

基于多层注意力表示的音乐推荐模型

李 琳, 唐守廉

(北京邮电大学经济管理学院, 北京 100876)

摘 要: 目前的音乐推荐方法只能挖掘用户与歌曲之间的一般性关系,无法区分不同用户对同一首歌曲的差异化偏好.为此,本文提出了基于多层注意力表示的音乐推荐模型,利用用户属性信息和歌曲内容信息从多维度学习歌曲表征,挖掘用户与歌曲之间的偏好关系.为了区分用户对歌曲多域特征的差异化偏好,设计了用户特征依赖的注意力网络;为了区分不同历史行为对用户偏好的差异性,挖掘用户行为的时序依赖关系,设计了歌曲依赖的注意力网络.最后,利用 Softmax 函数计算用户对候选歌曲的偏好分布并产生推荐.在 30Music 和 MIGU 数据集上的实验结果表明,相比目前的推荐模型,本文提出的模型在 Recall 和 MRR 均得到了显著提升.

关键词: 特征表示; 注意力网络; 时序关系; 音乐推荐

中图分类号: TP311 **文献标识码:** A **文章编号:** 0372-2112 (2020)09-1672-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.09.002

Hierarchical Attention Representation Model for Music Recommendation

LI Lin, TANG Shou-lian

(School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Current music recommendation models can mine the general preference between users and music, which are unable to distinguish the differential preference of different users towards the same song. Therefore, we propose a hierarchical attention representation model (HARM) to improve the music recommendation quality. HARM utilizes attributes of users and contents of music to learn music representation from the perspective of multi-dimension and mine the preference relationship between user and music. In order to mine users' differential preference on music's multi-field feature, a user feature-dependent attention network is designed. In addition, in order to mine the impacts of different historical behavior on user preference and learn sequential dependency of user's, a music-dependent attention network is designed. Finally, recommendation is generated by using a softmax function to calculate the preference distribution of users on candidate songs. The experimental results on 30Music and MIGU datasets shows that, comparing with the existent recommendation models, HARM can gain significant improvement on Recall and MRR.

Key words: feature representation; attention network; sequential relationship; music recommendation

1 引言

流媒体音乐的普及给用户带来了极大的自由度和对音乐内容的无限访问,如 Last.fm、Pandora、QQ 音乐、网易云音乐和咪咕音乐等.对这些音乐平台来说,如何将歌曲推荐给目标用户,让用户从海量曲库中快速找到自己喜爱的歌曲,提升用户的收听体验,是面临的主要难题之一.近年来,随着海量用户收听行为数据的采集,推荐系统在社会化多媒体平台中得到了成功应用,

因此,流媒体音乐平台也不同程度地引入了推荐系统,根据用户收听行为挖掘用户的音乐偏好,为用户推荐可能喜欢的歌曲.

基于近邻的协同过滤是最早被用于歌曲推荐中的推荐算法,依据计算用户或歌曲在行为上的近邻关系产生推荐,取得了较好的推荐效果.然而,在流媒体音乐平台中的用户行为数据通常服从长尾分布,20%的热门歌曲聚集了 80% 的收听行为^[1],导致平台在特定场景下给用户推荐了大量不喜欢的热门歌曲,即流行度偏

差,而那些很少被用户收听或新上架的歌曲很难被推荐给目标用户.此外,相比于海量的音乐曲库来说,用户收听过的歌曲数量是极少的,容易造成用户行为数据的稀疏性问题.为了解决这些问题,研究者提出了基于模型的协同过滤方法^[2,3],利用中间隐含因子矢量连接用户和歌曲,通过点积操作预估用户对歌曲的喜好程度.虽然这种方法在一定程度上减少了数据稀疏性问题和流行度偏差,但仍是基于用户单一交互行为的基础上,忽略了用户对歌曲内容特征的偏好差异.

流媒体音乐平台除了包含大量用户行为数据外,还记录了歌曲的元数据,如歌手、歌词、流派、封面和音频等,可以利用这些元数据设计出新的推荐模型,提高模型的推荐精度.内容过滤方法^[4]弥补了协同过滤的不足,通过挖掘用户收听过的歌曲内容特征推荐与之相似的歌曲.而内容过滤忽略了用户之间的“协同规律”,因为相似的歌曲具备相似的性质,相似的用户也具有相似的喜好.因此,如何充分结合两类方法的优势构造更加高效的推荐算法成为学术界和工业界亟待解决的难题.早期粗暴式的融合模型是以贪婪策略选择用户预期评分最高的歌曲产生较为安全的推荐列表,这种策略会带来长期的次优问题,因为用户的音乐偏好是基于现有的知识进行估计,忽略了那些未知知识对用户偏好的影响,比如时序依赖关系、上下文信息等.近年来,国内外研究者又提出了基于会话和上下文的混合推荐算法^[5,6],充分挖掘未知知识对音乐偏好的影响,提升推荐质量.这些方法只从全局层面考虑用户在当前会话中的上下文信息和时序关系,缺乏对细粒度特征的理解.比如,用户对收听过的歌曲列表中不同歌曲具有不同的兴趣点,且不同用户对同一首歌曲所关注的内容也存在差异.

