ELSEVIER

Computers & Geosciences 33 (2007) 20–30

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

A data mining approach for heavy rainfall forecasting based on satellite image sequence analysis ☆

Yubin Yang[a,b,*], Hui Lin[b], Zhongyang Guo[b], Jixi Jiang[c]

[a]*State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China*
[b]*Joint Laboratory for Geoinformation Science, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong*
[c]*National Satellite Meteorological Center, China Meteorological Administration, Beijing 100081, China*

Received 3 March 2005; received in revised form 26 May 2006; accepted 26 May 2006

Abstract

Investigating the evolvement process of Mesoscale Convective Systems (MCSs) over the Tibetan Plateau using satellite remote sensing image sequence is a very important and effective method of forecasting heavy rainfall. This paper presents a spatial data mining approach, by which a possible heavy rainfall forecast can be made, based on MCS tracking using remote sensing satellite images. Firstly, an automatic method for object tracking from the satellite image sequence is proposed, aiming at identification of the qualified MCSs, their characteristics and their moving trajectories. Then, a novel two-phase spatial data mining framework is designed to enable the deduction of the correlations and causalities between MCS activities and possible heavy rainfall occurrences. The proposed approach proves to be capable of lifting the heavy burden of manual rainfall forecasting from the shoulders of meteorologists, by automatically analyzing and interpreting massive, meteorological remote sensing data sets to assist weather forecasting.

Keywords: Image sequence analysis; Spatial data mining; Remote sensing; Weather forecast; Meteorology

1. Introduction

Meteorological satellite data have been operationally used in weather services for more than 30 years. During this period, forecasting of severe weather based on satellite remote sensing data has been a challenging task. Early warnings of severe weather, made possible by timely and accurate forecasting will help prevent casualties and damages caused by natural disasters. This is particularly significant and urgent in China's Yangtze River Basin, which has so often suffered from flooding as to cause the flood control situation in China to become increasingly grave and urgent. For example, the unprecedented, severe flood in the Yangtze River Basin in 1998 resulted in the deaths of 4150 people and damage to property of approximately 32 billion US dollars. Since almost all floods are caused by intensive heavy rainfalls, the responsible authorities have a key and clear mandate to be able to

☆This work is supported by the National Natural Science Foundation of PR China under Grant No. 60505008 and the RGC grant from Hong Kong Research Grant Council under Grant No. CUHK4132/99H.

*Corresponding author. Joint Laboratory for Geoinformation Science, The Chinese University of Hong Kong, Shatin N.T., Hong Kong. Tel.: +852 2609 6010; fax: +852 2603 7470.

E-mail address: yangyubin@ieee.org (Y. Yang).

0098-3004/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2006.05.010

provide both accurate and advance forecasting of possible heavy rainfall.

Meanwhile, the study of the life cycles, moving trajectories and evolvement trends of Mesoscale Convective Systems (MCSs) remains a challenging and important issue facing the meteorological community, because these phenomena often cause severe weather such as heavy rainfalls, thunderstorms and hurricanes (Houze et al., 1990). In China, the MCSs over the Tibetan Plateau were recently revealed as a major factor resulting in heavy rainfalls occurring in the Yangtze River Basin, which directly caused the severe floods in South China, such as in 1991, 1994, 1998 and 1999 (Jiang and Fan, 2002).

In order to improve on current severe flood control measures in China, it is necessary to accurately track and characterize the active MCSs over the Tibetan Plateau, using the satellite remote sensing images. A meteorological analysis of all MCSs can then be performed, by taking into account their environmental, physical variables, such as temperature, wind divergence and water vapor flux divergence. As a result, the correlations and causalities between the MCS evolvement process and heavy rainfall occurrences can be deduced from the historical remote sensing scenarios, and be represented as known knowledge to assist prediction of potential occurrences of heavy precipitation.

Unfortunately, meteorologists continue to manually track, characterize and analyze MCSs. In this, so-called, "expert-eye-scanning" technique, meteorologists carry out extensive manual work to discover the moving trajectories and evolvement trends of MCSs from the satellite remote sensing images, using their professional experience and knowledge (Arnaud et al., 1992). However, the volumes of satellite image data can be huge, making this method inadequate for tracking MCSs covering wide ranges and long time periods. The method is time consuming, ineffective and often yields unstable and variable results from the different experts involved, affecting the reliability and practicability of heavy rainfall forecasting.

To address the above problems, this paper aims to provide meteorologists with an automatic, spatial data mining method based on MCS tracking and analysis in the satellite image sequence, with which possible heavy rainfalls can be predicted so that effective flood control measures can be taken. The basic principle behind the method is formulated from a recent observation that the eastward movement and propagation of MCSs over the Tibetan Plateau is the crucial factor leading to the heavy rainfalls in the Yangtze River Basin (Jiang and Fan, 2002). The method seeks to model and uncover the latent patterns of MCS activities and their evolvement trends over the Tibetan Plateau, using data mining and knowledge discovery techniques. Firstly, using the image sequences of Temperature of Black Body (TBB) acquired from the Geostationary Meteorological Satellite (GMS-5), the qualified MCSs are automatically identified and tracked by image processing and computer vision techniques. An automatic object-tracking approach of investigation of moving MCS trajectories is developed for this purpose. Then, the High-resolution Limited Area Analysis and Forecasting System (HLAFS) remote sensed data around the geographical location of each MCS are used and a novel two-phase spatial data mining process adopting the C4.5 decision tree algorithm (Quinlan, 1993) is integrated to discover the hidden knowledge that helps reveal the correlations and causalities between the moving trajectories of MCSs and the observed environmental, physical variables. The discovered knowledge is represented in two forms, i.e., derivation rules and environmental, physical model graphs, which can reveal evolvement trends of the MCSs causing heavy rainfall occurrences in the Yangtze River Basin.

The rest of this paper is organized as follows. Firstly, in Section 2, the satellite data sources used in the study are introduced. Section 3 presents a framework of the proposed meteorological data mining approach. MCS tracking techniques and MCS data mining techniques are detailed in Sections 4 and 5, respectively. The experimental results are illustrated in Section 6. Finally, concluding remarks are provided in Section 7.

2. Data sources

To track MCS and discover useful information and knowledge crucial to heavy rainfall forecast, the collection of large amounts of satellite data with high spatial and temporal resolutions is indispensable. For this purpose, satellite remote sensing images of the TBB data, taken by the GMS-5 satellite, and data of the HLAFS were provided by the National Satellite Meteorological Center, China Meteorological Administration, for use in this study. The data cover the time period from June