SEASONAL HEAVY RAIN FORECASTING METHOD

Jaekwang KIM

School of Convergence / Convergence program for social innovation,
Sungkyunkwan University, Seoul 03063, South Korea

ABSTRACT

In this study, we study the technique for predicting heavy / non-rain rainfall after 6 hours from the present using the values of the weather attributes. Through this study, we investigated whether each attribute value is influenced by a specific pattern of weather maps representing heavy and non-heavy rains or seasonally when making heavy / non-heavy forecasts. For the experiment, a 20-year cumulative weather map was learned with Support Vector Machine (SVM) and tested using a set of correct answers for heavy rain and heavy rain. As a result of the experiment, it was found that the heavy rain prediction of SVM showed an accuracy rate of up to 70%, and that it was seasonal variation rather than a specific pattern that influenced the prediction.

KEYWORDS

Prediction Method, Forecasting, Machine learning, Feature extraction.

1. INTRODUCTION

Prediction of dangerous weather forecasts such as heavy rain and heavy snow is a very important research field because it has a profound effect on many people in various fields [1]. A typical method of predicting heavy rain today is to predict the weather using the situation at that time based on the past weather map similar to the current weather map. However, this method has a problem in that it is expensive to find a pattern similar to the current weather map in the vast amount of past weather maps. To this end, a data mining study that effectively analyzes past weather forecasts is needed [3]. In this study, based on past weather data, we learned whether heavy rain would occur or not after 6 hours, and predicted whether heavy rain would occur after 6 hours when current weather data was input. In addition, as an analytical study for this, we compared and analyzed whether each attribute representing weather is highly influenced by the feature pattern of weather map according to heavy / nonheavy weather or seasonally. Prediction of heavy / nonheavy rainfall after 6 hours through experiments showed accuracy of up to 70%, and it was confirmed that each property was more affected by the season. The rest of the paper is organized as follows: Section 2 includes an explanation of previous weather chart for forecasting. In Section 3, Heavy rain prediction experiments are shown to verify the proposed method. Section 4 presents a result of seasonal heavy rain situation prediction experiment. Finally, this paper is concluded in Section 5.

2. PREVIOUS WEATHER CHART

For the past weather forecast, the weather forecast with ECMWF 1.5-degree resolution in 6-hour increments (00, 06, 12, 18 UTC) was used. Including UWND (East-West), VWND (North-South)

David C. Wyld et al. (Eds): SAIM, ACSIT, SNLP, ICITE, CoNeCo - 2021
pp. 11-16, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111002

with 6 properties and 4 isobar face layers of 850hPa, 700hPa, 500hPa, 200hPa for each property were used. The criterion for determining heavy rain is when the cumulative precipitation for over 6 hours rains more than 70 millimeters, and the criterion for determining heavy rain is defined as the cumulative precipitation for 6 hours is 5 millimeters or less. In addition, heavy / non-heavy rain days were extracted according to sequentially defined criteria among May, June, July, and September during which heavy rains are concentrated throughout the year. The total duration of the data used is 20 years from 1989 to 2009. The ECMWF 1.5- degree data consist of 60x31 grids representing East Asia. To predict the weather on the Korean Peninsula, it is unnecessary to observe the entire map of East Asia, and as shown in Figure 1, a 10x10 grid including the Korean Peninsula region and its surroundings was extracted and used.

