



# 基于协议匹配和数据压缩的采煤机数据管理技术研究

王忠宾 魏东 司垒 梁超权 谭超 赵亦辉

## 引用本文：

王忠宾, 魏东, 司垒, 等. 基于协议匹配和数据压缩的采煤机数据管理技术研究[J]. 煤炭科学技术, 2024, 52(11): 89–102.  
WANG Zhongbin, WEI Dong, SI Lei. Research on data management technology of shearer based on protocol matching and data compression[J]. Coal Science and Technology, 2024, 52(11): 89–102.

在线阅读 View online: <https://doi.org/10.12438/est.2024-1138>

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 基于GIS的工作面探煤数据管理方法研究与应用

Research and application of coal exploration data management method in working face based on GIS  
煤炭科学技术. 2022, 50(12): 254–261 <https://doi.org/10.13199/j.cnki.cst.2021-0511>

#### 智能化煤矿大数据平台架构及数据处理关键技术研究

Study on big data platform architecture of intelligent coal mine and key technologies of data processing  
煤炭科学技术. 2020, 48(7) <http://www.mtkxjs.com.cn/article/id/b0ab44ca-6def-4469-aa36-9945f0f926c9>

#### 透明工作面智能化开采大数据分析决策方法及系统研究

Research on big data analysis and decision system of intelligent mining in transparent working face  
煤炭科学技术. 2022, 50(2): 252–262 <http://www.mtkxjs.com.cn/article/id/62bf9949-9d39-4600-9c2c-da0b68533560>

#### 基于透明地质大数据智能精准开采技术研究

Research on intelligent and precision mining technology based on transparent geological big data  
煤炭科学技术. 2021, 49(1): 286–293 <https://doi.org/10.13199/j.cnki.cst.2021.01.026>

#### 煤炭工业监控大数据平台建设与数据处理应用技术

Platform construction and data processing application technology in coal industry monitoring big data  
煤炭科学技术. 2019(3) <http://www.mtkxjs.com.cn/article/id/c16e1546-17a5-4c6c-ab68-a9b956679113>

#### 基于5G+云边端协同技术的采煤机智能调高调速控制系统设计与应用

Research on cloud-edge-terminal collaborative intelligent control of coal shearer based on 5G communication  
煤炭科学技术. 2023, 51(6): 205–216 <https://doi.org/10.13199/j.cnki.cst.2022-1017>



关注微信公众号，获得更多资讯信息



移动扫码阅读

王忠宾, 魏东, 司垒, 等. 基于协议匹配和数据压缩的采煤机数据管理技术研究[J]. 煤炭科学技术, 2024, 52(11): 89–102.

WANG Zhongbin, WEI Dong, SI Lei, et al. Research on data management technology of shearer based on protocol matching and data compression[J]. Coal Science and Technology, 2024, 52(11): 89–102.

## 基于协议匹配和数据压缩的采煤机数据管理技术研究

王忠宾<sup>1</sup>, 魏东<sup>1</sup>, 司垒<sup>1</sup>, 梁超权<sup>1</sup>, 谭超<sup>1</sup>, 赵亦辉<sup>2</sup>

(1. 中国矿业大学 机电工程学院, 江苏 徐州 221116; 2. 西安煤矿机械有限公司, 陕西 西安 710200)

**摘要:** 采煤机作为综采工作面的关键装备, 其智能化水平的提升对提高矿山智能化体系建设具有重要作用。当前, 复合传感与精细化监控技术日益成为智能化采煤机研制的主流方向, 导致所需监控的数据量明显提升, 对采煤机数据管理系统提出了更高的要求。采煤机远程监控系统的应用效果受协议适配度与监控数据传输及储存能力影响严重。为了提高采煤机远程管控系统的适配性和数据管理的时效性, 降低采煤机远程监控系统开发及部署难度, 开展了通讯协议匹配解析和实时数据压缩存储相关研究。首先构建采煤机通讯协议树模型, 提出基于子树匹配的采煤机通讯协议相似度计算方法, 设计了采煤机通讯协议匹配算法, 实现了不同型号采煤机通讯协议点表自适应匹配与解析; 其次设计了基于相同状态字的开关传感量压缩算法和基于变长编码的模拟传感量压缩算法, 提高了采煤机实时数据的压缩比, 降低了数据存储子系统开销; 最后, 构建了基于协议匹配和数据压缩的采煤机远程监控系统, 并进行实验测试。测试结果表明: 本次实验中, 所提出的基于子树匹配的采煤机通讯协议相似度计算方法与专家参考值一致, 协议匹配吻合度可达100%; 所提出的面向采煤机的数据压缩算法对开关量和模拟量的压缩率分别达到了99.16%和91.80%。

**关键词:** 采煤机实时数据管理; 远程监控系统; 通讯协议匹配; 数据压缩存取

中图分类号: TD421 文献标志码: A 文章编号: 0253-2336(2024)11-0089-14

## Research on data management technology of shearer based on protocol matching and data compression

WANG Zhongbin<sup>1</sup>, WEI Dong<sup>1</sup>, SI Lei<sup>1</sup>, LIANG Chaoquan<sup>1</sup>, TAN Chao<sup>1</sup>, ZHAO Yihui<sup>2</sup>

(1. School of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116; 2. Xi'an Coal Mine Machinery Co., Ltd., Xi'an 710200, China)

**Abstract:** Shearer as the key equipment of fully mechanized mining face, its improvement of the intelligence level plays an important role in improving the construction of the coal-mine intelligence. At present, composite sensing and fine monitoring technologies have increasingly become the mainstream direction of the development of intelligent shearer, resulting in a significant increase in the amount of data required to monitor, and higher requirements for the data management system of shearer are put forward. The application effect of shearer remote monitoring system is seriously affected by the suitability of the communication protocol and the capacities for the transmission and storage of the monitoring data. In order to improve the adaptability of the shearer remote control system and the timeliness of data management, and reduce the difficulty of development and deployment of the shearer remote monitoring system, the relevant research was carried out on communication protocol matching analysis and real-time data compression storage. Firstly, the shearer communication protocol tree model is constructed, the similarity calculation method of shearer communication protocol based on sub-tree matching is proposed. Then the matching algorithm of shearer communication protocol is designed, and the adaptive matching and parsing of communication protocol point table of different types of shearer is realized. Next, the switching sensor data compression algorithm based on the same state word

收稿日期: 2024-08-05 策划编辑: 常琛 责任编辑: 黄小雨 DOI: 10.12438/cst.2024-1138

基金项目: 国家自然科学基金资助项目(52174152, 52204179); 江苏省自然科学基金资助项目(BK20221128)

作者简介: 王忠宾(1972—), 男, 江苏徐州人, 教授, 博士。E-mail: [wzbcme@163.com](mailto:wzbcme@163.com)

通讯作者: 魏东(1992—), 男, 辽宁阜新人, 讲师, 博士。E-mail: [weidongcme@cumt.edu.cn](mailto:weidongcme@cumt.edu.cn)

and the analog sensor data compression algorithm based on the variable length coding are designed to improve the compression ratio of the real-time data of the shearer and reduce the overhead of the data storage subsystem. Finally, a remote monitoring system of shearer based on protocol matching and data compression was constructed and tested. The experimental results show that: the similarity calculation method of shearer communication protocol based on sub-tree matching proposed in this paper is consistent with the expert reference value, and the protocol matching coincidence can reach 100%. The proposed data compression algorithm for shearer achieves 99.16% and 91.80% compression rates for switching and analog data, respectively.

**Key words:** shearer real-time data management; remote monitoring system; communication protocol matching; data compression access.

## 0 引言

根据国家统计局的数据显示,2023年全年一次能源生产总量48.3亿t标准煤,比上年增长4.2%,全年能源消费总量57.2亿t标准煤,比上年增长5.7%,煤炭消费量增长5.6%,煤炭消费量占能源消费总量比重为55.3%,这表明煤炭的主体能源地位短期内仍然难以改变<sup>[1]</sup>。2020年,国家发展改革委、国家能源局等八个重要部门共同发布了《关于加快煤矿智能化发展的指导意见》,这一重大政策的出台,标志着煤炭工业正迈向一个全新的智能化发展阶段,预示着行业将实现更为先进的技术应用和更高的生产效率<sup>[2]</sup>。与此同时,工业4.0时代的到来也给煤炭工业带来了新的机遇与挑战,而数据被视为工业4.0核心要素之一,其海量数据的收集与高效使用对于实现智能生产至关重要<sup>[3]</sup>。煤矿数据管理是煤矿智能化发展的核心问题之一<sup>[4]</sup>。

