

企业电力征信大数据价值挖掘与应用

辛保江, 李德文, 王兰兰

国网山东省电力公司潍坊供电公司, 山东 潍坊 261000

摘要

针对传统电力征信平台稳定性不足、测试准确性低等缺点, 研究设计了一个电力征信大数据平台。使用联机分析法对电力大数据进行分析, 并将其分为用户行为、费用细则、用户价值与个人信用四大类。以模块化结构为基础, 分别对数据采集模块、数据分析模块、用户交互模块进行优化设计, 采用KNN算法和交叉验证法对用电数据进行分类与决策处理, 得出区域的用电规律, 以此设计和调整配电方案。最后将提出的平台与传统电力征信平台进行对比, 实验结果表明, 提出的平台的稳定性和准确性都有所提升, 在测试过程中准确性高达98.9%。

关键词

大数据模型; 电力征信; 综合型算法; 网络架构; 系统测试

中图分类号: U495

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2021067

Value mining and application of big data in enterprise power credit investigation

XIN Baojiang, LI Dewen, WANG Lanlan

State Grid Shandong Electric Power Company Weifang Power Supply Company, Weifang 261000, China

Abstract

Aiming at the shortcomings of traditional power credit investigation platform such as insufficient stability and low test accuracy, a big data power credit investigation platform was designed. The online analytical method was used to analyze power big data, which was divided into four categories: user behavior, expense rules, user value and personal credit. Based on the modular structure, the data acquisition module, data analysis module and user interaction module were optimized respectively. The KNN and cross validation method were used to classify and process the power consumption data, and the regional power consumption law was obtained, so as to design and adjust the distribution scheme. Finally, the platform was compared with the traditional power credit investigation platform. The experimental results show that the stability and accuracy of the platform are improved, and the accuracy is as high as 98.9% during the test.

Key words

big data model, power credit investigation, comprehensive algorithm, network architecture, system testing

1 引言

中国是全世界唯一实现全民通电的国家^[1-5]。在电力技术高速发展的今天,电力成为一种不可或缺的资源,我国的电力消费一直秉承着“先使用后付款”的收费方式,但在用户拖欠电费、违约用电方面没有一个很好的解决方案^[3]。电力征信的应用能够有效约束信用不良的用户,减少违约用户的产生,提升供电公司的管理水平。

参考文献[1]设计了基于9项指标的电力征信检测平台,根据用户的用电量、用电缴费情况和违约用电记录等指标对用户进行分类,增强管理的条理性,但数据采集方式过于老旧,没有正确建立用户的用电模型。参考文献[2]提出利用大数据平台对电力征信用户数据进行提取与处理,这一做法能保证数据分析的客观性与准确性,但系统稳定性却没有得到保证。本研究针对电力征信平台进行创新,设计电力征信大数据平台,基于平台稳定性与测试准确性进行优化。

2 电力征信关键技术

2.1 用户用电模型构建

在大数据模型技术中,联机分析法是数据分析的主要方法,通过联机分析法,大数据信息能够在信息平台上进行多维传播^[6]。将联机分析法运用到电力征信领域,对电力用户的用电行为进行收集与分析,从而为电力企业供电计划的制订提供便利^[7-10]。用电数据处理流程如图1所示。

如图1所示,首先对用电用户的数据进行提取,将用户的用电行为进行分类,然后记录分类数据,同时对用户数据进行行为分析,最终审核数据是否有遗漏或者错误的地方,如果有,则需要对数据进行再次提取。传统记录主要提取用户的9类数据,本研究针对用户的数据分析,将数据提取分为四大类、八小类,使得数据处理更加条理化,同时加快数据采集工作的开展与完成。数据分类如图2所示。

如图2所示,将数据分为用户行为、费用细则、用户价值与个人信用四大类。此外,还可以将上述数据信息分为用电情况、费用情况、信用情况与用户评价,为了表达方便,本文使用“用户行为、费用细则、用

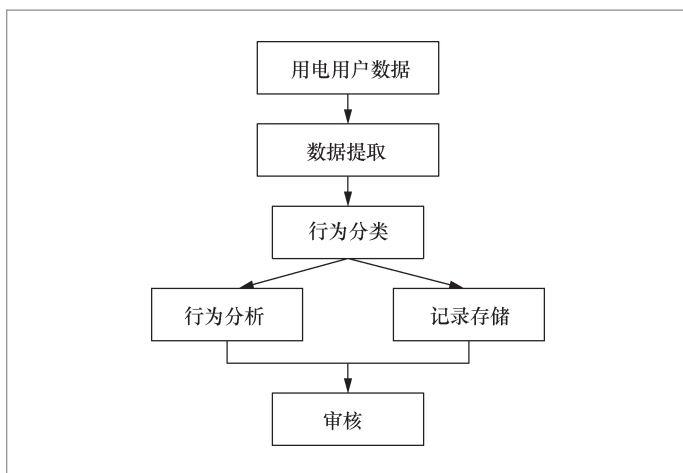


图1 用电数据处理流程

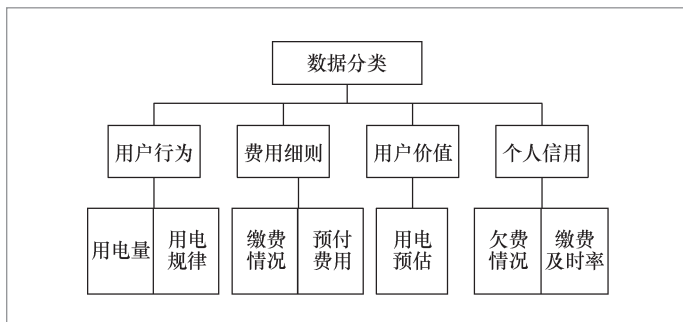


图2 数据分类

户价值、个人信用”进行说明。然后将用户行为细分为该用户的月度、季度、年度用电量与用电规律,将费用细则细分为用户的缴费情况和预付费用情况;将用户价值细分为对未来用户用电情况的预估;针对个人信用方向,记录用户欠费情况与缴费及时率,通过对此4项数据的采集,能够完整地构建用户的用电模型,有利于对未来电力分布、企业管理和用户信用进行综合分析。

对以上几类数据进行分析处理时,还可根据分析结果将用户群分为尊贵用户、普通用户、风险用户和失信用户,针对不同的用户提供不同的处理方式,从而加强电力企业对用户的管理力度,对电力企业的绩效分析起到促进作用^[11]。

