公共环境下的混合型音乐推荐系统的关键技术研究*

陈雅茜
(西南民族大学 计算机科学与技术学院, 成都 610041)

摘 要: 随着计算机网络和多媒体技术的迅速普及,数字音乐消费已经成为人们日常生活中的常见活动,音乐推荐系统也因此成为了推荐系统和电子商务领域的一大研究热点。在对现有音乐推荐系统调研的基础上,重点研究公共环境下的混合型音乐推荐系统的设计和实现;将音乐特征和语境信息相结合,提出了一种新颖的混合型音乐推荐算法。为保证实际应用环境中音乐消费行为的灵活性,系统实现了投票和 DJ 两种推荐模式。该系统在实验室及实地测试中均取得了较高的用户满意度。

关键词: 音乐推荐系统; 混合型推荐系统; 推荐系统
中图分类号: TP311 **文献标志码**: A **文章编号**: 1001-3695(2012)11-4250-04
doi:10.3969/j.issn.1001-3695.2012.11.063

Hybrid music recommender system in public environment

CHEN Ya-xi
(*School of Computer Science & Technology*, *Southwest University for Nationalities*, *Chengdu* 610041, *China*)

Abstract: With the rapid development of the Internet and multimedia technologies, digital music consumptions have become daily activities. Music recommender systems(MRS) have thus become one of the key research issues in the fields of recommender systems and E-commerce. Based on a thorough investigation of existing MRSs, this research explored hybrid MRS in public environment. Based on a combination of musical features and contextual information, this paper presented a hybrid recommendation algorithm. And it presented voted and DJ modes to ensure flexibility in a practical scenario. Satisfactory results were achieved in the laboratory test and field study.

Key words: music recommender system; hybrid recommender system; recommender system

近年来,随着计算机网络和数字音乐的不断普及,音乐推荐系统已成为电子商务领域的一大热门应用类别。它不仅帮助用户快速找到喜爱的音乐,向用户推荐适合其个人品味的音乐,同时还能通过推荐的方式进行音乐推广服务。随着个人音乐库规模的不断扩大以及在线听歌系统的快速普及,如何向用户推荐合适的音乐并为之构建播放列表成为了音乐推荐系统的核心任务[1]。音乐消费是多层次的、具有高度灵活性的复杂活动,因此音乐推荐系统是一个很有挑战性的课题。

尽管音乐推荐系统已得到了迅速发展,但现有系统大多针对个人用户。考虑到公共应用环境的推荐系统还比较少,因此,在对现有音乐推荐系统的详细调研的基础上,本课题重点研究公共环境下的混合型音乐推荐系统。

1 音乐推荐系统

1.1 个人音乐推荐系统

早期的音乐推荐系统大多数使用基于内容的推荐算法,通过音频信号分析得到歌曲的底层特征,如过零率、光谱质心、梅尔频率倒谱系数等[2,3],在此基础上提取音色、节奏、乐器等音乐特征,然后通过相似度对比将内容相似的歌曲纳入播放列表[4,5]。

由于音频信号处理是一项耗时的任务,因此通常不能实现特征的实时提取。另外,音乐理解是很主观的个人行为,从底层音频信号中提取出的音乐特征往往和用户的主观理解之间存在着一定的差距。因此研究者开始探索基于语境(context)信息的协同过滤(collaborative filtering)技术[5]。Yapriady 等人[6]通过用户的年龄、性别和国籍等个人信息来计算用户间的相似度,然后结合皮尔森相关系数来预测当前用户对某首歌曲的评分。Pampalk 等人[7]通过分析用户的听歌习惯来实现播放列表的实时更新:删除列表中与用户跳过歌曲相似的歌曲,添加和播放歌曲类似的歌曲。另外一些研究则着重探索推荐算法应遵循的基本规则。例如,PATS[8]通过限制同一首歌曲在列表中的出现次数来保证和谐与变化之间的平衡;Hansen 等人[9]则用列表中的同时出现率(co-occurrence)来计算歌曲间的相似度。

