



移动扫码阅读

韩东辉, 唐跃刚. SOM+k-means 两段聚类煤质大数据挖掘方法与应用[J]. 煤炭科学技术, xxxx, xx(x): x-xx.  
HAN Donghui, TANG Yuegang. Coal quality big data mining method and application based on SOM plus K-means two-stage clustering[J]. Coal Science and Technology, xxxx, xx(x): x-xx.

## SOM+k-means 两段聚类煤质大数据挖掘方法与应用

韩东辉, 唐跃刚

(中国矿业大学(北京)地球科学与测绘工程学院, 北京 100083)

**摘要:** 充分利用煤炭开发利用过程中积累的海量煤质数据, 挖掘其中隐含的信息, 可以产生新的信息, 应用于社会生产建设。不同地质条件煤炭资源优势区域会呈现一种数据分布聚类现象。选取山西省六大煤田太原组原煤水分( $M_{ad}$ )、灰分( $A_d$ )、挥发分( $V_{daf}$ )以及全硫( $S_{t,d}$ )四个参数, 将原始数据进行预处理后的数据采用 SOM+K-means 算法处理, 将读取的数据首先基于自组织神经网络 SOM 处理, 将得到的结果作为第二阶段 k-means 聚类分析算法进一步处理。依据国家相关标准将两类数据中按照原煤质量不同展布到地图上, 划定优势区域。数据挖掘结果表明, 第一类、第二类聚类中优质煤及中质煤优势区域面积所占比例分别为 90.1% 和 24.1%, 说明第一聚类相较于第二聚类原煤质量高。由此证明数据挖掘算法煤质大数据分析的可能性, 拓展了煤质数据的使用方式, 为煤质数据库使用与发展提供了新的思路。

**关键词:** 煤质大数据; 聚类分析; 信息挖掘; 资源评价

**中图分类号:** 文献标志码:A 文章编号: 0253-2336(2023)00-0001-09

### Coal quality big data mining method and application based on SOM plus K-means two-stage clustering

HAN Donghui, TANG Yuegang

(School of Earth Sciences and Surveying Engineering, China University of Mining and Technology-Beijing, Beijing 100083)

**Abstract:** In the process of developing and utilizing coal resources, a large amount of data is generated, and this data contains a lot of potentially valuable information. Making full use of the massive coal quality data accumulated in the process of coal development and utilization and mining the hidden information can generate new information and apply it to social production and construction. Areas with advantageous coal resources under different geological conditions will present a clustering phenomenon of data distribution. Four parameters of raw coal, including moisture ( $M_{ad}$ ), ash yield ( $A_d$ ), volatile matter ( $V_{daf}$ ) and total sulfur ( $S_{t,d}$ ) of the Taiyuan Formation in the six major coal fields in Shanxi Province are selected. The raw data is preprocessed using SOM+K-Means algorithm processing, and the read data is first processed based on the self-organizing neural network SOM, and the result obtained is used as the second stage k-means clustering analysis algorithm for further processing. According to the relevant national standards, the two types of data are displayed on the map according to the different quality of the raw coal, and the advantageous areas are delineated. The results of data mining show that the dominant areas of high-quality coal and medium quality coal account for 90.1% and 24.1% in the first and second clusters, respectively, indicating that the first cluster has higher quality coal than the second cluster. This proves the possibility of data mining algorithm coal quality big data analysis, expands the use of coal quality data, and further provides new ideas for the use and development of coal quality databases.

**Key words:** coal quality big data; cluster analysis; information mining; resource evaluation

收稿日期: 责任编辑: 周子博 DOI: 10.13199/j.cnki.cst.2021-1048

基金项目: 国家自然科学基金重点资助项目(42030807); 国家自然科学基金资助项目(41572146); 山西省国土资源厅资助项目(HXCT01-2018F005-77)

作者简介: 韩东辉(1994—), 男, 黑龙江省齐齐哈尔人, 硕士。E-mail: handonghui1994@163.com

通讯作者: 唐跃刚(1958—), 男, 重庆人, 教授, 博士生导师, 博士。E-mail: tyg@vip.163.com

## 0 引言

科学技术的跨越式发展以及过程工业中数据快速大量的积累导致在原始数据资料基础上获取更有价值的信息越来越重要<sup>[1-2]</sup>, 对这一趋势的研究对国民经济发展具有重要意义<sup>[3-5]</sup>。我国在煤炭资源勘查与生产开发利用过程中积累了大量的煤质数据, 部分学者通过数据挖掘分析了提高选煤厂生产效率的可能性<sup>[6-11]</sup>。探讨了利用煤炭销售运营大数据降低生产成本的途径<sup>[12]</sup>。利用煤炭开发过程中积累的海量煤质数据, 以一种新的数据信息提取方式发掘其中的价值尤为重要。不同于传统的煤质数据使用的方式, 基于聚类分析算法处理煤质数据, 对煤质数据进行数据挖掘, 探究多属性指标的分布规律, 拓展煤质数据库的使用方式, 为煤质数据的利用提供一种新的思路和方法。

## 1 原理与流程

### 1.1 SOM 神经网络法

自组织特征映射神经网络又被称为 SOM 神经网络或 SOM, 1981 年由 Kohonen 教授提出<sup>[13]</sup>, 其主要的设计思想来源于生物学。类似于人的大脑, 面对某一特定的信息输入, 处在相应区域位置也会产生相应的兴奋, 而且对于类似的输入信息, 其产生兴奋的区域是连续的。SOM 神经网络就是来源于模拟生物的神经网络机制, 其在数据降维<sup>[14]</sup>、可视化以及聚类分析中都有着广泛应用。SOM 神经网络具有对噪声数据抗干扰能力强, 而且在初始聚类时并不用指定聚类数与聚类中心等优点, 可以将高维数据进行转换、降维映射到低维度空间, 很好地处理非线性数据, 自动将数据分开。但是, SOM 聚类方式无法提供聚类后的精确聚类信息<sup>[15]</sup>。

