

doi:10.3772/j.issn.2095-915x.2015.06.016

# 音乐推荐系统性能调查探究

王东雨

(河北大学管理学院 保定 071002)

**摘要:** 随着国内移动音频行业的快速发展, 各类音频客户端推荐系统存在的问题也在不断显现。利用分析工具 R, 对国内某音乐客户端行业报告中数据进行分析, 探讨其推荐系统存在的问题、分析原因, 最后给出解决这些问题的若干措施, 以期国内移动音频行业能够重视其推荐系统的构建、优化和维护, 提高用户忠诚度和满意度, 更好的推动其业务的发展。

**关键词:** 移动音乐, 音乐推荐系统, 个案分析

**中图分类号:** G253,TP391

## Research on the Performance of Music Recommendation System

Wang Dongyu

(School of management, Hebei University, Baoding 071002, China)

**Abstract:** Along with the fast development of mobile audio industry, some problems of recommended system in audio clients appeared constantly. By means of the analytical tool R, this paper analyses the report data, which are derived from the domestic audio clients. Firstly, the paper found some existing problems of the recommended systems. Secondly, it analyses the forming reasons in different aspects. Finally, the paper proves into the countermeasures of

**作者简介:** 王东雨, (1991-), 男, 山东临沂, 河北大学, 硕士研究生。研究方向: 个性化推荐、智慧医疗, 389011848@qq.com。

above-mentioned problems. The purposes of the paper are to urge the domestic mobile audio industry to pay more attention to the establishing, optimizing and maintaining of the recommended systems, and to increase loyalty and satisfaction of the customers, so that the industrial development will be greatly upgraded.

**Key words:** Mobile music, music recommendation system, case analysis

## 1 引言

据国内首个《移动音频媒体价值白皮书》<sup>[1]</sup>显示:2015年上半年,我国移动音频行业的用户规模已达3.5亿,网民渗透率高达58.3%,与移动视频相比,其拥有的用户粘性更强、满意度更高。随着网络覆盖率的提高、资费的下降和智能手机的普及,使得民众接触互联网的时长大幅增加,智能手机的使用时长也在逐年超过PC、平板等设备,这些条件为移动音频行业的爆发提供了机遇。

随着移动音频行业的飞速发展,音频用户群体的规模也在不断壮大,海量的音频数据和用户对移动音频行业带来了巨大的挑战,如果推荐系统能够有效地进行音频数据挖掘、分析,不仅可以准确预测用户偏好,发现潜在用户,为用户提供喜欢的音频,提高用户体验,增加用户忠诚度;而且还可以科学的评价公司的决策,如广告投放业务的效果,预测决策产生的结果,指导公司的重大战略决策等<sup>[2]</sup>。

## 2 国内音乐推荐系统的现状

相比国外,国内的音乐推荐技术整体发展落后于欧美国家,大多数音乐网站采用的技术不够成熟、全面。随着大数据和云计算以及推荐算法的研究深入,国内各行业涌现出研发适合自身推荐系统的浪潮,就是在这种环境下,涌现出了诸多优秀的音乐推荐网站,有些甚至可以和国外的推荐系统相匹敌,比如网易云音乐、虾米网和豆

瓣网等,他们以自身拥有的用户数据,算法为核心竞争力,满足无目的听众的需求。目前,国内音乐推荐的主要应用的方法有基于内容的推荐、协同过滤推荐、基于关联规则的推荐、基于知识的推荐以及组合推荐,但是不管怎样,国内的推荐系统还处于快速发展时期,推荐系统都或多或少的面临冷启动、稀疏性、覆盖率和长尾等诸多问题,如何解决这些问题,成为今后国内推荐系统研究关注的焦点<sup>[3]</sup>。

## 3 音乐推荐系统实例调查探究

笔者以国内某音乐客户端的行业季度报告<sup>[4]</sup>和品牌竞争分析报告<sup>[5]</sup>中公开数据为研究对象,数据包括:用户ID,歌曲ID,歌曲下载时间,歌曲名,演唱者等信息,统计数字见表1,利用分析工具R,对其进行数据分析,以期发现其音乐推荐系统存在的问题,分析造成这些问题的原因,并给出完善的意见。