考虑到基于内容和语境分析技术各自的优点,一些研究开始尝试混合型(hybrid)推荐系统的设计。例如,Kaiji 等人[10]通过内容分析生成初始播放列表,分析歌词和标签计算音乐和用户品味间的相似度,然后对列表进行相应修改。

1.2 群体音乐推荐系统

考虑到音乐消费的公众社会性[1],一些音乐推荐系统支持多用户的实时音乐推荐。一些系统支持用户对播放内容的直接控制,例如,通过传感器捕捉到的用户身体摇动节奏来控制播放速度[11],或通过臂挂式红外设备自动识别多个用户的

收稿日期: 2012-03-01; **修回日期**: 2012-04-15 **基金项目**: 国家教育部留学回国人员科研启动基金资助项目;西南民族大学中央高校基本科研业务费专项资金资助项目(11NZYQN25);2012 年国家外国专家局资助项目(2012-45)

作者简介: 陈雅茜(1981-),女,四川雅安人,博士,CCF 会员,主要研究方向为推荐系统、信息检索、人机交互(yaxichen@swun.edu.cn).

个人爱好从而产生一个群体播放列表[12]。为了提高系统的普及性及用户的参与度,很多项目将手机或电脑等用户个人设备作为基本的交互设备。在 BluetunA[13] 中,系统通过蓝牙技术自动读取用户手机中存放的个人爱好信息,然后通过投票法产生群体播放列表。

这类隐式的交互方式不需要特别的用户投入,系统在用户的自然状态下就能收集到所需信息。但提前设定的用户爱好不能满足某些灵活多变的实际应用环境,因此一些研究者致力于显式用户交互的探索。在 AmbientDJ[14] 和 Jukola[15] 中,用户可以通过手机、PDA、电脑等多种设备进行歌曲选择,系统根据投票结果决定歌曲的播放顺序。

2　混合型音乐推荐系统

现有系统大多通过投票法来决定用户选择音乐的播放顺序。当用户投票数较多时,投票法具有很好的公平性。但在用户投票数较少或用户不愿投入过多精力的情况下,投票推荐模式的合理性和实用性将有所降低。因此本系统实现了投票和 DJ 两种推荐模式以保证公共环境中音乐消费的灵活性,并将内容分析和语境信息相结合,提出了一种新颖的混合型音乐推荐算法。

2.1　系统构架

如图 1 所示,服务器首先为音乐库生成一幅音乐地图并将其显示在公共显示屏上。为保证信息显示的完整性和可读性,同时降低用户进行音乐选择时耗费的认知成本(cognitive effort),地图内显示的是歌手信息而非歌曲信息[16]。根据音乐流派对歌手进行类别划分,歌手之间的距离表示其相似度。

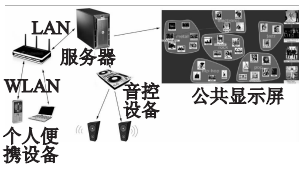

图1　系统构架

用户通过手机、笔记本电脑等个人便携设备访问音乐地图,可进行浏览、查询和选择等操作。服务器根据用户投票结果依次播放每位歌手的热门歌曲;屏幕右侧显示当前播放列表中的前四首歌曲信息,并在地图中通过红色箭头加以标注。当用户投票数较少或播放列表里相邻歌曲间差异较大时,服务器自动转换到 DJ 推荐模式,在相邻歌曲间自动插入过渡歌曲以保证音乐播放的不间断和歌曲切换的平滑性。

2.2　算法实现

笔者曾对音乐用户的消费习惯进行过调查[17],结果显示用户习惯用音乐流派来描述个人喜爱的音乐风格及歌手,用节奏、乐器等内容特征来描述具体歌曲。为了降低音乐选择时的用户投入、方便用户选择,本系统对歌手和歌曲分别使用了基于语境信息和内容分析的相似度度量方法。

1)基于语境信息的音乐地图

如 1.1 节所述,语境信息能有效地反映使用环境及用户反馈,因此广泛应用于协同过滤推荐算法中。对于音乐而言,标签(tag)是使用最为频繁的语境信息,并且歌曲的标签大多与音乐流派有关[18]。因此,本系统利用标签来计算歌手间的相似度,用以生成音乐地图。

