

一种 ERP 数据管理、治理方法的研究

徐 敏,薛晓茹,王国梁,齐 慧

(国网安徽省电力有限公司信息通信分公司,安徽 合肥 230061)

摘 要:企业资源计划(ERP)是指以信息技术为基础,针对物资、人员、财政和信息等资源,采用电子系统化的管理模式,为企业建立数据信息管理平台。为了提高 ERP 数据的管理和治理能力,提出基于模糊信息聚类分析的 ERP 数据优化检索和管理模型。建立 ERP 数据的分块存储结构模型,采用模糊特征标注方法进行 ERP 数据融合;建立 ERP 数据的语义本体结构模型,采用自适应的特征提取方法进行 ERP 数据的信息检索;提取 ERP 数据的关联特征量,采用模糊信息聚类方法进行 ERP 数据的信息融合,根据信息分布进行大数据融合聚类,实现 ERP 数据的优化管理和治理。仿真结果表明,采用该方法进行 ERP 数据治理的融合性较好,提高了数据管理的优化检索能力。

关键词:企业资源计划数据;模糊信息聚类;信息检索;数据融合

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2020)10-0021-05

doi:10.3969/j.issn.1673-629X.2020.10.004

Research on an ERP Data Management and Governance Method

XU Min, XUE Xiao-ru, WANG Guo-liang, QI Hui

(Information and Communication Branch, State Grid Anhui Electric Power Co., Ltd., Hefei 230061, China)

Abstract: Enterprise resource planning (ERP) refers to the establishment of a data and information management platform for enterprises based on information technology with the electronic systematic management mode for materials, personnel, finance and information. In order to improve the management and management ability of ERP data, an ERP data optimization retrieval and management model based on fuzzy information clustering analysis is put forward. The block storage structure model of ERP data is established, and the fuzzy feature tagging method is used for ERP data fusion. The semantic ontology structure model of ERP data is established, and the adaptive feature extraction method is used for ERP data information retrieval. The associated feature quantity of ERP data is extracted, and the fuzzy information clustering method is used for ERP data information fusion, and the big data fusion clustering is carried out according to the information distribution to realize the optimal management and governance of ERP data. The simulation shows that the proposed method has better integration of ERP data management, which improves the optimal retrieval ability of data management.

Key words: ERP data; fuzzy information clustering; information retrieval; data fusion

0 引言

随着现代企业资源计划的不断量化^[1],需要采用更精细化管理模型进行 ERP 数据分析。结合分区融合技术,进行的资源信息调度,采用大数据分析方法进行 ERP 数据分析,提高企业资源的智能化管理能力,相关的 ERP 数据治理和分析方法研究受到人们的极大关注^[2]。ERP 数据的管理和治理方法是建立在对企业信息资源的大数据融合基础上,通过建立 ERP 数据信息融合和自适应评估模型,采用模糊信息聚类方法,实现 ERP 数据的综合治理,但不文献中提到的方法还是传统方法,进行 ERP 数据综合治理模糊度都比较大。

文献[3]利用 MapReduce 模型对数据进行了融合,使更为相似的数据得到较好的融合,但是在数据调度时,未考虑数据的关联规则特征量,导致准确率降低,并未对数据进行去模糊处理,且计算过程较为复杂,增加了数据处理时间;文献[4]利用大数据技术,实现不同类型海量数据的快速调度,加快了数据处理进程,但是未考虑各种数据之间的模糊界限,使数据处理的准确性降低。

对此,该文提出基于模糊信息聚类分析的 ERP 数据管理和治理方法。首先对 ERP 数据的存储结构进行了分析,建立 ERP 数据的分块异构存储结构模型,采用关联维分布式调度的方法,提取 ERP 数据的关联

收稿日期:2019-11-26

修回日期:2020-03-30

基金项目:国网安徽省基金项目(22345878032)

