

# 模糊聚类分析方法在 IT 市场预测分析中的应用<sup>①</sup>

## A Fuzzy Clustering Using in IT Market predict and analysis

雷小宇 张亚鸣 樊晓平 廖志芳 (中南大学信息科学与工程学院 湖南长沙 410075)  
瞿志华 (美国中佛罗里达大学电子与计算机工程系 奥兰多 FL 32816, USA)

**摘要:** 研究模糊逻辑和聚类分析在 IT 市场中的预测分析问题。模糊聚类分析是一种应用广泛的模糊数学方法,可广泛应用于各个领域。IT 市场各类信息产品的价格行情是瞬息万变的,如何对市场做出及时、准确的判断和决策成为个 IT 公司重点关心的问题。本文首先阐述了模糊聚类分析的基本思想和一般的步骤,然后采用模糊聚类分析方法对笔记本市场进行有效的分类。结果表明模糊聚类分析法对 IT 市场预测分析准确、有效,且比其他方法预测的更接近实际的市场行情。

**关键词:** 模糊逻辑 聚类分析 市场预测 模糊矩阵 等价矩阵

### 1 引言

“物以类聚,人以群分”这一大自然生物群现象揭示了聚类的本质。其实,不管是自然界还是人类社会中的事物常常以相同或者相似的属性聚集成类。美国 Lotfi Zadeh 教授的文章《Fuzzy Sets》被大家公认为现代模糊理论的开山之作。近年来软计算科学,如 ANN,遗传算法, DNA 计算,进化计算和模糊理论等迅速发展,它们已经成为了人们的研究热点,它们能有效地解决常规方法难以很好解决的问题。目前,人们普遍认为模糊理论的现实意义主要有两个方面:一是控制,二是决策和分析。

聚类分析的概念主要是来自多元统计分析,它是一种数学方法。在人们的日常生活中,我们常常要将身旁的事物或者是数值的集合按照它们的性质、用途、属性等划分成多个区间或者多个组,这样的分类过程我们就称为聚类分析。而聚类划分的方法有经典的聚类分析划分方法(硬划分)和模糊的聚类分析划分方法(软划分)。近些年来,聚类分析被广泛地应用于天气预报、地震预测、地质探勘、模式识别、生物、医学、化学、运动员心理素质分类、河流水质污染程度等方面。本文先阐述模糊聚类分析的基本思想和一般步骤,然后讨论了模糊聚类分析在 IT 市场预测分析中的应用。

### 2 模糊聚类分析方法的基本思想和具体步骤

聚类分析的基本思想是用相似性尺度来衡量事物之间的亲疏程度,并以此来实现分类。模糊聚类分析的实质则是根据研究对象本身的属性来构造模糊矩阵,在此基础上根据一定的隶属度来确定其分类关系。模糊聚类分析是基于模糊等价关系进行的,其主要步骤为:

#### (1) 确定分类对象,抽取因素数据

设分类对象的全体为:  $y = \{x_1, x_2, \dots, x_n\}$ , 而每一个对象  $x_i$  由一组数据  $x_i = (x_{i1}, x_{i2}, \dots, x_{im}) \in R_m$  来表征。

#### (2) 建立模糊相似关系和模糊等价矩阵

用实数  $r_{ij} \in [0, 1]$  来刻画对象  $x_i$  与  $x_j$  之间的相关程度,而设模糊矩阵  $R = (r_{ij})_{m \times n}$ :  $r_{ij} * r_{ij} = 1$ 。这样,  $R$  是一个模糊相似矩阵。在实际中,关键是如何确定  $r_{ij}$  的值,常用的有很多种方法,比如相关系数法、最大最小法、算术平均最小法、几何平均最小法、绝对指数法、绝对值减法、欧氏距离法、夹角余弦等等。然后,再求得模糊等价矩阵  $R^*$ , 即模糊相似矩阵  $R$  自乘得  $R \circ R = R_2$ , 再自乘  $R_2 \circ R_2 = R_4$ , 直到  $R_{2k} = R_k$  为止,  $R_k$  便是一个模糊等价关系矩阵  $R^*$ 。

