Mine Engineering 矿山工程, 2023, 11(2), 142-148
Published Online April 2023 in Hans. https://www.hanspub.org/journal/me
https://doi.org/10.12677/me.2023.112018

Hans汉斯

基于集成学习的煤与瓦斯突出预测研究

张　杰[1]，邓　森[2]

[1]贵州省能源安全技术中心，贵州　贵阳
[2]贵州大学矿业学院，贵州　贵阳

收稿日期：2023年2月28日；录用日期：2023年3月30日；发布日期：2023年4月10日

摘　要

为了提升煤与瓦斯突出事故预测的准确性和可行性，使用主成分分析法对影响煤与瓦斯突出的12个影响因素的原始数据进行降维处理，进而得到包含原始数据85%信息量的8个主成分，以此8个主成分作为输入通过AdaBoost并以单层决策树作为弱分类器进行学习，建立起主成分分析法与AdaBoost相结合的煤与瓦斯突出预测模型。并选取实例利用64组数据为训练样本，16组为预测样本，通过混淆矩阵判断证明模型的稳定性。结果表明：基于AdaBoost算法以单层决策树为弱分类器的预测模型预测精度达到100%，且总体水平稳定，可为安全生产提供理论依据。

关键词

集成学习，决策树，煤与瓦斯突出，预测

Research on Prediction of Coal and Gas Outburst Based on Integrated Learning

Jie Zhang[1], Sen Deng[2]

[1]Guizhou Energy Security Technology Center, Guiyang Guizhou
[2]School of Mining, Guizhou University, Guiyang Guizhou

Received: Feb. 28th, 2023; accepted: Mar. 30th, 2023; published: Apr. 10th, 2023

Abstract

In order to improve the accuracy and feasibility of coal and gas outburst accident prediction, principal component analysis is used to reduce the dimensionality of the original data of 12 factors affecting coal and gas outburst, and then the information content containing 85% of the original data is obtained. The 8 principal components are used as input through Adaboost and the single-layer decision tree is used as a weak classifier to learn, and a coal and gas outburst prediction

文章引用: 张杰, 邓森. 基于集成学习的煤与瓦斯突出预测研究[J]. 矿山工程, 2023, 11(2): 142-148.
DOI: 10.12677/me.2023.112018

model combining principal component analysis and AdaBoost is established. And select examples to use 64 sets of data as training samples and 16 sets as prediction samples, and prove the stability of the model by judging the confusion matrix. The results show that the prediction accuracy of the prediction model based on the AdaBoost algorithm and the single-layer decision tree as the weak classifier reaches 100%, and the overall level is stable, which can provide a theoretical basis for safe production.

Keywords

Ensemble Learning, Decision Tree, Coal and Gas Outburst, Prediction

1. 引言

煤与瓦斯突出是一种类型的瓦斯特殊涌出的现象，即在压力作用下，破碎的煤与瓦斯由煤体内突然向采掘空间大量喷出的现象。煤与瓦斯突出是煤矿井下生产的一种强大的自然灾害，他严重威胁着煤矿的安全生产，具有极大的破坏性。准确的煤与瓦斯突出预测，对于及时撤出人员、减少伤亡具有重要的意义。

因此，如何快速准确地预测瓦斯突出的发生成为矿井安全生产的重中之重，近年来大批专家学者提出了自己的看法，其中周世宁等[1]通过对含瓦斯煤样进行三维受力分析，建立含瓦斯煤样蠕变行为模型，发现若条件具备各类煤层均有突出可能。蒋承林等[2]通过理论分析提出煤与瓦斯突出的失稳假说并通过实验进行了相关验证。匡芳君等[3]通过混沌粒子群算法对混合核支持向量机参数进行优化，建立可准确预测煤与瓦斯突出的改进混沌粒子群与支持向量机结合的模型。李鑫灵等将主成分分析法与支持向量机相结合，结合实例，建立了更为稳定准确的瓦斯涌出预测方法。温建强等[4]通过将 BP 神经网络与灰色理论相连并结合相关实例，得到预测瓦斯含量的模型。周西华等[5]对层次分析法进行改进并与 BP 神经网络结合，建立更为准确的煤与瓦斯突出预测模型。李映洁等[6]采用改进粒子群算法对最小二乘支持向量机进行参数寻优。但上述方法存在一定的局限，虽然在特定数据集预测相对准确，但泛化能力弱，受参数影响大，而研究表明 AdaBoost 泛化能力强，精度高，无需调参，适合二分类和多分类。通过主成分分析法进行简化降低了模型的训练时间，可提升模型运行效率。

