

基于深度数据挖掘和区块链技术的智库数据管控方法研究

韦冬妮, 齐彩娟, 张泽龙, 唐梦媛

(国网宁夏电力有限公司 经济技术研究院, 宁夏 银川 750002)

摘要:针对电力企业尚未专业、系统地开展智库建设,且缺乏完善的智库运营管理机制等问题,文中展开了基于深度数据挖掘和区块链技术的智库数据管控方法研究。文中构建的智库管控系统包括数据层、支撑层、业务层和表现层;该系统对智库数据进行深度数据挖掘,利用K-means聚类算法分类数据特征,并采用深度信念网络获得数据间的关联。利用区块链技术完成智库数据管控全过程,以保证数据的安全性。基于B/S协议和Matlab仿真计算平台对所提方法进行实验,测试结果表明,当数据量为8 MB时,其挖掘时间和准确率分别为1.5 s和93%,优于其他对比方法,而且安全性也更为理想。

关键词:智库数据管控;区块链技术;深度数据挖掘;深度信念网络;K-means聚类算法;B/S协议

中图分类号: TN248

文献标识码: A

文章编号: 1674-6236(2022)13-0114-05

DOI: 10.14022/j.issn1674-6236.2022.13.024

Research on think tank data management and control method based on deep data mining and blockchain technology

WEI Dongni, QI Caijuan, ZHANG Zelong, TANG Mengyuan

(Economic and Technological Research Institute, State Grid Ningxia Electric Power Co., Ltd., Yinchuan 750002, China)

Abstract: In view of the lack of professional and systematic think tank construction and perfect think tank operation and management mechanism in power enterprises, this paper studies the think tank data management and control method based on deep data mining and blockchain technology. The think tank management and control system is constructed, including data layer, support layer, business layer and presentation layer; The think tank data is deeply mined, and the K-means clustering algorithm is used to classify the data features, and the deep belief network is used to obtain the association between the data. Blockchain technology is used to complete the whole process of think tank data management and control to ensure data security. Based on B/S protocol and Matlab simulation platform, the proposed method is tested. The test results show that when the data volume is 8 MB, the mining time and accuracy are 1.5 s and 93% respectively, which are better than other comparison methods, and the security is also more ideal.

Keywords: think tank data management and control; blockchain technology; deep data mining; deep belief network; K-means clustering algorithm; B/S protocol

收稿日期: 2021-03-25 稿件编号: 202103229

基金项目: 国网宁夏电力企业智库建设方法研究项目(B007-300009250-00014)

作者简介: 韦冬妮(1975—),女,甘肃定西人,教授级高级工程师。研究方向: 企业战略研究。

-114-

智库作为国家软实力的重要组成部分,各行业需积极探索中国特色新型智库的组织形式和管理方式^[1]。随着大数据、人工智能、云计算、5G等新技术的不断成熟,将其应用于智库领域,能够提升数据管控效率并促进智库长足发展^[2-3]。目前,国外一流智库数据管理模式及其特点,对我国智库数据管理、决策咨询服务和可持续发展具有重大意义^[4]。

智库主要分为以自用为主和以整合外部资源为主的两种数据管控模式^[5-6],在国家电网“1+1+N”的特色专业智库体系建设思路下,省级电力公司企业因地制宜推进了本单位智库建设,积极完善智库运营管控机制并发挥决策支持作用。但当下部分省级供电公司尚未专业、系统地开展省级智库建设,在智库团队组建、平台搭建、成果转化等方面仍不成熟,亟需通过建立数字化、智能化的智库管控系统支撑智库体系建设^[7]。

