

# 融合内容表示的度量排序学习推荐模型

李 琳, 唐守廉

(北京邮电大学经济管理学院, 北京 100876)

**摘 要:** 融合内容语义信息的推荐模型可以有效缓解音乐推荐系统中的数据稀疏性和冷启动问题. 然而, 这些模型是通过最小化预测评分误差学习用户与音乐的全局关系, 忽略了用户和音乐隐式特征的细粒度差异. 此外, 内容语义特征是以推荐任务无关的无监督学习方式提取的, 从而导致不精确推荐. 为此, 本文提出了融合内容表示的度量排序学习推荐模型, 该模型是以个性化排序最优化为目标的概率图模型, 利用度量学习从全局和细粒度层面挖掘用户音乐偏好. 为了解决冷启动推荐问题, 本文建立了与推荐任务相关的监督学习策略训练内容语义特征提取模型. 在 KKBOX 和 MIGU 数据集上的实验结果表明, 提出的模型显著提升了冷启动音乐推荐的效果, 在不同稀疏度数据集上的鲁棒性得到了显著增强.

**关键词:** 内容表示; 度量学习; 排序学习; 音乐推荐

**中图分类号:** TP311 **文献标识码:** A **文章编号:** 0372-2112 (2020)08-1615-08

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.08.021

## Metric Ranking Learning Recommendation Model Based on Content Representation

LI Lin, TANG Shou-lian

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

**Abstract:** A model that incorporates semantic information can alleviate the sparse data and cold start problems in a music recommendation system. Current models learn the global relations between users and music by minimizing prediction score error. However, they ignore the fine-grained differences between the implicit features of users and music. Current models also extract semantic features by unsupervised learning that is irrelevant to recommending, leading to inaccurate recommendations. We propose a metric-ranking-learning recommendation model that incorporates content representation (CAM-RL). This model is a probabilistic graphical model that optimizes a personalized ranking and explores user music preferences through metric learning at both global and fine-grained levels. To solve the cold start recommendation problem, a supervised learning strategy in relation to the recommendation task is proposed to train the model of content semantic feature extraction. The results of trials using the KKBOX and MIGU datasets show that the proposed model significantly improves cold start music recommendations when compared with other algorithms; it is also more robust when using sparse datasets.

**Key words:** content representation; metric learning; ranking learning; music recommendation

### 1 引言

随着移动智能终端的普及和无线网络技术的发展, 诞生了各式各样的音乐服务平台, 如 Last.fm、Pandora、QQ 音乐、咪咕音乐等. 对于音乐服务平台来说, 如何将音乐推荐给目标用户, 让用户从海量曲库中找到自己感兴趣的音乐, 是提升用户收听体验的难题之一<sup>[1-3]</sup>. 推荐系统作为移动互联网时代非常重要的信

息检索工具, 也被应用到各大音乐服务平台, 通过推荐功能帮助用户发掘感兴趣的音乐. 近年来, 音乐服务平台采集到的大量用户行为数据为用户音乐偏好的研究提供了前所未有的契机, 极大地提升了音乐推荐的质量<sup>[1,2]</sup>.

协同过滤<sup>[4]</sup>作为经典的推荐算法, 最早被引入到音乐推荐系统中, 该算法通过计算用户-用户/物品-物品之间的相似性推断用户偏好. 然而, 该方法在实际应

用中面临严重的数据稀疏性和冷启动问题. 为了缓解数据稀疏性问题, 研究者提出了基于模型的协同过滤算法, 例如概率矩阵分解模型 (PMF)<sup>[5]</sup>、协同-主题概率张量分解模型 (CTTF) 等<sup>[6]</sup>. 这些方法利用观测到用户-物品交互行为学习用户和物品的隐式特征向量, 然后通过用户和物品隐式特征向量的内积产生推荐. 但是, 这些方法只能挖掘用户与音乐之间的一般性关系, 无法识别用户和物品隐式特征之间的细粒度关系. 此外, 基于矩阵分解的协同过滤方法也无法解决冷启动推荐问题, 在不同稀疏度数据集上的鲁棒性较差. 为了解决冷启动问题, 提高推荐模型的鲁棒性, 一些研究工作<sup>[7,8]</sup>挖掘物品的内容语义特征, 结合概率矩阵分解形成基于内容的推荐算法. 这些算法的核心是提取内容语义特征并巧妙地与协同过滤算法融合. 然而, 这些算法中的内容语义特征提取是与推荐任务无关的, 从而导致不准确的推荐.

为了解决音乐推荐中的数据稀疏性和冷启动问题, 提升冷启动音乐推荐的质量和鲁棒性, 本文提出了融合内容表示的度量排序学习推荐模型 (CAMRL). 该模型是一种概率排序学习方法, 通过将推荐系统中的评分预测问题转化为个性化排序问题, 充分利用观测到的和未观测到的用户行为进行模型训练, 以解决数据稀疏性问题, 进而提升模型的鲁棒性. 为了精确挖掘用户和音乐之间的关系, 模型将用户和音乐分别映射到相同的稠密语义空间, 利用距离度量指标构建用户与音乐之间的全局和细粒度关系. 针对音乐冷启动的推荐问题, 模型利用卷积神经网络 (CNN)<sup>[9]</sup>挖掘音乐的音频语义特征, 并以正则化方式与贝叶斯个性化排序框架 (BPR)<sup>[10]</sup>融合, 以推荐任务相关的监督学习方式挖掘音频语义特征.