采煤机作为煤矿综采工作面的核心装备之一,其智能化水平直接影响着煤矿智能化发展的进程<sup>[5,6]</sup>。在传统的煤矿工作模式中,采煤机操作人员通常需要依靠工作经验观察采煤机的工作情况,并且手动操控其运行。然而,由于煤矿工作面的极端环境,操作人员往往置身于危险之中。这种依靠人力经验的手动操控方式不仅存在着严重的安全隐患,还限制了生产效率的提升。面对这一挑战,降低工作面的人员数量、提高采煤机的智能化水平成为煤矿智能化发展的必由之路,智能化工作面技术为此提供了解决方案<sup>[7]</sup>。在智能化工作面系统中,远程监控系统<sup>[8,9]</sup>对采煤机的实时工况数据进行采集和存储,对采集到的数据进行解析播报从而了解采煤机的实时状态,实现对采煤机的实时监测,确保安全生产;对存储的采煤机历史数据分析,可以准确判断采煤机的运行状态与运行效率,为后续生产决策提供依据,优化采煤工艺与截割控制。因此,采煤机实时数据的管理成为实现综采智能化的关键问题之一<sup>[10]</sup>。

在采煤机截割过程中,采煤机机载控制系统以

毫秒级产生采煤机状态数据,通过高速工业总线传输到采煤机远程监控系统,采煤机工况数据的实时获取和数据存储是其中至关重要的环节<sup>[11-12]</sup>。采煤机远程监控系统采用高速工业现场总线确保及时准确地收集采煤机的工况数据,而由于采煤机通讯系统多根据煤矿需求进行定制化开发,即使采用相同的通讯协议,针对不同需求所开发的采煤机其通讯协议点也略有不同,导致采煤机远程监控系统的通讯程序不能适配所有型号的采煤机,造成采煤机远程通讯系统点表配置往往需要在井下进行,导致部署困难大。因此,设计一个通讯协议匹配算法,使采煤机远程监控系统能够迅速适配不同型号采煤机通讯具有重要的意义<sup>[13-14]</sup>;采煤机实时数据量庞大,且由于采煤机具有多种传感器,其数据类型复杂,因此研究高效的数据压缩存储算法以降低存储成本并提升数据管理效率同样具有重要的意义<sup>[15-16]</sup>。

在数据通讯方面,目前采煤机数据管理系统<sup>[17-18]</sup>主要存在以下问题:

1)缺少针对采煤机通讯协议格式应用层匹配的研究:不同型号的采煤机在综采工作面的设计需求上有所差异,它们所配备的传感器、智能分站及控制器也会有所不同,这种差异导致了采煤机数据通讯协议点表格式的不一致性,进而给采煤机远程监控系统的采集实时数据通讯模块带来了沉重的工作负担,使得系统的升级变得异常困难,需要解决通讯协议的自适应匹配问题。

2)缺少针对采煤机不同类型数据特征进行压缩的研究:采煤机配备多路CAN总线,CAN分站数量将近30个,各分站发送数据帧的间隔时间在10~250ms,数据量大且实时性强,同时采煤机监控数据复杂度高,不仅包含大量模拟量数据,还包含大量开关量数据。因此,需要一种专门针对采煤机实时数据的数据压缩算法来解决数据的压缩存储及访问问题。

针对以上不足,本文开展了通讯协议匹配解析和实时数据压缩存取方面的研究,构建了采煤机通讯协议树模型,提出了基于子树匹配的采煤机通讯

协议树相似度计算算法以及采煤机通讯协议匹配算法。同时,提出了针对开关量数据的相同状态字压缩算法以及针对模拟量数据的变长编码压缩算法,减少了存储或通信数据中的冗余,有效提高了有效数据密度。进而,设计并开发了一个集成通讯协议匹配和数据压缩存储的采煤机远程监控系统,可满足采煤机实时数据管理技术的需求,为生产决策提供可靠的数据支持。

## 1 基于子树匹配的采煤机通讯协议匹配算法研究

### 1.1 算法设计

采煤机实时工况数据采集流程如图1所示,首先,采煤机通过机载控制器采集传感器数据,机载控制器配有CAN通讯模块,通过CAN总线将数据定时发送给CAN网关,CAN网关将CAN数据帧转成

以太网数据帧发送给采煤机远程监控系统,随后,系统先进行协议匹配实现对CAN通讯协议的适配,并通过CAN通讯协议点表进行数据帧解析,最终发送给数据中心进行处理和分析<sup>[19]</sup>,从而实现对采煤机工况数据的实时采集和处理。

采煤机远程监控系统与采煤机之间的通讯需要依赖CAN通讯协议点表对数据帧进行数据解析。然而,由于不同型号的采煤机在不同综采工作面配置的传感器、智能分站及控制器存在差异,导致CAN通讯协议点表不一致。为了解决这一问题,本文针对不同版本的CAN通讯协议点表进行分析,发现点表之间大部分数据信息的定义是相同的,只存在针对不同采煤机特定功能实现上的部分差异。因此,可以通过计算其相似度来找出它们之间的差异,并进行相应的匹配替换,从而无需对系统程序进行大幅修改,保证了代码的可复用性。



图1 采煤机实时工况数据采集

Fig.1 Shearer real-time working condition data collection

如图2所示,基于子树匹配的采煤机通讯协议匹配算法包括以下步骤:首先,将原通讯协议点表泛化成通讯协议XML树结构并进行存储;然后,接收新通讯协议的测试数据帧,构建相应的数据帧子树;进而,采用子树匹配的算法来计算新通讯协议子树与原通讯协议树之间的子树相似度,找出原通讯协

议树中对应的最大相似子树;最后,对最大相似子树中存在的差异部分,进行替换操作处理,以适应新通讯协议格式,生成新通讯协议XML树,确保新的通讯协议格式能够被正确解析和应用。通过这一流程,能够有效地解决不同版本通讯协议之间的差异,并实现通讯协议的快速切换。



图2 采煤机通讯协议匹配算法过程

Fig.2 Shearer communication protocol matching algorithm process

### 1.2 CAN通讯协议XML树模型构建

采煤机CAN通讯协议树可结构分为根节点、子树和叶子节点,如图3所示。根节点在树的顶层,代表采煤机CAN通讯协议树。CAN通讯协议树根据CAN帧的CAN-ID划分为多个CAN帧子树,每个CAN帧子树可视为CAN通讯协议树的子树,其中每个CAN帧子树代表了一个CAN帧所携带的所有数据内容。每个CAN帧子树根据不同数据内容对数据节点进行划分,对采煤机模拟量数据以单独的数据节点作为叶子节点,或者对采煤机开关量数据将16位数据划分为不同的叶子节点,每个叶子节点代表了特定的数据信息。

采煤机CAN通讯协议XML树如图4所示。其中,图4a为采煤机CAN通讯协议XML树,其中以CAN通讯协议作为根节点,<CAN-ID>节点则代表CAN帧的标识信息,用来识别不同CAN帧。图4b为开关量CAN帧的子树结构,<数据1>、<数据2>、<数据3>、<数据4>四个节点分别对应了64个bit位数据段按照每16个bit位拆分的四个部分,每个部分代表一棵数据子树,方框内容表示每个数据子树节点的具体特征信息,即每个数据帧中各节点所对应的采煤机动作或者传感器数据信息的定义,而<bit15>到<bit0>表示每个数据子树节点下的每个bit位对应的具体特征信息。图4c为模拟量CAN帧的

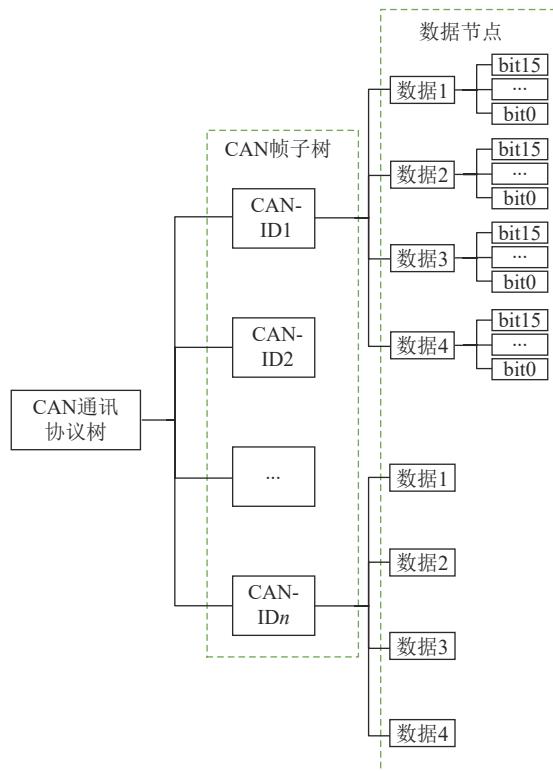


图3 采煤机 CAN 通讯协议树

Fig.3 Overall CAN communication protocol tree

子树,<数据 1>、<数据 2>、<数据 3>、<数据 4>节点代表 16 个 bit 位数据所代表的单一特征信息内容。

### 1.3 子树匹配相似度计算算法

#### 1.3.1 子树匹配相似度计算算法

在采煤机通讯协议树相似度计算中,将子树匹配定义为原通讯协议树  $T_{\text{old}}$  中的节点与现通讯协议树  $T_{\text{new}}$  中的节点的匹配过程。匹配的条件是节点之