经典的 SOM 包括输入层、输出层 2 层, 分别用来模拟感知外界信息输入的神经细胞以及神经元兴奋细胞, 输入层与输出层以权重向量  $W_{ji}$  相连, 较为常见的是将神经元分布在二维平面(图 1)<sup>[16]</sup>。其中, 第一层为输入层, 神经元节点个数与输入数据的维度相同; 第二层为竞争层, 神经元节点与输入层节点以权值向量形式相连接, 这种连接不限于竞争层节点与输入层节点的连接, 在竞争层之间和之内节点也相互连接, 导致模拟大脑皮层的竞争层神经元不仅与外界输入信息相关, 也与本身周围神经元之间的相互作用相关, 进而模拟生物神经行为。当外界输入一数据时, 竞争层的各神经元彼此之间相互竞争<sup>[17]</sup>, 进而产生一与输入模式最为相似的神经元成

为获胜神经元, 获胜神经元对周围神经元的影响由近及远逐渐减弱, 在这样的机制下达到聚类的目的。

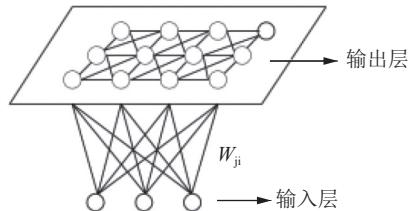


图 1 SOM 神经网络结构<sup>[16]</sup>  
Fig.1 SOM neural network structure

### 1.2 k-means 算法

k-means 算法即 k 均值聚类算法, 作为一种无监督的学习方式, 将对象分成不同的簇, 衡量数据是否在同一簇的指标一般为欧氏距离, 通过计算不同数据点与给定的初始聚类中心点之间的距离进而将不同的数据点划分, 并且不断地迭代这一过程, 直到满足算法的终止条件。K-means 是应用较为广泛的算法, 也是极具代表性的聚类分析算法, 属于划分聚类的一种, 是一种目前在生产生活中得到广泛使用的聚类分析算法<sup>[18]</sup>。

k-means 算法是把  $k$  作为参数, 把所有的数据点划分为  $k$  个簇<sup>[19]</sup>, 使得同一个簇内数据点之间的距离较小, 数据之间彼此相似, 不同簇之间距离尽量的大, 使得不同簇最大的不同。该方法使用的局限性在于其不能应用于类别属性的数据, 但数值型数据能得到很好的使用效果, 也存在着聚类结果与初始参数的选取有关以及易受噪声数据影响等缺点。

### 1.3 SOM+k-means 两阶段聚法

上述两类方法各有优势和局限。k-means 算法简单, 算法思想容易理解, 计算方便、快速, 但缺点同样十分明显。SOM 神经网络对噪声数据抗干扰能力强, 在初始聚类时并不用指定聚类数与聚类中心, 可以将高维数据进行转换、降维映射到低维度空间, 可以很好地处理非线性数据, 但无法提供聚类后的精确聚类信息。

结合上述 2 种方法的优点, 采用 SOM+K-means 两阶段聚类方式, 将 SOM 处理结果作为 K-means 算法的输入, 再利用 K-means 算法进一步聚类得到相应结果, 以解决两种算法在各自单独应用时的劣势。在进行 K-means 聚类时, 要首先指定  $k$  值的大小, 确定最佳聚类数  $K$  值的方式有很多, 以手肘法为主、轮廓系数法为辅的方式确定  $K$  值大小。

### 1.4 数据处理流程

SOM+k-means 两阶段聚类法煤质数据处理流程包括 3 个步骤:

### 第一步, 数据选取与预处理。

选取待分析研究数据, 本文主要选取山西省六大煤田太原组原煤煤质数据, 包括水分、灰分、挥发分、硫分 4 个指标。煤质数据录入过程中可能存在错误、冗余及缺失情况, 冗余或重复的数据采用适当的删除方式, 个别属性值缺失采用均值替代方式填充, 坐标缺失、属性值明显错误的数据予以剔除。数据处理完成之后需进行数据展布工作, 原始数据坐标要进行转化, 以适应后期处理。

### 第二步, SOM+k-means 算法数据处理。

将上一步预处理后的煤质数据, 利用 SOM+k-means 算法进行处理。算法主要是基于 Python 相关的程序包, 利用程序设计将输入数据进行聚类分析, 同一聚类中数据的相似程度最高, 相关指标的数值表现最为相似。存储聚类获得的不同簇类煤质数据,

以备为下一步数据展布分析。

### 第三步, 聚类数据展布及信息挖掘。

将聚类数据利用 Python 程序展布到地图上, 分析不同聚类的地域分布特点。将不同指标数据表现相似的同一簇聚类数据归纳为一类性质相似的煤炭资源, 结合不同聚类所划分的优势区域, 挖掘不同聚类信息, 揭示聚类煤炭资源分布特点。

## 2 煤质数据及处理结果

选用山西省六大煤田太原组原煤的水分、灰分、挥发分、硫分为数据基础, 数据来源于山西省煤炭地质勘查研究院, 同时也得到了相关项目的数据支持<sup>[20-21]</sup>。收集到的数据进行预处理之后, 去除不合理的数据, 得到的样本分布平面图如下所示(图 1), 样本在各煤田分布及属性概况见表 1。

