

基于用户兴趣变化和评论的协同过滤算法研究

董晨露 柯新生

(北京交通大学经济管理学院 北京 100044)

摘 要 传统协同过滤算法中,随着用户和商品数量的日益增多,用户-项目评分矩阵越来越稀疏。针对这一问题,提出了一种基于用户兴趣变化和评论的协同过滤算法。该算法将用户评论和遗忘曲线引入传统协同过滤算法中,将评论文本作为商品特征描述文本,使用主题模型计算商品主题特征,引入艾宾浩斯遗忘曲线来协同计算用户的评论分布及评论相似度。将用户评论相似度和用户评分相似度相结合,以得到最终的用户相似度,进而对商品评分进行预测。对网络爬取的真实数据进行验证,结果显示该算法能够在稀疏数据集上获得较好的推荐结果。

关键词 协同过滤,稀疏数据集,主题模型,用户兴趣变化,评论相似度

中图分类号 TP301 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.03.033

Study on Collaborative Filtering Algorithm Based on User Interest Change and Comment

DONG Chen-lu KE Xin-sheng

(School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China)

Abstract The user-item rating matrix is becoming more and more sparse with the increasing number of users and commodities in the traditional collaborative filtering algorithm. To solve this problem, a collaborative filtering algorithm based on user interest change and comment was proposed. The algorithm introduces user comment and forgetting curve into the traditional collaborative filtering algorithm. The comment text is used as the text of commodity feature description, the topic model is used to calculate the commodity topic features, and Ebbinghaus's forgetting curve is also introduced for the cooperative computing of user comment distribution and comment similarity. The similarity of user comment and the similarity of user rating are combined to get the final similarity, and then the rating of commodity is predicted. The algorithm was validated by real data crawled over the network. The experimental results show that the proposed algorithm can get better recommendation results in sparse data sets than the traditional collaborative filtering algorithm.

Keywords Collaborative filtering, Sparse data set, Topic model, User interest change, Comment similarity

1 引言

随着信息技术的快速发展,“信息过载”问题日益严重,推荐系统的出现很大程度地改善了这一状况。推荐算法作为推荐系统的核心,受到了国内外学者的广泛关注。在众多的个性化推荐算法中,协同过滤算法目前应用得最成功且最广泛。传统的协同过滤算法一般分为基于项目的协同过滤(Item-based Collaborative Filtering)和基于用户的协同过滤(User-based Collaborative Filtering)^[1-2]。通过用户-项目评分矩阵计算用户相似度或商品相似度,进而对用户未购买过的商品进行评分预测,并将预测评分高的商品推荐给用户。随着推荐系统中用户和商品数量的日益增多,用户-项目评分矩阵越来越稀疏,用户相似度或商品相似度的计算误差也越来越大,难以保证预测评分的准确率。

一般来说,用户在对商品打分的同时,还会填写相应的商品评论,用户在评论信息中往往会提及自己比较关注的商品特征,这是非常有价值的信息资源。将评论信息引入到推荐算法中,利用评论信息计算用户的评论相似度,能够有效提升推荐质量。另一方面,随着时间和环境的变化,用户的兴趣往往是不断变化的,通常情况下,越是最近购买的商品越能够反映用户当下的兴趣所在,而越早购买的商品对用户当下兴趣的反映程度则越低。然而,传统协同过滤算法只将用户-评分矩阵纳入到用户相似度的计算中,而未考虑不同的时间距离对用户兴趣的影响。将基于时间的用户兴趣变化引入到推荐算法中,对不同时间距离的用户兴趣赋予不同的权重,能够使用户兴趣的计算更加准确。

因此,本文提出一种基于用户兴趣变化和评论的协同过滤算法,即将用户评论信息作为商品特征信息来计算商品特征分布,通过商品特征分布计算用户评论分布,并使用艾宾浩

到稿日期:2016-12-30 返修日期:2017-03-16

董晨露(1993—),女,硕士生,主要研究方向为数据分析与数据挖掘,E-mail:15120622@bjtu.edu.cn;柯新生(1970—),男,博士,副教授,硕士生导师,主要研究方向为企业信息化与电子商务,E-mail:xske@bjtu.edu.cn(通信作者)。

斯遗忘曲线来模拟用户兴趣的变化,以得到更精确的用户评论分布,进而计算用户的评论相似度。将用户评论相似度和用户评分相似度结合起来作为用户最终的相似度,进而进行评分预测。

