

文章编号: 1671-251X(2023)05-0052-07

DOI: 10.13272/j.issn.1671-251x.18097

煤矿安全大数据特征及治理方法体系研究

陈孝慈¹, 李东海²

(1. 常州大学 商学院, 江苏 常州 213000;

2. 内蒙古师范大学 政府管理学院, 内蒙古 呼和浩特 010022)

摘要: 高效分析利用煤矿安全大数据, 对于提高煤矿的安全管理水平和生产效率具有重要意义。目前煤矿安全大数据治理存在数据特征不明、治理方法不清等问题, 针对该问题, 着重分析了煤矿安全大数据特征, 得出煤矿安全大数据具有 5V 特征, 即数据体量大 (Volume)、数据类型多 (Variety)、处理速度快 (Velocity)、价值密度低 (Value)、真实性 (Veracity), 此外还具有结构化程度不一致的特征。介绍了可应用于煤矿安全管理中的主要数据治理方法及模型, 并分为单变量方法、多变量统计分析方法、智能模式识别方法、系统动力学模型和综合集成模型五大类。从主体和客体的视角, 提出了煤矿安全大数据治理方法体系, 认为数据治理方法的选择必须与智慧矿山的主体、客体的数据治理目标相契合。基于主体的治理方法选择: 根据数据主体的需求、层次、担负的任务及安全管理目标确定数据治理具体内容; 基于客体的治理方法选择: 根据客体对象的时效性、吞吐量要求及安全管理目标确定数据治理具体内容。最后得出煤矿安全大数据治理方法的确定需要在统一目标和统一标准下, 根据作用域和作用对象的不同, 满足共性和个性需求。

关键词: 智慧矿山; 煤矿安全大数据; 数据特征; 数据治理目标; 数据治理方法选择; 时效性; 吞吐量

中图分类号: TD67

文献标志码: A

Research on the coal mine safety big data features and governance method system

CHEN Xiaoci¹, LI Donghai²

(1. Business School, Changzhou University, Changzhou 213000, China;

2. School of Government Management, Inner Mongolia Normal University, Hohhot 010022, China)

Abstract: Efficient analysis and utilization of coal mine safety big data is of great significance for improving the safety management level and production efficiency of coal mines. At present, there are some problems in coal mine safety big data governance, such as unclear data features and governance methods. In order to solve the above problems, this paper emphatically analyzes the features of coal mine safety big data. It is concluded that coal mine safety big data has 5V features, namely, large data volume (Volume), multiple data varieties (Variety), fast processing velocity (Velocity), low value density (Value), veracity (Veracity), and also has the features of inconsistent structural degree. This paper introduces the main data governance methods and models that can be applied to coal mine safety management. The methods are divided into five categories: single variable method, multivariate statistical analysis method, intelligent pattern recognition method, system dynamics model and comprehensive integration model. From the perspective of the subject and object, this paper puts forward a big data governance method system for coal mine safety. It is believed that the selection of data governance methods must be consistent with the data governance goals of the subject and object of intelligent mines. Selection of

收稿日期: 2023-04-03; 修回日期: 2023-04-25; 责任编辑: 胡娴。

基金项目: 国家自然科学基金项目 (42261032); 内蒙古自然科学基金项目 (2022LHQN07003)。

作者简介: 陈孝慈 (1991—), 男, 浙江宁波人, 讲师, 博士, 主要研究方向为可视化管理、知识可视化等, E-mail: nb.chenxc@qq.com。

通信作者: 李东海 (1984—), 男, 山西大同人, 讲师, 博士, 主要研究方向为能源经济, E-mail: donghai.1984@163.com。

引用格式: 陈孝慈, 李东海. 煤矿安全大数据特征及治理方法体系研究[J]. 工矿自动化, 2023, 49(5): 52-58.

CHEN Xiaoci, LI Donghai. Research on the coal mine safety big data features and governance method system[J]. Journal of Mine Automation, 2023, 49(5): 52-58.



扫码移动阅读

subject-based governance methods: the specific content of data governance is determined according to the needs, levels, tasks and security management objectives of data subjects. Selection of object-based governance methods: the specific content of data governance is determined according to the timeliness of object objects, throughput requirements and security management objectives. Finally, it is concluded that the determination of coal mine safety big data governance method needs to meet the common and individual needs according to the different scopes and objects under the unified goal and standard.

