

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

张宜浩¹⁾ 朱小飞²⁾ 徐传运¹⁾ 董世都¹⁾

¹⁾(重庆理工大学两江人工智能学院 重庆 400054)

²⁾(重庆理工大学计算机科学与工程学院 重庆 400054)

摘 要 目前,大多数推荐技术使用用户评分来推断用户偏好.当有充足的评分信息时,协同过滤技术表现良好.然而,评分数据普遍存在着稀疏性,或者难以让用户将其偏好表示为对物品的评分等级,故有效性受到限制.基于内容的推荐方法依据物品内容来寻找与目标用户喜欢的物品内容相似的物品.在目标用户没有充足的历史数据的情况下,该方法仍然不充分,其推荐效果也很有限.当前,融合多视图的兴趣偏好信息构建混合推荐系统是个性化推荐研究发展的趋势.混合推荐系统通过融合用户物品的交互评分、隐式反馈和辅助信息进行个性化推荐,故本文提出了一种新颖的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.针对用户评论、物品内容描述等短文本的情感及语义难以分析,单一推荐视图易导致对用户画像建模粗放等问题,本文利用词向量对用户评论的短文本进行分布式表征,并结合长短期记忆网络实现从上下文语义层面对用户评论的情感进行分析.同时,本文提出基于观点预过滤和基于用户评分嵌入的情感融合方法,设计了一种嵌入的网络结构对用户评论进行深层语义分析和情感计算,以解决用户评分与真实兴趣偏好存在较大偏差、评分等级分布极度不平衡等问题.此外,本文利用分布式的段落向量表征对物品内容描述的短文本进行相似度计算,并设计了候选物品相似性的计算方法及度量 K 个最近邻物品的方法,解决了推荐系统中物品内容信息不易挖掘和利用的问题.最后,本文提出了一种基于协同训练的融合用户评分、情感倾向和物品内容信息的混合推荐算法,实现对稀疏的用户评分矩阵的循环填充和修正,进而实现基于评分预测的 TopN 推荐.该方法解决了混合推荐系统中不同兴趣偏好的多推荐视图难以融合的问题,同时在一定程度上解决了推荐系统建模中缺乏足够的有标签数据问题.本文在亚马逊数据集上进行实验,与多种经典的和当前先进的推荐算法进行性能对比,采用平方误差、命中率和标准化折扣累积增益进行性能评价.实验结果表明,本文提出的算法在挖掘用户情感上效果显著;在 10 个推荐数据集上,系统的评分预测和 TopN 推荐指标皆有不同程度的显著改进.

关键词 混合推荐;分布式表征;情感分析;协同训练;评分矩阵

中图法分类号 TP391 DOI号 10.11897/SP.J.1016.2019.01316

Hybrid Recommendation Approach Based on Deep Sentiment Analysis of User Reviews and Multi-View Collaborative Fusion

ZHANG Yi-Hao¹⁾ ZHU Xiao-Fei²⁾ XU Chuan-Yun¹⁾ DONG Shi-Du¹⁾

¹⁾(School of Liangjiang Artificial Intelligence, Chongqing University of Technology, Chongqing 400054)

²⁾(College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054)

Abstract Currently, most recommender techniques use user ratings to infer user preferences. Collaborative filtering techniques perform well when there is sufficient rating information. However, their effectiveness is limited because of the rating sparsity problem, or the difficulty in letting

收稿日期:2018-05-20;在线出版日期:2019-03-05. 本课题得到国家自然科学基金(61702063)、重庆市基础科学与前沿技术研究重点专项(cstc2017jcyjBX0059)资助. 张宜浩,博士,副教授,中国计算机学会(CCF)会员,主要研究方向为推荐系统、机器学习、自然语言处理. E-mail: yhzhang@cqu.edu.cn. 朱小飞,博士,教授,中国计算机学会(CCF)会员,主要研究领域为大数据搜索与推荐、Web挖掘、机器学习. 徐传运,博士,副教授,中国计算机学会(CCF)会员,主要研究方向为机器学习、图像处理. 董世都,博士,副教授,中国计算机学会(CCF)会员,主要研究方向为机器学习、图像处理.

users express their preferences as scalar ratings on items. Content-based recommender methods rely instead on the content representations of items to locate items that have similar content to items the target user liked. However, these methods are still inadequate and its recommendation effect is limited, especially when the target user has little historical data. At present, it is a recent development trend to do personalized recommendation through fusing multi-view of interest preferences to build the hybrid recommendation model, which usually makes personalized recommendation with user-item interaction ratings, implicit feedback and auxiliary information in hybrid recommendation system. In this paper, a novel hybrid recommendation algorithm is proposed that based on deep sentiment analysis of user reviews and multi-view collaborative fusion. For these problems that it is difficult to analyze user reviews' sentiment and items content' semantics, and a single view of the recommended model lead to user profile is extensive, we use Word2vec to characterize the short texts of user reviews and combine long short-term memory networks to realize the sentiment analysis of the user review on the context semantic level. At the same time, a sentiment fusion method based on opinion pre-filtering and user rating embedding is proposed, and an embedded network structure is designed for deep semantic analysis and sentiment calculation of user's review. The proposed method will solve the problem that there is a great deviation between the user's rating and real interest preference, and also solve the extreme imbalance problem of the user rating distribution. In addition, we use the distributed vector representation of paragraph to characterize the short text of the item's text description, so as to realize the similarity calculation of the item's content. We design a method to measure the similarity of candidate items and calculate K nearest neighbor items, which solves the problem that the item's content information is not easy to mine and use in recommendation system. Finally, a fusion method of recommendation view based on collaborative training is proposed, which integrates user ratings, sentiment preferences and item's content information. It can fill and modify the sparse user ratings matrix, and then realize recommendation based on ratings prediction. It solves the problem that multi-recommendation views with different interests and preferences are difficult to fuse in hybrid recommendation system, and solves the problem of lack of sufficient labeled data for modeling in a certain degree. We conduct the experiments on Amazon product dataset, and compare our algorithm with a variety of classic and state-of-the-art recommendation algorithms. Specially, the results are evaluated in Mean Squared Error, Hit Radio, and Normalized Discounted Cumulative Gain. The experiment result shows that the algorithm proposed in this paper has a significant effect in mining user's sentiment. On the ten recommended datasets, our algorithm has also a significant improvement in the accuracy of the score prediction and TopN performance of the recommendation system in different degrees.