作者简介:徐 敏(1988-),男,硕士研究生,高级工程师,研究方向为大数据、数据中台、数据分析。

规则特征量,在云计算中心进行数据的高效索引。根据对 ERP 数据相似度特征的分析,进行数据融合。在信息检索时,建立 ERP 数据的语义本体结构模型,采用自适应的特征提取方法完成信息检索。与传统方法不同的是,该方法采用模糊 C 均值聚类方法,进行 ERP 数据的二元分组测试,建立高分辨信息融合模型,得到 ERP 数据信息检索的迭代模型,实现 ERP 数据的优化管理和治理。最后进行仿真测试分析,以验证该方法的有效性。

1 ERP 数据存储结构及特征分析

1.1 ERP 数据的存储结构

为了实现 ERP 数据管理、治理优化,首先需要了解 ERP 数据存储结构,构建 ERP 数据的分块存储结构模型,采用模糊特征标注方法进行资源信息调度^[5]。如果 ERP 数据分布满足凸组合模型 $X \sim S_\alpha(1, \beta, 0)$, $1 < \alpha < 2$, 即 ERP 数据管理过程中需要进行特征匹配,在 ERP 数据信息库中采用模糊关联特征调度方法进行信息匹配,构造 ERP 数据的线性组合模型为:

$$\begin{aligned} \min(f) &= \sum_{i=1}^m \sum_{j=1}^n C_{ij} X_{ij} \\ \text{s. t. } &\begin{cases} \sum_{j=1}^n X_{ij} = a_i, i = 1, 2, \dots, m \\ \sum_{i=1}^m X_{ij} = b_j, j = 1, 2, \dots, n \\ X_{ij} \geq 0, i = 1, 2, \dots, m, j = 1, 2, \dots, n \end{cases} \end{aligned} \quad (1)$$

结合分块信息融合方法,进行 ERP 数据的资源信息调度,采用模糊信息索引方法建立 ERP 数据的特征参数融合向量集。由于各个数据资源之间^[6],除了确切的界限,有些则是模糊的,模糊聚类分析就是将各种数据之间的模糊界限,按要求对这些数据进行分类分析^[7]。聚类分析是数据统计中一种多元化分析,根据数学方法确定数据的远近关系,从而客观地进行分类,按照分类结果进行系统计算。采用模糊信息聚类方法,得到 ERP 数据量化模型,记为 L_1, \dots, L_n 和 $P_1^{\min}, \dots, P_n^{\min}$, 根据 ERP 数据的存储节点分布,进行 ERP 数据的自适应融合。

在构建 ERP 数据的异构存储结构模型时,采用关联维分布式调度的方法^[8],提取 ERP 数据的关联规则特征量,在云计算中心进行数据的高效索引,采用日志来记录 ERP 数据的存储节点^[9],构建 ERP 数据的异构存储模型,得到梯度特征分布向量为 $\nabla^2 F(x)$ 。在云计算环境中进行 ERP 数据的分布式挖掘,采用存储节点的异构存储模型进行 ERP 数据信息融合,建立分布式 B 树进行 ERP 数据的分类处理,得到 ERP 数据的异构存储模型,如图 1 所示。

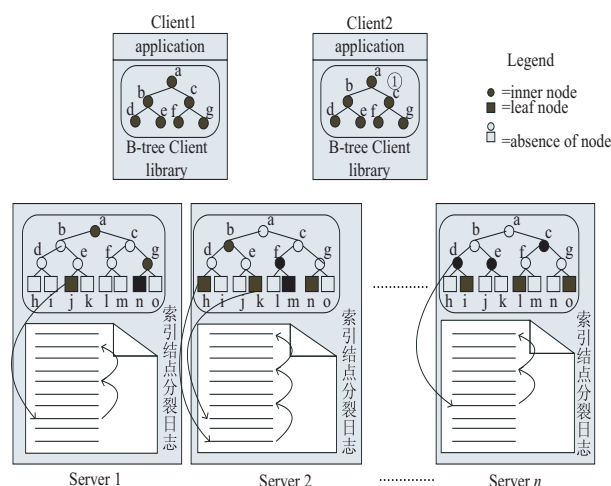


图 1 ERP 数据的异构存储模型

根据图 1 所示的存储结构模型,在服务器中进行信息调度,将所获得的信息数据储存在本地服务器中,采用回归树分析的方法进行 ERP 数据的综合管理,综合管理的交叉树结构模型如图 2 所示。

