

# 基于双注意力机制和迁移学习的跨领域推荐模型

柴玉梅<sup>1)</sup> 员武莲<sup>1)</sup> 王黎明<sup>1)</sup> 刘箴<sup>2)</sup>

<sup>1)</sup>(郑州大学 信息工程学院, 郑州 450001)

<sup>2)</sup>(宁波大学 信息科学与工程学院, 浙江宁波 315211)

**摘 要** 跨领域推荐可用于解决单一领域数据稀疏导致的推荐系统性能退化问题, 还可以缓解推荐系统中存在的用户冷启动问题。然而, 现有的方法大多利用用户对项目的评分进行建模, 忽略了评论文本所蕴含的信息。为此, 本文提出了一种基于双注意力机制和迁移学习的跨领域推荐模型, 首先通过 CNN 对评论文本建模, 提取用户和项目特征; 其次通过构造融合词的上下文关系的词注意力机制从评论文本中捕获词级别的信息, 以提升 CNN 对文本中重点信息的关注度; 然后通过构造特征突显机制从 CNN 提取到的用户特征和项目特征中捕获特征级别的信息; 最后引入迁移学习, 通过同时提取领域特有的特征和领域间的共享特征进行不同领域之间的联合建模, 进行评分预测。本文在 Amazon 数据集上进行了实验比较与分析, 首先对本文模型的推荐性能进行评估, 与现有的跨领域推荐模型相比, 在两种不同的跨领域数据集上平均绝对误差分别提升 6.1% 和 9.15%, 均方根误差分别提升 3.66% 和 7.01%; 然后对本文模型的知识迁移性能进行评估, 与现有的单领域推荐模型相比, 在不同数据集下均方误差分别提升 5.47% 和 10.35%; 最后通过实验验证了本文提出的注意力机制的有效性, 及在缓解数据稀疏问题和用户冷启动问题方面的优势, 也验证了模型的普适性。

**关键词** 推荐; 迁移学习; 跨领域; CNN; 评论文本

中图法分类号 TP391

## A Cross-Domain Recommendation Model Based on Dual Attention Mechanism and Transfer Learning

CHAI Yu-Mei<sup>1)</sup> YUN Wu-Lian<sup>1)</sup> WANG Li-Ming<sup>1)</sup> LIU Zhen<sup>2)</sup>

<sup>1)</sup>(School of Information Engineering, Zhengzhou University, Zhengzhou 450001)

<sup>2)</sup>( School of Information Science and Technology, Ningbo University, Zhejiang Ningbo 315211)

**Abstract** Cross-domain recommendation can be used to solve the problem of degrading the performance of the recommendation system caused by sparse data in a single domain, and it can also alleviate the cold start problem of users in the recommendation system. However, most of the existing methods use the user's rating data to model the item, ignoring the review text written by the user for the item and the rich user and item information it contains. In recent years, deep learning has been successfully applied to various fields. Inspired by this, this article proposes a Cross-Domain Recommendation Model based on the Dual Attention Mechanism and Transfer Learning (AMTR) based on the review text. Firstly, modeling review text through convolutional neural network, extract user and item features. Secondly, this paper constructs the word attention mechanism and feature highlighting mechanism that fused the context of words, and proposes a feature extraction network based on the dual attention mechanism. The word attention mechanism captures word-level information from the review text

投稿日期: 2019-09-28, 在线发布日期: 2020-03-20. 本课题得到国家自然科学基金 (No.U1636111) 资助。柴玉梅, 女, 1964年生, 教授, 主要研究领域为机器学习、数据挖掘、自然语言处理。E-mail: ieymchai@zzu.edu.cn。员武莲(通信作者), 女, 1995年生, 硕士研究生, 主要研究领域为自然语言处理。E-mail: yunwulll@163.com。王黎明(通信作者), 男, 1963年生, 博士, 教授, 主要研究领域为现代软件工程技术、分布式人工智能和数据挖掘等。E-mail: ielmwang@zzu.edu.cn。刘箴, 男, 1965年生, 博士, 研究员, 主要研究领域为虚拟现实、情感分析等。E-mail: liuzhen@nbu.edu.cn。

to increase CNN's attention to the key information in the text, while making the recommendation interpretable; The feature highlighting mechanism captures feature level information that is helpful for rating prediction from user features and item features extracted by CNN. Finally, based on the feature extraction network, transfer learning is introduced to jointly model between different domains to achieve knowledge transfer between domains. In this process, feature extraction networks in different domains are used to simultaneously extract domain-specific features and share features and combine with factorization machine to perform rating prediction to achieve cross-domain recommendation. In this paper, the experimental comparison and analysis is carried out on the Amazon dataset. The experimental results show that the performance of the AMTR model in terms of mean absolute error, root mean square error and mean square error is better than the comparison model. Firstly, evaluate the recommend performance of the AMTR model, compared with the existing cross-domain recommendation model, the mean absolute error of the AMTR model on both different cross-domain datasets increased by 6.1% and 9.15%, respectively, and the root mean square error is improved 3.66% and 7.01%. Secondly, the knowledge transfer performance of the AMTR model is evaluated. Compared with the existing single domain recommendation model, the mean square error of the AMTR model increased by 5.47% and 10.35% in different datasets. Then, the effectiveness of the attention mechanism proposed in this paper is verified through related experiments, and by controlling the number of reviews, the advantages of the proposed model in mitigating data sparseness and user cold start problems are verified.; Finally, the universality of the model is verified on various data sets. In addition, this paper also considers the impact of the difference in scoring scales on the model in this paper. This paper makes full use of review text in different domains, effectively mines user and item information in different domains through neural networks, improves the performance of cross-domain recommendation, and expands the new way of using review text information to a certain extent.

**Key words** recommendation; transfer learning; cross-domain; CNN; review text

## 1 引言

推荐系统<sup>[1]</sup>可以缓解当下大数据时代中存在的“信息过载”问题,它帮助用户从繁杂的信息中挑选出需求的信息,为用户提供个性化服务。传统的推荐都是基于单一领域,但随着数据的急剧增长,单一领域的信息难以满足用户的需求,而广泛应用于单领域的协同过滤方法<sup>[2-3]</sup>虽然简单有效,却存在评分数据稀疏和用户冷启动问题<sup>[4]</sup>,使得推荐性能下降,解决这种问题的方法之一是跨领域推荐。

跨领域推荐<sup>[5-6]</sup>旨在利用其它领域的用户偏好和项目特征等信息来帮助目标领域提高推荐的准确性,使得目标领域能对用户或项目进行更加充分的建模,有效的缓解了数据稀疏和用户冷启动问题。研究与事实也表明,来自不同领域的用户和项目之间有较大的关联,例如:用户在不同的领域也会表现出相同的爱好,如喜爱恐怖类电影的用户也会倾向于购买恐怖类的书籍。

现有的大部分跨领域推荐工作主要是利用用

户对项目的评分数据<sup>[7-8]</sup>,这种方法虽然一定程度上缓解了数据稀疏和用户冷启动问题,但在数据高度稀疏的情况下,仅依靠评分信息,推荐系统难以全面的对用户和项目建模,无法学习出可以进行知识迁移的有效信息,并且难以将源域的知识充分的迁移到目标域。

相比于评分仅代表用户对项目的总体态度来说,评论文本在反应用户喜好的同时,可以解释给出评分高低的原因,为用户和项目的建模提供丰富的信息,同时可以补充利用评分进行知识迁移过程中有效信息不足的缺点,提高跨领域推荐效果,这是以往跨领域推荐工作所忽视的。

然而评论文本结构复杂,很难直接被推荐系统所理解,因此如何利用自然语言处理技术帮助推荐系统解析评论,提取出评论文本包含的信息,同时使得推荐具有可解释性以及如何将这些信息在领域之间进行迁移是本文研究的重点。

受 Liu 等人<sup>[9]</sup>利用迁移学习<sup>[10]</sup>方法提高文本分类性能的启发,本文提出了一种端到端的基于双注意力机制和迁移学习的跨领域推荐模型 AMTR

(Attention Mechanism and Transfer Learning), 主要工作如下:

(1) 本文利用 CNN 分别对用户和项目对应的评论聚合文本进行特征提取, 相较于传统的特征提取方法能够提取更为充分地信息, 并对用户和项目的特征进行更好地建模。

(2) 本文提出两种不同的注意力机制, 分别从词级别和特征级别两种视角提取出评论文本中的重要信息。词注意力机制 (Words Attention Mechanism, WAM) 通过融合词的上下文信息, 衡量单词的重要性, 过滤出相对于全局重要的单词, 同时增加推荐的可解释性。特征突显机制 (Feature Highlighting Mechanism, FHM) 通过抑制相似通道, 捕捉特征通道中独立性较强的通道。从而提高模型对评论文本中的重要信息的关注, 使其具有更好的特征提取能力。

(3) 本文提出了一种新的跨领域推荐模型, 联合建模源域、目标域和共享域, 利用不同领域的神经网络同时提取领域特有的特征和领域间的共享特征, 有效地学习到了可以进行知识迁移的信息, 将源域的知识充分地迁移到目标域, 提高了推荐性能, 并缓解了数据稀疏问题和用户冷启动问题。

此外, 本文还针对用户之间的评分尺度差异问题进行了相应的研究和实验验证。

## 2 相关工作

跨领域推荐采用知识迁移的策略将与目标领域关联较为密集的辅助领域的知识迁移到目标领域, 提高目标领域的推荐性能, 有效的缓解了传统推荐系统中存在的数据稀疏和用户冷启动问题。Li 等人<sup>[11]</sup>提出了一个密码本迁移模型 CBT (CodeBook Transfer), 它假设不同领域共享一个评分模式并将其称为“密码本”, 将密码本从辅助领域迁移到目标领域缓解数据稀疏问题, 但是这种方法忽略了不同领域的评分模式具有差异性, 可能会导致负迁移。Gao 等人<sup>[12]</sup>在 CBT 基础上考虑不同领域拥有各自的特异性提出了 CLFM 模型, 在学习所有领域共享的评分模式的同时, 还学习领域特有的评分模式, 提高了跨领域推荐的准确度。Moreno 等人<sup>[13]</sup>考虑不同辅助领域和目标领域之间的关系, 根据不同领域的相关性自动学习不同程度的知识。

王俊等人<sup>[14]</sup>在考虑领域具有共享和特有的评分模式后, 考虑了项目的共享被评分模式, 提出一种三元桥迁移学习模型 (Triple-Bridge Transfer, TRBT), 增加了正迁移。Hu 等人<sup>[15]</sup>考虑领域和领域之间的潜在关系, 把不同领域的评分矩阵组成张量, 通过张量分解捕获特征。随后 Hu 等人又提出了 WITF (Weighted Irregular Tensor Factorization) 模型<sup>[16]</sup>, 利用所有用户的多领域反馈数据来学习用户和项目的跨领域先验, 从而提高推荐性能。Lian 等人<sup>[17]</sup>通过引入一个因子分解框架将协同过滤和基于内容的过滤结合, 随后将其嵌入到一个多视图神经网络实现跨领域推荐。Hu 和 Zhang 等人<sup>[18]</sup>采用神经网络学习各领域的潜在特征以及复杂的用户-项目交互关系, 实现不同领域之间的知识传递。

合理的使用评论文本进行评分预测从而提高推荐系统性能已被证明是有效的<sup>[19-20]</sup>。Zheng 等人<sup>[21]</sup>利用两个并行神经网络, 学习目标用户撰写的所有评论文本的潜在特征得到用户行为, 以及项目的所有评论文本的潜在特征得到项目属性, 然后结合这些潜在的特征提出 DeepCoNN (Deep Cooperative Neural Networks) 模型。Catherine 等人<sup>[22]</sup>在 DeepCoNN 模型上扩展一个 transform 层, 将用户和项目的潜在特征转换为目标评论的近似表示并用于评分预测, 构建出新的评分预测模型。Seo 等人<sup>[23]</sup>在 CNN 基础上提出了通过注意力机制建模评论中不同部分与用户行为和项目属性的关联度, 从而构造潜在表示。Tay 等人<sup>[24]</sup>提出了一种共同关注的多指针学习方案的深度学习推荐框架, 使最具信息性的评论能够用于评分预测同时实现更深层次的单词级交互。

