

文章编号: 1004-4574(2007) 04- 0049- 06

风暴潮灾情等级识别的模糊聚类 分析方法研究

孙 峥, 庄 丽, 冯启民

(中国海洋大学 环境科学与工程学院, 山东 青岛 266100)

摘要: 鉴于风暴潮灾情等级划分和识别过程中涉及的因素较多, 且具有较大的模糊性等特征, 描述了基于模糊等价关系的模糊聚类的概念, 及其在灾情等级科学划分和识别中的应用步骤。利用 1976-2002年期间对湛江地区造成灾害的多场台风风暴潮的灾情统计资料进行了实证分析, 结果表明该方法是实用的、有效的。

关键词: 风暴潮; 灾情等级; 模糊聚类; 模糊识别

中图分类号: X43

文献标识码: A

Fuzzy cluster analysis- based storm surge disaster gradation

SUN Zheng ZHUANG Li FENG Qimin

(College of Environmental Science and Engineering, Ocean University of China Qingdao 266100 China)

Abstract In view of that there are many fuzzy factors in the disaster gradation and identification of storm surge, this paper explicates the concept of fuzzy clustering analysis based on fuzzy equivalence relation and the steps of its application to disaster gradation and identification. An actual example analysis using disaster statistics of storm surges having struck Zhanjiang area from 1976 to 2002 demonstrates the feasibility and effectiveness of this method.

Keywords storm surge; disaster grade; fuzzy cluster; fuzzy identification

风暴潮是一种由热带风暴(例如台风、飓风)带来的自然灾害, 可以导致近海及沿岸浅水域水位猛烈增长。当风暴潮与天文潮迭加后的水位超过沿岸水位警戒线时, 造成海水外溢, 泛滥成灾, 给工业、农业、海业、盐业、交通运输、港湾建筑和人民生命财产带来巨大损失。风暴潮灾害是发生频率最高, 造成的经济损失最严重, 危害最大的海洋自然灾害^[1]。在我国, 几乎一年四季均有风暴潮灾发生, 并遍及整个中国沿海, 其影响时间之长, 地域之广, 危害之重均为西北太平洋沿岸国家之首^[2]。如果在灾情发生后, 我们能够对灾情进行快速评估, 确定灾情等级, 就能为救灾和减灾工作提供有力的决策支持, 尽可能降低各种损失。因此, 风暴潮灾情的等级划分与识别方法的研究工作重要而急迫。

从根本上来说, 风暴潮灾情等级评估实际是一个模式识别问题, 而灾情等级的科学划分是进行灾害等级识别的前提, 现有的风暴潮灾情等级划分方法多是采用“硬划分”或定性描述的方法, 带有一定的主观经验性^[3-5]。如在文献[3]中, 作者给出了表1的分类标准。在文献[4]中, 作者给出的分级标准如表2所示。

表 1 风 暴 潮 灾 害 等 级
Table 1 Disaster gradation of storm surge

灾害等级	特大潮灾	严重潮灾	较大潮灾	轻度潮灾
参考灾情	死亡千人以上或经济损失上亿元	死亡数百人或经济损失 0.2~1 亿元左右	死亡数十人以上或经济损失千万元左右	无死亡或死亡少量或经济损失数百万元以下
超警戒水位参考值	> 2 m	> 1 m	> 0.5 m	超过或接近

表 2 灾 度 等 级
Table 2 Gradation of disaster degree

级别	名称	死亡人数 /人	淹没田地 /667m ²	倒塌房屋 /间	经济损失 /亿元
1	轻量潮灾	100以下	10以下	0.1以下	0.5以下
2	一般潮灾	101~500	10~50	0.1~1	0.5~1
3	较大潮灾	501~1000	51~100	1~2	1~10
...					

