Bayesian Network Probability Model for Weather Prediction

Aye Nandar Hlaing, Thinn Thu Naing
UCSY
anandarhlaingucsy@gmail.com

Abstract

Bayesian networks, or belief networks, show conditional probability and causality relationships between variables. Weather forecasting is important for various areas. In this paper, weather forecasting system is presented based on Bayesian network (BN) model. This work applies BN to model the spatial dependencies among the different meteorological variables for weather prediction (rainfall and temperature) over Myanmar. In this work, regional and global weather data which are contributing to rainfall prediction of Myanmar are used. Then, inference ability of BN approximate inference algorithm in rainfall prediction is analyzed with experiments over independent test data sets. For model training and testing, collected historical records of weather stations between 1990 and 2006 are used. Prediction accuracy of the model is reported with empirical results.

1. Introduction

A Bayesian belief network is an expressive knowledge representation for uncertain reasoning that its graphical structure can capture explicit dependencies among domain variables [8, 9]. Techniques for probabilistic inference in belief networks [10, 12, 13, 14, 16] and for specification of belief networks [11, 15, 17, 18] have emerged in response to the diversity of belief-network applications. In the multivariate analysis of categorical data, applied statisticians have recognized the importance of graphical structures and fundamental notion of conditional independence [19]. Several researchers have developed exact and approximate BN inference algorithms for different distributions.

The use of Bayesian networks (BNs) as a method of representing uncertainty has grown and several weather forecast systems have begun to employ its use since it provides a concise way to represent conditional independence relationships. Abramson et al. [1996] apply Bayesian models in the realm of meteorology combining meteorological data and model with expert judgment, based on both experience and physical understanding, to forecast severe weather in Northeastern Colorado [3]. A.S.Cofiño, R.Cano et al. combine numerical atmospheric predictions with Bayesian network representation to model the spatial and temporal dependencies among the different stations for weather prediction [2].

C. Rafael, Carmen et al. (2004) present some applications of Bayesian networks in Meteorology. The resulting graphical models are applied to different meteorological problems including weather forecast and stochastic weather generation [21]. In this work, collections of historical weather data are used to build BN probability model based on the spatial dependencies between these variables and analyze inference ability for monthly rainfall and temperature prediction with experiments.

This paper is organized as follows: Bayesian Network is briefly explained in Section 2. In section 3, application context of the system is described and overview of the model is presented in Section 4. Some experimental results are reported in section 5 as inference ability of the system. Finally, conclusion and future work of this work is presented in section 6.

2. Bayesian Network

A Bayesian network, belief network or directed acyclic graphical model is a probabilistic graphical model that represents a set of random variables and their conditional independencies via a directed acyclic graph (DAG). Inference in BNs means computing the probability distribution of a set of query variables, given a set of evidence variables. Each node in BNs is annotated with quantitative probability information. The full specification is as follows:

1. A set of random variables makes up the nodes of the network. Variables may be discrete or continuous.

2. A set of directed links or arrows connects pairs of nodes. If there is an arrow from node X to node Y, X is said to be a parent of Y.

3. Each node X, has a conditional probability distribution $P(X_i \setminus Parents(X_i))$ that quantifies the effect of the parents on the node.

4. The graph has no directed cycles (and hence is a directed, acyclic graph, or DAG).

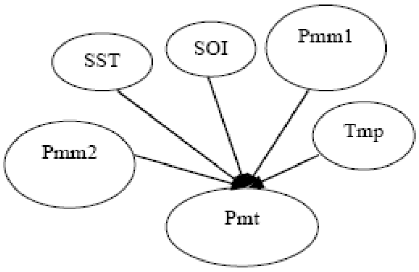

Figure 1: Simple Bayesian network model of Rainfall prediction model

SST –East India SST
SOI – Southern Oscillation Index
Pmm1–monthly precipitation of previous Year
Pmm2- precipitation of previous Month
Tmp - Temperature
Pmt - Precipitation amount

The topology of the network-the set of nodes and links specifies the conditional independence relationships that hold in the domain, in a way that will be made precise shortly. The intuitive meaning of an arrow in a properly constructed network is usually that X has a direct influence on Y. The dependency/independency structure displayed by an acyclic directed graph can be also expressed in terms of the Joint Probability Distribution (JPD) factorized as a product of several conditional distributions as follows:

$$Pr(y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} P(y_i|\pi_i).$$

The K2 algorithm is a simple greedy search algorithm for finding a high quality Bayesian network in a reasonable time [25]. The local K2 learning algorithm or simply LK2 which takes advantage of the spatial character of the problem to increase the speed and the meteorological significance of the obtained graphs. In order to increase the efficiency of the K2 search strategy, in local K2 algorithm, the set of candidate parents of node are modified to include only those nodes with similar climatology by computing the correlation of the observed records in different stations and obtaining the k-nearest neighbors for each station. In this way, complexity of the search process is reduced since the set of candidate parents is of constant size [26]. The LK2 learning algorithm is used to construct our BNs.