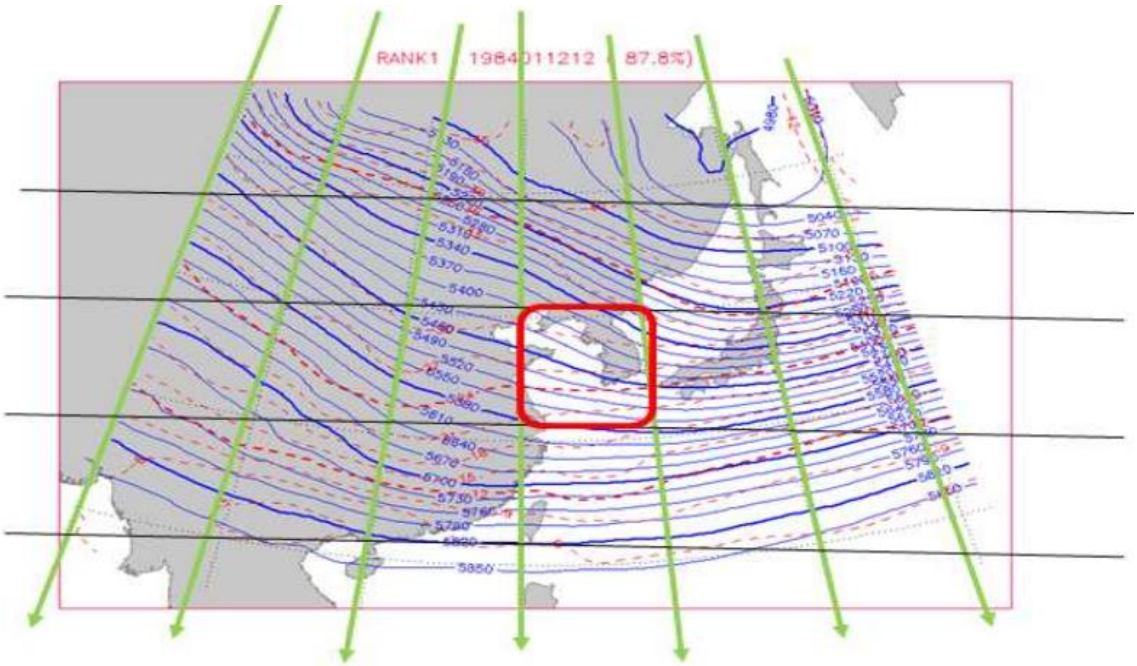

Figure 1. Extracting weather maps from the Korean Peninsula.

3. PROPOSED METHOD

3.1. Support Vector Machine: A Binary Classifier

SVM is one of the most widely used supervised learning techniques [5][6][7]. This model is similar to logistic regression in that it is driven by a linear function. One important point related to SVM is the use of the kernel trick [8]. Kernel trick allows many machine learning algorithms to be expressed only as inner product terms between samples.

Although the performance of SVM can vary greatly depending on the kernel function, there is a limitation in application that there is no theory on how to select an appropriate kernel function according to the data (Amari and Wu, 1999; Byun and Lee, 2002). . Therefore, it is necessary to go through the process of searching for parameters in order to select an appropriate kernel function according to the characteristics of the data. As such, parameter estimation is very important in using SVM as a prediction model. One thing to be aware of here is to set a parameter with a high degree of fit biased toward the learning data, which is called overfitting. That is, performance may be improved for previously used training data, but may decrease for new data. Therefore, in this study, the 10-Fold CV technique was used as a cross-validation (CV) method to solve this problem. This is a method that randomly divides data (x, y) into 10 sets, uses one as validation data, and uses the rest as learning data. After modeling, the average error value is used to determine the Evaluate the model's performance.

3.2. Seasonal Heavy Rain Forecasting System

Figure 2 shows a conceptual diagram of our heavy rain prediction system. The system is composed with two man-parts: one is the training part and the other is the forecasting part. In the training part, two different conditions are input as a training data. The SVMs are trained with these inputs. In the forecasting part, weather data is input to the trained SVM model and the output shows that the input data is heavy rain or not.

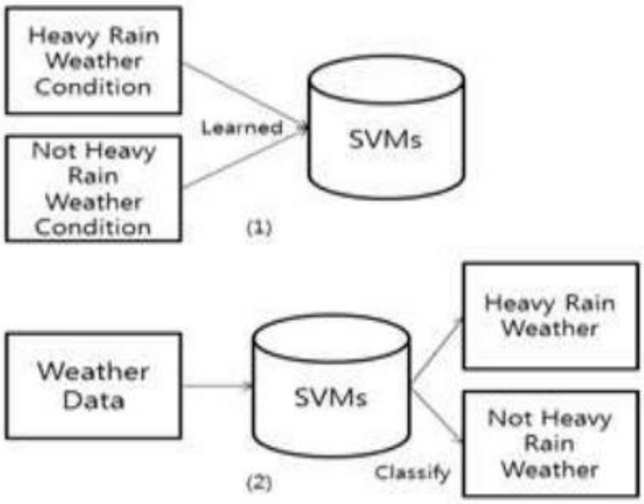

Figure 2. The Concept of the System.

