

回采工作面瓦斯涌出量的聚类分析及 归类预测模型的建立

谢东海^{1,2*}, 何爱宝³

(1. 湖南科技大学 资源环境与安全工程学院, 湖南 湘潭 411201;
2. 湖南科技大学 煤矿安全开采湖南省重点实验室, 湖南 湘潭 411201;
3. 湖南省煤业集团 资兴矿区管理局, 湖南 资兴 423404)

摘要: 为了预防瓦斯灾害的发生, 以实现回采工作面瓦斯涌出量的准确预测, 在对回采工作面瓦斯涌出量实测数据统计的基础上, 应用模糊 C 均值聚类算法进行回采工作面瓦斯涌出量的聚类分析, 得出各数据的聚类中心和样本数据对于分类的隶属度; 建立回采工作面瓦斯涌出量的归类预测模型, 对待测瓦斯涌出样本进行预测, 实现了瓦斯涌出量的归类预测; 最后, 用实例论证了该方法的可行性和有效性. 研究结果表明: 隶属度表征样本属于各个类别的程度, 同时也证明了瓦斯涌出量与各个影响因素之间的关系是非线性的; 采用归类预测模型对待测样本进行归类预测, 通过比较样本与聚类中心的关联度大小, 判定样本的归属类别, 避免了人为的主观性和盲目性; 该方法可行、实用, 是一种有效的瓦斯涌出量归类预测方法.

关键词: 瓦斯涌出量; 回采工作面; 模糊 C 均值聚类算法; 归类预测模型

中图分类号: TD713 **文献标志码:** A **文章编号:** 1672-9102(2017)01-0029-05

Cluster Analysis of the Gas Emission from Coal Face and Establishment of Its Categorization Prediction Model

XIE Donghai^{1,2}, HE Aibao³

(1. School of Resources, Environment and Safety Engineering, Hunan University of Science and Technology, Xiangtan 411201, China;
2. Hunan Provincial Key Laboratory of safe Mining Techniques of Coal Mines, Hunan University of Science and Technology, Xiangtan 411201, China;
3. Zixing Mining Area Management Bureau of Hunan Provincial Coal Group Company, Zixing 423404, China)

Abstract: Based on the statistical data of gas emission from coal faces, the gas emission data are analyzed by using the fuzzy C -mean cluster algorithm, and the cluster centers and the membership degrees of samples are obtained. The categorization prediction model for gas emission from the coal face is established, and is applied to the prediction of the gas emission. The categorization prediction of gas emission is achieved based on the combination of the fuzzy C -mean cluster algorithm with the categorization prediction model. Finally, the method is applied to the practical example. Practical application demonstrates that the relations of the gas emission and its influence factors are nonlinear; when the gas emission samples are predicted by using categorization prediction model, the adscription of each sample is determined by comparing the related degrees. As a result, the artificial subjectivity and blindness is avoided; the method has a good feasibility and a preferable practicability, thus it is an efficient categorization prediction method for gas emission.

收稿日期: 2015-09-21

基金项目: 国家自然科学基金资助项目 (51174086)

* 通信作者, E-mail: donghaixie@163.com

Key words: gas emission; coal mining face; fuzzy C - mean cluster algorithm; categorization prediction model

矿井瓦斯是煤矿安全生产过程中的最不安全因素,其与顶板、火灾、煤尘、水害并称为矿井的五大自然灾害^[1],容易造成群死群伤.据新华网的统计^[2],2014年全国煤矿共发生事故74起、死亡441人,其中瓦斯事故47起,死亡266人,分别占煤矿事故起数和死亡人数的63.5%和60.3%.为此,研究回采工作面瓦斯的涌出规律,对于预防因瓦斯而诱发的煤与瓦斯突出、瓦斯爆炸和瓦斯超限等煤矿事故,具有十分重要的理论意义和现实意义.

工作面瓦斯涌出预测是一个动态变化过程,受多个因素的影响,并且该过程不具有可逆性,这就限制了现场试验研究^[3,4].近年来,随着计算机技术及数学方法在煤矿中应用与发展,基于各种理论建立了许多预测方法,魏引尚等^[5]利用BP神经网络法进行工作面瓦斯涌出量预测,其预测值与实测值基本相当;樊保龙等^[6]提出使用免疫遗传算法预测工作面的瓦斯涌出量,预测值后,其与实测值的平均相对误差2.35%;付华等^[7]采用免疫遗传算法(IGA)优化的加权最小二乘支持向量机(LS-SVM)瓦斯涌出量预测的方法,通过预测后与实测值相比较,其最大相对误差3%;等等.虽然取得了许多进展,但仍未较好地解决这一问题.主要存在回采工作面瓦斯涌出系统的复杂性和人们研究的方向性等2个方面的原因.另外,人们一直致力于瓦斯涌出量值的预测研究,往往容易忽视瓦斯涌出系统的复杂性,其复杂性引发出非线性的问题^[8],靠单一指标就达不到预测的目的.但对于现场工程技术人员来说,越简单实用的预测方法越易于掌握.为此,采用模糊C均值聚类算法对回采工作面已知实测瓦斯涌出量样本进行聚类分析,得到聚类中心,建立归类预测模型,实现回采工作面待测瓦斯涌出量样本的归类预测,通过实测后与预测值进行比较,得出相对误差.这一研究内容对现场工程技术人员准确掌握具有简单易懂.对指导矿井设计和安全生产等,具有理论意义和现实意义.

