Application of Association Rule in Disaster Weather Forecasting

Zhongyang Guo[1,2], Xiaoyan Dai[1,2],and Hui Lin[3]

[1]Laboratory of Geographic Information Science, East China Normal University, Ministry of Education, Shanghai 200062, China
[2]Department of Geography,East China Normal University,Shanghai 200062,China
[3]Department of Geography and Resource Management & Joint Laboratory for GeoInformation Science
The Chinese University of Hong Kong,Hong Kong,China

Abstract

Recently, the evidences have indicated that the heavy rainfall in Yangtze River Basin is directly caused by Mesoscale Convective System (MCS) over the Tibetan Plateau in China. In this paper, the trajectories of MCS over the Tibetan Plateau are automatically tracked using GMS (Geostationary Meteorological Satellite) brightness temperature (Tbb) and High Resolution Limited Area Analysis and Forecasting System (HLAFS) data provided by China National Satellite Meteorological Center from June to August 1998. Based on these, the relationships between the trajectories of MCSs moving out of the Plateau and their environmental physical field values are analyzed using spatial association rule mining technique. The results indicate that at the level of 400hPa, the trajectories of MCSs, which move out of the Plateau, are mainly influenced by geopotential height, relative humidity, vorticity, divergence and vertical wind speed, while at the level of 500hPa, geopotential height, relative humidity, temperature, vertical wind speed and K index are the main factors which influence MCS to move out of the Plateau.

I. INTRODUCTION

With the development of technology obtaining spatial information, different spatial databases, which are related with meteorological phenomena, are expanding quickly. However, a lot of useful information hidden in these databases cannot be found using traditional methodology. Consequently, how to extract the useful information, which is related to disaster weather, from these databases and improve the accuracy of forecasting disaster weather has been regarded as one of the most important research aspects. On the other hand, spatial data mining, which can extract useful information and knowledge from amounts of data or large databases, is developed quickly (Frawley et al., 1991; Fayyad et al., 1993; Koperski and Han, 1995; Abraham and Roddick, 1998; Malerba et al., 2001; Aggarwal and Philip, 2002). Therefore, it can be used to solve the difficult problem existing in meteorology. For example, Lee and Liu (1999; 2000) developed the ATOMOSPHER system. In this system, 120 TC (Tropical Cyclone) cases in the period from 1985 to 1998 were used to test the effectiveness of the methods, and the results indicated that the system achieved a 97% accuracy rate for correct classification and an 86% correct prediction rate in TC tracking. Furthermore, Kitamoto (2001) used 34,000 typhoon images as a test bed to analyze other typhoon images and discover the statistical properties of typhoon cloud patterns from these images. Zhou et al. (1999) developed new neural network models to increase knowledge about typhoons, using related images. However, at present, forecasting disaster weather using spatial data mining technique is still at an initial stage.

Recently, the evidences have indicated that the heavy rainfall in Yangtze River Basin is directly caused by Mesoscale Convective System (MCS) over the Tibetan Plateau in China (Shi et al., 2000; Jiang and Fan, 2002). Consequently, it is important to abstract the environmental physical field characteristics which influence MCS to move out of the Plateau and develop the relationships between the trajectories of MCSs and their environmental physical field values using spatial data mining technique. Thus, not only can the trajectories rules of MCS over the Plateau be revealed, but also the accuracy of forecasting disaster weather can be improved and the damage caused by disaster weather can be decreased.

In this project, association rule mining, a spatial data mining technique, is used to determine the relationships between the trajectories of MCSs and their environmental physical field values over the Tibetan Plateau. Results indicate that it is feasible to predict the trajectories of MCSs based on their environmental physical field values over the Plateau. Therefore, it is of great value in revealing the thermal and dynamic actions over the Plateau and improving the accuracy of forecasting storm and strong convective weather using spatial data mining technique.

