

## 融合语义相似度的协同过滤推荐算法

王根生<sup>1,2</sup>, 潘方正<sup>2</sup>

(1. 江西财经大学计算机实践教学中心, 江西南昌 330013; 2. 江西财经大学人文学院, 江西南昌 330013)

**摘要:** 针对协同过滤推荐算法没有考虑推荐对象间语义关系的问题, 提出一种融合推荐对象语义相似度的改进型协同过滤推荐算法. 首先利用知识图谱表示学习算法将推荐对象的语义信息嵌入到一个低维语义空间; 然后计算推荐对象之间的语义相似度, 将该语义相似度融合到协同过滤推荐算法的相似度计算中, 弥补协同过滤推荐算法没有考虑推荐对象自身语义知识的缺陷. 实验结果表明, 该改进型算法相比传统协同过滤推荐算法, 具有更高的准确率、召回率和覆盖率.

**关键词:** 推荐算法; 协同过滤; 知识图谱; 表示学习; 语义相似度

**中图分类号:** TP391      **文献标识码:** A      doi: 10.3969/j.issn.0253-2778.2019.10.009

**引用格式:** 王根生, 潘方正. 融合语义相似度的协同过滤推荐算法[J]. 中国科学技术大学学报, 2019, 49(10): 835-841.

WANG Gensheng, PAN Fangzheng. Collaborative filtering recommendation algorithm based on semantic similarity[J]. Journal of University of Science and Technology of China, 2019, 49(10): 835-841.

## Collaborative filtering recommendation algorithm based on semantic similarity

WANG Gensheng, PAN Fangzheng

(1. Computer Practice Teaching Center, Jiangxi University of Finance and Economics, Nanchang 330013, China;

2. School of Humanities, Jiangxi University of Finance and Economic, Nanchang 330013, China)

**Abstract:** To solve the problem that collaborative filtering recommendation algorithm does not consider the semantic relationship between recommendation objects, an improved collaborative filtering recommendation algorithm based on semantic similarity of recommendation objects is proposed. First, the semantic information of the recommended object is embedded into a low dimensional semantic space by using the knowledge map representation learning algorithm; then the semantic similarity between the recommended objects is calculated and integrated into the similarity calculation of collaborative filtering recommendation algorithm, thus compensating for the shortcoming that the collaborative filtering recommendation algorithm does not consider the semantic knowledge of the recommendation object. The experimental results show that the improved algorithm has higher accuracy, recall and coverage than the traditional collaborative filtering recommendation algorithm.

**Key words:** recommendation algorithm; collaborative filtering; knowledge graph; representation learning; semantic similarity

收稿日期: 2019-05-15; 修回日期: 2019-09-28

基金项目: 国家自然科学基金(71461012), 江西省教育厅科技资助项目(GJJ181550), 江西省高校人文社科项目(GL19110), 深圳市哲学社会科学规划课题(SZ2019D050)资助.

作者简介: 王根生(通讯作者), 1974年生, 男, 博士/副教授. 研究方向: 数据挖掘. E-mail: wgs74@126.com

## 0 引言

随着互联网大数据的快速发展,用户面临的网络信息呈现指数级增长,推荐系统是解决信息过载问题的典型技术<sup>[1]</sup>,在电子商务、网络媒体、新闻广告、网络学习等领域得到了广泛应用.推荐算法根据不同的推荐引擎主要分为三类:基于内容过滤的推荐、协同过滤推荐和混合推荐<sup>[2]</sup>,协同过滤是目前应用最为广泛的一种推荐算法<sup>[3]</sup>.

协同过滤推荐算法基于用户-物品评分矩阵,依据评分矩阵计算用户或物品的相似度,进而实现物品推荐<sup>[4]</sup>.协同过滤推荐算法无需构建用户或物品的特征表示,没有领域限制,推荐结果具有较好的多样性和新颖性,缺点是没有考虑物品本身内涵特性或用户关系,导致算法的准确率不高.针对这个问题有学者提出了相关改进方案.如文献<sup>[5]</sup>提出一种基于物品内容和协同过滤混合的推荐算法,该算法首先基于物品内容过滤推荐算法发现用户既有兴趣,再利用协同过滤算法挖掘用户潜在兴趣,然后把用户既有兴趣和潜在兴趣进行融合,得到准确率较高的新闻推荐;文献<sup>[6]</sup>把社会关系信息融合到协同过滤矩阵分解推荐算法中,提高了算法的准确性;文献<sup>[7]</sup>提出一种融合用户评分、情感倾向和物品内容信息的混合推荐算法,从直接反馈、隐式反馈、辅助信息等多维度进行相似度计算,改善了推荐算法的性能.