Key words: intelligent mine; coal mine safety big data; data feature; data governance objective; data governance method selection; timeliness; throughput

0 引言

自20世纪末以来,矿山信息化建设经历了从数字矿山、感知矿山到智慧矿山的发展过程^[1-3]。以物联网、云计算、大数据、移动互联网等为代表的先进信息技术在煤炭企业广泛应用,大大提高了企业的科技水平,丰富了企业的信息获取手段,同时也对传统的管理模式造成极大的冲击。相关研究表明,在煤矿的生产作业过程中,受到来自于多源对象、多时空尺度、多专业领域的700多种安全威胁,如果以数据量来计算,每天产生的数据将超过300 GiB^[4]。煤矿安全数据已然脱离了人们对数据的传统认知,进入大数据的范畴。高效分析利用煤矿安全大数据,对于提高煤矿的安全管理水平和生产效率具有重要意义。

目前,煤炭企业对安全数据的应用大多止步于基本的统计分析,更多的安全知识淹没于海量数据中,数据丰富但知识贫乏的现象十分严重,煤炭企业信息处理加工能力的不足日益成为制约企业安全管理水平进一步提升的瓶颈^[5]。本文分析了煤矿安全数据特征,阐明了大数据情境下煤矿安全数据的特点,构建了一套完整的数据治理方法体系,以期煤矿安全大数据治理实践提供理论指导。

1 煤矿安全大数据特征

目前,大数据仍然是一个宽泛的概念,尚无明确的定义,但大数据的4V特征已经获得广泛共识^[6-8],即数据体量大(Volume)、数据类型多(Variety)、处理速度快(Velocity)、价值密度低(Value)。部分学者认为在4V的基础上,大数据还应当具有真实性(Veracity)^[9]。

煤矿安全数据是煤矿数据的有机组成部分^[10]。按照数据来源划分,可以将煤矿安全数据划分为自动感知数据和人工录入数据两大类。自动感知数据即通过传感器、通信网络自动获取的数据,一般为瓦斯浓度、顶板压力、地压等环境监测数据。人工录

入数据即通过人工巡查补录自动感知盲区的数据,如安全问题记录、制度法规等。多种来源、多种类型的煤矿安全数据共同构成了煤矿安全大数据,煤矿安全大数据也符合大数据的5V特征^[11-12]。

1) 数据体量大。在煤炭生产中产生了海量安全数据,主要包括以瓦斯浓度、粉尘浓度、矿井温度、矿井地压、水量、风量等为代表的环境监测信息,以人员定位、设备安置地点等为代表的空间信息,以采煤量、机电监测数据等为代表的生产作业信息,以安全隐患、设备点检数据等为代表的隐患信息,以及其他各种综合自动化、管理信息化和工程数字化系统所产生的数据。随着信息技术应用的深化,煤矿安全大数据的数据规模仍在不断增长。

2) 数据类型多。在煤炭生产中所产生的数据类型繁多,具有多源异构的特征。其中,既包括以数值型数据为代表的结构化数据,如通过传感器采集的井下环境信息、生产作业信息、人员定位信息,也包括以文本、图像视频数据等为代表的半结构化和非结构化数据,如通过人工录入的安全隐患文本及通过视频摄像获取的井下实况。

3) 数据处理速度快。煤矿安全监测本身有对数据实时性的要求,随着物联网、云计算、移动互联网等信息技术的应用,传统的信息壁垒已被打破,煤矿安全管理的各系统之间不再相互独立,系统不间断地产生大量数据,这些数据被快速处理、分析,形成可被接受的知识。

4) 价值密度低。相对于庞大的数据总量,数据中所蕴含的有效信息极为稀少,从海量数据中提取有效信息的过程无异于大海捞针,有效信息的提取需要耗费大量的资源。同时,考虑到信息的时间价值,信息蕴含的价值一直处于不断减少中,这与数据的快速增加形成了鲜明对比。

5) 数据真实性。一方面,煤矿安全大数据要求整体上的真实性,即全部或大部分数据必须真实有效,具有较高的信噪比,能够从中发现有价值的信息,并应用于生产实践中;另一方面,允许存在一定

