

# 基于聚类分析的历史文化村镇 动态监测系统数据库研究

武艳文

(1. 西安建筑科技大学信息网络中心, 陕西 西安 710055)

**摘 要:**在历史文化村镇动态监测系统的空间数据库中同类不同属性的信息聚类显得更加重要,一种确实可行的方法可以让规划的专业人员把注意力放在某一个特定的类上以作进一步的分析. 以陕西省党家村为例,对历史文化村镇空间数据库进行聚类分析,得出数据库中分布的一些深层的信息以及隐藏在数据中的聚类规则,并且可以与其他数据挖掘方法一起使用,从而可以为以后历史文化村镇保护规划的深入研究做铺垫.

**关键词:**聚类分析;历史文化村镇;动态监测系统;数据库

**中图分类号:**TP392

**文献标志码:**A

**文章编号:**1006-7930(2012)05-0756-05

历史文化村镇是不可再生的历史文化资源,是社会历史发展的见证和写照,是现代社会生活的载体. 历史文化村镇作为历史地区的类型之一,日渐受到关注. 建设和管理工作具有重大意义. 长远来看,保护历史文化村镇是延续历史文脉、实现社会稳定和可持续发展的需要.

传统保护规划是将历史文化村镇具象的生活片段抽象地加以记录,这是静态规划特征的蓝图式记录;基于 GIS 构建的历史文化村镇动态监测系统是通过对历史文化村镇进行数据库构建,借助网络技术以及地理信息系统等建立一套适应历史文化村镇保护规划数据库的技术框架,以提升保护规划的科学性和信息化程度,实现对历史文化遗产现状调查资料的输入、存储和管理,利用现状数据库绘制保护与发展规划所需的现状分析图,实现图形与属性数据的双向查询、检索、分析,为综合保护规划提供真实、完整、永续存在的综合数据库<sup>[1]</sup>.

而本文则采用计算机系统支持的方法进行空间地理数据管理,由计算机程序模拟常规的或专有地理分析方法,作用于空间数据,对信息进行聚类分析,适时提供多种空间的和动态的地理信息用以历史文化村镇保护规划.

## 1 聚类理论与方法

由于空间聚类能够发现隐藏在空间数据背后的空间分布规律,因此,将空间聚类应用于历史文化村镇动态监测以发现历史文化村镇分布的空间规律. 划分聚类能够将研究的所有样本进行聚类,使之归入某一级别. 级别内部能够表现出很好的连续性和整体性. 在定级时,对定级单元根据各因素的指标值,采用多因素综合法即简单加权求和法得到单元总分值. 采用单元格总分值进行级别划分的思路,使得一些指标的的决定性作用不能充分体现,空间聚类是在各指标的基础上进行聚类,考虑因素更为完善、合理.

### 1.1 划分聚类算法

K-Means 算法是最典型的划分聚类算法. 该算法基于簇中对象的平均值. 基本处理流程:从  $n$  个数据对象任意选择  $k$  个对象作为初始聚类中心,从而将全部对象划分为  $k$  个区域;并根据最小距离重新对相应对象进行划分;然后再计算每个所获新聚类的聚类中心(该聚类中所有对象的均值);这个过程不断重复,直到标准测度函数平方误差准则函数收敛为止. 一般都采用均方差作为测度函数.  $k$  个聚类具有

收稿日期:2012-06-07 修改稿日期:2012-10-10

基金项目:“十一五”国家科技支撑计划项目(2008BAJ08B02)

作者简介:武艳文(1980-),男,内蒙赤峰人,硕士,工程师,主要从事计算机系统设计、应用系统研究工作.

以下特点:各聚类本身尽可能的紧凑,而各聚类之间尽可能的分开<sup>[2]</sup>.