研究发现,从多维度进行相似度计算,是目前对协同过滤推荐算法改进的主要思路.随着知识图谱技术的发展,目前业界已经有大量开放的语义知识数据,如通用知识图谱 Freebase、OpenKN、DBpedia 等,特定领域知识图谱 HerbNet(中医领域)、WolframAlpha(数学领域)、BMKG(影视领域)<sup>[8]</sup>等.通过知识图谱表示学习算法可以将推荐对象所处领域的语义数据嵌入到一个低维语义空间.文献<sup>[9]</sup>提出一种基于知识图谱表示学习的协同过滤推荐算法,该算法将协同过滤推荐算法得到的邻近集合替换成知识图谱表示学习得到的语义邻近集合.本文在文献<sup>[9]</sup>的基础上进行改进,将语义相似度直接融合到协同过滤推荐算法的相似度计算中,从语义维度弥补协同过滤推荐算法没有考虑物品本身内涵特性的不足.

## 1 相关理论

### 1.1 协同过滤推荐算法

协同过滤推荐算法<sup>[10]</sup>利用用户群体对物品的评分数据得出物品或用户的相似性,主要分为基于用户的协同过滤(user-based CF)<sup>[11]</sup>和基于项目的协同过滤(item-based CF)<sup>[12]</sup>.假设有用户集合  $U = (u_1, u_2, \dots, u_m)$ ,  $m$  为用户个数,物品集合  $I = (i_1, i_2, \dots, i_n)$ ,  $n$  为物品个数,构建  $m \times n$  的用户-物品评分矩阵,如图 1 所示.

用户 \ 物品	$i_1$	$i_2$	$\dots$	$i_{n-1}$	$i_n$
$u_1$	$r_{11}$	$r_{12}$	$\dots$	$r_{1n-1}$	$r_{1n}$
$u_2$	$r_{21}$	$r_{22}$	$\dots$	$r_{2n-1}$	$r_{2n}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$u_{m-1}$	$r_{m-11}$	$r_{m-12}$	$\dots$	$r_{m-1n-1}$	$r_{m-1n}$
$u_m$	$r_{m1}$	$r_{m2}$	$\dots$	$r_{mn-1}$	$r_{mn}$

图 1 用户-物品评分矩阵

Fig. 1 User-item scoring matrix

$r_{kj}$  为用户  $k$  对物品  $j$  的评分 ( $1 \leq k \leq m, 1 \leq j \leq n$ ), 评分高低反映了用户对该物品的喜好程度.如果两个物品是相似的,那用户对其评价行为也是相似的.基于项目协同过滤的推荐算法就是利用所有用户对物品的评价向量来衡量物品的相似度,从而实现推荐.所有用户对物品  $j$  的评分向量表示为

$$S_j = (r_{1j}, r_{2j}, \dots, r_{mj})^T \quad (1)$$

式中,  $r_{kj}$  为用户  $k$  对物品  $j$  的评分 ( $1 \leq k \leq m$ ),  $m$  为用户数.物品相似度常用的计算方法有余弦相似度、Jaccard 相似度、皮尔逊相似度、对数似然相似度,其中余弦相似度应用最为广泛<sup>[13]</sup>,计算如公式为

$$\text{sim}_{\text{CF}}(i, j) = \cos(S_i, S_j) = \frac{S_i \cdot S_j}{\|S_i\| \cdot \|S_j\|} = \frac{\sum_{k=1}^m r_{ki} \cdot r_{kj}}{\sqrt{\sum_{k=1}^m r_{ki}^2} \cdot \sqrt{\sum_{k=1}^m r_{kj}^2}} \quad (2)$$

式中,  $\text{sim}_{\text{CF}}(i, j)$  为物品  $i$  和  $j$  协同过滤的余弦相似度,  $S_i$  和  $S_j$  分别为用户对物品  $i$  和  $j$  的评分向量.

