



计算机科学

COMPUTER SCIENCE

基于自注意力机制和迁移学习的跨领域推荐算法

方义秋, 张震坤, 葛君伟

引用本文

方义秋, 张震坤, 葛君伟. 基于自注意力机制和迁移学习的跨领域推荐算法[J]. 计算机科学, 2022, 49(8): 70-77.

FANG Yi-qiu, ZHANG Zhen-kun, GE Jun-wei. Cross-domain Recommendation Algorithm Based on Self-attention Mechanism and Transfer Learning[J]. Computer Science, 2022, 49(8): 70-77.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于知识图谱的层次粒化推荐方法](#)

Hierarchical Granulation Recommendation Method Based on Knowledge Graph

计算机科学, 2022, 49(8): 64-69. <https://doi.org/10.11896/jsjcx.210600111>

[基于双目叠加仿生的微换衣行人再识别](#)

Moderate Clothes-Changing Person Re-identification Based on Bionics of Binocular Summation

计算机科学, 2022, 49(8): 165-171. <https://doi.org/10.11896/jsjcx.210600140>

[基于神经架构搜索的点击率预测模型](#)

Click-Through Rate Prediction Model Based on Neural Architecture Search

计算机科学, 2022, 49(7): 10-17. <https://doi.org/10.11896/jsjcx.210600009>

[基于概率元学习的矩阵补全预测融合算法](#)

Fusion Algorithm for Matrix Completion Prediction Based on Probabilistic Meta-learning

计算机科学, 2022, 49(7): 18-24. <https://doi.org/10.11896/jsjcx.210600126>

[融合 RACNN 和 BiLSTM 的金融领域事件隐式因果关系抽取](#)

Implicit Causality Extraction of Financial Events Integrating RACNN and BiLSTM

计算机科学, 2022, 49(7): 179-186. <https://doi.org/10.11896/jsjcx.210500190>

基于自注意力机制和迁移学习的跨领域推荐算法

方义秋¹ 张震坤¹ 葛君伟²

1 重庆邮电大学计算机科学与技术学院 重庆 400065

2 重庆邮电大学软件工程学院 重庆 400065

(fangyq@cqupt.edu.cn)

摘要 传统的单领域推荐算法受限于用户和项目的稀疏关系,存在用户/项目冷启动的问题,并且,其仅以用户对项目评分进行建模,忽略了评论文本中所蕴含的信息。基于评论文本的跨领域推荐算法在辅助领域提取用户/项目的评论信息来缓解目标领域的的数据稀疏问题,以提高推荐的准确率。文中提出了结合自注意力机制和迁移学习的跨领域推荐算法 SAMTL(Self-Attention Mechanism and Transfer Learning)。与现有算法不同,SAMTL 充分融合了目标领域和辅助领域的知识。首先,引入自注意力机制建模用户的喜好信息;其次,通过交叉映射跨域传输网络实现借助一个领域的信息来提高另一个领域的推荐准确率;最后,在知识融合模块和评分预测模块整合两个域的信息,进行评分预测。在 Amazon 数据集上的实验表明,与现有的跨领域推荐模型相比,SAMTL 的 MAE 和 MSE 值更高,在 3 种不同的跨领域数据集上的 MAE 值分别提高了 8.4%,13.2% 和 19.4%,MSE 值分别提高了 6.3%,7.8% 和 5.6%。通过多项实验验证了自注意力机制和迁移学习的有效性,以及它们在缓解数据稀疏和用户冷启动问题方面的优势。

关键词: 推荐系统;跨领域;迁移学习;自注意力;评论文本

中图法分类号 TP391

Cross-domain Recommendation Algorithm Based on Self-attention Mechanism and Transfer Learning

FANG Yi-qiu¹, ZHANG Zhen-kun¹ and GE Jun-wei²

1 School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

2 School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract Traditional single-domain recommendation algorithm is limited by the sparse relationship between users and items, and there is a problem of user/item cold start, and only models the item ratings by users, ignoring the information contained in the review text. The cross-domain recommendation algorithm based on review text extracts user/item review information in the auxiliary domain to alleviate data sparseness in the target domain and improve the accuracy of recommendation. This paper proposes a cross-domain recommendation algorithm SAMTL that combines self-attention mechanism and transfer learning. Different from existing algorithms, SAMTL fully integrates the knowledge of the target domain and auxiliary domains. Firstly, the self-attention mechanism is introduced to model user's preference information. Then, by the cross-mapping cross-domain transmission network, the recommendation accuracy of another domain is improved with the help of information in one domain. Finally, the information of the two domains is integrated in the knowledge fusion and scoring prediction module to perform scoring prediction. Experiments on Amazon data set show that, compared with the existing cross-domain recommendation model, SAMTL has higher MAE and MSE values, and MAE increases by 8.4%, 13.2% and 19.4% on three different cross-domain data sets, MSE increases by 6.3%, 7.8% and 5.6% respectively. A number of experiments verify the effectiveness of self-attention mechanism and transfer learning, as well as the advantages in alleviating data sparsity and user cold start problems.

Keywords Recommendation system, Cross-domain, Transfer learning, Self-attention, Comment text

1 引言

推荐算法能有效缓解信息过载,常用的推荐算法包括

基于内容的单领域推荐算法^[1]、混合推荐算法^[2]和协同过滤算法。其中,最经典的算法是协同过滤算法^[3],其主要根据用户的历史反馈向用户推荐项目。然而,这些方法对于没有

到稿日期:2021-06-01 返修日期:2021-09-03

基金项目:国家自然科学基金(62072066)

This work was supported by the National Natural Science Foundation of China(62072066).