间的特征信息相似,若特征信息相同,则将相似度权重设为 1,否则为 0。

原通讯协议树为  $T_{\text{old}} = (R_{\text{old}}, S_{\text{old}})$ , 其中  $R_{\text{old}}$  是  $T_{\text{old}}$  的根节点,  $S_{\text{old}}$  是  $T_{\text{old}}$  的子树集合, 有  $m$  个子树,  $S_{\text{old}} = \{S_1, S_2, S_3, \dots, S_m\}$ ; 现通讯协议树为  $T_{\text{new}} = (R_{\text{new}}, S_{\text{new}})$ , 其中  $R_{\text{new}}$  是  $T_{\text{new}}$  的根节点,  $S_{\text{new}}$  是  $T_{\text{new}}$  的子树集合,  $T_{\text{new}}$  有  $n$  个子树,  $S_{\text{new}} = \{S_1, S_2, S_3, \dots, S_n\}$ 。则自顶向下完成匹配,首先,计算  $T_{\text{old}}$  和  $T_{\text{new}}$  的相似度:

$$\text{Sim}(T_{\text{old}}, T_{\text{new}}) = \frac{\sum_{i=1}^m \sum_{j=1}^n \text{Sim}(S_{\text{old},i}, S_{\text{new},j})}{\max(m, n)} \quad (1)$$

其中,  $\text{Sim}(S_{\text{old},i}, S_{\text{new},j})$  是  $T_{\text{old}}$  的子树  $S_{\text{old},i}$  与  $T_{\text{new}}$  的子树  $S_{\text{new},j}$  的子树的相似度,即:

$$\text{Sim}(S_{\text{old},i}, S_{\text{new},j}) = \frac{\sum_{k=1}^p \sum_{l=1}^q \text{Sim}(N_{\text{old},k}, N_{\text{new},l})}{\max(q, p)} \quad (2)$$

其中,  $N_{\text{old},k}$  和  $N_{\text{new},l}$  分别为  $S_{\text{old},i}$  和  $S_{\text{new},j}$  中的节点,  $\text{Sim}(N_{\text{old},k}, N_{\text{new},l})$  是节点  $N_{\text{old},k}$  和  $N_{\text{new},l}$  的特征信息相似度计算函数,其通过对  $N_{\text{old},k}$  与  $N_{\text{new},l}$  两个节点所包含的特征信息进行判断。若特征信息相似,则  $\text{Sim}(N_{\text{old},k}, N_{\text{new},l})$  的值为 1,否则为 0,而  $p$  和  $q$  分别为  $S_{\text{old},i}$  和  $S_{\text{new},j}$  中的节点个数。

#### 1.3.2 子树匹配相似度计算流程

首先对两棵树的根节点特征信息进行判断,如果其根节点特征信息不同,则直接输出两棵树的相似度为 0;如果其根节点特征信息相同,则遍历原通

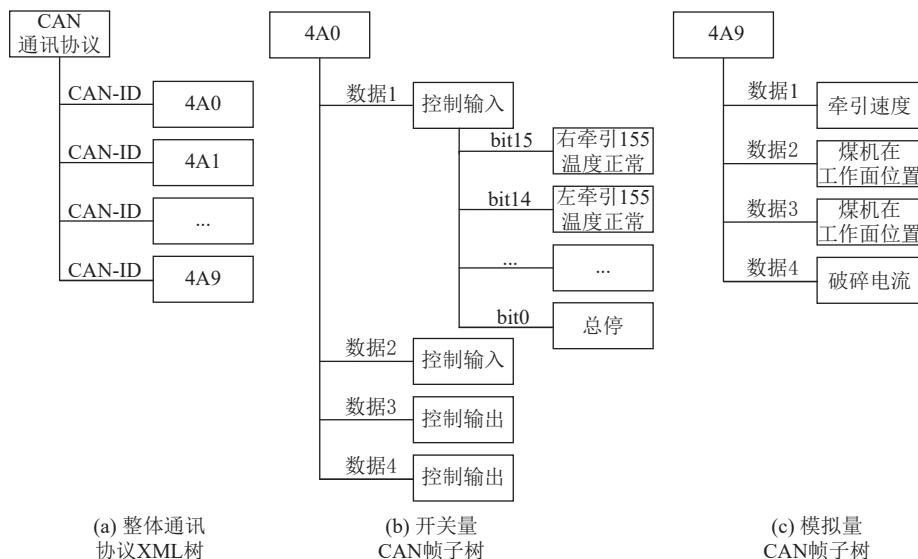


图4 采煤机 CAN 通讯协议 XML 树  
Fig.4 CAN communication protocol XML tree

讯协议树  $T_{\text{old}}$  和现通讯协议树  $T_{\text{new}}$  的子树集合。然后在现通讯协议树的子树集合中找到与原通讯协议树的子树特征信息相同的子树进行匹配, 对两个子树中的节点进行遍历, 判断它们的特征信息是否相同, 并依据前文定义进行权重计算, 即将特征信息相同, 则相似度权重为 1, 否则为 0, 返回整个子树的相似度  $\text{Sim}(T_{\text{old}}, T_{\text{new}})$ , 汇总所有子树的相似度, 加权求出两棵树的相似度。

## 1.4 通讯协议匹配算法

### 1.4.1 通讯协议匹配算法

由于在采煤机通讯协议匹配过程中, 在原始的通讯协议 XML 树被输入的基础上, 新通讯协议采用测试数据帧的方式输入, 通过 CAN 帧格式及其携带的特征信息来重新构造新的通讯协议 XML 树的过程需要耗费大量时间。为了提高通讯协议匹配效率, 本文采用直接对新通讯协议中的测试数据帧构建子树与原通讯协议树进行相似匹配, 找出新旧通讯协议中对应的最大相似子树, 并基于该最大相似子树对新协议中与原协议不同部分进行针对性替换, 即可实现在无需全面重新构造通讯协议 XML 树的情况下建立适配通讯协议。在此过程中, 定义新通讯协议中测试数据帧集为  $\text{Set}_{\text{new}} = \{\text{Set}_1, \text{Set}_2, \text{Set}_3, \dots, \text{Set}_m\}$ , 其中包含  $m$  个测试数据帧, 对每个测试数据帧直接与原通讯协议树  $T_{\text{old}}$  进行匹配, 根据当前测试数据帧携带的特征与原通讯协议树  $T_{\text{old}}$  的 CAN 帧子树下数据特征信息进行相似度计算, 如图 5 所示, A—P 为树节点。

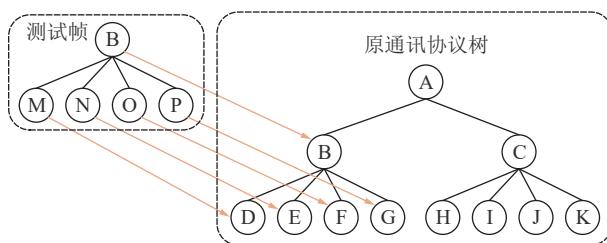


图 5 测试数据帧子树匹配

Fig.5 Test data frame subtree matching

### 1.4.2 通讯协议匹配流程

原通讯协议树  $T_{\text{old}}$  与新通讯协议中测试数据帧集为  $\text{Set}_{\text{new}}$  的匹配计算是一个遍历过程, 通过将新通讯协议测试数据帧中所携带的特征信息遍历原通讯协议树中的节点信息, 比较并记录比较节点, 从而对通讯协议树进行更新迭代。此过程中, 算法输入的是原通讯协议树  $T_{\text{old}}$  和包含  $m$  个测试数据帧的新通讯协议中测试数据帧集  $\text{Set}_{\text{new}}$ , 输出的是新通讯协议树  $T_{\text{new}}$ , 具体流程为:

首先, 持续接收系统发送的测试数据帧集合, 在判断原通讯协议树  $T_{\text{old}}$  所有节点是否遍历结束后, 根据系统原通讯协议树  $T_{\text{old}}$  内容解析当前测试数据帧, 即解析测试数据帧的报文识别码, 得到其 CAN 帧编码信息, 根据匹配节点占据 CAN 帧子树的匹配程度判断原通讯协议树  $T_{\text{old}}$  下的相同 CAN 帧子树是否匹配完成, 如果匹配未完成, 则进行解析测试数据帧的数据内容, 得到通过原通讯协议树  $T_{\text{old}}$  解析当前测试数据帧生成的特征信息, 并与当前测试数据帧所携带的特性信息对比, 如果特征信息相同, 即代表原通讯协议树  $T_{\text{old}}$  下的相同 CAN 帧子树内容可以适配于当前测试帧内容, 不需要替换, 否则, 需要替换原通讯协议树  $T_{\text{old}}$  下的相同 CAN 帧子树内容, 对比结束后需要记录当前所匹配的节点, 再继续接收下一测试数据帧判断, 当整个原通讯协议树  $T_{\text{old}}$  遍历结束, 即可以生成与新通讯协议适配的通讯协议树  $T_{\text{new}}$ 。