Monthly precipitation and average temperature for each weather station are predicted based on one of the BN approximate inference algorithm (Rejection Sampling algorithm). Rejection sampling is a general method for producing samples from a hard-to-sample distribution given an easy-to-sample distribution. In literature, however, this algorithm is simply unusable for complex problems since it rejects so many samples and the fraction of samples consistent with the evidence drops exponentially as the number of evidence variables grows. [1]

3. Application Context

Table1: Details of Some Weather station

No	Station Name	Region	Location
1	Sittwe	Coastal	20.1°N 92.8°E
2	Dawei	Coastal	14.1°N 97.6°E
3	Myitkyina	Upper	25.3°N 97.4°E
4	Haka	Upper	22.6°N 93.6°E
5	Homalin	Upper	24.8°N 94.9° E
6	Hpa-an	Lower	16.7°N 97.6°E
7	Bago	Lower	17.3°N 96.5°E
8	Monywa	Central	22.1°N 95.1°E
9	Mandalay	Central	21.9°N 96.1°E
10	Pathein	Delta	16.7°N 94.7°E

Myanmar is located between latitudes 09° 32'N and 28° 31'N and longitudes 92° 10' E and 101° 11' E. The location and topography of the country generate a diversity of climate conditions and seasonal changes in the monsoon wind directions crate summer, rainy and winter seasons. According to historical statistics, heavy monsoon rains in Myanmar in July and August have caused flooding along the rivers and their tributaries. It is, therefore, important to predict rainfall not only for hydrological purposes but also for various areas. In this work, we predict monthly rainfall amount for rainy season and monthly average temperature of Myanmar. Details of some weather station in the system are shown in Table1.

4. Overview of the System

In the first step, data of each weather station which are contributing to precipitation and temperature forecasting are collected and map collected variables in terms of BN. Then continuous quantities of collected data are quantized into N interval ranges (DISCRETIZATION). These interval values are defined on monthly amount of observation at each weather station.

In Predictor Analysis phase, state variables and evidence variables are defined for our probability model. As evidence variables of our model, weather data which are mainly contributing to precipitation in Myanmar are selected. All weather data are based on monthly time scale. The data used in this work consist of

(1) ***regional climate data***

• monthly total rainfall amount (in mm) of each weather station

• monthly average temperature of each weather station

(2) ***global climate data***

• Indian Ocean Dipole index (IOD)

The IOD affects the strength of monsoons over the Indian subcontinent. It is one aspect of the general cycle of global climate, interacting with similar phenomena like the El Niño-Southern Oscillation (ENSO) in the Pacific Ocean. In Myanmar, the southwest monsoon that starts during May in the India Ocean sometimes brings warm, moist air streams passing from the Indian Ocean.

• Southern Oscillation Index (SOI)

SOI is Measure of El Niño-Southern Oscillation (ENSO) which is a periodic change in the atmosphere and ocean of the tropical Pacific region. It is a ubiquitous influence on the monsoon circulation, playing a dominant role in the long range forecast of rainfall. It is accompanied by changes in the trade winds, and rainfall over the tropical Pacific Ocean and is related to many climatic anomalies around the globe.

• Sea Surface Temperature (SST)

SST plays a key role in setting up the land-ocean gradient which is important to the strength of the monsoon rainfall of our country since higher temperature during summer tends to favor a stronger monsoon rainfall. In this work, data from specific ocean areas (East India SST, 90°E-110°E 10°S) surrounding the Myanmar are collected.

To predict monthly average temperature, monthly average temperature of past two months and concerning monthly average temperature of last year are used as evidence variables.

BN models are constructed based on the spatial dependencies between weather variables defined in Predictor Analysis Process using LK2 algorithm. Table 2 shows evidence variables for rainfall prediction of our model.

Then, our probability model is trained with collected data for monthly rainfall and temperature prediction of each weather station.

Table2. Evidence Variables for rainfall Prediction

π_i	observable factors contributing to precipitation
π_1	East India SST (Previous Two Month)
π_2	Southern Oscillation Index (Previous Two Month)
π_3	Monthly precipitation amount of previous Year
π_4	Monthly precipitation amount of previous Month
π_5	Monthly average temperature

Then, our probability model is trained with collected data for monthly rainfall and temperature prediction of each weather station.

In Prediction process, posterior probability is calculated based on Rejection Sampling algorithm [1] that returns an estimate consistent with the evidence for weather prediction in BN model. The skill of the models is determined by the value of RMSE (root mean square error).

$$RMSE = \sqrt{1/n \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Details of some experimental results are reported in section 5.

