文章编号：1003-0077(2017)02-0194-10

基于情感分析和 LDA 主题模型的协同过滤推荐算法

彭 敏，席俊杰，代心媛，何炎祥

（武汉大学 计算机学院，湖北 武汉 430072）

摘 要：协同过滤推荐算法通常基于物品或用户的相似度来实现个性化推荐，但是数据的稀疏性往往导致推荐精度不理想。大多数传统推荐算法仅考虑用户对物品的总体评分，而忽略了评论文本中用户对物品各个属性面的偏好。该文提出一种基于情感分析的推荐算法 SACF(reviews sentiment analysis for collaborative filtering)，该算法在经典的协同过滤推荐算法的基础上，考虑评论文本对相似度计算的影响。SACF 算法利用 LDA 主题模型挖掘物品潜在的 K 个属性面，通过用户在各个属性面上的情感偏好计算用户相似度，从而构建推荐模型。基于京东网上评论数据集的实验结果表明，SACF 算法不但可以有效地改善传统协同过滤推荐算法中数据稀疏性的问题，而且提高了推荐系统的精度。

关键词：推荐系统；协同过滤；LDA；情感分析

中图分类号：TP391 **文献标识码**：A

Collaborative Filtering Recommendation Based on Sentiment Analysis and LDA Topic Model

PENG Min, XI Junjie, DAI Xinyuan, HE Yanxiang

(School of Computer, Wuhan University, Wuhan, Hubei 430072, China)

Abstract: Collaborative filtering achieves personalized recommendation based on the similarity between items or users. However, the data sparseness affects the calculation of similarity, leading to a low recommendation accuracy. Most of the traditional recommendation algorithms only consider the rate matrix between users and items, while ignoring the item reviews generated by users, that offer valuable information about the user's preferences to different attributes of the items. In this paper, we proposed a novel recommendation algorithm, called SACF (sentiment analysis collaborative filtering), which considers the impact of the review texts on the prediction of final score of items. By incorporating LDA topic model, SACF can extract K latent attribute aspects of the items and compute the user similarity according to the sentiment tendency in such attribute aspects. Our experimental results on Jingdong review dataset demonstrate that, the proposed method can not only alleviates the problem of data sparseness in collaborative filtering scheme, but also improves the recommendation accuracy.

Key words: recommender system; filtering recommendation; LDA; sentiment analysis

1 引言

随着互联网的普及和信息技术突飞猛进的发展，信息的过度丰富给信息筛选带来了巨大的挑战。个性化推荐系统的出现为用户提供了一种解决信息过载的工具。协同过滤算法是目前应用最广泛的推荐算法[1-2]，主要包括基于用户[3]和基于物品[4]的协同过滤。其基本原理是基于相似性，通过度量共同评分向量的相似度来寻找相似的用户或物品。然而，数据缺乏使得用户共同评论的物品较少，导致计算相似度不准确，最终影响推荐精度。

一方面，传统的协同过滤算法根据用户的总体评分，使得用户或者物品的相似度度量不够精确。以手机评论为例，用户 $u1$ 和 $u2$ 对同一款手机的总体评分均为 4 分，$u1$ 可能喜欢该款手机的外观而认

收稿日期：2015-09-06 定稿日期：2016-03-23

基金项目：国家自然科学基金(61472291, 61303115)

为价格过高，而 $u2$ 可能不太满意这款手机的外观但认为价格合适。虽然 $u1$ 和 $u2$ 对该款手机的总体评分相同，但是他们的偏好却不尽相同。由此可见评论文本中隐含着大量有关用户偏好的信息，若能充分利用这些信息，通过对评论文本进行情感分析，找到相似用户，可提高推荐精度。

另一方面，当数据集比较稀疏时，由于用户共同评分的物品较少，仅依赖用户对物品的总体评分来计算相似度也会导致相似度计算不准确。考虑到用户在评论中往往会对物品的多个属性面进行评价，本文将基于总体评分向量的相似度计算转化为基于属性面层次的相似度计算。通过扩展向量维度，在一定程度上解决数据的稀疏性问题，以提高推荐精度。

为此，本文提出一种新的协同过滤推荐算法 SACF，该算法采用一种新颖的基于情感分析和 LDA 主题模型的方法来预测用户对物品属性面的评分，并将用户基于属性面评分的相似度和基于总体评分的相似度进行融合，从而找到最相似的 N 个用户进行推荐。

本文工作的主要贡献包括以下几点：

（1）将传统协同过滤推荐算法中基于用户对物品总体评分的相似度计算转化为基于属性面层次的相似度计算，提高了相似度计算的准确性。

（2）提出了一种新颖的基于情感分析和 LDA 主题模型的算法 SACF，该算法通过提取物品的属性面，以及预测用户对属性面的评分来寻找相似用户。

（3）通过相关实验分析，证明本文所提出的推荐方法从一定程度上解决了数据稀疏性的问题，提高了推荐的精度。

2　相关工作

2.1　推荐系统

推荐系统（recommender system）大致可分为两大类：基于内容的推荐方法和协同过滤推荐。基于内容的推荐方法[5]是根据内容的相似性来发现相似物品。其优点是简单、有效，但是基于用户历史信息进行推荐，不能发掘新的用户感兴趣的信息。协同过滤推荐则是基于相似用户具有相似喜好这一假设来进行推荐。该类方法最大的问题是，在高维空间中基于稀疏数据计算的相似度并不准确。为解决数据稀疏性的问题，很多工作对传统的协同过滤推荐算法进行改进。