通信作者:张震坤(zhangzhenkun@cqupt.email.cn)

历史反馈的新用户,即冷启动用户而言,效果甚微。近年来,跨领域推荐引起了广泛关注^[4]。跨领域推荐^[5]旨在给定两个相关的域(如音乐和电影),用户可能在一个域(源领域)中具有历史交互,而在另一个域(目标领域)中没有历史交互。在目标领域,这些用户被视为冷启动用户,但是,这两个域是相关的,存在一部分共享知识,而这部分知识可以迁移到目标领域中,以缓解目标领域的数据稀疏问题,从而提高推荐系统的推荐准确度。

如果跨领域推荐算法只利用项目所包含的用户评分数据^[6]进行推荐预测,则效果不理想,只能在一定程度上缓解数据稀疏问题,因为评分数据包含的信息单一,难以从源领域学习到有效的知识对用户和项目进行建模,并且难以将源领域知识迁移到目标领域,以提高推荐精度。

跨领域推荐的核心任务是对两个相关域之间的用户进行偏好映射。为实现这一映射,现有的方法(如 EMCDR^[7], CDLFM^[8]和 RC-DFM^[9])将用户的偏好编码为单个向量,然后对整体进行跨域映射。如图 1 所示,现有解决方案分别在源领域和目标领域中学习用户/项目表示,然后基于重叠的用户来学习并进行跨域映射。但是源领域和目标领域用户之间的直接映射无法明确捕获用户在不同域中的各种细粒度偏好,例如,喜欢摇滚音乐的用户可能更喜欢观看科幻电影。

近年来,利用用户/项目的评论进行推荐的方法激增^[10]。本文提出了一种端到端的基于自注意力机制的跨领域推荐算法。本文的主要贡献如下:

(1)在源领域和目标领域分别对用户评论聚合文本和项目评论聚合文本进行特征提取,更好地为冷启动的用户和项目进行建模。

(2)使用自注意力机制从词级别特征提取评论文本中的重要信息,提高了模型对重要信息的关注。

(3)提出了一种新的跨领域推荐模型,该模型可以有效迁移所学习到的知识,提高推荐性能,缓解数据稀疏和用户冷启动问题。在 3 个数据集上的实验结果表明,SAMTL 算法取得了更好的推荐结果。

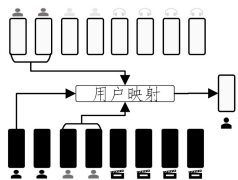


图 1 现有跨领域解决冷启动用户的方法

Fig. 1 Existing cross-domain solutions to cold-start users

2 相关工作

按照不同领域信息的处理方式可将现有的跨领域推荐技术分为聚合式推荐^[11]和迁移式跨领域推荐。Li 等^[12]提出的 CBT 算法从用户和项目的评分矩阵入手,分析了用户和项目自身独有的特征,把用户和项目分别进行聚类,通过聚类压缩得到用户和项目的评分模式并以“CodeBook”的形式存储,再将“CodeBook”迁移到目标领域中以缓解数据稀疏问题。RMGM 算法^[13]与 CBT 算法类似,通过构建和分解集群级别

的评分矩阵来预测目标矩阵中所缺失的部分。聚合式迁移算法通过嵌入式迁移技术将源领域的知识和目标领域的知识合并,其中最具代表性的算法是 TCF^[14]。该算法通过引入二进制评分来缓解目标领域出现的数据稀疏问题,提高了推荐准确性。Karatzoglou 等^[15]提出了一种融合机器学习的迁移学习方法,该方法从含有大量数据的源领域中提取知识,迁移到数据稀疏的目标领域中以解决数据稀疏问题,然后对目标领域生成推荐。Wang 等^[16]通过评分矩阵的合分解提取用户的潜在因子和共享评分模式,以及项目的潜在因子和共享评分模式,提出了一种三元桥迁移学习模型,增加了正迁移。Shi 等^[17]使用用户生成的标签作为桥梁连接源领域和目标领域,提出了基于标签推导的协同过滤跨领域推荐算法。用户生成的标签包含了用户的喜好信息,这些信息可以用来计算源领域和目标领域用户的相似度,这种相似度也可以融入基于矩阵分解的模型中,以提高推荐的准确性。Singh 等^[18]提出了一种基于协同矩阵分解的跨领域推荐系统,通过连接多个评估矩阵并跨域共享用户的偏好特征来实现跨域的知识集成。Elkahky 等^[19]提出一种基于 DSSM 模型的跨领域深度推荐模型,该方法将商品和用户的辅助信息映射到一个潜在空间,在该空间中,用户与其偏好商品之间的相似性将被最大化,再利用 DSSM 模型的匹配能力将编码到隐语义空间的用户和商品进行兴趣匹配。Hu 等^[20]提出了一种基于三元关系(用户-项目-域)的广义跨域三元分解(Cross-domain Triadic Factorization, CDTF)模型,该模型设计了两种 CDTF 算法,分别利用用户的显式和隐式反馈,以及基于遗传算法的权重参数来调整算法,以尽可能地平衡各个域之间的影响,更好地捕获用户因子和项目因子之间的相互作用。

3 相关知识

3.1 自注意力机制

本文的自注意力模型采用点积运算的方式来计算评论文本的重要性,其计算式如下:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{n}}\right)\mathbf{V} \quad (1)$$

其中, $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别是查询向量矩阵、键向量矩阵和值向量矩阵。这 3 个矩阵中每一行分别代表一个对应向量,一般通过把输入向量序列 \mathbf{X} 分别乘以 3 个矩阵 $\mathbf{W}^q, \mathbf{W}^k, \mathbf{W}^v$ 得到。

3.2 跨领域推荐

跨领域推荐的目的是丰富目标领域的数据,学习源领域中用户的喜好信息,最终提高目标领域推荐的准确性。跨领域推荐的任务定义如下。

U_S 和 U_T 是源领域和目标领域中具有用户偏好的用户集合, I_S 和 I_T 分别是源领域和目标领域中具有项目属性的项目集合,则跨领域推荐可以定义为:通过源领域的知识给目标领域的用户推荐目标领域的项目,即把 I_T 利用 $U_S \cup I_S$ 的知识推荐给 U_T 。

3.3 迁移学习

迁移学习^[21]指将在数据密度较高的领域学习到的知识迁移到数据密度较低的领域,即将在源领域中学习到的用户喜好信息迁移到目标领域,提高了目标领域推荐的准确性。

同时,迁移学习降低了传统机器学习中训练集和测试集要满足独立同分布的条件,提高了目标领域推荐的精度。迁移学习的任务定义如下。

假设源领域 D_S 和源任务 T_S , 目标领域 D_T 和目标任务 T_T , 当 $D_S \neq D_T$ 或 $T_S \neq T_T$ 时, 将 D_S 和 T_S 中学习到的 $f_S(\cdot)$ 迁移到目标领域中, 以提高目标预测函数 $f_T(\cdot)$ 的预测准确度, 迁移学习可以表示为 $\langle D_S, D_T, T_S, T_T, f_T(\cdot) \rangle$ 。

