

模糊聚类分析法在广西区域经济中的应用

Analysis on Guangxi District Economy Using Fuzzy Clustering Method

覃柏英,吕跃进,王玉燕

Qin Boying, Lü Yuejin, Wang Yuyan

(广西大学数学与信息科学学院,广西南宁 530004)

(Coll. of Math. & Info. Sci., Guangxi Univ., Nanning, Guangxi, 530004, China)

摘要:运用模糊聚类分析法对广西14个市、地区的综合实力进行模糊聚类分析。结果广西分为5类地区,第1类为南宁市、柳州市和桂林市;第2类为玉林市和南宁地区;第3类为梧州市、北海市、钦州市、贵港市、百色市和河池市;第4类为贺州市和来宾市;第5类为防城港市。分析结果与广西的实际情况比较吻合。

关键词:模糊聚类分析 模糊等价矩阵 区域经济

中图分类号:O159;C934 文献标识码:A 文章编号:1002-7378(2005)01-0010-03

Abstract: The fuzzy clustering analysis is used to classify the comprehensive strength of the 14 cities and areas of Guangxi. According to the analysis, Guangxi can be classified into 5 kinds of zones. The first zone includes the cities of Nanning, Guilin and Liuzhou. The second zone includes Yulin and Nanning prefecture. The third zone includes the cities of Wuzhou, Beihai, Qinzhou, Guigang, Baise and Hechi. The cities of Hezhou and Laibin belong to the fourth zone, and Fangchenggang belongs to the fifth zone. The analysis results are accorded with the actual situation of Guangxi.

Key words: fuzzy clustering analysis, fuzzy equivalent matrices, district economy

区域经济通常是指某一特定经济区域内部的社会经济活动和社会经济关系或联系的总和^[1]。它是一种以某种经济活动或以特定的经济极点(城市、生产基地)为中心的具有宏观经济意义的地域性综合经济体系,是经济区域的实体性内容。根据不同的目的和划分标准可以划分出不同的区域经济。区域经济本质上属于社会经济的范畴,有其自身的特点。对于国民经济和部门经济来说,区域经济具有综合性、区域性和不平衡性等特点。区域经济发展过程中的这些特点要求在对区域经济进行组织与管理时,必须坚持从实际出发,深入分析各个地区经济发展的历史过程与现状,贯彻因时、因地制宜的原则,对处于不同发展阶段的地区经济采取不同的管理策略和方法。区域经济问题是当代经济发展中的一个重要课题,而市(县)域经济是区域经济的一个重要层次,

在国民经济中处于基础性地位。

广西是一个以市(县)域经济为基础的省份,改革开放后,广西的市(县)域经济取得了一定的发展。但是由于各市(县)的情况千差万别,经济发展不平衡,为了对广西区域开发和对广西区域经济发展水平有一个比较具体、清晰的认识,本文试图用模糊聚类分析方法对广西14个具有代表性的市区进行聚类分析。

1 模糊聚类分析方法

对事物按一定要求进行分类的数学方法称为聚类分析方法^[2]。由于客观事物之间并没有一个截然区别的界限,因而分类时所依据的数据指标的变化也大都具有联系性,因此导致现实的分类往往伴随着模糊性。用模糊数学方法进行聚类分析就显得更自然、更符合客观实际。模糊聚类分析是用数学方法定量地确定研究对象的亲疏关系和相似性,从而能客观的分型划类,使其具有较强的分辨率和广泛的代表性,因而更具普遍意义。模糊聚类分析方法已经广泛应用于社会科学和自然科学等领域^[3~6]。

收稿日期:2004-09-08

修回日期:2004-10-14

作者简介:覃柏英(1979-),女,广西桂平人,硕士研究生,主要从事预测与决策分析研究。

1.1 有关定义^[2]

定义 1 设矩阵 $R = (r_{ij})_{n \times n}$, 如果满足 $0 \leq r_{ij} \leq 1, 1 \leq i \leq n, 1 \leq j \leq n$, 则称 R 为模糊矩阵^[3].

定义 2 设 $R = (r_{ij})_{n \times n}$ 是模糊矩阵, 假如其满足下面的条件:

- (1) $r_{ii} = 1, 1 \leq i \leq n$ (自反性),
- (2) $r_{ij} = r_{ji}, 1 \leq i, j \leq n$ (对称性),

则称 $R = (r_{ij})_{n \times n}$ 是模糊相似矩阵.

定义 3 设 $R = (r_{ij})_{n \times n}$ 是模糊相似矩阵, 如果有 $R^2 \subseteq R$ (传递性), 则称 $R = (r_{ij})_{n \times n}$ 是模糊等价矩阵.

定义 4 给定模糊矩阵 $R = (r_{ij})_{n \times n}$, 对任意的 $\lambda \in [0, 1]$, 若有矩阵 $R_\lambda = (r_{ij}^{(\lambda)})_{n \times n}$, 其中

$$r_{ij}^{(\lambda)} = \begin{cases} 1, & r_{ij} \geq \lambda \\ 0, & r_{ij} < \lambda \end{cases}, \text{ 则称 } R_\lambda \text{ 为 } R \text{ 的 } \lambda\text{-截矩阵.}$$

1.2 主要步骤

在模糊聚类分析中, 把研究的对象称为样本, 影响样本分类的因素为属性指标, 设有 n 个样本, 每个样本有 m 个属性指标, 模糊聚类分析的一般步骤为:

步骤 1 首先对样本进行调查, 建立样本属性指标矩阵.