为此,本文提出了基于多层注意力表示的推荐模型(HARM),利用用户属性信息和歌曲内容信息,从多维度学习歌曲嵌入式表示,挖掘用户对歌曲的偏好特征.具体来说,为了区分用户对歌曲多域特征的差异性偏好,设计了用户特征依赖的注意力网络;为了区分不同历史行为对用户偏好的差异性,并挖掘行为的时序依赖关系,设计了歌曲依赖的注意力网络.最后,利用Softmax函数计算用户对候选歌曲的偏好分布,并产生推荐.此外,为了能够提取音频内容特征,利用歌曲的流派标签数据预训练CNN网络,然后将其与推荐任务结合进行监督学习优化,用于学习歌曲音频表征.

2 相关工作

基于协同过滤和内容过滤的模型都是依赖用户与物品的交互行为数据,这就导致这类模型只能关注用户的长期、静态偏好.用户的行为序列中存在着极强的

关联性和因果性.序列挖掘是最早挖掘时序关系的方法,通过挖掘会话中物品之间的顺序关系推断用户的未来行为.马尔可夫链模型是最为经典的序列关系挖掘方法^[7],通过计算行为之间的单步转移概率预测用户的下一个行为.用户行为往往是受到多步依赖关系的影响,为了挖掘这种影响关系,高阶马尔可夫链模型^[8]考虑了更多先前行为的依赖关系.然而,由于多步依赖计算过程中状态数量较多,导致过大的计算量.此外,用户行为数据又是极为稀疏,导致不精确的推荐结果.为了解决多行为下计算复杂度问题,Tavakol等^[9]提出了基于物品属性序列的状态空间模型,缓解了状态空间爆炸的问题;Twardowski等^[10]提出了基于矩阵分解的时序推荐模型,根据会话和候选项之间的交互生成下一个推荐项;Kabbur等^[11]提出了FISM模型,对物品-物品的共线矩阵进行矩阵分解,不需要学习显式的用户表征.总的来说,这些建立在矩阵分解基础上的马尔可夫模型,只考虑了低阶情况下的作用关系,无法解决更高阶情况下的相互作用,同时,忽略了会话内和不同会话之间的行为顺序依赖关系.

近年来,循环神经网络(RNN)在自然语言处理领域中取得了非常成功的应用^[12],这种网络结构在处理时序数据时具有一定的记忆能力,有助于捕获句子中的上下文关系.Hidasi等^[13]首次将RNN应用到时序推荐任务中,设计了基于门控递归单元(GRU)的RNN模型(GRU4Rec).Tan等^[14]在GRU4Rec基础上通过对序列预处理和嵌入Dropout进行数据增强减少过拟合的问题.Quadrana等^[15]提出了分层的RNN模型,将跨会话信息纳入推荐模型中,捕获会话内和会话间的依赖关系.Chatzis等^[16]将变分推理加入到RNN中,以处理稀疏数据中带来的不确定性.除了RNN结构外,卷积神经网络(CNN)也被用于时序推荐任务.比如,Wu等^[17]提出了基于会话感知的推荐方法,将用户的不同类型行为组合并学习到一个会话表示,然后使用深度排序模型进行重新排序;Song等^[18]提出了基于深度神经网络的会话推荐模型,同时对静态和动态用户偏好进行建模;Tang等^[19]将最近的动作建模为时间和潜在维度之间的“图像”,使用CNN学习序列模式作为图像的局部特征,使得模型能够预测用户在“不久的将来”可能交互的前 N 个物品.与传统模型相比,基于深度学习的推荐方法可以挖掘用户行为的长短期依赖关系.然而,这些方法只能挖掘全局依赖关系,无法区分不同历史行为对用户偏好的差异性.

基于对人类视觉注意力特性而衍生的注意力机制在图像识别、自然语言处理等领域中得到了成功应用^[20].研究者也将其引入到推荐系统中,区分不同历史行为对用户偏好影响的差异性,凸显出那些对用户偏

好最相关的行为. 比如, Li 等^[21] 提出基于会话的时序推荐的编码-解码器框架, 通过注意机制消除来自意外行为的噪音影响; Liu 等^[22] 提出了短期记忆优先模型, 利用注意机制计算序列中用户交互的物品注意分数以及序列中最近物品之间的注意力相关性; Sachdeva 等^[23] 将注意力机制与 Bi-GRU 网络结合, 对用户的短期兴趣建模; Zhang 等^[24] 利用自注意机制从用户的历史交互中推断物品之间的关系, 得到用户交互轨迹中每个物品的权重. 这些方法均是基于用户-物品的交互行为数据, 在一定程度上提升了推荐模型的质量, 却忽略了用户属性和歌曲内容元数据的影响.

3 基于多层注意力表示的推荐模型

本文通过建立基于多层注意力表示的推荐模型, 提升歌曲推荐的质量. 所提出的模型包括歌曲编码、用户编码和预测三部分. 首先, 构建基于用户的注意力网络学习歌曲表征; 然后根据用户收听的歌曲表征, 构建基于歌曲的注意力网络学习用户偏好表征; 最后利用学习到的用户偏好表征和歌曲表征预测用户对候选歌曲的偏好概率, 并将得分较高的前 N 首歌曲推荐给用户. HARM 模型的框架设计如图 1 所示.