4 自注意力机制和跨域传输网络

SAMTL 模型结构如图 2 所示。SAMTL 模型由嵌入层、自注意力模块、交叉跨域传输网络和决策网络组成。其中, 嵌入层使用 GloVe^[22] 生成用户和项目的评论词向量; 自注意力模块提取用户的源领域特征、目标领域特征和项目的源领域特征、目标领域特征。交叉跨域传输网络中, 信息可进行双向传递, 从而达到使用辅助领域来提升目标领域推荐效果的目的。最后在决策网络进行不同领域的知识融合和评分预测。

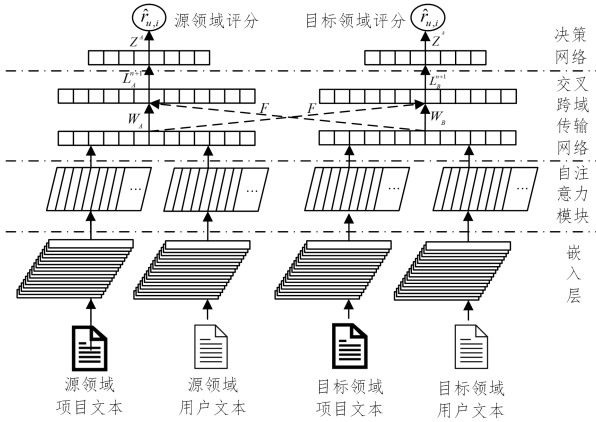


图 2 SAMTL 的体系结构

Fig. 2 Architecture of SAMTL

4.1 相关定义

本节主要介绍本文模型中输入相关数据的定义。

定义 1(输入数据表示) 输入模型的数据表示为一个三元组 $\{U, I, X_{UI}\}$, 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。

定义 2(评论聚合文本文档) 用户 U 对项目 I 填写的全部评论文本或某个项目 I 具有的用户 U 填写的全部评论文本。

定义 3(用户评论文本文档) 用户对所有项目撰写的评论文本的集合 D^U 一定程度上代表用户的偏好信息。用户 u 的评论文本文档可以表示为 $d_{1:a}^u = \{X_1, X_2, X_3, \dots, X_a\}$, 其中 a 表示用户撰写评论的数目。

定义 4(项目评论文本文档) 项目具有的所有用户填写的全部评论文本集合 D^I , 一定程度上代表项目的属性信息。项目 i 的评论文本文档可以表示为 $d_{1:b}^i = \{X_1, X_2, X_3, \dots, X_b\}$, 其中 b 表示项目拥有的评论的数目。

4.2 用户/项目评论聚合文本的词向量表示

本文将用户撰写的所有评论和项目包含的所有评论分别归纳为一个文档, 从中可以体现出用户的整体偏好以及项目的整体属性, 能够更加便捷地获取用户和项目的特征。

假设每条评论 X 包含 n 个单词, 则评论文本可以表示为:

$$X_{1:n} = x_1 \oplus x_2 \oplus x_3 \oplus \dots \oplus x_n \quad (2)$$

其中, \oplus 表示连接操作。

使用 GloVe 生成每个词的词向量矩阵, 每个词向量维度为 k , 评论 X 的句向量可表示为:

$$c = c_{x_1} \oplus c_{x_2} \oplus c_{x_3} \oplus \dots \oplus c_{x_n} \quad (3)$$

计算得到所有评论的句向量, 根据定义 3 可以得出用户评论文本文档 $d_{1:a}^u$ 的词向量, 记为 $V_{1:a}^u$, 其表达式为:

$$V_{1:a}^u = c_1 \oplus c_2 \oplus c_3 \oplus \dots \oplus c_a \quad (4)$$

根据定义 4 可以得出项目评论文本文档 $d_{1:b}^i$ 的词向量, 记为 $V_{1:b}^i$:

$$V_{1:b}^i = c_1 \oplus c_2 \oplus c_3 \oplus \dots \oplus c_b \quad (5)$$

评论聚合文本词向量生成的具体流程如算法 1 所示。

算法 1 评论聚合文本词向量生成算法

输入: 评论文本集合 $\{\{U, I, X_{UI}\}_{(1 \dots N)}\}$

输出: 用户评论聚合文本 d^u 和项目评论聚合文本 d^i

1. for 评论文本集中所有用户 $u(u \in U)$ do
2. if $u = \{U, I, X_{UI}\}[0]$ then
3. 将评论 X_{UI} 词向量化生成句向量 c_{UI}
4. 执行操作(1): $d^u = d^u \oplus c_{UI}$
5. end if
6. end for
7. for 评论文本集中的每一个项目 $i(i \in D)$ do
8. if $i = \{U, I, X_{UI}\}[0]$ then
9. 将评论 X_{UI} 词向量化生成句向量 c_{UI}
10. 执行操作(2): $d^i = d^i \oplus c_{UI}$
11. end if
12. end for
13. return d^u, d^i

其中, $\{U, I, X_{UI}\}[0]$ 表示评论文本集合的第 1 个元素, $\{U, I, X_{UI}\}[1]$ 表示评论文本集合的第 2 个元素。

4.3 自注意力机制网络

注意力机制关注部分细节, 而非基于全部信息。本质上, Attention 是从大量信息中筛选出少量重要信息, 并聚焦到这些重要信息上, 忽略大多数不重要的信息。本文在嵌入层和交叉跨域传输网络之间加入了自注意力模块。

由于卷积神经网络 (Convolutional Neural Network, CNN) 中卷积核尺寸的限制, 卷积操作覆盖的范围非常有限, 当特征距离稍远时, 就难以获取。而自注意力机制不仅能够建模序列中用户评论的上下文信息, 还能够通过计算序列中任意 2 个评论之间的相关程度来捕获行为的长期依赖关系和序列的全局信息, 缓解 CNN 只能覆盖局部信息的问题。