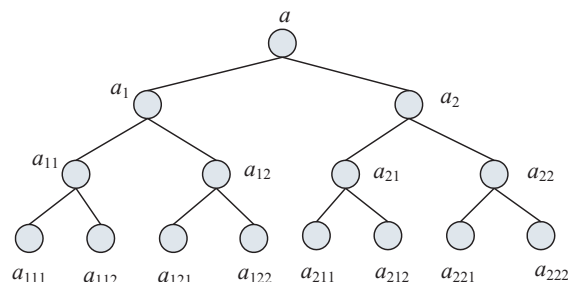


图 2 ERP 数据综合管理的交叉树模型

分析上述存储结构所获得的数据,采用模糊特征标注方法进行 ERP 数据信息检索,根据数据特征进行数据融合,建立 ERP 数据的语义本体结构模型,提高数据的综合管理和治理能力^[10]。

1.2 ERP 数据特征融合

建立 ERP 数据的语义本体结构模型,采用自适应的特征提取方法进行 ERP 数据的信息检索,建立描述 ERP 数据特征类别的约束特征函数为:

$$\begin{aligned} [\nabla^2 F(x)]_{kj} &= \frac{\partial^2 F(x)}{\partial x_k \partial x_j} = 2 \sum_{i=1}^N \left[\frac{\partial v_i(x)}{\partial x_k} \cdot \frac{\partial v_i(x)}{\partial x_j} + \right. \\ &\quad \left. v_i(x) \frac{\partial^2 v_i(x)}{\partial x_k \partial x_j} \right] = \\ &\quad 2J^T(x)J(x) + 2S(x) \end{aligned} \quad (2)$$

判断缓冲节点与服务器中存储相关性,进行 ERP 数据的分块区域调度^[11]。设 ERP 数据的关联调度的特征分布映射为 $f(k)$,在每次访问根节点的能耗开销计算公式为:

$$U_{i,j}(t) = \exp[-b[z_i(t) - z_j(t)]^2] \quad (3)$$

假设 ERP 数据有 6.4 万个 <key, pointer> 对,采用相关性的统计分析方法,得到 ERP 数据共享度为:

$$S_{i,j}(t) = \frac{p_{i,j}(t) - s(t)}{p_{i,j}(t)} U_{i,j}(t) \quad (4)$$

其中, $p_{i,j}(t)$ 为 ERP 数据共享的模糊相关性特征分布集, $s(t)$ 为 ERP 数据共享的模糊决策增量值。

初始化 ERP 数据的信息聚类中心, 存储 B 树节点的服务器中找到一个模糊聚类中心点, 得到 ERP 数据的特征分辨函数为 $F(x_i, A_j(L))$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, k$ 。采用空间规划结构重组的方法, 建立 ERP 数据非线性特征重组模型, 在每次访问 ERP 数据时, 统计数据的访问量, 结合定量递归分析方法, 得到数据的相似度特征为:

$$\gamma_i = \frac{\frac{1}{w} \sum_{l=0}^{w-1} [x_i(k-l) - \mu_i]^3}{\left(\frac{1}{w} \sum_{l=0}^{w-1} [x_i(k-l) - \mu_i]^2 \right)^{\frac{3}{2}}} \quad (5)$$

根据上述完成 ERP 数据相似度特征的分析, 利用上述相似度进行数据融合, 得到数据特征融合函数为:

$$R_h = \gamma_i \frac{\frac{1}{w} \delta_N \sum_{l=0}^{w-1} [\nabla^2 F(x)]_{kj}}{S_{i,j}(t)^2} \quad (6)$$

