ISSN:0975-9646

Kavita Pabreja / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (1) , 2012, 2996 - 2999

Clustering technique to interpret Numerical Weather Prediction output products for forecast of Cloudburst

Kavita Pabreja

Assistant Professor, Maharaja Surajmal Institute (an affiliate of GGSIP University), New Delhi, India
Research Scholar, Birla Institute of Technology and Science, Pilani, Rajasthan, India

***Abstract*— With the advent of digital computers and their continuous increasing processing power, the 'Numerical Weather Prediction' (NWP) models which solve a close set of equations of atmospheric model, have been adopted by most of the meteorological services to issue day to day weather forecasts. These forecasts are issued for public in general. But there are many limitations inherent to this technique viz. the actual weather event cannot be predicted directly by these models , so statistical regression techniques *viz.* Model Output Statistics (MOS) are used to derive the weather phenomenon from the NWP output products which itself requires long term consistent series of model forecasts. Due to the frequent revisions of the models, the long-term series of forecasts is not available. There is thus a strong need for searching alternative tools to MOS for interpretation of weather patterns provided by NWP models. Data mining is one such alternative that has been applied in this paper to interpret the forecast provided by European Center for Medium-range Weather Forecasting (ECMWF) model so as to infer the formation of cloudburst in advance.**

***Keywords*— Numerical Weather Prediction, Clustering, Cloudburst, Relative Humidity, Temperature, Convergence**

I. INTRODUCTION

Clouds are composed of microscopic droplets of liquid water (warm clouds), tiny crystals of ice (cold clouds), or both (mixed phase clouds). Cloud drops initially grow by the condensation of water vapor onto the drop when the super saturation of an air parcel exceeds a critical value [1]. In warm clouds, larger cloud droplets fall at a higher terminal velocity because the drag force on smaller droplets is larger than on large droplets. The large droplet can then collide with small droplet and combine to form even large drops. When the drops become large enough so that the acceleration due to gravity is much larger than the acceleration due to drag, the drops can fall to the earth as precipitation.

When the air reaches saturation and some of the resulting droplets come in contact with freezing nuclei (assuming they have reached the activation temperature) at atmospheric pressure level of 400hPa and 500hPa[2] causing combination of ice crystals and supercooled water droplets. From the perspective of the supercooled droplets, the air is in equilibrium at saturation, but from the perspective of the ice crystals, the air is supersaturated. Therefore, water vapor will sublimate on the ice crystals and the ice crystal grows through sublimation at the expense of the supercooled water droplet. At lower atmospheric pressure levels i.e. at 700hPa, when the temperature is high the clouds cannot sustain and fall.

In this paper, an attempt has been made to analyze the temperature at atmospheric pressure levels 400hPa, 500hPa and 700hPa; and also relative humidity at 850hPa and 925hPa in an area of 2.75° X 2.75° (approximately 275km X 275km) around the location of a real life case of cloudburst in Dhaka, Bangladesh. The flow pattern data used in the study is based on the output products of ECMWF model. The important ingredients to cloudburst formation are primary variables forecasted by the model *viz.* temperature and relative humidity, considered in this paper and a study has already been done that is based on derived parameters [3].

II Literature Review

There are many studies that support the applicability of Data mining techniques viz. **clustering, association rule mining and classification techniques** for weather forecasts. The use of **clustering** [4] is driven by the intuition that a climate phenomenon is expected to involve a significant region of the ocean or atmosphere, and that it is expected that such a phenomenon will be 'stronger' if it involves a region where the behavior is relatively uniform over the entire area. Shared Nearest Neighbor (SNN) clustering [5] has shown to find such homogeneous clusters. Based on **k-nearest neighbor approach**,[6] a quantitative snowfall forecast model has been developed for a station in Jammu and Kashmir, using surface meteorological data of the past 12 winters. The model predicts weather in terms of snow/no snow day and the amount of snowfall (snowheight in cm) for three consecutive days in advance. **Association analysis** has been used to derive spatio-temporal relationship hidden in earth science data [7].

