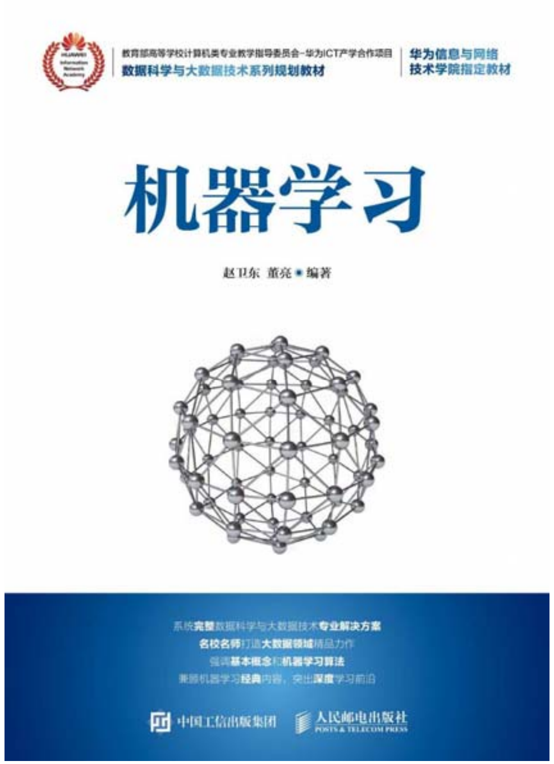



# 机器学习
# 贝叶斯网络


复旦大学 **赵卫东** 博士

wdzhao@fudan.edu.cn

## 章节介绍

- 贝叶斯网络(Bayesian network)，又称为信念网络(Belief network)，是一种通过有向无环图(Directed acyclic graph, DAG)表示一组随机变量及其条件依赖概率的概率图模型。概率图中，节点表示随机变量，有向边表示随机变量间的依赖关系，条件概率表示依赖关系的强度。没有父节点的节点用先验概率表达信息。两个节点若无连接则表示相互独立的随机变量。
- 贝叶斯网络中的节点可以表示任意问题，丰富的概率表达能力使能较好地处理不确定性信息或问题。贝叶斯网络中所有节点都是可见的，并且节点间的因果关系可以非常直观地观察到。这些特性都使得贝叶斯网络在众多智能系统中有相当重要的应用
- 本章首先介绍贝叶斯网络的基础知识，重点讲解贝叶斯的概率基础和朴素贝叶斯分类模型，并结合实际案例说明贝叶斯网络的应用

# 章节结构

- 贝叶斯理论概述
- 贝叶斯概率基础
  - 概率论
  - 贝叶斯概率
- 朴素贝叶斯分类模型
- 贝叶斯网络推理
- 贝叶斯网络的应用
  - 中文分词
  - 机器翻译
  - 故障诊断
  - 疾病诊断

# 贝叶斯理论概述

- 贝叶斯方法分析的特点是用概率表示不确定性，概率规则表示推理或学习，随机变量的概率分布表示推理或学习的最终结果
- 贝叶斯理论源于贝叶斯提出的贝叶斯定理。贝叶斯定理引入了先验概率，后验概率由先验概率和条件概率表达式计算出。假设有随机变量$x$和$y$，$p(x,y)$表示它们的联合概率，$p(x|y)$和$p(y|x)$表示条件概率，其中$p(y|x)$是后验概率，而$p(y)$称为 $y$的先验概率，$x$和 $y$的联合概率和条件概率满足下列关系：

$$p(x,y) = p(y|x)p(x) = p(x|y)p(y)$$

- 交换后得到：

$$p(y|x)=\frac{p(x|y)p(y)}{p(x)}$$

- 上述公式即为贝叶斯定理，它提供了从先验概率$p(y)$计算后验概率$p(y|x)$的方法

# 贝叶斯概率基础

- 概率论
  - 古典概率
  - 几何概率
  - 条件概率
  - 加法定理
  - 减法定理
  - 独立事件
  - 联合概率分布
  - 条件概率分布
- 贝叶斯概率
  - 先验概率
  - 后验概率
  - 全概率公式
  - 贝叶斯公式

## 朴素贝叶斯分类模型

- 朴素贝叶斯分类模型是一种简单的构造分类器的方法。朴素贝叶斯分类模型是将问题分为特征向量和决策向量两类，并假设问题的特征向量都是相互独立地作用于决策向量的，即问题的特征之间都是互不相关的。尽管有这样过于简单的假设，但朴素贝叶斯分类模型能指数级降低贝叶斯网络构建的复杂性，同时还能较好地处理训练样本的噪声和无关属性，所以朴素贝叶斯分类模型仍然在很多现实问题中有着高效的应用，例如入侵检测和文本分类等领域。目前许多研究学者也在致力于改善特征变量间的独立性的限制使得朴素贝叶斯分类模型可以应用到更多问题上