表1 数据统计表<sup>[4,5,6]</sup>

用户数量	歌曲数量	记录数
284, 393	38, 256	741, 493

### 3.1 长尾现象

从数据分析中,笔者发现用户的行为特征分为:down、listen,据图1user和song分析结果可以明显的看到“长尾现象”:即数据集中绝大部分用户只试听或下载系统中的少部分歌曲,而大部分歌曲仅代表小部分用户的个性化需求。

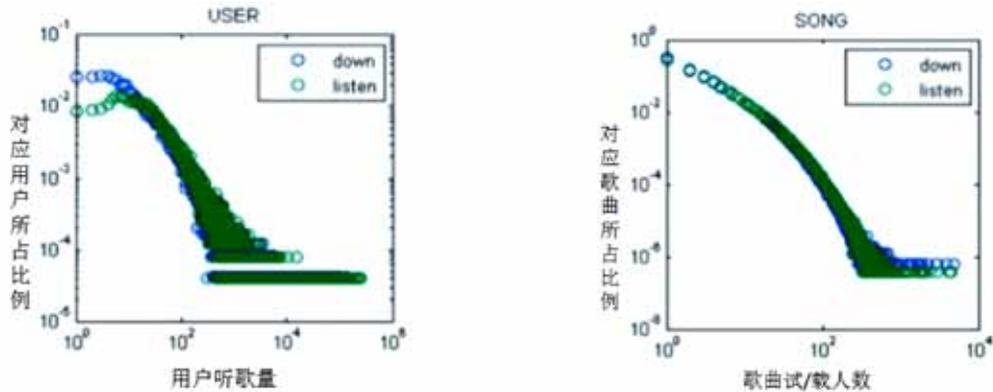


图 1：长尾现象

图 1 都呈双对数曲线分布，且所有数据在双对数曲线上都呈近似直线的形状，从而证明不管是歌曲试 / 载率还是用户的活跃度，都近似于长尾分布。

原因：用户大部分时候没有特别明确的需求，因为如果用户有着明确的需求，可以直接通过搜索引擎找到感兴趣的歌曲；歌曲数量巨大、关键词设置过多造成的信息过载严重和数据孤岛问题，都会使用户很难找到适合自身兴趣的歌曲；而大多数用户的直接选择 TopN 推荐<sup>[7]</sup>，系统会自动增大热门歌曲和非热门歌曲的流行度差距，让热门的歌曲更加热门，不热门的歌曲更加不热门，增加系统的马太效应，强化了推荐系统的长尾现象。

### 3.2 覆盖率

覆盖率为推荐系统能够推荐出来的物品占总物品集合的比例，描述一个推荐系统对物品长尾的发掘能力<sup>[8]</sup>。

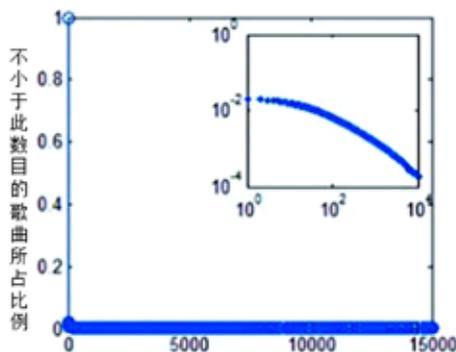


图 2：数据集歌曲覆盖率

为了更细致地描述数据集中歌曲的覆盖率，笔者分析了推荐列表中不同歌曲出现次数的分布，数据中仅有部分歌曲出现在推荐列表中，且相同歌曲在类似用户推荐列表中出现的次数也有些差异，在图中较陡峭的分布表述了歌曲覆盖率的情况，数据表明数据集中歌曲覆盖率较低，歌曲在推荐列表中出现次数的分布情况也表明推荐系统挖掘长尾的能力不足。