Keywords hybrid recommendation; distributed representation; sentiment analysis; collaborative training; scoring matrix

1 引言

随着电子商务和社交网络等信息技术的迅速发展,“信息超载”成为困扰人们网络生活的主旋律.个性化推荐作为一种帮助用户快速搜寻有用信息的有效工具,越来越受到人们的青睐.伴随而来的,各种

推荐算法如雨后春笋般出现.纵观推荐算法的发展历程,协同过滤和隐语义模型^[1]是推荐系统发展早期较流行的算法,其在过去十多年间得到了长足发展.鉴于深度学习技术在人工智能诸多应用中取得的显著成效,基于深度学习的推荐模型也逐渐成为研究者追逐的焦点^[2-3].目前,用户评分矩阵(Rating Matrix)仍然是大多数推荐系统利用的主要行为偏

好信息^[4],但基于用户评论^[5]、用户隐式反馈^[6]、物品内容信息^[7]的推荐越来越受到人们的关注,然而受文本挖掘、用户行为分析等方面的制约,这些方面研究取得的进展并不十分令人满意,但它们在解决推荐系统的推荐准确性、冷启动、可解释性等方面具有重要的潜力。

在推荐算法的发展历程中,一种传统的为研究者推崇的是协同过滤算法,其目标是将用户和物品间的二元关系转化为评分预测问题,然后依据用户对物品的评分进行协同过滤或排序^[8],进而产生推荐列表。随后大量的研究工作发现,由于受用户评分真实性的制约以及评分矩阵稀疏性的影响,依据用户评分产生的推荐结果并不能准确地体现用户的兴趣偏好^[8-9]。基于此,研究者们做了大量的后续工作。Zhang 等人^[10]通过比较用户评分和评论文本的情感倾向,指出用户评分并不能真实反映用户评论的情感倾向。扈中凯等人^[5]对用户评分进行统计分析,发现用户对物品的评分比较随意、且评分等级分布极度不均衡(如图 1 中的统计数据^[5],评分等级为 5 分的占 96.2%,评分等级为 4 分的占 3.5%,评分等级为 1~3 分的仅占 0.3%)。同时,本文对实验中来自亚马逊 1995 年至 2013 年的 Automotive 等 10 个数据集的 4 120 948 条用户评论数据进行统计分析,得出类似的结论,即用户的评分等级分布极度不均(评分等级为 1~5 分的分别占比 4.8%、4.4%、9.0%、21.7%、60.1%)。研究表明,这种评分分布极度不均衡的状况给协同过滤推荐造成了极大的困扰。陈龙等人^[11]对用户评论的情感倾向进行分析,发现商品评论的评分是一种弱标注标签,即评论中可能存在实际情感语义与评分不一致的情况(如一条 5 星级的评分对应的评论中仍然存在负面描述)。这种情感语义与评分不相符的标注数据称为噪声数据。研究

表明这种噪声数据会对协同过滤的推荐结果产生较大的负面影响。

在基于内容的推荐方面,物品内容的描述文本信息是一个重要的推荐依据。基于内容的推荐能有效解决系统的冷启动问题^[12],且不受打分稀疏性的约束,能够发掘隐藏的“暗信息”,具有良好的用户体验,因此受到广泛的关注。然而,针对物品内容的短文本自然语言描述(通常较短且零散),无足够的信息量供机器进行统计推断,这给物品内容的语义理解带来了巨大困难。

当前,利用深度学习技术融合多源异构数据^[13]、融合评分矩阵及评论文本^[14]、融合多特征的协同推荐^[15]成为研究的热点。本文在上述研究的基础上,针对推荐系统中用户评分分布的不均衡及多推荐视图不易融合的问题,提出了基于用户评论的深度情感分析与多视图协同融合的混合推荐方法(HRSM 算法),此处多视图即推荐系统中的多维度推荐因素。本文的混合推荐方法融合了用户评分矩阵、用户评论文本、物品的内容描述信息等三个推荐视图。与传统的加权融合和级联型等混合方法不同,本文设计了一种基于协同训练的推荐算法,实现用户评分的行为视图和物品描述的内容视图的融合。本文主要贡献在于提出了基于协同训练的多推荐视图融合的评分预测方法,并探讨了利用基于深度学习的自然语言处理技术对推荐系统中用户评论文本等辅助信息进行整合的技巧。本文的创新主要体现在以下 3 个方面:

(1) 提出基于协同训练的推荐视图融合方法,设计了一种基于协同训练的融合用户评分、情感偏好和物品内容信息的推荐算法,实现对稀疏的用户评分矩阵的循环填充和修正,进而实现基于评分预测的推荐。解决了混合推荐系统中不同兴趣偏好的多推荐视图不易融合的问题,同时在一定程度上解决了推荐系统建模中缺乏足够的有标签数据问题;

(2) 提出基于观点预过滤和基于用户评分嵌入的情感融合方法。设计一种嵌入的网络结构实现从上下文层面对用户评论进行深层语义分析和情感计算,并比较其在挖掘用户评论信息方面的效果。解决了推荐系统中用户原始评分与真实兴趣偏好存在偏差且评分等级分布不均衡的问题;

(3) 研究在推荐系统建模中融合多种自然语言处理技术的技巧。本文利用分布式的段落向量表征对物品内容描述的短文本进行相似度计算,并设计度量候选物品相似性的计算方法及计算 K 个最近