文献[6]首先根据物品之间的相似性初步预测用户对未购买物品的评分，然后再借助传统的协同过滤推荐算法进行推荐。该方法虽然能够缓解数据稀疏性的问题，但是初步预测评分的不准确将直接导致最终的推荐精度不高。文献[7]首先根据相似性项目进行聚类，然后在初始聚类的基础上进行交叉迭代调整，最终使得聚类簇达到较为稳定的状态，从而寻找到目标项目最近邻居并产生推荐。该方法虽然在一定程度上提高了推荐精度，但是聚类算法的准确度将成为推荐精度的瓶颈。

近年来，采用 LDA 主题模型和评论文本进行推荐的研究逐渐增多，但很少有研究将两者结合起来的。现有的利用 LDA 进行推荐的研究多数是针对文档进行推荐[8]，而不是将 LDA 应用于物品的评论文本。跟本文工作比较相近的是文献[9]和文献[10]，文献[9]依据隐语义模型[11]的基本原理，使用 LDA 直接将用户对物品的评分矩阵进行分解，通过降维的方法来提高相似度的计算精度。但是降维处理往往导致信息丢失，在物品空间维度很高或者用户评分矩阵比较稀疏的情况下，降维效果难以得到保证。文献[10]通过对评论文本进行情感分析，挖掘用户对物品各属性的情感偏好，然后通过聚集或者平均的方法预测用户对物品的总体评分，最后利用协同过滤的方法进行推荐。文献[12]则是在情感分析的基础上采用回归的方法预测总体的评分值。这些方法虽然利用评论文本缓解了数据稀疏性的问题，但是简单地通过聚合或者回归的方法来预测总体评分而忽略了用户在不同属性面上的相似性，导致相似度计算不够准确，进而影响了推荐精度。本文充分利用评分文本提供的信息，将相似度的计算转化到属性层面，使得相似度计算更加精确。

2.2　情感分析

现有的属性词和情感词提取方法主要分为有监督的学习方法[13-14]、半监督的学习方法[15]和无监督的学习方法[16-17]。文献[18]是最早也是当前最流行的属性词、情感词提取方法，它主要是基于关联规则来挖掘属性词和情感词。尽管属性词和情感词提取很早就开始研究，但是情感分析在推荐算法中的研究才刚刚起步。情感分析在推荐系统中的应用主要包括属性词和情感词的提取[14,18-19]及对属性面的评分预测[20-22]。文献[20]根据文献[19]的方法提取出评论文本中的属性词和情感词，然后基于统计方法预测用户对物品属性面的评分。该方法将一个

句子中所有的情感词都作为该句子中属性词的修饰词，导致计算的评分值不准确。

文献[21]着重提取属性词—情感词对，它首先基于“中文文本中副词后面紧跟着形容词”这一规则，利用副词的种子词库来提取情感词；然后通过预设的属性词相似度阈值进行过滤，用高频率属性词替代低频率属性词，得到属性面-情感词对；最后利用情感词的情感极性来预测用户对属性面的评分分值。该文中提取属性词—情感词对的方法存在一些缺陷：首先，在利用种子副词提取情感词以及根据已有情感词和属性词提取新的属性词和情感词时，它只考虑了词汇的位置信息，即词汇之间的距离，这会将那些原本不存在修饰关系的词对也提取出来，导致最终提取出的词对不够准确。其次，若一个情感词前面没有任何副词，该方法将不能提取出这样的情感词。

针对文献[21]中方法存在的缺陷，本文在该方法的基础上进行改进，提出一种新的情感词提取方法。首先，扩充情感词的种子词库，使得前面没有副词的情感词不会被遗漏；其次，利用依存句法分析，提取属性词-情感词对，从而过滤掉不相关的词对，提高了词对的提取精度；最后，在评分预测阶段，考虑程度副词和否定副词对属性词分值的影响，使得属性词的评分更加准确。

此外，本文利用 LDA 主题模型来获取物品潜在的属性面，并结合用户关注度，预测用户对物品属性面的评分。实验表明，本文提出的方法能够更准确合理地预测用户对属性面的评分。

3 问题描述

本节首先给出属性面和情感词的定义，然后形式化描述本文的主要任务。

属性面：在线评论时用户常使用不同的词汇描述相同的产品特征，如属性词“颜色”和“色彩”都是关于手机的颜色外观的特征描述词，属于同一个属性面。本文使用 LDA 主题模型将这些语义相近的属性词汇归纳为同一个属性面。

情感词：情感词表达了用户对物品属性面积极或者消极的情绪、态度和情感。例如，在评论语句“手机的外观很好看”中，“好看”一词表达了用户积极的情感极性。

属性词-情感词对：一个属性词和在评论句子中表达该属性词极性的情感词组成的词对。

基于以上定义，本文的问题可以描述为，对于给定的 N 个用户 $U=\{u_1,u_2,\cdots,u_N\}$ 对 L 个物品 $R=\{r_1,r_2,\cdots,r_L\}$ 的 M 个评论文本的集合 $D=\{d_1,d_2,\cdots,d_M\}$，首先，从每个评论文本的每个句子中抽取属性词和情感词，并根据属性词的极性得到用户对属性词评分的四元组集合{userid，itemid，feature，value}；然后利用 LDA 主题模型发现物品潜在的 K 个属性面 $F=\{f_1,f_2,\cdots,f_K\}$；其次，计算用户对属性面的评分四元组{userid，itemid，feature-aspect，value}；最后，构建用户相似度模型，根据用户最相似的 K 个邻居进行个性化推荐。

4 SACF 推荐算法

本文中提出的 SACF 推荐算法主要包括两个部分：(1)属性面的评分预测。对评论文本进行情感分析，提取属性词—情感词对，并预测用户对属性面的评分。(2)物品推荐。根据(1)中用户对属性面的评分计算用户之间的相似度，并将其与用户对物品总体评分的相似度融合，得到用户之间的综合相似度，进而根据最相似的 K 个邻居用户进行物品推荐。推荐算法流程图如图 1 所示。