其中, δ_N 表示 ERP 数据融合的谱信息。

利用上式完成数据特征融合, 通过该步骤为后续数据的管理奠定可行依据。

2 ERP 数据的管理和治理方法优化

2.1 ERP 数据的信息检索

根据上述 ERP 数据的资源信息调度, 采用模糊信息索引的方法, 在建立 ERP 数据的分块存储结构模型并进行特征提取的基础上, 进行 ERP 数据的管理和治理优化。该文提出基于模糊信息聚类分析的 ERP 数据优化检索和管理模型, 用 4 元组 (E_i, E_j, d, t) 来表示 ERP 数据共享调度的主特征量, 建立其关联函数为:

$$\text{CON}(t, c_i) = \log \frac{P(d|t, c_i)}{P(d|t)} = \log \frac{P(d|t, c_i)}{\sum_{i=m} P(d|t, c_i)} \quad (7)$$

其中, $P(d|t, c_i)$ 为 c_i 类 ERP 数据的特征分辨率, 在所有客户端进行 ERP 数据融合和信息检索^[12-13], 得到 ERP 数据治理的特征集 δ_k 变为 $\delta_{ik}(t)$:

$$\delta_{ik}(t) = G(V = k | U_i, \Theta(t)) \quad (8)$$

采用延迟更新的方法进行 ERP 数据的量化编码, 得到输出编码结果为:

$$x_m(t) = \sum_{i=1}^I s_i(t) e^{j\varphi_m} + n_m(t), \quad -p+1 \leq m \leq p \quad (9)$$

在数据的目标域空间, 对采集的 ERP 数据进行特征筛选, 得到 ERP 数据的信息熵满足:

$$\text{esup}^t(D) > \theta \quad (10)$$

采用模糊 C 均值聚类方法, 进行 ERP 数据的二元分组测试, 建立高分辨信息融合模型, 得到 ERP 数据信息检索的迭代模型为:

$$x_o^i = x_s^i + K d_i^{\max}(x_L^i - x_s^i) \quad (11)$$

其中, $K = 1 / \|x_L^i - x_s^i\|$ 。根据信息聚类结果, 进行 ERP 数据的动态规划, 采用自适应寻优方法, 得到 ERP 数据融合的动态规划模型为:

$$P_{i,j}^t = \begin{cases} P_{i-1,j-1}^t \times p_i + P_{i-1,j}^t \times (1 - p_i), & v_i = t \\ P_{i-1,j}^t, & v_i \neq t \end{cases} \quad (12)$$

其中, p_i 为 ERP 数据的空间谱密度函数, 取决区域 K 中的概率, $P_{i-1,j}^t$ 为数据的语义本体结构特征量, 结合空间信息资源调度的方法, 进行 ERP 数据的检索优化设计。

2.2 数据模糊聚类及优化治理输出

数据模糊聚类是指根据一组数据, 把其含义划分为不同的类别, 而各个类别之间并没有直接联系, 根据实际需把一些联系比较强的数据划分为一类, 是进行数据模糊采样的前提。在语义本体模型中执行 ERP 数据的样本统计, 得到 ERP 数据信息采样的 n 个统计变量, 遍历多个节点进行 ERP 数据综合治理, 得到综合治理的特征分解函数为:

$$\frac{dx}{dt} = ax - bx^2 + s(t) + \Gamma(t) \quad (13)$$

采用分布式的决策树模型进行 ERP 数据的统计分析, 建立 ERP 数据的统计检测模型为:

$$f(x) = \text{sgn} \left\{ z \sum_{i=1}^{l_1} \alpha_i^+ y_i K(x_i, x) + \sum_{i=1}^{l_2} \alpha_i^- y_i K(x_i, x) + b \right\} \quad (14)$$

其中, α_i^+ 表示 ERP 数据融合的模糊边缘特征向量, α_i^- 表示 ERP 数据融合的低频分量, l_1 和 l_2 分别表示数据的分布相似度和关联度函数。结合定量递归分析方法^[14], 得到 ERP 数据融合的梯度模型为:

$$\text{AVG}_X = \frac{1}{m \times n} \sum_{x=1}^n \sum_{y=1}^m |G_X(x, y)| \quad (15)$$

