

文章编号:1671-5896(2026)04-0991-07

基于爬虫技术的电商网页关联数据自适应挖掘算法

王 琪

(黑龙江财经学院 财经信息工程学院, 哈尔滨 150025)

摘要: 为有效挖掘和利用网页数据, 以爬虫技术为支持, 提出一种关联数据自适应挖掘算法。利用爬虫技术爬取电商网页, 突破数据获取限制, 将所需数据以结构化格式存储至数据库, 为后续处理提供有序数据基础。基于存储在数据库中的稀疏数据, 依据最小支持度并运用 Apriori 算法, 并通过剪枝策略, 在减少数据处理量的同时, 能在稀疏数据中精准找出频繁项集, 成功解决传统方法因数据稀疏难以获取频繁项集的问题。以树形结构为框架, 将 Apriori 算法得到的频繁项集作为结点, 通过自适应结合旧结点与新结点构建挖掘树, 直至无结点可结合, 使其生成的挖掘树全面呈现数据关联关系, 进而得到充分且精准的电商网页关联数据挖掘结果。经在大型电商网站的网页上验证后表明, 所提算法能确保网页爬取的完整性, 准确获得以购买行为为目标的频繁项集, 精准挖掘电商网页中的关联数据, 在为电商平台运营提供有益参考的同时, 给予商家和消费者更好的决策支持。

关键词: 电商网页数据; 爬虫技术; 网页数据爬取; 频繁项集; 树形结构; 自适应挖掘

中图分类号: TP391

文献标志码: A

Adaptive Mining Algorithm for Association Data of E-commerce Web Page Based on Crawler Technology

WANG Qi

(School of Finance and Information Engineering, Heilongjiang University of Finance and Economics, Harbin 150025, China)

Abstract: To effectively mine and utilize web data, a correlation data adaptive mining algorithm is proposed with the support of web crawling technology. Web crawling technology is used to crawl e-commerce web pages, breaking through data acquisition limitations, storing the required data in a structured format in a database, and providing an orderly data foundation for subsequent processing. Based on sparse data stored in the database, the Apriori algorithm is applied according to the minimum support degree. Through pruning strategy, this algorithm can accurately find frequent itemsets in sparse data while reducing data processing, successfully solving the problem of traditional methods being difficult to obtain frequent itemsets due to data sparsity. Using a tree structure as a framework, the frequent itemsets obtained by the Apriori algorithm are used as nodes. By adaptively combining old and new nodes, a mining tree is constructed until there are no nodes to combine. The generated mining tree comprehensively presents data association relationships, thereby obtaining sufficient and accurate e-commerce webpage association data mining results. After verification on the web pages of large e-commerce websites, it is found that the proposed algorithm can ensure the integrity of web crawling, accurately obtain frequent itemsets targeting purchasing behavior, and precisely mine associated data in e-commerce web pages. While providing useful references for e-commerce platform operation, it also provides better decision

收稿日期: 2025-02-07

基金项目: 黑龙江省教育科学规划重点课题基金资助项目(GJB1424318)

作者简介: 王琪(1977—), 女, 哈尔滨人, 黑龙江财经学院副教授, 主要从事计算机科学与技术研究, (Tel) 86-18646355970 (E-mail) wqgj321@163.com。

support for merchants and consumers.

Key words: E-commerce webpage data; crawling technology; Web data crawling; frequent itemsets; tree structure; adaptive mining

0 引 言

在电商平台的日常运营中,积累了难以估量的用户行为数据、详尽的商品信息以及海量的交易记录^[1-2]。由于电商网页数据具有多源异构、动态变化以及高度稀疏等特性,在利用传统的数据挖掘方法处理电商网页关联数据时,其高度稀疏性使传统算法难以有效捕捉数据中的潜在模式,极易遗漏有价值的关联信息,造成挖掘结果的不完整与不准确^[3]。因此,对电商网页关联数据自适应挖掘算法进行研究具有重要意义。

