

基于大数据融合的应急物资可追溯信息挖掘模型构建

刘佳¹ 陈雪莲² 连鸿波³

¹(国网江西省电力物资有限公司 江西 南昌 330000)

²(国网江西省电力有限公司信息通信分公司 江西 南昌 330000)

³(国网上海松江供电公司 上海 200000)

摘要 传统的应急物资信息方法受到应急物资信息类型以及来源的影响,导致追溯效果不佳。为此,构建基于大数据融合的应急物资可追溯信息挖掘模型。选取巴科斯-诺尔范式规范化描述对应急物资可追溯信息,采用大数据融合匹配应急物资信息追溯结果。通过 Dasarathy 信息融合模型设计三层递进模式,利用 EM 算法建立挖掘模型,实现对应急物资可追溯信息的有效挖掘。实验结果表明,采用该模型可有效挖掘应急物资可追溯信息,不同情况下挖掘精度均高于99%,并具有较高的实时性。

关键词 大数据融合 应急物资 可追溯信息 挖掘模型

中图分类号 TP311

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.04.051

CONSTRUCTION OF TRACEABLE INFORMATION MINING MODEL FOR EMERGENCY MATERIALS BASED ON BIG DATA FUSION

Liu Jia¹ Chen Xuelian² Lian Hongbo³

¹(State Grid Jiangxi Electric Power Materials Co., Ltd., Nanchang 330000, Jiangxi, China)

²(State Grid Jiangxi Information & Telecommunication Company, Nanchang 330000, Jiangxi, China)

³(State Grid Shanghai Songjiang Electric Power Supply Company, Shanghai 200000, China)

Abstract The traditional emergency material information method is affected by the type and source of emergency material information, which makes the tracing effect poor. Therefore, a traceability information mining model of emergency materials based on big data fusion is constructed. This paper selected Bakos Noel paradigm to describe the traceability information of emergency supplies, used big data fusion to match the traceability results of emergency materials information, designed a three-level progressive mode through Dasarathy information fusion model, and established a mining model using EM algorithm to realize the effective mining of traceable information of emergency supplies. The experimental results show that the model can effectively mine the traceability information of emergency materials, and the mining accuracy is higher than 99% in different situations, and has high real-time performance.

Keywords Big data fusion Emergency supplies Traceable information Mining model

0 引言

信息技术已广泛应用于工程、企业等众多领域管理中,有效的应急物资可追溯信息挖掘模型已成为工程以及企业处理安全事故,召回缺陷产品以及产品追

溯领域中重要的工具^[1]。目前欧美等发达国家产品质量追溯以及跟踪技术发展已较为完善,我国可追溯技术发展较晚,企业以及工程项目施工材料质量以及物资材料使用情况决定了工程质量^[2]。目前我国地震、泥石流等自然灾害频发,应急物资的高效调运与使用可有效提升救援效率。应急物资具有特殊性,需要充

分考虑物资质量,应急物资可追溯信息具有重要意义^[3-5]。人类生活已离不开信息化技术,移动设备、互联网技术等高科技技术令数据量不断增加,大数据融合技术应运而生,大数据具有海量、实时性、多源异构性、价值密度低等特征,海量数据处理为人们带来较大的挑战。

宋焕等^[5]提出一种政企联合下海上应急物资储备的位置-分配模型,通过成本最小为目标构建双层优化模型:第一级模型解决了海上应急物资储备的位置分配问题;第二级模型是在第一级模型结果的基础上求解联盟应急物资分配问题。该模型利用遗传算法实现了应急物资分配信息追溯,能够有效解决位置分配问题。任斌等^[6]提出基于前景理论的应急物资需求动态调整模型,通过心理感知价值设计应急物资运输方案,通过价值函数确定物资需求预测量,同时计算风险规避系数,能够提升决策质量。肖文等^[7]提出基于MapReduce计算模型的并行关联规则挖掘算法,采用MapReduce并行计算模型分析救援路径,通过并行关联规则挖掘算法分析应急物资数据信息,给出最佳配送方案,此方法能够提升物资配送速度,提升配送效率。以上方法都能提升配送效率,但是未考虑应急物资的安全性与可靠性,未进行物资追溯,无法实现有效追踪应急物资的来源。