4. HEAVY RAIN PREDICTION EXPERIMENT

For the prediction of heavy / non-heavy rainfall after 6 hours, a training set was constructed using a total of 200 dates of 100 heavy / non-heavy rains as defined above. The experiment was conducted with 5-fold cross validation consisting of 160 training data and 40 test data from the training set. Table 1 shows experimental results for each property and isobar. The attributes that showed the best results for each isostatic surface are shown in bold. The most predicted accuracy was 65-70% in most isostatic planes. Looking at the characteristics of the difference between isobars, GHGT and SHUM showed good results on all isobars, and TEMP also showed good results at low altitudes.

Table 1. Heavy / Non-Heavy Forecast Accuracy according to Each Feature.

hPa	GHGT	RHUM	SHUM	TEMP	UWND	UWNDxt
850	62.00	53.50	69.50	66.47	50.50	53.00
700	66.50	49.00	61.00	66.50	53.50	56.00
500	68.50	53.50	60.00	68.00	59.00	52.00
200	65.50	65.00	64.30	47.50	59.00	46.00

5. SEASONAL HEAVY RAIN SITUATION PREDICTION EXPERIMENT

For the prediction of the heavy rain situation according to the season, 100 rain and rain days meeting the criteria were selected in May and September based on the criteria defined above. 100 dates were selected for each of 50 dates. May, September, July, and August were grouped into two seasons, one summer and one summer. 5-fold cross validation was performed with 80 training data and 20 test data using 100 monthly dates consisting of a total of 4 sets. The experiment was conducted at 850hPa isostatic pressure. Accuracy, Precision, Recall, and F1 measures were used as indexes to evaluate the results. The results of experiments using dates classified in different seasons are much higher than the results obtained by using the data classified in the same season without distinction between heavy and non-heavy rains.

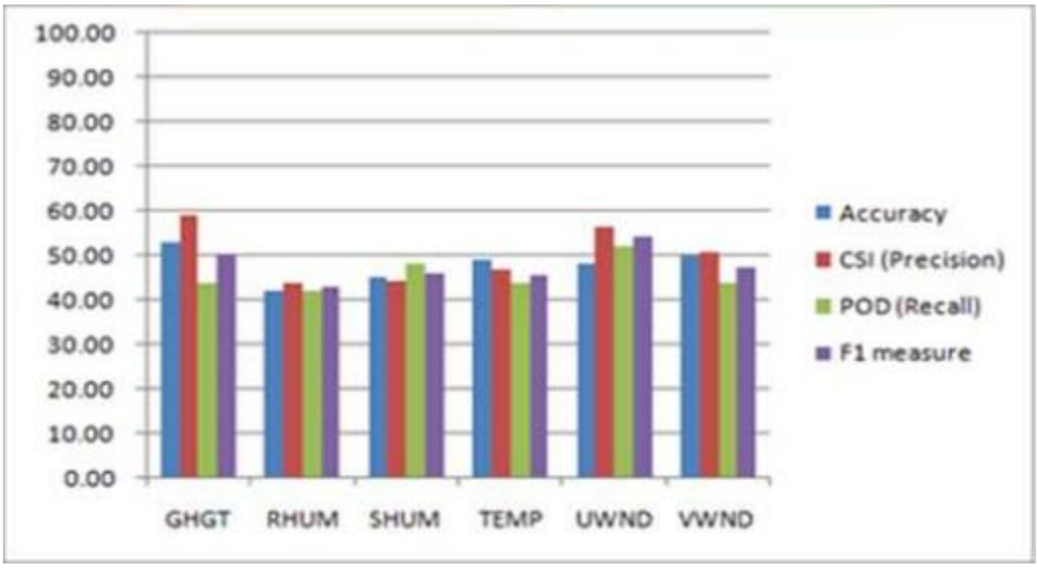

Figure 3. Seasonal heavy / non-rain rainfall forecast accuracy. (heavy rain 7, 8 vs. non-heavy rain 7, 8).