2 相关工作

近年来,协同过滤算法、基于内容的推荐算法、基于知识的推荐算法和混合推荐技术等受到了学术界的广泛关注^[3-4]。随着推荐系统中用户和商品数量的急剧增长,传统协同过滤算法的短板逐渐显现,如实时性、冷启动、数据过期和数据稀疏性等^[5-7]。本文主要讨论数据稀疏性问题。针对该问题,许多学者提出了相应的优化算法。其中一类重要的方法是混合协同过滤,如杨武等于2016年提出将基于内容的推荐算法与协同过滤算法相结合进行新闻推荐^[8],虽然能够在一定程度上缓解数据稀疏性问题,但这类方法所利用的信息受限于商品的类型,而且形式各异,有时难以获取。另一类方法是对用户或商品进行聚类,如郭伟光等利用基于用户模式相似的子空间聚类方法来改进协同过滤方法^[9],但该方法在稀疏数据集上的表现并不出色。与此同时,还有一些学者提出了将用户信任和协同过滤算法相结合的方法,如Jeong等提出将用户信任等级引入到协同过滤算法中^[10],但该方法对用户间的交互信息要求较高,更适合社交推荐等交互程度较高的推荐系统,并不适用于电子商务系统。

用户-项目评分矩阵的日益稀疏给推荐系统的推荐准确度带来了严重影响。随着评论信息价值的逐渐显现,学者们开始将评论引入到传统协同过滤算法中^[11]。Musat等^[12]提出了一种主题分布协同过滤技术,该技术从用户的评论文本中构建用户配置文件,并使用这些配置文件从用户的角度过滤其他评论文本。王全民等^[13]将评论挖掘技术引入到协同过滤算法中,并利用得到的物品预测评分对用户-评分矩阵进行填充以产生推荐。McAuley等^[14]利用LDA主题概率模型来计算商品主题分布,并通过与latent-factor模型相结合建立评论主题与评分维度间的关系,不过这种方法并未考虑用户对商品的情感倾向。高玮璠等^[15]利用LDA主题概率模型方法处理评论数据,以获取用户偏好与商品画像信息,并利用回归模型建立评论与评分间的关系,但该方法默认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。李伟霖等^[16]利用主题概率模型计算评论主题分布,利用评分计算评论态度影响因子,并利用评论态度影响因子来凸显评论主题分布中的突出特征,该方法虽然能够在一定程度上缓解数据稀疏性,但并未考虑用户兴趣迁移现象。

对用户兴趣随时间的变化情况进行研究,在推荐系统中也备受关注。Koren^[17]利用时序图对用户的长期兴趣和短期兴趣加以区分,以获取用户更为精确的当前兴趣。张艳梅等^[18]提出了一种融合指数遗忘权重和时间窗口的算法,将近期兴趣和反复出现的早期数据相结合作为衡量用户兴趣的重要因素。郑志高等^[19]提出了一种时间加权不确定近邻的协同过滤推荐算法,通过时间加权和引入近邻因子来进行推荐。陈海涛等^[20]将用户兴趣相似性融入到用户相似度的度

量中,根据用户对各类项目的购买数量比例矩阵和时间来衡量用户兴趣及其动态变化。蒋胜等^[21]将通过动态社会化标签得出的用户偏好相似度、通过用户背景计算得出的用户相似度以及基于时间权重的用户评分相似度结合起来进行推荐。赵海燕等^[22]通过标签流行度对用户偏好进行刻画,利用用户的时间行为信息获取用户兴趣相似度,并结合用户相似度和时间衰减项为用户产生推荐。总体来说,以上研究虽然大都考虑到了用户兴趣迁移现象,但大部分仍然无法缓解数据稀疏性问题。

综上所述,上述研究各有侧重,并且在各自侧重点的基础上提升了推荐质量,但这些研究都比较依赖用户评论信息或社交关系等具体情境信息,鲜有学者综合运用用户兴趣变化和用户评论来提高推荐系统的推荐质量。基于此,本文提出一种基于用户兴趣变化和评论的协同过滤算法,即将用户评论信息作为商品的特征信息,使用LDA主题模型获取商品特征分布,并利用商品特征分布获取用户的评论分布,基于用户兴趣变化现象和艾宾浩斯遗忘曲线引入时间权重函数,通过结合用户评论相似度和用户评分相似度得到用户最终相似度,进而对商品评分进行预测,以提升推荐质量。

