概率图模型

probabilistic graphical model
learning note For reading translation

我真的不懂忧郁

概率图模型

probabilistic graphical model learning note For reading translation

by

我真的不懂忧郁

Student Name	Student Number
First Surname	1234567

Instructor:	I. Surname
Teaching Assistant:	I. Surname
Project Duration:	Month, Year - Month, Year
Faculty:	Faculty of Aerospace Engineering, Delft

TU Delft

Preface

A preface...

我真的不懂忧郁
Delft, June 2024

Summary

A summary...

目录

Preface **i**

Summary **ii**

Nomenclature **iv**

1 背景 **1**

1.1 概述 1

1.2 极大似然估计 (maximum likelihood estimation) 2

1.3 频率派 vs. 贝叶斯派 3

2 贝叶斯网络 **4**

2.1 贝叶斯公式与乘法定理 5

2.2 贝叶斯网络 6

2.3 条件独立性 6

2.4 D-Separation 算法 10

2.5 贝叶斯网络在机器学习中的应用 10

2.6 贝叶斯网络模型 10

3 马尔可夫随机场 **11**

3.1 极大团 11

3.2 联合概率 11

3.3 条件独立性 13

3.4 因式分解与 Hammesley-Clifford 定理 13

3.5 马尔可夫随机场中的势函数 14

4 推断 Inference **15**

4.1 HMM 15

4.2 变量消除法 15

4.3 信念传播 15

4.4 BP 算法 16
4.5 Max Product Algorithm 16

5 概念补充 17
5.1 因子图 17
5.2 道德图 17

References 18

A Source Code Example 19

B Task Division Example 20

Nomenclature

If a nomenclature is required, a simple template can be found below for convenience. Feel free to use, adapt or completely remove.

Abbreviations

Abbreviation	Definition
ISA	International Standard Atmosphere
...	

Symbols

Symbol	Definition	Unit
V	Velocity	[m/s]
...		
ρ	Density	[kg/m^3]
...		

Chapter 1

背景

1.1. 概述

概率图模型（Probabilistic Graphical Model，PGM），简称图模型（Graphical Model，GM），是指一种用图结构来描述多元随机变量之间条件独立性的概率模型（注意条件独立性），从而给研究高维空间的概率模型带来了很大的便捷性。

对于一个 K 维的随机向量，其联合概率为高维空间中的分布，一般难以直接建模。假设每 $\mathbf{X} = [X_1, X_2, \cdots, X_K]^T$ 个变量为离散变量并且有 m 个取值，在不做任何独立假设条件下，则需要 $m^K - 1$ 个参数才能表示其概率分布。

参数数量是指数级的，这在实际应用中不可接受，而一种可以有效减少参数量的方法就是**独立性假设**。将 K 维随机向量的联合概率分解为 K 个条件概率的乘积

$$
\begin{aligned}
p(\mathbf{x}) &= P(\mathbf{X} = \mathbf{x}) \\
&= p(x_1)p(x_1|x_2)\cdots p(x_k|x_1, x_2, \cdots, x_{K-1}) \\
&= \prod_{k=1}^{K} p(x_k|x_1, \cdots, x_{k-1})
\end{aligned}
\tag{1.1}
$$

如果某些变量之间存在条件独立，则其参数量可以大幅度减少。假设有四个二值变量，在不知道这几个变量依赖关系的情况下，用一个联合概率表来记录每一种取值的概率需要 $2^4 - 1 = 15$ 个参数，则在已知 $\mathbf{X}_1$ 时，$\mathbf{X}_2$ 和 $\mathbf{X}_3$ 独立，即有

$$p(x_2|x_1, x_3) = p(x_2|x_1) \tag{1.2}$$

同理：$p(x_3|x_1, x_2) = p(x_3|x_1)$

因此假如我们知道更多独立性条件，可以大大简化联合概率分布的计算，而**概率图**则可以直观表示随机变量之间条件独立性关系。

概率图模型的三个基本问题

1. 表示问题：对于一个概率模型，如何通过图结构描述变量之间的依赖关系；

2. 学习问题：图模型的学习包括图结构的学习和参数的学习，即参数估计问题；
3. 推断问题：计算其他变量的条件概率分布；

有向图模型

有向图模型（Directed Graphical Model），也称为贝叶斯网络（Bayesian Network）或信念网络（Belief Network，BN），是一类用有向图来描述随机向量概率分布的模型。常见的有向图模型：很多经典的机器学习模型可以使用有向图模型来描述，比如朴素贝叶斯分类器、隐马尔可夫模型、深度信念网络等。

无向图模型

无向图模型，也称为马尔可夫随机场（Markov Random Field，MRF）或马尔可夫网络（Markov Network），是一类用无向图来描述一组具有局部马尔可夫性质的随机向量 $\mathbf{X}$ 的联合概率分布的模型。常见的无向图模型有：最大熵模型、条件随机场、玻尔兹曼机、受限玻尔兹曼机等。