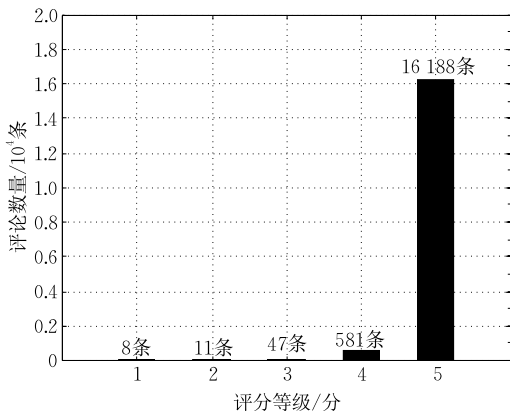


图 1 用户评分等级分布图

邻物品的方法,解决了推荐系统中物品内容的文本描述信息不易挖掘和利用的问题。

2 相关工作

在协同过滤推荐中,基于用户历史评分数据的推荐算法通常面临着数据稀疏的问题,而这种稀疏的用户评分通常会导致推荐质量下降^[16];此外,原始数据的不完整及算法本身处理数据的特殊性,也会导致最终推荐效果不理想^[17]。当前,基于用户评论、用户隐式反馈挖掘的推荐研究受到了高度的重视,但其受文本挖掘和情感分析等技术的困扰。田超等人^[18]通过整合网上商城的用户评论进行情感分析,实现 SuperRank 智能推荐系统原型。Shmueli 等人^[19]将用户评论信息作为协同过滤推荐的依据,并将其和内容信息并入因子模型中进行混合推荐。Zhang 等人^[20]研究用户评论作弊对推荐系统的制约,提出了一种基于可疑行为扩散的统一框架,使得系统人员不需要关心具体的作弊方法就能以较高的准确率识别作弊用户以及作弊行为。Wang 等人^[2]针对评分数据在应用中的稀疏性问题,提出结合物品内容的深度表示及协同训练构建评分矩阵。Wu 等人^[21]通过对 Stacked Denoising Autoencoders 进行扩展来提高评分预测的精确度。Chen 等人^[22]归纳总结了各种各样的基于评论的推荐方法,通过将用户生成的有价值的评论信息融入用户建模和推荐过程中,实现对用户评论推荐因素的挖掘,包括考虑评论的有用性、评论的主题、评论的总体观点、评论的内容、评论的情感等。为了充分挖掘推荐系统中用户评论的上下文信息,以及减少评论中无关信息对推荐准确性的影响。Zhang 等人^[23]提出了一种协同多级嵌入模型,它利用一个投影层将词嵌入模型整合进标准的评分矩阵模型中来解决上述两个局限。Zhang 等人^[24]提出一种深度协同神经网络(DeepCoNN),在网络的最后一层中耦合两个并行的神经网络来实现混合推荐,其中一个网络学习用户评论的行为信息,另外一个网络从用户评论中学习物品的属性信息。Chen 等人^[25]通过引入一种新颖的注意力机制挖掘用户评论的有用性信息,提出了基于神经注意回归模型的推荐系统,可以预测精确的评分及每条评论的有用性。Han 等人^[26]从异构信息网络中提取不同层面的特征,利用精心设计的深度神经网络来学习各个层面的潜在因子,然后将其融合到一个注意力机制中实现协同过滤推荐。

在推荐系统中,通过对真实数据分析也表明:用户评分往往与用户的兴趣偏好存在着较大的偏差,而用户评论等短文本信息可能更真实地反映了用户的兴趣偏好,故对用户评论的短文本的语义理解和情感挖掘是解决推荐系统中用户评分不真实、分布不均的重要途径^[5,18]。

在基于内容的推荐中,物品的内容信息是一个重要的推荐依据,它也是用来解决推荐系统中冷启动问题的重要途径,但这种推荐方法会受到信息获取技术的约束。在众多的物品内容表现形式中,一种重要的形式是商家对物品的自然语言描述,传统的自然言语处理技术(如 Bag of Words)将文本表示成一个 W 维的独热向量,该表示方法缺点非常明显,其假设所有对象都是相互独立的,容易受到数据稀疏问题的影响^[27];同时,这种自然语言处理技术不能从语义层面对物品的内容信息进行分析,因此对物品内容的短文本进行语义理解和相似度计算是解决问题的关键。在对短文本的语义理解方面,研究者做了大量的工作。Wang 等人^[28]利用语义聚类和卷积神经网络对短文本进行建模,并在模型中使用预训练的词嵌入来引入额外知识。王仲远等人^[29]归纳了当前比较流行的短文本语义理解模型:隐性模型、半显性模型、显性模型。Ma 等人^[30]提出一个新颖有效的框架,利用社会化媒体的内容信息来实现评分预测。Lian 等人^[31]提出了一个深度混合模型来提高个性化新闻推荐系统的表征学习能力。

基于内容的推荐是依据用户喜爱的物品内容信息找到相似物品进行推荐,当前较流行的做法是利用信息检索中的相关理论、方法与技术来实现对物品内容信息的建模。由于评分(rating)相对于物品(item)的稀疏性,基于模型的推荐通常缺乏足够的有标签数据^[32]。半监督学习作为一种同时利用有标签数据和无标签数据建模,来改进系统性能的学习策略,在推荐系统建模中得到较广泛的应用^[33-34]。鉴于推荐系统中各个推荐因素相对独立的特点,其很容易被划分为多个“充分冗余视图”,为利用协同训练策略构建推荐预测模型创造了天然的条件。Wu 等人^[33]利用无标签和有标签用户信息(User Profile)构建一个多分类模型,实现了一个半监督混合推荐系统(HySAD)。Zhang 等人^[34]提出了一个基于上下文感知的半监督协同训练方法,来解决推荐系统中的冷启动问题。Ding 等人^[35]将视图数据集成到基于隐式反馈的推荐系统中,以挖掘购买等主要反馈数据以外的隐藏偏好信息。Wang 等人^[36]提出一种

新的基于树增强的嵌入方法,来学习显式的决策规则和不可见的交叉特征,使推荐过程更加透明且有解释性.

3 基于协同训练的混合推荐系统模型

鉴于上述对推荐系统研究现状的论述,本文提出了一种基于用户评论的深度情感分析与多源推荐视图协同融合的混合推荐方法.一方面,我们通过挖掘用户评论的情感倾向,以实现用户对用户原始评分偏离用户真实兴趣偏好的纠正,采用观点预过滤(opinion pre-filtering)方法^[37]实现对用户的情感倾向和原始评分等级的综合度量,为基于物品的协同

过滤推荐模型提供更加精确的反映用户真实兴趣偏好的综合评分数据.另一方面,我们对物品内容描述的文本信息进行挖掘,利用神经网络的方法将其表示成为分布式的段落向量,实现对物品内容的相似度计算,进而构建基于物品内容的推荐模型.最后,本文利用协同训练策略实现对两个推荐视图的融合,并在协同训练中增加了基于置信度估计与聚类分析的数据选择策略,尽量消除迭代训练中加入到训练数据池中的数据分布偏差.在此基础上,利用协同训练模型输出的评分矩阵和物品的相似度,对初始推荐结果进行过滤和排序,从而得到最终推荐结果.基于协同训练的混合推荐系统框架如图2所示.

