

# 基于方面和胶囊网络的跨域评分预测模型

梁顺攀, 刘 伟, 郑智中, 原福永  
(燕山大学信息科学与工程学院, 河北秦皇岛 066004)

**摘要:** 用户评论可以反映用户对项目的偏好信息, 将用户在其他领域的偏好迁移到目标领域进行跨域推荐, 可以缓解目标域数据稀疏引起的冷启动问题. 本文针对传统的跨域推荐方法无法将完整的用户偏好进行迁移以及传统的方面提取方法预测精度不高两个问题, 提出基于方面和胶囊网络的跨域评分预测模型 ACN (Aspect and Capsule Network). ACN 模型使用胶囊网络挖掘评论文档的多个方面, 然后通过注意力机制筛选出对目标域最重要的特征, 迁移到目标域进行评分预测. 最后, 通过实验证明 ACN 模型分别在单一源域和多源域的情况下, 较基准模型最高有 2.3% 和 20.8% 的性能提升.

**关键词:** 跨域推荐; 评分预测; 胶囊网络; 方面提取; 注意力

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 0372-2112(2022)09-2102-08

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20210483

## Cross Domain Rating Prediction Based on Aspect and Capsule Network

LIANG Shun-pan, LIU Wei, ZHENG Zhi-zhong, YUAN Fu-yong

(College of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China)

**Abstract:** Users' reviews on items can reflect users' preference information. Migrating users' preferences in other fields to the target domain for cross domain recommendation can alleviate the cold start problem caused by sparse data in the target domain. Aiming at the two problems that the traditional cross domain recommendation methods can not migrate the complete user preferences and the prediction accuracy of the traditional aspect extraction methods is incorrect, this paper proposes a cross domain score prediction model based on aspect and capsule network (ACN). ACN uses capsule network to mine multiple aspects of review documents, and then selects the most important features for the target domain through attention mechanism, and migrate to the target domain for score prediction. Compared with the benchmark model, ACN has the highest performance improvement of 2.3% and 20.8% when using single source domain and multi-source domain.

**Key words:** cross domain recommendation; rating prediction; capsule network; aspect extraction; attention

## 1 引言

数据稀疏问题一直是推荐系统的挑战之一, 数据稀疏会导致推荐系统的推荐结果精度不高等问题, 文献[1]针对推荐系统中的高维稀疏矩阵提出了一种无约束非负潜在因子分析算法 UNLFA (Unconstrained Non-negative Latent Factor Analysis). 文献[2]研究了八种扩展的随机梯度下降算法, 提出八种新的潜在因子模型. 文献[3]提出了一种基于高维稀疏矩阵的深度潜在因子模型 DLFM (Deep Latent Factor Model), 通过一个非线性激活函数将多个潜在因子模型依次连接起

来. 上述几种方法利用矩阵分解有效的缓解了推荐系统中高维稀疏矩阵的数据稀疏问题, 但没有考虑到其他的辅助信息. 使用评论文档作为辅助信息可以缓解仅使用评分矩阵时产生的数据稀疏问题, 但用户在目标域冷启动<sup>[4]</sup>时, 就无法从目标域中提取相关用户的偏好.

跨域推荐能够缓解用户在目标域冷启动问题. 基于集合矩阵分解的关系学习模型 CMF (Collective Matrix Factorization)<sup>[5]</sup>将矩阵分解应用到跨域推荐中, 并且在多个源域的因子之间共享参数来实现跨域知识集成; 跨域推荐的嵌入映射框架 EMCDR (Embedding and

Mapping framework for Cross-Domain Recommendation)<sup>[6]</sup>通过多层感知机捕捉域间的非线性映射函数,得到来自不同领域的用户表示;跨域和跨系统的深度推荐框架 DCDCSR (Deep framework for both Cross-Domain and Cross-System Recommendations)<sup>[7]</sup>先使用矩阵分解模型生成用户和项目的潜在因子,然后使用深度神经网络来映射跨域的潜在因子,进一步扩展了 EMCDR. 基于共享知识模型的跨域推荐算法 SKP (Sharing Knowledge Pattern)<sup>[4]</sup>,通过分解用户-项目评分矩阵得到用户分组对项目分组的评分知识模型,并利用目标域的个性知识模型和各个领域共性知识模型得到推荐结果. 知识聚合和迁移相结合的跨领域推荐算法 ATCF (Aggregation and Transfer Collaborative Filtering algorithm)<sup>[8]</sup>,通过基于矩阵分解的矩阵拼接和矩阵填充方法,有效避免了负迁移. 上述方法考虑了用户对源域项目评分,忽略了评论文档可以更好的挖掘偏好信息这一特性.