程度的数据缺失、损坏等,如某一个或几个传感器在一段时间内未能上传数据或采集的数据具有明显错误、安全隐患记录的表述不一致等。

煤矿安全大数据除了具有大数据的 5V 特征,还有一个特征也值得注意,即数据的结构化程度。目前主要的数据分析处理技术一般源自信息科学或计算机科学领域,设计之初的处理对象是结构化数据。而一般认为,未来煤矿安全数据增长中的大部分将是非结构化或半结构化的数据,原因如下:

1) 在煤炭企业信息化建设的早期,企业的财务系统、库存系统、供应链系统等一般建设得比较完善,这些系统的数据大多是结构化数据,半结构化数据和非结构化数据的占比极小。随着煤矿信息化建设的深入,智能化管理、辅助决策系统的占比越来越大,因此产生的半结构化和非结构化数据也就越来越多。

2) 由于数据类型的不同所导致。对煤炭生产而言,数值型数据一般可以归结为结构化数据,文本、音频、图像等数据可以归结为半结构化或非结构化数据。一般情况下,1 个数字占用 1 byte,1 个字母占用 2 bytes,1 个汉字占用 2~3 bytes,而一般常见的 720P 视频(1 080×720)的码率一般为 5 000 Kibit/s。由此可看出半结构化或非结构化数据本身会占用较大的存储空间。

3) 结构化数据往往是对事物关键属性的描述,半结构化或非结构化数据则能更好地还原现场信息,这对于煤矿安全管理非常重要。例如,对于同一个安全隐患的描述,结构化的数值型数据所能提供的信息往往是非常有限的,现场人员的文字或语音记录能够更清晰地描述现场,而图形、视频具有最好的现场还原效果。因此,煤炭企业在实际生产中常以文字形式描述安全问题,在条件允许的情况下,则进一步提供图像或视频资料。

数据结构化层次如图 1 所示。图 1 中不同的数据是相互隔离的,但是现实中的数据往往是混杂在一起的,数据之间并没有明显界限。以安全隐患数据为例,安全隐患管理系统在建立时,虽然在安全隐患的“时间”“地点”“报告人”等属性中会设定统一格式的预设值,但在对安全隐患的详情描述中,并没有硬性规定。因此,安全隐患的呈报人员在向系统中输入安全隐患文本时,主要考虑的是将现实的安全隐患以最还原的方式呈递,使用更加口语化的方式来描述安全隐患。这样,在全部的安全隐患文本中,数据的结构化程度极为不一致。综上所述,这些复杂混合的煤矿安全数据处理难度极大,鲜有企业

能够有效利用,但同时这类数据也积累极多,数据极为丰富。

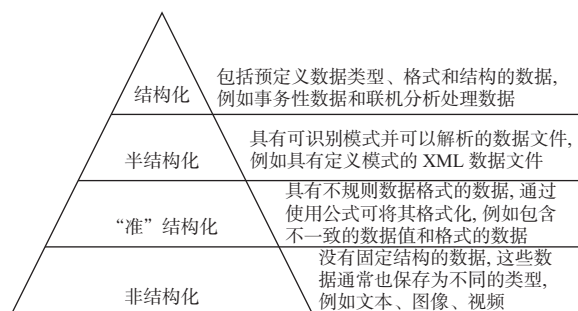


图 1 数据结构化层次

Fig. 1 Data structured hierarchy

2 煤矿安全大数据治理方法体系

2.1 主要数据治理方法

煤矿安全管理主要涉及煤矿安全风险辨识、隐患监测和安全预警等,初步统计,目前可以应用于煤矿安全管理中的主要数据治理方法及模型有 30 多种^[13-18],包括主成分分析(Principal Component Analysis, PCA)、判别分析、深度学习、系统动力学、支持向量机(Support Vector Machine, SVM)、模糊综合评判(Fuzzy Comprehensive Evaluation, FCE)、自回归滑动平均模型(Auto-Regressive Moving Average Model, ARMA)、层次分析法(Antalytic Hierarchy Process, AHP)、风险价值(Value at Risk, VAR)模型、Logistic 回归模型、自回归条件异方差(AutoRegressive Conditional Heteroskedasticity, ARCH)模型、灰色预测模型、粗糙集(Rough Set, RS)模型、模式识别模型、可拓模型、STV 横截面回归模型、FR 概率模型、KLR 信号分析法等。