## 2 相关工作

协同过滤是音乐推荐中最为常用的算法, 算法可分为基于内存和基于模型的推荐算法<sup>[6]</sup>. 基于内存的协同过滤推荐算法是基于计算的用户间或物品间相似性推断用户与物品的关系<sup>[4]</sup>. 基于模型的方法是利用机器学习技术, 如矩阵分解、张量分解等, 从用户对物品的反馈行为中推断用户和物品的潜在关系<sup>[11]</sup>. 其核心是将用户和物品分别映射到相同的稠密语义空间, 利用用户和物品在该空间中的内积建立用户对物品的偏好关系. 概率矩阵分解 (PMF) 是最早的基于模型的协同过滤方法, 利用用户对物品的评分数据学习用户和物品的潜在偏好关系<sup>[5]</sup>. 为了能够在推荐过程中同时推断用户喜欢和不喜欢的物品, Hu 等<sup>[4]</sup>还提出基于置信度的矩阵分解方法 (WRMF). 这些方法是以评分预测为优化目标, 忽略了推荐结果的排序关系. 为此, Stef-

fen 等<sup>[10]</sup>提出了个性化排序贝叶斯学习模型 (BPR), 将评分预测转化为个性化的排序学习. 最近, Hsieh 等<sup>[12]</sup>的研究认为用户对物品偏好是建立在全局和细粒度关系基础上的, 两个向量的内积操作无法区分细粒度上的差异, 提出了协同度量学习算法 (CML). 通过对这些文献的分析发现, 协同过滤算法可以直接应用于音乐推荐系统中, 但是面临严重的数据稀疏性和冷启动音乐的推荐问题.

为了缓解数据稀疏性问题, 提升用户或物品间相似度计算的准确性, 研究者提出了融合物品的内容相似性和行为相似性的解决方案<sup>[13]</sup>. 该方法依赖于统计相似度计算过程, 在数据极端稀疏时面临置信度不足的问题. 近年来, 随着多媒体内容数据 (包括文本、图片、视频、音频等) 的采集及内容特征提取技术的日益成熟和广泛应用, 国内外研究者还提出了将内容语义特征与基于模型推荐算法相融合的推荐策略, 显著改善了数据稀疏性和冷启动物品的推荐问题<sup>[14]</sup>. Wang 等<sup>[15]</sup>最早提出了协同主题模型 (CTR), 利用 LDA 主题模型从物品的文本内容描述中提取主题分布特征, 以正则化方式与概率矩阵分解框架融合, 通过交替训练策略学习用户和物品之间的关系. 鉴于 LDA 主题模型在短文本内容主题提取上的不足, Liu 等<sup>[6]</sup>提出了协同主题张量分解模型 (CTTF), 利用 BTM 模型提取短文本主题分布特征, 将提取到的主题特征与张量分解模型进行融合. 但无论是 LDA 还是 BTM 模型都只能够提取浅层的语义内容特征. 对此, Wang 等<sup>[7]</sup>提出了利用堆栈自编码 (SDAE) 提取物品文本内容的高层语义特征. 为了挖掘文本内部的情境信息和词序之间的关系, Kim 等<sup>[8]</sup>还提出了利用卷积神经网络 (CNN) 挖掘文本内容语义情境信息的方法, 与概率矩阵分解模型融合产生推荐. 这些方法在含有文本内容信息的物品推荐中取得了显著的推荐效果, 然而无法直接应用于音频内容特征的提取和推荐中.

为了解决音乐推荐中的数据稀疏性和冷启动推荐问题, Aaron 等<sup>[1]</sup>利用词袋模型和卷积神经网络 (CNN) 提取音乐的音频语义特征, 并以正则化方式与权重矩阵分解模型 (WMF) 进行融合. Liang 等<sup>[16]</sup>将神经网络训练成语义标注信息的内容模型, 并将其作为协同过滤模型的先验解决冷启动问题. 传统的基于内容的音乐推荐方法将音乐内容特征提取和用户偏好预测两个阶段隔离开, 使得特征的提取方法无法在音频中捕获所有相关信息, 为此, Wang 等<sup>[17]</sup>提出了基于内容的混合推荐方法, 利用深度信念网络和概率图模型将特征提取和偏好预测进行有机的融合. 这些方法以音频特征的提取为主, 忽略了音乐的文本特征, 为此, Oramas 等<sup>[2]</sup>提出了深度多模推荐方法, 利用预训练的模型分

别提取音乐的文本语义和音频语义特征,利用后向融合策略挖掘音乐的内容语义特征.这些方法虽然可以缓解数据稀疏性和冷启动问题,但语义内容特征的提取和推荐任务是相互分离的.此外,这些方法也无法同时挖掘用户和音乐之间的全局和细粒度关系,从而导致不精确的偏好推断.

### 3 融合内容表示的度量排序学习推荐模型

本文通过建立度量排序学习模型推断用户的音乐偏好,提升音乐推荐的质量和鲁棒性.为了便于模型表述,采用粗体大写字母表示矩阵,带下标的粗体大写字母表示向量,手写体大写字母表示集合,常规字母表示常量, $|\cdot|$ 表示集合中的元素个数.模型涉及到的主要符号及含义如表 1 所示.

表 1 主要符号及其含义

符号	含义
$\mathcal{U}, \mathcal{V}$	分别表示用户集合和音乐集合
$\mathbf{U}, \mathbf{Q}$	分别表示用户和音乐的隐式语义特征矩阵
$\mathbf{Z}$	音乐的音频语义特征矩阵
$\mathcal{D}_s$	用户对音乐偏好的偏序关系集合
$\mathbf{W}_l$	卷积神经网络第 $l$ 层权重参数矩阵

融合内容表示的度量排序学习推荐模型(CAMRL)是一种概率排序学习框架,将音乐推荐中的评分预测问题转化为个性化音乐排序问题,并充分利用挖掘的音乐内容语义特征解决冷启动推荐问题.提出的模型框架如图 1 所示.图 1 中的  $\mathbf{Q}_j$  表示用户收听过的音乐的隐式特征向量,  $\mathbf{Q}_k$  表示用户未收听过音乐的隐式特征向量,  $\mathbf{U}_i$  表示用户的隐式特征向量,  $\mathbf{Z}_j$  和  $\mathbf{Z}_k$  分别表示音乐  $v_j$  和  $v_k$  的音频语义特征向量.