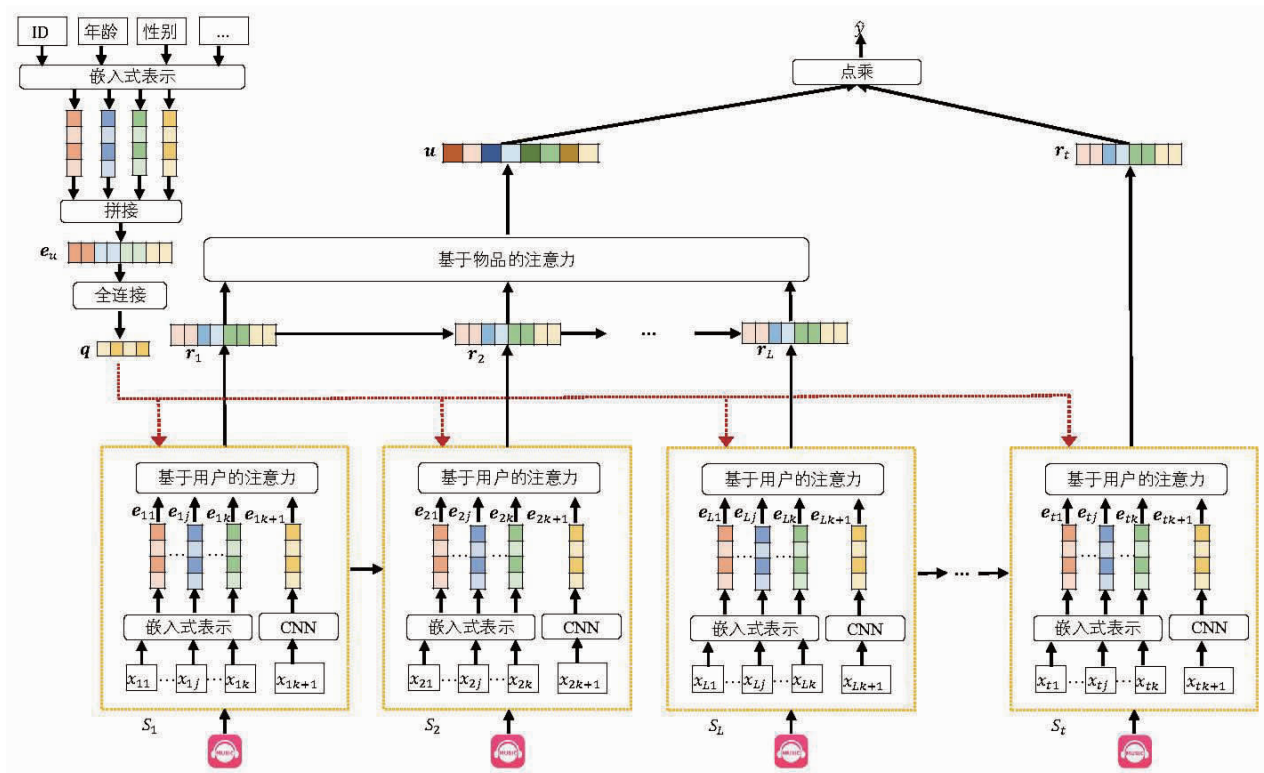


图1 HARM模型框架

3.1 歌曲编码

目前, 对歌曲进行的编码方法通常是利用嵌入式表示技术, 将用户收听歌曲的独热表示映射成稠密的语义特征向量, 并将其作为歌曲的编码特征. 这种粗粒度编码方法无法刻画歌曲的细粒度特征, 从而导致不精确的歌曲推荐. 用户对一首歌曲的偏好通常是受多域特征的影响, 例如, 歌曲的流派、歌手及音频旋律等. 此外, 不同用户往往会对同一首歌曲的偏好也呈现不同的关注度. 因此, 需要从细粒度层面区分用户对歌曲特征的差异性偏好. 为此, HARM 模型中的歌曲编码由三部分构成, 即离散特征的嵌入式表示、音频语义特征的提取和基于用户的注意力网络.

对于歌曲 i 的第 j 个域的离散特征值 x_{ij} , 首先利用

独热表示进行编码生成高维稀疏特征向量 $f_{ij} \in \mathbb{R}^{1 \times |f_j|}$, $|f_j|$ 表示第 j 个域的离散取值个数. 利用嵌入式表示技术将稀疏特征向量 f_{ij} 映射到稠密的语义空间, 即 $e_{ij} = f_{ij} E_j$, $E_j \in \mathbb{R}^{|f_j| \times N}$ 是第 j 个域的嵌入式表示矩阵, 通过模型训练优化得到, N_j 表示嵌入式维度. 对于音频内容, 本文用 x_{ik+1} 表示, 首先利用离散傅里叶变换将音频信号转化为频域, 为每首歌曲生成对应的彩色声谱, 声谱的深浅表示该频率下的声音大小, 长宽表示时间和频率. 由于每首歌曲的播放时间不同, 将每首歌曲的彩色声谱图分割成多个 $256 \times 256 \times 3$ 的图片. 最后, 采用迁移学习方式, 利用基于 ImageNet 数据集预训练好的 VGG16 网络模型从歌曲音频中提取语义内容表示 e_{ik+1} .

