Natural Language Processing CSCI 4152/6509 — Lecture 14 P0 Discussion (2); Bayesian Networks and HMM Inference

Instructors: Vlado Keselj
Time and date: 14:35 – 15:55, 18-Nov-2025
Location: Studley LSC-Psychology P5260

Previous Lecture

- P0 discussion (1): P-03, P-07, P-08, P-09, P-11, P-12, P-20, P-21, P-22, P-23
- HMM Brute-force approach
- HMM Inference: Viterbi algorithm

P0 Topics Discussion (2)

- Continued discussion of individual projects as proposed in P0 submissions (part 2)
- Projects discussed: P-01, P-04, P-06, P-13, P-14, P-15, P-16, P-17

HMM as Bayesian Network

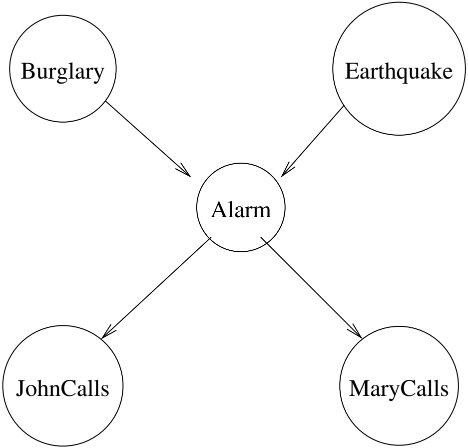
- **Viterbi** algorithm is an **efficient** way to solve a **special** problem:
  - completion with known observables and unknown hidden nodes of an HMM
- **General** approach:
  - Treat HMM as **Bayesian Network**
  - Apply **Product-Sum** (i.e., "Message-passing") algorithm for efficient inference

Bayesian Network Model

- Also known as: Belief Networks, or Bayesian Belief Networks
- A directed acyclic graph (DAG)
  - Each node representing a random variable
  - Edges representing causality (probabilistic meaning)
- Conditional Probability Table (CPT) for each node
- Bayesian Network assumption:

$$\mathrm{P}(\underbrace{V_1, V_2, \ldots, V_n}_{\text{full configuration}}) = \prod_{i=1}^{n} \mathrm{P}(V_i | \mathbf{V}_{\pi(i)})$$

Bayesian Network Example



Bayesian Network Assumption

- Bayesian Network Assumption for previous example:

$$\mathrm{P}(B, E, A, J, M) = \mathrm{P}(B)\mathrm{P}(E)\mathrm{P}(A|B, E)\mathrm{P}(J|A)\mathrm{P}(M|A)$$

- Probability of a complete configuration is a product of conditional probabilities
- Each node corresponds to one conditional probability: $\mathrm{P}(B)$, $\mathrm{P}(E)$, $\mathrm{P}(A|B, E)$, $\mathrm{P}(J|A)$, $\mathrm{P}(M|A)$
- CPTs (Conditional Probability Tables are model parameters)

Conditional Probability Tables

B	$\mathrm{P}(B)$
T	0.001
F	0.999

E	$\mathrm{P}(E)$
T	0.002
F	0.998

B	E	A	$\mathrm{P}(A\|B,E)$
T	T	T	0.95
T	T	F	0.05
T	F	T	0.94
T	F	F	0.06
F	T	T	0.29
F	T	F	0.71
F	F	T	0.001
F	F	F	0.999

A	J	$\mathrm{P}(J\|A)$
T	T	0.90
T	F	0.10
F	T	0.05
F	F	0.95

A	M	$\mathrm{P}(M\|A)$
T	T	0.70
T	F	0.30
F	T	0.01
F	F	0.99

Computational Tasks

- Evaluation:

$$\mathrm{P}(V_1 = x_1, \ldots, V_n = x_n) \quad = \quad \prod_{i=1}^{n} \mathrm{P}(V_i = x_i | \mathbf{V}_{\pi(i)} = \mathbf{x}_{\pi(i)})$$

- Simulation
- Learning from complete observations
- Inference in Bayesian Networks

Inference Example using Brute Force

$$P(B = T | J = T) = \frac{P(B = T, J = T)}{P(J = T)}$$

$$\begin{aligned}
P(B = T, J = T) &= \sum_{E,A,M} P(B = T, E, A, J = T, M) \\
&= \sum_{E,A,M} P(B = T) P(E) P(A | B = T, E) \\
&\quad P(J = T | A) P(M | A) \\
&\approx 8.49017 \cdot 10^{-4}
\end{aligned}$$

(continued)

$$\mathrm{P}(J = T) = \mathrm{P}(B = T, J = T) + \mathrm{P}(B = F, J = T)$$

$$\mathrm{P}(J = T) = \mathrm{P}(B = T, J = T) + \mathrm{P}(B = F, J = T) \approx$$
$$8.49017 \cdot 10^{-4} + 5.12899587 \cdot 10^{-2} = 0.0521389757$$

$$\mathrm{P}(B = T | J = T) = \frac{\mathrm{P}(B = T, J = T)}{\mathrm{P}(J = T)} \approx$$
$$\frac{8.49017 \cdot 10^{-4}}{0.0521389757} \approx 0.0162837299467699.$$

General Inference in Bayesian Networks

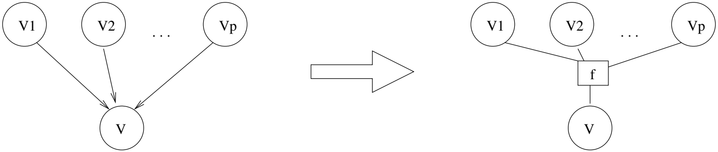
- In some Bayesian Networks inference is always expensive; e.g., joint distribution has a very large number of parameters
- Can we be more efficient if number of parent nodes is limited?
- Naïve Bayes or HMM has a limit of parents to 1
- If we limit number of parents to 2, this may already lead to an NP-hard inference problem
- Proof: a reduction from Circuit Satisfiability problem

Sum-Product Algorithms for Bayesian Networks

- Basic idea: optimizing sum-product calculation using graph structure
  Described in "Factor graphs and the Sum-Product Algorithm" by Kschishang, Frey, and Loeliger in 2000
- Algorithm overview:
  1. Construction of a factor graph
  2. Message-passing algorithms
- Construction of the factor graph
- Principles of message passing

Factor Graph

- Introduce factor nodes:



- Factor graph captures the structure of computation

Factor Graph Example