



计算机应用
Journal of Computer Applications
ISSN 1001-9081, CN 51-1307/TP

《计算机应用》网络首发论文

题目：基于双重最相关注意力网络的协同过滤推荐算法
作者：张文龙，钱付兰，陈洁，赵姝，张燕平
收稿日期：2020-06-12
网络首发日期：2020-10-23
引用格式：张文龙，钱付兰，陈洁，赵姝，张燕平. 基于双重最相关注意力网络的协同过滤推荐算法[J/OL]. 计算机应用.
<https://kns.cnki.net/kcms/detail/51.1307.TP.20201023.1009.002.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于双重最相关注意力网络的协同过滤推荐算法

张文龙^{1,2}, 钱付兰^{1,2*}, 陈洁^{1,2}, 赵姝^{1,2}, 张燕平^{1,2}

(1.安徽大学 计算机科学与技术学院, 合肥 230601;

2.计算智能与信号处理教育部重点实验室(安徽大学), 合肥 230601)

(*通信作者电子邮箱 qianfulan@hotmail.com)

摘要: 基于项目的协同过滤从用户的历史交互项目中学习用户偏好, 根据用户的偏好推荐相似的新项目。现有的协同过滤方法认为用户所交互的一组历史项目对用户的影响是相同的, 并且将所有历史交互项目在对目标项目作预测时的贡献看作是相同的, 导致这些推荐方法的准确性受限。针对上述问题, 提出了一种基于双重最相关注意力网络的协同过滤推荐算法, 该算法包含两层注意力网络。首先, 使用项目级注意力网络为不同历史项目分配不同的权重来捕获用户历史交互项目中最相关的项目; 然后, 使用项目交互级注意力网络感知不同历史项目与目标项目之间交互的关联度; 最后, 通过两层注意力网络的使用来同时捕获用户在历史交互项目上和目标项目上的细粒度偏好, 更好地进行下一步推荐工作。在 MovieLens 和 Pinterest 两个真实数据集上进行实验, 实验结果表明, 所提算法在推荐命中率上与基准模型 DeepICF 相比分别提升了 2.3 个百分点和 1.5 个百分点, 验证了该算法为用户进行个性化推荐时的有效性。

关键词: 推荐系统; 协同过滤; 深度学习; 隐式反馈; 注意力机制

中图分类号: TP183

文献标志码: A

Collaborative filtering recommendation algorithm based on dual most relevant attention network

ZHANG Wenlong^{1,2}, QIAN Fulan^{1,2*}, CHEN Jie^{1,2}, ZHAO Shu^{1,2}, ZHANG Yanping^{1,2}

(1. College of Computer Science and Technology, Anhui University, Hefei Anhui 230601, China;

2. Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education, (Anhui University), Hefei Anhui 230601, China)

Abstract: Item-based collaborative filtering learns user preferences from the user's historical interaction items and recommends similar new items based on the user's preferences. The previous collaborative filtering methods assume that a set of historical items that user has interacted with has the same impact on user, and all historical interaction items are considered to have the same contribution to the prediction of target item, which limits the accuracy of these recommendation methods. In order to solve the problems, a dual most relevant attention network recommendation algorithm was proposed, which contained two layers item level attention and item interaction level attention. Firstly, the item-level attention network was used to assign different weights to different historical items to capture the most relevant items in the user historical interaction items. Then, the item interaction level attention network was used to perceive the correlation degree of the interaction between the different historical items and the target item. Finally, the fine-grained preferences of users on both the historical interaction items and the target items could be captured through two layers attention network, so as to made better recommendations for the next step. The experiments were conducted on two real datasets MovieLens and Pinterest. The experimental results show that, the proposed algorithm improves the recommendation hit rate by 2.3% and 1.5% respectively compared with the benchmark model DeepICF, which verifies the effectiveness of the algorithm when making personalized recommendations for users.

Keywords: recommender system; collaborative filtering; deep learning; implicit feedback; attention mechanism

0 引言

互联网和移动技术的出现使人们可以随时随地获取信息。随着社交媒体、电子商务和各种生活服务应用等网络服

务的普及, 人们的生活方式发生了深刻的变化。为了能够在海量数据中快速有效地获取对用户最有价值的数据, 推荐系统应运而生^[1]。个性化推荐的目标是从历史的用户-项目交互(例如评分和浏览等)中了解用户偏好, 并根据这些用户-项目

收稿日期: 2020-06-12; 修回日期: 2020-09-16; 录用日期: 2020-09-27。

基金项目: 国家自然科学基金资助项目(61702003); 安徽省自然科学基金资助项目(1808085MF175)。

作者简介: 张文龙(1996—), 男, 安徽阜阳人, 硕士研究生, 主要研究方向: 深度学习、推荐系统; 钱付兰(1978—), 女, 安徽蚌埠人, 副教授, 博士, 主要研究方向: 粒计算、社交网络、推荐系统; 陈洁(1982—), 女, 安徽巢湖人, 副教授, 博士, 主要研究方向: 智能计算、机器学习、三支决策; 赵姝(1979—), 女, 安徽巢湖人, 教授, 博士, 主要研究方向: 粒计算、商空间理论、机器学习; 张燕平(1962—), 女, 安徽合肥人, 教授, 博士, 主要研究方向: 智能计算、粒计算、商空间理论。