3 基于用户兴趣变化和评论的协同过滤算法的设计

3.1 用户评论分布的计算

用户在评论信息中往往会提及自己比较关注的商品特征,据此可以认为评论信息能够作为商品特征的表征。本文使用隐含狄利克雷分布(Latent Dirichlet Allocation, LDA)主题模型来获取商品特征。LDA是一种文档主题生成模型,也是一个三层贝叶斯概率模型。其认为一篇文档中的每个词都是因为经过了“以一定的概率选择了某个主题,并从该主题中以一定概率选择了某个词语”的过程而得到的。LDA能够用于识别大规模文档集或语料库中潜藏的主题信息,是一种非监督机器学习技术。若要生成一篇文档,则该文档中每个词语的概率公式如下:

$$p(\text{词语} | \text{文档}) = \sum_{\text{主题}} p(\text{词语} | \text{主题}) \times p(\text{主题} | \text{文档}) \quad (1)$$

LDA主题模型的输入为大规模的文档集合,输出为文档在主题上的概率分布和主题在单词上的概率分布。在使用LDA主题模型之前,需要对评论信息进行处理,以得到LDA主题模型的输入文档。处理过程主要分为中文分词和停用词过滤两个方面。中文分词是指将中文句子切分成单个词语的排序列表,词语之间没有明显的界限。英文以空格和单词为分割点和基本单元,单词间有明显的界限。分词的准确性从源头上影响着推荐的质量,因此中文分词是评论挖掘的关键步骤之一。停用词是指在中文中出现频率较高却缺少实际意义的词,如“至于”“了”“那个”等,这些词对文本意义的表达作用甚微,并且在处理时还会占用大量的资源,因此需要在评论预处理过程中对其进行过滤。

理论上,用户在评论信息中既会提及其喜好的、重点关注的商品特征,又会提及其不喜欢的商品特征。然而我们在研究过程中发现,大部分用户倾向于在评论信息中提及其喜欢的商品特征,只有极少量用户会提及其不喜欢的商品特征,因此引入用户厌恶分布会大大增加用户评论相似度的误差,进

而降低推荐的精确度。因此,本文使用用户喜好分布直接作为用户的评论分布:

$$c'(u) = \frac{\sum_{i \in I_u} \theta_i (r_{u_i} - r_{\min})}{\sum_{i \in I_u} (r_{u_i} - r_{\min})} \quad (2)$$

其中, $c'(u)$ 表示用户的 K 维评论分布; I_u 表示用户 u 评分过的商品集合; θ_i 表示使用LDA主题模型得到的商品 i 的 K 维特征分布, r_{u_i} 表示用户 u 给商品 i 的评分; r_{\min} 表示系统评分范围的最低值。由上述公式可以看出,用户 u 对商品 i 的评分 r_{u_i} 越高,商品 i 的特征分布向量 θ_i 占该用户的喜好分布的比例就越大。

3.2 用户兴趣变化的引入

人们的兴趣总是随着时间的推移和环境的变迁而动态变化的,在电子商务系统中表现为越是最近购买的商品越能够反映用户当前的兴趣状态,越早的购买记录对当前的兴趣状态的反映程度越低。德国心理学家 Ebbinghaus 研究发现,遗忘在学习之后立即开始且呈非线性变化,遗忘的速度由快至慢。可以看出,用户的兴趣变化与遗忘规律有着很多相似之处,将遗忘规律加以应用,通过时间权重函数为不同时期的评分赋予相应的权重来模拟用户不同时期的兴趣变化,能够有效地将用户兴趣变化引入到协同过滤算法中。本文使用以下公式对时间权重进行度量:

$$w(u, i) = e^{-\frac{a(t_i - t_u)}{T}} \quad (3)$$

其中, $w(u, i)$ 为用户 u 对商品 i 评分时间的权值; t_i 表示用户最近一次对商品打分的时间; t_u 表示用户 u 对商品 i 打分的时间; T 表示用户在推荐系统中最早打分和最晚打分的时间差; a 是一个可变参数,取值范围为 $(0, 1)$,用于调整权重的大小。由上述公式可知, $w(u, i)$ 越小,表示用户 u 对商品评分的时间越早,该评分对推荐结果的影响越小; $w(u, i)$ 越大,表示评分时间越晚,该评分对推荐结果的影响也就越大。

将该时间权重函数引入到用户兴趣变化的计算中,得到新的用户评论分布公式为:

$$c(u) = \frac{\sum_{i \in I_u} \theta_i (r_{u_i} - r_{\min}) w(u, i)}{\sum_{i \in I_u} (r_{u_i} - r_{\min}) w(u, i)} \quad (4)$$