在数据挖掘的研究中,由于不同领域内差异显著的数据特性对数据挖掘的泛化能力提出了挑战,因此,人们致力于研究高效的数据挖掘算法。韩永印等^[4]针对社交网络隐式用户行为数据,基于决策树设计数据挖掘方法,根据行为数据的测试分支子集属性权重完成挖掘。但在面对电商网页数据的稀疏性时,该方法可能无法有效处理,难以精准获取频繁项集,导致挖掘不充分。陈万志等^[5]针对用户行为轨迹数据,通过改进 Prefix Span 算法,得到序列模式数据挖掘算法,经迭代去除冗余项集完成挖掘。但改进的 Prefix Span 算法并没有针对电商网页数据稀疏性设计有效的应对机制,难以突破数据稀疏性的限制,无法充分挖掘其中的关联信息,进而导致挖掘结果不充分。夏小雅等^[6]针对多源开源协作行为数据,通过整合数据存储模块、标签数据模块等,设计出 Open Digger 一站式数据挖掘系统。但该方法在处理稀疏数据时,难以精准定位和获取频繁项集,造成电商网页关联数据挖掘不充分。曹俊彬等^[7]针对军用航空器事故数据,通过融合地理信息系统与支持向量机、神经网络等多种机器学习算法,构建出用于挖掘事故关键质量特性的模型。但该模型现有的数据处理方式无法满足电商网页数据挖掘的特殊需求,使在面对电商网页数据稀疏性时,难以通过合适的数据处理流程获取频繁项集,最终影响电商网页关联数据挖掘效果。

爬虫技术^[8]通过自动化地抓取、解析网页内容,能以结构化数据^[9]为基准完成数据转化,给予后续挖掘更有效的操作对象。为此,以爬虫技术为主要支持,笔者提出一种用于电商网页的关联数据自适应挖掘算法。利用爬虫技术爬取电商网页中的所需数据,有助于高效、批量且自动化地采集海量精准的电商信息;基于 Apriori 算法获取频繁项集,能精准定位高频共现元素,助力高效挖掘隐藏的关联规则;以树形结构为框架进行自适应挖掘,有利于动态适配电商数据的持续更新与变化,提升挖掘精准度与效率。

1 基于爬虫技术的电商网页数据爬取

电商网页数据具有高度稀疏等特性,使传统的数据挖掘方法在处理电商网页关联数据时面临诸多挑战。爬虫技术能突破数据获取的障碍,按照预定规则自动、高效地从电商网页中抓取各类数据,将分散在不同页面、格式下的信息进行整合,为后续的数据挖掘提供丰富且全面的数据基础。通过爬取大量电商网页数据,能在一定程度上缓解数据稀疏性问题,为准确获取频繁项集、充分挖掘关联数据创造有利条件。爬虫技术在电商网页上的具体爬取流程如图 1 所示。

图 1 中,统一资源定位符(URL: Uniform Resource Locator)队列的模式为先进先出^[10],页面读取时,队列长度 L 的变化如下:

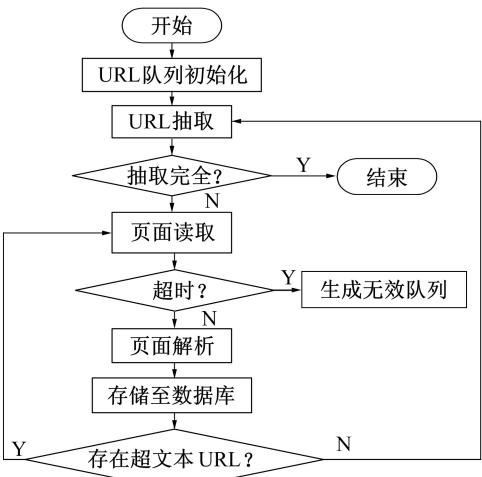


图 1 电商网页数据爬取流程
Fig. 1 Flow chart of E-commerce web page data crawling

$$L_{t+1} = L_t + l_t - 1, \quad (1)$$

其中 L_t 、 L_{t+1} 分别为 t 时和 $t+1$ 时的队列长度, l_t 为 t 时加入队列的链接数量, “-1” 为从队列头部选取下一个链接供爬虫抓取, 直至 $L=0$ 。

对超时读取次数 n_o 存在如下限制:

$$n_o \leq n_z - n_s, \quad (2)$$

其中 n_s 为成功读取次数, n_z 为总读取次数。

爬虫读取页面时, 若对象是多媒体数据^[11], 则存储至数据库; 若是无结构文本^[12] 或半结构化数据^[13], 则先使用工具解析, 然后按照键值对的形式组织提取信息, 转化为结构化数据后存储。

