人工智能原理与应用

华东理工大学

钟伟民 彭鑫 姜庆超 宋冰

课时内容

第4章 不确定性知识系统

4 不确定性知识系统

不确定性知识系统的主要内容包括不确定性知识表示与不确定性推理，但由于不确定性知识表示方法与所采用推理方法有关，故本章重点讨论不确定性推理问题。

- 不确定性推理概述
  a. 不确定性推理的含义
  b. 不确定性推理的基本问题
  c. 不确定性理的类型
- 可信度推理
- 主观Bayes推理
- 证据理论
- 模糊推理
- 概率推理

不确定性推理的含义

■ 不确定性推理的含义

- **什么是不确定性推理**

　　不确定性推理泛指除精确推理以外的其它各种推理问题。包括不完备、不精确知识的推理，模糊知识的推理，非单调性推理等。不确定性推理过程实际上是一种从不确定的初始证据出发，通过运用不确定性知识，最终推出具有一定不确定性但却又是合理或基本合理的结论的思维过程。

- **为什么要采用不确定性推理**

1. 所需知识不完备、不精确所需知识描述模糊、
2. 多种原因导致同一结论、解题方案不唯一

4 不确定性推理的基本问题

■ 不确定性的表示

- **知识的不确定性的表示**

考虑因素： 问题的描述能力，推理中不确定性的计算

含义： 知识的确定性程度，或动态强度

表示：

1. **用概率**，在[0,1]区间取值，越接近于0越假，越接近于1越真；
2. **用可信度**，在[-1,1]区间取值，大于0接近于真，小于0接近于假；
3. **用隶属度**，在[0,1]区间取值，越接近于0隶属度越低，反之越高。

■ 证据不确定性的表示

- **知识的不确定性的表示**

证据的类型：

1. **按证据组织：**基本证据，组合证据；
2. **按证据来源：**初始证据，中间结论；

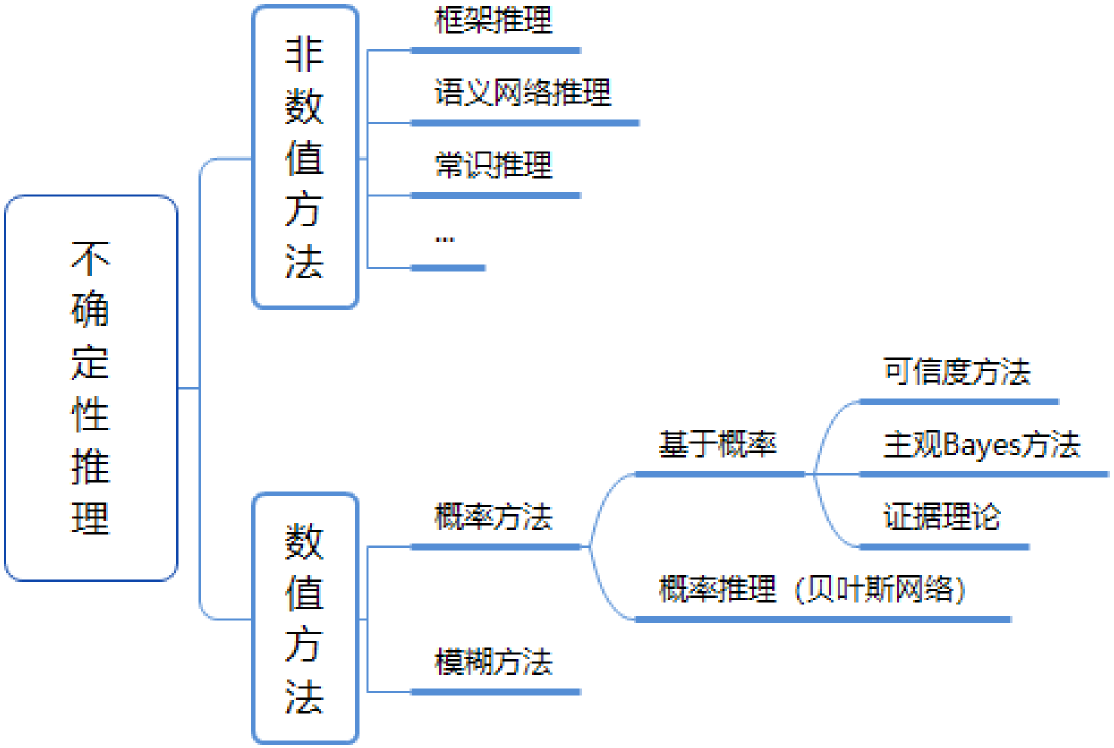
表示方法：概率，可信度，隶属度等；

基本证据：常与知识表示方法一致，如概率，可信度，隶属度等；

组合证据：

1. **组合方式：**析取的关系，合取的关系。
2. **计算方法：**基于基本证据，最大最小方法，概率方法，有界方法等

■ 不确定性推理的类型


不确定性推理
非数值方法
框架推理
语义网络推理
常识推理
...
数值方法
概率方法
基于概率
可信度方法
主观Bayes方法
证据理论
概率推理（贝叶斯网络）
模糊方法

4 不确定性知识系统

- 不确定性推理概述
- 可信度推理
  a. 可信度的概念
  b. 可信度推理模型
  c. 可信度推理的例子
- 主观Bayes推理
- 证据理论
- 模糊推理
- 概率推理

■ 可信度的概念

可信度是指人们根据以往经验对某个事物或现象为真的程度的一个判断，或者说是人们对某个事物或现象为真的相信程度。

例如，沈强昨天没来上课，理由是头疼。就此理由，只有以下两种可能：一是真的头疼了，理由为真；二是没有头疼，理由为假。但就听话人而言，因不能确切知道，就只能某种程度上相信即可信度。

可信度具有一定的主观性，较难把握。但对某一特定领域，让该领域专家给出可信度是可行的。

- 不确定性推理概述
- 可信度推理
- 主观Bayes推理
  a. 主观Bayes方法的概率论基础
  b. 主观Bayes方法的推理模型
  c. 主观Bayes推理的例子
- 证据理论
- 模糊推理
- 概率推理

■ 全概率公式

定理3.1 设事件$A_1, A_2, \ldots, A_n$满足:

(1)任意两个事件都互不相容，即当$i \neq j$时，有$A_i \cap A_j = \Phi \ (i = 1,2, \ldots, n; j = 1,2, \ldots n)$;

(2) $P(A_i) > 0 \ (i = 1, 2, \ldots, n)$;

(3) $D = \cup_{i=1}^{n} A_i$

则对任何事件B由下式成立:

$$P(B) = \sum_{i=1}^{n} P(A_i) \times P(B|A_i)$$

该公式称为全概率公式，它提供了一种计算P(B)的方法。

主观Bayes方法的概率论基础

■ Bayes公式

定理3.2 设事件$A_1, A_2, \dots, A_n$满足定理3.1规定的条件，则对任何事件B有下式成立:

$$P(A_i|B) = \frac{P(A_i) \times P(B|A_i)}{\sum_{j=1}^{n} P(A_j) \times P(B|A_j)} \quad i = 1, 2, \dots, n$$

该定理称为Bayes定理，上式称为Bayes公式。

其中，$P(A_i)$是事件A_i的先验概率，$P(B|A_i)$是在事件A_i,发生条件下事件B的条件概率;$P(A_i|B)$是在事件B发生条件下事件A_i的条件概率。

如果把全概率公式代入Bayes公式，则有:

$$P(A_i|B) = \frac{P(A_i) \times P(B|A_i)}{P(B)} \quad i = 1, 2, \dots, n$$

即

$$P(A_i|B) \times P(B) = P(B_i|A) \times P(A_i) \quad i = 1, 2, \dots, n$$

这是Bayes公式的另一种形式。

Bayes定理给处了用逆概率$P(B|A_i)$求原概率$P(A_i|B)$的方法

4 不确定性知识系统

- 不确定性推理概述
- 可信度推理
- 主观Bayes推理
- 证据理论
- 模糊推理
  - a. 模糊集及其运算
  - b. 模糊关系及其运算
  - c. 模糊知识表示
  - d. 模糊概念的匹配
  - e. 模糊推理的方法
- 概率推理

4 不确定性知识系统

- 不确定性推理概述
- 可信度推理
- 主观Bayes推理
- 证据理论
- 模糊推理
- 概率推理（概念理解即可）
  a. 贝叶斯网络的概念及理论
  b. 贝叶斯网络推理的概念和类型
  c. 贝叶斯网络的精确推理
  d. 贝叶斯网络的近似推理

■ 贝叶斯网络的定义

贝叶斯网络是由美国加州大学的珀尔（J.Pearl）于1985年首先提出的一种模拟人类推理过程中因果关系的不确定性处理模型。它是概率论与图论的结合，其拓扑结构是一个有向无环图。

- 定义

设$X=\{X_1, X_2, \dots, X_n\}$是任何随机变量集，其上的贝叶斯网络可定义为BN=$(B_s, B_p)$。其中:

① B_s是贝叶斯网络的结构，即一个定义在X上的有向无环图。并且，其中的每一个节点X_i都惟一地对应着X中的一个随机变量，并需要标注定量的概率信息；每条有向边都表示它所连接的两个节点之间的条件依赖关系。若存在一条从节点X_j到节点X_i的有向边，则称X_j是X_i的父节点，X_i是X_j的子节点。

■ 贝叶斯网络的定义

- **定义**

设$X=\{X_1, X_2, \ldots, X_n\}$是任何随机变量集，其上的贝叶斯网络可定义为BN= (B_s, B_p)。其中:

② B_p为贝叶斯网络的条件概率集合， $B_p=\{\mathrm{P}(X_i|\mathrm{par}(X_i))\}$。其中，$\mathrm{par}(X_i)$表示$X_i$的所有父节点的相应取值，$\mathrm{P}(X_i|\mathrm{par}(X_i))$是节点$X_i$的一个条件概率分布函数，它描述$X_i$的每个父节点对$X_i$的影响，即节点$X_i$的条件概率表。

从以上定义可以看出，**贝叶斯网络中的弧是有方向的，且不能形成回路，因此图有始点和终点**。在始点上有一个初始概率，在每条弧所连接的节点上有一个条件概率。

■ 贝叶斯网络的全联合概率分布表示

全联合概率分布亦称为全联合概率或联合概率分布，它是概率的一种合取形式，其定义如下。

- **定义**

设$X=\{X_1, X_2, \ldots, X_n\}$为任何随机变量集，其全联合概率分布是指对每个变量取特定值时x_i $(i=1,2,\ldots,n)$时的合取概率，即

$$P(X_1=x_1 \wedge X_2=x_2 \wedge \cdots \wedge X_n=x_n)$$

其简化表示形式为$P(x_1, x_2, \ldots, x_n)$

■ 贝叶斯网络的全联合概率分布表示

由全联合概率分布，再重复使用乘法法则

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

可以把每个合取概率简化为更小的条件概率和更小的合取式，直至得到如下全联合概率分布表示:

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

$$= P(x_n | x_{n-1}, x_{n-2}, \ldots, x_1) P(x_{n-1} | x_{n-2}, x_{n-3}, \ldots, x_1) \ldots P(x_2 | x_1) P(x_1)$$

$$= \prod_{i=1}^{n} P(x_i | x_{i-1}, x_{i-2}, \ldots, x_1)$$

这个恒等式对任何随机变量都是成立的，该式亦称为**链式法则**。

■ 贝叶斯网络的全联合概率分布表示

根据贝叶斯网络的定义，对子节点变量X_i，其取值x_i的条件概率仅依赖于X_i的所有父节点的影响。按照前面的假设，我们用$\mathrm{par}(X_i)$表示X_i的所有父节点的相应取值x_i，$\mathrm{P}(X_i|\mathrm{par}(X_i))$是节点$X_i$的一个条件概率分布函数，则对$X$的所有节点，应有如下联合概率分布：

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|\mathrm{par}(X_i))$$

这个公式就是贝叶斯网络的联合概率分布表示。

可见，贝叶斯网络的联合概率分布要比全联合概率分布简单得多。贝叶斯网络能够大大降低计算复杂度的一个重要原因是其具有局部化特征。

4 贝叶斯网络的概念及理论

■ 贝叶斯网络的全联合概率分布表示

所谓**局部化特征**，是指每个节点只受到整个节点集中少数别的节点的直接影响，而不受这些节点外的其它节点的直接影响。

贝叶斯网络是一种线性复杂度的方法。即在贝叶斯网络中，一个节点仅受该节点的父节点的直接影响，而不受其它节点的直接影响。

例如，在一个包含有n个布尔随机变量的贝叶斯网络中，如果每个随机变量最多只受k个别的随机变量的直接影响，则贝叶斯网络最多可由$2^k n$个数据描述。

■ 贝叶斯网络的条件依赖关系表示

从贝叶斯网络的局部化特征可以看出，贝叶斯网络能实现简化计算的最根本基础是**条件独立性**，即一个节点与它的祖先节点之间是条件独立的。下面从网络拓扑结构去定义下面**两个等价的条件独立关系的判别准则**:

(1) 给定父节点，一个节点与非其后代的节点之间是条件独立的。

(2) 给定一个节点，该节点与其父节点、子节点和子节点的父节点一起构成了一个**马尔科夫覆盖**，则该节点与马尔科夫覆盖以外的所有节点之间都是条件独立的。

■ 贝叶斯网络的构造

- **依据贝叶斯网络的联合概率分布表示，其构造过程如下**

(1) 首先建立不依赖于其它节点的根节点，并且根节点可以不止一个。

(2) 加入受根节点影响的节点，并将这些节点作为根节点的子节点。此时，根节点已成为父节点。

(3) 进一步建立依赖于已建节点的子节点。重复这一过程直到叶节点为止。

(4) 对每个根节点，给出其先验概率；对每个中间节点和叶节点，给出其条件概率表。

4 贝叶斯网络推理的概念及类型

■ 贝叶斯网络推理的概念

贝叶斯网络推理是指利用贝叶斯网络模型进行计算的过程，其基本任务就是要在给定一组证据变量观察值的情况下，利用贝叶斯网络计算一组查询变量的后验概率分布。

假设，用X表示某查询变量，E表示证据变量集$\{E_1, E_2,\ldots, E_n\}$，s表示一个观察到的特定事件，Y表示一个非证据变量（亦称隐含变量）集$\{y_1, y_2,\ldots, y_m\}$，则:

全部变量的集合$V=\{X\}\cup E\cup Y$，其推理就是要查询后验概率$P(X|s)$。

■ 贝叶斯网络推理的类型

步骤: 贝叶斯网络推理的一般步骤是，首先确定各相邻节点之间的初始条件概率分布；然后对各证据节点取值；接着选择适当推理算法对各节点的条件概率分布进行更新；最终得到推理结果。

类型: 贝叶斯网络推理的算法可根据对查询变量后验概率计算的精确度分为**精确推理**和**近似推理**两大类。

精确推理是一种可以精确地计算查询变量的后验概率的一种推理方法。它的一个重要前提是要求贝叶斯网络具有单连通特性，即任意两个节点之间至多只有一条无向路径连接。

但现实世界中复杂问题的贝叶斯网络往往不具有单连通性，而是多连通的。

■ 贝叶斯网络推理的类型

事实上，**多连通贝叶斯网络的复杂度是指数级的。因此，精确推理算法仅适用于规模较小的贝叶斯网络推理。而对复杂的多连通贝叶斯网络，则应该采用近似推理方法**。

近似推理算法是在不影响推理正确性的前提下，通过适当降低推理精确度来提高推理效率的一类方法。

常用的近似推理算法主要有**马尔科夫链蒙特卡洛**(Markov Chain Monte Carlo, MCMC) 算法等。

■ 推理方法

贝叶斯网络精确推理的主要方法包括**基于枚举的算法、基于变量消元的算法和基于团树传播的算法**等。其中，最基本的方法是基于枚举的算法，它使用全联合概率分布去推断查询变量的后验概率：

$$P(X \mid s) = \alpha P(X, s) = \alpha \sum_{Y} P(X, s, Y)$$

其中，各变量的含义如前所述，X表示查询变量；s表示一个观察到的特定事件；Y表示隐含变量集$\{y_1, y_2, \ldots, y_m\}$；a是归一化常数，用于保证相对于X所有取值的后验概率总和等于1。

■ 推理方法

为了对贝叶斯网络进行推理，可利用贝叶斯网络的概率分布公式:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|\mathrm{par}(X_i))$$

将上式中的P(X,s,Y)改写为条件概率乘积的形式。这样，就可通过先对Y的各个枚举值求其条件概率乘积，然后再对各条件概率乘积求总和的方式去计算查询变量的条件概率。

4 贝叶斯网络的近似推理

■ 推理方法

马尔科夫链蒙特卡洛(即MCMC)算法是目前使用较广的一种贝叶斯网络似推理方法。它通过对前一个世界状态作随机改变来生成下一个问题状态，通过对某个隐变量进行随机采样来实现对随机变量的改变。

- **例子**

我们知道，学习情绪会影响学习效果。假设有一个知识点，考虑学生在愉快学习状态下对该知识点的识记、理解、运用的情况，得到了如下图所示的多连通贝叶斯网络。如果目前观察到一个学生不但记住了该知识，并且还可以运用该知识，询问这位学生是否理解了该知识。

END