交互推荐相关的项目,这种方法被称为协同过滤(Collaborative Filtering, CF)^[2-3]。

在基于协同过滤的方法中,最成功的就是通过矩阵分解(Matrix Factorization, MF)^[4]的方法,将用户和项目映射到同一个潜在空间中,使用潜在特征向量来表示用户或项目。然后,将用户对项目的交互建模为其潜在向量的内积。然而,实际中的评分矩阵往往是非常稀疏的,这导致基于协同过滤的方法在学习合适的潜在表示时性能显著下降。最近,学习有效表征的强大方法之一便是深度学习(Deep Learning, DL)。因此,随着大规模评分和丰富的附加信息,将深度学习与协同过滤算法相结合,来学习用户和项目的潜在表示,可以充分挖掘数据的有效信息。由此,一些研究已经开始在传统协同过滤方法的基础上利用深度学习来进行推荐任务。Wang等^[5]提出协同深度学习(Collaborative Deep Learning, CDL)算法结合用户交互数据与项目内容辅助信息来学习隐式表示,缓解了协同过滤类方法学习用户与项目的交互往往过于稀疏导致推荐效果不好的问题。He等^[6]用神经网络来代替内积形式来学习用户和项目之间的交互,从而实现了矩阵分解类方法的扩展。Barkan等^[7]使用表示学习的方法,将每个项目表示为一个固定维度的向量,从而隐式地得到项目间的相互关系,使得在计算项目相似性方面更加高效,进而提升了模型的泛化能力。Zheng等^[8]引入光谱图理论结合在用户-项目二分图上进行光谱卷积操作来提升协同过滤算法的推荐效果。

尽管上述方法在推荐的准确性上取得了良好的性能,但它忽略了用户不同历史交互项目的重要性,并且缺乏对历史项目与目标项目之间交互关联度的区分,导致无法实现用户偏好的细粒度建模。直观地说,对于每个用户,他对所交互的一组历史项目中不同项目的偏好应该是不同的。例如,用户分别购买了衣服、手机和相机三种产品,并不意味着用户对这三种产品有着相同的偏好。同样,假定用户交互过的所有历史项目对目标项目的预测都有同等的贡献也是不现实的。通常,用户以前与几个项目交互过,这将直接影响用户对某个项目的决策。例如,当用户决定是否购买手机套时,影响最大的应该是先前购买的手机,而不是相机或服装产品。于是,引入深度学习中的注意力机制^[9],注意力机制已在众多领域中取得了成功的应用,其核心思想便是从原始信息中捕获与当前任务最相关的部分并进行强化。因此,通过注意力机制可以为较重要的项目和项目交互关系分配更大的权重来解决上述提到的两个问题,进而提升模型的推荐效果。

本文为了解决协同过滤类算法将所有项目看作同等重要的限制,利用注意力机制可有效区分不同项目重要性的特性,将项目因子相似模型(Factored Item Similarity Model, FISM)^[10]与注意力网络进行深度融合,提出了一种基于双重最相关注意力网络的协同过滤(Dual Most Relevant Attention Collaborative Filtering, DMRACF)算法。该算法利用深度神经网络来自动学习项目之间的复杂交互函数,并设计一个由项目级注意力和项目交互级注意力组成的双重最相关注意力

网络,为较重要的项目和项目交互关系分配更大的权重,进而捕获用户更细粒度下的偏好,实现对用户偏好的细粒度建模,这有利于推荐效果的提升。

1 相关工作

基于项目的协同过滤以其优异的性能在推荐系统的构建中得到了广泛的应用。早期的模型,如基于项目的K近邻(Item-based K-Nearest Neighbors, ItemKNN)^[2]只是简单地使用统计方法来估计项目之间的相似度,如余弦相似度和皮尔逊相关系数。但是,这些方法需要针对特定数据集进行手动调优,导致算法可扩展性方面受限。为了弥补这些方法的不足,出现了一些基于机器学习的方法,通过构造目标函数来自动学习项目之间的相似关系。在这些基于机器学习的方法中,最具代表性的模型是稀疏线性方法(Sparse Linear Method, SLIM)^[11]和FISM。具体来说,SLIM构造了一个基于回归的目标函数,并对其进行优化以学习项目相似度矩阵;但其训练成本较高,不可能挖掘项目之间的传递关系。与SLIM不同,FISM将两个项目之间的相似性表示为它们的低维向量的内积。尽管FISM实现了最先进的推荐性能,但它有一个限制,即假设与用户交互的所有历史项目在建模用户对目标项目的偏好时有同等的贡献。