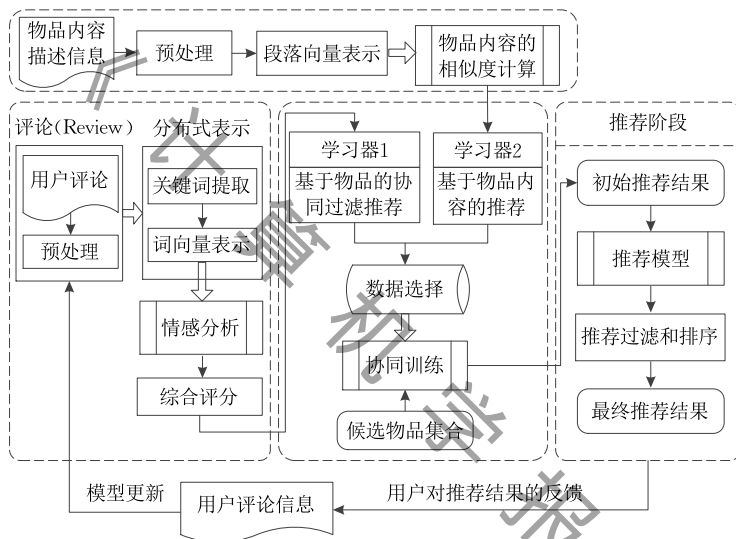


图2 基于协同训练的混合推荐系统框架

3.1 用户评论的情感分析

3.1.1 用户评论文本的分布式向量表示

通过对推荐系统中的用户评论文本进行统计分析,发现其呈现形式通常是关键词和短文本.研究表明,这些短文本信息通常与长文本的处理方法不尽相同.短文本具有长度短、语法不规则的特点,且亦无足够的信息量来供研究者进行统计和推断.传统的诸如词性标注、句法分析等自然语言处理技术在短文本分析方面基本无能为力.早期对短文本的分析和应用主要通过枚举或关键词匹配的方式,对文本的语义理解基本避而不谈,而自动化的短文本理解通常需要依赖额外的知识.本文利用基于词向量的关键词表示方法,解决了传统稀疏表示方式的维数灾难,且无法表示语义信息的问题.同时也挖掘了词之间的关联属性,从而提高了关键词语义表示的准确度.

词向量(Word2vec)作为一种进行高效率词嵌套学习的预测模型^[38],其包括连续词袋模型(CBOW)和Skip-Gram模型两种变体.CBOW通过窗口范围内的词语预测中心词出现的概率,而Skip-Gram则是基于中心词预测窗口范围内词语出现的概率,其训练目标就是找出对预测句子或文档中的周围词语有用的词语的向量表示.假如对于一个给定句子, w_1, w_2, \dots, w_T 表示句子中的词语,Skip-Gram模型的目标函数 $g(w)$ 就是求最大化平均对数概率.

$$g(w) = \frac{1}{T} \sum_{i=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{i+j} | w_i) \quad (1)$$

在式(1)中, c 表示训练文本的数量, c 越大,可能会使得模型的准确率越高.Skip-Gram模型使用层次Softmax函数来定义 $p(w_{i+j} | w_i)$.层次Softmax使用 W 个字作为叶子的输出层的二叉树表示,并且对于每个节点明确表示其子节点的相对概率,利用随

机游走算法分配每个单词的概率。

Word2vec 可以自动从大规模无标注用户评论中学习句法和语义信息,实现对用户评论中关键词的表征.利用 Word2vec 对用户评论的短文本信息进行向量表示,主要分为以下两个步骤:

(1) 根据收集的用户评论文本数据,利用 Skip-Gram 或 CBOW 训练词向量模型,将每个词表示成 K 维向量实数值;

(2) 对于用户评论的短文本,在分词的基础上利用 TF-IDF 等算法抽取 Top- N 个词表示文本的情感,然后从词向量模型中查找得到抽取的 Top- N 个词的 K 维向量表示。

在得到每个关键词的 K 维实数向量表示后,一种较为普遍的做法是利用加权平均的方式对关键词的向量进行处理,将其等价于用户评论文本的向量表示,以实现评论信息的情感分析.这种加权平均的处理方法忽略了词语间的排列顺序对情感预测模型的影响.因为基于 Word2vec 的词向量表示只是基于词的维度进行“语义分析”,而对词向量进行加权平均的处理方式并不具备上下文的“语义分析”能力,故本文构建基于词向量和长短期记忆网络的情感计算模型来实现对用户评论的情感分析。

3.1.2 基于词向量和长短期记忆网络的情感计算

在文本信息处理中,常用的方法是循环神经网络(RNN).然而,RNN 在处理长序列时会导致优化时出现梯度消失的问题.为解决这一问题,研究人员提出了门限(Gated RNN),其中最著名的就是长短期记忆网络(LSTM).研究也表明:在很多任务上,采用 LSTM 结构的神经网络比标准 RNN 网络表现更好。

LSTM 利用“门”结构来去除或增加信息到细

胞状态.它通过在神经元中增加输入门、遗忘门和输出门三个“门”结构来达到增强或遗忘信息的目的,使得自循环的权重是变化的.基于 LSTM 的模型在参数固定的情况下,通过动态改变不同时刻的累积,可以有效避免 RNN 网络结构容易出现的梯度膨胀,甚至梯度消失等问题.在 LSTM 网络结构中,每个 LSTM 单元的计算公式如式(2)~(7)所示:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (4)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (5)$$

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \quad (6)$$

$$h_t = O_t * \tanh(C_t) \quad (7)$$

在式(2)~(7)中, f_t 表示遗忘门, i_t 表示输入门, O_t 表示输出门; \tilde{C}_t 表示前一时刻细胞的状态, C_t 表示当前细胞的状态, h_{t-1} 和 h_t 分别表示前一时刻单元的输出和当前单元的输出。

本文采用基于 Word2vec 和 LSTM 的用户评论的情感分析方法如图 3 所示.首先利用 Word2vec 将矩阵形式的输入编码为较低维度的一维向量,以保留大多数有用信息;然后利用 LSTM 算法训练用户评论文本的情感分类模型,实现对用户评论的评分等级预测.同时,为了兼顾用户评分和评论信息对真实情感的交互影响,本文采用基于观点预过滤的方法和基于用户评分嵌入的方法分别对用户评分和情感预测评分进行融合.前者是利用 LSTM 网络得到预测评分后,同原始用户评分进行加权求和,基于用户评分嵌入的方法则是将 LSTM 网络向量与用户评分信息进行结合,将结果作为最后一层的输入,直接输出最终的综合评分。

