

采用评论挖掘修正用户评分的改进协同过滤算法

王红霞, 陈健, 程艳芬

(武汉理工大学 计算机科学与技术学院, 湖北 武汉 430063)

摘 要: 针对当前电子商务网站用户评分过于集中而区分度不明显, 以及整数评分可信度不高导致协同过滤推荐效果较差的问题, 提出一种改进协同过滤算法. 利用改进的词性路径模板算法挖掘评论中包含的产品特征和情感词, 分析并建立评论特征偏好向量; 依据评论特征偏好向量计算评论中包含的情感态度, 利用用户评论中包含的情感态度对评分进行修正, 使得修正后的评分更接近于用户的真实评分意愿; 利用修正后的评分计算评分相似度, 与偏好相似度结合产生推荐. 实验结果表明, 该算法有效地增加了评分区分度与可信度, 提高了最近邻居的质量, 从而提高了推荐结果的准确度.

关键词: 评论挖掘; 情感态度; 评论特征偏好向量; 评分修正; 协同过滤

中图分类号: TP 391 **文献标志码:** A **文章编号:** 1008-973X(2019)03-0522-11

Improved collaborative filtering algorithm to revise users' rating by review mining

WANG Hong-xia, CHEN Jian, CHENG Yan-fen

(School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430063, China)

Abstract: An improved collaborative filtering algorithm was proposed aiming at the problems that user ratings for current e-commerce websites are too concentrated, the distinguishing degree is not obvious, and the credibility of the integer rating is not very well. Firstly, the improved part-of-speech path template algorithm was used to mine the product features and sentiment words contained in user's reviews, and then the preference vector of the review was analyzed and established. Secondly, the emotional attitudes included in the review were calculated according to the review's preference vector, and the emotional attitudes were used to revise the user's rating, so that the revised rating was closer to the user's true score willingness. Finally, the revised rating similarity and the preference similarity were combined to produce recommendations. The experimental results show that the proposed algorithm can effectively increase the classification and credibility of user's ratings, and thus improves the quality of the nearest neighbors, as well as the accuracy of the recommended results.

Key words: review mining; emotional attitude; comment feature preference vector; rating revising; collaborative filtering

随着电子商务的快速发展, 互联网上商品的种类和数量也在不断增加, 用户很难在众多商品中找到自己想要购买或者感兴趣的商品, 产生了“信息过载”. 推荐系统是解决“信息过载”的一个有效手段^[1-3], 其中协同过滤算法的应用最为广泛.

该算法通过分析用户的历史数据, 构建个人兴趣偏好, 利用兴趣相似的其他用户向用户推荐可能喜欢的商品. 这不仅大大减轻了用户的负担, 还可以发现用户潜在的兴趣爱好. 目前已有许多电子商务网站运用推荐系统来提高产品的销售量,

收稿日期: 2018-02-06. 网址: www.zjujournals.com/eng/fileup/HTML/201903013.htm

作者简介: 王红霞 (1977—), 女, 副教授, 博士, 从事模式识别、图像分类研究. orcid.org/0000-0003-1213-7760. E-mail: 99575522@qq.com

通信联系人: 程艳芬, 女, 副教授, 博士. orcid.org/0000-0003-0472-4543. E-mail: 995132428@qq.com

如亚马逊、天猫、京东商城等。

通过挖掘用户评论建立用户兴趣偏好,再结合用户评分改进传统的协同过滤算法是近年来推荐领域研究的热点^[4-7]。Ganu等^[8-9]提出了一种基于支持向量机(support vector machine, SVM)的多标签文本分类器,对评论进行分类,通过评论生成文本评分,利用文本评分产生推荐。但在该方法中,需要人工注释大约3 400条句子的主题类别和情感分类。Mcauley等^[10]通过结合latent-factor模型与**文档主题生成模型**(latent dirichlet allocation, LDA),提出了潜在主题因素(hidden factors as topics, HFT)模型,利用LDA方法获取商品的评论主题分布,再将评论主题分布与latent-factor模型相结合,从而建立评论主题与评分维度之间的关系。该方法将评论主题分布与潜在评分维度视为一致并建立转换,缺乏较好的解释性与扩展性。Dehkordi等^[11]将用户评论以及偏好相似的其他用户评论作为隐式反馈加入到协同过滤算法中,一定程度上提高了推荐结果的准确率。张付志等^[12]将评论信息和环境信息引入传统的协同过滤算法,利用评论信息和环境信息来改进评分相似度,从而产生推荐。该方法仍以用户评分为主要依据来计算用户之间的相似度,在用户评分比较集中时无法很好地区分用户之间的相似度。谭云志等^[13]利用评论学习用户的喜好分布和厌恶分布,共同组成用户偏好相似度,再与评分相似度协同产生推荐。该方法只是将偏好相似度作为基础评分项的评分修正项,基础评分项则通过现有的任何评分预测算法得到。李伟霖等^[14]利用主题模型产生评论主题分布,利用评分生成评论态度影响因子来放大主题分布中的突出特征,建立了更加准确的用户偏好与物品特征。

关于评论情感打分机制的研究,王伟军等^[15]通过引入语言学的话语标记理论构建评论话语标记语库,设计了包含可信度、评论立场、情感态度的评分修正体系解决评论和评分的偏差问题。文献^[15]针对的是评论中的话语标记,如坦言型、阐发性、理据性、断言性和评价性等,没有考虑评论中对产品特征的情感评价,忽略了用户评论中产品特征以及情感评价这一关键信息,且需要通过问卷以及人工方式建立话语标记库。郑修猛等^[16]通过评论分词、提取特征、情感分析等技术将评论转化为对项目的评分,从而缓解了评分矩阵过于稀疏的问题。Yang等^[17]通过提取评论中的情

感词,将其划分为积极和消极两类,然后计算出评论得分作为相似度的补充,从而改进了协同过滤算法。

由于电子商务网站的评分数据存在过于集中且整数评分可信度不高的问题,本文旨在根据评分和评论之间的关联关系,挖掘评论中提及的产品特征和情感描述,建立评论级的特征偏好向量,依据评论的特征偏好向量计算评论的情感态度,再利用情感态度修正相应评分,从而提高评分的区分度和可信度。

