Bayesian Networks

School of Electronic and Computer Engineering
Peking University

Wang Wenmin

Contents

Contents

About Uncertain Knowledge 关于不确定性知识

- Evolution of an intelligent agent: problem solving, reasoning, planning and learning.
  智能体的进化：问题求解、推理、规划以及学习。
- Agents may need to handle uncertainty, due to partial observability and non-determinism.
  智能体可能需要处理不确定性，由于部分可观察性和不确定性问题。
- To make decision with uncertainty, we need
  在不确定性的情况下做出决策，我们需要
  - Probability theory,
    概率论，
  - Utility theory,
    效用论，
  - Decision theory.
    决策论。

Example: An uncertainty problem 一个不确定性问题

A_{90} = home to airport 90 minutes by taxi before flight departs.

从家里打车在航班起飞前90分钟到机场

- Question: 问题

  "Will A_{90} get me to the airport on time?"

  A_{90} 能使我准时到达机场吗？

- Answer: 答案

  The taxi agent concludes either:

  出租汽车公司给出两个结论中的一个：

  - the risks falsehood: "A_{90} will get us there in time".

    有风险的谎言：A_{90}将使我们及时到达机场。

  - the weaker conclusion: "A_{90} will get us there in time, if there is no traffic jam, I don't get into an accident, the car doesn't break down, and ..."

    A_{90}将使我们及时到达机场，如果没有交通堵塞、不出交通事故、汽车不出故障的话，……

Rational Decisions 理性决策

- Probability theory 概率论

  for dealing with degrees of belief.

  是用于处理置信度的理论

- Utility theory 效用论

  the quality of being useful.

  是有效性的质量

  - to represent and reason with preferences, every state has a degree of usefulness/utility.

    用偏好来表现和推理，每个状态都具有“有效性/效用”的度量值。

- Decision theory 决策论

  the general theory of rational decisions.

  是理性决策的通论

  - Decision theory = probability theory + utility theory

    决策论 = 概率论 + 效用论

Algorithm of a Decision-theoretic Agent 一种决策论智能体的算法

```
function DECISION-THEORETIC-AGENT(percept) returns an action
    persistent: belief_state, probabilistic beliefs about the current state of the world
                action, the agent's action

    update belief_state based on action and percept
    calculate outcome probabilities for actions,
          given action descriptions and current belief_state
    select action with highest expected utility,
          given probabilities of outcomes and utility information
return action
```

A decision-theoretic agent that selects rational actions.

一个选择理性动作的决策论智能体

Bayes' Rule 贝叶斯规则

- Product rule 乘积规则
  - Two ways to factor a joint distribution over two variables:
    两个变量联合分布的两种计算方法：

$$P(a \wedge b) = P(a \mid b)\, P(b) \quad \text{and} \quad P(a \wedge b) = P(b \mid a)\, P(a)$$

- Bayes' rule 贝叶斯规则

$$P(b \mid a) = \frac{P(a \mid b)\, P(b)}{P(a)}$$

- This rule underlies most modern AI for probabilistic inference.
  这个规则成为大多数现代人工智能概率推理的基础。
- Why is Bayes' rule useful 为什么贝叶斯定理有用
  - Often we have good probability estimates for three terms to compute the fourth.
    我们常常需要根据三个项的概率估计值去计算第四个。

Example: Inference with Bayes' Rule 贝叶斯规则进行推理

- Often we perceive as evidence the *effect* of some unknown cause, and would like to determine that *cause*. In that case, Bayes' rule becomes
  我们往往想根据一些未知原因的证据，来查明其原因。这样，贝叶斯定理就变成了

$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}$$

- Knows $P(symptoms \mid disease)$ and want to derive a diagnosis, $P(disease \mid symptoms)$.
  已知 $P(symptoms$ 症状 $\mid disease$ 疾病$)$，想要得出一个诊断 P(疾病 ｜ 症状)。

$P(s \mid m) = 0.7$ // *conditional probability that meningitis causes a stiff neck* 脑膜炎导致颈部僵硬的条件概率