2.2 模块化设计优化

传统电力征信平台通过模块化的设计将硬件进行组合,但这种模块化设计在数据处理过程中出现错误数据的概率大于需求的概率阈值^[12]。因此,在传统数据采集过程中需要分配人力资源对数据进行预先审查,这一过程中人力与时间耗费巨大,且人工审核的准确率也无法达到预期^[13]。因此针对这一缺点,本研究在传统模块化的基础上进行优化创新,主要将数据采集模块与网络模块进行交互,使得网络数据能够对采集到的数据进行校正;将设备管理模块中的硬件检测部分应用于数据分析模块,提高数据分析的速度,提升数据分析的效率^[14];在用户交互模块中,添加软件后门进行二次开发,便于及时应对电力资源需求变化过快的情况,提高平台稳定性^[15]。平台模块化设计如图3所示。

本研究所提电力征信大数据平台的创新点为针对硬件方面的模块化创新,其中数据存储方面为固定存储,当网络模块检

测到错误数据时,可将数据存储模块中存储的同类数据进行替代,针对替代数据进行分类。在实验中可以明显发现,此方法虽然会在一定程度上产生误差,但是其对结果的影响远没有输入错误数据造成的误差大,因此本研究针对模块化设计的创新具有技术优越性。

面对传统电力征信平台的功能与优缺点,本研究对电力征信平台进行改进。下面将从网络架构设计与算法设计两部分对本研究平台的创新点进行分析。

3 电力征信大数据平台设计

3.1 网络架构设计

本文设计的平台从电力行业的基础出发,与目前主流的大数据技术框架相吻合。针对传统电力征信平台计算准确性不足、电力征信大数据分析系统对数据分析的客观性不够的缺点,本研究通过对电力征信平台的网络架构进行重新整合,将数据监测与营销分析、运行管理等作为服务层,面向电力企业提供真实的业务服务^[11-17]。同时在数据计算方面采用专门针对电力信息分析的综合型算法,通过综合型算法得到用户的用电模型,将此模型通过数据接口应用于用电用户(即与用电相关的企业用户),为他们提供良好的用电方案与维护方案。本研究设计的平台整合了电力信息接入、用户数据分类治理、用户模型搭建与模型分析展示等功能,集成了多图形的可视化建模界面与二次开发接口拓展^[18]。整体网络架构如图4所示。

在网络架构中,对数据进行过滤,主要通过通过对数据的筛选与对关联数据的整

合,将用电数据通过表格和饼状图等可视化视图直观地展现出来。数据过滤流程如图5所示。

如图5所示,针对源表,即用户表的信息,可以选择将用户表作为基础,将用户表分解为区域表和设备表,区域表记录一个区域间的用电数据,将不同区域进行对比,能够很直观地看出区域与区域之间的用电差距;设备表主要记录电流与电压的监测值,根据电流与电压的变化预测用电量的变化趋势,具体如图6所示。

3.2 算法设计

针对本研究设计的电力征信大数据平台的网络架构,设计一种将数据进行分类、聚类、回归与集成的算法,即创新的综合型算法^[19]。给定一个电力征信平台的用户信息集,首先使用分类聚类算法将此信息集分为用户行为、费用细则、用户价值与个人信用四大类,这里采用 K 近邻(K -nearest neighbor, KNN)算法^[20-24]。 K 值根据电力信息数据集规模的不同进行选取,比如一个区域内的用电度数与用电时间不同, K 值也会随其不同而发生变化,在 K 值的选择上使用交叉验证的方式,即将不同的样本集交叉起来验证 K 值的方差,从而求得最佳的 K 值, K 值交叉验证如图7所示。

从图7可以明显看出,开始时随着 K 值的增大,误码率由高向低变化;当 K 值大于10之后,误码率开始变大,因此要选择样本能承受的最佳 K 值。此算法除了要注意 K 值的选择,还要考虑点与点之间的距离的计算,给定两个电力信息的数据点 (x_1, y_1) 与 (x_2, y_2) ,则这两点之间的距离 ρ 的计算式为:

$$\rho = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

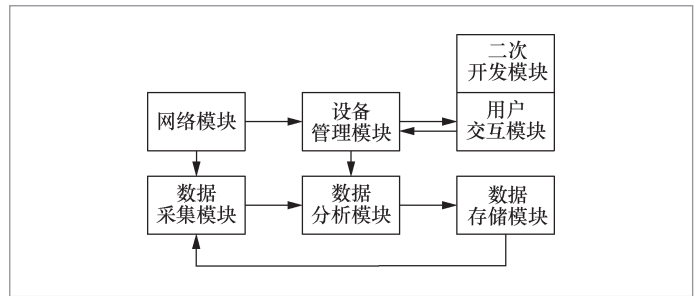


图3 平台模块化设计

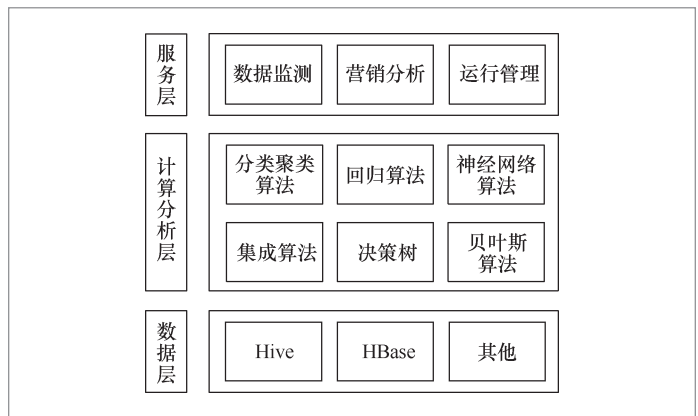


图4 整体网络架构

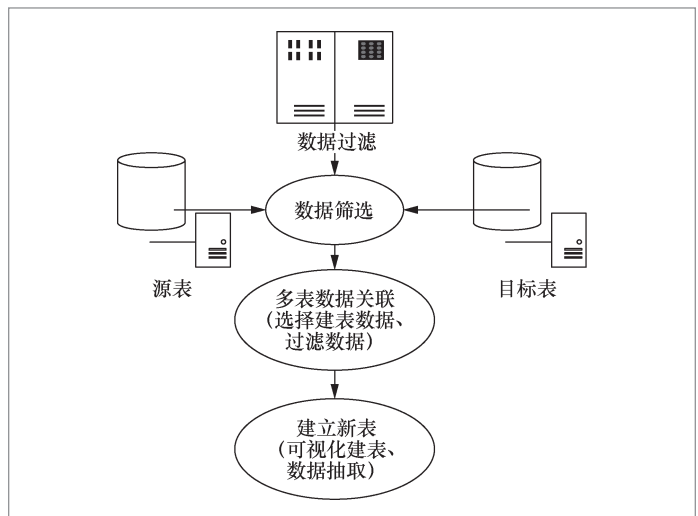


图5 数据过滤流程

通过式(1)可以计算出二维平面内点与点之间的距离,但电力信息一般为多维度的数据,因此需要将距离计算式拓展到多维空间,即:

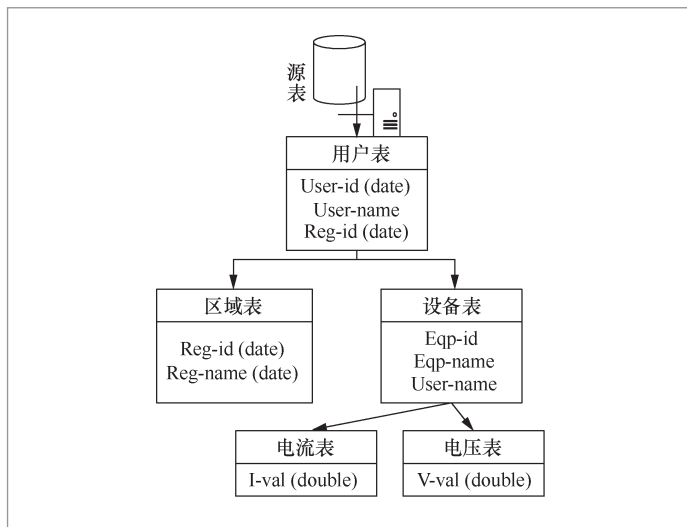


图6 源表数据

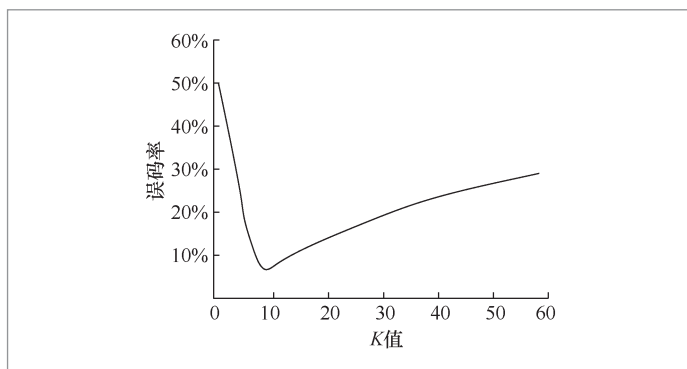


图7 K值交叉验证

$$\rho(x_i, y_i) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

确定好K值并计算完成电力数据集的点间距离后,就能进行数据的分类,将数据分为前文所述四大类后,下一步进行数据的决策,通过数据的决策将分类后的信息进行处理。这里给定电力征信数据集为 D ,计算电力征信数据集 D 的Gini系数,对于电力征信数据集中的每一个用电用户数据 A ,用户数据 A 中包含 a_1 、 a_2 、 a_3 这3种分组数据信息,然后对这3种分组进行处理和计算,得到该电力征信数据集的不同集合。

$$\text{Gini}(D, A) = \frac{a_1}{A} \text{Gini}(a_1) + \frac{a_2}{A} \text{Gini}(a_2) + \frac{a_3}{A} \text{Gini}(a_3) \quad (3)$$

在式(3)中,Gini系数 $\text{Gini}(D, A)$ 指不同分组情况下电力征信用户数据 A 在整个电力征信数据集 D 中的关联集合。假定将电力征信用户数据分为 N 个类别,测试用户数据中的任一随机数据属于第 n 类的概率为 P_n ,则针对电力征信用户模型分类概率因素的Gini系数为:

$$\text{Gini}(P) = \sum_{n=1}^n P_n(1 - P_n) = 1 - \sum_{n=1}^n P_n^2 \quad (4)$$

在式(4)中,将给定区域内的电力征信用户数据,按照用户行为、费用细则、用户价值与个人信用4个方面展开,根据Gini系数对用户未来一段时间的用电情况进行判定,通过对电力征信大数据的分类与决策,得出区域的用电规律,通过此用电规律可以调整整个电力企业服务覆盖区域的配电方案。通过这一设定,在区域内进行雷电预警测试,可根据配电决策树来演化雷电预警概率^[19-24],如图8所示。

4 实验设计与分析

4.1 实验环境及数据

本研究采用的硬件计算机操作系统为64位的Microsoft Windows 10, CPU为Inter(R)Core(TM)i7,主频为2.59 GHz,内存为16 GB。

X市常住人口约1 035万人,用电量在230万kW·h左右,因此本研究采用X市中心区域的用电情况作为模拟对象。在实验过程中将X市中心区域作为测试区域,通过对市中心区域的用电情况进行模拟,比

较本研究用电力征信大数据平台与传统电力征信平台的优缺点,验证本文平台的技术优越性。

4.2 实验设计与实验过程

为了验证本文所设计的平台的技术优越性,设置对照实验,先对平台搭建进行仿真,主要通过对电力征信大数据平台中的网络架构与数据分类模型进行搭建,对比参考文献[1]提出的基于9项指标的电力征信检测平台(后文称为一号平台)与参考文献[2]提出的利用大数据平台对电力征信用户数据进行提取与处理的平台(后文称为二号平台),判断三者对电力征信大数据信息处理过程中的稳定性与准确性。

设计两种实验分别验证本文平台与一号平台、二号平台之间的平台稳定性和测试准确性。针对平台稳定性的测试,本文选用高压测试,使用X市某地停电事故的错误数据集(该数据集为造成停电的各种数据信息集合,比如负荷停电、电力加载、异物等数据信息)对平台分别进行测试,记录平台结果的误码率,对平台测试的稳定性进行分析,同时标定应急用户、重要用户与普通用户,观察系统给出的3种用户的处理方式是否合理。针对准确性的测试,将X市近3年的用电情况输入平台,再将各个平台预测的未来一周用电情况与现实生活中的用电情况进行对比,比较三者的准确性。