表 1 样本在各煤田的分布及属性概况

Table 1 The distribution and attribute overview of the samples in each coal field

煤田	数量	水分( $M_{ad}$ )			灰分( $A_d$ )			挥发分( $V_{daf}$ )			全硫( $S_{t,d}$ )		
		最大值	最小值	平均值	最大值	最小值	平均值	最大值	最小值	平均值	最大值	最小值	平均值
沁水	585	4.63	0.04	1.16	47.86	7.01	21.85	31.44	2.95	13.58	11.48	0.30	3.34
河东	277	6.40	0.21	1.18	49.20	7.86	19.85	42.54	7.98	24.16	7.42	0.32	2.23
大同	38	7.03	0.97	3.63	41.32	2.95	27.52	42.62	29.46	38.06	2.43	0.07	1.08
宁武	63	2.93	1.06	1.82	22.02	12.55	22.27	38.54	14.23	35.47	2.22	1.45	2.01
霍西	95	2.84	0.08	1.06	38.94	6.63	23.19	39.18	4.94	26.36	6.47	0.81	2.69
西山	55	6.83	0.26	1.02	40.31	9.73	19.61	22.68	6.28	16.04	9.51	0.42	2.58

将处理的煤质数据进行 SOM 神经网络处理。由于后续进行第二阶段聚类, SOM 神经网络不需要完全收敛, 只需要得到大致聚类结果, 故将迭代的次数设为 350 次<sup>[22]</sup>。经由 SOM 处理后聚类数较多, 将这一步的结果输入到 k-means 算法中进行再次聚类,  $k$  值对应的轮廓系数越大,  $k$  值越合理。从肘部法折线图看出, 当  $k$  为 2 时, 折线上对应的点明显为折线的肘部(图 2), 聚类数为 2 时, 轮廓系数最大(图 3)。

上面分析显示 2 为最佳  $k$  值, 最终将数据聚为两类(表 2)通过 K-means 将数据聚为两类。其中第 1 簇数据样本共 263 个, 占样本总数的 24.6%; 第 2 簇共 804 个样本, 占比 75.4%。

## 3 讨 论

对比两类相关属性信息, 与第 2 类相比, 第 1 类具有较高的水分含量、灰分产率、挥发分产率以及相对较低的全硫含量(表 3)。为划分更细致, 且与国际标准更加接轨, 灰分指标不采用最新标准划分<sup>[23]</sup>, 而

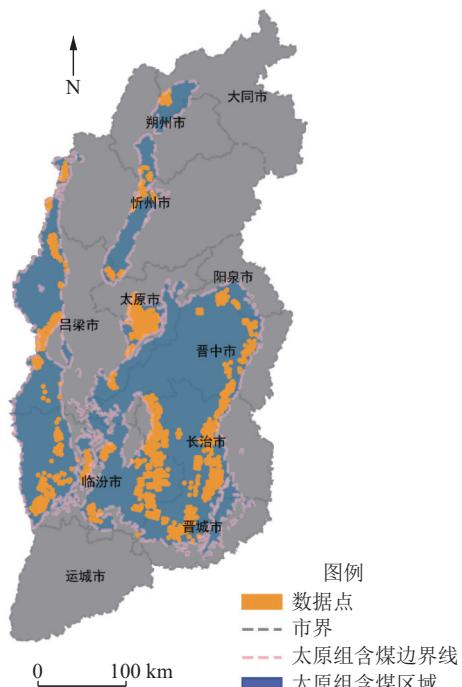


图 2 样本平面分布  
Fig.2 Sample plane distribution

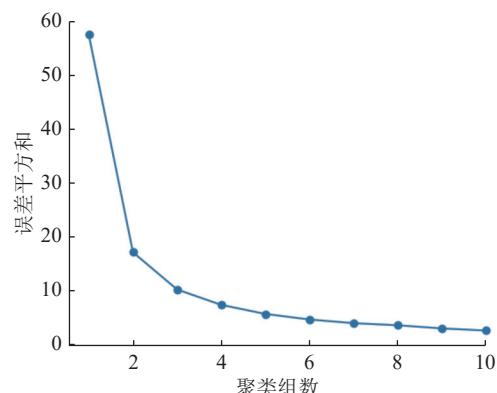


图 3 肘部法  $k$  值折线图  
Fig.3 Elbow method line chart of  $k$

表 2 数据样本聚类簇分布

Table 2 Data sample cluster cluster distribution

簇号	数目	占比
1	263	24.6%
2	804	75.4%

是根据国标 1994<sup>[24]</sup> 将煤中灰分划分为 6 个级别, 根据相关标准划分<sup>[25-26]</sup>, 第一类煤炭主要是中高灰分、中高挥发分、中硫分煤, 其相较于第二类具有较高的水分值, 第二类相对第一类为较低水分、中灰分、低挥发分高硫煤。第 1 类中, 特低灰煤到高灰煤均有分布, 可能原因是该类数据样本来自不同煤田的不同煤层, 簇内该属性值表现差异较大。第 2 类挥发分产率相对较低, 数值范围波动较小, 可能潜在原因是该簇内数据主要来自沁水煤田, 属性值表现相对较为集中, 样本变质程度较高<sup>[27]</sup>。沁水煤田受海侵影响较大, 各样本受海侵程度不同, 全硫属性值波动较大。

基于杨淑婷等<sup>[28-29]</sup>提出的分级方法, 将原煤质量等级划分为 6 级。分析发现两个聚类簇中均分布有优质煤、中质煤、低质煤以及劣质煤, 分布面积见表 4。将不同聚类簇煤质数据投影到平面, 原煤质量等级优势区域(图 3, 图 4)。