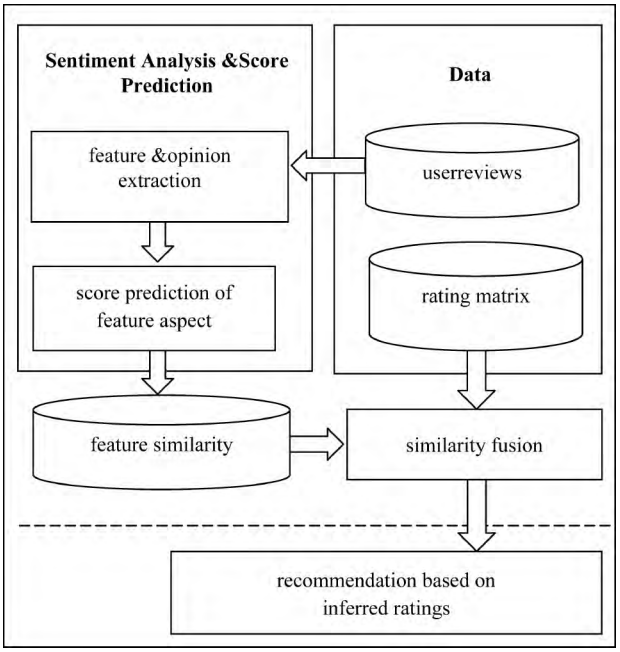

图 1 推荐算法流程图

4.1 概率主题模型

概率主题模型可用于识别大规模文档集中潜在的主题信息，这些主题能够很好地被理解。其中，狄

利克雷分布(latent dirichlet allocation，LDA)是目前被广泛应用的一种概率主题模型。

LDA 是一种无监督的机器学习技术，它是一个“文档—主题—词汇”的三层贝叶斯模型，其中每一篇文档 $d \in D$ 表示为 K 个主题的概率分布 θ_d，每一个主题 k 又表示成 N 个词所构成的概率分布 ϕ_k。

LDA 主题模型包含每个主题的词汇分布概率 ϕ_k、每个文档的主题分布概率 θ_d，以及每个词汇的主题分配序列 $z_{d,j}$，参数 $\{\theta,\phi\}$ 和主题分配序列 z 可通过吉布斯采样获得。LDA 主题模型以如下概率产生文本集合 D，如式(1)所示。

$$p(D \mid \theta,\phi,z) = \prod_{d\in D}\prod_{j=1}^{N_d}\theta_{z_{dj}}\phi_{z_{dj},w_{dj}}, \quad (1)$$

其中，$\theta_{z_{dj}}$ 表示特定主题的产生概率，$\phi_{z_{dj},w_{dj}}$ 表示词汇 w_{dj} 属于该主题的概率。

本文采用 LDA 主题模型将情感分析中提取出来的属性词聚集成属性面。LDA 主题模型建立在文档的基础上，因此，这里的首要任务是构建文档集合。考虑到需要获得每个用户对各个属性面的关注度，因此本文首先将用户 u 的所有评论文本集合作为主题模型中的文档 d_u，并使用从情感分析中提取出的属性词构建文档的空间向量。然后通过 LDA 主题模型挖掘出潜在的 K 个属性面，最终获得文档—主题分布矩阵 θ 和主题—词汇分布矩阵 ϕ。

4.2　属性面评分预测

在本文中，属性面的评分预测方法主要包括三个步骤：(1)从评论文本中提取属性词—情感词对；(2)根据情感词和修饰情感词的副词预测对应属性词的分值；(3)利用 LDA 主题模型将属性词聚集成潜在属性面，并预测属性面的评分。下面我们将详细阐述这三个步骤。

4.2.1　属性词—情感词对提取

在评论文本中，不同用户采用不同的属性词和情感词来表达观点。采用人工标注的提取方法显然是不合适的，这将耗费大量的时间并且召回率很低。本文受到文献[21]中属性词—情感词对提取方法的启发，并在原有方法的基础上进行改进，提出一种新的属性词—情感词对提取方法。

文献[21]中，作者依据“紧跟在副词后面的形容词是情感词”这一规则，利用副词的种子词库提取情感词。然而，这一规则将会漏掉那些前面没有副词的情感词，致使提取出的情感词-属性词对召回率下降。此外，文献[21]中提取情感词和属性词的方法是基于词距的，即在一个句子中根据两个词汇之间的距离来判断他们是否具有依赖关系，这样往往会把那些没有依赖关系或者具有不正确依赖关系的词对找出来，从而导致属性词—情感词对的提取准确率下降。针对这两点缺陷，本文在提取过程中加入了情感词的种子词库，首先采用句法分析器对句子进行依存句法解析，然后根据所需要的依存关系进行过滤。在整个提取过程中主要涉及两种依存关系：一是副词修饰形容词的状中关系(ADV)，二是形容词修饰名词的定中关系(ATT)。词汇极性判断的方法[23-24]有很多，本文采用文献[23]中的方法对新提取的情感词进行极性判断。