$P(m) = 1/50000$ // *prior probability that a patient has meningitis* 病人患脑膜炎的先验概率

$P(s) = 0.01$ // *prior probability that any patient has a stiff neck* 任何病人患有颈部僵硬的先验概率

$$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.7 \times 1/5000}{0.01} = 0.0014$$ // *a stiff neck to have meningitis* 颈部僵硬患有脑膜炎的概率

About Bayesian Networks 贝叶斯网络

- A probabilistic graphical model (a type of statistical model)

  一种概率图模型（一种统计模型的类型）

- With a directed acyclic graph (DAG), it represents:
  - a set of random variables, and conditional dependencies between the variables.

  采用一种有向无环图 (DAG)，它表示：一组随机变量，以及变量之间的条件相关性。

- Its specification: 它的规范如下：
  - 1) a set of nodes, each corresponds to a random variable, 2) a set of directed links to those nodes, and 3) a conditional probability distribution for each node given its parents:

    1）一组节点，每个节点对应于一个随机变量，2）一组这对这些节点的有向连接，以及 3）每个节点在给定双亲下的条件概率分布：

$$\mathbf{P}(X_i \mid Parents(X_i))$$

About Bayesian Networks 贝叶斯网络

- The name of Bayesian networks is the most common one, but there are many synonyms, including:

  贝叶斯网络这个名称是最常用的，但还有许多同义词，包括：

  - belief network, 信念网络
  - probabilistic network, 概率网络
  - causal network. 因果网络

- A Bayesian network represents a set of random variables and their conditional dependencies.

  一个贝叶斯网络表示一组随机变量和他们的条件依赖关系。

  - E.g., it could represent the probabilistic relationships between diseases and symptoms.