表 3 不同聚类簇煤质属性统计结果

Table 3 Statistical results of coal quality attributes of different clusters

簇号	水分( $M_{ad}$ )				灰分( $A_{ad}$ )				挥发分( $V_{daf}$ )				硫分( $S_{ad}$ )			
	最大值	最小值	均值	方差	最大值	最小值	均值	方差	最大值	最小值	均值	方差	最大值	最小值	均值	方差
1	7.03	0.23	1.69	1.80	47.86	2.95	24.00	52.58	42.62	21.52	33.9	22.70	6.27	0.32	1.62	0.88
2	6.83	0.04	1.07	0.55	44.93	7.14	20.78	50.52	28.54	2.95	14.21	15.82	11.48	0.30	3.27	3.03

表 4 原煤质量级别统计及其分布

Table 4 Statistics and distribution of raw coal quality grades

煤田	第1簇/km <sup>2</sup>						第2簇/km <sup>2</sup>					
	I: 稀缺特好	II: 特好	III: 优质	IV: 中质	V: 低质	VI: 劣质	I: 稀缺特好	II: 特好	III: 优质	IV: 中质	V: 低质	VI: 劣质
沁水	—	—	—	—	—	—	—	—	2 669.69	3 727.73	11 413.64	9 148.09
河东	—	—	1 301.07	1 526.00	192.72	53.61	—	—	—	899.48	2 009.09	1 320.17
霍西	—	—	—	223.24	123.31	53.10	—	—	—	—	—	600.24
西山	—	—	—	—	—	—	—	—	431.85	391.80	872.94	166.17
宁武	—	—	818.63	1 197.99	118.88	43.17	—	—	—	—	—	—
大同	—	—	227.12	358.47	32.91	—	—	—	—	—	—	—
合计			2 346.82	3 305.70	467.82		—	—	3 101.54	5 019.01	14 295.67	11 234.67

注: “—”为未统计到

2 个聚类簇数据挥发分产率、全硫含量存在较大差异, 是产生聚类现象的主要原因。煤中硫分含量主要受成煤植物以及海水的影响, 浅海相以及海陆交互相沉积煤田煤中全硫含量一般较高<sup>[30]</sup>。山西省本溪组以上太原组 10 号煤层以下为层序 I, 太原组中上段为层序 II, 两个层序海侵方向不同, 层序 I 海侵来自山西省东北方向, 层序 II 海侵改变为东南

方向<sup>[31]</sup>。由此, 可能导致海侵时间出现差异以及沉积环境的不同, 造成了全硫含量的差异。两簇类数据挥发分产率相差较大的潜在原因可能是数据样本的煤化程度不同(图 5)<sup>[32-34]</sup>。对比分析两聚类煤质特征与图 5 可以发现, 沁水煤田分布有大量的煤级较高的煤炭, 从质量等级优势区域分布图与图 5 的分布关系可以看出, 第 2 类数据样本大量分布于沁

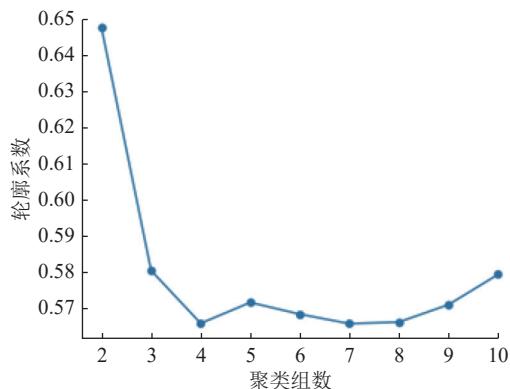


图 4 轮廓系数  $k$  值折线图  
Fig.4 Silhouette Coefficient line char of  $k$

水煤田, 主要分布的地区为中高煤级煤炭大量分布的区域, 煤化程度较高, 挥发分数值表现较低。第 1 类数据样本分布区域煤级较低, 挥发分数值较高。

分析发现, 第一聚类中, 优质煤与中质煤面积所占比例在 90% 以上, 主要分布在大同煤田潘家窑勘查区, 宇武煤田马家湾、曲立勘查区, 河东煤田杨家湾、杨家塔勘查区和三交详查区附近, 以及霍西煤田万安、乔家湾勘查区, 在西山煤田以及沁水煤田只有零星数据点分布。第二聚类中低质煤与劣质煤面积所占比例接近 80%, 主要分布在河东煤田郭家沟、白额勘查区, 吉县车城勘查区, 西山煤田草庄头勘查区,