在实际场景下,不同用户对同一首歌曲的关注角度存在较大差异,比如,一些用户关注歌曲的演唱者,另一些用户关注歌曲的风格或旋律.此外,即使同一首歌的相同特征对不同用户偏好的贡献也是存在显著差异.因此,区分歌曲的不同域特征对用户的偏好贡献也可以提升歌曲推荐的质量.为了能够区分歌曲各个域的特征对不同用户推荐的重要性,本文提出了基于用户的注意力网络,基于用户的特征表示识别并强化不同域特征的贡献.用户的特征表示是由用户的多个特征所构成的,为此,首先将用户的ID、年龄、性别等特征映射到稠密的表示空间,然后进行拼接得到用户嵌入特征向量 $\mathbf{e}_u \in \mathbb{R}^{1 \times D}$, D_e 表示特征向量的维度.然后利用全连接层计算用户偏好向量 $\mathbf{q} \in \mathbb{R}^{D_e \times 1}$, D_q 表示用户偏好向量的维度.计算公式为:

$$\mathbf{q} = \text{ReLU}(\mathbf{V} \times \mathbf{e}_u + \mathbf{b}_u) \quad (1)$$

其中, $\mathbf{V} \in \mathbb{R}^{D_e \times D}$, $\mathbf{b}_u \in \mathbb{R}^{D_e \times 1}$ 为模型参数.

在该网络中,将用户偏好向量 \mathbf{q} 作为查询,歌曲中每个域的嵌入式表示作为键-值对,利用注意力机制计算每个域特征的权重.具体来说,歌曲 i 的第 j 个域的权重 a_j 的形式化计算方法为:

$$a_j = \frac{\exp(\tilde{a}_j)}{\sum_{k=1}^{K+1} \exp(\tilde{a}_k)} \quad (2)$$

$$\tilde{a}_j = \mathbf{e}_{ij} \tanh(\mathbf{W}_q \times \mathbf{q} + \mathbf{b}_q) \quad (3)$$

其中 K 为域的个数, $\mathbf{W}_q \in \mathbb{R}^{N_j \times D_e}$, $\mathbf{b}_q \in \mathbb{R}^{N_j \times 1}$ 为映射参数.第 i 首歌曲的特征表示 \mathbf{r}_i 可以表示为基于注意力权重的不同域特征加权,表示为:

$$\mathbf{r}_i = \sum_{k=1}^{K+1} a_k \mathbf{e}_{ik} \quad (4)$$

对用户收听序列中的所有歌曲按照上述方法进行编码,从而得到用户收听序列的歌曲特征表示序列为 $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_L]$, L 表示序列长度, $\mathbf{R} \in \mathbb{R}^{N_j \times L}$.

3.2 用户编码

用户编码的目的是从用户收听的音乐特征表示中学习用户的偏好表征.很多方法可以用来学习用户的偏好表征,包括基于加权池化的方法^[25]、基于注意力机制的方法^[26]和基于循环神经网络的方法^[14]等.本文利用基于用户的注意力网络学习用户偏好表征.具体来说,利用自注意力网络学习用户收听歌曲之间的细粒度时序依赖关系.为了能够捕获用户收听行为的时序依赖关系,对用户收听序列中的每个位置 l 进行编码,用 \mathbf{c}_l 表示,则收听序列的所有位置编码用矩阵 $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_L]$ 表示, $\mathbf{C} \in \mathbb{R}^{N_j \times L}$.因此,用户收听序列中带有位置编码的歌曲表征可表示为 $\mathbf{F} = \mathbf{C} \oplus \mathbf{R}$,其中, \oplus 为元素级别的加法操作, $\mathbf{F} \in \mathbb{R}^{N_j \times L}$.

将序列中的歌曲特征表示 \mathbf{F} 作为自注意力机制的

查询 \mathbf{Q} 、键 \mathbf{K} 和值,利用全连接网络将 \mathbf{F} 进行非线性变换,则音乐收听行为之间的关系权重矩阵表示为:

$$\mathbf{A} = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{N_j}}\right) \quad (5)$$

$$\mathbf{Q} = \eta(\mathbf{F}^T \mathbf{W}^Q + \mathbf{B}_q) \quad (6)$$

$$\mathbf{K} = \eta(\mathbf{F}^T \mathbf{W}^K + \mathbf{B}_k) \quad (7)$$

其中, $\eta(\cdot)$ 为非线性激活函数,本文利用 ReLU 函数增加非线性能力; $\mathbf{W}^Q \in \mathbb{R}^{N_j \times N_j}$, $\mathbf{W}^K \in \mathbb{R}^{N_j \times N_j}$, $\mathbf{B}_q \in \mathbb{R}^{L \times N_j}$, $\mathbf{B}_k \in \mathbb{R}^{L \times N_j}$ 为模型参数; $\sqrt{N_j}$ 对点乘操作进行缩放.式(5)的输出为 $L \times L$ 的权重矩阵.在此基础上,注意力网络的输出为 $L \times N_j$ 的矩阵,定义为:

$$\mathbf{O} = \mathbf{A}\mathbf{F}^T \quad (8)$$

为了能够得到用户全局偏好表示,本文将式(8)在每个时刻的行为表示进行拼接.然后利用全连接网络学习用户全局偏好表示,表示为:

$$\mathbf{u}_i = \text{concat}(\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_L) \mathbf{W}^U \quad (9)$$

其中, concat 为向量拼接函数, $\mathbf{W}^U \in \mathbb{R}^{L \times N_j}$.

3.3 预测与训练

对于候选集中的目标歌曲 S_i ,首先利用3.1中的方法对目标歌曲 t 进行编码,得到歌曲的编码表示 \mathbf{r}_t .本文利用 Softmax 函数计算用户对目标歌曲的偏好概率,表示为:

$$P(S_i | \mathbf{u}_i) = \frac{\exp(\mathbf{u}_i \mathbf{r}_i^T)}{\sum_t \exp(\mathbf{u}_i \mathbf{r}_t^T)} \quad (10)$$

其中, $|V|$ 是候选集中歌曲数量, \mathbf{u}_i 是用户偏好表征向量.