4.3 实验结果

为了测试平台稳定性,将测试数据输入平台中,观察3个平台给出的反应。将平台的误码率绘制成折线图,如图9所示。

误码率在本文中是衡量平台测试数据在一定的时间内实现数据传输的精

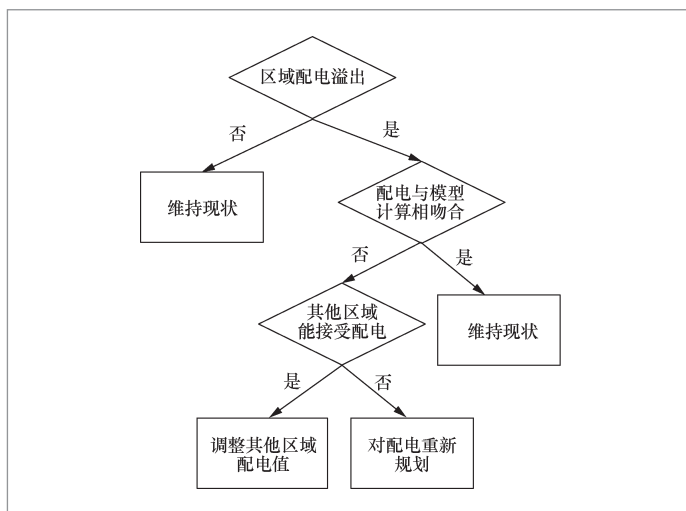


图8 配电决策树

确性指标,计算式为:误码率=传输中的误码/传输的总码数 $\times 100\%$ 。从图9可看出,一号平台误码率较高,在6.5 h的测试结束后,一号平台误码率达到8%,这是因为一号平台没有将错误的处理结果及时筛选出来,导致错误的结果继续向下游运算,进而产生更大的错误;二号平台在1.1 h之前与本研究平台误码率相同,但在1.1 h以后,由于算法处理等问题无法及时更新错误运算,导致误码率逐步提升,在6.5 h时达到4.5%;本文平台由于更新了大数据分类机制,能够有效地控制误码率增长,因此在6.5 h时,误码率仍在1.5%左右,且上升趋势稳定,这表明本文的数据分类方法稳定性较高。

针对准确性的设计,本文将近3年的用电数据作为训练集,先对3个平台进行训练,再对未来一周X市的用电情况进行,将预测结果与真实结果进行对比,引入准确性(在一定实验条件下,经过多次测定后的平均值与真实值相符合的程度),然后对准确性进行分析。具体用电分析情况见表1。

由表1可知,本文所用方案在用电量与缴费预测中都明显优于一号平台与二号平

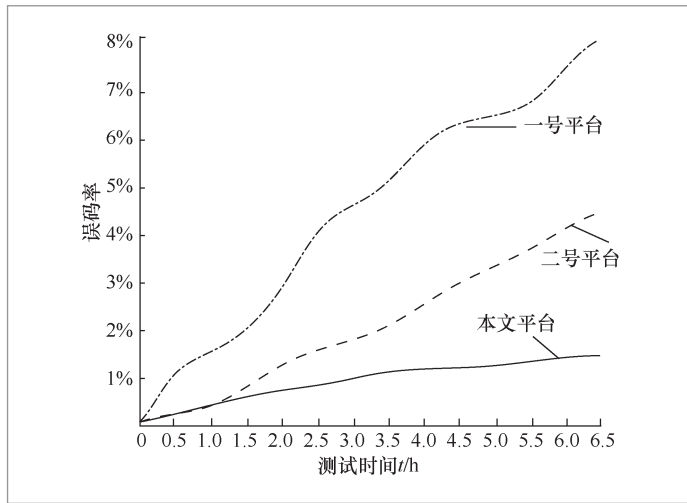


图9 平台误码率测试

表1 用电分析情况

对比项	用电量 预测值/ 万kW·h	用电量 真实值/ 万kW·h	缴费 预测值/ 万元	缴费 真实值/ 万元	准确性
一号平台	7.0	5.2	4.2	3.51	74.2%
二号平台	5.6	5.2	4.0	3.51	85.3%
本文平台	5.0	5.2	3.49	3.51	98.9%

台。在用电预测中,本文平台的预测值与真实值仅相差0.2万kW·h;在缴费预测中,本文平台的预测值与真实值仅相差0.02万元,准确率高达98.9%。该数据信息是经过多次测定后计算出的平均值。

经过上述两种实验的分析,本文平台在稳定性与测试准确性上都明显优于参考文献[1]与参考文献[2]设计的平台,验证了本文平台的技术优越性。

5 结束语

本文基于大数据挖掘进行数据研究,对传统电力征信平台进行改良。利用模块化结构技术,构建电力征信大数据平台的网络架构,以此构建出用户用电模型与区域用电模型,再通过综合型大数据分类决策算法对用

电用户的用电情况进行分类判断,进一步改进了传统电力资源管理平台稳定性不足与准确性低的缺点,同时为电力征信行业的大数据分析提供了理论与实践依据。

参考文献:

- [1] 鲜钊. 基于数据驱动的电力调度故障智能告警系统设计[J]. 自动化与仪器仪表, 2021(1): 144-147.
XIAN Z. Design of intelligent alarm system for power dispatch fault based on data drive[J]. Automation & Instrumentation, 2021(1): 144-147.
- [2] 刘博, 钱勇, 沈阿美. 大数据技术视域下电力配电网智能运维管控系统研究[J]. 工业加热, 2021, 50(1): 46-48.
LIU B, QIAN Y, SHEN A M. Research on intelligent operation and maintenance control system of power distribution network from the perspective of big data technology[J]. Industrial Heating, 2021, 50(1): 46-48.
- [3] 郑斌, 裘炜浩, 侯素颖, 等. 基于电力大数据分析的电力消费指数研究与应用[J]. 电气应用, 2021, 40(1): 66-71.
ZHENG B, QIU W H, HOU S Y, et al. Research and application of power consumption index based on power big data analysis[J]. Electrotechnical Application, 2021, 40(1): 66-71.
- [4] 潘亮亮, 刘志远, 马静. 智能电网电力大数据高性能处理方法优化[J]. 信息技术, 2021, 45(1): 152-156.
PAN L L, LIU Z Y, MA J. Optimization of high performance processing method for smart grid power big data[J]. Information Technology, 2021, 45(1): 152-156.
- [5] 郭文怡. 电力行业信用体系建设与发展[J]. 中国电力企业管理, 2019(28): 64-68.
GUO W Y. The construction and development of the credit system of the electric power industry[J]. China Power