随着深度学习技术在计算机视觉和自然语言处理方面取得了巨大的成功^[12],人们正在努力将深度学习技术引入到推荐系统中。通常,深度学习被应用于特征提取和预测。Xue等^[13]利用深度神经网络来捕获项目间的高阶交互关系,提出了一个基于项目的协同过滤深度学习算法(Deep Item-based Collaborative Filtering, DeepICF)。邓凯等^[14]将物品相似性因子模型与深度神经网络相结合,提出了一个基于物品的同一推荐模型(Item-based Unified Collaborative Filtering, UICF)。最近深度学习中的注意机制的有效性在机器翻译中被证明是非常有效的^[15],它的成功主要是基于一个合理的假设,即人类的识别并不倾向于一次性处理整个信号,相反,人们只关注整个感知空间中有选择性的部分。事实上,这一假设不仅适用于计算机视觉和自然语言处理领域,在许多现实世界的情况下都是合理的。于是基于注意力机制的推荐模型近年来得到了发展,并且取得了突出的表现。He等^[16]提出了一个神经注意力项目相似性(Neural Attentive Item Similarity, NAIS)模型,该模型在学习项目-项目之间的相似性方面使用注意力机制进行建模。Chen等^[17]将隐因子模型与组件级注意力机制相结合对多媒体推荐中的隐式反馈建模,效果良好。Zhuang等^[18]提出了一个注意力驱动的推荐算法,其使用注意力机制来估计用户对不同项目特征的注意力分布,提高了模型的可解释性。

2 基于双重最相关注意力网络算法

首先, 本文提出的基于双重最相关注意力网络的协同过滤算法框架如图 1 所示。整个模型包括输入层和嵌入层, 对交互层、池化层、深度交互层、输出层以及由项目级注意力和项目交互级注意力组成的双重最相关注意力网络。

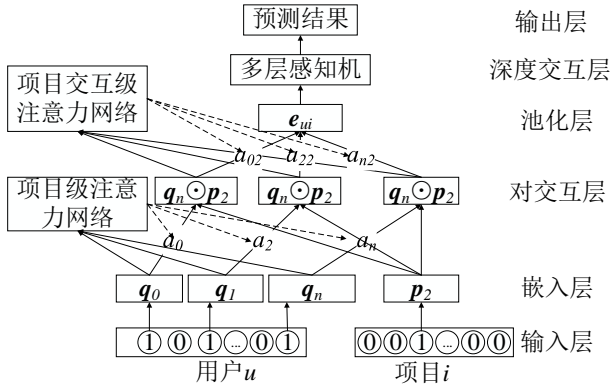


图1 DMRACF 模型框架

Fig. 1 Framework of DMRACF

在输入层中, 目标项目 i 的表示是使用 ID 特征的 one-hot 编码 $\mathbf{x}_{\text{item_input}} = [0, 0, 1, \dots, 0, 0]^m$, 用户 u 的表示是对 u 的交互项目集 R_u^+ 的 ID 特征进行 multi-hot 编码 $\mathbf{x}_{\text{user_input}} = [1, 0, 1, \dots, 0, 1]^m$ 。其中, m 表示编码的维数, 其大小为项目的总个数。由于此编码方式产生的特征空间非常大, 容易造成维度过大。因此, 在输入层之上引入一个嵌入层, 它是一个全连接神经网络用于将输入层稀疏的特征向量转换成稠密的特征向量, 有效地降低了特征空间的维数。通过公式 (1) 将目标项 i 映射到一个嵌入向量 $\mathbf{p}_i \in \mathbf{A}^k$ 上来表示目标项目, 其中 $k \ll m$ 为嵌入向量的维数。对于用户 u 与之交互的每个历史项 $j \in R_u^+$, 将其映射到一个嵌入向量 $\mathbf{q}_j \in \mathbf{A}^k$ 。最后, 根据嵌入层的输出, 可以得到一组向量 $\mathbf{Q}_u = \{\mathbf{q}_j | j \in R_u^+\}$, 一个向量 \mathbf{p}_i 分别表示用户 u 和目标项 i 。

$$\begin{cases} \mathbf{q}_j = \text{embedded}(\mathbf{x}_{\text{user_input}}) \\ \mathbf{p}_i = \text{embedded}(\mathbf{x}_{\text{item_input}}) \end{cases} \quad (1)$$

其中: $\mathbf{x}_{\text{item_input}}$, $\mathbf{x}_{\text{user_input}}$ 表示项目 and 用户的 one-hot 编码输入; embedded 表示全连接神经网络。

在得到一组表示用户 u 的向量 \mathbf{Q}_u 之后, 为了获得用户对历史交互项目的偏好, 使用一维注意力机制来为不同的历史交互项目分配不同的权重, 计算方法如下:

$$\mathbf{z}_j = \mathbf{h}^T \text{ReLU}(\mathbf{W}(\mathbf{q}_j) + \mathbf{b}) \quad (2)$$

$$\alpha_j = \frac{\exp(\mathbf{z}_j)}{\sum_{j \in R_u^+} \exp(\mathbf{z}_j)} \quad (3)$$

$$\mathbf{q}_j = \alpha_j \bullet \mathbf{q}_j \quad (4)$$

其中: \mathbf{W} 和 \mathbf{b} 分别是将输入映射到隐层的权值矩阵和偏置向量; \mathbf{h}^T 是将隐层映射到输出注意力权值的映射向量; α_j 可抽象地看作是用户 u 对其交互过的历史项目 j 的偏好。