给定训练集 $\chi = \{\langle H_i^L, \mathbf{u}_i, S_i \rangle\}$,联合概率分布可表示为:

$$P_\theta(\chi) = \prod_{x \in \chi} \frac{\exp(\mathbf{u}_i \mathbf{r}_i^T)}{\sum_t \exp(\mathbf{u}_i \mathbf{r}_t^T)} \quad (11)$$

参数 $\theta = \{\mathbf{E}, \mathbf{V}, \mathbf{W}_q, \mathbf{W}^Q, \mathbf{W}^K, \mathbf{W}^U, \mathbf{b}_q, \mathbf{B}_q, \mathbf{B}_k\}$ 可以通过最小化条件 log 似然概率学习得到,即:

$$\mathcal{L} = \sum_{x \in \chi} P(S_i | \mathbf{u}_i) - \lambda \theta \quad (12)$$

直接对式(10)进行优化是非常耗时的,因为需要对所有歌曲计算 Softmax 函数.因此,借鉴工作^[27],本文利用负采样技术提升模型优化速度,同时近似接近原始的目标函数 \mathcal{L} .改造后的目标损失函数为:

$$\mathcal{L} = \sum_{x \in \chi} \log \sigma(\mathbf{u}_i \mathbf{r}_i^T) + k \cdot \mathbb{E}_{G_l} [\log \sigma(\mathbf{u}_i \mathbf{r}_l^T)] - \lambda \theta \quad (13)$$

其中, σ 是激活函数, k 是根据噪声分布 G_l 采样到的负样本个数, G_l 是用用户收听歌曲分布频率表示.最后利用随机梯度下降法对模型进行训练.

4 实验部分

4.1 实验数据

本文选取了 30Music 和 MIGU 两个数据集验证所提出的模型. 30Music 数据集^[1]是基于 Last. fm 扩展的数据集, 该数据集包含了用户、歌曲的属性信息, 以及完整的会话信息. MIGU 数据集是咪咕音乐提供的私有数据集, 包含模型所需要的歌曲元数据和用户特征数据. 为了便于实验, 对原始数据集进行清洗. 对于 30Music 数据集, 在原始会话信息中删除那些缺少用户或歌曲属性信息的播放事件, 只保留播放事件大于 10 的会话信息. 然后在 Last. fm 上抓取歌曲对应的音频文件. 最终生成 313941 个会话, 4267295 条播放记录, 包含 13535 用户和 27302 首歌曲, 平均每个会话的播放事件为 13 个. 对于 MIGU 数据集, 只保留播放事件中用户或歌曲属性信息完整的记录, 其次, 保证会话中相邻播放事件的时间间隔小于 200s, 并保留了收听时长小于 30% 或者 20s 的播放事件. 最终生成 1374236 个会话, 17621958 个播放记录, 包括 41407 用户和 1926590 首歌曲, 平均每个会话播放事件为 12 个.

4.2 对比方法

本文选取以下几种先进的推荐算法与提出的模型进行比较.

BPR^[3]: 是利用用户隐式反馈数据构造用户偏好的偏序关系对, 通过贝叶斯网络学习用户和物品潜在特征向量, 是一种特殊的概率矩阵分解方法.

GRU4Rec^[13]: 首次将 RNN 应用于推荐系统领域, 模型引入了会话并行、基于小批量的输出采样和排序损失函数.

Improved-GRU4Rec^[14]: 是在 GRU4Rec 模型基础上增加了数据扩充和考虑用户行为的时间变化等策略来提高模型性能.

NARM^[21]: 是一种混合编码器, 利用注意机制对用户的连续行为进行建模, 从当前会话中捕获用户的意图, 然后将意图组合为一个统一的会话表示计算每个候选选项的推荐分数.

SASRec^[26]: 基于自注意力的序列模型, 该模型首先识别用户的操作历史中相关物品, 然后用其预测下一个偏好的物品.

4.3 评估方法

本文采用 Recall 和 MRR 指标来衡量提出的模型和对比模型的推荐效果.

Recall: 定义为推荐列表中的歌曲与用户喜欢的所有歌曲之间的比例.

$$\text{Recall}@N = \frac{1}{|D^{\text{Test}}|} \sum_{u \in D^{\text{Test}}} \frac{|S_u^{\text{pre}} \cap D_u^{\text{Test}}|}{|D_u^{\text{Test}}|} \quad (14)$$

其中, S_u^{pre} 表示为用户 u 推荐的前 N 首歌曲列表, D_u^{Test} 表示测试集中的歌曲列表, $|D_u^{\text{Test}}|$ 表示用户 u 在测试集中的样本个数.

MRR: 由于 Recall 只度量用户喜欢的歌曲是否出现在推荐列表中, 未考虑排名因素. MRR 是一种排序测量指标, 定义为:

$$\text{MRR} = \frac{1}{|U|} \sum_i \frac{1}{\text{rank}_i} \quad (15)$$

其中, $|U|$ 表示用户的个数, rank_i 表示用户 u 的推荐列表中第 i 个歌曲在真实列表中所在的排名位置.

实验在 4 核高性能 NVIDIA GPU 服务器上进行模型训练, 每个 GPU 含有 24GB 显存. 对比方法和本文提出的模型均是基于 Python 的 Tensorflow 框架所实现的, 其模型超参数通过交叉验证法得到.

