

doi: 10.12052/gdutxb.220009

基于用户长短期偏好和音乐情感注意力的音乐推荐模型

吴亚迪, 陈平华

(广东工业大学 计算机学院, 广东 广州 510006)

摘要: 针对现有音乐推荐在用户偏好建模时忽略用户长期偏好, 或对用户记录统一建模时忽略历史信息与当前情境联系的问题, 提出一种基于用户长短期偏好和音乐情感注意力的音乐推荐模型。首先将用户听歌记录切分为多个历史序列和当前序列, 利用多个长短期记忆网络分别进行特征提取, 得到用户长短期偏好: 对于历史音乐序列, 提出序列时段的概念, 并进行序列时段加权计算, 得到长期偏好; 对于当前序列, 利用平均池化提取当前情景音乐特征, 得到短期偏好。其次, 从音乐声学信号中学习音乐情感特征, 应用注意力机制计算音乐情感因子。最后, 将音乐情感因子融入用户长短期偏好, 得到一个音乐推荐列表。在Last.fm真实数据集上的实验结果表明, 模型的NDCG@10达到了0.5435, 优于现有方法; 消融实验和特征贡献分析进一步验证了模型的有效性。

关键词: 音乐推荐; 用户偏好; 音乐情感; 长短期记忆网络; 注意力机制

中图分类号: TP391

文献标志码: A

文章编号: 1007-7162(2023)04-0037-08

A Music Recommendation Model Based on Users' Long and Short Term Preferences and Music Emotional Attention

Wu Ya-di, Chen Ping-hua

(School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: Users' long-term preferences and the relationship between historical information and current situation are usually ignored in existing user preference modeling and user record unified modeling. To address this, in this paper, we propose a music recommendation model based on users' long-term and short-term preferences and music emotional attention. Firstly, we divide the user's listening record into multiple historical and current sequences, and learn the features by multiple long and short-term memory networks, respectively, to obtain user's long and short-term preferences. For the historical music sequences, we propose the concept of sequence period is proposed to obtain the long-term preferences by calculating the weights of the sequence period. For the current sequence, we use the average pooling to extract the music features of the current scene to obtain the short-term preference. Secondly, we learn the emotional characteristics of music from the acoustic signals, and use the attention mechanism to calculate the emotional factors of music. Finally, we integrate the music emotion factor into the user's long-term and short-term preference to generate a music recommendation list. The experimental results on the last.fm real data set show that the proposed model achieves 0.5435 in term of the NDCG@10, which is better than the existing methods. The ablation experiment and characteristic contribution analysis further demonstrate the effectiveness of the model.

Key words: music recommendation; user preference; sequence analysis; long-short term memory; attentional mechanism

收稿日期: 2022-01-27

基金项目: 广东省重点领域研发计划项目(2021B0101200002, 2020B0101100001); 广东省科技计划项目(2020B1010010010)

作者简介: 吴亚迪(1997-), 女, 硕士研究生, 主要研究方向为推荐系统、数据挖掘, E-mail: 603001605@qq.com

通信作者: 陈平华(1967-), 男, 教授, 主要研究方向为云计算、大数据、推荐系统, E-mail: pinghuachen@163.com

近年来,随着数字多媒体技术的发展,音乐以更加灵活的形式出现在人们的身边,影响着人们的生活。尽管人们可以获取大量的数字音乐,但是要从完整的数据库中挑选出最喜欢的音乐对用户来说是困难和耗时的。在这种情况下,缓解这一问题的主要解决方案是根据用户的历史音乐行为建立用户的音乐偏好模型,为用户推荐能够满足用户兴趣的音乐。音乐推荐提高了音乐服务的智能化程度,准确率高的音乐推荐系统可以大大提高用户欣赏音乐的便利性。

目前市场上很多音乐平台都有音乐推荐功能,如Last.fm、虾米音乐、网易云音乐、豆瓣电台等。这些音乐推荐系统首先建立自己的音乐库,然后分析歌曲的特点和用户的收听习惯,为用户做出推荐,协同过滤方法和基于内容的方法或者两者的结合是这些推荐系统中常用的推荐算法。例如,Last.fm根据用户的播放行为和收藏记录,找到若干与其兴趣爱好最相似的用户最喜欢的音乐,推荐给目标用户;网易云音乐根据用户的歌曲播放列表、歌曲播放列表的分类和相似性来推荐歌曲。但大多数现有音乐平台会忽略用户的长期偏好,或在推荐时未充分考虑到历史信息和当前情境中包含的时间关联以及音乐情感属性。

用户偏好是用户在考量商品或服务时做出的理性和有倾向性的选择,其中,用户听歌记录在音乐推荐的用户偏好建模中扮演着重要的角色。用户将播放的下一首音乐受其最近播放过的音乐的影响,即短期偏好;长期偏好表示从用户历史记录中挖掘出的总体兴趣。长期偏好通常是稳定的,而短期偏好往往随着时间的推移而频繁变化^[1]。用户的音乐偏好同时受用户的总体兴趣和当前状态影响,而这期间的时间关联性及音乐情感因素不容小觑。

