

文章编号: 2095-2163(2021)03-0113-02

中图分类号: TN911.1-34

文献标志码: A

Apriori 算法在无线网络数据智能挖掘中的应用

刘云萍

(太原工业学院 计算机工程系, 太原 030008)

摘要:为解决传统挖掘方法进行数据挖掘时,存在规定时间范围内数据挖掘量少,导致挖掘效率不高的问题,提出 Apriori 算法在无线网络数据智能挖掘中的应用研究。通过数据挖掘关联规则设计、基于 Apriori 算法的无线网络数据文本分类和无线网络数据离群点智能过滤,实现无线网络数据智能挖掘。通过实验证明,所提挖掘方法与传统方法相比挖掘效率得到明显提升。

关键词: Apriori 算法; 无线网络; 关联规则; 智能挖掘

Application of Apriori algorithm in intelligent data mining of wireless network

LIU Yunping

(Computer Engineering Department, Taiyuan Institute of Technology, Taiyuan 030008, China)

[Abstract] In order to solve the problem that the amount of data mining in the specified time range is small when the traditional mining method is used for data mining, which leads to low mining efficiency, this paper proposes the application research of Apriori algorithm in wireless network data intelligent mining. Through the design of data mining association rules, using text classification of wireless network data based on Apriori algorithm and intelligent filtering of wireless network data outliers, intelligent mining of wireless network data is realized. Experiments show that the mining efficiency of the proposed method is significantly improved compared with the traditional method.

[Key words] Apriori algorithm; wireless network; association rules; intelligent mining

0 引言

当前,数据挖掘技术各个领域中都拥有着十分广泛的应用,并逐渐成为具有较高价值意义的工程类工具和手段。大部分数据挖掘过程中都会产生海量的数据,因此导致计算量庞大,很难实现对数据的精准挖掘,同时还会浪费大量的时间^[1]。目前的挖掘方法有着严重的局限性,在实际应用过程中多会出现数据存储中心宽带消耗量大、数据易发生泄漏等问题^[2]。面对该问题现状,相关领域的研究人员逐渐将针对无线网络中数据的挖掘方法作为主要研究内容。研究可知,Apriori 算法能够通过逐层迭代的方式找出不同环境当中数据之间的关联,并进一步推导出其规则。当前 Apriori 算法应用效果十分良好。基于此,本文拟开展 Apriori 算法在无线网络数据智能挖掘中的应用研究。

1 基于 Apriori 算法的无线网络数据智能挖掘方法设计

1.1 无线网络数据挖掘关联规则设计

利用 Apriori 算法对无线网络数据文本进行分

类前,要明确数据挖掘关联规则。在实际挖掘时,则要经历 2 次无线网络数据库。第一次是在开始挖掘频繁候选集阶段,在此阶段当中,主要完成对频繁项集的生成;第二次是挖掘频繁候选集阶段^[3]。无线网络数据挖掘关联规则设计流程可阐释表述如下。

步骤 1 将选取的无线网络数据样本进行分段处理,将其输入到集群中的各个节点上,计算每一项的支持度。执行相应的程序后,获取到本地数据,并输入一个数据记录^[4]。

步骤 2 是 Apriori 算法中最关键的一步。在 mapper 中输入,将其与步骤 1 中的数值序列进行比较,即可获得相应的结果。

步骤 3 按照上述方法来分配节点,将不同的无线网络数据信息放置在不同的节点之上^[5]。在此过程中,会出现同一时间,对应的频繁项集不处于同一节点上。针对这一问题,运用 Apriori 算法中默认的对关键数值排序的功能,得到无线网络数据挖掘关联规则。

1.2 基于 Apriori 算法的无线网络数据文本分类

为实现对无线网络中海量数据的挖掘,将编程思想作为基础,结合 Apriori 算法的设计原理,将无

基金项目:太原工业学院教研项目(JG201906)。

作者简介:刘云萍(1980-),女,硕士,讲师,主要研究方向:数据挖掘。

收稿日期:2020-10-26

线网络环境下的海量数据进行重构,并完成对无线网络数据文本的分类。在实际应用中,单纯使用 Apriori 算法很难达到数据挖掘对无线网络文本分类的要求,因此,当挖掘过程中文本经常出现特定特征下的数据时,会导致挖掘方法在应用后产生大量的公共频繁词汇^[6]。至此,为了能够有效避免这一问题的产生,本文在实际执行 Apriori 算法时,引入 tf-idf 算法,对无线网络环境中每一个数据的权重进行合理分配,再以此为基础,对无线网络数据文本进行分类,以此能够有效保证分类结果的准确性,同时也能够进一步加强 Apriori 算法应用的实际意义。

