 8:26	1/8/2015 14:17	14.8896
La Jolla	CA	92037	215597	10012618	1/8/2015 12:10	1/8/2015 14:52	13.189
La Jolla	CA	92037	200987	8505163	1/8/2015 11:43	1/8/2015 14:51	4.87195
San Diego	CA	92037	200786	3280234	1/8/2015 12:33	1/8/2015 13:48	3.03735
La Jolla	CA	92037	216136	8415385	1/8/2015 10:56	1/8/2015 12:24	4.42383
La Jolla	CA	92037	216134	5787689	1/8/2015 9:20	1/8/2015 12:16	9.7128935
San Diego	CA	92037	200786	3280234	1/6/2015 18:42	1/8/2015 11:42	0.587402
La Jolla	CA	92037	215940	10014736	1/6/2015 9:40	1/6/2015 10:30	4.03259
La Jolla	CA	92037	216137	10018646	1/7/2015 19:25	1/8/2015 0:30	7.34814
La Jolla	CA	92037	211655	4692957	1/7/2015 18:31	1/7/2015 20:15	8.22656
La Jolla	CA	92037	213549	D496D5	1/7/2015 12:11	1/7/2015 15:05	13.0625
La Jolla	CA	92037	216135	8415385	1/7/2015 12:59	1/7/2015 15:28	7.02512
La Jolla	CA	92037	216137	6319260	1/7/2015 8:24	1/7/2015 14:20	16.1016
La Jolla	CA	92037	216134	294900	1/6/2015 13:00	1/6/2015 17:12	17.3535
La Jolla	CA	92037	200987	8505163	1/6/2015 12:08	1/6/2015 15:47	5.1748
La Jolla	CA	92037	216132	8463297	1/6/2015 10:45	1/6/2015 15:35	12.0212

As can be seen, the first few columns represent charging locations and vehicle identifications. The last three columns represent connection time, disconnection time and cumulative energy.

Table 1 shows the sample raw data for EV charging events on UCSD campus from 2012 to 2014. The goal of this paper is to forecast the arrival and departure time of electric vehicles. Therefore, we only need to consider EV's connection time, disconnection time and cumulative energy.

For instance, figure 2 can be shown if EV's arrival and departure information were plotted of year 2012.

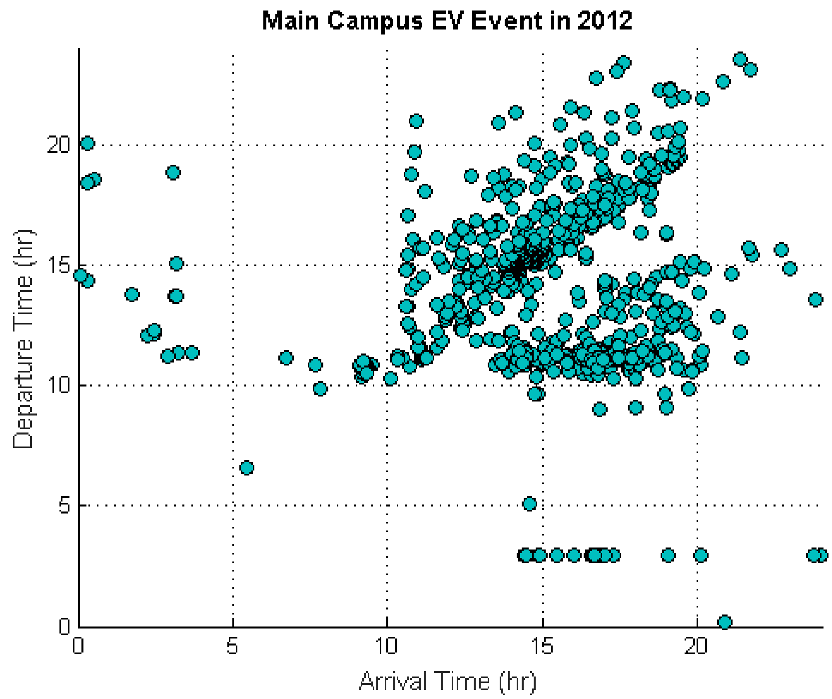

Figure 2: Arrival & Departure Time of All EVs on UCSD Campus of year 2012

Figure 2 shows an example of arrival and departure time of all types of electrical vehicles on UCSD in 2012. As it is showed from the figure, the primary arrival time and departure time of EVs are distributed approximately between 10 am to 8 pm.

