

基于 Apriori 算法的音乐推荐算法研究

严海卫, 林春花, 徐欢潇, 陈枢茜

(南通理工学院, 江苏 南通 226000)

摘要: 随着信息冗余程度日益加深, 筛选信息难度也日益加大, 关联算法的出现很大程度上缓解了这种问题。听音乐是人们生活中的一种放松方式, 然而用户和音乐之间存在着信息不对称的问题, 将 Apriori 算法应用于音乐推荐, 利用用户的偏好和行为数据来为用户提供个性化的服务。Apriori 算法根据用户之间或者物品之间的关联度来预测用户对未知物品的评分或喜好程度。使用 SPSS Modeler 工具可以进行建模, 从而将 Apriori 算法应用于音乐推荐研究。

关键词: 音乐推荐; Apriori 算法; 关联度; SPSS Modeler 工具

DOI:10.16184/j.cnki.comprg.2025.02.040

1 概述

音乐是人类生活中不可或缺的艺术元素, 它可以给人们带来愉悦和放松, 也可以表达人们的情感和个性。随着互联网和数字技术的发展, 音乐资源变得越来越丰富和多样, 用户可以通过各种在线平台和应用来获取和享受音乐。然而, 这也带来了一个问题, 即如何在海量的音乐中找到符合自己口味和需求的音乐。此时可以使用音乐推荐系统帮助用户提高筛选效率。音乐推荐系统是一种利用用户的偏好和行为数据来为用户提供个性化的音乐推荐服务的系统^[1]。它可以帮助用户发现新的或潜在感兴趣的音乐, 也可以增加用户对平台或应用的满意度。音乐推荐系统是推荐系统研究领域中的一个重要且具有挑战性的子领域, 它涉及多种因素, 例如, 用户的心理状态、情境背景、社会关系, 以及音乐本身的特征和风格等^[2]。

Apriori 算法应用于音乐推荐领域, 可以根据用户之间或者物品之间的关联度来预测用户对未知物品的评分或喜好程度。例如, 如果爱听轻音乐和古典乐, 那么很有可能爱听民谣。Apriori 算法进行逐层的搜索, 采用迭代的方法找出数据库中项集与项集的关系^[3], 从而形成规则, 借助 SPSS Modeler 工具建模, 将 Apriori 算法应用于音乐推荐研究, 实现音乐推荐服务。

2 音乐推荐

随着大数据的兴起, 音乐产业面临着巨大的变革。个性化音乐推荐成为数据挖掘方向炙手可热的研究领域, 音乐推荐算法应用于音乐网站, 可以使音乐制作人和音乐公司更好地推广和销售自己的作品, 开拓新的商业模式^[4]; 可以使用户拥有更好的体验, 帮助用户更好地享受音乐、提高生活质量。音乐推荐算法的研究具有重要的意义, 不仅可以提高人们的生活质量, 也可以促进音乐产业的发展^[5], 个性化音乐推荐成为很多学者的

研究课题。推荐系统已经渐渐成为许多软件平台的特色功能, 无论是在在线购物、音乐和视频流媒体、社交媒体方面, 还是在新闻和内容聚合方面, 推荐系统都扮演着至关重要的角色。推荐系统收集用户相关数据, 分析用户的历史行为和偏好, 以及其他相关信息, 为用户提供个性化的推荐内容, 从而提高用户使用的频率, 促进销售和传播。随着互联网在社会各方面的普及, 推荐系统也应用到很多行业, 音乐推荐系统成为其中的一个研究领域。

音乐推荐有其独特性, 音乐表达的内容非常丰富, 内涵深刻, 虽然不像文学作品那样让人一望而知, 但是其独特的个性与魅力能够震撼人们的心灵。通过音乐的独特个性, 可以表达出人类的情绪, 例如, 喜、怒、哀、乐等, 陶冶人们的情操。音乐推荐需要配合音乐的特性, 结合用户的心理, 理解用户的喜好。用户听歌的曲目灵活性强、种类繁多, 导致个性化推荐技术难度加大, 需要根据用户的听歌习惯收集大量数据, 进行个性化推荐^[6]。

3 Apriori 算法

Apriori 算法是关联规则挖掘算法, 属于经典算法, 应用非常广泛。关联规则挖掘是研究数据挖掘时比较常用的方法, 最早由 Agrawal 等提出, 用于购物篮分析问题。使用关联规则挖掘能够发现交易数据库中不同商品

基金项目: 南通理工学院校级项目: 基于 Apriori 算法的学生课程成绩关联性分析 (项目编号: 2022XK(Z)29)。

作者简介: 严海卫 (1990—), 女, 工程师, 硕士, 研究方向为软件开发; 林春花 (1995—), 女, 讲师, 硕士, 研究方向为无线传感器网络; 徐欢潇 (1991—), 女, 讲师, 硕士, 研究方向为网络安全; 陈枢茜 (1992—), 女, 讲师, 硕士, 研究方向为人工智能。