原因：冷启动和马太效应。用户冷启动：当新用户到来时，系统中没有用户的行为数据，推荐系统无法根据其历史行为预测其兴趣，而用户也无法通过有效方式搜索到喜欢的歌曲；歌曲冷启动：新曲入库之后，需要时间进行新曲的分析、归类，与用户偏好进行匹配，推荐系统短时间内无法知晓如何将歌曲推送给可能对其感兴趣的用戶；马太效应增加了 TopN 热门歌曲的流行度，使得热门歌曲拥有更多的曝光率，周而复始，使得推荐系统覆盖率降低。

### 3.3 时间分布和歌曲属性

在数据分析中，发现公开数据集中用户听歌时间是间断性分布，即在不同的时间用户活跃程度不同（图 3）；且不同活跃性的用户所听歌曲也不同，笔者从专辑、歌手及推荐结果的新颖性三方面<sup>[9]</sup>对不同活跃性的用户所听歌曲进行分析（图 4）。

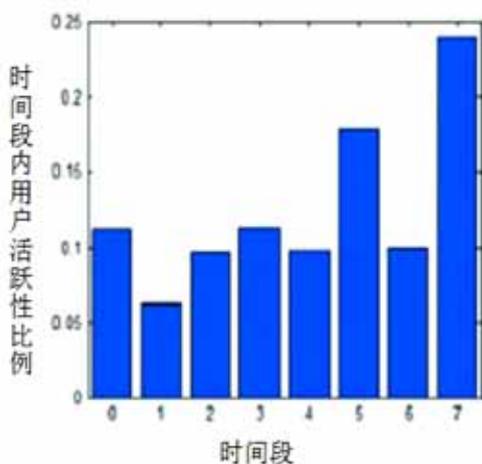
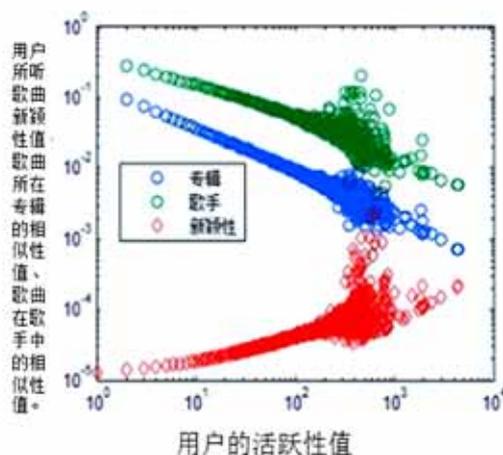


图3：听歌时段分布图



4：歌曲因素分析

图3中，将24小时划分为8段，时间段内的动态，代表用户听歌期间的活跃性，且不同时间段用户活跃性不同。从图中用户听歌的时间段可以看出，主要以休息碎片时间为主，另据《移动音频媒体价值白皮书》<sup>[1]</sup>显示，用户使用移动音频高峰期出现在每天的7点、12点、18点、22点，与笔者的数据分析结果有些不同，一方面是笔者收集样本数据相对于整个移动音频行业来说，数据量过于微小，另一方面笔者选取的样本数据仅来自行业报告中模糊的用户音乐数据，而《移动音频媒体价值白皮书》<sup>[1]</sup>还包括有声小说、综艺娱乐、新闻资讯、电影、养生保健等移动音频。

原因：用户活跃程度存在明显群化现象。活跃性低的用户，可能所听歌曲或是歌手比较固定，或是往往根据自己的情境来选择适合当下情境的

歌曲；活跃性高的用户，可能是对音乐有特殊爱好或是音乐相关专业，对各种风格的音乐或是歌手有着广泛的涉猎，只是单纯的享受各种音乐带来的听觉享受<sup>[10]</sup>；另外用户职业对其听歌行为存在一定影响，根据用户ID中的注册信息来看，上班族听歌时间集中在上班前、下班后的某些时段，而且时间比较集中，学生族听歌时间相对于上班族比较零散<sup>[11]</sup>，成碎片化分布于图3中的各个时间段；不同人群的不同生活习惯，如作息时间，也会造成用户听歌时间分布的不同，甚至性别差异、文化差异、年龄差异，地区差异都会影响用户的此类行为。