## 3 加入双注意力机制的特征提取网络

为了从评论文本中获得用于评分预测的用户特征和项目特征。本文首先构造特征提取网络进行文本特征提取, 将 CNN 作为特征提取的基础, 相较于传统的机器学习方法, CNN 具有提取复杂特征的优势, 可以自动提取评论文本中的语义特征, 不依赖于人工, 从而有效的避免了人工提取特征的不足, 学习到更鲁棒的特征表示。为了使模型有更好的特征提取能力, 提出词级别的 WAM 以及特征级别的 FHM。如图 1 所示, 特征提取网络首先将用户和项目对应评论文本的词向量通过 WAM, 加强对

评论文本的重点信息的关注, 其次使用 CNN 提取特征, 最后通过 FHM 捕获更多重要信息。

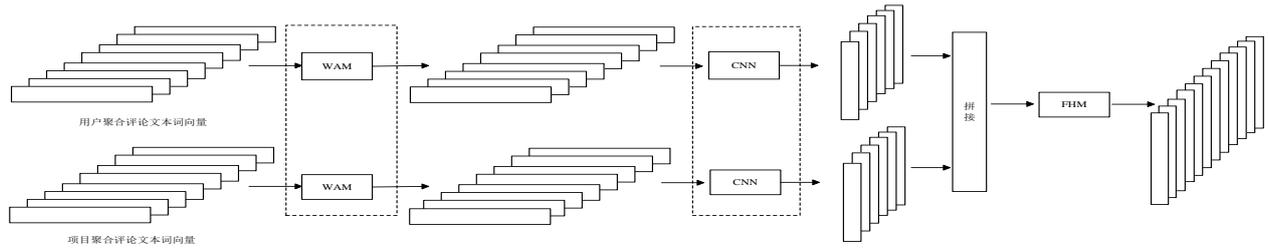


图1 特征提取网络

### 3.1 相关定义

为了方便阅读, 本节主要介绍一些模型中用到的定义。

**定义 1.** 数据表示. 输入模型的数据表示可以描述为一个四元组  $\{U, I, X_{UI}, R_{UI}\}$ , 其中  $U$  为用户、 $I$  为项目、 $X_{UI}$  为用户  $U$  对项目  $I$  所做的评论、 $R_{UI}$  为用户  $U$  对项目  $I$  的评分。

**定义 2.** 评论聚合文本. 某个用户  $u$  为项目  $I$  撰写的所有评论文本或某个项目  $i$  拥有的用户  $U$  为其撰写的所有评论文本。

**定义 3.** 用户行为. 一个用户的行为可以从它为项目撰写的评论中体现, 可以描述为一个二元组  $\{u, D^u\}$ , 其中  $u$  表示用户,  $D^u$  表示用户  $u$  撰写评论的集合,  $D^u = \{X_1, X_2, X_3, \dots, X_m\}$ , 其中  $m$  表示用户撰写评论的个数。

**定义 4.** 项目属性. 一个项目的属性可以从用户为其撰写的评论集合中体现, 可以描述为一个二元组  $\{i, D^i\}$ , 其中  $i$  表示项目,  $D^i$  表示用户为项目  $i$  撰写评论的集合,  $D^i = \{X_1, X_2, X_3, \dots, X_p\}$ , 其中  $p$  表示项目拥有的评论的个数。

### 3.2 用户和项目评论聚合文本的词向量表示

模型使用自然语言处理中词向量<sup>[25]</sup>表示文本的方法将评论文本词向量化, 然后通过神经网络提取评论文本的语义信息。

不同于传统词袋模型, 为了保留单词的顺序, 设每条评论  $X = \{x_1, x_2, x_3, \dots, x_n\}$ 。通过 Glove<sup>[26]</sup> 获得评论  $X$  中每个词的词向量, 得到评论  $X$  的句向量  $e = \{e_{x_1}, e_{x_2}, e_{x_3}, \dots, e_{x_n}\}$ 。根据定义 3, 对于用户  $u$  撰写的评论来说, 其撰写  $m$  条评论, 将这些评论聚合为一个文档  $d_{1,m}^u$ , 记为:

$$d_{1,m}^u = e_1 \oplus e_2 \oplus e_3 \oplus \dots \oplus e_m \quad (1)$$

根据定义 4, 项目  $i$  的评论聚合文本的表示为文档  $d_{1,p}^i$ , 其中  $p$  条评论分别来自不同的用户  $u$ ,  $d_{1,p}^i$

记为:

$$d_{1,p}^i = e_1 \oplus e_2 \oplus e_3 \oplus \dots \oplus e_p \quad (2)$$

评论聚合文本的词向量生成的具体过程如算法 1 所示。

**算法 1.** 评论聚合文本词向量生成

输入: 评论文本集合  $\{\{U, I, X_{UI}, R_{UI}\}_{(U, I)}\}$

输出: 用户评论聚合文本  $d^u$ 、项目评论聚合文本  $d^i$

1. FOR 评论文本集合中的每一个用户  $u \in U$  DO
2.   IF  $u = \{U, I, X_{UI}, R_{UI}\}[0]$  THEN
3.     将评论  $X_{UI}$  词向量化生成句向量  $e_{UI}$
4.     执行操作(1):  $d^u = d^u \oplus e_{UI}$
5.   END IF
6. END FOR
7. FOR 评论文本集合中的每一个用户  $i \in I$  DO
8.   IF  $i = \{U, I, X_{UI}, R_{UI}\}[1]$  THEN
9.     将评论  $X_{UI}$  词向量化生成句向量  $e_{UI}$
10.     执行操作(1):  $d^i = d^i \oplus e_{UI}$
11.   END IF
12. END FOR
13. RETURN  $d^u, d^i$

其中,  $\{U, I, X_{UI}, R_{UI}\}[0]$  表示元组的第 1 个元素,  $\{U, I, X_{UI}, R_{UI}\}[1]$  表示元组的第 2 个元素。

### 3.3 融合上下文关系的WAM

受 Seo 等人<sup>[23]</sup>采用局部和全局注意力层学习单词权重的启发, 本文在词向量层和卷积层之间加入 WAM。WAM 通过融合词的上下文关系, 在原始评论文本中得到不同词在当前文本中的重要性, 使得 CNN 的训练重点集中在评论文本中与评分预测信息关联度高的部分, 由此加强对评论文本的重点部分的关注, 并且削弱了影响较小的词, WAM 如图 2 所示。

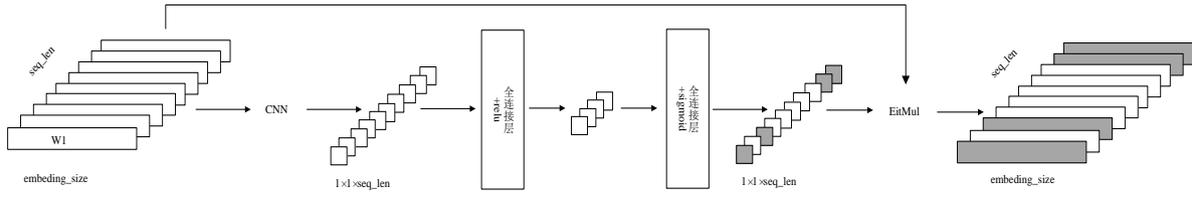


图2 WAM示意图

以用户评论聚合文本的词向量经过 WAM 为例，首先将用户评论聚合文本的词向量矩阵经过一个卷积层进行卷积操作。卷积层具有  $w$  个神经元，每个神经元  $j$  对应一个窗口大小为  $q$  的卷积核。将用户  $u$  的词向量文档  $d_{1:m}^u$  与卷积层的每个卷积核  $K_j$  进行卷积运算，压缩后的特征如下：

$$WA_j^u = f(d_{1:m}^u * K_j + WA_j) \quad (3)$$

本文将 WAM 的卷积操作视为一种参数可学习的特征压缩方法，因此将该卷积核的窗口大小与输出通道个数设置为 1。其中卷积核为  $K_j \in R^{C \times 1}$ ，\* 表示卷积操作， $WA_j$  表示偏置量， $f$  是  $ReLU$ <sup>[27]</sup> (Rectified Linear Units) 激活函数。

将压缩后的特征输入到权重生成块中得到对应词的权重因子，权重生成块包含两个全连接层。受 word2vec<sup>[28]</sup> 的启发，使用全连接层进行上下文关系的融合，将文本中每个词的上下文信息融合到隐层的每个结点。

第一层全连接层使用  $ReLU$  为激活函数，压缩长度为  $L$  的特征，输出长度为  $L/r$  的特征，其表示如下：

$$fc_1 = \text{relu}(W_d * WA_j^u + b_d) \quad (4)$$

其中  $W_d \in R^{L \times (L/r)}$  表示权重矩阵，其中  $L$  为词向量矩阵的宽， $r$  为超参数（本文设置  $r$  为 16），表示中间的特征衰减系数， $b_d \in R^{L/r}$  表示偏置量。

第二层全连接层使用  $sigmoid$  为激活函数，将经过第一层全连接层衰减后的特征维度复原，输出长度为  $L$  的权重向量，即每个单词对应的权重因子  $fc_2$ 。

$$fc_2 = \text{sigmoid}(W_u * fc_1 + b_u) \quad (5)$$

其中  $W_u \in R^{(L/r) \times L}$  表示权重矩阵， $b_u \in R^L$  表示偏置量。

将得出的权重因子与原始用户评论聚合文本的词向量逐通道相乘，得到带有权重因子的向量矩

阵，从而融合上下文关系，突出不同单词的重要性，向量矩阵表示如下：

$$WAM^u = d_{1:m}^u \otimes fc_2 \quad (6)$$

### 3.4 基于CNN的文本特征提取网络

本文采用 CNN 进行文本特征的自动提取，在 WAM 后利用卷积层以及最大池化进行特征提取。卷积层对经过 WAM 后的词向量文档进行卷积操作后得到新的特征。用户  $u$  的词向量文档  $WAM_{1:m}^u$  首先进行卷积运算，每个卷积操作后产生一个特征图，产生的特征图的公式如下：

$$z_j^u = f(WAM_{1:m}^u * K_j + b_j) \quad (7)$$

其中卷积核为  $K_j \in R^{C \times q}$ ， $q$  ( $q \in \{3, 4, 5\}$ ) 表示卷积核的窗口大小，\* 表示卷积操作， $b_j$  表示偏置量， $f$  是  $ReLU$  激活函数。

随后对特征图的每个通道进行最大池化操作，得出每个通道中最大的值  $o_j$ 。

$$o_j^u = \max \{z_1^u, z_2^u, \dots, z_{\sum_{i=1, \dots, m}^{n_i - q + 1}}^u\} \quad (8)$$

最大池化后，卷积特征被缩减为一个固定大小的向量，其维度和通道的个数相同。将  $w$  个神经元对应特征拼接得到特征，记为：

$$O^u = \{o_1^u, o_2^u, o_3^u, \dots, o_w^u\} \quad (9)$$

随后将用户特征和项目特征进行拼接，构建用户-项目特征，得到特征  $O$  如下：

$$O = O^u \oplus O^i \quad (10)$$

其中  $O^i$  表示项目  $i$  经网络提取得到的特征。

### 3.5 引入FHM

受 Hu 等人<sup>[29]</sup> 的启发，提出关注特征层通道之间相关性的 FHM。FHM 通过建模特征通道之间的关系，抑制相似通道，自适应的增加特征通道中独立性较强的通道，提高网络性能，FHM 如图 3 所示。