The technique of **Sequential Pattern Mining** has been used in [8] where the mining of frequently occurring patterns related on time-axis is done so as to forecast tornado. **Spatio-temporal data mining** of large geophysical datasets has been experimented [9] by design, implementation and application of CONtent-based Querying in Space and Time (CONQUEST) that has been built under the auspices of National Aeronautics and Space Administration (NASA) High Performance Computing and Communications (HPCC) program. Application of the system has been demonstrated for the detection of blocking features and cyclone. Spatial data mining technique based on the **aggregation techniques** and **spatial relations** has been used by authors [10] to finding pressure trough features in weather data sets. The trajectories of Mesoscale Convective system (MCS) over Tibetan Plateau in China, are automatically tracked using Geostationary Meteorological Satellite (GMS) brightness temperature and High Resolution Limited Area Analysis And Forecasting System (HLAFS) data provided by China National Satellite Meteorological Centre from June to August 1998[11]. Based on these, the relationship between the trajectories of MCSs moving out of the plateau and their environmental physical field values are analyzed by authors using **spatial association rule mining** technique. Most of these weather events are either mesoscale or synoptic scale, but none of the studies are meant for the disastrous sub-grid scale weather events like cloudbursts by applying data mining techniques.

III DATASETS USED

The forecasts made by ECMWF model (T-799) for the weather variables Relative Humidity and temperature has been considered. These forecasts provided by Indian Meteorological Department that has been used for analysis is as follows:

- Forecast made on 0000GMT 28 July 09, valid for 1800GMT 28 July 09.
- Forecast made on 0000GMT 28 July 09, valid for 0000GMT 29 July 09.

An attempt has been made to analyse the temperature at atmospheric pressure levels 400hPa, 500hPa and 700hPa; and also relative humidity at 850hPa and 925hPa in an area of 2.75° X 2.75° around the location of cloudburst in Dhaka. The cloudburst occurred between 1:00am and 7:00am on 29 July, 2009 at Dhaka, Bangladesh (23.5°N and 90.25°E), so the forecast valid for 1800GMT 28July 09 and 0000GMT 29July 09 has been considered. The area under consideration is from 20.75°N, 87.5°E to 26.25°N, 93°E. The output product of ECMWF model is in GRIB format and has been converted to (.csv) by using National Digital Forecast Database (NDFD) GRIB2 decoder program of National Oceanic and Atmospheric Administration (NOAA), downloaded from Internet [12].

IV TECHNIQUE USED

A. Analysis of Relative Humidity

The relative humidity (RH) at 850hPa and 925hPa for the mentioned window under analysis has been selected from the forecast data. The clusters using k-means clustering technique of WEKA tool [13] have been generated for relative humidity. Two clusters have been generated by selection of RH value as the criterion for clustering, one cluster corresponds to 100% RH and other for rest of the values of RH. These clusters have been generated for the two atmospheric pressure levels viz. 850hPa and 925hPa. The 3-dimensional visualization of these clusters for forecast valid for 1800GMT 28July 09 corresponding to RH at 850hPa and at 925hPa are shown in Fig. 1. Similarly, the 3-dimensional visualization of these clusters for forecast valid for 1800GMT 28July 09 corresponding to RH at 850hPa and at 925hPa are shown in Fig. 2. Here, the clusters corresponding to 100% RH only has been shown.