其数学描述为:

聚类数据集合  $X = \{x_1, x_2, \dots, x_n\}$ , 数据空间  $R_q, c_j (j = 1, 2, \dots, k)$  表示聚类的  $k$  个类别,  $c_i (c_1, c_2, \dots, c_k)$  表示初始的聚类中心, 两个对象之间的欧式距离

$$d(x_i, x_j) = \sqrt{(x_i - x_j)^T (x_i - x_j)} \quad (1)$$

聚类中心为

$$c_j = \frac{1}{n_j} \sum_{x \in c_j} x \quad (2)$$

K-means 算法的核心思想是通过迭代把数据对象划分到不同的簇中, 以求目标函数最小化, 从而使生成的簇尽可能的紧凑和独立.

$$E = \sum_{i=1}^k \sum_{j=1}^{n_j} \text{dist}(x_j, c_i) \quad (3)$$

## 1.2 基于 GIS 的 K-mean 聚类

基于 GIS 的 K-mean 聚类首先通过 GIS 组件读取空间数据, 将空间实体的坐标信息读入数据库中, 与属性数据一块存储. 然后用 K 均值空间聚类算法进行聚类. 聚类系统的总体框图如下所示:

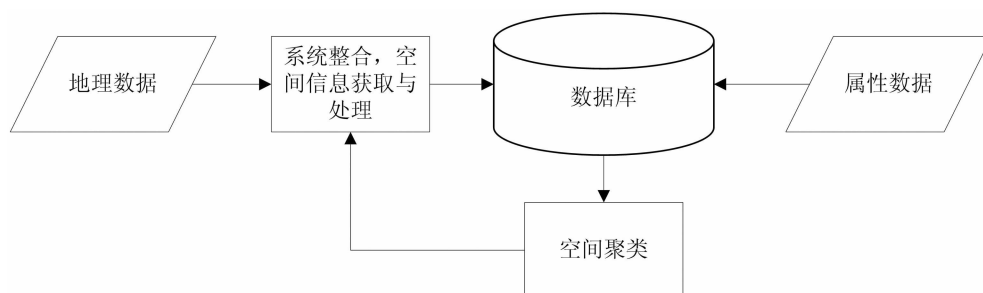


图1 空间聚类系统总体设计图

Fig.1 Diagram of spatial cluster system

## 2 应用实例

### 2.1 研究区界定

选取位于陕西韩城市区东北 9 公里, 距黄河仅 3 公里的党家村为研究区, 该研究区是国内迄今为止保存最好的明清建筑村寨, 是中国北方典型的传统民居村落. 国内外专家称赞为“民居瑰宝”、“东方人类居住村寨的活化石”, 现为国保单位, 具有典型性, 为研究方法的检验和分析提供了条件.

### 2.2 定级指标体系确定

评价指标体系框架根据中国历史文化名镇名村评价理论研究确定<sup>[3]</sup>, 分为目标层  $A_n$ , 综合评价层  $B_n$  和要素层  $C_n$ . 在确定目标层 ( $A_1$ ) 为中国历史文化名镇名村保护评价体系以后, 综合评价层 ( $B_1 \sim B_3$ ) 和要

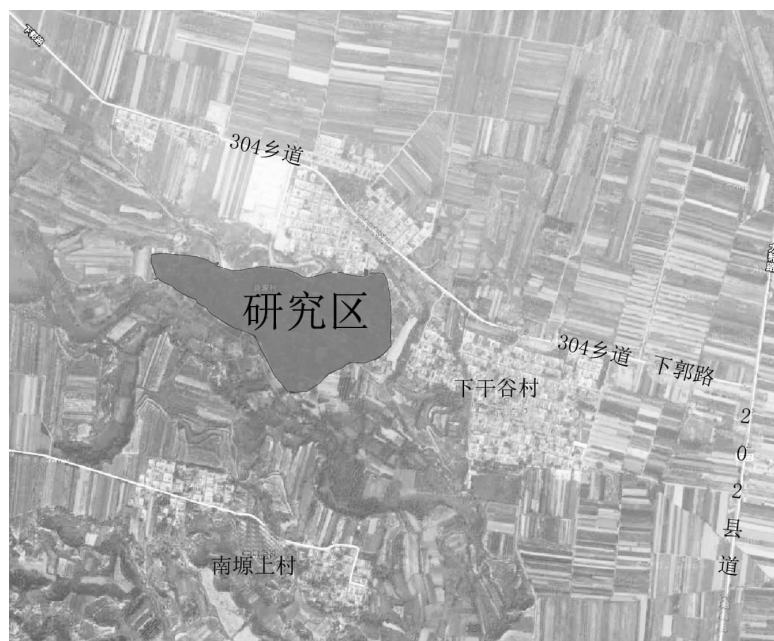


图2 基于空间聚类的历史文化村镇定级研究区示意

Fig.2 Diagram of Rating-in-research areas of historical towns by spatial cluster method