同时出现率是信息检索也是音乐推荐领域常用的特征[19]。令 T_i 表示歌手 i 的所有标签,即 $T_i=(t_{i1},t_{i2},\cdots,t_{im})$,则歌手 A_i 和 A_j 间的相似度为

$$S_t(A_i,A_j)=\frac{|T_i\cap T_j|}{|T_i\cup T_j|} \tag{1}$$

得到歌手间的相似度后,用式(2)设定地图中歌手间的距离,其中 $D_{\max}$ 为地图中允许的歌手间最大距离,常数 α 和 β 由歌手总数决定。

$$D_{i,j}=\frac{D_{\max}}{\alpha}\times e^{\beta\times \mathrm{sim}_{i,j}} \tag{2}$$

歌手在地图中的最终位置通过三种常见的力反馈系统实时计算得到[20]

a)n-body force 模拟 n 个物体的互相排斥效果。对于 n 个粒子,给定初始位置 $q_j(0)$ 和速度 $\dot{q}_j(0)$($q_j(0)\neq q_k(0)$):

$$m_j\,\ddot{q}_j=G\sum_{k\neq j}\frac{m_jm_k(q_k-q_j)}{|q_k-q_j|^3}\quad j=1,2,\cdots,n \tag{3}$$

其中:m_j 为质心,q_j 为质心在三维空间内的坐标,G 为重力常量。n-body force 用以生成歌手在地图中的初始位置,相似歌手自动形成聚集效果(见图 1 中右侧地图)。

b)Drag force 代表拖动一个对象所使用的牵引力。当地图内的歌手位置发生变化时,可以用该力模拟各节点的运动。其中,ρ 为质心密度,v 为物体运动速度,A 为参考区域,C_D 为拖动系数。

$$F_D=\frac{1}{2}\rho v^2C_DA \tag{4}$$

c)特别相似的歌手之间的 n-body force 较小,可能形成位置重叠,因此利用 spring force 来保持歌手间的最小距离。其中,x 为移动距离,k 为弹力常量。

$$F=-kx \tag{5}$$

为了方便用户在音乐地图中的浏览,系统对歌手进行了基于音乐流派的聚类。

2)基于内容分析的 DJ 推荐模式

如 2.1 节介绍,当用户投票数较少或播放列表里相邻歌曲间差异较大时,推荐算法将从投票模式切换到 DJ 模式,自动插入相似歌曲以保证音乐播放不间断和歌曲切换的平滑性。首先为每首歌曲提取以下四类音乐特征[3]:

a)过零率(zero crossing rate)表示语音信号波穿过横轴的次数。在过零率基础上计算出过零均值、标准差等统计特征。

$$\mathrm{ZCR}_n=\frac{1}{2}\sum_{m=0}^{N-1}|\mathrm{sgn}[x_n(m)]-\mathrm{sgn}[x_n(m-1)]|$$

$$gn[x]=\begin{cases}1 & x\geqslant 0\\ -1 & x<0\end{cases} \tag{6}$$

b)快速傅里叶变换(fast Fourier transform)是常用的信号频域分析方法,在此基础上计算幅度频谱(magnitude spectrum)和功率谱(power spectrum)。

$$\mathrm{FFT}(x)=\sum_{n=0}^{N-1}x(n)W_N^{nk}\quad k=0,1,2,\cdots,N-1 \tag{7}$$

c)均方根(root mean square)用于计算信号强度。

$$X_{\mathrm{rms}}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2} \tag{8}$$

d)节拍直方图(beat histogram)反映了节奏信息[2]。对于小波变换(wavelet transformation)的子频带(式(9))进行低通滤波、下采样等操作,再对其求和及作自相关函数分析(式(10)),将得到的前五个峰值的 BPM(beats per minute)值相加