4.2.2　属性词的分值预测

经过上一个步骤中属性词—情感词对的提取，我们得到用户 u_i 对物品 r_j 的评论四元组{userid，itemid，feature，opinion}，然后依据情感词的极性和修饰情感词的副词来预测用户对属性词的评分。由于我们需要计算用户对属性词的评分，所以需要设置情感词极性的初始分值，即情感基数。情感基数可以是任意不为 0 的实数(在后文的计算中将对其进行归一化处理，不影响最终的评分预测)，这里为了方便计算，我们设定情感词的情感基数为 1，即积极的情感词为 1，消极的情感词为 −1。很多情感分析的研究[20-21]在进行情感分值预测的时候仅考虑了情感词而忽略了副词的修饰成分，或者简单地将副词划分为“高、低”两个等级。而文献[25]的研究工作，考虑到程度副词和否定副词的影响，将程度副词分为七个等级，并根其表达的强烈程度设定相应的副词修饰百分比。副词修饰百分比即在初始情感基数基础上增加的比例，属性词分值＝(1＋副词修饰百分比)×情感基数，相应的副词修饰百分比如表 1 所示。

表 1　程度副词修饰百分比

程度副词	举例说明	修饰百分比/%
Most	最高 非常 势必……	70
Very	大力 大量 很……	50
Really	越来越 日益……	30
Pretty	基本 适当 刚好……	10
Somewhat	稍显 轻微 略有……	0
Slightly	有些 有点 较为……	−10
否定副词	不 没 并非……	80

最终属性词的分值由情感基数和对应副词的修饰百分比相乘得到，考虑到程度副词和否定副词以

及两种副词的组合形式对情感词的修饰，我们得到如表2所示的计算方法。

表2 属性词打分计算方法

修饰类型	计算方式	举例说明	情感分值
S=PW	SV(PW)	好	1.0
S=NW	SV(NW)	差	−1.0
S=Neg+PW	P(Neg)×SV(PW)	不好	−0.8
S=Neg+NW	P(Neg)×SV(NW)	不差	0.8
S=Neg+Neg+PW	P(Neg)×P(Neg)×SV(PW)	不是不好	0.64
S=Neg+Neg+NW	P(Neg)×P(Neg)×SV(NW)	不是不差	−0.64
S=Int+PW	(1+P(Int))×SV(PW)	特别好	1.7
S=Int+NW	(1+P(Int))×SV(NW)	特别差	−1.7
S=Neg+Int+PW	P(Neg)×(1−P(Int))×SV(PW)	不是特别好	−0.24
S=Neg+Int+NW	P(Neg)×(1−P(Int))×SV(NW)	不是特别差	0.24
S=Int+Neg+PW	(1+P(Int))×P(Neg)×SV(PW)	特别不好	−1.36
S=Int+Neg+NW	(1+P(Int))×P(Neg)×SV(NW)	特别不差	1.36

注：(1) PW为正向情感词，NW为负向情感词，SV为情感词的情感打分，Neg为否定副词，Int为程度副词，P为副词的修饰百分比。
(2) SV(好)=0.8，SV(差)=−0.8，P(很)=0.5，P(特别)=0.7。

需要指出的是当程度副词和否定副词出现顺序不同时，计算方法有所差异。当否定副词出现在程度副词之前时，实际上削弱了否定副词带来的负面效应。因此在这种情况下我们将程度副词的修饰百分比进行转向；而当程度副词出现在否定副词之前时，则直接按副词出现顺序叠加副词修饰效果即可。在如表2所示的例子中，我们依据表2的打分规则得到“不是特别好”的情感打分为0.24，“特别不好”的情感打分为−1.36，这是符合实际生活中的情感表达倾向的。给定评论四元组集合{userid, itemid, feature, opinion}，最终求得用户 u_i 对物品 r_j 的 w_n 属性词的分值为 S_{ijn}。

4.2.3 属性面的分值预测

本文将提取出的属性词作为LDA主题模型中的特征词汇，利用LDA算法将相关词汇归属到相应的主题，从而属性词被聚集成 K 个属性面，其中每个面表现为属性词的概率分布，从而将属性词和属性面相关联，然后根据主题-词汇分布得到用户 u_i 对物品 r_j 的 f_k 属性面的评分，如式(2)所示。

$$V_{ijk} = \sum_{n=1}^{N} S_{ijk}\phi_{kn} \tag{2}$$

其中，ϕ_{kn} 为主题—词汇分布中词汇 w_n 属于主题 f_n 的概率。

若一个用户在评论文本中对某个属性面的评论越频繁，表明该用户越关注该属性面，因而，在预测用户对物品属性面的评分中，若能考虑该用户关注度的影响，将会使得预测结果更准确，我们在实验部分也给出了相关证明。事实上，LDA主题模型中得到的文档—主题的分布矩阵 θ_{ik} 即用户 i 对物品第 k 个属性面的关注度。因此，在式(1)的基础上，考虑用户关注度对属性面评分的影响，得到用户 u_i 对物品 r_j 的 f_k 属性面评分，如式(3)所示。

$$S_{ijk} = \theta_{ik} V_{ijk} \tag{3}$$

其中，θ_k 为文档 d_{ui} 在主题 k 上的分布概率，即用户 u_i 对 f_k 属性面的关注度。

整合式(2)和式(3)，得到用户 u_i 对物品 j 的 f_k 属性面综合评分，如式(4)所示。

$$S_{ijk} = \theta_{ik} \sum_{n=1}^{N} V_{ijk}\phi_{kn} \tag{4}$$

4.3 用户相似度计算

在上文中已经得到用户 u 对物品 j 的第 k 属性面的评分四元组集合 $\{u, j, k, S_{u,j,k}\}$，本文使用余弦相似度计算用户 u 和用户 v 在评论文本上的相似度如式(5)所示。

$$\mathrm{sim}_{\mathrm{feature}}(u,v) = \frac{\sum_{j\in(R_u\cap R_v)}\sum_{k\in(K_{uj}\cap K_{vj})} S_{ujk}S_{vjk}}{\sqrt{\sum_{j\in R_u}\sum_{k\in K_{uj}} S_{ujk}^2}\sqrt{\sum_{j\in R_v}\sum_{k\in K_{vj}} S_{vjk}^2}}, \tag{5}$$