为了解决以上问题,本文构建基于大数据融合的应急物资可追溯信息挖掘模型,将大数据融合技术应用用于应急物资可追溯信息挖掘模型构建中,利用大数据融合技术充分整合海量应急物资可追溯信息,便于可追溯信息高效挖掘。利用可追溯对象批次映射关系建立可追溯信息挖掘模型,该模型可对应急物资的整体供应链实现充分跟踪与追溯,实时监控应急物资,充分实现数据共享,有效完善应急物资监控体系。

1 模型构建

1.1 可追溯信息规范化描述

应急物资需要以最快的速度配送到指定位置,有较强的时效性要求。应避免在分配流程、配送线路等问题上花费过多的时间,同时应急物资需要尽可能的合理,不要过少,同时也不宜过多影响配送效率。又因为是应急物资,所以需要根据具体情况给出具体方案,运输工具及运输规模都不是固定的,同时为了避免在运输过程中的突发状况,需要给出备选方案,并且要在72小时内将物资配送到救灾现场,保证受灾群众可以得到及时救治。为此需要尽可能地提升应急物资配送

效率。同时,应急物资的安全性也需要考虑在内,因此需要构建应急物资可追溯信息挖掘模型,保证物资的安全性及可靠性。

应急物资的可追溯信息是指应急物资供应过程中全部阶段的物资批次、生产批号、物资原料等信息,其中包括质检、采购、入库以及出库等应急物资全部流通环节。利用KZS表示可追溯对象,选取巴科斯-诺尔范式BNF规范化描述追溯批次以及应急物资活动情况。

1) 可追溯对象。将应急物资的可追溯对象视作单元控制存在可追溯性的对象,应急物资可追溯对象具有自身属性信息以及表示编号。

利用巴科斯-诺尔范式表示可追溯对象公式如下:

$$KZS ::= \langle KZS_{id}, KZS_{attribute} \rangle \quad (1)$$

式中:KZS与 KZS_{id} 分别表示随机阶段应急物资活动和批次以及应急物资活动或批次的位移标识编码; $KZS_{attribute}$ 表示可追溯对象的属性信息集合。

2) 关联关系。利用GL表示不同可追溯对象的追溯活动与追溯活动以及追溯活动与追溯批次间关联关系,差异环节通过追溯批次将活动对象互相关联。

利用巴科斯-诺尔范式表示关联关系公式如下:

$$GL ::= \langle KZS_{d_1}, KZS_{d_2} \rangle \quad (2)$$

式中: KZS_{d_1} 与 KZS_{d_2} 均表示应急物资供应链内可追溯对象数量为2时的唯一性编码。

利用一个以及多个关联对象可组成应急物资关联关系集合如下:

$$GL = (gl_1, gl_2, \dots, gl_n) \quad (3)$$

式中: gl_i 为关联关系,且 $i=1, 2, \dots, n$ 。

3) 映射关系。应急物资的跟踪映射关系可体现应急物资经多次转换后相应的映射关系以及物资批次间关联关系^[8-9],应急物资映射关系如图1所示。

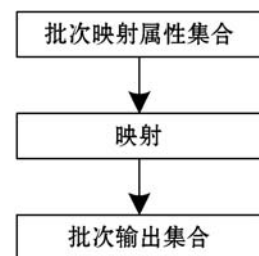


图1 应急物资的批次映射关系

应急物资批次的映射关系并非全部为一对一映射,主要包括以下三种映射关系:

1) 一对一映射。用 N 表示应急物资批次映射属性的数据集,不同供应环节的批次属性数据集分别用 $A \in N$ 以及 $B \in N$ 表示,当 $a_i \in A$ 时,存在 $b_i \in B$ 令 $b_i \in G(a_i)$,此时,从 A 至 B 内的批次属性数据的一对一映射为 G ,用 $G:A \rightarrow B$ 表示。