得到节拍直方图。在此基础上提取和、最大值等统计数据。

$$W(j,k)=\sum_{j}\sum_{k}x(k)2^{-j/2}\psi(2^{-j}n-k) \tag{9}$$

$$y[k]=\frac{1}{N}\sum_{n=0}^{N-1}x[n]x[n+k] \tag{10}$$

e)通过光谱分析得到光谱质心(spectral centroid)、光谱波动(spectral flux)等特征[2,3]。

系统根据提取的音乐特征进行歌曲间相似度的计算。令 $M=\{m_1, m_2,\cdots,m_N\}$ 为 N 首歌曲集合,其中每首歌曲 m_i 由大小为 K 的特征向量 F_{iK} 表示,即 $m_i=(F_{i1}, F_{i2},\cdots, F_{iK})$。歌曲 i 和 j 的第 k 个特征的相似度 S_k 定义见公式(11),其中 R_k 为该特征在所有歌曲中的分布范围,即 $R_k=|k_{\max}-k_{\min}|$。

$$S_k=1-\frac{|F_{ik}-F_{jk}|}{R_k} \tag{11}$$

歌曲 i 和 j 的内容相似度 S 定义如下:

$$S(m_i,m_j)=\sum_{k=1}^{K}S_k \tag{12}$$

3 系统测试

为了测试音乐地图中投票推荐和 DJ 推荐两种模式的可用性,本文进行了两轮系统测试,第一轮实验室评估后,在某餐吧内完成了第二轮实地测试。

3.1 实验室评估

实验采用参与者内重复测量的析因分析法(repeated measures within participants factorial design)。选取了 20 位歌手共计 90 首热门流行歌曲,所有歌手信息分别显示在音乐地图和对应的文字列表中。分别对顺序模式(即"先到先服务"模式)、投票模式、DJ 模式等三种推荐模式进行测试,因此所有用户(自带手机)参与了 6 组测试(2 种显示方式 ×3 种推荐模式)。参与者在每组测试中进行音乐的自由选择,每组测试的出现顺序进行了平衡处理以减少学习效应(learning effect)。自变量为显示方式(地图、列表)和推荐模式(顺序模式、投票模式、DJ 模式),因变量为用户评分(采取 5 分制李开特量表,5 分为最高分)。对因变量采用了单向重复测量的方差分析(one-way repeated measures ANOVA)并且所有 p 值均经过 Bonferroni 修正。

用户首先填写包含个人信息、音乐消费习惯的调查问卷;经介绍本系统的各项功能后,用户进行了 2 min 的自由使用,然后在接下来的 6 组测试中进行了音乐的自由选择。用户在每组测试完成以后对显示方式(地图、列表)和推荐列表质量进行评分,在所有测试完成后选择一种最喜爱的推荐模式。

八名年龄在 22 ~ 28 岁之间的用户(5 名男性,3 名女性)参与了本次测试。用户对本系统的总体印象较好,各组测试中的用户评分如图 2 所示。相对于文字列表,用户认为音乐地图中的歌手布局合理、传达了丰富的信息且更具娱乐性,因此在三种推荐模式中的得分均较高,特别在投票模式中的得分显著高于列表($P<0.05$)。对于三种推荐模式所产生的播放列表质量,投票模式得分最高。

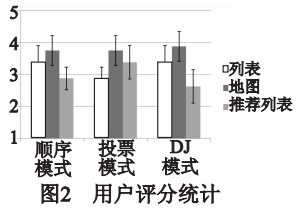

图2 用户评分统计

由于投票模式简单直观且有良好的可预测性,因此大多数用户(5/8)选择投票模式为最喜爱的推荐模式。但是,所有用户均肯定了 DJ 模式的重要性,他们认为 DJ 模式确实提高了歌曲间切换的平滑度。实验室环境不同于自然的社交氛围,用户注意力高度集中在歌曲选择上,而且投票模式中的用户掌控性比 DJ 模式高,这些可能是投票模式评价高于 DJ 模式的主要原因。为了测试本系统在实际应用环境中的表现,笔者进行了第二轮的实地测试。

3.2 实地测试

本次测试在某餐吧内进行(图 3),测试音乐库包含 77 位歌手共计 170 首热门歌曲。测试在完全无干扰的环境中进行,从上午 11 点开始持续到下午 4 点,共持续五个小时。后台日志文件记录了客户端和服务器端的所有活动。研究者以旁观者的身份观察该环境内的顾客反应并在餐吧出口处对离开的顾客进行了随机调查。