该文提出了一种智库数据管控方法,通过深度数据挖掘算法获得数据层信息的特征及相关联系,以供业务层使用。整个数据管控过程在区块链中进行,确保智库数据的安全可靠。

1 智库管控系统设计

智库管控系统是基于企业智库的运营机制与各核心要素的管理需求,采用B/S体系架构开发设计的。其主要分为数据层、支撑层、业务层和表现层4个层次,整体架构如图1所示。

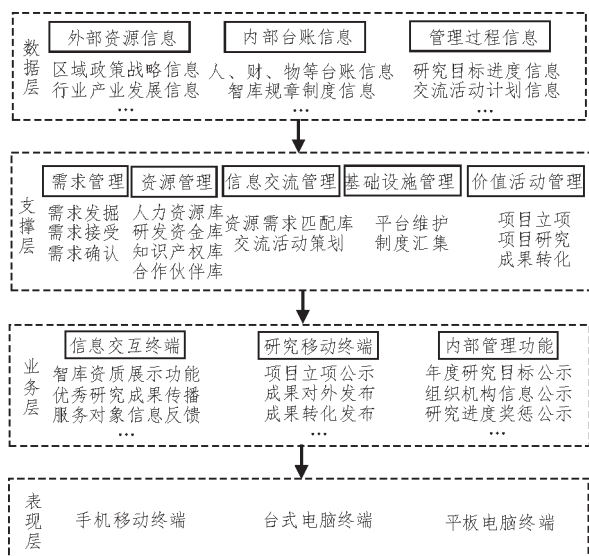


图1 智库管控系统的整体架构

其中,数据层采用Oracle、SQL等数据库获取外

部资源信息、内部台账信息、管理过程信息等,依托电力物联网采集内外部数据打通能源电力发展与区域经济发展链接的渠道^[8]。收集和储存智库服务对象的需求信息、区域经济及行业产业发展最新动向、内部管理台账以及在研究与管理过程中产生的各类相关信息,实现省级电力公司智库业务层面和管理层面信息的全面集约。

支撑层明确智库管控的具体操作包括:需求的发掘、接收、确认;人力资源、研发资金、知识产权的台账化管控;资源需求匹配与交流活动策划等信息交流管控;平台硬件和制度建设等基础设施管理;项目立项、项目研究、成果转化等课题研究价值活动管控;智库阶段性发展目标、智库组织机构管理等^[9]。

业务层是全面支撑省级电力公司企业智库业务的功能模块,通过搭建信息交互、研究推进、内部管理三大业务平台,实现内外部业务交互促进。

表现层是将系统架构部署于各类终端,以方便服务对象、社会大众、研究员、智库管理人员等不同角色进行平台操作。如:服务对象通过手机移动终端、电脑桌面开展信息交互;智库管理人员通过电脑桌面或平板查看研究推进情况等。

2 智库数据管控方法

在所提的智库数据管控方法中,通过深度数据挖掘算法获得数据层信息的特征及相关联系,供业务层使用。整个数据管控过程在区块链中进行,确保智库数据的安全可靠。

2.1 区块链技术

区块链可称为分布式账本,其中数据和交易不受任何第三方的控制。区块链是由Satoshi Nakamoto作为比特币的核心部分首次引入,随着技术的发展,出现了适用于不同目标的区块链。但所有区块链均存在一部分共同元素,总结如下:

1)复制账本:区块链网络中的所有节点安全地存储在交易历史数据中,最新的事务被打包到一个区块,其中所有事务均可参与网络所有节点之间的分布和复制^[10]。

2)点对点网络:所有节点共享一个公共账本,集中控制参与者不通过互联网接入。换言之,所有节点均通过对等网络连接,事务与区块通过该网络同步。

3)共识:在区块插入链之前,网络中所有节点需要就区块内交易的有效性和顺序达成共识。最具代

表性的共识算法有工作证明(Proof of Work, PoW)、权益证明(Proof of Stake, PoS)和实用拜占庭容错算法(Practical Byzantine Fault Tolerance, PBFT)等。

4)密码学:区块链网络的安全性是基于密码学知识实现的。在区块链网络中,交易的完整性应包括数字签名和专有数据结构(如Merkle树、Merkle Patricia树)。此外,数字签名保证交易的真实性,非对称密码体制保证交易的隐私性^[11]。