步骤 2 把样本的统计指标的数据标准化. 以便于分析和比较, 这步称为正化.

记被分类样本的全体为 $X = \{x_1, x_2, \dots, x_n\}$, 而每一个样本 x_i 有 m 个指标属性, 记为: $x_i = (x_{i1}, x_{i2}, \dots, x_{im}) \in R_+^m, (i = 1, 2, \dots, n)$.

标准化: (1) 效益型属性: $\bar{x}_{ij} = \frac{x_{ij} - x_{\min j}}{x_{\max j} - x_{\min j}}, (1 \leq i \leq n, 1 \leq j \leq m)$;

(2) 成本型属性: $\bar{x}_{ij} = \frac{x_{\max j} - x_{ij}}{x_{\max j} - x_{\min j}}, (1 \leq i \leq n, 1 \leq j \leq m)$,

其中, $x_{\min j}$ 是第 j 项指标的最小值; $x_{\max j}$ 是第 j 项指标的最大值.

步骤 3 建立模糊相似矩阵. 用数 $r_{ij} \in [0, 1]$ 来刻画对象 x_i, x_j 之间的相关程度, 而设模糊矩阵为 $R = (r_{ij})_{n \times n}, r_{ii} = 1, r_{ij} = r_{ji}$. 其中确定 r_{ij} 的值要根据实际情况决定, 一般可从下面几种方法来规定.

$$(1) \text{ 数量积法: } r_{ij} = \begin{cases} 1, & i = j \\ (\sum_{k=1}^n \bar{x}_{ik} \bar{x}_{jk}) / M, & i \neq j \end{cases}$$

其中 M 为一适当常数, 满足

$$M \geq \sum_{k=1}^n \bar{x}_{ik} \bar{x}_{jk}$$

$$(2) \text{ 夹角余弦法: } r_{ij} = \frac{\sum_{k=1}^n \bar{x}_{ik} \bar{x}_{jk}}{\sqrt{\sum_{k=1}^n \bar{x}_{ik}^2} \sqrt{\sum_{k=1}^n \bar{x}_{jk}^2}}$$

$$\sqrt{\sum_{k=1}^n \bar{x}_{jk}^2},$$

$$(3) \text{ 最大、最小法: } r_{ij} = \frac{\sum_{k=1}^n (\bar{x}_{ik} \wedge \bar{x}_{jk})}{\sum_{k=1}^n \bar{x}_{ik} \vee \bar{x}_{jk}}$$

$$(4) \text{ 几何平均最小法: } r_{ij} = \frac{\sum_{k=1}^m \bar{x}_{ik} \wedge \bar{x}_{jk}}{\sum_{k=1}^m \sqrt{\bar{x}_{ik} \cdot \bar{x}_{jk}}}$$

$$\sqrt{\bar{x}_{ik} \cdot \bar{x}_{jk}}$$

步骤 4 改造模糊相似矩阵为模糊等价矩阵 ($R = (r_{ij})_{n \times n}$ 必为模糊等价矩阵才可聚类). 由步骤 2 得到的模糊矩阵一般只满足自反性与对称性, 不满足传递性. 因此有必要把 $R = (r_{ij})_{n \times n}$ 改造为模糊等价矩阵. 由文献[3], 改造过程利用求模糊相似矩阵传递闭包的平方法.

步骤 5 得出模糊等价矩阵后, 选取分类水平 $\lambda, \lambda \in [0, 1]$, 计算 λ 模糊截矩阵 $R_\lambda = (r_{ij}^\lambda)_{n \times n}$, 确定分类结果. 当 R 确定后, 给定任一 $\lambda \in [0, 1]$, 若 $r_{ij} \geq \lambda$, 则样本 x_i 和 x_j 属于同一类. 随着 λ 从 1 降到 0 时, $R_\lambda = (r_{ij}^\lambda)_{n \times n}$ 所确定的分类随之由细到粗变化; 这样就可以得到较为合理的分类关系.

2 模糊聚类分析法在广西区域经济中的应用

本文就广西 2002 年 14 个市、地区的综合实力进行分类. 进行聚类分析的关键是考虑属性指标与样本的选取. 因此, 首先建立一套比较科学和完整的指标体系. 通过相关分析, 结合广西的实际, 经过比较综合, 建立以 (A_1) 国内生产总值, (A_2) 人均国内生产总值, (A_3) 财政一般预算收入, (A_4) 人均财政一般预算收入, (A_5) 城乡居民储蓄存款余额, (A_6) 人均城乡居民储蓄存款余额, (A_7) 农村居民人均纯收入为主要指标的综合测评指标体系^[7]. 运用这一指标体系, 对广西 14 个市、地区的经济实力进行测算, 区分不同发展水平的市、地区经济.