自注意力机制对外部信息依赖较少, 更能捕捉到内部信息的特征, 自注意力机制如图 3 所示。给定用户的嵌入向量构成的矩阵 $V_u = [v_{u,1}, v_{u,2}, v_{u,3}, \dots, v_{u,a}]$, $v_{u,i} \in V_{1:a}^u$, 其中 a 表示用户的个数。由用户向量构成的 query, key 和 value 矩阵作为自注意力机制的输入, 输出是带有权重的 value 矩阵。具体算法步骤如下。

步骤 1 将 query 和 key 投影到同一空间中得到 Q 和 K 。

$$Q = \text{ReLU}(V_u W_Q) \quad (6)$$

$$K = \text{ReLU}(V_u W_K) \quad (7)$$

其中, $\mathbf{W}_Q \in \mathbf{V}_{a \times a} = \mathbf{W}_K \in \mathbf{V}_{a \times a}$ 分别为 query 和 key 的权重矩阵, 使用 ReLU 非线性激活函数进行学习。

步骤 2 将 1 中的结果相乘并进行 softmax 归一化处理, 输出 $a \times a$ 的矩阵 \mathbf{S}_u 。

$$\mathbf{S}_u = \text{softmax} \left(\frac{\mathbf{QK}^T}{\sqrt{n}} \right) \quad (8)$$

其中, $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别表示自注意力模块中的 query, key 和 value。 \mathbf{QK}^T 采用点积运算的方式是为了计算 query 与某个 key 之间的相似性, 其中 scale 操作 \sqrt{n} 是为了避免与 \mathbf{K} 的某个点积结果过大, 导致归一化后出现 0 或 1 的情况。为了避免 key 和 query 由相同向量导致的高匹配分数, 在 scale 之后使用 mask 来隐藏矩阵的对角线元素。 softmax 表示对 $(\mathbf{QK}^T / \sqrt{n})$ 运算之后的结果进行归一化处理, 使计算后得到的数值变成权重之和为 1 的概率分布。对每一个样本, 有一个 n_q 维的查询向量, 构成一个 $a \times n_q$ 维的查询向量矩阵 \mathbf{Q} , 可以认为查询向量是样本的特征。然后, 对数据中的每条信息(向量), 都有一个 n_k 维的键向量和 n_v 维的值向量, 构成一个键值对。

步骤 3 将步骤 2 中得到的权重矩阵 \mathbf{S}_u 与 value 值相乘, 得到用户自注意力模块的最终加权值 \mathbf{Z}_u 。

$$\mathbf{Z}_u = \mathbf{S}_u \mathbf{V}_u \quad (9)$$

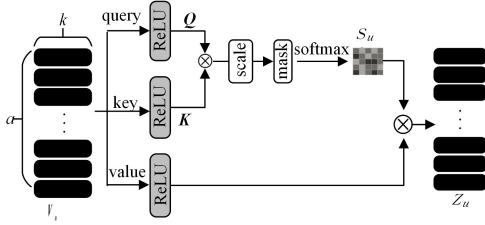


图 3 自注意力机制结构

Fig. 3 Self-attention mechanism structure

4.4 跨域传输网络

4.4.1 基于交叉映射的感知网络

本节提出了一种基于交叉映射的跨域传输网络^[23]来实现域跟域之间的信息传输。其主要任务是在两个域之间通过在同一层网络中共享同一个矩阵来实现域和域之间的双向信息传递, 从而达到使用辅助域来提升目标领域推荐效果的目的。

跨域传输网络(Cross-domain Transport Network, CTN)的架构如图 4 所示, 可以看出 CTN 是一个单层交叉映射感知网络, 它以来自源领域和目标领域两个域的用户矩阵作为输入, 对其中的第 n 层和第 $n+1$ 层的具体描述如下。

(1)第 n 层

令 \mathbf{L}_A^n 表示 CTN 在域 A 中的第 n 层的输出, \mathbf{L}_B^n 表示 CTN 在域 B 中第 n 层的输出, $\mathbf{\Gamma}_A^n$ 表示同时考虑两个域信息的在域 A 的第 n 层的输出, 即域 A 的第 $n+1$ 层的输入。 \mathbf{L}_A^{n+1} 表示在域 A 的第 $n+1$ 层的输出。类似地, $\mathbf{\Gamma}_B^n$ 和 \mathbf{L}_B^{n+1} 也是同样的释义。CTN 在第 n 层域 A 和域 B 的输出公式如下:

$$\mathbf{\Gamma}_A^n = \mathbf{W}_A^n \mathbf{L}_A^n + \mathbf{F}^n \mathbf{L}_B^n \quad (10)$$

$$\mathbf{\Gamma}_B^n = \mathbf{W}_B^n \mathbf{L}_B^n + \mathbf{F}^n \mathbf{L}_A^n \quad (11)$$

其中, \mathbf{W}_A 和 \mathbf{W}_B 分别表示域 A 和域 B 的权重矩阵, 矩阵 \mathbf{F} 的作用是控制域 A 与域 B 之间信息流的传输。跨域传输的信息会同时从域 A 传向域 B 和从域 B 传向域 A 中。因此, CTN 可以实现知识的双向传输, 保证模型在每个域都可以

接收到用户在其他辅助域的相关信息。

(2)第 $n+1$ 层

跨域传输模块通过一个交叉映射网络来实现域跟域之间的信息传输。域 A 和域 B 通过交叉连接进行耦合:

$$\mathbf{L}_A^{n+1} = \sigma(\mathbf{\Gamma}_A^n) = \sigma(\mathbf{W}_A^n \mathbf{L}_A^n + \mathbf{F}^n \mathbf{L}_B^n) \quad (12)$$

$$\mathbf{L}_B^{n+1} = \sigma(\mathbf{\Gamma}_B^n) = \sigma(\mathbf{W}_B^n \mathbf{L}_B^n + \mathbf{F}^n \mathbf{L}_A^n) \quad (13)$$

其中, $\sigma(\cdot)$ 表示激活函数(在实验中, 使用 ReLU 作为模型的激活函数)。由式(12)可知, 域 A 中 \mathbf{L}_A^{n+1} 会接收到两部分信息: 一部分是权重矩阵 \mathbf{W}_A 控制的本地信息 \mathbf{L}_A^n ; 另一部分是由矩阵 \mathbf{F} 控制的来自域 B 的信息(域 B 同理)。本文将矩阵 \mathbf{F} 称为传输矩阵, 因为矩阵 \mathbf{F} 控制着来自其他域的信息传输。

