

可拓聚类预测方法预测卷烟销售量

沈航, 邹平

(昆明理工大学 管理与经济学院, 云南 昆明 650093)

摘要: 利用可拓聚类预测方法对卷烟销售量进行了预测, 用卷烟销售量和第 1、第 3 产业产值的历史数据构造分类经典域和节域物元, 应用物元和可拓集合中的关联函数建立预测模型, 通过聚类分析得到预测结果. 利用我国云南某地区的实际数据进行分析计算, 其结果表明可拓聚类预测方法应用于卷烟销售量预测上是有效的.

关键词: 烟草销售; 可拓集合; 聚类分析; 关联函数

中图分类号: F271 **文献标识码:** A **文章编号:** 1007 - 855X(2006)03 - 0095 - 04

Extension Classified Prediction Used in Predicting Sales Volume of Tobacco

SHEN Hang, ZOU Ping

(Faculty of Management and Economics, Kunming University of Science and Technology, Kunming 650093, China)

Abstract: Extension classified prediction is used to predict sales volume of tobacco; sales volume of tobacco and the historical production value of the first industry and the third industry are used to establish classified classic field and node field element; the dependent function of material element and extension set is applied to establish prediction model. The prediction results can be obtained by means of classified analysis. Through analyzing and calculating the real data of a certain region in Yunnan, the results show that extension classified prediction is effective in predicting sales volume of tobacco.

Key words: tobacco marketing; extension set; classified analysis; dependent function

0 引言

烟草行业是我国国民经济的支柱产业之一, 是我国的利税大户, 是我国国民经济中一个不可或缺的部分. 因此研究烟草销售预测技术具有很强的现实意义.

烟草作为一种特殊的垄断性的商品, 具有和其它普通商品不同的特征. 烟草销售量受到很多外界因素的影响, 而这些影响又很难进行确定和量化的描述, 目前应用于预测的方法很多, 主要有趋势外推预测模型、指数平滑法、时间序列预测模型或组合预测模型等, 上述模型或者忽略或者简化了这些影响, 因此预测结果往往并不能反映真实的情况. 本文利用一种比较新的预测模型——可拓聚类预测模型对烟草销售量进行预测分析, 可对相关管理者决策提供帮助.

1 烟草销售量可拓聚类预测的建模机制

烟草作为一种特殊的由国家进行专卖管理的商品, 其价格的制定、生产、物资供应、分销和零售等各环节都由国家进行严格的控制. 因此其销售量的大小更多的受到外部经济环境的影响. 可拓学的物元理论可以把多个环境变量综合起来作为一个物元进行分析, 使能够结合多个环境变量进行预测成为可能. 如果把不同经济环境下卷烟销售量的变化作为一种预测结果, 就可以用聚类预测方法进行预测. 根据事物关于特征的量值来判断事物隶属于某集合的程度与可拓集合的基本思想是一致的. 而关联函数则使聚类预测方法更为精细化. 可拓聚类预测方法利用了关联函数可以取负值的特点, 使聚类分析方法能全面地分析对象

收稿日期: 2005 - 06 - 01.

第一作者简介: 沈航 (1976 ~), 男, 在读硕士研究生. 主要研究方向: 决策理论与方法. E - mail: shh3000@163.com

属于某集合的程度.同时,也有助于从变化的角度来分析变化中的事物.可拓聚类预测方法首先通过聚类分析划分集合 P 的若干子集,构造它们的经典域物元和节域物元,并确定待测物元.然后根据关联函数值确定待测样本隶属于哪个子集,从而得到聚类预测结果.

2 可拓聚类预测的物元模型

设 $I_i (i=1, 2, \dots, m)$ 是 P 的 m 个子集, $I_i \subset P, (i=1, 2, \dots, m)$, 对任何待测对象 $p \in P$, 用下面步骤判断 p 属于哪个子集 I_i , 并计算 p 隶属于任一子集 I_i 的程度.