# 朴素贝叶斯分类模型

- 假设问题的特性向量为X，Xi={X1, X2, ...,Xn}是特征属性之一，并且X1, X2, ...,Xn之前相互独立，那么p(X|Y)可以分解为多个向量的积，即有

$$p(\mathrm{X}|\mathrm{Y}) = \prod_{i=1}^{n} p(X_i|Y)$$

- 那么这个问题就可以由朴素贝叶斯分类器来解决，即

$$p(\mathrm{Y}|\mathrm{X}) = \frac{p(Y)\prod_{i=1}^{n} p(X_i|Y)}{p(X)}$$

- 其中p(X)是常数，先验概率p(Y)可以通过训练集中每类样本所占的比例进行估计。给定Y=y，如果要估计测试样本X的分类，由朴素贝叶斯分类得到y的后验概率为：

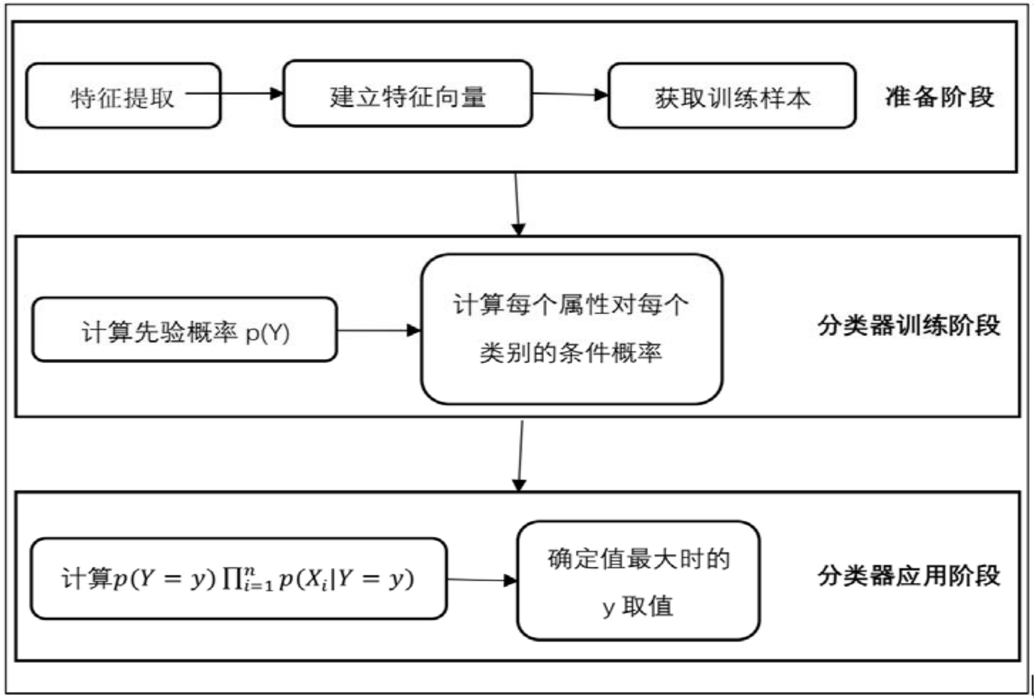
$$p(\mathrm{Y}=\mathrm{y}|\mathrm{X}) = \frac{p(Y=y)\prod_{i=1}^{n} p(X_i|Y=y)}{p(X)}$$

## 朴素贝叶斯分类模型

- 因此最后只要找到使$p(Y=y)\prod_{i=1}^{n}p(X_i|Y=y)$最大的类别y即可
- 从计算分析中可见，$p(X_i|Y)$的计算是模型关键的一步，这一步的计算视特征属性的不同也有不同的计算方法
  - 对于离散型的特征属性Xi，可以用类Y中的属性值等于Xi的样本比例来进行估计。
  - 对于连续性的特征属性Xi，通常先将Xi离散化，然后计算属于类Y的训练样本落在Xi对应离散区间的比例估计$p(X_i|Y)$。也可以假设$p(X_i|Y)$的概率分布，如正态分布，然后用训练样本估计其中的参数。
  - 而在$p(X_i|Y)=0$的时候，该概率与其他概率相乘的时候会把其它概率覆盖，因此需要引入Laplace修正。做法是对所有类别下的划分计数都加一，从而避免了等于零的情况出现，并且在训练集较大时，修正对先验的影响也会降低到可以忽略不计

# 朴素贝叶斯分类模型

- 朴素贝叶斯分类模型应用流程的三个阶段



## 朴素贝叶斯分类模型

- 朴素贝叶斯分类器还可以进行提升(Boosting)，提升方法的主要思想是学习多个分类器组成一个分类器序列，序列中后面的分类器对前面的分类器导致的错误分类的数据给予更高的重视，即调整前一个分类器分类错误的训练集的权值，并对训练集重新计算权值以调整下一个分类器，以此类推，最终得到提升后的强分类器
- 朴素贝叶斯分类模型结构简单，只有两层结构。由于特征向量间的相互独立，算法简单易于实现。同时算法有稳定的分类效率，对于不同特点的数据集其分类性能差别不大。朴素贝叶斯分类在小规模的数据集上表现优秀，并且分类过程时空开销小。算法也适合增量式训练，在数据量较大时，可以人为划分后分批增量训练