依据各类数据治理方法和模型的特点,可以将数据治理方法及模型大致分为单变量方法、多变量统计分析方法、智能模式识别方法、系统动力学模型和综合集成模型五大类。

1) 单变量方法。该方法基于少数指标的变化状态反映分析对象状态的变化,其基本特征:需要建立一个起中介作用的新指标,再将这一指标的值和类似对象的历史数据综合对比,最终确定对象状态;所选用指标的权重在实际分析时不存在差异。该方法一般只能保证一定的短期效果,基本不适用于非线性系统的分析,已经被逐步淘汰。

2) 多变量统计分析方法。与单变量方法相对,多变量判别方法往往采用多个指标作为自变量,其特点是以已有样本的统计分析结果为依据,发掘其中的规律,从而判定待检样本的状态。多变量分析

方法弥补了单变量方法的不足,增强了准确性和适用性,得到了广泛应用。判别分析方法、PCA、各类回归分析方法等都是典型的多变量统计分析方法。

3) 智能模式识别方法。该方法基于计算机自动分析样本特征,从而智能高效地将具有类似特征或相同属性的样本划分至特定的类别。智能模式识别与人工智能具有高度的相关性,本质上说,从简单的线性回归方法到以 SVM、RS、概率模型等为代表的机器学习算法和以神经网络为核心的深度学习算法都可以归纳至智能模式识别的范畴。

4) 系统动力学模型。系统动力学是研究系统内外部动态行为的一种仿真方法,适用于非线性、高阶次、多变量、多重反馈的复杂系统。系统动力学集合了系统论、控制论及信息论的精髓,主要从系统本身结构和系统内各要素的相互作用来解释或预测系统的行为,构建完善的系统动力学模型往往需要经历多次迭代。

5) 综合集成模型。由于各类模型具有各自的优缺点和适用情境,将多个辨识方法或模型通过集成技术融合在一起,既可以拓宽模型的应用场景,还能起到扬长避短的作用。综合集成模型的发展趋势:与人工智能技术深度结合,以充分发挥计算机的智能优势;建立更加全面的指标体系,避免单一指标的片面影响。

相关数据处理方法和模型也可以根据数值、文本、音频、图像、视频等数据类型进行初步划分^[19]:

1) 以数值型数据的处理为基础,将文本、音频等转换为数值数据。随着计算机技术的发展,数值处理方法已经成熟,数据的量已不再被计算能力掣肘,对于高维、多元数据也已经存在经过检验的合理有效的处理模型。相关模型可用于数据的异常检测、差异识别、波动预测等。

2) 图像、视频的处理,基本过程是特征提取→特征分析→特征对比→特征识别,即从图像、视频中以固定的范围获取特征,再利用模型对比分析。以图像为对象,则多次提取多个固定范围的特征;以视频为对象,则按固定时间频率截取图像流,重复图像特征提取过程。常应用于处理图像、视频的方法是深度学习。

需要指出的是,相关模型的类别划分并不一定具有清晰的界限。以 SVM 为例,既是多变量统计分析方法,又可以归属为早期的智能模式识别方法。从数据类型的角度来看,在相当长的时间内, SVM 用于图像、视频识别的准确率不高,但在图像识别的早期探索中其所起的作用不可忽视。自深度学习方法

获得广泛应用后,图像、视频的分析处理效率得到了极大提高。图像、视频分析可用于危险源识别、视频监控、安全场景分析。实际上,人工智能技术建立在深度学习基础上,具有广泛的适用性,体现出极大的发展潜力,煤矿安全大数据技术也会更多地向智能化迈进。

2.2 数据治理方法选择

煤矿安全大数据技术服务于智慧矿山整体建设,智慧矿山在本质上是由意识世界、物理世界和信息世界构成的多维复杂智能系统。因此,数据治理方法的选择必须与智慧矿山的主体、客体的数据治理目标相契合。

2.2.1 基于主体的治理方法选择

智慧矿山可以分解为涉及煤矿开发的一切事物及活动的“活动层”、包含整个煤炭产品生命周期的“价值链”和涉及各个组织层级的“组织级”,每一个维度所对应的智慧矿山在不同层次的主体,其本质是维系矿山运作的人^[20]。不同主体在智慧矿山中担负的任务不同,因而其主要目标也不尽相同。根据不同的主体,最终的数据治理目标可以较为清晰地界定。