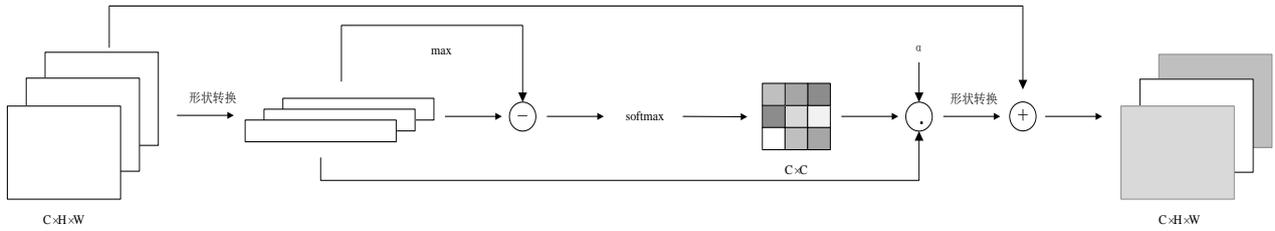


图3 FHM示意图

将词向量经过CNN之后产生大小为 $C \times H \times W$ 的特征图 $O$ 作为FHM的输入,随后进行形状转换,转换为 $C \times 1 \times (H \times W)$ 的特征图 $G$ 后,将 $G$ 中的通道与通道之间两两进行点乘操作,计算通道之间的相似性矩阵。FHM的目的是削弱特征通道之间相似程度较高的通道,突显相似程度较低的通道作为重要的特征通道。因此将相似性矩阵中最大的特征值减去其中的每一个元素,随后计算权重因子 $atte$ :

$$atte = \text{softmax}(\max(GG^T) - GG^T) \quad (11)$$

将 $atte$ 与原始的特征进行内积操作得到的新的特征,随后将新的特征与可学习参数 $\alpha$ 相乘,再与原始输入特征相加得出最终输出。

$$h = \alpha \cdot (atte \bullet O) + O \quad (12)$$

其中 $\bullet$ 表示向量的内积。

#### 4 跨领域推荐模型 AMTR

AMTR 模型目的是在对评论文本合理建模提

取有效信息的基础上,将源领域获取的知识用来丰富目标领域的的数据,实现跨领域的知识迁移,从而解决目标领域数据稀疏和用户冷启动问题,提高目标域推荐的准确性。因此本文在基于双注意力机制的特征提取网络基础上,通过联合建模源域、目标域和共享域三个不同的领域,利用特征提取网络分别学习源域和目标域中领域特有的特征以及共享特征,最后引入因子分解机 FM (Factorization Machine) 从而构建出跨领域推荐模型 AMTR。

如图4所示,AMTR模型由词向量、特征提取网络和决策网络组成,其中特征提取网络使用CNN同时提取评论文本中用户的源域特征、目标域特征和共享特征,以及项目的源域特征、目标域特征和共享特征,在CNN前后加入WAM和FHM。决策网络进行不同领域知识融合以及使用因子分解机进行评分预测。

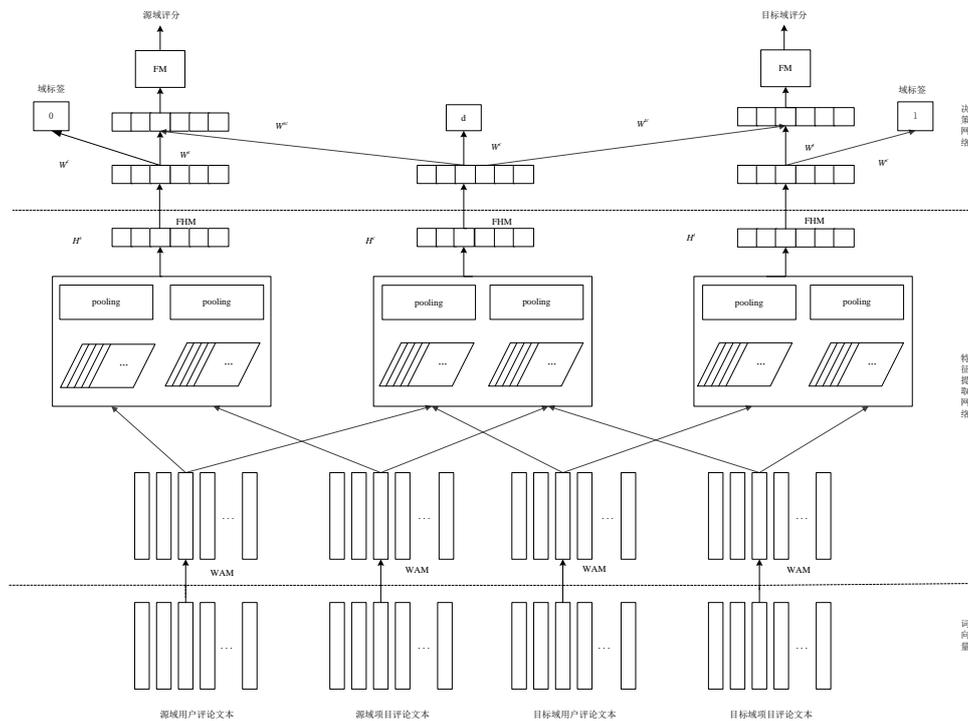


图4 AMTR模型

#### 4.1 引入迁移学习后的特征提取网络

本文模型通过建模不同的领域实现知识的迁移，因此特征提取网络分为三个部分：源域，目标域以及共享域。三个域的网络模块皆使用两个并行的 CNN 分别进行用户特征和项目特征的提取，其使用相同的网络结构，网络之间参数不共享。通过源域和目标域学习领域特有的特征，通过共享域学习领域间的共享特征。其中源域输入源域评论，目标域输入目标域评论，共享域输入源域和目标域的评论。源域、目标域和共享域的特征提取的具体过程如算法 2 所示。

**算法 2.** 源域、目标域以及共享域的特征提取  
输入：用户聚合评论文本文档  $d_{i,m}^u$ 、项目聚合评论文本文档  $d_{i,p}^i$ ，卷积核的窗口大小  $q$

输出：源域、目标域特征  $O^l$ ，共享域特征  $O^c$ ， $l \in \{s, t\}$

1. 将  $d_{i,m}^u, d_{i,p}^i$  送进 WAM 中，生成  $WAM_{i,m}^u, WAM_{i,p}^i$
2. 对  $WAM_{i,m}^u, WAM_{i,p}^i$  进行操作(7)，生成  $(Z_j^u)^l, (Z_j^i)^l$
3. 对  $(Z_j^u)^l, (Z_j^i)^l$  执行操作(8)
4. FOREACH  $j \in q$  DO
5. 进行操作(9)生成  $(O^u)^l \oplus = (O_j^u)^l, (O^i)^l \oplus = (O_j^i)^l$
6. END FOR
7. 进行操作(10)生成  $O^l = (O^u)^l + (O^i)^l$
8. 对  $WAM_{i,m}^u, WAM_{i,p}^i$  进行操作(7)，生成  $(Z_j^u)^c, (Z_j^i)^c$
9. 对  $(Z_j^u)^c, (Z_j^i)^c$  进行操作(8)
10. FOREACH  $j \in q$  DO
11. 进行操作(9)生成  $(O^u)^c \oplus = (O_j^u)^c, (O^i)^c \oplus = (O_j^i)^c$
12. END FOR
13. 进行操作(10)生成  $O^c = (O^u)^c + (O^i)^c$
14. RETURN  $O^l, O^c$

#### 4.2 知识融合以及评分预测的网络

通过特征提取网络得到三个领域的特征，随后进行三个领域知识的融合，将来自源域的特征和共享域的特征结合，来自目标域的特征和共享域的特征结合，实现知识的迁移。特征融合过程为：三个域的特征分别通过全连接层后进行拼接，得到最终的特征输出  $\hat{y}^k$ ：

$$\hat{y}^k = \begin{cases} f(w^{sc}h^c + w^s h^s + b^s), & \text{if } k=0 \\ f(w^{tc}h^c + w^t h^t + b^t), & \text{if } k=1 \end{cases} \quad (13)$$

其中  $k (k \in \{0,1\})$  是域标签，0 代表源域，1 代表目标域。 $w^{sc}$ 、 $w^{tc}$ 、 $w^s$ 、 $w^t$  和  $w^c$  分别为共享域和源域、共享域和目标域、源域、目标域和共享域的权重，而  $b^s$  和  $b^t$  分别为源域和目标域的偏置量， $f$  代

表  $ReLU$  激活函数。

随后引入 FM<sup>[30]</sup> 得到用户对项目的预测评分：

$$\hat{r} = \hat{w}_0 + \sum_{i=1}^{|\hat{y}^k|} \hat{w}_i \hat{y}_i^k + \sum_{i=1}^{|\hat{y}^k|} \sum_{j=i+1}^{|\hat{y}^k|} \langle \hat{v}_i, \hat{v}_j \rangle \hat{y}_i^k \hat{y}_j^k \quad (14)$$

其中  $\hat{w}_0$  是全局偏置量， $\hat{w}_i$  是  $\hat{y}^k$  中第  $i$  个分量的权重值。 $\langle \hat{v}_i, \hat{v}_j \rangle = \sum_{f=1}^{|\hat{v}|} \hat{v}_{i,f} \hat{v}_{j,f}$  为各分量之间的二阶交互。

#### 4.3 AMTR模型的损失函数

考虑到领域特有的特征可能进入共享特征空间，导致数据冗余，为保证它们之间的独立性，在共享特征后引入域鉴别器<sup>[9]</sup>。使用域鉴别器  $p(d|h^c)$ ，预测共享特征  $h^c$  上的领域标签  $d$ ，区分特征来自源域还是目标域，定义如下：

$$p(d|h^c) = \text{softmax}(w^c h^c + b^c) \quad (15)$$

其中  $w^c$  表示共享域权重， $b^c$  表示共享域偏置量。

为了使共享域难以区分特征是来自于源域还是目标域的特征空间，减少不同领域的概率分布的差异，防止领域特有的特征进入共享空间，本文在共享特征空间中加入对抗性损失<sup>[9]</sup>  $L_{adv}$  消除噪声特征，记为：

$$L_{adv} = \frac{1}{n} \sum_{i=1}^n \sum_{k=0}^1 p(d=k|h_i^c) \log p(d=k|h_i^c) \quad (16)$$

随后分别在源域和目标域的特征空间加入域鉴别损失函数，使得领域特有的特征空间可以更好的区分不同的领域，定义源域  $L_s$  和目标域  $L_t$  的负熵损失函数为：

$$L\{s,t\} = -\frac{1}{n_{\{s,t\}}} \sum_{i=1}^{n_{\{s,t\}}} \sum_{k \in \{s,t\}} I^{(d_i=k)} \log p(d=k|h_i^{\{s,t\}}) \quad (17)$$

为了学习领域的不变的特征，在已学到的共享特征  $h^c$  和领域特有的特征  $h^s$ 、 $h^t$  上增加正交约束<sup>[9]</sup>  $L_{orth}$  记为：

$$L_{orth} = \sum_{k \in \{s,t\}} H^c T H^k \quad (18)$$

AMTR 模型可以选择最小二乘误差( $L_2$ )和最小绝对偏置( $L_1$ ) 损失函数训练模型，实验显示  $L_2$  比  $L_1$  的表现稍好，因此本文利用  $L_2$  损失函数进行训练，实验结果详细见 5.5.6 节。模型的损失函数最终定义为：

$$\text{loss} = \sum_{k \in \{s,t\}} -\frac{1}{n_k} \sum_{j=1}^{n_k} \frac{1}{2} (r_{u,i}^k - \hat{r}_{u,i}^k)^2 + \frac{\lambda_1}{2} L_{adv} + \frac{\lambda_2}{2} L_s + \frac{\lambda_3}{2} L_t + \frac{\lambda_4}{2} L_{orth} + \frac{\lambda_5}{2} \|\Theta\|_F^2 \quad (19)$$

其中  $\lambda_1$ 、 $\lambda_2$ 、 $\lambda_3$ 、 $\lambda_4$  和  $\lambda_5$  为不同损失函数的正则

化参数,  $\Theta$  表示模型参数。  $r_{u,i}^k$  和  $\hat{r}_{u,i}^k$  为用户  $u$  对项目  $i$  的真实评分和预测评分。

#### 4.4 AMTR模型的联合训练

AMTR 模型训练的具体过程如算法 3 所示。采用联合训练的方式将源域和目标域中的数据依次送入网络中, 用域鉴别器判定特征的域标签, 将不同域传递出的损失值相加进行反向传播从而进行参数的更新。

##### 算法 3. AMTR 模型的训练

输入: 源域的训练数据  $D^s$  ( { 数据  $d^s$ , 标签  $l^s$  } ), 目标域的训练数据  $D^t$  ( { 数据  $d^t$ , 标签  $l^t$  } ) (  $|D^s| > |D^t|$  ), 模型  $f$ , 损失函数  $L$ , 学习率  $\varepsilon$ , 衰减率  $\rho_1, \rho_2$ , 稳定系数  $\delta$

输出: 决策层参数  $w^c, w^b, w^s, b^s, w^t, b^t, w^{sc}, b^{sc}, w^{tc}, b^{tc}$ , 特征层的网络参数  $\Theta_s, \Theta_t, \Theta_c$