区块链是按照时间顺序将数据区块进行连接,从而组成链式结构。每个区块由区块头和区块体组成^[12],如图2所示。

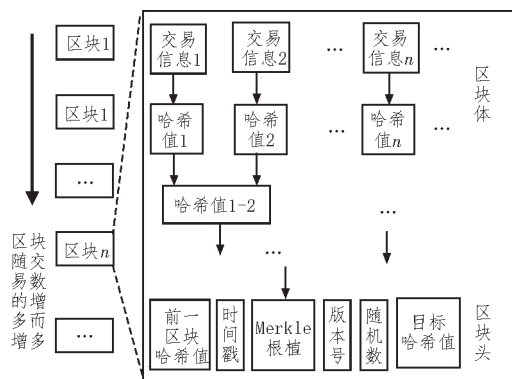


图2 区块链结构

其中,区块头主要是由数据区块内大多数的验证信息组成,如版本号、时间戳、区块的Merkle根;区块体主要为电力物联网采集的内外部数据^[13]。基于区块链进行数据管控,密钥自身与别的节点均间接交互,这样能够防止交互环节中遭遇网络攻击而造成密钥外泻,确保了数据的安全性^[14]。此外,数据只包括数字签名,根本上缩减了区块链的数据传输量。因此可减少构建区块链的耗时,保证数据的时效性。

2.2 深度数据挖掘

数据挖掘主要由采集、存储、剖析、挖掘、可视化及数据融合等环节组成。在数据深度挖掘过程中,首先是利用各种传感器获取海量多源的数据,且经过去噪等方式选取符合要求的数据构成数据集进行存储。然后对数据集进行处理和剖析,按照相应的规则将数据集分类,同时剖析数据之间与数据类型之间所存在的关联等^[15]。紧接着,将完成分类的数据进行深层次挖掘,采用各种学习算法获得数据之间的内部关联、隐藏的信息结构等。最终将关联、结构等进行可视化,直接的呈现方式便于用户理解,且将相联系的类型予以融合,以进行充分的数

据利用^[16]。

2.2.1 K-means聚类算法

K-means算法一般用于数据分类,以欧氏距离作为相似度的评价指标。以任意 k 个点作为初始聚类核心,再计算其他节点到 k 个聚类核心的距离,并将其划分至距离最小的类型中。随后统计各个类型中的均值,以完成类核心的迭代。重复此过程,直至平方误差准则函数趋于极小值。

设对象集合 $M=\{x_1, x_2, \dots, x_n\}$, $x_i=\{x_{i1}, x_{i2}, \dots, x_{in}\}$, 样本 x_i 与样本 x_j 的欧式距离计算公式如下:

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2} \quad (1)$$

平方准则误差函数如下:

$$\vartheta_c = \sum_{i=1}^k \sum_{j=1}^{t_i} \|x_j - n_i\|^2 \quad (2)$$

在第 i 类数据中心, t_i 为样本个数, n_i 为均值。

2.2.2 深度学习算法

采用K-means算法对智库数据特征进行聚类处理,从而完成智库数据特征的筛选,为智库数据的深度挖掘提供依据。在智库数据深度挖掘的过程中,利用深度置信网络(Deep Belief Network, DBN)学习算法分析数据间关系,以便将有关联的类别进行融合。

DBN作为一种非监督深度学习神经网络,主动学习输入数据并自动挖掘隐藏在已知数据中的信息,可被视为玻尔兹曼机(Restricted Boltzmann Machines, RBM)的堆栈,RBM是一个基于能量的生成型模型。每个RBM由可视层和隐藏层构成,可视层负责数据输入,而隐藏层负责特征提取。层与层之间全连接,同一层各单元彼此互不连接,DBN是由多个RBM堆叠而成的模型。

设RBM状态为 (v, h) ,其能量定义为:

$$E(v, h|\theta) = -\sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m v_i W_{ij} h_j \quad (3)$$

式中, v_i 、 h_j 分别为可视层与隐藏层的状态向量; θ 为RBM的参数; W_{ij} 为可视层和隐藏层的连接权值; a_i 、 b_j 分别为可视层与隐藏层的偏置向量。

一旦明确参数,根据能量函数能够获得 (v, h) 的联合概率分布:

$$P(v, h|\theta) = \frac{1}{\mathcal{Z}} e^{-E(v, h|\theta)} \quad (4)$$

式中, $\partial(\theta)$ 为归一化因子。

3 实验结果与分析

实验中智库系统采用 B/S 协议设计,数据挖掘算法采用 Matlab 编程实现,仿真实验环境配置为: Win10 操作系统,48 GB 内存,AMD A10-5750 M2.50 GHz CPU,并采用分类时间和准确率、安全性作为性能指标评估所提方法。