- Enterprise Management, 2019(28): 64-68.
- [6] 贺绍鹏, 谢晓非, 方茂欢. 电力行业垂直产业链一体化交易平台设计与场景研究[J]. 招标采购管理, 2019(2): 37-40.
HE S P, XIE X F, FANG M H. Design and scenario research on the integrated trading platform of the vertical industrial chain of the electric power industry[J]. Tendering & Purchasing Management, 2019(2): 37-40.
- [7] 刘梦雨. 强化电力行业信用体系建设提升市场主体信用资产价值[J]. 中国信用, 2020(5): 95-97.
LIU M Y. Strengthen the construction of the credit system of the power industry and enhance the value of the credit assets of market entities[J]. China Credit, 2020(5): 95-97.
- [8] 王元峰, 王林波, 王冕, 等. 基于大数据的配电设备运行状态监控与故障预警系统设计[J]. 自动化与仪器仪表, 2021(1): 148-151.
WANG Y F, WANG L B, WANG M, et al. Design of operation state monitoring and fault early warning system for distribution equipment based on big data[J]. Automation & Instrumentation, 2021(1): 148-151.
- [9] 吴敬法, 袁震. 基于蚁群算法的电力基建工程物联网全数据管理[J]. 信息技术, 2021, 45(1): 60-66.
WU J F, YUAN Z. Whole data management of Internet of things for power infrastructure projects based on ant colony algorithm[J]. Information Technology, 2021, 45(1): 60-66.
- [10] 李丽华. 大数据在电力信息安全中的实施对策[J]. 科技风, 2021(2): 195-196.
LI L H. Implementation countermeasures of big data in power information security[J]. Technology Wind, 2021(2): 195-196.
- [11] 张文静, 张晶, 邢金峰. 大数据在配电网投资成效分析中的研究应用[J]. 电力大数据, 2017, 20(10): 31-34.
ZHANG W J, ZHANG J, XING J F. Research and application of big data in the effectiveness analysis of distribution grid investment[J]. Power Systems and Big Data, 2017, 20(10): 31-34.
- [12] 田永明, 张润恩, 娄孝东, 等. 面向大数据分析下配网主动式运维分析[J]. 数字通信世界, 2021(1): 127-128.
TIAN Y M, ZHANG R E, LOU X D, et al. Analysis of active operation and maintenance of distribution network for big data analysis[J]. Digital Communication World, 2021(1): 127-128.
- [13] 姚丹靖, 褚燕. 基于大数据分析的电力用户多维价值识别精准营销投入产出模型研究[J]. 电力大数据, 2020, 23(6): 57-62.
YAO D J, CHU Y. Research on multi-dimensional value recognition and precision marketing input-output model of power users based on big data analysis[J]. Power Systems and Big Data, 2020, 23(6): 57-62.
- [14] 张翠翠, 孙佳丽, 洪德华, 等. 大数据自动化运维在电力企业中的应用探讨[J]. 现代工业经济和信息化, 2020, 10(12): 82-83.
ZHANG C C, SUN J L, HONG D H, et al. Application of big data automatic operation and maintenance in power enterprises[J]. Modern Industrial Economy and Informationization, 2020, 10(12): 82-83.
- [15] 袁小凯, 李果, 黄世平. 基于大数据技术的多变量短期电力需求预测研究[J]. 电网与清洁能源, 2020, 36(12): 30-34, 40.
YUAN X K, LI G, HUANG S P. Multi-variable short-term power demand forecasting research based on big data technology[J]. Power System and Clean Energy, 2020, 36(12): 30-34, 40.
- [16] 阴皓, 王督, 郑腾霄, 等. 基于物联网的电力营销管理信息系统的构架设计[J]. 自动化技术与应用, 2020, 39(12): 91-95.
YIN H, WANG D, ZHENG T X, et al. Architecture design of power marketing management information system based on Internet of things[J]. Techniques of Automation and Applications, 2020, 39(12): 91-95.

- [17] 孔庆波. 基于Hadoop平台的电力统计不良数据高效识别方法研究[J]. 电子设计工程, 2020, 28(24): 95-99.
KONG Q B. Research on high efficiency identification method of bad data in power statistics based on Hadoop platform[J]. Electronic Design Engineering, 2020, 28(24): 95-99.
- [18] 陈孝文, 郭威, 王岩, 等. 集中模式下电力信息化数据自动生成系统设计[J]. 电子设计工程, 2020, 28(24): 104-107.
CHEN X W, GUO W, WANG Y, et al. Design of data automatic generation system of electric power information under centralized mode[J]. Electronic Design Engineering, 2020, 28(24): 104-107.
- [19] 罗慧. 基于大数据的电力企业财务数据管理系统研究[J]. 电力大数据, 2019, 22(2): 59-65.
LUO H. Research on financial data management system of power enterprises based on big data[J]. Power Systems and Big Data, 2019, 22(2): 59-65.
- [20] 袁爱进, 岳滨楠, 闫鑫, 等. 工业大数据的应用与实践[J]. 大数据, 2017, 3(6): 27-41.
YUAN A J, YUE B N, YAN X, et al. Application and practice of industrial big data[J]. Big Data Research, 2017, 3(6): 27-41.
- [21] 李俊艳. 基于Hadoop平台的电力大数据聚类算法研究[J]. 数码世界, 2020(12): 71-72.
LI J Y. Research on power big data clustering algorithm based on Hadoop platform[J]. Digital Space, 2020(12): 71-72.
- [22] 王海洋, 赵忠强, 唐建华. 面向电力物联网的电力大数据应用[J]. 电力大数据, 2020, 23(2): 87-92.
WANG H Y, ZHAO Z Q, TANG J H. Power big data application oriented to power Internet of Things[J]. Power Systems and Big Data, 2020, 23(2): 87-92.
- [23] 张小陆. 基于B+树的电力大数据混合索引设计分析[J]. 电子设计工程, 2020, 28(22): 7-10, 15.
ZHANG X L. Design and analysis of hybrid index of power big data based on B+ tree[J]. Electronic Design Engineering, 2020, 28(22): 7-10, 15.
- [24] 孙小虎, 张亚平, 夏同飞, 等. 基于决策树自标识的电网大数据质量控制算法[J]. 电子设计工程, 2020, 28(22): 61-64, 69.
SUN X H, ZHANG Y P, XIA T F, et al. Power grid big data quality control algorithm based on decision tree self-identification[J]. Electronic Design Engineering, 2020, 28(22): 61-64, 69.

作者简介



辛保江(1986-),男,国网山东省电力公司潍坊供电公司副高级工程师,主要研究方向为电力大数据价值挖掘与应用。

李德文(1994-),男,国网山东省电力公司潍坊供电公司助理工程师,主要研究方向为电力大数据应用。

王兰兰(1986-),女,国网山东省电力公司潍坊供电公司经济师,主要研究方向为电力营销服务。

收稿日期: 2021-02-08

通信作者: 辛保江, manmiqian43@163.com

《大数据》2021年（第7卷）

总目次

◎专题：数据驱动的软件智能化开发

- 基于大数据的软件智能化开发方法与环境……………谢冰, 彭鑫, 尹刚, 李宣东, 魏峻, 孙海龙 1 [2021001]
- 基于大数据的软件项目知识图谱构造及问答方法……………邹艳珍, 王敏, 谢冰, 林泽琦 1 [2021002]
- 基于上下文的智能化代码复用推荐……………彭鑫, 陈驰, 林云 1 [2021003]
- 基于大数据的开源项目缺陷报告智能预检技术……………席圣渠, 徐锋, 陈鑫, 李宣东 1 [2021004]
- 基于领域知识的Docker镜像自动构建方法……………陈伟, 叶宏杰, 周家宏, 魏峻 1 [2021005]
- 数据驱动的软件开发者智能协作技术……………张建, 孟祥鑫, 孙海龙, 王旭, 刘旭东 1 [2021006]
- 面向智能化软件开发的开源生态大数据……………张洋, 王涛, 尹刚, 余跃, 黄井泉 1 [2021007]

