国 防 科 技 大 学 学 报
第 21 卷 第 1 期 JOURNAL OF NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY Vol. 21 No. 1 1999

条件随机变量与条件数据融合*

李兵 罗雪山 罗爱民

(国防科技大学系统工程与数学系 长沙 410073)

摘 要 本文讨论基于条件事件代数系统上的条件随机变量的表示问题, 并给出了在多个信息源提供的不同类型的条件数据时之统一表示方法, 以及利用条件随机变量分布函数性质, 对所得数据进行关联等。

关键词 条件事件代数, 条件随机变量, 条件数据融合

分类号 O236.

Conditional Ranolom Variables and Conditional Data Fusion

Li Bing Luo xueshan Luo aiming

(Department of systems Engineering and Mathematics, NUDT, Changsha, 410073)

Abstract In this paper we discuss an expression of conditional random variables at first, give their conditional distributions, and apply these to conditional data fusion.

Key words Conditional event algebra, Conditional random variable, Conditional data fusion

条件事件代数理论又称为条件信息理论, 它是针对知识表示与推理中的条件问题, 通过研究条件事件代数结构与条件推理, 达到充分利用条件信息来解决不确定性问题的。可以说, 凡是涉及到专家系统中的产生式规则 "If b, then a" 的问题都与条件事件代数问题相关。条件事件代数是在确保规则概率与条件概率相容的前提下, 把布尔代数上的逻辑运算推广到条件事件(规则)集合中而得到的代数系统, 目的是为智能系统中的条件推理建立一个数学基础, 也是计算机进行高层次推理与高层次专家系统所必需的工具[1,2]。故其理论与实践价值都是巨大的, 美国将其列为重要研究课题之一。

在数据融合中我们需对来自多信息源的信息进行综合处理和评估。我们知道随机变量在经典随机分析中是最为基本的工具, 同样在条件事件代数系统上, 引入条件随机变量亦是十分重要的, 条件随机变量所反映的是在不同场景下对观测目标的观测结果。本文讨论条件随机变量的表示及其分布问题, 并给出了在多个信息源提供的不同类型的条件数据时统一表示方法, 及利用条件随机变量分布函数性质对所得数据进行关联等。

1 条件随机变量及其分布

为了引入条件随机变量, 我们先给出一个引理

引理 设 D, E 为任两集合, 而 $R \subset P(D)$, $S \subset P(E)$ 为两布尔代数, 其中 $P(D)$, $P(E)$ 分别为 D, E 的幂集, $f: D \to E$ 为任一函数, 其函数像及函数逆像扩张分别为 $f: P(D) \to P(E)$ 与 $f^{-1}: P(E) \to P(D)$, 则可依下法得 $f^{-1}: (S\ S) \to (R\ R)$:

$$f^{-1}(c\ d) = (f^{-1}(c)\ f^{-1}(d)), \forall c, d \in S$$

设 (R^m, B^m) 为 m 维 Borel 空间, (Ω, F, p) 为一给定的概率空间, $V: \Omega \to R^n$, $W: \Omega \to R^m$ 为随机变量(向量)。由于它们象空间的维数不同, 为以下引入条件随机变量时便于处理, 我们用提升的办法对任意的 $a \in B^m$, $b \in B^n$, 定义乘积形式的条件事件如下:

$$[a\ b] \equiv (a \times b\ R^m \times b)$$

* 国家预研基金与国防科技大学试验技术研究项目资助
1998 年 6 月 22 日收稿
第一作者: 李兵, 男, 1963 年生, 副教授

并定义 $$[B^m\ \ B^n] \equiv \{[a\ \ b] : a \in B^m, b \in B^n\}$$
及联合映射 $V \times W : (\Omega, F) \to (R^m \times R^n, B^m \times B^n)$:
$$(V \times W)(x) \equiv (V(x), W(x))$$
其函数像扩张仍记为 $V \times W : P(\Omega) \to P(R^m) \times P(R^n)$, 从而可得映射 $V \times W : F \to B^m \times B^n$, 继而可得映射 $V \times W : (F\ \ F) \to (B^m \times B^n\ \ B^m \times B^n)$, 及其逆像 $(V \times W)^{-1} : (B^m \times B^n\ \ B^m \times B^n) \to (F\ \ F)$, 定义映射 $[V\ \ W]^{-1} : [B^m\ \ B^n] \to (F\ \ F)$ 为上面映射 $(V \times W)^{-1}$ 在 $[B^m\ \ B^n]$ 上的限制, 再将此映射视为引理中的 f, 则由引理可得映射 $[V\ \ W] : (F\ \ F) \to [B^m\ \ B^n]$, 我们称此映射为条件随机变量。这里是为了与无条件情形相对应而这样称呼的, 实际上确切地说, 它是一定义于概率空间 $((F\ \ F), P(F\ \ F), p)$ 上的随机子集。

定义条件随机变量 $[V\ \ W]$ 的分布如下:
$$p([V\ \ W] \text{ 在} [a\ \ b] \text{ 中}) \equiv p([V\ \ W]^{-1}[a\ \ b]) = p(V^{-1}(a)\ \ W^{-1}(b))$$
$$= p([V \text{ 在 } a \text{ 中}\ \ W \text{ 在 } b \text{ 中}])$$

对于多个条件随机变量的联合条件分布定义如下:

设 $V_j : \Omega \to R^{m_j}, W_j : \Omega \to R^{n_j}; \forall a_j \in B^{m_j}, b_j \in B^{n_j}, j = 1, 2$
$$p([V_1\ \ W_1] \text{ 在} [a_1\ \ b_1] \text{ 中}, [V_2\ \ W_2] \text{ 在} [a_2\ \ b_2] \text{ 中})$$
$$\equiv p([V_1\ \ W_1]^{-1}[a_1\ \ b_1]\ \ \ [V_2\ \ W_2]^{-1}[a_2\ \ b_2])$$
$$= p\left(V_1^{-1}(a_1)\ \ W_1^{-1}(b_1)\ \ \ V_2^{-1}(a_2)\ \ W_2^{-1}(b_2)\right)$$
$$= p\left(V_1 \times W_1 \times V_2 \times W_2 \text{ 在} [a_1\ \ b_1] \times [a_2\ \ b_2] \text{ 中}\right)$$

$\forall t \in R^n$, 对任给定具有 $p(b_1 \cdots b_n) > 0$ 的 $b_j \in A, j = 1, \cdots n$, 定义条件随机变量 $[V_1\ \ W_1] \cdots [V_n\ \ W_n]$ 的联合条件分布函数 $F_b : R^n \to [0, 1]$ 如下: $\forall t \equiv (t_1, \cdots t_n) \in R^n$
$$F_b(t) \equiv p((a_1(t_1)\ \ b_1) \cdots (a_n(t_n)\ \ b_n))$$
其中事件 $a_i(t_i) \equiv (V_i \le t_i), i = 1, \cdots, n$。

特别当 $n = 2$ 时, 由于
$$((V_1 \le t_1)\ \ b_1)\ \ \ ((V_2 \le t_2)\ \ b_2) = ((V_1 \le t_1)\ \ \ (V_2 \le t_2)\ \ (V_1 > t_1)\ \ \ b_1 \vee (V_2 > t_2)\ \ \ b_2 \vee b_1\ \ \ b_2)$$
从而
$$F_b = p(((V_1 \le t_1)\ \ b_1)\ \ \ ((V_2 \le t_2)\ \ b_2))$$
$$= p((V_1 \le t_1)\ \ \ (V_2 \le t_2)\ \ (V_1 > t_1)\ \ \ b_1 \vee (V_2 > t_2)\ \ \ b_2 \vee b_1\ \ \ b_2)$$
$$= p((V_1 \le t_1)\ \ \ (V_2 \le t_2)\ \ \ b_1\ \ \ b_2) / [p((V_1 > t_1)\ \ \ b_1 \vee (V_2 > t_2)\ \ \ b_2 \vee b_1\ \ \ b_2)]$$
其中 $p((V_1 > t_1)\ \ \ b_1 \vee (V_2 > t_2)\ \ \ b_2 \vee b_1\ \ \ b_2)$
$$= p(b_1\ \ \ b_2) + p((V_1 > t_1)\ \ \ b_1) + p((V_2 > t_2)\ \ \ b_2) - p((V_1 > t_1)\ \ \ b_1\ \ \ (V_2 > t_2)\ \ \ b_2)$$
$$- p((V_1 > t_1)\ \ \ b_1\ \ \ b_2) - p((V_2 > t_2)\ \ \ b_1\ \ \ b_2) + p((V_1 \le t_1)\ \ \ (V_2 \le t_2)\ \ \ b_1\ \ \ b_2)$$
$$= (c - \beta(t_1) - \gamma(t_2))$$

因此
$$F_b(t) = \alpha(t) / (c - \beta(t_1) - \gamma(t_2))$$
其中 $\alpha(t) \equiv p(a(t_1)\ \ \ a(t_2)\ \ \ b_1\ \ \ b_2)$ $\beta(t_1) \equiv p\left(a(t_1)\ \ \ b_1\ \ \ b_2'\right)$

$\gamma(t_2) \equiv p\left(a(t_2)\ \ \ b_1'\ \ \ b_2\right)$ $c \equiv p(b_1 \vee b_2)$

(b_i' 为 b_i 的逆事件)。

若 F_b 的概率密度函数 f_b 存在, 则有
$$f_b(t) = \partial^2 F_b(t) / \partial_1 \partial_2 = A(t) / B(t)$$
其中 $B(t) \equiv c - \beta(t_1) - \gamma(t_2)$
$$B(t) \equiv 2B(t)\ \ \ \alpha(t)\ \ \ \beta'(t_1)\ \ \ \gamma'(t_2) + B(t)^2\ \ \ (\partial\alpha(t)/\partial_1)\ \ \ \gamma'(t_2) + (\partial\alpha(t)/\partial_2)\ \ \ \beta'(t_1)$$
$$+ B(t)^2\ \ \ \partial^2\alpha(t) / \partial_1\partial_2$$
由此可见, 即使每个 $p(\alpha(t_j)\ \ b_j)$ 关于变元 t_j 是服从高斯分布的 $(j = 1, 2, \cdots, n)$, f_b 一般并不是高斯分布的联合密度形式。但当 $b_1 = b_2$ 时为高斯分布的联合密度形式。

在数据融合中, 对已得到的信息即一组事件 $b=(b_1, b_2, \cdots b_n)$, 求出其联合条件分布函数后, 便可用概率统计方法, 得出关于此条件下的条件信息, 并由此进行相应的推断和决策。

2 条件数据融合

设有两个不同的信息源, 它们均可用来对未知参数 $\theta \in D \subset R$ 进行估计(其中 D 为参数空间), 我们分以下几种情形进行讨论如何对条件数据进行融合。

(1) 两个信息源依无条件随机变量形式进行随机描述

设 $V_1, V_2: \Omega \to D$, $\forall a_1, a_2 \in B$, 可依下式计算联合概率

$$p(V_1 \text{ 在 } a_1 \text{ 中}, V_2 \text{ 在 } a_2 \text{ 中}) = p(V_1^{-1}(a_1) \quad V_2^{-1}(a_2))$$

从而应用对角化方法将上式限制在 $a_1 = a_2 = \{x\}$, $x \in D$ 即可得所需结果。

以下给出一种建立模糊集理论与概率论之间转换的方法。

对任给定的函数 $f: D \to [0, 1]$, 对某概率空间 (Ω, A, p), 存在(无穷多个) 随机集 $S: \Omega \to P(D)$, 使得 S 是等价于 f 的单点覆盖函数, 即

$$p(\tilde{\omega}: x \in S(\tilde{\omega})) = f(x), \forall x \in D$$

这样的随机集是易于构造的, 例如可取典则网随机集或随机水平集(或随机截集) $S_f(U): \Omega \to [0, 1]$, 其中 $U: \Omega \to [0, 1]$ 为一服从均匀分布的随机变量

$$S_f(U) \equiv f^{-1}(U, 1) = \{x: x \in D, f(x) \geq U\}$$

这样可将对模糊集的研究转换为对随机集的研究。

(2) 两个信息源依条件随机变量形式进行随机描述

设 $(V_j \ W_j)$ 为条件随机变量, 其中 $V_j: \Omega \to D$, $W_j: \Omega \to B_j \subset R^{m_j}$, $j = 1, 2$, 则可由前面的定理计算联合概率:

$$p(\{(V_1 \ W_1) \text{ 在}[a_1 \ b_1] \text{ 中}\} \cap \{(V_2 \ W_2) \text{ 在}[a_2 \ b_2] \text{ 中}\}), \text{其中 } a_j \in D, b_j \in B_j, j = 1, 2\text{。}$$

(3) 信息源 1 是基于自然语言的, 其隶属函数为 $g: D \to [0, 1]$; 信息源 2 的描述为无条件随机变量 $V: \Omega \to D$

首先将 g 用与之等价的网随机集 $S_g(U): \Omega \to P(D)$ 替代, 其中 $U: \Omega \to [0, 1]$ 为一服从均匀分布且与 V 统计独立随机变量, $S_g(U) \equiv g^{-1}[U, 1] = \{x: x \in D, g(x) \geq U\}$。注意由假设可知: $S_g(U)$ 与 V 是统计独立的。

其次, 为了应用上与 $S_g(U)$ 相对应, 将 V 用其扩张 $\hat{V}: \Omega \to P(D)$ 替代, $\hat{V}$ 为 D 的随机子集, 具有概率分布如下:

$$p(\hat{V} = a) = p(V \text{ 在 } a \text{ 中}) / \sum_{b \in P(D)} p(V \text{ 在 } b \text{ 中}), \forall a \in P(D)$$

这样对于信息源 1 与 2, 通过上述转换, 只需考虑 $S_g(U)$ 与 $\hat{V}$ 即可。

(4) 两个信息源均是基于自然语言

此时同上类似, 只需将两模糊从属函数都转换成网随机集的形式即可。

(5) 两个信息源中的一个或者全部为条件自然语言

此时只需综合前面几种情形作法即可。

还可讨论考虑更为一般的情形。设有 n 个信息源 s_j, $j = 1, \cdots, n$, 其中的某些可能是基于自然语言的, 也可能是基于随机的; 某些是条件的(基于自然语言或随机的), 也可能是无条件的。

为了统一以便于讨论, 我们将基于自然语言的如 s_j, $j = 1, \cdots, m$, 转化为随机形式: 网随机集 $S(s_j)$; 由于无条件情形为条件情形的特殊形式, 故不妨将全体 $S(s_j)$ 表示为条件随机集。

这样可由前面所讨论的条件随机变量的方法, 得到联合条件概率分布 $F(S(\underline{s})): D^n \to R$ 或联合条件概率密度 $f(S(\underline{s})): D^n \to R^+$, (这里记 $\underline{s} \equiv (s_1, \cdots, s_n)$), 由此对基于信息源 $\underline{s}$ 的决策问题讨论, 可转化为基于 $\mathrm{diag}(f(S(\underline{s}))): D \to R^+$ 上的决策问题讨论(其中 $\mathrm{diag}(f)(x) \equiv f(x_1, \cdots, x_n) / \int_{x \in D} f(x, \cdots, x)\, dx$

为f 的对角化)。从而最终可应用标准概率模型中的最优决策、决策误差等方法来处理这方面的问题。

令θ表示两目标的相关水平, $0\leq\theta\leq 1$。当$\theta=0$时表示二者无关联; 当$\theta=1$表示二者可确认为一致。只有通过观察数据$Y=(Y_1, Y_2)$才有可能掌握θ的值, 其中的Y_1, Y_2分别表示目标1, 2的所对应的属性值, 例如对目标舰船1与2进行观察, 记

$$Y_i=\{第 i 个目标的空间状态、速度、悬挂旗的颜色、形状\}$$

其中的每一项均为其部分属性特征。通过匹配规则将目标舰船1与2进行关联分析。例如用某类距离函数$M_1(Z_{11}, Z_{12})$对属性1状态进行分析, 其中Z_{1i}表示第i目标的真实状态; 用某类自然语言$M_4(Z_{41}, Z_{42})$对属性3颜色进行分析。这样, 首先通过各个相应属性的关联分别得到$\theta_1, \theta_2, \theta_3, \theta_4$, 故下面不再将$\theta$视为一维的数, 而记$\theta=(\theta_1, \theta_2, \theta_3, \theta_4)$。通过综合评判最后得到两目标的关联水平$Q(\theta)$。若$Q(\theta)=0$表示二者无关联; 若$Q(\theta)=1$表示二者可确认为一致。

一般地, 设$\theta\in D\subset R^m$为所关心的未知参数, (其中参数空间D为已知集, D^n为其乘积空间), $f_j: D\to R^+$为概率密度函数(已知), 它是对第j个信息源的描述($j=1, 2, \cdots, n$), $f: D^n\to R^+$为联合概率密度函数(已知), 一般可认为(f_j)为相互统计独立, 从而f为(f_j)的乘积。

由于各信息源所提供的信息的重要程度不一致, 故常要进行加权处理。设n为正整数, 权重w_j, λ_j满足: $0\leq w_j, \lambda_j\leq 1$; $1=w_1+w_2=\lambda_1+\cdots+\lambda_n$。$f: D^n\to R^+$为刻划参数$\theta\in D$的$n$个信息源的联合概率密度函数, $h: D^n\to D$为从联合空间到参数空间自身的过程, 即融合n个信息源的过程, 它相当于统计中的一个估计量。定义h的损失函数$L(h)$如下:

$$L(h)\equiv\int_{\underline{x}\in D^n}\left[\left(w_1\sum_{j=1}^{n}\lambda_j\left(x_j-h(x)\right)^2+w_2\sum_{1\leq i<j\leq n}(x_i-x_j)^2\quad h(\underline{x})^2\right)\right]f(\underline{x})d\underline{x}$$

即它为$h(\underline{x})$与所有可能的分量x_j间的平方距离的期望的加权组合。

我们以使损失函数$L(h)$达到最小的h作为某种最优的评判。例如对于前面的关联问题, 取$Q=h$即可得到$0\leq Q(h)\leq 1$。

参考文献

1　IEEE Transaction on S. M. C., 1994, 24(12)

2　Goodman I R. . The role of conditional event algebra in the modeling of C^3 systems. ADA262 985/5GAR

3　李兵. 条件事件的表示. 国防科技大学学报. 1998, 20(3)