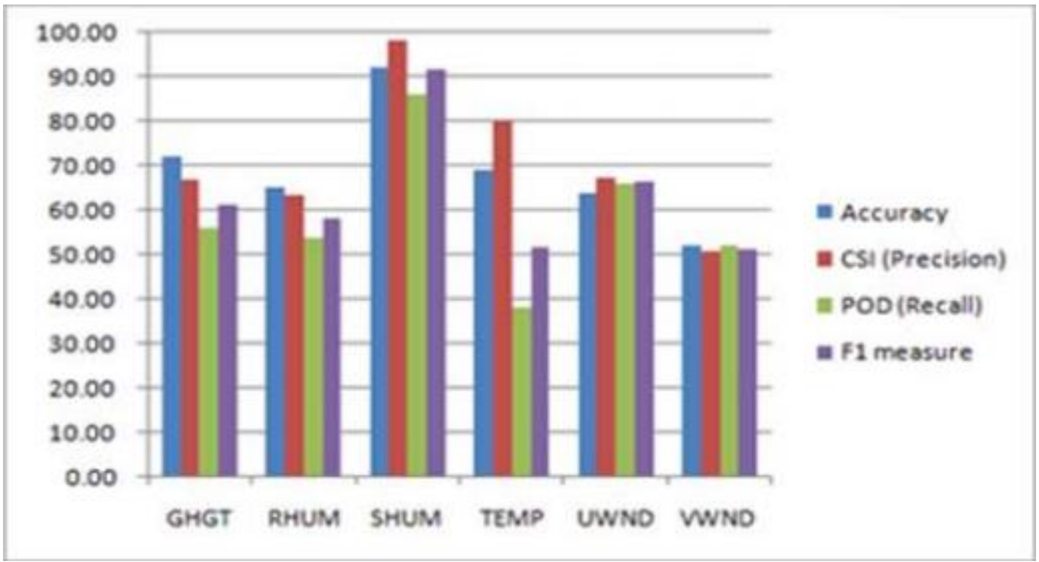

Figure 4. Seasonal heavy / non-rain rainfall forecast accuracy. (heavy rain 7, 8 vs. non-heavy rain 5, 9).

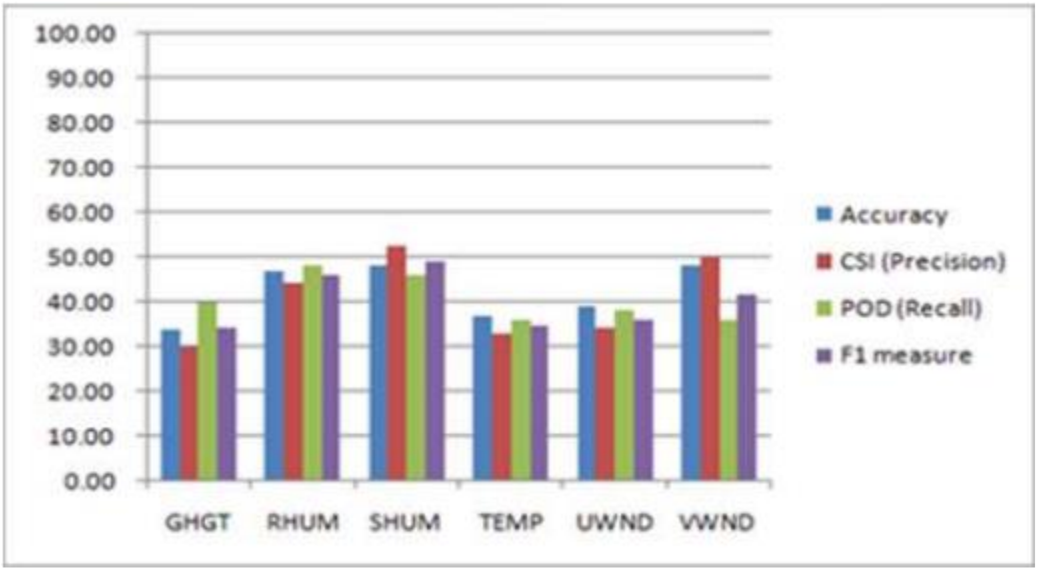

Figure 5. Seasonal heavy / non-rain rainfall forecast accuracy. (heavy rain 5, 9 vs. non-heavy rain 5, 9).