之间的联系规则，从而产生商业价值。这些规则是顾客购买商品的行为模式的体现，使用这些规则可以指导商家合理地安排采购、管理库存及设计货架等。这项研究打开了关联规则研究的大门，诸多学者及研究人员开始对关联规则的挖掘问题进行大量研究，通过研究得出了很多有效结论，用于指导各行业的布局发展。研究人员对关联规则的挖掘理论进行进一步探索，改进原有的算法，不断设计新算法，例如，并行关联规则挖掘、数量关联规则挖掘。在研究过程中，他们坚持不懈，付诸良多，提高了挖掘规则算法的适用性、合理性、有效性等。关联规则数据挖掘是从大量数据中的项集之间发现潜在的联系，假如项集中两项或多项属性之间存在关联，那么可以依据项集中其他属性值来预测其中一项属性。通过关联规则挖掘技术，发现用户音乐数据库中音乐歌曲之间的联系，得出用户选择音乐歌曲的关联模式，例如，收听一首音乐歌曲对收听另一首音乐歌曲的影响。

Apriori算法的运行过程包括连接和剪枝。该算法中项集的概念即为项的集合，项集的频率是包含项集的事务数。数据库中存在多个项集，如果其中某项集满足最小支持度，则称其为频繁项集，依据频繁项集产生的关联规则，得出每条规则的置信度和支持度，采用满足最小置信度阈值的规则。

算法的流程大致如下。

- (1) 扫描数据集，获取所有项的支持度计数（频繁1项集）。
- (2) 根据最小支持度阈值筛选出频繁1项集。
- (3) 根据频繁1项集生成候选2项集，并扫描数据集计算支持度。
- (4) 根据最小支持度阈值筛选出频繁2项集。
- (5) 重复以上步骤，逐步生成频繁 k 项集，直到无法生成更多频繁项集为止。以下为根据频繁 $k-1$ 项集获取频繁 k 项集的部分代码。

```
map< set<string>, int > Apriori::getLk(int k, set<set<string> > Lk0) {
    map< set<string>,int> Lk,Ck;
    set< set<string> > CkSet = aprioriGen(k-1, Lk0);
    //Ck的set-string，再集成set
    for(set< set<string> >::iterator i=CkSet.begin(); i!=CkSet.end(); i++){ //每一个set<string>
        string head = *(*i).begin();
        set<int> temp = mp[head];
        for(set<string>::iterator j = ++(*i).begin(); j!=(*i).end(); j++){
```

```
set<int> t = mp[*j]; //是 set<string>里的每一个string，找到对应的int
temp=intersec(temp,t);
}
if(temp.size()>=minSup)
    Lk[*j]=(int)temp.size();
}
return Lk;
}
```

根据频繁项集生成的关联规则，得出置信度，筛选出满足音乐推荐场景的最小置信度阈值的规则，通过不断迭代生成频繁项集。Apriori算法可以高效地挖掘出数据集中的频繁项集以及关联规则，Apriori算法步骤如图1所示。

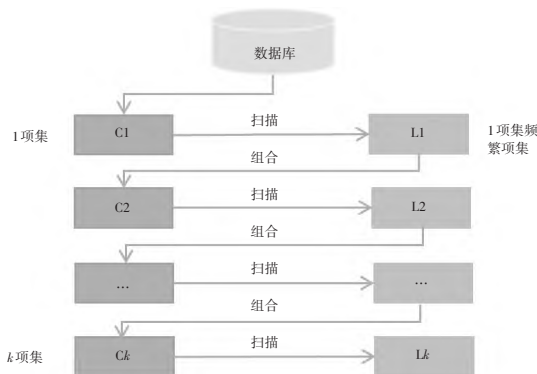


图1 Apriori算法步骤

假设音乐1、音乐2和音乐3为3个不同的音乐数据项，数据集 D 是由这3个数据项所组成的集合，记为 $D=\{\text{音乐1,音乐2,音乐3}\}$ ，存在如表1所示的9位用户的听歌记录，0表示没听过，1表示听过。

表1 音乐数据

| | 音乐1 | 音乐2 | 音乐3 |
|-----|-----|-----|-----|
| 用户1 | 0 | 1 | 0 |
| 用户2 | 0 | 1 | 0 |
| 用户3 | 1 | 1 | 1 |
| 用户4 | 1 | 1 | 0 |
| 用户5 | 0 | 1 | 1 |
| 用户6 | 1 | 1 | 0 |
| 用户7 | 1 | 1 | 0 |
| 用户8 | 1 | 1 | 1 |
| 用户9 | 1 | 0 | 1 |