1 评分问题剖析

通过统计知名电子商务网站的用户评分数据发现,用户的评分数据非常集中且为整数分值,这样的评分数据区分度不明显,可靠性和可信度不高,导致协同过滤推荐效果较差。

1)关于用户评分过于集中的问题。以京东商城为例,统计14大类、142小类产品的评分好评率。结果显示,好评率最低为94.3%,最高达到98.6%,平均好评率为96.2%。受评分规则(评分为1~5的整数)的限制,且产品整体质量较好,用户的评分非常集中,几乎都是5分,区分度非常不明显。

2)关于整数评分可信度不高的问题。由于电子商务网站的评分是1~5的整数,用户只能尽量选择与其真实评分意愿最为接近的整数作为评分结果,这样往往不能真实地表达用户的评分意愿。从实验数据集中随机抽取评价同一产品且评分为5分的4条评论如下。用户一:“物流简直飞速,头一天晚上下的单,第二天中午就收到货,都有点不敢相信。手机非常漂亮,速度也很快,完全没有卡顿现象,配置也是杠杠滴,完全物超所值。一句话,棒,太棒了”;用户二:“内存大,电池耐用,速度快,还不错哦”;用户三:“没有他们评价的那么好,只能说对得起999这价格,按键不灵敏,凑合”;用户四:“外观还可以,就是反应有点不理想,用用看吧!发货很快,快递员很好”。可以看出,在用户对产品较为满意的情况下,都会选择5分作为评分结果。用户评论内容中包含的情感倾向表明:用户的真实评分意愿如果只是单一用5这个整数来表示,明显不太精确。

本文旨在挖掘用户评论中对产品特征的情感描述,建立评论级的特征偏好向量,依据评论特征偏好向量求出评论的情感态度,利用情感态度

修正用户相应评分,突破整数评分的限制,把整数评分细化到小数层面,将过于集中的评分离散化,提高评分的区分度;同时使修正后的评分更加接近用户的真实评分意愿,提高评分的可信度。本文的主要研究内容如下:

1)通过挖掘用户评论,提取出描述产品的特征词和相应的情感描述词,建立评论级的特征偏好向量,从而使评论的情感态度描述更加准确;

2)依据评论的特征偏好向量计算出评论的情感态度,利用情感态度修正相应评分,提高评分的可信度和区分度;

3)爬取京东商城真实的交易数据和评论作为实验数据集,通过改进的词性路径模板算法提取评论中的特征词及相应的情感词,依据情感态度评价算法量化评论的情感态度,并通过情感态度对评分进行修正,解决评分过于集中以及可信度不高的问题。

2 采用评论挖掘修正用户评分的改进协同过滤算法

2.1 算法整体框架

在电子商务网站中,用户对产品的反馈通常包含一个评分和一段评论。评分是用户对产品的综合评价;而评论则包含了用户的详细评价,是对评分结果的解释,因此评分和评论是息息相关的。针对评分过于集中以及可信度不高的问题,从评分对应的评论中挖掘分析用户的情感态度,利用评论的情感态度修正评分,达到提高评分的区分度与可信度,从而提高推荐结果准确度的目的。该算法的整体框架如图1所示,主要处理过程分为以下5个步骤。

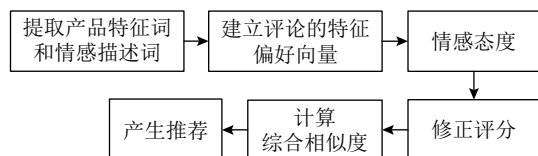


图1 采用评论挖掘修正用户评分算法的整体框架

Fig.1 Overall framework of improved collaborative filtering algorithm based on revising user's ratings by mining reviews

1)提取评论中的产品特征词和情感词。将评论泛化到词性层面,提取出评论中包含的产品特征词和情感词,过滤没有产品特征词的评论。

2)建立用户特征偏好。考虑到不同的用户对产品关注的特征不同,有必要为用户建立个人特征偏好,用户特征偏好描述了用户对产品不同特征的偏好程度。

3)建立评论特征偏好向量。不同评论中包含的产品特征词不同,表达的情感倾向也不相同,因此给出“评论特征偏好向量”的概念,针对每一条评论,依据评论中提取的产品特征和情感词建立该条评论的评论特征偏好向量。

4)修正用户评分。先将情感词进行量化,然后依据量化后的分值与评论特征偏好向量计算出评论的情感态度,利用情感态度修正评分并计算修正后的评分相似度。

5)计算综合相似度。将偏好相似度和修正后的评分相似度加权求和得到综合相似度。

6)产生推荐。根据综合相似度计算待预测用户的K-最近邻,利用最近邻居集来预测用户对产品的评分,选取预测评分超过阈值的产品作为推荐结果。

2.2 提取产品特征词和情感词

产品特征表示用户对产品某个方面的关注点,情感词表达了用户对该关注点的情感评价,例如:很好、差等。提取产品特征词和情感词的第一步是分词。目前常用的中文分词算法有LTP-cloud、ICTCLAS、FudanNLP、CRF++、StanfordNLP等,采用ICTCLAS分词系统对评论进行分词处理。ICTCLAS是目前分词精度最高的汉语词法分析系统,分词精度高达98.45%。评论分词完成后,需要提取其中包含的特征词和情感词。商品特征词和情感词提取主要有基于词频和共现、机器学习、句法依存、改进的LDA方法等。Wu等^[18]利用依存分析发现,评论中商品特征与情感词之间的联系,并使用Tree-kernel SVM方法将特征词和情感词组合分类,分为“相关”和“不相关”两类,扈中凯等^[19]将语言泛化到词性层面,利用统计方法寻找特征词和情感词之间的频繁联系,得到词性路径模板,依据词性路径模板提取评论的特征词和情感词,取得了较好的效果。本研究采用文献^[19]提出的词性路径模板算法,并针对其不足加以改进,提取评论中的产品特征词和情感词,并将提取出的产品特征词 w 和对应的情感词 h 用特征情感词对 $f=(w,h)$ ^[19]表示。

2.2.1 改进的词性路径模板提取算法 文献^[19]假设特征词和情感词之间存在某些句法联系,将语言粒度泛化到词性层面,采用统计方法寻找这些

频繁的联系,最终提取出 12 个高频词性路径模板,再通过匹配评论中每个句子的模板,提取出相应的特征词与情感词,这是一种“硬匹配”.实验发现:“硬匹配”容易漏掉某些特征情感词对,针对此不足,提出“软匹配”思想,依据词性路径模板对评论进行“模糊匹配”,提取评论中的特征词和情感词.改进的词性路径模板提取算法步骤如下.