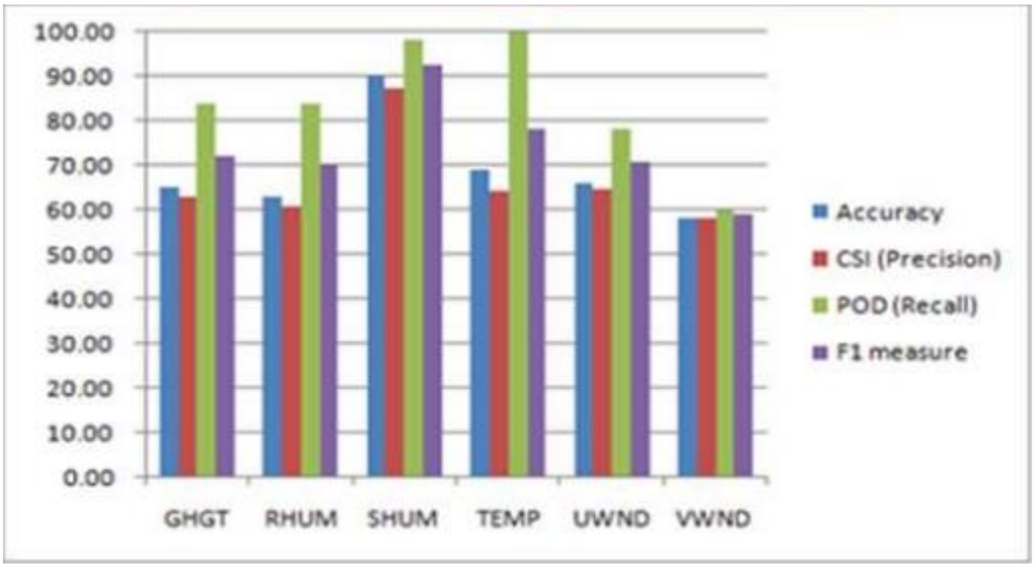

Figure 6. Seasonal heavy / non-rain rainfall forecast accuracy. (heavy rain 7, 8 vs. heavy rain 5, 9).

6. CONCLUSIONS

In this paper, we study the technique for predicting heavy / non-rain rainfall after 6 hours from the present using the values of the weather attributes. We investigated whether each attribute value is influenced by a specific pattern of weather maps representing heavy and non-heavy rains or seasonally when making heavy / non-heavy forecasts. In the experiments, a 20-year cumulative weather map was learned with Support Vector Machine (SVM) and tested using a set of correct answers for heavy rain and heavy rain. By considering the experiment results, it was found that the heavy rain prediction of SVM showed an accuracy rate of up to 70%, and that it was seasonal variation rather than a specific pattern that influenced the prediction. In the future, combining our proposed method with the latest research related to recommendation systems or knowledge processing, we expect to be able to expand expressions efficiently [9][10].

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF2019R1C1C1008174) and by the MSIT(Ministry of Science and ICT), Korea, under the ICT Creative Consilience program(IITP-2020-0-01821) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation).

REFERENCES

[1] Lee, J., Kim, J., Lee, J.-H., "Parameter Evaluation to Classify Heavy Rain using SVMs", ISIS 2011.

[2] Korea Meteorological Administration

[3] http://www.kma.go.kr

[4] Lee, J., et al. "Development of a Similar Diary Search Algorithm using Principal Component Analysis", *Proceedings of the Korean Meteorological Society Conference*, pp. 214-215.

[5] Boser, B.E., Guyon, I.M., and Vapnik, V.N., "A training algorithm for optimal margin classifiers", *In COLT '92: Proceeding of the fifth annual workshop on Computational learning theory*, ACM, New York, NY, USA, pp. 144-152, 1992.

[6] Cortes, C., and Vapnik, V., "Support vector networks", *Machine Learning*, Vol. 20, pp. 273-297, 1995.

[7] Goodfellow, I., Bengio, Y., and Courville, A., "Deep learning", *MIT press*, 2016.

[8] Kwon, H., Kim, M., and Kim, W., "A development of water demand forecasting model based on Wavelet transform and Support vector machine", *Journal of Korea Water Resources Association*, Vol. 45, No. 11, pp. 1187-1199, 2012.

[9] Kim, J., & Lee, J.-H., "A Novel Recommendation Approach Based on Chronological Cohesive Units in Content Consuming Logs", *Information Sciences*, Vol. 470, pp. 141-155, Jan. 2019.

[10] Kim, J., "A Document Ranking Method with Query-related Web Context", *IEEE ACCESS*, Vol. 7, pp. 150168-150174, Oct. 28, 2019.

AUTHOR

Jaekwang KIM

He was born in Seoul, South Korea, in 1980. He received the B.S., M.S., and Ph.D. degrees from Sungkyunkwan University (SKKU), Suwon, South Korea, in 2004, 2006 and 2014, respectively, where he is a currently an assistant professor with the College of Convergence and Convergence program for social innovation. His research interests include networks, security, medical health, education based on various machine intelligent algorithms. He received the Best Presentation Paper Award at ICUIMC, Suwon, in January 2009, and was elected as a Bronze Medal Superior Student from the Brain Korea 21 Enterprise Department, Sungkyunkwan University.