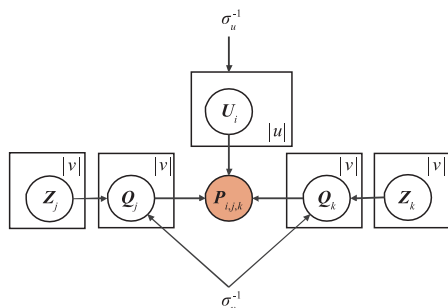


图1 CAMRL模型的图表示

设用户集合  $\mathcal{U} = \{u_1, u_2, \dots, u_{|\mathcal{U}|}\}$ , 音乐集合  $\mathcal{V} = \{v_1, v_2, \dots, v_{|\mathcal{V}|}\}$ . 对于每一个用户和音乐,分别将其映射到相同的稠密语义空间.用户  $u_i$  的隐式特征向量  $\mathbf{U}_i \sim N(0, \sigma_u^{-1}I)$ , 所有用户的隐式特征向量构成矩阵  $\mathbf{U} \in R^{|\mathcal{U}| \times K}$ . 对于音乐  $v_j$ , 利用深度卷积神经网络从音乐  $v_j$  的音频内容中提取内容语义特征向量  $\mathbf{Z}_j$ , 则音乐  $v_j$  的

隐式特征向量  $\mathbf{Q}_j \sim N(\mathbf{Z}_j, \sigma_v^{-1}I)$ , 所有音乐的隐式特征向量构成矩阵  $\mathbf{Q} \in R^{|\mathcal{V}| \times K}$ . CAMRL 模型的核心是学习用户的音乐偏好函数  $f(x_{ij}; \boldsymbol{\theta})$ , 使得可以对用户未收听过的音乐偏好进行排序, 产生 Top-N 个音乐给目标用户, 其中  $\boldsymbol{\theta} = \{\mathbf{U}, \mathbf{Q}\}$ .

#### 3.1 相对偏好模型

构建用户偏好模型的目的是从海量音乐中识别用户最可能喜欢的音乐.也就是说,给定任意两个音乐  $v_j$  和  $v_k$ ,  $f(x_{ij}; \boldsymbol{\theta})$  函数能够判断用户  $u_i$  对  $v_j$  和  $v_k$  的偏好差异.利用用户和音乐隐式特征向量的内积估计用户对音乐的偏好是最为常用的方法<sup>[2,18,19]</sup>.这种方法是从全局层面建立用户与音乐的一般性偏好,无法从细粒度层面挖掘用户和音乐间的内在关系,会导致不精确的估计.用户对音乐的偏好往往是由音乐特征向量中部分特征所决定的,因此,本文利用距离度量指标建立用户与音乐之间的全局和细粒度关系.为了便于表述,利用二元组  $x_{ij} = \langle u_i, v_j \rangle$  表示用户  $u_i$  收听了音乐  $v_j$ .基于欧式距离度量指标构建的用户偏好模型为:

$$f(x_{ij}; \boldsymbol{\theta}) = \frac{e^{-U_i - Q_j^2}}{Z(u_i)} \quad (1)$$

其中,  $U_i - Q_j^2 = \sum_{k=1}^K (\mathbf{U}_{ik} - \mathbf{Q}_{jk})^2$ ,  $K$  为隐式空间维度,  $\mathbf{U}_{ik} - \mathbf{Q}_{jk}$  衡量用户  $u_i$  和音乐  $v_j$  在第  $k$  维隐式特征上的差异,  $Z(u_i) = \sum_{n=1}^{|\mathcal{V}|} e^{-\|\mathbf{U}_i - \mathbf{Q}_n\|^2}$  为归一化因子.

为了建立用户对音乐偏好的偏序关系,将二元组  $x_{ij} = \langle u_i, v_j \rangle$  扩展为三元组  $x_{ijk} = \langle u_i, v_j, v_k \rangle$ , 表示相对于音乐  $v_k$ , 用户  $u_i$  更偏好音乐  $v_j$ .为了便于表示,用符号  $j >_i k$  表示三元组的偏序关系.从而,基于构建的偏好函数  $f(x_{ij}; \boldsymbol{\theta})$ , 用户  $u_i$  对音乐  $v_j$  和  $v_k$  偏序关系可以表示为:

$$g(j >_i k | \boldsymbol{\theta}) \Leftrightarrow f(x_{ij}; \boldsymbol{\theta}) > f(x_{ik}; \boldsymbol{\theta}) \quad (2)$$

其中  $v_k$  表示用户  $u_i$  未收听过的音乐.本文利用随机采样法从海量用户未收听过的音乐中为每个用户收听过的音乐  $v_j$  筛选一个  $v_k$  作为负样本,则为所有用户构造的偏序关系集合用

$$\mathcal{X}_s = \{ \langle u_i, v_j, v_k \rangle | u_i \in \mathcal{U}, v_j \in \mathcal{Q}_{u_i}^+, v_k \in \mathcal{Q} \setminus \mathcal{Q}_{u_i}^+, \}$$

表示,  $\mathcal{Q}_{u_i}^+$  表示用户  $u_i$  收听过的音乐集合,  $\mathcal{Q} \setminus \mathcal{Q}_{u_i}^+$  表示用户  $u_i$  未收听过的音乐集合.结合式(1)和(2),可得到如下偏序关系模型为:

$$g(j >_i k | \boldsymbol{\theta}) \Rightarrow m + D_{ij} - D_{ik} \quad (3)$$

其中  $D_{ij} = \mathbf{U}_i - \mathbf{Q}_j^2$ ,  $D_{ik} = \mathbf{U}_i - \mathbf{Q}_k^2$ ,  $m$  表示软边界间隔大小,限定了  $D_{ij}$  与  $D_{ik}$  的最小间隔距离.