1)评论分词,加入词性标注,去掉标点和中文字,生成词性句子,类似于“/v/a/u/a”;

2)根据词性路径模板对 1)生成的词性句子进行模糊匹配,找出词性句子中所有匹配模板的结果句子,匹配遵循以下规则:a)不要求模板和标注句子完全匹配;b)允许有“间隔”词性的结果句子和模板匹配.按照以上规则,找出所有的结果句子.以 1)中的结果为例:假设词性路径模板为“/v/a”,则可以提取到的特征情感词对有“/v/a”(第一个“/a”)和“/v/a”(第二个“/a”).

3)对特征情感词对进行过滤,利用“互信息”来判断特征词是否和产品主题相关,判断情感词和情感是否相关,过滤和主题不相关或者非情感词的特征情感词对.

随机选取测试数据集中的一条评论,对比文献[19]的词性路径模板算法和改进后的词性路径模板算法,对比结果如表 1 所示.由表 1 可以看

出,改进后的词性路径模板算法可以提取出文献[19]中的词性路径模板算法无法提取的特征情感词对.

2.3 建立评论特征偏好向量并修正评分

为了清晰地描述提出的算法,在具体叙述如何建立评论特征偏好向量以及修正评分之前,先作如下定义.

定义 1 浮动因子 δ .表示评分修正的幅度大小.若原评分为 v ,则修正后的评分在 $[v-\delta, v+\delta]$ 上,浮动因子 δ 的大小限制了评分修正的幅度.考虑用户给出的整数评分都会尽可能表达自己的真实评分意愿,数值上理解为给出的评分是距离真实评分意愿最近的整数,这符合四舍五入的原则.因此,可以认为用户的真实评分在原整数评分上下浮动 0.5 分的区间内,假设原整数评分为 x ,则真实评分在 $[x-0.5, x+0.5]$ 内.此时,浮动因子 $\delta=0.5$.整数评分修正后的范围如图 2 所示.

定义 2 评论特征偏好向量 S .算法依据评论中提取的产品特征词和情感词为评论建立特征偏好向量.以评论 t 为例,则评论 t 的特征偏好向量记为 S_t .

定义 3 评论情感态度 E .评论内容表现出的整体情感倾向是用户对产品评分的直接依据,由评论特征偏好向量计算得出.

表 1 词性路径模板算法提取效果对比
Tab.1 Comparison of extraction effects of speech path template algorithm

评论内容	词性标注	词性句子	文献[19]算法	改进后的算法
手机外观不错	手机/n 外观/n 不错/a	/n/n/a	—	(外观,不错)
手机颜色也很漂亮	手机/n 颜色/n 也/d 很/d 漂亮/a	/n/n/d/d/a	—	(颜色,漂亮)
后盖是金属的	后盖/n 是/v 金属/n 的/u	/n/v/n/u	—	—
有重量	有/v 重量/n	/v/n	—	—
铃声大	铃声/n 大/a	/n/a	(铃声,大)	(铃声,大)
屏幕大	屏幕/n 大/a	/n/a	(屏幕,大)	(屏幕,大)
操作真的简单	操作/v 真/a 的/u 简单/a	/v/a/u/a	—	(操作,简单)
适合中老年使用	适合/v 中老年/n 使用/v	/v/n/v	—	—

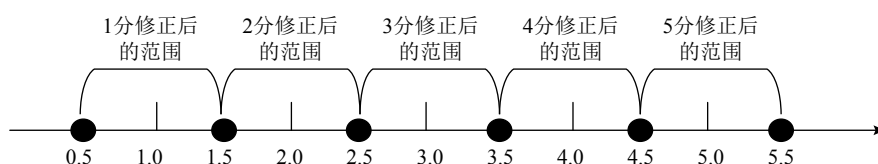


图 2 整数评分修正后的范围
Fig.2 Range of revised integer rating

2.3.1 建立用户特征偏好 考虑到不同用户对产品有不同的特征偏好,有必要为用户建立个人特征偏好,用户特征偏好描述了用户对产品不同特征的偏好程度.可通过 TF-IDF 算法确定每个用户的特征偏好.词频 (term frequency, TF) 是某一个给定的词语在给定文件中出现的频率.逆向文档频率 (inverse document frequency, IDF) 是一个词语普遍重要性的度量.

用户评论语料库中不同特征词的频数表明了用户对该特征的关注程度,较高关注程度的特征词在用户特征偏好中的权重较大.建立用户特征偏好的步骤如下.

1) 将每个用户的所有评论组成该用户的评论文档 T , 例如, 用户 j 的所有评论组成评论文档 T_j .

2) 计算文档中所有特征词的 TF-IDF 值. 以特征词为例, 对于文档 T_j 中的特征词 w_i , 词频表示为

$$f_{i,j} = n_{i,j} / \sum_k n_{k,j}. \quad (1)$$

式中: $n_{i,j}$ 为特征词 w_i 在文档 T_j 中出现的次数, $\sum_k n_{k,j}$ 为文档 T_j 所有特征词出现次数的和.

对 $f_{i,j}$ 进行归一化处理, 得到

$$\text{TF}_{i,j} = \theta + \frac{(1-\theta)f_{i,j}}{\max(f_{i,j})}. \quad (2)$$

其中, 默认 $\theta = 0.5$. 通过下式计算 IDF:

$$\text{IDF}_{i,j} = \log \left(\frac{|T|}{|\{j \in T : i \in T_j\}| + 1} \right). \quad (3)$$

其中, j 为包含特征词 w_i 的评论数量.

TF-IDF 的计算公式为

$$\text{TI}_{i,j} = \text{TF}_{i,j} \times \text{IDF}_{i,j}. \quad (4)$$

3) 将文档 T_j 中所有特征词的 TF-IDF 计算结果组成用户的特征偏好, 记为 $W = \{\text{TI}_{1,j}, \text{TI}_{2,j}, \dots, \text{TI}_{i,j}\}$.

2.3.2 建立评论特征偏好向量 接下来详细叙述建立评论特征偏好向量和修正评分的算法步骤. 以用户 u 为例, 建立评论特征偏好向量的算法伪代码如下.

算法 1 建立评论特征偏好向量算法.