从两个表格可以看出,文献[3]选取死亡人数,经济损失和超警戒水位 3 个因素来衡量灾情的大小,文献[4]有所不同,作者给出了 4 个特征因素并用灾度概念来表示灾情的大小。但是,无论他们的选取的因素和分级级数如何变化,上述分类方法都忽视了两点:其一,灾情级别差异的中介过渡性。我们知道,在现实世界中,一组事物根据其亲疏程度和相似性是否形成一个类群,或一个事物是否属于某一个类别,其界限往往是不分明的,具有很大的模糊性。类别的差异存在中介过渡性。以上分类方法恰恰忽视了这一点。例如对于灾情级别,从死亡人数标准来看,500人和 501人仅差一人却截然划分到不同的级别,显然不符合人们的认知习惯。其二,多因素的相互影响。由于多因素间关系并无规律可循(例如呈现正比例关系),就造成利用不同的灾情指标得到不同的灾情等级的情况,致使有些灾情等级划分陷入两难境地,这些矛盾都是上述方法无法克服的。其他相关文献分类方法类似,在此不一一列举。

模糊聚类理论按照聚类样本集对于全体类别加权广义欧式权距离平方和最小的原则对样本的指标特征值进行聚类,体现了“物以类聚”的自然分类原则。本文从风暴潮灾情历史数据出发,尝试构建了基于模糊等价关系的风暴潮灾情等级模糊聚类分析识别模型,即通过模糊聚类,将样本划归不同灾害等级,科学合理的建立起模糊模式,在此基础上利用模糊优先比相似选择法对待识别样本进行灾情等级识别。本文最后结合实例说明了其应用过程。

1 模糊聚类分析的概念和分析步骤

1.1 基于模糊等价关系的模糊聚类分析

聚类分析是多元统计分析中的一个分支,它是一种硬划分,把每个待辨识的对象严格地划分到某个类中,具有非此即彼的性质。由于现实的分类往往伴随着模糊性,所以用模糊理论进行聚类分析显得更符合客观实际。

定义 1: 论域 U, V 的直积 $U \times V$ 的一个模糊子集 $R \in F(U \times V)$ 称为 U 与 V 之间的一个二元模糊关系,记为 uRv 。

定义 2 设 R 是论域 U 上的一个模糊关系,对于任意 $u, v \in U$, 如果 R 满足:

- (1) 自反性 $R(u, u) = 1$;
- (2) 对称性 $R(v, u) = R(u, v)$ 。

则称 R 是 U 上的一个模糊相似关系。其中 $R(u, v)$ 表示 U 的元素 u 与 v 的相关程度。

当 $U = (u_1, u_2, \dots, u_n)$ 时,模糊相似关系 R 可以用模糊矩阵 $R = (r_{ij})_{n \times n}$ 表示。在这种情况下, R 为主对角线元素均为 1 的对称模糊矩阵。

定义 3 设 R 是论域 U 上的一个模糊相似关系, 对于任意 $u, v \in U$, 如果 R 满足传递性, 即 $R^2 \leq R$, 则称 R 是 U 上的一个模糊等价关系。

定义 4 包含模糊矩阵的最小的模糊传递矩阵称为 R 的传递闭包, 记为 $t(R)$ 。

定理 1 若 R 是论域 U 上的一个模糊相似关系, 则 $t(R)$ 是 U 上的一个模糊等价关系。

定理 2 对于任意 n 行 n 列的模糊相似矩阵 R , 必定存在自然数 $k \leq n$, 使得 $t(R) = R^k$, 并且对于一切大于 k 的自然数总有 $R^l = R^k$ 。

由定理 1 和定理 2 知, 通过 R 的多次合成可以求得 R 的模糊等价关系 $t(R)$ 。

1.2 模糊优先比相似识别

采用模糊优先比相似选择法^[6]进行识别, 即以成对样本与一个待识别固定样本相比较, 以确定成对样本中哪一个样本与固定样本更相似, 从而得到与固定样本相似程度较大的样本, 方法简介如下:

设给定样本集合 $X, X = \{x_1, x_2, x_3, \dots, x_n\}$, x_k 为给定固定样本, 任意 $x_i, x_j \in X (i, j = 1, 2, 3, \dots, n)$ (为方便说明, 假定每个样本包含一个特征因素, 其值亦用 x_i 表示)。

(1) 计算海明距离 $D_i = |x_i - x_k|, D_j = |x_j - x_k|$

(2) 计算相似优先比 $v_{ij} \in [0, 1]$, 构建模糊优先比矩阵 $V = (v_{ij})_{n \times n}$, 其中 $v_{ij} = \frac{D_j}{(D_i + D_j)}$, 且规定 $v_{ii} = 1$ 。

(3) 利用不同置信水平的 $\lambda \in [0, 1]$ 截割 V , 得到截矩阵 $V_\lambda = (v_{ij}^\lambda)$, 其中 $v_{ij}^\lambda = \begin{cases} 1 & v_{ij} \geq \lambda \\ 0 & v_{ij} < \lambda \end{cases}$ 。 λ 从 1 逐渐