#### 3.2 音频语义特征提取

目前,一些研究文献<sup>[2,7,18]</sup>借鉴迁移学习思想,利用基于 ImageNet 数据集预训练好的 VGG16 网络模型从音乐音频中提取语义内容特征,通过提取语义内容特征

向量与用户潜在特征向量的内积产生推荐. 然而, 这种与推荐任务无关的无监督学习方式会导致不准确的语义特征提取结果, 因为预训练好的 VGG16 网络只能提取一般化的内容语义特征. 本文提出了推荐任务相关的音频语义特征提取方法. 首先, 根据音乐所归属的类别标签, 预训练 6 层的简单卷积神经网络; 然后将网络模型第 5 层输出作为音乐潜在特征表示的均值融合到 BPR 模型框架中, 以推荐任务驱动的方式优化音乐语义特征提取网络模型. 本文采用的音乐内容语义特征模型如图 2 所示.

由于音乐是由音频信号所构成的, 首先利用离散傅里叶变换将音频信号转化为频域, 为每首音乐生成

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. 由于每首音乐的播放时间不同, 对于每首音乐的彩色声谱图, 将其分割成多个  $256 \times 256 \times 3$  的图片, 每张图片均对应着音乐所属类别的标签. 最后, 将构成的图片数据集作为卷积神经网络的输入, 预训练神经网络模型. 移除预训练好的网络模型最后一层, 则神经网络的语义内容特征输出用  $Z_j$  表示, 将其作为音乐潜在特征表示的均值融合到 BPR 模型中,  $Z_j$  的维度等于音乐的潜在语义特征维度. 在冷启动音乐的推荐过程中, 将  $Z_j$  代替  $Q_j$  计算用户  $u_i$  对音乐  $v_j$  的偏好.

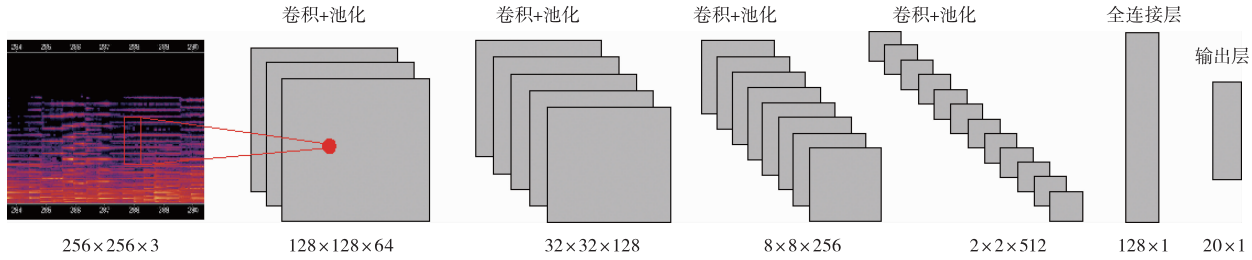


图2 内容语义特征提取模型

### 3.3 联合训练

为了能够对提出的模型参数进行优化, 假设用户对每首音乐的收听行为都是相互独立的. 基于 BPR 框架<sup>[10]</sup>, 采用最大化后验概率进行模型优化:

$$\varphi^* = \operatorname{argmax}_{\varphi} \prod_{(u_i, v_j, v_k) \in \mathcal{D}_i} P(j >_i k | \theta) P(\varphi) \quad (4)$$

其中  $\varphi = \{\theta, \mathbf{W}_l\}$ ,  $\theta = \{U, Q\}$ ,  $P(j >_i k | \theta)$  是用户  $u_i$  对音乐  $v_j$  和  $v_k$  相对偏好函数  $g(j >_i k | \theta)$  的概率化表示, 即:

$$P(j >_i k | \theta) = \sigma(m + D_{ij} - D_{ik}) \quad (5)$$

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (6)$$

假设  $\varphi$  中的参数取值服从高斯先验概率分布, 则模型的最终目标函数可表示为:

$$\begin{aligned} \varphi^* = \operatorname{argmax}_{\varphi} & \sum_{(u_i, v_j, v_k) \in \mathcal{D}_i} \ln(\sigma(m + D_{ij} + D_{ik})) \\ & - \frac{\lambda_U}{2} \sum_i \|U_i\|^2 - \frac{\lambda_V}{2} \sum_j \|Q_j - Z_j\|^2 \\ & - \frac{\lambda_V}{2} \sum_k \|Q_k - Z_k\|^2 - \frac{\lambda_W}{2} \sum_l \|\mathbf{W}_l\|^2 \end{aligned} \quad (7)$$

其中,  $\mathbf{W}_l$  表示卷积神经网络第  $l$  层的权重参数.

由于式(7)的每次计算均需要对全量用户-音乐偏好对集合  $D_i$  求和, 直接对式(7)求导并更新参数是极为耗时的. 为此, 模型利用随机梯度下降法进行求解参数. 求解过程中先固定卷积神经网络的参数  $\mathbf{W}_l$ , 然后利用随机梯度下降法求解参数  $U$  和  $Q$ , 然后再固定  $U$  和  $Q$ , 利用后向传播算法求解以下函数更新参数  $W$ .

$$\begin{aligned} \mathbf{W}^* = \operatorname{argmax}_{\mathbf{W}} & \left( \frac{\lambda_U}{2} \sum_i \|U_i\|^2 - \frac{\lambda_V}{2} \sum_j \|Q_j - Z_j\|^2 \right. \\ & \left. - \frac{\lambda_V}{2} \sum_k \|Q_k - Z_k\|^2 - \frac{\lambda_W}{2} \sum_l \|\mathbf{W}_l\|^2 \right) \end{aligned} \quad (8)$$