多视角深度神经网络学习方法 MVDNN (Multi-View Deep Neural Network)<sup>[9]</sup>将用户和项目的评论信息映射到一个潜在空间,最大化了用户和喜欢的项目之间的相似度;基于联合张量因子分解的评论跨域推荐模型 RB-JTF (Review-Based cross-domain recommendation through Joint Tensor Factorization)<sup>[10]</sup>提出了一个辅助域和目标域的联合张量分解模型,利用评论文本中提取的方面因子来提高跨域推荐的性能;基于内容的深度融合模型 R-DFM (Review and content based Deep Fusion Model)<sup>[11]</sup>使用融合评论的堆叠式去噪自动编码器来训练源域和目标域中的用户或项目因素,然后利用多层感知器将其在两个域之间传递,从而提高冷启动用户推荐性能;通过方面转换为冷启动用户提供跨域推荐模型 CATN (Cross-domain recommendation framework via Aspect Transfer Network)<sup>[12]</sup>使用 Aspects<sup>[13]</sup>以及注意力<sup>[14,15]</sup>机制提取用户和项目的方面,并且增加了相似用户的评论作为辅助文档增强推荐的准确率. 以上这些方法虽然使用了评论文档作为辅助信息,提高了推荐的准确度,但其使用的深度模型不能准确地分出用户的细粒度偏好,并且只使用了单一源域,故这些方法的推荐准确度不高.

用户对项目的评论信息可以反映出用户对项目多个层面的评价,项目表现出的多个层面即“方面”<sup>[13]</sup>. 方面能反映用户评论对项目属性的情感程度,将“方面”引入跨域推荐模型能够更好的进行知识迁移. 胶囊网络<sup>[16-18]</sup>将神经网络中的神经元扩展成神经向量,并用胶囊来表示,胶囊的长度代表属性的概率,而胶囊的方向代表了该属性的具体描述,通过这一特性就可以保存评论文档中的用户和项目方面. 胶囊网络的动态路由机制可以选择性的聚集一些低层的胶囊(属性),

进而生成高层的胶囊(属性),这些特性都使胶囊网络比其他神经网络对空间和语序更加敏感. 此外,胶囊网络的动态路由机制与神经网络的反向传播机制不同,仅通过少量迭代就可以获得较好的结果. 以上这些特性说明,胶囊网络可以更好地挖掘文本的细粒度特征.

基于以上背景,本文提出了一个基于方面和胶囊网络的跨域评分预测模型:ACN. 为了更好的学习跨域知识,本文将一对相关数据集设定互为源域和目标域,交替进行学习训练. 本文主要贡献总结如下:

(1) 基于胶囊网络可以挖掘细粒度偏好这一特性,本文使用胶囊网络挖掘评论文档、提取用户和项目的方面,并且通过基于方面的自注意力机制,获取对于用户和项目表示中更重要的方面.

(2) 本文将胶囊网络的映射矩阵修改为共享矩阵,使路由过程输出的每个偏好胶囊都位于同一个向量空间中,保证迁移到目标域的方面处在同一向量空间.

(3) 本文使用不同类型的源域输入模型进行实验,证明了在数据稀疏和冷启动的前提下,ACN 可以学习并迁移不同源域、多源域的知识.

## 2 模型结构

本文提出的 ACN 模型结构如图 1 所示. ACN 设置为一个双塔结构网络,分别用来处理用户评论文档和项目评论文档,最后通过知识迁移层将从源域中获取的偏好迁移到目标域,从而预测用户对项目的评分.

ACN 模型分为四层:(1)初始胶囊层:将评论文档嵌入表示,经过一次卷积操作初步提取文本特征,将评论向量转化为初始化胶囊.(2)主胶囊层:将初始化的胶囊输入胶囊网络,通过迭代路由过程挖掘用户或项目的方面.(3)方面注意力层:将输出胶囊进行方面注意力表示,针对不同源域的侧重点筛选出不同的方面.(4)知识迁移层:将两个域的方面表示匹配,进行最终的评分预测. 由于用户网络与项目网络结构相同,故下文主要以用户网络为例进行详细介绍.

ACN 模型中主要用到的符号及其含义如表 1 所示.

### 2.1 初始胶囊层

以用户  $u$  为例,将  $u$  在源域的评论文档定义为  $D_u \in \mathbb{R}^L$ ,其中每一项代表用户评论文档中的单词索引, $L$  是评论长度. 将  $D_u$  进行词嵌入后得到词嵌入矩阵  $E_u = [e_1 \cdots e_g \cdots e_L]$ ,其中  $e_g \in \mathbb{R}^d$  代表第  $g$  个词的维度为  $d$  的嵌入向量. 将  $E_u$  经过一次一维卷积,并初始化为胶囊网络的输入胶囊矩阵.

选用  $n$  个宽度为  $d$ ,窗口大小为  $w_1$  的卷积核来提取文本特征,构成初始化胶囊,则第  $k$  个卷积核对  $E_u$  的卷积操作定义为:

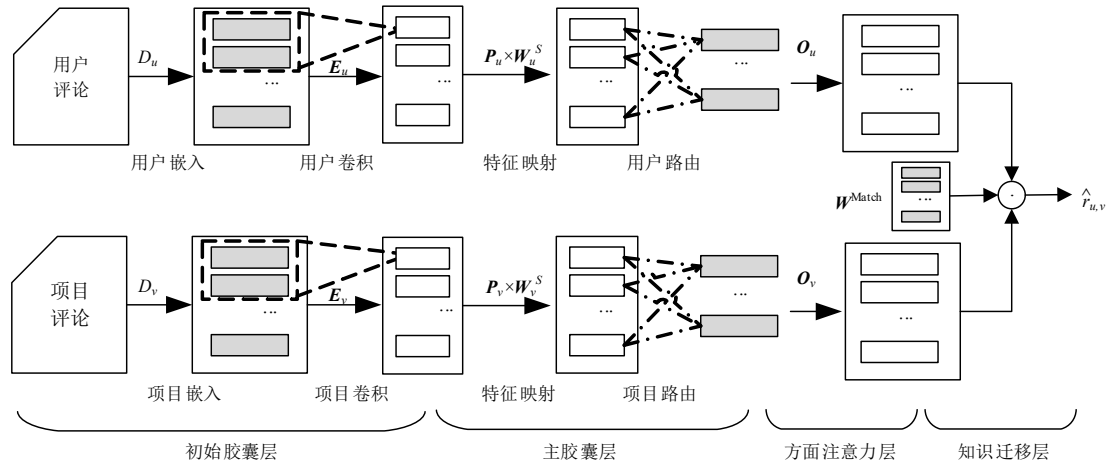


图1 ACN模型结构

表1 模型符号定义

符号	定义	符号	定义
$U$	用户集合	$S$	方面对之间的语义匹配矩阵
$V$	项目集合	$P_u, P_v$	用户和项目的初始胶囊
$M$	用户或项目的方面数量	$O_u, O_v$	用户和项目的高层胶囊
$R$	用户项目交互的集合	$A_u, A_v$	用户和项目的方面矩阵
$L$	评论长度	$E_u, E_v$	用户和项目的词嵌入矩阵
$w$	窗口大小	$W^k$	第 $k$ 个卷积权重矩阵
$r_{u,v}$	评分矩阵中用户 $u$ 对项目 $v$ 的评分	$W^{\text{Match}}$	方面匹配映射矩阵
$\hat{r}_{u,v}$	用户 $u$ 对项目 $v$ 的预测评分	$W^Q, W^K, W^{\text{Attn}}$	非线性变换矩阵和注意力权重
$h$	胶囊维度	$Q_u, K_u$	用户 $u$ 的注意力机制查询项和键项
$D_u, D_v$	用户和项目的评论文档	$b$	偏置向量
$W_u^S$	共享矩阵	$q$	高层胶囊模糊表示

$$p_k = \text{ReLU}(E_u * W^k + b^k) \quad (1)$$

其中  $*$  代表一维卷积,  $W^k$  是第  $k$  个卷积权重矩阵,  $b^k$  是第  $k$  个卷积偏置. 将  $n$  个卷积核卷积结果进行拼接, 得到  $L-w_1+1$  个维度为  $n$  的初始化胶囊  $P_u = [p_1 \cdots p_k \cdots p_{L-w_1+1}]$ , 其中  $p_k \in \mathbb{R}^n$  代表第  $k$  个初始化胶囊.

## 2.2 主胶囊层

将初始化胶囊输入到胶囊网络作为低层胶囊, 首先对低层胶囊进行特征映射, 预测高层胶囊在偏好空间的位置. 由于使用不同的映射矩阵会将胶囊映射到不同的偏好空间, 导致高层胶囊存在于不同空间, 影响最终的知识迁移过程, 当使用的源域与目标域不相关或者使用多个源域时, 这种影响尤为严重. 本文在每对低层胶囊和高层胶囊之间使用路由全局共享的映射矩阵  $W_u^S$ , 将用户或项目的每个低层胶囊映射到相同的向量空间, 得到高层胶囊的模糊表示  $q_i$ .

$$q_i = W_u^S \cdot p_i \quad (2)$$

其中  $p_i$  代表第  $i$  个低层胶囊,  $W_u^S$  是共享映射矩阵, 代表低层胶囊和高层胶囊的联系.

得到  $q_i$  后, 对其进行动态路由过程. 首先通过低层

胶囊和高层胶囊之间的连接权重  $c_{ij}$  加权计算高层胶囊的候选向量  $s_j$ , 连接权重  $c_{ij}$  可以在动态路由过程通过  $b_{ij}$  计算得出.

$$c_{ij} = \text{softmax}(b_{ij}) \quad (3)$$

$$s_j = \sum_i c_{ij} \cdot q_i \quad (4)$$

其中  $b_{ij}$  表示每对低层胶囊和高层胶囊之间的一致性分数, 初始值为 0.

将高层胶囊的候选向量  $s_j$  经过非线性 squash 函数, 得到标准化的高层胶囊.

$$o_j = \text{squash}(s_j) = \frac{s_j}{\|s_j\|} \cdot \frac{\|s_j\|^2}{1 + \|s_j\|^2} \quad (5)$$

squash 函数第一项的作用是将  $s_j$  单位化, 即将  $s_j$  压缩到  $[0, 1]$ , 第二项是一个大于 0 小于 1 的值, 分母中的数字代表函数的挤压程度, 当  $\|s_j\|^2$  大于挤压程度时, 第二项则会大于 0.5, 表示该胶囊代表的特征存在几率大, 而当  $\|s_j\|^2$  小于挤压程度时, 第二项则会小于 0.5, 表示该胶囊代表的特征存在几率小. 这样可以放大  $s_j$  的重要部分, 缩小  $s_j$  的不重要部分.

