

灭火指挥数据挖掘研究

施伟荣

(上海市宝山区公安消防支队司令部, 上海 201901)

摘 要: 根据火灾和灭火大数据中挖掘数据的关联性, 将具有相关联的属性值和频繁出现的属性值提取出来, 建立关联规则, 并通过大数据挖掘 Apriori 算法准确提高灭火指挥数据应用的准确性. 讨论 Apriori 关联规则算法的改进, 降低对数据库的扫描次数, 减少后选集的生成, 提高灭火指挥数据的数据挖掘性能和效率, 满足灭火指挥数据建设需求.

关键词: 灭火指挥; 数据挖掘; Apriori 算法; 改进

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2016)04-0141-03

Research on Fire Fighting Command Data Mining

SHI Wei-rong

(Baoshan District Public Security Fire Brigade Headquarters, Shanghai 201901, China)

Abstract: From the fire and fire the relevance of data mining in large data, to have associated attribute values and frequent attribute values extracted and establish the association rules, and through data mining Apriori algorithm accuracy improve fire accuracy of command data applications. This paper discusses the improvement of Apriori algorithm of association rules, reduce the times of database scanning, reduced after Anthology of generation, improve fire fighting command data data mining performance and efficiency, meet the fire command data construction demand.

Key words: Fire fighting command; Data mining; Apriori algorithm; Improvement

1 引言

在现代化消防工作中, 应用大数据获取准确火灾信息和灭火信息, 并对其进行挖掘处理提高灭火效率和能力已经成为消防工作中的重要内容. 随着火灾数据及灭火数据的日益增加, 从灭火指挥处警平台的数据库中快速提炼有效信息变得非常困难, 由此提出对大数据进行数据挖掘, 从大量的信息、不完整的信息、模糊的信息及含有噪声的信息中提取关键的数据、隐含的数据, 是大数据挖掘应用的关键^[1]. Apriori 关联规则数据挖掘算法是利用大数据的关联性进行数据挖掘, 经过各领域的应用证明, 该方法具有较好的数据处理能力. 但是, Apriori 算法有一个关键的缺点, 就是在进行数据关联计算时, 会多次扫描数据库原始数据, 不仅生成大量的候选集,

而且效率也不高, 并不能达到人们对数据挖掘的理想需求. 所以, 本文以灭火大数据为研究对象, 提出一种 Apriori 改进算法, 通过对灭火数据的结果检验证明改进后的 Apriori 算法能够改善数据挖掘性能, 提高数据挖掘效率, 为提高消防灭火指挥能力提供可用依据.

2 灭火指挥数据关联规则

在灭火指挥数据库中, 关联规则的配置是用户指定灭火指挥数据的最小支持度和最小置信度后进行数据挖掘寻找数据间的关联规则, 关联规则从两个方面进行配置.

2.1 发现灭火指挥数据中的频繁项目集

给定最小支持度 minsupport, 寻找所有灭火指挥数据中频繁出现的项目集, 该项目集满足在灭火

指挥数据库上的支持度,即项目集 I_i 在数据库 D 中所占的百分比. 我们可以设 $I = \{i_1, i_2, \dots, i_m\}$ 为一个灭火大数据集合, 灭火指挥数据库 $D = \{t_1, t_2, \dots, t_n\}$, TID 为灭火指挥数据集的唯一标识, t 为灭火数据, 满足 $t \subseteq I$. 设灭火数据集 $I_i \subseteq I$ 在灭火指挥数据库 D 的支持度 $\text{Support}(I_i) = \|\{t \in D \mid I_i \subseteq t\} \parallel I \parallel D \parallel$.

2.2 生成灭火数据关联规则

关联规则是用户给定置信度后再每个最大频繁项目集中寻求灭火指挥数据置信度不小于最小置信度的关联规则^[2]. 设 $I_1 \Rightarrow I_2$ 的关联规则需要满足一定的置信度得出, 则需满足的置信度 $\text{confidence}(I_i \Rightarrow I_2) = \text{support}(I_1 \Rightarrow I_2) / \text{support}(I_1)$, 其中 $I_1, I_2 \subseteq I, I_1 \cap I_2 = \Phi$.