    例如：它可以表示疾病与症状之间的概率关系。

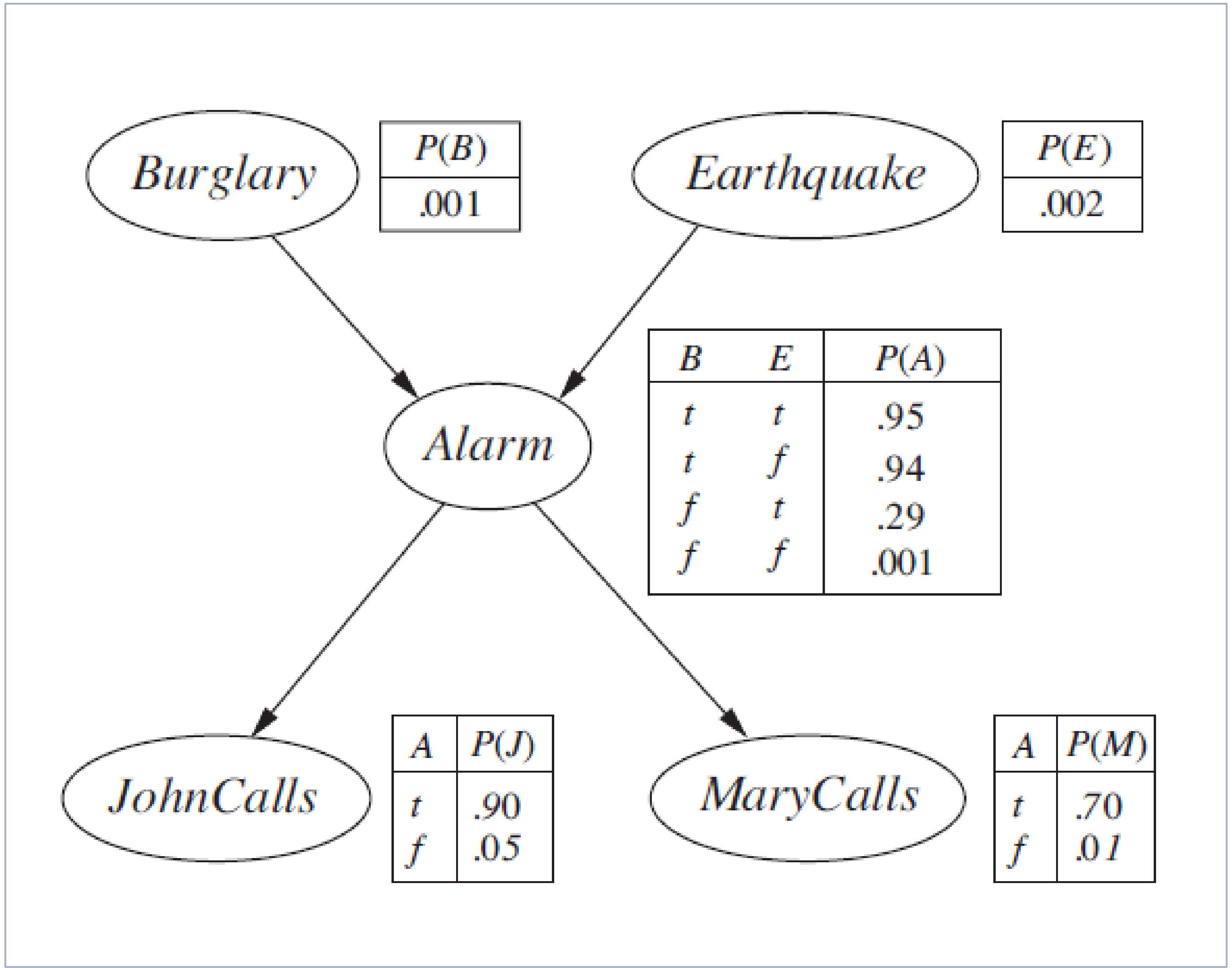
Two Views 两个观点

- The two views to understand semantics of Bayesian networks:
  理解贝叶斯网络语义的两个观点：
  - 1st: to view the network as a representation of the joint probability distribution.
    第一、将该网络视为一种联合概率分布的表示。
  - 2nd: to view it as an encoding of a collection of conditional independence statements.
    第二、将其视为一组条件独立语句的一种编码。
- The two views are equivalent, but: 这两个观点是等价的，但是：
  - 1st view: helpful in understanding how to construct networks,
    第一种观点：有助于理解如何构建网络，
  - 2nd view: helpful in designing inference procedures.
    第二种观点：有助于设计推理过程。

Example: A typical Bayesian network 一个典型的贝叶斯网络

- A burglar alarm installed at home, used to detect a burglary or minor earthquakes.
  房子里安装了一个防盗报警器，用于检测被盗或地震。
- Two neighbors, John and Mary, who have promised to call you at work when they hear the alarm.
  有两个邻居，John和Mary，他们已答应当听到报警时，就给你的办公室打电话。
- Variables 变量： *Burglar*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*.
- Network topology reflects the knowledge:
  网络的拓扑结构要反应如下知识：
  - A *Burglar* or an *Earthquake* can set the *Alarm*.
    盗窃或者地震会导致报警。
  - The *Alarm* can cause *Mary* or *John* to call.
    报警会引起Mary或John打电话。

Example: A typical Bayesian network 一个典型的贝叶斯网络



A Bayesian network with the conditional probability tables (CPTs).

一个具有条件概率表 (CPTS) 的贝叶斯网络。

Where 其中

- *B, E, A, J, M* stand for *Burglary, Earthquake, Alarm, JohnCalls, MaryCalls*;
  B, E, A, J, M 代表 *Burglary, Earthquake, Alarm, JohnCalls, MaryCalls*
- *t* and *f* stand for *true* and *false*;
  t 与 *f* 代表 *true* 和 *false*
- Each row shows one number p for $X_i = t$, (the number for $X_i = f$ is just $1 - p$).
  每一行显示 $X_i = t$ 的概率值 p，$X_i = f$ 时概率值则为 $1\text{-}p$