1. 正交初始化参数  $\Theta = \{ w^c, w^b, w^s, b^s, w^t, b^t, w^{sc}, b^{sc}, w^{tc}, b^{tc}, \Theta_s, \Theta_t, \Theta_c \}$

2. 迭代次数  $epoch = 0$

3. 一阶矩和二阶矩变量  $s = 0, r = 0$ ;

4. 时间步  $step = 0$

5. WHILE  $epoch \leq Maxepoch$  DO

6. FOREACH  $(d^s, l^s) \in D^s$  DO

7. 从  $D^t$  中随机选取一个目标域训练数据  $d^t, l^t$

8. 计算梯度  $g \leftarrow \frac{1}{m} \nabla_{\Theta} L(f(d^s, d^t, \Theta), l^s, l^t)$

9. 更新参数  $\Theta \leftarrow \Theta - \varepsilon \frac{\rho_1 s + (1 - \rho_1) g}{1 - \rho_1^{step}},$

$$\sqrt{\frac{\rho_1 r + (1 - \rho_1) g^2}{1 - \rho_1^{step}} + \delta}$$

$s \leftarrow \rho_1 s + (1 - \rho_1) g, r \leftarrow \rho_1 r + (1 - \rho_1) g^2$

10.  $step += 1$

11. END FOR

12.  $epoch += 1$

13. END WHILE

14. RETURN  $\Theta$

#### 4.5 模型时间复杂度分析

本文提出的 AMTR 模型中, 主要的时间成本包含以下两个部分: 特征提取网络和决策网络。

其中特征提取网络包含 WAM、CNN 和 FHM。对于 WAM 来说, 设置句子的长度为  $n$ , 词向量的维度为  $d$ 。进行卷积操作的时间复杂度为  $O(nd)$ , 通过两个全连接层的时间复杂度为  $O(2 * n^2 / r)$ ,  $r$  为特征衰减系数。

通过 CNN 的时间复杂度为  $O(\sum_i (q_i)^2 * d * n * Cout)$ , 其中  $Cout_i$  为输出通道数。通过 FHM 的时间复杂度为  $O(\sum_i Cout_i)$ 。

对于决策网络来说, 特征合并层的时间复杂度为  $O(\sum_i (Cout_i) * n')$ ,  $n'$  表示决策参数, 且  $n' < n$ 。FM 的时间复杂度为  $O(k(n')^2)$ , 其中  $k$  为 FM 的系数。因此 AMTR 模型总的时间复杂度为:

$$O = (nd + 2 * n^2 / r + \sum_i (q_i)^2 * d * n * Cout + \sum_i Cout_i + \sum_i (Cout_i) * n' + k(n')^2)$$

## 5 实验与分析

### 5.1 数据集

为保证实验结果的可靠性, 本文选用 Amazon 数据集<sup>①</sup>进行实验评估。该数据集主要包含用户对网站商品的评价信息及商品元数据, 总共有 142,800,000 条评论, 评分范围为 1 到 5 的整数, 时间跨度为 1996 年 5 月至 2014 年 7 月。

本文在 Amazon 数据集中选择“电影”、“图书”和“CD”作为实验数据, 将其分为“电影-图书”和“电影-CD”两个跨领域组合进行实验分析。首先对数据进行稠密度的预处理, 保留三组数据中具有 10 条以上评论的用户, “图书”和“电影”中超过 120 条评论的项目, “CD”中超过 30 条评论的项目, 随后从“电影-图书”和“电影-CD”跨领域组合中选择有交互的用户。“电影”的稠密度相比于“图书”和“CD”的稠密度更大, 因此在跨领域组合中将“电影”作为源域, “图书”和“CD”分别作为目标域。其次对评论文本进行分词、删除停用词、用 Nltk<sup>②</sup>进行词形的还原等处理, 数据集统计信息如表 1 所示。

表 1 数据集统计信息

域	电影-图书		电影-CD	
	电影	图书	电影	CD
#用户	2,000		2,000	
#项目	5,846	5,638	5,873	6,975

① <http://jmcauley.ucsd.edu/data/amazon/>

② <https://www.nltk.org/>

#评论	215,862	121,135	214,038	140,422
稠密度	1.84%	1.07%	1.82%	1.006%

## 5.2 评价指标

平均绝对误差  $MAE$ (Mean Absolute Error)、均方根误差  $RMSE$ (Root Mean Square Error) 和均方误差  $MSE$ (Mean Square Error) 是评估推荐性能的三个指标，它们通过计算预测评分和真实评分之间的误差来衡量推荐结果的准确性  $MAE$ 、 $RMSE$  和  $MSE$  定义分别为：

$$MAE = \frac{1}{N} \sum_{i=1}^N |r_{u,i} - \hat{r}_{u,i}| \quad (20)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_{u,i} - \hat{r}_{u,i})^2} \quad (21)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (r_{u,i} - \hat{r}_{u,i})^2 \quad (22)$$

其中  $N$  表示测试数据的数量， $r_{u,i}$  表示测试数据真实的评分， $\hat{r}_{u,i}$  表示通过 AMTR 模型后得到的  $r_{u,i}$  的预测值。 $MAE$ 、 $RMSE$  和  $MSE$  三个指标越小代表推荐准确性越高。 $MAE$ 、 $RMSE$  和  $MSE$  的评价结论一致，因为  $MSE$  是  $MAE$  的平方， $RMSE$  是  $MSE$  开根号后的值。为了与对比模型在同一评估方法下进行实验对比，本文将选择对比模型采用的评价指标来评估评分预测的精度。

## 5.3 对比模型

为了评估 AMTR 模型的性能，将与以下模型进行对比：

(1)CBT<sup>[11]</sup>：一种跨领域推荐的经典模型。将用户和项目的特征因子进行聚类得到评分模式矩阵，同时假设领域之间的评分模式相似，在领域间进行评分模式共享。

(2)CLFM<sup>[12]</sup>：一种聚类层次的潜在因子模型。不仅学习领域之间共享的评分模式，还学习每个领域特有的评分模式，仅利用共享评分模式进行知识迁移。

(3)CDTF<sup>[15]</sup>：一种典型的基于张量分解的跨领域推荐模型。利用用户-项目-领域三元关系，通过张量的三因式分解捕捉特征，从而进行知识迁移。

(4)WITF<sup>[16]</sup>：一种基于张量因式分解模型。它利用从多领域学习到的显式偏好数据和隐式偏好数据作为先验知识从而提高目标域的推荐性能。

(5)CCCFNet<sup>[17]</sup>：一种多视图的神经网络学习框

架。将协同过滤和基于内容的过滤结合进行跨领域推荐。

(6)MF<sup>[31]</sup>：一种传统的基于协同过滤思想的方法。它将评分矩阵进行分解，得到用户矩阵和项目矩阵，从而预测缺失评分。

(7)DeepCoNN<sup>[21]</sup>：一种利用神经网络进行评论文本建模的代表性推荐模型。通过并行 CNN 联合学习评论文本中用户和项目的特征从而进行评分预测。

(8)D-ATT<sup>[23]</sup>：一种基于注意力机制的推荐模型。通过在 CNN 编码层之前引入注意力机制过滤评论文本中重要单词得到用户和项目特征进行评分预测。

## 5.4 实验设置

本文模型采用两种不同的方式进行训练与测试。方式 1 将经过预处理的实验数据集中源域数据全部用于训练，目标域数据随机分为训练集和测试集，其中 80% 的数据用于训练，20% 用于测试。方式 2 将源域数据全部用于训练，目标域数据随机分为 50% 的训练集和 50% 的测试集。

在实验环节，实验配置为 Ubuntu16.04、CPU i7-8700HQ、16G 内存、NVIDIA GTX 1050Ti 4G，实验所需环境为 Python2.7.13、Pytorch0.4.0、Nltk、Scipy、Gensim=3.2.0、Numpy=1.15。本文使用 Glove 将评论文本向量化，词向量维度  $L$  为 100，卷积核数量设置为 150，采用多个不同的卷积核进行卷积操作，窗口大小分别为 3、4、5。为了缓解过拟合现象，将 0.5 的 dropout<sup>[32]</sup> 应用于决策网络的全连接层。损失函数的参数设置为  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0.04$ ， $\lambda_5 = 0.00064$ 。本文模型中 FM 的权重因子维度设置为 5，输入 FM 的用户特征维度和项目特征维度设置为 50。采用自适应矩估计 Adam<sup>[33]</sup> (Adaptive Moment Estimation) 训练模型进行优化，它是一个基于随机梯度的优化器，具有自适应估计，其学习率设置为 0.001，其余参数设置和 Pytorch<sup>®</sup> 相同。对比模型的参数均为最优值，CBT 模型的用户和项目的聚类的数量均设置为 50；CLFM 模型的用户和项目的聚类的数量均设置为 50，共享公共子空间的维度设置为 40；CDTF 模型的潜在因子向量的维度设置为 25， $\lambda_u = 0.05$ ， $\lambda_v = 0.001$ ， $\lambda_c = 0.001$ ；WITF 模型的潜在因子向量的维度设置为 25， $\lambda_v = 1$ ，

$\lambda_v=1$ ,  $\lambda_c=1$ ; CCCFNet 模型的参数按论文最优值取值; MF 算法的潜在因子向量的维度设置为 20, 正则化参数设置为 0.001; DeepCoNN 模型的卷积核数目设置为 150, 权重衰减设置为 0.0064; D-ATT 模型按照原论文的最优参数取值。

本文对 AMTR 的重要参数进行了研究, 实验表明不同的卷积核数目对模型的性能有不同的影响。在实验中, 设置卷积核数目为 50、100、150、200、250。在“电影-CD”数据集得到的实验结果如图 5 所示。

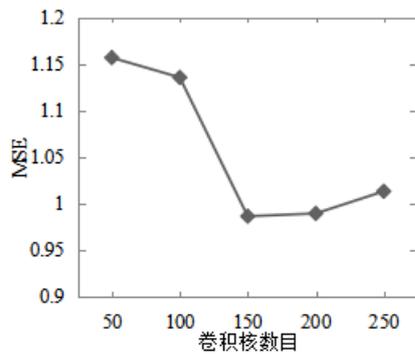


图 5 卷积通道数目对 MSE 值的影响

从图中可以看出, 当卷积核数目为 150 时, 模型在数据集上 MSE 取最小值, 而随着卷积核数目的继续增加, 模型的 MSE 持续变大, 模型的性能随着卷积核数目的增加反而变差, 因此本文设置模型的卷积核数目为 150。而 DeepCoNN 模型与本文一样采用深度学习的方法, 利用 CNN 进行文本特征的提取从而进行用户和项目的建模。在 DeepCoNN 模型测试卷积核数目对模型性能影响的时候, MSE 同样呈先下降后上升的趋势。因此可以看出卷积核数目对模型性能有较大的影响。

## 5.5 实验结果与分析

本文实验的目的是验证融合评论文本信息进行跨领域推荐的 AMTR 模型优于仅利用评分信息的跨领域推荐模型以及仅在单领域进行推荐的模型。因此通过验证如下内容, 评估本文模型的有效性: (1)验证融合文本信息带来的推荐性能提升; (2)验证迁移学习在不同数据集上带来的推荐性能提升; (3)验证注意力机制带来的推荐性能提升; (4)验证本文模型解决数据稀疏和用户冷启动问题的能力; (5)验证本文模型的普适性。此外, 还对调整用户评分尺度差异问题进行了相应实验验证。

### 5.5.1 推荐准确度比较

为了验证模型的推荐准确度, 实验评估了本文

提出的 AMTR 模型与现有的跨领域推荐模型 CBT、CLFM、CDTF、WITF 以及 CCCFNet 在“电影-图书”和“电影-CD”两种数据集下的结果, 并将结果进行比较。对于 AMTR 模型, 使用方式 1 进行训练与测试。由于上述对比模型大多采用 MAE 和 RMSE 作为实验评估指标, 为与其评估方法一致, 采用上述指标。不同模型在不同数据集下的 MAE 和 RMSE 如表 2 所示。

表 2 不同数据集下 MAE 和 RMSE 的比较

模型	电影-图书		电影-CD	
	MAE	RMSE	MAE	RMSE
CBT	0.8546	1.0286	0.8777	1.0631
CLFM	0.9035	1.0787	0.9101	1.1115
CDTF	1.1543	1.2535	1.1522	1.2238
WITF	1.0800	1.2790	1.0341	1.2003
CCCFNet	0.9462	1.1991	0.8713	1.0958
AMTR	0.7936	0.9920	0.7798	0.9930