## 4 数据集与评估方法

### 4.1 数据描述与处理

为了验证提出的 CAMRL 模型的有效性, 使用 KKBOX 数据集和 MIGU 数据集进行实验分析. KKBOX 数据集是 WSDM 2018 竞赛中公开的由亚洲音乐服务商 KKBOX 公司所提供的音乐收听行为数据集, 该数据集包含用户 ID、音乐 ID、音乐名称、歌手、一个月内是否重复收听 4 个字段, 涉及 26722 个用户对 21761 首音乐的收听行为, 共计 5661451 记录数据, 数据稀疏度为 0.974%. 由于 KKBOX 数据集并未提供本文所需要的音频内容数据, 因此, 我们采用网络爬虫技术从咪咕音乐、网易云音乐等各大音乐网站抓取音乐的音频文件. MIGU 数据集是由中国移动咪咕音乐公司提供的用户音乐收听行为数据, 该数据集是由作者从咪咕音乐后台业务系统日志中提取的, 包含用户的听歌、收藏、点赞、设置彩铃、分享等行为数据和音乐的音频文件. 采集到的 MIGU 数据集包括 101479 个用户对 18156 首音乐的收听行为数据, 共计 11780581 条记录, 数据稀疏度为 0.639%.

对于 KKBOX 数据集和 MIGU 数据集, 删除没有音频文件的用户行为记录. 对于音乐音频文件, 利用 FFPEMG 工具将音乐的音频文件转换彩色声谱图, 按照

从右向做的顺序将声谱图分割成若干个  $256 \times 256 \times 3$  的图片,丢弃剩余不足 256 长度的声谱图。

## 4.2 对比方法与参数设置

本文选取以下几种比较先进的推荐算法与 CAMRL 进行比较。

**BPR<sup>[10]</sup>**:BPR 是利用用户的隐式反馈构造用户偏好的偏序关系对,通过构造排序型的贝叶斯网络学习用户和物品的潜在特征向量,是一种特殊的概率矩阵分解方法。

**WRMF<sup>[4]</sup>**:WRMF 是一种加权的矩阵分解算法,该算法通过对正样本和负样本赋予不同的权重,采用驻点方式解决推荐中的 One-Class 协同过滤问题。

**CML<sup>[12]</sup>**:CML 是利用距离度量指标对用户偏好、用户-用户和物品-物品相似关系联合建模,学习用户和物品间的内在关系。

**VBPR<sup>[20]</sup>**:VBPR 是利用预训练好的深度神经网络模型提取物品的视觉特征,并融入到 BPR 算法中学习用户和物品之间的关系。在本文的实验中,利用预训练好的深度神经网络模型提取音乐声谱图的内容特征。

**DeepMusic<sup>[1]</sup>**:DeepMusic 是利用深度卷积神经网络提取歌曲音频中的特征作为歌曲的特征,然后采用 WMF 模型进行隐式评分预测。

**CDL<sup>[7]</sup>**:CDL 是同时利用评分数据和物品文本信息的混合推荐算法,该算法利用堆栈式自编码(SDAE)提取物品内容特征,然后将提取到的内容特征融合到概率矩阵分解框架。在本文的实验中,利用 SDAE 提取音乐声谱图的内容特征。

**ConvMF<sup>[8]</sup>**:ConvMF 是融合内容特征的概率矩阵分解模型,该算法利用卷积神经网络(CNN)提取物品内容特征,然后以正则化方式融合到概率矩阵分解框架。在本实验中,利用 CNN 提取音乐声谱图的内容特征。

## 4.3 评估指标

为了评估 CAMRL 模型在 KKBOX 和 MIGU 数据集上的效果,实验采用 Recall 和 MAP 指标进行衡量<sup>[6]</sup>。具体计算公式为:

$$\text{Recall}@N = \frac{1}{|D^{\text{Test}}|} \sum_{u \in D^{\text{Test}}} \frac{|S_u^{\text{pre}} \cap D_u^{\text{Test}}|}{|D_u^{\text{Test}}|} \quad (9)$$

$$\text{MAP}@N = \frac{\sum AP^u@N}{|D^{\text{Test}}|} \quad (10)$$

$$AP^{(u,r)}@N = \frac{1}{|D_u^{\text{Test}}|} \sum_{r=1}^K (P_i(r) \times \text{rel}_i(r)) \quad (11)$$

其中, $AP^u@N$ 表示用户  $u$  的 Top-N 推荐结果在测试集中的平均准确度; $S_u^{\text{pre}}$ 表示为用户  $u$  推荐的音乐列表, $D_u^{\text{Test}}$ 表示测试集中的音乐列表, $|D_u^{\text{Test}}|$ 表示用户  $u$  在测试集中的样本个数; $r$ 表示给定的截断排序, $P_i(r)$ 表示在给定截断排序  $r$  条件下,用户  $u$  的第  $i$  个测试样本准

确度; $\text{rel}_i(\cdot)$ 表示给定排序情况下的二元相对函数。

实验是在 4 核高性能 NVIDIA GPU 服务上进行模型训练,每个 GPU 含有 24GB 显存。设计的模型利用 Python 和 Tensorflow 深度学习框架实现。为了得到准确的评估结果,每次实验重复进行 5 次,取实验结果的平均值作最终评价结果。对于模型的超参数,采用交叉验证法寻找最优的参数;对于对比方法的超参数,通过交叉验证法在本文的实验数据集上寻找最优的参数,具体参数设置为:对于 BPR 和 WRMF 的正则化超参数  $\lambda$  设为 0.0001;对于 CML,  $m = 1, \lambda_f = 1, \lambda_j = 10$ ;对于 CDL,  $a = 1, b = 0, K = 100, \lambda_u = 10, \lambda_v = 100$ ;对于 VBPR,  $K = 100, \lambda_u = \lambda_p = \lambda_j = 1, \lambda_b = \lambda_c = 0$ ;对于 DeepMusic,  $K = 50, \lambda = 0.01$ ;对于 ConvMF,  $K = 100, \lambda_u = 10, \lambda_v = 1$ 。

