

基于情景建模的移动互联网 音乐推荐系统研究

薛 婧

(西安邮电大学, 西安 710121)

摘要 近年在当下移动互联网中,数字音乐得到了大力研发,而基于大数据算法的数字音乐个性化推荐正在蓬勃发展,对于这类研究需要借助计算机技术、多媒体技术来补充。在探索过程中,线上音乐的传统二元推荐算法正逐渐被情景推荐所取代,通过构建情景模型的音乐推荐算法有利于用户获取音乐,可以增长互联网音乐的有效投递,通过传统音乐推荐算法与改进的情景模型的推荐算法对比,有效验证了算法的有效性,克服了数据稀疏性的问题。结合交互的音乐方式,使该模型根据实时的场景对用户信息推送,对于移动互联网数字音乐的研究具有一定的实际价值和积极意义。

关键词 大数据算法; 移动互联网; 兴趣分类; 个性化推荐; 消息推送

中图分类号 TP399 文献标识码 A 文章编号 1009-2552(2021)01-0018-05

DOI 10.13274/j.cnki.hdzj.2021.01.004

Mobile internet music recommendation system based on scenario modeling

XUE Jing

(Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

Abstract: In recent years, in the current mobile Internet, digital music has been vigorously developed, and the personalized recommendation of digital music based on big data algorithm is booming, for this kind of research needs to be supplemented by computer technology and multimedia technology. In the process of exploration, online music is gradually replaced by the traditional binary recommendation algorithm by the scenario recommendation algorithm. The music recommendation algorithm based on the scenario model is conducive to users' access to music and can increase the effective delivery of Internet music. Through the comparison between the traditional music recommendation algorithm and the improved scenario model recommendation algorithm, the effectiveness of the algorithm is effectively verified and overcome Data sparsity. Combined with the interactive music mode, the model can push the user information message according to the real-time scene, which has certain practical value and positive significance for the research of mobile Internet digital music.

Key words: big data algorithm; mobile Internet; interest classification; personalized recommendation; message push

0 引言

随着“互联网+”的蓬勃发展,互联网为各种

基金项目:西安市2019年度社会科学规划基金项目(19S96)

作者简介:薛婧(1982-),女,硕士研究生,讲师,研究方向为音乐教学、音乐软件研究。

行业带来了全新的模式^[1]。数字音乐的发展依赖于大数据的使用,海量的信息采用一般的方式处理已经变得非常棘手。同时由于网络资源的分享越来越方便,导致了信息爆炸式增长,信息资源的泛滥,对用户造成了一定的困扰,使用户无法得到

有效的数据。为了减轻用户的信息过载问题^[2],信息分类、搜索引擎、推荐系统相继而生。音乐已经变成人们生活中必不可少的部分。基于大数据算法的用户音乐推荐,即可以满足用户的需求和对音乐的偏好,也可以减少信息冗余的问题。在数字音乐发展领域,推荐音乐系统已经逐渐由传统的“用户-项目”二元关系发展为情景感知的关联推荐算法^[3]。并且尝试避免以牺牲超高时间复杂度的代价获得较好的推荐效果,而是以针对多维数据稀疏性的隐式转换处理^[4],基于时序概率矩阵分解的情景感知算法,保证用户的行为选择正确性。

1 相关技术概述

1.1 数据挖掘

大数据时代,海量的用户数据不容易处理,无法形成对用户行为有效的管理,而诞生的数据挖掘技术可以对这些数据提取出有效价值,便于企业制定合理有效的营销模式。在分析的过程中,会使用多种计算方法,数据挖掘的计算方法通常有模式识别、机器学习、统计学等方法^[5],并且由于新兴的神经网络方法的优越性,大量企业开始技术的投入。互联网企业率先使用数据挖掘技术,有效地制定推广策略,为企业的广告精准投放和用户精准推荐起到了关键性的作用,也节省了大量的成本。数据挖掘的具体实现过程如图1所示。在数据挖掘的处理过程中,获取海量的数据集是首要的一步,由于不完整的数据集合不符合格式的数据集,需要对数据进行清洗和分类处理。

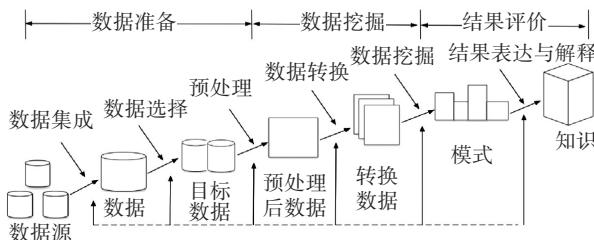


图1 数据挖掘的过程

1.2 聚类算法

对于获取到的数据集合进行分类,可以采用无监督的学习方式。这种方式能将大量数据划分成不同的簇^[6],在同一个集合中,数据之间具有良好的相似性。数据间的距离与数据相似性成负相