为此,本文提出一种混合音乐推荐模型来解决音乐推荐中的上述问题。针对用户听歌记录,结合序列时段性对用户长短期偏好分别进行建模,同时利用音乐声学信号,对音乐情感因子进行分析,通过特征贡献分析揭示序列时段、音乐情感与用户音乐偏好之间的内在关联性。在大型数据集上的大量实验表明,本文模型显著优于现有的音乐推荐模型或方法。本文的主要贡献如下。

(1) 利用多个长短期记忆网络 (Long-Short Term Memory, LSTM) 对历史序列片段和当前序列分别进行建模,维护序列依赖关系;

(2) 针对历史信息和当前信息采取不同操作,提出序列时段相似性度量历史序列与用户当前所处时

间段的联系,收集历史信息中最有用的信息;

(3) 基于深度学习方法从声学信号中提取音乐情感特征,并利用注意力机制 (Attentional Mechanism, AM) 分析音乐情感因子;

(4) 融合用户长期和短期偏好,并结合音乐情感,进行系统建模,并通过大型数据集上的大量对比实验和消融实验验证了本文方法。

1 相关工作

1.1 基于协同过滤的音乐推荐

目前,音乐推荐的研究主要采用协同过滤的方法,包括基于内容的方法、基于上下文的方法、基于图模型的方法等。

基于内容的音乐推荐方法将与内容特征相似的音乐添加到用户的播放列表中。例如,Bogdanov等^[2]从音频数据中提取声学特征,并与用户喜欢的音乐进行比较,将声学特征相似的音乐作品分成一类,推荐给用户;Oord等^[3]使用卷积神经网络来预测音乐音频的潜在因素。

基于上下文的音乐推荐方法^[4]在传统推荐方法的基础上增加相关上下文信息。例如,Cheng和Tang^[5]将声学信号与用户个性相结合,Cheng等^[6]试图捕捉用户特定信息对音乐偏好的影响。

基于图模型的音乐推荐方法^[7],将用户的历史音乐记录转化为图模型上的边和节点,将用户对音乐的偏好行为转化为边的权重。根据已有信息度量用户节点与音乐节点之间的相关性,将相关性高的节点添加到用户喜爱的音乐列表中。例如,Oramas等^[8]提出了基于知识图谱的音乐推荐方法,建立了2种显式的图形特征映射。

这些音乐推荐方法的基本思想是基于用户对音乐的历史偏好行为,通过不同的相似度度量方法,直接为用户生成推荐音乐列表。这些方法只能对用户的静态偏好建模,不考虑用户对音乐行为的历史偏好所反映的用户兴趣的波动,影响了推荐音乐的准确性。

1.2 基于序列分析的音乐推荐

现有的研究主要集中在对短期的用户偏好(用户实时行为)动态进行建模,这些研究的思想是认为用户的状态在短时间内保持稳定,用户可能听的下一首歌曲与用户正在听的歌曲具有相似的特征。例如,Lee等^[9]将用户-物品交互和音乐音频端到端结合起来,提出深度内容-用户嵌入模型;Sachdeva等^[10]提出一种注意力神经网络,利用音乐序列及特征学习

用户短期偏好。

除了近期听歌行为所反映的短期偏好动态外,用户的听歌记录还表现出长期的周期性规律^[11]。然而,只有少数研究同时考虑了长期和短期偏好,其中,Wang等^[12]基于音乐听歌记录和元数据,提出一种情境感知的音乐推荐方法;Shen等^[13]采用分层注意力网络,对用户的长期品味和短期偏好进行建模。

上述工作虽然取得了一定的进展,但往往采用同一方法对全部用户-音乐交互记录进行统一建模,忽略了历史序列与当前情境的时段相关性或长期和短期偏好对音乐推荐的不同贡献。

1.3 基于音乐情感的音乐推荐

由于音乐的内容是负载情感的,情感语境也会影响音乐偏好,一些研究在对用户偏好进行建模时也考量了音乐的情感特征。例如,Dhahri等^[14]根据隐式用户反馈构建个性化的情感感知音乐地图;Ashu Abdul等^[15]结合深度卷积神经网络方法和加权特征提取方法,从音乐数据中感知情绪,在加权特征提取方法中,通过生成隐含的用户对音乐的评分来提取用户数据与音乐情感之间的相关性;Bhaumik等^[16]利用生成式对抗网络,从每首歌的特定对象中提取情感样本。

2 模型总体框架

本章将介绍本文提出的nLSTMs+AM模型。模型分为3个模块:第1个模块为基于nLSTMs结构的用户长短期偏好建模模块,利用多个LSTM网络对用户听歌序列进行建模,学习用户长期和短期偏好;第2个模块为音乐情感分析模块,采用注意力机制学习音乐情感在推荐中的贡献,即音乐情感因子;第3个模块为推荐模块,利用全连接层和softmax将音乐情感因子融入用户长期偏好和短期偏好,得到最终top-N推荐。

当用户 u 在时间戳 t 播放某音乐 m 超过1 min时,记录一条用户-音乐交互记录 (u, m, t) ,目标是在下一个时间戳向用户 u 推荐最优的 N 首音乐。模型总体框架如图1所示。