II. METHODS AND DATA

GMS (Geostationary Meteorological Satellite) brightness temperature (Tbb) and High Resolution Limited Area Analysis and Forecasting System (HLAFS) data provided by China National Satellite Meteorological Center from June to August 1998 are used in this project. Among these data, the spatial resolution of Tbb is 0.5° lat ×0.5°long, and the time resolution of Tbb is one hour. And, the spatial resolution of HLAFS is 1° lat × 1° long and the time resolution of HLAFS is twelve hours.

1082-4006/04/1001-68$5.00

On the other hand, the study area is from 27°N to 40°N, 80°E to 105°E, and the study levels include 400hPa and 500hPa. Furthermore, HLAFS values, including geopotential height, temperature, vorticity, divergence, water vapor flux divergence, vertical wind speed, pseudo-equivalent potential temperature, K index and relative humidity are used in this project. The study focuses on MCS that cover at least 3 connected pixels having Tbb $\leq$ -32°C in each Tbb image, and last for at least 3 consecutive hours.

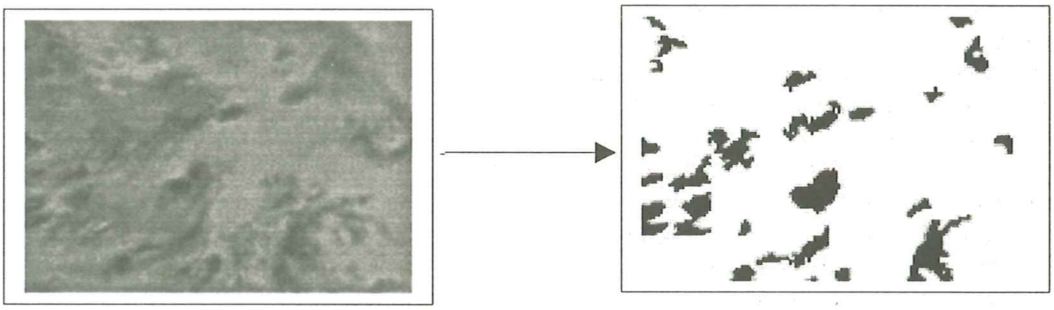
Figure 1(a) shows the meteorological satellite cloud picture over the Tibetan Plateau at 2UTC on 11 August 1998. Figure 1(b) shows the distribution of MCS that are extracted from Figure 1(a) and satisfy the conditions mentioned above. It can be found that there are twenty-seven MCSs over the Plateau at this UTC. What is more, for each MCS, it can also be found that the shape and area are various. On the other hand, Figures 2 (a) ~ (c) show the distribution of MCS over the Tibetan at 4UTC, 5UTC and 6UTC on 13 August 1998,

respectively. From it, for each MCS, it can be seen that the movement speed and movement direction is different because the area and shape of each MCS is different. Consequently, two hours later, the area and shape of each MCS have changed in a different way. Furthermore, some MCSs have been combined, while some MCSs have disappeared. Therefore, it is important to predict MCS trajectories through abstracting the environmental physical field conditions that influence MCS movement.

III. DATA PREPROCESSING

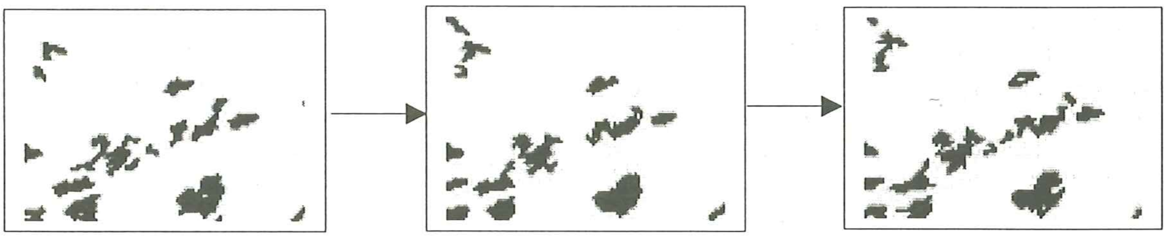
In the course of data preprocessing, MCS is firstly tracked using the method provided by Arnaud et al. (1992). And, in this project, if an MCS moves across 105^{0}E, then the MCS is defined as 'move out of the Plateau', otherwise, it is defined as 'stay in'. Results indicate that there are 749 MCSs over the Tibetan Plateau from June to August 1998. While the total number of MCS, which move out of the Plateau, is 55. Among these, the number of MCSs moving out of the Plateau and to the east (E), south- east (SE) and north- east (NE), is 41, 5 and 9, respectively. Therefore, the ratio of moving to E, SE and NE, is 74.5%, 9.1% and 16.4%, respectively. Figure 3 shows the trajectories of MCS moving out of the Plateau in August 1998, it can be seen that the initiation locations of most of MCSs are between 100^{0}E and 105^{0}E. Consequently, in the course of association rule mining, the initiation location of each MCS is defined near 100^{0}E.