第 1 步, 确定经典域和节域, 令

$$R_i = (I_i, C, X) = \begin{bmatrix} I_i & c_1 & X_{i1} \\ & c_2 & X_{i2} \\ & \dots & \dots \\ & c_n & X_{in} \end{bmatrix} = \begin{bmatrix} I_i & c_1 & a_{i1}, b_{i1} \\ & c_2 & a_{i2}, b_{i2} \\ & \dots & \dots \\ & c_n & a_{in}, b_{in} \end{bmatrix}, i=1, 2, \dots, m. \text{ 其中 } C_1, C_2, \dots, C_n$$

是 I_i 的 n 个不同的特征, 而 $X_{i1}, X_{i2}, \dots, X_{in}$ 分别为 I_i 关于特征 C_1, C_2, \dots, C_n 的取值范围, 即经典域, 并且有 $X_{ij} = [a_{ij}, b_{ij}] (j=1, 2, \dots, n)$. 令

$$R_p = (P, C, X) = \begin{bmatrix} P & C_1 & X_{p1} \\ & C_2 & X_{p2} \\ & \dots & \dots \\ & C_n & X_{pn} \end{bmatrix} = \begin{bmatrix} P & C_1 & a_{p1}, b_{p1} \\ & C_2 & a_{p2}, b_{p2} \\ & \dots & \dots \\ & C_n & a_{pn}, b_{pn} \end{bmatrix}$$

其中 $X_{p1}, X_{p2}, \dots, X_{pn}$ 分别是 P 关于特征 C_1, C_2, \dots, C_n 的取值范围, 即 P 的节域. 记 $X_{pi} = [a_{pi}, b_{pi}] (j=1, 2, \dots, n)$.

待测样本形成的物元为

$$R_x = (p, c, x) = \begin{bmatrix} p & c_1 & x_1 \\ & c_2 & x_2 \\ & \dots & \dots \\ & c_n & x_n \end{bmatrix}$$

式中 x_1, x_2, \dots, x_n 分别为待测样本的 n 个因子的观测值.

第 2 步, 根据距的定义, 确定关联函数值待测样本各因子与各类的关联度按下式计算

$$K_i(x_j) = \begin{cases} \frac{- (x_j, X_{ij})}{|X_{ij}|} & x_j \in X_{ij} \\ \frac{(x_j, X_{ij})}{(x_j, X_{pj}) - (x_j, X_{ij})} & x_j \notin X_{ij} \end{cases} \quad (1)$$

$$(x_j, X_{ij}) = \left| x_j - \frac{1}{2} (a_{ij} + b_{ij}) \right| - \frac{1}{2} (b_{ij} - a_{ij})$$

$$(x_j, X_{pj}) = \left| x_j - \frac{1}{2} (a_{pj} + b_{pj}) \right| - \frac{1}{2} (b_{pj} - a_{pj})$$

第 3 步, 确定权系数并计算待测样本与各类的隶属程度. 权系数的确定方法可根据实际情况的需要采用专家评价法, 层次分析法, 比重权数法等等. 这里采用比重权数法确定权系数. 所谓比重权数, 是指根据某指标在所有被评价对象上的观测值比重差异大小来确定的一种数量权数, 它用该指标的比重差异信息来衡量其重要性大小. 即对于每一个要进行判别的类来说, 待测样本每个因子的权系数由其与相对应特征的经典域最大值的比值占这一类中各因子与其相对应特征值经典域最大值的比值之和的比例确定. 即:

$$w_{ij} = \frac{x_{ij}/b_{ij}}{\sum_{j=1}^n (x_j/b_{ij})} \quad (2)$$

2) 式中, 符号 j 代表因子, $(j=1, \dots, n)$; i 代表类别, $i=1, 2, \dots, m$.

则待测样本 p 对 i 类的关联度为:

$$K_i(p) = \max_{j=1}^n K_{ij}(x_j) \tag{3}$$

第 4 步, 对待测样本 p 所属类别的判定

若 $K_i = \max K_i(p)$, $i = 1, 2, \dots, m$, 则判定样本 p 属于 i 类.

