

基于粗糙集的主成分聚类方法

花遇春, 杨璇, 熊文丹

长安大学理学院, 陕西 西安

收稿日期: 2022年4月26日; 录用日期: 2022年5月24日; 发布日期: 2022年5月31日

摘要

针对当前大规模高维数据难处理以及传统的二支聚类造成误分类等问题。本文首先根据主成分分析方法对数据进行降维处理, 并将粗糙集理论与聚类相结合, 提出了基于粗糙集的主成分聚类方法。最后, 对陕西省内11个城市的25个经济指标进行实验分析, 根据经济指标的聚类结果将城市进行三划分, 提高了聚类的准确率。

关键词

主成分分析, *K-Means*聚类, 粗糙集

Principal Component Clustering Method Based on Rough Set

Yuchun Hua, Xuan Yang, Wendan Xiong

School of Science, Chang'an University, Xi'an Shaanxi

Received: Apr. 26th, 2022; accepted: May 24th, 2022; published: May 31st, 2022

Abstract

In view of the difficulties in processing large-scale high-dimensional data and the misclassification caused by traditional two-way clustering, firstly, this paper reduces the dimension of the data according to the principal component analysis method, and combines the rough set theory with clustering, and the principal component clustering method based on rough set are proposed. Finally, 25 economic indicators of 11 cities in Shaanxi Province are experimentally analyzed. According to the clustering results of economic indicators, the cities are divided into three groups, which improves the accuracy of clustering.

Keywords

Principal Component Analysis, *K-Means* Clustering, Rough Set

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

聚类是将样本按其自身的属性聚成若干类，以保证类内样本相似度尽可能高，而类间样本相似度尽可能低。经典的聚类方法有 *K-Means* [1]和 *K-Medoids* [2]。Qian 等[3]从聚类标准、聚类表示及算法框架角度分析了多个流行的聚类算法，该文在国内聚类算法领域具有一定的地位。Johannes 等[4]从数据挖掘的角度分析了一些聚类方法。Hoppner 等[5]提出了模糊聚类的方法，基于模糊集中隶属度以及数据区间隶属度来划分不同类，姚一豫等[6]用区间集表示聚类结果中的一个类，通过离散化数据集刻画类差异性以及类间相似性。这些都是传统的二支聚类，样本要么属于这个类，要么不属于这个类，聚类具有清晰的边界，这样会造成很大的误分类。

针对上述问题，本文首先对高维数据通过主成分分析方法进行降维，有效地处理了目前大规模高维数据集，并将粗糙集与聚类相结合，将聚类划分为三个不相交的区域，将不确定性的对象划分到边界域中，提高聚类的准确率。最后，本文选取了陕西省内 11 个城市的 25 个经济指标进行研究。根据这些经济指标将 11 个城市进行粗糙聚类。通过对经济指标的研究分析，得到陕西省各地区城市发展现状，为今后各城市经济发展提供一些依据。

2. 基本概念

2.1. 主成分分析

主成分分析[7] [8] [9] (Principal Component Analysis, 简称 PCA)就是考虑各指标之间的相互关系，利用降维的方法将多个指标转换为少数几个互不相关的指标，从而使进一步研究变得简单的一种统计方法。主成分分析是 1933 年首先提出的一种“降维”的思想，在损失很少信息的前提下把多个指标转化为几个综合指标，称为主成分。主成分分析主要是使用降维的方法，使用较少的变量来代替原有的较多的变量。在变量转换过程中，采用了映射的原理。即较少的变量是原有较多属性变量的线性表示。具体内容如下：

假设原来的变量指标为 x_1, x_2, \dots, x_p ，它们的综合指标，即新变量指标为 $z_1, z_2, \dots, z_m (m \leq p)$ ，则

$$\begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \dots + l_{1p}x_p \\ z_2 = l_{21}x_1 + l_{22}x_2 + \dots + l_{2p}x_p \\ \vdots \\ z_m = l_{m1}x_1 + l_{m2}x_2 + \dots + l_{mp}x_p \end{cases} \quad (1)$$