为了捕获任意一个历史项目和目标项目之间的特征交互, 通过对用户 u 的历史交互项与目标项 i 之间的嵌入向量进行简单的元素乘积运算来获取项目对交互向量集 $\mathbf{V}_{ui} = \{\mathbf{q}_j \odot \mathbf{p}_i | j \in R_u^+ \setminus i\}$, 获取用户 u 的历史交互项目与目标项目 i 的二阶特征交互关系。

直观地说, 不同的历史项目对目标项目的预测有不同的贡献。因此, 将这组两两交互的向量输入到一个注意力网络中。通过学习, 对较重要的项目对给予较大的交互权重, 以感知历史项目与目标项目交互关系的不同重要性, 计算方法如下:

$$\mathbf{z}_{ij} = \mathbf{h}^T \text{ReLU}(\mathbf{W}(\mathbf{p}_i \odot \mathbf{q}_j) + \mathbf{b}) \quad (5)$$

$$\alpha_{ij} = \frac{\exp(\mathbf{z}_{ij})}{\left[\sum_{j \in R_u^+} \exp(\mathbf{z}_{ij}) \right]^\beta} \quad (6)$$

$$\mathbf{V}_{ui} = \{\alpha_{ij} \bullet (\mathbf{q}_j \odot \mathbf{p}_i)\} \quad (7)$$

其中, \mathbf{h}^T 为将隐层映射到输出注意力权值的映射向量。由于不同用户的历史交互项目数量变化很大, 使用一个超参数 β 用于平滑 softmax 函数中的分母项, 其取值范围在 [0, 1], 从而平衡了一些活跃用户交互的历史项目数过多导致注意力权值过小的问题, 同时使得注意力权值能在一个比较小的方差内。 α_{ij} 为历史项目 j 对目标项目 i 的注意力权值, 可视为用户 u 对项目 i 兴趣的抽象表示。

由于用户的历史交互项目的数量(用户的历史评分项的数量)变化很大, 对特征交互层的输出是一个可变大小的向量集。为了便于后续处理, 使用池化层对可变大小的向量进行操作, 生成一个固定大小的向量。这里, 为了表示不同的池化方式, 使用了一个超参数来控制, 计算方法如下:

$$f_{\text{pooling}}(\mathbf{V}_{ui}) = \frac{1}{|\mathbf{V}_{ui}|^\gamma} \sum_{\mathbf{v} \in \mathbf{V}_{ui}} \mathbf{v} \quad (8)$$

其中: γ 是一个控制池化方式的超参数, 其取值范围在 [0, 1]。当 γ 设置为 0 时, 就变成标准的和值池化; 当 γ 设置为 1 时, 就变成了标准的平均池化。

前一个池化层的输出是一个 k 维的向量 \mathbf{e}_{ui} ($\mathbf{e}_{ui} = f_{\text{pooling}}(\mathbf{V}_{ui})$), 其中包含历史项目和目标项目之间的二阶交互。为了对历史项目和目标项目之间的高阶特征交互进行建模, 在其上叠加了一个多层感知器 (MultiLayer

Perceptron, MLP), 实现了高阶建模, 获得了更深层次的复杂交互信息, 计算方法如下:

$$\begin{cases} \mathbf{e}_1 = \text{ReLU}(\mathbf{W}_1 \mathbf{e}_{ui} + \mathbf{b}_1) \\ \mathbf{e}_2 = \text{ReLU}(\mathbf{W}_2 \mathbf{e}_1 + \mathbf{b}_2) \\ \dots \\ \mathbf{e}_L = \text{ReLU}(\mathbf{W}_L \mathbf{e}_{L-1} + \mathbf{b}_L) \end{cases} \quad (9)$$

其中: \mathbf{W}_L 、 \mathbf{b}_L 、 \mathbf{e}_L 分别表示第 L 层的权重矩阵、偏置向量、激活函数和输出向量。使用 Relu 作为激活函数, 当网络变深时, 更能抵抗饱和问题^[19], 并在本文的实验中显示出良好的性能。

作为深度交互层的输出, 所得到的深度交互向量 \mathbf{e}_L 包含了项目之间从二阶到高阶之间的特征交互信息, 然后用简单的线性回归模型预测最终结果:

$$\hat{y}_{u,i} = \mathbf{h}^T \mathbf{e}_L + \mathbf{b} \quad (10)$$

其中 \mathbf{h}^T 和 \mathbf{b} 分别表示权值矩阵和偏置向量。

为了学习推荐模型, 指定了一个目标函数进行优化。当前主要存在的两种方法, 分别是平方损失和对数损失, 我们认为, 平方损失可能与隐式反馈数据不太吻合。这是因为对于隐式反馈数据, 目标值 y_{ui} 是一个二值化的 1 或 0, 表示 u 是否与 i 交互。可以将 y_{ui} 的值看作一个标签 1 表示项目 i 与 u 相关, 0 表示项目 i 与 u 不相关。预测分数 \hat{y}_{ui} 表示 i 与 u 相关的可能性有多大。要赋予这样的概率解释, 需要将输出 \hat{y}_{ui} 限制在 [0,1], 这可以很容易通过概率函数来实现。

$$L_{sqr} = \sum_{(u,i) \in N} (y_{ui} - \hat{y}_{ui})^2 \quad (11)$$

通过以上的说明可知, 当处理问题是二值数据 1 或 0 的隐式反馈时, 学习推荐模型可以看作是一个二分类任务, 因此使用式 (12) 所示的对数损失作为目标函数。通过最小化目标函数学习模型参数:

$$L = -\frac{1}{N} \left[\sum_{(u,i) \in R^+} \ln \sigma(\hat{y}_{ui}) + \sum_{(u,i) \in R^-} \ln(1 - \sigma(\hat{y}_{ui})) \right] + \lambda \|\Theta\|^2 \quad (12)$$

