Prediction of Next Day's Sea-Level Pressure by Support Vector Regression

Tin Nilar Lin† Hiroshi Watanabe†

†Graduate School of Fundamental Science and Engineering, Waseda University

1. INTRODUCTION

Today, prediction of weather or climate data plays an important role in disaster management. There are many influence factors to predict the weather or climate data accurately. Thus, the prediction model of weather data will be complex to get the result as accurate as possible and some factors is difficult to get exactly. In this paper, we focus on the prediction of sea level pressure for next day as accurate as possible with optimal features by applying Support Vector Regression (SVR).

2. Proposed Approach

We applied the valid SVR to get the good prediction accuracy. The aim of ε-SV regression is to find the linear function that has at most predefined deviation ε from the obtained target for all training data. That is, we neglect the error smaller than predefined threshold and any deviation higher than the threshold will be penalized [1], [2], [3].

In our study, we also find the optimal subsets of parameters of proposed approach by investigating combination of the penalty parameter C∈ (1, 2, … 5, 10, 20, ... 100) and free parameter $\gamma \in$ (0.01, 0.02, 0.03, … 0.1) using RBF kernel. The estimated optimal parameters for our prediction approach is C=1, $\gamma = 0.03$ and predefined error threshold ε=0.01.

3. MODEL EVALUATION AND EXPERIMENTAL RESULT

In this experiment, we rearrange the weather dataset from Kaggle data science repository [4], and create our own dataset with 21 features and 3300 instances to predict the sea level pressures of next day ahead. As a model evaluation, correlation coefficient (CE), mean absolute error (MAE) and root mean square error (RMSE) with 10 folds cross validation are used (1), (2), (3). First, we evaluated the CE value of each observed and predicted output to know the degree of relationship between them. Next, we arrange the features according to their relative sensitivity value and selected high rank features. Then, we filtered five unselected features that showed the lowest rank.

$$CE = 1 - \frac{\sum_{i=1}^{m}[O_i - P_i]^2}{\sum_{i=1}^{N}[O_i - \bar{O}_i]^2} \quad (1)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{m}[|O_i - P_i|] \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{m}[O_i - P_i]^2} \quad (3)$$

In these equation, O_i and P_i is the observed and predicated values. Finally, we compared the prediction performance before and after some features selection on our dataset as shown in Table I. The prediction results of sea level pressure for next day ahead by the SVR is shown in Fig. 1.

TABLE I Evaluation of Model Performance

Evaluation Index	Before feature reduced	After feature reduced
MAE	**0.6213**	**0.4125**
RMSE	2.8876	2.8780

After the selection of some optimal features, the prediction performance of the model becomes better without suffering the accuracy. The prediction of sea level pressure of next day is also close agreement with the observed value.

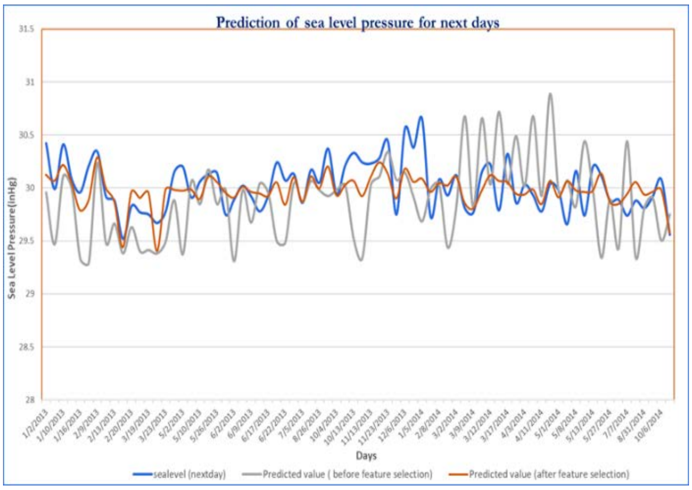

Fig 1. Comparison of prediction results

The prediction performance of next day becomes better with the smaller dataset and reduce the training time after selecting some optimal features. It is possible to reduce much training time when the number of instances for prediction is increased.

4. Conclusion

In this paper, we analyze the SVR with optimal features to make the prediction of sea level pressure for next day with good prediction accuracy. We adjust the SVR parameters to obtain the good prediction solution.

References

[1] Z.A. Sunkad and Soujanya, "Feature Selection and Hyperparameter Optimization of SVM for Human Activity Recognition," 3rd Int. Conf. on Soft Computing & Machine Intelligence, pp.104-109, Nov. 2016.
[2] T. Nguyen, Q. N. Huu, and M. J. Li, "Forecasting Time Series Water Levels on Mekong River Using Machine Learning Models," in 7th Int. Conf. on Knowledge and Systems Engineering, pp.292-297, 2015.
[3] Alex J. SmolaBernhard Schölkopf, "A tutorial on support vector regression," Statistics and Computing, vol. Volume 14, no. Issue 3, p. pp 199–222, August ,2004.
[4] "Kaggle," Kaggle Inc, [Online]. Available: https://www.kaggle.com /c/walmart-recruiting-sales-in-stormy-weather/data. [Accessed 15 August 2016].