素层( $C_1 \sim C_9$ ). 在此评价框架体系的基础上,进一步采用特尔菲法和层次分析法进行权重计算,最后得到具体指标体系如下表所示:

2.3 定级单元划分

定级单元划分采用网格法,设定网格大小为 1 ha,将研究区域划分为 22 个单元格. 研究区范围较小,网格划分区域大小能够满足实验的需要.

2.4 基于空间聚类分析的历史文化村镇多因素定级

获取网格的中心点的坐标,代表该网格单元的空间位置,再对表格 1 中的各变量进行标准化,并赋相应权重(见表 3).

表 1 历史文化村镇定级指标体系<sup>[4]</sup>

Tab. 1 Rating index system of historical towns

A 层因子	B 层因子	C 层因子	权值
中国历史文化名镇名村保护评价体系 $A_1$ 1.000	物质文化遗产 $B_1$ 0.527 2	环境风貌 $C_1$ 0.163 4	0.086 1
		街巷空间 $C_2$ 0.297 0	0.156 6
		古迹建筑 $C_3$ 0.539 6	0.284 5
	非物质文化遗产 $B_2$ 0.296 5	历史影响 $C_4$ 0.272 7	0.080 9
		文化民俗 $C_5$ 0.545 4	0.161 7
		生活延续 $C_6$ 0.181 9	0.059 3
	保护措施 $B_3$ 0.176 3	保护规划 $C_7$ 0.571 4	0.100 7
		保护修复 $C_8$ 0.142 9	0.025 1
		保障机制 $C_9$ 0.285 7	0.050 4

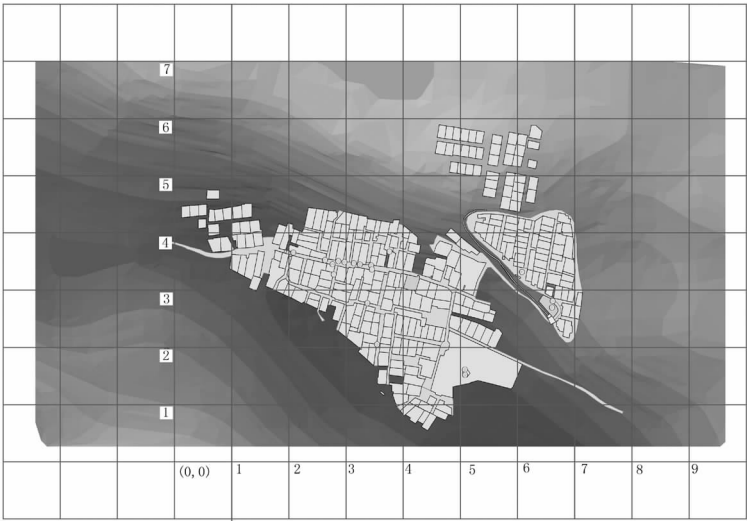


图 3 党家村定级单元划分

Fig. 3 Rating of Dang-jia Village

表 2 网格中心属性

Tab. 2 The property of grid center

网格中心	属性 $C_1$	属性 $C_2$	属性 $C_3$	属性 $C_4$	属性 $C_5$	属性 $C_6$	属性 $C_7$	属性 $C_8$	属性 $C_9$	横坐标 $X$	纵坐标 $Y$
1	0.086 1	0	0	0.080 9	0.161 7	0.059 3	0	0	0	0.5	4.5
2	0.086 1	0	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	0.5	3.5
3	0.086 1	0	0	0.080 9	0.161 7	0.059 3	0	0	0	1.5	4.5
4	0	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	1.5	3.5
5	0.086 1	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	2.5	4.50
6	0	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	2.5	3.5
7	0.086 1	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	2.5	2.5
8	0.086 1	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	3.5	4.5
9	0	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	3.5	3.5
10	0	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	3.5	2.5
11	0.086 1	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	3.5	1.5
12	0.086 1	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	4.5	3.5
13	0	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	4.5	2.5
14	0	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	4.5	1.5
15	0.086 1	0	0	0.080 9	0.161 7	0.059 3	0	0	0	4.5	0.5