## 5 实验结果与分析

### 5.1 冷启动音乐推荐性能对比

为了验证提出的模型在冷启动音乐推荐上的效果,采用文献[19,22,23]中的冷启动物品测试方法进行测试。首先,将两个数据集中被用户收听次数少于 5 次的音乐视为冷启动音乐,然后选择至少收听过一首冷启动音乐的用户作为本次实验的测试用户群体。对于每个用户的收听记录,将其收听的冷启动音乐记录作为测试集,剩余的收听记录作为训练集。实验的目的是判断冷启动的音乐是否在推荐的 Top-N 个音乐推荐列表中。

对比方法中只有 VBPR、DeepMusic、CDL 和 ConvMF 可以解决冷启动推荐问题,因此,实验仅将 CAMRL 与这三个模型进行实验对比,采用 Recall 和 MAP 指标进行衡量。为了避免单次实验引起的误差,采用 10 次重复实验的平均值作为最终的实验结果。实验结果如图 3 和表 2 所示。

表 2 不同推荐算法在冷启动音乐推荐上的 MAP 实验结果

数据集	VBPR	DeepMusic	CDL	ConvMF	CAMRL
KKBOX	0.0011	0.0039	0.0043	0.0062	<b>0.0085</b>
MIGU	0.0009	0.0012	0.0013	0.0041	<b>0.0052</b>

实验结果表明,在 KKBOX 和 MIGU 数据集上,本文提出的 CAMRL 模型在 Recall 和 MAP 指标上均显著优于对比模型。这是由于本文利用音乐所属类别标签进行预训练卷积神经网络,提取音乐的音频内容语义特征;然后将预训练模型倒数第二层的结果以正则化方式融合到贝叶斯网络框架,通过推荐任务相关的监督学习方式微调音频内容语义特征提取网络模型。此外,CAMRL 模型将推荐的评分预测问题转换为排序学习问题,可充分利用观测和未观测到的用户行为进行训练。由于 VBPR、CDL 和 DeepMusic 算法采用预训练好的深度神经网络提取内容语义特征,是与推荐任务

无关的,因此会产生有偏差的推荐结果. 虽然 ConvMF 以与推荐任务相关的方式微调预训练好的网络模型,但以评分预测为优化目标,面临严重的数据稀疏性问题. 此外,实验还发现 CAMRL 模型和对比模型在 KKBOX 数据集上的 Recall 和 MAP 指标显著高于在 MIGU

数据集上的指标,这是由于 KKBOX 数据集是经过严格清洗过的比赛数据,而 MIGU 数据集是作者直接从生产环境采集到的. 相比于 KKBOX 数据集,MIGU 数据集更接近真实商业环境.

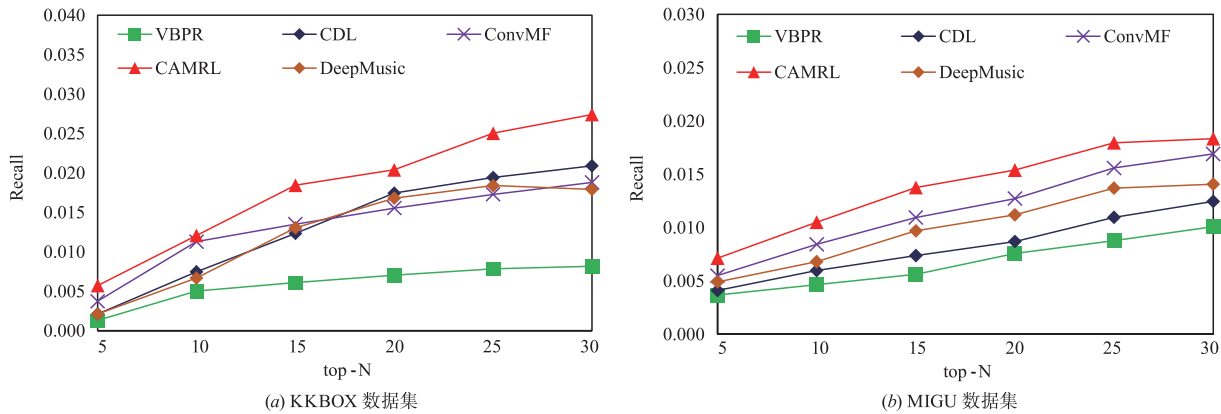


图3 不同推荐算法在冷启动音乐推荐上的Recall实验结果

### 5.2 不同稀疏数据集上的鲁棒性分析

为了验证 CAMRL 模型在不同稀疏度数据集上的鲁棒性,分别对 KKBOX 和 MIGU 数据集构造 4 种

稀疏数据集 { 20% , 40% , 60% , 80% }, 利用 Recall@10 和 MAP 两个指标进行衡量,实验结果如图 4 和图 5 所示.