其中，R_u 表示用户 u 的评论物品集合，K_{uj} 表示用户 u 对物品 j 评论的属性面集合，$K_{uj} \cap K_{vj}$ 表示两用户在物品 j 上评论的属性面的交集。

尽管评论文本中含有丰富的表达用户情感倾向的语义信息，但是并非所有文本信息都有价值。例如某些用户在评论中并没有显露出自己对物品属性面的情感偏好，对于这类用户，总体评分信息往往比评论文本更有价值。考虑到该因素，依据传统推荐方法的基本思想，我们同时考虑不同用户在总体评分上的相似性，其计算方式如式(6)所示。

$$\mathrm{sim}_{\mathrm{score}}(u,v)=\frac{\sum_{j\in(R_u\cap R_v)}W_{uj}W_{vj}}{\sqrt{\sum_{j\in R_u}W_{uj}^2}\sqrt{\sum_{j\in R_v}W_{vj}^2}},\tag{6}$$

其中 W_{uj} 表示用户 u 对物品 j 的总体评分值。

最后，将用户在评论文本上的相似度 $\mathrm{sim}_{\mathrm{feature}}(u,v)$ 与用户的总体评分相似度 $\mathrm{sim}_{\mathrm{score}}(u,v)$ 融合来度量用户之间的综合相似度 $\mathrm{sim}(u,v)$，如式(7)所示。

$$\mathrm{sim}(u,v)=\gamma\cdot\mathrm{sim}_{\mathrm{feature}}(u,v)+(1-\gamma)\cdot\mathrm{sim}_{\mathrm{score}}(u,v),\tag{7}$$

其中，γ 为两类相似度之间的平衡参数。

4.4　算法描述

最后，本文提出的基于情感分析和 LDA 主题模型的协同过滤推荐算法的整体过程呈现在算法 1 中。

算法 1：SACF 算法

输入：用户评分矩阵，用户对物品的评论文本集合
输出：推荐集合
步骤：
(1) 对评论文本进行情感分析，提取出所有的属性词-情感词对；
(2) 利用 LDA 主题模型，生成物品的 K 个潜在属性面，并结合公式(1)中的属性词-情感词对，预测用户对物品属性面的评分；
(3) 根据公式(2)中用户对物品属性面的评分，计算用户相似度矩阵；
(4) 融合用户的评分相似度，计算最终的用户相似度矩阵；
(5) 对任意用户 u，利用步骤(4)中的相似度矩阵选择与其最相似的 K 个邻居 N_k，依据式(8)对未评分物品 i 的评分进行预测。

$$r_{ui}=\frac{\sum_{k\in N_k}\mathrm{sim}(u,k)\cdot W_{ki}}{\sum_{k\in N_k}\mathrm{sim}(u,k)}\tag{8}$$

5　实验结果与分析

5.1　实验设置

为验证本文提出的 SACF 算法的有效性，文中使用中国知名的电商网站京东(www.jd.com)的手机评论数据集进行实验。该数据集总共包括 109 691 条评论信息，过滤掉评论数小于 5 的用户，最终得到 2 887 个用户对 868 个物品的 33 778 条评论。每条评论包括用户对物品的总体评分值(介于 1～5 分之间)和评论文本。实验采取五折交叉验证进行，每次采用约 80% 的原数据集作为训练集，20% 为测试集，重复五次后取平均结果。

此外，本文采用中国科学院的分词工具 ICTCLAS① 对评论文本进行分词和词性标注，采用哈尔滨工业大学 LTP[26] 进行句法分析。本文中副词和形容词的种子词库均来自于中国知网。

本文将 SACF 算法和以下基准推荐算法进行比较：

(1) CF(collaborative filtering)：传统的基于用户的协同过滤推荐算法。

(2) LDA-CF[8]：使用 LDA 主题模型对评分矩阵进行降维，从而预测用户对物品的评分。

(3) RI-CF(rating inference collaborative filtering)[12]：将评论文本进行情感分析，预测出用户在物品各个属性面上的评分，再使用其提出的回归方法计算用户对物品的总体评分。

(4) PORE(preference and opinion－based recommendation algorithm)[21]：通过分析评论文本中用户的偏好，结合物品和特征属性之间的关系，从而预测用户对物品的评分。

此外，为验证融合用户的总体评分相似度，以及用户对物品属性面关注度对推荐算法的影响，我们从 SACF 算法衍生出两组对比方法：

(5) SACF-WR(sentiment analysis collaborative filtering without rating)：仅考虑用户对属性面评分相似度的协同过滤，忽略 SACF 算法中用户对物品总体评分的影响(用户最终相似度由式(5)产生)。

(6) SACF-WC(sentiment analysis collaborative filtering without concern)：不考虑用户关注度的协同过滤方法，在计算用户对属性面评分时忽略

① http://ictclas.nlpir.org/

关注度因子(式(3)中 θ_{ik})的影响。

5.2 评估指标

在推荐系统的评估标准中,最常用的是平均绝对误差(mean absolute error,MAE),该指标可以直观地对推荐质量进行度量,并被大多数推荐系统相关研究所采用。本文采用平均绝对误差 MAE 衡量预测评分值的准确性,如式(9)所示。

$$\mathrm{MAE}=\frac{\sum_{i=1}^{N}|p_i-q_i|}{N} \tag{9}$$

其中,$\{p_1,p_2,\cdots,p_N\}$为推荐算法预测的用户评分集合,$\{q_1,q_2,\cdots,q_N\}$为实际的用户评分集合。MAE 值越小,表明推荐算法预测的用户评分与实际的用户评分之间的误差越小,推荐质量越高;反之,表明推荐质量越低。