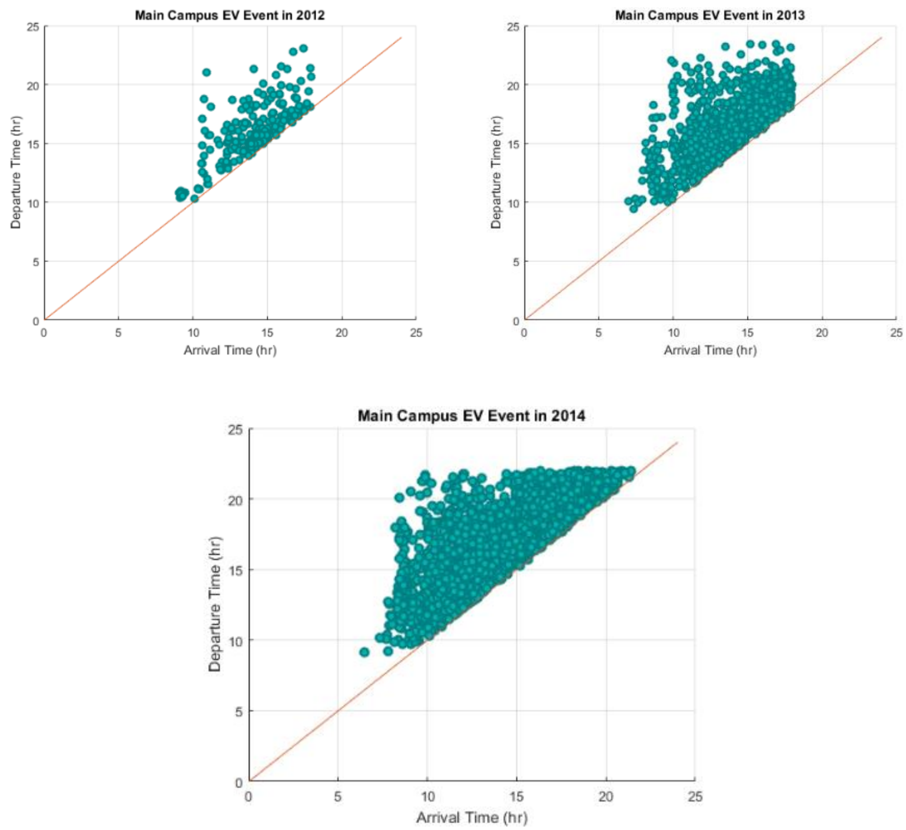

Figure 3: Arrival and Departure Time of Designated EVs on UCSD Campus of 2012, 2013 and 2014

Figure 3 indicates that the arrival and departure time of designated EVs on campus after applying all the restrictions above. It is clear that the number of electric vehicles arrived and departed on UCSD campus has increased from 2012 to 2014.

There are several methods to measure the accuracy of a forecasting model. In this paper, Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Mean Bias Error (MBE) are used to evaluate the performance of the method.

The formula for calculating the MAPE, RMSE [6,7] and MBE are:

$$MAPE = \frac{\sum_{1}^{N} |Observed - Predicted|}{Ave\_Observed} \times 100\%$$

$$RMSE = \sqrt{\frac{\sum_{1}^{N} (Observed - Predicted)^2}{N}}$$

$$MBE = \frac{\sum_{1}^{N} (Observed - Predicted)}{N}$$

Where N is the number of predicted values and Ave_Observed is the average value of observed data.

Because the data is collected from 2012, 2013 and 2014 on UCSD campus and the number of electric vehicles participated in the event has increased tremendously over those 3 years. Therefore, it is not appropriate to apply the machine learning method over those 3 years all together since those data follow different distribution pattern. Instead, those data were divided weekly within the same year.