1 C均值聚类算法

数据集 $X = \{x_1, x_2, \dots, x_n\} \subset R^S$ 为给定的 n 个模式空间中的一组有限观测数据样本集, $x_k = \{x_{k1}, x_{k2}, \dots, x_{ks}\}^T \in R^S$ 为观测数据样本 x_k 的特征矢量,对应特征空间一个点, x_{kj} 为特征矢量 x_k 第 j 特征的赋值,对数据集 X 进行聚类分析就是要产生 X 的 c 个划分.

C均值聚类算法是采用隶属度确定每个样本属于某一聚类程度的一种聚类算法^[9-10].C均值聚类算法采用聚类准则函数为类内加权误差平方和函数,即:

$$J = \sum_{i=1}^c \sum_{k=1}^n (\mu_{ki})^m \|x_k - w_i\|^2. \quad (1)$$

式中, μ_{ki} : 样本 x_k 对于第 i 类的隶属度; m : 表征对比度的参数, $1 < m < \infty$, m 越大,对比度越大; w_i : 第 i 类的聚类中心.并满足如下条件:

$$\mu_{ki} \in [0, 1], \sum_{i=1}^c \mu_{ki} = 1, 0 < \sum_{k=1}^n \mu_{ki} < n. \quad (2)$$

对于式(1),可用最小二乘法求得聚类中的参数 w_i 和 μ_{ki} . 首先分别对 w_i 和 μ_{ki} 求导,并令它们的导数为0,再代入条件式(2),即可得

$$\mu_{ki} = \frac{(1/\|x_k - w_i\|^2)^{\frac{1}{m-1}}}{\sum_{i=1}^c (1/\|x_k - w_i\|^2)^{\frac{1}{m-1}}} \quad (i = 1, 2, \dots, c; k = 1, 2, \dots, n); \quad (3)$$

$$w_i = \frac{\sum_{k=1}^n (\mu_{ki})^m x_k}{\sum_{k=1}^n (\mu_{ki})^m}. \quad (4)$$

C 均值聚类算法的应用步骤如下^[11]:

步骤 1 给定参数 m , 类别数 c , 容许误差 E_{\max} 的值, 令 $t = 1$;

步骤 2 初始化聚类中心: $w_i(1) (i = 1, 2, \dots, c)$;

步骤 3 按式(3)计算隶属度 $\mu_{ki}(t) (i = 1, 2, \dots, c; k = 1, 2, \dots, n)$;

步骤 4 按式(4)修正所有的聚类中心 $w_i(t+1) (i = 1, 2, \dots, c)$;

步骤 5 计算误差:

$$e = \sum_{i=1}^c \|w_i(k+1) - w_i(k)\|^2. \quad (5)$$

如果 $e < E_{\max}$, 则迭代计算结束; 否则, $k+1 \Rightarrow k$, 回到步骤(3)反复进行迭代计算, 直到满足 $e < E_{\max}$ 终止.

2 建立归类预测模型

设待测数据样本 $w_0 = \{w_0(1), w_0(2), \dots, w_0(s)\}$ 为标准数据序列, 由 C 均值聚类算法得出聚类中心序列 $w_i = \{w_i(1), w_i(2), \dots, w_i(s)\} (i = 1, 2, \dots, c)$ 为参评数据序列, 标准数据序列和参评数据序列一起组成评价数据序列 $w = \{w_0, w_1, \dots, w_c\}^T$, 若 w_0 与 w_i 的关联度最大, 则将待测数据样本 w_0 归为第 i 类. 具体步骤如下^[12]:

1) 数据初始化

$$z_i(j) = w_i(j) / \left[\frac{1}{c+1} \sum_{i=0}^c w_i(j) \right] \quad (i = 0, 1, \dots, c; j = 1, 2, \dots, s). \quad (6)$$

式中, $w_i(j)$: 序列 w_i 中第 j 特征上的赋值; 第 $z_i(j)$ 为初始化处理后的数据.