关的关系。进行聚类处理,首先假设 $X = \{x_1, x_2, \dots, x_n\}$ 是 d 维空间中的一组对象,其中, $x_i = \{x_{i1}, x_{i2}, \dots, x_{id}\}^T$, x_i 与 x_j 是两个数据对象。 $d(x_i, x_j)$ 是 x_i 与 x_j 数据间的距离。聚类分析具有几个典型的计算距离的函数。曼哈顿距离计算方法是常用的计算数据对象的方法。方差加权距离计算方法、欧式距离计算方法也是十分常见的^[7],在这里,通常曼哈顿的距离表示两个数据对象的非直线距离如式(1)所示,方差加权距离如式(2)所示,欧式距离如式(3)所示。

$$d(x_i, x_j) = \sum_{k=1}^m |x_{ik} - x_{jk}| \quad (1)$$

$$d(x_i, x_j) = \left[\sum_{k=1}^m \frac{(x_{ik} - x_{jk})^2}{s_k^2} \right]^{1/2} \quad (2)$$

$$d(x_i, x_j) = \left(\sum_{k=1}^m |x_{ik} - x_{jk}|^2 \right)^{1/2} \quad (3)$$

聚类分析是计算机技术和人工智能技术等结合的交叉学科,数据挖掘的基本思想是如图2所示,聚类可以分成基于划分的聚类、基于层次的聚类、基于密度的聚类、基于图论的聚类、基于网格的聚类和基于模型的聚类^[8]。对于不同的数据,可以采用不同的概率模型。

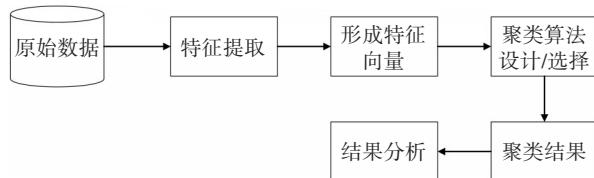


图2 传统聚类分析的过程

在移动互联网情境下,数字音乐的推荐系统主要由收集用户信息行为的记录模块、分析用户喜好的模型分析模块以及音乐推送模块组成。而这里最重要的就是音乐推荐算法。基于大数据算法,采用情景信息进行分析。通过收集用户当时的周围情景,信息收集的方式有所差异。按照通常的分类方法,包含显式获取、隐式获取以及推理获取三种信息获取方式^[9]。在基于情景的用户偏好提取,主要分为定量分析和定性分析技术。

2 基于情景过滤的移动互联网音乐研究

2.1 行为数据分析与关系模型

在基于情景过滤的数字音乐发展中,主要是利用改进的时序用户关系模型。由于音乐具有时

代性和流行性。理想条件的时序图如图3所示,该时序图具有很强的承接性。用户A与用户B对于音乐的选择在每个时间刻度具有极强的关联性,称之为时序相关。在预测用户B在 t_5 时刻的音乐选择,可以在用户A在 t_4 时刻的选择做出适当的推荐。

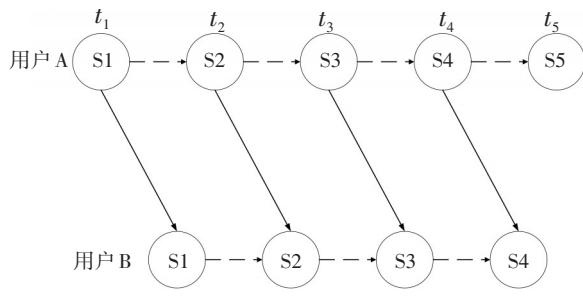


图3 理想条件下的时序推荐图

2.2 情景再现推荐

情景再现推荐算法有其独特性。情景过滤重耍保留核心的内容,去掉无关紧要的内容。而情景再现则是增加或者复原出真实情况。情景再现的基本思想是通过情景匹配的方式,分析用户的音乐历史行为,在此分析过程中,还寻找对用户影响最大的相邻用户,基于用户评分时间的先后顺序,构建基于时序的用户关系图。这类评分先后顺序的关系模型将行为影响者与音乐行为的接受者关联在一起^[10],情景再现推荐包含了情景过滤的过程,并且进行了适当的细化。情景过滤过程由情景预过滤和情景后过滤两部分组成^[11]。情景预过滤做法具有很好清洗数据的效果,过滤掉与当前情景不相关的数据集,构建合理的评分集,预测用户的偏好。在实现算法的同时,为了满足音乐用户的个性化需求,获取数据的方式采用程序手段得到,通过python语句获取结果,并整理最终数据,其原理图如图4所示。