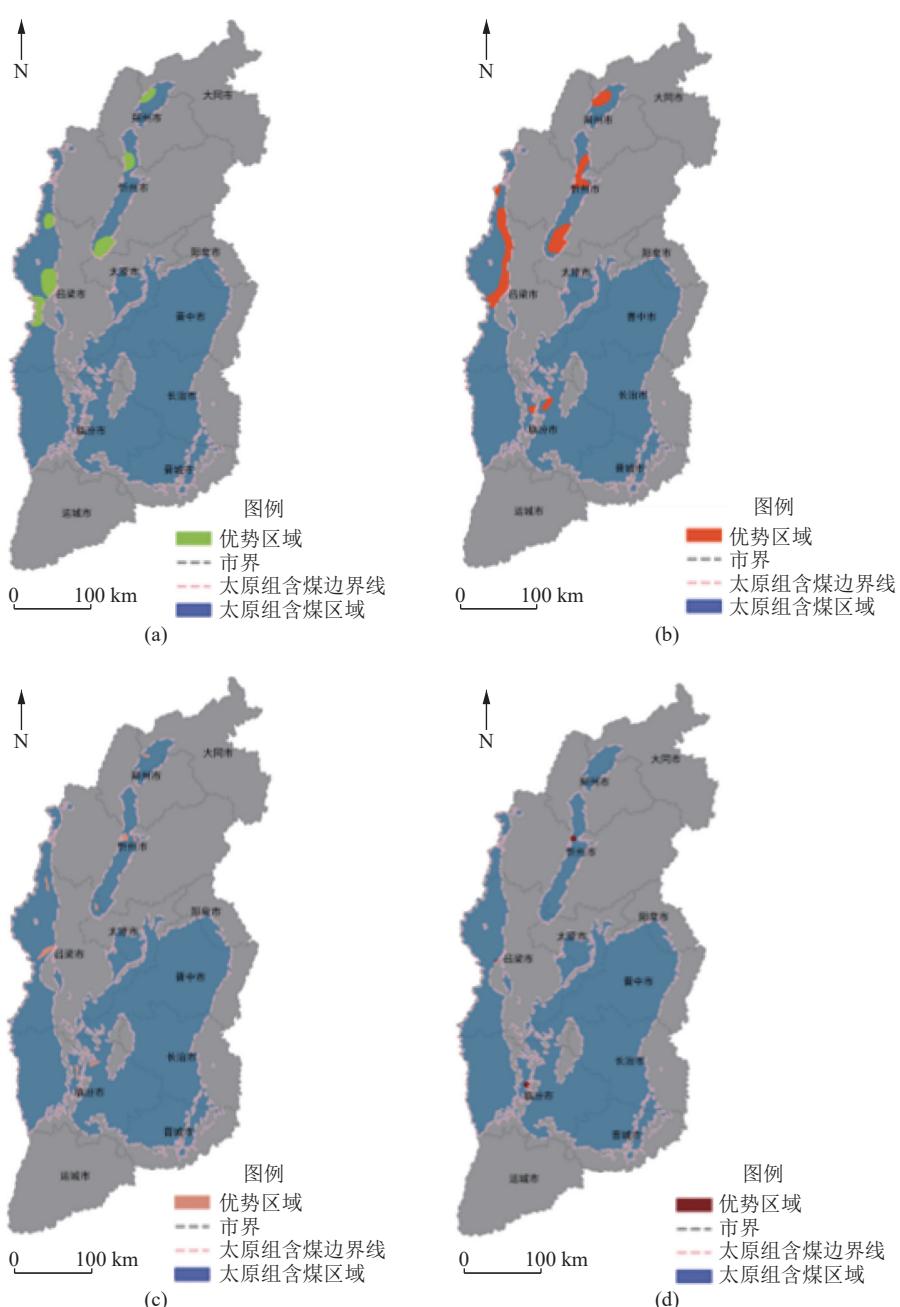


图 5 第一聚类原煤质量等级优势区分布  
Fig.5 The distribution of the dominant area of the first cluster raw coal quality grade

