PREDICTION OF TOMORROW'S RAINFALL USING BAYESIAN NETWORK

CSE516 - Probabilistic Graphical Model

TABLE OF CONTENTS

BACKGROUND

- Weather plays an important role in daily life
- Probabilistic Graphical Model can solve rain prediction problem easily
- BBN is an intuitive knowledge representation for handling uncertainty
- The main object of these networks is trying to understand the structure of causality relations.

BACKGROUND

- Mathematically Belief network should contain variable X= {X1,X2,.....,XN} of a Directed Acyclic Graph and probabilities can be calculated by the formula (X1,.....,XN) =

$$(X_1, \ldots, X_N) = \prod_{i=1}^{N} \mathrm{P}(X_i/\mathrm{Parents}(X_i))$$

- Here Parents(Xi) are the parents of Xi in a network. Further we can calculate the joint distribution and conditional probabilities indicated by the network. Bayesian belief network can be used as representation tool for the decision making under uncertainty.

MOTIVATION

- Social and Economic systems also depend on the weather forecast
- The weather requires wind, direction, speed, air pressure, date, location, etc.
- We will use a Bayesian network that will reduce the complexity and we will make inferences easily

CONTRIBUTION

- Used the Australian data of rain and related parameters of the last ten years.
- We have implemented the Bayesian Network.
- We have predicted tomorrow's rain.
- We are using fewer resources to predict tomorrow's rainfall and get reasonable accuracy as compared to other expensive methods and technologies.

METHODOLOGY

Data and its preprocessing

- We are using Australian data to predict tomorrow's rainfall. This data has a total of 23 variables.
- Here We are using only 11 variables to predict the rain.
- All variable data of different ranges are converted into the different states as high as 1, moderate as 2, low as 2, or high as 1 and low as 2.

Variables that we have used

- Season
- Mintmp
- Maxtmp
- Sunshine
- Humidity
- Pressure
- Cloud
- Wind Speed
- Evaporation
- Rainfall
- Rain Tommorow

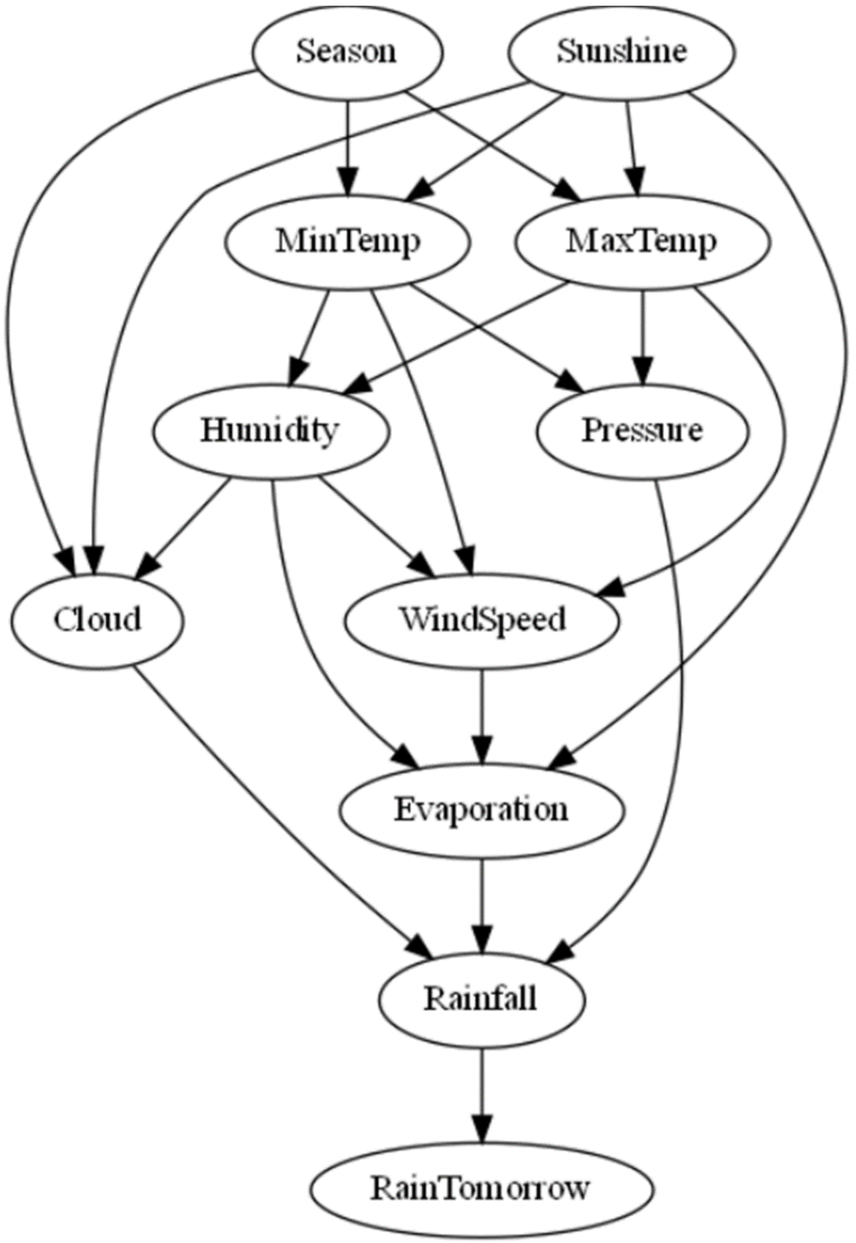
ANALYTICAL OVERVIEW

- With the help of these 11 nodes, we make 2 different Bayesian network

  a. Manual constructed bayesian network
  b. Automatically generated bayesian network