2.1 用户长短期偏好建模

在现实场景中,用户对音乐的偏好很大程度上取决于在推荐时的情境,其中时间是很重要的一个信号,例如,在上班族的日常生活中,她/他可能会在工作日的夜晚收听一些安静放松的音乐。然而当前的大部分研究并未考虑到历史序列与当前情境的时

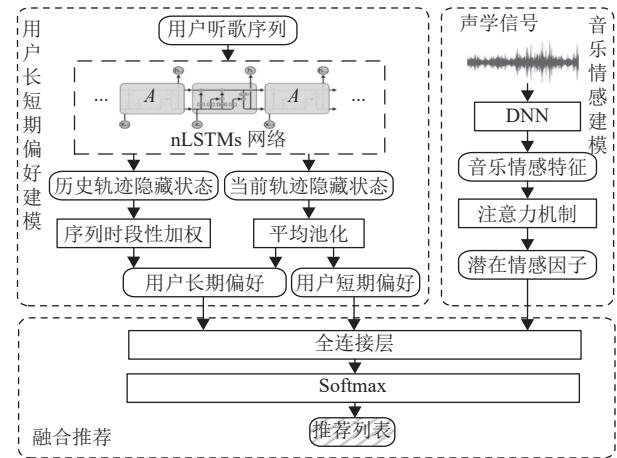


图1 nLSTMs+AM模型框架

Fig.1 The basic framework of the nLSTMs+AM model

间相关性。为解决上述问题,本节通过分析不同时间段的音乐相似性,对历史序列进行时段性加权操作,挖掘历史信息与当前情境的联系,结合多个LSTM网络分别对历史记录和当前记录进行解码,从用户长期偏好和短期偏好2个角度系统地对用户偏好进行建模,得到长短期偏好的不同表示。

图2描述了nLSTMs模块的结构,包括长期偏好建模和短期偏好建模。

设 $U = \{u_1, u_2, \dots, u_{|U|}\}$ 表示一组用户, $M = \{m_1, m_2, \dots, m_{|M|}\}$ 表示一组音乐,每首音乐通过ID进行编码。对于用户 $u \in U$,可以得到一个多段的听歌序列集 $S = \{S_1, S_2, \dots, S_n\}$,其中 S_n 为当前序列。序列 $S_i \in S$ 由用户 u 播放的一系列音乐组成,即 $S_i = \{m_1, m_2, \dots, m_{|S_i|}\}, m \in M$ 。

给定听歌记录 S ,对于目标用户 $u \in U$,本节的目标是针对用户的历史序列集 $\{S_1, S_2, \dots, S_{n-1}\}$ 和当前序列 $S_n = \{m_1, m_2, \dots, m_{t-1}\}$,对用户进行长期和短期偏好建模,其中 m_{t-1} 是 u 最近一次播放过的音乐。

2.1.1 基于时段性的历史序列表示

当对长期偏好建模时,一个直观的想法是根据当前情况从历史中有选择地收集最有用的信息。因此,设计了一个基于序列时段性的时间加权操作来学习目标用户历史序列的表示。

(1) nLSTMs网络:首先,给定用户 u ,利用 n 个维护序列依赖关系的LSTM网络对每个序列 $S_h \in \{S_1, S_2, \dots, S_n\}$ 中的所有音乐 $\{m_1, m_2, \dots, m_{|S_h|}\}$ 进行编码。

$$\mathbf{h}_t = \text{LSTM}(m_t, \mathbf{h}_{t-1}) \quad (1)$$

式中: \mathbf{h}_t 为LSTM的隐藏状态。这样, S_h 中的音乐 $\{m_1, m_2, \dots, m_{|S_h|}\}$ 被编码为 $\{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{|S_h|}\}$ 。