◎专题：大数据可视分析应用

- 大数据可视分析在海洋领域的应用
……………解翠, 李明悝, 陈萍, 李孝天, 宋键, 董军宇, 赵佳萌 2 [2021011]
- 足球比赛数据可视分析……………曹安琪, 陈泓宇, 谢潇, 巫英才 2 [2021012]
- 基于移动通信数据的城市可视分析研究
……………王桂娟, 周锐, 蔡梦杰, 汤勇, 李茸茸, 陈华容, 吴亚东 2 [2021013]
- 食品安全大数据可视化关联分析……………陈谊, 孙梦, 武彩霞, 孙小然 2 [2021014]
- 中国白酒文化可视化研究……………陈超, 吴亚东, 付朝帅, 董兴, 李攀, 褚琦凯, 王雪楠 2 [2021015]

◎专题：高性能计算虚拟数据空间

- 面向国家高性能计算环境的虚拟数据空间系统……………秦广军, 肖利民, 张广艳, 牛北方, 陈志广 2 [2021016]
- GVDS: 面向广域高性能计算环境的虚拟数据空间
……………肖利民, 宋尧, 秦广军, 周汉杰, 王超波, 韦冰, 魏巍, 霍志胜 2 [2021017]

- 分布式存储系统中的数据高效缓存方法……………杨青霖, 吴桂勇, 张广艳 2 [2021018]
- 国家高性能计算环境的虚拟数据空间运行支撑技术研究……………何小雨, 邓笋根, 栾海晶, 牛北方 2 [2021019]
- 面向非易失内存的MPI-I/O接口优化……………邓镇龙, 陈志广 2 [2021020]

◎专题: 基于大数据的知识图谱及其应用

- 基于主体掩码的实体关系抽取方法……………郑慎鹏, 陈晓军, 向阳, 沈汝超 3 [2021022]
- 实体摘要系统的解释性评测……………刘庆霞, 李俊宥, 程 龚 3 [2021023]
- 时态知识图谱补全的方法及其进展……………申宇铭, 杜剑峰 3 [2021024]
- 知识图谱推理: 现代的方法与应用……………王文广 3 [2021025]
- 知识图谱多跳问答推理研究进展、挑战与展望……………社会芳, 王昊奋, 史英慧, 王 萌 3 [2021026]
- 事件图谱的构建、推理与应用……………胡志磊, 靳小龙, 陈剑赞, 黄冠利 3 [2021027]
- 大规模知识图谱预训练模型及电商应用……………陈华钧, 张 文, 黄志文, 叶橄强, 文 博, 张 伟 3 [2021028]
- 基于金融知识图谱的会计欺诈风险识别方法……………陈 强, 代仕娅 3 [2021029]

◎专题: 新基建环境下的数据资产价值评估与定价

- 数据权属界定面临的问题困境与破解思路……………何 波 4 [2021034]
- 数据资产价值评估与定价: 研究综述和展望……………尹传儒, 金 涛, 张 鹏, 王建民, 陈嘉一 4 [2021035]
- 数据要素市场的组织形式和估值框架……………邹传伟 4 [2021036]
- 基于区块链的大数据交易模式研究与探索……………李 源, 高 宁, 孙 晶, 赵会群 4 [2021037]
- 基于区块链技术的数据资产管理机制……………赵 明, 董大治 4 [2021038]
- 基于博弈论和拍卖的数据定价综述……………张小伟, 江 东, 袁 野 4 [2021039]

◎专题: 国产环境下的大数据处理系统

- 面向多星多任务的大数据处理系统设计……………马福利, 石 涛, 陈 玲, 郑 岩, 熊森林 5 [2021045]
- 基于国产处理器架构的高能物理数据处理系统
……………程耀东, 程垚松, 毕玉江, 高 宇, 李海波, 汪 璐, 姚秋玲 5 [2021046]

从格点量子色动力学应用看国产超算环境的基础软件·····	宫 明, 蒋翔宇, 陈 莹, 刘朝峰	5 [2021047]
数值核反应堆大数据及其应用·····	汪 岸, 任 帅, 苗 雪, 董玲玉, 朱 迎, 陈丹丹, 胡长军	5 [2021048]
广域虚拟数据空间中边缘缓存系统的研究与实现·····	霍建同, 肖利民, 霍志胜, 徐耀文	5 [2021049]
面向大数据处理应用的广域存算协同调度系统 ·····	张晨浩, 肖利民, 秦广军, 宋 尧, 蒋世轩, 王继业	5 [2021050]

◎专题: 大数据支撑的智能应用

基于材料数值计算大数据的材料辐照机理发现 ·····	任 帅, 陈丹丹, 储根深, 白 鹤, 李慧昭, 何远杰, 胡长军	6 [2021056]
基于BERT阅读理解框架的司法要素抽取方法·····	黄 辉, 秦永彬, 陈艳平, 黄瑞章	6 [2021057]
结合案件要素序列的罪名预测方法·····	孙 倩, 秦永彬, 黄瑞章, 刘丽娟, 陈艳平	6 [2021058]
基于多输入模型及句法结构的中文评论情感分析方法·····	张宝华, 张华平, 厉铁帅, 商建云	6 [2021059]
大数据认知计算在内容安全管控中的应用·····	杜雪涛	6 [2021060]

◎专栏: 数据驱动和优化

基于样本的优化·····	张智杰, 孙晓明, 张家琳, 陈 卫	5 [2021051]
基于优化反馈的组合在线学习·····	孔 芳, 杨悦然, 陈 卫, 李 帅	5 [2021052]
强化学习在资源优化领域的应用·····	王金予, 魏欣然, 石文磊, 张 佳	5 [2021053]

◎专栏: 聚焦地方政府大数据

抢抓数字经济机遇 加快云南省数字化发展——云南省数字经济发展实践·····	洪正华	3 [2021033]
---------------------------------------	-----	-------------

◎研究

基于城市交通监控大数据的行程时间估计·····	李文明, 刘 芳, 吕 鹏, 于彦伟	1 [2021008]
联邦学习隐私保护研究进展·····	王健宗, 孔令炜, 黄章成, 陈霖捷, 刘 懿, 卢春曦, 肖 京	3 [2021030]
医学知识图谱构建关键技术及研究进展 ·····	谭 玲, 鄂海红, 匡泽民, 宋美娜, 刘 毓, 陈正宇, 谢晓璇, 李峻迪, 范家伟, 王晴川, 康霄阳	4 [2021040]