On the other hand, the shape of MCS is also a main factor to influence MCS movement. Therefore, in this project, ellipse equation using least squares method is used to determine the shape of MCS. And, the shape of MCS is classified into three types according to the ratio of long axis length and short axis length, i.e., if the ratio is [0.9,1.0] and [0.7,0.9), then the shape of MCS is defined as circle and ellipse, otherwise, it is defined as others.



(a) (b)

Figure 1. Meteorological satellite cloud picture and distribution of MCS over the Tibetan Plateau at 2UTC on 11 August 1998



(a) (b) (c)

Figure 2. Charts of trajectories of MCS over the Tibetan Plateau at 4~6 UTC on 13 August 1998

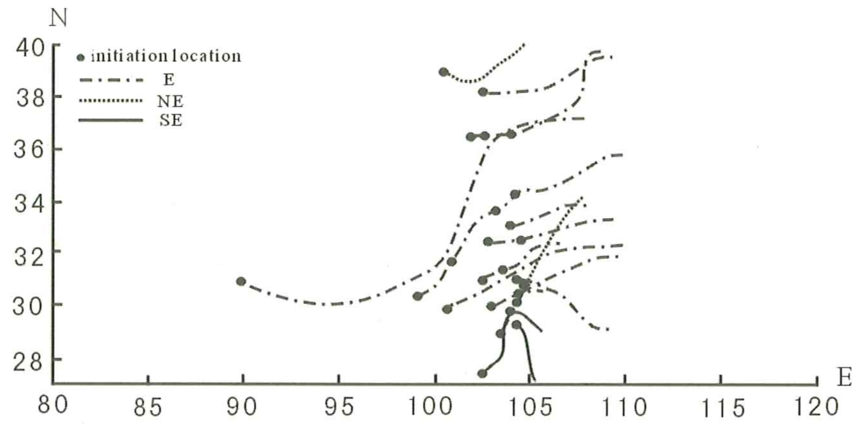


Figure 3. The trajectories of MCSs moving out of the Tibetan Plateau in August 1998

IV. METHODOLOGY

Usually, association rule can be represented as follows, $X \rightarrow Y(s, c)$, where X is a set of attribute values that determine the attribute Y, s is the support of the rule, it represents that the existence of $X \rightarrow Y$ is s% of records in the dataset, c is the confidence of the rule, it means that there are c% of the records in the dataset that contain X and also contain Y. However, in spatial association rule mining, at least one of the attribute values is a spatial attribute value. Furthermore, in this project, in the course of spatial association rule mining, $X \rightarrow Y$ is constrained that Y must be the attribute 'MCS movement direction'.

Now, an MCS is centered at (x_1, y_1) when it moves to near 100°E at the time $UTC=H_m$. Choose the HLAFS at $UTC=H_h$ (either equals to 00,12 or 24) where H_h is the nearest 00/12/24 UTC to H_m. If the MCS is centered at (x_2, y_2) at the time $UTC=H_m+2$, then three areas (A, B, C), which are near the MCSs at the time $UTC=H_m+2$, will be defined according to the direction in which the MCS is moving. Each area (A, B, C) is 3°lat×1°long. For each of the 9 HLAFS attributes, the average for A, B and C will be found respectively, then $D_{b\text{-}a}$ and $D_{b\text{-}c}$ are defined as the difference of the average in areas B and A and that of areas B and C respectively (see Figure 4).