广西 14 个市、地区代表样本的因素指标如表 1 所示. 取定样本后, 对样本的属性指标进行标准化, 用几何平均最小法确定 r_{ij} , 经过传递闭包法, 取不同的 λ 值就得到不同的分类. 当取 $\lambda \in [0.7837, 0.7949]$ 时, 聚类结果共分为 5 类(表 2).

从表 2 可以看出, 第 1 类包括南宁市, 柳州市, 桂林市. 这三个城市是广西最主要的城市. 南宁市位于广西南部, 是广西壮族自治区首府, 是广西政治、

表1 广西14个市、地区代表样本的因素指标(2002年)^[8]

市(地区)	GDP (万元)	人均GDP (元)	财政一般 预算收入 (万元)	人均财政 一般预算 收入 (元)	城乡居民 储蓄存款 余额 (万元)	人均城乡 居民储蓄 存款余额 (元)	农村居 民人均 纯收入 (元)
南宁市	3560670	12024	261766	879.27	3273428	10995	2524
柳州市	2852319	8207	227611	653.10	2203031	6321	1954
桂林市	3577507	7358	201928	414.11	2487711	5102	2194
梧州市	1531624	5275	81957	281.30	980759	3366	1897
北海市	1362296	9398	71906	494.23	858211	5899	2454
防城港市	692829	8879	35696	455.83	391921	1175	2164
钦州市	1546561	4664	69728	209.12	673410	2020	2441
贵港市	1232664	2745	65873	142.44	1032312	2232	2091
玉林市	2256110	3859	108701	185.14	1810003	3083	1959
百色市	1442420	3877	95401	260.40	699725	1907	1331
贺州市	1102567	5301	36652	175.76	426646	2046	1793
河池市	1446067	3851	94321	248.53	890266	2346	1419
来宾市	1035931	4324	53479	221.95	399028	1656	1769
南宁地区	1971745	3507	133194	236.37	1226459	2177	1849

表2 聚类分析结果

类型	市、地区
第1类	南宁市 柳州市 桂林市
第2类	玉林市 南宁地区
第3类	梧州市 北海市 钦州市 贵港市 百色市 河池市
第4类	贺州市 来宾市
第5类	防城港市

注:聚类分析结果是用 Matlab5.3 编程计算。

经济、文化、科技、金融、信息中心。柳州市位于广西中部,是广西重要的工业城市和区域经济中心。桂林市位于广西东北部,是著名的风景旅游城市和历史名城。旅游业发达。从广西的地理位置考虑,这三个城市从南到北贯穿广西,最能体现广西改革开放后取得的经济成果,所以把他们聚为一类是比较合理的。第2类有玉林市和南宁地区。玉林位于广西东南部,是桂东南交通中心,工业和农业都相对发达;南宁地区由首府南宁的周边市县构成,为南宁市的发展提供充足的物质资源。因此把它们聚为一类。第3类包括梧州市、北海市、钦州市、贵港市、百色市、河池市。从梧州市→贵港市→北海市→钦州市→百色市→河池市,从这条路线可以观察到,这一类是沿着桂东→桂南→桂西环绕桂中、桂北的经济圈,这一类都是广西重要的经济城市。比如,梧州市位于广西东部,毗邻广东,是广西重要的内河港口,口岸城市和著名的轻工业城市。北海市、钦州市位于广西南端,是重要沿海和港口城市。同理分析第4类,第5类、防城港市位于中国大陆海岸线西南端,广西南部沿海是西南地区出海通道的主要出海口,有西部最大的海港。聚类分析的结果与广西的实际情况比较吻

合,所以用模糊聚类分析方法得到的结论是合理、可行的。

3 结束语

应用模糊聚类分析法进行区域经济的划分是简便而可行的。通过聚类,可以具体了解城市与城市(地区与地区)经济发展情况之间的关系,哪些城市(地区)经济发展比较相似,哪些差异较大;这有利于管理者和决策者更好地进行区域经济管理、组织、协调工作。本文只是对广西部分有代表性的城市进行聚类分析,此法可以应用到更广泛的区域经济管理中去。

参考文献:

- [1] 张敦富. 区域开发研究[M]. 北京:中国轻工业出版社,2001. 2.
- [2] 胡继才,万福钧,吴珍权,等. 应用模糊数学[M]. 武汉:武汉测绘科技大学出版社,1998.
- [3] Yang M S. A survey of fuzzy clustering math computer[J]. Modelling, 1993, (18): 1-16.
- [4] 梁 静. 经济波动的模糊聚类分析[J]. 江苏统计, 2000, 3: 21-22.
- [5] 何 清. 模糊聚类分析理论与应用研究进展[J]. 模糊系统与数学, 1998, 12(2): 89-94.
- [6] 曾文艺,张 颜,宋雯彦. 研究生招生中的模糊聚类分析[J]. 北京师范大学学报(自然科学版), 2001, 37(4): 436-439.
- [7] 陈泽军. 广西县域经济发展问题研究[D]. [硕士论文]. 南宁:广西大学, 2003.
- [8] 广西统计局. 广西统计年鉴-2003[M]. 北京:中国统计出版社, 2003.

(责任编辑:黎贞崇)