其中， z_1, z_2, \dots, z_m 即为数据集的主成分，是原来变量指标的线性表示。 $l_{11}, l_{12}, \dots, l_{1p}$ 是系数矩阵。主成分和原始变量之间的关系是通过系数矩阵来体现的。主成分满足互不相关的要求，这在线性数学中满足逆矩阵相乘的条件。否则，则求不出系数矩阵。 z_1 是 x_1, x_2, \dots, x_p 的线性组合中的方差最大指标； z_2 是与 z_1 不相关的 x_1, x_2, \dots, x_p 的线性组合中的方差最大指标。依此类推， z_m 是 z_1, z_2, \dots, z_{m-1} 都不相关的 x_1, x_2, \dots, x_p 的

所有线性组合中的方差最大指标。这就使得主成分指标比原始指标具有某些更优越的性能。主成分分析不能看作是研究的结果，而应该在主成分分析的基础上继续采用其他方法解决问题。

2.2. 粗糙集

粗糙集(Rough Set) [10] [11]是波兰数学家 Pawlak 在 1982 年首先提出的一种处理不确定、不精确性知识的数学方法。粗糙集作为一种较新的软计算方法，近年来越来越受到重视，其有效性已在许多科学与工程领域的成功应用中得到证实，是当前国际上人工智能理论及其应用领域中的研究热点之一。在粗糙集中，近似集理论是非常重要的部分，通过近似集理论确定目标集的上、下近似，计算目标集的边界域及负域，并刻画出那些不确定的知识的粗糙程度。

定义 1 [12] [13] 设信息系统是一个四元组 $S = (U, At, V, f)$ 。 U 是非空有限的对象集合， At 是属性集合， $V = \bigcup_{a \in At} V_a$ 是属性值的集合， V_a 属性 a 的值域， f 是信息函数，且对任意 $x \in U$ ， $a \in At$ ， $f(x, a) \in V_a$ 。

定义 2 [12] [13] 给定信息系统 $S = (U, At, V, f)$ 。对于每一个属性子集 $A \subseteq At$ ，等价关系 R_A 定义如下：

$$R_A = \{(x, y) \in U \times U \mid \forall a \in A, f(x, a) = f(y, a)\}. \tag{2}$$

$U/R_A = \{[x]_{R_A} \mid x \in U\}$ 是论域 U 上的划分， $[x]_{R_A}$ 表示论域 U 中对象 x 关于等价关系 R_A 的等价类，为 U 中基于属性子集 A 的确定性知识。为方便表现，信息系统 $S = (U, At, V, f)$ 简记为 $S = (U, At)$ ， U/R_A 简记为 U/A ， $[x]_{R_A}$ 简记为 $[x]_A$ 。

例 1 表 1 给出信息系统 $S = (U, At, V, f)$ ，其中 $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ ， $At = \{a_1, a_2, a_3, a_4\}$ 。

Table 1. Information system

表 1. 信息系统

U	a_1	a_2	a_3	a_4
x_1	0.6	0.7	0	1
x_2	0.9	0.5	0.1	0.9
x_3	0.2	0.7	0.5	1
x_4	0.9	0.5	0.1	0.9
x_5	0.3	0.5	0.1	0.9
x_6	0.6	0.7	0	1

在表 1 中，属性集 At 产生的等价类为：

$$[x_1]_{At} = \{x_1, x_6\}, [x_2]_{At} = \{x_2, x_4\}, [x_3]_{At} = \{x_3\}, [x_5]_{At} = \{x_5\},$$

属性子集 $A_1 = \{a_2, a_3, a_4\}$ 产生的等价类为：

$$[x_1]_{A_1} = \{x_1, x_6\}, [x_2]_{A_1} = \{x_2, x_4, x_5\}, [x_3]_{A_1} = \{x_3\},$$

属性子集 $A_2 = \{a_2, a_4\}$ 产生的等价类为：

$$[x_1]_{A_2} = \{x_1, x_3, x_6\}, [x_2]_{A_2} = \{x_2, x_4, x_5\}.$$

定义 3 [12] [13] 给定信息系统 $S = (U, At, V, f)$ 。对任意 $X \subseteq U$ ， $A \subseteq At$ ，目标集 X 关于属性子集 A 的上近似 $\bar{R}_A(X)$ 和下近似 $\underline{R}_A(X)$ 分别定义为：

$$\begin{aligned} \bar{R}_A(X) &= \{x \mid x \in U, [x]_A \cap X \neq \emptyset\}, \\ \underline{R}_A(X) &= \{x \mid x \in U, [x]_A \subseteq X\}. \end{aligned} \tag{3}$$

