

Beijing Forest Studio
北京理工大学信息系统及安全对抗实验中心



贝叶斯网络

贝叶斯网络

硕士研究生 苏霞

2020年06月21日



- 背景简介
- 基本概念
- 算法原理
- 优劣分析
- 应用总结
- 参考文献



- 预期收获
 - 1. 深入理解贝叶斯网络的原理。
 - 2. 了解朴素贝叶斯和马尔可夫链。

- **逆概问题**

- **正向概率**：假设袋子里有 N 个白球， M 个黑球，摸出黑球的概率是多大
- **逆向概率**：假设我们事先不知道袋子里黑白球的比例，摸出好几个球后观察这些球的颜色，那么我们可以对袋子里的黑白球的比例作出什么样的推测

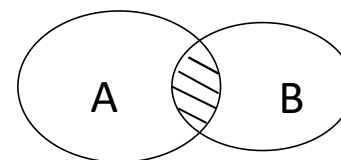


- 频率派
 - 认为总体分布参数 θ 是固定的，而样本 X 是随机的
- 贝叶斯派
 - 认为 θ 是随机的，而样本 X 是固定的
 - 先验分布 $\pi(\theta)$ + 样本信息 $X \Rightarrow$ 后验分布 $\pi(\theta|X)$
 - 先验分布：总体分布参数 θ 的一个概率分布,一般来源于经验跟历史资料
 - 后验分布：在给定样本 X 的情况下的 θ 条件分布

- 概率复习

- 条件概率：事件A在事件B已经发生条件下的发生概率

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



- 联合概率：两个事件共同发生的概率

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

$$P(x_1, x_2, \dots, x_n)$$

$$= P(x_1|x_2, x_3, \dots, x_n)P(x_2|x_3, x_4, \dots, x_n) \cdots P(x_{n-1}|x_n)P(x_n)$$

- 概率复习
 - 绝对独立

$$P(A, B) = P(A)P(B)$$

- 条件独立：事件 A 和 B 对于给定事件 C 是条件独立的

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

$$\Leftrightarrow P(A|B, C) = P(A|C)$$

$$\Leftrightarrow P(B|A, C) = P(B|C)$$

- 贝叶斯公式
 - 公式表达

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

- 应用：拼写检查

A screenshot of a Google search interface. The search bar contains the text "Julw". Below the search bar, there are navigation tabs for "网页" (Web), "图片" (Images), "视频" (Videos), "新闻" (News), "地图" (Maps), "更多" (More), and "搜索工具" (Search Tools). The "网页" tab is selected. Below the tabs, it says "找到约 733,000 条结果 (用时 0.53 秒)". The main content area shows "显示的是以下查询字词的结果: July" and "仍然搜索: Julw". Below this, there is a search result snippet for "结构之法算法之道- 博客频道- CSDN.NET" with the URL "blog.csdn.net/v_JULY_v". The snippet text reads: "从头到尾彻底理解KMP 作者: July 时间: 最初写于2011年12月, 2014年7月21日晚10点 全部删除重写此文。 1. 引言本KMP原文最初写于2年多前的2011年12月, 因 ... 程序员面试、算法研究、编程艺术 - 程序员如何快速准备面试中的算法 - 目录视图 - 尾页".