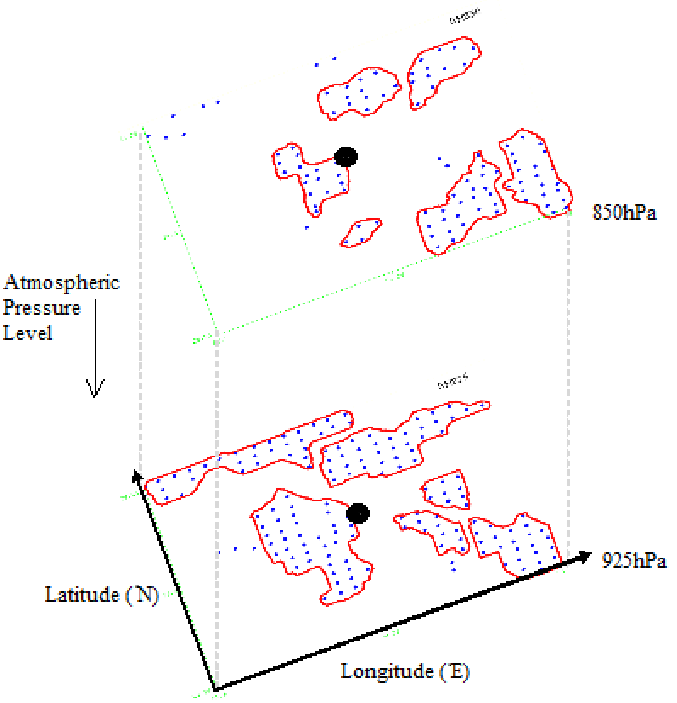

Fig. 1 3-dimensional visualization of clusters corresponding to Relative Humidity of 100% at 850hPa and 925hPa for forecast valid for 1800GMT 28July 09, location of cloudburst marked as black circle.

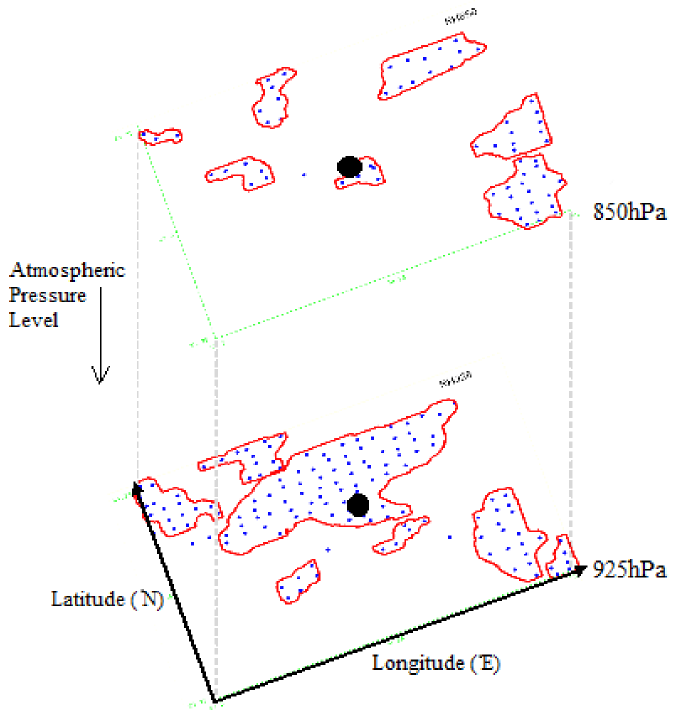

Fig. 2. 3-dimensional visualization of clusters corresponding to Relative Humidity of 100% at 850hPa and 925hPa for forecast valid for 0000GMT 29July 09, location of cloudburst marked as black circle.

B. Analysis of Temperature

The minimum and maximum temperature at 400hPa, 500hPa and 700hPa for the mentioned window under analysis has been selected from the forecast data. From the observations of forecast data, the findings are as follows in table 1.

TABLE I. FORECAST OF TEMPERATURE PRODUCED BY ECMWF MODEL

	Temperature at atmospheric pressure level 400hPa	Temperature at atmospheric pressure level 500hPa	Temperature at atmospheric pressure level 700hPa
Forecast valid for 1800GMT 28July 09	-11.7°C to -9.4°C	-2.8°C to -0.5°C	10.5°C to 12.8°C
Forecast valid for 0000GMT 29July 09	-11.7°C to -8.3°C	-2.8°C to 0°C	10.6°C to 12.8°C