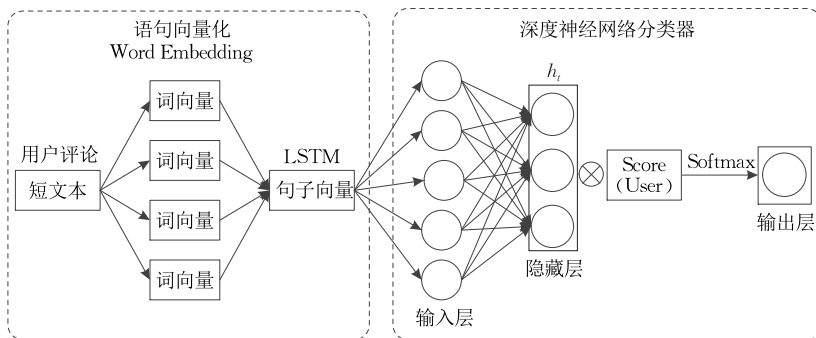


图 3 基于用户评分嵌入的情感分析方法

基于观点预过滤^[37](opinion pre-filtering)的方法,利用 Word2vec 和 LSTM 对用户评论文本建模

进行情感分析,预测得到每个用户对物品评论的情感倾向分数 $Score_r$,对用户的原始评分进行加权求

和得出综合评分 $Score_c$.

$$Score_c = \alpha Score_r + (1 - \alpha) Score_o \quad (8)$$

在式(8)中, $Score_r$ 表示用户对物品评论的情感预测评分, $Score_o$ 表示用户对物品的原始评分, α 是两个评分间权重的平衡因子.

基于用户评分嵌入的方法是在对用户评论信息进行情感分析的基础上,将得到的 LSTM 输出向量与用户评分信息进行结合(如式(9)所示),然后将上述结果作为最后一层(全连接层)的输入,并通过 softmax 激活函数直接输出最终的综合情感评分.

$$H_i = h_i \otimes Score(User_i) \quad (9)$$

3.2 基于物品内容的相似度计算

在推荐系统中,对物品内容的自然语言描述较短且大多是不完整的句子,通常也不遵循语法规则.本文利用段落向量(Paragraph Vector)^[39]对物品内容描述的短文本进行分布式表示.段落向量是一种基于神经网络的隐性短文本理解模型,它将短文本向量当作“语境”用于辅助推理,在极大似然估计中,文本向量亦被作为模型参数进行更新.同基于 Word2vec 的文本向量表示方法相比,它在模型训练过程中对段落也增加了编码.普通的词一样,段落编码也是被先映射成一个向量(即段落编码向量).在计算中,段落编码向量和词向量累加或者连接起来,作为输出层 Softmax 的输入.在对物品内容描述文本的训练过程中,段落编码保持不变,相当于在每次预测单词概率时,其都整合了整个句子的语义信息.在预测阶段,我们给物品内容的描述文本分配一新的段落编码,同时保持词向量和输出层 Softmax 的参数不变.最后,我们利用梯度下降法训练新的物品内容描述文本,直至其收敛,从而得到物品内容的低维向量表示.物品内容的段落向量分布式表征如图 4 所示.

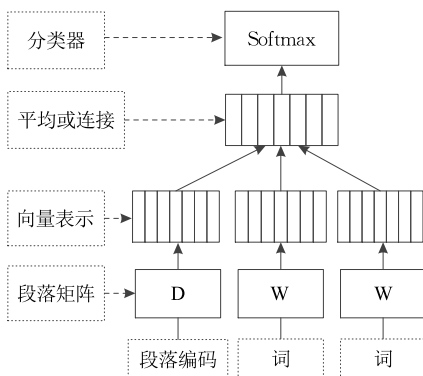


图 4 物品内容的段落向量分布式表征

在得到物品内容的惟一 d 维分布式向量表示后,就可以利用相似度计算的方式得到每两个物品内容之间的相似度和距离.本文利用余弦公式度量两个物品间的相似度,同时利用马氏距离计算两个物品内容自然语言描述的距离.假设两个物品内容自然语言描述的段落向量表示为 $PV_a = (x_{11}, x_{12}, \dots, x_{1d})$ 和 $PV_b = (x_{21}, x_{22}, \dots, x_{2d})$,其中 d 表示两个段落向量的维度.则它们间的相似度和距离分别定义式(10)和式(11):

$$\begin{aligned} sim(PV_a, PV_b) &= \frac{PV_a \cdot PV_b}{\|PV_a\|^2 \cdot \|PV_b\|^2} \\ &= \frac{\sum_{i=0}^{i=d} x_{1i} x_{2i}}{\sqrt{\sum_{i=0}^{i=d} x_{1i}^2} \sqrt{\sum_{i=0}^{i=d} x_{2i}^2}} \end{aligned} \quad (10)$$

$$dis(PV_a, PV_b) = \sqrt{(PV_a - PV_b)^T S^{-1} (PV_a - PV_b)} \quad (11)$$

其中 S 是特征向量 PV_a 和 PV_b 的协方差矩阵.

3.3 基于协同训练的推荐视图融合

在构建混合推荐系统时,本文利用用户综合评分视图构建基于物品的协同过滤推荐模型;与此同时,利用物品内容的自然语言描述视图构建基于物品内容的推荐模型;最后基于协同训练策略实现两个推荐视图的融合.在数据选择方面,利用基于置信度估计与聚类分析的数据选择算法对数据进行过滤,而后加入到另一个分类器的训练数据池中,进行下一轮训练,如此迭代.基于协同训练的推荐视图融合的框架如图 5 所示.

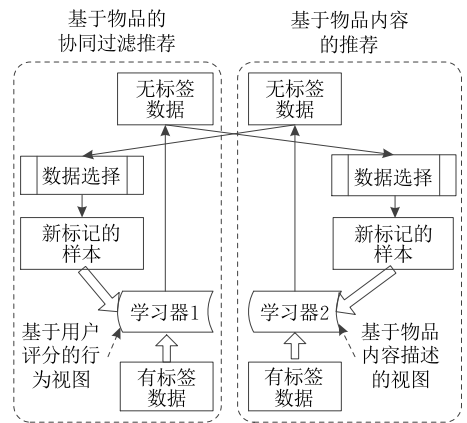


图 5 基于协同训练的推荐视图融合

3.3.1 基于协同训练的混合推荐算法

基于协同训练的混合推荐算法是在用户对物品评分的基础上构建初始评分矩阵;然后利用观点预

过滤的方法度量综合评分,从而更新评分矩阵;最后设计一个基于协同训练的混合推荐算法,依据综合评分矩阵和物品内容描述的向量相似度来循环地填充和优化评分矩阵,进而实现推荐和排序.基于协同训练的混合推荐算法流程如图6所示.

