

Introduction to Artificial Intelligence



COMP307

Uncertainty and Probability 3: Introduction to Bayesian Network

Yi Mei

yi.mei@ecs.vuw.ac.nz

Outline

- Rules from previous lectures
- What is Bayesian Networks
- Why Bayesian Networks
- Cause — Effect
- Summary



Thomas Bayes ([/ˈbeɪz/](#); c. 1701 – 7 April 1761)

Rules from Previous Lectures

- **Product Rule**
 - $P(X,Y) = P(X)*P(Y | X) = P(Y)*P(X | Y)$
- **Sum Rule:**
 - $P(X) = \sum_y P(X, Y)$
- **Normalisation Rule**
 - $\sum_x P(X) = 1, \sum_x P(X|Y) = 1$
- **Independence**
 - $X \perp Y, P(X|Y) = P(X), P(X, Y) = P(X) * P(Y)$
 - $X \perp Y | Z, P(X | Y, Z) = P(X|Z), P(X, Y|Z) = P(X|Z) * P(Y|Z)$
- **Bayes Rule**
 - $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
 - $P(Y|X_1, \dots, X_n) = \frac{P(X_1|Y)...P(X_n|Y)P(Y)}{P(X_1,...,X_n)}$ [assume conditional independence]

A Lazy Detective

- Early one morning, Mr. Boddy was found dead



from both the maid and the butler, the only two individuals who could have possibly committed the crime.



- Possible weapons:
 - Vacuum cleaner pipe / Candle stick
- Possible times:
 - Evening / Midnight
- Possible murderer:
 - Maid / Butler

Variable	Domain
Weapon	{VCP, CS}
Time	{Evening, Midnight}
Murderer	{Maid, Butler}