最后,在动态路由的每次迭代中, $\mathbf{b}_{ij}$ 通过以下公式进行更新:

$$\mathbf{b}_{ij} = \mathbf{b}_{ij} + \text{squash}(s_j) \cdot \mathbf{q}_i \quad (6)$$

动态路由过程收敛后,胶囊网络输出  $M$  个高层胶囊  $\mathbf{o}_k \in \mathbb{R}^e, k \in \{1, 2, \dots, M\}$ ,它们组成了用户  $u$  的胶囊矩阵  $\mathbf{O}_u$ ,代表了用户  $u$  的  $M$  个方面的偏好特征.

$$\mathbf{O}_u = [\mathbf{o}_1 \mathbf{o}_2 \dots \mathbf{o}_M] \quad (7)$$

### 2.3 方面注意力层

用户在不同的领域会关注不同的方面,只有突出用户在源域中更加关注的方面才能对模型的学习更有帮助,因此需要将用户更加关注的方面赋予更高权重.

以用户  $u$  为例,使用缩放点积注意力计算每个胶囊对每个方面的注意力权重,并对胶囊进行自注意力加权.通过共享参数非线性变换矩阵将胶囊矩阵投影到同一空间,得到注意力机制查询项  $\mathbf{Q}_u$  和键项  $\mathbf{K}_u$ .

$$\mathbf{Q}_u = \text{ReLU}(\mathbf{O}_u \mathbf{W}^Q) \quad (8)$$

$$\mathbf{K}_u = \text{ReLU}(\mathbf{O}_u \mathbf{W}^K) \quad (9)$$

其中  $\mathbf{W}^Q, \mathbf{W}^K \in \mathbb{R}^{d \times d}$  为非线性变换矩阵,ReLU 是非线性映射激活函数.

然后,通过查询项和键项计算胶囊的注意力权重  $\mathbf{W}^{\text{Attn}}$ ,使用胶囊维度开方  $\sqrt{h}$  对注意力进行缩放,可以避免查询项和键项之间产生较高的匹配分数,影响结果.

$$\mathbf{W}^{\text{Attn}} = \text{softmax}\left(\frac{\mathbf{Q}_u \mathbf{K}_u^T}{\sqrt{h}}\right) \quad (10)$$

最后,通过注意力权重对胶囊矩阵  $\mathbf{O}_u$  进行加权,计算出用户  $u$  对应的方面矩阵  $\mathbf{A}_u$ .

$$\mathbf{A}_u = \mathbf{W}^{\text{Attn}} \mathbf{O}_u \quad (11)$$

同理,将项目  $v$  在目标域的评论文档输入模型的前三层,可以获得项目  $v$  对应的方面矩阵  $\mathbf{A}_v$ .

### 2.4 知识迁移层

评论文档经过模型的前三层,就得到了用户  $u$  和项目  $v$  的方面矩阵  $\mathbf{A}_u$  和  $\mathbf{A}_v$ ,在知识迁移层中,将  $\mathbf{A}_u$  和  $\mathbf{A}_v$  中两个方面之间进行语义匹配来计算用户  $u$  对项目  $v$  的预测评分.

首先计算  $\mathbf{A}_u$  和  $\mathbf{A}_v$  之间每个方面对之间的语义匹配,公式如下:

$$\mathbf{S} = \mathbf{A}_u^T \mathbf{W}^{\text{Match}} \mathbf{A}_v \quad (12)$$

其中  $\mathbf{W}^{\text{Match}}$  是方面匹配映射矩阵,它编码了源域中每个方面和目标域中每个方面之间偏好迁移的重要性.  $\mathbf{S}_{p,q}$  反映了相应方面对  $(p, q)$  之间的匹配程度.

最后,将  $\mathbf{S}$  中每一对语义匹配求和并取其平均值,考虑到用户和项目偏置,计算最终的预测评分,公式定义如下:

$$\hat{r}_{u,v} = \frac{1}{M \cdot M} \sum_{p=1}^M \sum_{q=1}^M \mathbf{S}(p, q) + \mathbf{b}_u + \mathbf{b}_v \quad (13)$$

其中,  $\mathbf{b}_u$  是用户偏置项,  $\mathbf{b}_v$  是项目偏置项.

### 2.5 训练及优化

本文将 ACN 模型训练过程的损失函数定义为:

$$\text{Loss} = \frac{1}{\|\mathbf{R}\|} \sum_{(u,v) \in \mathbf{R}} (\hat{r}_{u,v} - r_{u,v})^2 + \mathbf{b}_r \quad (14)$$

其中  $\mathbf{b}_r$  是评分偏置项.为了更好地学习跨域迁移知识,本文设置一对相关领域互为源域和目标域,在训练过程中依次执行两个学习流程.本文采用 Adam<sup>[19]</sup> 优化器对损失函数进行优化,Adam 优化器可以在训练过程中自动调整各个参数的学习速率,并且比普通的 SGD 优化器更准确、收敛更快.

## 3 实验

### 3.1 数据集和评估标准

在本文的实验中,使用 Amazon5-cores<sup>[20]</sup> 中的 Book、Movie (Movies and TV) 和 Music (CDs and Vinyl) 三个彼此相关以及 Toy (Toys and Games)、Tool (Tools and Improvement) 和 Beauty 三个彼此不相关的数据集.每个数据集中包含用户编号、项目编号、用户对项目的评分、用户对项目的评论、评论投票等 9 种数据类型,本文使用用户编号、项目编号、评论和评分 4 种数据类型.各个数据集统计如表 2 所示.