2) 一对多映射。应急物资的分批供需直接造成应急物资的分批一对多问题^[10], 一对多映射即一个原像与多个原像相对应。

3) 多对一映射。应急物资的合批是指一个原像对应多个原像。多对一映射存在非唯一的逆映射条件, 多个批次的应急物资合批前批次可合并至一个子集, 合并的子集以一个像呈现, 获取可逆映射的新映射。

1.2 大数据融合技术

应急物资信息较为碎片化, 可追溯性较差, 为此引入 Dasarathy 信息融合模型, 该模型能够通过信息的不同功能对海量物流信息融合处理, 实现数据层、语义层及服务层信息的融合, 通过决策模块实现信息挖掘, 提升物流信息的可追溯性。

将海量的应急物资可追溯信息多源碎片通过大数据融合技术处理后融合为符合应急物资可追溯信息挖掘决策需求的信息^[11-13]。为符合应急物资可追溯信息挖掘决策需求, 需要利用应急物资可追溯信息领域知识模型融合, 通过海量应急物资可追溯信息具有针对性的结构化以及描述规范化^[14], 令融合后应急物资可追溯信息为数据挖掘提供基础。

基于大数据融合技术的应急物资可追溯信息模型结构如图 2 所示。

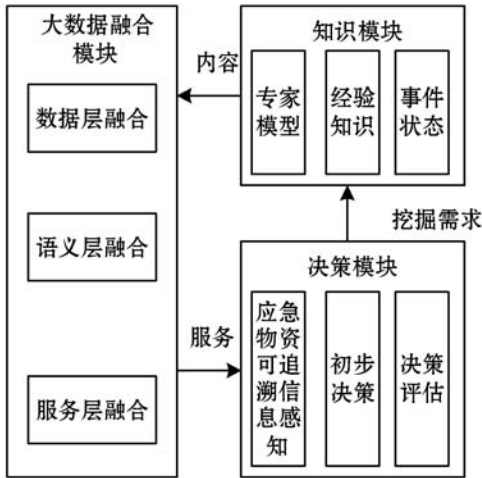


图2 应急物资可追溯信息模型结构图

应急物资可追溯信息模型包括大数据融合模块、知识模块以及决策模块。大数据融合模块是应用于应急物资可追溯信息的大数据融合技术核心处理部分。大数据融合模块服务于决策模块, 利用知识模块辅助大数据融合模块实施处理海量应急物资可追溯信息。

大数据融合模块利用 Dasarathy 信息融合模型作为应急物资可追溯信息的大数据融合技术架构^[15], 该模型主要由数据层、语义层以及服务层三层递进模式构建。海量的碎片化应急物资可追溯信息通过数据层

融合输入大数据融合模块, 经过数据层融合后将符合统一元数据体现标准的数据输出, 数据层通过多源获取、集中存放等过程实现数据的充分融合。将数据层融合结果输入语义层, 通过知识模型映射、语义特征提取、语义细粒度融合以及制定融合规则等步骤实现大数据融合。将初步分类结构的语义化信息作为服务层融合输入, 通过可视化、时间、空间等融合处理获取应急物资可追溯信息挖掘的具体信息服务, 利用决策模块感知应急物资可追溯信息。初步决策并进行评估, 决策主体通过服务层融合处理后直观展示决策信息, 将决策信息挖掘需求传输至知识模块。根据专家模型进行客观计算, 输出知识模块内容结构的语义化信息, 转化为指令传到数据写入端然后生成事件, 再将应急物资可追溯信息内容传输至大数据融合模块, 实现应急物资可追溯信息融合。大数据融合技术以逐层细化、组成深入方式深加工处理应急物资可追溯信息。

1.3 可追溯信息挖掘模型构建

利用高效的挖掘模型可有效从海量大数据中挖掘应急物资的可追溯信息, 将规范化描述的海量应急物资可追溯信息利用大数据融合技术处理, 并建立数据挖掘模型提取有用信息。

选取数据挖掘中常用的 EM 算法建立大数据融合的应急物资可追溯信息挖掘模型。EM 算法通过迭代实现应急物资可追溯信息的极大似然估计, 通过设置应急物资可追溯信息挖掘模型中止阈值, 已知应急物资可追溯信息概率分布, 获取应急物资可追溯信息挖掘模型, 挖掘得出可追溯信息的大概值。