一般地, 基于属性子集 A 和目标集 X , 可得目标集 X 与下近似的关系: $\underline{R}_A(X) \subseteq X$, 即下近似 $\underline{R}_A(X)$ 中的元素完全属于目标集 X 。同样可得目标集 X 与上近似的关系: $X \subseteq \bar{R}_A(X)$, 即目标集 X 中的元素一定属于上近似 $\bar{R}_A(X)$ 。

对于目标集 $X \subseteq U$, 若 $\bar{R}_A(X) = \underline{R}_A(X)$, 则目标集 X 称为可定义集; 若 $\bar{R}_A(X) \neq \underline{R}_A(X)$, 则目标集 X 称为粗糙集。如果给定属性子集 A 和目标集 X , 论域 U 被划分为如下三个不相交的区域[14]:

- 1) $POS_A(X) = \underline{R}_A(X)$, 称为目标集 X 关于属性子集 A 的正域;
- 2) $BND_A(X) = \bar{R}_A(X) - \underline{R}_A(X)$, 称为目标集 X 关于属性子集 A 的边界域;
- 3) $NEG_A(X) = U - \bar{R}_A(X)$, 称为目标集 X 关于属性子集 A 的负域。

因此, 基于属性子集 A 和目标集 X , 论域 U 被划分为三个不相交的区域, 如图 1 所示。

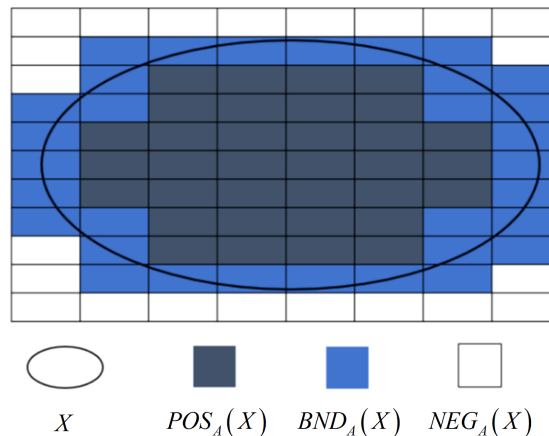


Figure 1. Three disjoint regions based on Pawlak rough set
图 1. 基于 Pawlak 粗糙集的三个不相交区域

2.3. 聚类分析

K -Means 算法最早由 Steinhaus, Lloyd 和 MacQueen 于 20 世纪五六十年代在不同的科学领域独立发现的[1], 是典型的基于距离的聚类算法之一。它采用距离作为相似性的评价指标, 即认为两个样本的距离越近, 其相似度就越大[15]。具体算法见表 2:

Table 2. K -Means clustering
表 2. K -Means 聚类

算法 1. K -Means 聚类

输入: 数据集 C , 聚类数 K

输出: $\pi_K = \{C_1, \dots, C_K\}$

步骤 1: 在 C 中随机选取 K 个样本, 记作 $\{\mu_1, \dots, \mu_K\}$, 作为聚类中心;

步骤 2: 计算剩余的每个样本 x_j 与每个聚类中心 $\mu_i (1 \leq i \leq K)$ 的距离 d_{ji} : $d_{ji} = \|x_j - \mu_i\|_2$, 并将其样本归到距离 x_j 最近的中心 μ_l 所在的类 $C_l = \{x_j \mid d_{jl} \leq d_{ji}, l \neq i\}$;

步骤 3: 重新计算已经得到的各个类的聚类中心 μ'_i : $\mu'_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$, 令 $\mu_i = \mu'_i$;

步骤 4: 迭代步骤 2 和步骤 3, 直到新的聚类中心不再发生变化或迭代次数达到最大为止。

计算所有距离的算法复杂度为 $O(k * n)$ ，因此算法 1 的复杂度为 $O(n)$ 。

3. 基于粗糙集的主成分聚类及实证分析

3.1. 基于粗糙集的主成分聚类

本节将粗糙集中的下、上近似与 *K-Means* 聚类相结合，将聚类中心改为：

$$RC(i) = e_{lower} \times \frac{1}{N_{lower}} \times \sum_{X \in C_i} X + e_{upper} \times \frac{1}{N_{upper}} \times \sum_{X \in \bar{C}_i} X \quad (4)$$