3.1 数据管控性能分析

在所提的智库数据管控系统中,所有数据均是通过区块链技术进行动态更新,以提高数据的安全性。为了验证所提方法在挖掘时间与准确率两个方面的性能,将其与文献[3]、文献[6]、文献[10]进行对比分析,对比结果如图3所示。

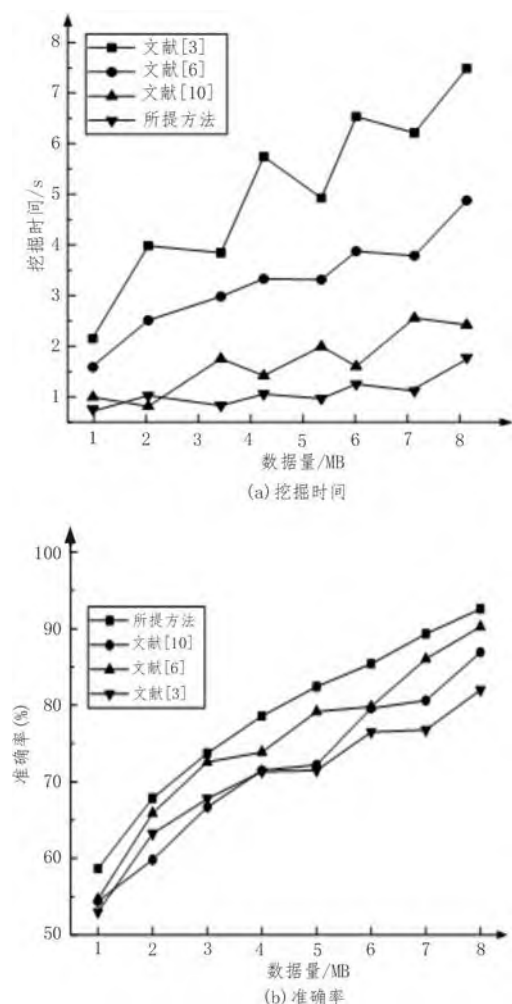


图3 不同智库数据管控方法的挖掘时间和准确率的对比结果

从图3中可看出,挖掘时间随着数据量的增加而增加,但所提方法的挖掘时间明显低于其他对比方法。由于其采用区块链技术,可分布式进行挖掘且数

据挖掘中采用 DBN 算法进一步节约了挖掘时间。当数据量为 8 MB 时,所提方法耗用的时间仅为 1.6 s。

此外,所有方法的分析准确率随着数据量的增加均有所提升。相比于其他方法,所提方法在区块链的基础上,利用 DBN 算法进行深度挖掘,能够更准确地获得数据特征并将其进行应用。因此当数据达到 8 MB 时,分析准确率达到 93%。从综合挖掘时间与准确率两方面来看,所提方法的数据管控能力是最佳的,能有效应用于智库系统。

3.2 安全性分析

安全性是智库数据管控的一个重要指标,将所提方法与文献[3]、文献[6]、文献[10]进行对比分析。在不同数据量下,各方法的安全性能如表1所示。

表1 不同数据管控安全性对比结果

| 文件大小/kB | 安全性(%) | | | |
|---------|--------|-------|--------|------|
| | 文献[3] | 文献[6] | 文献[10] | 所提方法 |
| 4 | 86 | 90 | 92 | 98 |
| 8 | 84 | 89 | 91 | 97 |
| 16 | 79 | 83 | 86 | 92 |
| 32 | 70 | 75 | 81 | 90 |
| 64 | 68 | 71 | 79 | 88 |
| 128 | 60 | 65 | 72 | 84 |
| 512 | 59 | 61 | 68 | 80 |