利用爬虫进行用户情景时序数据的收集,以虾米音乐作为爬虫对象,抓取17785个音乐用户的9245429条最近收听的音乐时序集。对于在数据稀疏性上的显式与隐式的用户反馈对比,通过python抓取音乐软件中的4358首音乐。收集这些音乐的82194条用户评论作为研究数据,上万条的数据量符合大数据处理的所要求的数量级别。根据不同的用户ID,将其分析出对应评论的用户数为79024个。不考虑无任何评论的用户,

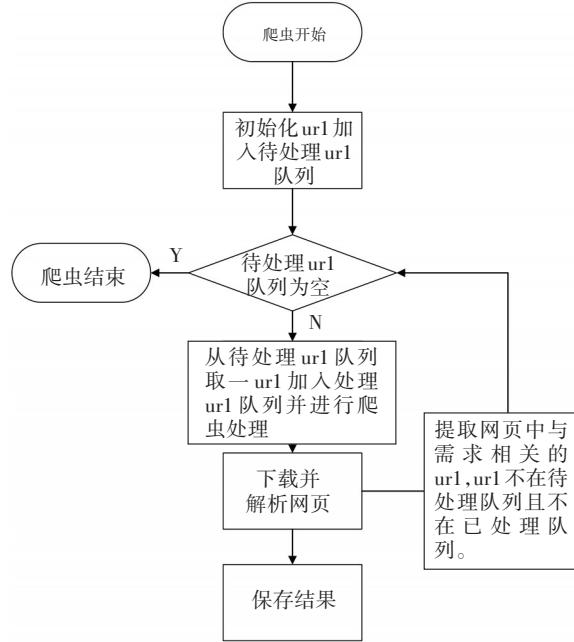


图4 爬虫原理图

表1 隐式反馈与显式反馈的数据对比

类别	用户数	用户反馈数	用户平均反馈数
显式用户评论	79024	82194	1.0401
隐式用户评论	17785	9245429	519.87

其对比差异性如表1所示,从而看出隐式用户反馈更具有有效性。

为了获取用户听音乐时的天气状况,需要获取当时的信息,主要根据用户主页的地址信息,再按照关键词匹配,根据城市数据库中的资料,获得用户城市名字。然后再进一步分析全国主要城市、县的历史天气记录,然后基于Jsoup技术通过Jquery操作方法提取网页资源,解析Html文本内容。

为了合理评价推荐算法的准确度,采用RMSE将预测评价与真实评价的偏差作为指标。其表示方法如式(4)所示,在测试中, RMSE越小, 算法准确度越好, TestSet 表示测试集数据。

$$RMSE = \sqrt{\frac{\sum_{(u,i) \in TestSet} (\hat{r}_{ui} - r_{ui})^2}{|TestSet|}} \quad (4)$$

2.3 实验设计与分析

为了验证时序最近邻算法的优越性,通过采用对比实验的方法,将传统最近邻推荐与时序最近邻推荐的算法准确度进行差异比较,依据python语句爬取的互联网音乐数据,以 k 表示最近邻数, k 的取值从10开始,以10位差值的等差数

列。其对比如表2所示。

表2 不同最近邻下的RMSE对比表

类别	$k=10$	$k=20$	$k=30$	$k=40$	$k=50$	$k=60$
传统最近邻推荐算法	0.7857	0.7785	0.7748	0.7691	0.7723	0.7759
时序最近邻推荐算法	0.7839	0.7731	0.7663	0.7567	0.7592	0.7665

如图5所示,可以看出随着 k 值从10增长至40,在这个范围内,最近邻数的增加导致两种算法RMSE数值呈现递减的状态。为了更好地比较何种算法具有优越性,可以由函数单调性的角度,看出时序最近邻推荐的单调递减性更好。当取值 $k=40$ 时,二者对应的RMSE下降效果最明显,两者算法都取得了最小值。最小值点的位置寿命,两种算法已达到最佳的音乐推荐效果。在最邻近数大于40的范围内,两种算法的RMSE取值逐渐增加。在最近邻数的整体参考范围内,图5中的曲线图可以清晰地说明,RMSE的取值无论是上升还是下降的状态,都可以得出结论:时序最近邻推荐效果始终要比传统最近邻推荐效果更好。对于传统推荐算法来说,时序最近邻推荐更佳。从而证明了时序用户关系模型的有效性与准确度。