其中: N 表示整个训练集的总数, 包括正样本 R^+ 和负样本 R^- ; σ 表示 sigmoid 函数, 它将预测值约束在 [0,1] 范围内; \hat{y}_{ui} 为预测结果, 其值表示用户 u 与项目 i 交互的可能性; λ 为防止过拟合的正则化参数; Θ 为整个模型的可训练参数。

综上所述, 得到算法 1 所示的 DMRACF 算法。

算法 1 DMRACF。

使用高斯分布初始化模型参数 Θ 。

foreach $i, j \in R$ do

由式(1)计算用户向量组表示 \mathbf{Q}_u 和项目项目表示 \mathbf{q}_j ;

end for

foreach $p_i \in \mathbf{Q}_u$ do

由式(2)、(3)、(4)计算用户 u 的项目级注意力表示;

end for

foreach $i \in R$ do

foreach $j \in R$ do

由式(5)、(6)计算用户 u 对项目 i 的注意力权值 a_{ij} ;

由式(7)、(8)、(9)计算用户 u 对项目 i 的偏好表示;

由式(10)计算最终预测值 $\hat{y}_{u,i}$;

end for

根据 $\hat{y}_{u,i}$ 选取 Top- N 个项目推荐给用户;

end for

foreach $\theta \in \{U, \Theta\}$

$$\text{更新 } \theta \leftarrow \theta + \left(\frac{\exp^{-\hat{y}_{ui}}}{1 + \exp^{-\hat{y}_{ui}}} \cdot \frac{\partial \hat{y}_{ui}}{\partial \theta} + \lambda \theta \right);$$

直到模型收敛;

返回 U, Θ ;

end for

算法中, R 表示项目集, U 表示项目级注意力和项目交互级注意力的参数集。

3 实验与结果分析

3.1 数据集描述与评价指标

为了验证本文提出的方法的性能, 在两个真实数据集上进行了评估。MovieLens 是一个被广泛用于研究协同过滤算法性能的电影评分数据集^[20]。在本文实验中, 选择的是 MovieLens-1M, 包含 100 万个评分, 其中每个用户至少有 20 个评分。Pinterest 是一个用于图像推荐的数据集, 其中包含 9916 名用户对 55187 张图片的评分。数据集详细信息如表 1 所示。

表1 实验数据集的描述

Tab. 1 The description of dataset in experiment

数据集	交互数	用户数	项目数	稀疏度
MovieLens-1M	1 000 209	3 706	6 040	4.47%
Pinterest	1 500 809	9 916	55 187	0.27%

本文中使用一种被广泛使用的 leave-one-out 评价方法来研究项目推荐的性能^[21]。首先, 根据每个用户的时间戳对用户-项目交互进行排序。然后, 将最新的交互作为每个用户的测试数据, 并利用与用户对应的其余交互进行训练。对每个测试项目(正样本)随机抽取 99 个没有被相应用户交互的项目(负样本)进行抽样, 以便在这 100 个项目构成的集合进行模型的评估。因此, 可以减轻在评估期间为每个用户对所有项目进行排序的耗时问题。

在评价指标方面,采用命中率(Hit Ratio@k, HR@k)和归一化折损累积增益(Normalized Discounted Cumulative Gain@k, NDCG@k)来评估本文的模型生成的排名列表的性能^[22]。在实验部分,设置两个指标的 $k=10$ 。如果测试项目出现在排名前10的列表中,那么HR@10的度量就能够直观地进行度量,并且NDCG@10说明了排名的质量,即在为命中位置靠前的分配更高的分数。两个指标越高,推荐性能越好。

3.2 对比算法

为了评估本文提出的模型的有效性,还研究了以下几种协同过滤类方法的性能。

1)ItemPop(Item Popularity)。该算法是一个非个性化的推荐算法,因为它根据商品的受欢迎程度(通过交互次数的数量来衡量)来对商品进行排名。

2)ItemKNN。该方法是最基本的基于项目的协同过滤,通过历史项目与目标项目之间的相似性来给出推荐,在实验中采用余弦相似度来计算项目之间的相似性。

3)BPR(Bayesian Personalized Ranking)^[23]。该方法利用贝叶斯个性化排序损失来优化MF模型,从而实现了MF模型的优化。

4)eALS(element-wise Alternating Least Square)^[20]。该方法也学习了一个MF模型,但优化了一个不同的点态回归损失,该损失将所有缺失的数据视为具有较小权重的负反馈。