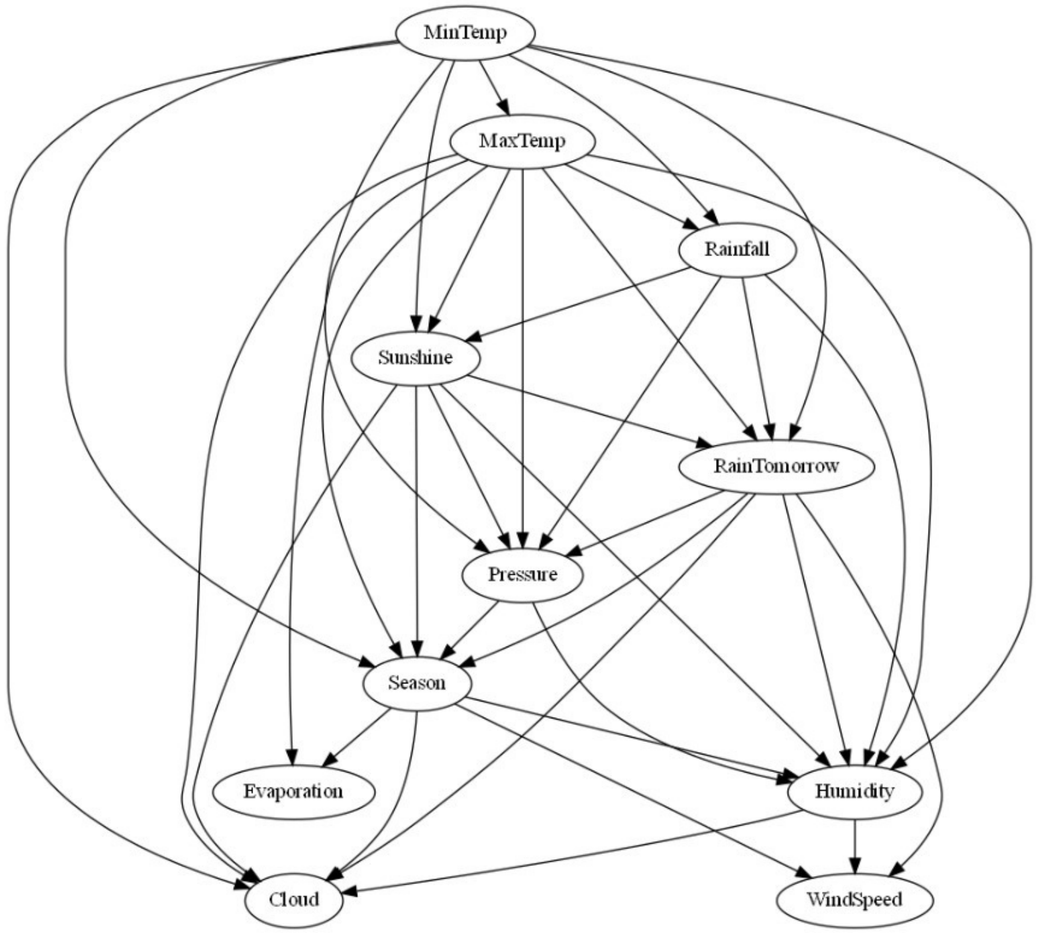
MANUAL CONSTRUCTED BAYESIAN NETWORK

- Identify the relevant node and structure dependencies between different nodes.
- Tomorrow’s rain is directly or indirectly dependent on present-day atmospheric phenomena.
- Only doing Parameter learning to Find CPT.
- Do variable elimination to find the probability of tomorrow's rain.



AUTOMATICALLY GENERATED BAYESIAN NETWORK

- Do Structure learning on 11 nodes of data
- Get the Bayesian network from structure learning
- Use Hill climbed algorithm for structure learning
- Do parameter learning to get CPT
- Use the bayesian estimator for parameter learning
- Use the variable elimination to get the probability of the given query.



PARAMETER LEARNING

- Parameter learning is the process of quantifying the conditional probability between nodes
- we use the Bayesian estimator for this model.
- The Bayesian estimator is maximize the posterior probability.

STRUCTURE LEARNING

- Structure learning is the process of finding the Bayesian network.
- we use Hill climb K2 algorithm for finding the dependencies between various nodes.
- Hill climb is a greedy algorithm to find the bayesian network.

HILL CLIMB

- It is a local search Algorithm
- It works for a greedy approach
- It includes changes in the state
- It reduces space complexity.
- It is also known as the Heuristic search

Variable Elimination

- **It is an efficient method for deriving inferences rather than joint distribution.**
- **Final conditional probability can be calculated in two ways.**
- **For our bayesian network, we calculated rain tomorrow with evidence as Evaporation, WindGustSpeed.**

RESULTS AND INTERFERENCE

- Pygame
- Python
- Baysiannetwork
- variable elimination

Model -1

- Hand made
- 11 nodes

Model - 2

- HC - K2 score
- 11 nodes

Models	Accuracy
Model - 1	0.7872
Model - 2	0.7916
Model - 3	0.8172

- Probability of tomorrow Rain
- Evidences
- Understand networks

Model - 3

- HC - K2 score
- 20 nodes

CONCLUSION

- We are using 11 nodes but increasing the number of nodes can increase the accuracy of the prediction of rainfall
- We can increase the parameter to increase the accuracy

- For scope of improvement, Is rain very high or low or moderate that can be determined.
- We can use more RSMAX2(General 2 - phase Restricted Maximizatio) and H2PC (Hybrid HPC)

THANK YOU