2.3 Case Study

A. Data Configurations

This paper only focused on the Electric Vehicles that are used for daily commuting on UCSD campus. Therefore, the range of arriving and departing time has been restricted from 6 am to 9 pm. Overnight parked vehicles were eliminated if arrival time is greater than departure time. In addition, the cumulative energy is negligible if the difference between arrival and departure time is very small. In this case, any cumulative energy less than <0.5> kwh will be eliminated. In addition, any electric vehicles arrived and departed before 6am or after 9pm will also be eliminated. The outcome of the model is a matrix including all information about EV's arrival and departure time between 6 am to 9 pm on UCSD campus for year 2012, 2013 and 2014.

B. Modeling

Once the historical data model for one year is completed, the next step is to format those data as inputs to the Support Vector Machine (SVM).

The data are classified as the following features:

1. *Week*: number of the week (1-51)
2. *Day*: number of the day (1-7) starting with Monday
3. *Hour*: 6 am to 9 pm of each day, one hour increment
4. *Arrival Time*: number of electric vehicles arrived within each hour
5. *Previous Arrival Time*: number of electric vehicles arrived within each hour of previous hours
6. *Departure Time*: number of electric vehicles departed within each hour
7. *Previous Departure Time*: number of electric vehicles departed within each hour of previous hours

This Table contains a summary of the above features for one week

Table 2: Input Parameters for the SVM Model (One Week)

Week	1 … 1
Day	1 … 7
Hour	6 … 21 … 6 … 21
Arrival Time	A_6 … A_{21} … A_6 … A_{21}
Previous Arrival Time	PA_6 … PA_{21} … PA_6 … PA_{21}
Departure Time	D_6 … D_{21} … D_6 … D_{21}
Previous Departure Time	PD_6 … PD_{21} … PD_6 … PD_{21}

The entire sample of the historical data is divided into three parts for each year: training, testing and validation. The majority of data sample was used to train the model in order to maximize the accuracy of the data as much as possible. Once the model is trained, the testing data for 5 weeks are provided to the model. The final output is the forecast of a single week. The accuracy of the forecast will be evaluated by applying the validation set.

Forecasting for just one week is not sufficient to evaluate the model. Therefore, three forecast runs were generated for different weeks in the years and the values of RMSE, MAPE and MBE were averaged. For each forecast run, 1 week as validation dataset, 5 weeks as testing dataset and the remainder of the year as training dataset.

3. RESULTS AND DISCUSSIONS

The results of the proposed model are compared for each year. The training part of the SVM model is the percentage of arrival and departure time of EVs on UCSD campus for each week of each year. In this case, the training and testing part of the SVM model are the arrival and departure time for 50 weeks. In this paper, 3 forecast runs were generated for different weeks of the year and the errors were averaged. For those forecast runs, 5^{th}, 15^{th} and 45^{th} week were selected as validation sets, respectively. Therefore, the datasets of those 3 weeks were excluded from training and testing dataset. 5^{th}, 15^{th} and 45^{th} week are appropriate weeks to be selected as validation data because they represent typical winter, spring and fall quarter of the year. Once the model of 50 weeks is trained, the forecast is produced. The forecast is then compared with the arrival & departure event for 5^{th}, 15^{th} and 45^{th} week, respectively.