5)MLP^[6]。该方法利用深度神经网络代替简单的内积,从数据中学习用户与项目之间的非线性交互。

6)FISM。该方法是当前最先进的基于项目的协同过滤算法,其为每个项目都生成两个向量表示,分别用来表示历史交互项目和目标项目。

7)NAIS。该方法利用注意力机制来学习物品相似度,进而对FISM进行扩展。

8)DeepICF。该方法使用深度神经网络来同时捕获项目间二阶和高阶交互关系,是一种基于深度学习的项目协同过滤方法。

3.3 参数设置

用高斯分布随机初始化模型参数,其中均值和标准差分别为0和0.01。对于嵌入大小 k ,在实验中对[8,16,32,64]的值进行了评估。注意力因子尺寸与每次的嵌入尺寸相同。超

参数 β 和 γ 的取值范围在[0,1],在实验中 β 和 γ 的取值分别为0.5和0。防止模型过拟合的L2正则化系数 λ 在 $[1E^{-5},1E^{-4},1E^{-3},0]$ 之间进行调整。使用优化器Adagrad对目标函数进行优化,从而完成对整个模型参数的训练。学习率设置为0.01,迭代次数为80次。整个模型的实现是基于TensorFlow,所有的实验都是在NVIDIA的Tesla GPU上进行。

3.4 结果分析

由于嵌入尺寸控制基于嵌入的方法的建模能力,为基于嵌入的方法(BPR、eALS、MLP、FISM)设置了相同的嵌入尺寸16进行实验结果的比较。表2展示了不同推荐算法在两个数据集上的准确性。观察表2可以发现,与其他方法相比,DMRACF在两个数据集上都取得了最好的性能。NAIS和DMRACF均在模型中使用了注意力机制,MovieLens数据集上NAIS相较于当前最先进的算法FISM在两种评价指标下分别提高了4.8个百分点、6.2个百分点,而本文提出的DMRACF提升的结果则是6个百分点和8.7个百分点。这表明引入注意机制可以帮助模型进一步提取数据中更为复杂的特征信息,进而提高模型的表达性和准确率。其中,DMRACF的准确率最高,说明设计的双重最相关注意力网络可以有效地捕捉到更细粒度的用户偏好,使得本文的模型具有更强大的表示能力。两种基于项目的协同过滤方法FISM和ItemKNN使用相同的预测模型,但FISM的预测结果远远超过了ItemKNN,在MovieLens上的HR和NDCG分别提高了6.7个百分点和10.1个百分点。两种方法的关键区别在于项目相似度的估计方法不同,导致FISM方法的结果优于ItemKNN方法。基于用户的协同过滤模型(BPR、eALS和MLP)与基于项目的协同过滤模型(FISM)在不同数据集上具有不同的表现。具体来说,基于用户的协同过滤模型在MovieLens数据集上的性能优于FISM,而FISM在Pinterest数据集上的性能优于基于用户的协同过滤模型。观察表1可知,Pinterest数据集具有较高的稀疏性(相对于MovieLens数据集),本文认为基于项目的协同过滤方法在高度稀疏的数据集上比基于用户的协同过滤方法表现得更好。

表2 不同方法在嵌入尺寸为16时的推荐准确度

Tab. 2 Recommendation accuracy of different methods at embedding size 16

方法	MovieLens		Pinterest	
	HR@10	NDCG@10	HR@10	NDCG@10
ItemPop	0.4536	0.2543	0.2739	0.1409

ItemKNN	0.6227	0.3587	0.7857	0.4832
BPR	0.6674	0.3907	0.8628	0.5406
eALS	0.6689	0.3977	0.8755	0.5449
MF-BPR	0.6664	0.3973	0.8691	0.5401
MLP	0.6841	0.4103	0.8648	0.5385
FISM	0.6647	0.3949	0.8663	0.5522
NAIS	0.6969	0.4194	0.8764	0.5612
DeepICF	0.6881	0.4113	0.8706	0.5608
DMRACF	0.7046	0.4285	0.8835	0.5711

双重最相关注意力网络的设计是本文模型的核心部分, 为了验证不同层级注意力模块对本文模型的影响。将 DMRACF 与它的两个变体进行比较, 即 DMRACF_IT(只具有项目级注意力)和 DMRACF_II(只具有项目交互级注意力)。从表 3 可以看出, 与 DMRACF_IT 相比, DMRACF_II 具有更好的性能。据统计, DMRACF_II 相对于 DMRACF_IT

在 MovieLens 上的 HR 和 NDCG 分别提高了 1.4 个百分点和 3.5 个百分点。这表明项目交互级的注意力比项目级的注意力对本文的模型贡献更大。另外, DMRACF 借助双重最相关注意力网络, 最终取得了最优的效果(与 DMRACF_IT、DMRACF_II 相比)。这进一步验证了注意力机制的引入和双重最相关注意力网络的设计对模型的有效性。

表3 DMRACF 和它两个变体的实验结果比较

Tab. 3 Comparison of experimental results between DMRACF and its two variants

方法	MovieLens		Pinterest	
	HR@10	NDCG@10	HR@10	NDCG@10
DMRACF_IT	0.6891	0.4133	0.8806	0.5661
DMRACF_II	0.6989	0.4250	0.8813	0.5682
DMRACF	0.7046	0.4285	0.8835	0.5711