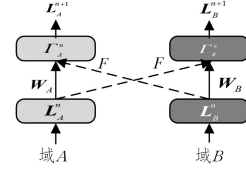


图 4 交叉跨域传输网络

Fig. 4 Cross-domain transmission network

4.4.2 知识融合和评分预测

自注意力机制分别在源领域和目标领域提取用户和项目的特征后, 再由全连接层将从源领域提取的特征和从目标领域提取的特征进行融合, 并映射到同一特征空间中, 由此得到最终特征 \mathbf{Z}^k :

$$\mathbf{Z}^k = \begin{cases} f(\mathbf{W}_s \mathbf{L}_s^{n+1} + b_s), & \text{if } k = s \\ f(\mathbf{W}_t \mathbf{L}_t^{n+1} + b_t), & \text{if } k = t \end{cases} \quad (14)$$

其中, f 为 ReLU 激活函数^[24]; \mathbf{W}_s 表示源领域权重; \mathbf{W}_t 表示目标领域权重; b_s 表示源领域的偏执量; b_t 表示目标领域的偏执量; k 表示领域标签, $k = s$ 为源领域, $k = t$ 为目标领域。

$$\mathbf{S}_{u,i} = \mathbf{Z}^T \mathbf{WZ}' \quad (15)$$

其中, $\mathbf{S}_{u,i}$ 表示源领域和目标领域具有的共同特征。

最后, 得到用户对项目的预测评分 $\hat{r}_{u,i}$:

$$\hat{r}_{u,i} = \frac{1}{M * M} \sum_{p=1}^M \sum_{q=1}^M \mathbf{S}_{u,i}(p, q) + b_u + b_i \quad (16)$$

其中, M 表示源领域和目标领域的序列个数, b_u 和 b_i 分别表示用户偏移量和项目偏移量。

4.4.3 损失函数

本文使用平方损失函数作为目标函数对各个域进行如下训练:

$$L = \frac{1}{|\mathbf{S}|} \sum_{(u,i) \in \mathbf{S}} (r_{u,i} - \hat{r}_{u,i})^2 + \frac{\lambda}{2} \|\Theta\|^2 \quad (17)$$

其中, $r_{u,i}$ 和 $\hat{r}_{u,i}$ 为用户 u 对项目 i 的真实评分和预测评分, \mathbf{S} 表示源领域和目标领域的训练序列, 为了防止过拟合问题, 加入正则化项 $\frac{\lambda}{2} \|\Theta\|^2$, λ 为正则化参数, Θ 表示模型参数。模型中所有的参数都是以端到端的方式进行学习。

5 实验结果及分析

5.1 实验数据集

为了保证实验结果的可信度, 本文选取亚马逊公司的公开数据集进行实验。亚马逊公司的公开数据集主要包括了用户

对商品的评论信息和商品的属性信息,其中评论信息一共有142800000条,评分范围为1~5的整数,时间跨度是1996—2014年。

本文在亚马逊公司数据集中选择“电影”、“图书”和“音乐”作为实验数据,在每个域中,删除没有评论文本的交互记录,然后过滤掉交互少于10个的用户和交互少于30个项目。数据集统计信息如表1所列。由于这3个域彼此相关,首先构建了3个跨域方案:“电影”和“图书”、“电影”和“音

表1 3种跨领域场景统计

Table 1 Statistics of three cross-domain scenarios

Domain	Reviews	Overlap users	Overlap users ratio	Verification user	Test users	$\theta=100\%$	$\theta=70\%$	$\theta=50\%$	$\theta=30\%$
D_s	Movie	1 697 533	21.83%	1 214	1 823	3 037	2 125	1 519	911
D_t	Book	8 898 041	4.80%						
D_s	Movie	1 697 533	10.00%	556	835	1 391	943	696	417
D_t	Music	1 097 592	25.17%						
D_s	Book	8 898 041	1.35%	340	512	853	597	427	256
D_t	Music	1 097 592	15.43%						

注: θ 代表重叠用户的比例

5.2 评价指标

本文选用平均绝对误差 MAE 和均方误差 MSE 作为评价指标,推荐精度越高,代表预测分数越接近真实分数,即 MAE 和 MSE 两个指标越小,代表推荐准确性越高。MAE 和 MSE 的定义分别为:

$$MAE = \frac{1}{N} \sum_{i=1}^N |r_{u,i} - \hat{r}_{u,i}| \quad (18)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (r_{u,i} - \hat{r}_{u,i})^2 \quad (19)$$

其中, N 表示测试数据的数量, $r_{u,i}$ 表示测试数据的真实评分, $\hat{r}_{u,i}$ 表示模型计算后得到的对 $r_{u,i}$ 的预测值。

5.3 对比模型

为了评估 SAMTL 模型的性能,将其与以下模型进行了对比。

(1)D-ATT^[25]。一种基于注意力机制的推荐模型。通过注意力机制提取重要单词并获取到用户和项目的特征后再输入到卷积神经网络编码层进行评分预测。

(2)ANR^[26]。一种基于跨域联合矩阵分解模型。通过构建用户对不同项目的评分矩阵,将用户评论和项目类型相关联进行评分预测。

(3)CDTF^[20]。一种基于张量分解的跨领域推荐模型。通过张量的三因式分解来捕捉用户-项目-领域的三元关系,从而进行知识迁移。

(4)MF^[27]。一种基于矩阵分解的推荐算法。将评分矩阵分解为用户和项目矩阵,由此来预测用户和项目矩阵中缺失的部分。

(5)DFM^[9]。一种使用堆叠式降噪自动编码器(SDAE)将审阅文本和项目内容与辅助域和目标领域中的评分矩阵融合在一起,使用多层感知器在两个域之间转移用户潜在因素由此来解决数据稀疏性和冷启动问题。

(6)EMCDR^[7]。一种跨域推荐的嵌入和映射框架。使用多层感知器来捕获跨域的非线性映射,以及使用具有大量数

据的实体来学习映射功能。

乐”以及“图书”和“音乐”。在每种情况下,选择具有更多用户的域作为源领域,另一个为目标领域。从所有重叠用户中抽取50%用户作为冷启动用户,即这部分用户在目标领域中的交互被隐藏,但是他们的交互信息可以用来验证以及测试,测试用户占30%,验证用户占20%,剩下的50%的重叠用户用于训练。为了模拟不同比例的重叠用户,设定剩余的50%的重叠用户按照一定的比例(100%,70%,50%,30%)来构成训练集。