表 2 数据集统计

数据集	用户	项目	评论	用户平均评论	项目平均评论
Book	126666	63202	3494976	27.6	55.3
Movies and TV	27822	12287	779376	28.0	63.4
CDs and Vinyl	11053	7710	296188	26.8	38.4
Toys and Games	19412	11924	167597	8.6	14.1
Tools and Improvement	16638	10217	134476	8.1	13.2
Beauty	22363	12101	198475	8.9	16.4

由于跨域推荐模型能够缓解数据稀疏所引起的冷启动问题,本节在源域和目标域的重叠用户中,随机选择 50% 的用户作为推荐系统的冷启动用户,在模型验证和测试过程均不需要他们在目标域中的交互历史(测试集包含 30%,验证集包含 20%),并将剩下的 50% 的重叠用户用于模型训练过程,实验中训练集、测试集和验证集的比值为 5:3:2.

本文除了将 ACN 模型与基准模型进行纵向对比,还将重叠用户以  $\eta \in \{100\%, 50\%, 20\%, 10\%, 5\%\}$  的比例构建训练集,将本文模型和基准模型同时在这五种比例的数据集中进行训练,并将结果进行横向对比,证明 ACN 缓解数据稀疏引起冷启动问题的有效性.

在实验过程中以 MSE(均方误差)作为模型实验效

果的评判标准, MSE 越低, 模型性能越高. MSE 的公式定义如下:

$$\text{MSE} = \frac{1}{\|\mathbf{R}\|} \sum_{(u,v) \in \mathbf{R}} (\hat{r}_{u,v} - r_{u,v})^2 \quad (15)$$

为了验证模型的有效性, 实验选取以下 5 个基准模型与 ACN 进行对比:

(1) CMF<sup>[5]</sup>: 对源域和目标域均进行因子分解, 在因子之间共享参数来实现跨域推荐.

(2) EMCDCR<sup>[6]</sup>: 使用多层感知机学习不同域间的非线性映射函数, 并且将源域知识迁移给目标域来生成推荐.

(3) R-DFM<sup>[11]</sup>: 使用堆叠式去噪自动编码器来挖掘源域和目标域的特征, 然后利用多层感知器将其在两个域之间传递.

(4) CATN<sup>[12]</sup>: 通过方面门机制提取用户和项目的方面, 并且增加用户辅助文档的最新跨域推荐模型.

(5) ANR<sup>[21]</sup>: 挖掘用户和项目方面的最新单域推荐模型. 本文使其直接在目标域中参与实验.

### 3.2 模型对比

本文将 Book、Movie 和 Music 三个数据集两两组合, 分别组成三对源域和目标域 (分别是 Book→Movie、Movie→Music 和 Book→Music), 对应三个跨域推荐任务. 本节在这三个跨域推荐任务中对本文模型 ACN 以及基准模型进行了对照试验, 实验结果如表 3 所示.

基于矩阵分解方面建模方法 (如 CMF), CMF 通过对源域和目标域中的评分矩阵进行矩阵分解来学习用户和项目的特征, 但忽略了其他有用的信息, 故其迁移给目标域的知识不够完整, 所以在基准模型中性能是最差的. ACN 模型在知识迁移层也是矩阵相乘, 但这两个矩阵并非原始矩阵分解而来, 而是通过胶囊网络分别对卷积得到的用户和项目这两个矩阵进行处理, 而且在 ACN 模型中矩阵相乘的目的并非为了还原原始矩阵并得到原始矩阵中的丢失值, 而是为了迁移到目标域进行评分预测, 这与矩阵分解本质上是不同的.

EMDCDR 使用了多层感知机捕获源域和目标域之间的非线性映射, 效果比 CMF 使用的矩阵分解要好, 但还是没有利用到其他辅助信息, 故与其他模型相比性能较差; R-DFM 扩展了堆叠式降噪自动编码器, 并且将评论文本和项目描述内容与评分矩阵相融合, 在学习到的用户和项目的潜在因素中保留了更多的语义信息, 但自动编码器并不能准确地提取出文本中的细粒度特征, 故在一些数据集上甚至不如仅使用评分矩阵的两个模型, 在其他数据集上的表现也并不是最好的; ANR (Aspect-based Neural Recommender) 使用注意力机制对评论文档进行方面表示学习, 在目标域的评论文档中可以挖掘到对推荐有重要作用的信息, 在单域和跨域模型中均有较好的效果, 但其挖掘到的特征并不完整, 且当用户在目标域冷启动时, 就无法从目标域中提取出相关用户的偏好; CATN 是目前最新的跨域推荐模型, 它使用方面门以及注意力机制提取用户和项目的方面, 并且为冷启动用户额外增加了辅助文档, 挖掘出的偏好更加完整, 性能在本文选取的基准模型中最好, 但 CATN 方面提取的方法是通过反向传播对模型进行更新, 没有引入其他监督信息, 导致挖掘出的偏好容易产生过拟合, 并且 CATN 没有考虑到多源域评论文档对目标域推荐的影响; ACN 在三个推荐任务中的性能均高于其他模型, 且在三个推荐任务中相对于性能最高的基准模型获得了 2.3%、0.8% 和 2.2% 的提升.

仅通过纵向与基准模型对比不能充分体现 ACN 缓解数据稀疏引起冷启动问题的能力, 因此本节将模型在每个推荐任务上的效果进行横向对比. 如图 2 所示.