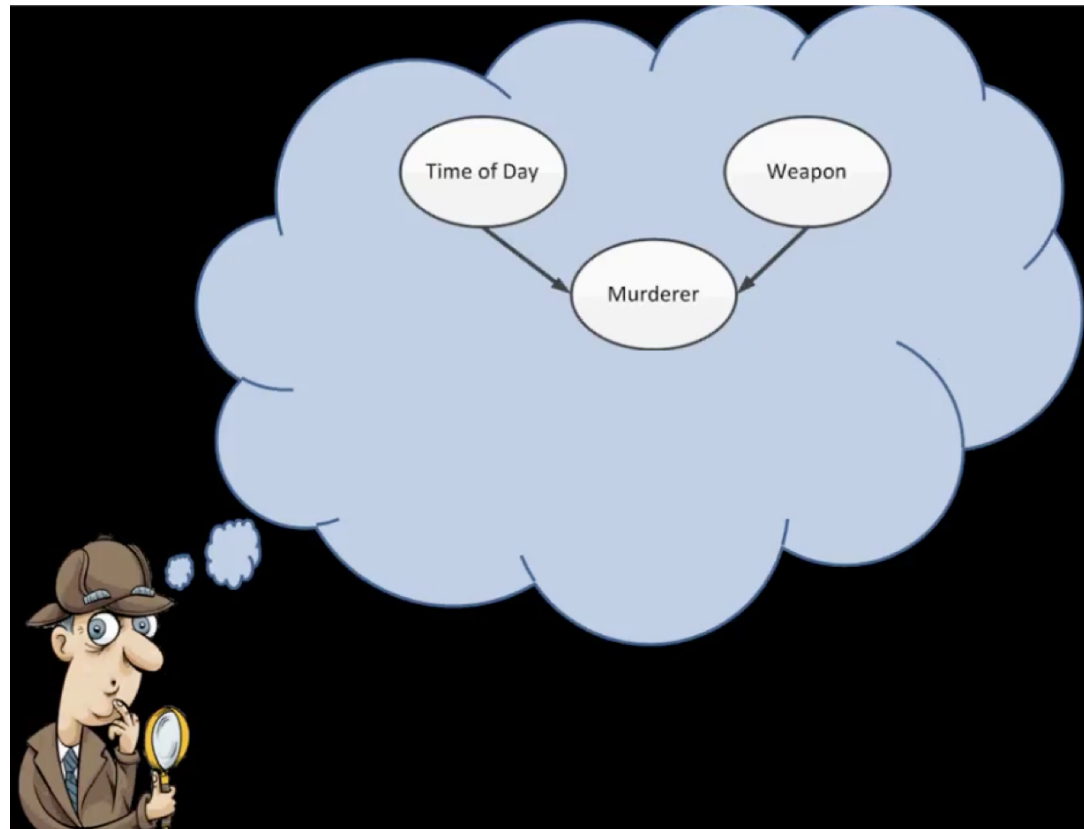
A Lazy Detective

- Not wanting to stay in the scene any longer
- Not wanting to gather more information
- Reasoning based on only the current information
- If the **weapon** is **vacuum cleaner pipe**, then the **murderer** is very likely to be the **maid**
- If the **weapon** is **candle stick**, then the **murderer** is very likely to be the **butler**
- If the murder **time** is **evening**, then the murderer is likely to be the **maid**
- If the murder **time** is **midnight**, then the **murderer** is likely to be the **butler**
- ...

Conditional Dependencies

- Model the conditional dependencies between the variables as a directed acyclic graph (DAG)
 - Time of Day -> Murder
 - Weapon -> Murder
 - Time of Day and Weapon are independent

- What is missing?



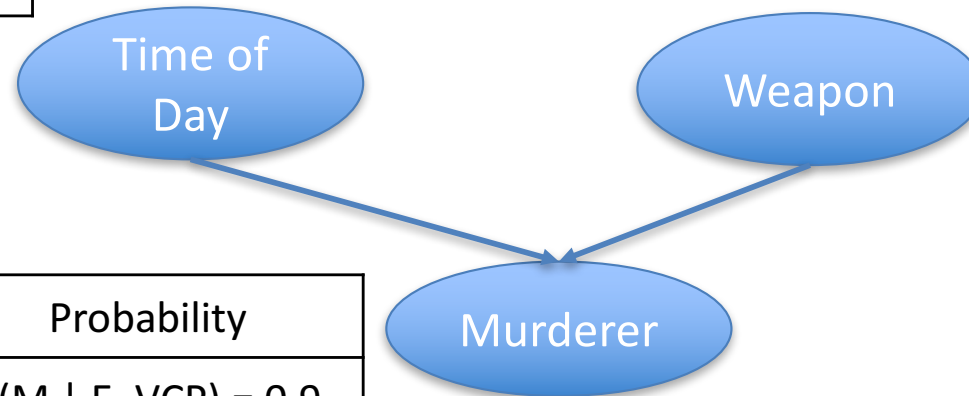
Bayesian Networks

- Beliefs as conditional probabilities
 - Report from the lab:
 - $P(\text{Time} = \text{evening}) = 0.05$, $P(\text{Time} = \text{midnight}) = 0.95$
 - $P(\text{Weapon} = \text{VCP}) = 0.8$, $P(\text{Weapon} = \text{CS}) = 0.2$
 - From detective
 - $P(\text{Murderer} = \text{maid} \mid \text{Time} = \text{evening}, \text{Weapon} = \text{VCP}) = 0.9$
 - $P(\text{Murderer} = \text{maid} \mid \text{Time} = \text{evening}, \text{Weapon} = \text{CS}) = 0.55$
 - $P(\text{Murderer} = \text{maid} \mid \text{Time} = \text{midnight}, \text{Weapon} = \text{VCP}) = 0.35$
 - $P(\text{Murderer} = \text{maid} \mid \text{Time} = \text{midnight}, \text{Weapon} = \text{CS}) = 0.05$
 - $P(\text{Murderer} = \text{butler} \mid \text{Time} = \text{evening}, \text{Weapon} = \text{VCP}) = 0.1$
 - $P(\text{Murderer} = \text{butler} \mid \text{Time} = \text{evening}, \text{Weapon} = \text{CS}) = 0.45$
 - $P(\text{Murderer} = \text{butler} \mid \text{Time} = \text{midnight}, \text{Weapon} = \text{VCP}) = 0.65$
 - $P(\text{Murderer} = \text{butler} \mid \text{Time} = \text{midnight}, \text{Weapon} = \text{CS}) = 0.95$