从表 2 可以看出, 不管是基于 MAE 指标, 还是 RMSE 指标, 本文提出的 AMTR 模型在不同数据集的实验中均优于对比模型, 并且取得了最优的结果。在“电影-图书”数据集上, AMTR 模型的 MAE 为 0.7936, 相对对比模型中表现最好的 CBT 模型提高 6.1%; RMSE 为 0.9920, 提高 3.66%。另外在“电影-CD”数据集上, 其 MAE 为 0.7798, 相对对比模型中 MAE 最好的 CCCFNet 模型提高了 9.15%; RMSE 为 0.9930, 相对对比模型中 RMSE 最好的 CBT 模型提高 7.01%。该结果表明利用评论文本的 AMTR 模型相对于利用评分的模型能更有效提高评分预测的准确度。

### 5.5.2 知识迁移的有效性

为了验证 AMTR 模型知识迁移的有效性, 表明模型可以通过迁移学习提高目标域的推荐精度, 本文将 AMTR 与单领域的推荐模型 MF、DeepCoNN 以及 D-ATT 进行了比较。对于 AMTR 模型采用方式 1 进行训练和测试, 其中三个单领域模型只在“图书”和“CD”数据集上进行测试。由于对比模型大多采用 MSE 作为评价指标, 因此本实验采用 MSE 作为评价指标, 结果如图 6 和图 7 所示。

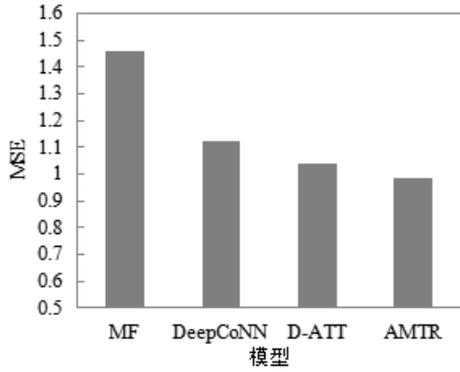


图6 “电影-图书”数据集中的 MSE 比较

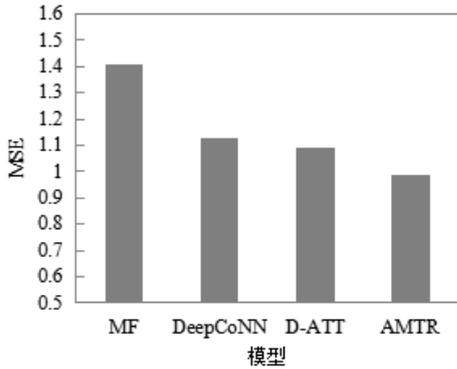


图7 “电影-CD”数据集中的 MSE 比较

从图6和图7的结果可以看出,在两种数据集上,本文提出的AMTR模型均优于对比模型。在“电影-图书”数据集上,AMTR模型的MSE为0.9841,相比于MF、DeepCoNN和D-ATT模型分别提升了47.72%、14.1%、5.47%。在“电影-CD”数据集上,AMTR模型的MSE为0.9861,相比于MF、DeepCoNN和D-ATT模型分别提升了42.49%、14.18%、10.35%。结果表明,相比于只利用单领域数据来说,利用辅助领域数据可以帮助目标域提高推荐性能。表明了AMTR模型具有较好的知识迁移性能,可以提高目标域的推荐精度。

与此同时,利用评论文本的模型DeepCoNN和D-ATT以及AMTR的MSE相较于利用评分的MF在“电影-图书”数据集分别提升了33.62%、42.25%、47.72%,在“电影-CD”数据集分别提升了28.31%、32.14%、42.49%,由此可见利用评论文本相比于利用评分可以更好地提高推荐的性能。

为了进一步验证AMTR模型能够有效地进行知识迁移。本文设计了一组新的对比实验进行验证,本文模型采用方式2进行训练与测试,单领域推荐模型的数据集随机分为50%作为训练集,剩余50%作为测试集。实验结果如图8和图9所示。

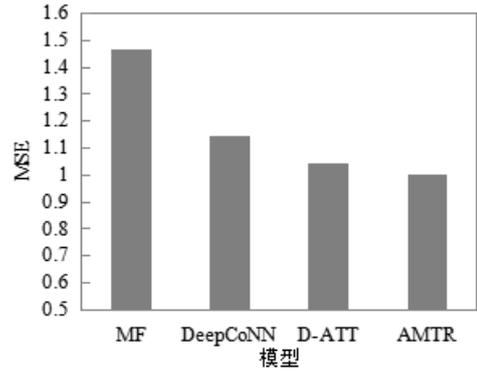


图8 “电影-图书”数据集中的 MSE 比较

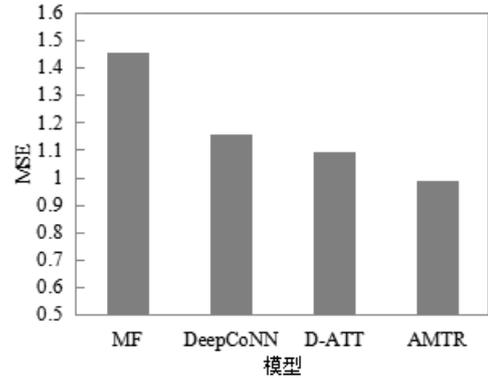


图9 “电影-CD”数据集中的 MSE 比较

通过图8和图9可以看出,本文的AMTR模型在不同的数据集下仍取得了最优性能。在“电影-图书”数据集上,AMTR模型的MSE为1.0003,相比于MF、DeepCoNN和D-ATT模型分别提升了46.41%、14.2%、4.55%。在“电影-CD”数据集上,AMTR模型的MSE为0.9884,相比于MF、DeepCoNN和D-ATT模型分别提升了47.06%、16.82%、10.81%。结果表明本文模型可以充分利用源域数据提高目标域的推荐精度的问题,具有较好的知识迁移性能。

### 5.5.3 WAM和FHM的有效性

本文加入词级别和特征级别的注意力机制,用于捕获更多的重要信息,使得模型有更好的特征提取能力。为了研究WAM和FHM对模型性能的影响,本文在“电影-图书”和“电影-CD”跨领域数据集上设置了一组对比实验。分别在不加入WAM和FHM的AMTR-W&F模型,只加入FHM的AMTR-W模型,只加入WAM的AMTR-F模型以及AMTR模型上进行对比,结果如表3所示。

表3 不同注意力机制下的 MSE 比较

数据集	AMTR-W&F	AMTR-W	AMTR-F	AMTR
电影-图书	1.1423	1.0546	1.0945	<b>0.9841</b>

电影-CD	1.1089	1.0730	1.0925	<b>0.9861</b>
-------	--------	--------	--------	---------------

从表3可以看出加入注意力机制的模型的性能整体优于不加入注意力机制的模型。在“电影-图书”数据集下,加入WAM和FHM的本文模型AMTR的MSE为0.9841,相较于AMTR-W&F、AMTR-W、AMTR-F分别提升了15.82%、7.05%、11.04%。在“电影-CD”数据集下,本文模型的MSE为0.9861,相较于AMTR-W&F、AMTR-W、AMTR-F分别提升了12.28%、8.69%、10.64%。由此可以表明WAM和FHM可以有效的捕获出重要信息,使模型有更好的特征提取能力,提高了模型的性能。

为了更直观的说明WAM可以通过词的输出权重表示不同词的重要性,因此以两个可视化的案例来说明本文提出的WAM的有效性,随机从“图书”和“电影”的测试集中选择一条用户对某个项目的评论语句作为测试对象,该评论不做任何除词干化之外的操作,保留该句的每一个词的权重,将权重值大于平均值的词标为斜体加下划线,对远大于平均值的词加粗显示,以此为标准,绘制如下图10所示的注意力分布情况。

“图书”领域下用户的评论
This is a good <b>book</b> for <i>children</i> of all ages, <i>describing</i> the <i>stories</i> behind some of the most <i>popular ballets</i> . It is <i>written</i> in <b>easy-to-understand language</b> and has <i>beautiful illustrations</i> .
“电影”领域下用户的评论
I really <i>enjoyed</i> this <i>movie</i> ! It is <i>fun</i> , and <b>action packed</b> . The <i>characters are fresh</i> , and the <i>cinematography feels good</i> . It was <i>totally</i> made for <b>Blue - ray</b> .

图10 注意力分布可视化结果

“图书”领域下的评论总体是用户对书籍内容本身,以及阅读感受的分析评价。“This is a good book”表明用户对“书”的积极情感;“book for *children* of all ages”表达了物品“书”本身的特征,是关于“儿童”的书籍;“*describing* the *stories* behind some of the most *popular ballets*”表明书中“描述”的“故事”是关于“芭蕾舞的”;“*written* in *easy-to-understand language* and has *beautiful illustrations*”表明书用“易于理解的语言”进行“编

写”以及具有“精美的插图”。从这句话可以得出“书”的特征有“儿童”“描述”“芭蕾舞的故事”“编写”“易于理解的语言”和“精美的插图”,同时也可以得知用户对书的“描述”“编写”“语言”和“插图”等方面特征的关注。用户虽然只有一个“good”描述了对这本书的感受,但是从“易于理解的语言”和“精美的插图”等方面可以看出用户对这本书的评价呈优。

“电影”领域的评论中,可以看出用户对电影整体的偏好,认为它有趣,动作丰富,人物鲜活,摄影技术很好,是蓝光的。体现出用户对电影关注在“电影给予的感觉”、“动作”、“角色”和“摄像”,这些关注的评价呈优,也体现了电影本身的特征“有趣”、“动作”、“角色”、“摄像”和“蓝光”。WAM很好的关注了enjoy、movie、fun、action packed、characters are fresh、cinematography feels good、Blue - ray(喜欢、电影、有趣、动作丰富、人物鲜活、摄影技术很好、蓝光)。且“action”和“Blue - ray”具有更好的权重,说明它们占据的关注度更高。

从上可以看出,WAM很好的关注了文本中的重要信息,并为这些重要的信息赋予了高权重,因此,神经网络可以通过WAM加强对文本内容重点部分的关注,使得模型有更好的特征提取能力,同时WAM提取重要信息的有效性得到证明,使得推荐更具有解释性。

#### 5.5.4 缓解数据稀疏问题

为验证AMTR模型可以缓解目标域中数据稀疏问题,本节设置两组对比实验,分别与跨领域推荐模型和单领域推荐模型进行对比,通过控制测试集中用户的评论数进行实验。本文模型采用方式1进行训练与测试。针对测试集中每个用户的所有评论,分别从中随机挑选 $\alpha$ 倍的评论作为测试集, $\alpha$ 为10%、30%、50%、70%以及100%。