基于上述原因，通过将主成分分析法于 AdaBoost 相结合，得到了主成分分析法和 AdaBoost 相结合的煤与瓦斯突出预测模型，通过对原始数据进行主成分分析，将处理后的数据输入以单层决策树为弱模型的 Adaboost 模型对煤与瓦斯突出进行预测，提升了运算效率以及预测精准率可以更好地为安全生产提供指导。

2. 算法原理

神经网络也称为人工神经网络(ANN)，由节点层组成，包含一个输入层、一个或多个隐藏层和一个输出层。每个节点也称为一个人工神经元，它们连接到另一个节点，具有相关的权重和阈值。如果任何单个节点的输出高于指定的阈值，那么该节点将被激活，并将数据发送到网络的下一层。否则，不会将数据传递到网络的下一层。深度学习中的“深度”指的只是神经网络中层的深度。由三层以上组成的神经网络(包含输入和输出)可视为深度学习算法或深度神经网络。只有两层或三层的神经网络只是基本神经

网络，而集成学习指的是通过将多个基学习器结合，通常都会获得比单一学习器显著优越的泛化性能。为了使模型运行更加快捷，使用主成分分析法对原始数据进行处理，主成分分析是对于原先提出的所有变量，将重复的变量删去多余，建立尽可能少的新变量，使得这些新变量是两两不相关的，而且这些新变量在反映课题的信息方面尽可能保持原有的信息。

2.1. 主成分分析法

主成分分析法是机器学习中一种非监督学习方法，该方法利用正交变换将线性相关的若干组数据转换成少数几个由线性无关的变量称为主成分。主成分的个数通常小于原始数据的个数，因此主成份分析法属于降维方法。

主成分分析法计算公式如下所示：假设是 m 维随机变量，其均值向量是其协方差矩阵考虑由 m 维随机变量的线性变换其中求主成分的方法如下第一步，在 x 所有线性变换中，在条件下，求方差最大，得到第二主成分；第 k 步，在与不相关的 x 的所有变换中，在条件下，方差最大，则为第 k 主成分，主成分分析算法适用于对复杂数据进行降维。

2.2. AdaBoost 算法

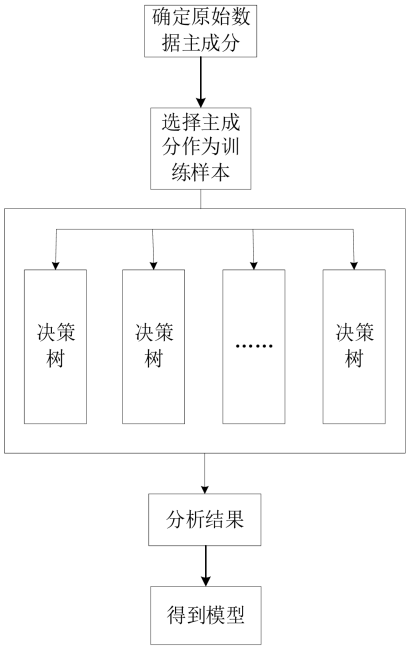

Figure 1. Algorithm flow chart
图 1. 算法流程图

AdaBoost 算法是一种常用的统计学习方法，应用广泛且十分有效，它通过改变训练样本权重，学习多个分类器，将这些分类器进行线性组合，从而达到提高分类效果的目的，AdaBoost 算法的使用条件是必须满足基础模型的准确率要达到 50%以上。

AdaBoost 算法步骤如下[7]：

1) 假设训练数据具有均匀的权值分布，即每个训练样本在基本分类器中具有相同的。输入训练集 $T=\{(x_1,y_1)(x_2,y_2),\cdots,(x_N,y_N)\}$，其中 $x_i \in X$，X 属于实例空间，$y_i \in [-1, 1]$。

2) 初始化训练网络权值分布 $D_1=(w_{1i},\cdots,w_{1N}), w_{1i}=\frac{1}{N}, i=1,2,3,\cdots,N$ 使用具有权值分布 D_m，$m \in 1, 2, 3, \cdots$,