V. DISCUSSION

In this project, the parameters, Geopotential height (H)$\rightarrow 10^{-1}$m, Temperature (T)$\rightarrow$°C, Relative humidity (RH)$\rightarrow$%, Vorticity (VOR)$\rightarrow 10^{-6}s^{-1}$, Divergence (DIV)$\rightarrow 10^{-6}s^{-1}$, Vertical wind speed ($\omega$)$\rightarrow 10^{-5}hPa{\cdot}s^{-1}$, Water vapor flux divergence (IFVQ)$\rightarrow 10^{-10}g/cm^2h{\cdot}Pa{\cdot}s$, Pseudo-equivalent potential temperature (θSE)$\rightarrow$°C, K index (K), area $\rightarrow km^2$, the average lowest temperature tbb $\rightarrow$ °C, position (longitude and latitude) and shape of MCS are included.

On the other hand, in the course of spatial association rule mining, HLAFS values are divided into ten segments between their maximum values and minimum values. Furthermore, ADT 1.0 software is used in this project.

Table 1 and Table 2 show the association rules of influencing MCS movement out of the Tibetan Plateau and to the east at the levels of 400hPa and 500hPa respectively. However, the association rules, which influence MCS to move to the east-south and east-north, are not found.

In Table 1 and Table2, H_{ba} means the difference of geopotential height in area B and area A, E (4,1.0,0.018) means the number of MCS, which satisfies this rule, is 4. And, the movement direction is to the east, the confidence is 1.0, the support is 0.018. Other represents are similar.

From Table 1, it can be seen that at the level of 400hPa, the MCS trajectories, which move out of the Plateau and to the east, is not related with temperature, average lowest temperature, area and location of MCS. Furthermore, the association rules, which the confidence is 1.0, are mainly related with geopotential height, relative humidity, vorticity, divergence, vertical velocity and K index. By and large, at this level, the MCS which move out of the Plateau and to the east are mostly