① 基金项目:国家自然科学基金项目(编号:69975003);湖南省自然科学基金项目(编号:06JJ50143)

(3) 分类

分类,人们把相似而又不完全相同的现象或者事物分成一组的过程。现在的主要分类方法有三种:第一种是等价闭包法,即通过计算相似矩阵的传递闭包将对象分类;第二种是最大树方法;第三种方法是编网法。

由于模糊聚类分析是基于模糊等价关系进行的,在此我们列出几个相关的定义:

定义1 模糊相似关系:设 R 是 X 中的模糊关系,若 R 具有自反性(注1)和对称性,则称 R 为模糊相似关系。

定义2 模糊等价关系:设 R 是 X 中的模糊关系,若 R 具有自反性、对称性和传递性,则称 R 为模糊相似关系。

定义3 设  $R = (r_{ij})_{m \times n}$ , 对任意的  $\lambda \in [0, 1]$ , 称  $R_\lambda = (r_{ij}(\lambda))_{m \times n}$ , 为模糊矩阵 R 的  $\lambda$  截矩阵, 其中当  $r_{ij} \geq \lambda$  时,  $r_{ij}(\lambda) = 1$ ; 当  $r_{ij} < \lambda$  时,  $r_{ij}(\lambda) = 0$ 。

3 运用模糊聚类方法分析、预测 IT

产品市场前景

IT 市场产品销量受着许多方面的制约,能对它造成影响的因素就非常多,且这些因素并非是确切的数值,而是一些如有人认认为某款笔记本不好看,具体不好看到什么程度无法形容和描述,通过简单的数学统计分析是不能准确得出数据,所以我们运用模糊聚类方法来处理相应的数据并做出预测。

3.1 确定分类对象,抽取因素数据,建立论域 U

我们以 IBM、LENOVO、SONY、ACER、HASEE 等五个品牌笔记本在华南地区的受欢迎程度为聚类样本,对其包括价格、外观设计、综合性能、品牌美誉度等四项指标进行受欢迎程度的模糊聚类分析。

五个样本:IBM、LENOVO、SONY、ACER、HASEE; 四个考察指标:价格、外观设计、综合性能、品牌美誉度。

所以,论域  $U = \{u_1, u_2, u_3, u_4, u_5\} = \{IBM、LENOVO、SONY、ACER、HASEE\}$ ; 指标集  $V = \{\text{价格、外观设计、综合性能、品牌美誉度}\}$

根据广泛调查笔记本消费者得到的评价结果是:

IBM 的笔记本电特点是价格昂贵,外观设计比较经典,但是款式不多,机器的综合性能很好,在业界口碑非常好;

SONY 和 ACER 的价格是偏高的,它们的外观设计非常不错,而且款式也很多,综合性能比较突出,品牌美誉度 SONY 稍稍强于 ACER;

LENOVO 联想的机型价格中档,外观设计一般,不过款式比较多,品牌美誉度在中国区比较好;

HASEE 价格是五个样本中最低的一个,外观设计一般,散热性能不好,品牌美誉度一般。针对以上评价列出评价值(见表-1)。

表 1 各品牌笔记本评价结果

指标 样本	价格1	外观设计2	综合性能3	品牌美誉度4
IBM (u <sub>1</sub> )	2(u <sub>11</sub> )	4(u <sub>12</sub> )	5(u <sub>13</sub> )	5(u <sub>14</sub> )
LENOVO (u <sub>2</sub> )	4(u <sub>21</sub> )	4(u <sub>22</sub> )	4(u <sub>23</sub> )	4(u <sub>24</sub> )
SONY (u <sub>3</sub> )	3(u <sub>31</sub> )	5(u <sub>32</sub> )	4(u <sub>33</sub> )	4(u <sub>34</sub> )
ACER (u <sub>4</sub> )	3(u <sub>41</sub> )	5(u <sub>42</sub> )	4(u <sub>43</sub> )	3(u <sub>44</sub> )
HASEE (u <sub>5</sub> )	5(u <sub>51</sub> )	3(u <sub>52</sub> )	3(u <sub>53</sub> )	1(u <sub>54</sub> )