## 1.5 算法实验

从陕西某采煤机研发企业收集两组原始 CAN 通讯协议点表数据, 将其中一组数据设置为旧版本采煤机通讯协议点表, 用于原始系统的保留; 另一组设置为新版本采煤机通讯协议点表, 用于进行通讯协议匹配实验。由于构建新版本采煤机通讯协议 XML 树需要将所有数据内容读取完成才能进行协议匹配, 不符合采煤机远程监控系统快速匹配的思想, 因此构建了测试数据帧集, 该 CAN 帧信息只包含一个特征信息, 对测试数据帧集构建数据帧子树模式进行匹配, 同时对重复的测试数据帧舍弃, 进行采煤机通讯协议匹配算法的实验验证。

为了客观评价基于子树匹配的采煤机通讯协议匹配算法中的相似性算法的准确率, 以旧版本采煤机通讯协议树与新版本采煤机通讯协议树的 CAN 帧子树相似度进行评估。本次实验中, 由来自西安煤矿机械有限公司的 5 名从事采煤机远程控制系统开发 10 年以上的技术专家联合评估, 得到专家提供的相似度, 记为专家参考值。通过子树匹配相似度算法计算得出的相似度与简单树匹配算法计算的出的相似度、专家提供的相似度进行比较, 并绘制散点图以直观地观察实验结果, 如图 6 所示。

在图 6 中, 横坐标表示 CAN 帧子树, 纵坐标表示原通讯协议树与新通讯协议树的 CAN 帧子树相似度, 其中正方形代表采用简单树匹配算法求出的相似度, 三角形代表本文算法得出的相似度, 圆形代表专家提供的相似度。可以看出, 简单树匹配算法计算得出的相似度与专家提供的相似度并不能保持

一致,这是因为简单树匹配算法只考虑树的节点数量作为相似度的判断依据,而本文提出的子树匹配相似度算法在简单树匹配算法基础上添加了对树的节点特征信息的相似度判断,在本次实验中,其计算出的相似度与专家提供的相似度一致,即本文相似度计算算法可以找出新版本采煤机通讯协议子树对应的旧版本通讯协议最大匹配子树,对解决旧版本的采煤机通讯协议树与新版本的采煤机通讯协议树的匹配问题是有效的。

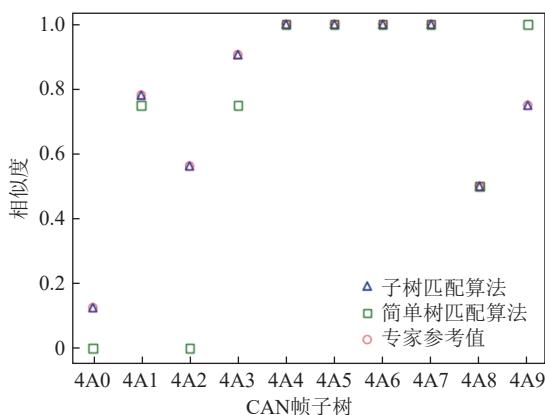


图 6 相似度散点图

Fig.6 Similarity scatter plot

由于新版本的采煤机通讯协议树不会直接以 XML 树形式进行匹配,本文通过对新版本的测试数据帧在原通讯协议树中的匹配,成功地确定了每个原通讯协议树与测试数据帧的对应子树中的差异点,这些差异点代表每个 CAN 帧子树中需要进行适配的位置,其结果如图 7 所示,其横坐标表示 CAN 帧子树,纵坐标表示每个 CAN 帧中的 64 个 bit 位位置,其中蓝色的散点代表开关量数据 CAN 帧子树需要替换的 bit 位,红色直线代表模拟量数据 CAN 帧子树需要替换的字节位。算法得出的差异点与专家提供的相似度反推每个 CAN 帧子树差异点比较,两者吻合度达到 100%。进一步将差异点替换可以实现将原通讯协议树构造生成新通讯协议树,从而避免了对整个通讯协议点表进行大规模替换所可能带来的误差以及系统程序代码不适应于新版本通讯协议出现不可复用的情况,提高了通讯效率。

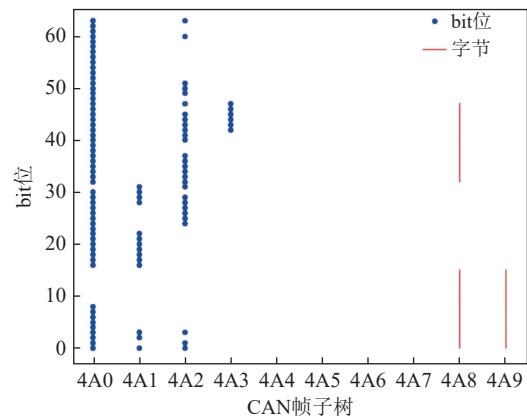


图 7 原 CAN 通讯协议树中 CAN 帧子树的差异点

Fig.7 The difference point of CAN frame subtree in the original CAN communication protocol tree

## 2 基于相同状态字与变长编码的采煤机实时数据压缩存取算法研究

### 2.1 采煤机实时数据压缩存取算法设计

采煤机配备多路 CAN 总线, CAN 分站数量将近 30 个,各分站发送数据帧的间隔时间在 10~250 ms,鉴于采煤机几乎整天都在运转,采集并存储其数据的成本相对较高,因此,设计采煤机实时数据压缩存取算法,以减轻采煤机远程监控系统的存储负担<sup>[20]</sup>。

采煤机实时数据包括开关量以及模拟量两种数据。对于开关量而言,其在数据段中只占据 1 个 bit 位,因此其数据值只存在 0 或 1 两种情况,同时由于采煤机运动状态的连续性,其数据值大多会在一段时间存在多个连续的 0 或 1。对于模拟量而言,通过对采煤机实时工况数据研究,其解析后的数据值的位数往往不超过 4 位,同时根据 CAN 通讯协议点表可知,其原始格式并不存在小数点,通常需要通过点表设定的倍率值去反应其真实数据范围。

针对采煤机实时数据中的两种不同形态,设计一种采煤机实时数据压缩存取算法,如图 8 所示。首先针对采煤机实时数据,根据其不同形态进行数据分类,并以设定的一段时间范围内的数据送入数据压缩模块进行压缩处理;然后针对开关量数据,采用基于相同状态字压缩算法,将一段时间内相同的状态字压缩成状态数值、第一个时间戳、第二个时间

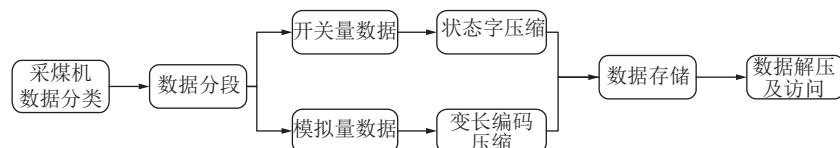


图 8 采煤机实时数据压缩存取过程

Fig.8 Shearer real-time data compression and access process

截以及最后一个时间截;同时针对模拟量数据,采用基于变长编码压缩算法,将数据通过编码方式重新压缩成长整型进行存储,并携带其压缩时间截起点以及数据倍率;最后对两种不同形态的数据进行解压以供提取访问。

## 2.2 基于相同状态字的采煤机开关量数据压缩算法

### 2.2.1 算法设计

针对采煤机开关量数据以0和1状态字代表动作状态及故障状态且具有连续性的特点,设计了一种相同状态字压缩算法。将开关量数据中连续出现的相同状态字段划分为一类,称为连续状态字段。具体而言,对于一个采煤机时间序列 $S$ ,可以将其切分为多个序列段,而对于一个如式(3)的序列段:

$$S_{i,j} = [(t_i, v_i), \dots, (t_j, v_j)], j - i \geq 2, \quad (3)$$

如果对于任意的 $k \in [i, j]$ ,满足式(4)的条件:

$$\begin{cases} v_k = v_i \\ t_k - t_{k-1} = t_{i+1} - t_i, \end{cases} \quad (4)$$

则可以称其为连续状态字段,可以采用开始时间 $t_{\text{start}}$ 、结束时间 $t_{\text{end}}$ 、时间间隔 $\Delta t$ 、数据值value四个参数进行来进行表示,即可完全重构相同的时间序列,否则无法进行解压缩操作。同时,观察到整个连续状态字段的完整描述从本质上仅包含在该段的第一个、第2个和最后一个数据点中,因此,理论上,一个连续状态字段可以仅通过这3个数据点完全表示,其具体操作过程如图9所示。在压缩过程中,把相同状态字中多余的冗余点去重,只保留连续状态字段的第一个、第2个和最后一个数据点用于记录重构序列所需要的必要信息,实现带有连续状态字段的时间序列数据的压缩。