网格中心	属性 C <sub>1</sub>	属性 C <sub>2</sub>	属性 C <sub>3</sub>	属性 C <sub>4</sub>	属性 C <sub>5</sub>	属性 C <sub>6</sub>	属性 C <sub>7</sub>	属性 C <sub>8</sub>	属性 C <sub>9</sub>	横坐标 X	纵坐标 Y
16	0.086 1	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	5.5	4.5
17	0.086 1	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	5.5	3.5
18	0.086 1	0	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	5.5	2.5
19	0.086 1	0	0	0.080 9	0	0	0.100 7	0.025 1	0.050 4	5.5	1.5
20	0.086 1	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	6.5	4.5
21	0	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	6.5	3.5
22	0.086 1	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	6.5	2.5

表 3 用于空间聚类的变量及权重  
Tab. 3 Variables and weighing for spatial cluster

属性变量(权重 0.5)									空间变量(权重 0.5)	
环境风貌	街巷空间	古迹建筑	历史影响	文化民俗	生活延续	保护规划	保护修复	保障机制	横坐标 X	纵坐标 Y
0.086 1	0.156 6	0.284 5	0.080 9	0.161 7	0.059 3	0.100 7	0.025 1	0.050 4	1	1

则距离公式为：

$$d_{xy} = \frac{1}{2} \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$
$$+ \frac{1}{2} (B_1 \sqrt{0.163\ 4(C_{1i} - C_{1j})^2 + 0.297\ 0(C_{2i} - C_{2j})^2 + 0.539\ 6(C_{3i} - C_{3j})^2}$$
$$+ B_2 \sqrt{0.272\ 7(C_{4i} - C_{4j})^2 + 0.545\ 4(C_{5i} - C_{5j})^2 + 0.181\ 9(C_{6i} - C_{6j})^2}$$
$$+ B_3 \sqrt{0.571\ 4(C_{7i} - C_{7j})^2 + 0.142\ 9(C_{8i} - C_{8j})^2 + 0.285\ 7(C_{9i} - C_{9j})^2})$$

(4)

将每个网格中心看做一个类,类间距离选最短距离,如公式(4),只要两类之间的距离小于阈值,就将两类合并成一类。

其中  $d_{xy}$  表示  $\omega_i$  类中的样品  $x$  与  $\omega_j$  类中的样品  $y$  之间的距离.若  $\omega_i$  类是由  $\omega_m, \omega_n$  两类合并而成,则

$$\omega_i = \omega_m \cup \omega_n, \text{ if } \min\{d_{uv}\} < T$$
$$u \in \omega_m, v \in \omega_n$$

(5)

$T$  为判断两个类是否合并的距离阈值.所有的类间距离遍历完后聚类完毕。

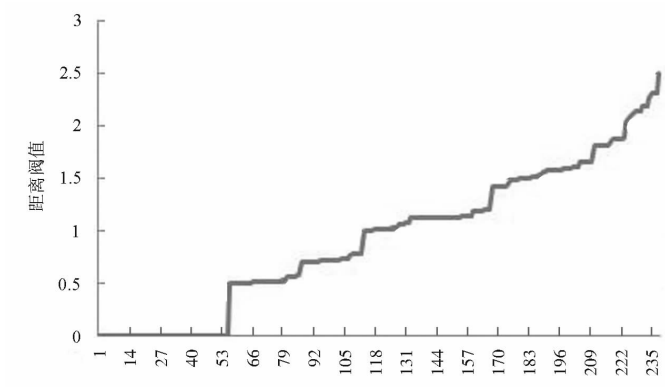


图 4 基于空间聚类的多因素综合评价的距离阈值分布图

Fig. 4 Diagram of distance threshold distribution  
with different factors and evaluating progress