设 EM 算法中包含数据集维度与数量分别为 d 以及 m , 设置应急物资可追溯信息挖掘模型中止阈值为 ε 。

用 y 表示应急物资可追溯信息, 该数据从属于 k 聚类的概率公式如下:

$$q(y) = \frac{q'(y)f_v(y|\gamma \cdot \sum_{y=d}^k)}{\sum_{i=1}^n q'(y)f_v(y)} \quad (4)$$

式中: $q'(y)$ 表示实际从属于 k 聚类的可追溯应急物资数据量; $f_v(y)$ 表示应急物资数据总信息量, 依据式(4)获取应急物资可追溯信息挖掘模型如下:

$$\begin{cases} q(y) = \frac{\sum_{y=d} q(y) \cdot y}{\sum_{y=d} q'(y)} \\ \sum k = \frac{\sum_{y=d} q(y) \cdot (y - k)}{\sum_{y=d} q'(y)} \quad k = 1, 2, \dots, n \end{cases} \quad (5)$$

设定挖掘中止条件为 $|L(q) - L(q')|$, 当挖掘结果符合中止条件时, 结束应急物资的可追溯信息挖掘。

在上述基础上, 给出应急信息可追溯信息挖掘流程图。

根据图 3 应急物资可追溯信息挖掘流程可知, 采集应急物资可追溯信息, 利用巴科斯-诺尔范式规范化处理应急物资可追溯信息, 采用大数据融合处理, 构建 Dasarthy 信息融合模型, 通过不同功能融合处理海量应急物资可追溯信息, 实现数据层、语义层及服务层信息融合的三层递进模式, 得到应急物资可追溯信息融合结果, 运用 EM 算法迭代应急物资可追溯信息极大似然估计, 获取应急物资可追溯信息挖掘模型, 设置应急物资可追溯信息挖掘模型中止阈值, 挖掘应急物资可追溯信息的大概值, 实现应急物资可追溯信息挖掘。



图 3 可追溯信息挖掘流程

2 实验分析

为检测本文构建基于大数据融合的应急物资可追溯信息挖掘模型挖掘应急物资可追溯信息有效性, 选取 MATLAB 软件仿真模拟 2013 年 7 月 22 日甘肃定西市地震后应急物资需求情况, 设置应急物资可追溯信息数据量大小为 10 GB, 检测采用本文模型挖掘应急物资可追溯信息有效性。

采用本文模型挖掘 20 个应急物资可追溯信息结果如表 1 所示。

表 1 本文模型挖掘结果

应急物资名称	生产批号	主要原料	物流信息
止血绷带	205463854	纱布	大连市-定西市
医用手套	4352184	乳胶	定州市-定西市
瓶装水	34516258	水	九江市-定西市
棉衣	52431658	棉布	四平市-定西市
棉被	246158743	棉花	大同市-定西市
罐头	23541685	食品	武汉市-定西市
压缩食品	1852643	食品	烟台市-定西市
消毒水	1524876	乙醇	任丘市-定西市
5 号电池	2564871	电解液	日照市-定西市
7 号电池	2568746	电解液	常州市-定西市
手电筒	2368451	金属	深圳市-定西市
对讲机	56484352	塑胶	汕头市-定西市
帆布	13648524	麻织物	石家庄市-定西市
编织袋	41568978	聚乙烯	乐山市-定西市
防护口罩	56351843	无纺布	无锡市-定西市
帐篷	526841266	尼龙布	厦门市-定西市
垃圾袋	2568415	聚乙烯	邯郸市-定西市
餐具	5236481565	金属	上海市-定西市
保温毯	5645135462	玻璃棉	黄石市-定西市
防护眼镜	5284562	PVC	杭州市-定西市

