

THE COMPARISON BETWEEN RAINFALL OF TYPHOON MANGKHUT AND ITS WEIBO DISTRIBUTION

Xinyue Li*

NO.2 High School of East China Normal University

Just as our life experience has told us, weather forecasts can only be regarded as references but not the future, whereas modern meteorology is a thriving subject that hasn't been probed into. Take numerical weather forecast as an example, there have only been several decades since people began calculating and working out the prediction for weather variations. In this case, there still are a lot of internal laws of natural phenomenon that haven't been comprehended by researchers. What's more, the global climate is becoming more and more changeable, making it impossible for us to search every nook and cranny in the world for weather data despite those well-developed technologies. Therefore, we have to try our best to obtain information of weather through as many resources as possible. Social media, as a platform to which everybody can contribute their perspectives, serves as a suitable source of weather data.

The overall objective of this project is to determine how processed data from Weibo can work for the real-time distribution of rainfall during Typhoon Mangkhut. Specifically, this project collected about 130000 weibos which is sent in Guangzhou from the platform during 9/16/2018 00:00 ~ 9/17/2018 00:00 (24 hours in all) and introduced an operational infrastructure for processing data from Weibo. This project choose to use hierarchical attention networks for the document classification in Chinese weibo. This project uses the document-level classification, that is, the document is composed of the sentence, and the sentence is composed of the word, so the natural has a hierarchical relationship. The

word-level feature is used to extract the word-level feature of the network to obtain the feature vector representing the sentence; then the sentence-level vector is input into the network to extract the sentence-level feature to obtain the final document-level feature, and then the document feature is passed through a linear representation and softmax. In order to assign different weights to different words and different sentences, the paper designs a hierarchical architecture attention mechanism to improve the performance of the model. The result turned out to be 65% accuracy on a dev set (16% of data) after 3 epochs. Results reported in the paper are 71% on Yelp'15 and No systemic hyper parameter optimization was performed.

By processing the collected data, it is possible for us to visualize the result of the rain distribution form the data of social media. By comparing the result from social media and the real-time distribution from the authorized data, we can get to the conclusion that the social media can work for the fill of the blanks of Global Precipitation Measurement (GPM) data in worldwide range.

Furthermore, since this project mostly focus on the comparison between rainfall of Typhoon Mangkhut and its weibo distribution, we can aim on using social media in the weather forecast, which means we can collect the data other than text message about the specific event but the description of weather features like humidity, wind, visibility that are important for the prediction. With the help of social media, the modern meteorology can work better.

* *Corresponding author address:* Xinyue Li
NO.2 High School of East China Normal University,
555 Chenhui Road, Pudong, Shanghai,
201203, China; email: Sarah_L726@163.com