1.2. 极大似然估计 (maximum likelihood estimation)

给定概率分布 D，已知其概率密度函数（连续分布）或者概率质量函数（离散分布）为 f_D，以及一个分布参数 θ，我们可以从这个分布中抽一个具有 n 个值的采样 $X_1, X_2, \cdots, X_n$，利用 f_D 计算出其似然函数

$$L(\theta|x_1,\cdots,x_n) = f_\theta(x_1,\cdots,x_n) \tag{1.3}$$

若 D 是离散分布，f_θ 即是在参数为 θ 时观测到这一采样的概率；若其是连续分布，f_θ 则为 $X_1, X_2, \cdots, X_n$ 联合分布的概率密度函数在观测值处的取值。

从数学上来说，我们可以在 θ 的所有可能取值中寻找一个值使得似然函数取到最大值。这个使可能性最大的 $\widehat{\theta}$ 值即称为 θ 的最大似然估计。由定义，最大似然估计是样本的函数。

相对熵

最大似然估计可以从相对熵推导而来。相对熵衡量了使用一个给定分布 Q 来近似另一个分布 P 时的信息损失，对于离散随机变量

$$D_{KL}(P||Q) = \sum_i P(i) log \frac{P(i)}{Q(i)} \tag{1.4}$$

其中 P 是真实分布，Q 是近似分布。在最大似然估计的情景下，假设分布拥有一系列参数 θ，我们希望通过样本得到参数的估计值 $\hat{\theta}$，我们可以利用相对熵来评估估计的好坏

$$D_{KL}(p_\theta(x)||p_{\hat{\theta}}(x)) = \sum_{x\in E} p_\theta(x) log \frac{p_\theta(x)}{p_{\hat{\theta}}(x)} \tag{1.5}$$

根据数学期望的定义，上式可以改写

$$D_{KL}(p_\theta(x)||p_{\hat{\theta}}(x)) = \mathbb{E}_\theta[log(\frac{p_\theta(x)}{p_{\hat{\theta}}(x)})] = \mathbb{E}_\theta[log\ p_\theta(x)] - \mathbb{E}_\theta[log\ p_{\hat{\theta}}(x)] \tag{1.6}$$

KL 值越大，参数估计越坏，因此，需要通过改变估计参数 $\hat{\theta}$ 的值来获得最小的值，所对应的参数极为最佳估计参数

$$\hat{\theta}_{best} = arg \min_{\hat{\theta}} D_{KL}(p_{\theta}(x)||p_{\hat{\theta}}(x)) \tag{1.7}$$

假设有 n 个样本，根据大数定律，

$$\mathbb{E}_{\theta}[log\ p_{\hat{\theta}(x)}] \rightsquigarrow \frac{1}{n}\sum_{i=1}^{n} log\ p_{\hat{\theta}}(x) \tag{1.8}$$

因此我们可以通过下式去估计

$$D_{KL}(p_{\theta}(x)||p_{\hat{\theta}}(x)) = \mathbb{E}_{\theta}[log\ p_{\theta}(x)] - \frac{1}{n}\sum_{i=1}^{n} log\ p_{\hat{\theta}}(x_i) \tag{1.9}$$

对于一个已知的分布，其参数 θ 是确定的。因此，$\mathbb{E}_{\theta}[log\ p_{\theta(x)}]$ 为常数。因此，我们可以通过最小化 KL 值获得最佳估计参数：

$$\hat{\theta} = arg \min_{\theta} \mathbb{E}_{\theta}[log\ p_{\hat{\theta}(x)}] - \frac{1}{n}\sum_{i=1}^{n} log\ p_{\hat{\theta}}(x_i) \tag{1.10}$$

只要求和项最大，那么整体就最小，这个优化问题等价于

$$\begin{aligned} & arg \max_{\theta} \frac{1}{n}\sum_{i=1}^{n} log\ p_{\hat{\theta}}(x_i) \\ & \Rightarrow arg \max_{\theta} log[\prod_{i=1}^{n} p_{\hat{\theta}}(x_i)] \\ & \Rightarrow arg \max_{\theta} \prod_{i=1}^{n} p_{\hat{\theta}}(x_i) \end{aligned} \tag{1.11}$$

因此，要得到最佳参数估计值，只需要最大化 $\prod_{i=1}^{n} p_{\hat{\theta}}(x_i)$，这就是最大似然函数。

1.3. 频率派 vs. 贝叶斯派

频率派和贝叶斯派的区别是是否允许先验估计。

Frequentist	Bayesian
频率论方法通过大量独立实验将概率解释为统计均值（大数定律）	贝叶斯方法则将概率解释为信念度（degree of belief）（不需要大量的实验）
频率学派把未知参数看作普通变量（固定值），把样本看作随机变量	贝叶斯学派把一切变量看作随机变量
频率论仅仅利用抽样数据	贝叶斯论善于利用过去的知识和抽样数据