输入: 用户 u 的评论集合 T_u ;

输出: 评论特征偏好向量的集合 S_u .

Segmentation(T_u); /*对评论进行分词*/

$F_u \leftarrow \text{Fetch}(T_u)$; /*提取特征情感词对*/

$N \leftarrow \text{addAll}(F_u)$; /*所有特征词数量求和*/

for each $w \in F_u$

$p(w) \leftarrow n_w/N$;

$W_u \leftarrow p(w)$; /*建立用户特征偏好*/

end for

for each $t \in T_u$

$P \leftarrow \text{addAll}(W_u)$; /*特征词的 TD-IDF 值求和*/

for each $w \in t$

$q(w) \leftarrow p(w)/P$;

$S_t \leftarrow q(w)$; /*建立评论特征偏好向量*/

end for

$S_u \leftarrow S_t$;

end for

1) 提取特征情感词对. 获取用户 u 的所有评论组成评论集合 T_u , 根据第 2.2 节对 T_u 进行分词并提取评论中的特征词和情感词, 组成特征情感词对集合 F_u .

2) 建立用户特征偏好. 根据第 2.3.1 节建立用户 u 特征偏好, 记为 W_u .

3) 建立评论特征偏好向量. 评论特征偏好表征了每条评论对其中提到的产品特征的偏好程度. 将特征偏好细粒度到评论层面, 是为了更细致地描述每条评论的特征偏好, 从而得到更加准确的评论情感态度. 对于 T_u 中的任意一条评论 t , 若 t 中出现的特征词有 w_1, w_2, \dots, w_k , 则对应的评论特征偏好向量表示为 $S_t = \{q(w_1), q(w_2), \dots, q(w_k)\}$,

$$q(w_i) = W_{u,i} / \sum_{j=1}^k W_{u,j}. \quad (5)$$

其中, $W_{u,i}$ 为用户 u 的特征词 w_i 的 TF-IDF 值. $\sum_{j=1}^k W_{u,j}$ 为评论 t 中所有特征词 w_1, w_2, \dots, w_k 的 TF-IDF 之和, $i = 1, 2, \dots, k; j = 1, 2, \dots, k; W_{u,i} \in W_u$.

2.3.3 修正评分 建立评论特征偏好向量后, 根据评论特征偏好向量修正评分, 仍以用户 u 为例, 根据评论特征偏好向量修正评分的算法伪代码如下.

算法 2 利用评论特征偏好向量修正评分算法.

输入: S_u , 原评分 v 的集合;

输出: 修正后的评分 v' 的集合.

for each $t \in T_u$

for each $h \in t$

$r(u) \leftarrow \text{quantify}(u)$; /*量化情感词*/

$R_t \leftarrow r(h)$; /*建立情感词分值向量*/

end for

$E_t \leftarrow S_t \times R_t$; /*计算情感态度*/

$M \leftarrow (M_{\max} - M_{\min}) / (O_{\max} - O_{\min}) \times$

$(E_t - O_{\min}) + M_{\min}$;

$v' \leftarrow v + M$;

end for

1) 量化情感词. 对于 T_u 中的任意一条评论 t , 若 t 中出现的特征词有 w_1, w_2, \dots, w_k , 对应的情感词记为 h_1, h_2, \dots, h_k . 采用台湾大学 NTUSD 简体中文情感词典来判断情感词的极性^[20], NTUSD 是目前中文词极性判断准确度最高的中文情感词典. 量化规则如下: 将带有积极情绪的情感词量化为 5 分, 消极词量化为 1 分, 中性词量化为 3 分. 依据该规则对情感词 h_1, h_2, \dots, h_k 进行量化, 量化后的分值 $r(h_1), r(h_2), \dots, r(h_k)$ 组成向量 R_t , 记为 $R_t = \{r(h_1), r(h_2), \dots, r(h_k)\}$.

2) 计算评论情感态度. 评论情感态度表示评论对所评价产品表现出的整体情感倾向, 由评论特征偏好向量计算得出

$$E_t = S_t \cdot R_t. \quad (6)$$

计算得到的情感态度 $E_t \in [1, 5]$.

3) 依据情感态度修正评分. 利用 2) 得到的情感态度 E_t 修正评分之前, 需要将其作一次变换, 得到最终用于修正评分的修正值. 根据浮动因子 δ 的定义可得 $\delta = 0.5$, 即评分修正的幅度为 0.5, 修正值须在区间 $[-0.5, 0.5]$ 上. 而情感态度 $E_t \in [1, 5]$, 因此需要将 E_t 映射到修正值所在的区间 $[O_{\min}, O_{\max}]$. 通过改进 min-max 标准化公式来完成映射, 可实现将原区间 $[O_{\min}, O_{\max}]$ 上任意数值映射到目标区间 $[M_{\min}, M_{\max}]$ 上:

$$M = \frac{M_{\max} - M_{\min}}{O_{\max} - O_{\min}} \times (E_t - O_{\min}) + M_{\min}. \quad (7)$$

式中: $[O_{\min}, O_{\max}] = [1, 5]$ 为情感态度 E_t 的区间, $[M_{\min}, M_{\max}] = [-0.5, 0.5]$ 为修正值的区间, M 为将 E_t 映射到修正值区间后的结果, 即修正值. 式 (7) 得到的 M 为最后的修正值:

$$v' = v + M. \quad (8)$$

2.4 计算综合相似度

综合相似度包括评分相似度和偏好相似度, 是两者加权求和的结果. 采用改进的相关相似度计算公式^[19]来计算评分相似度:

$$\text{sim}^s(i, j) = \frac{1}{\gamma} \min(|I|, \gamma) \times \text{sim}^l(i, j). \quad (9)$$

式中: 通过设定阈值 γ 与共同评价过的项目集合 I 进行比较, 可以避免因共同评价项目较少时较大误差的情况. $\text{sim}^l(i, j)$ 为用户 i 和用户 j 的相关相似度:

$$\text{sim}^l(i, j) = \frac{\sum_{c \in I} [(V'_{i,c} - \bar{V}_i)(V'_{j,c} - \bar{V}_j)]}{\left[\sum_{c \in I} (V'_{i,c} - \bar{V}_i)^2 \right]^{1/2} \left[\sum_{c \in I} (V'_{j,c} - \bar{V}_j)^2 \right]^{1/2}}. \quad (10)$$

式中: I 为用户 i 和用户 j 共同评价过的项目集合, $V'_{i,c}$ 和 $V'_{j,c}$ 分别为用户 i 和用户 j 对共同评分项目 c 修正后的评分, \bar{V}_i 和 \bar{V}_j 分别为用户 i 和用户 j 对所评分项目的修正后的评分均值.