Bayesian Networks

Time	Probability
Evening	$P(T = E) = 0.05$
Midnight	$P(T = M) = 0.95$

Weapon	Probability
VCP	$P(W = VCP) = 0.8$
CS	$P(W = CS) = 0.2$



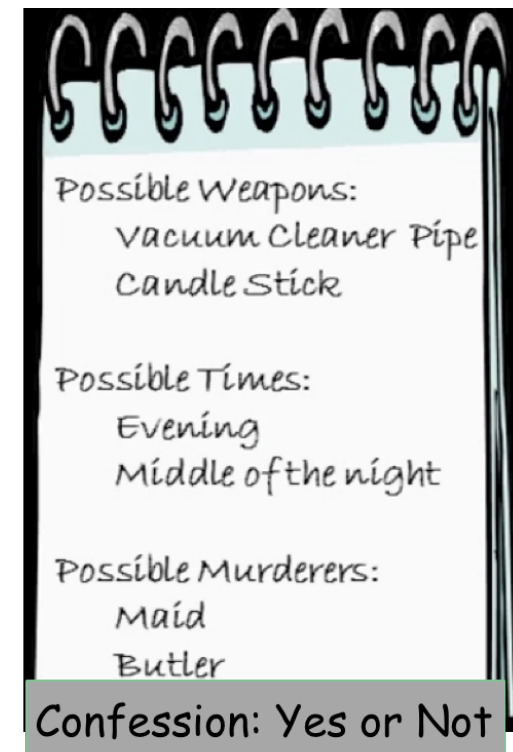
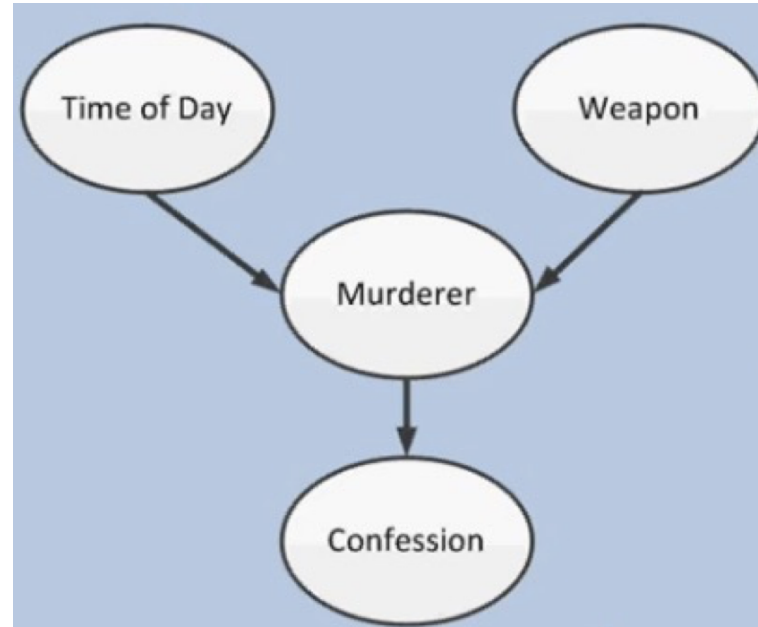
T	W	Murderer	Probability
E	VCP	Maid	$P(M E, VCP) = 0.9$
E	CS	Maid	$P(M E, CS) = 0.55$
M	VCP	Maid	$P(M M, VCP) = 0.35$
M	CS	Maid	$P(M M, CS) = 0.05$
E	VCP	Butler	$P(B E, VCP) = 0.1$
E	CS	Butler	$P(B E, CS) = 0.45$
M	VCP	Butler	$P(B M, VCP) = 0.65$
M	CS	Butler	$P(B M, CS) = 0.95$