### 1.2 知识图谱表示学习

2012 年,Google 提出了知识图谱的概念,用于

构建基于语义的智能化搜索引擎. 知识图谱使用“实体-关系-实体”三元组描述现实世界中的实体和实体之间的语义关系, 通过关系构成网状的知识结构<sup>[14]</sup>. 表示学习的目的是把对象嵌入到低维实数向量空间<sup>[15]</sup>, 知识图谱表示学习则针对知识图谱中的实体和关系进行向量空间映射, 得到的向量可以有效表示实体和关系的语义联系<sup>[16]</sup>. 近年来, 以深度学习为代表的表示学习方法受到了广泛关注<sup>[17]</sup>, 知识图谱表示学习常见的模型有距离模型、能量模型、单层神经网络模型、矩阵分解模型、双线性模型、张量神经网络模型和翻译模型等<sup>[18]</sup>. 以 TransE<sup>[19]</sup> 为代表的翻译模型, 参数简单、计算复杂度低, 在大规模知识图谱上性能显著, 是目前主流的学习模型. 对于每个 3 元组  $(h, r, t)$ , 其中  $h, t$  分别表示头实体和尾实体,  $r$  为头尾实体间的关系, TransE 将  $h, t$  和  $r$  分别表示为嵌入向量  $v_h, v_t$  和  $v_r, v_r$  为向量  $v_h$  和  $v_t$  间的平移, 也称为向量  $v_h$  到  $v_t$  的翻译, 三者之间的关系表示为

$$v_h + v_r \approx v_t \quad (3)$$

TransE 模型要使训练得到的向量  $v_h, v_t$  和  $v_r$  使公式(3)无限接近, 其间的误差越小, 说明头尾两个实体间越可能存在关系  $r$ , 所以 TransE 模型的损失函数为

$$f(v_h, v_r, v_t) = \|v_h + v_r - v_t\|_2^2 \quad (4)$$

式中,  $\|\cdot\|_2$  为向量的 2 范数, 即欧几里得距离. 针对所有 3 元组样本的总体代价函数表示为

$$L = \sum_{(h,r,t) \in S} \sum_{(h',r',t') \in S'} \max(0, f(v_h, v_r, v_t) - f(v_{h'}, v_{r'}, v_{t'}) + \gamma) \quad (5)$$

式中,  $S$  为知识图谱中所有 3 元组集合, 称为正样本;  $S'$  为集合  $S$  的负采样, 即对  $S$  中每个存在的 3 元组随机替换掉其头实体或尾实体, 得到一个新的 3 元组, 且该 3 元组不属于  $S$ , 则称这类 3 元组的集合为负样本;  $\gamma$  为正负样本间的距离. 训练时为了使代价函数  $L$  最小化, 就要使得正样本的损失函数  $f(v_h, v_r, v_t)$  趋近于 0, 负样本的损失函数  $f(v_{h'}, v_{r'}, v_{t'})$  趋近无穷大.

## 2 融合语义相似度的协同过滤推荐算法

针对协同过滤推荐算法仅利用用户历史评价数据而没有考虑推荐对象间语义关系的问题, 本文提出一种融合推荐对象语义相似度的改进型协同过滤

推荐算法, 把推荐对象的语义相似度融合直接融合到协同过滤推荐算法的相似度计算中, 从语义视角维度弥补协同过滤推荐算法没有考虑物品本身内涵特性的不足. 算法模型如图 2 所示.