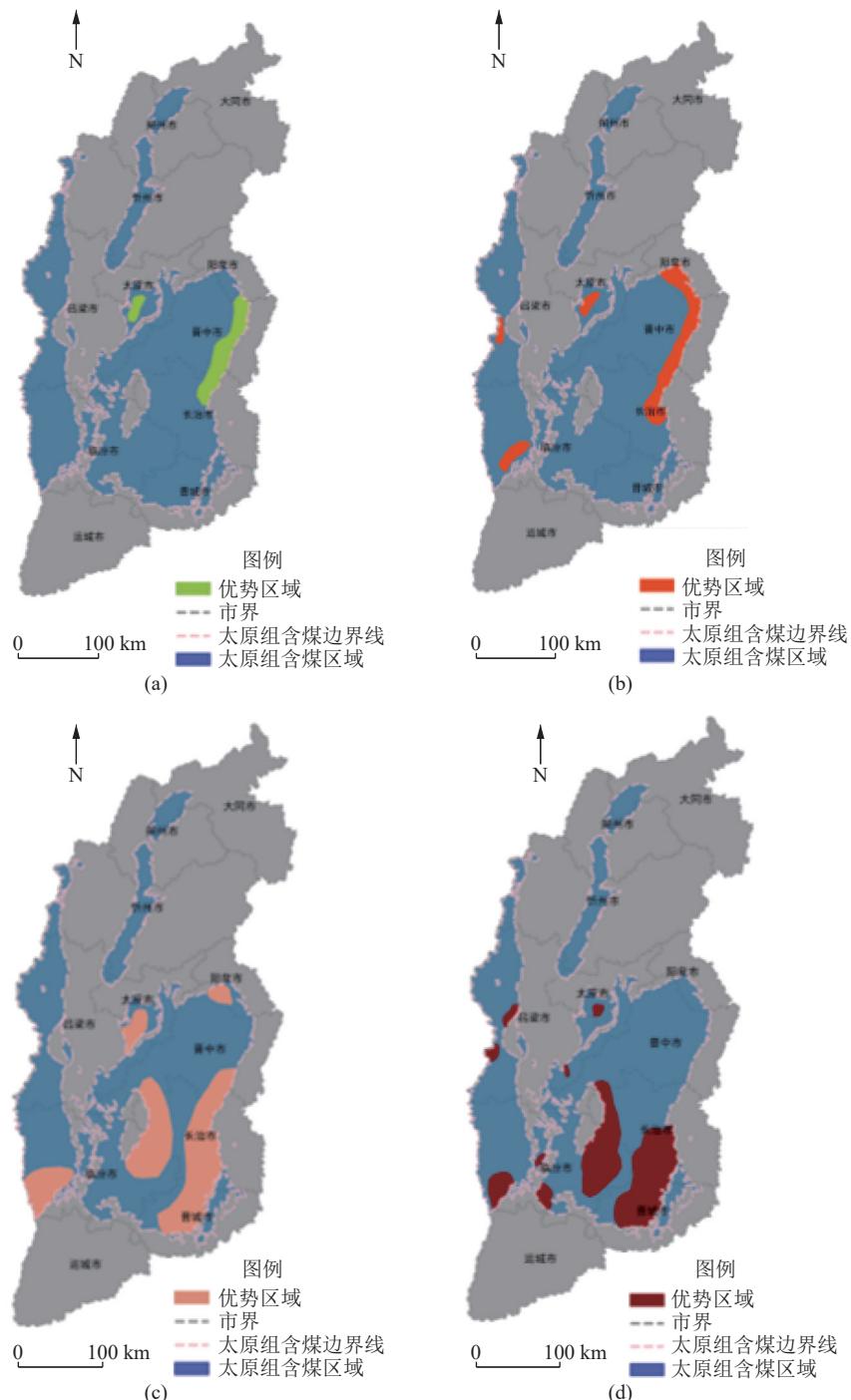


图 6 第二聚类原煤质量等级优势区分布

Fig.6 The distribution of the dominant area of the second cluster raw coal quality grade

霍西煤田白壁关及连村勘查区;中质煤主要分布在沁水煤田西侧以及东侧大部分地区。可以认为,山西省太原组第1聚类中原煤质量相对较好,第2聚类中原煤质量等级相对较差。

整理得到2个聚类簇原煤质量等级面积统计如下(表5)。

#### 4 结 论

1. 首次将山西省六大煤田太原组原煤煤质数据

6

基于聚类分析算法处理得到2类聚类数据,进一步探究两种聚类原煤的质量等级,揭示了第一聚类煤炭质量优于第二聚类。

2. 将2种聚类中不同的原煤质量数据进行展布,显示第一类煤炭资源主要分布在大同煤田潘家窑勘查区、宁武煤田马家湾、曲立勘查区,河东煤田杨家湾、杨家塔、三交详查区以及霍西煤田万安、乔家湾勘查区等地区。第二类煤炭资源主要分布在河东煤田郭家沟、白额以及吉县车城勘查区,西山煤田草庄

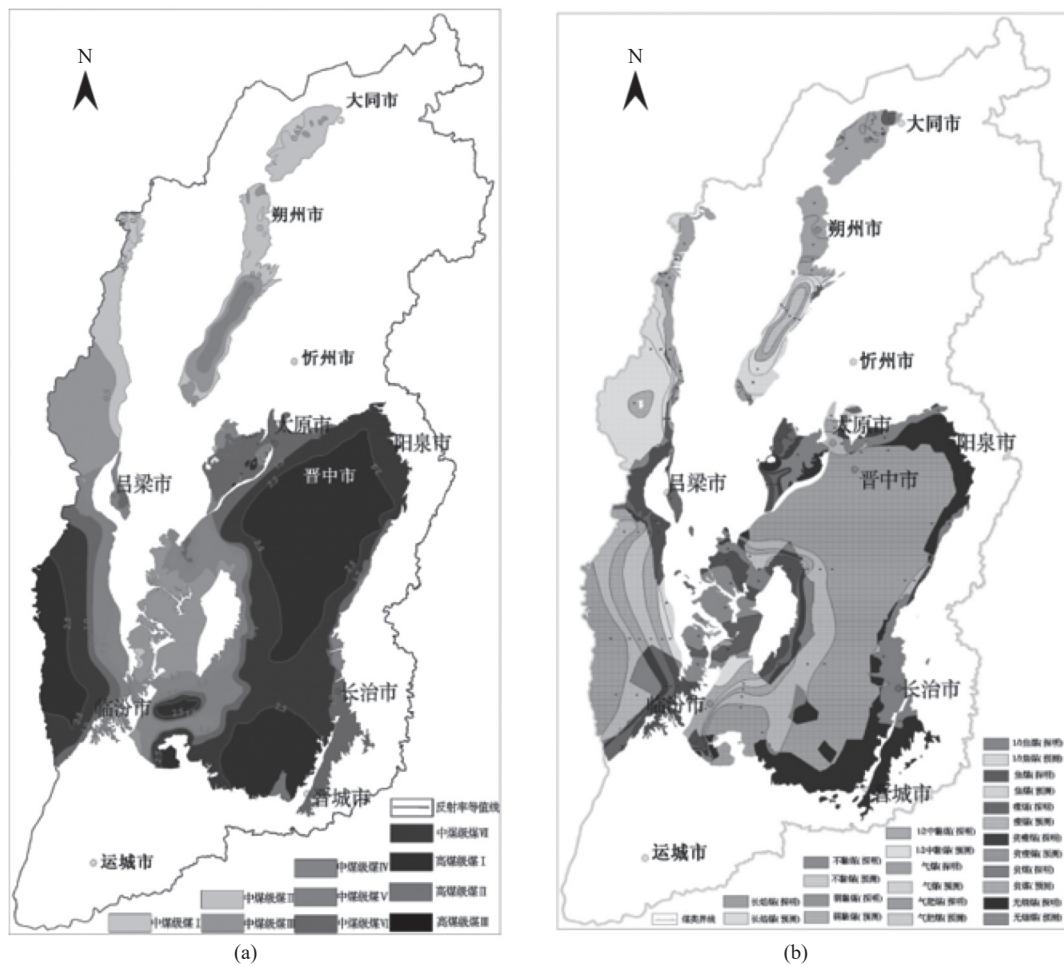


图 7 山西省太原组煤级与煤类分布<sup>[32-34]</sup>

表 5 两个聚类簇原煤质量等级面积统计表  
 Table 5 Two clusters of raw coal quality grade area statistics