为评估爬虫技术的有效性和数据完整性, 假设爬虫总共抓取了 m 类数据, 每类数据的数量分别为 $D_1, D_2, \dots, D_m$, 则总的爬取数据量 D 为

$$D = \sum_{i=1}^m L_{t+1} D_i. \quad (3)$$

通过上述步骤, 爬虫技术不仅能有效突破电商网页数据获取的限制, 还能将所需数据以结构化格式存储至数据库, 为后续的数据处理提供坚实基础。

2 基于频繁项集的电商网页关联数据自适应挖掘

以爬虫爬取的数据为依据, 根据从中获取的频繁项集, 构建关联数据挖掘树, 自适应挖掘出电商网页中的关联数据。

2.1 电商网页爬取数据的频繁项集获取

在电商网页数据中, 由于数据稀疏性的问题, 传统的数据挖掘方法难以准确找出频繁项集, 从而影响了数据分析效果。为此, 借助 Apriori 算法, 并通过剪枝策略, 在减少数据处理量的同时, 能在稀疏数据中精准识别频繁出现的项集。基于存储在数据库中的稀疏数据, 运用 Apriori 算法依据最小支持度找出频繁项集, 从而显著提高数据分析的准确性和效率, 为电商网站的商品推荐、市场趋势预测等提供有力支持。

基于 Apriori 算法^[14] 获取爬虫爬取数据中的频繁项集, 具体流程如下:

1) 扫描数据库, 基于总的爬取数据量 D , 找出爬取数据中的所有 1-项频繁项集。

2) 从 2-项开始, 通过循环得到 k -项频繁项集。

3) 剪枝^[15] ($k-1$)-项频繁项集。若 k -项频繁项集前 λ 个项目构成的子集是 d_λ , 其在 map 集合中的出现次数是 n_{d_λ} , 则 ($k-1$)-项频繁项集的剪枝条件如下:

$$n_{d_\lambda} < k - \lambda. \quad (4)$$

4) 利用 Apriori 算法, 从剪枝后的项集中得到 k -项候选项集, 求解各候选项频度^[16]。从数据库映射^[17] k -项候选项集, 若该项集是三维集, 映射所得候选项 i 的对应数组为 D_1^i, D_2^i, D_3^i , 候选项集中有 n_c 个重复数据, 则利用

$$R_y^i = n_c (D_1^i \wedge D_2^i \wedge D_3^i) \quad (5)$$

进行合取运算, 得到 3 个数组的“与”结果 R_y^i 。结合 R_y^i 与 t 时候选项 i 的频度 P_t^i , 由

$$P_{t+1}^i = R_y^i + P_t^i \quad (6)$$

推算出 $t+1$ 时该候选项的频度 P_{t+1}^i 。

5) 基于最小支持度^[18], 判定候选项的属性。若含有候选项 i 的数据量为 n_i , 数据库中爬取的总数据量为 N , 则候选项 i 的支持度 S_i 为

$$S_i = n_i / |N|. \quad (7)$$

由此得到最小支持度 $\min\{S_i\}$, 则候选项 i 的属性判定条件如下:

$$P_{t+1}^i \geq \min\{S_i\}. \quad (8)$$

当候选项 i 的频度 P_{i+1}^i 与最小支持度 $\min\{S_i\}$ 使式(7)成立时,其属于频繁项集。

2.2 电商网页关联数据自适应挖掘

在找出频繁项集后,虽然已经解决了数据稀疏性导致难以获取频繁项集的问题,但数据关联关系仍然不够直观和全面,难以充分揭示数据间的复杂联系。为此,通过构建挖掘树的方法对其进行分析。以树形结构为框架,将 Apriori 算法得到的频繁项集作为结点,通过自适应结合旧结点与新结点逐步构建挖掘树,直至无法再结合新的结点。此时生成的挖掘树能全面呈现数据间的关联关系,从而得到充分且精准的电商网页关联数据挖掘结果。这种方法不仅能提高数据的可视化程度,还能更准确揭示数据间的内在联系,为电商网站的商品推荐和市场趋势预测提供更强有力的支持。

基于自适应关联规则挖掘方法^[19],以树形结构^[20]为框架,设计出电商网页关联数据的自适应挖掘算法。具体实现流程如下。

1) 当候选项 i 的频度 P_{i+1}^i 与最小支持度 $\min\{S_i\}$ 使式(7)成立时,取得所有频繁项集,经 k -means 算法归类后,得到电商网页数据中频繁项属性扩展后全部的一类关联规则。

