SPATIO-TEMPORAL DATA MINING ON MCS OVER TIBETAN PLATEAU USING SATELLITE METEOROLOGICAL DATASETS

Yu-Bin Yang[1,2], *Hui Lin*[2]

1. State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
2. Institute of Space and Geoinformation Science, The Chinese University of Hong Kong, Hong Kong, China

1. MOTIVATION

This paper 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 of all clouds in order to discover the hidden correlations between the satellite observed data and the eastward evolvement trends of MCS clouds. Moreover, the meteorological environment factors that may cause the eastward evolvement trends of MCS clouds are also analyzed and conceptually modeled. An MCS Cloud Analysis System (MCAS) is also successfully designed and implemented by applying the above approaches, through which it is possible for meteorologists to forecast potential heavy rainstorms more easily.

2. MCS IDENTIFICATION AND TRACKING

Since the satellite image data are spatio-temporal, all MCS clouds need to be detected and tracked correctly and efficiently from complete image sequences. Subsequently, it is necessary to extract some representative attributes of the tracked MCS from the corresponding data collections in order to characterize them. To address the above problems, we propose a fast tracking and characterization method of multiple moving clouds from meteorological satellite images based on feature similarities. The method is based on the fact that in a relatively small time-span, the deformation of an MCS is progressive and detectable, which means that at two consecutive satellite images the same MCS will maintain a relatively similar moving speed, shape, area and texture.

The complete flow of our method is described as follows. Firstly, image processing techniques are applied on the satellite remote sensed images to make MCS segmentation, by which all the MCS structures in the images are labeled. Then, a group of important features of each MCS are extracted as its representation, based on which feature similarities of different MCS in consecutive images are computed to track the same MCS along the time axis. Finally, in the characterization stage, the qualified MCSs are categorized into four types according to their evolvement trends indicated on the satellite images: (1) MCS moving East (E) out of the Tibetan Plateau; (2) MCS moving Northeast (NE) out of the Tibetan Plateau; (3) MCS moving Southeast (SE) out of the Tibetan Plateau; and (4) MCS staying in the Tibetan Plateau (STAY-IN). The categorization is implemented by computing the direction angle values from the movement trajectory of each MCS, and is crucial to meteorologists for predicting and evaluating the potential occurrences of heavy rainfalls.

3. MCS DATA MINING

Due to the heterogeneity of our data sources, appropriate attributes of each MCS, including TBB value, HLAFS values and feature values, should be firstly integrated to constitute a consistent input dataset serving for data mining purposes. We designed a data integrator to integrate the relevant data of TBB images and HLAFS variables into a unified dataset. Then, we implemented a spatial data mining approach directly from the integrated dataset to yield the meteorological knowledge.

Firstly, a data mining process is designed and implemented by adopting the C4.5 decision tree algorithm [1], outputting a set of knowledge primitives represented as decision rules. The aim of the data mining step is to analyze the meteorological activities of MCS clouds and further discover the hidden correlations between the satellite data and the eastward evolvement trends of MCS clouds, which is crucial for subsequent rainstorm forecasting. Secondly, we proposed an MCS conceptual modeling method based on dependency network technique [2], in order to select out the relevant HLAFS variables which powerfully influence the evolvement trends and moving trajectories of MCS clouds. Then, a set of environmental model graphs can be established and plotted to illustrate how the selected attributes interact with each other and finally influence the evolvement trends of MCS.

It should be noted that the C4.5 rule is based on an individual MCS cloud. We can predict the evolvement trend of each newly identified MCS cloud from the mined decision rules. The conceptual model is different from those rules. It targets the attributes in the integrated dataset where all the correlated attributes of the same type of MCS clouds are selected from the generated dependency network. Their values are then spatially averaged to provide a meteorological relevance analysis for the MCS evolvement trends. Both of them are integrated and complementary components for analyzing the evolvement trends of MCS and evaluating their influence on heavy rainfalls.

4. RESULTS

The experimental data is taken from the satellite remote sensed images of Geostationary Meteorological Satellite (GMS-5) over Tibetan Plateau in the summer of 1998. The weather satellite GMS-5 transmits infrared remote sensed images every hour. From June to August of 1998, there are totally (30+31+31)×24=2,208 images. Firstly, we compare our MCS tracking approach with the area-overlapping tracking method proposed by Arnaud et al. [3]. The experimental results show that the cloud tracking accuracy of our method creates an average 17% enhancement over the tracking method proposed by Arnaud et al, and is very close to the tracking accuracy of meteorologists.

Experiments are designed and implemented to test the proposed data mining method and generated conceptual model. After the MCS tracking and characterization stage, there are totally 320 MCSs that qualify for data mining step, among which 50 clouds have moved out of the Tibetan Plateau (105^0E): 37 clouds for "E", 9 clouds for "NE" and 4 clouds for "SE". 70% of all the identified MCSs, i.e. 224 clouds, are used as train samples and the remaining 30% are kept for testing. A set of decision rules for predicting MCS evolvements are generated by using the C4.5 decision tree algorithm and a set of meteorological environment models corresponding to the relevant HLAFS attributes are also automatically constructed and plotted according to dependency network techniques.

The experimental results indicate that it is feasible to model and predict evolvement trends of MCSs on the Tibetan Plateau based on their meteorological environment attribute (HLAFS data) values and TBB images from the satellite databases. Moreover, our data mining approach may help meteorologists reveal the hidden connections of MCS evolvements to heavy rainfalls occurring in China's Yangtze River Basin.

5. ACKNOWLEDGMENTS

This research has been funded by the National Natural Science Foundation of P.R. China under Grants 60875011 and 60505008, the Natural Science Foundation of Jiangsu Province under Grant BK2007520, the Key Program of the National Natural Science Foundation of P.R. China under Grant 60723003 and the RGC grant from Hong Kong Research Grant Council under Grant CUHK4132/99H. The authors also thank their collaborators in the National Satellite Meteorological Center of China Meteorological Administration and at the Hong Kong Observatory, for their stimulating discussion on domain-specific knowledge and provision of data sources.

REFERENCES

[1] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufman, San Francisco, CA, 1993.

[2] D. Heckerman, D. Chickering, C. Meek, R. Rounthwaite, and C. Kadie, "Dependency Networks for Inference, Collaborative Filtering, and Data Visualization", *Journal of Machine Learning Research*, 1(9), pp. 49-75, 2000.

[3] Y. Arnaud, M. Desbios, and J. Maizi, "Automatic tracking and characterization of African convective systems on meteosat pictures", *Journal of Applied Meteorology*, 31(5), pp. 443-453, 1992.