DOI: 10.13656/j.cnki.gxkx.1998.02.015
广西科学 Guangxi Sciences 1998, 5 (2): 129~ 130

先验概率值的一种学习方法*

A Method of Learning Prior Probability

陈 军 张师超**
Chen Jun Zhang Shichao
(广西财政专科学校 南宁市明秀路 530003)
(Guangxi Training School of Public Finance, Mingxiulu, Nanning, Guangxi, 530003)

摘要 近年来,已提出了一些非确定性信息的理论,如概率理论,证据理论,可能性理论,基于 Bayesian公式的似然推理模型等。这些方法均在不同程度上要求专家给出一些先验概率值,从而影响了它们的应用程度。所以,建立一种有效的获取先验概率值的方法是有必要的。本文给出一种从实际例子中自动获取先验概率值的方法。

关键词 概率推理 不确定性推理 模糊推理 信度网络
中图法分类号 TP 18

Abstract In recent years, some theory of handling uncertainty information, such as probability theory, evidence theory, possibility theory and plausible reasoning model, are established. Because the methods are all based on prior probability, their applicability is constrained. Hence, it is necessary to establish an efficiency way of acquiring prior probability. An automatic approach that the prior probabilites are acquired from examples is suggested in this paper.

Key words Probabilistic reasoning, uncertainty reasoning, fuzzy reasoning, belief network

人类的问题求解和决策活动,通常是在该问题的信息是部分的或近似的环境中进行. AI (人工智能)研究者一直试图在计算机专家系统中模拟人类这一能力. 现在使用的模拟方法大多数缺少理论基础. 近年来,已提出了一些非确定性信息的理论,像概率理论,证据理论,可能性理论等. 这些方法均在不同程度上要求专家给出一些先验概率值,从而影响了它们的应用程度.

最近, J. Pearl建立了基于 Bayesian公式的似然推理模型,它提供了研究不确定推理的基本框架[1]. 不幸的是,该推理模型是基于向量与矩阵相乘的方法,虽然它有坚实的理论基础,但其复杂度相当高,甚至是困难的.另一方面,它要求大量的先验概率值. 由于这两方面的原因,该模型并不实用. 所以,建立一种有效的获取先验概率值的方法是有必要的,本文给出一种从实际例子中自动获取先验概率值的方法.

1 基本概念

用 $A, B, \cdots$ 大写字母表示随机变量,小写字母 $a, b, \cdots$ 表示随机变量的赋值.用点对 $B = (V, P)$ 表示一个 Bayesian网,其中 V是网中随机变量的集合,P是网边上的条件概率的集合. $P(A = a| C_1 = c_1, \cdots, C_n = c_n) \in P$ 当且仅当 $C_1, \cdots, C_n$是 A的交结点,且有一条边连接 C_i 和 A, $i = 1, \cdots, n$.

定义 1 给定一个随机变量 A, A的所有可能取值的集合称为 A的范围,记为 $R(A)$,也称 $R(A)$中的每一个值为 A的点值.

定义 2 随机变量 A的所有点值组成一个向量 $(x_1, x_2, \cdots, x_n)$. A的任意一状态可用点值向量中点值的概率值描述,即 $(P(x_1) = a_1, P(x_2) = a_2, \cdots, P(x_n) = a_n)$ 为 A的一个状态.该状态可简记为 $(a_1, a_2, \cdots, a_n)$. 所有状态组成了 A的状态空间,记为 $S(A)$.

$S(A)$中的元素个数是无穷的,我们感兴趣的通常是 $S(A)$的一个样本空间 \$ (A).设 \$ (A)中有 L个元素,称 L为 \$ (A)的容度并记为 $l($\$ $(A))$.

1997-06-05收稿.

* 得到国家自然科学基金和国家 863计划的资助.

** 广西师范大学数学与计算机系,桂林市育才路 3号, 541004
(Dept. of Maths. & Computer Sci, Guangxi Normal University, 3 Yucailu, Guilin, Guangxi, 541004)

定义 3 对于规则 $X\rightarrow Y$,设 $X=(x_1,x_2,\cdots,x_n)\in S(X)$为$X$的观察值或状态,则可根据 $M_{X|Y}$确定 Y的一个状态 $b=(b_1,b_2,\cdots,b_m)$,即

$$(a_1,a_2,\cdots,a_n)M_{Y|X}=(b_1,b_2,\cdots,b_m)^{-1}$$

或 $aM_{Y|X}=b^{-1}$ (1)

其中,$M_{Y|X}=(c_{ij})_{n*m}$.这就是 Pearl的似然推理模型.

2 学习模型

设 $D=(X_k,Y_k)$是一个实例集合,这里,$k=1,2,\cdots,K$; $X_k\in S(X)$, $Y_k\in S(Y)$.那么,对于 D中任意一个例子,由(1)式有,

$$a_1c_{1i}+a_2c_{2i}+\cdots+a_nc_{ni}=b_i$$

$i=1,2,\cdots,m$.那么,对于 D中的所有元素要求,

$$f(c_{1i},c_{2i},\cdots,c_{ni})=\sum_{d\in D}(a_{j1}c_{1i}+a_{j2}c_{2i}+\cdots+a_{jn}c_{ni}-b_{ji})^2$$

的值尽量小,或者

$$f(c_{1i},c_{2i},\cdots,c_{ni})=\sum_{j=1}^{K}(a_{j1}c_{1i}+a_{j2}c_{2i}+\cdots+a_{jn}c_{ni}-b_{ji})^2$$

的值尽量小,这里,$(a_{j1},a_{j2},\cdots,a_{jn})\in S(X)$,$(b_{j1},b_{j2},\cdots,b_{jm})\in S(Y)$.将上式简记为

$$f(c_{1i},c_{2i},\cdots,c_{ni})=\sum(a_{j1}c_{1i}+a_{j2}c_{2i}+\cdots+a_{jn}c_{ni}-b_{ji})^2$$

根据数学分析中极值原理可知,要 $f(c_{1i},c_{2i},\cdots,c_{ni})$尽量小,即对 $f(c_{1i},c_{2i},\cdots,c_{ni})$分别求关于 $c_{1i},c_{2i},\cdots,c_{ni}$的偏导数并令它们等于零,即

$$\begin{cases}\dfrac{\partial f}{\partial c_{1i}}=2\sum(a_{j1}c_{1i}+a_{j2}c_{2i}+\cdots+a_{jn}c_{ni}-b_{ji})a_{j1}=0\\ \dfrac{\partial f}{\partial c_{2i}}=2\sum(a_{j1}c_{1i}+a_{j2}c_{2i}+\cdots+a_{jn}c_{ni}-b_{ji})a_{j2}=0\\ \cdots\\ \dfrac{\partial f}{\partial c_{ni}}=2\sum(a_{j1}c_{1i}+a_{j2}c_{2i}+\cdots+a_{jn}c_{ni}-b_{ji})a_{jn}=0\end{cases}$$

经整理可得

$$\begin{cases}c_{1i}\sum(a_{j1})^2+c_{2i}\sum(a_{j1}a_{j2})+\cdots+c_{ni}\sum(a_{j1}a_{jn})-b_{ji}a_{j1}=0\\ c_{1i}\sum(a_{j2}a_{j1})+c_{2i}\sum(a_{j2})^2+\cdots+c_{ni}\sum(a_{j2}a_{jn})-b_{ji}a_{j2}=0\\ \cdots\\ c_{1i}\sum(a_{jn}a_{j1})+c_{2i}\sum(a_{jn}a_{j2})+\cdots+c_{ni}\sum(a_{j})^2-b_{ji}a_{jn}=0\end{cases}$$

设上面方程组关于 $c_{1i},c_{2i},\cdots,c_{ni}$的系数矩阵为 A,如果 A的行列式 $d=|A|\neq 0$,那么该方程组有唯一解,其解为

$$c_{1i}=\frac{d_1}{d},c_{2i}=\frac{d_2}{d},\cdots,c_{ni}=\frac{d_n}{d}$$

其中 d_i是把矩阵 A中第 i列换成方程组的常数项 $b_{ji}a_{j1},b_{ji}a_{j2},\cdots,b_{ji}a_{jn}$所成的矩阵的行列式.

由此得到的值 $c_{1i},c_{2i},\cdots,c_{ni}$是关于点值 y_i在条件 x_i下的先验概率值.要保证它的概率意义,必须按如下方式处理,

$$c_{ji}=c_{ji}/(c_{1i}+c_{2i}+\cdots+c_{ni})$$

这里,$i=1,2,\cdots,m$; $j=1,2,\cdots,n$.

3 算法

获取先验概率值的算法如下.

```
Begin
 输入实例集合 D;
  For i: = 1 To m
   Begin
    计算 d;
    c: = 0;
    For j: = 1 To n
     Begin
      计算 dj;
      cji: = dj /d;
      c: = c+ cji
     End;
     For j: = 1 To n
      Begin
       cji = cji /c;
       输出 cji
      End
     End
    End.
```

上面的方法综合了一个样本的所有点值的概率意义,它比概率论中单点值统计先验概率值的方法要优越,而且,它需要的样本数量不多.

参考文献

1 Pearl J, Fusion. Propagation and structuring in belief networks. Artif Intell, 1996, 29: 241~288.

2 张师超等.不确定性推理技术.桂林:广西师范大学出版社,1996,6.

(责任编辑:黎贞崇)