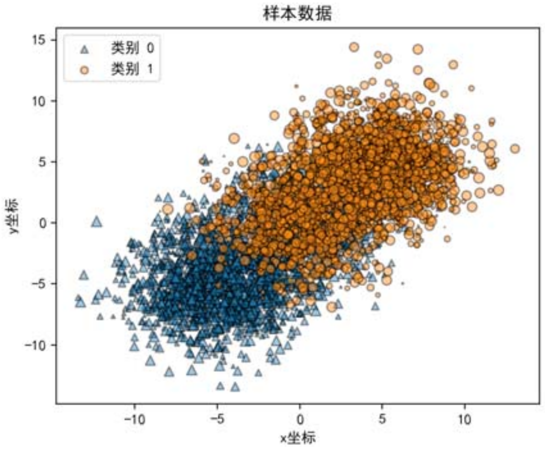
# 朴素贝叶斯分类模型

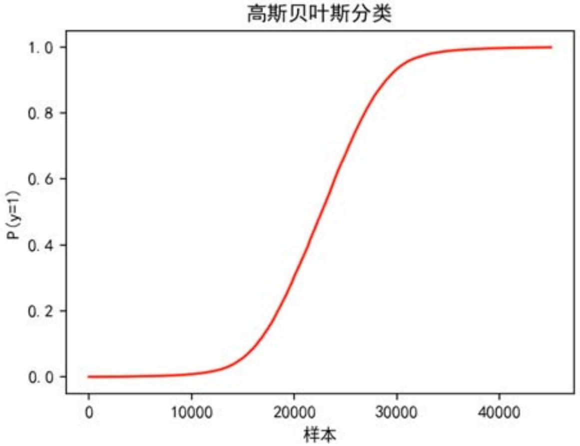
- 以下是应用sklearn库中朴素贝叶斯（高斯）分类模型进行分析的示例代码。数据源是通过sklearn中的聚类生成器（make_blobs）生成的50000个随机样本，每个样本的特征数为2个，共有三个类簇，样本集的标准差是1.0，随机数种子为42

```
centers = [(-5, -5), (0, 0), (5, 5)]
X,y=make_blobs(n_samples=50000,n_features=2,cluster_std=1.0,centers=centers,shuffle=False, random_state=42)
y[:n_samples // 2] = 0
y[n_samples // 2:] = 1
sample_weight = np.random.RandomState(42).rand(y.shape[0])
X_train, X_test, y_train, y_test,sw_train,sw_test=train_test_split(X, y, sample_weight, test_size=0.9, random_state=42)
clf = GaussianNB()
clf.fit(X_train, y_train)
prob_pos_clf = clf.predict_proba(X_test)[:, 1]
target_pred = clf.predict(X_test)
score = accuracy_score(y_test, target_pred, normalize = True)
print("accuracy score:",score)
```

# 朴素贝叶斯分类模型

- 通过GuassianNB算法fit之后，对测试集X_test进行预测，结果存在prob_pos_clf中




- 输出分类结果：
- accuracy score: 0.8335

# 朴素贝叶斯分类模型

- 贝叶斯垃圾邮件过滤器
- 传统的垃圾邮件过滤方法是关键词过滤，但这种方法过于绝对，很容易出现误判的情况。贝叶斯过滤会同时考虑关键词在正常邮件和垃圾邮件中出现的概率，并且学习用户的偏好，可以减少误判的可能性。
- 假设收到一封电子邮件$E$，邮件由$n$个关键词构成。设$X$=1表示邮件是正常邮件，$X$=0表示邮件是垃圾邮件。那么判定新邮件是否为垃圾邮件的问题可以表示为比较下列两式值的问题：

$$p(X=1|E)=\frac{p(X=1)p(E|X=1)}{p(E)}$$

$$p(X=0|E)=\frac{p(X=0)p(E|X=0)}{p(E)}$$

# 朴素贝叶斯分类模型

- 其中$p(X=1)$和$p(X=0)$可以很容易地在邮箱里查出，所以只需要计算$p(E|X=1)$和$p(E|X=0)$。这里可以简单假设$E$中$n$个关键词是互不相关，即将问题转化为朴素贝叶斯分类模型。所以就有：
$$p(E|X=1)=p(E_1|X=1)*p(E_2|X=1)*\cdots*p(E_n|X=1)$$
- 等式右边的每个分式的计算都是很容易的，于是就可以很容易地得到上文需要的两个概率值。可以预先设定好垃圾邮件的概率阈值，比较$p(X=0|E)$和$p(X=1|E)$即可实现自动的垃圾邮件标识与过滤。

# 贝叶斯网络推理

- 不确定性推理是机器学习的重要研究内容之一。用概率论方法进行不确定推理的一般流程是首先将问题抽象为一组随机变量与其联合概率分布表，然后根据概率论公式进行推理计算，但这个流程复杂度高

## 警铃问题

- 欧阳老师的家中安置了一套智能监控设备，在家中遭受盗窃或发生火灾时，设备会发出刺耳的警铃声，欧阳老师的邻居是小明和小红，假设两个邻居一般都在家中，他们听到欧阳老师家中的警报时会给欧阳老师打电话，但警报响的时候两个邻居可能会听不见。某天，出门在外的欧阳老师接到了小明的电话，小明说听到了欧阳老师家中的警报声。欧阳老师想知道家中遭受盗窃的可能性有多大？
- 按照上文所述的一般流程，该问题包含了5个变量，将其分别定义为：警报（A）、遭受盗窃（B）、发生火灾（C）、接到小明电话（D）、接到小红电话（E），每个变量均有“yes”和“no”两种可能取值。假设欧阳老师对5个变量的联合分布概率p(A, B, C, D, E)的判断如下表所示

# 警铃问题

| B | C | A | D | E | 概率 | | B | C | A | D | E | 概率 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| yes | yes | yes | yes | yes | 1.2E–4 | | no | yes | yes | yes | yes | 3.6E-3 |
| yes | yes | yes | yes | no | 5.1E–5 | | no | yes | yes | yes | no | 1.6E-3 |
| yes | yes | yes | no | yes | 1.3E–5 | | no | yes | yes | no | yes | 4.0E-4 |
| yes | yes | yes | no | no | 5.7E–6 | | no | yes | yes | no | no | 1.7E-4 |
| yes | yes | no | yes | yes | 5.0E–7 | | no | yes | no | yes | yes | 7.0E-6 |
| yes | yes | no | yes | no | 4.9E–7 | | no | yes | no | yes | no | 6.9E-4 |
| yes | yes | no | no | yes | 9.5E–8 | | no | yes | no | no | yes | 1.3E-4 |
| yes | yes | no | no | no | 9.4E–6 | | no | yes | no | no | no | 1.3E-2 |
| yes | no | yes | yes | yes | 5.8E–3 | | no | no | yes | yes | yes | 6.1E-4 |
| yes | no | yes | yes | no | 2.5E–3 | | no | no | yes | yes | no | 2.6E-4 |
| yes | no | yes | no | yes | 6.5E–4 | | no | no | yes | no | yes | 6.8E-5 |
| yes | no | yes | no | no | 2.8E–4 | | no | no | yes | no | no | 2.9E-5 |
| yes | no | no | yes | yes | 2.9E–7 | | no | no | no | yes | yes | 4.8E-4 |
| yes | no | no | yes | no | 2.9E–5 | | no | no | no | yes | no | 4.8E-2 |
| yes | no | no | no | yes | 5.6E–6 | | no | no | no | no | yes | 9.2E-3 |
| yes | no | no | no | no | 5.5E–4 | | no | no | no | no | no | 9.1E-1 |

## 警铃问题

- 问题可转化为求$p(\mathrm{B}=\mathrm{y}|\mathrm{D}=\mathrm{y})$的概率。根据联合概率分布表，可以计算出边缘概率分布：

$$p(B,D)=\sum_{A,C,E} p(A,B,C,D,E)$$

- 得到下表所示的边缘概率分布结果。

| B | D | $p(\mathrm{B},\mathrm{D})$ |
|---|---|---|
| yes | yes | 0.000115 |
| yes | no | 0.000075 |
| no | yes | 0.00015 |
| no | no | 0.99966 |

## 警铃问题

- 根据条件概率公式得到：

$$p(B=y|D=y)=\frac{p(B=y,D=y)}{p(D=y)}=\frac{p(B=y,D=y)}{p(B=y,D=y)+p(B=n,D=y)}$$

$$=\frac{0.000115}{0.000115+0.00015}\approx 0.61$$

- 上述过程即利用联合概率进行不确定性推理的一个例子。注意到这个过程的复杂度相当高，包含n个变量的联合概率有$2^n$个项，其中有$2^n-1$个独立参数，上述问题就有$2^5-1=31$个独立参数。当n增加时，独立参数的个数将以指数倍增长，并且这些独立参数的获取、存储和运算同时将指数级复杂。于是如何降低复杂度提高运算效率显得尤为关键。引入条件独立以分解联合分布。

## 警铃问题

- 运用条件概率的链式规则，可以得到：

$$p(B,C,A,D,E) = p(B)p(C|B)p(A|B,C)p(D|B,C,A)p(E|B,C,A,D)$$

- 注意到，遭到盗窃（B）和发生火灾（C）可以认为是互相无关的，于是上式中p(C|B)即可以简化为p(C)。此外接到小明电话（D）、接到小红电话（E）实际只与警报（A）有关，于是有$p(D|B,C,A) = p(D|A)$，$p(E|B,C,A,D) = p(E|A)$。所以上式可以简化为：

$$p(B,C,A,D,E) = p(B)p(C)p(A|B,C)p(D|A)p(E|A)$$

- 现在的独立参数减少，复杂度降低了两倍多。
- 将上述分解过程一般化，假设有$n$个变量组成的联合分布$p(x_1,\ldots,x_n)$，运用条件概率的链式规则，可以得到：

$$p(x_1,\ldots,x_n) = p(x_1)p(x_2|x_1)\ldots p(x_n|x_1,\ldots,x_{n-1}) = \prod_{i=1}^{n} p(x_i|x_1,\ldots,x_{i-1})$$

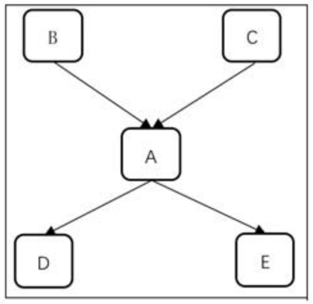
## 警铃问题

- 对于任意的$x_i$，假设存在集合φ$(X)$，φ$(X) \sqsubseteq \{x_1, \dots, x_n\}$，使得在φ$(X)$确定下，$X$与$\{x_1, \dots, x_n\} -$ φ$(X)$中任意元素条件独立，即有$\mathrm{P}(X|x_1, \dots, x_{i-1}) = \mathrm{p}(X|\varphi(x_i))$，于是有：

$$p(x_1, \dots, x_n) = \prod_{i=1}^{n} p(x_i|\varphi(\mathrm{x}))$$

- 假设φ(X)中最多有$m$个元素，一般而言有$m < n$，在$X$只有两个值时，联合概率的独立参数个数最多为$n2^m$个，相比分解前的$2^n - 1$个参数，复杂度已经有下降。在$m \ll n$时，复杂度优化效果更明显
- 从上文的分解结果可见，$X$的分布只依赖于φ$(X)$中变量的取值，而与$\{x_1, \dots, x_n\}$中的其它变量条件独立。Pearl在1986年提出用一个有向无环图来表示这种依赖关系和条件独立性，即变量$X$作为图中的节点，而φ$(X)$中节点都有一条有向边指向$X$

# 警铃问题




| A | B | C | $p(A\|B,C)$ |
|---|---|---|---|
| yes | yes | yes | 0.95 |
| no | yes | yes | 0.05 |
| yes | yes | no | 0.94 |
| no | yes | no | 0.06 |
| yes | no | yes | 0.29 |
| no | no | yes | 0.71 |
| yes | no | no | 0.001 |
| no | no | no | 0.999 |

## 警铃问题

- 如何定量地描述图中节点间的依赖关系呢？
- 根据联合概率分布表和基本公式，可以计算出每个节点的条件概率表。有向无环图和条件概率表就构成了一个贝叶斯网络

| B | p(B) |
|---|---|
| yes | 0.01 |
| no | 0.99 |

| C | p(C) |
|---|---|
| yes | 0.02 |
| no | 0.98 |

| D | A | p(D\|A) | | E | A | p(E\|A) |
|---|---|---|---|---|---|---|
| yes | yes | 0.9 | | yes | yes | 0.7 |
| no | yes | 0.1 | | no | yes | 0.3 |
| yes | no | 0.05 | | yes | no | 0.01 |
| no | no | 0.95 | | no | no | 0.99 |

## 贝叶斯网络的表示

- 贝叶斯网络是使用有向无环图来表示变量间依赖关系的概率图模型。网络中每个节点表示一个随机变量，每一条边表示随机变量间的依赖关系，同时每个节点都对应一个条件概率表(Condition Probability Table，CPT)，用于描述该变量与父变量之间的依赖强度，也就是联合概率分布。
- 贝叶斯网络可以形式化表示。一个贝叶斯网络由结构G和参数$\theta$两部分构成，结构G为有向无环图，图中每一个节点对应一个随机变量。若两个随机变量间有依赖关系，则用一条边将其相连。参数$\theta$定量地表示了变量间的依赖关系，例如若变量$x_i$在G中的父变量集为$y_i$，则$\theta$中有每个变量的条件概率表，即

$$\theta_{x_i|y_i} = p(x_i|y_i)$$

## 贝叶斯网络的构建

- 贝叶斯网络的构建一般有三种方式
  - 根据问题和领域专家知识手工构建
  - 通过对数据进行分析得到贝叶斯网络
  - 结合了领域专家知识和数据分析得到贝叶斯网络
- 贝叶斯网络由有向无环图结构和对应的条件概率表构成，所以手工构建的过程也包括了确定网络结构和确定网络参数两个环节
- 确定网络结构通常的流程是确定能描述问题的一组随机变量$\{x_1, \dots, x_n\}$，对这组随机变量以某种顺序依次添加到结构G中，每一次在添加$x_i$时，需要确定$x_i$在图中依赖的节点集$\varphi(x_i)$，对$\varphi(x_i)$中的节点，添加一条指向$x_i$的有向边
- 网络参数在手工构建时一般通过数据统计分析和专家知识获得,常通过假设条件分布具有某种规律以减少网络参数的个数

## 贝叶斯网络的学习

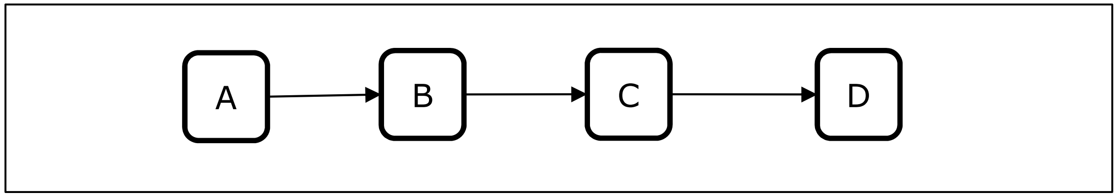
- 贝叶斯网络学习是对数据进行统计分析获取贝叶斯网络的过程。学习包括了参数学习和结构学习两部分。参数学习是在网络结构已知的情况下确定参数即条件概率表中的值。结构学习则既需要确定网络结构G以定性反映变量间的依赖关系，又需要确定网络参数以定量得到条件概率表中的值
- 在对贝叶斯网络进行参数学习时，我们已经知道了网络结构G和G中所有节点或部分节点的状态值，这些状态值就是需要进行学习的数据集

# 贝叶斯网络推理

- 贝叶斯网络的推理是指在已知网络结构G和参数Θ下，给定某些证据或变量的值通过概率论的方法求目标变量值的过程。贝叶斯网络的推理主要包括两种，一种为自顶向下的推理，一种为自底向上的推理
- 推理主要运用的方法有精确推理和近似推理两种，分别有一些算法来解决实际问题。不同情况下有不同因素影响推理，贝叶斯网络拓扑结构和推理任务是两大主要复杂度来源。网络的大小、变量的类型和分布情况、推理任务的类型和相关证据的特征都会影响推理过程和结果，实际应用中也应灵活选择推理方法

## 精确推理

- 精确推理最简单的方法即计算全局的联合概率，但直接对联合概率进行计算的效率很低，常常采用变量消元法分别联合概率的求解达到简化计算的目的。变量消元法利用链式乘积法则和条件独立性对联合概率计算表达式进行变换，改变基本运算的次序改变消元的次序，最终达到减少计算量的目的。该方法的基本思想可以通过一个简单例子描述，假设有如下所示的简单贝叶斯网络



## 精确推理

- 现需要求P(D)，根据已有知识，可以得到：

$$\mathrm{P(D)} = \sum_{A,B,C} P(A,B,C,D) = \sum_{A,B,C} P(A)P(B|A)P(C|B)P(D|C)$$

- 现在对上式做基本运算次序的改变，有：

$$\mathrm{P(D)} = \sum_{C} P(D|C) \sum_{B} P(C|B) \sum_{A} P(A)P(B|A)$$

- 现在的计算量相比改变次序前已经有了较大的降低
- 注意到，上面简单的改变次序使运算局部化，计算只涉及到与某个变量相关的部分，在变量依赖关系复杂的网络中，这种运算局部化可能将指数级降低运算复杂度。上面过程的变量消元次序为{A,B,C,D}，若按照{D,C,B,A}的次序消元，复杂度就不会得到任何减少
- 降低复杂度的关键是找到一个最优的变量消元次序

## 近似推理

- 在贝叶斯网络节点很多或依赖关系很复杂时，精确推理的复杂度很高，通常需要降低推理的复杂度，在问题的因果关系在网络中可独立于某一块存在时，可以将这一部分结构提取出来用精确推理的方法推理。在不能利用局部独立时，就需要降低计算的精度，即采用近似推理的方法。
- 随机抽样算法是最常用的近似推理方法。该方法又被认为蒙特卡洛算法或随机仿真。算法的基本思想上根据某种概率分布进行随机抽样以得到一组随机样本，再根据这一组随机样本近似地估计需要计算的值

# 贝叶斯网络的应用

- 贝叶斯网络经过长期的发展，现已被应用到人工智能的众多领域。包括模式识别、数据挖掘、自然语言处理、辅助智能决策等等。其中针对很多领域核心问题的分类问题，大量卓有成效的算法都是基于贝叶斯理论设计
- 在医疗领域，贝叶斯网络用于医疗诊断
- 在工业领域，贝叶斯网络用于对工业制品的故障检测和性能分析
- 在军事上也被应用于身份识别等各种战场推理
- 在生物农业领域，贝叶斯网络在基因连锁分析、农作物推断、兽医诊断、环境分析等等问题上都有大量的应用
- 在金融领域可用于构建风控模型
- 在企业管理上可用于决策支持
- 在自然语言处理方面可用于文本分类、中文分词、机器翻译

## 中文分词

- 中文分词问题可以描述为给定一句话，将其切分为合乎语法和语义的词语序列。一个经典的中文分词案例是对“南京市长江大桥”的分词。正确的分词结果为“南京市/长江大桥”，错误的分词结果是“南京市长/江大桥”。下面我们使用贝叶斯算法来解决这一问题。
- 设完整的一句话为X，Y为组成该句话的词语集合，共有n个词语。于是分词问题可以转化为求下列式子最大值的问题：

$$\mathrm{p(Y|X)} = \frac{p(Y)p(X|Y)}{p(X)}$$

## 中文分词

- 所以只需找到$p(Y)p(X|Y)$的最大值。由于任意的分词情况下由词语序列生成句子是精确的，所以可以忽略$p(X|Y)$，只需找到$p(Y)$的最大值即可。按照联合概率公式对$p(Y)$进行展开，有
$$p(Y) = p(Y_1, Y_2, \dots, Y_n) = p(Y_1) * p(Y_2|Y_1) * p(Y_3|Y_1, Y_2) * \cdots$$
- 这样的展开子式是指数级增长的，并且数据稀疏的问题也会越来越明显。所以假设每个词语只会依赖于词语序列中该词前面出现的k个词语，即k元语言模型(k-gram)。这里我们假设k=2，于是就有
$$\mathrm{p(Y)} = p(Y_1) * p(Y_2|Y_1) * p(Y_3|Y_2) * \cdots$$
- 回到上面到问题，正常的语料库中，“南京市长”与“江大桥”同时出现的概率一般为0，所以这一分词方式会被舍弃，“南京市/长江大桥”的分词方式会是最终的分词结果

# 机器翻译

- 基于统计的方法是机器翻译常用的实现，这种方法的核心算法即是贝叶斯方法。统计机器翻译问题可以描述为，给定某种源语言的句子X，其可能的目标语言翻译出的句子Y，p(Y|X)代表该种翻译句子符合人类翻译的程度，所以即找到使p(Y|X))最大的Y即可。根据贝叶斯公式，

$$\mathrm{p(Y|X)} = \frac{p(Y)p(X|Y)}{p(X)}$$

- 所以需要找到使得$p(Y)p(X|Y)$最大的Y。对于p(Y)，在中文分词案例中可以利用k元语言模型(k-gram)计算出。而对于$p(X|Y)$，通常利用一个分词对齐的平行语料库，具体言之，将英文“you and me”翻译为汉语，最佳的对应模式为“你和我”，此时有

$$p(X|Y) = \mathrm{p(you|你)p(and|和)p(me|我)}$$

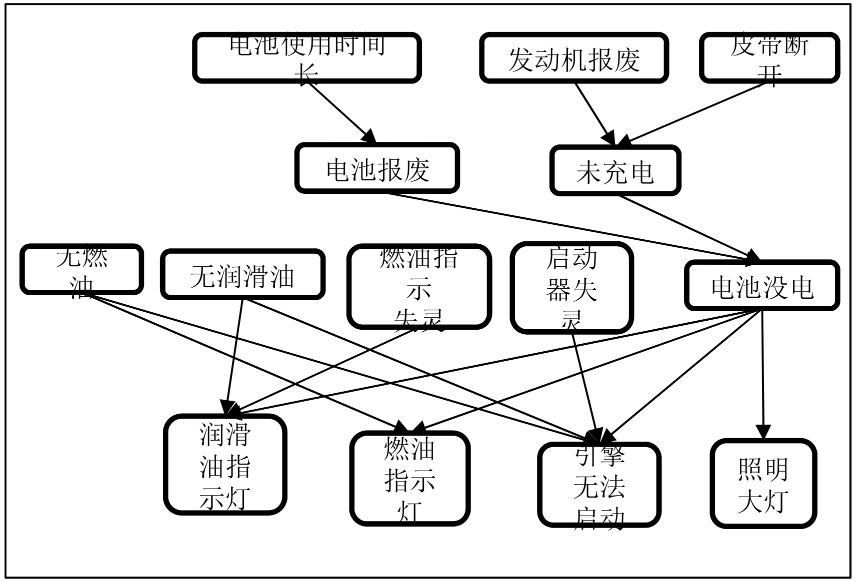
- 上式中右边各项都可以很容易地计算出，所以便可以通过分词对齐的方法计算出$p(X|Y)$的值，最终找到使得p(Y|X)最大的B，便是X最佳的翻译方式

## 故障诊断

- 故障诊断是为了找到某种设备出现故障时故障的所在部件，在工业领域，自动的故障诊断装置能节省一线工作人员大量的预判断时间。基于规则的系统可以被用于故障诊断，但是起不能处理不确定性问题，在实际环境中难以灵活应用。贝叶斯网能较好地描述可能的故障来源，在处理故障诊断的不确定问题上有优秀的表现。多年来研究人员开发出了多种基于贝叶斯网络的故障诊断系统，包括对汽车启动故障的诊断、波音飞机的故障诊断、核电厂软硬件的故障诊断等等

## 故障诊断

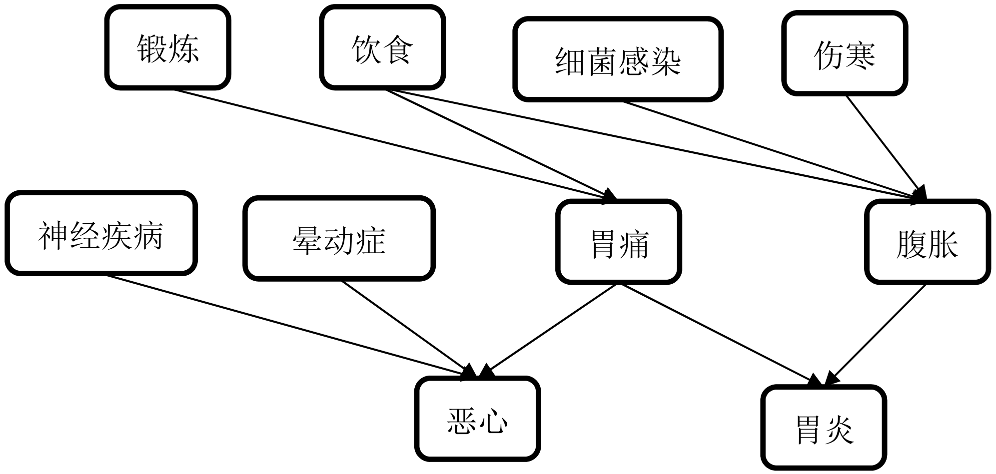
- Heckerman等人在1995提出了一种汽车发动机诊断系统，该系统用于诊断汽车无法正常启动的原因，可见原因有多种，所以可以利用前文提到的诊断推理的方法，找到后验概率最大的故障原因其网络结构如下