3 烟草销售量可拓聚类预测的物元模型

3.1 影响烟草销售量环境因素的确定

根据对我国云南某地区的卷烟销售状况和当地经济发展统计数据进行分析, 考虑到数据获取的权威性和数据的可操作性, 选取 1990 ~ 2003 年间云南某地区烟草销售量、国内生产总值 (GDP)、居民人均可支配收入、居民人均消费支出、第 1 产业产值、第 2 产业产值、第 3 产业产值、城乡居民储蓄余额、居民平均工资 9 个指标进行分析, 对 2004 年烟草销售量进行预测, 并与实际情况进行比较. 利用主成分分析方法和 R 型聚类分析法进行研究后可以得出结论, 对于当地卷烟销售量增长率影响较大的环境因素有第 1 产业产值增长率和第 3 产业产值增长率. 其结果和当地实际经济发展状况调查结果和烟草商品的价格需求弹性特点相符合. 因此选择它们作为影响当地卷烟销售量的环境因素.

3.2 历史数据资料处理

考察云南某地卷烟销售量、第 1 产业产值、第 3 产业产值历史数据 (数据来源: 烟草销售额来自省烟草公司, 其余数据来自当地统计局), 以表 2 中 1991 ~ 2003 年间的历史数据作为可拓聚类预测样本, 把 2004 年作为待测年. 见表 1, 2

由于卷烟属于国家管制销售的产品, 其价格的波动不如其它商品明显, 因此销售额和销售量基本上是呈正比关系的. 从销售额上即可看出销售量的变化情况. 根据经验, 一般一箱卷烟价格平均约为 9 000 元人民币.

表 1 云南某地区卷烟销售量、第 1 产业产值、第 3 产业产值历史数据

Tab 1 History data of tobacco sales, first industry values, third industry values in someplace of Yunnan Province

年份	卷烟销售 额 / 万元	第 1 产业 产值 / 万元	第 3 产业 产值 / 万元
1990	4 265.36	8 253	26 627
1991	4 986.57	9 680	30 961
1992	5 963.24	9 272	32 378
1993	5 632.14	8 893	33 233
1994	5 863.25	9 089	35 398
1995	6 443.28	9 936	37 214
1996	7 318.46	10 694	40 361
1997	8 481.59	11 802	42 581
1998	7 984.88	12 390	45 625
1999	6 269.05	12 734	48 795
2000	8 359.58	13 546	53 205
2001	7 749.83	24 247	96 338
2002	8 579.44	24 938	102 842
2003	11 167.74	25 776	110 164

表 2 云南某地区卷烟销售量、第 1 产业产值、第 3 产业产值年增长率

Tab 2 Increasing rate of tobacco sales, first industry values, third industry values in someplace of Yunnan Province

年份	卷烟销售 额	第 1 产业 产值	第 3 产业 产值
1991	1.236 95	1.172 907	1.162 767
1992	1.195 86	0.957 851	1.045 767
1993	0.944 48	0.959 124	1.026 407
1994	1.041 03	1.022 04	1.065 146
1995	1.098 93	1.093 19	1.051 302
1996	1.135 83	1.076 288	1.084 565
1997	1.158 93	1.103 61	1.055 004
1998	0.941 44	1.049 822	1.071 487
1999	0.785 12	1.027 764	1.069 479
2000	1.333 47	1.063 766	1.090 378
2001	0.927 06	1.015	1.069
2002	1.107 05	1.028 5	1.067 5
2003	1.301 69	1.033 6	1.071 2

由表 2 可以看出, 卷烟销售量增长率 R_{ratio} 的变化在 0.78 ~ 1.33 之间, 将样本数据按卷烟销售量的年增长率 R_{ratio} 分为 5 类: $I_1: 0.78 < R_{ratio} \leq 0.89$; $I_2: 0.89 < R_{ratio} \leq 1.00$; $I_3: 1.00 < R_{ratio} \leq 1.11$;

$I_4: 1.11 < R_{ratio} \leq 1.22$; $I_5: 1.22 < R_{ratio} \leq 1.33$

再计算出各类样本的两个特征 (因子): 第 1 产业产值增长率、第 2 产业产值增长率的平均值.

