Iowa State University — Lecturer: Aditya Ramamoorthy and Zhengdao Wang

EE 520: Topics in Communications: Multi-User Information Theory

October 28, 2008 — Scribe: Kun Qiu

Lecture 19

In this lecture, we study the graphical probability model.

1 Introduction

Graphical model is a good tool for providing a compact representation of joint probability distributions. There are two main kinds of graphical models: directed and undirected. Directed graphs are also called Bayesian Networks or Dependency Graphs; undirected graphical models are also known as Markov Networks or Markov Random Fields (MRFs).

2 Directed Graphic Models

A directed graph $G(V, E)$ consists of nodes (or vertices) V and arrows (or directed edges) E connecting some of the nodes. We only consider directed graphs with no directed cycles: Directed Acyclic Graph (DAG). There could be cycles if the directions are disregarded.

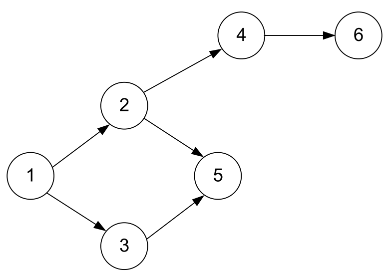

Figure 1: A DAG example.

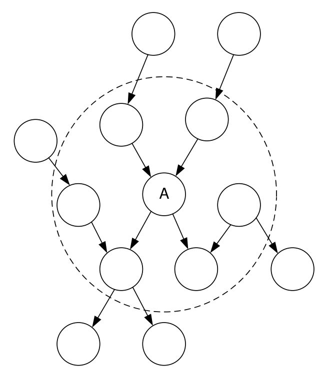

Figure 2: Markov blanket.

Definition 1. *Here are some basic definitions of DAG:*

- *Vertices set V: each vertex represents a random variable, say $X_i (i \in V)$.*
- *for a subset of vertices $A \subset V$, $X_A = \{X_i | i \in A\}$;*
- *edge set E is a set of ordered pairs of vertices, if there is an arrow from vertex i to j, then $(i, j) \in E$;*
- *the parents of vertex i is $\pi_i = \{j | j \to i \in E\}$;*
- *the children of vertex i is $c_i = \{j | i \to j \in E\}$;*
- *the descendants of i is $\delta_i = \{j \in V | \text{there is a direct path from } i \text{ to } j\}$;*

- *if X_A are independent of X_B given X_C ($A, B, C \subset V$), we write $X_A \perp X_B | X_C$;*
- *A sequence of vertices constituting an undirected path i to j has a* collider *at k if there are two arrows along the path pointing to k.*

Look at the DAG in figure 1, we have $V = \{1, 2, 3, 4, 5, 6\}$ and $E = \{(1, 2), (1, 3), (2, 4), (2, 5), (3, 5), (4, 6)\}$. For vertex 2, it has a single parent 1, two children 4 and 5, three descendants $\{4, 5, 6\}$.5 is a collider in the graph.

Definition 2. *For a DAG G with vertices $V = \{1, 2, \cdots, N\}$, we say it has conditional independence property if for $i = 1, 2, \cdots, N$*

$$P(X_i | X_{\pi_i}, X_A) = P(X_i | X_{\pi_i}),$$

where $A \subset V - \delta_i$. This is the defining conditional independence property.

This property says that given its parents, a vertex is conditionally independent of any other vertices that are neither this node's parents nor its descendants. A DAG with the above property follows the following theorem.