- 基于SVD++隐语义模型的信任网络推荐算法·····陈佩武, 束方兴 4 [2021041]
- 一种基于sort-last架构的非结构网格并行体可视化算法
·····范 良, 张晓蓉, 吴亚东, 陈 呈, 王 昉 4 [2021042]
- 基于深度学习的大脑性别差异分析·····温景熙, 于胡飞, 辛 江, 唐 艳 4 [2021043]
- 基于分布式缓存加速容器化深度学习的优化方法·····张 凯, 车 漾 5 [2021054]
- 基于法律裁判文书的法律判决预测·····张 虎, 潘邦泽, 谭红叶, 李 茹 5 [2021055]
- 基于特征选择的局部敏感哈希位选择算法·····周文桦, 刘华文, 李恩慧 6 [2021061]
- 面向非易失性内存的持久索引数据结构研究综述·····王永锋, 陈志广 6 [2021062]
- 大数据定价方法的国内外研究综述及对比分析·····刘 枏, 郝雪镜, 陈俞宏 6 [2021063]
- 水环境模型与大数据技术融合研究·····马金锋, 饶凯锋, 李若男, 张 京, 郑 华 6 [2021064]

◎应用

- 大数据技术在疫情精准防控中的应用·····李 刚, 郑 佳, 尹华山, 黄文超 1 [2021009]
- 智慧医院建设背景下的电子病历分析利用框架·····徐良辰, 郭崇慧 4 [2021044]
- 时间频率科学数据管理控制与应用·····章 宇, 袁海波, 王燕平, 董绍武, 张继海 6 [2021065]
- 基于ISM的大数据在建筑领域中的应用障碍分析·····纪颖波, 赵子豪, 姚福义 6 [2021066]
- 企业电力征信大数据价值挖掘与应用·····辛保江, 李德文, 王兰兰 6 [2021067]

◎论坛

- 数据跨境监管初探·····朱扬勇, 熊 贇 1 [2021010]

◎动态

- 广东省数字政府网络安全评估体系与实践
·····高尚省, 郭 勇, 高智伟, 钟世敏, 刘丕群, 刘 婷 2 [2021021]
- MadFS: 高性能超算缓存文件系统·····陈 康, 武永卫, 郑纬民 3 [2021031]
- AIPerf: 大规模人工智能算力基准测试程序·····翟季冬 3 [2021032]

BIG DATA RESEARCH

Content

2021 (Vol.7)

Big data based intelligent software development methodology and environment	<i>XIE Bing, PENG Xin, YIN Gang, LI Xuandong, WEI Jun, SUN Hailong</i>	1	[2021001]
Software knowledge graph construction and Q&A technology based on big data	<i>ZOU Yanzhen, WANG Min, XIE Bing, LIN Zeqi</i>	1	[2021002]
Context-based intelligent recommendation for code reuse	<i>PENG Xin, CHEN Chi, LIN Yun</i>	1	[2021003]
Big-data based intelligent bug triage techniques for open-source projects	<i>XI Shengqu, XU Feng, CHEN Xin, LI Xuandong</i>	1	[2021004]
An approach to automatically building Docker images by using domain knowledge	<i>CHEN Wei, YE Hongjie, ZHOU Jiahong, WEI Jun</i>	1	[2021005]
Data driven intelligent collaboration of software developers	<i>ZHANG Jian, MENG Xiangxin, SUN Hailong, WANG Xu, LIU Xudong</i>	1	[2021006]
Big data of open source ecosystem for intelligent software development	<i>ZHANG Yang, WANG Tao, YIN Gang, YU Yue, HUANG Jingquan</i>	1	[2021007]
Travel time estimation based on urban traffic surveillance data	<i>LI Wenming, LIU Fang, LYU Peng, YU Yanwei</i>	1	[2021008]
Application of big data technology in precise prevention and control of epidemic situation	<i>LI Gang, ZHENG Jia, YIN Huashan, HUANG Wenchao</i>	1	[2021009]
Primary exploration of transborder data flow supervision	<i>ZHU Yangyong, XIONG Yun</i>	1	[2021010]
Application of big data visual analysis in the marine field	<i>XIE Cui, LI Mingkui, CHEN Ping, LI Xiaotian, SONG Jian, DONG Junyu, ZHAO Jiameng</i>	2	[2021011]
Visual analysis for soccer match data	<i>CAO Anqi, CHEN Hongyu, XIE Xiao, WU Yingcai</i>	2	[2021012]
A survey on mobile communication data based urban visual analysis	<i>WANG Guijuan, ZHOU Rui, CAI Mengjie, TANG Yong, LI Rongrong, CHEN Huarong, WU Yadong</i>	2	[2021013]
Visual associations analysis of big data in food safety	<i>CHEN Yi, SUN Meng, WU Caixia, SUN Xiaoran</i>	2	[2021014]

- Research on visualization of Chinese Baijiu culture
 *CHEN Chao, WU Yadong, FU Chaoshuai, TONG Xing, LI Pan, CHU Qikai, WANG Xuenan* 2 [2021015]
- Virtual data space system for national high-performance computing environment
 *QIN Guangjun, XIAO Limin, ZHANG Guangyan, NIU Beifang, CHEN Zhiguang* 2 [2021016]
- GVDS: a global virtual data space for wide-area high-performance computing environments
 *XIAO Limin, SONG Yao, QIN Guangjun, ZHOU Hanjie,*
WANG Chaobo, WEI Bing, WEI Wei, HUO Zhisheng 2 [2021017]
- An approach to buffering data efficiently in distributed storage systems
 *YANG Qinglin, WU Guiyong, ZHANG Guangyan* 2 [2021018]
- Study of technique support on the operation of virtual data space in national high performance
 computing environment..... *HE Xiaoyu, DENG Sunge, LUAN Haijing, NIU Beifang* 2 [2021019]
- An optimization of MPI-IO interface for non-volatile memory
 *DENG Zhenlong, CHEN Zhiguang* 2 [2021020]
- Practice of digital government network security index evaluation system in Guangdong Province
 *GAO Shangsheng, GUO Yong, GAO Zhiwei, ZHONG Shimin, LIU Piquan, LIU Ting* 2 [2021021]
- An entity relation extraction method based on subject mask
 *ZHENG Shenpeng, CHEN Xiaojun, XIANG Yang, SHEN Ruchao* 3 [2021022]
- An interpretive evaluation of entity summarization system
 *LIU Qingxia, LI Junyou, CHENG Gong* 3 [2021023]
- Temporal knowledge graph completion: methods and progress
 *SHEN Yuming, DU Jianfeng* 3 [2021024]
- Knowledge graph reasoning: modern methods and applications
 *WANG Wenguang* 3 [2021025]
- Progress, challenges and research trends of reasoning in multi-hop knowledge graph
 based question answering..... *DU Huiyang, WANG Haofen, SHI Yinghui, WANG Meng* 3 [2021026]
- Construction, reasoning and applications of event graphs
 *HU Zhilei, JIN Xiaolong, CHEN Jianyun, HUANG Guanli* 3 [2021027]
- Large scale pre-trained knowledge graph model and e-commerce application
 *CHEN Huajun, ZHANG Wen, WONG Chi-Man, YE Ganqiang, WEN Bo, ZHANG Wei* 3 [2021028]
- Recognition method of accounting fraud risk based on financial knowledge graph
 *CHEN Qiang, DAI Shiya* 3 [2021029]
- Research advances on privacy protection of federated learning
WANG Jianzong, KONG Lingwei, HUANG Zhangcheng, CHEN Linjie, LIU Yi, LU Chunxi, XIAO Jing 3 [2021030]
- MadFS: a high performance burst buffer file system
 *CHEN Kang, Wu Yongwei, ZHENG Weimin* 3 [2021031]
- AIPerf: large-scale AI system benchmark..... *ZHAI Jidong* 3 [2021032]