##### 实验1. 与跨领域推荐模型的比较

本实验在不同的 $\alpha$ 下分别对跨领域模型进行测试,结果如图11和图12所示。

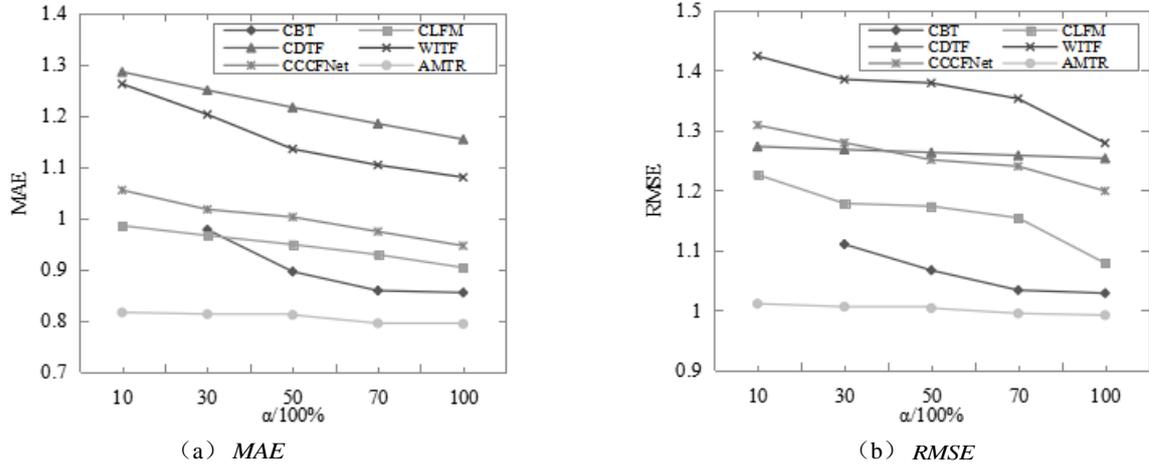


图 11 “电影-图书”数据集上，不同评价标准的实验结果比较

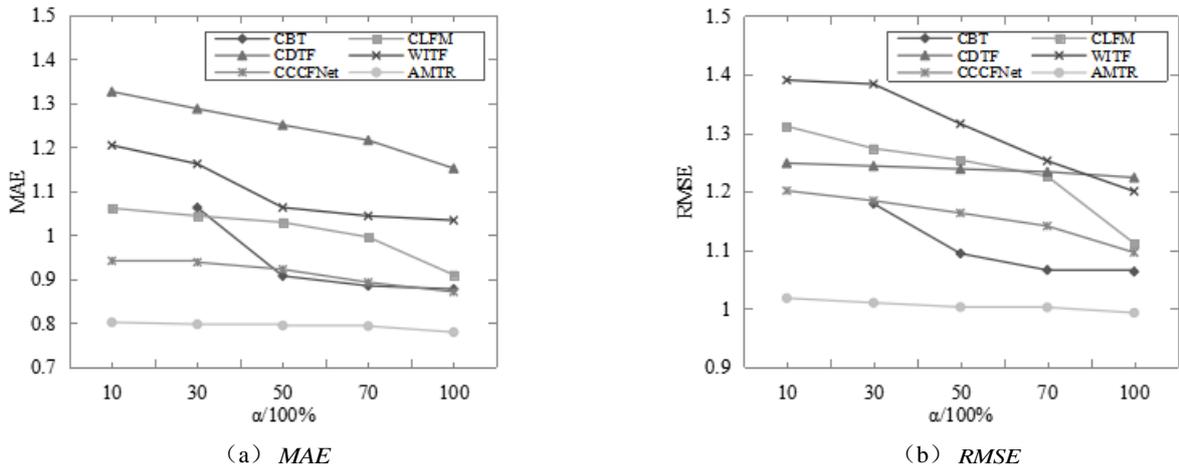


图 12 “电影-CD”数据集上，不同评价标准的实验结果比较

$\alpha$  越小，表明领域中的数据越稀疏，数据稀疏问题越明显。由图 11 和图 12 可以看出无论  $\alpha$  取何值，AMTR 模型在“电影-图书”和“电影-CD”两个数据集上 MAE、RMSE 普遍优于对比模型，取得了较好的效果。其中，在  $\alpha$  为 10% 时，CBT 模型由于数据太少无法得出实验结果。

图 11 表示“电影-图书”数据集下的实验结果。首先从横向比较可以看出，随着  $\alpha$  的增加，用户的评论数量增加，各个模型的推荐效果均逐步在提升。例如，当  $\alpha$  为 100% 时，AMTR、CBT、CLFM、CDTF、WITF、CCCFNet 模型相比于  $\alpha$  为 10% 时，MAE 分别提升了 2.24%、-、8.19%、13.17%、18.24%、10.87%；RMSE 分别提升了 1.91%、-、14.76%、1.97%、14.51%、10.98%；相比于  $\alpha$  为 50% 时分别提升了 1.75%、4.11%、4.52%、6.23%、5.52%、5.61%；RMSE 分别提升了 1.17%、3.82%、9.5%、0.98%、10.2%、5.2%。由此可见，本文的 AMTR 模型相较于对比模型取得了更为稳定的效果，说明 AMTR 模

型受数据稀疏问题的影响较小。同时表明了 AMTR 模型相对于对比模型可以挖掘更多的有利于评分预测的用户和项目信息，即使在数据较为稀疏的情况下，模型通过评论文本进行知识迁移后的信息可以帮助模型取得较为稳定的推荐效果，更好地缓解数据稀疏问题。

从纵向比较可看出，本文模型在各种不同  $\alpha$  下相较于对比模型呈现的效果为最优。例如，当  $\alpha$  为 10% 时，AMTR 模型的 MAE 为 0.8160，与 CLFM、CDTF、WITF、CCCFNet 模型相比分别提高了 16.94%、47%、44.64%、23.89%；RMSE 为 1.0111，分别提高了 21.52%、26.21%、41.3%、29.78%。 $\alpha$  为 50% 时，AMTR 模型的 MAE 为 0.8111，相比于 CBT、CLFM、CDTF、WITF、CCCFNet 模型分别提高了 8.46%、13.76%、40.55%、32.41%、19.12%；RMSE 为 1.0037，分别提高了 6.31%、17%、25.96%、37.55%、24.74%。 $\alpha$  最小时，本文模型效果远高于对比模型，充分说明 AMTR 模型相较于对比模型可

以更好地缓解数据稀疏问题。此外,在 $\alpha$ 为任意值的情况下 AMTR 模型相比于考虑评分的模型效果提升显著,说明评论文本相较于评分蕴含更多的用户和项目信息,能更好的缓解数据稀疏问题。

图 12 表示“电影-CD”数据集下的实验结果。首先从横向比较可以看出,随着 $\alpha$ 的增加,各个模型的推荐效果均不断变好。例如,当 $\alpha$ 为 100%时,AMTR、CBT、CLFM、CDTF、WITF、CCCFNet 模型相比于 $\alpha$ 为 30%时,MAE 分别提升了 1.79%、18.52%、13.4%、13.55%、12.83%、6.74%,RMSE 分别提升了 1.7%、11.61%、16.19%、1.96%、18.31%、8.84%;相比于 $\alpha$ 为 70%时,MAE 分别提升了 1.38%、0.74%、8.59%、6.4%、0.99%、2.17%,RMSE 分别提升了 0.92%、0.27%、11.44%、0.97%、5.2%、4.52%。由此可见,本文的 AMTR 模型效果较为稳定,相比于对比模型能够更好地缓解数据稀疏问题。

从纵向比较可看出,本文模型在各种不同 $\alpha$ 下效果呈最优。例如,当 $\alpha$ 为 30%时,AMTR 模型的 MAE 为 0.7977 与 CBT、CLFM、CDTF、WITF、CCCFNet 模型相比分别提高了 26.52%、24.64%、49%、36.47%、14.1%;RMSE 为 1.0100,分别提高了 16.92%、26.34%、23.34%、37.34%、17.42%。 $\alpha$ 为 70%时,AMTR 模型的 MAE 为 0.7936,与 CBT、CLFM、CDTF、WITF、CCCFNet 模型相比分别提高了 9.15%、20.24%、42.26%、25.04%、9.94%;RMSE 为 1.0022,分别提高了 6.36%、22.37%、23.13%、25.01%、13.88%。

由此可见,本文所提出的 AMTR 模型在一定程度上缓解了数据稀疏问题对推荐结果带来的不利影响,更好的提高了推荐的性能。

## 实验 2. 与单领域推荐模型比较

本实验在不同的 $\alpha$ 下分别对单领域模型的表现情况进行了测评,结果如图 13 和图 14 所示。

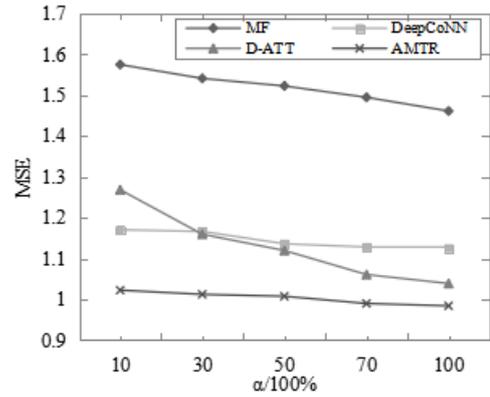


图 13 “电影-图书”数据集中的 MSE 比较

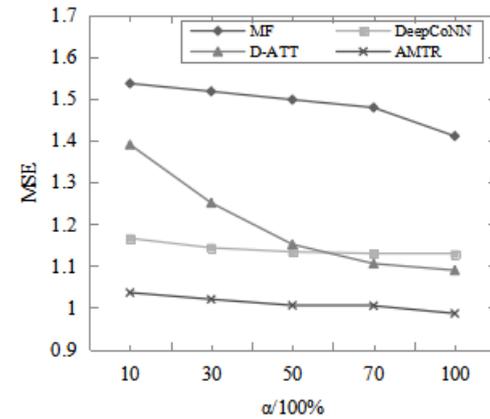


图 14 “电影-CD”数据集中的 MSE 比较

图 13 和图 14 结果表明,即使在不同的数据集下,本文模型相较于对比模型取得了较高的性能。首先从横向比较可以看出,随着 $\alpha$ 的增加,各个模型效果逐步提高。例如,在“电影-图书”数据集下, $\alpha$ 为 100%时,AMTR 模型的 MSE 为 0.9841,与 $\alpha$ 为 10%、30%、50%、70%相比分别提高了 3.83%、2.81%、2.34%、0.57%。在“电影-CD”数据集下, $\alpha$ 为 100%时,AMTR 模型的 MSE 为 0.9861,与 $\alpha$ 为 10%、30%、50%、70%相比分别提高了 4.99%、3.39%、1.94%、1.84%。本文 AMTR 模型效果随着 $\alpha$ 增加效果变化不明显,表明数据稀疏问题对模型性能影响不大,AMTR 模型的稳定性较好。

其次,从纵向可以看出,在各种不同 $\alpha$ 下,AMTR 模型呈现的效果最好。例如,在“电影-图书”数据集下, $\alpha$ 为 10%时,AMTR 模型的 MSE 为 1.0224,与 MF、DeepCoNN、D-ATT 模型相比分别提高了 55.27%、14.79%、24.6%。 $\alpha$ 为 50%时,AMTR 模型的 MSE 为 1.0075,分别提高了 51.52%、12.86%、11.21%。在“电影-CD”数据集下, $\alpha$ 为 30%时,AMTR 模型的 MSE 为 1.0200,与 MF、DeepCoNN、D-ATT 模型相比分别提高了 49.81%、

12.34%、23.11%。 $\alpha$ 为70%时, AMTR 模型的  $MSE$  为 1.0045, 分别提高了 47.49%、12.53%、10.1%。结果表明, 数据越稀疏, AMTR 模型的性能越优于对比模型, 表明了本文 AMTR 模型相较于仅利用单领域信息的推荐模型能够更有效地缓解数据稀疏问题。且利用评论文本的 DeepCoNN、D-ATT、AMTR 的效果相较于利用评分的 MF 的效果更好, 表明了评论文本相对于评分蕴含着更丰富的用户和项目信息, 可以更好地缓解数据稀疏问题。

### 5.5.5 缓解用户冷启动问题

用户冷启动<sup>[34]</sup>是推荐系统存在的一个不可忽

视的问题, 当新用户行为消息较少时, 系统无法掌握其喜好, 推荐性能会下降, 跨领域推荐可以缓解这种问题。

为验证 AMTR 模型可以缓解目标域中用户冷启动问题, 本节设置两组对比实验, 将 AMTR 模型与对比模型分别在“电影-图书”和“电影-CD”数据集上进行对比。从“图书”和“CD”数据集中挑选出 400 个不与训练集用户重叠的新用户, 同时将新用户的评论数目控制为 1、3、5 进行测试。实验结果如图 15 和 16 所示。