- 贝叶斯公式
 - 应用：拼写检查

$$P(c|w) = \frac{P(w|c)P(c)}{P(w)}$$

c 表示拼写正确的情况， w 表示拼写错误的情况， w 已发生时， $P(w)$ 相同，只需比较：

$$P(w|c)P(c)$$

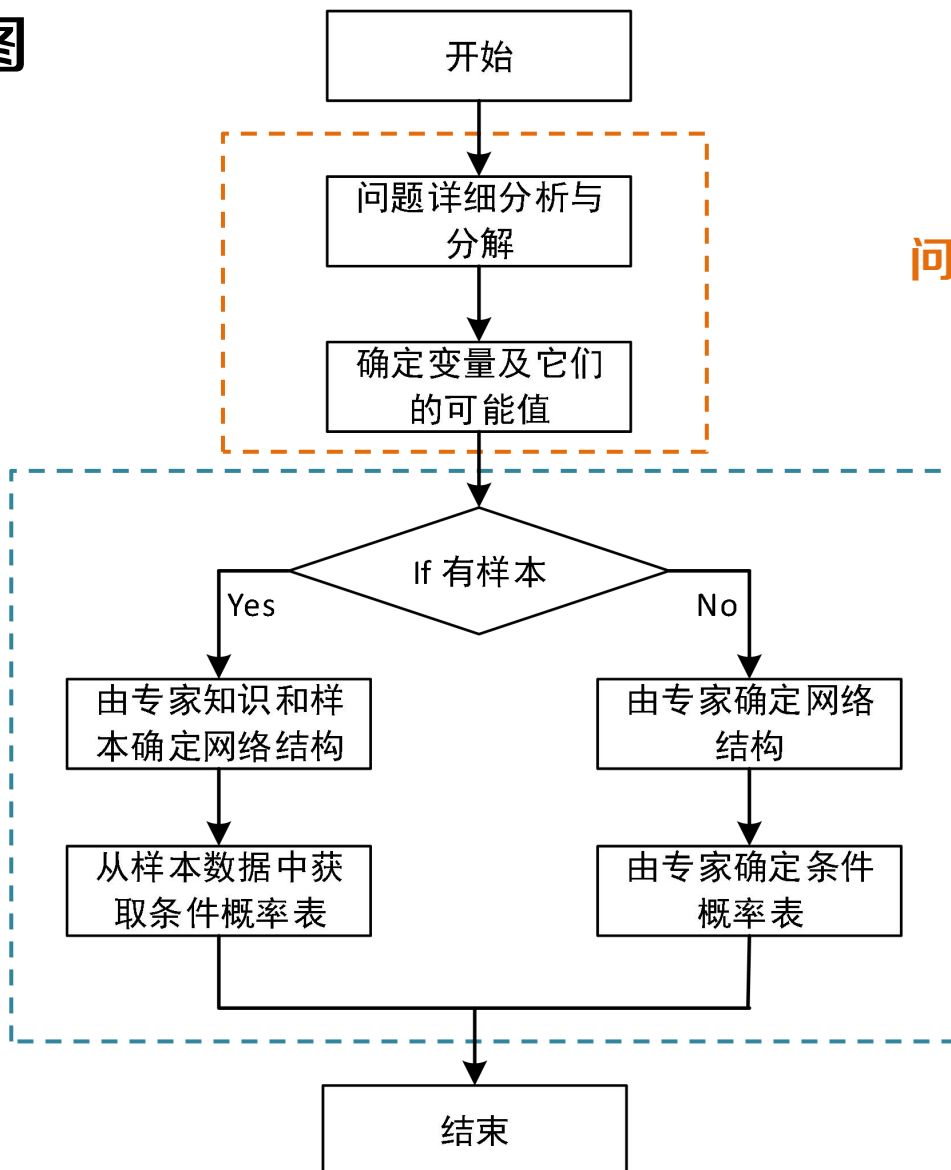
$P(c)$ 表示某个正确的词的出现概率， $P(w|c)$ 表示在试图拼写 c 的情况下，出现拼写错误 w 的概率

July -> julw Julie->julw

T	表达随机变量之间复杂的概率不确定性，发现数据间的潜在不确定性
I	先验知识或数据
P	1.确定随机变量间的拓扑关系 2.参数学习 3.进行概率推断
O	推断结果