5.3 实验结果分析

5.3.1 情感词—属性词对提取

为了验证本文提出的情感词-属性词对提取方法(opinion feature extraction)的有效性,我们从数据集中随机选取 3 000 个评论文本,然后人工标注出文本中的情感词和属性词,最终得到 122 个属性词和 153 个情感词。我们将 HU 和 LIU[18] 和 LIU 和 HE[21] 的方法作为基准方法,与本文的情感词-属性词对提取方法进行比较。提取属性词和情感词的准确率(提取出正确的情感词和属性词的百分比)、召回率(提取出正确的情感词和属性词占标注词汇的百分比),以及 F 值 = (2 × precision × recall)/(precision + recall)如表 3 和表 4 所示。

表 3 情感词提取结果

算 法	准确率/%	召回率/%	F 值
opinion feature extraction	53.77	62.45	0.58
LIU and HE(2013)	53.28	45.21	0.49
HU and LIU(2004)	52.12	38.26	0.44

表 4 属性词提取结果

算 法	准确率/%	召回率/%	F 值
opinion feature extraction	64.33	27.18	0.38
LIU and HE(2013)	47.23	28.41	0.35
HU and LIU(2004)	21.57	48.36	0.29

从表 3 可以发现,尽管本文方法和对比方法提取情感词的准确率相当,但是本文方法具有较高的召回率,原因是本文方法利用副词将那些没有属性词修饰的情感词也提取出来。除此之外,我们加入了情感词的种子词库,从而保证前面没有副词的情感词不会被遗漏。

同时,表 4 结果表明,本文中提取属性词的方法具有最高的准确率,这是因为对比方法采用基于词距的方式来提取属性词,而本文的算法采用句法分析器对句子进行依存句法解析,根据所需要的依存关系对属性词进行过滤,从而提高了属性词提取的准确率。

5.3.2 参数分析

在本文的推荐算法中,LDA 主题模型中主题的个数 K 和相似度的融合参数 γ 对推荐精度有较大的影响,因此,我们针对这两个参数进行如下分析。

(1) 主题数目的影响。

为探讨 LDA 主题模型中主题个数 K 对实验结果的影响,本文设置主题个数 K 的取值从 5 到 40 进行相关实验,结果如图 2 所示。这里设置用户属性面评分相似度和总体评分相似度的融合参数 $\gamma=0.5$,LDA 主题模型参数 $\alpha=0.5$,$\beta=0.1$,最邻近用户数 $N=20$。结果显示,当主题个数为 25 时效果最好。原因是当 K 的取值过小($K<25$)时,LDA 不能准确地区分评论文本中潜在的属性面,进而不能够详细地挖掘出用户在各个面上的偏好;当 K 的取值过大($K>25$)时,属性面之间的耦合度过高,进而影响各个面上的评分预测精度。此外,随着属性面的增多,用户在属性面上的评分矩阵将变得稀疏,进而

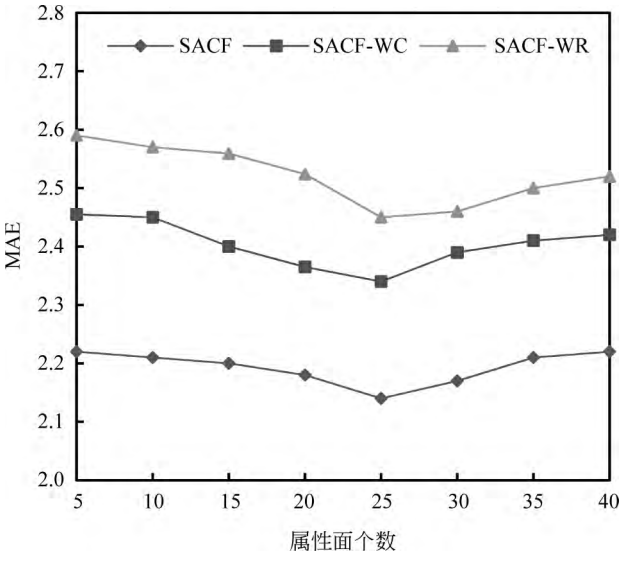

图 2 属性面个数 K 取不同值下的 MAE

影响相似度的计算精度。

（2）融合参数的影响。

为测试融合参数对实验结果的影响，我们将 γ 的值从 0 变化到 1，实验结果如图 3 所示。这里，设定 LDA 主题个数 $K=25$，最邻近用户数 $N=20$。结果显示当 $\gamma=0.9$ 时效果最佳。MAE 之所以不是在 $\gamma=1$ 时取得最小值，是因为在相似度的融合中总体评分的相似度对推荐精度有一定影响，但是从评论文本中通过情感分析得到的用户相似度对算法的影响更大，这进一步证明了本文的观点。

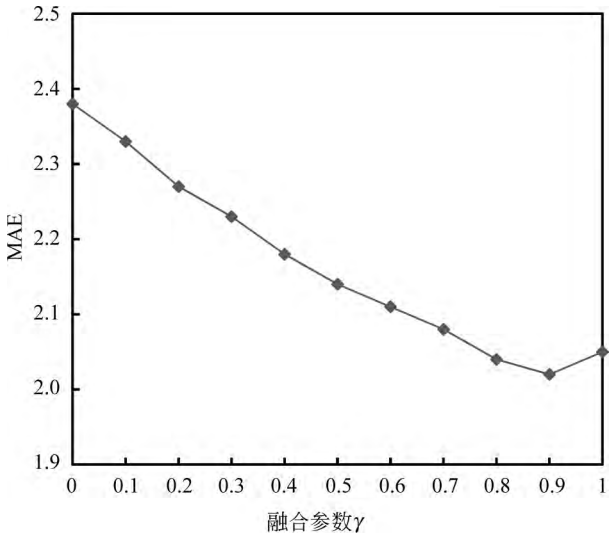

图 3　不同平衡因子 γ 下的 MAE

5.3.3　本文方法和基准方法

在对比实验中，设置 SACF 算法主题数目 $K=25$，融合参数 $\gamma=0.9$，计算出融合相似度后，采用基于用户的推荐方法进行实验，实验结果如图 4 所示。