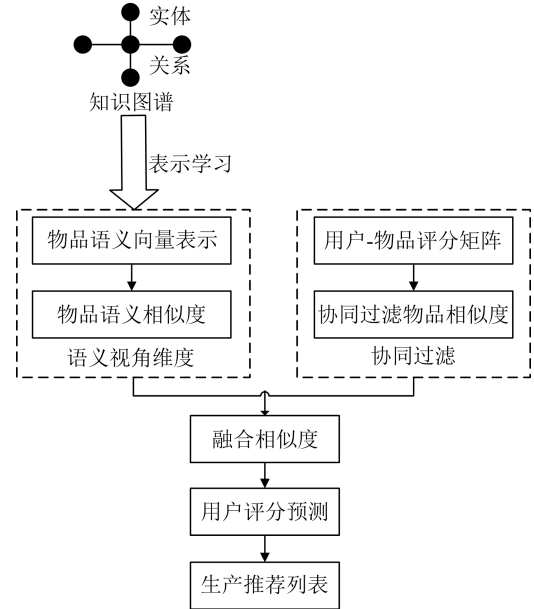


图 2 算法模型

Fig. 2 Algorithm model

### 2.1 物品语义相似度

使用 TransE 算法把知识图谱中的实体和关系嵌入到了一个  $d$  维的语义空间, 物品语义向量表示为

$$I_i = (E_{1i}, E_{2i}, \dots, E_{di})^T \quad (6)$$

式中,  $I_i$  表示物品  $i$  的语义向量,  $E_{ki}$  为语义向量第  $k$  维上的值 ( $1 \leq k \leq d$ ). TransE 算法进行训练时损失函数是基于欧几里得距离, 为了保持一致性, 物品语义的相似度同样采用该距离作为衡量, 计算公式为

$$d(I_i, I_j) = \sqrt{\sum_{k=1}^d (E_{ki} - E_{kj})^2} \quad (7)$$

为了计算统一将其规约到  $(0, 1]$  之间, 规约计算

$$\text{sim}_{\text{semantics}}(i, j) = \frac{1}{1 + d(I_i, I_j)} \quad (8)$$

$\text{sim}_{\text{semantics}}(i, j)$  值越大, 说明物品  $i$  和  $j$  语义越相似.

基于物品语义相似度的推荐其实是一种基于物品内容特征的推荐算法. 如电影, 其主要特征包括电影类型、导演、编剧、主演、制片国家、片长等, 正是这些信息构成了一部电影的语义信息. 在知识图谱中, 这些特征信息组成了“实体-关系-实体”3 元组, 如 3

元组“十面埋伏-导演-张艺谋”、“英雄-导演-张艺谋”,表示电影《十面埋伏》、《英雄》由张艺谋导演,电影间相关联的信息越多,语义信息越相似,知识图谱表示学习算法得出的电影向量也越相近,用户对其的喜爱程度也应该是越相似的,这就是基于物品语义相似度推荐的核心思想。

## 2.2 融合相似度

根据知识图谱表示算法得到物品语义相似度  $\text{sim}_{\text{semantics}}$ ,再把该语义相似度与协同过滤推荐算法中的相似度以线性加权的方式进行融合,融合算法公式表示为

$$\text{sim}(i, j) = \alpha \cdot \text{sim}_{\text{semantics}}(i, j) + (1 - \alpha) \text{sim}_{\text{CF}}(i, j) \quad (9)$$

式中,  $\text{sim}(i, j)$  代表物品  $i$  和  $j$  的融合相似度;  $\text{sim}_{\text{semantics}}$  为物品语义相似度,其计算见公式(8);  $\text{sim}_{\text{CF}}(i, j)$  为协同过滤物品相似度,其计算见公式(2);  $\alpha$  为权重因子,取值范围为  $[0, 1]$ . 当  $\alpha$  为 0 时,本文算法就成了基于项目的协同过滤推荐算法;  $\alpha$  为 1 时为基于语义内容推荐算法;  $\alpha \in (0, 1)$  时为融合推荐算法. 协同过滤推荐算法是间接站在用户对物品评分的角度计算物品相似度,该相似度的准确性一方面依赖于评分数据的准确性,另一方面还依赖于数据的稠密度. 如果数据过于稀疏,该算法的准确性无法保证,这就是协同过滤推荐算法所面临的“矩阵稀疏”问题. 基于语义内容推荐算法是直接站在物品本身特征的角度计算物品相似度,该算法的准确性依赖于物品特征的描述,特征描述越准确、全面,物品相似度的计算也越准确,从而推荐结果也越有实用价值. 融合两种算法的相似度,从间接和直接两个角度进行衡量,可以弥补单个角度的不足。

## 2.3 评分预测与推荐列表生产

根据融合相似度预测用户对一个未评价过物品的评分,预测评分计算公式为

$$p_{ui} = \frac{\sum_{j \in (N(u) \cap S(i, k))} (\text{sim}(i, j) * r_{uj})}{\sum_{j \in (N(u) \cap S(i, k))} \text{sim}(i, j)} \quad (10)$$