Grasp the opportunity of digital economy to accelerate the digital development of Yunnan Province: digital economy development practice in Yunnan Province.....	<i>HONG Zhenghua</i>	3	[2021033]
Issues faced by the determination of data ownership and solutions	<i>HE Bo</i>	4	[2021034]
Assessment and pricing of data assets: research review and prospect	<i>YIN Chuanru, JIN Tao, ZHANG Peng, WANG Jianmin, CHEN Jiayi</i>	4	[2021035]
Organizational forms and valuation framework of data factor market	<i>ZOU Chuanwei</i>	4	[2021036]
Research and exploration of big data transaction model based on blockchain	<i>LI Yuan, GAO Ning, SUN Jing, ZHAO Huiqun</i>	4	[2021037]
Data asset management mechanism based on blockchain technology	<i>ZHAO Ming, DONG Dazhi</i>	4	[2021038]
A survey of game theory and auction-based data pricing	<i>ZHANG Xiaowei, JIANG Dong, YUAN Ye</i>	4	[2021039]
Key technologies and research progress of medical knowledge graph construction	<i>TAN Ling, E Haihong, KUANG Zemin, SONG Meina, LIU Yu, CHEN Zhengyu, XIE Xiaoxuan, LI Jundi, FAN Jiawei, WANG Qingchuan, KANG Xiaoyang</i>	4	[2021040]
A recommender algorithm based on SVD++ model under trust network	<i>CHEN Peiwu, SHU Fangxing</i>	4	[2021041]
A sort-last architecture based parallel volume visualization algorithm for unstructured grid	<i>FAN Liang, ZHANG Xiaorong, WU Yadong, CHEN Cheng, WANG Fang</i>	4	[2021042]
Analysis of gender differences in the brain based on deep learning	<i>WEN Jingxi, YU Hufei, XIN Jiang, TANG Yan</i>	4	[2021043]
Analysis and utilization framework of electronic medical records under the background of smart hospital construction	<i>XU Liangchen, GUO Chonghui</i>	4	[2021044]
Design of big data processing system supporting multi-satellites and multi-tasks	<i>MA Fuli, SHI Tao, CHEN Ling, ZHENG Yan, XIONG Senlin</i>	5	[2021045]
Data processing system for HEP based on domestic processor architecture	<i>CHENG Yaodong, CHENG Yaosong, BI Yujiang, GAO Yu, LI Haibo, WANG Lu, YAO Qiuling</i>	5	[2021046]
Software infrastructures for Chinese supercomputers from the perspective of lattice QCD applications	<i>GONG Ming, JIANG Xiangyu, CHEN Ying, LIU Zhaofeng</i>	5	[2021047]
Big data of numerical nuclear reactor and its application	<i>WANG An, REN Shuai, MIAO Xue, DONG Lingyu, ZHU Ying, CHEN Dandan, HU Changjun</i>	5	[2021048]
Research and implementation of edge cache system in global virtual data space across WAN	<i>HUO Jiantong, XIAO Limin, HUO Zhisheng, XU Yaowen</i>	5	[2021049]

- A wide-area collaborative scheduling system oriented to big data processing applications
 *ZHANG Chenhao, XIAO Limin, QIN Guangjun, SONG Yao, JIANG Shixuan, WANG Jiye* 5 [2021050]
- Optimization from samples
 *ZHANG Zhijie, SUN Xiaoming, ZHANG Jialin, CHEN Wei* 5 [2021051]
- Combinatorial online learning based on optimizing feedbacks
 *KONG Fang, YANG Yueran, CHEN Wei, LI Shuai* 5 [2021052]
- Applications of reinforcement learning in the field of resource optimization
 *WANG Jinyu, WEI Xinran, SHI Wenlei, ZHANG Jia* 5 [2021053]
- Method of accelerating deep learning with optimized distributed cache in containers
 *ZHANG Kai, CHE Yang* 5 [2021054]
- Legal judgment prediction based on legal judgment documents
 *ZHANG Hu, PAN Bangze, TAN Hongye, LI Ru* 5 [2021055]
- Discovery of irradiation mechanism based on big data of material simulation
 *REN Shuai, CHEN Dandan, CHU Genshen, BAI He, LI Huizhao, HE Yuanjie, HU Changjun* 6 [2021056]
- Legal element extraction method based on BERT reading comprehension framework
 *HUANG Hui, QIN Yongbin, CHEN Yanping, HUANG Ruizhang* 6 [2021057]
- Charge prediction method combined with case elements sequence
 *SUN Qian, QIN Yongbin, HUANG Ruizhang, LIU Lijuan, CHEN Yanping* 6 [2021058]
- Chinese comment sentiment analysis method based on multi-input model and syntactic structure
 *ZHANG Baohua, ZHANG Huaping, LI Tieshuai, SHANG Jianyun* 6 [2021059]
- Applications of big data cognitive computing in content security governance..... *DU Xuetao* 6 [2021060]
- Algorithm of locality sensitive hashing bit selection based on feature selection
 *ZHOU Wenhua, LIU Huawen, LI Enhui* 6 [2021061]
- A survey of persistent index data structures on non-volatile memory
 *WANG Yongfeng, CHEN Zhiguang* 6 [2021062]
- A review and comparative analysis of domestic and foreign research on big data pricing methods
 *LIU Nan, HAO Xuejing, CHEN Yuhong* 6 [2021063]
- Research on the integration of water environment model and big data technology
 *MA Jinfeng, RAO Kaifeng, LI Ruonan, ZHANG Jing, ZHENG Hua* 6 [2021064]
- Management control and application of the time-frequency scientific data
 *ZHANG Yu, YUAN Haibo, WANG Yanping, DONG Shaowu, ZHANG Jihai* 6 [2021065]
- Analysis of application barriers for big data in construction field based on ISM
 *JI Yingbo, ZHAO Zihao, YAO Fuyi* 6 [2021066]
- Value mining and application of big data in enterprise power credit investigation
 *XIN Baojiang, LI Dewen, WANG Lanlan* 6 [2021067]