(2) 序列时段性加权操作:在现实场景中,音乐的流行程度随着时间的推移而变化。例如,人们可能

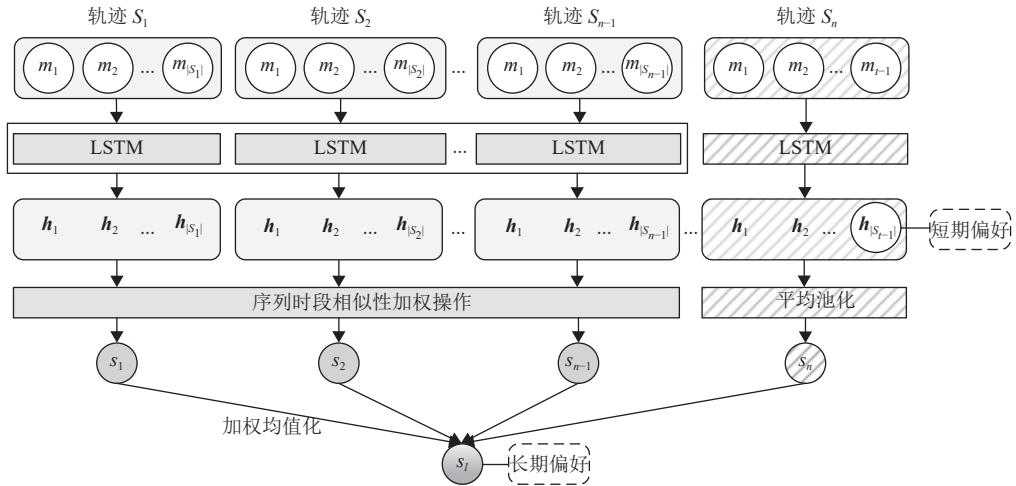


图2 用户长短期偏好建模结构

Fig.2 The structure of the user long- and short-term preference model

在午夜变得伤感,播放的音乐也可能转变为抒情的。因此,每个序列 S_h 的总体表示 s_h 应该包含这些信息,以捕获时间敏感性。为实现这一点,对于长期偏好建模,提出了时段相似性加权操作。具体来说,将1天划分为4个时间段,那么1周可划分为28个时间段。对于每个时间段*i*,构造一个音乐集 $H_i = \{m_1, m_2, \dots, m_{|H_i|}\}$,其中 $m \in M$ 是在时间段*i*内至少一个用户播放的音乐。然后,计算第*i*和第*j*个时间段的时间相似度 $\tau_{i,j}$ 。

$$\tau_{i,j} = \frac{|H_i \cap H_j|}{|H_i \cup H_j|} \quad (2)$$

从直观上看,2个时间段的音乐重叠越多,它们的相似度就越高。对于某个历史序列 S_h ,根据每首音乐的播放时间,可以推导出 $|S_h|$ 首音乐的播放时间段序列,用 $\{p_1, p_2, \dots, p_{|S_h|}\}$ 表示,其中 $p \in \{1, 2, \dots, 28\}$ 。

根据目标用户的当前时间段*c*,某个历史序列 S_h 的总体表示 s_h 为

$$s_h = \sum_{t=1}^{|S_h|} w_t \mathbf{h}_t, w_t = \frac{\exp(\tau_{c,p_t})}{\sum_{j=1}^{|S_h|} \exp(\tau_{c,p_j})} \quad (3)$$

式中: τ_{c,p_j} 为 S_h 中当前时间段*c*与第*j*次播放的音乐的时间段之间的时间相似性。

通过上述加权操作,某段历史序列表示 s_h 与用户当前时间段及最近听歌序列建立联系。到目前为止,可以用 $\{s_1, s_2, \dots, s_{n-1}\}$ 来表示*n*-1段历史序列。

2.1.2 长短期偏好建模

对于当前轨迹 S_n ,利用一个单独的LSTM来对最近播放的音乐进行显式建模,则当前轨迹的最终隐藏状态 h_{t-1} 可表示为用户短期偏好。此外,对于 S_n ,序列时段性加权操作被平均池代替。

$$s_n = \frac{1}{|S_n|} \sum_{t=1}^{|S_n|} \mathbf{h}_t \quad (4)$$

其基本原理是,在轨迹 S_n 中新播放的音乐更能代表用户最近的偏好,而平均池保存了 S_n 中所有音乐的信息,这对长期和短期偏好建模都有好处。

在学习了所有轨迹 $S = \{S_1, S_2, \dots, S_n\}$ 的表示 $\{s_1, s_2, \dots, s_n\}$ 之后,通过加权均值化操作来计算用户长期偏好 s_l 。

$$s_l = \frac{1}{|S_n|} \sum_{i=1}^{|S_n|} \mathbf{W}_i s_i \quad (5)$$

式中: \mathbf{W}_i 为可训练的权重矩阵。

2.2 音乐情感注意建模

音乐是一种表达情感的艺术形式,其情感表达作为音乐的高阶属性,是用户偏好的重要影响因素。本节利用深度学习方法对音频信号中提取的声学特征进行建模,得到音乐情感特征,并通过注意力机制分析音乐情感在推荐中的权重,以获得更准确的用户音乐偏好。

2.2.1 音乐特征提取

音乐的声学测量已被广泛研究,在这项工作中,首先使用开源多媒体特征提取器openSMILE来提取988维的声学特征集,包括低级别描述符(如强度、响度、音调等),以及这些描述符的 δ 系数和一些函数值(如范围、平均值、偏度、峰度)。

接着,采用一个深度神经网络(Deep Neural Network, DNN)对声学特征进行训练,以隐式和非线性的方式对音乐的情感建模(开心、喜欢、惊讶、愤怒、悲伤、恐惧、讨厌等),生成7维音乐情感向量。DNN的输入是所有声学特征向量的拼接,DNN的过程形式上可以表达为

$$z_{h+1} = \sigma(\mathbf{W}_h z_h + \mathbf{b}_h) \quad (6)$$

式中: \mathbf{W}_h 、 \mathbf{b}_h 为参数, \mathbf{z}_h 为第 h 层的隐藏层状态。

2.2.2 情感潜在因子建模

首先将所有特征向量(音乐情感特征、用户ID、音乐ID、时间特征)转化为 d 维嵌入向量 \mathbf{x}_i' , 其中时间特征根据1天(早上、下午、晚上和午夜)和1周(工作日和周末)的时间对每个用户-音乐交互的时间戳进行分类。

$$\mathbf{x}_i' = \mathbf{W}_{x_i} \mathbf{x}_i + \mathbf{b}_{x_i} \quad (7)$$

式中: \mathbf{W}_{x_i} 、 \mathbf{b}_{x_i} 为嵌入参数(对于独热稀疏向量, $\mathbf{b} = 0$), \mathbf{x}_i 为各特征向量。