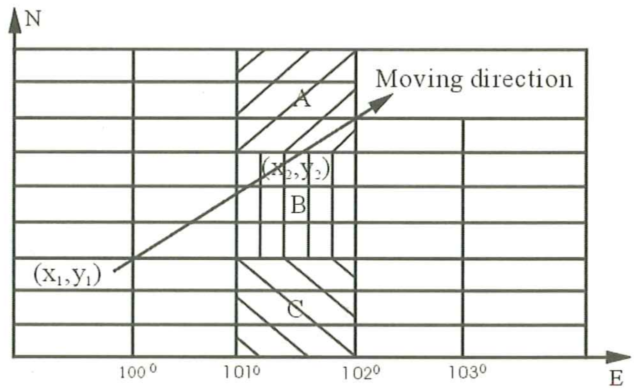


Figure 4. Interesting area of spatial association rule mining

Table 1. The association rules of influencing MCS trajectories in 400hPa level

Index	Rules
1	$H_{ba}=28\sim35.8\Lambda RH_{ba}=22.92\sim32.38\Lambda DIV_{bc}=-10.5\sim-2.8\rightarrow E(4,1.0,0.018)$
2	$H_{ba}=12.4\sim20.2\Lambda H_{bc}=-11.6\sim-5.7\Lambda DIV_{ba}=0.5\sim7.0\Lambda DIV_{bc}=-2.8\sim4.9\rightarrow E(4,1.0,0.018)$
3	$H_{bc}=-29.3\sim-23.4\Lambda VOR_{bc}=-13.6\sim-1.5\Lambda \omega_{ba}=-138\sim-37.4\rightarrow E(3,1.0,0.013)$
4	$RH_{bc}=-23.4\sim-13.8\Lambda VOR_{ba}=3.6\sim15.7\Lambda DIV_{bc}=-2.8\sim4.9\rightarrow E(3,1.0,0.013)$
5	$VOR_{ba}=3.6\sim15.7\Lambda DIV_{bc}=-2.8\sim4.9\Lambda K_{ba}=6\sim12.5\rightarrow E(3,1.0,0.013)$
6	$DIV_{bc}=-10.5\sim-2.8\Lambda \omega_{bc}=-117.6\sim-9.0\Lambda IFVQ_{bc}=-58.8\sim-30.5\rightarrow E(4,0.8,0.018)$
7	$RH_{bc}=-4.2\sim5.4\Lambda VOR_{ba}=-8.5\sim3.6\Lambda DIV_{bc}=-10.5\sim-2.8\Lambda \theta SE_{bc}=0.4\sim2.3\Lambda K_{bc}=1\sim15\rightarrow E(3,0.75,0.013)$
8	$DIV_{bc}=-6\sim0.5\Lambda \omega_{bc}=-117.6\sim-9\Lambda$Shape is others$\rightarrow E(5,0.71,0.022)$
9	$H_{bc}=-11.6\sim-5.7\Lambda DIV_{ba}=0.5\sim7.0\Lambda DIV_{bc}=-2.8\Lambda 4.9\rightarrow E(5,0.71,0.022)$

Table 2. The association rules of influencing MCS trajectories in 500hPa level

Index	Rules
1	$RH_{ba}=-9.66\sim-1.39\Lambda VOR_{ba}=-5\sim10.8\Lambda \omega_{ba}=27.8\sim130.1\rightarrow E(4,1.0,0.018)$
2	$H_{ba}=13\sim19\Lambda RH_{ba}=-5\sim10.8\Lambda \omega_{ba}=27.8\sim130.1\rightarrow E(4,1.0, 0.018)$
3	$H_{ba}=13\sim19\Lambda T_{bc}=5.6\sim12.3\Lambda RH_{ba}=-9.66\sim-1.39\Lambda K_{bc}=1\sim15\rightarrow E(3,1.0,0.013)$
4	$H_{ba}=13\sim-19\Lambda RH_{bc}=-0.02\sim7.36\Lambda \omega_{bc}=-110.9\sim4.8\Lambda K_{bc}=1\sim15\Lambda$ Shape is others$\rightarrow E(3,1.0, 0.013)$
5	$T_{ba}=34.3\sim40\Lambda \omega_{ba}=-176.8\sim-74.5\Lambda K_{bc}=-13\sim1\rightarrow E(3,1.0, 0.013)$
6	$RH_{bc}=15.15\sim23.42\Lambda$The lowest $Tbb=-48.8\sim-44.2\rightarrow E(4,0.8, 0.018)$
7	$H_{ba}=13\sim19\Lambda RH_{ba}=-9.66\sim-13.9\Lambda DIV_{bc}=-4\sim2.8\rightarrow E(3,0.75, 0.013)$
8	$H_{ba}=13\sim19\Lambda T_{bc}=-27.9\sim-21.2\Lambda K_{ba}=-0.5\sim6\rightarrow E(3,0.75, 0.013)$
9	$RH_{ba}=15.15\sim23.42\Lambda VOR_{bc}=-42\sim-30.9\rightarrow E(3,0.75, 0.013)$

influenced by geopotential height, relative humidity, vorticty, divergence and vertical velocity. On the other hand, the support of each association is relatively small, it is only 0.02 or so, all the same, association rule mining results indicate that it is feasible to forecast the trajectories of MCSs based on their environmental physical field values.

Table 2 shows the association rules that influence MCS movement out of the Plateau at the level of 500hPa. From it, it can be found that the MCS trajectories, which move out of the Plateau and to the east, are not related with θSE, water vapor flux divergence, location and area of MCS. Furthermore, the trajectories of MCS are less influenced by vorticity, divergence, average lowest temperature and shape of MCS. By and large, at this level, the trajectories of MCS are mainly influenced by geopotential height, relative humidity, temperature, vertical velocity and K index. On the other hand, at this level, the support of each association rule is similar to that of 400hPa, the values are between 0.01 and 0.02.