在一个数据集中，音乐1和音乐2同时出现的概率称为支持度，为音乐1和音乐2同时在数据集中出现的次数除以数据的总个数，其计算公式如公式（1）所示：

$$\text{Support}(\text{音乐1} \Rightarrow \text{音乐2}) = \frac{P(\text{音乐1} \cap \text{音乐2})}{P(\text{all})} = \frac{5}{9} \quad (1)$$

在一个数据集中，音乐1出现的同时音乐2也出现

的概率称为置信度，为音乐1和音乐2同时出现的次数除以音乐1出现的次数，其计算公式如公式（2）所示：

$$\text{Confidence}(\text{音乐1} \Rightarrow \text{音乐2}) = \frac{P(\text{音乐1} \cap \text{音乐2})}{P(\text{音乐1})} = \frac{5}{6} \quad (2)$$

4 音乐推荐分析

4.1 SPSS Modeler 工具

SPSS Modeler是由IBM公司开发的，用于数据挖掘和预测建模。它具备领先的可视化数据科学和机器学习（ML）解决方案，旨在通过加快数据科学家的操作任务进度来缩短实现价值的时间。世界各地的组织都将它用于数据准备和发现、预测分析、模型管理和部署，以及ML，最终实现数据资产变现。SPSS Modeler能够将数据自动转换为最佳格式，从而实现最准确的预测建模，具备强大的图形引擎，让洞察分析变得生动形象，利用智能图表推荐器能够从数十个选项找出数据的理想图表，并采用可视化呈现方式来展示分析结果。SPSS Modeler提供了成熟的Apriori模型，可应用于音乐推荐。

在SPSS Modeler中使用Apriori算法进行关联分析的步骤如下。导入数据集，将需要进行关联分析的数据集导入SPSS Modeler中。配置节点，在SPSS Modeler的流程图中，找到关联分析节点（Apriori算法节点），将其添加到流程图中。设置参数，对于Apriori算法节点，可以设置支持度和置信度的阈值。一个音乐集出现在音乐数据集中的频率称为支持度，关联规则的可靠性则为置信度。根据具体的需求，可以调整这些参数。执行分析，配置好参数后，运行流程图，SPSS Modeler将会执行Apriori算法进行关联分析。最后进行结果解释，分析完成后，SPSS Modeler会生成关联规则和关联度等结果，可以根据需要进行结果解释和进一步分析。

4.2 SPSS Modeler 建模

准备若干条用户使用音乐软件听歌的数据^[7]，使用SPSS Modeler软件进行音乐推荐的建模，部分实验数据如表2所示。

表2 实验数据

| 用户 | 音乐1 | 音乐2 | 音乐3 | 音乐4 | 音乐5 | 音乐6 | 音乐7 | 音乐8 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 用户1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| 用户2 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| 用户3 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| 用户4 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |
| 用户5 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 |

在SPSS Modeler软件的源选项栏中添加Excel源，导入Excel数据。数据流运行成功，生成的数据表或模

型结果会显示在流管理器的相应选项卡中，否则，SPSS Modeler会给出数据流运行错误的提示信息，用户需对节点参数进行重新调整后再运行数据流。在数据流运行过程中，若发现问题，则可强行中断数据流的运行。数据读取完成后进行字段过滤，对异常值进行处理，设置字段类型，将字段“用户”的角色设置为记录标识，将剩余的字段的角色设置为任意。在建模选项中选择Apriori模型，建模流程如图2所示。

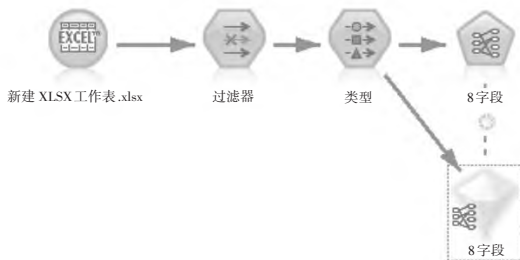


图2 Apriori算法建模流程

经过多次建模和调整，最佳的支持度为30%、最佳置信度为70%，模型设定值如图3所示。建模完成后运行得出实验结果，实验结果如图4所示。



图3 模型设定值

| 后项 | 前项 | 支持度百分比 | 置信度百分比 |
|-----|-----|--------|--------|
| 音乐1 | 音乐4 | 30.612 | 80.0 |
| 音乐5 | 音乐8 | 38.776 | 78.947 |
| 音乐1 | 音乐4 | 42.857 | 76.19 |
| 音乐5 | 音乐4 | 32.653 | 75.0 |
| 音乐5 | 音乐6 | 32.653 | 75.0 |
| 音乐5 | 音乐3 | 51.02 | 72.0 |
| 音乐5 | 音乐6 | 42.857 | 71.429 |