通过图 2 可以看出, 在数据逐渐稀疏的情况下, 基准模型 EMCDCR、R-DFM 和 CATN 的 MSE 大幅上升, 而本文模型 ACN 在 Book→Movie 和 Book→Music 上的 MSE 涨幅明显小于以上三个基准模型, 并且涨幅与基准模型 CMF 以及 ANR 基本一致. 产生这样效果的原因是 CMF 仅使用评分数据作为模型输入, 而 ANR 则是单

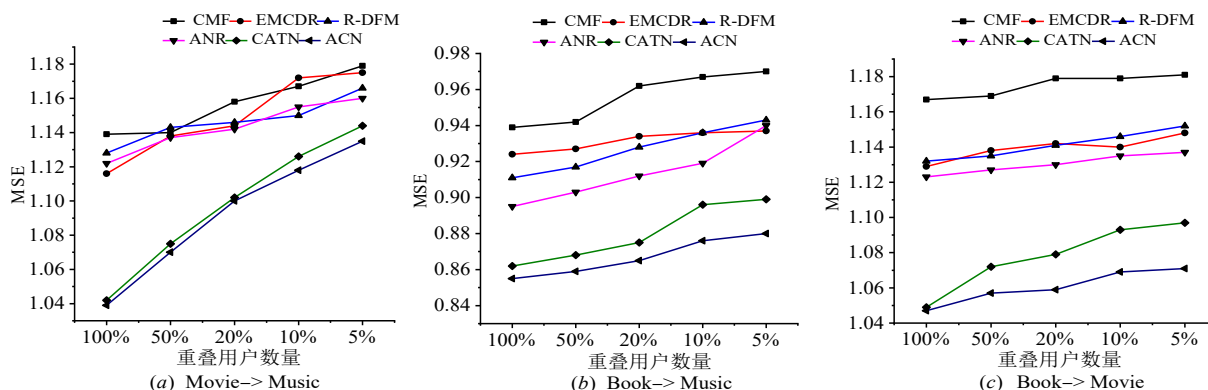


图 2 推荐性能随用户数的变化

表 3 各模型实验结果对比

	重叠用户数	CMF	EMCDR	R-DFM	ANR	CATN	ACN
Book	100%	1.167	1.129	1.132	1.123	1.049	<b>1.047</b>
	50%	1.169	1.138	1.135	1.127	1.072	<b>1.057</b>
	20%	1.179	1.142	1.141	1.130	1.079	<b>1.059</b>
Movie	10%	1.179	1.140	1.146	1.135	1.093	<b>1.069</b>
	5%	1.131	1.148	1.152	1.137	1.097	<b>1.071</b>
Movie	100%	1.139	1.116	1.128	1.122	1.042	<b>1.039</b>
	50%	1.140	1.138	1.143	1.137	1.075	<b>1.070</b>
	20%	1.158	1.144	1.146	1.142	1.102	<b>1.100</b>
Music	10%	1.167	1.172	1.150	1.155	1.126	<b>1.118</b>
	5%	1.173	1.175	1.166	1.160	1.144	<b>1.135</b>
	100%	0.939	0.924	0.911	0.895	0.862	<b>0.855</b>
Book	50%	0.942	0.927	0.917	0.903	0.868	<b>0.859</b>
	20%	0.962	0.934	0.928	0.912	0.875	<b>0.865</b>
	10%	0.967	0.936	0.936	0.919	0.896	<b>0.876</b>
5%	0.970	0.937	0.943	0.940	0.899	<b>0.880</b>	

域推荐模型,其效果与重叠用户的稀疏度无关.证明了ACN缓解冷启动问题的有效性.

综上所述ACN可以更准确地挖掘用户和项目的方面,并且可以将将在源域中学习到的知识迁移到目标域,能更有效地提高冷启动用户在目标域预测评分的准确率.

### 3.3 不相关源域对模型效果的影响

本节将 Toy、Tool 和 Beauty 这 3 个不相关的数据集组成三对源域和目标域 (Toy->Tool、Toy->Beauty 和 Tool->Beauty),对应于三个跨域推荐任务.将本文模型 ACN、CATN 以及 ANR 依次在三个跨域推荐任务中进行对照试验,证明 ACN 使用不相关的源域和目标域进行迁移信息的有效性.实验结果统计如图 3 所示.

ACN 使用与目标域不相关的数据集作为源域时,在三个推荐任务上的性能均高于单域推荐模型 ANR,并且性能表现略强于跨域推荐模型 CATN.实验结果一方面证明 ACN 在使用不相关源域评论文档时,仍然可以挖掘出有用的信息并迁移给目标域,提升推荐的准确度.另一方面通过与 ANR 的性能对比,证明 ACN 可以有效缓解用户在目标域冷启动的问题.

### 3.4 多源域 ACN

本节对 ACN 在单一源域和多个源域进行训练时的性能差异进行比较.本节将三个相关的数据集 (Book、Music 和 Movie) 和三个不相关的数据集 (Toy、Tool 和 Beauty) 进行组合,形成两个跨域推荐任务:Book+Movie->Music 和 Toy+Tool->Beauty,并在这两个推荐任务上对 ACN 模型进行训练,将实验结果与 ACN 使用单

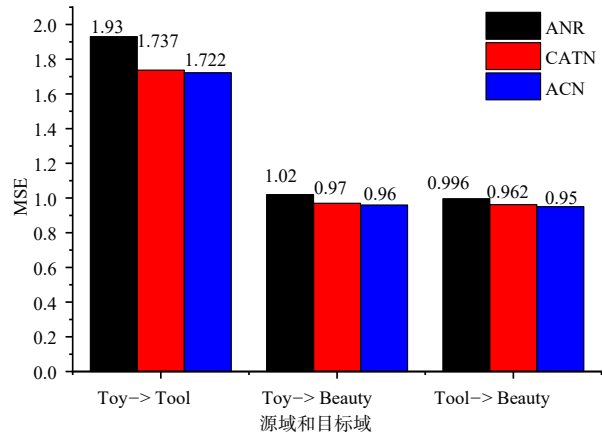


图 3 使用与目标域不相关源域时模型的性能

一源域时的性能进行对比,实验结果如图 4 和图 5 所示.

