

# 大数据挖掘在采煤工作面安全预警中的应用

沈向辉 孟凡强

北京赛迪工业和信息化工程监理中心有限公司 北京 100048

**摘要:** 采煤工作面安全监测与管理过程中积累了大量数据, 这些大数据对安全预警具有重大价值, 但传统数据挖掘方法的局限性, 数据没有很好的得到应用。为了充分利用这些大数据潜在价值, 应用大数据挖掘关联技术, 构建基于大数据挖掘采煤工作面安全预警管理方法, 以工作面多源大数据为基础, 通过挖掘数据间关联关系, 实现采煤工作面安全风险的识别与预警。按照三类危险源的分类方式, 分析和辨识了采煤工作面三类危险源隐患, 而且采用“7W1H”方法对工作面各类隐患数据进行结构转换, 为下一步大数据挖掘奠定基础。论述了大数据在安全预警中所用到技术原理及数据挖掘过程, 尤其详细阐述了安全隐患预警信息多维层次关联规则挖掘方法。最后并以山东某矿E3402综采工作面安全监测数据为例, 通过对数据挖掘获得的强关联规则分析, 研究结果表明了大数据挖掘技术对采煤工作面安全预警的有效性。

**关键词:** 采煤工作面; 安全预警; 大数据; 数据挖掘; 多维层次关联

## 引言

采煤工作面作为煤炭的主要生产场所, 通常在煤矿开采的自然区域及突发条件共同形成的环境中进行作业, 环境中的瓦斯、粉尘、辐射、热、爆炸、火、水、噪音和照明等, 影响着在采煤工作面的作业人员健康, 并会引发事故, 如瓦斯爆炸、火灾、冒顶、透水等灾害, 时刻危及着矿工的生命安全。

近几年, 在学术领域开始了利用数据挖掘方法对煤矿安全监测数据的研究, 如张治斌<sup>[1]</sup>等将关联规则技术应用于煤矿安全监测系统中, 分析出瓦斯是影响煤矿安全主要因素。丁振对工作面环境参数进行关联规则分析, 挖掘参数之间的强规则。对某特定工作面瓦斯浓度进行回归预测分析。贾慧娟等应用关联规则算法中的Apriori算法于煤矿安全预警系统中, 对煤矿安全监测系统数据进行挖掘。因此本文采用大数据挖掘技术, 深入挖掘采煤工作面安全隐患数据的价值, 更好的实现采煤工作面安全预警管理。

## 1 理论基础

预警指的是在警情没有发生时, 就要提前预测、预报, 也就是运用技术、装备及理论的基础上, 通过对事物发展规律的总结、分析和认识, 针对事物现有的状态、特征信息, 从而判断、描述和预测事物的变化趋势, 并与预先制定的目标、指标标准对照, 发现异常情况及时按照预设方式实行预告和警示, 使得预警管理主体能够有时间采取相应对策和处理措施<sup>[1]</sup>, 最大程度减轻危险发生带来的损失或规避风险。采煤工作面安全预警是依据煤矿安全生产法规, 应用现代化技术装备和手

段, 收集工作面各类危险源的隐患数据, 并针对安全生产活动进行评估、审核、整理、分析及监测, 发出不同程度的安全预警信号。预先设定风险管理阈值, 将采集到的隐患数据经处理分析后, 与阈值进行比较, 对不同的控制行为作出决策, 将各种可能转换为风险的隐患提前规避, 或将已发生的风险损失降低到最小。

数据挖掘(Data Mining, DM)从大量的、不完全的、模糊的、随机的数据中, 提取隐含的有价值的信息, 一个重要过程就是从数据中挖掘知识的过程, 又称为数据库中知识发现(Knowledge Discovery in Database, KDD), 是从数据中发现模式的算法的最好的方式。

采煤工作面生产系统中包括以下子系统: 采煤系统、运输系统、通风系统、排水系统、供电系统及液压支架支护系统、安全监测系统<sup>[2]</sup>。

## 2 大数据技术的应用

采煤工作面安全预警大数据技术应用, 重点关注的数据采集以及挖掘分析。由于安全预警大数据结构复杂, 包括结构化、半结构化及非结构化数据, 普通的信息化处理系统难以实现数据的处理, 需要利用大数据处理技术才能对其进行有效存储和处理。

在采煤工作面安全预警大数据挖掘过程中, 通过MapReduce平台批量处理数据, 抽取、转换、加载, 转换成可识别的状态数据, 大量的数据不断积累, 集聚为煤矿安全预警的经验大数据。

