Probabilistic Graphical Models

Huynh Xuan Phung - Coursera

Contents

Chapter 1

Representation

1.1 Introduction

Model

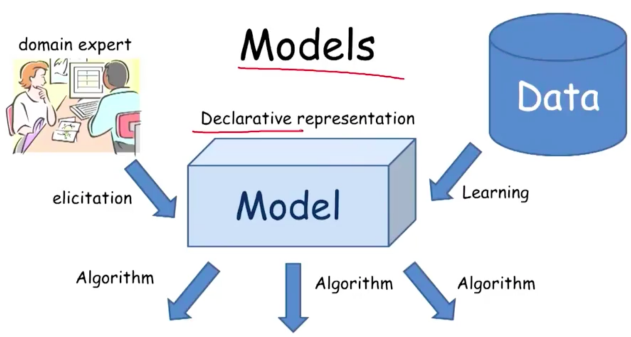

Figure 1.1: Model is a declarative representation of our understanding of the world

It is important because the same representation, that same model can be used in the the context of one algorithms that might answer different kind of questions. Or the same question in more efficient way.

We can construct methodologies the elicit these models from a human extrict, or learn from data or combination.

Uncertainty

— Partial knowledges of state of the world

— Noisy observations

— Phenomena not covered by our model

— Inherent stochasticity

Probability Theory

— Declarative representation with clear semantics

— Powerful reasoning patterns: conditioning decision making

— Established learning methods

Complex Systems

Graphical Models: Bayesian Networks, Markov Networks (directed or undirected graphs)

Graphical Representation

— Intuitive and compact data structure

— Efficient reasoning using general purpose algorithms

— Sparse parameterization

—— feasible elicitation : by hand

—— learning from data automatically

1.1.1 Distribution

Joint Distribution: P(I,D,G)

Conditioning: observation 1 value of variable —> Reduction —> Renormalization $P(I, D, g^1)$ —> $P(I, D|g^1)$

Marginalization: $\sum_I P(I, D) = P(D)$. Example: you have thrown two 6-sided dice, D_1 and D_2. $P(D_1, D_2)$ is a joint probability distribution. The probability that $D_2 = 1$ is equals to $\sum_{i=1}^{6} P(D_1 = i, D_2 = 1)$.

1.1.2 Factors

A factor is a function or table $\phi(X_1, ..., X_k)$

$\phi : Val(X_1, ..., X_k) \to R$

Scope = $X_1, ..., X_k$

Joint distribution is a factor

Unnormalized measure is a factor

Conditional Probability Distribution (CPD) is a factor P(G—I,D): G in columns while I,D in rows.

Why factors?

— Fundamental building block for defining distributions in high-dimensional spaces

— Set of basic operations for manipulating these probability distributions

1.2 Bayesian Network (Directed Models)

1.2.1 Semantics and Factorization

What does random variable depend on?

Draw nodes, edges, each node with a factor is CPD (conditional probability distribution)

A bayesian network is:

— A directed acyclic graph (DAG) G whose nodes represent the random variables

— For each node X_i a CPD $P(X_i|Par_G(X_i))$

The BN represent a joint distribution via the chain rule for Bayesian Networks

$P(X_1, ..., X_n) = \prod_i P(X_i|Par_G(X_i))$

BN is a Legal distribution : $\sum P = 1$ and $P > 0$

1.2.2 Reasoning Pattern

Causal Reasoning: reasoning going down

Evidential Reasoning: reasoning going up

Inter-causal Reasoning: The probability of class is hard, if we observe the "C" grade and change the posterior probability of high intelligence is goes up

1.2.3 Flow of Probabilistic Influence

When can X influence Y?

X $\rightarrow$ Y

X $\leftarrow$ Y

Active Trails:

A trail X_1 – ... – X_k is active if: it has no v-structures $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$

When can X influence Y given evidence about Z?

A trail X_1 — ... — X_k is active given Z if:

— for any v-structure $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ we have that X_i or one of its descendants $\in$ Z

— no other X_i is in Z

Assignment

1. How many parameter to present a CPD. If X have m possibilities the P(X) needs m-1 independent parameters

IF Y have k possibilities, Z have l possibilities then P(X—Y,Z) has (m-1)*k*l independent parameters.

2. Inter-causal reasoning

To calculate the required values, we can apply Bayes' rule. For instance,

$\frac{P(A=1|T=1,P=1)=P(A=1,T=1,P=1)}{P(T=1,P=1)}$

$= \frac{P(A=1,T=1,P=1)}{(P(A=0,T=1,P=1)+P(A=1,T=1,P=1))}$.

We can then use the chain rule of Bayesian networks to substitute the correct values in, e.g.,

$P(A = 1, T = 1, P = 1) = P(P = 1) * P(A = 1) * P(T = 1|P = 1, A = 1)$

This example of inter-causal reasoning meshes well with common sense: if we see a traffic jam, the probability that there was a car accident is relatively high. However, if we also see that the president is visiting town, we can reason that the president's visit is the cause of the traffic jam; the probability that there was a car accident therefore drops correspondingly.

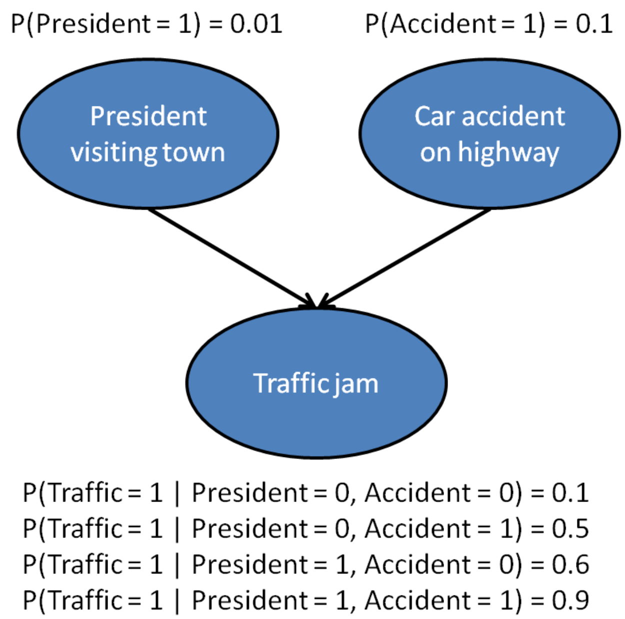
P(President = 1) = 0.01
P(Accident = 1) = 0.1
President visiting town
Car accident on highway
Traffic jam
P(Traffic = 1 | President = 0, Accident = 0) = 0.1
P(Traffic = 1 | President = 0, Accident = 1) = 0.5
P(Traffic = 1 | President = 1, Accident = 0) = 0.6
P(Traffic = 1 | President = 1, Accident = 1) = 0.9

Figure 1.2: