第21卷　第3期　　华侨大学学报(自然科学版)　　Vol.21　No.3
2000年7月　　Journal of Huaqiao University (Natural Science)　　Jul.2000

文章编号　1000-5013(2000)03-0228-06

应用简易模糊推理的多级判别与预测

陈　治　典

(华侨大学管理信息科学系, 泉州 362011)

摘要　将模糊数学与多元统计相结合, 提出一种模糊推理的简易方法, 可避免计算多维模糊关系矩阵的困难. 所述方法能更多地利用相关因子(线性的及非线性的)信息, 具有处理模糊信息的能力, 不受样本数据概率分布的限制, 因而提高了多级判别与预报的效果. 它可用作趋势及定值预报, 能更广泛地应用于实际问题.

关键词　模糊推理, 简易方法, 多级判别, 预测

中图分类号　O 159 : O 212.4　　文献标识码　A

通常的模糊推理方法必须解决两个问题. (1) 模糊关系生成规则. 在实际问题中会遇到复合(即多输入)多重模糊命题, 即"若 A_{11} 且 A_{12} 且…且 A_{1m} 则 B_1, 否则, 若 A_{21} 且 A_{22} 且…且 A_{2m} 则 B_2, 否则, …, 若 A_{n1} 且 A_{n2} 且…且 A_{nm} 则 B_n". 其中 A_{ij} 皆为 X_j 上的模糊集, B_i 皆为 Y 上的模糊集($i=1,2,\cdots,n; j=1,2,\cdots,m$). 这时 R 可定义为 $R=\bigcup_{i=1}^{n}(A_{i1}\times A_{i2}\times\cdots\times A_{im}\times B_i)$, 按常用的 Mamdani 算法[1], R 的从属函数为

$$\mu_R(x_1,x_2,\cdots,x_m,y)=\bigvee_{i=1}^{n}\bigwedge_{k=1}^{m}[\mu_{A_{ik}}(x_k)\wedge\mu_{B_i}(y)],\quad x_k\in X_k(k=1,2,\cdots,m),\ y\in Y. \tag{1}$$

若直接按式(1)计算, 即使诸 X_i 及 Y 都是有限集, R 也表现为一个 $m+1$ 维矩阵, 计算相当麻烦. 当然, 可采用一些简化的方法来克服这一困难[2]. 不过, 无论如何, 计算量总是非常大的. (2) 模糊推理合成规则. 由模糊蕴涵关系 R 和推理的小前提 A' (在多输入问题中为 $A'_1, A'_2, \cdots, A'_m$), 用合成运算得出 Y 上的模糊集 $B'=A'\circ R$ 作为近似推论. 这又得作大量的计算. 在上述式(1)中, 运算采用 Zadeh 算子"∧, ∨". 若用别的算子计算则更繁. 其实模糊推理是一种近似推理, 即使计算得很精确, 小前提与推理结论之间也只是一种近似的对应关系. 若能用一种简单的方法由小前提直接得出其所对应的近似推论, 则可免除许多繁琐的计算. 本文所提出的方法, 可达此目的.

1　方法步骤

1.1　对观测资料的整理

设研究对象为 y, 影响 y 的重要因素有 $x_1, x_2, \cdots, x_m$. 已知 y 及各影响因素的 n 个观测值

收稿日期　1999-09-29　　作者简介　陈治典(1940-), 男, 副教授

基金项目　华侨大学优秀课程建设基金资助项目

$$y_1, y_2, \cdots, y_n, \quad x_{i1}, x_{i2}, \cdots, x_{in}, \quad i = 1, 2, \cdots, m.$$

将 y 的观测值按从小到大的顺序排列, 依次记作 $y_{(1)}, y_{(2)}, \cdots, y_{(n)}$. 将各影响因素的观测值按对应于 $y_{(1)}, y_{(2)}, \cdots, y_{(n)}$ 的次序重排, 依次记作 $x_{i(1)}, x_{i(2)}, \cdots, x_{i(n)}$ ($i= 1, 2, \cdots, m$). 若选定的因素确实与 y 的关系密切, 则两者的观测值重排后的大小顺序应基本相同或相反(可去掉那些跟 $y_{(1)}, y_{(2)}, \cdots, y_{(n)}$ 大小顺序对应很差的因素).

1.2 对研究对象数据的分类(分级)

根据研究需要及研究对象 y 的实际情况, 将 y 的数据适当地分成 k 类(级), 使各类之间有明显的差异, 分别记作 $Y_{(1)}, Y_{(2)}, \cdots, Y_{(k)}$. 由于客观事物的属性往往具有模糊性, $Y_{(1)}, Y_{(2)}, \cdots, Y_{(k)}$ 均应视为某个论域 Y 中的模糊子集. 从理论上或经验上求出各类的聚类中心(比如以该类的平均取值为聚类中心). 设第 g 类的聚类中心为 $\overline{y}^{(g)}$, 则 $Y_{(g)}$ 的从属函数可定义为

$$\mu_{Y_{(g)}}(y) = 1 - \frac{\left|y - \overline{y}^{(g)}\right|}{\max\limits_{1\leq s\leq k}\left|y - \overline{y}^{(s)}\right|}, \qquad y \in Y, g = 1, 2, \cdots, k. \tag{2}$$

将已知的观测资料按 y 之值从小到大的顺序重新排列后, 就形成了所谓“有序样品”, 即可按文[3]的方法进行聚类, 或者直接按最大隶属原则[4]聚类(但这必须先定义了各类标准模式的隶属函数才行). 最简单的是直接按应用部门的规定将 y 的观测值归类, 比如将降雨量少于某值的归于“偏少”类, 将降雨量介于某区间中的归于“正常”类, 将降雨量大于某值的归于“偏多”类. 设将 y 的已知观测值按大小顺序分成了下面 k 类:

$$Y^{(1)} = \{y_{(1)}, y_{(2)}, \cdots, y_{(n_1)}\}, \qquad Y^{(2)} = \{y_{n_1+1)}, y_{n_1+2}, \cdots, y_{n_1+n_2)}\}, \cdots,$$

$$Y^{(k)} = \{y_{(n-n_k+1)}, y_{(n-n_k+2)}, \cdots, y_{(n)}\}.$$

1.3 对各因子(影响因素)数据的分类

假定各因子的数据均含有对 y 的相应数据分类的信息. 按 y 的分类, 将同一类的 y 值及对应的各因子数据归在一起列成表 1, 以 G_g 表示第 g 类. 计算各个因子的各类平均值作为该因子该类的标准模式:

$$\overline{x}_i^{(g)} = \frac{1}{n_g}\sum_{x_{i(l)}\in G_g} x_{i(l)}, \qquad g = 1, 2, \cdots, k, \quad i = 1, 2, \cdots, m\left(\sum_{g=1}^{k} n_g = n\right) \tag{3}$$

(也可根据理论或经验求出各因子的各类标准模式). 由于事物属性的模糊性, 各因子的各类也应视为相应论域中的模糊子集. 以 X_i^g 表示因子 x_i 的第 g 类, 它是 x_i 值域 X_i 上的模糊集, 其从属函数可定义为

$$\mu_{X_i^{(g)}}(x_i) = 1 - \frac{\left|x_i - \overline{X}_i^{(g)}\right|}{\max\limits_{1\leq s\leq k}\left|x_i - \overline{x}_i^{(s)}\right|} \qquad (x_i \in X_i, g = 1, 2, \cdots, k). \tag{4}$$

对于因子 x_i 的任一观测值 x_{ij}, 若有某个 $l(1\leq l\leq k)$ 使得

$$\mu_{X_i^{(l)}}(x_{ij}) = \max_{1\leq g\leq k}\mu_{X_i^{(g)}}(x_{ij}), \tag{5}$$

则判断 x_{ij} 属于 x_i 的第 l 类. 将各个因子的所有已知数据和对应的 y 的数据的类别, 都列入同一个表中, 如表 2 所示.

1.4 确定模糊推理法则及进行模糊推理

每一个因子数据的类别与 y 之值类别的对应就是一条模糊蕴涵命题. 同时考虑 m 个因子的数据的类别与 y 之值类别的对应, 就可看作一条复合多重模糊蕴涵命题, 比如“若 $X_1^{(l)}$ 且

表 1　同一类的 y 值及对应的各因子数据

类别	序号	y	x_1	x_2	…	x_m
第	1	$y_{(1)}$	$x_{1(1)}$	$x_{2(1)}$	…	$x_{m(1)}$
	2	$y_{(2)}$	$x_{1(2)}$	$x_{2(2)}$	…	$_{m(2)}$
	⋮	⋮	⋮	⋮	⋮	⋮
1	n_1	$y_{(n_1)}$	$x_{1(n_1)}$	$x_{2(n_1)}$	…	$x_{m(n_1)}$
类			$\overline{x}_1^{(1)}=\frac{1}{n_1}\sum_{l=1}^{n_1}x_{1(l)}$	$\overline{x}_2^{(1)}=\frac{1}{n_1}\sum_{l=1}^{n_1}x_{2(l)}$	…	$\overline{x}_m^{(1)}=\frac{1}{n_1}\sum_{l=1}^{n_1}x_{m(l)}$
第	n_{1+1}	$y_{(n_1+1)}$	$x_{1(n_1+1)}$	$x_{2(n_1+1)}$	…	$x_{m(n_1+1)}$
	n_{1+2}	$y_{(n_1+2)}$	$x_{1(n_1+2)}$	$x_{2(n_1+2)}$	…	$x_{m(n_1+2)}$
	⋮	⋮	⋮	⋮	⋮	⋮
2	n_1+n_2	$y_{(n_1+n_2)}$	$x_{1(n_1+n_2)}$	$x_{2(n_1+n_2)}$	…	$x_{m(n_1+n_2)}$
类			$\overline{x}_1^{(2)}=\frac{1}{n_2}\sum_{l=n_1+1}^{n_1+n_2}x_{1(l)}$	$\overline{x}_2^{(2)}=\frac{1}{n_2}\sum_{l=n_1+1}^{n_1+n_2}x_{2(l)}$	…	$\overline{x}_m^{(2)}=\frac{1}{n_2}\sum_{l=n_1+1}^{n_1+n_2}x_{m(l)}$
⋮	⋮	⋮	⋮	⋮	⋮	
第	$n-n_k+1$	$y_{(n-n_k+1)}$	$x_{1(n-n_k+1)}$	$x_{2(n-n_k+1)}$	…	$x_{m(n-n_k+1)}$
	$n-n_k+2$	$y_{(n-n_k+2)}$	$x_{1(n-n_k+2)}$	$x_{2(n-n_k+2)}$	…	$x_{m(n-n_k+2)}$
	⋮	⋮	⋮	⋮	⋮	⋮
k	n	$y_{(n)}$	$x_{1(n)}$	$x_{2(n)}$	…	$x_{m(n)}$
类			$\overline{x}_1^{(k)}=\frac{1}{n_k}\sum_{l=n-n_k+1}^{n}x_{1(l)}$	$\overline{x}_2^{(k)}=\frac{1}{n_k}\sum_{l=n-n_k+1}^{n}x_{2(l)}$	…	$\overline{x}_m^{(k)}=\frac{1}{n_k}\sum_{l=n-n_k+1}^{n}x_{m(l)}$
	当前值	待推断	x_{10}	x_{20}	…	x_{m0}

表 2　y 及各因子数据的分类

年份	X_1 的类	X_2 的类	X_3 的类	X_4 的类	X_5 的类	Y 的类	年份	X_1 的类	X_2 的类	X_3 的类	X_4 的类	X_5 的类	Y 的类
1965	2	1	1	1	1	1	1975	2	2	2	3	2	2
1976	1	2	2	1	1	1	1985	2	2	2	2	2	2
1977	1	1	1	1	1	1	1969	1	2	2	3	2	2
1974	1	1	1	1	1	1	1964	2	2	2	3	3	2
1968	1	1	1	1	1	1	1984	1	1	1	1	1	2
1966	1	1	1	1	1	1	1980	2	2	1	2	2	2
1988	1	1	1	2	1	1	1978	2	2	2	3	2	2
1981	1	1	1	2	3	1	1989	2	3	3	2	3	2
1979	1	1	1	1	1	1	1987	2	2	3	1	2	2
1962	1	1	2	1	1	1	1990	3	3	2	2	3	2
1970	1	1	1	1	2	1	1973	2	2	2	1	1	2
1982	2	1	1	2	2	1	1972	2	2	2	3	3	2
1986	1	2	2	1	1	1	1963	2	3	2	2	1	2
1967	2	1	1	1	1	1	1971	3	3	3	3	3	3
1960	1	1	1	1	1	1	1961	3	3	3	3	3	3
1983	2	2	2	2	2	2							

$X_2^{(1)}$ 且…且 $X_m^{(1)}$ 则 $Y_{(1)}$”，“若 $X_1^{(2)}$ 且 $X_2^{(2)}$ 且…且 $X_m^{(2)}$ 则 $Y_{(2)}$”，…．由所有已知的观测数据便可得出 n 条复合多重模糊蕴涵命题，其中有许多是相同的(表 2)．我们从中选出互不相同的命题列成表(表 3)，便建立了模糊推理法则．直接按照这些推理法则便可进行模糊推理了．设各因子的当前值为 $x_{10}, x_{20}, \cdots, x_{m0}$，按式(4)，(5) 求出它们各自归属的类别．然后，对照上述模糊推理法则表，就推断出研究对象 y 应归属的类别．如果还要进一步推断 y 的具体数值，可按文[3]所提出的方法进行．

表 3　模糊推理法则

X_1 的类	X_2 的类	X_3 的类	X_4 的类	X_5 的类	Y 的类	X_1 的类	X_2 的类	X_3 的类	X_4 的类	X_5 的类	Y 的类
2	1	1	1	1	1	1	2	2	3	2	2
1	2	2	1	1	1	2	2	2	3	3	2
1	1	1	1	1	1	2	2	1	2	2	2
1	1	1	2	1	1	2	3	3	2	3	2
1	1	1	2	3	1	2	2	3	1	2	2
1	1	2	1	1	1	3	3	2	2	3	2
1	1	1	1	2	1	2	2	2	1	1	2
2	1	1	2	2	1	2	3	2	2	1	2
2	2	2	2	2	2	3	3	3	3	3	3
2	2	2	3	2	2						

2　推理实例

已知福建省泉州市气象台 1959 年到 1990 年的实测资料，推断永春县 1991 年 9 月中旬降雨量(推断对象用 y 表示)．

用适当的方法[3]得到 5 个跟 y 关系密切的因子(组合因子)，分别记作 x_1, x_2, x_3, x_4, x_5．y 及各因子的数据列入表 4．各因子与 y 的相关系数，列在表中末一行．将所有数据按 y 之值从小到大的顺序重新排列后，便形成表 5．表 5 的数据构成“有序样本”，可按文[3]的方法或按通常采用的 Fisher 最优分割法[5]对 y 作合理分类．但更简捷的是根据气象部门的要求，将降雨量分为三类(级)：第一类(偏少) 为 $y \leq 44$ mm；第二类(正常) 为 44 mm$< y \leq 156$ mm；第三类(偏多) 为 $y > 156$ mm．这正好对应着类内离差平方总和在 275 mm^2 以下的 Fisher 最优分割．对 y 的已知观测值分类结果为 $Y^{(1)} = \{y_{(1)}, y_{(2)}, \cdots, y_{(15)}\}$，$Y^{(2)} = \{y_{(16)}, y_{(17)}, \cdots, y_{(29)}\}$，$Y^{(3)} = \{y_{(30)}, y_{31)}\}$．在表 5 中标上各已知数据的类别．按式(3) 计算各因子的各类均值(作为该类的标准模式)，也填入表 5．再按式(4)，(5) 求出各因子所有已知观测值的类别，将它们与 y 的对应类别一起填入表 2．表 2 的每一行便可视为一条复合多重模糊蕴涵命题．从表 2 中选出互不相同的 19 条命题，列成表 3，这就是我们所得出的模糊推理法则*．将各因子的当前值(见表 4 倒数第 2 行) 代入式(4)，(5)，求出其类别皆为第一类，即知模糊推理的小前提是 $X_1^{(1)}$，$X_2^{(1)}$，$X_3^{(1)}$，$X_4^{(1)}$，$X_5^{(1)}$．对照表 3，可推断对应的 y 值(即 1991 年 9 月中旬永春县的降雨量)，属于第一类(偏少，即少于 44 mm)，与实际情况(27.6 mm) 相符．

* 其中 1984 年的数据为矛盾个例，可能是该年的数据不准．用别的方法亦不能判别．

表4　研究对象 y 与各因子的观测值及相关系数

年份	y	x_1	x_2	x_3	x_4	x_5	年份	y	x_1	x_2	x_3	x_4	x_5
1960	43.9	43.5	44.0	72.5	40.5	28.5	1977	1.7	41.5	29.5	57.0	46.5	56.0
1961	212.9	143.5	119.5	121.5	122.0	114.0	1978	75.3	107.0	96.0	83.5	125.5	95.5
1962	31.9	59.0	80.0	82.0	78.5	81.5	1979	23.1	58.5	74.5	61.0	67.5	62.5
1963	155.3	101.5	113.0	84.5	89.0	68.5	1980	73.3	93.5	93.0	67.0	95.5	86.5
1964	67.1	110.0	97.5	91.5	117.5	121.0	1981	23.0	70.5	76.0	58.0	82.5	113.0
1965	0.0	82.0	38.0	59.5	47.0	42.5	1982	33.2	84.5	66.5	63.0	83.0	93.5
1966	9.8	45.0	73.0	62.0	60.5	53.5	1983	54.2	87.5	92.5	94.0	94.0	85.5
1967	43.0	102.0	65.0	71.5	58.5	62.0	1984	67.4	78.0	64.5	46.5	46.5	66.0
1968	3.9	39.5	35.5	61.5	56.0	61.0	1985	61.1	113.5	87.5	87.5	86.5	97.5
1969	62.9	77.5	107.0	104.5	116.5	97.5	1986	40.7	71.0	94.5	104.5	81.0	75.0
1970	32.3	36.0	71.2	65.5	63.5	101.0	1987	105.9	101.5	83.0	114.5	79.5	94.0
1971	174.0	126.5	123.5	141.0	122.0	116.5	1988	18.6	63.0	65.0	54.0	88.5	72.5
1972	154.8	95.5	98.0	106.0	125.5	123.5	1989	80.1	102.0	116.0	114.0	109.0	112.0
1973	150.5	91.5	94.5	105.5	73.5	81.0	1990	109.9	125.0	122.0	112.5	98.5	133.0
1974	2.4	42.0	47.5	45.5	45.0	38.0	当前值	待推断	81.0	92.0	61.5	59.0	33.0
1975	55.1	100.5	93.5	102.0	112.0	99.5	相关系数	–	0.781	0.764	0.736	0.678	0.616
1976	0.0	66.5	86.5	85.0	69.5	75.0							

表5　按 y 从小到大的顺序重排已知数据

类别	年份	y	x_1	x_2	x_3	x_4	x_5	类别	年份	y	x_1	x_2	x_3	x_4	x_5
1	1965	0.0	82.0	38.0	59.5	47.0	42.5	2	1985	61.1	113.5	87.5	87.5	86.5	97.5
1	1976	0.0	66.5	86.5	85.0	69.5	75.0	2	1969	62.9	77.5	107.0	104.5	116.5	97.5
1	1977	1.7	41.5	29.5	57.0	46.5	56.0	2	1964	67.1	110.0	97.5	91.5	117.5	121.0
1	1974	2.4	42.0	47.5	45.5	45.0	38.0	2	1984	67.4	78.0	64.5	46.5	46.5	66.0
1	1968	3.9	39.5	35.5	61.5	56.0	61.02	2	1980	73.3	93.5	93.0	67.0	95.5	86.5
1	1966	9.8	45.0	73.0	62.0	60.5	53.5	2	1978	75.3	107.0	96.0	83.5	125.5	95.5
1	1988	18.6	63.0	65.0	54.0	88.5	72.5	2	1989	80.1	102.0	116.0	114.0	109.0	112.0
1	1981	23.0	70.5	76.0	58.0	82.5	113.02	2	1987	105.9	101.5	83.0	114.5	79.5	94.0
1	1979	23.1	58.5	74.5	61.0	67.5	62.5	2	1990	109.9	125.0	122.0	112.5	98.5	133.0
1	1962	31.9	59.0	80.0	82.0	78.5	81.5	2	1973	150.5	91.5	94.5	105.5	73.5	81.0
1	1970	32.3	36.0	71.5	65.5	63.5	101.0	2	1972	154.8	95.5	98.0	106.0	125.5	123.5
1	1982	33.2	84.5	66.5	63.0	83.0	93.5	2	1963	155.3	101.5	113.0	84.5	89.0	68.5
1	1986	40.7	71.0	94.5	104.5	81.0	75.0	2	均值	–	98.89	97.00	93.82	97.79	97.21
1	1967	43.0	102.0	65.0	71.5	58.5	62.0	3	1971	174.0	126.5	123.5	141.0	122.0	116.5
1	1960	43.9	43.5	44.0	72.5	40.5	28.5	3	1961	212.9	143.5	119.5	121.5	122.0	114.0
1	均值	–	60.30	63.13	66.83	64.53	67.70	3	均值	–	135.0	121.5	131.3	122.0	115.3
2	1983	54.2	87.5	92.5	94.0	94.0	85.5	待定	当前	待推断	81.0	92.0	61.5	59.0	33.0
2	1975	55.1	100.5	93.5	102.0	112.0	99.5								

若还要进一步推断(预报) y 的具体数值, 可将各因子的当前值归在一起写成五维向量 $\boldsymbol{X}_0 = (x_{10}, x_{20}, x_{30}, x_{40}, x_{50})$. 它是五维向量空间 $\mathscr{X}^5$ 中的元素, $\boldsymbol{X}_0$ 与 $Y^{(1)}$ 中的数值之间的关系也是一种模糊关系. 对于 $Y^{(1)}$ 中的任一值 $y_j\,(1\le j\le 15)$, 推断它的样本全体应视为 $\mathscr{X}^5$ 上的模

糊子集, 记作 A_j. 其从属函数规定为

$$\mu_{A_j^{(1)}}(\boldsymbol{X}) = 1 - \frac{\sum_{i=1}^{5} r_i \left| x_i - x_{i(j)} \right|}{\max\limits_{1 \le l \le 15} \left\{ \sum_{i=1}^{5} r_i \left| x_i - x_{i(l)} \right| \right\}}, \qquad \boldsymbol{X} = (x_1, x_2, \cdots, x_5) \in \mathscr{X}_5, \tag{6}$$

式中 $j = 1, 2, \cdots, 15$, r_i 为因子 x_i 与 y 的相关系数. 以各因子的当前值 $\boldsymbol{X}_0 = (x_{10}, x_{20}, x_{30}, x_{40}, x_{50})$ (具体数值见表 4 倒数第 2 行) 代入式(6), 计算得

$$\mu_{A_{11}^{(1)}}(\boldsymbol{X}_0) = \max_{1 \le j \le 15} \mu_{A_j^{(1)}}(\boldsymbol{X}_0).$$

对应于 1979 年的数据, 故推断(预报) 永春县 1991 年 9 月中旬的降雨量在 23. 1 mm 左右(实况为 27. 6 mm) .

参　考　文　献

1　Mamdani E H. Applications of fuzz algorithms for simple dynamic plant[J]. Proc. IEE., 1974, (121): 1 585～1 588

2　张文修, 梁广锡 . 模糊控制与系统[M]. 西安: 西安交通大学出版社, 1998. 55～70

3　陈治典 . 组合因子的模糊 AID 预报法[J]. 数理统计与应用概率, 1993, 8(4) : 14～22

4　汪培庄 . 模糊集合论及其应用[M]. 上海: 上海科学技术出版社, 1983. 42～52

5　Fisher W D. On grouping for maximum homogeneity[J] . J. Amer. Statist. Assoc. , 1958, 53: 789～798

A Multi-Level Discrimination and Prediction by Applying Simple Fuzzy Reasoning

Chen Zhidian

(Dept. of Manag. Info. Sci., Huaqiao Univ., 362011, Quanzhou)

Abstract　Combining fuzzy mathematics with multivariate statistccs, the author advances a simple and easy method of fuzzy reasoning by which the difficulty of computing multi-dimensional relational matrix can be avoided. The above-mentioned method will make full use of linear and non linear information of correlation factors. It is capable of treating fuzzy information free from the restriction of probability distribution of sample data. Thus the effects of multi-level discrimination and prediction can be promoted; and the trend and the definite value can be predicted. This is a method which will be still more widely applied to the actual problems.

Keywords　fuzzy reasoning, easy method, multi-level discrimination, prediction