2) 令当前的网页频繁项集类关联规则单元为树的结点,相对稳定的类关联规则单元为旧结点,新发现的类关联规则单元为新结点。根据必有一个新结点的准则,自适应两两结合树的结点。

3) 计算结合点的支持度和置信度。若两个结点分别是 x 、 y ,则结合点的支持度 $S(x\Rightarrow y)$ 与置信度 $C(x\Rightarrow y)$ 分别由

$$S(x\Rightarrow y) = \frac{|x \cup y|}{|N|},$$

(9)

$$C(x\Rightarrow y) = \frac{|x \cup y|}{|x|}$$

(10)

解得。

若 $S(x\Rightarrow y)$ 、 $C(x\Rightarrow y)$ 与最小支持度 $\min\{S(x\Rightarrow y)\}$ 、最小置信度 $\min\{C(x\Rightarrow y)\}$ 均满足:

$$S(x\Rightarrow y) = \min\{S(x\Rightarrow y)\},$$

(11)

$$C(x\Rightarrow y) = \min\{C(x\Rightarrow y)\}.$$

(12)

则将对对应结合点纳入最终的类关联规则集合。当再无结合点时,得到最终的关联数据挖掘树,即电商网页关联数据的自适应挖掘结果。

3 电商网页关联数据自适应挖掘算法实例验证

3.1 实验

针对某个大型电商网站的网页,对其关联数据自适应挖掘进行实验。该平台网站的结构为超文本标记语言,经 Chrome 开发者工具分析,识别出不同页面的 URL 模式,据此构造爬虫请求。使用 Python 的 Scrapy 框架,编写爬虫程序;采用 MongoDB 数据库,存储爬取到的数据,并通过 matplotlib 实现挖掘结果的可视化展示。挖掘算法的运行平台架构如图 2 所示。

经调试后,得到电商网页关联数据自适应挖掘算法涉及的关键参数的最优值,以确保算法的高效性与准确性。参数值如表 1 所示。

表 1 挖掘算法相关参数设置

Tab. 1 Parameter settings related to mining algorithms

部分	名称	型号	部分	名称	型号
Apriori 算法	最小支持度	0.5		请求并发量/个	16
	最小置信度	0.6		同域名请求并发量/个	9
	项集最大长度	2		数据清洗管道优先级	300
Scrapy 框架	下载请求时间间隔/s	2		数据存储管道优先级	400

3.2 电商网页数据的爬虫爬取结果分析

从目标电商网站的网页上，以西装为对象进行爬虫爬取，得到上万个店铺的相关数据，包括销售量、商品名称、商品价格等近百万条数据。其中，爬取的部分网页数据如图 3 所示。

由图 3 可见，爬虫技术能自动执行数据收集任务，并且通过精确定位网页中的目标信息，能确保数据爬取的准确性和完整性，有效避免手动输入可能出现的错误。同时，爬虫技术凭借较强的适应性，可以根据不同的需求和目标网站的结构进行定制，更好地处理该电商网页上复杂的数据收集任务。电商网页数据的爬取结果表明，该方法能为关联数据挖掘提供丰富、准确的数据支持。

数据分类	变量名称	数据描述
商品本身属性	销售量	800 笔
商品本身属性	商品名称	海澜之家男士商务西装 春秋款 正装
商品本身属性	商品价格	取值范围为 500 - 2000 元
商品本身属性	款式风格	欧式、修身、单排扣
所属店铺信息	店铺所在地	上海
所属店铺信息	店铺名称	海澜之家官方旗舰店
所属店铺信息	店主电话	138xxxx5678 (假设)

数据分类	变量名称	数据描述
商品本身属性	销售量	500 笔
商品本身属性	商品名称	雅戈尔男士羊毛西装 冬季款 正装
商品本身属性	商品价格	取值范围为 1000 - 3000 元
商品本身属性	款式风格	中式、宽松、双排扣
所属店铺信息	店铺所在地	宁波
所属店铺信息	店铺名称	雅戈尔官方旗舰店
所属店铺信息	店主电话	139xxxx8910 (假设)