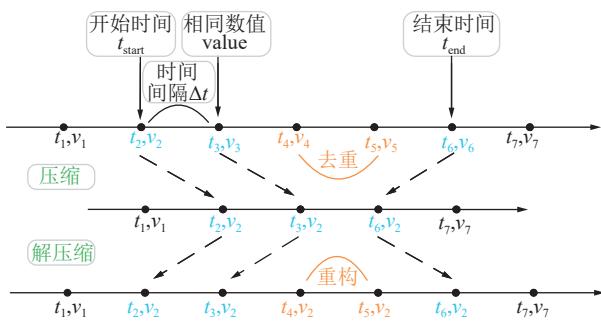


图9 相同状态字压缩算法操作过程

Fig.9 Operation process of the same status word compression algorithm

### 2.2.2 算法流程

基于相同状态字的采煤机开关量数据压缩算法的总体逻辑为:相同状态字压缩算法通过接收一个未压缩的时间序列数据流 $U$ ,经处理后输出一个压缩

的时间序列数据流 $R$ ,算法流程如下(图10):

在相同状态字数据压缩算法的开始阶段将压缩后的时间序列数据流 $R$ 、已压缩的前一个数据点的时间截 $t_c$ 、数值 $v_c$ 、连续出现次数 $l_c$ 、时间间隔 $l_t$ 初始化为空集、error、0、0、0;然后,算法遍历未压缩的数据流 $U$ ,逐个处理其中的数据点,如果当前数据点 $n_i$ 的值与前一个数据点的值相同且时间间隔与前一个数据点记录的时间间隔相同时,算法更新压缩数据点的时间截 $t_c$ ,并更新连续出现次数 $l_c$ ,使其自增一次。以上操作的目的是将连续出现的相同状态字的数据点合并为一个时间段,减少数据的重复存储。而当遇到不同状态字的数据点时,算法会判断连续出现次数 $l_c$ 是否大于0,如果大于0则代表前一个数据点是否是连续状态字段的结束点,并将其推入压缩的数据流 $R$ 中,同时当前数据点 $n_i$ 推入压缩的数据流 $R$ 中作为新的时间段的开始点并更新当前数据点信息,将已压缩的前一个数据点的时间截 $t_c$ 、数值 $v_c$ 、连续出现次数 $l_c$ 、时间间隔 $l_t$ 重置为0;最后,算法会处理剩余的数据点,将其推入压缩的数据流中,以确保当前时间段的所有数据都实现压缩存储。由于算法需要循环遍历数据流数据,循环迭代的处理逻辑都在常数时间内完成,因此算法总体时间复杂度为 $O(n)$ 。

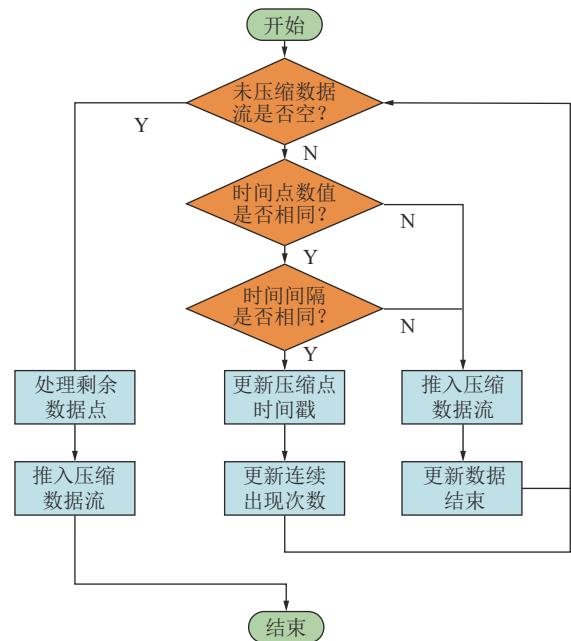


图10 相同状态字压缩算法流程

Fig.10 Flow chart of the same status word compression algorithm

## 2.3 基于变长编码的采煤机模拟量数据压缩与算法

### 2.3.1 算法设计

采煤机模拟量数据解析后的数据位数往往不超

过4位(小数值数据),根据CAN通讯协议点表可知,原始数据不存在小数点,需要通过点表设定的倍率值去反应其真实数据。据此设计了一种变长编码压缩算法,首先,针对需要进行压缩的采煤机时序数据段中的每个数据,采用按位编码的方式进行处理,将每个数据转换成一个相对较短的编码序列;然后,将每个数据的编码结果拼接在一起,形成一个完整的编码数据流;最后,将多个编码数据组合成一个长整数型数据。

变长编码旨在利用小数值数据在二进制bit位存储中通常存在大量高位0的特点,以这些空闲的bit位来存储其他小数值数据,可以有效减少存储空间的浪费,实现数据的紧凑存储。为了实现这一目的,需要建立一个编码数字映射表,用于指定0~9每个原始数字的编码规则,使得数据能够被高效地进行压缩操作,具体的编码数字映射见表1。

表1 编码数字映射

Table 1 Encoding digital mapping table

原始数字	编码数字
0	000
1	001
2	010
3	011
4	100
5	101
6	1100
7	1101
8	1110
9	1111

传统的二进制编码中,0~9数字涉及到10位数字,需要使用4个bit位进行存储。本文对原始数字0~5采用3个bit位的编码数字,高两位以00、01、10进行区分,对原始数字6~9,编码数字为4个bit位,高两位以11进行区分,这种编码规则不仅实现了编码单元的唯一性,还减少了以4个bit位进行存储每个数字所浪费的存储空间。考虑到时间序列中可能存在不同数据值位数的情况,算法以时间序列数据值中的最大位数作为标准编码位数,对于不满足编码位数的数据,需要在进行高位补0的操作。此外,在编码时需要标识出整个编码序列的标准编码位数,因此还需要设计一个标识数字来实现这个规则,标识数字编码映射见表2。

### 2.3.2 算法流程

变长编码压缩算法通过接收一个未压缩的时间

序列数据流U,并输出一个压缩的时间序列数据流R,算法流程如下:首先求出时间序列数据的最大位数并记录设计标识数字,求出需要存储的时间序列数据量的长度size,并把指针置为0,然后将编码压缩数据流U<sub>result</sub>设置为空列表,推入两位标识数字,把压缩输出的时间序列数据流R设置为空集。接下来,算法在一个循环中遍历未压缩数据流,在每次迭代中,算法获取当前指针下的数据点,并进行编码处理。然后,算法判断存储长度加上新加入的数据是否超过64位,如果没有,则将数据推入编码压缩数据流中,并更新存储长度和指针;如果超过64位,则将编码压缩数据流放入压缩后的时间序列流R中,并清空编码压缩数据流推入两位标识数字。最后,算法将剩余的编码压缩数据流存入压缩输出的时间序列流R,并返回结果。算法需要先遍历求出时间序列长度以及找出最大位数,然后还需要重新循环中需要遍历处理每个元素,因此算法总体时间复杂度为O(nm),需要对时间序列数据长度进行控制以降低处理时间。

表2 标识数字编码映射

Table 2 Identifies the digital encoding mapping table

标准编码位数	标识数字
0	00
1	01
2	10
3	11

## 2.4 算法实验分析

### 2.4.1 数据准备与数据压缩评价

采集采煤机在截割过程中的实时数据,进行数据预处理以及数据转化,根据CAN通讯协议点表进行分类处理,如果数据帧指示内容以一个bit位单位进行表示,则将其归类为开关量数据,如果数据帧指示内容以双字节位单位进行标识,则将其归类为模拟量数据。

选取采煤机转速、温度、位置、压力以及左牵状态(采煤机向左牵引)、左升状态(采煤机左滚筒上升)进行数据压缩存取算法研究。

在对开关量数据进行测试时,以采煤机左牵状态和左升状态两种情况为研究对象,开关量数据为状态字,其数据特征为0或1,开关量数据测试集如图11所示。

在对模拟量数据进行测试时,以采煤机转速、温度、位置、压力四种内容为研究对象,其中转速的波

动范围为[90,130],温度的波动范围为[15.0,18.0],位置的波动范围为[3 500,3 575],压力的波动范围是[4.5,7.5],温度以及压力数据是由CAN帧解析得到的数据通过CAN通讯协议点表中定义0.1的倍率生成,在进行数据压缩可以将其还原成整数,不必考虑小数点的位置,因此四种数据的数值位分别可以定义为3、3、4、2,据此对不同数位的压缩比进行测试,模拟量数据测试集如图12所示。