2) 计算评价指标间的关联系数 $\xi_{0i}(j)$

$$\xi_{0i}(j) = \frac{\Delta_{\min} + p\Delta_{\max}}{\Delta_{0i}(j) + p\Delta_{\max}}. \quad (7)$$

式中, $\Delta_{0i} = |w_0(j) - w_i(j)| (i = 0, 1, \dots, c; j = 1, 2, \dots, s)$: 标准序列与第 i 个参评序列对应的第 j 个指标差的绝对值; $\Delta_{\min} = \min \min \Delta_{0i}$: 二级最小差; $\Delta_{\max} = \max \max \Delta_{0i}$: 二级最大差; p : 分辨系数, 取数值区间为 $[0, 1]$, $p = 0.5$ 时有较高的分辨率.

3) 计算排序及关联度 R_{0i}

$$R_{0i} = \frac{1}{s} \sum_{j=1}^s \xi_{0i}(j) \quad (i = 0, 1, \dots, c). \quad (8)$$

若 $R_{0i} = \max \{R_{01}, R_{02}, \dots, R_{0c}\}$, 则将待测数据样本 w_0 归为第 i 类.

3 应用举例

通过多方面的考察, 影响回采工作面瓦斯涌出量的主要因素有: 煤层埋深 H , 煤层厚度 M , 煤层瓦斯含量 Q , 工作面推进速度 V , 工作面平均日产量 S 等 5 个方面. 选取金竹山矿业公司土朱煤矿为研究对象. 该矿井含煤 7 层, 可采 4 层, 为 2~5 层煤. 煤层平均厚度分别为: 2 煤厚度 0.0~3.0 m, 平均厚 1.0 m; 3 煤厚度 0.8~7.0 m, 平均厚 2.0 m; 4 煤厚度 0.0~2.2 m, 平均厚 0.8 m; 5 煤厚度 0.5~3.8 m, 平均厚 1.37 m. 煤层倾角 $50^\circ \sim 90^\circ$, 一水平在 -50 m 水平, 北翼较陡, 南翼较缓; 二水平在 -250 m 水平南翼较缓, 北翼较陡, 且煤层厚度在深部有加大的趋势. 矿井最大相对瓦斯涌出量为 $8.52 \text{ m}^3/\text{min}$, 最大瓦斯压力为 1.53 MPa, 进入二水平的瓦斯涌出量还会加大. 工作面采用炮采机装. 工作面瓦斯预测在 2011 年以来一直没有出现过超值现象, 也没有发生瓦斯事故. 本次研究分析采用该矿的实测资料(如表 1)进行. 首先, 应用 C 均值聚类算法对表 1 中的瓦斯涌出样本进行聚类分析, 将已知的 11 个样本分为 3 类. 各样本的隶属度如表 2 所示, 样本 4, 5, 10 属于第 1 类; 样本 2, 6, 7, 9 属于的 2 类; 样本 1, 3, 8, 11 属于第 3 类. 各类的聚类中心如表 3 所示. 各样本绝

对瓦斯涌出量的实测值:第1类中的样本4,5,10的实测值依次为1.52,1.59,2.47 m³/min;第2类中的样本2,6,7,9的实测值依次为3.18,2.12,3.47,2.77 m³/min;第3类中的样本1,3,8,11的实测值依次为5.87,7.80,6.31,7.21 m³/min.由此可见,C均值聚类数据分析结果与实际工作面观测值的相对误差分别为第1类5.3%,第2类9.5%,第3类8.3%,已能满足在回采工作面进行瓦斯涌出量预测了,具有良好的实用性.

表1 瓦斯涌出样本实测数据

样本	H/m	M/m	Q/(m ³ /t)	v/(m/d)	S/(t/d)
1	604	6.20	4.03	2.64	2 099
2	531	2.90	3.35	3.68	1 369
3	634	6.25	4.80	2.92	2 340
4	408	2.00	1.92	4.42	1 134
5	411	2.00	2.15	4.16	1 067
6	456	2.20	2.40	4.51	1 272
7	563	3.00	3.68	3.53	1 358
8	590	5.90	4.21	2.85	2 157
9	516	2.80	3.22	3.45	1 239
10	432	2.30	2.58	4.67	1 378
11	640	6.30	4.67	2.75	2 221

表2 各样本对聚类中心的隶属度

样本	1	2	3
1	0.000 455	0.001 027	0.998 520
2	0.009 988	0.986 780	0.003 230
3	0.013 580	0.026 823	0.959 600
4	0.875 040	0.116 770	0.008 186
5	0.954 180	0.040 054	0.005 769
6	0.088 396	0.901 660	0.009 943
7	0.056 904	0.909 420	0.033 675
8	0.025 927	0.068 077	0.906 000
9	0.000 829	0.998 960	0.000 210
10	0.840 490	0.145 770	0.013 740
11	0.000 355	0.000 775	0.998 870

表3 各指标值的聚类中心

分类	1	2	3	4	5
1	410.51	2.008 6	2.057 5	4.282 5	1 198.4
2	503.13	2.667 8	3.075 8	3.916 7	1 297.5
3	617.89	6.233 3	4.429 8	2.784 9	2 198.2

表2中可以得知,同一个数据样本和各个类别数据之间都有关联,只是隶属程度不同,这充分说明了回采工作面瓦斯涌出量与影响因素之间具有复杂的非线性关系.在聚类分析的同时,得出了各类数据的聚类中心,这为回采工作面瓦斯涌出量的归类预测打下了一定的技术基础.