### 3.4 依赖性

在数据集中，歌曲存在两个权重较大属性：album 和 artist。

属性	Strong null model	Weak null model	Temporal null model
album	0.000208	0.005351	0.011823
artist	0.002603	0.024991	0.036152

图5：数据中专辑和歌手属性分布

在图5可以看出，专辑和歌手在所有播放间、歌曲间和相似播放间的相似性值之间的差异还是存在巨大差距的，表明相比于专辑，用户更倾向

于歌手属性。

原因：多数情况下，用户不会直接选择倾向专辑选项，一方面是专辑中的所有歌曲不一定都

是符合用户兴趣的歌曲，另一方面用户有自己的兴趣习惯<sup>[12]</sup>，往往是根据自己的兴趣、情景和网络对专辑中歌曲的评论来选择歌手属性来倾听，此外歌手属性中除了收录了专辑中所有的歌曲，还收录了歌手单独发行的单曲，这是专辑中没有的优势。

此外还有多样性、新颖性、信任性、惊喜性等都是评价推荐系统的指标<sup>[13]</sup>，但是公开的数据样本中，除笔者描述的因素外其他因素的分析效果不明显或是样本数据没有显现，所以笔者没有对其他因素展开详述。

## 4 推荐系统优化措施

由于报告中提取的数据有限和数据精确度等问题，分析的问题没有面面俱到，但是也可以看出移动音频行业快速发展的同时推荐系统存在的问题，针对上述存在的问题笔者给出了若干解决措施：

### 4.1 长尾问题

传统音乐客户端做的关注点是如何给用户推送热点的前 20% 的音乐。而对于个人而言更多的长尾剩余的 80% 其实一样有丰富好听的音乐，却因为缺乏好的推荐机制而被淹没在浩瀚的信息流中<sup>[14]</sup>。推荐系统解决的就是将长尾的 80% 主动的推送给用户，并且可以筛选长尾中个人感兴趣的部分。这样整条音乐链就不会出现断层现象，不会每次都听重复的音乐而产生听觉疲劳。

基于以上可以根据“歌曲基因”理论进行歌曲分类，结合根据用户感兴趣的歌曲，分析其“音乐基因”相似度，根据基因相似度为用户推荐符合其喜好的歌曲<sup>[15]</sup>；系统定期自动抓取用户注册信息、听歌记录等历史行为，分析用户的行为习惯以及用户对歌曲的反馈，在此基础上计算用户

之间在歌曲上的喜好相似度，从而给用户推荐和他有相似听歌爱好的其他用户喜欢的歌曲<sup>[16]</sup>；另外可以根据用户对音乐的共同喜好，为其推荐有着共同曲风爱好的用户。从图 1 中可以看出长尾歌曲和长尾用户占很大比重，如果能够有效的推荐长尾中的歌曲，增加长尾用户的粘性，将会为其赢得更广阔的市场。

### 4.2 覆盖率低问题

覆盖率描述推荐系统对物品长尾的发掘能力。一般通过所有推荐物品占总物品的比例和所有物品被推荐的概率分布来计算。比例越大，概率分布越均匀则覆盖率越大。通过图 2 可以看出，推荐系统的覆盖率比较低，推荐效果不是很理想，可能原因是歌曲基因解析存在问题或用户特征提取不准确，致使用户和歌曲没有实现精准的双向匹配，用户找不到符合其偏好的歌曲，使推荐系统不能准确将歌曲推荐给用户，致使仅有少量的歌曲最终达到用户的推荐列表<sup>[17]</sup>。