5 实验结果与分析

5.1 对比分析

为了评估 HARM 模型与对比模型的精度, 分别在 30Music 和 MIGU 数据集上进行了测试, 利用 Recall、MRR 衡量推荐的质量, 推荐歌曲数量 N 的取值从 5 到 30, 间隔为 5. 为了排除单次实验结果的不稳定性因素, 本文重复了 10 次实验, 并取实验结果的平均值作为最终评估结果.

图 2 显示了 HARM 模型与对比模型在 30Music 和 MIGU 数据集上的 Recall 实验结果, 实验结果表明, 与传统模型 (BPR) 相比, Recall@20 在 30Music 数据集上提升了 49%, 在 MIGU 数据集上提升了 44%; 与时序模型 (GRU4Rec、Improved-GRU4Rec) 相比, Recall@20 在 30Music 数据集上分别提升了 44% 和 32%, 在 MIGU 数据集上分别提升了 33% 和 24%; 与注意力模型 (NARM、SASRec) 相比, Recall@20 在 30Music 数据集分别提升了 28% 和 21%, 在 MIGU 数据集上分别提升了 24% 和 17%.

还对模型进行 MRR 指标测评, 各模型在 30Music 和 MIGU 数据集上 MRR 指标的性能表现如图 3 所示, HARM 模型的 MRR 数值均优于对比方法. 与 SASRec 相比, 在 30Music 和 MIGU 数据集上分别提高了 10% 和 5.6%. 这是由于在音乐场景中, 用户偏好往往受多域特征的影响, 不同的用户还会对同一首歌曲的特征也存在不同的关注度. 本文提出的 HARM 模型能够更加细致地区分不同用户对同一首歌曲的偏好差异, 精细刻画歌曲特征表示, 从而提升歌曲推荐的质量.

此外, 从两个数据集的实验结果来看, HARM 模型在 30Music 数据集的指标数值均高于 MIGU 数据集, 这是由于 MIGU 数据集是从真实生产环境采集到的, 数据稀疏度更高, 而 30Music 数据集是经过特定清洗后得到的.