图 1.1: 频率派和贝叶斯派

Chapter 2

贝叶斯网络

贝叶斯网络 (Bayesian network)，又称信念网络 (Belief Network)，或有向无环图模型 (directed acyclic graphical model)，是一种概率图模型，于 1985 年由 Judea Pearl 首先提出。它是一种模拟人类推理过程中因果关系的不确定性处理模型，其网络拓朴结构是一个有向无环图 (DAG)。我们将有因果关系（或非条件独立）的变量或命题用箭头来连接（换言之，连接两个节点的箭头代表此两个随机变量是具有因果关系，或非条件独立）。若两个节点间以一个单箭头连接在一起，表示其中一个节点是“因 (parents)”，另一个是“果 (children)”，两节点就会产生一个条件概率值。例如，假设节点 E 直接影响到节点 H，即 E→H，则用从 E 指向 H 的箭头建立结点 E 到结点 H 的有向弧 (E,H)，权值 (即连接强度) 用条件概率 P(H|E) 来表示，如下图所示：

图 2.1: 贝叶斯网络

主要历程

年份	事件	相关论文/Reference
1763	Thomas Bayes提出贝叶斯定理的基本形式	Bayes, T. (1763). An Essay towards solving a Problem in the Doctrine of Chances. The Philosophical Transactions. 53: 370–418.
1960	Ruslan L. Stratonovich在他的论文中描述了隐马尔科夫模型（,hidden markov model，HMM）中使用的向前和向后递归以及边缘平滑概率的计算	Stratonovich, R.L. (1960). Conditional Markov Processes.Theory of Probability and its Applications. 5 (2): 156–178.
1975	Baker提出了DRAGON系统，将HMM应用于语音识别	Baker, J.(1975). The DRAGON system—An overview.IEEE Transactions on Acoustics, Speech, and Signal Processing. 23: 24–29.
1979	Dawid通过对统计几个不同领域的有条件独立的研究阐述了条件独立理论所形成的统计推断理论的框架概念	Dawid, A. P.(1979). Conditional Independence in Statistical Theory.Journal of the Royal Statistical Society, Series B. 41 (1): 1–31.
1985	Pearl提出贝叶斯网络的概念	Pearl J. (1985). Bayesian Networks: A Model of Self–Activated Memory for Evidential Reasoning.Proceedings of the 7th Conference of the Cognitive Science Society, University of California, Irvine, CA. pp. 329–334.
1995	贝叶斯分类在60年代初被以不同的名称引入文本检索社区	Russell, Stuart; Norvig, Peter (2003) [1995]. Artificial Intelligence: A Modern Approach (2nd ed.). Prentice Hall.
2001	Rish使用Monte Carlo模拟对几类随机生成的问题进行系统的分类准确性研究，证明了朴素贝叶斯的良好性能	Rish, I. (2001). An empirical study of the naive Bayes classifier. IJCAI Workshop on Empirical Methods in AI.

图 2.2: 贝叶斯网络发展主要事件

2.1. 贝叶斯公式与乘法定理

乘法定理

事件 A 在事件 B 发生的情况下发生的条件概率表示为

$$P(A|B) = \frac{P(AB)}{P(B)} \tag{2.1}$$

因此乘法定理写成

$$P(AB) = P(A|B)P(B) \tag{2.2}$$

贝叶斯公式

贝叶斯公式可以理解成：*后验概率 = (似然性 * 先验概率)/标准化常量*，数学表达为

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B|A)}{P(B)} \tag{2.3}$$

链式法则

对于非相互独立的随机事件的乘法是相对复杂的，根据乘法定理，我们可以得到 n 个随机变量的联合概率

$$P(x_1, \cdots, x_n) = P(x_1)P(x_2|x_1)\cdots P(x_n|x_{n-1}, \cdots, x_1) \tag{2.4}$$

举个例子：假设事件 X_1, X_2, X_3 三个事件，联合概率分布为

$$P(X_1X_2X_3) = P(X_1)P(X_2|X_1)P(X_3|X_1X_2) \tag{2.5}$$

可以发现 $P(X_1)P(X_2|X_1) = P(X_1X_2)$，这刚好就是 X_1X_2 同时发生的概率，也就是 $P(X_3|X_2X_1)$ 中的条件。

2.2. 贝叶斯网络

贝叶斯网络用有向图来描述 n 个随机事件的概率分布，假设事件 a 发生情况下 b 发生的概率为 $p(b|a)$，b 发生情况下 c 发生的概率为 $p(c|b)$，a 发生的情况下 c 发生的条件概率表示为 $p(c|a)$，通过有向图表述为