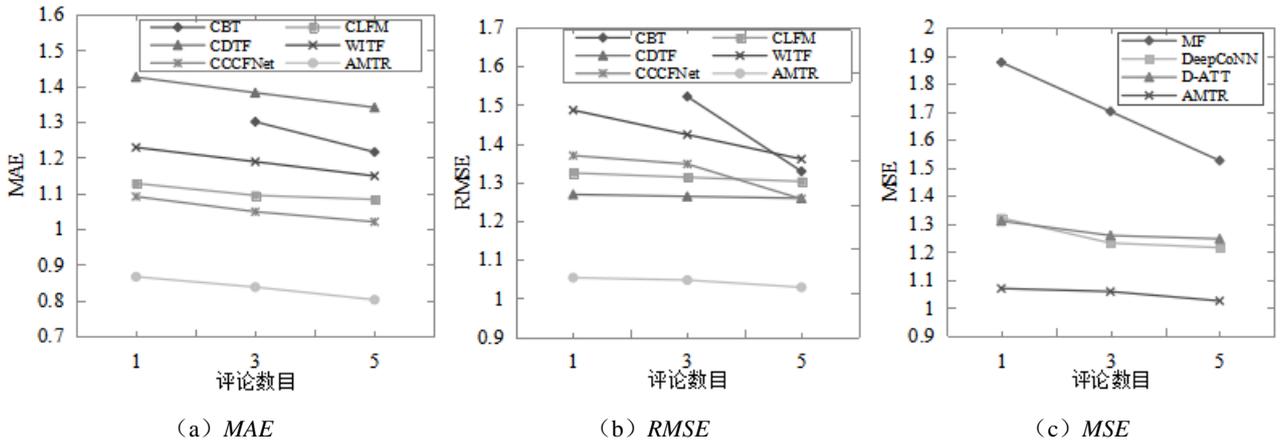


图 15 “电影-图书”数据集上, 各模型在不同评价标准下的实验结果比较

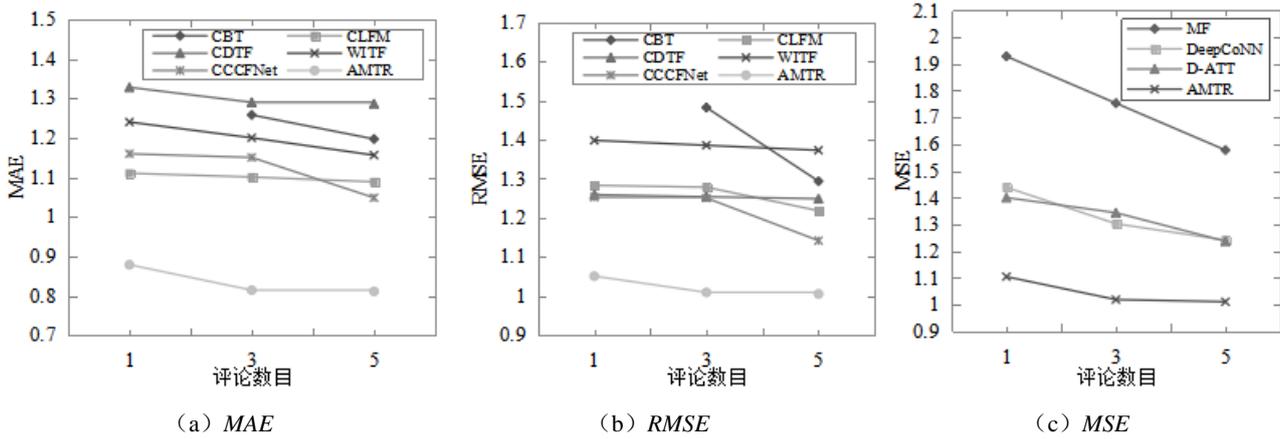


图 16 “电影-CD”数据集上, 各模型在不同评价标准下的实验结果比较

图 15 和 16 显示了在“电影-图书”和“电影-CD”两个数据集上, 不同推荐模型在不同评价指标下受用户冷启动问题影响的结果。通过观察可以发现, 随着评论数目的增多, 各个模型效果逐渐提升。本文模型在各种评价指标下相比于其它模型均取得了较好的效果, 特别是在评论最少, 用户冷启动问题最明显的情况下, 效果更为明显。验证了本文模型缓解用户冷启动问题方面的优势。

### 5.5.6 损失函数的优化研究

为了使模型取得更优的结果, 本文在“电影-图书”以及“电影-CD”两种数据集上进行  $L_1$  和  $L_2$  损失函数的实验对比, 结果如表 4 所示。

表 4  $L_1$  损失函数及  $L_2$  损失函数下的  $MSE$  比较

数据集	$L_1$ 损失函数	$L_2$ 损失函数
电影-图书	1.2473	<b>0.9841</b>
电影-CD	1.0864	<b>0.9861</b>

从表4可以看出模型在不同数据集上,  $L_2$  损失函数取得的效果均优于  $L_1$  损失函数的效果, 表现出良好的性能, 因此本文选择  $L_2$  损失函数进行模型的训练。

### 5.5.7 模型时间效率分析

本文计算了各个模型在不同数据集上的时间效率, 考虑到模型是否存在训练过程、数据集的大小以及不同数据的处理方式都会影响模型的时间。因此本文将一个用户对一个项目的评分预测时间作为模型的消耗时间, 各模型的时间消耗如表5所示。

表5 推荐模型消耗时间的比较 (单位: s)

数据集	电影-图书	电影-CD
CBT	0.0132	0.0074
CLFM	0.0030	0.0025
CDTF	0.0025	0.0026
WITF	0.0102	0.0123
CCCFNet	0.0208	0.0206
MF	0.0033	0.0017
DeepCoNN	1.005	0.561
D-ATT	0.016	0.014
AMTR	1.083	0.679

从表5可以看出, 利用评分进行推荐的跨领域模型 CBT、CLFM、CDTF、WITF、CCCFNet 和单领域模型 MF 速度较快, 其中 MF 速度最快, 因为它采用梯度下降的方法进行近似计算, 缩短了时间。采用 CNN 处理评论文本、利用 FM 进行评分预测的本文模型以及 DeepCoNN 模型速度较慢。D-ATT 虽然也是针对评论文本, 但是它关注于评论中的某个方面, 特征提取较快, 降低了运行时间。

针对时间消耗以及推荐准确性的实验结果, 本文选择了 CBT、MF 以及 DeepCoNN 进行时间复

杂度的分析, 更完善的分析了模型的时间效率。

MF 的时间复杂度为:  $O(T * M * N * K)$ 。其中  $T$  为迭代次数,  $M$  为用户个数,  $N$  为项目个数,  $K$  为因子个数。

CBT 的时间复杂度为:  $O(T + nk + ml + ml + T(pk + ql) + q * n)$ 。其中  $k$  和  $l$  表示用户和项目聚类的数量,  $n$  和  $m$  为辅助评价矩阵的行数和列数,  $p$  和  $q$  为目标评价矩阵的行数和列数,  $T$  为迭代次数。

DeepCoNN 时间复杂度为:

$$O = (\sum_i (q_i)^2 * d * n * Cout + \sum_i (Cout_i) * n' + k(n')^2)$$

从以上可以看出, 无论是跨领域还是单领域, 融合评论文本信息的模型相比于仅利用评分信息的模型, 时间复杂度更为复杂, 时间消耗更高。其次, 深度学习比传统方法的时间复杂度更为复杂, 时间消耗较高。主要可以从以下两点进行解释, 首先评论文本相比于评分来说更为复杂, 处理评论文本所需的时间比评分要长, 过程也更为复杂。其次由于深度学习的内部复杂的推理过程使得其特征提取过程相较于传统的机器学习方法更为复杂, 但是深度学习具有良好的特征提取能力, 可以更好地提取特征。虽然本文模型时间复杂度较为复杂, 时间消耗相对较多, 但是本文模型有效的提高了评分预测的准确度, 提升了推荐性能。

### 5.5.8 模型普适化研究

由于上述实验挑选的“电影”、“图书”和“CD”三个领域数据集的语义关联性较大, 不足以验证本文模型的普适性, 因此本文挑选了关联性较小的数据集来验证本文模型的普适性。实验以“电影”为源域, 选取与“电影”关联性较小的“电子产品”、“衣服”和“玩具”作为目标域。实验参数与“5.4 实验设置”中的实验参数一致。与跨领域和单领域推荐模型的实验对比结果分别如表6和表7所示。

表6 不同数据集下的 MAE 和 RMSE 比较

模型	电影-电子产品		电影-衣服		电影-玩具	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
CBT	1.1088	1.1649	1.0609	1.2437	-	-
CLFM	0.8916	1.1413	0.9001	0.9976	0.9012	0.9906
CDTF	0.9378	1.1083	0.9785	1.0488	0.9187	1.0952
WITF	0.8649	1.0763	0.8880	0.9481	0.8960	0.9570
CCCFNet	0.8870	0.9897	0.8901	1.1403	0.8599	1.1059
AMTR	0.7476	0.9752	0.6887	0.8601	0.8377	0.9778

表7 不同数据集下的 MSE 比较

模型	电影-电子产品	电影-衣服	电影-玩具
MF	1.3107	1.2622	1.5252
DeepCoNN	1.2143	1.1922	1.2671
D-ATT	1.1006	1.0778	1.0923
AMTR	0.9511	0.8816	0.9560

从表中可以看出,无论在何种数据集下,本文模型仍取得最佳效果。实验结果表明,即使在关联性较小的数据集下,本文模型相较于其它模型,评分预测的准确度更高,充分验证了本文模型的普适性。

本文在实验中挑选了语义关联性较小的域进行迁移,但当两个域关联较小时会产生负迁移,而本文模型整体呈优,原因在于本文通过以下几点有效的提取出了有利于迁移的共享特征。(1) 本文利用神经网络提取特征的有效性,进行源域和目标域的共享特征的提取。(2) 在共享特征空间中加入对抗性损失,使共享域难以区分特征来自于源域还是目标域的特征空间,减少不同领域的概率分布的差异,以防止领域特有的特征进入共享空间,更好地寻找网络中的域不变表示,学习更健壮的共享特征。(3) 在共享特征上引入域鉴别器,以防止领域特有的特征进入共享特征空间。(4) 在共享特征和领域特有的特征上增加正交约束,学习领域不变的特征。(5) 在源域、共享域和目标域中采用相同的网络结构,但是参数不共享,使不同网络根据自身数据的特征学习到充足的知识。

通过以上的方式,本文模型可以提取出更纯粹的、更有利于迁移的共享特征,并减少迁移过程中领域特有的特征对共享特征的影响,在有效进行知识迁移的同时缓解负迁移。因此即使在领域关联性较小的情况下,模型结果呈优。

为了更好的验证模型的普适性,另选择了 Yelp<sup>④</sup>数据集和中文数据集进行验证。其中, Yelp 数据集是一个由餐馆评论信息组成的大型数据集,来自 Yelp Dataset Challenge,包含 100 万条以上的评论,评分范围为 1 到 5 的整数。由于 Yelp 和 Amazon 同属于英文文本,因此直接在利用 Amazon 数据集训练好的模型上进行测试,实验参数均采用“5.4 实验设置”中的参数。

④ <https://www.yelp.com/dataset/challenge>

由于中文和英文的差异,中文数据集不能直接用于测试,因此选用“电影”<sup>⑤</sup>和“餐馆”<sup>⑥</sup>两个领域进行了新的跨领域实验,“电影”数据集来源于“豆瓣电影”,其包含 28 部电影,超 70 万用户,超 200 万条评分/评论数据;“餐馆”数据集是 Yongfeng Zhang 教授为 WWW 2013, SIGIR 2013, SIGIR 2014 会议论文而搜集的数据,其包含 24 万家餐馆,54 万用户,440 万条评论/评分数据,评分范围均为 1 到 5 的整数。实验过程中除了采用腾讯词向量<sup>⑦</sup>进行文本的处理,其它部分与英文场景下的一致,词向量维度设置为 100。结果如表 8 所示。

表8 不同数据集下 MSE 的比较

模型	Yelp		中文数据集
	“电影-图书”	“电影-CD”	“电影-餐馆”
DeepCoNN	1.6788	1.8121	1.3377
AMTR	<b>1.0271</b>	<b>1.0662</b>	<b>0.9822</b>

本文模型在“图书-电影”和“图书-CD”上的 MSE 分别为 0.9861 和 0.9841。参考上述 MSE,从表 8 中看出,本文模型无论在哪个数据集下取得的效果均呈优。在 Yelp 数据集上,模型的 MSE 在不同跨领域数据集上分别为 1.0271 和 1.0662,相比于本文模型在“电影-图书”和“电影-CD”的 MSE 稍差,但相较于其它利用英文文本的模型性能更好,效果更稳定。实验结果表明,本文模型具有很好的普适性。

在中文数据集上的实验结果表明,本文模型可以适用于中文场景,有较好的普适性。且“电影”和“餐馆”的关联性较小,更好地验证了本文模型的有效性。无论是中文文本还是英文文本,对于神经网络而言,文本均需要进行词向量化处理,输入的都是一组向量的表示。因此只要模型性能较好,取得的效果也会较好。