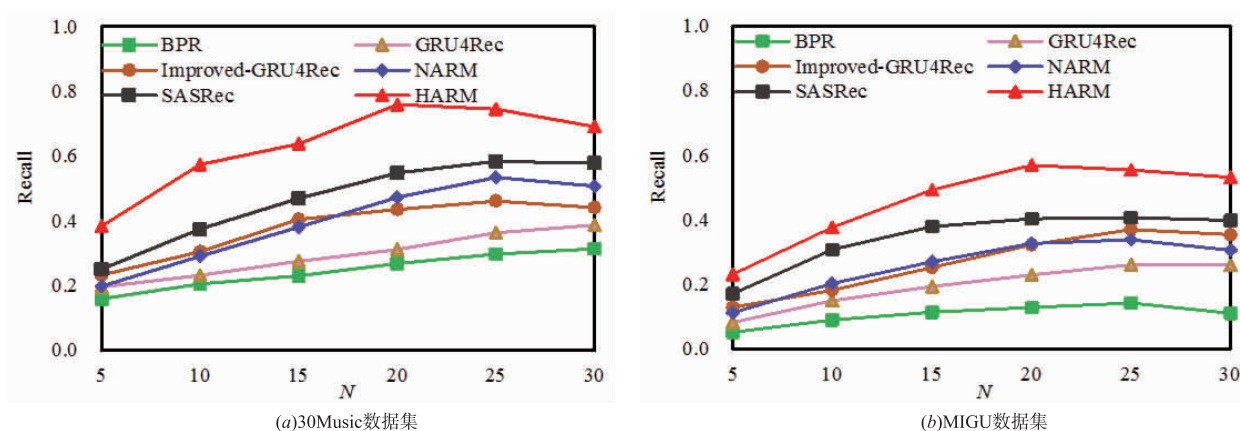


图2 HARM模型与对比模型在30Music和MIGU数据集上的Recall结果

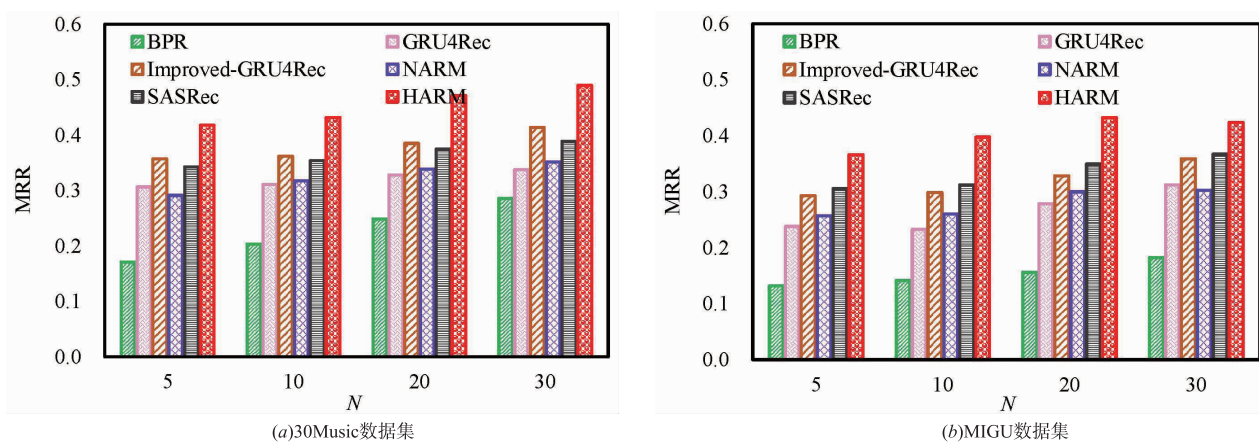


图3 HARM模型与对比模型在30Music和MIGU数据集上的MRR结果

5.2 参数敏感度

模型的潜在因子维数 K 是影响模型准确度的主要参数. 本文在单独的验证集上做交叉验证实验, 寻找最优超参数. 实验将潜在因子维数 K 取 $\{20, 60, 100, 200, 260, 300, 400, 500, 600\}$, 分别计算 $\text{Recall}@5$ 、 $\text{Recall}@10$ 、 $\text{Recall}@20$ 和 $\text{Recall}@30$, 实验结果如图 4 所示.

实验结果表明, HARM 模型的召回率随着 K 的增

大先略有增大, 但当随着 K 值超过 260 时, Recall 值开始呈下降趋势. 可以看出, 当 K 很小时, HARM 模型不能很好地表示用户的音乐偏好和歌曲的语义特征. 但当 K 超过阈值 ($K = 260$) 时, 模型已可以处理各种复杂关系的数据, 但模型的准确性开始下滑. 此时, 增加 K 无疑会提高模型的性能, 但也会增加模型的训练时间, 导致收益递减.

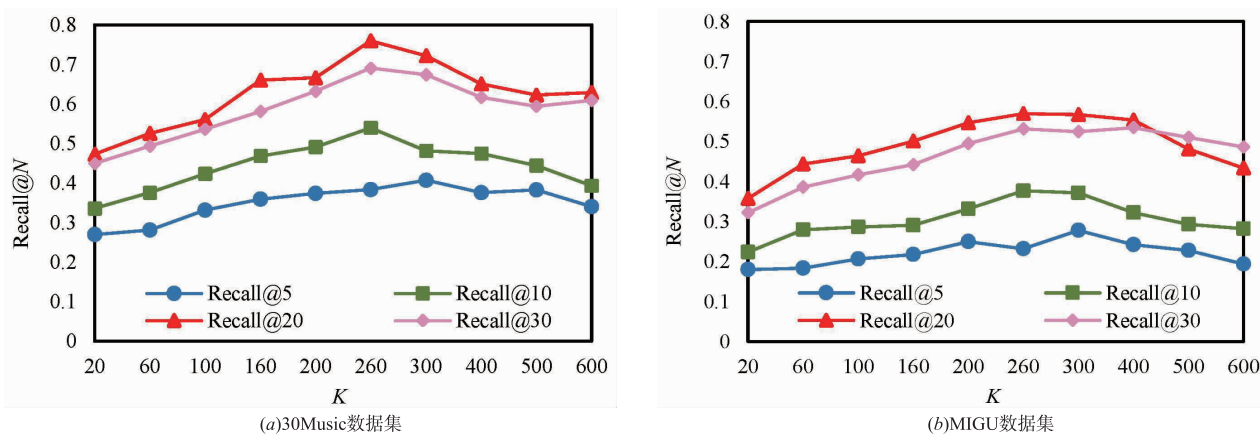


图4 潜在因子 K 对模型的影响

6 结论

本文提出了基于多层注意力表示的音乐推荐模型,该模型构建了基于用户的注意力网络学习歌曲的表征,然后根据用户收听的歌曲表征,构建基于歌曲的注意力网络学习用户的偏好表征,最后利用学习到的用户偏好表征和歌曲表征预测用户对候选歌曲的偏好概率,并将得分较高的前 N 首歌曲推荐给用户. 在 30Music 和 MIGU 数据集上的实验结果表明,本文提出的 HARM 模型在 Recall 和 MRR 上的实验结果均优于对比方法. 在 30Music 和 MIGU 数据集上, HARM 模型的 Recall 指标比 BPR 分别提升了 49% 和 44%;模型的 MRR 指标比排名第二的 SASRec 分别高出了 10% 和 5.6%.