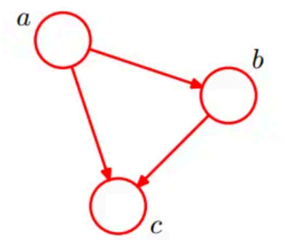


图 2.3: 事件 abc 的贝叶斯网络

图的节点表示事件，图的边表示条件概率。从图中可以看到因为 a 导致 b，a 和 b 导致 c，所以有

$$P(abc) = P(a)P(b|a)P(c|ba) \tag{2.6}$$

下图是一个贝叶斯网络的一般形态。

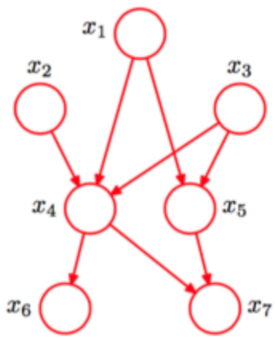


图 2.4: 一个”正常”的贝叶斯网络

2.3. 条件独立性

如果事件 A 和事件 B 是相互独立的，则乘法公式满足

$$P(AB) = P(A)P(B) \tag{2.7}$$

贝叶斯网络结构暗示条件独立性。例如如下简单的贝叶斯网络

$$\textcircled{A} \longrightarrow \textcircled{B}$$

图 2.5: A 导致 B，A 和 B 存在因果关系，因此不存在条件独立性

联合概率分布为

$$P(AB) = P(A)P(B|A) \tag{2.8}$$

如果 A 和 B 是相互独立的两个事件

图 2.6: A 和 B 不存在因果关系，即体现条件独立性

此时概率分布就是相对独立事件的乘法公式。

条件独立性的判定

对于一般性的贝叶斯网络我们如何去判定节点之间的条件独立性呢？考虑一下三种贝叶斯网络的结构在什么情况下能推出 $P(AB) = P(A)P(B)$

1. *tail-to-tail*：父节点已知则子节点独立：；

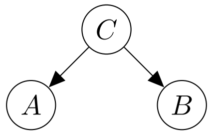


图 2.7: C 未知，A 和 B 被 C 的影响，所以不独立

如果 C 发生的概率未知，根据贝叶斯网络结构，联合概率分布为

$$P(ABC) = P(C)P(A|C)P(B|C) \tag{2.9}$$

由于不知道 $P(C)$，无法得出 $P(AB) = P(A)P(B)$。

如果 C 发生的概率已知，那么根据 $P(AB|C) = P(ABC)/P(C)$[1]，联合概率分布为

$$P(AB|C) = P(A|C)P(B|C) \tag{2.10}$$

所以此时 A 和 B 的条件独立。所以，在 C 给定的条件下，A、B 被阻断 (blocked)，是独立的，称之为 *tail-to-tail* 条件独立。

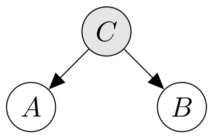


图 2.8: 若 C 是先验概率，A 和 B 之间被 C 阻断，因此不存在因果关系，即 A 和 B 条件独立

2. *head-to-tail*：中间节点已知则两边节点独立；

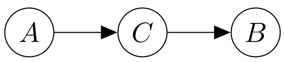


图 2.9: C 的概率未知的情况下，A 可以通过 C 去影响 B，因此 A 和 B 存在因果关系

联合概率分布为

$$P(ABC) = P(A)P(C|A)P(B|C) \tag{2.11}$$

C 的概率未知的情况下，A 可以通过 C 去影响 B，因此 A 和 B 存在因果关系。

如果已知 C 发生的概率 $P(C)$，根据贝叶斯公式

$$\begin{aligned} &P(ABC) = P(A)P(C|A)P(B|C) \\ &\Rightarrow P(AB|C) = \frac{P(ABC)}{P(C)} = \frac{P(A)P(C|A)P(B|C)}{P(C)} \\ &\Rightarrow P(AB|C) = \frac{P(ABC)}{P(C)} = \frac{P(AC)P(B|C)}{P(C)} \\ &\Rightarrow P(AB|C) = P(A|C)P(B|C) \end{aligned} \tag{2.12}$$