Theorem 1. *A DAG with vertices $V = \{1, 2, \cdots, N\}$ has the conditional independence property if and only if*

$$P(X_V) = \prod_{i=1}^{N} P(X_i | X_{\pi_i}).$$

Proof: Since graph G has no directed cycles, it must be possible to choose an ancestral ordering to denote each vertex such that the descendants of each node appear after itself. Let the ordered vertices sequence to be $\{X_{(1)}, X_{(2)}, \cdots, X_{(N)}\}$. (In the DAG example of figure 1, the ordered sequence could be $\{1, 2, 4, 6, 3, 5\}$.) By chain rule, we have

$$\begin{aligned} P(X_V) &= \prod_{i=1}^{N} P(X_{(i)} | X_{(1)}, X_{(2)}, \cdots, X_{(i-1)}) && (1) \\ &= \prod_{i=1}^{N} P(X_{(i)} | X_{\pi_{(i)}}) && (2) \\ &= \prod_{i=1}^{N} P(X_i | X_{\pi_i}). && (3) \end{aligned}$$

In (2), we have used the conditional independence property.

Definition 3. *Let i and j be distinct vertices of a DAG and C be a set of vertices not containing i or j. Then, i and j are* **d-connected** *by C if there is an undirected path P between i and j such that:*

- *every collider in P has a descendant in C;*
- *no other vertex on P is in C.*

If i and j are not d-connected given C, they are **d-separated** *by C. If A,B, and C are non-overlapping sets of vertices in the DAG and A and B are not empty, then A and B are d-separated by C if, for every $i \in A$ and $j \in B$, i and j are d-separated by C.*

Theorem 2. *A and B are d-separated by C if and only if $X_A \perp X_B | X_C$.*

Definition 4. *The* **Markov blanket** *for a vertex A is the set of vertices composed of A's parents, its children, and its children's parents. It is denoted as ∂A.*

Figure 2 is an example of Markov blanket. A set of nodes B in the graph is conditionally independent of A when conditioned on the set ∂A, that is

$$P(X_A | X_B, X_{\partial A}) = P(X_A | X_{\partial A}).$$

Now, we introduce the Bayes ball algorithm to decide d-separation. Let us start with the three canonical graphs.

Figure 3: Markov chain (left: unconditioning, right: conditioning on C).

Figure 4: Hidden cause (left: unconditioning, right: conditioning on C).

Figure 5: Explaining away/collider (left: unconditioning, right: conditioning on C).

We want to move ball from A to C following the rules as below:

- In the Markov chain in figure 3, the ball passes through C if we do not condition on C, and is bounced back otherwise;
- In the hidden cause situation in figure 4, the ball passes through C if we do not condition on C, and is bounced back otherwise;
- In the explaining away situation in figure 5, the ball is bounced back by C if we do not condition on any decendent of C, and passes through otherwise.

With the rules and canonical graphs in mind, the Bayes ball algorithm is described as follows:

1. First, shade the nodes C that are conditioned on;
2. Start the ball within the nodes in set A and bounce it around the graph according to the rules stated before;

3. Finally, evaluate the results: if the ball can reach a node in B, then A and B are d-connected by C; if the ball cannot reach B, then the nodes in A and B are d-separated by C.

Some examples. In figure 6, 1 is d-connected with 3 by 4, because we can pass the ball through the route $1 \rightarrow 2 \rightarrow 4 \rightarrow 2 \rightarrow 3$. In figure 7, 1 is d-separated with $3, 4$ by 2, since there is no route to pass the ball from 1 to $3, 4$.

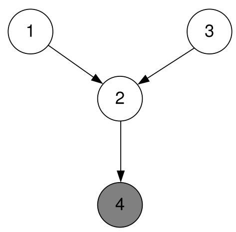

Figure 6: Bayes ball algorithm example 1.

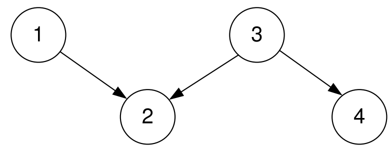

Figure 7: Bayes ball algorithm example 2.

REFERENCES

1. Pearl 1988 Book: "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference"

2. M. Jordan Berkeley CS281A/STAT241A notes. (9/2/04, 9/7/04)

3. M. Jordan Tutorial, "Graphical Models"

4. B. Frey 1998 Book: "Graphical Models for Machine Learning and Digital Communication"