P	尽可能使贝叶斯网络表达式和全联合概率等价
C	存在邻域专家及相应数据
D	如何根据数据和专家知识高效、准确的建立贝叶斯网络
L	

- 算法流程图



问题分析：节点选择

模型设计：结构和概率表

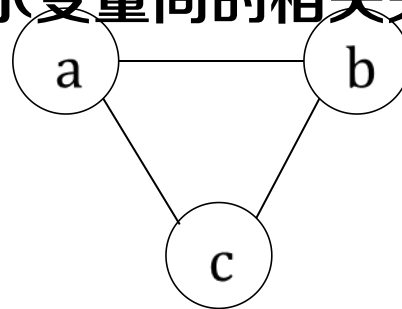
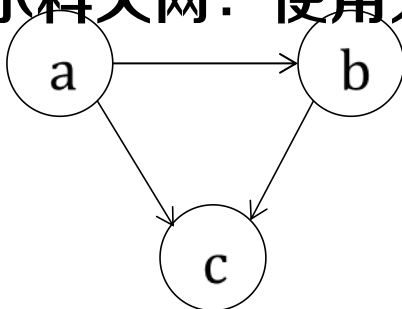
- **概率图模型 (Probabilistic Graphical Model, PGM)**

- 概率图模型是指一种用图结构来描述多元随机变量之间条件独立关系的概率模型。

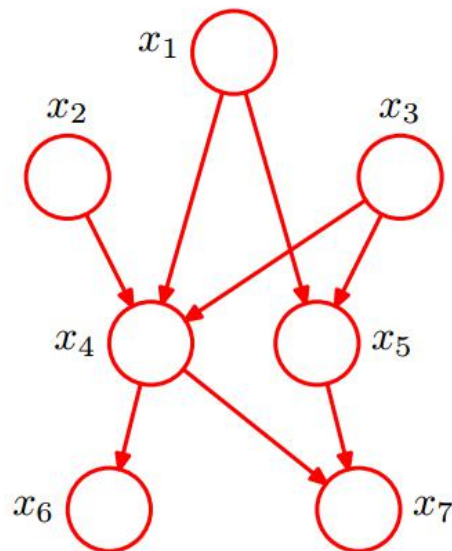
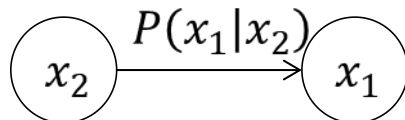
- 根据边的性质不同进行分类

- 贝叶斯网：使用有向无环图表示变量间的依赖关系

- 马尔科夫网：使用无向图表示变量间的相关关系

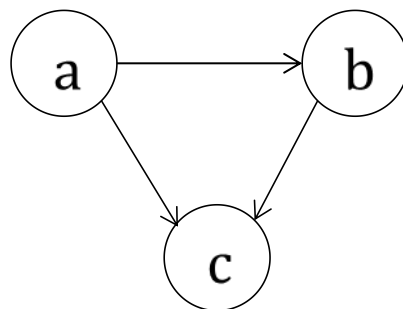


- 贝叶斯网络(Bayesian Network, BN)
 - 节点：随机变量
 - 有向边：依赖关系 (不一定是因果关系)
 - 每个节点有条件概率分布 $P(x_i|parents(x_i))$ ，对应一个条件概率表 (CPT)
 - $P(x_1, x_2, \dots, x_n) = \prod P(x_i|parents(x_i))$



- 一个简单的贝叶斯网络

$$P(a, b, c) = P(c|a, b)P(b|a)P(a)$$



- 贝叶斯网络(Bayesian Network, BN)

- 全联合分布

$$P(x_1, x_2, \dots, x_n) = P(x_1|x_2, x_3, \dots, x_n) \cdots P(x_{n-1}|x_n)P(x_n)$$