据的实体来学习映射功能。

(7)CBT^[12]。一种跨领域推荐的经典模型。假设不同领域间具有类似的评分模式,将用户和项目的特征因子通过聚类的方式合并为评分模式矩阵,在不同领域间进行评分模式的共享。

5.4 实验设置

本文实验环境配置如下:操作系统为 ubuntu16.04;内存为128 GB;处理器为 Intel (R) Xeon (R) Gold 5218 CPU 2.30 GHz;图形处理器为两张 Quadro RTX 4000 GPU,磁盘为4TB机械硬盘;开发框架为 tensorflow2.2。预处理数据集中所有的用户和项目文档:1)删除停用词和文档中出现频率较高的词(相对文档频率高于0.45);2)根据其 tf-idf (term frequency-inverse document frequency)分数,选择前20000个单词作为词汇,并从原文档中删除其它单词;3)截断长文档和填充短文档,使它们的长度大约等于500个单词。利用 Google News 中经过训练的300维词向量嵌入来获取每个单词的初始嵌入向量,卷积核数量设置为200,窗口大小为3。在训练过程中,采用 dropout 策略随机忽略一小部分值,设置 dropout 比率为0.6,学习率为0.001,进行模型训练。

本文研究了卷积核数目对 SAMTL 模型性能的影响。分别设置卷积核数量为50,100,150,200,250。在“音乐-图书”数据集中得到的实验结果如图5所示。

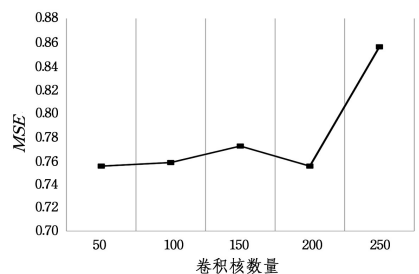


图5 卷积核数量对 MSE 值的影响

Fig. 5 Influence of convolution kernel number on MSE value

从图5可以看出,卷积核数量在50~200时,模型的MSE值存在波动;卷积核数量在200时,MSE的值达到最小;卷积核数量到达250时,MSE值急剧增大,模型性能衰减,因此本文设置卷积核数量为200。

5.5 实验结果及分析

为了验证SAMTL模型的推荐精度,将SAMTL模型与现有的跨领域推荐模型ANR,CDTF,DFM以及EMCDR在“电影-图书”“电影-音乐”和“图书-音乐”3个数据集上进行比较,并采用MAE和MSE作为实验评估指标,实验结果如表2所列。

表2 不同方法在不同数据集上MAE和MSE的值

Table 2 MAE and MSE of different methods in different datasets

$D_s \rightarrow D_t$	Movie \rightarrow Book		Movie \rightarrow Music		Book \rightarrow Music	
	MAE	MSE	MAE	MSE	MAE	MSE
ANR	0.8673	1.1241	0.9152	1.1223	0.8204	0.8953
CDTF	1.1532	1.2801	1.1521	1.2241	0.9312	0.9831
DFM	0.9254	1.1404	0.8562	1.1371	0.8814	0.9234
EMCDR	0.9344	1.1301	0.8932	1.1170	0.8542	0.9241
SAMTL	0.8002	1.0572	0.7563	1.0361	0.6869	0.8482

从表2可知,SAMTL模型在3个数据集上的MAE指标和MSE指标均优于对比模型,且都是最优结果。SAMTL模型在“电影-图书”数据集上的MAE值和MES值分别为0.8002和1.0572,比对比模型中效果最好的ANR模型分别提高了8.4%和6.3%。

在“电影-音乐”数据集上,其MAE值为0.7563,比对比模型中表现最好的DFM模型提高了13.2%;MSE值为1.0361,比对比模型中MSE最好的EMCDR提高了7.8%。在“图书-音乐”数据集上,本文模型的MAE值为0.6869,比对比模型中MAE值最高的ANR模型提高了19.4%;MSE值为0.8482,比ANR模型提高了5.6%。结果表明,SAMTL模型具有更高的评分预测准确度。

5.6 迁移学习的有效性

为验证SAMTL模型具有有效的知识迁移以及表明迁移学习可以提高目标领域推荐精度的影响,将SAMTL模型与MF和D-ATT两个单领域的推荐模型进行对比,两个单领域推荐模型只在“电影”和“图书”数据集上进行测试。本实验也采用MSE作为评价指标,结果如图6和图7所示。

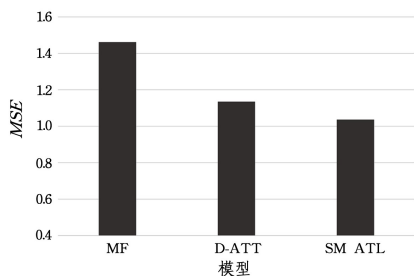


图6 “音乐”-“电影”数据集上的MSE比较

Fig. 6 MSE comparison between “music” and “movie” datasets

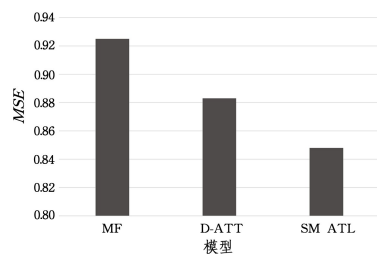


图7 “音乐”-“图书”数据集上的MSE比较

Fig. 7 MSE comparison between “music” and “book” datasets

图6和图7的结果表明,本文提出的SAMTL模型优于对比模型。在“音乐”-“电影”数据集上SAMTL模型的MSE值为1.0362,相比MF模型和D-ATT模型分别提升了41.1%和9.6%;在“音乐”-“图书”数据集上,SAMTL模型的MSE值为0.8482,相比MF和D-ATT模型分别提升了9.1%,4.1%。从实验结果可以看出,使用了辅助领域数据的SAMTL模型在目标领域的推荐精度要高于单领域推荐模型,表明SAMTL模型具有更好的知识迁移性能。