式中,  $p_{ui}$  代表用户  $u$  对物品  $i$  的预测评分;  $N(u)$  为用户  $u$  评价过的所有物品集合,  $S(i, k)$  表示与物品  $i$  融合相似度最大的  $k$  个邻近物品,  $N(u) \cap S(i, k)$  表示两个物品集合的交集,该交集结果为预测评分的参考;  $\text{sim}(i, j)$  为物品  $i$  和  $j$  的融合相似度,计算见公式(9);  $r_{uj}$  为用户  $u$  对物品  $j$  的评分。

基于预测评分越高,用户对其越感兴趣的原则,

利用公式(10)计算用户对所有物品的预测评分,然后按评分结果按高低排序,优先选择前  $\text{top-N}$  个物品推荐给用户。

## 3 实验及分析

### 3.1 实验数据

本实验选取电影作为推荐实验对象,实验数据来源于网络爬虫爬取的豆瓣影评数据. 初始数据共包括 94 534 个用户对 8 872 部电影的 404 972 条评价数据,初始评价矩阵数据太过稀疏,基于用户评分的相似度算法无法衡量物品真实情况,所以删除评价数据少于 10 条的电影和发表评论不足 10 条的用户,适当降低评价矩阵的稀疏性,最后保留了 7 815 个用户,1 593 部电影,214 920 条评论. 用户对电影的喜爱程度通过其对电影的星级评价来衡量,星级等级分为 1~5 星,星级越大说明用户对该电影越喜欢. 通过观察分析,本文把评价为 4 和 5 星的认为是用户喜爱的电影,1~3 星的为用户不喜欢的电影。

知识图谱本文选用由清华大学知识工程试验研究室于 2016 年发布的双语影视知识图谱(BMKG),该知识图谱包含 72 万多个与影视相关的实体,91 个属性,1 300 多万条 3 元组;融合了百度百科、豆瓣电影及 LinkedMdb 等多个中英文影视数据. 为了减少知识图谱表示学习算法的训练时间,文本从 BMKG 中只抽取和实验数据相关的知识。

### 3.2 评价指标

基于推荐算法的推荐结果,本文使用准确率(Precision),召回率(Recall),覆盖率(Coverage) 3 个指标进行算法性能衡量,三者的计算分别为

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (11)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (12)$$

$$\text{Coverage} = \frac{N_d}{N} \quad (13)$$

公式(11)、(12)中,TP、FP、FN 为混合矩阵中的值,具体如表 1 所示。

表 1 混合矩阵

Tab. 1 Mixed matrix

推荐算法	用户喜欢	用户不喜欢
推荐	TP	FP
未推荐	FN	TN

公式(13)中,  $N$  为实验中所有电影的个数,  $N_d$

为推荐算法推荐给用户不同电影数目。覆盖率越高,说明算法对冷门物品具有很好的推荐能力,推荐结果具有多样性和新颖性。

为了对算法的性能进行更准确的衡量,本文使用  $k$ -交叉验证的方式进行验证, $k$  值取 5,即随机把试验数据均分成 5 份,每次挑选其中一份作为测试集,其他 4 份作为训练集,一共进行 5 次测试,使用 5 次测试的平均结果值作为算法最终的评价。

### 3.3 实验结果及分析

实验具体步骤如下:

Step1 根据影视知识图谱(BMKG)抽取与实验数据相关的知识;

Step2 使用知识图谱表示学习算法 TransE 对抽取出的知识进行训练,得出电影实体的语义向量表示;

Step3 根据训练数据集构建用户-电影评价矩阵;

Step4 根据电影的语义向量,计算电影间的语义相似度,具体计算见公式(8);

Step5 根据用户-电影评价矩阵,计算协同过滤中基于用户行为的电影相似度,具体计算见公式(2);

Step6 融合 Step4 和 Step5 的计算结果,得出最终电影相似度,具体计算见公式(9);

Step7 根据 Step6 得出的融合相似度,找出每个电影最相近的  $k$  个邻近电影;

Step8 计算测试数据集中每个用户对所有电影的预测评分,具体计算见公式(10);