这相当于 C 已知的情况下，A 和 B 的关系被截断了，因此 A 和 B 在 C 概率已知的情况下是条件独立的。

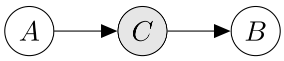


图 2.10: C 是先验概率，A 和 B 之间被 B 阻断，因此不存在因果关系，即 A 和 B 条件独立

3. *head-to-head*：子节点未知则父节点独立；

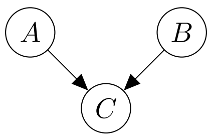


图 2.11: C 是先验概率，A 和 B 之间被 B 阻断，因此不存在因果关系，即 A 和 B 条件独立

C 概率未知的情况下，贝叶斯网络的联合概率分布为

$$P(ABC) = P(A)P(B)P(C|AB) \tag{2.13}$$

因为 $P(C|AB)$ 和 $P(A)$ 和 $P(B)$ 独立性无关，所以可以简化为

$$P(AB) = P(A)P(B) \tag{2.14}$$

如果 $P(C)$ 给定，则

$$P(AB|C) = \frac{P(A)P(B)P(C|AB)}{P(C)} \tag{2.15}$$

上面式子因为约不掉 $P(C)$，所以没办法推出 $P(AB) = P(A)P(B)$，因此不满足条件独立性。

因此在 head to head 的情况下，如果没有观测到共同子节点 C 的概率，则父节点 A 和 B 相互独立，反之则不满足独立条件。

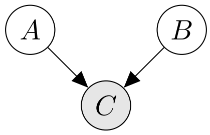


图 2.12: C 是先验概率，C 已知的情况下父节点通过子节点关联，A 和 B 条件不独立

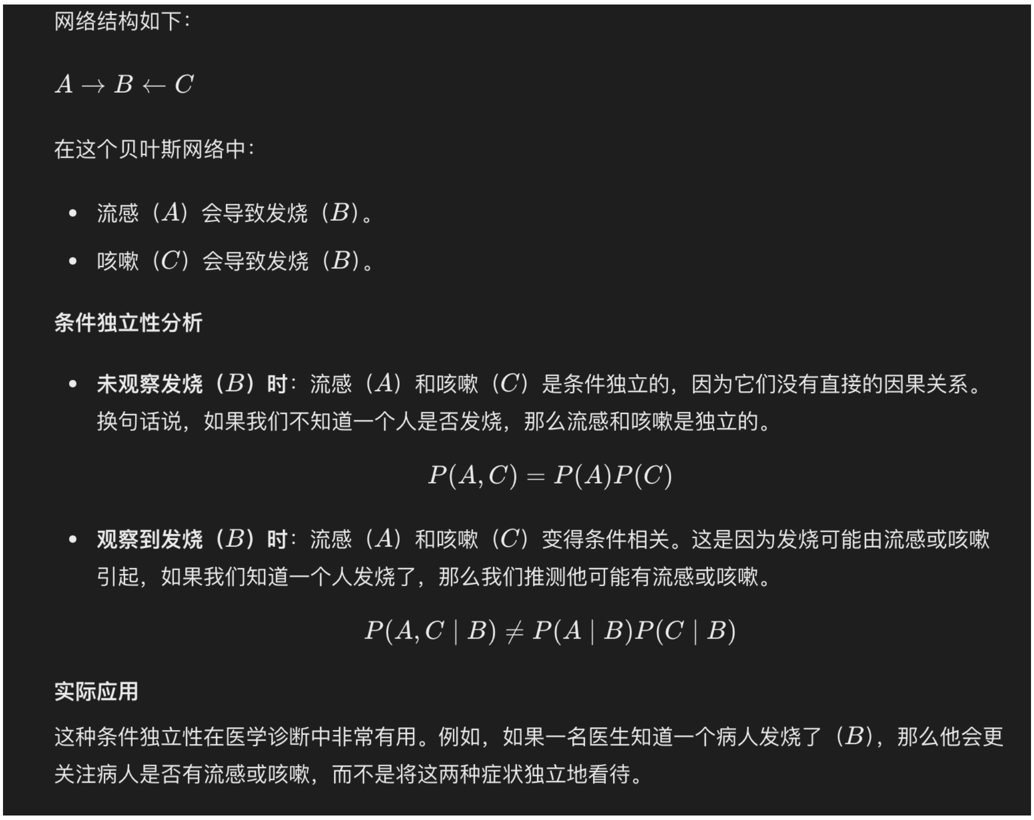
可以这么理解：老父亲还在，则兄弟关系还在，老父亲出问题了，兄弟联系断了。如果两个节点是直接连接的，它们肯定是非条件独立的，是直接因果关系。其中父节点是“因”，子节点是“果”。