3.3 用户相似度计算和评分预测

计算用户相似度的方法有很多,如余弦相似度、修正余弦相似度、皮尔逊相似度等。前文得到了用户的评论分布,它是一个 K 维向量,选用余弦相似度的方法来计算用户的评论相似度。

用户 u 与用户 v 的评论相似度为:

$$sim_c(u, v) = \frac{\vec{c}_u \cdot \vec{c}_v}{\|\vec{c}_u\| \cdot \|\vec{c}_v\|} \quad (5)$$

其中, $sim_c(u, v)$ 为用户 u 与用户 v 的评论相似度, \vec{c}_u 和 \vec{c}_v 分别表示用户 u 和用户 v 的评论分布。

不同用户的打分水平存在差异,例如,如果用户 u 和用户 v 对某一商品都非常满意,且满分为10分,则用户 u 可能打9分或10分,而用户 v 可能打8分或9分。为了消除这种评分水平的差异,选用皮尔逊相似度方法对用户评分相似度进行计算。

用户 u 与用户 v 的评分相似度:

$$sim_r(u, v) = \frac{\sum_{i \in I_{uv}} (r_{u_i} - \bar{r}_u)(r_{v_i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{u_i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{uv}} (r_{v_i} - \bar{r}_v)^2}} \quad (6)$$

其中, $sim_r(u, v)$ 为用户 u 与用户 v 的评分相似度, I_{uv} 表示用户 u 和用户 v 共同打过的商品集合, r_{u_i} 和 r_{v_i} 分别表示用户 u 和用户 v 对商品 i 的评分, \bar{r}_u 和 \bar{r}_v 分别表示用户 u 和用户 v 的评分均值。

将用户评论相似度和用户评分相似度进行线性组合,从而得到用户最终相似度。选取组合参数时需要在实验中使用不同数据进行测试,以得到最优参数。

$$sim(u, v) = b \cdot sim_c(u, v) + (1 - b) \cdot sim_r(u, v) \quad (7)$$

其中, b 为可变参数,取值范围为 $(0, 1)$,用于调整评分相似度和偏好相似度的比例大小。 b 越大,评分相似度占比越高; b 越小,偏好相似度占比越高。

对用户相似度进行排序,选取与目标用户相似度最高的若干个用户作为目标用户的最近邻集合,使用传统协同过滤算法中计算预测评分的公式来计算用户 u 对商品 i 的预测评分:

$$\hat{P}(u, i) = \bar{r}_u + \frac{\sum_{v \in L_u} (r_{v_i} - \bar{r}_v) \cdot sim(u, v)}{\sum_{v \in L_u} |sim(u, v)|} \quad (8)$$

其中, L_u 表示用户 u 的最近邻集合; \bar{r}_u 表示用户 u 在推荐系统中的评分均值; r_{v_i} 表示用户 v 对商品 i 的评分; $sim(u, v)$ 表示用户 u 和用户 v 之间的相似度。

4 实验

4.1 数据准备

实验数据的真实性、完整性对实验的成功至关重要。通过对比网络上的各种实验数据,最终选取豆瓣电影的短评作为实验数据。使用Python编写网络爬虫,爬取了豆瓣电影上1620部电影的190多万条短评。通过对数据进行初步整理,从原始数据集中提取了1762个用户对1430部电影的72140条评论,时间跨度为2007年1月到2017年8月,评论信息包括电影ID、用户名、评论内容、评分以及评论发表时间,如图1所示。

1291588,岁月的童话 おもひでぼろぼろ(1991),爱丽丝,4,人物笑起来的时候都好丑...,2009-07-17

1291588,岁月的童话 おもひでぼろぼろ(1991),独角兽,4,此中有真意,悠悠已忘言.,2009-11-21

1291588,岁月的童话 おもひでぼろぼろ(1991),littletwo,5,若干年后女孩在失眠的夜里想起,曾经有个木讷的男生站在她回家的路口,问她:你比较喜欢哪一个?阴天还是晴天.....,2008-02-23

1291588,岁月的童话 おもひでぼろぼろ(1991),A-sun*,4,片尾实在太棒了啊.,2010-02-14

1291588,岁月的童话 おもひでぼろぼろ(1991),恶魔的步调,5,那是一个吃菠萝都可以拿来作为一个场景的年代。现在与过去的记忆这样的虚实表达手法如果从如今来看的话也许我们会说有着今敏样的标签,但自然今敏是个不切实际的艺术家,而高畑勋是个通俗故事者,这也是他与宫崎骏的不同之处,高畑勋总是带着这样的现实主义.,2009-07-28