Bayesian Networks

- Bayesian networks (BNs): a **graphical** representation of a **probabilistic dependency model**
 - also known as **Belief networks** (or Bayes nets for short)
 - Belong to the family of **probabilistic graphical models** (GMs).
 - Other GMs: Markov network ...
- These **graphical structures** are used to **represent knowledge about an uncertain domain**.
 - each **node** in the graph represents a **random variable**,
 - the **edges** between the nodes represent **probabilistic dependencies** among the corresponding random variables.
 - The **conditional dependencies** in the graph are often **estimated** by using known statistical and computational methods.
- BNs combine principles from graph theory, probability theory, computer science, and statistics.

Bayesian Networks

- Each **node** or **variable** may take one of a number of **possible states or values**.
- The **belief** in, or certainty of, each of these values is determined from the belief in **each possible value of every node directly connected to it** and **its relationship** with each of these nodes.
- The **belief** in each state of a node is **updated** whenever the belief in each state of any directly connected node **changes**.



Semantics of Bayesian Networks

- A set of **nodes**, one for a **variable** X
- A **directed, acyclic** graph
 - Each edge shows the **direct** influence between **parent** and **child**
 - A **child** depends on its parents
- A **conditional probability table** for each node
 - a collection of distributions over X , one for **each combination of parents values**
分布在 X 上的一组分布，每个分布对应于父值的组合
 - $$P(X \mid a_1, \dots, a_n)$$
 - (usually) description of a “causal” process

A Bayes Net = Topology (graph) + Local Conditional Probabilities

贝叶斯网络 = 拓扑(图) + 局部条件概率

Why Bayesian Networks

- Several advantages for data analysis:
 - the model encodes **dependencies among all variables**, it readily handles situations where some data entries are missing.
 - a Bayesian network can be used to learn **causal** relationships, and hence can be used to gain **understanding about a problem** domain and to **predict the consequences of intervention**.
 - the model has both a **causal and probabilistic semantics**, it is an ideal representation for **combining prior knowledge** (which often comes in causal form) and data.
 - Bayesian statistical methods in conjunction with Bayesian networks offer an efficient and principled approach for avoiding the **overfitting of data**.

数据分析的几个优点:

- 模型对所有变量之间的依赖关系进行编码，它很容易处理某些数据条目丢失的情况。
- 贝氏网路可以用来学习因果关系，因此可以用来了解问题领域和预测干预的后果。
- 该模型具有因果语义和概率语义(causal & probabilistic semantics)，是结合先验知识combing prior knowledge (经常以因果形式出现)和数据的理想表示。
- 与贝叶斯网络结合的贝叶斯统计方法bayesian statistical methods为避免数据的过度拟合提供了一个有效和有原则的方法。

Cause and Effect

- A **Cause** is why something happens. An **Effect** is what actually happens (results of the Cause).
- A patient got a flu, and had a fever (high temperature).
 - Flu (disease) is the cause.
 - High temperature (symptom) is the effect.
- **Causal Reasoning**: solving a problem where only cause is known 因果推理: 解决一个只知道原因的问题
 - $P(\text{Effect} \mid \text{Cause})$
- **Diagnostic Reasoning**: reasoning about Cause when Effect is known 诊断推理: 在已知结果的情况下对原因进行推理
 - $P(\text{Cause} \mid \text{Effect})$
- **Inter-causal Reasoning**: reasoning about the interactions between multiple causes influences 相互因果推理: 关于多种原因影响之间相互作用的推理

Cause and Effect

- Bayesian rules:
 - $P(\text{Cause} \mid \text{Effect}) = P(\text{Effect} \mid \text{Cause}) * P(\text{Cause}) / P(\text{Effect})$
 - $P(\text{Cause})$ often called **prior**
 - $P(\text{Cause} \mid \text{Effect})$ is known as the **posterior**
 - $P(\text{Effect} \mid \text{Cause})$ is known as the **likelihood**

The diagram illustrates Bayes' theorem with the following components and labels:

- Posterior:** $P(\text{Cause} \mid \text{Effect})$ (indicated by a downward arrow from the left box)
- Likelihood:** $P(\text{Effect} \mid \text{Cause})$ (indicated by an upward arrow from the first term in the numerator)
- Prior:** $P(\text{Cause})$ (indicated by an upward arrow from the second term in the numerator)
- Marginal likelihood or normalisation:** $P(\text{Effect})$ (indicated by a downward arrow from the denominator)

$$P(\text{Cause} \mid \text{Effect}) = \frac{P(\text{Effect} \mid \text{Cause}) \times P(\text{Cause})}{P(\text{Effect})}$$

Cause and Effect

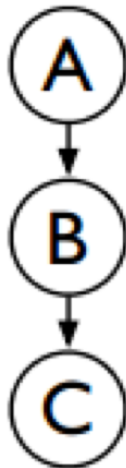
- Different conditional dependencies 不同的条件依赖性

Direct cause



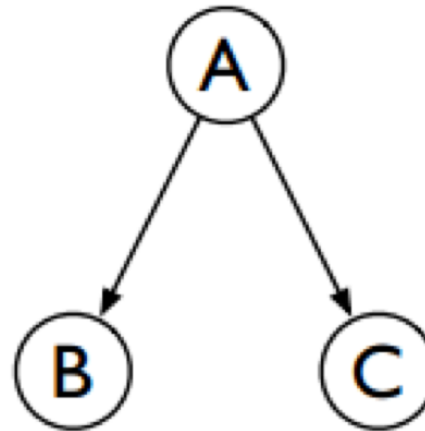
$$P(B|A)$$

Indirect cause



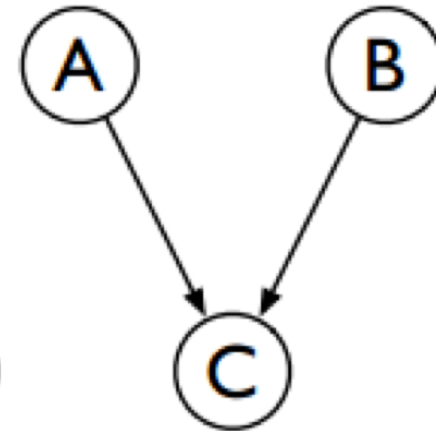
$$P(B|A) \\ P(C|B)$$

Common cause



$$P(B|A) \\ P(C|A)$$

Common effect



$$P(C|A,B)$$

- Common effect (multiple causes or “explaining away”. Suppose that there are exactly two possible causes of a particular effect, represented by a v-structure)

- The causes are independent if the effect is unknown
- Are the causes independent if the the effect is known?

