Precipitation forecasting using satellite images and SVMs*

Eslam Hussein[0000−0001−6040−8516] and Mehrdad Ghaziasgar[0000−0002−7713−677X]

Department of Computer Science, University of the Western Cape, South Africa
eslamhuss34@gmail.com mghaziasgar@uwc.ac.za

Abstract. The prediction of rainfall is important for planning; it can help individuals plan their days ahead; more importantly it can help governments prepare for potential disasters. This research aims to investigate a data-driven approach to rainfall intensity prediction using support vector machines (SVMs) and sequences of daily satellite precipitation images as input. The primary aim of the work is to accurately predict one day ahead, but is also extended to predict several days into the future.

Keywords: Machine Learning· Rainfall · Satellite · Sequence · Images.

1 Introduction

Weather forecasting aims to predict the state of the atmosphere at a specific location and time in the future. The accurate prediction of the weather and climate is essential to the production of crops [1]. Weather can also cause natural disasters, such as typhoon in Mozambique in April 2019. An accurate weather forecasting system can provide early warnings which can help individuals and governments to better prepare for such events.

Recently, a number of papers studied the use of different machine learning techniques for weather prediction [[2], [3], [4], [5]]. Due to space constraints, we only explain Boonyuen et al's [6] work here. Boonyuen et al proposed the use of the standard “Inception 3” model to forecast rain fall up to three days ahead. Satellite image data taken every ten minutes was used as input. The system used one image as input to predict up to three days ahead on a binary classification scale of “rain” versus “no rain” with an accuracy of 63%. This accuracy represents an 13% advantage over random guessing which statistically yields a 50% accuracy for two-class problem. The authors recommended the use of more than one image as input, and to attempt to predict up to one week ahead. They also proposed to extend the classification scale to more classes, such as light rain, moderate rain, and heavy rain, which can serve as a more useful source of information.

* Supported by the Openserve/Telkom/Aria Technologies Centre-of-Excellence at UWC and the use of the ilifu cloud computing facility.

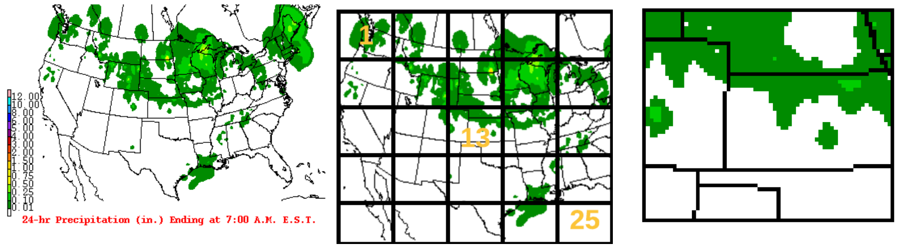

Fig. 1. Satellite precipitation images: (left) An image from the National Centers for Environmental Prediction (NCEP) data set; (centre) 5 grid overlay; (right) Grid cell 13.The NCEP data were provided by the NOAA at /https://www.esrl.noaa.gov/psd/

This research aims to investigate precipitation forecasting on the NCEP data set using SVMs.The data set consists of images of the United States (US) taken at 7am daily. One image of the data set is shown on the left of Figure 1. The parameters that will be investigated are: i) the length of the input image sequence i.e. the number of input days $\in \{2, 4, 6, 8\}$ and ii) the size of the input images $\in \{100\%, 50\%, 25\%\}$. Furthermore, we propose to divide up the US using a 5×5 grid seen in the centre of Figure 1; the US is very large and it is expected that the weather patterns that we aim to model and predict are more region-specific. We therefore aim to develop a separate model for, and investigate each of the two parameters mentioned previously on, each grid cell. Finally, we aim to predict a a multi-class output where we classify rainfall into three classes $\in \{\text{no rain}, \text{light rain}, \text{heavy rain}\}$, and to predict k days ahead $\in \{1, 2, 3, 4, 5, 6, 7, 14, 30\}$.

2 Preliminary Results

As a start, we ran an experiment on grid cell 13, indicated in the centre of Figure 1 and shown more clearly on the right of the figure, using sequences of 2 days as input and image sizes of 50%. The NCEP data was class-reduced into three classes, and all sequences of length 2 which started with a minimum threshold of precipitation were determined and used. This resulted in 2548 sequences for training and 284 sequences for testing. Training involved selecting the best cost parameter C of the SVM model using 10-fold cross-validation on the training set. Table 1 illustrates the prediction F1-score on the testing set for 1–30 days ahead (DA) on this grid cell.

Noting that the statistical accuracy of an untrained system on a 3-class problem is about 33%, the results are very encouraging. We are able to predict one DA with a 22% advantage over an untrained system. The F1 score fluctuates between 40% and 48% from 2 DA to 14 DA. Finally, we are able to predict at a 4% advantage over an untrained system even for a full month ahead.

Table 1. Testing F1-Score for predicting k days ahead (DA).

F1-Score of Days Ahead (DA) (%)								
1DA	2DA	3DA	4DA	5DA	6DA	7DA	14DA	30DA
55	46	48	46	41	40	48	46	37

References

1. M. Reyniers, *Quantitative Precipitation Forecasts Based on Radar Observations: Principles, Algorithms and Operational Systems.* Institut Royal Météorologique de Belgique Brussel, Belgium, 2008.
2. H. S. Jang, K. Y. Bae, H.-S. Park, and D. K. Sung, "Solar power prediction based on satellite images and support vector machine," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1255–1263, 2016.
3. S. S. Aung, Y. Senaha, S. Ohsawa, I. Nagayama, and S. Tamaki, "Short-term prediction of localized heavy rain from radar imaging and machine learning," *IEIE Transactions on Smart Processing & Computing*, vol. 7, no. 2, pp. 107–115, 2018.
4. B. Klein, L. Wolf, and Y. Afek, "A dynamic convolutional layer for short range weather prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4840–4848, 2015.
5. X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems*, pp. 802–810, 2015.
6. K. Boonyuen, P. Kaewprapha, and P. Srivihok, "Daily rainfall forecast model from satellite image using convolution neural network," in *2018 IEEE International Conference on Information Technology*, pp. 1–7, 2018.