可以看出, 采用本文模型可有效挖掘应急物资的生产批号、主要原料以及物流信息三种可追溯信息, 实验结果有效验证本文模型挖掘应急物资可追溯信息的有效性。

2.1 溯源时间对比

采用基于前景理论的挖掘模型、基于 MapReduce 计算的挖掘模型以及本文挖掘模型对应急物资可追溯信息的溯源时间进行检测, 结果如图 4 所示。

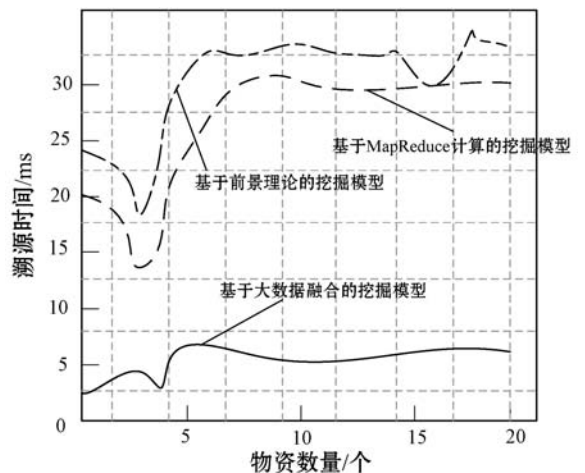


图 4 不同模型的溯源时间

分析图 4 可知,不同物资数量下溯源时间不同。当物资数据为 5 个时,基于前景理论挖掘模型的溯源时间为 32 ms,基于 MapReduce 计算挖掘模型的溯源时间为 24 ms,基于大数据融合挖掘模型的溯源时间为 7 ms。当物资数据为 20 个时,基于前景理论挖掘模型的溯源时间为 36 ms,基于 MapReduce 计算挖掘模型的溯源时间为 30 ms,基于大数据融合挖掘模型的溯源时间为 7.5 ms。这是因为本文方法利用大数据融合匹配应急物资信息追溯结果,利用 EM 算法建立了挖掘模型,实现了应急物资可追溯信息的有效挖掘。说明本文模型具有较高的挖掘实时性,本文模型将大数据融合技术应用于应急物资可追溯信息挖掘中,有效提升本文模型于海量数据中的挖掘性能。

2.2 挖掘精度对比

为进一步验证基于大数据融合挖掘模型的平均挖掘精度,统计结果如图 5 所示。

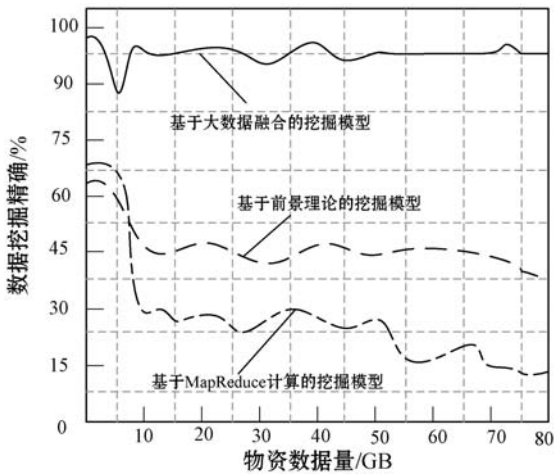


图 5 不同模型的挖掘精度

分析图 5 可知,不同物资数量下挖掘精度存在差异。当物资数据为 10 GB 时,基于前景理论挖掘模型的数据挖掘精度为 45%,基于 MapReduce 计算挖掘模型的数据挖掘精度为 29%,基于大数据融合挖掘模型的数据挖掘精度为 96%。当物资数据为 50 GB 时,基于前景理论挖掘模型的数据挖掘精度为 46%,基于 MapReduce 计算挖掘模型的数据挖掘精度为 28%,基于大数据融合挖掘模型的数据挖掘精度为 95%,本文模型平均挖掘精度随着应急物资可追溯信息数据量的增加而有所降低,但仍可以保持较高的平均挖掘精度。应急物资可追溯信息数据量为 80 GB 时,本文模型仍可以保持 97.5% 的挖掘精度。这是由于本文方法通过 Dasarathy 信息融合模型设计数据层融合、语义层融合以及服务层融合三层递进模式,使应急资源信息一一对应,有效提升了应急信息数据的挖掘精度。实验结果再次验证本文模型的挖掘准确性。