1.3 无线网络数据离群点智能过滤

在上述相关行为的支撑下,根据不同数据点的所属簇组进行研究^[7]。分析不同簇组的核心数据值,对其簇组核心值进行计算,评估计算结果,以此作为评价数据离群点是否为冗余数值。综合上述分析,对数据离群点簇组核心值的计算公式可写为:

$$\hat{\pi}_i^d(t) = \frac{1}{d} \left(\sum_{T=t+w-d}^{t-1} 1_{[S(T)=s_i]} \right). \quad (1)$$

其中, $\hat{\pi}$ 表示核心数据; d 表示对数据群的挖掘关联规则; i 表示簇组数据群点排列组合行为的发生次数; s 表示冗余数据对于数据挖掘行为造成的风险数值; T 表示一个挖掘周期; t 表示一个智能过滤周期; w 表示影响无线网络数据离群点智能过滤的多种因素。

2 对比实验

为验证本文提出的基于 Apriori 算法的无线网络数据智能挖掘方法在实际应用中的效果,选择某区域内的无线局域网作为实验环境。对比本文方法与传统挖掘方法的性能。本文选用数据挖掘量作为评价 2 种方法的指标,设置 2 种方法的挖掘时间分别为 100 ms、500 ms、1 000 ms、2 000 ms 和 3 000 ms,在保证 2 种方法的挖掘时间相同的情况下,对其挖掘结果进行记录。2 种挖掘方法的实验结果对比见表 1。

根据表 1 中的数据结果可以看出,随着挖掘时间的增加,本文挖掘方法的数据挖掘量增加幅度不断变大,而传统挖掘方法增加幅度较缓。在 3 000 ms 挖掘时间条件下,传统挖掘方法并未完成相应的挖掘任务,挖掘量仍然停留在 2 000 ms 时间时的结果,已经无法在该无线局域网环境下对更多的数据进行挖

掘。因此,通过对比实验证明,本文提出的基于 Apriori 算法的无线网络数据智能挖掘方法效率更高,满足海量数据的挖掘需要。

表 1 2 种挖掘方法实验结果对比表

Tab. 1 Comparison table of experimental results of two mining methods

挖掘时间/ms	本文挖掘方法 数据挖掘量/bit	传统挖掘方法 数据挖掘量/bit
100	263.54	102.45
500	596.47	215.34
1 000	1 256.84	423.64
2 000	3 564.28	526.24
3 000	5 986.45	526.24

3 结束语

本文提出基于 Apriori 算法的无线网络数据智能挖掘方法设计研究。将该方法付诸应用能够有效解决挖掘效率低、挖掘量有限的问题。但由于研究时间的有限,本文仅针对无线网络中的无线局域网进行了研究,关于无线广域网在使用本文挖掘方法时会有一定概率出现数据泄露问题,后续还将进行更加深入的研究。

参考文献

- [1] 兰济乐,阮叶萍,蒋东晓. 基于数据挖掘和网络药理学的失眠中药配方规律及作用机制研究[J]. 浙江中医药大学学报, 2020, 44(12): 1230-1238, 1241.
- [2] 周文静,朱彦,郭怡含,等. 基于数据挖掘和网络药理学的丹参配伍规律及其防治冠心病作用机制研究[J]. 中国中医基础医学杂志, 2020, 26(11): 1643-1648.
- [3] 江佳琳,何静玲,黎小斌. 基于数据挖掘及网络药理学探讨中药灌肠治疗慢性盆腔炎的核心中药及机制[J]. 广州中医药大学学报, 2021, 38(2): 379-385.
- [4] 邵臧杰,王盼盼,李红,等. 基于数据挖掘及复杂网络技术探讨周仲瑛教授治疗哮喘发作期经验[J]. 中国中医急症, 2020, 29(10): 1744-1748, 1753.
- [5] 黄羚,刘铁钢,白辰,等. 基于数据挖掘和网络药理学的谷晓红教授治疗小儿急性上呼吸道感染的研究[J]. 世界中医药, 2020, 15(20): 3007-3013.
- [6] 韦凌霞,丁茂鹏,王志旺,等. 基于数据挖掘及网络药理学方法探讨含当归中药复方抗肝纤维化的用药规律及作用机制[J]. 中药新药与临床药理, 2020, 31(11): 1323-1331.
- [7] 王光耀,许光兰,陈小丽,等. 基于数据挖掘和网络药理学的古方治疗肺结核的用药规律及机制分析[J]. 世界科学技术-中医药现代化, 2020, 22(7): 2256-2268.