MODULE 10
Bayes Classifier

LESSON 20
Bayesian Belief Networks

Keywords: Belief, Directed Acyclic Graph, Graphical Model

Bayesian Belief Network

- A Bayesian network is a graphical model of a situation which represents a set of variables and the dependencies between them by using probability.
- The nodes in a Bayesian network represent the variables and the directional arcs represent the dependencies between the variables. The direction of the arrows show the direction of the dependency.
- Each variable is associated with a conditional probability table which gives the probability of this variable for different values of the variables on which this node depends.
- Using this model, it is possible to perform inference and learning.
- Bayesian networks that model a sequence of variables varying in time are called dynamic Bayesian networks.
- Bayesian networks with decision nodes which solve decision problems under uncertainly are Influence diagrams.
- The graphical model will be a directed acyclic graph with the nodes representing the variables. The arcs will represent the dependencies between nodes. If there is a directed arc between A and B, then A is called the parent of B and B is a child of A.
- A variable which has no parent is a variable which does not depend on any other variable. It is a variable which is independent and is not conditioned on any other variable.
- The conditional probability table associated with each node gives the probability of this variable for different values of its parent nodes.
- If a node does not have any parents, the conditional probability table is very simple as there are no dependencies.
- The joint distribution of the variable values can be written as the product of the local distribution of each node and its parents. In other words, if $f_1, f_2, ..., f_d$ are the variables, the joint distribution

$P(f_1, f_2, ..., f_d)$ is given by

$P(f_1, f_2, ..., f_d) = \prod_{j=1}^{d} P(f_j \mid parents(f_j)$

Using the above equation, it is possible to get the joint distribution of the variables for different values.

- As as example, let us consider the following scenario.

  Lakshman travels by air if he is on an official visit. If he is on a personal visit, he travels by air if he has money. If he does not travel by plane, he travels by train but sometimes also takes a bus.

The variables involved are :

1. Lakshman travels by air(A)
2. Goes on official visit(F)
3. Lakshman has money(M)
4. Lakshman travels by train(T)
5. Lakshman travels by bus(B)

This situation is converted into a belief network as shown in Figure 1.

In the graph, we can see the dependencies with respect to the variables. The probability values at a variable are dependent on the value of its parents. In this case, the variable A is dependent on F and M. The variable T is dependent on A and variable B is dependent on A. The variables F and M are independent variables which do not have any parent node. So their probabilities are not dependent on any other variable. Node A has the biggest conditional probability table as A depends on F and M. T and B depend on A.

Once the graph is drawn which is a directed acyclic graph (DAG), the probabilities of the variables depending on the values of their parent node has to be entered. This requires us to know the problem at hand and estimate