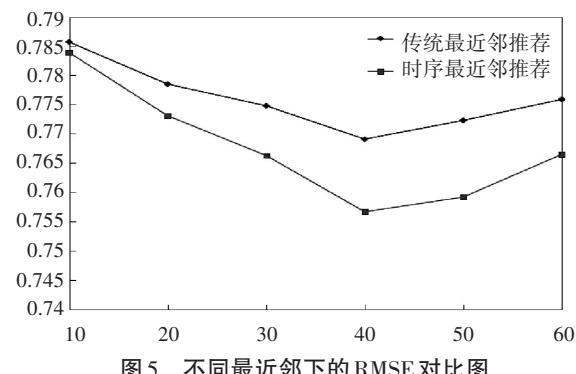


图5 不同最近邻下的RMSE对比图

3 基于情景建模的移动互联网音乐推荐

3.1 概率矩阵分解原理

概率矩阵分解以数据概率为研究对象,具有较好的理论基础^[12]。概率矩阵分解原理是利用模型的训练将“用户-项目”矩阵进行分解,分解原理的依据相对复杂一些,主要是依赖于潜在特征矩阵的概率分布特性,利用项目与用户矩阵评分项目的乘积预测推荐概率,同时参考不同场景下的音乐记录的差异,映射到各个场景的对应评分下^[13]。

3.2 基于时序概率矩阵分解的实验设计与分析

选择合理的音乐用户数据集,考虑涉及的情景因素,将数据集依据用户的类别进行划分。从0到1000。用基于时间概率矩阵分解的场景感知推荐与场景再现推荐的RMSE比较。

从图6显示实验结果,显示出在极为严重的数据稀疏范围内,两种算法表现的差异尤为明显。基于时序概率矩阵分解的情景感知推荐具有更好的数据稀疏性处理方式。数据稀疏性相对严重的场景再现推荐预滤波方法的准确度有所减少,不利于算法的实现效果。基于时序概率矩阵分解的情景感知推荐与情景再现推荐的RMSE的使用显示出相同的下降趋势。其对比的差异范围逐渐缩小,逐渐趋于一致。综上所述的比较,可以分析出基于时间序列概率矩阵分解的态势感知明显优于场景再现的推荐效果,更加适合在大数据情境下使用。

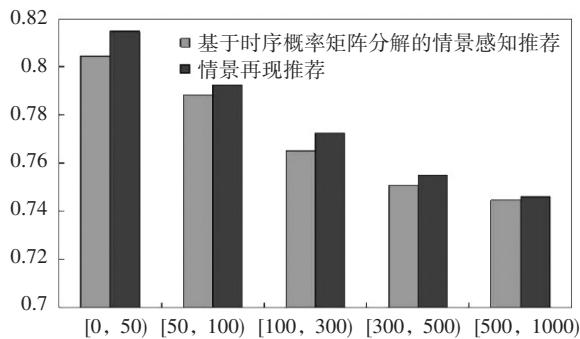


图6 基于时序概率矩阵分解的情景感知推荐与情景再现推荐的RMSE对比

4 结束语

基于大数据算法的移动互联网数字音乐为音乐爱好者提供了更好的音乐推荐技术支持。基于大数据算法的发展,通过分析数据挖掘理论和国内外研究现状,阐述了相比于传统音乐推荐算法的最邻近推荐算法自身具备的优势。并且设计情景再现的算法模型,尤其有针对性地采用基于时序概率矩阵分解算法,该算法的准确度提高,数据稀疏性较好,适用于互联网数字音乐的推广。基于理论模型的计算,采用基于时序概率矩阵分解的训练模型和构建方式,模拟出音乐用户的习惯特征,从而更好地实现智能推荐的效果,让音乐软件的用户在互联网中进行沉浸式欣赏音乐,通过对设计的比较,能够更好地对软件设计进行改

进。在后续的研究中,应加入更人性化的界面设计,采用更大规模模型样本来实现人工智能,使之准确度更加具有普适性,保证在操作度方面具备良好的稳定性,也可以将多种人工智能算法融合用来改进算法,获得更通用的操作方法,这是下一步研究的重点方向。

参考文献:

- [1] 石路路, Alexis Huet, 吴冬华. 一种基于众包式的大数据算法分析移动APP对LTE无线网络资源的消耗与利用[J]. 江苏通信, 2016, 32(3):15-18.
- [2] 于森, 庄洪杰. 基于移动医疗大数据平台下深度最优匹配算法的机会网络转发机制[J]. 网络安全技术与应用, 2016(9):52,54.
- [3] 闫娜. 基于大数据吞吐效益评估的网络数据综合调控算法研究[J]. 计算机与数字工程, 2016, 44(7):1304-1308.
- [4] 唐德权, 黄金贵, 史伟奇. 基于大数据平台的动态车辆路径调度算法[J]. 计算机工程, 2018, 44(1):74-78.
- [5] 王晖楠, 魏娇. 基于人工智能识别的音乐片段指纹检索技术研究[J]. 自动化与仪器仪表, 2019(5):119-122.
- [6] 罗莉. 基于改进BigFIM算法的网络信息大数据高频数据项挖掘算法研究[J]. 激光杂志, 2016, 37(7):135-140.
- [7] 李颜汐. 基于情感特征的背景音乐分类方法[J]. 现代电子技术, 2017, 40(15):115-118.
- [8] 邓永莉, 吕愿愿, 刘明亮, 等. 基于中高层特征的音乐情感识别模型[J]. 计算机工程与设计, 2017, 38(4):1029-1034.
- [9] 常远, 杜慧婕, 于锐宸, 等. 音乐大数据的应用研究[J]. 明日风尚, 2018(21):110-111.
- [10] 刘定一, 应毅. 基于大数据的网络舆情预测分析[J]. 现代信息科技, 2019, 3(12):100-101, 103.
- [11] 陈建峡, 朱季骐, 王鹰适, 等. 一种基于SVM的博客大数据分类算法及应用[J]. 湖北工业大学学报, 2016, 31(4):70-74.
- [12] 任华, 张玲, 叶煜. 数字化校园中用户网络行为大数据的分析与监控[J]. 计算机与数字工程, 2017, 45(9):1814-1818, 1823.
- [13] 孟虎, 梁晓蓓, 杨以雄, 等. 大数据背景下基于LMBP算法的供应链绩效评价与优化[J]. 数据分析与知识发现, 2018, 2(11):37-45.

(责任编辑:丁明)

(上接第17页)出了更高的要求^[10-11]。海洋测绘是为智慧海洋、智慧海洋中心城市建设提供基础地理信息底层数据,还是海洋资源开发、海洋生态保护、海岸线保护与管理、海域管理必不可少的、最基础的工作。为此,简要地对我国海洋测绘市场、海洋测绘技术进行综述。

参考文献:

- [1] 周兴华. 海洋测绘产业发展报告,中国地理信息产业发展报告(2020)[R]. 测绘出版社, 2020.
- [2] 海洋测绘市场现状与发展趋势[EB/OL]. (2017-10-25 [2010-11]. <http://www.tiuchina.com/yjbg/169.html>.
- [3] 龚强. 应用前沿科技融合测绘地理信息技术助力国土空间规划编制与实施[J]. 黑龙江自然资源, 2020(3).
- [4] 杨元喜, 徐天河, 薛树强, 等. 我国海洋大地测量基准与海洋导航技术研究进展与展望[J]. 测绘学报, 2017, 46(1):1-8.
- [5] 吴自银. 高分辨率海底地形地貌[M]. 北京:科学出版社, 2017.
- [6] 赵建虎, 陆振波, 王爱学. 海洋测绘技术发展现状[J]. 测绘地理信息, 2017, 42(6).
- [7] 暴景阳, 瞿国君, 许军. 海洋垂直基准及转换的技术途径分析[J]. 武汉大学学报:信息科学版, 2016, 41(1):52-57.
- [8] 张全德, 范京生. 我国卫星导航定位技术应用及发展[J]. 导航定位学报, 2016, 4(3):82-88.
- [9] 阳凡林, 康忠志, 独知行, 等. 海洋导航定位技术及其应用与展望[J]. 海洋测绘, 2006, 26(1):71-74.
- [10] 滕惠忠, 辛宪会, 李军, 等. 卫星遥感水深反演技术的发展与应用[C]. 第二届高分辨率对地观测学术年会. 北京, 2013.
- [11] 龚强, 李德江, 龚天慧, 等. 论构建地理信息共享平台的计算思维[J]. 测绘工程, 2018, 27(1):1-4.

(责任编辑:丁一元)