数据分类	变量名称	数据描述
商品本身属性	销售量	300 笔
商品本身属性	商品名称	报喜鸟男士时尚西装 夏季款 轻薄正装
商品本身属性	商品价格	取值范围为 300 - 1500 元
商品本身属性	款式风格	韩版、修身、单排扣
所属店铺信息	店铺所在地	温州
所属店铺信息	店铺名称	报喜鸟官方旗舰店
所属店铺信息	店主电话	136xxxx2345 (假设)

数据分类	变量名称	数据描述
商品本身属性	销售量	1200 笔
商品本身属性	商品名称	罗蒙男士休闲西装 春秋款 混纺正装
商品本身属性	商品价格	取值范围为 400 - 1800 元
商品本身属性	款式风格	美式、宽松、单排扣
所属店铺信息	店铺所在地	宁波
所属店铺信息	店铺名称	罗蒙官方旗舰店
所属店铺信息	店主电话	133xxxx7890 (假设)

图 3 电商网页爬虫爬取数据示意图

Fig.3 Schematic diagram of data crawling by e-commerce web crawlers

3.3 电商网页爬取数据频繁项集获取结果分析

从电商网页上爬虫爬取的西装数据中，获取以购买行为为目标的频繁项集，具体如图 4 所示。

其中，5 个字母分别表示马甲、裤子、衬衫、领带、外套；频繁项集为多次出现的商品组合，而非频繁项集则为较少出现的商品组合。

由图 4 可见，经多次迭代和优化，该方法能有效解决海量商品数据的高效处理和分析问题，成功地从爬虫爬取的网页数据中提取出一系列高频出现的商品组合，即频繁项集，直观展示出哪些商品在电商平台上经常被消费者购买，进而揭示出消费者在购买西装时的关联购买行为和市场需求趋势。



图 2 挖掘算法运行平台架构示意图

Fig.2 Schematic diagram of the architecture of the mining algorithm running platform

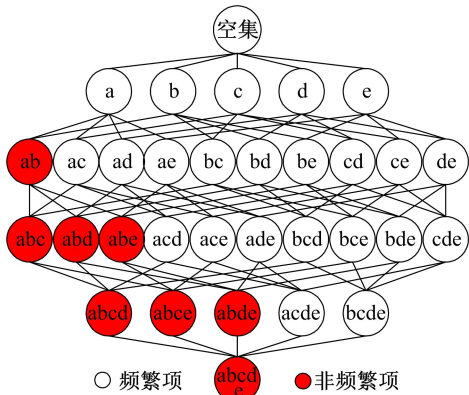


图 4 电商网页爬取数据频繁项集示意图

Fig.4 Schematic diagram of frequent itemsets for crawling data from e-commerce web pages

3.4 电商网页关联数据自适应挖掘结果分析

针对爬取西装商品数据的频繁项集,电商网页关联数据自适应挖掘结果如图 5 所示。由图 5 可见,所提方法基于自适应关联规则挖掘方法所构建的挖掘树,不仅能成功挖掘出电商网页中的关联数据,直观给出商品间的关联模式,而且还可以深入理解消费者的购物行为和偏好,准确捕捉消费者行为和商品关联。该方法有助于优化商品布局、推荐算法和营销策略,为电商网站提供宝贵的业务洞察,提升电商网站的竞争力和用户体验。

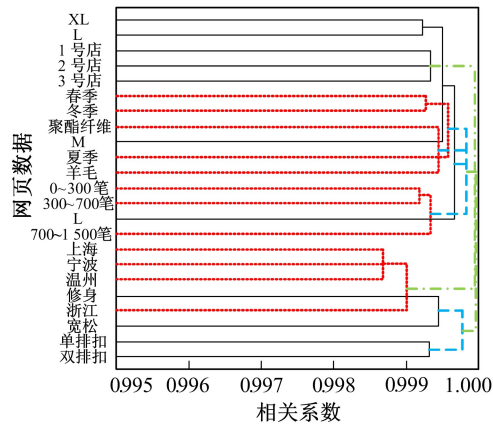


图 5 电商网页关联数据自适应挖掘结果示意图
Fig. 5 Schematic diagram of adaptive mining results for e-commerce webpage association data