M 的训练数据集学习，得到基本的分类器 $G_m(x)$，并计算 $G_m(x)$ 在训练集上的误差率 $e_m=\sum_{i=1}^{N}P\left(G\left(x_i\right)\neq y_i\right)=\sum_{i=1}^{N}w_{mi}I\left(G_m\left(x_i\neq y_i\right)\right)$，计算 $G_m(x)$ 系数 $a_m=\frac{1}{2}\log\frac{1-e_m}{e_m}$，更新训练集权值分布 $D_{m+1}=\left(w_{m+1,1},\cdots,w_{m+1,i},\cdots,w_{m+1,N}\right)$，$w_{m+1,i}=\frac{w_{mi}}{Z_m}\exp\left(-a_m y_i G_m\left(x_i\right)\right),i=1,2,3,\cdots,N$，这里 Z_m 是规范因子 $Z_m=\sum_{i=1}^{N}\exp\left(-a_m y_i G_m\left(x_i\right)\right)$ 它使 D_{m+1} 成为一个概率分布。

3) 构成基本分类器的线性组合 $f\left(x\right)=\sum_{m=1}^{M}a_m G_m\left(x\right)$，得到最终的分类器 $G_m\left(x\right)=\text{sign}\left(f\left(x\right)\right)=\text{sign}\left(\sum_{m=1}^{M}a_m G_m\left(x\right)\right)$，图 1 为算法流程。

3. 实例分析

3.1. 煤与瓦斯突出的影响因素

通过查阅得到煤与瓦斯突出的主要影响因素：瓦斯含量(A)、瓦斯压力(B)、瓦斯放散初速度(C)、煤的坚固系数(D)、煤层埋深(E)、煤的破坏类型(F)、瓦斯浓度变化率(G)、煤层厚度变化率(H)、顶岩性变化(I)、与地质构造带距离(J)、煤的硬度变化率(K)、最大钻屑量(L)来预测煤与瓦斯突出的发生各影响因素原始数据见表 1 [8]。

Table 1. Raw data of main influencing factors of coal and gas
表 1. 煤与瓦斯主要影响因素原始数据

A	B	C	D	E	F	G	H	I	J	K	L	
7.97	1.26	19.6	0.4	393	3	13.27	23.65	0	1.42	23.96	2.1	0
8.04	105	8.8	0.5	345	4	13.75	2.37	1	4.56	54.56	1.8	0
8.23	1.34	10.1	0.4	489	2	23.42	134.22	1	6.73	15.43	5.9	1
…	…	…	…	…	…	…	…	…	…	…	…	…
10.54	2.15	9.57	0.8	469.03	1	16.27	89.18	1	5.78	73.36	3.68	1
13.71	4.74	6.52	0.6	594.17	2	12.06	13.93	1	5.47	0	7.47	1
10.89	2.32	8.62	0.5	518.45	2	56.39	96.74	0	5.69	15.29	5.05	1
9.51	1.7	5.57	0.7	586.81	1	12.85	141.27	1	5.77	61.33	9.36	1
7.33	0.96	8.03	0.5	314.74	4	12.54	2.16	1	4.16	49.77	1.64	0
4.39	0.52	9.67	0.3	400.49	3	14.82	53.82	0	1.87	0	1.55	0

3.2. 原始数据的主成分分析

对煤与瓦斯突出主要影响因素相关原始数据进行分析。根据表 1 中数据可知共有 12 个特征，对其进行主成分分析，根据上述主成分分析法的原理和方法，计算各成分的方差贡献率及累计贡献，结果如表 2 所示。表 2 为各成分的方差贡献率及累计贡献率，主成分贡献率是指主成分的方差在所考察的随机变量的总方差中所占的比例，主要用以度量主成分对于原变量变异性的解释能力。主成分累积贡献率是选择有效主成分的重要依据。表 3 为主成分分析结果。选取前 q 个特征值的累积百分数大于等于 80%的因子为主成分，依照表 2 结果，选取公共因子 8 个，覆盖将近原信息量的 93%。

Table 2. Variance contribution rate and cumulative contribution rate
表 2. 方差贡献率及累计贡献率

	总计	百分比	累积(%)
1	4.83	40.30	40.30
2	2.05	17.10	57.40
3	1.17	9.74	67.15
4	0.84	7.01	74.17
5	0.68	5.71	79.89
6	0.56	4.68	84.57
7	0.48	4.00	88.58
8	0.42	3.57	92.15
9	0.37	3.12	95.28
10	0.29	2.48	97.76
11	0.15	1.30	99.06
12	0.11	0.93	100