1291588,岁月的童话 おもひでぼろぼろ(1991),城南草木生,5,你才是岁月神偷.,2010-11-27

图1 爬取的数据集示例

Fig. 1 Example of crawling data set

为了避免沉默用户对推荐结果产生影响,保证清理后的数据集中每个用户至少对10部电影发表过评论和评分,每部电影也至少有10条评论和评分。数据集的稀疏度一般用公式 $1 - \frac{\text{评分总数}}{\text{用户数} \times \text{项目数}}$ 来计算,因此本文实验所用数据集的稀疏度为97.1%,可见本文的数据集是非常稀疏的。数据集的基本信息统计如表1所列。

表1 数据集信息统计

Table 1 Statistics of data sets

用户数量	电影数量	评分数	评分范围	最少评分记录	稀疏度/%
1762	1430	72140	1~5	10	97.1

4.2 评价指标

1) 推荐效果评价指标

准确率(Precision)、召回率(Recall)和平均绝对偏差(Mean Absolute Error, MAE)都可以作为衡量推荐算法推荐效果的指标。MAE值是预测评分与真实评分之间的平均绝对误差,是衡量推荐质量最常用的指标。MAE的计算公式如下:

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (9)$$

其中, p_i 表示第 i 个项目的预测评分; q_i 表示第 i 个项目的真实评分; N 为测试集中待预测项目的个数。MAE值越小,说明推荐误差越小,其值与推荐质量成反比

2) 主题概率模型评价指标

Perplexity能够衡量主题概率模型对文本的预测能力,是用来衡量语言模型生成性能的测量指标。其值表示文本在主题概率模型生成的集合中的似然估计,值越小表示文本的似然估计越高。Perplexity值与主题概率模型的生成性能成反比,计算公式如下:

$$Perplexity = \exp\left\{-\frac{\sum_{m=1}^M \log \# p(r_m)}{\sum_{m=1}^M N_m}\right\} \quad (10)$$

其中, $p(r_m)$ 为第 m 篇文档 r_m 的词语集合在主题概率模型下的生成概率, N_m 为第 m 篇文档 r_m 的总词数。本文的算法中使用用户评论表征商品特征,将每一个商品的所有评论整合到一篇文档中,因此这里的一篇文档即是一个商品的所有评论信息。

4.3 实验结果及分析

实验使用交叉验证的方法,将实验所用数据集中的80%作为训练集,20%作为测试集。训练集中的数据用于算法中各个步骤的计算和对测试集中的电影评分进行预测;测试集中的数据用于衡量算法的推荐质量。

4.3.1 LDA主题数

LDA参数的选择关系到商品特征提取的质量。LDA参数主要有超参数 α, β 和主题数 K 。超参数 α 取经验值 $50/K$, β 为0.01,主题数 K 分别取值为5,10,15,20,25,30。观察不同主题数下 Perplexity 值的变化,实验结果如图2所示。

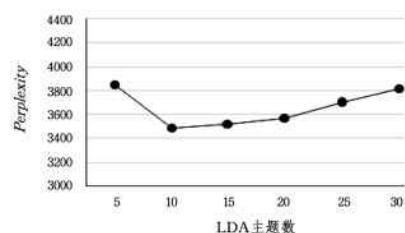


图2 LDA主题数对 Perplexity 值的影响

Fig. 2 Influence of LDA subject number on Perplexity value

从图2看出,随着LDA主题数 K 从5依次增加,Perplexity值先减小再增大,当主题数 K 为10时,Perplexity的值最小,此时本文采用的LDA主题概率模型对评论文本的生成性能最优,因此选取 $K=10$ 作为本文LDA主题模型的主题数目。

4.3.2 时间权重因子

时间权重因子决定了不同时间段下评论的权重。目前已有研究成果表明,当最近邻的数量为30到60时,推荐效果较为良好^[23],因此选取最近邻数量为50来观察时间权重因子 a 和用户相似度权重因子 b 的取值对MAE值的影响。