4 结 语

电商平台的不断涌现和商品种类的日益丰富,使电商网页数据呈现出爆炸式增长。这些数据中蕴含着丰富的市场信息和用户行为特征,而商家深入挖掘这些数据有助于了解市场需求、优化产品设计和制定有效的营销策略。从数量庞大的电商网页数据中迅速且准确地发现有价值的信息,已成为急需解决的关键问题。因此,借助爬虫技术对目标网页链接和内容的自动识别与抓取能力,笔者提出一种电商网页关联数据自适应挖掘算法。今后该算法可以在以下几个方面对其进行研究和改进:

- 1) 通过优化爬虫策略和数据解析方法,进一步提高数据抓取的效率;
- 2) 针对不同电商网页的结构和内容差异,通过引入机器学习算法,进一步增强算法的自适应能力;
- 3) 除了电商网页数据挖掘外,将该算法拓展到社交媒体数据分析、新闻网页内容挖掘等其他领域的数据挖掘任务中。

参考文献:

[1]TANG W, LI G. Enhancing Competitiveness in Cross-Border E-Commerce through Knowledge-Based Consumer Perception Theory: An Exploration of Translation Ability [J]. Journal of the Knowledge Economy, 2023, 15(3): 14935-14968.

[2]GUANGKUAN D, JIANYU Z, YING X, et al. Revisiting E-Commerce Platforms' Strategies of Exercising Channel Power: A Contingency Perspective [J]. Journal of Business & Industrial Marketing, 2024, 39(10): 2239-2256.

[3]UMAMAHESWARI N, KUMUTHA K, KALAIVANI S M, et al. Backpropagation Shuffled Leaping Neural Network Implementation in Classification of DDoS Packet Flow Traffic in Data Mining [J]. SN Computer Science, 2024, 5(8): 1131-1131.

[4]韩永印, 王侠, 王志晓. 基于决策树的社交网络隐式用户行为数据挖掘方法 [J]. 沈阳工业大学学报, 2024, 46(3): 312-317.

HAN Y Y, WANG X, WANG Z X. Data Mining Method Based on Decision Tree for Implicit User Behavior in Social Network [J]. Journal of Shenyang University of Technology, 2024, 46(3): 312-317.

[5]陈万志, 赵帅, 方圆, 等. 改进 PrefixSpan 的行为轨迹数据挖掘算法 [J]. 辽宁工程技术大学学报(自然科学版), 2023, 42(4): 506-512.

CHEN W Z, ZHAO S, FANG Y, et al. Behavioral Trajectory Data Mining Algorithm with Improved PrefixSpan [J]. Journal of Liaoning Technical University (Natural Science Edition), 2023, 42(4): 506-512.

[6]夏小雅, 赵生字, 韩凡宇, 等. 面向开源协作数字生态的信息服务与数据挖掘 [J]. 计算机科学, 2024, 51(10): 187-195.

XIA X Y, ZHAO S Y, HAN F Y, et al. Data Mining and Information Service for Open Collaboration Digital Ecosystem [J]. Computer Science, 2024, 51(10): 187-195.

[7]曹俊彬, 邵航, 姜坤, 等. 军用航空器事故关键质量特性的数据挖掘模型 [J]. 计算机工程与设计, 2024, 45(2): 562-570.

CAO J B, SHAO H, JIANG K, et al. Data Mining Model for Critical-to-Quality Characteristics of Military Aviation Accidents [J]. Computer Engineering and Design, 2024, 45(2): 562-570.