注意力机制在将不同部分压缩成单一表示时区分它们的重要性^[15]。首先对特征进行加权求和:

$$\mathbf{e}_{\text{att}} = \sum_{i=1}^n \alpha_i \mathbf{x}_i' \quad (8)$$

式中: α_i 为嵌入特征 \mathbf{x}_i' 的注意力分数。在这里, 注意力分数通过一个2层注意力网络进行计算。

$$\alpha_i' = \mathbf{w}^T \sigma(\mathbf{W} \mathbf{x}_i' + \mathbf{B}) + \mathbf{b} \quad (9)$$

式中: \mathbf{W} 、 \mathbf{B} 是第1层参数, \mathbf{w} 、 \mathbf{b} 是第2层参数。式(4)的注意力分数进一步采用softmax归一化处理。

$$\alpha_i = \frac{\exp(\alpha_i')}{\sum_{i=1}^n \exp(\alpha_i')} \quad (10)$$

这样, 最终得到了音乐的情感潜在因子, 即 \mathbf{e}_{att} , 如图3所示。

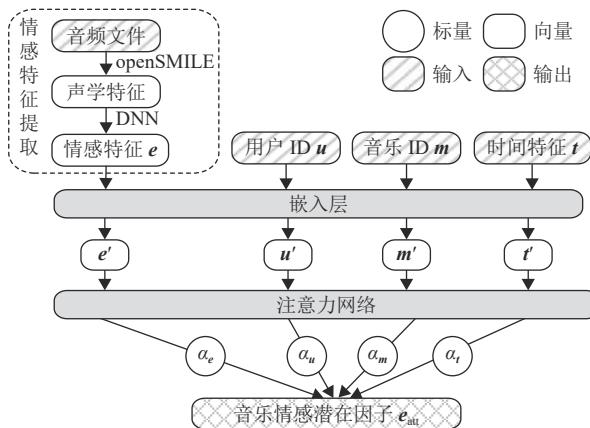


图3 音乐情感建模结构

Fig.3 The structure of the music emotion model

2.3 融合推荐

在得到音乐情感因子及用户长期、短期偏好的表示之后, 计算 $|M|$ 首音乐上的概率分布 \mathbf{p} :

$$\mathbf{p} = \text{softmax}(\mathbf{W}_p(\mathbf{e}_{\text{att}} \oplus \mathbf{s}_l \oplus \mathbf{h}_{t-1})) \quad (11)$$

式中: \oplus 为全连接层, \mathbf{W}_p 为所有音乐的可训练投影矩阵, \mathbf{e}_{att} 为音乐情感因子, \mathbf{s}_l 、 \mathbf{h}_{t-1} 分别为用户的长期和

短期偏好。

因此, 目标用户在下一个时间戳 t 最有可能播放的音乐就是概率最大的那个, 则目标函数可以表示为对数似然:

$$L = -\sum_{k=1}^N \ln(p_k) \quad (12)$$

式中: N 为训练样本总数, p_k 为模型对第 k 个训练样本生成音乐的概率。考虑到音乐推荐中负反馈的自然缺少, 对于每个正样本 (u, m^+, t) , 随机选择 u 从未交互的音乐 m^- , 形成负样本 (u, m^-, t) 。负样本在每次训练遍历中选取, 在不同遍历中可能有所不同。

3 实验及分析

本节通过在真实数据集上大量的对比实验和消融实验来评估nLSTMs+AM模型。

3.1 实验设置

3.1.1 数据集

本文实验数据来源于Last.fm (<http://www.last.fm>) 的1K users数据集, Last.fm是一家英国的网络电台和音乐社区, 其向开发者提供了丰富的API, 其中包括1K users。该数据集作为有上下文信息的隐性反馈数据集的代表, 包含了近1000位用户的所有音乐播放记录与播放时间, 以及用户和音乐的相关信息。

对于该数据集, 过滤播放用户少于10个的不受欢迎的音乐; 将用户1 d内的所有播放记录视为一个单一的轨迹, 并删除少于3个播放记录的轨迹; 轨迹少于5条的非活跃用户也会被过滤掉。经过预处理后, 数据包含301 080条用户-音乐交互记录, 包含934个用户和40 868首音乐。

将每个用户的前80%的轨迹用于训练, 其余用于测试。虽然某个用户音乐记录可能涉及几个交互记录, 但保证它们都属于训练集或测试集。

考虑到时间戳, 每个测试用例只有一个正样本, 每个正测试样本与99个随机抽样的负样本配对, 每种方法预测100个样本的偏好得分。

3.1.2 对比模型

本文将提出的nLSTMs+AM与以下通用的基于深度学习的推荐方法进行比较:

LSTM: 这是RNN模型的一种变体, 在处理序列数据方面表现良好。

DeepFM^[18]: 它结合了因子分解机和神经网络。

AFM^[19]: 它通过一个神经注意力网络来辨别不同特征交互的重要性。

此外, 还考虑了以下面向音乐的推荐方法:

c-IFRM^[20]:一种针对用户音乐播放记录,结合聚类和潜在因子模型的混合音乐推荐方法。

DCUE^[9]:深度内容-用户嵌入模型的简称,一种利用用户-物品交互和音乐音频内容的混合方法。

STABR^[10]:一种利用注意力神经网络从音乐序列及特征中学习用户短期偏好的方法。

MEM^[12]:根据用户的收听记录,通过音乐嵌入来对用户的全局和上下文音乐偏好进行建模。

PEIA^[13]:一种对用户的长期品味和短期偏好进行建模的注意力模型。

3.1.3 评价指标

为了评估每个方法的性能,采用以下3个在音乐推荐方法中常用的评价指标:

Recall@K。召回率,表示正确预测出正样本占实际正样本的概率。

HR@K。命中率,在Top-N推荐中,是一种常用的衡量召回率的指标,计算公式为

$$HR@K = \frac{\text{NumberofHits}@K}{GT} \quad (13)$$

式中:GT为所有的测试集合,NumberofHits@K为每个用户Top-K列表中属于测试集合的个数的总和。

NDCG@K。归一化折损累计增益,用来评价Top-K列表与用户真实交互列表的差距。高关联度的结果比一般关联度的结果更影响最终的指标得分;有高关联度的结果出现在更靠前的位置的时候,指标得分会越高。

在本文中,选择K=10进行评估。

3.1.4 参数设置

将所有基于深度学习方法的嵌入维度和隐藏状态设置为500。模型中所有的参数采用高斯分布随机初始化,并利用梯度下降优化算法Adam进行优化,批量大小为256,学习率为0.01。

3.2 实验结果

3.2.1 推荐性能评估

不同的推荐方法的结果如表1所示。在每一列中,最好的结果用黑体突出显示。

从统计数据中得出以下观察结果:

(1)在实验数据集上,本文提出的nLSTMs+AM在每个指标上都持续显著地优于所有对比模型,nLSTMs+AM的Recall@10、HR@10和NDCG@10分别达到0.7048,0.8117,0.5435。定量评价结果清楚地证明了该方法的优越性。

(2)基于序列分析的方法(STABR、MEM、PEIA)虽然细节结构不同,但总体上优于基于相似性度量

表1 不同方法的推荐性能

Table 1 Experimental results obtained by different methods on the Last.fm dataset

方法	Rec@10	HR@10	NDCG@10
LSTM	0.4959	0.6335	0.4022
DeepFM	0.5686	0.6976	0.4626
AFM	0.5992	0.7536	0.4924
c-IFRM	0.5852	0.7242	0.4901
DCUE	0.6436	0.7530	0.4978
STABR	0.6390	0.7619	0.4942
MEM	0.6648	0.7783	0.5118
PEIA	0.6901	0.7941	0.5369
nLSTMs+AM	0.7048	0.8117	0.5435

的方法(DeepFM、AFM、c-IFRM),这说明利用听歌记录对用户偏好进行动态建模可以提高推荐性能。

(3)在对比模型方法中,MEM和PEIA表现最好。由于这两种方法都考虑了长期和短期偏好,它说明了在短期偏好建模之外,从历史信息建模长期偏好的重要性。然而,在没有区别考虑历史和当前序列并挖掘两者相关性的情况下,MEM和PEIA很难使推荐精度最大化,这是nLSTMs+AM优于这两种方法的关键因素。

(4)与其他模型相比,基于注意力机制的方法(AFM和PEIA)表现较优,体现了以不同的权重区分特征的重要性。

(5)DeepFM、AFM均优于LSTM。这显示了建模不同音乐之间的关系的优势。此外,c-IFRM仅以上下文信息作为输入,其性能不如DCUE、STABR、MEM。相比之下,其他的方法将音乐特征或音乐情感融合到序列建模中,改进了性能。

3.2.2 用户长短期偏好模块评估

为了验证nLSTMs+AM中长期、短期偏好建模和序列时段性加权操作对性能增益的贡献,本部分进一步对模型的简化版本进行消融测试。

S_Model:这个版本删除了本文模型的长期模块,仅利用单个LSTM网络对用户听歌记录进行建模得到用户短期偏好。

L_Model:这个版本删除了本文模型的短期模块,只启用了长期模块。

L_Model':这个版本删除了本文模型的短期模块,且历史序列片段的时段性加权操作以平均池化操作代替。

Model':这个版本采用平均池化操作代替本文模型中的序列时段性加权操作。

nLSTMs+AM降级版本的结果如表2所示。

表2 不同nLSTMs+AM变体的性能

Table 2 Performance of different nLSTMs+AM variants

方法	Rec@10	HR@10	NDCG@10
S_Model	0.5854	0.7381	0.5150
L_Model'	0.6487	0.7617	0.5274
Model'	0.6741	0.7836	0.5312
L_Model	0.6924	0.7981	0.5396
nLSTMs+AM	0.7048	0.8117	0.5435