5. Experimental Study

Table 3: RMSE rating table for Mandalay station

No	Period of Training Data	Period of Testing Data	RMSE
1	1990-2005	2006	2.8966
2	1990-2000	2001-2006	2.6342
3	1990-1995 2001-2006	1996-2000	2.7982

In our experiment, the collected historical weather datasets are divided into training data and testing data. Firstly, datasets of 1990-2005 are used as training datasets and datasets of 2006 are used as testing datasets. Secondly, we use 1990-2000 datasets as model training and 2001-2006 for model validation. Finally, datasets of 1990-1995 and 2001-2006 are used as training datasets and datasets of 1996-2000 are used as testing datasets. We analyzed the prediction accuracy of the model on different states of query variable by comparing RMSE values.

The model can give acceptable accuracy as respect to all experiments and some experimental results are presented in Table 3.

6. Conclusion & Future Work

In this work, BN probability model can give acceptable accuracy in terms of experimental results. Further analysis is still needed for determining the efficiency of the model, for sensitivity of number of evidence variables in particular. We intend to analyze performance comparison of other inference algorithms and the sensitivity of the parameters estimation to the different query and evidence states. We also intend to update and predict weather time series data over time using dynamic Bayesian network unrolling and inference algorithms.

7. References

[1] S.Russell, P.Norvig, "Artificial Intelligence: A Modern Approach", 2nd Edition, Prentice Hall, Inc, 2003.

[2] A.S.Cofiño, R.Cano et.el, "Bayesian Network for Weather Prediction", Proceedings of the 15th European conference on Artificial Intelligence, 695-700(2002).

[3] B. Abramson, J.M. Brown, W. Edwards, A. Murphy, and R.L. Winkler, 'Hailfinder: A Bayesian System for Forecasting Severe Weather'. International Journal of Forecasting, 12, 57–71, (1996).

[4] T. Cavazos, 'Using Self-Organizing Maps to Investigate Extreme Climate Event: An Application to Wintertime Precipitation in the Balkans', Journal of Climate, 13, 1718–1732, (2000).

[5] R. Neapolitan. Probabilistic Reasoning in Expert Systems. John Wiley & Sons, Inc., New York, NY, 1990.

[6] J. Pearl. Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann, San Mateo, CA, 1988.

[7] R. Chavez and G. Cooper. A randomized approximation algorithm for probabilistic inference on bayesian belief networks. Networks, 20:661685, 1990.

[8] G. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. Machine Learning, 9:309347, 1992.

[9] P. Dagum and R.M. Chavez. Approximating probabilistic inference in Bayesian belief networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(3):246255, 1993.

[10] P. Dagum and E. Horvitz. A Bayesian analysis of simulation algorithms for inference in belief networks. Networks, 23:499516, 1993.

[11] M. Henrion. An introduction to algorithms for inference in belief nets. Uncertainty in Artificial Intelligence 5, pages 129138. Elsevier, Amsterdam, The Netherlands, 1990.

[12] J. Pearl and T. Verma. A theory of inferred causation. Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference, pages 441-452. Morgan Kaufmann, April 1991.

[13] S. Lauritzen and D. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society B, 50(19), 1988.

[14] J. Pearl and T. Verma. A statistical semantics for causation. Statistics and Computing, 2:9195, 1992.

[15] P. Spirtes, C. Glymour, and R. Scheines. Causility, Prediction, and Search. Springer-Verlag, New York, 1993.

[16] J. Whittaker. Graphical Models in Applied Multivariate Statistics. John Wiley & Sons, New York, NY, 1990.

[17] E. Castillo, J. M. Guti´errez, and A. S. Hadi, Expert Systems and Probabilistic Network Models, Springer-Verlag, New York, 1997.

[18] D. Heckermean, 'A Tutorial on Learning with Bayesian Networks', Tech. Rep. MSR-TR-95-06, Microsoft Research, March (1996).

[19] G.F. Cooper, and E., Herskovitz, 'A Bayesian method for the induction of probabilistic networks from data', Machine Learning, 9, 309–347, (1992).

[20] G.W. Brier, 'Verification of forecasts expressed in terms of probabilities' Monthly Weather Review, 78, 13–22, (1950).

[21] R.Cano, C.Sordo et.el., Applications of Bayesian Networks in Meterology in Advances in Bayesian Networks, Gamez et al.eds., 309-327, Springer, 2004.

[22] URL: http://en.wikipedia.org/wiki/Indian_Ocean_Dipole

[23] URL: http://www.bom.gov.au/climate/glossary/soi.shtml

[24] URL: http://en.wikipedia.org/wiki/Bayesian_network

[25] Cooper G.F., Herskovitz E. (1992) A Bayesian method for the induction of probabilistic networks from data. Machine Learning 9, 309–347

[26] R.Cano,C.Sordo et al.,Applications of Bayesian Networks in Meteorology. in Advances in Bayesian Networks, Gámez et al. eds., 309-327, Springer, 2004.