共同影响（多重原因或“解释消失”）。假设一个特定的效应有两个可能的原因，用一个v结构表示

——如果效应未知，原因是独立的

——如果效应已知，原因是独立的吗？

Bayesian Networks Example

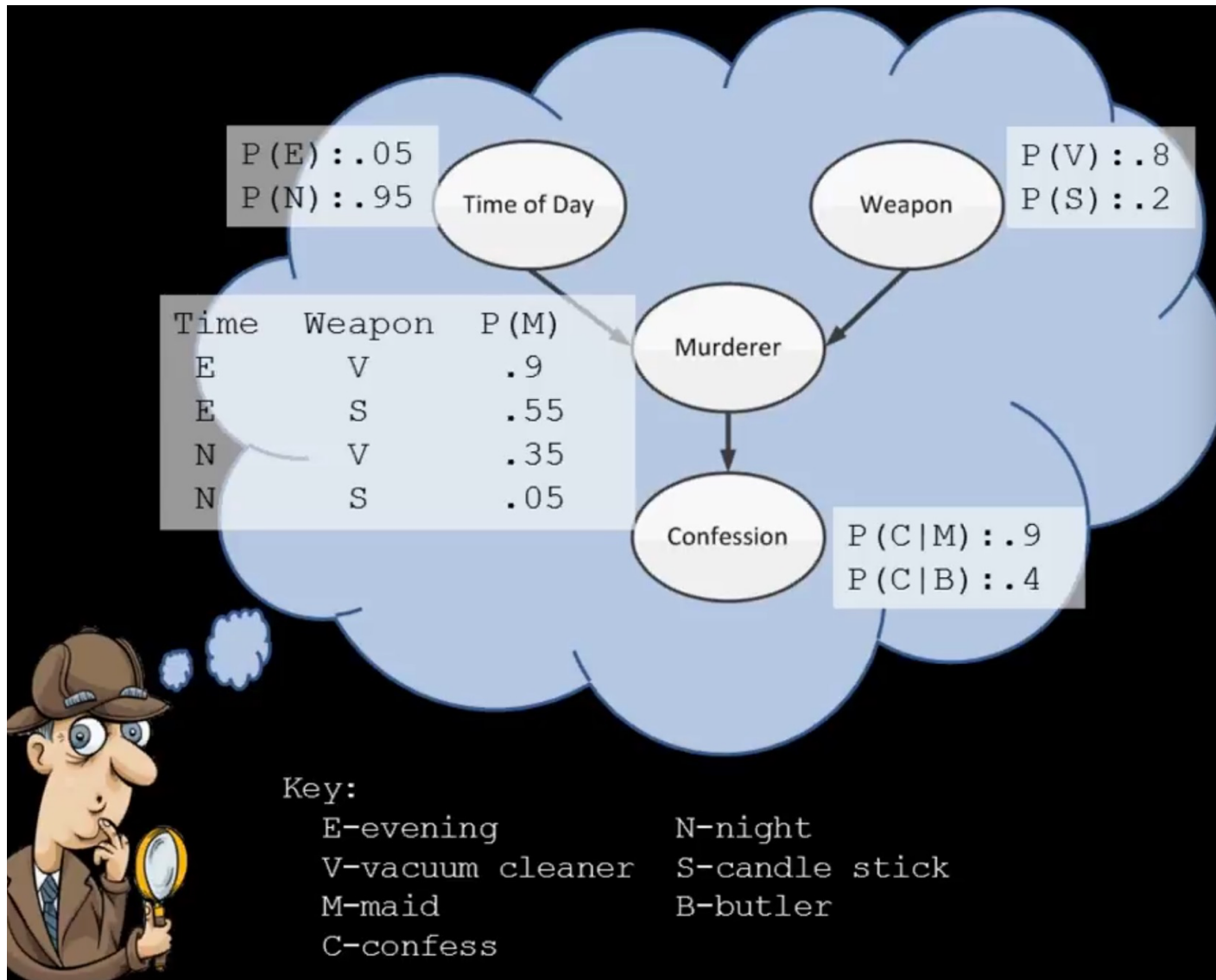
- Calculate the probability of the **murderer** being the **maid**

$$\begin{aligned}P(M) &= P(M, E, VCP) + P(M, E, CS) + P(M, M, VCP) + P(M, M, CS) \\&= P(M|E, VCP)P(E, VCP) + P(M|E, CS)P(E, CS) + \\&\quad P(M|M, VCP)P(M, VCP) + P(M|M, CS)P(M, CS) \\&= P(M|E, VCP)P(E)P(VCP) + P(M|E, CS)P(E)P(CS) + \\&\quad P(M|M, VCP)P(M)P(VCP) + P(M|M, CS)P(M)P(CS) \\&= 0.9 \times 0.05 \times 0.8 + 0.55 \times 0.05 \times 0.2 + \\&\quad 0.35 \times 0.95 \times 0.8 + 0.05 \times 0.95 \times 0.2 \\&= 0.317\end{aligned}$$

- Which rules are used here?
- What is $P(B)$?
- Time and Weapon are independent if the murderer is unknown. Are they still independent if the murderer is known?

Bayesian Networks Example

- Calculate the probability that **the murderer confesses**



Summary

- Bayesian networks
 - A directed acyclic graph
 - Represent conditional dependencies between variables
 - Conditional distribution tables
- Cause and effect
 - Different relationships
- How to calculate probabilities in a Bayesian network
- Next lecture: how to build a BN