等级	第一聚类簇		第二聚类簇	
	面积/km <sup>2</sup>	占比/%	面积/km <sup>2</sup>	占比/%
优质煤	2 346.82	37.40	3 101.54	9.21
中质煤	3 305.70	52.70	5 019.01	14.91
低质煤	467.82	7.50	14 295.67	42.48
劣质煤	149.88	2.40	11 234.67	33.40

头勘查区,霍西煤田白壁关、连村勘查区以及沁水煤田东西两侧大部分地区。

3. 结果证明了数据挖掘算法煤质数据分析的可能性,拓展了煤质数据的使用方式,为煤质数据库使用与发展提供了新的思路。

致谢 由衷感谢山西省煤炭地质勘查研究院、山西地质矿产研究院领导及相关工作人员对本次数据采集工作的帮助。十分感谢实验室师兄师姐几个

月来不辞辛苦现场录入山西省六大煤田煤质数据对本论文的数据支持。

#### 参考文献(References):

- [1] 马小平,代伟.大数据技术在煤炭工业中的研究现状与应用展望[J].*工矿自动化*,2018,44(1): 50–54.  
MA Xiaoping, DAI Wei. Research status and application prospect of big data technology in coal industry[J]. *Industry and Mine Automation*, 2018, 44(1): 50–54.
  - [2] 黄少芳,刘晓鸿.地质大数据应用与地质信息化发展的思考[J].*中国矿业*,2016,25(8): 166–170.  
HUANG Shaofang, LIU Xiaohong. Thinking about the application of geological big data and geological information development[J]. *China Mining Magazine*, 2016, 25(8): 166–170.
  - [3] 李朝奎,严雯英,肖克炎.地质大数据分析与应用模式研究[J].*地质学刊*,2015,39(3): 352–357.  
LI Chaokui, YAN Wenying, XIAO Keyan. Analysis and application mode of geological big data[J]. *Journal of Geology*, 2015, 39(3): 352–357.
  - [4] 俞立平.大数据与大数据经济学[J].*中国软科学*,2013(7): 177–183.  
YU Liping. Big Data and Big Data Economics[J]. *China Soft Science*,