降为 0 若首次出现 V_λ , 它的第 i 行元素全部为 1, 则第 i 行元素对应样本 x_i 与 x_k 最相似, 其序号记为 l_i 。划去 i 行 i 列, 重复步骤 (3), 可以求出其他样本的序号。按序号从小到大排序, 序号越小, 样本与固定样本越接近。当样本包含多个特征因素时, 可首先取各样本一个因素按上述方法求其序号, 然后将每个样本的诸因素序号求和, 最后序号和排序, 序号和越小, 该样本与固定样本越接近。

1.3 聚类分析识别的基本步骤

(1) 确定分类对象, 抽取因素数据

设分类对象的全体为 $U = \{u_1, u_2, \dots, u_n\}$, 而每一个对象 u_i 有一组特征数据:

$u_i = (u_{i1}, u_{i2}, \dots, u_{im}) \in (R^+)^m, i = 1, 2, \dots, n$ 来表征。

(2) 建立模糊相似关系

用数 $r_{ij} \in [0, 1]$ 来刻画对象 u_i, u_j 之间的相似程度, 组成模糊关系矩阵 $R = (r_{ij})_{n \times n}$ 。 r_{ij} 有多种构建方法, 可以根据需要选用。

(3) 改造模糊相似关系为模糊等价关系

利用传递闭包法将模糊相似关系 R 改造成 $t(R)$, 即将 R 改造成了模糊等价关系。

设 $t(R) = (r'_{ij})_{n \times n}$ 。

(4) 模糊聚类

对模糊等价关系进行聚类处理, 给定不同的置信水平的 $\alpha \in [0, 1]$, 求矩阵 $t(R)_\alpha$, 得到普通分类关系。聚类原则为: u_i 与 u_j 在 α 水平上属于同类, 当 $r'_{ij} \geq \alpha$ 时, u_i 与 u_j 归为一类^[6]。

(5) 模糊识别

在模糊聚类的基础上, 利用模糊优先比相似选择法对待识别样本进行灾情等级的识别。

2 风暴潮灾情等级识别模糊聚类方法的应用实例

2.1 选取样本, 抽取因素数据

本文根据 1976-2002 年以来对湛江地区造成灾害的多场台风风暴潮的灾情统计资料^[7], 构成论域 U , 选取死亡人数、受灾面积、直接经济损失 3 项指标作为特征因素, 对风暴潮灾情按其严重程度进行等级划分。划分后每一个等级成为一个模式类, 每一模式类都包含了一些灾度相同的风暴潮灾害个例, 从而可以将分属于各模式类的风暴潮灾害个例看成是一个模式样本集。由于湛江地区经济发展较快、物价波动较大, 不同年份之间的经济损失不具有可比性, 以当年风暴潮灾害直接经济损失的数据难以准确反映灾害损失的增长趋势, 因此我们对原始数据进行了可比性处理, 全部统一到 1975 年的物价水平。同时为了消除不同量纲

- 当 $\alpha = 0.991$ 时, 分类为 {2, 7}, {4, 5, 9, 13, 14}, {1}, {3}, {6}, {8}, {10}, {11}, {12}, {15}
- 当 $\alpha = 0.988$ 时, 分类为 {2, 4, 5, 7, 9, 13, 14}, {1}, {3}, {6}, {8}, {10}, {11}, {12}, {15}
- 当 $\alpha = 0.986$ 时, 分类为 {2, 4, 5, 6, 7, 9, 13, 14}, {1}, {3}, {8}, {10}, {11}, {12}, {15}
- 当 $\alpha = 0.977$ 时, 分类为 {2, 4, 5, 6, 7, 9, 13, 14, 15}, {1}, {3}, {8}, {10}, {11}, {12}
- 当 $\alpha = 0.950$ 时, 分类为 {2, 4, 5, 6, 7, 8, 9, 11, 13, 14, 15}, {1}, {3}, {10}, {12}
- 当 $\alpha = 0.939$ 时, 分为 4 类, 即 {2, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15}, {1}, {3}, {12}
- 当 $\alpha = 0.891$ 时, 分为 3 类, 即 {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15}, {1}, {12}
- 当 $\alpha = 0.879$ 时, 分为 2 类, 即 {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15}, {12}
- 当 $\alpha = 0.318$ 时, 分为 1 类, 即 {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15}