## 3 大数据挖掘过程

### 3.1 结构化处理

采煤工作面日常维护、维修以及相关的维护设备材

料、事故处理和管理人员等数据多以文本形式存储，这些非结构化的文本存储着丰富的有价值的信息。数据挖掘模型难以对这些非结构化信息直接处理，需要对这些文本进行结构化处理，形成结构化数据，以便满足数据挖掘要求。本文采用基于规则的大数据平台下非结构化或半结构化数据预处理，主要步骤包括逻辑结构抽取、文档预处理、数据抽取、数据组织等。

### 3.2 数据清洗与集成

在收集到的原始数据通常存在大量的异常数据，如关键指标偏离期望值、关键属性信息缺失、错误类型数据等，这些数据如果不清理将严重影响信息的质量。采用平滑噪声数据、填补遗漏缺失数据、剔除异常数据进行处理，并将数据结构标准化处理，使不同的数据融合到一个集合。

数据集成过程中，以下问题需要考虑：

#### (1) 模式集成问题

实体识别问题，即多个数据源的实体相互匹配，如采煤工作面初始数据库采煤机\_ID与运行数据库中采煤机\_number所指是否为同一实体。解决这个问题可以通过数据库或数据仓库包含的元数据，避免模式集成时发生错误。

#### (2) 冗余问题

在数据集成中，冗余问题经常出现。若一个属性可以从其它属性中推导出来，那这个属性就是冗余属性，属性命名不一致也会导致冗余。利用相关分析，可以判别属性之间的相互关系，如属性X、Y间相互关系可根据下式分析得到。

若

$$r_{X,Y} = \begin{cases} >0 & \text{属性X、Y正关联；} r_{X,Y} \text{越大，二者正关联关系越密；} \\ =0 & \text{属性X、Y相互独立；二者之间没有关系；} \\ <0 & \text{属性X、Y负关联；越大，二者关联关系越密。} \end{cases}$$

#### (3) 数据值冲突检测与消除

对于同一个实体，其来自不同数据源的属性值或许不同。如采煤工作面运输顺槽某处巷道挂网破裂导致片帮，当班安全员记录以挂网脱落面积记录，矿安监人员以片帮脱落的煤体积计算，这样同一条隐患来自采煤工区、矿安监站记录描述不同，出现数据值冲突。

##### 3.1.3 数据选择与变换

数据规范化是将有关属性数据落入到比较小范围之中，目的是消除数值型属性因大小不一而造成挖掘结果的偏差。本文采用最大-最小规范化方法，该方法是对原始数据进行线性变换。假定 $\min_Y$ 和 $\max_Y$ 分别是属性Y的最小值与最大值，最大-最小规范化计算公式为：

$$v' = \frac{v - \min_Y}{\max_Y - \min_Y} (\text{new\_max}_Y - \text{new\_min}_Y) + \text{new\_min}_Y \quad (6)$$

上式中， $v$ 是属性Y的一个值，且 $v \in [\min_Y, \max_Y]$ ； $v'$ 是 $v$ 映射后的值， $v' \in [\text{new\_min}_Y, \text{new\_max}_Y]$ 。

## 4 应用案例分析

以E3402综采工作面安全监测系统2018年1月份报表数据作为研究对象，根据该矿地质构造、煤层特性上分析，选取瓦斯、一氧化碳、风速、温度、日产量5项参数，原始监测数据如下表所示：

表1 原始监测数据

序号	瓦斯浓度	一氧化碳	温度	风速	日产量
	$\text{g.m}^{-3}$	$\text{mg.m}^{-3}$	$^{\circ}\text{C}$	$\text{m.s}^{-1}$	t
1	0.02	0.00	21.25	1.17	910
2	0.01	0.11	17.29	1.38	890
3	0.06	0.20	17.92	1.37	1020
4	0.12	0.01	17.00	1.37	1080
5	0	0.15	16.20	1.39	845
6	0.05	0.17	16.23	1.34	920
7	0.11	0.00	13.90	1.39	980
	...	...	...	...	