图4 实验结果

5 结语

通过分析实验结果,得出下列结论。

(1) 音乐1与音乐4、音乐5存在较高的关联度,如果有用户喜欢音乐1,则可以推荐音乐4和音乐5给用户。

(2) 与音乐5关联度高的音乐比较多,包括音乐1、音乐3、音乐4、音乐6和音乐8,通过实验结果分析可以得出,音乐5受大众欢迎度很高,音乐平台可以推荐音乐5。

(3) 通过Apriori建模没有挖掘出音乐2与音乐7的关联规则,因此音乐2与音乐7无法关联到其他音乐,无法进行相关的音乐推荐。

音乐推荐是一个复杂而有趣的研究课题,也是一个具有广泛应用价值和社会影响力的技术领域。然而,音乐推荐系统也面临着一些挑战和问题,需要不断地进行研究和改进。研究数据为用户播放音乐数据,无法代表用户心理状态,因此研究有局限性,需要结合用户心理情景做进一步研究和改进。由于用户的偏好和行为是随着时间、情境、心理等因素而变化的,而Apriori算法只

利用了历史数据,忽略了用户的实时状态,因此会影响Apriori算法的预测准确性和覆盖率,以及推荐系统的适应性和灵活性。针对以上不足需进一步展开研究,提出一些改进和扩展的方法,实现更理想的音乐推荐效果。

参考文献

- [1] 丁明波. 基于深度学习的个性化音乐推荐方法研究[D]. 青岛: 中国石油大学硕士学位论文, 2017.
- [2] 郭静菡. 应用于音乐节目分类的Apriori挖掘算法设计[J]. 现代电子技术, 2019, 42(10): 90-94.
- [3] 李臻. 面向用户偏好的个性化音乐智能推荐系统[J]. 计算机与通信技术, 2023, 42(19): 118-112.
- [4] 余莉娟. 基于深度学习的个性化音乐推荐算法研究[J]. 微型电脑应用, 2020, 36(10): 140-143.
- [5] 华泽, 叶雨航. 协同过滤算法及在个性化音乐推荐中的应用[J]. 现代计算机, 2021(22): 43-46.
- [6] 姚谊. 基于社会化标签的音乐推荐技术研究[D]. 杭州: 杭州电子科技大学, 2016.
- [7] 裴玫. 国内音乐数据库研究特征的现状与内容分析[J]. 河南图书馆学刊, 2023, 43(1): 90-93.

(上接第22页)

据。对于劳动之后的评价考核管理,主要依据劳动过程中学生或者教师上传的材料进行,通过线上评分并对接教务系统,将成绩匹配到学生最终成绩单中。

5 结语

劳动实践课预约系统采用低代码平台进行开发,结合微服务架构的办事大厅快速上线,实现了劳动实践课开展过程中劳动实践项目的维护、劳动实践项目预约,以及预约结果查询等相关功能,为课程管理提供了有力的信息化能力支撑。系统无论是在PC端还是在移动端,都有良好的适配性,并且具有可扩展性。课程的开展及系统的开发对积极探索具有中国特色的劳动教育模式,创新体制机制,促进学生形成正确的世界观、人生观、价值观有良好的促进作用,对创新和改革高校教学管理具有十分重要的现实意义。

参考文献

- [1] 胡晓. 新时代劳动教育的现实转变与课程建设[J]. 湖南科技学院学报, 2022, 43(4): 77-79.
- [2] 张尧雨. 低代码在数据平台开发中的运用[J]. 软件和集成电路, 2024(Z1): 2-5.
- [3] 苏伟, 国建勋, 冯宽. 低代码开发平台发展现状及标准化研究[J]. 信息技术与标准化, 2024(Z1): 17-21.

- [4] MARCUS W. 无/低代码软件开发的兴起——无需经验[J]. Engineering, 2020, 6(9): 960-964.
- [5] FRANK B. Mendix 低代码平台助力安全设计[J]. 软件和集成电路, 2022(1): 70-71.
- [6] 高健, 丁之林. 中国式高校劳动教育新道路探究[J]. 黑龙江教师发展学院学报, 2024, 43(4): 1-5.
- [7] 黄河, 弓晓波, 王晓红. 基于低代码的综合管理平台设计与开发[J]. 电脑知识与技术, 2024, 20(4): 49-52.
- [8] 邱奕超. 低代码平台在高职院校流程引擎优化中的应用与实践——以J校为例[J]. 信息系统工程, 2024(4): 169-172.
- [9] 罗栩. 基于低代码开发平台的财务人员数智化转型案例研究[J]. 企业改革与管理, 2024(6): 120-122.