其中， e 表示权重， $upper$ 表示粗糙集的上近似集， $lower$ 表示粗糙集的下近似集。 N_{lower} 和 N_{upper} 分别表示第 i 个类的下近似集和上近似集的样本个数。本文运用一种基于主成分分析的粗糙集聚类算法。即对数据先进行主成分分析，然后进行粗糙集聚类，得到聚类中心，每个类都包含有上近似集和下近似集，下近似集中的对象表示确定属于该类的对象，上近似集的对象表示可能属于该类的对象。基于粗糙集的主成分聚类分析算法如表 3：

Table 3. Principal component clustering algorithm based on rough set

表 3. 基于粗糙集的主成分聚类算法

算法 2. 基于粗糙集的主成分聚类算法

输入：数据集 C ，聚类数 K ，阈值 α

输出：样本集合的聚类 C^*

步骤 1：在 C 中随机选取 K 个样本，记作 $\{\mu_1, \dots, \mu_K\}$ ，作为聚类中心；

步骤 2：计算剩余的每个样本 x_j 与每个聚类中心 $\mu_i (1 \leq i \leq K)$ 的距离 d_{ji} ： $d_{ji} = \|x_j - \mu_i\|_2$ ，若 $d_{ji} \geq \alpha$ ，将样本划分到类簇 C_i 的上近似中，若 $d_{ji} < \alpha$ ，将样本划分到类簇 C_i 的下近似中；

步骤 3：重新计算已经得到的各个类的聚类中心 μ'_i ： $\mu'_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$ ，令 $\mu_i = \mu'_i$ ，得到聚类结果 $C^{(t)}$ ，聚类中心为 $\{\mu_1, \dots, \mu_K\}$ ；

步骤 4：迭代步骤 2 和步骤 3，直到新的聚类中心不再发生变化或迭代次数达到最大为止，输出 $C^* = C^{(t)}$ 。

计算所有距离的算法复杂度为 $O(k * n)$ ，因此算法 2 的复杂度为 $O(n)$ 。

3.2. 实验结果及分析

3.2.1. 数据集

本文研究的数据是陕西 11 个城市的 25 个经济指标数据。数据来源于由国家统计局主编的《陕西统计年鉴》第二部分。陕西省的 11 个城市分别为：西安市、铜川市、宝鸡市、咸阳市、渭南市、杨凌示范区、汉中市、安康市、商洛市、延安市、榆林市。25 个经济指标为：年底常住人、城镇非私营单位就业人员年末人数、生产总值、房地产开发投资、地方一般性支出、一般预算支出、城镇非私营单位就业人员工资总额、城镇非私营单位就业人员平均工资、城镇居民人均可支配收入、农村居民人均可支配收入、农林牧渔业总产值、粮食产量、棉花产量、油料产量、规模以上工业总产值、邮电业务总量、固定电话、移动电话、社会消费品零售总额、进出口总值、出口总值、实际利用外直接投资额、卫生机构数、卫生机构床位数、卫生技术人员。如表 4 所示。

Table 4. The basic descriptions of data sets**表 4.** 数据基本描述

指标	关中	西安
年底常住人口(万人)	2459.12	1,020.35
城镇非私营单位就业人员年末人数(万人)	327.98	187.46
生产总值(亿元)	16,090.29	9321.19
房地产开发投资(亿元)	3282.09	2464.78
地方一般预算收入(亿元)	997.55	702.56
一般预算支出(亿元)	2622.85	1247.02
城镇非私营单位就业人员工资总额(亿元)	2592.43	1718.33
城镇非私营单位就业人员平均工资(元)		92,584
城镇居民人均可支配收入(元)		41,850
农村居民人均可支配收入(元)		14,588
农林牧渔业总产值(亿元)	1989.14	499.32
粮食产量(万吨)	689.58	139.90
棉花产量(吨)	6520	122
油料产量(万吨)	12.03	0.67