其中, m, n 分别是 ERP 数据分布的维数和采样节点。

设 D 为不确定数据库, T_i 为数据的特征信息采样延迟, 由此得到 ERP 数据的特征重组输出为:

$$\text{sgn}(z_R^2(k) - R_{\text{MDMMA}_R}) = \text{sgn}(z_R^2(k) - \hat{e}_R^2(k)) \quad (16)$$

$$\text{sgn}(z_I^2(k) - R_{\text{MDMMA}_I}) = \text{sgn}(z_I^2(k) - \hat{e}_I^2(k)) \quad (17)$$

其中, $\hat{e}_R^2(k)$ 表示 ERP 数据融合的梯度分布序列, $z_R^2(k)$ 为原始训练集中的信噪比, $z_I^2(k)$ 为 ERP 数据的模糊隶属度函数, $\hat{e}_I^2(k)$ 为融合误差。根据上述模型

设计,采用模糊信息聚类方法进行 ERP 数据的信息融合,根据企业资源计划信息分布进行大数据融合聚类^[15],实现 ERP 数据的优化管理和治理。

3 仿真测试分析

3.1 实验环境及数据时域分布

为了验证该方法在实现 ERP 数据管理和治理优化中的应用性能,结合上述过程中的信息融合方法,进行 ERP 数据的综合管理。采用 Matlab 进行实验测试分析,设定对 ERP 数据采样的节点数为 220, Sink 节点为 12,数据采样的单组长度为 1 200,分块融合聚类的分辨率为 0.83,相似度系数为 1.25。根据上述仿真参数设定,进行 ERP 检索,得到的数据时域分布如图 3 所示。

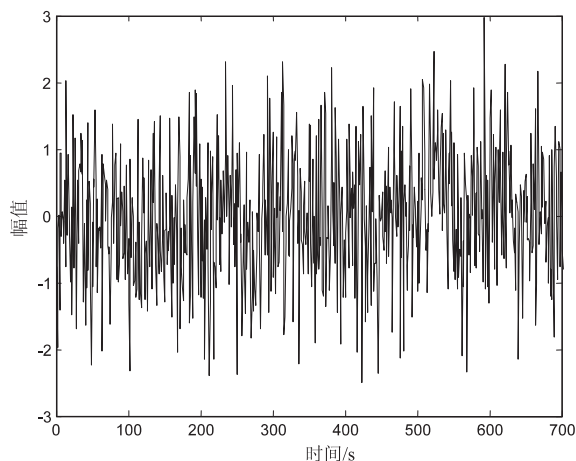


图 3 数据时域分布

以图 3 的数据时域分布为研究对象,结合企业大数据融合聚类以及数据的优化治理输出,进行 ERP 数据管理和综合治理,随机采用企业金融、人力资源和综合资源三种事物集。

3.2 实验指标

为全面验证该方法的鲁棒性,分别检测其吞吐量及数据召回率。其中,数据吞吐量计算公式为:

$$F = \text{AVG}_x \frac{V_u R}{T_i} \quad (18)$$

其中, V_u 表示并发数, T_i 表示响应时长, R 表示平均每个用户发出的请求数量。若吞吐量越高,则表示该方法数据传输性能越好。

数据召回率计算公式为:

$$Z(h) = \frac{\text{AVG}_x}{N_z} \times 100\% \quad (19)$$

其中, N_z 表示总数据数量。若数据召回率越高,则表示方法具有较高的数据查全性能。

3.3 实验分析

为验证该方法的有效性及其可行性,本次实验将随机采用企业金融、人力资源和综合资源三种事物集分

别测试其输出的吞吐量,得到的测试结果如图 4 ~ 图 6 所示。

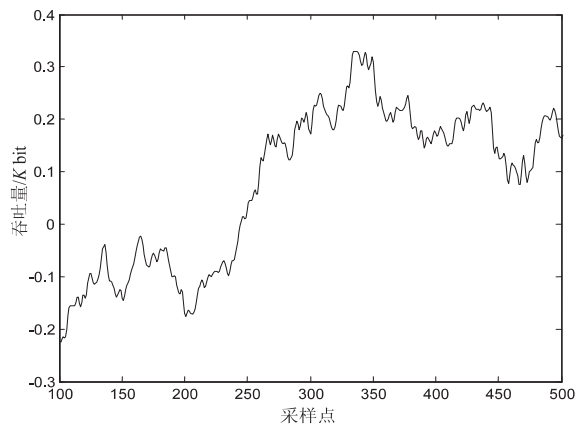


图 4 企业金融事物集下的吞吐量

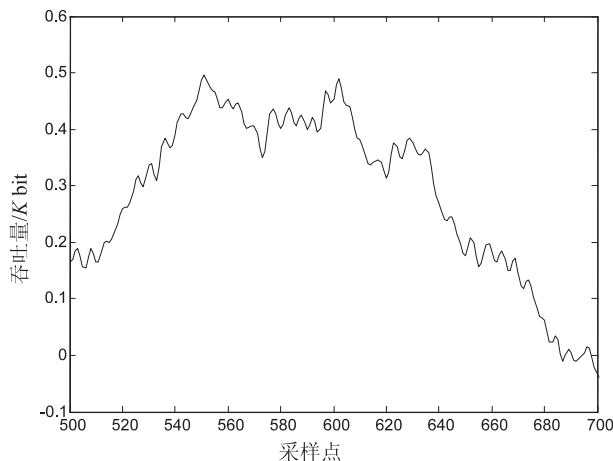


图 5 人力资源事物集下的吞吐量

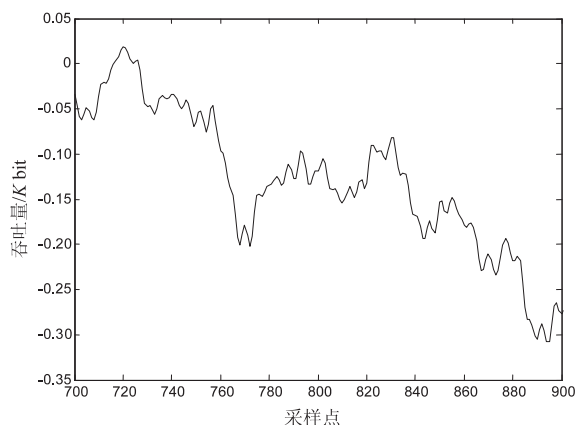


图 6 综合资源事物集下吞吐量

该方法通过构建 ERP 数据的分块存储结构模型,在服务器中进行信息调度,经数据信息检索后对 ERP 进行数据融合,分析数据时域分布情况,测试在不同事物集下的输出吞吐量。分析上述仿真结果可知,采用该方法能在不同的 ERP 数据管理模式提高资源调度吞吐量,根据资源调度吞吐量的结果,测试数据的召回率,得到的对比结果见表 1。

由表 1 可知,随着迭代次数的增加,三种方法的召回率都随之增高,但文中方法的召回率始终高于文献

[3]和文献[4],显示出了良好的召回性能。文中方法根据数据特征进行数据融合,通过建立描述 ERP 数据特征类别的约束特征函数来构建 ERP 数据的语义本体结构模型,从而完成特征数据融合,提高了资源调度吞吐量,对数据进行了有效管理,为后续处理奠定了基础。在进行 ERP 数据的信息检索时,采用模糊 C 均值聚类方法,建立高分辨信息融合模型,通过 ERP 数据的二元分组测试,得到 ERP 数据信息检索的迭代模型,加快了数据处理进程。在输出过程中,根据信息聚类结果,进行 ERP 数据的动态规划,采用自适应寻优方法,提高了数据召回率,强化了数据管理的优化检索,实现 ERP 数据的优化管理和治理。

表 1 数据召回率测试

迭代次数	文中方法	文献[3]	文献[4]
100	0.934	0.893	0.877
200	0.984	0.901	0.883
300	1	0.923	0.912