#### 5.5.9 用户评分尺度差异问题对推荐性能的影响

用户之间的评分的尺度差异是指在没有统一评分标准的情况下,评分数据会因不同人的主观标准而出现差异,例如一个用户打出的最低分为 1 分,

⑤ [https://github.com/SophonPlus/ChineseNlpCorpus/blob/master/datasets/dmcs\\_v2/intro.ipynb](https://github.com/SophonPlus/ChineseNlpCorpus/blob/master/datasets/dmcs_v2/intro.ipynb)

⑥ [https://github.com/SophonPlus/ChineseNlpCorpus/blob/master/datasets/yf\\_dianping/intro.ipynb](https://github.com/SophonPlus/ChineseNlpCorpus/blob/master/datasets/yf_dianping/intro.ipynb)

⑦ <https://github.com/InsaneLife/ChineseNlpCorpus>

最高分为5分,而另一个用户给出的最低分为2分,最高分为4分,这种差异对评分预测结果有所影响,从而影响推荐效果。

本文针对上述问题进行研究,提出了一种调整评分尺度差异的数据处理方法 ASSD (Adjust Score Scale Difference),首先统计训练集中所有用户的评分集合进行去最大值及最小值的操作,将剩余评分数据的评分范围记为  $\{\min(A), \max(A)\}$ , 并取平均值记为  $mean(A)$ 。同时统计每个用户的评分集合进行相同操作,将其评分范围记为  $\{\min(u), \max(u)\}$ , 平均值记为  $mean(u)$ , 其中最大值和最小值均只去除一个。然后通过上述值将不同用户的评分范围映射到相同的评分范围,得出新的评分  $I_{u,i}$ 。最后将  $I_{u,i}$  代替原来的评分值作为新的标签值用于训练,从而调整不同用户之间存在的尺度差异。评分  $I_{u,i}$  记为:

$$I_{u,i} = \begin{cases} r_{u,i} * \frac{\max(A)}{\max(u)}, & \text{if } r_{u,i} > mean(u) \\ r_{u,i} * \frac{\min(A)}{\min(u)}, & \text{if } r_{u,i} < mean(u) \\ r_{u,i} * \frac{mean(A)}{mean(u)}, & \text{if } r_{u,i} = mean(u) \end{cases} \quad (23)$$

本文将采用 ASSD 方法的模型称为 AMTR+, 实验仍在“图书-电影”和“图书-CD”两个数据集上进行,采用方式1进行训练与测试,实验参数均采用“5.4 实验设置”中的参数,与 AMTR 模型进行比较的实验结果如表9所示。

表9 不同数据集下 MSE 的比较

模型	电影-图书	电影-CD
AMTR	0.9841	0.9861
AMTR+	<b>0.9186</b>	<b>0.9180</b>

从表中可以看出,加入 ASSD 方法的 AMTR+ 模型相较于 AMTR 模型效果呈优。实验结果表明,用户之间的评分尺度差异对推荐性能有一定的影响,调整这种尺度差异有利于提高推荐效果。

## 6 总结

本文提出一种基于双注意力机制和迁移学习的跨域推荐模型 AMTR, 利用源域的信息提高目标域推荐的准确性,有效地缓解了数据稀疏和用户冷

启动问题。该模型首先利用 CNN 提取评论文本中的用户和项目特征用于评分预测;其次通过构造两种注意力机制,从词级别和特征级别提取出评论文本中的重要信息,使得模型具有更好的特征提取能力;最后利用不同领域的神经网络同时提取领域特有的特征和领域间共享的特征,进行联合建模,将特征与 FM 结合进行评分预测,从而有效地学习到了可以进行知识迁移的信息,将源域的知识充分的迁移到目标域。

实验结果表明,AMTR 模型在公开基准数据集上的性能表现总体呈优,能够取得较好的评分预测结果。实验验证了本文提出的注意力机制的有效性,同时验证了 AMTR 模型缓解了数据稀疏和用户冷启动问题以及模型具有的普适性。而本文对用户评分尺度差异研究结果表明调整这种差异会使推荐性能有所提高。因此,在今后的工作中会对用户评分尺度差异问题进行更深入的研究。

## 参考文献

- [1] Resnick P, Varian H R. Recommender systems. Communications of the Acm, 1997, 40(3): 56-68
- [2] Sarwar B M, Karypis G, Konstan J A, et al. Item-based collaborative filtering recommendation algorithms//Proceedings of the 10th International Conference on World Wide Web. Hong Kong, China, 2001:285-295
- [3] Zhao Z D, Shang M S. User-based collaborative-filtering recommendation algorithms on hadoop//Proceedings of the 2010 Third International Conference on Knowledge Discovery and Data Mining. Phuket, Thailand, 2010:478-481
- [4] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749
- [5] Cremonesi P, Tripodi A, Turrin R. Cross-domain recommender systems//Proceedings of IEEE 11th International Conference on Data Mining Workshops. Vancouver, BC, Canada, 2011: 496-503
- [6] Khan M M, Ibrahim R, Ghani I. Cross Domain Recommender Systems: A Systematic Literature Review. ACM Computing Surveys, 2017, 50(3):1-34
- [7] Pan W, Xiang E W, Liu N N, et al. Transfer learning in collaborative filtering for sparsity reduction//Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence. Atlanta, Georgia, USA, 2010: 230-235
- [8] Li B, Yang Q, Xue X. Transfer learning for collaborative filtering via a rating-matrix generative model//Proceedings of the 26th Annual

- International Conference on Machine Learning. Montreal, Quebec, Canada, 2009, 617-624
- [9] Liu P, Qiu X, Huang X, et al. Adversarial multi-task learning for text classification//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver, Canada, 2017: 1-10
- [10] Pan S J, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345-1359.
- [11] Li B, Yang Q, Xue X. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction //Proceedings of the 21st International Joint Conference on Artificial Intelligence. Pasadena, USA, 2009: 2052-2057
- [12] Gao S, Luo H, Chen D, et al. Cross-domain recommendation via cluster-level latent factor model//Proceedings of the 2013th European Conference on Machine Learning and Knowledge Discovery in Databases. Prague, Czech Republic, 2013: 161-176
- [13] Moreno O, Shapira B, Rokach L, et al. TALMUD-transfer learning for multiple domains//Proceedings of the 21st ACM international conference on information and knowledge management. Maui, HI, USA, 2012: 425-434
- [14] Wang Jun, Li Shi-Jun, Yang Sha, et al. A new transfer learning model for cross-domain recommendation. *Chinese Journal of Computers*, 2017, 40(10) :2367-2380(in Chinese)  
(王俊, 李石君, 杨莎等. 一种新的用于跨领域推荐的迁移学习模型. *计算机学报*, 2017, 40(10): 2367-2380)
- [15] Hu L, Cao J, Xu G, et al. Personalized recommendation via cross-domain triadic factorization//Proceedings of the 22nd international conference on World Wide Web. Rio de Janeiro, Brazil, 2013: 596-606
- [16] Hu L, Cao L, Cao J, et al. Learning informative priors from heterogeneous domains to improve recommendation in cold-start user domains. *ACM Transactions on Information Systems*, 2016, 35(2): 1-37
- [17] Lian J, Zhang F, Xie X, et al. CCCFNet: A content-boosted collaborative filtering neural network for cross domain recommender systems//Proceedings of the 26th International Conference on World Wide Web Companion. Perth, Australia, 2017: 817-818
- [18] Hu G, Zhang Y, Yang Q, et al. Conet: Collaborative cross networks for cross-domain recommendation//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino, Italy, 2018: 667-676
- [19] Zhang Y Huang Li-Wei, Jiang Bi-Tao, Lv Shou-Ye, et al. Suvery on deep learning based recommender systems. *Chinese Journal of Computers*, 2018, 41(7): 1619-1647(in Chinese)  
(黄立威, 江碧涛, 吕守业等. 基于深度学习的推荐系统研究综述. *计算机学报*, 2018, 41(7):1619-1647)
- [20] Cheng Z, Ding Y, Zhu L, et al. Aspect-aware latent factor model: rating prediction with ratings and reviews//Proceedings of the 2018 World Wide Web Conference. Lyon, France, 2018: 639-648
- [21] Zheng L, Noroozi V, Yu P S. Joint deep modeling of users and items using reviews for recommendation//Proceedings of the 10th ACM International Conference on Web Search and Data Mining, Cambridge, UK, 2017: 425-434
- [22] Catherine R, Cohen W W. TransNets: learning to transform for recommendation//Proceedings of the 11th ACM Conference on Recommender Systems. Como, Italy, 2017: 288-296
- [23] Seo S, Huang J, Yang H, et al. Interpretable convolutional neural networks with dual local and global attention for review rating prediction//Proceedings of the 11th ACM Conference on Recommender Systems. Como, Italy, 2017: 297-305
- [24] Tay Y, Luu A T, Hui S C, et al. Multi-pointer co-attention networks for recommendation//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, United Kingdom, 2018: 2309-2318
- [25] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch. *The Journal of Machine Learning Research*, 2011, (12): 2493-2537
- [26] Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation//Proceedings of the 19th Conference on Empirical Methods in Natural Language Processing. Doha, Oatar, 2014: 1532-1543
- [27] Hinton G. E. Rectified linear units improve restricted boltzmann machines//Proceedings of the 27th International Conference on Machine Learning. Haifa, Israe, 2010: 807-814
- [28] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality//Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada, 2013: 3111-3119
- [29] Hu J, Shen L, Sun G. Squeeze-and-excitation networks//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA, 2018: 7132-7141
- [30] Rendle S. Factorization Machines//Proceedings of the 10th IEEE International Conference on Data Mining. Sydney, Australia, 2010: 995-1000
- [31] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *IEEE Computer*, 2009, 42(8): 30-37
- [32] Srivastava N, Hinton G E, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 2014, 15(1): 1929-1958
- [33] Kingma D, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv: 1412.6980*, 2014
- [34] Schein A I, Popescul A, Ungar L H, et al. Methods and metrics for cold-start recommendations//Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval. Tampere, Finland, 2002:253-260



**Chai Yu-Mei**, born in 1964, M. S., professor. Her main research interests include machine learning, data mining and natural language processing.

**Wang Li-Ming**, born in 1963, Ph. D., professor. His main research interests include modern software engineering techniques, distributed artificial intelligence and data mining.

**Liu Zhen**, born in 1965, Ph. D., researcher. His main research interests include virtual reality and sentiment analysis.

**Yun Wu-Lian**, born in 1995, M. S. candidate. Her main research interests include natural language processing.

### Background

Traditional recommendation system has problems such as data sparsity and cold-start of users. Cross-domain recommendation is one of the methods to solve these problems in recent years, which uses knowledge from other domains to rich the target domain data to improve recommendation performance, and providing more reasonable and personalized services. Nowadays, cross-domain recommendation primarily uses rating information for knowledge transfer. In the case of highly sparse data, the use of rating information to achieve cross-domain recommendations makes it difficult for the recommendation system to model users and projects, and it is difficult to learn effective knowledge to transfer, which has greater limitations. The review text is different from ratings. It can fully reflect the user's preferences and the characteristics of the item, and can explain the reason why the user gives the item a rating. The rating can only represent the user's overall attitude to the item. Review texts providing a wealth of information for the modeling of users and projects, which was ignored by previous cross-domain recommendation efforts. Recent research has further shown using reviewers for recommendation could achieve better results.

Therefore, this paper proposes a cross-domain recommendation model based on double attention mechanism and transfer learning for the review text. Firstly, rationally model the review text using natural language processing techniques, extract more information that is beneficial to the rating prediction, and use the parallel CNN to model the user

and the project separately, the extracted features are used for rating prediction. Secondly, in the process of CNN feature extraction, by constructing dual attention mechanism, the important information in the review text is better captured from the word level and feature level. Finally, in order to solve the problem of data sparsity and cold-start of users, consider the domain-specific and shared features, introduce transfer learning based on the feature extraction network, and learn shared and domain-specific features by building source, target, and shared domains to achieve knowledge transfer, while combine with traditional recommendation techniques for cross-domain recommendations, and introduce domain discriminators to enhance the model's knowledge transfer capabilities. At the same time, this paper conducted experiments on the Amazon dataset. The experimental results show that the performance of the AMTR model on the public dataset is generally superior, and it can achieve better rating prediction results than the comparative model and effectively solves the problem of data sparsity and cold-start of users.

This work is supported by Nation Natural Science Foundation of China(No.U1636111). This paper uses the deep learning method to achieve cross-domain recommendation based on the review text, which is an effective study to improve the quality of the recommendation system.