2.3 查全率对比

在上述基础上,追加应急物资可追溯信息的查全率检测实验,结果如图 6 所示。

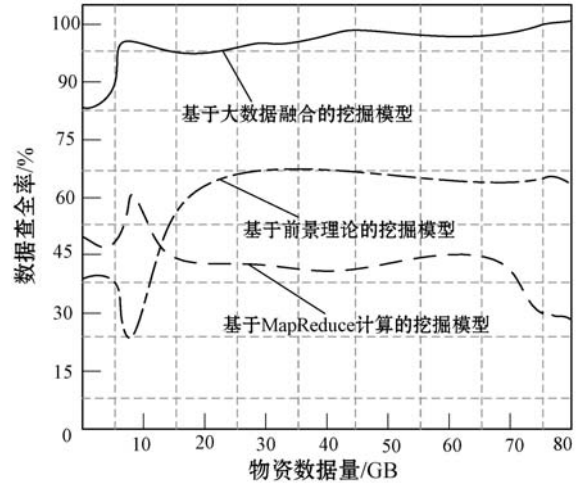


图 6 不同模型的数据查全率

分析图 6 实验结果可知,当物资数据为 20 GB 时,基于前景理论挖掘模型的数据查全率为 67%,基于 MapReduce 计算挖掘模型的数据查全率为 44%,基于大数据融合挖掘模型的数据查全率为 96%。当物资数据为 70 GB 时,基于前景理论挖掘模型的数据查全率为 65%,基于 MapReduce 计算挖掘模型的数据查全率为 45%,基于大数据融合挖掘模型的数据查全率为 97%。所设计挖掘模型的应急物资可追溯信息在不同数据量大小均可保持较高的查全率。本文模型在应急物资可追溯信息数据量大小为 80 GB 时,查全率仍可高达 98%。这是因为所提方法利用了巴科斯-诺尔范式对应急物资可追溯信息进行规范化描述,使应急物资信息追溯匹配过程更加精准,并且使用了大数据融合技术实现了数据层融合、语义层融合以及服务层融合三层递进模式,实现应急物资可追溯信息的有效挖掘。

3 结 语

应急物资的可追溯信息挖掘具有较高要求,通过对应急物资可追溯信息的标准化、规范化处理,有利于应急物资可追溯信息挖掘模型有效建立。将大数据融合技术应用于可追溯信息挖掘模型建立中,有效简化追溯流程以及追溯时间,并提升应急物资可追溯信息挖掘精确性以及挖掘速度。通过仿真实验验证本文模型不仅可有效提升应急物资可追溯信息挖掘精度,对于应急物资可追溯信息的自动化、数字化技术具有理论意义以及使用价值。