从图 4 可以看出，当最邻近用户数 N 小于 60 时，三种方法得到的 MAE 值均较高；随着 N 的增加，MAE 值均随之减小；而当 N 大于 120，算法的 MAE 值均趋于稳定。当最邻近用户数 N 相同时，SACF 算法具有最小的 MAE 值，这说明本文提出的方法能够有效地提高推荐质量。而同样采用了 LDA 主题模型，SACF 算法得到的 MAE 值比 LDA-CF 小，说明采用评论文本确实可以提高推荐精度。然而同样利用了评论文本信息的 RI-CF 算法效果却不如 SACF 算法，这说明通过 LDA 主题模型来获取用户在物品各个属性面上的相似度的方法是有效的。

此外，本文将 SACF 衍生得到的两组方法的实验结果进行了比较。通过对比 SACF-WR 和 CF 的

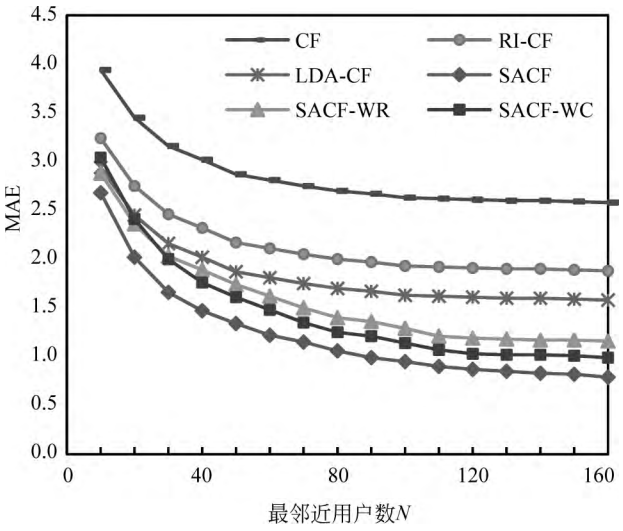

图 4　不同推荐算法的 MAE 值比较

实验结果，可以发现，基于属性面相似度的推荐比传统的基于用户总体评分的协同过滤推荐算法更加准确，这说明基于属性面的相似度度量方法更能够精确地发现相似用户，从而提高了推荐精度。比较 SACF 和 SACF-WR 的实验结果可以看出，融合之后的相似度计算方法比仅使用情感分析相似度方法的推荐效果更好，这说明融合用户总体评分相似度可以提高推荐精度。此外，SACF 和 SACF-WC 的实验结果表明，考虑用户对属性面关注度的影响可提高算法的推荐精度。

由于基准算法 PORE[21] 不属于协同过滤推荐算法这一类别，因此我们将其与本文的算法 SACF 单独进行实验对比。实验中 PORE 的属性特征包括四个评分特征和十个情感特征，设置 SACF 主题数目 $K=25$，融合参数 $\gamma=0.9$，最邻近用户数 $N=160$。五折交叉验证得到的 MAE 值及平均 MAE 值如表 5 所示。

表 5　PORE 和 SACF 的 MAE 值

推荐算法	MAE1	MAE2	MAE3	MAE4	MAE5	平均 MAE
PORE	1.02	1.03	1.05	1.02	1.02	1.028
SACF	0.78	0.80	0.79	0.78	0.79	0.788

实验结果显示，SACF 算法得到的平均 MAE 值较低。可见，对于此类基于文本情感分析的推荐算法，属性特征和情感词的提取对最终的推荐精度有很大影响，而前文中情感分析的实验部分显示，在属性特征和情感词的提取中，SACF 具有较高的 F 值，从而决定了其更好的推荐效果。此外，区别于 PORE 采用词汇的相似度来合并属性词，SACF 采

用LDA主题模型将属性词聚集成更加精确的面，因此提高了推荐精度。

5.3.4 案例分析

为说明本文中LDA主题模型生成的K个属性面的有效性，我们将属性面以属性词的形式进行展示，即根据主题—词汇分布中属性词的分布概率，选择前五个最相关的属性词汇表示每个属性面。最终，从25个属性面中选取最重要十个属性面，结果如表6所示。

表6 属性面词汇分布

属性面1	属性面2	属性面3	属性面4	属性面5	属性面6	属性面7	属性面8	属性面9	属性面10
电池	外观	钱	质量	物流	声音	软件	感觉	信号	图片
待机	屏幕	价格	性价比	硬件	音乐	游戏	按键	CPU	色彩
时间	大屏	性价比	正品	服务	歌曲	功能	手感	内存卡	产品
耗电量	分辨率	功能	系统	质量	质量	Wifi	字体	智能	视频
功能	界面	系统	行货	功能	价格	网	键盘	系统	菜单

从表6可以看出，每个属性面都包含了与该属性特征相关的属性词，我们从属性词的分布大致可以识别出每个属性面所表达的含义。例如，从属性面1的词汇分布中可以看出该属性面与手机电池相关。

6 总结

本文提出了一种基于情感分析的协同过滤推荐算法SACF，该算法将传统的协同过滤中用户对物品总体评分的相似度扩展为融合了属性面和评分的相似度。该算法充分利用了评论文本的丰富信息，通过情感分析和LDA主题模型，挖掘物品潜在的属性面，并预测用户对物品属性面的评分，从而在属性面的层次上计算用户的相似度矩阵。实验结果表明，SACF算法使得用户相似度的计算更加准确，从而提高了推荐精度。此外，SACF算法也在一定程度上解决了数据的稀疏性问题。在下一步工作中，我们将改进情感分析算法，使得属性词-情感词对的提取和属性面评分计算更加准确，从而进一步提高推荐算法的精度。

参考文献

[1] Pu P, Chen L, Hu R. A user-centric evaluation framework for recommender system[C] //Proceedings of the 5th ACM Conference Recommender System. New York: ACM Press, 2011: 157-164.