3.2.2. 实验结果

1) 传统聚类分析

a) 使用 SPSS 进行的数据分析，得出初始聚类中心见表 5。

Table 5. Initial cluster center table**表 5.** 初始聚类中心表

	初始聚类中心		
	中心		
	1	2	3
年底常住人口	1020.350	237.910	527.810
城镇非私营单位就业人员年末人数	187.460	19.330	41.690
生产总值	9321.190	837.210	1828.470
房地产开发投资	2464.780	27.280	284.880
地方一般性支出	702.560	20.960	86.010
一般预算支出	1247.020	298.570	467.310
城镇非私营单位就业人员工资总额	1718.330	107.820	251.280

Continued

城镇非私营单位就业人员平均工资	92,584.00	56,628.00	61,623.00
城镇居民人均可支配收入	41,850.00	25,503.00	33,674.00
农村居民人均可支配收入	14,588.00	10,025.00	12,775.00
粮食产量	139.90	49.15	226.86
农林牧渔业总产值	499.32	193.07	577.69
农林牧渔业总产值	499.32	193.07	577.69
棉花产量	122	765	6357
油料产量	0.67	2.13	6.34
规模以上工业总产值	6252.54	1325.25	2004.48
邮电业务总量	1432.95	129.40	341.81
固定电话	276.41	18.80	65.50
移动电话	1731.94	180.59	501.48
社会消费品零售总额	5140.93	214.62	718.81
进出口总值	3243.06	17.90	14.08
出口总值	1730.21	17.64	11.69
实际利用直接投资额	665,666	5061	5048
卫生机构床位数	72,549	14,986	33,279
卫生机构数	7011	2782	4251
卫生技术人员	136,333	21,156	52,033

b) 表 6 为迭代历史, 该表给出了迭代过程中类中心的变动量, 可以看出本次聚类过程进行了 2 次就收敛了。

Table 6. Iteration history

表 6. 迭代历史记录

迭代	迭代历史记录		
	聚类中心中的变动		
	1	2	3
1	0.000	10,302.436	11,679.885
2	0.000	0.000	0.000

c) 表 7 为聚类成员, 该表给出了样本观测量所属类别以及与所属类中心的距离。从表 7 中可以看出, 将 11 个城市分成了 3 类, {西安}为第一类, {铜川市、杨凌示范区、安康市、商洛市、延安市}为第二类, {宝鸡市、咸阳市、渭南市、榆林市、汉中市}为第三类。

Table 7. Cluster member
表 7. 聚类成员

个案号	聚类	距离
1	1	0.000
2	2	7573.2
3	3	6012.333
4	3	8612.773
5	3	11,679.885
6	2	19,711.903
7	3	11,143.668
8	2	3643.108
9	2	10,302.436
10	2	9208.111
11	3	15,386.653

由于以上部分的分析都是在表中展示了聚类中心之间的距离以及各类城市的聚类类别，数据可视性较差。图 2 是本文根据 *K-Means* 聚类方法在 R 语言上绘制聚类成像图，使得各聚类成员所属类别以及各类别之间的距离更为清楚和可视化。

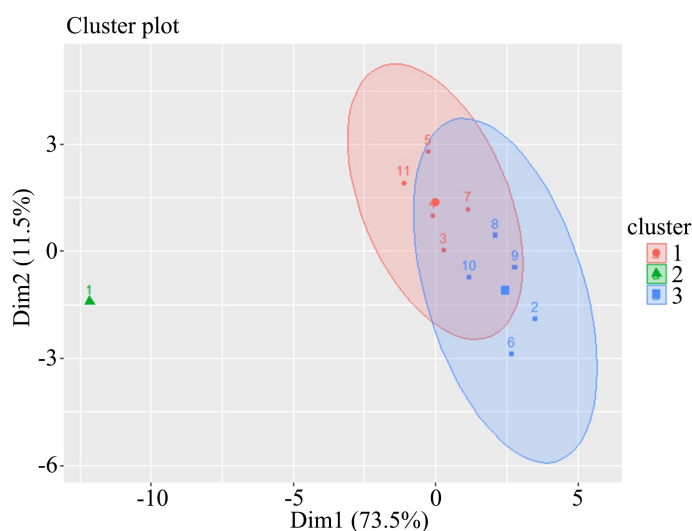


Figure 2. Clustering member based on *K-means*
图 2. 基于 *K-means* 聚类成员图

2) 主成分聚类分析

由经典聚类分析的聚类结果得到：25 个属性代表 11 个城市的经济指标，数据信息冗余量较大，而且很多信息在计算过程中对计算结果带来了干扰。所以需要先进行主成分分析以达到降维的目的。