- ence, 2013(7): 177–183.
- [ 5 ] 杨汝岱. 大数据与经济增长[J]. *财经问题研究*, 2018(2): 10–13.  
YANG Rudai. Big data and economic growth[J]. *Research on Financial and Economic Issues*, 2018(2): 10–13.
- [ 6 ] 杨学瑜, 顾合英. 数据挖掘在选煤中的应用探讨[J]. *中国煤炭*, 2004, 30(5): 49–49.  
YANG Xueyu, GU Heying. Discussion on the Application of Data Mining in Coal Preparation[J]. *China Coal*, 2004, 30(5): 49–49.
- [ 7 ] 匡亚莉, 邓建军, 刘怀宇. 人工智能技术在选煤领域的应用[J]. *中国矿业大学学报*, 2001(6): 558–563.  
KUANG Yali, DENG Jianjun, LIU Huaiyu. Application of artificial intelligence technology in the field of coal preparation[J]. *Journal of China University of Mining & Technology*, 2001(6): 558–563.
- [ 8 ] 王若成. 基于数据挖掘的煤矿计算机安全监测体系建立[J]. *煤炭技术*, 2013, 32(1): 124–125.  
WANG Ruocheng. Computer-based Data Mining of Coal Mine Safety Monitoring System is Established[J]. *Coal Technology*, 2013, 32(1): 124–125.
- [ 9 ] 田玉山. 基于数据挖掘技术的煤矿远程监控系统研究[J]. *煤炭技术*, 2012, 31(12): 77–79.  
TIAN Yushan. Research on Remote Monitoring System of Coal Mine Based on Data Mining Technology[J]. *Coal Technology*, 2012, 31(12): 77–79.
- [ 10 ] 王艳亮. 基于数据挖掘的矿山企业智能安全监管系统模型研究[J]. *煤炭技术*, 2011, 30(002): 82–84.  
WANG Yanliang. Intelligent Model of Safety Control for Mining Enterprises Based on Data Mining Control[J]. *Coal Technology*, 2011, 30(002): 82–84.
- [ 11 ] 宫晓曼, 滕荣华. 基于神经网络的数据挖掘在煤矿选煤中的应用[J]. *煤炭技术*, 2013(9): 127–128.  
GONG Xiaoman, TENG Ronghua. Application of Data Mining Based on Neural Network for Coal Quality in Mine Area[J]. *Coal Technology*, 2013(9): 127–128.
- [ 12 ] 吴 乔, 罗 键, 林金有. 煤炭企业生产成本联机分析处理和数据挖掘研究[J]. *工矿自动化*, 2014, 40(4): 10–14.  
WU Qiao, LUO Jian, LIN Jinyou. Research of online analytical processing and data mining for production cost of coal enterprises[J]. *Industry and Mine Automation*, 2014, 40(4): 10–14.
- [ 13 ] KOHONEN T. Self-organized formation of topologically correct featuremaps[J]. *Biological Cybernetics*, 1982, 43(1): 59–69.
- [ 14 ] 范小平, 张立强. 基于SOM的多维信息可视化研究[J]. *应用基础与工程科学学报*, 2011(3): 379–388.  
RUI Xiaoping, ZHANG Liqiang. Visualization of multi-dimensional information based on SOM[J]. *Journal of Basic Science and Engineering*, 2011(3): 379–388.
- [ 15 ] 梁斌梅. 自组织特征映射神经网络的改进及应用研究. [J]计算机工程与应用, 2009(31): 137–138.  
LIANG Binmei. Study on improvement and application of self-organizing map neural network. Computer Engineering and Applications, 2009(31): 137–138.
- [ 16 ] 周 欢, 李广明, 张高煜. SOM+K-means两阶段聚类算法及其应用[J]. *现代电子技术*, 2010, 33(16): 113–116.
- ZHOU Huan, LI Guangming, ZHANG Gaoyu. SOM+K-means Two-phase Clustering Algorithm and Its Application[J]. *Modern Electronics Technique*, 2010, 33(16): 113–116.
- [ 17 ] KOHONEN T, Honkela T. Kohonen network[J]. Scholarpedia, 2007, 2(1): 83–100.
- [ 18 ] KANUNGO, T, Mount, D. M, Netanyahu, N. S. An efficient k-means clustering algorithm: analysis and implementation[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 24(7): 0–892.
- [ 19 ] Hartigan J A, Wong M A. Algorithm AS 136: A k-means clustering Algorithm[J]. *Journal of the Royal Statistical Society. Series C ( Applied Statistics)*, 1979, 28(1): 100–108.
- [ 20 ] 王学军, 王海生, 屈晓荣, 等. 山西省煤炭资源潜力评价[M]. 北京: 煤炭工业出版社, 2017.  
WANG Xuejun, WANG Haisheng, QU Xiaorong, et al. Evaluation of coal resource potential in Shanxi Province[M]. China Coal Industry Publishing House, 2017.
- [ 21 ] 王学军, 张正喜, 王海生, 等. 山西煤质评价及煤的工业利用[M]. 北京: 煤炭工业出版社. 2014.  
WANG Xuejun, ZHANG Zhengxi, WANG Haisheng, et al. Evaluation of coal quality and industrial utilization of coal in Shanxi[M]. China Coal Industry Publishing House, 2014.
- [ 22 ] 张发明. 一种融合SOM与K-means算法的动态信用评价方法及应用[J]. *运筹与管理*, 2014, 23(6): 186–192.  
ZHANG Faming. A dynamic credit evaluation method and application combining SOM and K-means algorithm[J]. *Operations Research and Management Science*, 2014, 23(6): 186–192.
- [ 23 ] GB/T 15224.1-2018, 煤炭质量分级 第一部分: 灰分[S].
- [ 24 ] MTT849-2000煤的挥发分分级标准[S].
- [ 25 ] GB/T 15224.1-2010, 煤炭质量分级 第二部分: 硫分[S].
- [ 26 ] GB/T 15224.1-1994, 煤炭质量分级煤炭灰分分级[S].
- [ 27 ] 许凯玉. 山西省煤炭资源洁净等级划分研究[D]. 北京: 中国矿业大学(北京), 2019.  
XU Kaiyu. Study on Classification of Clean Grade of Coal Resources in Shanxi Province[D]. China University of Mining & Technology, Beijing, 2019.
- [ 28 ] 杨淑婷, 唐跃刚, 解锡超, 等. 煤炭资源洁净等级评价研究[J]. *洁净煤技术*, 2011, 17(1): 5–8,11.  
YANG Shuting, TANG Yuegang, XIE Xichao, et al. Research on clean potential assessment of coal resources[J]. *Clean Coal Technology*, 2011, 17(1): 5–8,11.
- [ 29 ] 杨淑婷. 中国煤炭资源洁净潜势评价研究[D]. 北京: 中国矿业大学(北京), 2015.  
YANG Shuting. Cleaness Potential Evaluation Study on China Coal Resource[D]. China University of Mining & Technology, Beijing, 2015.
- [ 30 ] 陈 鹏. 中国煤中硫的赋存特征及脱硫[J]. *煤炭转化*, 1994, 017(2): 1–9.  
CHEN Peng. The Occurrence Characteristics and Desulfurization of Sulfur in Chinese Coal[J]. *Coal Conversion*, 1994, 017(2): 1–9.
- [ 31 ] 邵龙义等. 山西省石炭二叠纪含煤地层沉积环境与聚煤规律研究[R]. 2011.

- SHAO Longyi. Study on sedimentary environment and coal accumulation law of Carboniferous Permian coal bearing strata in Shanxi Province[R]. 2011.
- [ 32 ] 唐跃刚, 程爱国, 王海生, 等. 山西省太原组和山西组煤质特征分析[J]. 煤炭科学技术, 2013, 41(7): 10–15.
- TANG Yuegang, CHEN Aiguo, WANG Haisheng, *et al.* Coal Quality Characteristic Analysis of Taiyuan Formation and Shanxi Formation in Shanxi Province[J]. Coal Science & Technology, 2013, 41(7): 10–15.
- [ 33 ] 解锡超, 张正喜, 唐跃刚, 等. 山西省太原组和山西组煤的煤岩特征分析[J]. 中国煤炭地质, 2011, 23(8): 66–69.
- XIE Xichao, ZHANG Zhengxi, TANG Yuegang. Taiyuan and Shanxi Formations Coal Petrologic Feature Analyses in Shanxi Province[J]. *Coal Geology of China*, 2011, 23(8): 66–69.
- [ 34 ] 解锡超, 张庆辉, 唐跃刚等. 山西太原组和山西组煤类分布特征及变质作用分析[J]. 中国煤炭地质, 2011, 23(8): 78–81.
- XIE Xichao, ZHANG Zhengxi, TANG Yuegang. Taiyuan and Shanxi Formations Coal Rank Distribution and Metamorphism Analysis in Shanxi Province[J]. *Coal Geology of China*, 2011, 23(8): 78–81.