在设定用户相似度权重 b 为0.2,0.4,0.6的条件下,当时间权重因子 a 取不同值时,本文算法对测试集中的项目的预测评分与其实际评分的MAE值如图3所示。

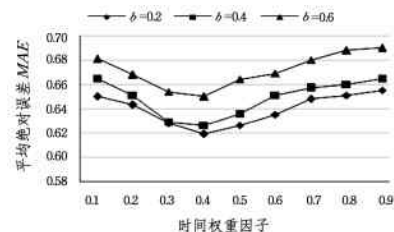


图3 时间权重因子对 MAE 值的影响

Fig. 3 Influence of time weighting factor on MAE value

从图3看出,在不同的相似度权重因子 b 下,随着时间权重因子 a 从0.1增至0.9,本文提出的算法对测试集中项目的预测评分曲线呈现先下降后升高的趋势。当 a 取0.4时,算法的MAE值最小,此时推荐质量最好。为了获得更加精确的结果,在接下来的实验中将时间权重因子的值取为0.3。

4.3.3 用户相似度因子

根据前文的实验,在LDA主题概率模型的主题数 K 为10、时间权重因子 a 为0.4的情况下,观察MAE值随用户相似度权重因子 b 的变化而变化的情况,得到如图4所示的实验结果。

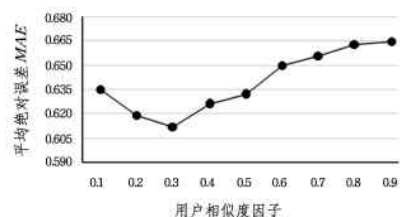


图4 用户相似度因子对 MAE 值的影响

Fig. 4 Influence of user similarity factor on MAE value

从图 4 看出,随着用户相似度因子 b 的不断增大,本文算法对测试集项目的预测评分的 MAE 值先减小再增大,当 b 取 0.3 时,MAE 取得最小值,此时算法的推荐质量最好,即用户评论相似度占 30%,用户评分相似度占 70% 时,本文算法的推荐质量最高。该结果也说明了用户评论相似度的引入能够有效降低预测评分的平均绝对偏差,达到更好的推荐效果。

4.4 对比实验

为了验证所提算法的质量,设计了一组实验来对各算法(所提算法(TTCF)、未考虑用户评论信息和时间因素传统的基于用户的协同过滤算法(UCF)、文献[16]提出的使用主题模型对用户评论进行利用但未考虑时间因素的算法(UCFRR))在不同最近邻数量下的 MAE 值进行对比,结果如图 5 所示。

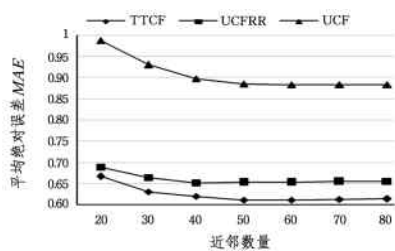


图 5 不同算法的近邻数对 MAE 值的影响

Fig 5 Influence of the neighbor number of different algorithms on the MAE value

从图 5 看出,使用主题模型对评论信息加以利用的协同过滤算法(TTCT 和 UCFRR)的 MAE 值在不同近邻数量下均远小于传统的协同过滤算法的 MAE 值;而在评论信息的基础上进一步将时间权重引入到协同过滤算法中能够为不同时间阶段的数据赋予不同的权重,从而进一步减小 MAE 值,提高推荐质量。

综上所述,本文提出的算法在稀疏数据集上能够获得较好的推荐效果。

结束语 本文通过将用户评论信息和用户兴趣变化因素引入到传统协同过滤算法中,提出了基于用户兴趣变化和评论的协同过滤算法。将用户评论信息作为商品特征信息的表征,使用 LDA 主题模型提取商品的特征分布,并以此来获取用户的评论分布。考虑到用户兴趣的变化情况,将基于艾宾浩斯遗忘曲线的时间权重函数引入到用户评论分布的计算过程中,以获取更为精确的用户评论分布和用户评论相似度。将用户评论相似度与用户评分相似度线性结合以获取用户最终相似度,进而对商品评分进行预测。使用网络爬取的豆瓣电影的真实数据进行验证,实验结果显示,本文提出的基于用户兴趣变化和评论的协同过滤算法在稀疏数据集下能够获得较好的推荐效果。下一步将尝试在本文算法的基础上,将其他非评分信息或方法(如上下文信息或用户信任度等)与已有的信息结合,从而实现更为精确的推荐。

参考文献

[1] BREESE J S, HECKERMAN D, KADIE C. Empirical analysis of predictive algorithms for collaborative filtering[C]//Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence(UAI98). 1998:43-52.