Table 3. Principal component analysis results
表 3. 主成分分析结果

y1	y2	y3	y4	y5	y6	y7	y8
−2.13	−1.52	2.11	−3.37	0.84	−2.78	0.33	2.56
−1.38	6.91	1.24	−0.36	−0.8	−1.84	−0.45	−0.11
1.33	−0.01	0.04	−1.25	−1.28	1.6	4	−0.93
…	…	…	…	…	…	…	…
1.97	0.89	−0.97	−0.22	−1.5	−0.27	−5.45	2.45
3.02	0.64	−0.69	−1.5	−0.76	1.3	−0.26	−3.28
−2.14	−1.5	2.05	−3.32	0.86	−2.74	0.35	2.38
−0.47	0.59	−2.13	0.47	0.81	−0.35	−2.15	0.12
−1.87	−1.23	0.97	0.77	0.35	−0.08	−1.67	−4.14
−1.68	−0.71	0.75	1.31	−3.03	−0.71	1.73	0.86
−2.18	−0.6	0.07	0.61	−2.01	3.9	−0.97	0.53
−2.39	1.36	−4.46	1.18	3.48	1.07	0.12	1.95
−2.24	−1.31	1.69	−3.01	1.04	−2.45	0.48	1.19
−1.35	1.8	−3.4	−1.13	−1.14	−0.8	0.37	−1.74
−1.99	−1.07	0.71	0.8	0.51	0.07	−1.39	−4.87
−1.85	−0.59	0.51	1.33	−2.69	−0.49	1.81	−0.2
2.1	0.85	−0.83	−0.2	−1.71	−0.33	−5.66	2.97
1.68	−1.5	2.02	1.25	1.06	0.9	0.98	1.17
2.11	−1.1	0.94	−1.17	5.72	2.72	1.21	−0.3
1.85	0.54	−0.78	0.48	−0.84	−2.84	−0.42	0.19

Continued

2.11	−1.89	2.03	5.07	1.89	−1.32	−1.24	1.24
1.45	−0.07	0.18	−1.27	−1.45	1.56	4.05	−0.52
2.29	−0.41	1.21	−2.77	−1.88	4.07	−1.2	2
1.5	−0.51	−0.48	3.37	−0.35	−2.58	0.29	−0.48
1.66	0.72	−1.24	0.63	−1.58	−0.46	3.78	0.69
2.35	0.54	−1.03	−2.84	1.2	−2.3	0.07	−1.14
2.06	0.86	−0.87	−0.21	−1.65	−0.31	−5.6	2.81
1.53	−1.42	1.74	1.46	1.15	1.44	1.03	1.68
2.98	0.66	−0.73	−1.5	−0.69	1.29	−0.24	−3.38
2.07	−1.07	0.88	−1.17	5.72	2.71	1.21	−0.42
1.81	0.56	−0.82	0.47	−0.78	−2.8	−0.4	0.06
2.06	−1.87	1.98	5.03	1.92	−1.3	−1.22	1.09
1.42	−0.05	0.13	−1.26	−1.4	1.57	4.03	−0.65
1.62	−0.57	−0.37	3.46	−0.54	−2.65	0.25	−0.04
1.79	0.68	−1.13	0.67	−1.77	−0.54	3.83	1.14
2.47	0.48	−0.89	−2.88	1.02	−2.38	0.03	−0.72
2.2	0.82	−0.74	−0.19	−1.87	−0.36	−5.82	3.33
1.78	−1.55	2.13	1.27	0.99	0.9	0.97	1.5
3.13	0.61	−0.59	−1.5	−0.91	1.3	−0.3	−3.03
−1.37	1.81	−3.42	−1.12	−1.08	−0.77	0.39	−1.87
−2.01	−1.05	0.67	0.81	0.53	0.09	−1.35	−4.97

3.3. 主成分分析与 AdaBoost 的预测模型

将原始数据进行主成分分析，并将得到的主成分作为新的特征输入 AdaBoost 模型进行预测。采用 Python 3.6 进行编写，将单层决策树作为弱分类器，弱分类器数量 $n = 3$，初试情况为均匀分布，即所有样本都为 $1/n$，对每个弱分类器进行训练，更新每个弱分类器的权重，训练完成后采用最终的分类器对数据进行预测。随机抽选 16 组做为预测组，预测结果如表 4 所示。