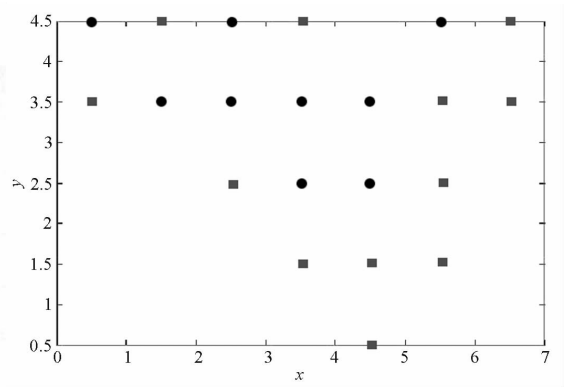


图 5 基于加权距离的聚类结果图

Fig. 5 Diagram of weighing distance based clusters

图 4 是用表 2 中的 GIS 数据进行网格化之后,计算出的网格中心点之间的距离分布<sup>[5]</sup>,很明显距离 0.5 是一个分界,在本实验中取  $T=0.5$ 。

按照公式(4)对网格中心进行聚类,最终形成两类,如图 5 所示,圆形区域表示根据中国历史文化名镇名村评价体系,这些区域应该加以重点保护,是专业人员应该特别研究关注的部分.方形区域则进行

一般保护,给专业人员提供普通的研究信息.

### 3 结 语

在本文中通过对历史文化村镇动态监测数据库中以党家村为例的属性数据进行空间聚类,通过对历史文化村镇进行动态监测分析,发现历史文化村镇分布的空间规律.首先研究了中国历史文化名镇名村评价理论,形成一个历史文化村镇定级指标体系,可以对村镇进行定级.将研究区域进行网格化,将所有样本网格进行空间聚类,最终形成两类,可以作为是否保护该村镇的参考,考虑因素更为完善、合理,从而可以为以后历史文化村镇保护规划的深入研究做铺垫.

### 参考文献 References

- [1] 于 洋,雷振东,杨 铭.历史文化村镇保护规划数据库构想:以陕西党家村为例[J].理想空间,2010(10):16-19.  
YU Yang, LEI Zhen-dong, YANG Ming. Conservation planning of historic towns and villages database conception-take the Shaanxi DangJiacun village as the example [J]. Ideal Space, 2010(10):16-19.
- [2] 杨淑莹.图像模式识别:VC++ 技术实现[M].北京:清华大学出版社,2005.  
YANG Shu-ying, Image pattern recognition: VC++ technology to achieve [M]. Beijing: Tsinghua University Press, 2005.
- [3] 赵 勇.中国历史文化名村名镇保护理论与方法[M].北京:中国建筑工业出版社,2008.  
ZHAO Yong. Chinese famous historical and cultural towns and villages protection theory and method[M]. Beijing: China Building Industry Press, 2008.
- [4] 周铁军,黄一滔,王雪松.西南地区历史文化村镇保护评价体系研究[J].城市规划学刊,2011(6):109-116.  
ZHOU Tie-jun, HUANG Yi-tao, WANG Xue-song. An Analysis of the Conservation Evaluation System of Historic Towns and Villages in Southwest China [J]. Urban Planning Forum, 2011(6):109-116.
- [5] 窦 智.城市功能区划分空间聚类算法研究[D].成都:四川师范大学,2010.  
DOU Zhi. City function area division of spatial clustering algorithm [D]. Chengdu: Sichuan Normal University, 2010.

## Dynamic monitoring database research of historical towns based on cluster analysis

WU Yan-wen

(Information & Network Center, Xi'an Univ. of Arch. & Tech., Xi'an 710055, China)

**Abstract:** In the dynamic monitoring system spatial database of historical towns, the information clusters with same kind but different properties, are especially important. A feasible and practical way is to make professional planners pay attention to a certain information cluster to be engaged in further analysis. This paper, taking Dangjiacun village in Shaanxi Province as an example, analyzes the spatial database with cluster method to obtain some deep-rooted information and cluster rules behind the database, in order to make preparation for the research that follows.

**Key words:** cluster analysis; historical towns; dynamic monitoring system; database