根据表 2 的 1, 3 产业的产值和卷烟销售量的年增长率和表 3 的各类样本的因子均值的资料构造出了

各个类别的经典域物元别为:

$$R_1 = \begin{bmatrix} I_1 & c_1 & 1.03214, 1.08588 \\ & c_2 & 1.03746, 1.11280 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} I_2 & c_1 & 0.96222, 1.05373 \\ & c_2 & 1.03029, 1.08097 \end{bmatrix}$$

$$R_3 = \begin{bmatrix} I_3 & c_1 & 1.00856, 1.08725 \\ & c_2 & 1.05256, 1.07006 \end{bmatrix}$$

$$R_4 = \begin{bmatrix} I_4 & c_1 & 0.98803, 1.16729 \\ & c_2 & 1.03388, 1.14016 \end{bmatrix}$$

$$R_5 = \begin{bmatrix} I_5 & c_1 & 0.98829, 1.16729 \\ & c_2 & 1.03388, 1.14016 \end{bmatrix}$$

节域物元为:

$$R_p = \begin{bmatrix} P & c_1 & 0.988298, 1.10454 \\ & c_2 & 1.03952, 1.10355 \end{bmatrix}$$

式中 P 代表全体类别, 而待测物元为

$$R_x = (p, C, X) \begin{bmatrix} I_p & c_1 & 1.04 \\ & c_2 & 1.072 \end{bmatrix}$$

3.3 计算待测样本 P 对各类的关联度

根据公式 (1), (2) 计算待测样本与各类的关联度和因子的权系数, 然后用下式计算待测样本 P 对各类的关联度.

$$K_i(p) = \sum_{j=1}^n w_j K_i(x_j)$$

3.4 结果分析

表 4 给出了该地区 2004 年的卷烟销售量增长率的聚类预测结果属于 I_4 类, 即年增长率在 1.11 ~ 1.22 之间. 而该年实际卷烟销售量增长率为 1.166, 说明用该方法预测卷烟销售量是有效的.

采用回报方法, 检验上述预测结果如表 5.

由结果可以看出, 总共对 14 组判别中, 判对 11 个, 判错 3 个, 预测正确率为 85%, 结果还是令人满意的.

4 结束语

烟草销售量的变化受到诸多因素的影响, 如果要想获得比较精确的预测结果, 就必须综合考虑这些因素. 可拓学的物元理论可以把多个环境变量综合起来作为一个物元进行分析, 使能够结合多个环境变量进行预测成为可能. 由于可拓聚类预测模型中关联函数的计算比较简单, 物元模型能综合多个环境因素指标进行统一分析, 因此其预测结果具有更好的精确度. 而且在构造各样本的经典域和节域物元时, 可以根据需要用不同的方法对其加以确定, 以得到更加拟合现实数据的模型, 因此该方法具有灵活性. 文中的实际计算也证明了可拓聚类预测方法的可行性与优越性.

参考文献:

[1] 蔡文. 物元模型及其应用 [M]. 北京: 科学文献技术出版社, 1994.
 [2] DAV D KW, NG, CA IW. Treating Non - Compatible Problem from Matter Element Analysis to Extenics [J]. ACM SIGCE Bulletin, 1997, 22(3): 1 - 9.
 [3] 高洁. 可拓聚类预测方法及其在邮电业务总量预测中的应用 [J]. 系统工程, 2000, 18(3): 73 - 77.
 [4] 李林志. 物元模型在股市预测中的应用 [J]. 系统工程理论与实践, 1998, 18(1): 121 - 124.

表 3 卷烟销售量增长率的每类样本数及因子均值
 Tab 3 Sample numbers and ingredient average value in per cluster of increasing rate in tobacco sales

类别	每类样本数	因子均值	
		\bar{X}_1	\bar{X}_2
I_1	1	1.027764	1.069479
I_2	3	1.007982	1.055631
I_3	3	1.04791	1.061316
I_4	4	1.077664	1.087026
I_5	2	1.048683	1.080789

表 4 待测样本与各类别的关联度及聚类预测结果
 Tab 4 The result of dependant degree and cluster prediction of samples prepared to predict and every type

$K_1(p)$	$K_2(p)$	$K_3(p)$	$K_4(p)$	$K_5(p)$	判定类型
-0.041	0.158	0.191	0.328	0.248	I_4

表 5 回报检验

Tab 5 Retest

实际组	期望组				
	1	2	3	4	5
1	1	0	0	0	0
2	0	2	1	0	0
3	0	0	3	0	0
4	0	0	1	3	0
5	0	0	0	1	1