与实验室测试结果一致,在用户回访中顾客对音乐地图的实用性和娱乐性评价均较高。他们普遍认为个人设备的引入提高了系统的灵活性,用户选择与 DJ 模式的结合保证了用户只需较小投入就能实现对播放音乐的总体控制。顾客没有访问该系统的原因主要有:手机上无 WLAN 配置;没有感兴趣的音乐。特别有趣的是,大多数顾客由于对当前播放列表感到满意而未对其进行任何修改。

图 4 所示的日志文件[21]显示,在完全没有任何宣传的情况下,大约 20% 的顾客自发地使用手机、笔记本电脑等个人设备访问了本系统。与实验室测试中的积极参与非常不同,音乐选择并非顾客的核心活动,因此进行了音乐选择的顾客并不多。为了保证歌曲的连续播放,系统大多数时间运行在 DJ 模式下。这一结果不仅凸显了 DJ 模式在实际应用环境中的实用性,也反映了用户行为在实验室和实际环境中的差异,从而体现了进行实地系统测试的必要性。

图3 实地测试场地

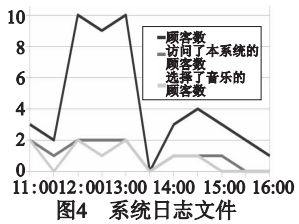

图4 系统日志文件

3.3 讨论

对本系统进行的两次测试结果证明,用户的行为和需求在实验室和实际环境下差异明显,因此多用户交互系统测试应将两种评估方法相结合以获取丰富的定性和定量数据。现有评估方法侧重于用户评分、反馈等定量数据,应继续探讨播放列表质量的量化标准,从而实现可控条件下对推荐算法的定量分析。

本系统在歌手和歌曲层次上分别应用了不同的相似度计算方法,测试结果证明基于标签相似度的音乐地图和基于内容分析的 DJ 推荐模式均受到用户欢迎。考虑到语境信息所反映的丰富内容,如何在混合型音乐推荐系统中引入用户反馈、天气、消费时段等更多类别的语境信息是一个值得深入探讨的课题。

实验室测试中用户的注意力高度集中于音乐的选择与播放上,因此对简单直观的投票模式的评价较高。但在多数应用

环境中音乐选择并非用户的核心活动,此时 DJ 模式的优势得到凸显。由于本系统支持投票模式和 DJ 模式的自动切换,因此能较好地保证主动及非主动情况下的用户需求。

实验结果显示个人便携设备的引入可提高系统灵活性和用户兴趣度。目前本系统中的公共显示屏主要用于信息的公布、显示,尚不支持用户个人设备和公共显示屏之间的直接交互。后续研究将重点探索多用户环境下的个人设备与多触点触摸界面 Tabletop[22] 及大型公共显示屏等非传统交互设备之间新颖的交互模式。

4　结束语

音乐推荐及音乐播放列表的生成是受众多因素影响的复杂任务,特别涉及到多用户的使用环境,如何满足公共环境下用户的实时音乐消费需求是一个富有挑战性的课题。本文首先调查了典型的音乐推荐系统,在此基础上实现了一个支持公共环境下实时推荐的混合型音乐推荐系统,用户通过手机等个人设备进行实时音乐选择,系统通过对歌手、歌曲的语境及内容分析而产生音乐推荐并为之生成播放列表。为了满足灵活多变的用户需求,系统实现了投票和 DJ 两种推荐模式。推荐算法在两轮用户测试中均取得了令用户比较满意的推荐质量。两种推荐模式分别在主动和非主动用户交互中有着各自的优势,因此两者的结合更能保证多用户环境中音乐消费的灵活性。

在接下来的研究中笔者将在推荐算法优化、用户交互模式、推荐质量评估指标及实验设计等方面作进一步的探讨。音乐推荐系统的研究有助于提高音乐推荐质量,鼓励用户参与,从而提高音乐推荐系统在实际应用环境中的可用性。同时,相关研究成果也将对其他推荐系统以及多媒体应用系统的研究起到重要的借鉴作用。

参考文献:

[1] CUNNINGHAM S J, REEVES N, BRITLAND M. An ethnographic study of music information seeking: implications for the design of a music digital library [C]//Proc of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries. Washington DC: IEEE Computer Society, 2003:5-16.