[2] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C]//Proceedings of the 10th International World Wide Web Conference. 2001: 285-295.

[3] SUN Z, HAN L, HUANG W, et al. Recommender systems based on social networks[J]. Journal of Systems and Software, 2015, 99(C): 109-119.

[4] LU J, WU D, MAO M, et al. Recommender system application developments: a survey[J]. Decision Support Systems, 2015, 74(C): 12-32.

[5] SU X, KHOSHGOFTAAR T M. A survey of collaborative filtering techniques[M]. London: Hindawi Publishing Corp., 2009.

[6] WEI S Y, YE N, ZHANG S, et al. Collaborative filtering recommendation algorithm based on item clustering and global similarity[C]//5th International Conference on Business Intelligence and Financial Engineering(BIFE). IEEE, 2012: 69-72.

[7] ZHANG G Q, LEI C H, LV X M. The survey of E-commerce recommendation[J]. Journal of Intelligence, 2011, 30(12): 60-65. (in Chinese)

张光前, 雷彩华, 吕晓敏. 电子商务推荐的研究现状及其发展前景[J]. 情报杂志, 2011, 30(12): 60-65.

[8] YANG W, TANG R, LU L. News recommendation method by fusion of content-based recommendation and collaborative filtering[J]. Journal of Computer Applications, 2016, 36(2): 414-418. (in Chinese)

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.

[9] GUO W G, ZHANG L. Collaborative filtering method for personalized recommendation based on users pattern clustering[J]. Journal of Intelligence, 2011, 30(2): 160-163. (in Chinese)

郭伟光, 章蕾. 基于用户模式聚类的协同过滤个性化推荐方法[J]. 情报杂志, 2011, 30(2): 160-163.

[10] JEONG B, LEE J, CHO H. User credit-based collaborative filtering[J]. Expert Systems with Applications, 2009, 36(3): 7309-7312.

[11] GANU G, ELHADAD N, MARIAN A. Beyond the Stars: Improving Rating Predictions using Review Text Content[C]//WebDB. 2009: 1-6.

[12] MUSAT C C, LIANG Y, FALTINGS B. Recommendation using textual opinions[C]//IJCAI International Joint Conference on Artificial Intelligence. 2013: 2684-2690.

[13] WANG Q M, WANG L, CAO J Q. Improved collaborative filtering recommendation algorithm based on comments mining[J]. Computer Technology and Development, 2015, 25(10): 24-28. (in Chinese)

王全民, 王莉, 曹建奇. 基于评论挖掘的改进的协同过滤推荐算法[J]. 计算机技术与发展, 2015, 25(10): 24-28.

[14] 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. ACM, 2013: 165-172.

- change detection and recognition system for forestry[J]. International Journal of Remote Sensing, 1998, 19(6): 1079-1099.
- [5] LLADÓ X, GANILER O, OLIVER A, et al. Automated detection of multiple sclerosis lesions in serial brain MRI[J]. Neuro-radiology, 2012, 54(8): 787-807.
- [6] RADKE R J, ANDRA S, AL-KOFAHI O, et al. Image change detection algorithms: a systematic survey[J]. IEEE Transactions on Image Processing, 2005, 14(3): 294-307.
- [7] PEI L, REZA S M S, IFTEKHARUDDIN K M. Improved brain tumor growth prediction and segmentation in longitudinal brain MRI[C]// IEEE International Conference on Bioinformatics and Biomedicine. IEEE, 2015: 421-424.
- [8] MORAAL B, MEIER D S, POPPE P A, et al. Subtraction MR Images in a Multiple Sclerosis Multicenter Clinical Trial Setting [J]. Radiology, 2009, 250(2): 506-514.
- [9] PECOT T, KERVRANN C. Patch-based markov models for change detection in image sequence analysis[C]// The International Workshop on Local and Non-Local Approximation in Image Processing. 2008.
- [10] SEO H J, MILANFAR P. Using local regression kernels for statistical object detection[C]// IEEE International Conference on Image Processing. IEEE, 2008: 2380-2383.
- [11] SEO H J, MILANFAR P. Training-free, generic object detection using locally adaptive regression kernels[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(9): 1688-1704.
- [12] FU Y, YAN S, HUANG T S. Correlation Metric for Generalized Feature Extraction[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2008, 30(12): 2229-2235.
- [13] BOSCH M, HEITZ F, ARMSPACH J P, et al. Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution[J]. Neuroimage, 2003, 20(2): 643-656.
- [14] SIMÕES R, SLUMP C. Change detection and classification in brain MR images using change vector analysis. [C]// 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011: 7803-7807.
- [15] HSU Y Z, NAGEL H H, REKERS G. New likelihood test methods for change detection in image sequences[J]. Computer Vision Graphics & Image Processing, 1984, 26(1): 73-106.
- [16] MITRA A, DE A, BHATTACHARJEE A K. Detection of Progression of Lesions in MRI Using Change Detection[C]// Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2013. Springer International Publishing, 2014: 467-473.
- [17] PATRIARCHE J W, ERICKSON B J. Automated Change Detection and Characterization in Serial MR Studies of Brain-Tumor Patients[J]. Journal of Digital Imaging, 2007, 20(3): 203-222.
- [18] MA Y, LAO S, TAKIKAWA E, et al. Discriminant analysis in correlation similarity measure space[C]// Proceedings of the Twenty-Fourth International Conference on Machine Learning (DBLP). 2007: 577-584.
- [19] NIKA V, BABYN P, ZHU H. Change detection of medical images using dictionary learning techniques and principal component analysis[J]. Medical Imaging: Computer-aided Diagnosis, 2014, 9035(2): 024502.
- [20] GONG M, ZHAO J, LIU J, et al. Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks[J]. IEEE Transactions on Image Processing, 2016, 24(4): 2141-2151.