用户的偏好相似度采用余弦相似度来表示:

$$\text{sim}^p(s_i, s_j) = \cos(s_i, s_j) = \frac{s_i \cdot s_j}{\|s_i\| \|s_j\|}. \quad (11)$$

式中: s_i 和 s_j 分别为用户 i 和用户 j 的兴趣偏好向量, $\text{sim}^p(s_i, s_j)$ 为用户 i 和用户 j 的偏好相似度, $\text{sim}^p(s_i, s_j)$ 的值域在 $[0, 1]$ 上, 值越大, 则表明用户 i 和用户 j 的兴趣偏好越相似.

综合相似度的计算公式如下:

$$\text{sim}(i, j) = \theta \cdot \text{sim}^s(i, j) + (1 - \theta) \text{sim}^p(s_i, s_j). \quad (12)$$

式中: $\text{sim}^s(i, j)$ 和 $\text{sim}^p(s_i, s_j)$ 分别代表用户 i 和用户 j 的评分相似度和偏好相似度, 而权重因子 θ 的取值范围在 $[0, 1]$ 上.

2.5 评分预测和产生推荐

采用与目标用户相似度最高的 K 个用户作为目标用户的邻居集合来预测目标用户对产品的评分. 这 K 个用户与目标用户的相似度最高, 表明两者具有相似的爱好的, 于是 K 个用户的评分可以作为目标用户预测评分的参考评分. 评分预测的计算公式如下:

$$P_{i,A} = \bar{V}_i + \frac{\sum_{u \in U} (V'_{u,A} - \bar{V}_u) \text{sim}(i, u)}{\sum_{u \in U} \text{sim}(i, u)}. \quad (13)$$

式中: \bar{V}_i 为用户 i 对所评分项目的修正后的评分均值, U 为最近邻居集合, 包含 K 个用户, $V'_{u,A}$ 为用户 u 对项目 A 的修正后的评分, \bar{V}_u 为用户 u 对所评分项目的修正后的评分均值, $\text{sim}(i, u)$ 为用户 i 和用户 u 的综合相似度.

3 实验与分析

3.1 实验数据

编写爬虫程序爬取京东商城手机类商品的评分和评论信息, 选择手机类商品的评分和评论作

为实验数据集有以下优点: 1) 手机产品更新速度快, 用户会多次购买, 降低了评分矩阵的稀疏程度; 2) 手机产品的评论十分详细, 绝大部分评论会说明购买原因以及对产品特征的评价, 这有利于对用户进行偏好建模以及对评论特征的提取。

爬取的数据全部来源于京东商城手机类商品真实的交易评分数据和评论信息, 包括苹果系列、小米系列、荣耀系列、三星系列、魅族系列、夏普、OPPO、努比亚等, 累计爬取 45 万条评论数据。然后对评论数据进行数据清洗和预处理。数据清洗的主要目标是去除无效的评论数据, 主要包含以下步骤: 1) 过滤只有评分没有评论的数据; 2) 过滤评论中没有特征词或者特征词较少的数据。

数据清洗可以去除掉大量的无效评论数据, 最终提取出 27 209 个用户对 2 954 款手机产品的 28 862 条评论。评分信息统计结果显示, 5 分评价达到 91.6%, 表明使用原整数评分进行推荐的局限性; 评论数量信息统计结果显示, 用户评论次数分布如表 2 所示, N_{com} 为评论次数, N_{user} 为拥有对应评论次数的用户数量。可以看出, 经过预处理后的用户都发表过至少 1 条以上的评论, 这样就可以利用评论中包含的丰富用户情感信息建立准确的用户偏好, 保证算法的可行性。

实验采用交叉验证法, 其中 80% 的数据作为训练集, 20% 的数据作为测试集, 对训练结果进行验证; 并与传统的协同过滤算法以及近年来融合了评论和评分的产品推荐算法进行对比, 进一步证明算法改进的有效性。

3.2 评测标准

实验采用的推荐算法评价指标是准确率、召回率和多样性^[21]。准确率采用平均绝对误差 (mean

absolute error, MAE) 来衡量评分预测的准确度, 如果 $r_{u\alpha}$ 为用户 u 对商品 α 的真实评分, $r'_{u\alpha}$ 为用户 u 对商品 α 的预测评分, E_p 为测试集, 那么

$$\text{MAE} = \frac{1}{|E_p|} \sum_{(u,\alpha) \in E_p} |r_{u\alpha} - r'_{u\alpha}|. \quad (14)$$

MAE 越小表示误差越小, 预测的质量也就越好。不过, MAE 指标也有一定的局限性, 因为对 MAE 指标贡献比较大的往往都是很难预测评分的低分商品, 但是从测试数据集来看, 4 分和 5 分的比例占 94.27%, 因此 MAE 的局限可以忽略不计。

召回率表示一个用户喜欢的商品被推荐的概率, 定义为推荐列表中用户喜欢的商品与系统中用户喜欢的所有商品的比率。对于用户 u , 其召回率为

$$R_u = \frac{\sum_{u \in U} |I(u) \cap D(u)|}{|D(u)|}. \quad (15)$$

式中: $I(u)$ 为用户 u 的商品推荐列表, $D(u)$ 为测试集中用户 u 评分在 3 分以上的商品列表。将所有用户的召回率求平均得到系统整体的召回率:

$$R_U = \frac{1}{n} \sum_{u \in U} R_u. \quad (16)$$

推荐系统中的多样性有 2 种: 用户间的多样性 (Inter-user diversity), 衡量推荐系统对不同用户推荐不同商品的能力; 另一个是用户内的多样性 (Intra-user diversity), 衡量推荐系统对一个用户推荐商品的多样性。采用 Intra-user diversity 评价算法推荐结果的多样性:

$$I_u(L) = \frac{1}{L(L-1)} \sum_{\alpha \neq \beta} s(\alpha, \beta). \quad (17)$$