Representing Full Joint Distribution 表征全联合分布

- By the product of the elements of conditional distributions:
  采用条件分布元素的乘积：

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$

- To illustrate this, we can calculate the probability: Alarm has sounded, but neither a Burglary nor an Earthquake has occurred, and both John and Mary call.
  为了说明，我们可以计算概率：报警响起，但盗窃和地震都未发生，而John和Mary打电话。
- We multiply entries from the joint distribution:
  我们依据联合分布，将这些项相乘：

$$P(j, m, a, \neg b, \neg e) = P(j \mid a)\, P(m \mid a)\, P(a \mid \neg b \wedge \neg e)\, P(\neg b)\, P(\neg e)$$

$$= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.000628$$

where j, m, a, b, e stand for *JohnCalls*, *MaryCalls*, *Alarm*, *Burglary*, *Earthquake*.

其中 j, m, a, b, e 代表 *JohnCalls*, *MaryCalls*, *Alarm*, *Burglary*, *Earthquake*.

Constructing Bayesian Networks 构建贝叶斯网络

- First, rewrite joint distribution in terms of conditional probability, using product rule:
  首先，用条件概率公式对联合概率进行改写，使用如下乘积规则：

$$P(a \wedge b) = P(a \mid b)\, P(b)$$

$$P(x_1 , \ldots, x_n) = P(x_n \mid x_{n\text{-}1} , \ldots, x_1)\, P(x_{n\text{-}1} , \ldots, x_1)$$

- Then, repeat the process, reducing each conjunctive probability to a conditional probability and a smaller conjunction. We end up with one big product.
  然后，重复这个过程，将每个合取概率缩减为条件概率和较小的合取。最终得到一个大的乘积。

$$P(x_1 , \ldots, x_n) = P(x_n \mid x_{n\text{-}1} , \ldots, x_1)\, P(x_{n\text{-}1} \mid x_{n\text{-}2} , \ldots, x_1) \ldots P(x_2 \mid x_1)\, P(x_1)$$

$$= \prod_{i=1}^{n} P(x_i \mid x_{i-1} , \ldots, x_1)$$

This is called chain rule, for any set of random variables.
这被称为对任意一组随机变量的链式法则。

Constructing Bayesian Networks 构建贝叶斯网络

- The specification of the joint distribution is equivalent to the general assertion that, for every variable X_i in the network,

  联合分布的规格等同于一般性断言，即：对于网络中的每个变量X_i，

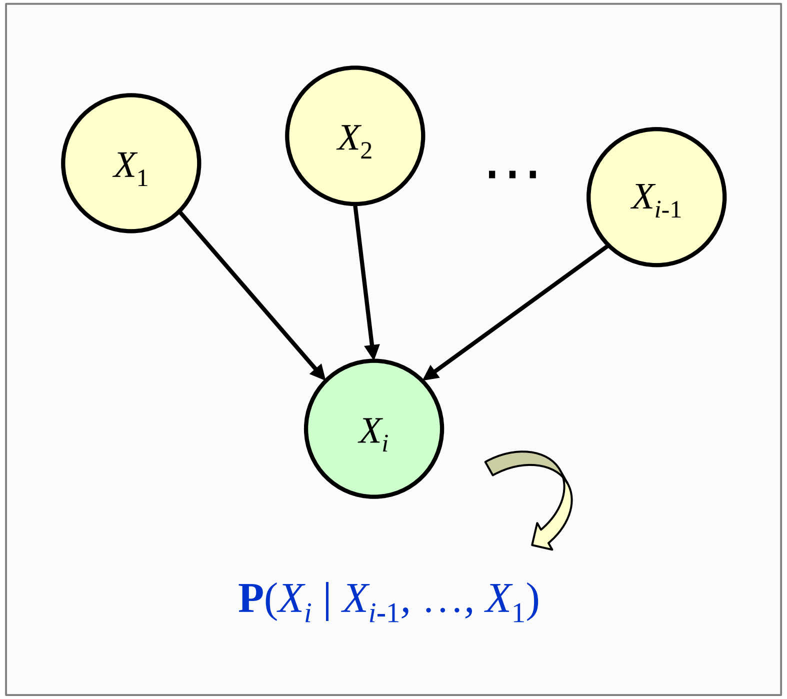
  $$\mathbf{P}(X_i \mid X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i \mid Parents(X_i))$$

  provided that 假设

  $$Parents(X_i) \subseteq \{X_{i-1}, \ldots, X_1\}$$

- The equation says: Bayesian network is a correct representation of domain, only if each node is conditionally independent of its other predecessors in the node ordering, given its parents.