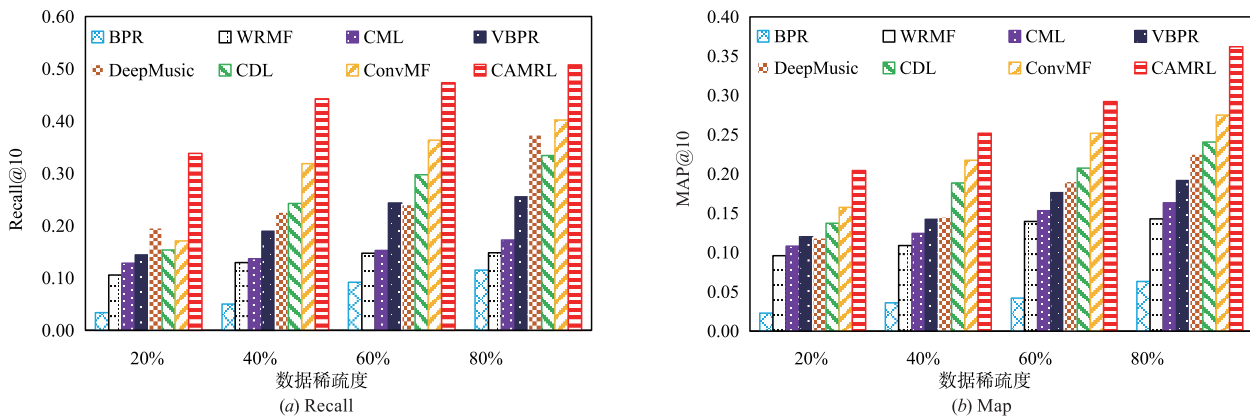


图4 各算法在不同稀疏程度的KKBOX数据集上的实验结果

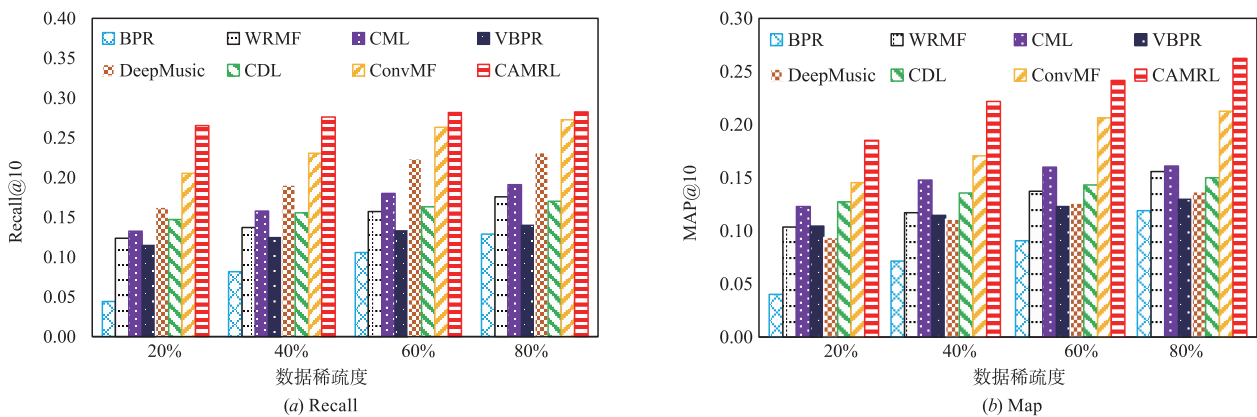
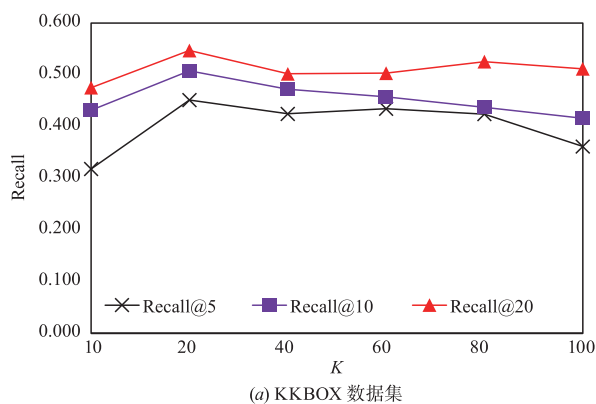


图5 各算法在不同稀疏程度的MIGU数据集上的实验结果

实验结果表明,无论是在 KKBOX 数据集还是 MIGU 数据集上,CAMRL 模型在不同稀疏数据集上的表现均优于对比方法,模型的鲁棒性得到了显著提升.这是由于本文提出的模型是一种排序学习模型,可以充分利用观测和未观测行为数据构造偏序关系,基于偏序关系数据对模型训练,从而使得模型训练更加充分;其次,构建了音乐内容语义特征的提取模型,提取的内容语义特征在一定程度上缓解了数据稀疏性问题.

### 5.3 模型参数分析

模型的潜在因子  $K$  是影响模型准确度的主要参数.为了寻找最佳的参数取值,将  $K$  的取值分别设为



{10,20,40,60,80,100},然后在 MIGU 和 KKBOX 数据集上评估对 Recall 指标的影响.实验结果如图 6 所示.

实验结果表明,随着潜在因子维度  $K$  的取值增大,Recall@5、Recall@10 和 Recall@20 的值也显著增大.当潜在因子维度过低时,无法充分学习用户和物品的隐含特征,导致算法无法精确预测用户对物品的偏好.当潜在因子维度  $K$  持续增加时,Recall 指标在两个数据集上的取值呈下降趋势,这是由于当潜在因子维度过大时,用户向量和物品潜在向量会包含过多的冗余特征,导致算法预测准确率下降.当潜在因子维度  $K$  的取值为 20 时取得最好的实验结果.因此,本文将  $K$  设为 20.