Step9 根据 Step8 的预测评分进行按高低排序,优先选择前  $N$  部电影推荐给用户,本实验  $N$  取 10;

Step10 统计准确率、召回率、覆盖率 3 个指标;

Step11 改变训练集和测试集,重复 Step3- Step10 的实验过程,一共重复 5 次;

Step12 统计 5 次实验的平均准确率、召回率及覆盖率。

#### (I) 不同知识图谱表示学习嵌入维度对比实验

在进行知识图谱表示学习时,不同的嵌入维度会对实验产生影响,分别选取维度为 50、100、150、200 四组进行实验对比,相似度融合权重因子  $\alpha$  设定为 0.7,相似度最大邻近个数  $k$  设定为 30,最终实验对比结果如图 3 所示。

通过对比发现,4 组不同的嵌入维度中,维度为 100 的准确率、召回率、覆盖率较好。

#### (II) 不同邻近个数 $k$ 值对比实验

选取不同邻近个数  $k$  值也是影响算法性能的一个重要参数,本文分别选取 10、20、30、40、50 作为邻近数进行对比试验,知识图谱表示学习嵌入维度设定为 100,融合权重  $\alpha$  设定为 0.7,实验结果如图 4 所示。

通过实验对比发现,邻近  $k$  为 30 时,算法的准确

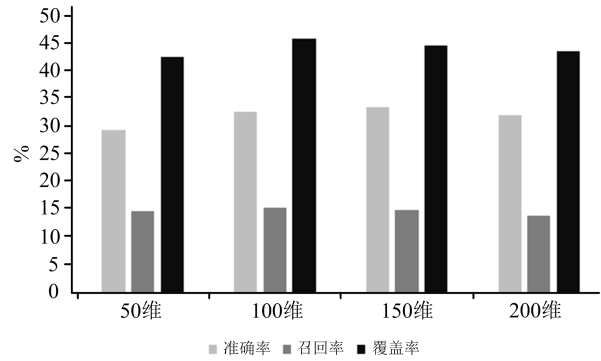


图 3 不同嵌入维度下的实验结果对比

Fig. 3 Comparison of experimental results in different embedding dimensions

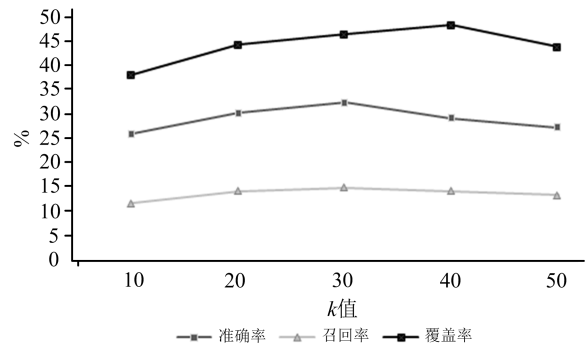


图 4 不同邻近  $k$  下的实验结果对比

Fig. 4 Comparison of experimental results under different  $k$  values

率和召回率达到最佳, $k$  为 40 时覆盖率达到最佳。

#### (III) 不同融合权重因子比实验

融合权重因子  $\alpha$  控制了语义相似度和协同过滤相似度在最终融合相似度中的比例,是本文算法的一个关键因子。 $\alpha$  取值范围为  $[0, 1]$ ,实验从  $\alpha$  取 0 开始,并每次递增 0.1,到 1 结束,一共 11 次。知识图谱表示学习嵌入维度设定为 100,相似度最大邻近个数  $k$  设定为 30,实验结果如图 5 所示。

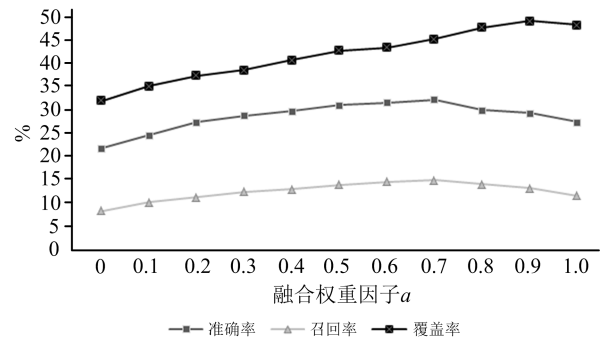


图 5 不同融合权重  $\alpha$  下实验结果对比

Fig. 5 Comparison of experimental results under different  $\alpha$  values

通过实验发现,当融合达到一定比例时,本文算法的准确率、召回率、覆盖率比单个基于语义内容推荐算法( $a=1$ )和基于项目的协同过滤推荐算法( $a=0$ )要高,融合权重  $a$  为 0.7 时准确率和召回率达到最佳值, $a$  为 0.9 时覆盖率达到最佳,证明本文算法是有效的。