  该等式说明：贝叶斯网络是域的一个正确表示，仅当给定其父节点，每个节点条件独立于节点序中其它前趋节点时。



Constructing Bayesian Networks 构建贝叶斯网络

- We can satisfy this condition with this methodology:
  我们可以用如下方法来满足该条件：
  - 1. *Nodes*: First determine the set of variables to model the domain, $\{X_1, \ldots, X_n\}$.
    节点：先确定要对域建模的变量集，$\{X_1, \ldots, X_n\}$。
  - 2. *Links*: For $i = 1$ to n do: 链接：从$i = 1$至n，做：
    - Choose, from $X_1, \ldots, X_{i\text{-}1}$, a minimal set of parents for X_i.
      从 $X_1, \ldots, X_{i\text{-}1}$ 中选择 X_i 的父节点的最小集。
    - For each parent insert a link from the parent to X_i.
      对每个父节点插入一个从该父节点至 X_i 的链接。
    - Write down the conditional probability table (CPT),
      记录 该条件概率表 (CPT)，

$$\mathbf{P}(X_i \mid Parents(X_i)).$$

The parents of node X_i should contain all nodes in $X_1, \ldots, X_{i\text{-}1}$ that directly influence X_i.
节点 X_i 的父辈应该包含直接影响 X_i 的 $X_1, \ldots, X_{i\text{-}1}$ 中的所有节点。

Example: A typical Bayesian network 一种典型的贝叶斯网络

- Suppose we have completed the network, except for the choice of parents for *MaryCalls*.

  假设除了*MaryCalls*父节点的选择之外，我们已经完成了该网络。

- *MaryCalls* is not directly influenced by a *Burglary* or an *Earthquake*. Her calling behavior only through their effect on the alarm.

  *MaryCalls*并非直接受盗窃或地震的支配，她打电话的行为只受它们对报警器的影响。

- Also, given the alarm state, whether John calls has no influence on Mary's calling.

  并且，给定报警状态，John打电话与否并不影响Mary的电话。

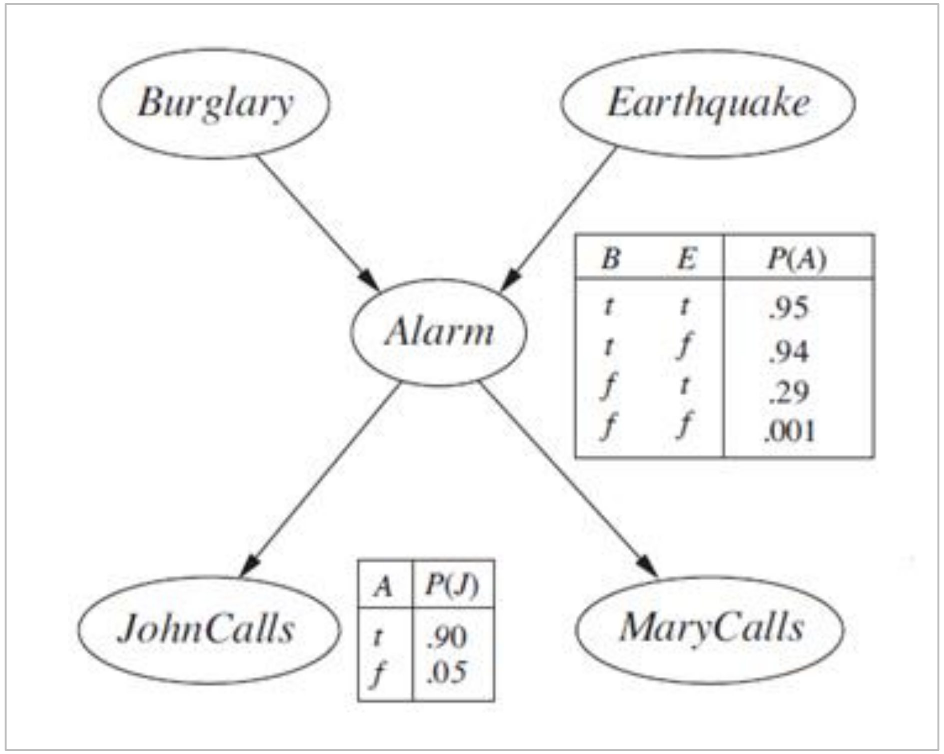
- Therefore, the following *conditional independence statement* holds:

  因此，如下条件独立语句成立：

$$\mathbf{P}(MaryCalls \mid JohnCalls, Alarm, Earthquake, Burglary) = \mathbf{P}(MaryCalls \mid Alarm)$$

Compactness 紧凑性

- A Bayesian network is far more *compact* than full joint distribution. Its compactness is a general property of locally structured (also called sparse) systems.
  贝叶斯网络远比全联合分布紧凑。它的紧凑性是局部结构化（也称为稀疏）系统的一般特性。
- In a locally structured system, each subcomponent interacts directly with only a bounded number of other components.
  在局部结构化系统中，每个子成分仅直接与其它成分的有限数量打交道。
- In Bayesian networks, a CPT (conditional probability table) for a Boolean variable with k parents has 2^k rows.
  贝叶斯网络中，具有k个父节点的布尔变量的CPT（条件概率表）有2^k行。

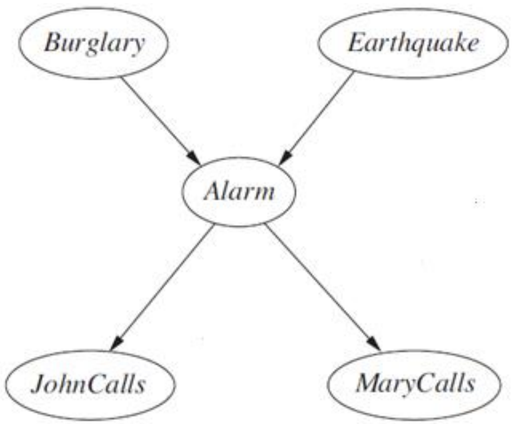


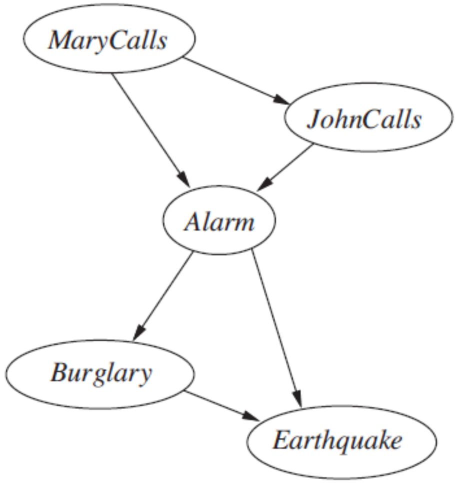
Compactness 紧凑性

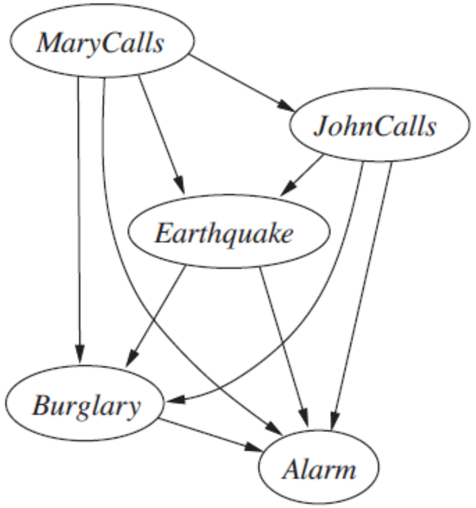
- Assume there are n Boolean variables: 假设有n个布尔变量：
  - Bayesian networks: 贝叶斯网络
    if each variable has no more than k parents, then the complete network can be specified by $n2^k$ numbers.
    如果每个变量的父节点不超过k个，则全部网络可以用 $n2^k$ 个数指定。
  - Full joint distribution: 全联合分布
    it contains 2^n numbers.
    它包含 2^n 个数。
- E.g., suppose $n = 30$ nodes, each with $k = 5$ parents, then
  例如：设节点$n = 30$，每个具有 $k = 5$ 个父节点，则
  - Bayesian network: $n2^k = 30 \times 2^5 = 960.$
  - Full joint distribution: $2^n = 2^{30} = 1{,}073{,}741{,}824.$