在得到聚类中心之后,应用归类预测模型对选取的未知(待测)样本进行预测,预测结果见表4.表4中的未知样本1,2,3的预测类别依次是第2类、第3类、第1类,事后的实际瓦斯涌出量测量值分别为3.88,6.75,1.48 m³/min,都处于各自的聚类中心附近,其相对误差分别为3.47%,3.37%,4.20%,说明预测结果可以用于工程实际.为此,在对已知回采工作面瓦斯涌出数据样本进行聚类分析时,建立了回采工作面瓦斯涌出量的归类预测模型,对回采工作面未知瓦斯涌出数据样本进行归类预测是可行的,预测结果是可靠的.

表4 未知样本的预测类别及实测值

样本	影响因素					关联度			预测类别	实测涌出量/(m ³ /min)
	H/m	M/m	Q/(m ³ /t)	v/(m/d)	S/(t/d)	R ₀₁	R ₀₂	R ₀₃		
1	550	2.9	3.61	4.02	1 495	0.700 600	0.917 596	0.612 441	2	3.88
2	607	6.1	4.34	2.77	2 167	0.503 723	0.606 587	0.942 256	3	6.75
3	420	1.8	2.14	4.13	953	0.975 647	0.801 675	0.532 413	1	1.48

4 结论

1)对于新的回采工作面瓦斯涌出量预测可以采用C均值聚类算法进行.该法基于矿井长期生产过程中实测了大量的瓦斯涌出量数据,通过聚类分析,建立聚类中心和归类预测模型.较好地反映了预测回采

工作面的瓦斯涌出量与各个影响因素之间的非线性关系。

2) 采用归类预测模型对未知样本进行归类预测,通过比较样本与聚类中心的关联度大小,来判定未知样本的归属类别,避免了人为的主观性和盲目性,减少了工作中的误差。

3) 采用 C 均值聚类算法与归类预测模型进行分步预测是可行的。该方法简便,且误差较小,易于现场工程技术人员掌握,具有较好的推广应用价值。

参考文献:

- [1] 国家安全生产监督管理总局,国家煤矿安全监察局.煤矿安全规程[M].北京:煤炭工业出版社,2011.
- [2] 新华网.2014 年全国煤矿事故简况及事故分类[EB/OL].http://news.xinhuanet.com/2015-02/12/c_1114355030.htm.
- [3] 国家安全生产监督管理总局,国家煤矿安全监察局.防治煤与瓦斯突出规定[M].北京:煤炭工业出版社,2009.
- [4] 国家安全生产监督管理总局.AQ1029-2007 煤矿安全控制系统及检测仪器使用管理规定[S].北京:煤炭工业出版社,2007.
- [5] 魏引尚,刘云飞.基于 MonteCarlo 方法改进的 BP 神经网络对回采工作面瓦斯涌出量预测[J].煤炭工程,2014,46(12): 84-87.
- [6] 樊保龙,白春华,李建平.基于 LMD-SVM 的采煤工作面瓦斯涌出量预测[J].采矿与安全工程学报,2013,30(6): 946-952.
- [7] 付华,姜伟,单欣欣.基于耦合算法的煤矿瓦斯涌出量预测模型研究[J].煤炭学报,2012,37(4):654-657.
- [8] 梁冰,秦冰,孙维吉,等.基于综合评判模型的煤与瓦斯突出危险程度预测[J].中国安全生产科学技术,2015,11(7): 129-134.
- [9] Pal N R, Bezdek J C. On Cluster Validity for the Fuzzy C-means Model[J]. IEEE Trans, Fuzzy Systems, 1995,3(3): 370-379.
- [10] Howon C, Jordan B J. On the Optimal Choice of Parameters in a Fuzzy C-means Algorithm[C]//IEEE International conference on fuzzy Systems. Tow and Contry Hotel San Diego, California, 1992: 349-353.
- [11] 肖健华.智能模式识别方法[M].广州:华南理工大学出版社,2006.
- [12] 刘金海,冯涛,袁坚.基于非线性灰色归类模型的岩爆预测方法[J].地下空间与工程学报,2005,1(6):821-824.