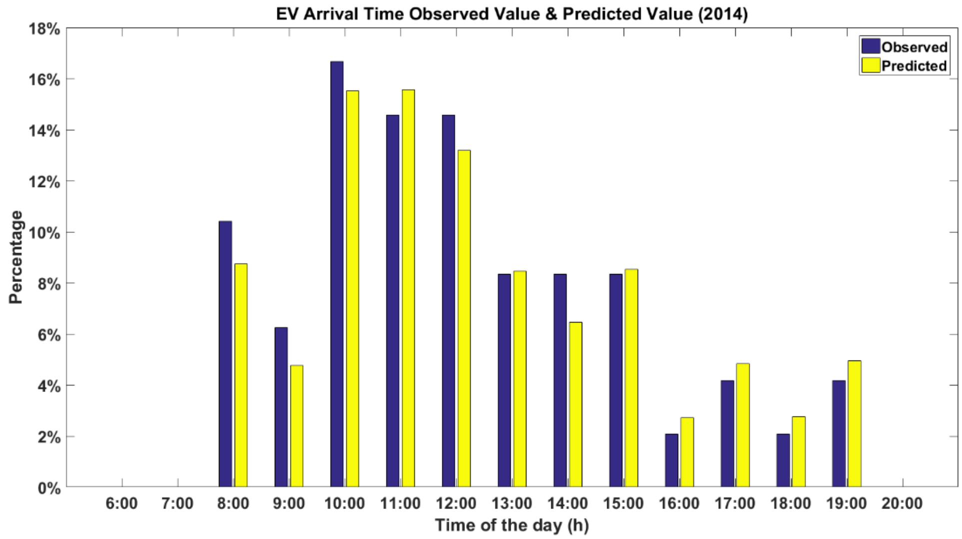

Figure 4: Observed & Predicted EV Arrival Time Comparison of 2014 on UCSD Campus (Averaged value of 5th, 15th and 45th week)

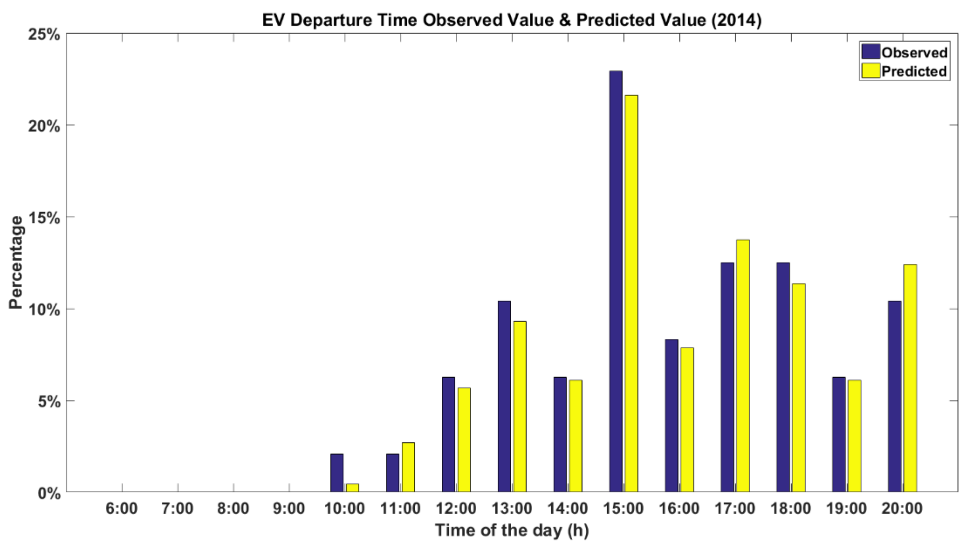

Figure 5: Observed & Predicted EV Departure Time Comparison of 2014 on UCSD Campus (Averaged value of 5th, 15th and 45th week)

Table 3: MAPE, RMSE & MBE Performance for Year 2012, 2013 & 2014 by Using Vector Support Machine method

Year / Index	2012 (arrival)	2012 (departure)	2013 (arrival)	2013 (departure)	2014 (arrival)	2014 (departure)
MAPE (%)	3.173	6.250	3.918	3.478	1.465	1.375
RMSE	0.033	0.059	0.023	0.038	0.012	0.011
MBE	0.014	0.028	0.028	0.022	0.007	0.006

Table 4: MAPE, RMSE & MBE Performance for Year 2012, 2013 & 2014 by Using Reference Forecast

Year / Index	2012 (arrival)	2012 (departure)	2013 (arrival)	2013 (departure)	2014 (arrival)	2014 (departure)
MAPE(%)	4.333	7.001	4.086	4.498	1.484	1.970
RMSE	0.055	0.063	0.107	0.110	0.010	0.083
MBE	0.020	0.0311	0.055	0.107	0.007	0.008

As can be seen from the tables above, both MAPE and RMSE show the tendency of decreasing over 2012 to 2014. In addition, it showed that the distribution of EV's arrival and departure time in 2014 covers more from 6am to 9pm than in year 2012 and 2013.