Node Ordering 节点排序

- We will get a compact Bayesian network only if choose node ordering well (Fig. a).
  只有当选择一个好的节点排序（图a），我们才会得到一个紧凑的贝叶斯网络。
- What happens if we choose the wrong order (Fig. b and c).
  如果我们选择一个差的排序会发生什么（图b和c）。







(a) A typical Bayesian network. 一个典型的贝叶斯网络
(b) A bad node ordering 一个差的节点排序： *MaryCalls, JohnCalls, Alarm, Burglary, Earthquake.*
(c) A very bad node ordering 一个更差的节点排序： *MaryCalls, JohnCalls, Earthquake, Burglary, Alarm.*

Conditional Independence Relations 条件独立关系

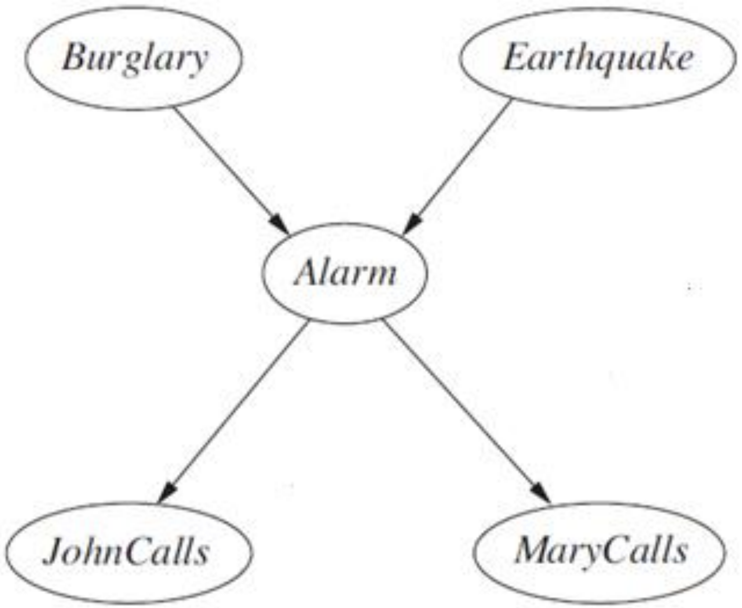
☐ Numerical semantics: 数值语义

a node is conditionally independent of its other predecessors, given its parents.

一个节点，给定其父节点后，条件独立于其它前趋节点。

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$

Example 举例



$$\begin{aligned} & P(j, m, a, \neg b, \neg e) \\ &= P(j \mid a)\, P(m \mid a)\, P(a \mid \neg b \wedge \neg e)\, P(\neg b)\, P(\neg e) \\ &= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 \\ &= 0.000628 \end{aligned}$$

where j, m, a, b, e stand for 其中j, m, a, b, e代表

JohnCalls, MaryCalls, Alarm, Burglary, Earthquake.