VI. CONCLUSION

In this project, spatial association rule mining technique is used to extract the information of influencing MCS movement out of the Plateau, the results indicate that it is feasible to determine the relationships between the trajectories of MCS and their environmental physical field values using spatial association rule mining technique. However, the factors which influence MCS to move out of the Plateau are very complex. Therefore, in future, other spatial data mining techniques, such as classification rule mining, clustering rule mining, characteristic rule mining, et al., will be used to analyze the relationships between the trajectories of MCS and their environmental physical field values.

ACKNOWLEDGEMENT

The work described in this paper was partially supported by the National Natural Science Foundation of China (Project No. 40371080), Key Foundation Supported by Ministry of Education (Project No. 104083), and Foundation of WuHan University State Key Laboratory of Information Engineering in surveying, mapping and Remote Sensing (Project No. WKL (03) 0103).

REFERENCE

[1] Abraham, T. and Roddick J. F., 1998, Opportunities for knowledge discovery in spatio-temporal information systems. *Australian Journal of Information System,* 5(2): 3-12.

[2] Aggarwal, C. C. and Philip, S. Y., 2002, Redefining clustering for high-dimensional applications. *IEEE Transactions on Knowledge and Data Engineering,* 14(2): 210-225.

[3] Arnaud, Y, Michel, D., and Joel, M., 1992, Automatic Tracking and characterization of African Convective Systems on Meteosat Pictures. *Journal of Applied Meteorology*, 31(5): 443-453.

[4] Fayyad, U., Weir, N., and Djorgovski, S., 1993, Automated cataloging and analysis of ski survey image databases: the SKICAT System. In *Proc. of the Second Int. Conf. on Information and Knowledge Management*, Washington DC, 527-536.

[5] Frawley, W., Piatetsk, S. G., and Matheus, C. J., 1991, *Knowledge Discovery in Databases: An Overview. in Knowledge Discovery in Databases*. MIT Press.

[6] Jiang, J. X. and Fan, M., 2002, Convective clouds and mesoscale systems over the Tibetan Plateau in summer. *Journal of Atmospheric Sciences (in Chinese)*, 26: 263-270.

[7] Kitamoto, A., 2001, Data mining for typhoon image collection. *Proceedings of the 2nd International Workshop on Multimedia Data Mining*, pp. 68-77.

[8] Koperski, K., Han, J., 1995, Discovery of spatial association rules in geographic information databases. In Egenhofer M J and Herring J R(Eds.): *Advances in Spatial Databases*, LNCS 951, Springer-Verlag, pp. 47-66.

[9] Lee, R. S. T. and Liu, J. N. K., 1999, An automatic satellite interpretation of tropical cyclone patterns using elastic graph dynamic link model. *IJPRAI*, 13(8): 1251-1270.

[10] Lee, R. S. T. and Liu, J. N. K., 2000, ATMOSPHERE - Automatic Track Mining and Objective Satellite Pattern Hunting System Using Enhanced RBF and EGDLM. *AAAI/IAAI*, pp. 603-608.

[11] Malerba, D., Esposito, F., and Lisi, F. A., 2001, Mining spatial association rules in census data. In *Proc. 4th Int. Seminar on New Techniques and Technologies for Statistics* (NTTS 2001), Crete, pp. 18-22.

[12] Shi, C. X., Jiang, J. X., and Fang, Z. Y., 2000, A study on the features of severe convective cloud clusters causing serious flooding over Changjiang River Basin in 1998. *Climate and Environmental Research (in Chinese)*, 5(3): 279 -286.

[13] Zhou, Z., Chen, S., and Chen, Z., 1999, Mining typhoon knowledge with neural networks. In *Proc. 11th Int. Conf. on Tools with Artificial Intelligence*, IEEE, pp. 325-326.