## 疾病诊断

- 疾病诊断是从一系列历史经验和临床检验结果中对病人患有疾病种类和患病程度的判断。机器学习在疾病诊断领域有较多的应用，在上世纪70年代就有基于规则设计的产生式专家系统用于对疾病进行诊断，但是该类型系统不能处理不确定性使其诊断正确率远低于临床医生。后来研究人员基于贝叶斯网络设计了新的疾病诊断系统以处理不确定性问题，新系统的诊断准确程度已可与专业临床医生相当。下面是一个对胃部疾病建模的简单贝叶斯网络的部分，网络结构与条件概率不一定符合真实情况，这里关键是对贝叶斯网络的应用予以阐释。假设我们已经根据历史诊断经验得到了如下图所示贝叶斯网络结构

# 疾病诊断



## 疾病诊断

- 其对应的部分条件概率表为：
- 对“锻炼”与“饮食”节点：

| | | | | |
|---|---|---|---|---|
| 锻炼=“是” | 0.5 | | 锻炼=“否” | 0.5 |
| 饮食=“健康” | 0.4 | | 饮食=“亚健康” | 0.6 |

- 对“胃痛”节点：

| | 胃痛=“是” | 胃痛=“否” |
|---|---|---|
| 锻炼=“是”，饮食=“健康” | 0.2 | 0.8 |
| 锻炼=“是”，饮食=“亚健康” | 0.45 | 0.55 |
| 锻炼=“否”，饮食=“健康” | 0.55 | 0.45 |
| 锻炼=“否”，饮食=“亚健康” | 0.7 | 0.3 |

## 疾病诊断

- 对“腹胀”节点：

| | “腹胀”=“是” | 腹胀=“否” |
|---|---|---|
| 饮食=“健康” | 0.2 | 0.8 |
| 饮食=“亚健康” | 0.6 | 0.4 |

- 对“恶心”节点：

| | 恶心=“是” | 恶心=“否” |
|---|---|---|
| 胃痛=“是” | 0.7 | 0.3 |
| 胃痛=“否” | 0.2 | 0.8 |

- 对“胃炎”节点：

| | 胃炎=“是” | 胃炎=“否” |
|---|---|---|
| 胃痛=“是”，腹胀=“是” | 0.8 | 0.2 |
| 胃痛=“是”，腹胀=“否” | 0.6 | 0.4 |
| 胃痛=“否”，腹胀=“是” | 0.4 | 0.6 |
| 胃痛=“否”，腹胀=“否” | 0.1 | 0.9 |

## 疾病诊断

- 现在我们可以利用该贝叶斯网络对患者进行诊断，假设现在只基于给定的条件概率表中有节点进行判断，不考虑没有条件概率的节点。现有患者A，我们对其状况毫不知情，我们需要先判断其是否患有“胃痛”。该问题即转化为求$\mathrm{p}(胃痛=“是”)$的概率。求解过程为：设$\mathrm{x} \in \{是，否\}$表示锻炼情况的两个取值，$\mathrm{y} \in \{健康，亚健康\}$表示饮食情况的两个取值，于是有

$$
\begin{aligned}
&\mathrm{p}(胃痛=“是”)\\
&=\sum_x\sum_y p(胃痛=“是”|锻炼=x，饮食=y)p(锻炼=x，饮食=y)\\
&=0.5\times0.4\times0.2+0.5\times0.6\times0.45+0.5\times0.4\times0.55+0.5\times0.6\times0.7\\
&=0.495
\end{aligned}
$$

所以在没有先验信息情况下患者有“胃痛”病症的可能性为49.5%。

## 疾病诊断

- 假设病人告诉我们他有“恶心”的症状，我们判断其是否患有“胃痛”，问题转化为求p(胃痛 = “是”|恶心 = “是”)的概率。根据贝叶斯概率公式，有

$$p(\text{胃痛}=\text{“是”}|\text{恶心}=\text{“是”})=\frac{p(\text{胃痛}=\text{“是”},\text{恶心}=\text{“是”})p(\text{胃痛}=\text{“是”})}{p(\text{恶心}=\text{“是”})}$$

- 于是需要计算$p(\text{恶心}=\text{“是”})$，设$x\in\{\text{是，否}\}$表示胃痛情况的两个取值，根据全概率公式，有

$$p(\text{恶心}=\text{“是”})=\sum_{x}p(\text{恶心}=\text{“是”}|\text{胃痛}=x)p(\text{胃痛}=x)$$

$$=0.7\times0.495+0.2\times0.505=0.4475$$

- 将值带入上文公式中即有：
- $p(\text{胃痛}=\text{“是”}|\text{恶心}=\text{“是”})=\frac{0.7\times0.495}{0.4475}=0.7743$所以在已知患者有“恶心”症状的情况下，患者有“胃痛”病症的可能性为**77.43%**


THANK
YOU