这一部分利用主成分分析的结果作为聚类分析的数据源，对数据进行基于粗糙集的聚类分析。将这 11 个城市(示范区)进行基于粗糙集的聚类分析，得到最终的聚类中心和聚类的上近似集和下近似集中心。如表 8 所示：

Table 8. Final clustering center
表 8. 最终聚类中心

主成分		聚类		
		1	2	3
第一类	下近似集	2.83850	-0.32549	-0.18670
	上近似集	2.83850	0.26113	0.06385
第二类	下近似集	-0.83391	-0.16343	0.65931
	上近似集	-0.83391	1.13054	1.65006
第三类	下近似集	-0.36979	-0.01308	0.15379
	上近似集	-0.36979	0.23015	1.54236
第四类	下近似集	-0.35833	0.45166	-0.93442
	上近似集	-0.35833	2.32935	-1.72544

根据聚类的上近似集中心和下近似集中心，可以得到的基于主成分降维后的粗糙集聚类结果如表 9 所示：

Table 9. Number of objects in the cluster
表 9. 聚类中的对象个数

类数		对象个数
第一类	下近似集	1
	上近似集	1
第二类	下近似集	4
	上近似集	7
第三类	下近似集	6
	上近似集	8

根据聚类的上、下近似集的聚类结果可得基于粗糙近似集的主成分聚类结果，如表 10 所示：

Table 10. Principal component clustering results based on rough approximation set
表 10. 基于粗糙近似集的主成分聚类结果

类数		聚类结果
第一类	下近似集	西安
	上近似集	西安
第二类	下近似集	铜川市、汉中市、安康市、商洛市
	上近似集	铜川市、汉中市、安康市、商洛市、宝鸡市、咸阳市、延安市
第三类	下近似集	宝鸡市、咸阳市、渭南市、杨凌示范区、延安市、榆林市
	上近似集	宝鸡市、咸阳市、渭南市、杨凌示范区、延安市、榆林市、铜川市、汉中市