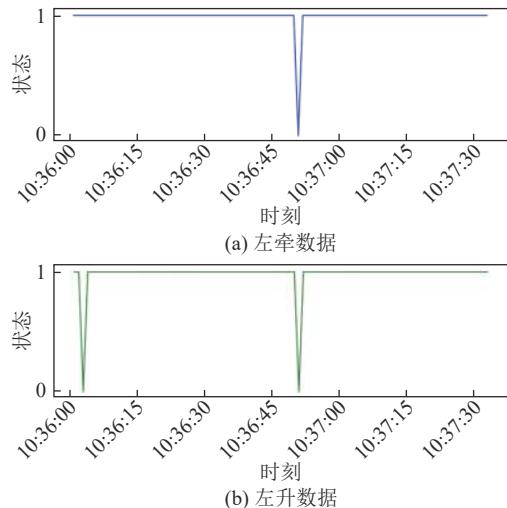


图11 开关量数据  
Fig.11 Switch data

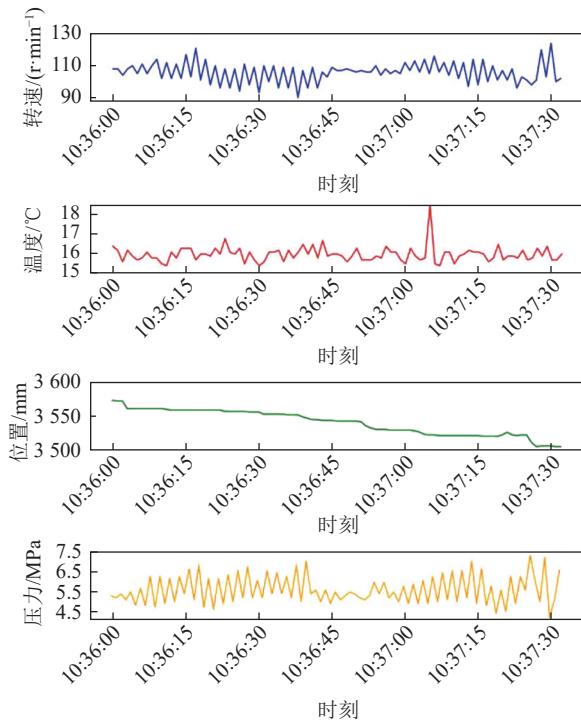


图12 模拟量数据  
Fig.12 Analog data

分别对采煤机开关量数据应用状态字压缩算法

以及对模拟量数据应用变长编码压缩算法进行数据压缩,通过压缩比来评估压缩算法的性能。压缩比计算如式(5)所示:

$$r = \frac{U}{C} \quad (5)$$

式中:  $r$ 为压缩比,  $U$ 为原始数据大小,  $C$ 为压缩后数据大小,压缩比越高,说明压缩算法效果越好,压缩后的文件大小越小。

#### 2.4.2 开关量数据压缩算法

分别使用状态字压缩算法、LZ78算法以及Gzip压缩算法对采煤机开关量数据进行数据压缩,得到的压缩比如图13所示,可以看出,在处理左牵数据和左升数据时,状态字压缩算法表现出了较好的压缩性能。针对于左牵数据以及左升数据拥有不同的压缩比例,其原因在于状态字压缩算法对于左牵数据进行压缩只需要记录三段数据内容,而左升数据进行压缩需要记录五段数据内容,因此左牵数据的压缩比要高于左升数据。这表明在时间序列数据中,若连续状态字段的长度越长,即数据在长时间稳定的状态下,将获得更好的压缩比。

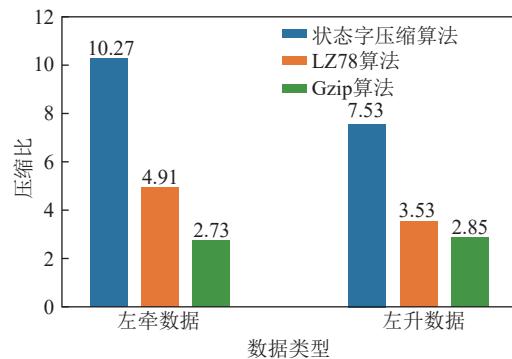


图13 采煤机开关量数据压缩算法对比  
Fig.13 Comparison of compression algorithms for shearer switch data

#### 2.4.3 模拟量数据压缩算法

分别使用变长编码压缩算法、LZ78算法以及Gzip压缩算法对采煤机模拟量数据进行数据压缩,得到的压缩比如图14所示,可以看出,变长编码压缩算法在处理采煤机模拟量数据时表现出较好的压缩性能。针对于转速、温度、位置、压力拥有不同的压缩比例,其原因在于变长编码压缩算法主要是针对数据的位数进行压缩。转速的波动范围为[90,130],算法认定其最大位数为3,温度的波动范围为[15.0,18.0],其数据由CAN帧解析的数值通过CAN通讯协议点表中定义0.1的倍率生成,在进行数据压缩可以将其还原成整数,因此其可以看为最

大位数也为3,位置的波动范围为[3 500,3 575],其最大位数为4,压力的波动范围是[4.5,7.5],同样其数据由CAN帧解析的数值通过CAN通讯协议点表中定义0.1的倍率生成,在进行数据压缩可以将其还原成整数,因此其可以看为最大位数为2。

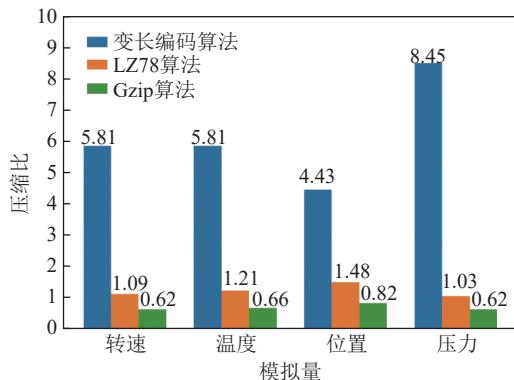


图14 采煤机模拟量数据压缩算法对比

Fig.14 Comparison of compression algorithms for shearer analog data

变长编码压缩算法采用长整型进行存储,其位数为64个bit位,减去标识数据的2个bit位即为62位,同时,通过编码数字映射表可以得出原始数据的每个数值位占据3~4个bit位,即可以判断出针对不同最大位数n而言,其压缩比的取值范围为[62/4n,62/3n],由此可以判断出,针对数值位越小的数据段,变长编码压缩算法具有更好的压缩比。此外,采用将一个数据(int型数据)拆分成多个数字,并将每个数字依据2.3.1中定义压缩成3~4个bit位,进而将多个数字整合成一个long型存储,这使得最大压缩比可以突破3/4≈1.33,实现更佳的模拟信号压缩效果。

### 3 采煤机远程监控系统

#### 3.1 采煤机远程监控系统架构和功能设计

采煤机远程监控系统架构如图15所示,主要包括采煤机数据采集模块、采煤机数据存储模块以及采煤机数据展示模块。采煤机数据采集模块负责实时获取采煤机上的传感器数据;采煤机数据存储模块用于高效地存储采集的数据,确保数据的安全性和完整性;采煤机数据展示模块将存储的数据进行处理和分析,并以直观的方式展示给用户,帮助用户全面了解采煤机的运行情况。

采煤机远程监控系统功能模块如图16所示,具体内容如下:

1)数据采集模块的主要功能包括协议匹配、数

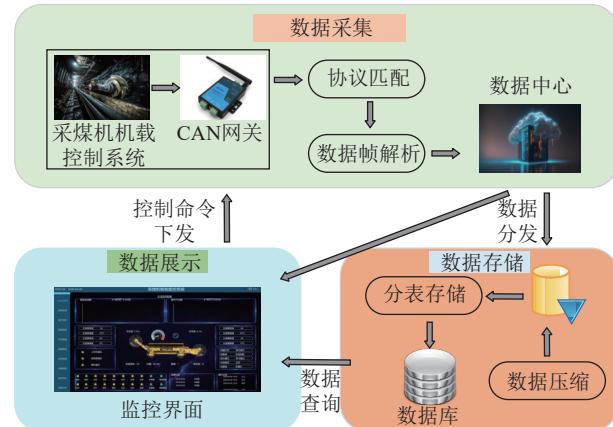


图15 采煤机远程监控系统架构

Fig.15 System architecture design

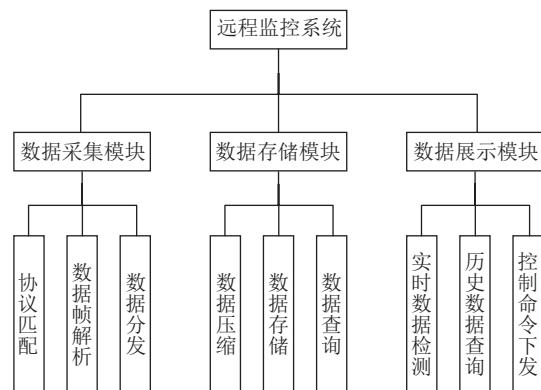


图16 系统功能设计

Fig.16 System function design

据帧解析和数据分发。首先,通过采煤机通讯协议匹配算法,将发送的新CAN通讯协议数据帧与系统保存的原CAN通讯协议XML树进行子树相似度计算,快速定位存在差异的CAN帧子树并进行适配修正;然后,通过数据帧解析,通过适配之后的CAN通讯协议XML树对采煤机实时工况数据帧进行解析,生成解析后的采煤机数据字典;最后,通过多线程的方式,实现对解析后的数据字典的实时分发,确保数据能够及时准确地传输到数据存储模块和数据展示模块。