[2] TZANETAKIS G, COOK P. Musical genre classification of audio signals[J]. IEEE Trans on Speech and Audio Processing, 2002, 10(5):293-302.

[3] McENNIS D, McKAY C, FUJINAGA I, *et al*. jAudio: a feature extraction library[C]//Proc of International Society for Music Information Retrieval Conference. 2005:600-603.

[4] POHLE T, PAMPALK E, WIDMER G. Generating similarity-based playlists using traveling salesman algorithms[C]//Proc of the 8th International Conference on Digital Audio Effects. 2005:220-225.

[5] AUCOUTURIER J J, PACHET F. Music similarity measures: what's the use? [C]//Proc of International Society for Music Information Retrieval Conference. 2002.

[6] YAPRIADY B, UITDENBOGERD A L. Combining demographic data with collaborative filtering for automatic music recommendation[C]//Proc of the 9th International Conference on Knowledge-based Intelligent Information and Engineering Systems. 2005:201-207.

[7] PAMPALK E, POHLE T, WIDMER G. Dynamic playlist generation based on skipping behavior[C]//Proc of International Society for Music Information Retrieval Conference. 2005:634-637.

[8] PAUWS S, EGGEN B. PATS: realization and user evaluation of an automatic playlist generator[C]//Proc of the 3rd International Conference on Music Information Retrieval. 2002.

[9] HANSEN D L, GOLBECK J. Mixing it up, recommending collections of items[C]//Proc of SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press, 2009:1217-1226.

[10] KAIJI K, HIRATA K, NAGAO K. A music recommendation system based on annotations about listeners' preferences and situations[C]//Proc of the 1st International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution. 2005:231-234.

[11] FELDMEIER M, PARADISO J A. Giveaway wireless sensors for large-group interaction[C]//Proc of SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press, 2004:1291-1292.

[12] CROSSEN A, BUDZIK J, HAMMOND K J. Flytrap: intelligent group music recommendation[C]//Proc of the 7th International Conference on Intelligent User Interfaces. New York: ACM Press, 2002:184-185.

[13] BAUMANN S, BASSOLI A, JUNG B, *et al*. BluetunA: let your neighbour know what music you like[C]//Proc of SIGCHI Conference on Human Factors in Computing Systems. 2007:1941-1946.

[14] KOMNINOS A, CAMERON N, FERON C, *et al*. AmbientDJ: enabling interaction between people and socially aware environments[C]//Proc of Workshop on Mobile Interaction with the Real Would. 2008.

[15] O'HARA K, LIPSON M, JANSEN M, *et al*. Jukola: democratic music choice in a public space[C]//Proc of the 5th IEEE International Workshop on Distributed Intelligent Systems. 2004:145-154.

[16] BERWEIN F. Party Jukebox: supporting group playlist generation in a public environment[D]. Munich: University of Munich, 2009.

[17] CHEN Ya-xi, BORING S, BUTZ A. How last.fm illustrates the musical world: user behavior and relevant user-generated content[C]//Proc of International Workshop on Visual Interfaces to Social and Semantic Web. 2010.

[18] LI Rui, BAO Heng-hua, YU Yong, *et al*. Towards effective browsing of large scale social annotations[C]//Proc of the 16th International Conference on World Wide Web. New York: ACM Press, 2007:943-952.

[19] HASSAN M Y, HERRERO S V. Improving tag-clouds as visual music retrieval interfaces[C]//Proc of International Conference on Multidisciplinary Information Sciences & Technologies. 2006.

[20] HEER J, CARD S K, LANDAY J A. Prefuse: a toolkit for interactive information visualization [C]//Proc of SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press, 2005:421-430.

[21] LAMBERS F. Party Jukebox2: Supporting co-located musical consumption in apublic environment[D]. Munich: University of Munich, 2010.

[22] HILLIGES O, IZADI S, WILSON A D, *et al*. Interactions in the air: adding further depth to interactive tabletops[C]//Proc of the 22nd Annual ACM Symposium on User Interface Software and Technology. New York: ACM Press, 2009:139-148.