Table 4. Comparison of predicted results with the real situation
表 4. 预测结果与真实情况的对比

y1	y2	y3	y4	y5	y6	y7	y8	实际值	预测值
−2.39	1.36	−4.46	1.18	3.48	1.07	0.12	1.95	0	0
−1.84	−1.02	0.17	0.91	0.39	1.73	1.65	1.03	0	0
−1.34	13.42	8.75	1.94	2.2	0.97	0.54	−0.02	0	0
3.13	0.61	−0.59	−1.5	−0.91	1.3	−0.3	−3.03	1	1
−2.14	−1.5	2.05	−3.32	0.86	−2.74	0.35	2.38	0	0
1.42	−0.05	0.13	−1.26	−1.4	1.57	4.03	−0.65	1	1
1.53	−0.52	−0.45	3.4	−0.41	−2.6	0.27	−0.35	1	1

Continued

2.07	−1.07	0.88	−1.17	5.72	2.71	1.21	−0.42	1	1
−1.68	−0.71	0.75	1.31	−3.03	−0.71	1.73	0.86	0	0
1.81	0.56	−0.82	0.47	−0.78	−2.8	−0.4	0.06	1	1
1.41	−0.47	−0.56	3.31	−0.23	−2.53	0.31	−0.78	1	1
1.66	0.72	−1.24	0.63	−1.58	−0.46	3.78	0.69	1	1
1.33	−0.01	0.04	−1.25	−1.28	1.6	4	−0.93	1	1
1.62	−0.57	−0.37	3.46	−0.54	−2.65	0.25	−0.04	1	1
−1.35	1.8	−3.4	−1.13	−1.14	−0.8	0.37	−1.74	0	0
1.97	−1.81	1.88	4.93	1.97	−1.24	−1.16	0.75	1	1

由分析可知使用主成分分析法对数据样本进行处理，并基于决策树使用 Adaboost 模型预测样本精度可达 100%，模型具有良好的煤与瓦斯突出预测的准确性，适用于煤与瓦斯突出预测研究

4. 结论

(1) 采用主成分分析法，对 AdaBoost 模型输入进行降维处理，用较少的输入特征代替原有的数据，从而构建的主成分分析法和 AdaBoost 结合的方法，可以提高预测的准确度以及运行速度。

(2) 随机挑选 16 组训练样本数据对改进的 AdaBoost 模型进行学习、训练，并将预测值与实际值对比，正确率为 100%，证明训练完成的改进 AdaBoost 模型具有良好的预测效果。

(3) 采用经主成分分析法改进的 AdaBoost 模型从训练样本中随机选取 16 组作为预测样本，正确率为 100%，证明了基于主成分分析与 AdaBoost 结合的方法用于煤与瓦斯突出预测是可行的，并且预测结果具有良好的准确性。

参考文献

[1] 周世宁, 何学秋. 煤和瓦斯突出机理的流变假说[J]. 中国矿业大学学报, 1990, 19(2): 1-8

[2] 蒋承林, 俞启香. 煤与瓦斯突出机理的球壳失稳假说[J]. 煤矿安全, 1995(2): 17-25.

[3] 匡芳君, 徐蔚鸿, 张思扬. 基于改进混沌粒子群的混合核 SVM 参数优化及应用[J]. 计算机应用研究, 2014, 31(3): 671-674+687.

[4] 温建强, 张岩, 高帅帅, 高望. 基于灰色理论-BP 神经网络预测瓦斯含量[J]. 能源技术与管理, 2020, 45(1): 44-45+55.

[5] 周西华, 郭坤, 白刚, 宋东平. 改进的 AHP 结合 BP 神经预测煤与瓦斯突出[J]. 物探化探计算技术, 2019, 41(1): 121-127.

[6] 李映洁, 杨永国. 基于改进 PSO 优化 LS-SVM 参数的煤与瓦斯突出预测研究[J]. 煤炭技术, 2017, 36(9): 129-131.

[7] 吕晓玲, 宋捷. 大数据挖掘与统计机器学习[M]. 北京: 中国人民大学出版社, 2016: 239.

[8] 高参天. 基于 SFES-PSO-BP 算法的矿井突出预测系统研究[D]: [硕士学位论文]. 厦门: 厦门理工学院, 2019.