- [8] 刘多林, 吕苗. Scrapy 框架下分布式网络爬虫数据采集算法仿真 [J]. 计算机仿真, 2023, 40(6): 504-508.
LIU D L, LÜ M. Simulation of Distributed Web Crawler Data Collection Algorithm under Scrapy Framework [J]. Computer Simulation, 2023, 40(6): 504-508.
- [9] GAVIN R, THORSTEN W, MARKUS S, et al. TomoTwin: Generalized 3D Localization of Macromolecules in Cryo-Electron Tomograms with Structural Data Mining [J]. Nature Methods, 2023, 20(6): 871-880.
- [10] 李远航, 王劲林, 韩锐. 信息中心网络中一种基于内容热度的分区缓存替换方法 [J]. 电子设计工程, 2023, 31(6): 133-138, 143.
LI Y H, WANG J L, HAN R. A Partition Cache Replacement Method Based on Content Popularity in Information Centric Networking [J]. Electronic Design Engineering, 2023, 31(6): 133-138, 143.
- [11] 周艺腾, 唐鑫, 金路超. 基于自适应 MSB 可逆信息隐藏的图像云数据密文安全去重机制 [J]. 计算机科学, 2024, 51(12): 352-360.
ZHOU Y T, TANG X, JIN L C. Adaptive MSB Reversible Data Hiding Based Security Deduplication for Encrypted Images in Cloud Storage [J]. Computer Science, 2024, 51(12): 352-360.
- [12] 杨阳蕊, 朱亚萍, 刘雪梅, 等. 水利工程文本中抢险实体和关系的智能分析与提取 [J]. 水利学报, 2023, 54(7): 818-828.
YANG Y R, ZHU Y P, LIU X M, et al. Intelligent Analysis and Joint Extraction of Rescue Entities and Relationships in Water Project Texts [J]. Journal of Hydraulic Engineering, 2023, 54(7): 818-828.
- [13] 张磊, 焦晶, 李勃昕, 等. 融合机器学习和深度学习的大容量半结构化数据抽取算法 [J]. 吉林大学学报(工学版), 2024, 54(9): 2631-2637.
ZHANG L, JIAO J, LI B X, et al. Large Capacity Semi Structured Data Extraction Algorithm Combining Machine Learning and Deep Learning [J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(9): 2631-2637.
- [14] SOMU K, VELU C. A Novel Prediction of Sales and Purchase Forecasting for Festival Season of Hypermarkets with Customer Dataset Using Apriori Algorithm Instead of FP-Growth Algorithm to Improve the Accuracy [J]. Electrochemical Society Transactions, 2022, 107(1): 12647-12659.
- [15] 韩虎, 孔博, 何勇禧, 等. 基于剪枝策略的知识增强方面级情感分析 [J]. 华中科技大学学报(自然科学版), 2024, 52(11): 140-146.
HAN H, KONG B, HE Y X, et al. Aspect Based Sentiment Analysis Based on Knowledge Enhancement of Pruning Strategy [J]. Journal of Huazhong University of Science and Technology (Nature Science Edition), 2024, 52(11): 140-146.
- [16] 张敏, 沈嘉裕. 突发公共卫生事件中政务短视频主题与用户行为的关联演化研究 [J]. 情报杂志, 2023, 42(3): 181-189.
ZHANG M, SHEN J Y. Research on the Evolution of the Association between Governmental Short Video Themes and User Behavior in Public Health Emergency [J]. Journal of Intelligence, 2023, 42(3): 181-189.
- [17] 朱宇婧, 陈芳, 王学昭. 基于商业管制清单-专利网络映射的关键核心技术识别研究——以工业软件为例 [J]. 数据分析与知识发现, 2024, 8(10): 1-13.
ZHU Y J, CHEN F, WANG X Z. Identifying Core Technologies Based on the Commerce Control List-Patent Network Mapping: Case Study of Industrial Software [J]. Data Analysis and Knowledge Discovery, 2024, 8(10): 1-13.
- [18] 贺帆, 刘漫丹, 钟超. 基于动态最小支持度的增量频繁序列挖掘 [J]. 华东理工大学学报(自然科学版), 2024, 50(2): 257-263.
HE F, LIU M D, ZHONG C. Dynamic Minimum Support Incremental Frequent Trajectory Mining Algorithm [J]. Journal of East China University of Science and Technology (Natural Science Edition), 2024, 50(2): 257-263.
- [19] 杨建平, 向月, 刘俊勇. 面向配电网投资决策的小样本关联规则自适应迁移学习方法 [J]. 中国电机工程学报, 2022, 42(16): 5823-5834, 6159.
YANG J P, XIANG Y, LIU J Y. Adaptive Transfer Learning of Small Sample Correlation Rules for Distribution Network Investment Decision [J]. Proceedings of the CSEE, 2022, 42(16): 5823-5834, 6159.
- [20] 刘昌春, 张凯, 包美凯, 等. 基于语境与文本结构融合的中文拼写纠错方法 [J]. 南京大学学报(自然科学), 2024, 60(3): 451-463.
LIU C C, ZHANG K, BAO M K, et al. Research on Chinese Spelling Correction Based on the Integration of Context and Text Structure [J]. Journal of Nanjing University (Natural Sciences), 2024, 60(3): 451-463. (责任编辑: 刘东亮)