3.2 建立模糊相似关系

建立模糊相似关系,即求得相似关系矩阵 R,这里我们采用绝对值减数法来建立模糊相似关系矩阵,其公式为

$$r_{ij} = \begin{cases} 1 & \Leftrightarrow i=j \\ 1 - c \sum_{k=1}^m |x_{ik} - x_{jk}| & \Leftrightarrow i \neq j \end{cases} \quad \text{其中 } c \text{ 取适当}$$

数,使得满足  $0 \leq r_{ij} \leq 1$ 。

这里选取  $c=0.1, m=4, i, j=1, 2, 3, 4, 5$ 。

通过计算,我们可以得到:

$$\begin{aligned} r_{11} &= r_{22} = r_{33} = r_{44} = r_{55} = 1, \\ r_{12} &= 1 - c * [ |u_{11} - u_{21}| + |u_{12} - u_{22}| + |u_{13} - u_{23}| + |u_{14} - u_{24}| ] = 1 - 0.1 * [ |2 - 4| + |4 - 4| + |5 - 4| + |5 - 4| ] = 0.6 \\ r_{13} &= 0.6, r_{14} = 0.5, r_{15} = 0 \end{aligned}$$

同理,我们可以得到矩阵

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} \\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} \end{bmatrix}$$

$$= \begin{Bmatrix} 1 & 0.6 & 0.6 & 0.5 & 0 \\ 0.6 & 1 & 0.8 & 0.7 & 0.4 \\ 0.6 & 0.8 & 1 & 0.9 & 0.2 \\ 0.5 & 0.7 & 0.9 & 1 & 0.3 \\ 0 & 0.4 & 0.2 & 0.3 & 1 \end{Bmatrix}$$

然后再根据模糊相似关系矩阵建立模糊等价矩阵

$$R_2 = R * R = \begin{Bmatrix} 1 & 0.6 & 0.6 & 0.6 & 0.4 \\ 0.6 & 1 & 0.6 & 0.8 & 0.4 \\ 0.6 & 0.6 & 1 & 0.9 & 0.4 \\ 0.6 & 0.8 & 0.9 & 1 & 0.4 \\ 0.4 & 0.4 & 0.4 & 0.4 & 1 \end{Bmatrix}$$

类似,我们可以得到:

$$R_4 = \begin{Bmatrix} 1 & 0.6 & 0.6 & 0.6 & 0.4 \\ 0.6 & 1 & 0.8 & 0.8 & 0.4 \\ 0.6 & 0.8 & 1 & 0.9 & 0.4 \\ 0.6 & 0.8 & 0.9 & 1 & 0.4 \\ 0.4 & 0.4 & 0.4 & 0.4 & 1 \end{Bmatrix}$$

$$R_8 = R_4 \circ R_4 = \begin{Bmatrix} 1 & 0.6 & 0.6 & 0.6 & 0.4 \\ 0.6 & 1 & 0.8 & 0.8 & 0.4 \\ 0.6 & 0.8 & 1 & 0.9 & 0.4 \\ 0.6 & 0.8 & 0.9 & 1 & 0.4 \\ 0.4 & 0.4 & 0.4 & 0.4 & 1 \end{Bmatrix} = R_4 = R *$$

### 3.3 分类

根据模糊等价矩阵  $R^*$ , 我们选取聚类置信水平  $\lambda = 1, 0.9, 0.8, 0.6, 0.4$ , 根据定义 3 求得实际截矩阵  $R_\lambda$ 。

3.3.1 当  $\lambda = 1$  时,  $\lambda$  截矩阵为:

$$R_1 = \begin{Bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{Bmatrix}, \text{ 此时, 论域 } U \text{ 可以分为五}$$

类  $\{IBM\}, \{LENOVO\}, \{SONY\}, \{ACER\}, \{HASEE\}$ 。

3.3.2 当  $\lambda = 0.9$  时,  $\lambda$  截矩阵为:

$$R_{0.9} = \begin{Bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{Bmatrix}, \text{ 此时, 论域 } U \text{ 则可以分为}$$