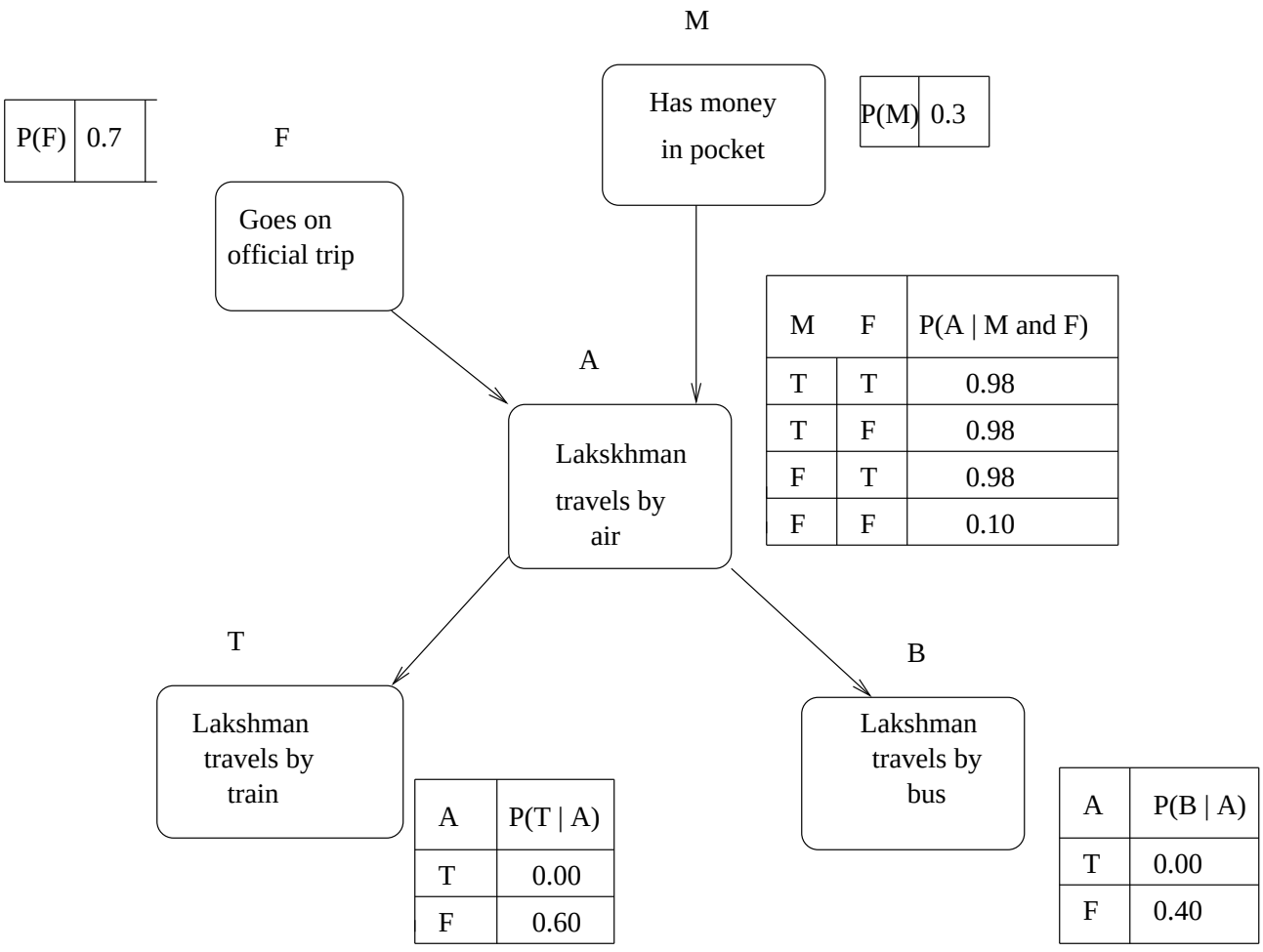

Figure 1: Bayesian Belief Network

these probabilities. So for each node the Conditional Probability Table has to be entered.

First we take the independent nodes. Node F has a probability of $P(F) = 0.7$. Node M has a probability of $P(M) = 0.3$.

We next come to node A. The conditional probability table for this node can be represented as

F	M	P(A | F,M and P)
T	T	0.98
T	F	0.98
F	T	0.90
F	F	0.10

The conditional probability table for T can be represented as

A	P | A
T	0.0
F	0.6

The conditional probability table for B is

A	P | A
T	0.0
F	0.40

Using the bayesian belief network, we can get the probability of a combination of these variables. For example, we can get the probability that Lakshman travels by train, does not travel by air, goes on an official trip and has money. In other words, we are finding $P(T, \neg A, F, M)$. The probability of each variable given its parent is found and multiplied together to give the probability.

$P(T, \neg A, M, P) = P(T \mid \neg A) * P(\neg A \mid F \; and \; M) * P(F) * P(M)$
$= 0.6 * 0.98 * 0.7 * 0.3 = 0.123$

Assignment

1. Let the probability that a road is wet P(w) = 0.3. Let probability of rain, P(R) = 0.3. Given that 90% of the time when the roads are wet, it is because it has rained, and it has rained, calculate the *posterior* probability that the roads are wet.

2. Let *blue, green, and red* be three classes of objects with prior probabilities given by P(blue) = 0.3, P(green) = 0.4, P(red) = 0.3. Let there be three types of objects: *pencils, pens, and paper*. Let the class-conditional probabilities of these objects be given as follows. Use Bayes classifier to classify pencil, pen, and paper.

P(pencil|green) = 0.3	P(pen|green) = 0.5	P(paper|green) = 0.2
P(pencil|blue) = 0.5	P(pen|blue) = 0.2	P(paper|blue) = 0.3
P(pencil|red) = 0.2	P(pen|red) = 0.3	P(paper|red) = 0.5

3. Consider a two-class (Tasty or nonTasty) problem with the following training data. Use Naive Bayes classifier to classify
$Cook = Asha,\ Health - Status = Bad,\ Cuisine = Continental$

Cook	Health-Status	Cuisine	Tasty
Asha	Bad	Indian	Yes
Asha	Good	Continental	Yes
Sita	Bad	Indian	No
Sita	Good	Indian	Yes
Usha	Bad	Indian	Yes
Usha	Bad	Continental	No
Sita	Bad	Continental	No
Sita	Good	Continental	Yes
Usha	Good	Indian	Yes
Usha	Good	Continental	No

4. Consider the following dataset with three features f_1, f_2, and f_3. Consider the test pattern $f_1 = a$, $f_2 = c$, $f_3 = f$. Classify it using NNC and Naive Bayes Classifier.

f_1	f_2	f_3	Class Label
a	c	e	No
b	c	f	Yes
b	c	e	No
b	d	f	Yes
a	d	f	Yes
a	d	f	No

5. The profit a businessman makes depends on how fresh the provisions are. Further, if there a festival approaching, his profit increases. On the other hand, towards the end of the month, his sales come down. If he makes enough profit, he celebrates Diwali in a big way. Draw the belief network and suggest the likely conditional probability tables for all variables. Using this data, find the probability that the businessman celebrates Diwali big given that the provisions are fresh.

References

V. S. Devi and M. N. Murty (2011) *Pattern Recognition: An Introduction* Universities Press, Hyderabad.

R.O. Duda, P.E. Hart and D.G. Stork (2001) *Pattern Classification*, Second Edition, Wiley-Interscience.

S. Russell and P. Norvig (2003) *Artificial Intelligence : A Modern Approach*, Pearson India.

P. Domingos and M. Pazzani (1997) On the optimality of the simple Bayesian classifier under zero-one loss, *Machine Learning*, Vol. 29, pp. 103-130.

J. Pearl, (1988) *Probabilistic reasoning in intelligent systems, Morgan Kauffman.*

I. Rish *(2001) An empirical study of the naive Bayes classifier, IJCAI workshop on Empirical Methods in Artificial Intelligence*

D. Hackerman *(1996) Bayesian networks for knowledge discovery, In U.M. Fayyad, G.P. Shapiro, P. Smyth and R. Uthurusamy, editors Advances in Knowledge Discovery and Data Mining, MIT Press.*
R.E. Neopolitan, *(2003) Learning Bayesian Networks, Prentice Hall.*