参考文献

- [1] Turrin R, Quadrana M, Condorelli A, et al. 30music listening and playlists dataset[A]. Proceedings of the 9th ACM Conference on Recommender Systems [C]. New York: Association for Computing Machinery, 2015. 1 – 4.
- [2] 任开旭, 王玉龙, 刘同存, 等. 融合多维语义表示的概率矩阵分解模型[J]. 电子学报, 2019, 47(9): 1848 – 1854.
Ren K X, Wang Y L, Liu T C, et al. A probabilistic matrix factorization model based on multidimensional semantic representation learning[J]. Acta Electronica Sinica, 2019, 47(9): 1848 – 1854. (in Chinese)
- [3] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback[A]. Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence[C]. Virginia: AUAI Press, 2009. 452 – 461.
- [4] Van den Oord A, Dieleman S, Schrauwen B. Deep content-based music recommendation[A]. Advances in Neural Information Processing Systems[C]. New York: NIPS, 2013. 2643 – 2651.
- [5] Dias R, Fonseca M J. Improving music recommendation in session-based collaborative filtering by using temporal context[A]. 2013 IEEE 25th International Conference on Tools with Artificial Intelligence [C]. Piscataway: IEEE, 2013. 783 – 788.
- [6] Wang D, Deng S, Xu G. Sequence-based context-aware music recommendation[J]. Information Retrieval Journal, 2018, 21(2 – 3): 230 – 252.
- [7] He R, Kang W C, McAuley J. Translation-based recommendation[A]. Proceedings of the Eleventh ACM Conference on Recommender Systems [C]. New York: Association for Computing Machinery, 2017. 161 – 169.
- [8] He R, McAuley J. Fusing similarity models with markov chains for sparse sequential recommendation [A]. 2016 IEEE 16th International Conference on Data Mining (ICDM) [C]. Washington: IEEE Computer Society, 2016. 191 – 200.
- [9] Tavakol M, Brefeld U. Factored MDPs for detecting topics of user sessions[A]. Proceedings of the 8th ACM Conference on Recommender Systems [C]. New York: Association for Computing Machinery, 2014. 33 – 40.
- [10] Twardowski B. Modelling contextual information in session-aware recommender systems with neural networks [A]. Proceedings of the 10th ACM Conference on Recommender Systems [C]. New York: Association for Computing Machinery, 2016. 273 – 276.
- [11] Kabbur S, Ning X, Karypis G. Fism: factored item similarity models for top-n recommender systems[A]. Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]. New York: Association for Computing Machinery, 2013. 659 – 667.
- [12] 潘博, 于重重, 张青川, 等. 基于词性与词序的相关因子训练的 word2vec 改进模型[J]. 电子学报, 2018, 46(8): 1976 – 1982.
Pan B, Yu C C, Zhang Q C, et al. The improved model for word2vec based on part of speech and word order[J]. Acta Electronica Sinica, 2018, 46(8): 1976 – 1982. (in Chinese)
- [13] Hidasi B, Karatzoglou A, Baltrunas L, et al. Session-based recommendations with Recurrent Neural Networks [OL]. <https://arxiv.org/abs/1511.06939>, 2015.
- [14] Tan Y K, Xu X, Liu Y. Improved recurrent neural networks for session-based recommendations[A]. Proceedings of the 1st Workshop on Deep Learning for Recommender Systems [C]. New York: Association for Computing Machinery, 2016. 17 – 22.
- [15] Quadrana M, Karatzoglou A. Personalizing session-based recommendations with hierarchical recurrent neural networks[A]. Proceedings of the Eleventh ACM Conference on Recommender Systems [C]. New York: Association for Computing Machinery, 2017. 130 – 137.
- [16] Chatzis S P, Christodoulou P, Andreou A S. Recurrent latent variable networks for session-based recommendation [A]. Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems [C]. New York: Association for Computing Machinery, 2017. 38 – 45.
- [17] Wu C, Yan M. Session-aware information embedding for e-commerce product recommendation[A]. Proceedings of the 2017 ACM Conference on Information and Knowledge Management [C]. New York: Association for Computing Machinery, 2017. 2379 – 2382.
- [18] Song Y, Elkahky A M, He X. Multi-rate deep learning for

- temporal recommendation [A]. Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval [C]. New York; Association for Computing Machinery, 2016. 909 – 912.
- [19] Tang J, Wang K. Personalized top-n sequential recommendation via convolutional sequence embedding [A]. Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining [C]. New York; Association for Computing Machinery, 2018. 565 – 573.
- [20] 冯建周, 宋沙沙, 王元卓, 等. 基于改进注意力机制的实体关系抽取方法 [J]. 电子学报, 2019, 47 (8): 1692 – 1700.
Feng J Z, Song S S, Wang Y Z, et al. Entity relation extraction based on improved attention mechanism [J]. Acta Electronica Sinica, 2019, 47 (8): 1692 – 1700. (in Chinese)
- [21] Li J, Ren P, Chen Z, et al. Neural attentive session-based recommendation [A]. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management [C]. New York; Association for Computing Machinery, 2017. 1419 – 1428.
- [22] Liu Q, Zeng Y, Mokhosi R, et al. STAMP: short-term attention/memory priority model for session-based recommendation [A]. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining [C]. New York; Association for Computing Machinery, 2018. 1831 – 1839.
- [23] Sachdeva N, Gupta K, Pudi V. Attentive neural architecture incorporating song features for music recommendation [A]. Proceedings of the 12th ACM Conference on Recommender Systems [C]. New York; Association for Computing Machinery, 2018. 417 – 421.
- [24] Zhang S, Tay Y, Yao L, et al. Next Item Recommendation with Self-Attention [OL]. <https://arxiv.org/abs/1808.06414v2>, 2018.
- [25] Zhou G, Zhu X, Song C, et al. Deep interest network for click-through rate prediction [A]. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining [C]. New York; Association for Computing Machinery, 2018. 1059 – 1068.
- [26] Kang W C, McAuley J. Self-attentive sequential recommendation [A]. 2018 IEEE International Conference on Data Mining (ICDM) [C]. Washington; IEEE Computer Society, 2018. 197 – 206.
- [27] Goldberg Y, Levy O. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method [OL]. <https://arxiv.org/abs/1402.3722v1>, 2014.

作者简介



李琳 (通讯作者) 男, 1977 年出生于天津, 现为北京邮电大学博士研究生, 主要研究为数据挖掘, 推荐算法, 服务科学与工程。
E-mail: li_mg@outlook.com



唐守廉 男, 1952 年出生于上海, 北京邮电大学经济管理学院教授、博士生导师, 主要研究为政府电信管制政策研究, 企业发展战略研究, 企业市场营销策划。
E-mail: tangshoulian@263.net