条件独立性的意义

条件独立性可以简化链过长；条件独立性允许将复杂的联合概率分布分解成更简单的条件概率分布。在实际应用中，计算高维联合概率分布是非常困难的，而条件独立性提供了简化这些计算的方法。例如，在贝叶斯网络（Bayesian Network）中，通过利用条件独立性，联合概率分布 $P(X_1, X_2, \cdots, X_N)$ 可以分解为一系列的条件概率：

$$P(X_1, X_2, \cdots, X_N) = \prod_{i=1}^{N} P(X_i|\pi(X_i)) \tag{2.16}$$

其中 $\pi(X_i)$ 表示 X_i 的父节点集合。利用条件独立性关系，这种分解显著降低了计算复杂度。



网络结构如下：

$A \to B \leftarrow C$

在这个贝叶斯网络中：

- 流感（A）会导致发烧（B）。
- 咳嗽（C）会导致发烧（B）。

条件独立性分析

- **未观察发烧（B）时**：流感（A）和咳嗽（C）是条件独立的，因为它们没有直接的因果关系。换句话说，如果我们不知道一个人是否发烧，那么流感和咳嗽是独立的。

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

- **观察到发烧（B）时**：流感（A）和咳嗽（C）变得条件相关。这是因为发烧可能由流感或咳嗽引起，如果我们知道一个人发烧了，那么我们推测他可能有流感或咳嗽。

$$P(A, C \mid B) \neq P(A \mid B)P(C \mid B)$$

实际应用

这种条件独立性在医学诊断中非常有用。例如，如果一名医生知道一个病人发烧了（B），那么他会更关注病人是否有流感或咳嗽，而不是将这两种症状独立地看待。

图 2.13: chat-GPT 给的例子

2.4. D-Separation 算法

D 分离算法又称为有向分离算法，是一种用来判断变量是否条件独立的图形化方法。对于一个 DAG，D 分离可以快速判断两个节点之间是否条件独立。

在前面小节中已经列举了条件独立性判定可能的三种情况：*tail-to-tail*、*head-to-tail*、*head-to-head*，基于三种条件独立性判定情况，*D* 分离算法通过路径遍历来确定是否路径被阻断。下面举例子说明

例子

马尔可夫毯

对于一个人来讲和全世界的关系等于和自己周围的血缘关系，和别人没关系；

2.5. 贝叶斯网络在机器学习中的应用

2.6. 贝叶斯网络模型

朴素贝叶斯 (Native Bayes)

单一；
贝叶斯假设

$$P(X|Y) = \prod_{i=1}^{K} p(x_i|y=1) \tag{2.17}$$

概率图模型

高斯混合模型

和时间相关的

主要两个算法:*Markcov Chain* 和高斯过程 *(*无限维高斯分布*)*

连续的

高斯贝叶斯网络

动态系统

混合和时间，最常见的是 HMM，LDS，

Chapter 3

马尔可夫随机场

马尔可夫随机场 *(Markov Random Field, 简称 MRF)* 是一种无向图模型，图中每一个结点表示一个或者一组变量，结点之间的边表示两个变量之间的依赖关系。马尔科夫随机场又一组势函数 *(potential function)*，也称为因子（*factor*），这是定义在变量子集上的非负实函数，主要用于定义概率分布函数。

3.1. 极大团

下图是一个简单的马尔可夫随机场，对于图中结点的一个子集，如果任意两个结点之间有边链接，则称该节点子集为一个**团 (clique)**，若在一个团中加入另外任意一个结点都不在形成团，则该团称为**极大团 (maximal clique)**。即极大团就是不能被其他团所包含的团。

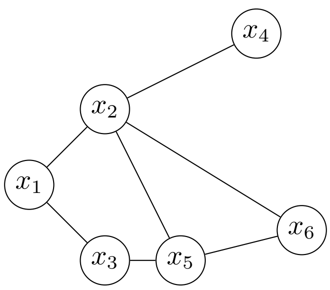


图 3.1: 一个简单的马尔可夫随机场例子

上例中，$\{x_1, x_2\}$，$\{x_1, x_3\}$，$\{x_3, x_5\}$，$\{x_2, x_5\}$，$\{x_2, x_6\}$，$\{x_5, x_6\}$，$\{x_2, x_4\}$ 和 $\{x_2, x_5, x_6\}$ 都是团，但是除了 $\{x_2, x_5\}$，$\{x_2, x_6\}$，$\{x_5, x_6\}$，以外都是极大团。显然，每个节点至少出现在一个极大团中。

3.2. 联合概率

无向图的联合概率表示

在马尔可夫随机场中，多个变量之间的联合概率分布能基于团分解为多个因子的乘积，**每个因子仅仅只和一个团相关**，具体来说，对于 n 个变量 $x=\{x_1,x_2,\cdots,x_n\}$，所有团构成集合 $\mathcal{C}$，与团 $\mathcal{Q}\in\mathcal{C}$ 对应的变量集合记为 $x_\mathcal{Q}$，则联合概率 $P(x)$ 定义为

$$P(x)=\frac{1}{Z}\prod_{\mathcal{Q}\in\mathcal{C}}\psi_\mathcal{Q}(x_\mathcal{Q}) \tag{3.1}$$

其中 $\psi_\mathcal{Q}$ 为与团对应的是势函数，用于对团 $\mathcal{Q}$ 中的变量关系进行建模，Z 为规范化因子，以确保 $P(x)$ 定义的概率是有意义的，表达式如下

$$Z=\sum_x\prod_{\mathcal{Q}\in\mathcal{C}}\psi_\mathcal{Q}(x_\mathcal{Q}) \tag{3.2}$$