式中: $s(\alpha, \beta)$ 为商品 α 和商品 β 的相似度, L 为推荐在用户 u 的前 L 个商品。

3.3 实验结果与分析

3.3.1 评分修正结果与分析 采用提出的算法对用户评分进行修正, 由于数据量较大, 只列出了部分数据, 分别如图 3 和表 3 所示。图 3 选取人数最集中的 5 分数据, 将修正后的用户评分分布给予展示, S 为修正后的评分, N_p 为评分对应的人数, 可以看出, 修正后的小数评分分布范围更广, 更加离散化, 区分程度明显提高。对随机抽取的两条用户评论进行评分修正, 以及修正评分与用户真实意愿的对应关系, 即评分可信度的展示如表 3 所示。从表中修正结果来看, 提出的算法可以很好地根据用户评论的情感态度修正评分, 使修

表 2 评论次数与对应用户数量

Tab.2 Number of reviews and corresponding users

N_{com}	N_{user}	N_{com}	N_{user}
1	26 173	9	7
2	783	10	4
3	135	11	4
4	58	12	1
5	9	13	2
6	7	19	1
7	18	21	1
8	6	—	—

正后的评分更加接近用户的真实评分意愿,提高了评分的区分度和可信度.

3.3.2 推荐结果与分析 通过几个对比实验,从

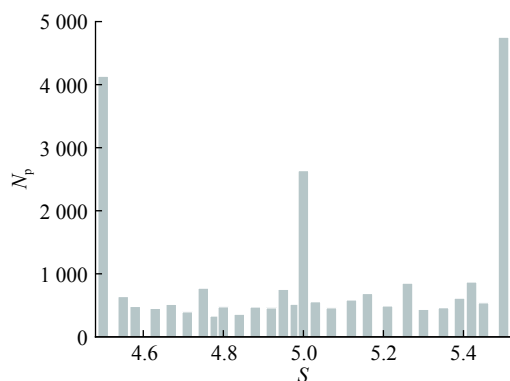


图3 修正后的用户评分分布

Fig.3 Distribution of revised user rating distribution

MAE、召回率和多样性等多方面来论证算法的有效性,并且讨论权重因子 θ 的最优值.实验中的对比算法分别为经典算法和近年来结合了评论与评分的优秀推荐算法.基于评分数据的协同过滤算法记为CF,文献[12]提出的融合用户评论和环境信息的协同过滤推荐算法记为CFIRC,文献[19]提出的基于用户评论挖掘的产品推荐算法记为PP-CF,本文算法记为RR-CF.

1)在仅考虑评分的情况下,将RR-CF算法与CF算法进行对比.通过改变最近邻居数 N 得到不同的MAD值,对比结果如图4所示.从图4可以看出,随着最近邻居数 N 的增加,RR-CF算法的MAD曲线趋于稳定且始终在CF算法的MAD曲线下方,表明相比于原评分,使用修正后的评分可以提高评分预测的准确度,证明了评分修正的有效性.

(2)将RR-CF算法与近年来结合了评论与评分的推荐算法进行对比.为了进一步证明算法的有效性,将RR-CF与融合了评分和评论信息的CFIRC、PP-CF进行对比.通过改变最近邻居数 N 得到不同的MAD值、召回率 R 以及多样性 μ ,实验结果分别如图5~7所示.

由图5可以看出,RR-CF算法的MAD曲线始终处于CFIRC、PP-CF和CF的MAD曲线下方,表明RR-CF算法的准确率要优于CFIRC、PP-CF和CF算法.RR-CF算法的平均MAD值为0.071,CFIRC和PP-CF算法的平均MAD值分别为0.083和0.082;CFIRC算法在计算评论整体的极性和强度时没有考虑不同用户情感评价尺度不同的问题,容易产生“情感失衡”,导致用户偏好不够精确;而RR-CF算法通过用户特征偏好和情感偏好解决了情感评价尺度问题,而且利用评论特征偏好向量求出的情感态度相较于其他算法要更加准确.图6中,RR-CF算法的平均召回率为0.047,CFIRC和PP-CF算法的平均召回率分别为0.045和0.041.由此看出,RR-CF算法的召回率均优于CFIRC、PP-CF和CF算法的召回率;同时,RR-CF算法与CFIRC算法的稳定性相当,优于PP-CF和CF算法.由图7可以看出,多样性随着最近邻居数 N 的增加而增加,而且RR-CF算法的多样性也稍优于CFIRC、PP-CF和CF算法,进一步增加了向用户推荐商品多样性的能力.

3)讨论权重因子 θ 的最优值.

通过2)的实验得到,在实验数据集中,RR-CF算法的最近邻居数的最优值为30,而PP-CF算法最近邻居数的最优值为45.分别取RR-CF和PP-CF算法的最优、最近邻居数,在测试权重因子 θ 取不同数值的情况下,进行MAE值对比,实验结果如图8所示.

由图8可以看出,RR-CF算法的MAE曲线始终在PP-CF算法的MAE曲线下方,表明RR-CF算法的评分预测准确度始终要优于PP-CF算法,进一步证明了算法改进的有效性.从图8还可以看出,当权重因子 θ 取值为0~1.0时,PP-CF算法的MAE曲线先下降、后上升,在 $\theta=0.3$ 时,MAE值达到最小.当 $\theta=0.3$ 时,用户的综合相似度中,评分相似度占30%,偏好相似度占70%.这是合理的,原评分过度集中,导致评分相似度区分不明显,偏好相似度反而能更好地区分用户喜好.因

表3 原评分与修正后评分的可信度对比

Tab.3 Credibility comparison of original rating and revised rating

评论内容	原评分	提取特征情感词对	情感态度	修正后的评分
整体比较满意,就感觉下面功能键不好,反应慢	5	{(整体,满意),(反应,慢)}	1.98	4.70
手机不错,外观漂亮,运行流畅,相机可以,给京东物流一个赞	5	{(手机,不错),(外观,漂亮),(运行,流畅),(相机,可以),(物流,赞)}	4.88	5.47

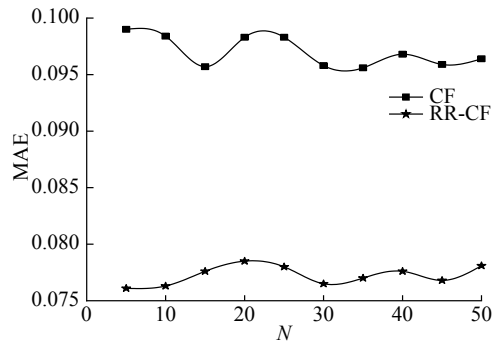


图 4 仅考虑评分时所提算法与传统协同过滤算法的 MAE 值对比

Fig.4 Comparison of MAE values between proposed algorithm and traditional collaborative filtering algorithm considering only score