基于主体的数据治理方法选择过程如图 2 所示。主要步骤:① 目标定位。实施数据治理的前置条件是必须完成对主体目标的确定,最初的主体目标应当是整个智慧矿山运行实施中所设定的整体性目标,这一目标会在层级链三大维度上形成映射。② 目标分解。基于层级链参考模型汇聚不同主体的目标需求,管理者能够清晰地确定这个整体目标在智慧矿山中的定位,在层级链参考模型中定位目标后,可以将目标分解至对应的维度,形成层目标、链目标和级目标,三类目标实质上反映了不同群体需求的最大公约。③ 方法确定。基于三类目标衍生出了不同的判别标准,再结合煤矿整体的原生标准,形成方法过滤器,进而评判全部备选方法,最后选用最适合的方法。

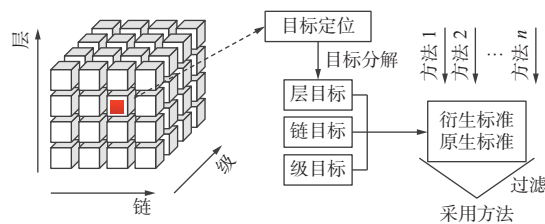


图2 基于主体的数据治理方法选择过程

Fig. 2 Subject-based selection process of data governance methods

煤矿安全大数据的治理是矿业科学与数据科学深度融合的交叉研究领域,要求数据治理主体不仅

具备矿业领域专业性知识,还应当具有科学的数据治理思维。从组织架构层面来看,需要在首席数据官(Chief Data Officer, CDO)的主导下,设立专业的数据治理部门,统一数据治理意识,协调多元主体综合治理。

2.2.2 基于客体的治理方法选择

智慧矿山客体即煤矿安全数据治理的对象,广义上说,一切与煤矿安全相关的数据都属于数据治理客体,但是部分数据来源于煤矿之外,时效性差且处理难度大,难以用于数据治理过程中制定有效的数据采集策略、数据存储策略和数据开发策略。因此,煤矿安全数据治理的对象应限定于煤矿范围内全域全量数据、与数据相关的技术和业务。

时效性和吞吐量是数据治理的两大主要指标,大多数应用场景的煤矿安全数据都可以依此划分为热数据、温数据和冷数据^[21],如图 3 所示。煤矿日常生产作业产生的监测、感知和控制数据是典型的热数据,其体现出数据元素联机实时、数据顺序随机、数据大小不限及数据超时易损的特性,对这些数据的治理工作本身是一个连续的过程,必须保证相关方法的快速、高效、健壮。温数据一般是经营管理和物料监测数据,这些数据大部分并非来自生产作业一线,但数据异常往往是不安全的征兆。冷数据则是对应煤炭企业存储的历史数据,这类数据一般用于查看,而不需要进行处理。

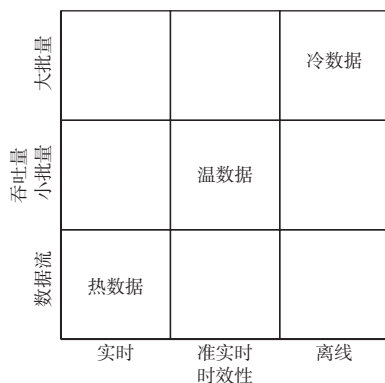


图 3 煤矿安全数据分类

Fig. 3 Classification of coal mine safety data

基于客体的治理方法选择过程如图 4 所示。先根据当前数据治理目标进行方法的初筛,再根据所要治理对象的性质进行进一步筛选,最终采用的方法必须满足数据时效性和吞吐量的需求。

智慧矿山客体划分应尽可能实现“物以类聚”,从而减少客体内部的个体差异,避免源数据表述存在差异。同时,煤矿安全大数据构成复杂,来源多样,而方法模型本身存在适用性限制,因此客体的划