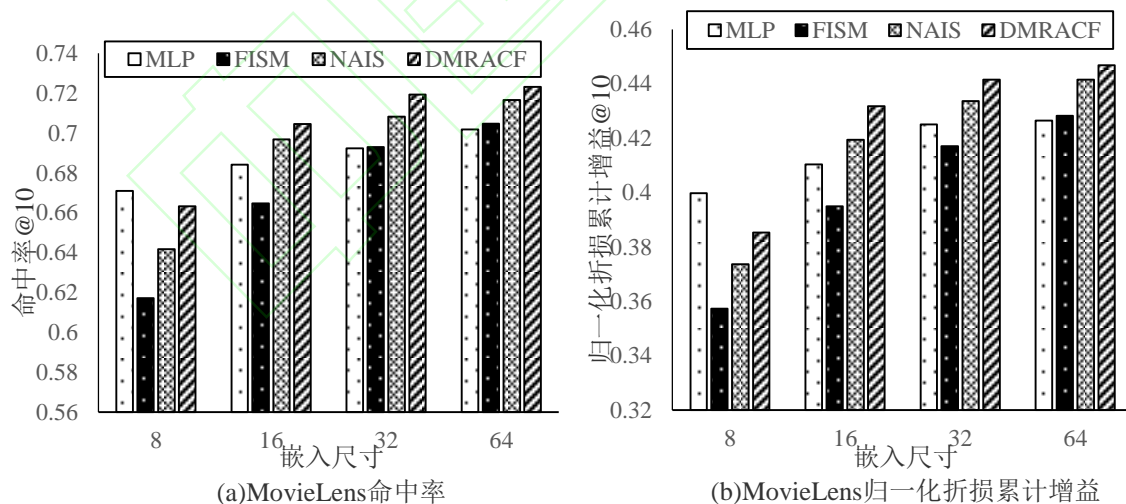


图2 MovieLens 数据集上基于嵌入的方法在不同嵌入尺寸上的性能比较

Fig. 2 Performance comparison of embedding-based methods on MovieLens dataset at different embedding sizes

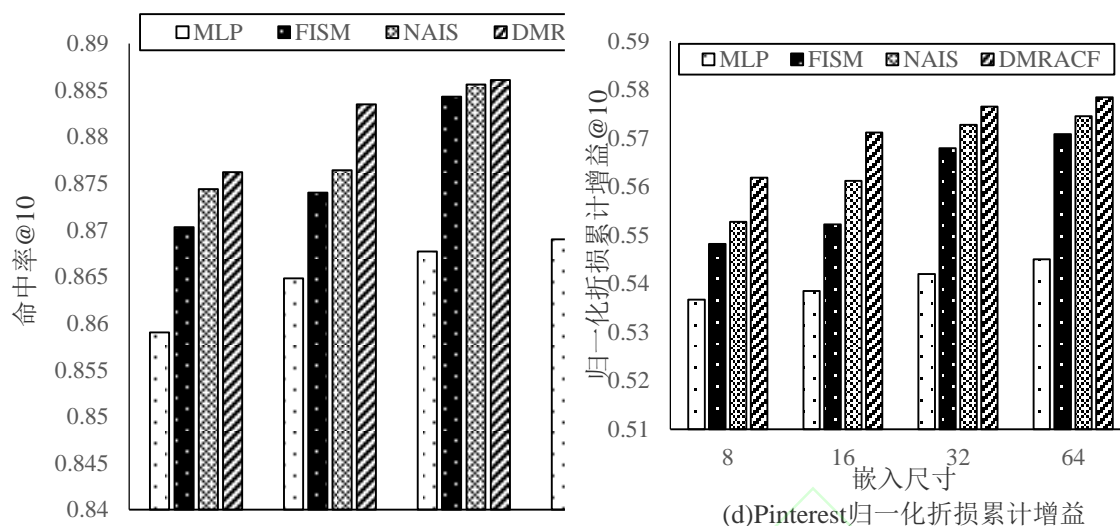


图3 Pinterest 数据集上基于嵌入的方法在不同嵌入尺寸上的性能比较

Fig. 3 Performance comparison of embedding-based methods on Pinterest dataset at different embedding sizes

图2和图3显示了在两个数据集上HR和NDCG对于不同嵌入尺寸的性能表现。从图2和图3中可以看出,一般情况下,嵌入尺寸为8、32和64的推荐性能趋势与嵌入尺寸为16的推荐性能趋势相似。本文提出的DMRACF方法在大多数情况下都取得了最优的表现,除了嵌入尺寸为8时,MLP在MovieLens上的性能优于DMRACF。在MovieLens这个相对密集的数据集上(与Pinterest相比),基于用户的非线性方法(在本例中为MLP)能够在较小的嵌入尺寸下取得更强的表示能力。

4 结语

本文提出了一种新的基于双重最相关注意力网络的协同过滤算法,称为DMRACF,用于Top-N推荐。DMRACF通过设计一个双重最相关注意力网络,不仅强化了与用户最相关的历史项目,而且还挑选出了对预测最重要的历史项与目标项之间的交互关系,进而解决了协同过滤类算法将所有历史项目看作同等重要的限制,提高了推荐的准确性。实验表明,DMRACF的性能优于其他协同过滤类方法。未来,计划将DMRACF从以下方向进行扩展。首先,这项工作的重点是纯协同过滤系统,只使用了用户和项目的ID属性进行建模,但实际上存在许多用户和项目的其他可用属性(如:类别、职业)。未来将研究DMRACF在使用这些可用辅助信息时的有效性。