基于极大团定义联合概率

显然如果变量比较多，那么团的数量就会很多，这就意味着联合概率就会有很多乘积项，显然会给计算带来负担，但是我们注意到 *如果 $\mathcal{Q}$ 不是最大团，那么它势必会被一个极大团 $\mathcal{Q}^*$ 所包含，这意味着变量 $x_\mathcal{Q}$ 之间的关系不仅体现在势函数 $\psi_\mathcal{Q}$ 中，函体现在 $\psi_{\mathcal{Q}^*}$*。于是由极大团来定义联合概率，假定所有极大团构成的集合为 $\mathcal{C}^*$，则

$$P(x)=\frac{1}{Z^*}\prod_{\mathcal{Q}\in\mathcal{C}^*}\psi_{\mathcal{Q}^*}(x_{\mathcal{Q}^*}) \tag{3.3}$$

这样来说我们就不需要去统计每个团的势函数的积。

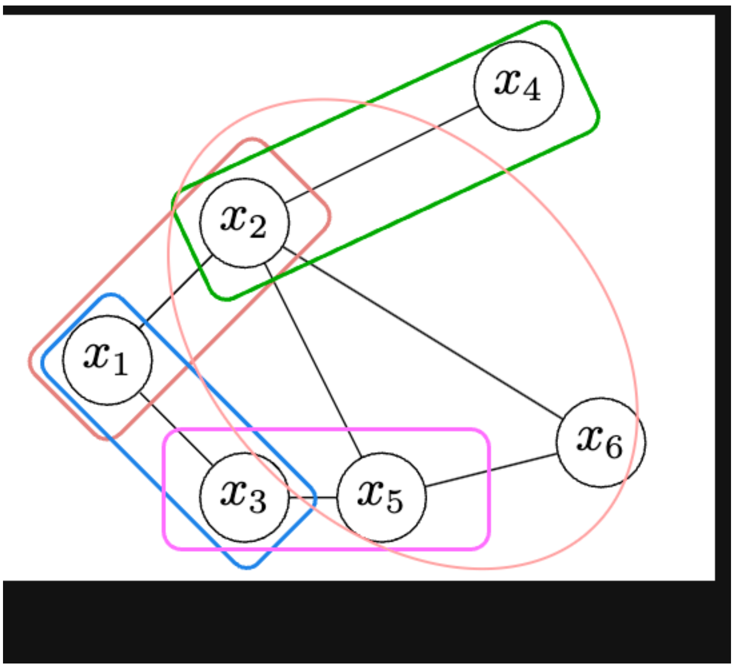


图 3.2: 基于极大团的联合概率分布

3.3. 条件独立性

在马尔可夫随机场中如何得到条件独立性呢？同样借助"分离"的概念，如下图所示，如果结点集 A 的结点到 B 中的结点都必须经过结点集 C 中的结点，则称结点集 A 和 B 被结点集 C 分离，C 称为分离集。

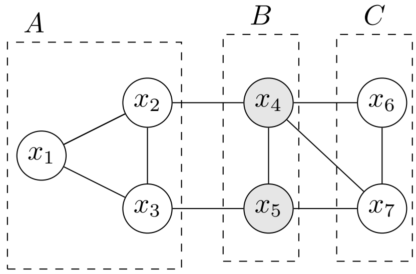


图 3.3: x_4 和 x_5 组成的团是分离集

全局马尔可夫

给给定两个变量的子集的分离集，则这两个子集条件独立。

局部马尔可夫

给定某变量的邻接变量，则该变量条件独立于其他变量，形式地说，令 V 为图的结点集合，$n(v)$ 为结点 v 在图上的邻接结点，$n^*(v) = n(v) \cup \{v\}$，有 $x_v \perp x_{V/n^*(v)}|x_n(v)$。

成对马尔可夫

给定某变量的邻接变量，两个非邻接变量条件独立，形式化地说，令图的结点集和边集分别为 V 和 E，对途中两个结点 u 和 v，若 $\langle u,v\rangle \notin E$，则 $x_u \perp x_{V/\langle u,v\rangle}$