从图8和图9可以看出,利用评论文本的模型CBT和D-ATT以及SAMTL的MSE值相比利用评分的MF在“音乐”-“电影”数据集上分别提升了32.4%,28.8%和41.1%;在“音乐”-“图书”数据集上分别提升了6.2%,4.8%和9.1%。由此可知,与使用评分相比,使用评论文本可以提高模型的推荐性能。

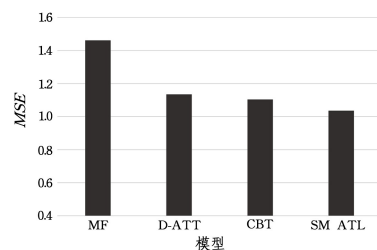


图8 “音乐”-“电影”数据集上的MSE比较

Fig. 8 MSE comparison between “music” and “movie” datasets

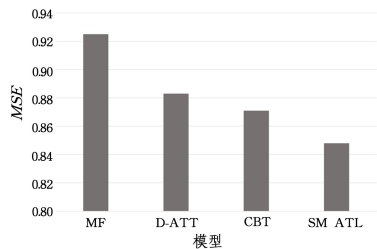


图9 “音乐”-“图书”数据集上的MSE比较

Fig. 9 MSE comparison between “music” and “book” datasets

本文设计了一组新的对比实验来进一步的验证SAMTL模型能有效地进行知识迁移。将单领域推荐模型的数据集的50%作为训练集,剩下的50%作为验证集。实验结果如图10和图11所示。

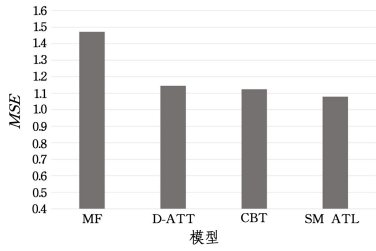


图 10 “音乐”-“电影”数据集中的 MSE 比较

Fig. 10 MSE comparison between “music” and “movie” datasets

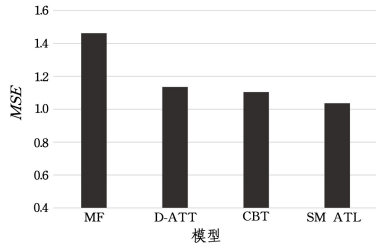


图 11 “音乐”-“电影”数据集中的 MSE 比较

Fig. 11 MSE comparison between “music” and “movie” datasets

从图 10 和图 11 可知, SAMTL 模型在“音乐”-“电影”和“音乐”-“图书”数据集中的 MSE 值均为最低, 说明 SAMTL 具有最佳的推荐精度; 在数据集“音乐”-“电影”中, SAMTL 模型的 MSE 值为 1.0793, 相比 MF, D-ATT 和 CBT 分别提升了 36.4%, 6.1% 和 4.2%; 在“音乐”-“图书”数据集上, SAMTL 模型的 MSE 值为 0.8732, 相比 MF, D-ATT 和 CBT 分别提升了 9.2%, 4.5% 和 2.4%。从结果可知, SAMTL 模型可以通过迁移学习利用源领域的知识来提高目标领域的推荐精度, 表明 SAMTL 模型具有较好的知识迁移性能。

5.7 缓解数据稀疏的问题

本文通过与单领域模型和跨领域模型的对比来验证 SAMTL 模型可以缓解目标领域数据稀疏的问题。在实验中, 通过控制测试集用户的评论数来调整数据的稀疏程度, 从用户的评论中随机挑选 α 倍的评论作为测试集, α 分别设为 30%, 50%, 70% 以及 100%。

(1) 与跨领域推荐模型的比较

本实验对跨领域推荐算法采用不同的 α 进行测试, 结果如图 12 和图 13 所示。

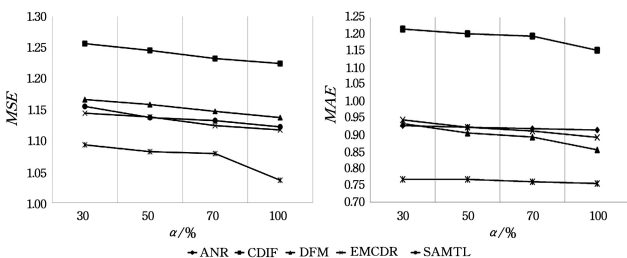


图 12 “音乐”-“电影”数据集中, 不同评价标准的实验结果比较

Fig. 12 Comparison of experimental results of different evaluation criteria in “music” and “movie” datasets

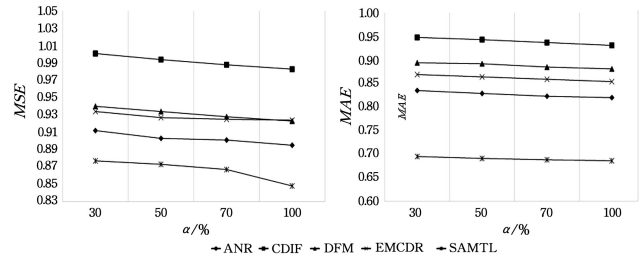


图 13 “音乐”-“图书”数据集上, 不同评价标准的实验结果比较

Fig. 13 Comparison of experimental results of different evaluation criteria in “music” and “book” datasets

其中, α 表示数据的稀疏程度, α 越小代表数据越稀疏。从图 12 和图 13 可以看出, 无论 α 如何变化, SAMTL 模型在“音乐”-“电影”和“音乐”-“图书”两个数据集上的 MAE 和 MSE 值均最小, 取得了较好的效果。

图 12 给出了“音乐”-“电影”数据集上的实验结果, 可以看出, 随着 α 的增加, 即数据稀疏程度的降低, 各个模型的推荐效果均有提升。例如, 当 α 为 100% 时, 相比 α 为 30% 时, ANR, CDIF, DFM, EMCDCR 和 SAMTL 模型的 MAE 值分别提高了 1.4%, 5.5%, 9.1%, 5.8% 和 1.6%。相比 α 为 50% 时, 其 MAE 值分别提高了 0.9%, 4.3%, 5.8%, 3.4% 和 1.6%。由此可见, SAMTL 模型相比对比模型可以获取到更多有利于预测评分的用户信息和项目信息。