4 结束语

ERP 融合了企业信息技术和实践经验,随着企业的进步,不断对其数据资源进行优化,加强了员工之间协同合作,提高了企业的发展水平,为增强企业竞争力做出了巨大贡献。研究 ERP 数据管理、治理方法对企业管理有着重要的作用,通过对企业的 ERP 数据管理,提高企业的资源信息管理和调度能力。该文提出基于模糊信息聚类分析的 ERP 数据优化检索和管理模型。通过建立描述 ERP 数据特征类别的约束特征函数来构建 ERP 数据的语义本体结构模型,并提取 ERP 数据的关联规则特征量,实现在云计算中心进行数据的高效索引,采用日志来记录 ERP 数据的存储节点,在所有客户端进行 ERP 数据融合和信息检索。采用模糊 C 均值聚类方法,建立高分辨信息融合模型,通过 ERP 数据的二元分组测试,得到 ERP 数据信息检索的迭代模型,加快了数据处理进程。在输出过程中,根据信息聚类结果,采用自适应寻优方法,实现 ERP 数据的优化管理和治理。经实验证明,该方法能在不同的 ERP 数据管理模式,提高资源调度吞吐量,数据治理融合及召回率较高。由此可知,用该方法进行资源数据管理的检索能力较好,具有较好的 ERP 数据

的管理和治理能力。

参考文献:

- [1] 王慧健,刘 峥,李 云,等. 基于神经网络语言模型的时间序列趋势预测方法[J]. 计算机工程,2019,45(7):13-19.
- [2] 唐新宇,张新政,赵月爱. 云计算中基于群体智能算法的大数据聚类挖掘[J]. 重庆理工大学学报:自然科学版,2019,33(4):128-133.
- [3] 宋人杰,余 通,陈宇红,等. 基于 MapReduce 模型的大数据相似重复记录检测算法[J]. 上海交通大学学报,2018,52(2):214-221.
- [4] 黄思杰. 基于大数据的智能 ERP 系统设计[J]. 现代电子技术,2018,41(24):94-97.
- [5] 刘久彪. 空间数据库反向最近邻聚类方法[J]. 吉林大学学报:理学版,2019,57(2):387-392.
- [6] 于繁华,姚亦飞,逯启荣. 教育大数据下基于离群检测的教学预警模型[J]. 吉林大学学报:信息科学版,2018,36(4):459-464.
- [7] LI Tan, ZHAO Chengyong. Recovering the modular multi-level converter from a cleared or isolated fault[J]. IET Generation, Transmission & Distribution, 2015, 9(6):550-559.
- [8] GUO Chunyi, LIU Wei, ZHAO Chengyong, et al. Small-signal dynamics and control parameters optimization of hybrid multi-infeed HVDC system[J]. International Journal of Electrical Power & Energy Systems, 2018, 98(12):409-418.
- [9] DOU Q, CHEN H, YU L Q, et al. Multi-level contextual 3D CNNs for false positive reduction in pulmonary nodule detection[J]. IEEE Transactions on Biomedical Engineering, 2017, 64(7):1558-1567.
- [10] 汪 晖,尹世玉,王 颖,等. 保护性约束信息化管理系统的构建及应用[J]. 中华护理杂志,2019,54(6):850-854.
- [11] 蒋 义,熊先青,刘海波,等. 基于订单流导向的木塑企业 ERP 管控技术研究[J]. 林产工业,2018,45(5):45-50.
- [12] 黄小东,杨志伟. 物流信息管理系统在数字化铸造车间的应用[J]. 铸造,2017,66(7):723-726.
- [13] 李玉灵,单丹微,杨以雄,等. 基于需求预测的服装 O2O 货品管理探析[J]. 毛纺科技,2018,46(7):1-7.
- [14] 陈思雨,凌 宁,任蒙蒙,等. 基于 JMP/SAP 的六西格玛管理系统设计及其在 SMED 中的应用[J]. 科技管理研究,2017,37(18):186-193.
- [15] 袁维华,熊自明,褚靖豫,等. 基于 BIM 的南京地铁运营资产管理信息系统[J]. 现代隧道技术,2019,56(2):30-39.