- 贝叶斯网络

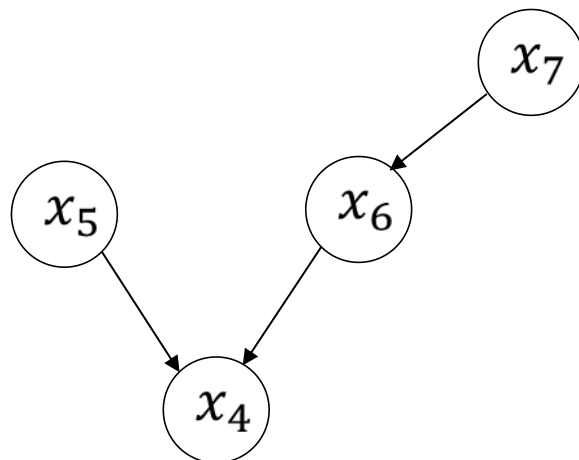
$$P(x_1, x_2, \dots, x_n) = \prod P(x_i|\text{parents}(x_i))$$

$$P(x_i|\text{parents}(x_i)) = P(x_i|x_{i+1}, x_{i+2}, \dots, x_n)$$
$$\Leftrightarrow P(x_i|x_{i+1}, x_{i+2}, \dots, x_n)$$

条件独立: $P(A|B, C) = P(A|C)$

每一个节点在已知父节点的情况下, 与其它非后代节点独立

- 贝叶斯网络(Bayesian Network, BN)

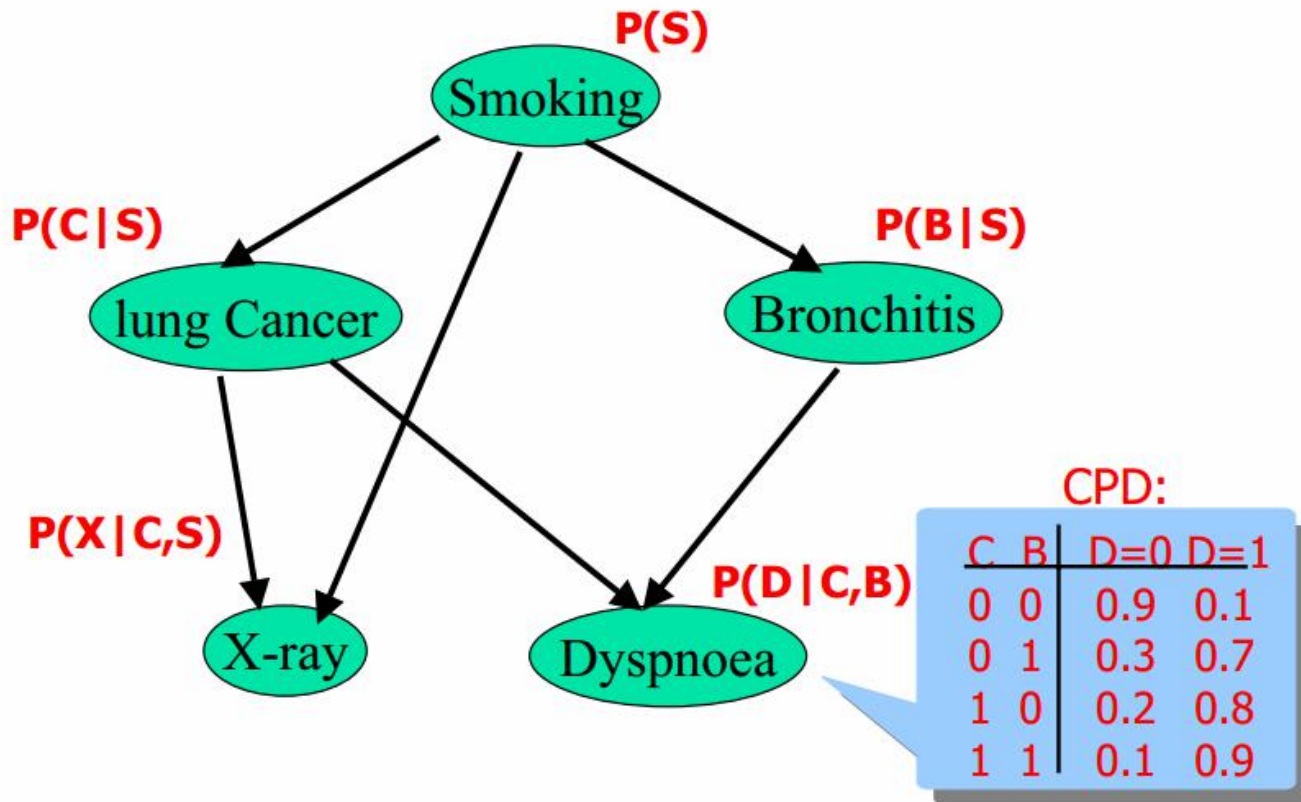


x_4 和 x_7 在已知 x_5 和 x_6 的条件下是相互独立的, 所以有:

$$P(x_4|x_5, x_6, x_7) = P(x_4|x_5, x_6) = P(x_4|parents(x_4))$$

$$P(x_4, x_5, x_6, x_7) = \prod P(x_i|parents(x_i))$$

- 一个实例



$$P(S, C, B, X, D) = P(S) P(C|S) P(B|S) P(X|C,S) P(D|C,B)$$

$$1+2+2+4+4=13 \text{ vs } 2^5-1$$

- **贝叶斯网络学习**

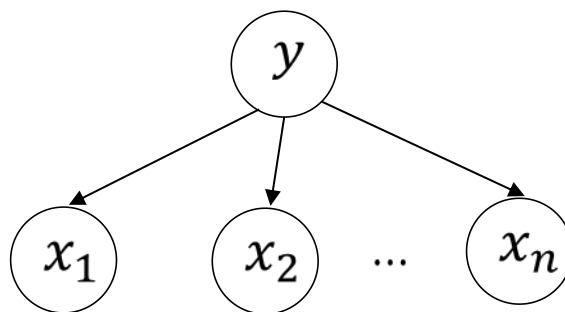
- **结构学习**

- **目标：网络拓扑结构的学习**
 - **方法：基于打分搜索的方法、基于依赖关系分析的方法**

- **参数学习**

- **目标：已知网络结构的条件下，学习每个节点的概率分布表**

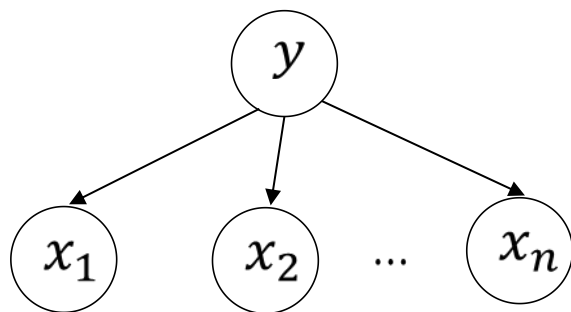
- 特殊的贝叶斯网络
 - 朴素贝叶斯



$$P(x_1, x_2, \dots, x_n, y) = P(x_1|y)P(x_2|y) \cdots P(x_{n-1}|y)P(y)$$

属性条件独立性假设：对已知类别，假设所有属性相互独立

- 特殊的贝叶斯网络
 - 例子：垃圾邮件判断

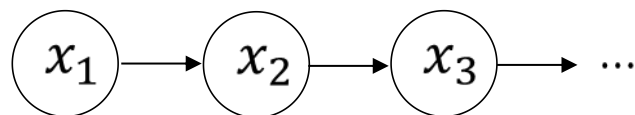


是否为垃圾邮件

邮件中出现的字符，如“钱”、“你好”等

$$P(y = 1|x_1, x_2, \dots, x_n) = \frac{P(x_1, x_2, \dots, x_n, y = 1)}{P(x_1, x_2, \dots, x_n)}$$
$$= \frac{P(x_1|y = 1)P(x_2|y = 1) \cdots P(x_{n-1}|y = 1)P(y = 1)}{P(x_1, x_2, \dots, x_n)}$$