应用Apriori算法，先根据预设 $\text{minsup}=25\%$ 找出最大项集，再根据 $\text{mincon}=70\%$ 得到强规则。选择以下5组强规则分析：

表2 数据挖掘分析表

序号	前件	后件	支持度	置信度	提升度
1	U1	V2	0.429	0.751	1
2	V2	W2	0.429	1	1.4
3	V2	X3	0.429	0.867	1.01
4	<U1,V2>	Y1	0.286	1	3.5
5	<U1,V2,W2,X3>	Y1	0.286	0.667	2.3

表2描述为：

(1)  $U1 \Rightarrow V2$ ,  $\text{support}=0.429$ ,  $\text{confidence}=0.751$ ,  $\text{lift}=1$  瓦斯浓度在第1等级 (0—0.05]，推出一氧化碳浓度在第2等级 (0.10—0.15]；

(2)  $V2 \Rightarrow W2$ ,  $\text{support}=0.429$ ,  $\text{confidence}=1$ ,  $\text{lift}=1.4$ ，瓦斯浓度在第2等级 (0.05—0.10]，推出温度在第2等级 (14—20]；

(3)  $V2 \Rightarrow X3$ ,  $\text{support}=0.429$ ,  $\text{confidence}=0.867$ ,  $\text{lift}=1.01$ ，瓦斯浓度在第2等级 (0.05—0.10]，推出风速在第3等级 (1.35—~]；

(4)  $\langle U1,V2 \rangle \Rightarrow Y1$ ,  $\text{support}=0.286$ ,  $\text{confidence}=1$ ,  $\text{lift}=3.5$ ，瓦斯浓度在第1等级 (0—0.05]，一氧化碳浓度在第2等级 (0.10—0.15]，推出煤日产量在第一等级 (0—

900];

(5)  $\langle U1, V2, W2, X3 \rangle \Rightarrow Y1$ , support=0.286, confidence=0.667, lift=2.3, 瓦斯浓度在第1等级(0—0.05], 一氧化碳浓度在第2等级(0.10—0.15], 温度在第2等级(14—20], 风速在第3等级(1.35—~], 推出煤日产量在第一等级(0—900]。

#### 5 对上面关联规则进一步作如下解释、分析

(1) 规则(1)表示瓦斯、一氧化碳二者之间关联, 瓦斯浓度处于第1等级比较低, 说明工作面通风应该比较好, 但一氧化碳浓度处于第2等级, 说明偏高, 可解释为随着采煤过程进行不断氧化产生一氧化碳, 但lift=1, 从理论上说U1、V2相互独立, 没有关联, 这与关联规则以及煤矿生产实际情况都不符。分析原因, 所选研究对象为低瓦斯矿井, 监测系统采集到的瓦斯浓度都很低, 在划分等级时级差不大, 以至于所得结论与分析结果矛盾。结合其他规则分析, 瓦斯浓度、一氧化碳浓度关联性较高<sup>[4]</sup>, 可同步采取措施防治瓦斯、一氧化碳浓度升高。

(2) 规则(2)和(3)表示一氧化碳浓度与工作面温度以及风速之间关系, 工作面温度偏高, 会影响到一氧化碳浓度也升高, 因此在采煤工作面通风要好, 降低工作面温度, 减少一氧化碳的产生与集聚。

规则(4)和(5)表明瓦斯浓度、一氧化碳浓度、风速与工作面煤炭日产量之间关系。煤矿安全生产要求“以风定产”, 意思是煤矿生产过程中必须保证充分的风量供给。该工作面风速在第3等级(1.35—~], 通风状

况良好, 而日产量在第1等级(0—900], 表明产量与通风相匹配, 符合安全生产规定, 同时通风状况好了, 工作面温度低, 瓦斯、一氧化碳浓度也低, 工作面处于安全状态。

#### 6 结论

本文采用“7W1H”方法对采煤工作面隐患数据分析, 并应用大数据挖掘技术实现安全预警, 主要得出以下结论: 根据采煤工作面事故隐患特征实际情况, 可以将非结构化或半结构化隐患记录进行转换, 对每个维度进行了层次分析, 解决了非结构化文本隐患大数据的结构化转换问题, 为采煤工作面安全预警知识发现与大数据挖掘奠定了基础; 应用大数据关联规则挖掘、多维层次规则等数据挖掘技术, 构建安全风险预警方法, 挖掘隐患微观因素间的关联关系, 实现整个工作面安全风险的识别与预警, 对预防煤矿事故的发生起到积极作用。

#### 参考文献:

- [1]申海军. 煤矿采煤工作面安全管理研究. 山西能源学院学报, 2015(3):95-96.
- [2]梁文寿, 徐文焯. 矿山采矿环境影响评价应重点关注的几个问题. 轻工科技, 2010, 26(11):80-81.
- [3]赵斯佳, 骆祖江. 聚类分析模糊综合评判法在矿井突水危险性评价预测中的应用. 中国煤炭地质, 2015, 27(7):82-87.
- [4]张治斌, 姜亚南, 闫俊豪. 关联规则挖掘技术在煤矿安全预警中的应用研究. 煤矿机电, 2009(2):5-6.