本文方法应用对象是应急物资可追溯信息,需要

及时获得反馈,因此,接下来的研究重点是提升溯源用时。

参 考 文 献

- [1] Alkadri M F, Turrin M, Sariyildiz S. A computational workflow to analyse material properties and solar radiation of existing contexts from attribute information of point cloud data [J]. *Building and Environment*, 2019, 155(1):268-282.
- [2] 李维刚,杨威,赵云涛,等.融合大数据与冶金机理的热轧带钢力学性能预报模型[J].*钢铁研究学报*,2018,30(4):302-308.
- [3] 王剑波,李腾飞.基于大数据挖掘的高速公路货车能耗统计方法研究与应用[J].*公路*,2019,64(4):228-232.
- [4] Ai Y, Lu J, Zhang L. Location-allocation model of maritime emergency supplies repertory under joint of government and enterprise[J]. *Fifth International Conference on Transportation*, 2015,21(1):883-892.
- [5] 宋焕,王瑞梅,胡好.食品供应链中溯源信息共享的演化博弈分析[J].*哈尔滨工业大学学报(社会科学版)*,2017,19(2):111-118.
- [6] 任斌,朱昌锋,钟校.基于前景理论的应急物资需求动态调整模型[J].*中国安全科学学报*,2018,28(3):179-184.
- [7] 肖文,胡娟,周晓峰.基于 MapReduce 计算模型的并行关联规则挖掘算法研究综述[J].*计算机应用研究*,2018,35(1):13-23.
- [8] Wang X, Zhai Y, Lin Y, et al. Mining layered technological information in scientific papers: A semi-supervised method [J]. *Journal of Information Science*,2019,45(6):779-793.
- [9] 盛虎宜,刘长石,鲁若愚.震后初期应急物资短缺情景下的定位-路径问题研究[J].*运筹与管理*,2019,28(6):41-47.
- [10] 黎茵.有机农产品的食源性风险与可追溯体系[J].*北京交通大学学报(社会科学版)*,2017,16(4):46-53.
- [11] 孟燕萍,申慢慢.考虑灾后道路恢复情况下动态应急物资选址问题[J].*重庆交通大学学报(自然科学版)*,2019,38(1):89-96.
- [12] Hong S H, Lee S K, Yu J H. Automated management of green building material information using web crawling and ontology[J]. *Automation in Construction*, 2019, 102(7):230-244.
- [13] Christensen T H, Rommes E. Don't blame the youth: The social-institutional and material embeddedness of young people's energy-intensive use of information and communication technology [J]. *Energy Research & Social Science*, 2019, 49(1):82-90.
- [14] 杨建亮,侯汉平.基于自然灾害的大众应急物资快速投送问题研究[J].*北京交通大学学报(社会科学版)*,2017,16(4):72-79.
- [15] 宋英华,葛艳,杜丽敬,等.考虑车辆等待的应急物资调配

方案优化研究[J].*控制与决策*,2019,34(10):2229-2236.

(上接第343页)

模型。该模型通过对压力监测点历史数据的学习,发现运行中管网监测点与出厂压力的变化规律,并融合时间特征、特征权重筛选的控制点、经验调度控制点来预测各水厂出厂压力。对比结果表明,采用特征权重筛选与经验调度相结合选择压力控制点的方法具有较高且稳定的预测精度。

水厂压力预测模型作为供水智能调度系统的重要组成部分,未来与管网水力模拟系统相组合,根据对未来工况的预判生成水厂调度指令,在保障水量、水压、水质的前提下,实现智能化供水调度。

参 考 文 献

- [1] 李树平,赵子威,周艳春,等.多厂协同供水优化调度分析[J].*给水排水*,2021,47(6):148-154.
- [2] 金晓静,王金辉,毛丽萍.基于管网模型的多水源供水科学调度系统实践与应用[J].*给水排水*,2020,46(S1):942-944.
- [3] Marchi A, Simpson A R, Lambert M F. Pump operation optimization using rule-based controls[J]. *Procedia Engineering*,2017,186:210-217.
- [4] 黄廷林,戴雪峰,张卉,等.改进 PSO 算法在多水源供水系统优化调度中的应用[J].*中国给水排水*,2013,29(23):64-68.
- [5] 王彤,张浩祥,徐杰,等.城市供水系统两级优化调度研究[J].*人民黄河*,2019,41(7):81-86.
- [6] 李俊禹,刘书明,吴雪,等.基于动态剪枝的城市供水管网优化调度算法[J].*环境工程*,2022,40(6):226-232,153.
- [7] 蒲政衡,赵平伟,冯偲懋,等.基于深度学习的供水管网实时智能调度研究[J].*给水排水*,2022,48(11):166-172.
- [8] 林峰,李旭,曾翰,等.某大型城市智能供水调度优化算法研究与模拟实践[J].*中国给水排水*,2023,39(9):109-115.
- [9] 赵美玲,张巧珍,朱俊,等.基于在线模型的供水管网优化调度系统设计[J].*中国给水排水*,2022,38(16):35-39.
- [10] 周志华.机器学习[M].北京:清华大学出版社,2016:73-91.
- [11] Meng Q, Ke G L, Wang T F, et al. A communication-efficient parallel algorithm for decision tree [C]//30th Conference on Neural Information Processing Systems,2016:1279-1287.
- [12] Ke G L, Meng Q, Finley T, et al. LightGBM: A highly efficient gradient boosting decision tree [C]//31st Conference on Neural Information Processing Systems,2017:3149-3157.