从表1中可看出,所提方法的安全性能明显高于其他对比方法,当文件大小为 512 kB 时,仍不低于 80%。随着文件大小的增加,安全性在不断降低,这是因为文件越大,需要管控的数据越多,也就面临更高的安全风险,因此智库系统对管控方法的安全性提出了更高的要求。文献[6]中采用知识管理进行数据管控,但缺乏相应的数据安全防护。文献[10]基于大数据规则挖掘和区块链通信技术进行交互,但应用场合缺乏智库研究,不适用于智库中海量的数据。而所提方法采用区块链技术保证智库数据管控的全程安全,并结合深度挖掘技术应对系统海量的数据,因此所提方法的安全性能最为理想。

4 结束语

智库作为经济高质量发展引擎、行业高质量发展支撑,对决策者与社会产生着重要影响,是新型智库体系的重要组成部分。该文在构建的智库管控系统中,利用 K-means 聚类算法分类数据特征,并采用 DBN 进行智库数据的深度挖掘。同时整个智库管控过程在区块链网络中进行,保证数据的安全性。实

验结果表明,该文所提方法的挖掘时间、准确率和安全性均优于对比方法,因此建立智库推动国家能源电力及经济高质量发展具有重要意义。

虽然区块链技术能保证智库数据管控过程安全、高效,但该技术也存在一定的弊端,例如区块链是不可变的,只能进行追加操作。因此区块链的存储空间会不断增长,其分布式存储的特点造成了资源浪费。此外,网络拥塞、块大小、同步机制等问题也是接下来研究的热点。

参考文献:

- [1] 江红玲.基于DIIS理论方法的国防科技智库研究体系建设[J].中国电子科学研究院学报,2020,15(8):750-755.
- [2] Lacan A.Think tank:meaningful management in a postmodern society[J].Global Business and Organizational Excellence,2019,38(3):6-10.
- [3] 潘卓齐.基于关联数据的科技智库数据共享机制研究[D].广州:华南理工大学,2019.
- [4] 陈海贝,卓翔芝.动态惩罚机制下智库行为及其监管策略的演化博弈分析[J].淮北师范大学学报(自然科学版),2019,40(2):60-68.
- [5] Ma W, Ma L, Zhang J, et al.Theoretical framework and realization pathway of agricultural green development[J].Chinese Journal of ECO-Agriculture, 2020,28(8):1103-1112.
- [6] 曾燕.协同工作流驱动的知识管理系统研究与设计[D].南京:南京理工大学,2017.
- [7] Seffers G I.Think tank pushes for cybersecurity renaissance[J].Signal,2019,73(8):24-26.
- [8] 纪志伟.基于数字孪生的泛在电力物联网模型研究[J].电力学报,2020,35(3):274-279.
- [9] 牛洪杰.基于综合指数法的工程项目管理成果评价研究[D].保定:华北电力大学,2019.
- [10] 王国军,胡静,汪瑾.基于区块链分布优化的大数据融合QM模型研究[J].微型电脑应用,2020,36(12):113-116.
- [11] Turesson H K, Kim H, Laskowski M, et al.Privacy preserving data mining as proof of useful work: exploring an AI/blockchain design[J].Journal of Database Management,2021,32(1):69-85.
- [12] Chinnasamy P, Deepalakshmi P, Praveena V, et al. Blockchain technology:a step towards sustainable development[J].International Journal of Innovative Technology and Exploring Engineering,2019,9(2): 1034-1040.
- [13] 黄虹,文康珍,刘璇,等.泛在电力物联网背景下基于联盟区块链的电力交易方法[J].电力系统保护与控制,2020,48(3):22-28.
- [14] 田道坤,彭亚雄.在区块链中基于混合算法的数字签名技术[J].电子科技,2018,31(7):19-23.
- [15] Nuzzolese E.Electronic health record and blockchain architecture:forensic chain hypothesis for human identification[J].Egyptian Journal of Forensic Sciences,2020,35(10):1-5.
- [16] 杨井荣,侯向宁.正负关联规则数据挖掘算法研究[J].计算机技术与发展,2020,30(11):64-68.

欢迎投稿! 欢迎订阅! 欢迎刊登广告!

国内刊号: CN61-1477/TN

国际刊号: ISSN 1674-6236

在线投稿系统: <http://mag.ieechina.com>

dzsjgc@vip.163.com (广告)

地址: 西安市劳动南路210号5-1-3信箱

邮政编码: 710082