#### (IV)和传统协同过滤推荐算法对比

为了验证本文改进型协同过滤推荐算法的有效性,分别和传统基于用户协同过滤(user-based CF)、基于项目协同过滤(item-based CF)进行实验对比。本文算法的知识图谱表示学习嵌入维度设定为 100,融合权重  $a$  设定为 0.7,3 种算法的相似度最大邻近个数  $k$  设定为 30,推荐个数 top- $N$  设定为 10,实验结果如图 6 所示。

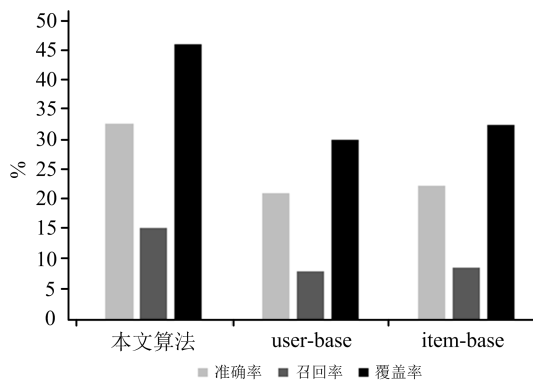


图 6 和传统协同过滤算法实验结果对比

Fig. 6 Comparison of experimental results with traditional collaborative filtering algorithm

通过实验对比发现,本文算法比传统协同过滤推荐算法具有更高的准确率、召回率和覆盖率。

## 4 结论

本文针对协同过滤推荐算法仅利用用户历史评价数据而没有考虑推荐对象间语义关系的问题,提出一种融合推荐对象语义相似度的改进型协同过滤推荐算法。该算法既利用了用户对物品的评价信息,又使用了物品本身内在的语义信息,从多个角度考虑物品相似度的计算,弥补了协同过滤推荐算法没有考虑推荐对象自身语义知识的缺陷。实验结果证明,该算法提高了传统协同过滤推荐算法的准确率、召回率和覆盖率。本文算法也有一定的局限性,一方面算法依赖于开源的知识图谱,导致算法具有一定的领域限制;另一方面,本文算法没有考虑到用户兴趣漂移问题,因为用户的兴趣很有可能随着时间的

