

# Chapter 18

## Rainstorm Forecasting By Mining Heterogeneous Remote Sensed Datasets

**Yu-Bin Yang**

*State Key Laboratory for Novel Software Technology, China*

**Hui Lin**

*The Chinese University of Hong Kong, Hong Kong*

### ABSTRACT

*This chapter presents an automatic meteorological data mining system based on analyzing and mining heterogeneous remote sensed image datasets, with which it is possible to forecast potential rainstorms in advance. A two-phase data mining method employing machine learning techniques, including the C4.5 decision tree algorithm and dependency network analysis, is proposed, by which a group of derivation rules and a conceptual model for metrological environment factors are generated to assist the automatic weather forecasting task. Experimental results have shown that the system reduces the heavy workload of manual weather forecasting and provides meaningful interpretations to the forecasted results.*

### INTRODUCTION

Meteorological satellite data and images have been operationally utilized in weather services for more than 30 years. Since the inception of weather forecasting based on satellite remote sensed data, meteorologists have faced the challenge of using this tool to minimize the potential damage caused during adverse weather conditions by collecting and analyzing these images. Based on this analysis, responsible officers are able to take necessary action to minimize the potential damage caused

by weather-related disasters. This is particularly significant and urgent for China, particularly for the Yangtze River Basin, which suffers from frequent flooding that endangers life, disrupts transportation and commerce and results in serious economic losses. For example, the unprecedented, severe flood in the Yangtze River Basin in 1998 resulted in the deaths of 4,150 people and damage to property of approximately 32 billion US dollars. Since almost all floods are caused by heavy rainfall, advance forecasting of these adverse weather conditions has been a key factor in attempts to mitigate casualties and damages caused by floods.

DOI: 10.4018/978-1-61520-757-2.ch018

Meanwhile, the study of “Mesoscale Convective Systems” (MCS) clouds, including the study of their life cycles, trajectories and evolution trends, remains an important and challenging issue in meteorology, by which severe weather situations, or, “mesoscale weather events” are often caused, such as strong winds, heavy rain falls, thunderstorms and hurricanes (Houze & Smull, 1990; Li & Plale, 2008).

In China, the MCS clouds over the Tibetan Plateau have recently been revealed to be the major factor in some disastrous weather phenomena occurring in the Yangtze River Basin. Meteorological observations have already shown that the eastward evolution trends of MCS clouds over the Tibetan Plateau is the crucial factor leading to heavy rain falls in the Yangtze River Basin, which may be the direct cause of severe floods in Southern China (Jiang & Fan, 2002). Therefore, it is vital that we discover the evolution trends of MCS over the Tibetan Plateau from the available meteorological satellite data (such as the meteorological environment parameters including temperature, wind divergence and water vapor flux divergence) and image collections, in order to effectively and efficiently predict and evaluate the potential occurrences of heavy rainstorms. The hidden correlations between the satellite data and the eastward evolution trends of MCS clouds may then be discovered and revealed from similar historical remote sensed scenarios, if accurate data analysis and mining can be made. Most importantly, the meteorological environment factors causing eastward evolution trends of MCS clouds can also be effectively modeled and explained, through which we may be able to forecast the future eastward evolution trends of MCS clouds and subsequently locate and forecast potential heavy rainstorms.

Unfortunately, meteorologists still performed tracking, characterization and analysis tasks of MCS manually on most occasions. This so-called “expert-eye-scanning” tracking technique is both time and labour intensive, as expert meteorolo-

gists work to identify the movement trajectories and evolution trends of MCS from the satellite remote sensed images (Arnaud & Desbios, 1992). Further, the quantity of satellite image data is so huge that this method is inadequate for the task of tracking MCSs in wide ranges and over long periods, despite its relatively high accuracy. It is too time-consuming, too ineffective, and the observation results are often unstable and vary between experts, which decreases the reliability and practicability of heavy rainfall forecasting. On the other hand, satellite remote sensed data used for this purpose have heterogeneous characteristics due to different data types, sensor properties and satellite specifications. Some data are infrared remote sensed images and the others are numeric observed or predicted data, describing different environmental properties of the activities of MCS clouds. Not only are the currently available data of different data types, but they also differ considerably in data quality, spatial resolution and temporal granularity. Since they have different spatial and temporal resolutions, it is necessary to model seamless integration of heterogeneous satellite data before MCS cloud analysis and mining is carried out.

To address the above issues, this chapter aims at presenting an automatic meteorological data mining approach based on analyzing and mining heterogeneous remote sensed image datasets, with which it is possible to forecast potential rainstorms in advance. Firstly, automatic MCS cloud detection and tracking methods are proposed to identify the geo-referenced cloud objects in satellite remote sensed images. Next, a data integration modeling mechanism is designed to extract meaningful properties of those detected clouds, by integrating the heterogeneous image data and observed data into a unified view. Finally, based on the integrated global data schema, a two-phase data mining method employing machine learning techniques, the C4.5 decision tree algorithm and dependency network analysis, is proposed to analyze and forecast the meteorological activities

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/rainstorm-forecasting-mining-heterogeneous-remote/42370](http://www.igi-global.com/chapter/rainstorm-forecasting-mining-heterogeneous-remote/42370)

## Related Content

---

### A Look back at the PAKDD Data Mining Competition 2006

Nathaniel B. Norieland Chew Lim Tan (2007). *International Journal of Data Warehousing and Mining* (pp. 1-11).

[www.irma-international.org/article/look-back-pakdd-data-mining/1780](http://www.irma-international.org/article/look-back-pakdd-data-mining/1780)

### Discovering Frequent Embedded Subtree Patterns from Large Databases of Unordered Labeled Trees

Yongqiao Xiaoand J. F. Yao (2005). *International Journal of Data Warehousing and Mining* (pp. 70-92).

[www.irma-international.org/article/discovering-frequent-embedded-subtree-patterns/1752](http://www.irma-international.org/article/discovering-frequent-embedded-subtree-patterns/1752)

### A State-of-the-Art in Spatio-Temporal Data Warehousing, OLAP and Mining

Leticia Gómez, Bart Kuijpers, Bart Moelansand Alejandro Vaisman (2011). *Integrations of Data Warehousing, Data Mining and Database Technologies: Innovative Approaches* (pp. 200-236).

[www.irma-international.org/chapter/state-ofa-art-spatio-temporal/53077](http://www.irma-international.org/chapter/state-ofa-art-spatio-temporal/53077)

### A Machine Learning-Based Lexicon Approach for Sentiment Analysis

Tirath Prasad Sahuand Sarang Khandekar (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 836-851).

[www.irma-international.org/chapter/a-machine-learning-based-lexicon-approach-for-sentiment-analysis/308522](http://www.irma-international.org/chapter/a-machine-learning-based-lexicon-approach-for-sentiment-analysis/308522)

### Image Classification of Crop Diseases and Pests Based on Deep Learning and Fuzzy System

Tongke Fanand Jing Xu (2020). *International Journal of Data Warehousing and Mining* (pp. 34-47).

[www.irma-international.org/article/image-classification-of-crop-diseases-and-pests-based-on-deep-learning-and-fuzzy-system/247919](http://www.irma-international.org/article/image-classification-of-crop-diseases-and-pests-based-on-deep-learning-and-fuzzy-system/247919)