为确定样本的一个最优分类, 应合理选定 α 值。选取 α 值的方法有多种, 如专家经验法和 F- 统计量法。本文结合专家经验和当地实际情况, 选取 $\alpha = 0.891$ 。此时, 样本被分为 3 类, 即 {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15}, {1}, {12}, 我们分别定义其对应灾害等级为小灾、中灾和大灾。我们看到, 经过这样分类, 小灾次数 13 次, 占样本的 86.7%, 而中灾和大灾分别为 6.6%, 即小灾频繁, 间或有较大灾情出现, 分类结果符合该地区实际灾情, 说明 α 取值是合理的。

2.5 模糊识别

经过模糊聚类, 我们将样本分成 3 类, 作为模糊识别的 3 个识别模式。选用 0220 号风暴潮 (归一化值见表 4) 为待识别对象, 选用灾害等级为中灾的 7619 号风暴潮, 灾害等级为小灾的 8303 号风暴潮和灾害等级为大灾的 9615 号风暴潮为识别模式, 采用模糊优先比相似选

表 4 0220 号风暴潮归一化值

Table 4 Normalized data of storm surge No. 0220

编号	死亡人数	受灾面积	直接经济损失
0220	0	0.079	0.02

择法对 0220 号风暴潮进行灾情等级识别。按照第 1.2 节所述的步骤进行计算, 0220 号风暴潮和 7619 号风暴潮, 8303 号风暴潮及 9615 号风暴潮灾情相似优先比序号和排序值分别为 6.3 和 9.3 取最小值 3.3 即 0220 号风暴潮与 8303 号风暴潮灾情最相似, 灾情等级判定为小灾, 与实际灾情相符合。当然, 通过这一方法可以对任意风暴潮灾害事件的灾情等级进行识别。

3 讨论与结论

在模糊聚类分析中, 对于不同的置信水平的 $\alpha \in [0, 1]$, 可以得到不同的分类, 从而形成一个动态聚类过程, 对于全面了解样本的分类情况是比较形象和直观的。并且, 模糊聚类与“硬化分”相比, 充分考虑了等级差异的中介过渡性, 较好避免了主观性, 更加符合人们的认知习惯。同时, 在模糊聚类基础上采用模糊优先比相似选择法进行灾害等级的识别简单易行。因此, 利用该方法对风暴潮灾情进行分级和识别, 实用有效。

值得指出的是, 基于模糊聚类进行灾害等级划分与识别, 首先要给出灾害模糊识别模式, 这个模式的选择具有一定的区域性, 那么这个方法的适用性只限于本区域, 具有一定的局限性。这个区域可以是一个地区, 或者一个国家。区域选的过大, 数据的离散性越大, 得到的结果越不可靠, 因此应该考虑致灾因素, 经济发展模式, 人口密度可大致相比的区域为好。这样, 该识别模式较好的反映了该地区的风暴潮灾害特征, 能够更准确的进行灾害等级的识别。

参考文献:

- [1] 乐肯堂. 我国风暴潮灾害风险评估方法的基本问题 [J]. 海洋预报, 1998, 15(3): 38-44
- [2] 王喜年. 全球海洋的风暴潮灾害概况 [J]. 海洋预报, 1993, 10(1): 30-36
- [3] 史建辉, 王名文, 王永信, 等. 风暴潮和风暴潮灾害分级问题的探讨 [J]. 海洋预报, 2000, 17(2): 12-15
- [4] 许启望, 谭树东. 风暴潮灾害经济损失评估方法研究 [J]. 海洋通报, 1998, 17(1): 1-12
- [5] 李培顺. 青岛地区的台风风暴潮灾度预报研究 [J]. 海洋预报, 1998, 15(3): 72-78
- [6] 张跃, 邹寿平, 宿芬. 模糊数学方法及其应用 [M]. 北京: 煤炭工业出版社, 1992: 227-230, 280-282
- [7] 叶雯, 刘美南, 陈晓宏. 感知器算法在台风风暴潮灾情等级评估中的应用 [J]. 中山大学学报, 2004, 43(2): 117-120
- [8] 王新洲, 史文中, 王树良. 模糊空间信息处理 [M]. 武汉: 武汉大学出版社, 2003: 64