3 Apriori 算法在灭火指挥数据关联中的应用

Apriori 在灭火指挥数据挖掘计算中通过验证与推测分析火灾类型数据、相似火灾指挥处理数据等为决策提供决策依据. 在灭火数据中, A 数据 $\rightarrow B$ 数据, 支持度 $P(A \cap B)$, A 数据与 B 数据并集, 置信度在 A 发生的同时 B 数据发生的概率 $P(AB)/P(A)$, 例如: 对消防站 A 与消防站 B 的执勤车辆调度情况进行分析, 在消防站 A 的某块辖区, 消防站 $A \Rightarrow$ 消防站 B , 支持度 3%, 置信度 24%. 这说明在该区域, 消防站 A 作为主管中队先到场后, 消防站 B 出车到场应援的概率为 3%, 而消防站 B 作为应援中队先到场后, 主管中队消防站 A 后到场的概率为 24%. 这些数据为消防站点布局的合理性和辖区划定分配提供了科学准确的大数据决策支撑.

4 Apriori 算法改进

4.1 Apriori 算法改进方法

Apriori 算法最大的缺点是会产生大量的候选集, 对数据挖掘的准确度会产生影响. 此外, 数据挖掘过程反复扫描数据库, 降低了数据挖掘的效率.

本文提出的 Apriori 算法改进思路是以频繁 1-项集为候选项目集, 增加判断条件 $L_k.\text{count} \geq \min \text{sup port}$ 后得到频繁集 $k+1$ 项集, 再进行 Apriori 关联计算, 计算原理:

将数据项 A 增加至 $I(A \notin I)$ 得到的结果集 $(A \cup I)$ 频度 $\leq I$ 的频度, 如果 I 不满足所设定的最小支持度门限值, 则过滤不计算, 由此得出推论, 采用改进后的 Apriori 算法扫描数据库次数较少, 并且挖掘出的候选集数据更加精炼, 在大规模数据挖掘

中, 性能和效率均有巨大的改善^[3].

4.2 改进 Apriori 算法在灭火指挥数据关联中的应用

在改进的 Apriori 算法中, 灭火指挥 K 维数据项集 X 为高频数据项集的条件要求, 其所包含的 $k-1$ 维子集均为最大数据项集, 数据项集的组成为 Y_k-1 . 当 K 维数据项集 T_k 的任意一个 $k-1$ 维子集不是最大数据项集, 则 K 维数据项集本身也不是最大数据项集. T_k 是灭火 K 维数据项集, 当所有的 $k-1$ 维高频数据项集集合 L_k-1 中包含 T_k 的 $k-1$ 维子集的个数小于 k , T_k 则不是 k 维的最大数据项集^[4].

扫描灭火指挥数据项集合 T , 找出频繁 1-项集合 L ; 对 L 中的灭火数据项按照支持计数进行降序排列, 创建 FP 树. for 灭火指挥数据项集 T 中的每一个数据 t , 找出 t 中频繁 1-项集合 L_t , Insert_FP 创建灭火指挥数据项分支, $L_S = \text{Search_FP}$ 找出所有频繁项集, 在 L_S 中产生关联规则集合 R_S .

举例说明: 某大型商务综合体内发生火灾, 起火点位于地上 3 层, 着火时间为白天营业时间. 火灾扑救完毕后, 对相关灭火指挥数据实施处理. 灭火指挥数据项集合 T 包括: 火警信息、火场 GIS 信息、单位电子预案、水源分布、战术应用、交通路线等. 通过对着火点的数据挖掘找出频繁项集合 L , 根据支持计数进行降序排列创建 FP 树如图 1 所示.

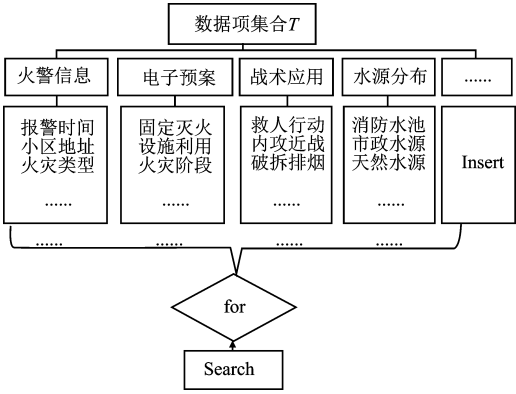


图 1 支持计数进行降序排列 FP 树

依据支持计数进行排序后 for 灭火数据项集合中的每个数据 t , 通过 insertFP 创建指挥数据项分支, $L_S = \text{Search_FP}$ 找出频繁集, 在改进的 Apriori 算法中, 确定该大型商务综合体灭火指挥 K 维数据项集, 设定 X 为高频数据项集火警信息、电子预案、战术应用、水源分布等的关键条件, 确定报警时间、火灾类型、火灾阶段、救人、内攻、破拆、水源、交通等

最数据项集,以此为出发点进行数据关联,将可能存在的各类灭火指挥信息通过关联规则找出关联结果,据此调整现场处置流程和灭火作战预案,优化灭火战术应用.