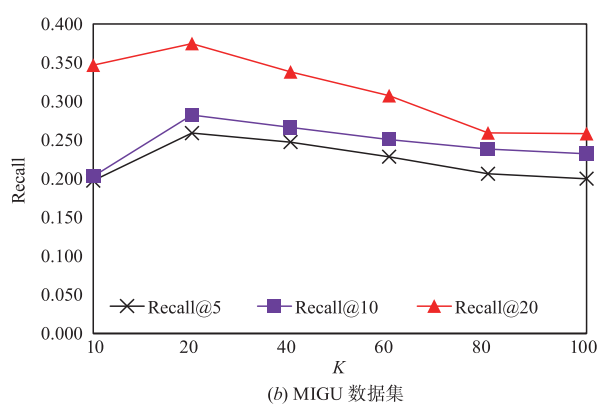


图6 潜在因子 $K$ 对模型的影响

## 6 结论

针对音乐推荐中的数据稀疏性和冷启动问题,本文提出了融合内容表示的度量排序学习推荐模型.该模型是一种概率排序学习方法,采用度量学习方法挖掘用户对音乐的全局和细粒度偏好关系,利用观测到的和未观测到的用户行为对模型训练,达到缓解数据稀疏性,提升模型鲁棒性的目的.针对冷启动音乐的推荐问题,通过构建与推荐任务相关的卷积神经网络模型提取音频语义特征,并以正则化方式融入到概率排序学习模型中.在 KKBOX 和 MIGU 数据集上的测试结果表明,提出的模型显著提升了在不同稀疏度的数据集上的精度和鲁棒性.与 ConvMF 相比,模型的冷启动音乐的 Recall@10 在 KKBOX 和 MIGU 数据集上分别提升了 24.61% 和 6.9%.

### 参考文献

- [1] Van den Oord A, Dieleman S, Schrauwen B. Deep content-based music recommendation [A]. Proceedings of the 26th International Conference on Neural Information Processing Systems [C]. ACM Press, 2013. 2643 - 2651.
- [2] Oramas S, Nieto O, Sordo M, et al. A deep multimodal ap-

proach for cold-start music recommendation [A]. Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems [C]. Como Italy: ACM, 2018. 32 - 37.

- [3] Chen J, Ying P, Zou M. Improving music recommendation by incorporating social influence [J]. Multimedia Tools and Applications, 2019, 78(3): 2667 - 2687.
- [4] Hu Y, Koren Y, Volinsky C. Collaborative filtering for implicit feedback datasets [A]. Proceeding of the 2008 Eighth IEEE International Conference on Data Mining [C]. ACM Press, 2008. 263 - 272.
- [5] Salakhutdinov R, Mnih A. Probabilistic matrix factorization [A]. Proceedings of the 20th International Conference on Neural Information Processing Systems [C]. ACM Press, 2007. 1257 - 1264.
- [6] Liu T, Liao J, Wang Y, et al. Collaborative tensor-topic factorization model for personalized activity recommendation [J]. Multimedia Tools and Applications, 2019, 78(12): 16923 - 16943.
- [7] Wang H, Wang N, Yeung D Y. Collaborative deep learning for recommender systems [A]. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]. ACM Press, 2015. 1235 - 1244.

- [8] Kim D, Park C, Oh J, et al. Convolutional matrix factorization for document context-aware recommendation[A]. Proceedings of the 10th ACM Conference on Recommender Systems[C]. ACM Press, 2016. 233 – 240.
- [9] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large – Scale Image Recognition[A]. Computer Vision and Pattern Recognition[C]. San Diego: Computer Science, 2014.
- [10] 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]. ACM Press, 2009. 452 – 461.
- [11] Liu B, Xiong H, Papadimitriou S, et al. A general geographical probabilistic factor model for point of interest recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(5): 1167 – 1179.
- [12] Hsieh C K, Yang Y, Cui Y, et al. Collaborative metric learning[A]. Proceedings of the 26th International Conference on World Wide Web[C]. ACM Press, 2017. 193 – 201.
- [13] 黄贤英, 熊李媛, 李沁东. 基于改进协同过滤算法的个性化新闻推荐技术[J]. 四川大学学报(自然科学版), 2018, 55(01): 49 – 55.  
Huang X Y, Xiong L Y, Qin-Dong L I. Personalized news recommendation technology based on improved collaborative filtering algorithm[J]. Journal of Sichuan University, 2018, 55(01): 49 – 55. (in Chinese)
- [14] 任开旭, 王玉龙, 刘同存, 李炜. 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)
- [15] Wang C, Blei D M. Collaborative topic modeling for recommending scientific articles[A]. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining[C]. ACM Press, 2011. 21 – 24.
- [16] Liang D, Zhan M, Ellis D P W. Content-aware collaborative music recommendation using pretrained neural networks[A]. Process of the 16th International Society for Music Information Retrieval Conference[C]. 2015. 295 – 301.
- [17] Wang X, Wang Y. Improving content-based and hybrid music recommendation using deep learning[A]. Proceedings of the ACM International Conference on Multimedia[C]. ACM Press, 2014. 627 – 636.
- [18] Liao J, Liu T, Liu M, et al. Multi-context integrated deep neural network model for next location prediction[J]. IEEE Access, 2018, 6: 21980 – 21990.
- [19] Liu T, Liao J, Wu Z, et al. A geographical-temporal awareness hierarchical attention network for next point-of-interest recommendation[A]. Proceedings of the 2019 ACM on International Conference on Multimedia Retrieval[C]. ACM Press, 2019. 7 – 15.
- [20] He R, McAuley J. VBPR: visual Bayesian personalized ranking from implicit feedback[A]. Proceedings of the 30th AAAI Conference on Artificial Intelligence[C]. ACM Press, 2016. 144 – 150.
- [21] Wang H, Wang N, Yeung D Y. Collaborative deep learning for recommender systems[A]. in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C]. ACM Press, 2015. 1235 – 1244.
- [22] Li J, Lu K, Huang Z, et al. Two birds one stone: on both cold-start and long-tail recommendation[A]. Proceedings of the 2017 ACM on Multimedia Conference[C]. ACM Press, 2017. 898 – 906.
- [23] Zhang Y, Yin H, Huang Z, et al. Discrete deep learning for fast content-aware recommendation[A]. Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining[C]. ACM Press, 2018. 717 – 726.

### 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