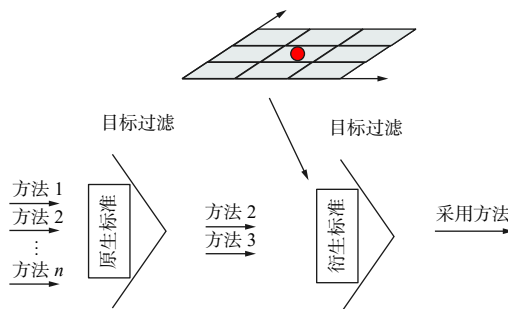


图 4 基于客体的数据治理方法选择过程

Fig. 4 Object-based selection process of data governance methods

分显得尤为重要。

2.3 煤矿安全大数据治理方法体系逻辑框架

综合分析 2 种数据治理思想可以发现,采用什么方法来治理数据,其关键在于标准(包括原生标准和衍生标准)的建立。其中,原生标准是在大的集团、企业层面的共性需求,而衍生标准是针对主体或客体的个性化需求。事实上,在数据治理的实践中,2 种思想并不冲突。虽然不同的视角下数据治理的层次、主要需求等可能有所不同,但在方法论层次上,所有活动仍有其共性。煤矿安全大数据治理方法体系逻辑框架如图 5 所示。

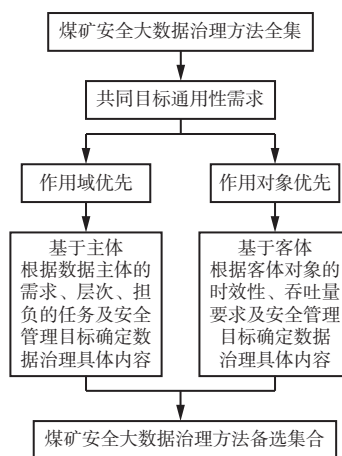


图 5 煤矿安全大数据治理方法体系逻辑框架

Fig. 5 Logical framework of the big data governance method system for coal mine safety

2.4 煤矿安全大数据治理方法体系应用示例

煤炭企业在智慧矿山建设中获取了大量的数据,其中图像视频数据量大、处理难度大,一直以来未能得到有效利用。以解决智能矿山建设中的图像识别问题为例,分析煤矿安全大数据治理方法体系的现实应用,如图 6 所示。

假设某煤炭企业能够在多种场景中部署应用 SVM、人工神经网络(Artificial Neural Network, ANN)、AHP、RS 和 ARMA 模型,那么企业的数据治理方法全集就由上述模型构成。假设在工作面巷道

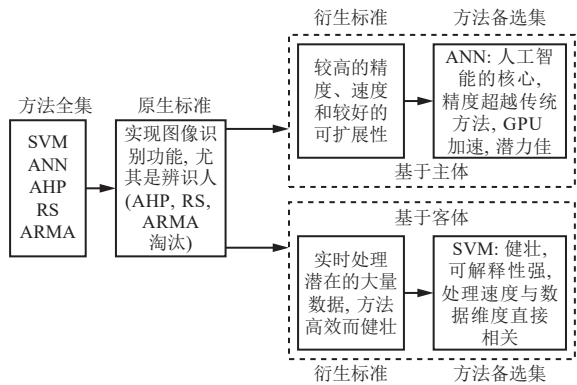


图6 煤矿安全大数据治理方法体系应用示例

Fig. 6 Application of the big data governance approach system for coal mine safety

入口设置有视频监控设备,企业的数
据治理目标是能够快速有效地识别工作
面人员的出入,获取工作面作业人员数
量,那么原生标准即可定义为“实现图
像识别功能,尤其是辨识人”。在方法
全集中,只有SVM和ANN能够满足需求。
SVM和ANN的主要特性见表1。

表1 SVM和ANN的主要特性
Table 1 The main features of SVM and ANN

主要特性	SVM	ANN
数据要求	适用于一定规模的高维数据	适用于大规模的数据集
解释性	结果易解释	结果较难解释
鲁棒性	对噪声和离群点具有一定鲁棒性	对噪声和离群点较敏感
精度	精度较高,与特征选取相关	精度高,与训练数据相关

从数据主体的角度看,对于工作面出入人员的
辨识,应当属于作业与管理层、矿井级和开采链。
从层的角度看,要求识别达到一定的标准;从级的角
度看,识别效果最好能够在未来不断提高,有与其
他系统联动的潜力;从链的角度看,识别需要达
到一定的速度。因此,基于主体的衍生标准即可
定义为“较高的精度、速度和较好的可扩展性”。
由表1可知,ANN更加符合该标准。