[2] Knijnenburg B P, Willemsen M C, Gantner Z, et al. Explaining the user experience of recommender system [J]. User Modeling and User-Adapted Interaction, 2012, 22(4): 441-504.

[3] Sarwar B M, Karypis G, Konstan J A, et al. Analysis of recommendation algorithms for ecommerce[C] // Proceedings of the 2nd ACM Conference on Electronic Commerce. New York: ACM Press, 2000: 158-167.

[4] Sarwar B M, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C] //Proceeding of the 10th International Conference on World Wide Web. New York: ACM Press, 2001: 285-295.

[5] Mooney R J, Roy L. Content-based book recommending using learning for text categorization[C] //Proceedings of the ACM international conference on digital libraries. New York: ACM Press, 2000: 195-204.

[6] 邓爱林，朱扬勇，施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(09): 1621-1628.

[7] 王明文，陶红亮，熊小勇. 双向聚类迭代的协同过滤推荐算法[J]. 中文信息学报, 2007, 22(4): 61-65.

[8] Chang T M, Hsiao W F. LDA-based personalized document recommendation[C]// Proceedings of the PACIS 2013, Paper 13.

[9] 廉涛，马军，王帅强等. LDA-CF：一种混合协同过滤方法[J]. 中文信息学报，2014，28(2)：129-135.

[10] C W ki Leung, S C fai Chan, F lai Chung. Integrating collaborative filtering and sentiment analysis: A rating inference approach// Proceedings of the ECAI-Workshop on Recommender Systems, 2006, 62-66.

[11] Hofmann T, Puzicha J. Latent class models for collaborative filtering[C] //Proceedings of the 16th IJCAI, 1999: 688-693.

[12] Ganu G, Kakodkar Y. Improving the quality of predictions using textual information in online user reviews[J]. Information Systems 38(1), 1-15, 2013.

[13] Moraes R, Valiati J F, Gaviao Neto W P. Document-level sentiment classification an empirical comparison

between SVM and ANN[J]. Expert System with Applications, 2013, 40(2): 621-633.

[14] Sayeedunnissa S F, Hussain A R, Hameed M A. Supervised Opinion Mining of Social Network Data Using a Bag-of-Words Approach on the Cloud[C] //Proceedings of 7th International Conference on Bio-Inspired Computing: Theories and Applications (BICTA 2012). India: Springer, 2013: 299-309.

[15] 黄诗琳,郑小林,陈德人. 针对产品命名实体识的半监督学习方法[J]. 北京邮电大学学报, 2013, 36(002): 20-23.

[16] Chen C C, Chen Z Y, Wu C Y. An unsupervised approach for person name bipolarization using principal component analysis [J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(11): 1963-1976.

[17] Paltoglou G, Thelwall M. Twitter, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2012, 3(4): 66.

[18] Hu M, Liu B. Mining and summarizing customer reviews[C] //Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM Press, 2004: 168-177.

[19] Qiu G, Liu B, Bu J, et al. Opinion word expansion and target extraction through double propagation[C] //Computational Linguistics. 2011: 9-27.

[20] Wang Yuanhong, Liu Yang, Yu Xiaohui. Collaborative Filtering with Aspect-Based Opinion Mining: A Tensor Factorization Approach[C] //Proceedings of the IEEE International Conference on Data Mining. Piscataway NJ: IEEE, 2012: 1152-1157.

[21] Liu Hongyan, He Jun, Wang Tingting, et al. Combining user preferences and user opinions for accurate recommendation[J]. Electronic Commerce Research and Applications, 2013, 12(1): 14-23.

[22] 刘丽佳,郭剑毅,周兰江等. 基于 LM 算法的领域概念实体属性关系抽取[J]. 中文信息学报, 2014, 28(6): 216-222.

[23] Liu H, Yang H, Li W, et al. CRO: A system for online review structurization[C] //Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM Press, 2008: 1085-1088.

[24] 宋晓雷,王素格,李红霞等. 基于概率潜在语义分析的词汇情感倾向判别[J]. 中文信息学报, 2011, 25(2): 89-93.

[25] M Taboada, J Brooke, M Tofiloski, et al. Lexicon-based methods for sentiment analysis[J]. Computational Linguistics, 2011, 37(2): 267-307.

[26] Che Wanxiang, Li Zhenghua, Liu Ting. LTP: A Chinise Language Technology Platform [C]//Proceedings of 23rd International Conference on Computational Linguistics: Demonstrations. New York: ACM, 2010: 13-16.

彭敏(1973—),教授,主要研究领域为自然语言处理、信息检索、分布式计算等。
E-mail: pengm@whu.edu.cn

代心媛(1991—),硕士,主要研究领域为自然语言处理、数据挖掘、情感分析。
E-mail: tracy_day@whu.edu.cn

席俊杰(1989—),硕士,主要研究领域为自然语言处理、数据挖掘、推荐系统。
E-mail: xijunjie@whu.edu.cn