通过消融实验,可以观察到:

(1) L_Model' 总是比S_Model表现得更好,原因是L_Model' 可以更好地捕捉用户的长期周期性,这清楚地展示了建模用户长期偏好的好处。

(2) L_Model相对于L_Model'、Model' 均进一步提高了推荐性能,原因是L_Model可以根据用户的当前状态更好地捕捉用户历史信息中相关性较大的部分,这验证了序列时段性分析的有效性。

(3) 虽然S_Model的竞争力不如其他变体模型,但它仍然可以获得比表1中的LSTM和c-IFRM等对比模型更好的预测性能,这是由于S_Model的有效性来自于注意力模型,它捕捉了音乐与短期偏好之间的情感关系。

(4) 完整的模型nLSTMs+AM是上述变体模型的组合,在数据集上的性能是最好的,表明长期、短期偏好和序列时段性分析都对用户下一首音乐的选择有积极的影响。

3.2.3 音乐时段与情感贡献分析

在本小节中,通过用户偏好分析评估时间段特征和注意力机制中音乐情感特征的有效性。

针对音乐情感特征和时间特征,分析1 d中音乐偏好的时间变化。具体来说,关注用户播放音乐的情感向量,计算4个时间段中7维情感的平均值。如图4所示,人们在白天更倾向于选择更多表达“喜欢”和“愤怒”的振奋人心的音乐,而在夜晚更倾向于选择表达“恐惧”和“悲伤”的压抑的音乐,这与人们的日常情绪变化情况一致。

结合表3中特定用户的音乐行为示例,可以明显看出nLSTMs+AM模型能够较准确地动态估计出不同特征的重要性(α_e 、 α_t 分别为情感特征和时间特征的注意力分数)。从第1条和第2条记录中可以看出,Maroon 5欢快的音乐经常为用户所喜爱。然而,在第3条记录中,随着用户在午夜变得伤感,播放的音乐转变为抒情的音乐。在这种情况下,用户对音乐的选择同时受到情感和时间的影响,而注意力模型可以以不同的注意力分数 α_e 、 α_t 自适应地预测情况。

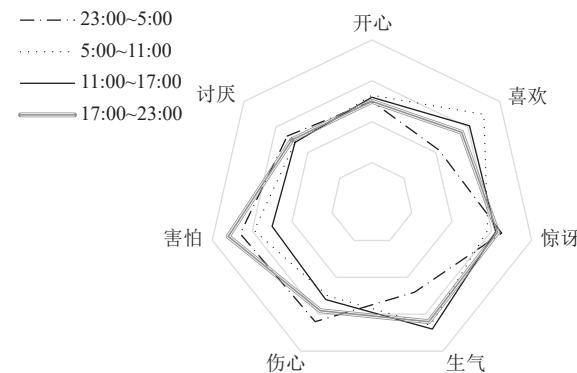


图4 音乐情感特征分析

Fig.4 Analysis of the emotional characteristics of music

表3 用户音乐行为示例

Table 3 Examples of the user music behaviors

时间段	交互音轨	音乐情感类型	注意力分数	
			α_e	α_t
2012-12-17 19:25	Lucky Strick	开心	0.69	0.31
2012-12-19 15:10	Sugar	喜欢	0.88	0.12
2012-12-25 23:46	Five Hundred Miles	悲伤	0.42	0.58

3.2.4 历史序列长度的影响分析

nLSTMs+AM模型的一个关键贡献是考虑了历史信息对用户长期偏好进行建模。该部分实验根据用户历史轨迹的数量将他们的轨迹分成7组,第1组用户拥有4条或更少的历史轨迹,而拥有10条或更多历史轨迹的用户属于最后一组,然后将每一组作为模型输入分别评估性能,结果如图5所示。

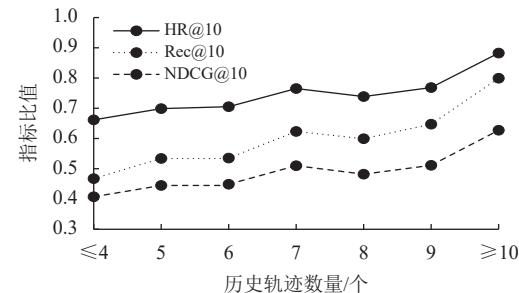


图5 历史序列长度影响

Fig.5 Influence of the historical sequence length

图5显示了随着历史序列长度的增长,模型性能的整体上升趋势。因为揭示人的周期性规律需要时间,历史序列长度越长,本文的长期偏好建模模块可以提取的信息就越多。

4 总结

本文提出了一个新的音乐推荐模型,即nLSTMs+AM。应用多LSTM网络对用户长短期偏好进行建模,应用注意力机制对音乐情感因子进行分析。从实

验结果可以看出,通过深度学习框架进行特征提取和基于历史信息与当前情境相关性的长短期偏好学习,可以构建具有良好性能的音乐推荐系统。经实验验证本文提出的音乐推荐模型有效可行。值得注意的是,在音乐推荐中,音乐情感不仅能通过声学信号表现出来,还可以通过音乐评论等信息体现出来,同时,用户情绪也在短期内影响用户音乐偏好。在未来的研究中,计划通过引入音乐评论及用户实时分享(如音乐软件社交动态分享)等文本信息来解决这些问题。