5 结果分析

将改进后的 Apriori 算法与未改进的 Apriori 算法进行比较,以 2 200 条记录为测试数据,C++实现算法,算法执行时间如表 1 所示.

表 1 2 200 条记录运行时间

最小支持度	0.35	0.3	0.25	0.2	0.15	0.1	0.07	0.05
Apriori 运行时间	1.66	1.66	3.12	5.88	6.52	7.56	15.42	19.58
改进后 Apriori 运行时间	...	1.2	2.3	4.62	4.94	5.94	12.30	15.72

如表 1 所示,改进后的 Apriori 算法在各最小支持度中都要比未改进的 Apriori 算法执行时间要段,速度的改变与最小支持度相关联,最小支持度越小则算法运行时间越大,最小支持度差异较小时会得到较大的强项集 L_k .

5.1 性能比对分析

对灭火指挥数据进行改进后的 Apriori 算法计算挖掘,与未改进的 Apriori 算法计算挖掘性能比较,使用最小支持度 0.2,试验数据采用灭火指挥数据库中的 300~2 200 个项目序列,运行时间比对如图 2 所示.

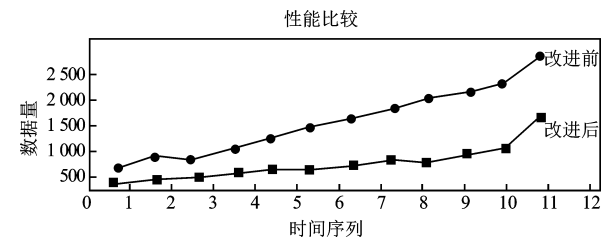


图 2 算法改进前后性能比对

如图 2 所示,随着灭火指挥数据库中的数据项目不断增加,改进后的 Apriori 算法要比改进前的 Apriori 算法在候选集数据量上明显降低,执行性能明显提高.

5.2 效率比对分析

运行时间比对如图 3 所示.

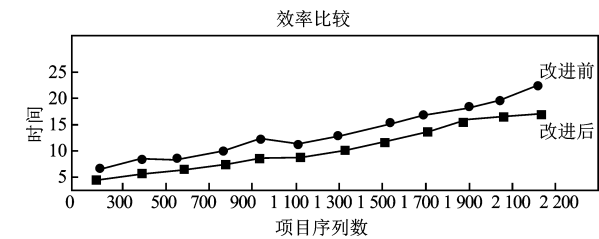


图 3 算法改进前后效率比对

如图 3 所示,随着灭火指挥数据库中的数据项目不断增加,改进后的 Apriori 算法要比改进前的 Apriori 算法在执行时间上明显降低,执行效率明显提高.

6 结束语

灭火指挥数据挖掘改进 Apriori 算法是提高灭火指挥数据应用的有效方法,通过实验数据的比对分析,结果在同一灭火指挥数据库挖掘中,执行性能和执行效率均比未改进前有较明显提高.对于灭火指挥数据库的挖掘要从数据频繁集合和关联规则入手,通过已有数据属性的特征分析为正确处理火情提供决策依据,在灭火指挥大数据应用中具有非常重要的现实意义.

参考文献:

[1] 中国消防手册. 灭火救援基础[J]. 上海科学技术出版社,2006,9(12):1587-1641.

[2] 田敬波. 消防大数据应用剖析[J]. 电信技术,2014,30(8):67-69.

[3] 张恺,郑晶. 一种基于云计算的新的关联规则 Apriori 算法[J]. 甘肃联合大学学报:自然科学版,2012,8(6):16-19.

[4] Wu Xindong, Zhu Xingquan, Wu Gongqing, et al. Data mining with big data[J]. Knowledge and Data Engineering, 2014,26(1):134-136.

作者简介:

施伟荣 男,(1979-),硕士研究生,工程师.研究方向为灭火救援技术研究. E-mail:swr_fire@aliyun.com