基于以上，可以采取给歌曲打标签的方式，让用户和系统共同来为歌曲设置标签，要注意标签的设置方式，使标签信息能够充分表示歌曲的具体属性，可以根据标签数据聚类用户和实现定向推荐，科学设置推荐概率的权重，给予对应的权重值，观察用户一定时期内点击率的变化，根据评测和反馈来调整权重设置，以优化用户推荐，提高歌曲推荐数量<sup>[18]</sup>；还可以建立曲目和用户间关系的数据库，使近似歌曲或是近似用户被推荐时，其他音乐也会在智能排序后推送给用户，增大其他歌曲被推送机会<sup>[19]</sup>；还可以优化混合型推荐技术，多方面抓取、分析用户行为、偏好，针对其兴趣对歌曲进行个性化推荐<sup>[20]</sup>。

### 4.3 间断性分布问题

对于用户听歌时间的间断性分布，可以在不

同的时间段抓取用户特征,利用地理信息系统结合用户所在的上下文环境,实现精准情景推荐<sup>[21]</sup>;另外可以结合时间常识进行推荐,在不同的时间段推荐用户不同的主题歌曲以适应不同的上下文环境,优化用户体验效果:如17-18时多为下班时间,挖掘以往用户在此时间段的行为特征,结合时间环境、地理信息系统和智能交通系统等进行不同场景的推荐,并不定期的对用户的体验进行调查,以优化体验、提高推荐精确度<sup>[22]</sup>。

对于不同用户听歌行为的不同,可以根据用户特征如职业信息等对用户进行多种场景的聚类,将用户聚类为多个不同群体,然后在相同的群体内部根据群内差异在进行细化用户分组,为其推荐群体内相似用户的歌曲,根据系统抓取的用户偏好推荐符合其偏好的歌曲<sup>[23]</sup>。

#### 4.4 依赖性问题的

由于不同用户对歌曲属性依赖性不同,可以通过历史数据分析获取用户对歌曲属性的依赖性,从中得知用户对哪种属性更加依赖,并对该类用户进行分析、聚类,然后结合用户偏好特征的相似属性,对其依赖属性中的歌曲匹配,进行智能排序推荐,为用户推荐符合其偏好的属性歌曲和相同属性偏好的用户,再者可以分析用户的个人收藏或是用户自建曲库,分析歌曲来源路径,根据来源路径信息及曲库歌曲偏好,为用户推荐歌曲<sup>[24]</sup>。

## 5 结语

目前国内还没有专业的第三方数据公司出具针对移动音频的用户数据报告,公司制定相关战略只是根据自身拥有的用户数据做得出的针对本客户端内用户的战略,行业内公布的可用数据只有少量部分,以至于移动音频推荐系统的实战研

究大部分还仅限于企业内部,笔者收集到的音频数据也仅是企业公布的有限数据,上述分析仅能反映推荐系统存在的部分问题,但推荐系统的建设还存在诸多问题,除了上述问题之外还有很多问题亟待解决,如客户端在内容创新上大同小异,没有找到自家门户的个性,同质产品泛滥,用户粘性、用户体验差;系统内广告内容没有实现“私人定制”,还需设计更有效的用户数据分析方法和推荐算法,加上业内缺乏衡量移动音频广告效果的行业标准,使得广告个性化推荐结果无法有效评估;对于用户数据的保密性,也没有相关国家标准和行业标准,使得用户隐私问题变得严峻;此外用户行为习惯和用户反馈的收集、分析、模型的构建都还有很长的一段路要走。