从数据客体的角度看,工作面作业的合规与安
全生产息息相关,因此需处理的数据是典型的热
数据,数据量的大小与系统监测点的设置相关。
但是,由于矿井环境所限,相关图像可能清晰
度不足,数据噪声较大。另外,图像数据在构成
上较类似,即数据的维度在达到一定数据量后
不再会显著增长。因此,基于客体的衍生标准
即可定义为“实时处理一定量的数据,方法要
高效而健壮”。由表1可知,SVM更加符合该
标准。

结合数据主体和客体2个方面,可得到针对图
像识别问题的备选方法集合。

3 煤矿安全大数据治理策略

1) 注重顶层设计,明确治理目的。顶层设
计必须从行业层面出发,形成行业层面的统
一标准、共通共享的数据模型、性能优良
的数据算法和成熟可靠的培训体系,在高
层管理者的统一管理下共同服务于煤矿安
全这一大目标。其中,标准贯穿始终,是煤
炭行业内安全大数据治理的共同准则。数
据模型是数据分析的基本框架。数据算法
作为安全数据处理的工具,应当科学有效。
培训体系则是为了提升煤炭行业从业人
员的整体素质。

2) 增进数据共享,实现协同作业。作为煤
炭行业转型升级的重要环节,煤矿安全大
数据治理必然牵涉数据、网络、模型架构
等一系列计算机资源,标准、规范、制度
等一系列管理资源,以及供-产-销协同
联动和人-财-物协同配置。因此,进行
煤矿安全大数据治理时,必须坚持以数据
、技术、业务等核心资源的沉淀和复用及
核心能力的共享和价值创造为依据,综合
管控,脱离传统安全管理局部管理的理
念,解决信息化不完全带来的信息孤岛问
题。

3) 推动国家领导,落实层次责任。煤矿
安全大数据建设是一项长期、复杂的系统
性工程。从国家层面看,煤矿安全大数据
治理对保障国家能源数据安全、实现现代
化能源体系建设等具有重要意义,国家
应加强建设跨区域的信息化平台,推动数
据开放和共享,建设完善的信息化标准体
系。从行业层面看,煤炭行业的安全大数
据建设尚处于各自为政状态,行业缺乏科
学的理论探索,企业间缺少深层交流,需
要以煤炭行业协会和学会为主导,担负起
科学研究和实践探索的责任,打通企业间
联系,同时为国家层面的安全大数据建设
出谋划策。从企业层面看,当前各类大数
据平台层出不穷,但是不同平台间数据不
通、标准不一,建设质量难以保证,企业
应贯彻煤矿数据治理体系和煤矿数据标
准,与科研机构紧密合作,因地制宜展开
企业级安全数据治理,联合构建通用的数
据平台。

4 结语

煤矿大数据的应用尚处于萌芽状态,但其
拥有广阔的前景,煤矿安全大数据颠覆性
地重塑了现实世界的煤矿安全与信息世界
的智慧矿山建设的关系,使人们得以从多
种视角分析安全问题产生的根本原因,具
有科学性、先进性和前瞻性的特点。本
文分析了煤矿安全大数据特征,梳理了主
要的数据分析方法模型,结合智慧矿山的
层级链参考模型,提出

了基于主体和基于客体的数据治理方法选择模型,指出煤矿安全大数据治理必须在统一标准之下,结合国家、企业的实际需求,解决现实存在的数据问题。

参考文献(References):