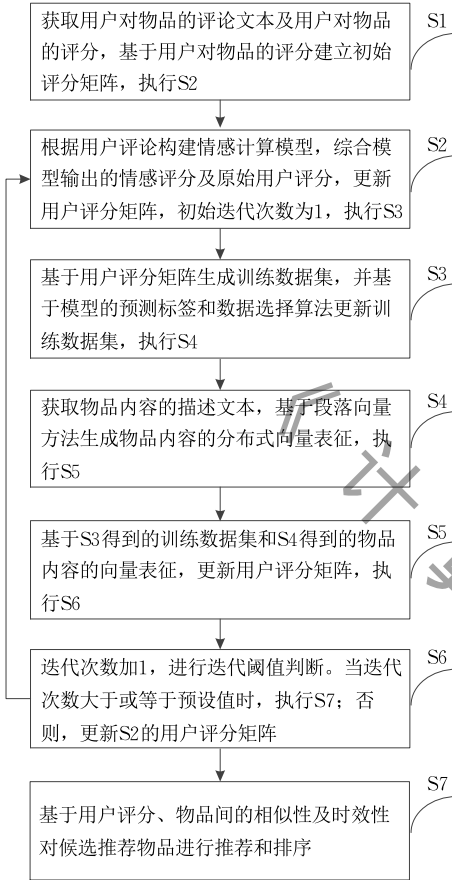


图6 基于协同训练的混合推荐算法流程

在推荐系统中,用户 u 对物品 i 的评分记为 $R_u(i)$; 对应的评分矩阵为 $\mathbf{R}_{m \times n}(U, I)$, 其中行向量 m 表示用户的个数,列向量 n 表示物品的个数.在基于物品的协同过滤推荐模型中,输入用户的原始评分矩阵 $\mathbf{R}_{m \times n}(U, I)$, 其中 $R_u(i) \in \{1, 2, 3, 4, 5\}$, 以及情感分析模型预测的虚拟评分矩阵 $\tilde{\mathbf{R}}_{m \times n}(U, I)$, 其中 $\tilde{R}_u(i) \in \{1, 5\}$, 1 表示用户情感为负面, 5 表示用户情感为正面, 输出为数据集 D_{train} . 基于物品的协同过滤推荐算法的描述如算法1所示.

算法1. 基于物品的协同过滤推荐算法.

输入: 用户对物品的评分矩阵 $\mathbf{R}_{m \times n}(U, I)$, 情感计算模型预测的虚拟评分 $\tilde{\mathbf{R}}_{m \times n}(U, I)$

输出: 基于物品协同过滤推荐的训练数据集 D_{train}

1. 根据用户评分矩阵, 抽取针对用户 u 的训练数据 $D_i = \{\mathbf{R}(i)^T | \mathbf{R}(i) \in \mathbf{R}_{m \times n}(U, I), R_u(i) \neq \emptyset, i \in [1, n]\}$, 其类别标签为 $L(i) = R_u(i) \in \{1, 2, 3, 4, 5\}$;

// 在 $m \times n$ 的评分矩阵中, 行向量表示用户, 列向量表示物品. 其中 $\mathbf{R}(i)$ 表示评分矩阵的列向量, $R_u(i)$ 表示用户 u 对物品 i 的评分

2. 更新训练数据评分 $R_u(i)$: // 利用观点预过滤的方法计算用户对物品的综合评分. 其中 $Time_{\text{cur}}$ 表示当前时间, $Time_{R_u(i)}$ 表示用户对物品的评论时间, 时间只取年份

$$\tilde{R}_u(i) = \{\tilde{\mathbf{R}}_{m \times n}(U, I) \in \{1, 5\} | m = u, n = i\},$$

$$R_u(i) = \frac{\alpha \tilde{R}_u(i) + (1 - \alpha) R_u(i)}{Time_{\text{cur}} - Time_{R_u(i)}}.$$

3. 更新训练数据集: // 将评分 ≥ 4 的标记为正类, 加入数据池 $D_i(+)$ 中, 将评分 ≤ 2 的标记为负类, 加入数据池 $D_i(-)$ 中

$$D_L = \{D_i(+), D_i(-)\},$$

$$D_i(+) = \{\mathbf{R}(i)^T | \mathbf{R}(i) \in \mathbf{R}_{m \times n}(U, I), L(i) \geq 4\},$$

$$D_i(-) = \{\mathbf{R}(i)^T | \mathbf{R}(i) \in \mathbf{R}_{m \times n}(U, I), L(i) \leq 2\}.$$

4. 训练基于物品的协同过滤推荐模型, 用分类器 h_1 对候选数据 $D = \{\mathbf{R}(i)^T | \mathbf{R}(i) \in \mathbf{R}_{m \times n}(U, I), R_u(i) = \emptyset\}$ 进行预测, 得到预测标签 $L(i)$;

$$D'_L \leftarrow \{(D(i), L(i))\}.$$

5. 利用基于置信度估计与聚类分析的数据选择算法对数据进行筛选, 返回预加入训练数据池的数据. // D_L 表示一次迭代中原有的数据, D'_L 表示一次迭代中增加的数据(数据的标签为协同过滤模型的预测评分)

$$\text{Return } D_{\text{train}} = \{D_L \cup D'_L\}.$$

在算法1中, 利用基于物品的协同过滤推荐方法, 来填充用户评分矩阵的缺省值; 同时更新用户 u 的训练数据集. 在情感分类模型中, 一般分为细粒度(5级分类)和粗粒度(2级分类), 考虑到2级情感分类模型的准确率远高于5级情感分类模型^[38], 故本文的推荐算法中采用2级情感分类. 分别将用户情感为正面和负面的评分设置为5分和1分; 然后利用观点预过滤的方法对用户情感评分和原始评分进行综合度量; 最后利用基于物品的协同过滤模型实现对评分矩阵的预测和填充, 并利用基于置信度估计与聚类分析的数据选择算法对数据进行筛选, 将增量数据加入用户 u 的训练数据集.

在基于物品内容描述模型中, 利用 K 最近邻算法来计算物品内容描述的距离, 通过物品的余弦相似度以及 K 个最近邻物品(马氏距离)的评分来更新或填充用户评分和缺省值, 将其利用到基于物品内容的推荐模型中进行下一步迭代. 基于物品内容的推荐算法描述如算法2所示.

算法2. 基于物品内容的推荐算法.