In addition, a reference forecast was computed to compare against the forecast generated by SVM. This comparison will establish how well the more sophisticated and difficult-to-implement forecast using SVM compares against a very simple and easy-to-implement forecast. In this paper, 5^{th}, 15^{th} and 45^{th} week were selected as validation datasets. One of the best reference forecast is persistence. Therefore, 4^{th}, 14^{th} and 44^{th} week were selected to forecast 5^{th}, 15^{th} and 45^{th} week of year 2012, 2013 and 2014. The errors were averaged and MAPE, RMSE and MBE were calculated. As can be seen from 2 tables above, the forecast results generated by SVM tend to have relatively smaller error than the forecast results generated by reference forecast.

One of the primary reasons is that the number of electric vehicles participated on UCSD campus has dramatically increased over these 3 years. Larger training model will have relatively smaller MAPE error and better performance.

4. CONCLUSIONS

Support Vector Machine (SVM), an artificial intelligent supervised learning model with associated learning algorithm used for data prediction, was used in this paper. A short-term Electric Vehicle arrival and departure forecasting model is also presented in this paper. Data sets of EV's arrival and departure time on UCSD campus of year 2012, 2013 and 2014 were trained as a training model. After the model was trained, the SVM model provided a forecast for weekly EV's arrival and departure time distribution. The forecast results were compared within each year, respectively and proves that larger training model will have relatively smaller MAPE error and better performance.

5. FUTURE WORK

The primarily focus in this paper was one week short term forecasting method. The limitation of this method is that one week period is the maximum precision can be accessed. One day precision forecasting was not able to be accessed in this case. UCSD campus is relatively small area to evaluate electric vehicle events in San Diego and there is only very limited amount of EVs participated on weekdays. Therefore, we anticipate larger error if daily forecasting method was used in this paper. In future work, more datasets will be acquired for weekdays to generate daily forecast.

In addition, the proportion of training, testing and validation data is another factor in terms of forecasting accuracy. In this paper, 45 weeks of training data, 5 weeks of testing data and 1 week of validation data were evaluated. In future work, different proportion of training, testing and validation dataset maybe attempted to obtain different forecasting accuracy and evaluate the performance [10].

REFERENCES

[1] N.p., n.d. Web. <https://en.wikipedia.org/wiki/Plug-in_electric_vehicles_in_the_United_States>.

[2] Frangoul, Anmar. "Electric vehicles: California leads the charge." N.p., n.d. Web.

[3] Ayre, James. "40% Of US Electric Car Sales Are In California." N.p., 21 Nov. 2015. Web. <https://cleantechnica.com/2015/11/21/40-of-us-electric-car-sales-are-in-california/>.

[4] Xydas, E. S., C. E. Marmaras, L. M. Cipcigan, A. S. Hassan, and N. Jenkins. "Forecasting Electric Vehicle charging demand using Support Vector Machines." 2013 48th International Universities' Power Engineering Conference (UPEC) (2013): n. pag. Web.

[5] Hsu, Chih-Wei, Chih-CHuang Chang, and Chih-Jen Lin. "Support Vector Machines for Classification." Support Vector Machines Information Science and Statistics (2016): 285-329. Web

[6] Amjad, Shaik et al. "Evaluation Of Energy Requirements For All-Electric Range Of Plug-In Hybrid Electric Two-Wheeler". *Energy* 36.3 (2011): 1623-1629. Web.

[7] Bayram, Islam Safak et al. "On The Evaluation Of Plug-In Electric Vehicle Data Of A Campus Charging Network". (2016): 6. Print.

[8] Tortos, Jairo Quiros and Luis(Nando) Ochoa. "A Statistical Analysis Of EV Charging Behavior In The UK". 2015. 6. Print.

[9] Bayram, Islam Safak et al. "Capacity Planning Frameworks For Electric Vehicle Charging Stations With Multiclass Customers". *IEEE Transactions on Smart Grid* 6.4 (2015): 1934-1943. Web.

[10] Clement-Nyns, K., E. Haesen, and J. Driesen. "The Impact Of Charging Plug-In Hybrid Electric Vehicles On A Residential Distribution Grid". IEEE Transactions on Power Systems 25.1 (2010): 371-380. Web.