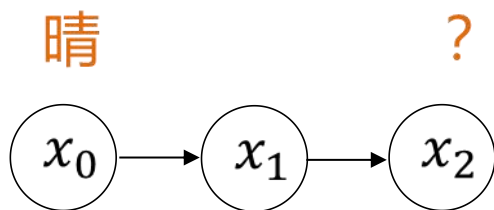
- 特殊的贝叶斯网络
 - 马尔可夫链



$$P(x_1, x_2, x_3 \dots) = P(x_1)P(x_2|x_1)P(x_3|x_2) \dots$$

x_t 只和 x_{t-1} 有关

- 特殊的贝叶斯网络
 - 例子：天气



x_{t-1}	x_t	$P(x_t x_{t-1})$
晴	晴	0.9
晴	雨	0.1
雨	晴	0.3
雨	雨	0.7

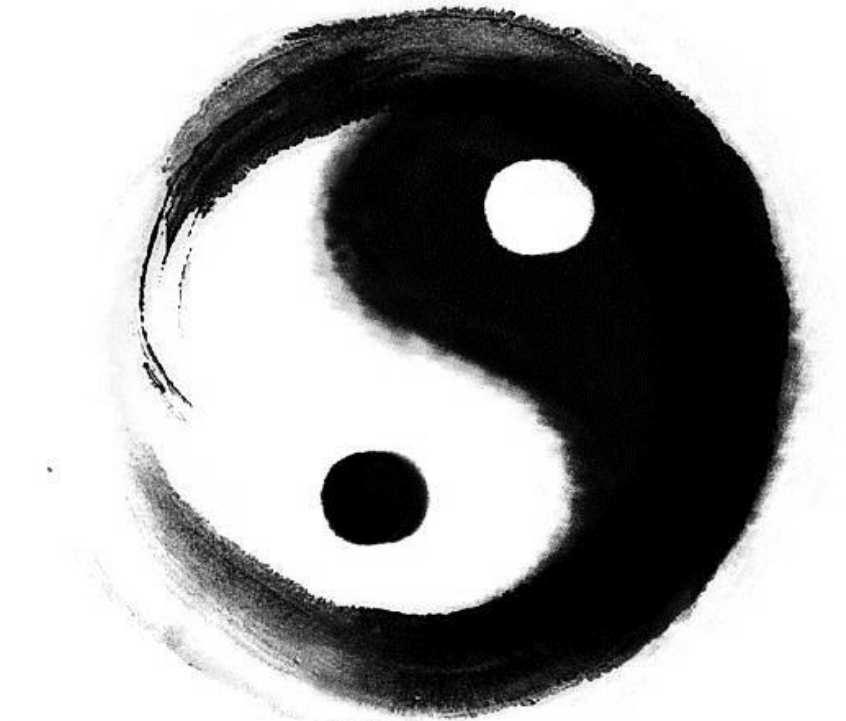
$$P(x_2 = \text{晴}) = \sum_{x_1} P(x_1, x_2 = \text{晴}) = \sum_{x_1} P(x_1)P(x_2 = \text{晴}|x_1)$$

- **贝叶斯网络优点**
 - 变量之间的关系表示形式更加直观
 - 只模拟那些具有依赖关系的事件的概率，节省了大量的计算
- **贝叶斯网络缺点**
 - 如何根据数据和专家知识高效、准确的建立贝叶斯网络
 - 目前对于这一类学习问题，主要有基于打分搜索的方法和基于依赖分析的学习方法，前者存在搜索空间巨大，可能收敛于局部最优解等问题，后者则存在节点之间的独立性或条件独立性判断困难等问题

- 用途
 - 诊断: $P(\text{病因}|\text{症状})$
 - 预测: $P(\text{症状}|\text{病因})$
 - 分类: $\max_{\text{class}} P(\text{类别}|\text{数据})$



- Thomas Bayes "An essay towards solving a Problem in the Doctrine of Chances" (贝叶斯定理原始论文) <http://www.sbs-bvs.be/bsn57/bsn57-6.pdf>
- <http://pan.baidu.com/s/1o69Lp1K>
- https://blog.csdn.net/zdy0_2004/article/details/41096141



谢谢!

大成若缺，其用不弊。大盈若冲，其用不穷。大直若屈。大巧若拙。大辩若讷。静胜躁，寒胜热。清静为天下正。