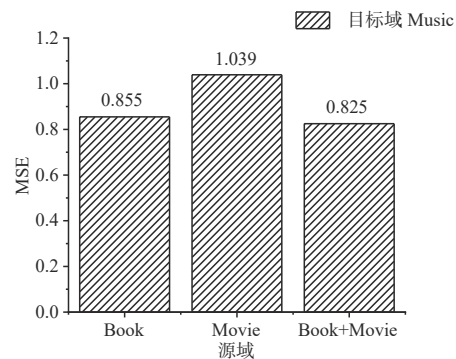


图 4 相关多源域和单源域实验结果对比

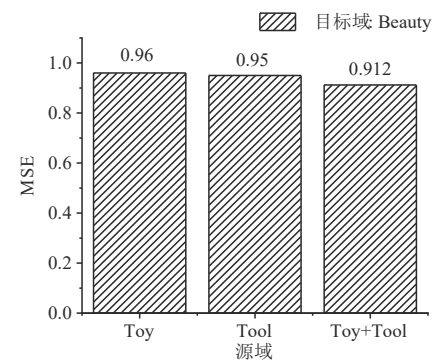


图 5 不相关多源域和单源域实验结果对比

从图 4 可以看出,ACN 使用多个源域进行跨域推荐的效果远远好于使用单个源域的效果.当目标域是 Music 时,ACN 使用 Book 与 Movie 的多源域比使用单源域的性能提高了 3.5% 和 20.6%;当目标域是 Beauty 时,ACN 使用 Toy 和 Tool 的多源域比使用单源域的性能提高了 5.0% 和 4.0%.这说明无论源域与目标域类别相关或是不相关,使用多个源域都可以大幅度提升模型

的性能. 与 3.2 节的实验对比, 使用多源域的 ACN 比使用单源域的 CATN 性能提高了 4.3% 和 20.8%. 本节实验证明了 ACN 不仅可以迁移多个相关源域的知识到目标域, 增强用户偏好的偏好信息, 当源域与目标域不相关时, 还可以通过整合从多个源域中学习到的知识, 提高目标域推荐的准确度.

### 3.5 消融实验

为了验证 ACN 模型使用胶囊网络挖掘评论文档多个方面以及通过自注意力机制筛选特征的合理性, 使用 ACN 的两个子模型进行消融试验, 两个子模型定义如下:

ACN-cnn: 将 ACN 模型的卷积胶囊层替换为一次卷积和池化操作的 CNN, 用于证明胶囊网络可以更有有效的挖掘评论文档的多个方面.

ACN-noattn: 去掉 ACN 的自注意力机制, 注意力权重用维度相同数值全是 1 的矩阵代替, 用于证明自注意力机制能够更好的筛选出对目标领域最重要的特征.

将 ACN-cnn 和 ACN-noattn 在 Movie→Music 数据集上, 重叠用户以 100%, 50%, 20%, 10%, 5% 的比例进行实验, 并与 ACN 模型进行对比, 实验结果如表 4 所示.

表 4 各模型实验结果对比

	重叠用户	ACN-noattn	ACN-cnn	ACN
Movie	100%	1.054	1.114	<b>1.039</b>
	50%	1.082	1.131	<b>1.070</b>
↓	20%	1.130	1.196	<b>1.100</b>
Music	10%	1.131	1.185	<b>1.118</b>
	5%	1.136	1.189	<b>1.135</b>

首先, 使用胶囊网络和自注意机制的 ACN 模型的性能高于使用 CNN 和自注意力机制的 ACN-cnn 模型, 说明胶囊网络能够更好地挖掘评论文档的多个方面, 而且 ACN 模型的性能也要优于使用胶囊网络的 ACN-noattn 模型, 这说明自注意力机制能够更好的筛选出对目标领域最重要的特征. 其次, 使用胶囊网络的 ACN-noattn 模型明显要好于使用自注意力机制的 ACN-cnn 模型, 即不使用胶囊网络模型效果下降的更多, 说明 ACN 模型中胶囊网络比自注意力机制更为重要. 纵向来看, 在数据逐渐稀疏的情况下, ACN-noattn 的 MSE 要明显低于 ACN-cnn 的 MSE, 这说明胶囊网络能够更好的应对数据稀疏引起的冷启动问题, 从而我们模型在面对冷启动问题上表现得更好.