参考文献:

- [1] SUN K, QIAN T, CHEN T, *et al*. Where to go next: modeling long-and short-term user preferences for point-of-interest recommendation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2020: 214-221.
- [2] BOGDANOV D, HARO M, FUHRMANN F, *et al*. Semantic audio content-based music recommendation and visualization based on user preference examples [J]. *Information Processing & Management*, 2013, 49(1): 13-33.
- [3] OORD A, DIELEMAN S, SCHRAUWEN B. Deep content-based music recommendation [J]. *Advances in Neural Information Processing Systems*, 2013, 26: 2643-2651.
- [4] 彭嘉恩, 邓秀勤, 刘太亨, 等. 融合社交和标签信息的隐语义模型推荐算法[J]. *广东工业大学学报*, 2018, 35(4): 45-50. PENG J E, DENG X Q, LIU T H, *et al*. A recommendation algorithm of latent factor model fused with the social and tag information [J]. *Journal of Guangdong University of Technology*, 2018, 35(4): 45-50.
- [5] CHENG R, TANG B. A music recommendation system based on acoustic features and user personalities[C]//Trends and Applications in Knowledge Discovery and Data Mining. Aucland: Springer Cham, 2016: 203-213.
- [6] CHENG Z, SHEN J, NIE L, *et al*. Exploring user-specific information in music retrieval[C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: Association for Computing Machinery, 2017: 655-664.
- [7] CHEN X. The application of neural network with convolution algorithm in western music recommendation practice [J]. *Journal of Ambient Intelligence and Humanized Computing*, 2020, 11(3): 1-11.
- [8] ORAMAS S, OSTUNI V C, NOIA T D, *et al*. Sound and music recommendation with knowledge graphs [J]. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2016, 8(2): 1-21.
- [9] LEE J, LEE K, PARK J, *et al*. Deep content-user embedding model for music recommendation[EB/OL]. arXiv: 1807.06786(2018-07-18) [2022-01-13].<https://doi.org/10.48550/arXiv.1807.06786>.
- [10] SACHDEVA N, GUPTA K, PUDI V. Attentive neural architecture incorporating song features for music recommendation[C]//Proceedings of the 12th ACM Conference on Recommender Systems. New York: Association for Computing Machinery, 2018: 417-421.
- [11] 杨明极, 刘畅, 宋泽. 注意力机制与改进RNN的混合音乐推荐算法研究[J]. *小型微型计算机系统*, 2020, 41(10): 2235-2240. YANG M J, LIU C, SONG Z. Research on music recommendation algorithm based on attention mechanism and improved RNN [J]. *Journal of Chinese Computer Systems*, 2020, 41(10): 2235-2240.
- [12] WANG D, DENG S, ZHANG X, *et al*. Learning to embed music and metadata for context-aware music recommendation [J]. *World Wide Web*, 2018, 21(5): 1399-1423.
- [13] SHEN T, JIA J, LI Y, *et al*. PEIA: Personality and emotion integrated attentive model for music recommendation on social media platforms[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2020: 206-213.
- [14] DHAHRI C, MATSUMOTO K, HOASHI K. Mood-Aware music recommendation via adaptive song embedding[C]// 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) . Barcelona: IEEE, 2018: 135-138.
- [15] ABDUL A, CHEN J, LIAO H Y, *et al*. An emotion-aware personalized music recommendation system using a convolutional neural networks approach [J]. *Applied Sciences*, 2018, 8(7): 1103.
- [16] BHAUMIK M, ATTAH P U, JAVED F. Emotion integrated music recommendation system using generative adversarial networks [J]. *SMU Data Science Review*, 2021, 5(3): 4.
- [17] 赵永建, 杨振国, 刘文印. 基于双向条目注意网络的推荐系统[J]. *广东工业大学学报*, 2020, 37(4): 27-34. ZHAO Y J, YANG Z G, LIU S Y. DIAN: dual-aspect item attention network for item-based recommendation [J]. *Journal of Guangdong University of Technology*, 2020, 37(4): 27-34.
- [18] GUO H, TANG R, YE Y, *et al*. DeepFM: a factorization-machine based neural network for CTR prediction[EB/OL]. arXiv: 1703.04247(2017-03-13) [2022-01-13].<https://doi.org/10.48550/arXiv.1703.04247>.
- [19] XIAO J, YE H, HE X, *et al*. Attentional factorization machines: learning the weight of feature interactions via attention networks[EB/OL]. arXiv: 1708.04617(2017-08-15) [2022-01-13].<https://doi.org/10.48550/arXiv.1708.04617>.
- [20] JIN Y, HAN C. A music recommendation algorithm based on clustering and latent factor model[C]//2019 International Conference on Computer Science Communication and Network Security. Sanya: EDP Sciences, 2020: 1-9.