- [1] 谭章禄,王美君. 智慧矿山数据治理概念内涵、发展目标与关键技术[J]. 工矿自动化, 2022, 48(5): 6-14.
TAN Zhanglu, WANG Meijun. Research on the concept connotation, development goal and key technologies of data governance for smart mine[J]. Journal of Mine Automation, 2022, 48(5): 6-14.
- [2] 谭章禄,吴琦. 智慧矿山理论与关键技术探析[J]. 中国煤炭, 2019, 45(10): 30-40.
TAN Zhanglu, WU Qi. Analysis and discussion of smart mine theory and key technologies[J]. China Coal, 2019, 45(10): 30-40.
- [3] 王国法,王虹,任怀伟,等. 智慧煤矿2025情景目标和发展路径[J]. 煤炭学报, 2018, 43(2): 295-305.
WANG Guofa, WANG Hong, REN Huaiwei, et al. 2025 scenarios and development path of intelligent coal mine[J]. Journal of China Coal Society, 2018, 43(2): 295-305.
- [4] 王龙康. 煤矿安全隐患层次分析与预警方法研究[D]. 北京:中国矿业大学(北京), 2015.
WANG Longkang. The hierarchy analysis for hidden danger and research on early warning method in coal mine[D]. Beijing: China University of Mining & Technology-Beijing, 2015.
- [5] 陈孝慈. 煤矿安全隐患管理知识发现研究[D]. 北京:中国矿业大学(北京), 2021.
CHEN Xiaoci. Research on knowledge discovery of coal mine safety hidden peril management[D]. Beijing: China University of Mining & Technology-Beijing, 2021.
- [6] PROVOST F, FAWCETT T. Data science and its relationship to big data and data-driven decision making[J]. *Big Data*, 2013, 1(1): 51-59.
- [7] GOLDSTON D. Big data: data wrangling[J]. *Nature*, 2008, 455: 15.
- [8] REICHMAN O J, JONES M B, SCHILDHAUER M P. Challenges and opportunities of open data in ecology[J]. *Science*, 2011, 331(6018): 703-705.
- [9] GUPTA U G, GUPTA A. Vision: a missing key dimension in the 5V big data framework[J]. Journal of International Business Research and Marketing, 2016, 1(3): 46-52.
- [10] 何敏. 智能煤矿数据治理框架与发展路径[J]. 工矿自动化, 2020, 46(11): 23-27.
HE Min. Framework and development path of data governance in intelligent coal mine[J]. Industry and Mine Automation, 2020, 46(11): 23-27.
- [11] 王国法,刘峰,庞义辉,等. 煤矿智能化——煤炭工业高质量发展的核心技术支撑[J]. 煤炭学报, 2019, 44(2): 349-357.
WANG Guofa, LIU Feng, PANG Yihui, et al. Coal mine intellectualization: the core technology of high quality development[J]. Journal of China Coal Society, 2019, 44(2): 349-357.
- [12] 谭章禄,王美君. 智能化煤矿数据归类与编码实质、目标与技术方法[J]. 工矿自动化, 2023, 49(1): 56-62, 72.
TAN Zhanglu, WANG Meijun. The essence, goal and technical method of intelligent coal mine data classification and coding[J]. Journal of Mine Automation, 2023, 49(1): 56-62, 72.
- [13] WU Xindong, KUMAR V, QUINLAN J R, et al. Top 10 algorithms in data mining[J]. *Knowledge and Information Systems*, 2008, 14: 1-37.
- [14] FREITAS A A. A review of evolutionary algorithms for data mining[M]//MAIMON O, ROKACH L. Data mining and knowledge discovery handbook. New York: Springer, 2010: 371-400.
- [15] GRABMEIER J, RUDOLPH A. Techniques of cluster algorithms in data mining[J]. *Data Mining and Knowledge Discovery*, 2002, 6: 303-360.
- [16] 赵勇,林辉. 大数据革命[M]. 北京:电子工业出版社, 2014.
ZHAO Yong, LIN Hui. Big data revolution[M]. Beijing: Publishing House of Electronics Industry, 2014.
- [17] 李航. 统计学习方法[M]. 2版. 北京:清华大学出版社, 2019.
LI Hang. Statistical learning methods[M]. 2nd ed. Beijing: Tsinghua University Press, 2019.
- [18] 刘年平. 煤矿安全生产风险预警研究[D]. 重庆:重庆大学, 2012.
LIU Nianping. Research on risk early warning in mine safety production[D]. Chongqing: Chongqing University, 2012.
- [19] NASSAJI H. Qualitative and descriptive research: data type versus data analysis[J]. *Language Teaching Research*, 2015, 19(2): 129-132.
- [20] 谭章禄,吴琦. 基于层级链参考模型的智慧矿山建设问题分析[J]. 矿业科学学报, 2022, 7(2): 257-266.
TAN Zhanglu, WU Qi. Analysis of the problems of smart mine construction based on the layer-level-chain reference model[J]. Journal of Mining Science and Technology, 2022, 7(2): 257-266.
- [21] 谭章禄,王美君. 智能化煤矿数据治理概念模型及技术架构研究[J]. 矿业科学学报, 2023, 8(2): 242-255.
TAN Zhanglu, WANG Meijun. Research on the conceptual model and technical architecture of data governance for intelligent coal mine[J]. Journal of Mining Science and Technology, 2023, 8(2): 242-255.