四类  $\{IBM\}, \{LENOVO\}, \{SONY, ACER\}, \{HASEE\}$ 。

3.3.3 当  $\lambda = 0.8$  时,  $\lambda$  截矩阵为:

$$R_{0.8} = \begin{Bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{Bmatrix}, \text{ 此时的论域 } U \text{ 可以分为}$$

三类  $\{IBM\}, \{LENOVO, SONY, ACER\}, \{HASEE\}$ 。

3.3.4 当  $\lambda = 0.6$  时,  $\lambda$  截矩阵为:

$$R_{0.6} = \begin{Bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{Bmatrix}, \text{ 此时论域 } U \text{ 可以分为两}$$

类  $\{IBM, LENOVO, SONY, ACER\}, \{HASEE\}$ 。

3.3.5 当  $\lambda = 0.4$  时,  $\lambda$  截矩阵为:

$$R_{0.4} = \begin{Bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{Bmatrix}, \text{ 此时论域 } U \text{ 就归为一类}$$

$\{IBM, LENOVO, SONY, ACER, HASEE\}$ 。

### 3.4 结论

到目前为止, 我们通过模糊聚类分析得到聚类分析数据, 见表 2。

表 2 聚类分析结果

$\lambda$ 值	聚类数	聚类结果
1	5	$\{IBM\}, \{LENOVO\}, \{SONY\}, \{ACER\}, \{HASEE\}$
0.9	4	$\{IBM\}, \{LENOVO\}, \{SONY, ACER\}, \{HASEE\}$
0.8	3	$\{IBM\}, \{LENOVO, SONY, ACER\}, \{HASEE\}$
0.6	2	$\{IBM, LENOVO, SONY, ACER\}, \{HASEE\}$
0.4	1	$\{IBM, LENOVO, SONY, ACER, HASEE\}$

我们得到的聚类结果与我们从华南 IT 市场得到的笔记本出货量的数据基本符合。当品牌的样本数量越大, 模糊聚类分析方法的优点就越能体现。

### 4 讨论

通过模糊聚类分析我们可以对笔记本市场做出了判断, 给 IT 渠道商一个具有说服力的参考依据, 同时也给各大笔记本生产商在产品定位和设计相关方面提供

了一些很好的反馈意见。当然影响决策的因素还有很多:比如国家政策影响,当地的经济状况,出口数量的增减,季节性的变化,不同地域的消费者对产品信任和依赖程度等等。这些要比分析简单几个笔记本品牌的市场复杂的多,同时需要的计算量也比较大。

## 5 结束语

模糊聚类分析方法是一种还在不断发展的分析方法,它与其它软计算方法有着紧密的联系。目前,模糊聚类分析方法的研究热点还有:基于遗传算法动态模糊聚类方法;基于粗糙集模型的模糊聚类方法;神经网络与模糊聚类的综合研究;灰度聚类与模糊聚类的综合研究;免疫算法、进化算法和模糊聚类综合研究等等。

### 参考文献

1 Zadeh L A. Fuzzy Sets. Inf Control, 1965(8): 338 - 35.

- 2 谌红, 模糊数学在国民经济中的应用, 武汉, 华中理工大学出版社, 1993. 12: 160 - 178.
- 3 孙才志、王敬东、潘俊, 模糊聚类分析最佳聚类数的确定方法研究, 模糊系统与数学 Vol. 15, No. 1 Mar. 2001.
- 4 邓赵红、王士同, 鲁棒性的模糊聚类神经网络, 软件学报, Vol. 16, No. 8. 2005.
- 5 刘耀年、张文生、张玉霞, 一种电力系统短期负荷预测的新方法, 电力系统及其自动化学报, Vol. 13, No. 4 Aug. 2001.
- 6 郑岩、黄荣怀、战晓芬、周春光, 基于遗传算法的动态模糊聚类, 北京邮电大学学报, Vol. 28, No. 1 Feb. 2005.
- 7 张敏、于剑, 基于划分的模糊聚类算法, 软件学报, Vol. 15, No. 6. 2004.
- 8 刘宇红、刘桥、任强, 基于模糊聚类神经网络的语音识别方法, 计算机学报, Vol. 29, No. 10 Oct. 2006.