输入: 用户对物品的预测评分 $D'_L = \{(D(i), L(i))\}$, 物品内容描述的向量表示 $\mathbf{PV}(Item)$, 训练数据集 D_{train}

输出: 评分矩阵 $\mathbf{R}_{m \times n}(U, I)$

1. 根据训练数据集 D_{train} , 获取用户的物品集合:

$$D_{it} = \{\mathbf{PV}(i) \in \mathbf{PV}(\text{Item}) \mid L(i) \geq 4 \cup L(i) \leq 2\}.$$

2. 选取 $2p$ 个用户评分为 Φ 的物品候选集, 分别计算候选物品与数据集 D_{it} 中物品的距离和相似度:

FOREACH $D_i \in D_{2p}$ do

FOREACH $D_j \in D_{it}$ do

$$\text{dis}(D_i, D_j) = \sqrt{(D_i - D_j)^T \mathbf{S}^{-1} (D_i - D_j)}$$

$$\text{sim}(D_i, D_j) = \frac{D_i \cdot D_j}{|D_i| |D_j|}.$$

3. 选择 K 个最近邻物品 $\{D_1, D_2, \dots, D_K\}$; // a_q 标记 K 个最近邻物品中每个评分等级的个数, dis_{s_q} 表示物品 D_i 与 K 个最近邻物品的平均距离

FOREACH $q \in \{1, 2, 3, 4, 5\}$ do

if $(q-1 \leq L(K) \leq q)$

$a_q = ++$;

$$\text{dis}_{s_q} = \frac{1}{a_q} \sum_1^{a_q} \text{sim}(D_i, D_j).$$

4. 根据物品内容相似度更新评分矩阵的值:

FOREACH $D_i \in D_{2p}$ do

$\text{Score} = \arg \max(a_j \mid j \in \{1, 2, 3, 4, 5\})$

$R_u(i) = \text{Score} * \text{dis}(q).$

5. 针对不同用户 u 的 $R_u(i)$ 值得, 更新物品 i 的评分:

$$L(i)' \leftarrow \mathbf{R}(i).$$

6. 更新训练数据的评分:

$$D'_L = \{(D(i), L(i))\} \leftarrow \{(D(i), L(i)')\}.$$

7. 利用基于置信度估计与聚类分析的数据选择算法对 D'_L 进行数据分布分析, 并返回筛选的数据.

Return $\mathbf{R}_{m \times n}(U, D).$

混合推荐方法将多种推荐技术进行混合相互弥补缺点, 从而获得更好的推荐效果. 与传统的混合推荐技术(如加权融合、混合推荐、级联型推荐)不同, 本文在构建混合推荐系统时, 采用了协同训练策略构建基于物品的协同过滤推荐(算法 1)和基于物品内容的推荐(算法 2)的混合模型. 在协同训练模型的每一次迭代训练过程中, 利用计算的综合评分数据训练评分预测模型, 实现对评分矩阵的填充和更新; 然后根据更新后的评分矩阵和物品的内容描述信息(评分 ≥ 4 和评分 ≤ 2 的物品分别放入用户喜欢和不喜欢的物品训练池中), 训练得到基于物品内容的推荐模型, 从而对评分矩阵进行填充和更新, 将其作为基于物品的协同过滤推荐模型的输入, 进行下一次迭代训练. 相较于加权融合混合推荐需要不断调整各推荐结果的权值、混合推荐的排序困难、以及级联型推荐的分阶段过程, 本文提出的基于协同训练的混合推荐方法, 在每次迭代训练中充分利用了用户对物品的评分信息(Item Profile 视图)和物品

的内容描述信息(物品的元数据视图), 实现对两种推荐视图的融合, 达到了较好的混合推荐效果.

3.3.2 协同训练中的数据选择

本文在构建协同训练模型时, 增加了数据选择策略对欲加入训练池的数据进行筛选. 规定用户的每一评分等级为数据中的一个分类类别; 在数据池中的训练数据为有标签数据, 待预测的数据为无标签数据. 在数据选择策略中, 不仅考虑样本属于某一类别的置信度分数, 同时也要求选择的样本在每一个(聚类)簇内是均匀分布的, 可以避免选择的训练数据在高斯分布上存在较大的估计偏差. 基于置信度估计与聚类分析的数据选择算法描述如算法 3 所示.

算法 3. 基于置信度估计与聚类分析的数据选择算法.

输入: 训练样本集 (x, y)

输出: 样本集 (x, y) 的选择标签 $Flag = \{\text{true}, \text{false}\}$

1. 初始化

(a) 将训练数据中每一评分等级作为一类, 计算训练样本与每一类 c 的相似度作为该样本的置信度估计度量, 记为 $f(c; x).$

2. 训练数据划分

(a) 利用 Kmeans 聚类算法将候选数据划分为 k 个子簇 $D_1, D_2, \dots, D_k.$

(b) 计算:

n_k^l : 聚类子簇 D_k 中的有标签样本的数目;

n_k^u : 聚类子簇 D_k 中的无标签样本的数目;

$n_{k,c}^l$: 属于类 c 且属于聚类子簇 D_k 中的有标签样本数目;

$n_{k,c}^u$: 属于类 c 且属于聚类子簇 D_k 中的无标签样本数目.

(c) 计算每一个聚类簇的先验概率:

$$P(D_k) = (n_k^l + n_k^u) / \sum_k (n_k^l + n_k^u).$$

(d) 计算针对每一类别和聚类簇的概率:

$$P(c | D_k) = \frac{\alpha P_l(c | D_k) + (1 - \alpha) P_u(c | D_k) + \beta}{Z},$$

其中 $P_l(c | D_k) = n_{k,c}^l / n_k^l, P_u(c | D_k) = n_{k,c}^u / n_k^u, \alpha$ 是 $P_l(c | D_k)$ 和 $P_u(c | D_k)$ 的平衡因子, β 是一个数值极小的常量, Z 是使 $P(c | D_k)$ 成为概率函数的归一化因子.

3. 数据选择

(a) 根据概率 $P(D_k)$ 选择聚类簇 D_k , 在聚类簇 D_k 内依据概率 $P(c | D_k)$ 选择类别 $c.$

(b) 求聚类簇 D_k 和类 c 的样本的交集, 选择具有最高置信度估计的前 m 个样本(令其 $Flag = \text{true}$).

(c) 重复(a)~(b)步直至筛选完所有候选预加入训练池的数据.