图3 根据基于粗糙近似集的主成分聚类方法在 R 语言上绘制聚类成像图, 使得各聚类成员所属类别以及各类别之间的距离更为清楚和可视化。

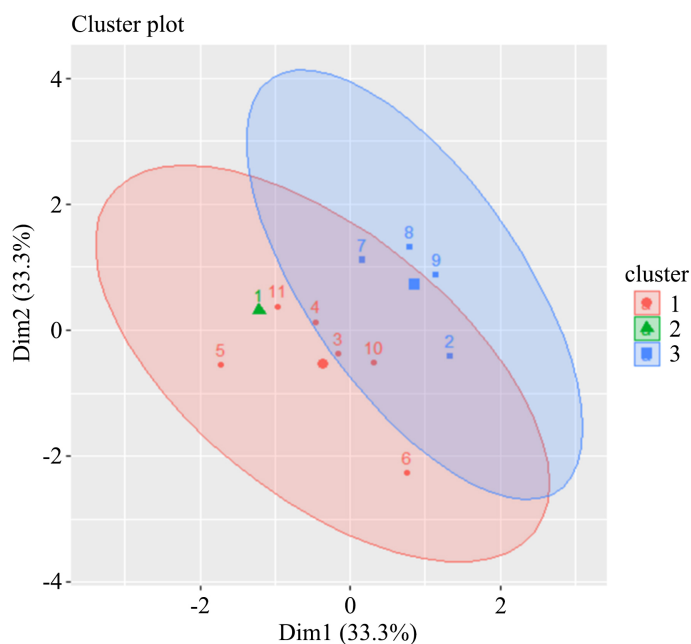


Figure 3. Principal component clustering member based on rough approximation set

图3. 基于粗糙近似集的主成分聚类成员图

4. 总结

本文提出了基于粗糙集的主成分聚类方法。首先通过主成分分析将高维数据降维, 解决了高维数据难处理的问题。进一步将粗糙集与聚类相结合, 将数据集划分为不相交的三部分, 降低了误分类的概率。最后对陕西 11 个城市的 25 个经济指标数据进行实验, 将 11 个城市划分为三类, {西安}为第一类, {汉中市、安康市、商洛市、延安市}为第二类, {宝鸡市、咸阳市、渭南市、榆林市}为第三类。但从聚类的结果中可以看到存在一定不合理性。例如, 咸阳市毗邻西安市, 实际的经济水平高于其他城市。在基于粗糙集的主成分聚类结果中, 将 11 个城市划分为三类, 第一类仍然为西安, 陕西省省会城市经济发展水平远高于其他城市。第二类为中等发展城市, 第三类为不发达城市。采用本文方法得到的聚类结果更为合理, 层次更为鲜明, 同时也与实际情况较为契合。

参考文献

- [1] Jain, A.K. (2009) Data Clustering: 50 Years beyond K-Means. *Pattern Recognition Letters*, **31**, 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- [2] Kaufmann, L. and Rousseeuw, P.J. (1987) Clustering by Means of Medoids. In: Dodge, Y., Ed., *Statistical Data Analysis Based on the L1-Norm & Related Methods*, North-Holland, Amsterdam, 405-416.
- [3] Qian, W.N. and Zhou, A.Y. (2002) Analyzing Popular Clustering Algorithms from Different Viewpoints. *Journal of Software*, **32**, 432-445.
- [4] Grabmeier, J. and Rudolph, A. (2002) Techniques of Cluster Algorithms in Data Mining. *Data Mining and Knowledge Discovery*, **6**, 303-360. <https://doi.org/10.1023/A:1016308404627>
- [5] Hoppner, F., Klawonn, F., Kruse, R., et al. (1999) *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*. Wiley Press, Chichester, 1-48.

-
- [6] Yao, Y.Y., Lingras, P., Wang, R.Z., *et al.* (2009) Interval Set Cluster Analysis: A Re-Formulation. *Proceedings of the 12th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, Delhi, 16-18 December 2009, 398-405. https://doi.org/10.1007/978-3-642-10646-0_48
- [7] Nie, B., Du, J., Liu, H., *et al.* (2009) Crowds' Classification Using Hierarchical Cluster, Rough Sets, Principal Component Analysis and Its Combination. *2009 International Forum on Computer Science-Technology and Applications*, Chongqing, 25-27 December 2009, 287-290. <https://doi.org/10.1109/IFCSTA.2009.75>
- [8] Zhao, L. and Guo, Z. (2011) Face Recognition Method Based on Adaptively Weighted Block-Two Dimensional Principal Component Analysis. *Third International Conference on Computational Intelligence, Communication Systems and Networks, CICSyN 2011*, Bali, 26-28 July, 2011, 22-25.
- [9] Zhang, X. and Ren, X. (2011) Two Dimensional Principal Component Analysis Based Independent Component Analysis for Face Recognition. *2011 International Conference on Multimedia Technology*, Hangzhou, 26-28 July 2011, 934-936. <https://doi.org/10.1109/ICMT.2011.6002199>
- [10] Liu, G., Song, X. and Zhao, X. (2007) Rough Sets and Zadeh's Extension Principles. *Proceedings-2007 IEEE International Conference on Granular Computing*, Fremont, 2-4 November 2007, 180-185. <https://doi.org/10.1109/GrC.2007.19>
- [11] 苗夺谦, 李道国. 粗糙集理论、算法与应用[M]. 北京: 清华大学出版社, 2008.
- [12] Pawlak, Z. and Skowron, A. (2007) Rudiments of Rough Sets. *Information Sciences*, **177**, 3-27. <https://doi.org/10.1016/j.ins.2006.06.003>
- [13] Pawlak, Z. (1984) Rough Classification. *International Journal of Man-Machine Studies*, **20**, 469-483. [https://doi.org/10.1016/S0020-7373\(84\)80022-X](https://doi.org/10.1016/S0020-7373(84)80022-X)
- [14] Yao, Y.Y. (2015) Rough Sets and Three-Way Decisions. *International Conference on Rough Sets & Knowledge Technology*, Tianjin, 20-23 November 2015, 62-73. https://doi.org/10.1007/978-3-319-25754-9_6
- [15] 王平心, 刘强, 杨习贝, 等. 基于动态邻域的三支聚类分析[J]. 计算机科学, 2018, 45(1): 62-66+89.