proposition 3.3.1: 三种马尔可夫是相互等价的

3.4. 因式分解与 Hammesley-Clifford 定理

definition 3.4.1: *(团)* 团是一个关于节点的集合，集合中的节点之间相互都是连通的（都有边）。

definition 3.4.2: *(最大团)* 团中没办法添进去新的节点。

基于最大团形式的因式分解。

$$p(x) = \frac{1}{z}\prod_{i=1}^{K}\varphi(x_{ci}) \tag{3.4}$$

其中 c_i 是最大团，x_{ci} 最大团随机变量集合，$\varphi(x_{ci})$ 势函数，必须为正。

有一个定理保证基于最大团的因式分解是马尔可夫随机场

theorem 3.4.1: *(Hammesley-Clifford 定理)*

Markov Random Field ⇔ Gibbs Distribution

吉布斯概率分布和玻尔兹曼分布（这部分最好提前去了解一下统计物理里面的概念）

吉布斯分布在概率分布的角度还具备了最大熵的概念。

马尔可夫随机场

3.5. 马尔可夫随机场中的势函数

势函数用来定量刻画变量集 x_Q 中变量之间的相关关系，他应该是非负函数，且在所偏好的变量取值上有较大函数值。例如

$$\psi_{AC}(x_A, x_C) = \begin{cases} 1.5, & if\ x_A = x_C \\ 0.1, & ohterwise \end{cases}$$

$$\psi_{BC}(x_B, x_C) = \begin{cases} 0.2, & if\ x_B = x_C \\ 1.3, & ohterwise \end{cases}$$

则说明该模型偏好变量 x_A 和 x_C 有相同的取值，x_B 和 x_C 有不同的取值。换言之 x_A 和 x_C 正相关，x_B 和 x_C 负相关。

常用的势函数还有指数函数

$$\psi_Q(x_Q) = e^{-H_Q(x_Q)} \tag{3.5}$$

Chapter 4

推断 Inference

推断分为近似推断，精确推断和采样方法；

4.1. HMM

动态规划问题

4.2. 变量消除法

有向图变量消去

无向图变量消去

4.3. 信念传播

信念传播 (*Belief Propagation*) 算法将变量消去法中的求和操作看作一个消息传递的过程，较好地解决了求解多个边缘分布时重复计算的问题。具体来说，变量消去法通过求和操作

$$m_{ij}(x_j) = \sum_{x_i} \varphi(x_i, x_j) \prod_{k \in n(i)/j} m_{ki}(x_i) \tag{4.1}$$

在信念传播算法中，一个结点仅仅在接受到来自其他所有结点的消息后才能向另一个结点发送消息，且结点的边缘分布正比于他所接受到的消息的乘积

$$P(x_i) \propto \alpha \prod_{k \in n(\epsilon)} m_{ki}(x_i) \tag{4.2}$$

为了解决重复问题
变量消除法

4.4. 近似推断

MCMC

变分推断

无向树的因式分解

4.5. BP 算法

BP 算法就是为了解决 VE+caching，重复计算问题?
反向传播算法
树或者图的遍历
BP（Sequential Implementation）

- Get root ;
- Collect Message;
- Distribute Message ;

4.6. Max Product Algorithm

MP 算法是 BP 算法的改进。

Chapter 5

概念补充

5.1. 因子图

有向图有因式分解，无向图有最大团。

引入因子节点。

5.2. 道德图

道德图：有向图（树）→ 无向图（引入环），

References

[1] I. Surname, I. Surname, and I. Surname. "The Title of the Article". In: *The Title of the Journal* 1.2 (2000), pp. 123–456.

Chapter A

Source Code Example

Adding source code to your report/thesis is supported with the package `listings`*. An example can be found below. Files can be added using* `\lstinputlisting[language=<language>]{<filename>}`.

```python
"""
ISA Calculator: import the function, specify the height and it will return a
list in the following format: [Temperature,Density,Pressure,Speed of Sound].
Note that there is no check to see if the maximum altitude is reached.
"""

import math
g0 = 9.80665
R = 287.0
layer1 = [0, 288.15, 101325.0]
alt = [0,11000,20000,32000,47000,51000,71000,86000]
a = [-.0065,0,.0010,.0028,0,-.0028,-.0020]

def atmosphere(h):
    for i in range(0,len(alt)-1):
        if h >= alt[i]:
            layer0 = layer1[:]
            layer1[0] = min(h,alt[i+1])
            if a[i] != 0:
                layer1[1] = layer0[1] + a[i]*(layer1[0]-layer0[0])
                layer1[2] = layer0[2] * (layer1[1]/layer0[1])**(-g0/(a[i]*R))
            else:
                layer1[2] = layer0[2]*math.exp((-g0/(R*layer1[1]))*(layer1[0]-layer0[0]))
    return [layer1[1],layer1[2]/(R*layer1[1]),layer1[2],math.sqrt(1.4*R*layer1[1])]
```

Chapter B

Task Division Example

If a task division is required, a simple template can be found below for convenience. Feel free to use, adapt or completely remove.

表 **B.1:** Distribution of the workload

	Task	Student Name(s)
	Summary	
Chapter 1	Introduction	
Chapter 2		
Chapter 3		
Chapter *		
Chapter *	Conclusion	
	Editors	
	CAD and Figures	
	Document Design and Layout	