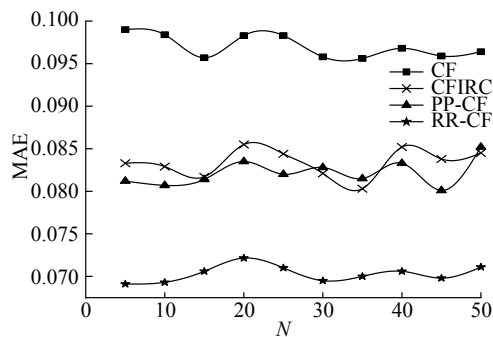


图 5 采用评论挖掘的不同协同过滤算法的 MAE 值对比

Fig.5 Comparison of MAE values among different collaborative algorithms based on review mining

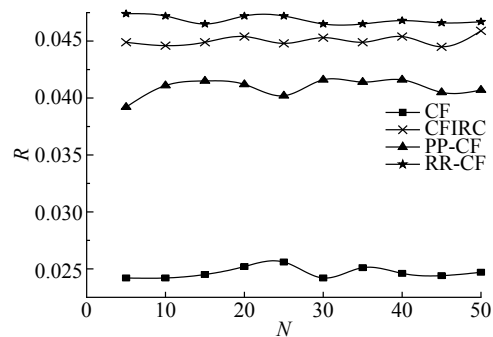


图 6 采用评论挖掘的不同协同过滤算法的召回率对比

Fig.6 Comparison of recall rate among different collaborative algorithms based on review mining

此在综合相似度中, 偏好相似度的作用大于评分相似度. 而 RR-CF 算法的 MAE 曲线先下降、后上升, 在 $\theta=0.5$ 时, MAE 值达到最小; 即在综合相似度中, 当评分相似度和偏好相似度各占 50% 时, MAE 值达到最小. 图 8 中的 2 条曲线对比说明, 经过 RR-CF 算法修正用户评分后, 评分区分度和可信度增加, 从而在综合相似度中占据的比

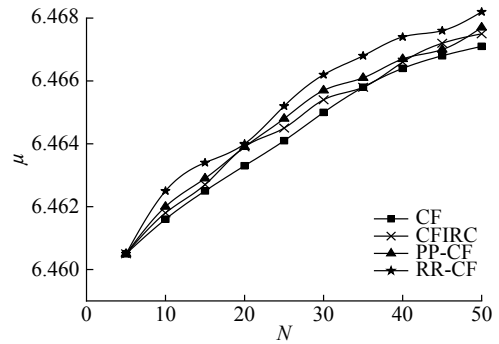


图 7 采用评论挖掘的不同协同过滤算法的多样性对比

Fig.7 Comparison of diversity values between different collaborative algorithms based on review mining

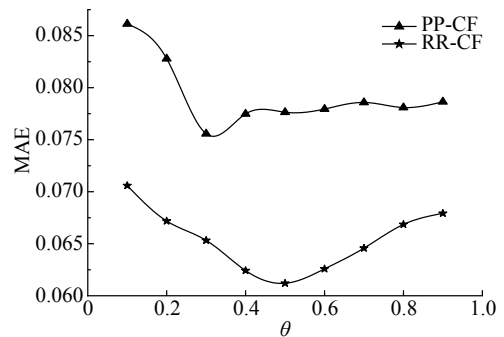


图 8 不同权重因子对 MAE 值的影响

Fig.8 Effect of different weighting factors on MAE value

例也相应增加。

4) 讨论权重因子 θ 对召回率的影响。

通过 2) 的实验得到, 在实验数据集中, RR-CF 算法最近邻居数的最优值为 30, PP-CF 算法最近邻居数的最优值为 45. 取算法 RR-CF 和 PP-CF 各自最优的最近邻居数, 讨论 θ 对召回率 R 的影响, 实验结果如图 9 所示。

由图 9 可以看出, 随着权重因子 θ 的增加, RR-CF 算法的召回率呈先上升、后下降的趋势, 在权重因子 θ 取值为 0.1~0.3 时, PP-CF 算法的召回率要优于 RR-CF 算法, 而在权重因子取值 0.4~0.9 时, RR-CF 算法的召回率要优于 PP-CF 算法, 在权

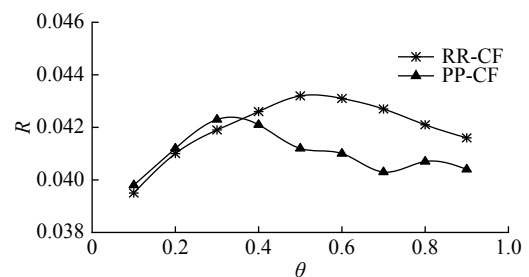


图 9 不同权重因子对召回率的影响

Fig.9 Effect of different weight factors on recall rate

重因子 $\theta=0.5$ 时, RR-CF 算法的召回率最大, 为 4.32%. 而在权重因子 $\theta=0.3$ 时, PP-CF 算法的召回率最大, 为 4.23%.

4 结 语

本研究采用评论挖掘修正用户评分的改进协同过滤算法, 通过挖掘分析评论的情感态度来修正评分, 突破整数评分的限制, 将整数评分细化到小数层面, 提高了评分的区分度, 解决了评分过于集中的问题; 同时使得修正后的评分更加接近用户的真实评分意愿, 提高了评分的可信度, 解决了整数评分可信度不高的问题. 实验结果表明, 所提算法的推荐结果准确度优于现在流行的采用评论挖掘的算法, 为目前各大电子商务网站存在的评分数据过于集中、数据可靠性低, 导致推荐效果较差的问题提供了一个可行的解决方案, 具有较高的实用价值.

本研究仍然存在一些可以改进的地方, 今后的工作将从以下几个方面展开: 1) 寻找更好的情感词评价算法, 使得情感词的评价更加规范化. 本文只是简单地评价情感词的极性并作评分, 今后还可以考虑修饰情感词的语气词, 使得情感词的评分更加标准化, 从而提高预测准确度. 2) 提高分词算法的准确度. 本文使用的分词算法是基于所有文档的 ICTCLAS 分词算法, 语义库包含了很多词语, 但是回归到具体产品, 今后可以建立产品自身特有的语义库, 从而加快分词算法的分词速度和准确度.