V RESULTS AND DISCUSSIONS

A real life case of cloudburst has been discussed using DM k-means clustering technique. There is a very strong relative humidity field at atmospheric pressure level 925hPa in the clusters of forecast valid for 1800GMT 28July 09 and 0000GMT 29July 09, shown in Figure 1 and Figure 2 respectively. It is observed that this very large region of relative humidity is an early signal of formation of cloudburst. From the 3-D visualizations of these clusters, we can observe that high relative humidity of 100% is present even at atmospheric pressure level of 850hPa around the area of occurrence of cloudburst. The presence of high RH is supported by the extremely low temperature -11.7°C to -9.4°C at higher atmospheric height (pressure level of 400hPa) which causes the water vapours to sublimate on the ice crystals. Because of these two conducive features and the additional support of high temperature at lower atmospheric height (pressure level of 700hPa), the clouds cannot sustain and result in cloudburst. Hence the presence of advance signal of formation of cloudburst is indicated.

VI CONCLUSIONS

In this paper, the derivation of sub-grid scale weather systems from NWP model output products is demonstrated. Such signals are not possible through normal MOS technique. This study has demonstrated that Intelligent systems can be a good alternative for unstable MOS. Data mining, specially clustering when applied on divergence and relative humidity can provide an early indication of formation of cloudburst. This study is an effort towards providing timely and actionable information of these events using data mining techniques in supplement with NWP models that can be a great benefit to society.

ACKNOWLEDGEMENT

The author wishes to acknowledge the support of **Indian Meteorological Department** in providing the output products of ECMWF model. Also this study would not have been completed without the guidance and support of **Dr. Rattan K. Datta**, Former Advisor – Deptt. of Science & Technology, Former President - Indian Meteorological Society and Computer Society of India.

REFERENCES

[1] Study material available from University Corporation for Atmospheric Research http://www.ucar.edu/communications/gcip/m8clclchange/m8pdfc1.pdf

[2] Wikipedia, , the free encyclopedia http://en.wikipedia.org/wiki/Cloud

[3] K. Pabreja, R.K. Datta, "A data warehousing and data mining approach for analysis and forecast of cloudburst events using OLAP-based data hypercube", Accepted by *International J. Data Analysis Techniques and Strategies, Inderscience Publishers.*

[4] R.C. Dubes., A.K. Jain, *Algorithms for Clustering Data*. Prentice Hall, 1988.

[5] L. Ert¨oz, M. Steinbach, V. Kumar, "Finding Clusters of Different Sizes, Shapes, and Densities in Noisy,High Dimensional Data", In *Proc. of the 3rd SIAM International Conference on Data Mining*, San Francisco, CA, USA,2003.

[6] D. Singh, A. Ganju, A. Singh, "Weather prediction using nearest-neighbour model" *Current Science*, April 2005, vol. 88, no. 8, 25: 1283-1289.

[7] V. Kumar, M. Steinbach, P.N. Tan, C. Potter, S. Klooster, A. Torregrosa, "Finding Spatio-temporal Patterns in earth science data" in *Proc. 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, USA, 2001, Aug 26-29, 1-12.

[8] J. Han, M. Kamber, "Mining Complex Types of Data", In: Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, San Francisco, USA. 2001, 395-450.

[9] P. Stolorz, H. Nakamura, "Fast Spatio-temporal Data Mining of Large Geophysical Datasets", in *Proc. of the First International Conference on Knowledge Discovery and Data Mining*, AAAI Press, Montreal, Canada, 1995, Aug 20-21, 300-305.

[10] X. Huanga, F. Zhaob, "Relation-based aggregation: finding objects in large spatial datasets", *USA Intelligent Data Analysis*, 2000, 4: 129-147.

[11] Z. Guo, X. Dai, H. Lin, "Application of association rule in disaster weather forecasting", *The international Association of Chinese Professionals in Geographic Information Science,* 2004.

[12] NDFD GRIB2 decoder program of NOAA from Internet Available from: www.nws.noaa.gov/mdl/degrib/download.php

[13] Weka 3- Data mining with open source machine learning software . Available from:- http://www.cs.waikato.ac.nz/ml/weka/