参考文献

- [1] 黄立威, 江碧海, 吕守业, 等. 基于深度学习的推荐系统研究综述[J]. 计算机学报, 2018, 41(7): 1619-1648. (HUANG L W, JIANG B T, LV S Y, et al. Survey on deep learning based recommender systems[J]. Journal of Computers, 2018, 41(7): 1619-1648.)
- [2] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C]//Proceedings of the 10th ACM International Conference on World Wide Web. New York: ACM, 2001: 285-295.
- [3] YUAN F J, KARATZOGLOU A, ARAPAKIS L, et al. A simple convolutional generative network for next item recommendation[C]//Proceedings of the 12th ACM International Conference on Web Search and Data Mining. New York: ACM, 2019: 582-590.
- [4] KOREN Y, BELL R M, VOLINSKY C. Matrix factorization techniques for recommender systems [J]. IEEE Computer, 2009, 42(8): 30-37.
- [5] WANG H, WANG N, YEUNG D Y. Collaborative deep learning for recommender systems[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2015: 1235-1244.
- [6] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering[C]//Proceedings of the 2017 International Conference on World Wide Web. New York: ACM, 2017: 173-182.
- [7] BARKAN O, NOAM K. Item2vec: Neural item embedding for collaborative filtering[C]//Proceedings of the 26th IEEE International Workshop on Machine Learning for Signal Processing. Piscataway: IEEE, 2016: 1-6.
- [8] ZHENG L, LU C, JIANG F, et al. Spectral collaborative filtering [C]//Proceedings of the 12th ACM International Conference on Recommender System. New York: ACM, 2018: 311-319.
- [9] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 30th Annual Conference on Neural Information Processing Systems. New York: ACM, 2017: 5998-6008.
- [10] KABBUR S, NING X, KARYPIS G. FISM: factored item similarity models for top-n recommender systems [C]//Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2013: 659-667.
- [11] NING X, KARYPIS G. Slim: Sparse linear methods for top-n recommender systems[C]//Proceedings of the 11th IEEE International Conference on Data Mining. Piscataway: IEEE, 2011: 497-506.
- [12] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [13] XUE F, HE X N, WANG X, et al. Deep item-based collaborative filtering for top-n recommendation[J]. ACM Transactions on Information Systems, 2019, 37: 247-254.
- [14] 邓凯, 黄佳进, 秦进, 等. 基于物品的统一推荐模型[J]. 计算机应用, 2020, 40(2): 530-534. (DENG K, HUANG J J, QIN J, et al. Item-based unified recommendation model[J]. Journal of Computer Applications, 2020, 40(2): 530-534.)
- [15] LUONG M T, PHAM H, MANNING C D. Effective approaches to

- attention-based neural machine translation[C]//Proceedings of the 20th ACL International Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2015: 1412–1421.
- [16] HE X, HE Z, SONG J, et al. NAIS: neural attentive item similarity model for recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30 (12): 2354–2366.
- [17] CHEN J, ZHANG H, HE X, et al. Attentive collaborative filtering: Multimedia recommendation with item- and component-level attention [C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2017: 335–344.
- [18] CHEN J, ZHUANG F, HONG X, et al. Attention-driven factor model for explainable personalized recommendation[C]//Proceedings of the 41th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2018: 909–912.
- [19] GULCEHRE C, CHO K, PASCANU R, et al. Learned-norm pooling for deep feed forward and recurrent neural networks[C]//Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases. New York: ACM, 2014 : 530–546.
- [20] HARPER F M, KONSTAN J A. The movie lens datasets: History and context[J]. ACM Transactions on Interactive Intelligent Systems, 2016, 5(4): 19.
- [21] FARSEEV A, SAMBORSKII I, FILCHENKOV A, et al. Cross-domain recommendation via clustering on multi-layer graphs[C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2017: 195–204.
- [22] DENG H Z, HUANG L, WANG C D, et al. DeepCF: a unified framework of representation learning and matching function learning in recommender system [C]//Proceedings of the 31th AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2019 : 61–68.
- [23] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback [C] //Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. New York: ACM, 2009 : 452–461.

This work is partially supported by the National Natural Science Foundation of China (61702003), the Natural Science Foundation of Anhui Province of China (1808085MF175).

ZHANG Wenlong, born in 1996, M. S. candidate. His research interests include deep learning, recommendation system.

QIAN Fulan, born in 1978, Ph. D., associate professor. Her research interests include granular computing, social network, recommendation system.

CHEN Jie, born in 1982, Ph. D., associate professor. Her research interests include intelligent computing, machine learning, three-way decision.

ZHAO Shu, born in 1979, Ph. D., professor. Her research interests include granular computing, quotient space theory, machine learning.

ZHANG Yanping, born in 1962, Ph. D., professor. Her research interests include intelligent computing, granular computing, quotient space theory.