参考文献 (References):

- [1] YU X, LIU Y, HUANG X, et al. Mining online reviews for predicting sales performance: a case study in the movie domain [J]. **IEEE Transactions on Knowledge and Data engineering**, 2012, 24(4): 720-734.
- [2] TAN Y, ZHANG Y, ZHANG M, et al. A unified framework for emotional elements extraction based on finite state matching machine [M] // **Natural Language Processing and Chinese Computing**. Berlin Heidelberg: Springer, 2013: 60-71.
- [3] ZHANG Y, LAI G, ZHANG M, et al. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis [C] // **Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval**. Gold Coast: ACM, 2014: 83-92.
- [4] KIRAN M V K, VINODHINI R E, ARCHANAA R, et al. User specific product recommendation and rating system by performing sentiment analysis on product reviews [C] // **4th International Conference on Advanced Computing and Communication Systems (ICACCS)**. Coimbatore: IEEE, 2017: 1-5.
- [5] CHEN G, CHEN L. Augmenting service recommender systems by incorporating contextual opinions from user reviews [J]. **User Modeling and User-Adapted Interaction**, 2015, 25(3): 295-329.
- [6] XU X, DUTTA K, DATTA A. Functionality-based mobile app recommendation by identifying aspects from user reviews [J]. **Journal of the Association for Information Science and Technology**, 2018, 69(2): 242-255.
- [7] RESCHKE K, VOGEL A, JURAFSKY D. Generating recommendation dialogs by extracting information from user reviews [C] // **Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics**. Sofia: ACL, 2013: 499-504.
- [8] GANU G, MARIAN A, ELHADAD N. Ursa-User review structure analysis: understanding online reviewing trends [R/OL]. Rutgers DCS Technical Report, No.668, April 2010. http://spidr-ursa.rutgers.edu/resources/TR_LRE.pdf.
- [9] GANU G, KAKODKAR Y, MARIAN A L. Improving the quality of predictions using textual information in online user reviews [J]. **Information Systems**, 2013, 38(1): 1-15.
- [10] MCAULEY J, LESKOVEC J. Hidden factors and hidden topics: understanding rating dimensions with review text [C] // **Proceedings of the 7th ACM conference on Recommender systems**. New York: ACM, 2013: 165-172.
- [11] DEHKORDI Y H, THOMO A, GANTI S. Incorporating user reviews as implicit feedback for improving recommender systems [C] // **The 4th IEEE International Conference on Big Data and Cloud Computing**. Sydney: IEEE, 2014: 455-462.
- [12] 张付志, 刘赛, 李忠华, 等. 融合用户评论和环境信息的协同过滤推荐算法[J]. 小型微型计算机系统, 2014, 35(2): 228-232.
ZHANG Fu-zhi, LIU Sai, LI Zhong-hua, et al. Collaborative filtering recommendation algorithm incorporating user's reviews and contextual

- information [J]. **Journal of Chinese Computer Systems**, 2014, 35(2): 228–232.
- [13] 谭云志, 张敏, 刘奕群, 等. 基于用户评分和评论信息的协同推荐框架[J]. 模式识别与人工智能, 2016, 29(4): 359–366.
- TAN Yun-zhi, ZHANG Min, LIU Yi-qun, et al. Collaborative recommendation framework based on ratings and textual reviews [J]. **Pattern Recognition and Artificial Intelligence**, 2016, 29(4): 359–366.
- [14] 李伟霖, 王成良, 文俊浩. 基于评论与评分的协同过滤算法[J]. 计算机应用研究, 2017, 34(2): 361–364.
- LI Wei-lin, WANG Cheng-liang, WEN Jun-hao. Collaborative filtering recommendation algorithm based on reviews and ratings [J]. **Application Research of Computers**, 2017, 34(2): 361–364.
- [15] 王伟军, 宋艳秋, 张婷婷, 等. 基于话语标记理论的在线商品评分修正方法[J]. 情报学报, 2016, 35(4): 358–368.
- WANG Wei-jun, SONG Yan-qiu, ZHANG Ting-ting. A correction method for score deviation of online product reviews based on discourse markers theory [J]. **Journal of The China Society for Scientific and Technical Information**, 2016, 35(4): 358–368.
- [16] 郑修猛, 陈福才, 黄瑞阳, 等. 面向协同推荐的评论文本情感打分机制研究[J]. 信息工程大学学报, 2017, 18(4): 464–469.
- ZHENG Xiu-meng, CHEN Fu-cai, HUANG Rui-yang, et al. Sentiment rating mechanism research on review text for CF recommendation systems [J]. **Journal of Information Engineering University**, 2017, 18(4): 464–469.
- [17] YANG H, SUH Y. Sentiment analysis of online customer reviews for product recommendation: comparison with traditional CF-based recommendation [C] // **The Korea Society of Management Information Systems**. Seoul: KMIS, 2015: 801–805.
- [18] WU Y, ZHANG Q, HUANG X, et al. Phrase dependency parsing for opinion mining [C] // **Proceedings of the 2009 Conference on Empirical Methods in Natural Language**. Singapore: EMNLP, 2009: 1533–1541.
- [19] 扈中凯, 郑小林, 吴亚峰, 等. 基于用户评论挖掘的产品推荐算法[J]. 浙江大学学报: 工学版, 2013, 47(8): 1475–1485.
- HU Zhong-kai, ZHENG Xiao-lin, WU Ya-feng, et al. Product recommendation algorithm based on users' reviews mining [J]. **Journal of Zhejiang University: Engineering Science**, 2013, 47(8): 1475–1485.
- [20] PANG B, LEE L. Opinion mining and sentiment analysis [J]. **Foundations and Trends in Information Retrieval**, 2008, 2(1-2): 1–135.
- [21] 朱郁筱, 吕琳媛. 推荐系统评价指标综述[J]. 电子科技大学学报, 2012, 41(2): 163–175.
- ZHU Yu-xiao, LU Lin-yuan. Evaluation metrics for recommender systems [J]. **Journal of University of Electronic Science and Technology of China**, 2012, 41(2): 163–175.