2)数据存储模块主要负责数据压缩、数据存储、数据查询功能。首先,通过数据压缩算法,根据时间段对采煤机实时数据进行分段,对分段后的采煤机数据进行开关量数据以及模拟量数据的分类,对开关量采用相同状态字压缩算法以及对模拟量采用变长编码压缩算法,减少数据的存储空间,增加存储效率;然后,将压缩后的数据通过分表方式存储到数据库中,以进一步减轻数据库的存储压力,确保数据的安全性和可靠性;最后,提供数据查询接口,支持用户根据需求检索数据,并在数据查询时进行解压操作将数据还原成原始数据以供用户进行数据分析处理。

3) 数据展示模块主要体现在采煤机远程监控系统界面上, 其功能包括实时数据播报、历史数据查询和控制命令下发。通过实时数据播报, 对数据中心分发的采煤机数据字典进行实时的播报展示给用户, 从而实现对采煤机实时数据的监测; 通过历史数据查询, 用户可以根据条件查询数据库中的历史数据, 了解采煤机过去的运行状态, 并支持对采煤机进行进一步的控制; 同时, 用户还

CAN-ID	描述	注释
4A0 数 据 2 (4905)	控制器状态 数据1(0CH) 控制器输入状 态指示, =0灯 灭, =1灯亮, 对应采煤机主 画面	一个状态位功能定义如下: bit15 =0牵变绝缘低, =1牵变绝缘正常 bit14 =0破碎绝缘低, =1破碎绝缘正常 bit13 =0右截割绝缘低, =1右截割绝缘正常 bit12 =0左截割绝缘低, =1左截割绝缘正常 bit11 bit10 bit9 =0无操作, =1牵变停止 bit8 =0无操作, =1牵变启动 bit7 =0无操作, =1破碎停止 bit6 =0无操作, =1牵变启动 bit5 =0无操作, =1右截停止 bit4 =0无操作, =1右截启动 bit3 =0无操作, =1左截停止 bit2 =0无操作, =1左截启动 bit1 =0无操作, =1顺序停止 bit0 =0无操作, =1牵变启动 bit15 保留位 bit14 保留位 bit13 保留位 bit12 保留位 bit11 保留位 bit10 保留位 bit9 =0从变频节点, =1从变频节点故障 bit8 =0主变频节点, =1主变频节点故障 bit7 =0右泵155温度断开, =1右泵115温度正常 bit6 =0左泵155温度断开, =1左泵115温度正常 bit5 =0右牵155温度断开, =1右牵115温度正常 bit4 =0左牵155温度断开, =1左牵115温度正常

(a) 原CAN通讯协议点表

可以通过控制命令下发功能, 实现对采煤机动作控制。

### 3.2 系统测试

#### 3.2.1 协议匹配功能测试

采煤机远程监控系统将一张 CAN 通讯协议点表生成为通讯协议 XML 树存储在系统中作为第一版本的通讯协议, 以供系统进行数据帧解析操作, 其映射实例如图 17 所示。

```

<communicate type="can">
  <id type="4A0">
    <数据1 type="控制输入">
      <bit15 type="牵变绝缘正常"/>
      <bit14 type="破碎绝缘正常"/>
      <bit13 type="右截绝缘正常"/>
      <bit12 type="左截绝缘正常"/>
      <bit11 type="保留位"/>
      <bit10 type="保留位"/>
      <bit9 type="牵变停止"/>
      <bit8 type="牵变启动"/>
      <bit7 type="破碎停止"/>
      <bit6 type="破碎启动"/>
      <bit5 type="右截停止"/>
      <bit4 type="右截启动"/>
      <bit3 type="左截停止"/>
      <bit2 type="左截启动"/>
      <bit1 type="顺序停止"/>
      <bit0 type="顺序启动"/>
    </数据1>
    <数据2 type="控制输入">
      <bit15 type="保留位"/>
      <bit14 type="保留位"/>
      <bit13 type="保留位"/>
      <bit12 type="保留位"/>
      <bit11 type="保留位"/>
      <bit10 type="保留位"/>
      <bit9 type="从变频节点故障"/>
      <bit8 type="主变频节点故障"/>
      <bit7 type="右泵155温度正常"/>
      <bit6 type="左泵155温度正常"/>
      <bit5 type="右牵155温度正常"/>
      <bit4 type="左牵155温度正常"/>
    </数据2>
  </id>
</communicate>

```



(b) 原CAN通讯协议XML树

图 17 CAN 通讯协议 XML 树

Fig.17 CAN communication protocol XML tree

根据新 CAN 通讯协议点表进行 CAN 数据帧设计, 并将 CAN 数据帧分为测试数据帧以及验证数据帧, 其中测试数据帧用于进行通讯协议匹配测试, 而

验证数据帧用于进行对新生成的 CAN 通讯协议 XML 树进行验证其数据解析的正确性, 表 3 给出新 CAN 通讯协议数据帧设计。

表 3 CAN 数据帧

Table 3 CAN data frame

CAN帧	CAN帧信息
b'\x08\x00\x00\x04\x00\x08\x00\x00\x00\x00\x00\x00'	右牵155温度正常
b'\x08\x00\x00\x04\x00\x04\x00\x00\x00\x00\x00\x00'	左牵155温度正常
b'\x08\x00\x00\x04\x00\x02\x00\x00\x00\x00\x00\x00'	顺停
b'\x08\x00\x00\x04\x00\x01\x00\x00\x00\x00\x00\x00'	顺启
b'\x08\x00\x00\x04\x00\x00\x08\x00\x00\x00\x00\x00'	左截割停止
b'\x08\x00\x00\x04\x00\x00\x04\x00\x00\x00\x00\x00'	左截割启动
b'\x08\x00\x00\x04\x00\x01\x04\x00\x00\x00\x00\x00\x00'	顺启、左截割启动
b'\x08\x00\x00\x04\x00\x02\x04\x00\x00\x00\x00\x00\x00'	顺停、左截割启动
...	...

系统进行协议匹配时, 会根据系统配置的原通讯协议 XML 树对数据帧进行解析, 如对数据帧 b'\x08\x00\x00\x04\x00\x01\x04\x00\x00\x00\x00\x00\x00' 的解析结果为 4A0 数据帧上数据 1 内容的左截

绝缘正常以及保留位状态字从 0 变成 1, 而数据帧携带的数据内容为顺启、左截割启动, 从而对其 bit 位进行差异定位以及替换操作。而当系统实现协议匹配之后, 会生成新的 CAN 通讯协议 XML 树, 对新

CAN 通讯协议 XML 树与其通讯协议点表进行对比测试,结果见表 4,匹配成功率可达 100%。

### 3.2.2 数据压缩存储功能测试

对采煤机远程监控系统进行一小时的采煤机实时数据归档测试,表5中给出了采煤机实时数据中部分数据的测试数据集,其中采煤机远程监控系统以250 ms进行一次数据字典推送给数据存储模块进行存储,测试数据集提取了采煤机实时数据里面比较具有标志性的内容,包括象限编号、煤机在工作面位置、牵引变压器温度、左滚筒采高四个模拟量数据以及右牵、输入左升两个采煤机状态的开关量数据。

系统通过对测试数据集分别进行简单存储以及压缩算法存储进行对比,当进行简单存储时,存储模块每一次接收到实时数据都会存入数据库中,而进行压缩算法存储时,系统会收集1 min内的数据,将其进

表 5 实时数据测试数据集

时刻	数据字典
2024-04-06 18:06:14.846 413	{'象限编号': 1, '煤机在工作面位置': 3 973, '右牵': 1, '牵引变压器温度': 16.3, '输入左升': 0, '左滚筒采高': 0}
2024-04-06 18:06:15.103 265	{'象限编号': 1, '煤机在工作面位置': 3 974, '右牵': 1, '牵引变压器温度': 15.6, '输入左升': 0, '左滚筒采高': 0}
2024-04-06 18:06:15.367 207	{'象限编号': 1, '煤机在工作面位置': 3 975, '右牵': 1, '牵引变压器温度': 16.3, '输入左升': 0, '左滚筒采高': 0}
2024-04-06 18:06:15.629 675	{'象限编号': 1, '煤机在工作面位置': 3 975, '右牵': 1, '牵引变压器温度': 15.7, '输入左升': 0, '左滚筒采高': 0}
2024-04-06 18:06:15.890 272	{'象限编号': 1, '煤机在工作面位置': 3 976, '右牵': 1, '牵引变压器温度': 15.1, '输入左升': 0, '左滚筒采高': 0}
2024-04-06 18:06:16.178 315	{'象限编号': 1, '煤机在工作面位置': 3 977, '右牵': 1, '牵引变压器温度': 16.8, '输入左升': 0, '左滚筒采高': 0}
...	...