发展而发生改变,以往的历史数据具有时效性,因此根据历史数据的时间动态性,建立相关的时间序列推荐模型是值得进一步研究的方向。

#### 参考文献(References)

- [1] 魏慧娟,戴壮红,宁勇余. 基于最近邻居聚类的协同过滤推荐算法[J]. 中国科学技术大学学报, 2016, 46(9): 736-742.  
WEI Huijuan, DAI Muhong, NING Yongyu. Collaborative filtering recommendation algorithm based on nearest neighbor clustering [J]. Journal of University of Science and Technology of China, 2016, 46(9): 736-742.
- [2] 陈洁敏,汤庸,李建国,等. 个性化推荐算法研究[J]. 华南师范大学学报, 2014, 46(5): 8-15.  
CHEN Jiemin, TANG Yong, LI Jianguo, et al. Survey of personalized recommendation algorithm [J]. Journal of South China Normal University, 2014, 46(5): 8-15.
- [3] WU M L, CHANG C H, LIU R Z. Integrating content-based filtering with collaborative filtering using co-clustering with augmented matrices [J]. Expert Systems with Applications, 2014, 41(6): 2754-2761.
- [4] 冷亚军,陆青,梁昌勇. 协同过滤推荐技术综述[J]. 模式识别与人工智能, 2014, 27(8): 720-734.  
LENG Yajun, LU Qing, LIANG Changyong. Survey of recommendation based on collaborative filtering [J]. Pattern Recognition and Artificial Intelligence, 2014, 27(8): 720-734.
- [5] 杨武,唐瑞,卢玲. 基于内容的推荐与协同过滤融合的新闻推荐方法[J]. 计算机应用, 2016, 36(2): 414-418.  
YANG Wu, TANG Rui, LU Ling. News recommendation method by fusion of content-based recommendation and collaborative filtering [J]. Journal of Computer Application, 2016, 36(2): 414-418.
- [6] 郭云飞,方耀宁,扈红超. 基于 Logistic 函数的社会化矩阵分解推荐算法[J]. 北京理工大学学报, 2016, 36(1): 70-74.  
GUO Yunfei, FANG Yaoning, HU Hongchao. A social matrix factorization recommender algorithm based on logistic function [J]. Transactions of Beijing Institute of Technology, 2016, 36(1): 70-74.
- [7] 张宜浩,朱小飞,徐传运,等. 基于用户评论的深度情感分析和多视图协同融合的混合推荐方法[J]. 计算机学报, 2019, 42(6): 1316-1333.  
ZHANG Yihao, ZHU Xiaofei, XU Chuanyun, et al. Hybrid recommendation approach based on deep sentiment analysis of user reviews and multi-view collaborative fusion [J]. Chinese Journal of Computers, 2019, 42(6): 1316-1333.

- [ 8 ] 王巍巍, 王志刚, 潘亮铭, 等. 双语影视知识图谱的构建研究[J]. 北京大学学报, 2016, 52(01):25-34.  
WANG Weiwei, WANG Zhigang, PAN Liangming, et al. Research on the construction of bilingual movie knowledge graph [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2016, 52(01):25-34.
- [ 9 ] 吴玺煜, 陈启买, 刘海, 等. 基于知识图谱表示学习的协同过滤推荐算法[J]. 计算机工程, 2018, 44(02):226-232, 263.  
WU Xiyu, CHEN Qimai, LIU Hai, et al. Collaborative filtering recommendation algorithm based on representation learning of knowledge graph [J]. Computer Engineering, 2018, 44(02):226-232, 263.
- [10] BOBADILLA J. Collaborative filtering based on significances[J]. Information Sciences, 2012, 185(1): 1-17.
- [11] ZHOU T Q, CHEN L N, SHEN J. Movie recommendation system employing the user-based CF in Cloud computing [C]// IEEE International Conference on Computational Science and Engineering. Guangzhou, China: IEEE, 2017:46-50.
- [12] ZHANG H Y, GANCHEV I, NIKOLOV N S, et al. A trust-enriched approach for item-based collaborative filtering recommendations [C]// IEEE, International Conference on Intelligent Computer Communication and Processing. Cluj-Napoca, Romania: IEEE, 2016: 65-68.
- [13] 肖文强, 姚世军, 吴善明. 一种改进的 top-N 协同过滤推荐算法 [J]. 计算机应用研究, 2018, 35(1): 105-108, 112.  
XIAO Wenqiang, YAO Shijun, WU Shanming. Improved top-N collaborative filtering recommendation algorithm [J]. Application Research of Computers, 2018, 35(1):105-108, 112.
- [14] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3):582-600.  
LIU Qiao, LI Yang, DUAN Hong, et al. Knowledge graph construction techniques[J]. Journal of Computer Research and Development, 2016, 53(3):582-600.
- [15] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]// International Conference on Neural Information Processing Systems. Daegu, Korea: Curran Associates Inc. 2013:3111-3119.
- [16] FAN M, ZHOU Q, ZHENG T F, et al. Distributed representation learning for knowledge graphs with entity descriptions [J]. Pattern Recognition Letters, 2017, 93(1):31-37.
- [17] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: A review and new perspectives [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(8): 1798-1828.
- [18] 刘知远, 孙茂松, 林衍凯, 等. 知识表示学习研究进展 [J]. 计算机研究与发展, 2016, 53(2):247-261.  
LIU Zhiyuan, SUN Maosong, LIN Yankai, et al. Knowledge representation learning: A review [J]. Journal of Computer Research and Development, 2016, 53(2):247-261.
- [19] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data [C]// International Conference on Neural Information Processing Systems. Daegu, Korea: Curran Associates Inc. 2013: 2787-2795.