(上接第 217 页)

- [15] GAO Y F, YU W Z, CHAO P F, et al. Analyzing reviews for rating prediction and item recommendation [J]. Journal of East China Normal University (Natural Science), 2015(3): 80-90. (in Chinese)
高玮璠, 余文喆, 晁平复, 等. 基于评论分析的评分预测与推荐 [J]. 华东师范大学学报 (自然科学版), 2015(3): 80-90.
- [16] LI W L, WANG C L, WEN J C. Collaborative filtering recommendation algorithm based on reviews and ratings[J]. Application Research of Computers, 2017, 34(2): 1-7. (in Chinese)
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): 1-7.
- [17] KOREN Y. Collaborative filtering with temporal dynamics[J]. Communications of the ACM, 2010, 53(4): 89-97.
- [18] ZHANG Y M, WANG L. Research on social tagging recommendation algorithm incorporated with user interest change [J]. Computer Engineering, 2014, 40(11): 318-321. (in Chinese)
张艳梅, 王璐. 适应用户兴趣变化的社会化标签推荐算法研究 [J]. 计算机工程, 2014, 40(11): 318-321.
- [19] ZHENG Z G, LIU J, WANG P, et al. Time-weighted uncertain nearest neighbor collaborative filtering algorithm[J]. Computer Science, 2014, 41(8): 7-12. (in Chinese)
郑志高, 刘京, 王平, 等. 时间加权不确定近邻协同过滤算法 [J]. 计算机科学, 2014, 41(8): 7-12.
- [20] CHEN H T, SONG S S, LI T Q. Improved user based collaborative filtering recommendation algorithm [J]. Information Studies: Theory & Application, 2015, 38(9): 100-103. (in Chinese)
陈海涛, 宋姗姗, 李同强. 基于用户的改进的协同过滤推荐算法 [J]. 情报理论与实践, 2015, 38(9): 100-103.
- [21] JIANG S, WANG Z Q, XIU Y, et al. Collaborative filtering recommendation method based on dynamic social behavior and user's background information [J]. Computer Science, 2015, 42(3): 252-255. (in Chinese)
蒋胜, 王忠群, 修宇, 等. 基于动态社会行为和用户背景的协同推荐方法 [J]. 计算机科学, 2015, 42(3): 252-255.
- [22] ZHAO H Y, HOU J D, CHEN Q K. Collaborative filtering recommendation algorithm combining time weight and trust relationship [J]. Application Research of Computers, 2015, 32(12): 3565-3568. (in Chinese)
赵海燕, 侯景德, 陈庆奎. 结合时间权重与信任关系的协同过滤推荐算法 [J]. 计算机应用研究, 2015, 32(12): 3565-3568.
- [23] HUANG C G, YIN J, WANG J, et al. Uncertain neighbors' collaborative filtering recommendation algorithm [J]. Chinese Journal of Computers, 2010, 33(8): 1369-1377. (in Chinese)
黄创光, 印鉴, 汪静, 等. 不确定近邻的协同过滤推荐算法 [J]. 计算机学报, 2010, 33(8): 1369-1377.