## 4 结论

本文提出了一个基于方面和胶囊网络的跨域评分预测模型 ACN, 使用胶囊网络挖掘源域和目标域中的用户或项目方面. 为了更好地契合跨域推荐任务, 本文设置了全局共享映射矩阵来表示低层胶囊和高层胶囊

之间的映射关系, 使模型能更准确地迁移并融合从源域和目标域挖掘出的方面. 经过与各种基准模型的实验对比, 证明了本文的 ACN 模型可以更准确地提取并迁移评论文档中的特征, 并且提升评分预测的准确率. 此外, 通过 ACN 使用不相关源域与其他基准模型的对比实验, 证明 ACN 使用与目标域不相关的源域, 仍可以学习到有用的信息. 本文还将多个源域进行组合并对照试验, 证明了与单个源域相比, 多个源域可以提升模型的性能.

在未来的工作中考虑通过其他上下文信息完善模型对推荐结果的可解释性.

### 参考文献

- [1] LUO X, ZHOU M C, LI S, et al. Algorithms of unconstrained non-negative latent factor analysis for recommender systems[J]. IEEE Transactions on Big Data, 2021, 7(1): 227-240.
- [2] LUO X, WANG D X, ZHOU M C, et al. Latent factor-based recommenders relying on extended stochastic gradient descent algorithms[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(2): 916-926.
- [3] WU D, LUO X, SHANG M S, et al. A deep latent factor model for high-dimensional and sparse matrices in recommender systems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(7): 4285-4296.
- [4] 李林峰, 刘真, 魏港明, 等. 基于共享知识模型的跨领域推荐算法[J]. 电子学报, 2018, 46(8): 1947-1953.  
LI L F, LIU Z, WEI G M, et al. Cross-domain recommendation algorithm based on sharing knowledge pattern[J]. Acta Electronica Sinica, 2018, 46(8): 1947-1953. (in Chinese)
- [5] SINGH A P, GORDON G J. Relational learning via collective matrix factorization[C]//Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2008: 650-658.
- [6] MAN T, SHEN H W, JIN X L, et al. Cross-domain recommendation: an embedding and mapping approach[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2017: 2464-2470.
- [7] ZHU Feng, WANG Yan, CHEN Chaochao, et al. A deep framework for cross-domain and cross-system recommendations[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: AAAI Press, 2018: 3711-3717.
- [8] 刘真, 田靖玉, 苑宝鑫, 等. 基于知识聚合和迁移的跨领

- 域推荐算法[J]. 电子学报, 2020, 48(10): 1928-1932.
- LIU Z, TIAN J Y, YUAN B X, et al. Cross-domain recommendation algorithm based on knowledge aggregation and transfer[J]. Acta Electronica Sinica, 2020, 48(10): 1928-1932. (in Chinese)
- [9] ELKAHKY A M, SONG Y, HE X D. A multi-view deep learning approach for cross domain user modeling in recommendation systems[C]//WWW' 15: Proceedings of the 24th International Conference on World Wide Web. New York: ACM, 2015: 278-288.
- [10] SONG T H, PENG Z H, WANG S Z, et al. Review-based cross-domain recommendation through joint tensor factorization[C]//International Conference on Database Systems for Advanced Applications. Cham: Springer, 2017: 525-540.
- [11] FU W J, PENG Z H, WANG S Z, et al. Deeply fusing reviews and contents for cold start users in cross-domain recommendation systems[C]//The Thirty-Third AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2019: 94-101.
- [12] ZHAO C, LI C L, XIAO R, et al. CATN: cross-domain recommendation for cold-start users via aspect transfer network[C]//SIGIR' 20: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020: 229-238.
- [13] CHENG Z Y, DING Y, ZHU L, et al. Aspect-aware latent factor model: Rating prediction with ratings and reviews[C]//WWW' 18: Proceedings of the 2018 World Wide Web Conference. New York: ACM, 2018: 639-648.
- [14] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[EB/OL]. (2014-09-01). <https://arxiv.org/abs/1409.0473>.
- [15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2017: 6000-6010.
- [16] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2017: 3859-3869.
- [17] 郑启航, 王章权, 刘半藤, 等. 基于加性间距胶囊网络的家庭活动识别方法研究[J]. 电子学报, 2020, 48(8): 1580-1586.
- ZHENG Q H, WANG Z Q, LIU B T, et al. Research on family activity recognition method based on additive margin capsule network[J]. Acta Electronica Sinica, 2020, 48(8): 1580-1586. (in Chinese)
- [18] 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)
- [19] KINGMA D P, BA J. ADAM: A method for stochastic optimization[EB/OL]. (2014-12-22). <https://arxiv.org/abs/1412.6980>.
- [20] MCAULEY J, TARGETT C, SHI Q F, et al. Image-based recommendations on styles and substitutes[C]//Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2015: 43-52.
- [21] CHIN J Y, ZHAO K Q, JOTY S R, et al. ANR: aspect-based neural recommender[C]. Proceedings of the 27th ACM International Conference on Information and Knowledge Management. New York: ACM, 2018. 147-156

#### 作者简介



梁顺攀 男, 1976年6月出生于广东省江门市, 现为燕山大学副教授、硕士生导师, 研究方向为推荐系统. 在国内外发表学术论文20余篇.  
E-mail: liangshunpan@ysu.edu.cn

刘伟 男, 1996年11月出生于河北省新乐市, 燕山大学硕士生, 研究方向为推荐系统.

郑智中 男, 1997年2月出生于河北省保定市, 现为燕山大学硕士生, 研究方向为推荐系统.

原福永 男, 1958年12月出生于辽宁省丹东市, 燕山大学教授. 研究方向为推荐系统.