图 13 给出了“音乐”-“图书”数据集上的实验结果, 可以看出, 随着 α 增加, 各个模型的推荐效果均有提升。例如, 当 α 为 100% 时, 相比 α 为 30% 时, ANR, CDIF, DFM, EMCDCR 和 SAMTL 模型的 MAE 值分别提高了 1.8%, 1.8%, 1.5%, 1.8% 和 1.3%。相比 α 为 50% 时, 其 MAE 值分别提升了 1.1%, 1.3%, 1.2%, 1.2% 和 0.7%。由此可见, SAMTL 模型相比对比模型可以获取到更多有利于预测评分的用户信息和项目信息。

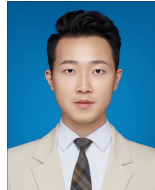
结束语 本文提出了一种基于自注意力机制和迁移学习的跨领域推荐算法模型 SAMTL。该模型将源领域学习到的知识迁移到目标领域来提高推荐精度, 有效缓解了用户冷启动和数据稀疏带来的问题。SAMTL 模型首先利用卷积神经网络从评论文本中提取用户和项目的特征信息, 用于评分预测; 然后在用户级和项目级使用自注意力机制, 从词级别和特征级别提取评论文本的重要信息, 充分考虑用户偏好和目标项目之间的关系; 最后提取源领域和目标领域特有的特征进行联合建模, 从而有效地将源领域信息迁移到目标领域, 提高了推荐质量。虽然 SAMTL 模型取得了较好的结果, 但其仍具有一定的局限性。在用户和商品建模上, 本文只使用了用户评论信息和商品评论信息, 但还存在许多用户和商品的其他可用信息(如用户职业、用户浏览记录、商品价格等), 未来将加入这些可用信息以及增加相似用户的信息来提高推荐的准确性。

参考文献

- Using Learning for Text Categorization[J]. arXiv:cs/9902011, 1999.
- [2] BALABANOVIC M,SHOHAM Y. Fab:content-based,collaborative recommendation[J]. Communications of the ACM,1997,40(3):66-72.
- [3] SALAKHUTDINOV R,MNIH A,HINTON G. Restricted Boltzmann machines for collaborative filtering[C]//Proceedings of the 24th International Conference on Machine Learning. 2007:791-798.
- [4] ZHU F,WANG Y,CHEN C,et al. A deep framework for cross-domain and cross-system recommendations[C]//IJCAI International Joint Conference on Artificial Intelligence. 2018:3711-3717.
- [5] KHAN M M,IBRAHIM R,GHANI I. Cross domain recommender systems:a systematic literature review[J]. ACM Computing Surveys(CSUR),2017,50(3):1-34.
- [6] PAN W, XIANG E, LIU N, et al. Transfer learning in collaborative filtering for sparsity reduction [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2010:230-235.
- [7] MAN T, SHEN H, JIN X, et al. Cross-Domain Recommendation: An Embedding and Mapping Approach[C]//IJCAI. 2017:2464-2470.
- [8] WANG X,PENG Z,WANG S,et al. Cross-domain Recommendation for Cold-start Users via Neighborhood based feature mapping[C]//International Conference on Database Systems for Advanced Applications. Cham:Springer,2018:158-165.
- [9] FU W,PENG Z,WANG S,et al. Deeply Fusing Reviews and Contents for Cold Start Users in Cross-Domain Recommendation Systems[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2019,33:94-101.
- [10] LI C,QUAN C,PENG L,et al. A capsule network for recommendation and explaining what you like and dislike[C]// Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2019:275-284.
- [11] LONI B,SHI Y,LARSON M,et al. Cross-domain collaborative filtering with factorization machines[C]//European Conference on Information Retrieval. Cham:Springer,2014:656-661.
- [12] LI B,QIANG Y,XUE X. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction[C]// Twenty-First International Joint Conference on Artificial Intelligence. 2009:2052-2057.
- [13] LI B,YANG Q,XUE X. Transfer learning for collaborative filtering via a rating-matrix generative model[C]//Proceedings of the 26th Annual International Conference on Machine Learning. 2009:617-624.
- [14] PAN W, XIANG E, YANG Q. Transfer learning in collaborative filtering with uncertain ratings[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2012:662-668.
- [15] KARATZOGLOU A,AMATRIAIN X,BALTRUNAS L,et al. Multiverse recommendation:N-dimensional tensor factorization for context-aware collaborative filtering[C]//Proceedings of the Fourth ACM Conference on Recommender Systems. 2010:79-86.
- [16] WANG J,LI S J,YANG S,et al. A new transfer learning model for cross-domain recommendation[J]. Chinese Journal of Computers,2017,40(33):1-15.
- [17] SHI Y,LARSON M,HANJALIC A. Tags as bridges between domains:Improving recommendation with tag-induced cross-domain collaborative filtering [C] // International Conference on User Modeling, Adaptation, and Personalization. Berlin:Springer,2011:305-316.
- [18] 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. 2008:650-658.
- [19] ELKAHKY A M,SONG Y,HE X. A multi-view deep learning approach for cross domain user modeling in recommendation systems[C]//Proceedings of the 24th International Conference on World Wide Web. 2015:278-288.
- [20] HU L,CAO J,XU G,et al. Personalized recommendation via cross-domain triadic factorization[C]// Proceedings of the 22nd International Conference on World Wide Web. 2013:595-606.
- [21] WU J,XU J,DING T. Fine-grained image classification algorithm based on ensemble methods of transfer learning [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition),2020,32(3):452-458.
- [22] PENNINGTON J,SOCHER R,MANNING C D. Glove:Global vectors for word representation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing(EMNLP). 2014:1532-1543.
- [23] HU G,ZHANG Y,YANG Q. Conet:Collaborative cross networks for cross-domain recommendation[C]// Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018:667-676.
- [24] NAIR V,HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//ICML. 2010:807-814.
- [25] SEO S,HUANG J,YANG H,et al. Interpretable convolutional neural networks with dual local and global attention for review rating prediction[C]//Proceedings of the eleventh ACM Conference on Recommender Systems. 2017:297-305.
- [26] CHIN J Y,ZHAN K,JOTY S,et al. ANR:Aspect-based neural recommender[C]// Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018:147-156.



FANG Yi-qiu, born in 1963, master, associate professor, graduate supervisor. Her main research interests include cloud computing and big data.



ZHANG Zhen-kun, born in 1997, master. His main research interests include big data and so on.