#### 参考文献

- [1] 赵英淑. 国内首个《移动音频媒体价值白皮书》[EB/OL] (2015-10-22) [2015-10-26]. <http://finance.huanqiu.com/roll/2015-10/7863077.html>.
- [2] Murakami T, Mori K, Orihara R. A Method to Enhance Serendipity in Recommendation and Its Evaluation[J]. Transactions of the Japanese Society for Artificial Intelligence, 2009, 24(5): 428-436.
- [3] 陈雅茜. 音乐推荐系统及相关技术研究[J]. 计算机工程与应用, 2012, 48(18): 9-16.
- [4] 网易云音乐发布 2014 中国移动音乐用户行为报告 [EB/OL]. (2014-12-19) [2015-12-13] <http://tech.163.com/14/1218/15/ADOQR50600092594.html>.
- [5] 网易云音乐竞品分析报告 [EB/OL] (2015-02-28) [2015-12-13]. <http://www.jianshu.com/p/9651a78be201>.
- [6] 网易云音乐体验报告 [EB/OL] (2015-06-01) [2015-12-13]. [http://wenku.baidu.com/link?url=6hylvFD3TGQ6z5uU-fwJ7BkC1L\\_bGWLMID-dD2MTka4gIVihe](http://wenku.baidu.com/link?url=6hylvFD3TGQ6z5uU-fwJ7BkC1L_bGWLMID-dD2MTka4gIVihe)

- Mx1QCeDFZ5FVZ3uK9e9q7rykapUNlwa\_Y-sSomQfugFgV61aM9RfyeUqH67.
- [7] 郝立燕,王靖.基于项目流行度的协同过滤 TopN 推荐算法[J].计算机工程与设计,2013,34(10):3497-3501.
- [8] 项亮.推荐系统实践[M].北京:人民邮电出版社,2012:27-28.
- [9] Popescu G, Pu P. What's the Best Music You Have?: Designing Music Recommendation for Group Enjoyment in Group Fun [C]// Proceedings of the 2012 ACM Annual Conference Extended Abstract on Human Factors in Computing Systems Extended Abstract. New York: ACM, 2012: 1673-1678.
- [10] Wang X, Rosenblum D, Wang Y. Context-Aware Mobile Music Recommendation for Daily Activities [C]// Proceedings of the 20th ACM International Conference on Multimedia. New York: ACM, 2012: 99-108.
- [11] 洪颖.音乐推荐系统中音乐情境信息源研究[J].图书馆工作与研究,2014(223):72-75.
- [12] Han Xiaogang, Wei Wei, Miao Cunyan, et al. Context-aware personal information retrieval from multiple social network [J]. IEEE Computational Intelligence Magazine, 2014(5):18-28.
- [13] 朱郁筱,吕琳媛.推荐系统评价指标综述[J].电子科技大学学报,2012,41(2):163-175.
- [14] 克里斯·安德森著.长尾理论 2.0.乔江涛等,译.北京:中信出版社,2009.
- [15] 李瑞敏,闫俊,林鸿飞.基于音乐基因组的个性化移动音乐推荐系统[J].计算机应用与软件,2012,29(9):27-30.
- [16] 范波,程久军.用户间多相似度协同过滤推荐算法[J].计算机科学,2012,39(1):23-26.
- [17] 王文娟.以淘宝网为例探讨网络零售中长尾理论的运用[J].东南传播,2010(6):116-118.
- [18] 王卫平,王金辉.基于 Tag 和协同过滤的混合推荐方法[J].计算机工程,2011,37(14):34-36.
- [19] 李瑞敏,闫俊,林鸿飞.基于用户-标签-项目语义挖掘的个性化音乐推荐[J].计算机研究与发展,2014,51(10):2270-2276.
- [20] Olivier H, Waselewsky M, Pinkwart N. Music Tagge: Exploiting User Generated Game Data for Music Recommendation [M]// Human-Computer Interaction. Users and Applications. Berlin: Springer, 2011: 678-687.
- [21] 洪颖.音乐推荐系统中音乐情境信息源研究[J].图书馆工作与研究,2014(9):72-75.
- [22] 陈雅茜.公共环境下的混合型音乐推荐系统的关键技术研究[J].计算机应用研究,2012,29(11):4250-4253.
- [23] Bogdanov D, Haro M, Fuhrmann F, et al. Semantic Audio Content-Based Music Recommendation and Visualization Based on User Preference Examples [J]. Information Processing and Management, 2013, 49(1):13-33.
- [24] Hurley N, Zhang M. Novelty and Diversity in Top-N Recommendation—Analysis and Evaluation [J]. ACM Transactions on Internet Technology, 2011, 10(4): Article No. 14.