Conditional Independence Relations 条件独立关系

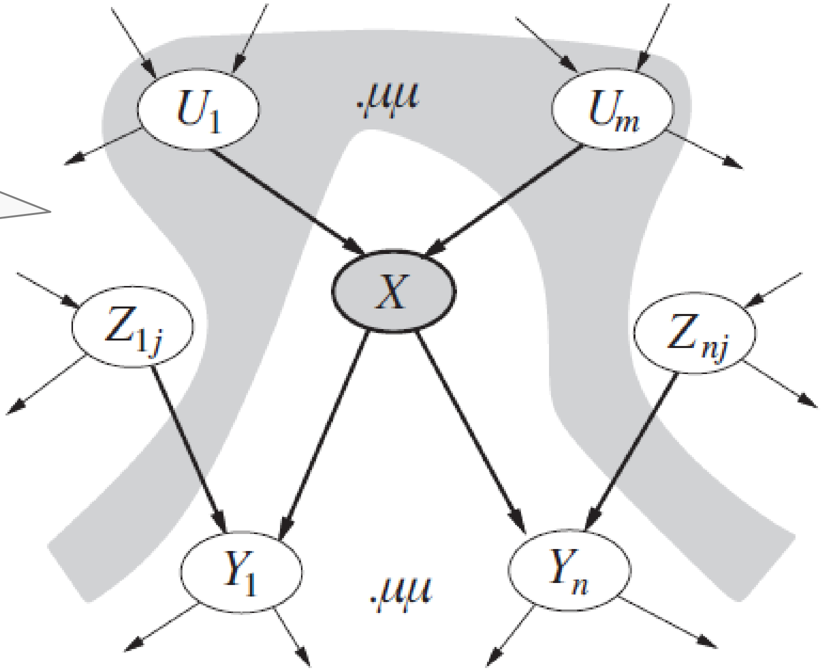
- Topological semantics: 拓扑语义

  each node is conditionally independent of its non-descendants, given its parents.

  每个节点，给定其父节点后，条件独立于它的非后继节点。

A node X is conditionally independent of its non-descendants Z_{ij}, given its parents U_i (shown in the gray area).

节点X，给定其父节点U_i（灰色区域所示）后，条件独立于它的非后继节点Z_{ij}。



Numerical semantics ⇔ Topological semantics

数值语义 ⇔ 拓扑语义

Conditional Independence Relations 条件独立关系

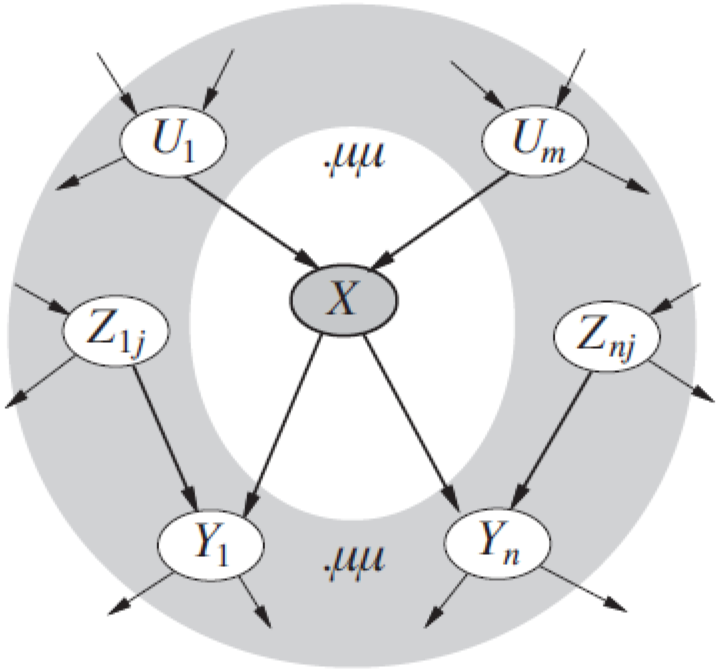
- Markov blanket 马尔科夫覆盖

  a node is conditionally independent of all other nodes in the network, given its Markov blanket (i.e. parents, children, and children's parents).

  一个节点，给定其马尔科夫覆盖后，条件独立于网络中的所有其它节点（即：父节点、子节点、以及子节点的其它父节点）。

A node X is conditionally independent of all other nodes in the network given its Markov blanket (the gray area).

一个节点X，给定马尔科夫覆盖（灰色区域）后，条件独立于网络中的所有其它节点。



Thank you for your attention!

AI