表 4 协议匹配测试结果

新CAN通讯协议XML 树对应CAN-ID	新通讯协议点表CAN-ID 匹配结果
4A0	Y
4A1	Y
4A2	Y
4A3	Y
4A4	Y
4A5	Y
4A6	Y
4A7	Y
4A8	Y
4A9	Y

行数据分类以及针对不同数据采用不同压缩算法进行压缩后再存入数据库中,其测试结果如图 18 所示。

图 18 存储结果

两种类型的数据分别采用简单存储以及算法压缩存储之后,针对其不同的存储方式采用单表存储行数以及单表占硬盘空间大小进行压缩效果性能的测试,压缩性能测试结果见表6,可知针对采煤机模拟量数据采用基于变长编码压缩算法进行数据存储,相对比直接存储的压缩率可以达到91.8%,针对采煤机开关量数据采用基于状态字压缩算法进行数据存储,相对比直接存储的压缩率可以达到99.16%。

表6 压缩性能测试结果

Table 6 Compression performance test results

存储方式	模拟量 数据简 单存储	模拟量数据 变长编码压 缩存储	开关量 数据简 单存储	开关量相 同状态字 压缩存储
数据表行数	54 712	228	27 356	114
数据表大小/kB	6 470.38	533	3 249.69	27.1

综上所述,采煤机远程监控系统采用采煤机实时数据压缩存取算法进行数据压缩,有效减小了数据库单表存储空间、减轻存储压力,同时由于行数较少可以更加快速地定位到查询目录以提供较快的查询效率,符合计算机“以空间换时间”的优化策略要求,可以缩小程序的内存占用,从而提升采煤机数据管理效率。

## 4 结 论

1)针对不同型号的采煤机数据通讯协议点表不一致的问题,以CAN通讯协议点表为主要研究对象,构建了采煤机通讯协议树模型,提出了基于子树匹配的采煤机通讯协议树相似度计算算法以及采煤机通讯协议匹配算法,并对算法进行实验分析。实验结果表明:所提出的基于子树匹配的采煤机通讯协议相似度计算方法与专家参考值一致,匹配结果吻合度可达100%,满足通讯协议匹配的准确性要求。

2)针对采煤机配备多路总线、数据帧发送间隔短造成实时数据量庞大的问题,对采煤机实时数据进行了数据分类,提出了针对开关量数据的相同状态字压缩算法以及针对模拟量数据的变长编码压缩算法,并对算法进行实验分析。实验结果表明:相比于LZ78算法以及Gzip算法两常用的数据压缩方法,本文所提出的采煤机实时数据压缩存取算法可以提供较高的压缩比。

3)设计并开发了一个集成通讯协议匹配和数据压缩存储的采煤机远程监控系统,并对其进行系统测试。测试结果表明:所研发的采煤机远程监控系统在通讯协议匹配方面的成功率可达100%;在数据

压缩方面,面向采煤机开关量和模拟量的压缩率分别达到了99.16%和91.8%,满足采煤机实时数据管理技术的需求。该系统的建立提高了使煤矿企业能够更好地监测和管理采煤机运行状态,为生产决策提供可靠的数据支持。

## 参考文献(References):

- [1] 国家统计局.中华人民共和国2023年国民经济和社会发展统计公报[N].人民日报,2024-03-01(010).
- [2] 王国法.煤矿智能化最新技术进展与问题探讨[J].煤炭科学技术,2022,50(1):1-27.  
WANG Guofa. New technological progress of coal mine intelligence and its problems[J]. Coal Science and Technology, 2022, 50(1): 1-27.
- [3] 连东辉,袁瑞甫,崔科飞,等.综采工作面自动化生产工艺研究与应用[J/OL].煤炭科学技术,1-11[2024-10-31]. <http://kns.cnki.net/kcms/detail/11.2402.TD.20240613.1512.005.html>.
- [4] LI C Q, CHEN Y Q, SHANG Y L. A review of industrial big data for decision making in intelligent manufacturing[J]. Engineering Science and Technology, An International Journal, 2022, 29: 101021.
- [5] 徐卫鹏.我国采煤机技术创新发展研究[J].煤矿机械,2021, 42(1): 27-29.  
XU Weipeng. Study on innovation and development of shearer technology in China[J]. Coal Mine Machinery, 2021, 42(1): 27-29.
- [6] 赵亦辉,赵友军,周展.综采工作面采煤机智能化技术研究现状[J].工矿自动化,2022,48(2): 11-18, 28.  
ZHAO Yihui, ZHAO Youjun, ZHOU Zhan. Research status of intelligent technology of shearer in fully mechanized working face[J]. Industry and Mine Automation, 2022, 48(2): 11-18, 28.
- [7] 王国法,杜毅博,徐亚军,等.中国煤炭开采技术及装备50年发展与创新实践:纪念《煤炭科学技术》创刊50周年[J].煤炭科学技术,2023,51(1): 1-18.  
WANG Guofa, DU Yibo, XU Yajun, et al. Development and innovation practice of China coal mining technology and equipment for 50 years: commemorate the 50th anniversary of the publication of coal science and technology[J]. Coal Science and Technology, 2023, 51(1): 1-18.
- [8] 葛世荣.采煤机技术发展历程(五):自动化技术[J].中国煤炭,2020,46(10): 1-15.  
GE Shirong. The development history of coal shearer technology (part five): automation technology[J]. China Coal, 2020, 46(10): 1-15.
- [9] 袁亮,张平松.煤炭精准开采地质保障技术的发展现状及展望[J].煤炭学报,2019,44(8): 2277-2284.

YUAN Liang, ZHANG Pingsong. Development status and prospect of geological guarantee technology for precise coal mining[J]. Journal of China Coal Society, 2019, 44(8): 2277–2284.

[10] 李福兴, 李璐熳. 面向煤炭开采的大数据处理平台构建关键技术[J]. 煤炭学报, 2019, 44(S1): 362–369.

LI Fuxing, LI Luxi. Key technologies of big data processing platform construction for coal mining[J]. Journal of China Coal Society, 2019, 44(S1): 362–369.

[11] 汪佳彪, 王忠宾, 张霖, 等. 基于以太网和 CAN 总线的液压支架电液控制系统研究[J]. 煤炭学报, 2016, 41(6): 1575–1581.

WANG Jiabiao, WANG Zhongbin, ZHANG Lin, et al. Research on electro-hydraulic control system of hydraulic support based on Ethernet and CAN-Bus[J]. Journal of China Coal Society, 2016, 41(6): 1575–1581.

[12] 薛霜思, 曹晖, 贾立新, 等. 矿用电机分布式远程智能在线监测系统设计[J]. 煤炭学报, 2023, 48(S1): 368–380.

XUE Shuangsi, CAO Hui, JIA Lixin, et al. Design of distributed remote intelligent online monitoring system for mining motors[J]. Journal of China Coal Society, 2023, 48(S1): 368–380.

[13] 张林强. 煤矿掘进机新型智能电控系统设计与研究[J]. 煤炭技术, 2024, 43(1): 245–248.

ZHANG Linqiang. Design and research of new intelligent electric control system for coal mine boring machine[J]. Coal Technology, 2024, 43(1): 245–248.

[14] 赵海伟. 基于 CANopen 的煤矿多设备协同控制方法[J]. 煤炭技术, 2022, 41(8): 221–223.

ZHAO Haiwei. CANopen-based coal mine multi-equipment linkage control method[J]. Coal Technology, 2022, 41(8): 221–223.

[15] 徐化岩, 初彦龙. 基于 influxDB 的工业时序数据库引擎设计[J]. 计算机应用与软件, 2019, 36(9): 33–36, 40.

XU Huayan, CHU Yanlong. Design of industrial time series database engine based on influxdb[J]. Computer Applications and Software, 2019, 36(9): 33–36, 40.

[16] 丁小欧, 于晨健, 王沐贤, 等. 基于相关性分析的工业时序数据异常检测[J]. 软件学报, 2020, 31(3): 726–747.

DING Xiaou, YU Shengjian, WANG Muxian, et al. Anomaly detection on industrial time series based on correlation analysis[J]. Journal of Software, 2020, 31(3): 726–747.

[17] 丁远. 基于无主通信方式的煤矿监控系统传感器关键技术[J]. 煤矿安全, 2023, 54(6): 207–211.

DING Yuan. Key technology of sensor in coal mine monitoring system based on hostless communication[J]. Safety in Coal Mines, 2023, 54(6): 207–211.

[18] 符大利. 基于数字孪生驱动的综采工作面远程监控系统[J]. 煤炭技术, 2022, 41(4): 175–178.

FU Dali. Remote Monitoring System of Fully Mechanized Mining Face Based on Digital Twin Drive[J]. Coal Technology, 2022, 41(4): 175–178.

[19] 苗青. 多种方式数据采集及通讯技术的需求分析[J]. 工业控制计算机, 2023, 36(10): 124–125, 142.

MIAO Qing. Requirement analysis of various data collection and communication technologies[J]. Industrial Control Computer, 2023, 36(10): 124–125, 142.

[20] 张丽霞. 数据开放平台无线传感网络能耗在线监测方法[J]. 信息技术, 2024, 48(3): 70–74, 82.

ZHANG Lixia. Online monitoring method for energy consumption of wireless sensor network on data open platform[J]. Information Technology, 2024, 48(3): 70–74, 82.