3.3.3 算法的时间复杂度分析

本文提出的基于协同训练的混合推荐算法, 其时间开销主要来源于用户评论的情感分析、物品内

容的相似度计算、基于物品的协同过滤模型及数据选择几个方面. 其中, 基于用户评论的情感分析是利用 LSTM 算法训练情感分类模型, 是通过离线训练生成, 不计入本文 HRSM 算法的时间复杂度. 假设在推荐系统中, 用户数为 m , 物品数量为 n , 物品内容的向量维度为 d , 则物品内容相似度计算的时间复杂度为 $O(d * n)$; 基于物品的协同过滤模型, 在最糟的情况下, 其时间复杂度为 $O(m * n)$, 但现实情况下, 由于用户评分数据的稀疏性, 其只需计算 $O(m+n)$ 次. 数据选择算法的时间复杂度为 $O(nlt)$, 其中 n 为物品数量, l 为代表聚类中心数, t 代表迭代次数, l 和 t 均为常数. 假设协同训练算法中, 协同训练的次数为 c (c 为常数), 则本文算法的计算次数为 $c * [(d * n) + (m * n) + n]$, 或 $c * [(d * n) + (m+n) + n]$. 由于在推荐系统中, 一般情况下用户数大于物品内容向量的维度 ($m > d$), 故本文提出的混合推荐算法的时间复杂度, 最糟的情况下为 $O(m * n)$, 即等同于基于领域的协同过滤推荐算法; 一般情况下为 $O(d * n)$, 即等同于基于内容的推荐算法.

4 实验结果与分析

4.1 实验数据集及评价指标

本文在 Amazon product data^[40] 上进行实验. 该数据集涵盖了从 1995 年 7 月至 2013 年 3 月来自亚马逊的约 3500 万条评论. 数据集包括: 用户信息、物品信息及纯文本评论信息. 其中, 用户信息由一些明确的用户元数据组成, 如名称、位置、用户等级、一些导出的统计信息等; 物品信息包括物品 Id、物品的标题、物品的价格等; 评论信息包括用户 Id、物品 Id、评论者、纯文本评论、物品评分、时间戳、用户的统计反馈等. 关于数据集的描述如表 1 所示.

表 1 Amazon product data 描述

| 数据集统计 | 数目 |
|----------------|-----------------|
| 评论条数 | 34 686 770 |
| 用户个数 | 6 643 669 |
| 物品个数 | 2 441 053 |
| 评论 > 50 条的用户个数 | 56 772 |
| 每条评论包含词的中位数 | 82 |
| 时间跨度 | 1995.07~2013.03 |

在基于用户评论的短文本信息的情感分析模型中, 本文利用信息检索中的准确率这一度量指标对情感分类实验进行评估. 在混合推荐模型中, 采用推荐系统中经典的评价指标均方误差 MSE (Mean-Squared Error). 假定测试集中的用户 u 和物品 i ,

r_{ui} 是用户 u 对物品 i 的实际评分, 而 \tilde{r}_{ui} 是预测评分, $|T|$ 为用户 u 对物品 i 的真实评分条数, 则推荐系统的评测指标 $MSE = \frac{1}{|T|} \sum_{u,i \in T} (r_{ui} - \tilde{r}_{ui})^2$.

4.2 情感分析实验

本文在 Amazon product data 上评测情感分类模型的性能. 我们从 Amazon 评论数据集中选取 Automotive、Baby 等 10 个数据集中的 12 000 条用户评论, 并人工标注每条评论文本的 Positive (标签为 5) 和 Negative (标签为 1) 情感标签, 用于模型训练和测试. 为了减少人工标注和校对的工作量, 本文从数据集中选取 6000 条评分为 5 或 4 (“overall”: 5.0 或 “overall”: 4.0) 的数据, 将其放入 Positive 的样本集中, 以供人工标注和校对; 同时选取 6000 条评分为 1 或 2 (“overall”: 1.0 或 “overall”: 2.0) 的数据, 将其放入 Negative 的样本集中, 以供人工标注和校对.

本文利用 Scikit-Learn 机器学习库构建情感分类模型, 选取 Scikit-Learn 中经典的 SVM 算法作为基分类算法. 在 SVM 算法中, 设置其核函数为高斯核 (kernel = ‘rbf’), 得到模型的平均准确率为 81.9%. 同时, 本文选取 TensorFlow + Keras 作为深度学习框架, 分别构建基于 LSTM 的情感分类模型和基于 LSTM 的用户评分嵌入方法的情感分类模型. 在评论文本中单词的分布向量表示方面, 利用 gensim 对 Amazon 数据集上所有的评论文本进行训练, 从而得到评论文本中单词的词向量表示. 在 $epochs = 20$ 时得到模型的最高平均准确率分别为 91.6% 和 91.8%. 情感分类模型的对比实验结果如表 2 所示.

表 2 情感分类模型的准确率

| 情感分类方法 | 评论情感 | 准确率/% |
|-------------------|----------|-------|
| SVM 方法 | Positive | 82.5 |
| | Negative | 81.3 |
| | Average | 81.9 |
| LSTM 方法 | Positive | 92.2 |
| | Negative | 91.0 |
| | Average | 91.6 |
| LSTM 方法 (评分嵌入) | Positive | 92.6 |
| | Negative | 91.0 |
| | Average | 91.8 |

从表 2 中的数据可以看出, 相比 SVM 算法, 利用 LSTM 训练的情感分类模型, 其准确率得到非常明显的提高, 平均准确率提高了近 10%. 本文也比较了基于 LSTM 的情感分类中嵌入用户评分的方式, 在相同的参数设置下, 其平均准确率与仅使用

评论文本训练模型相差不大,平均准确率提高了 0.2%。

为更清晰地展示 LSTM 算法训练得到的情感分类模型的性能,我们比较了在不同迭代次数中模型的准确率,分别设置 $epochs = \{1, 10, 20, 30, \dots, 100\}$, 采用 10 折交叉验证法对数据集进行评测,得到详细的情感分类模型性能指标如图 7 所示。

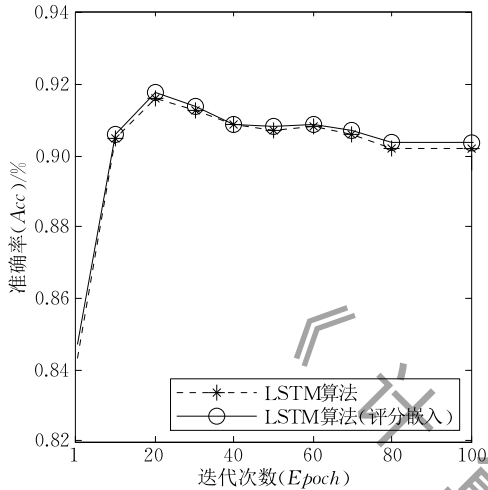


图 7 情感分类模型准确率

在图 7 基于 LSTM 和 Word2vec 的情感分类模型对比实验中,设置模型的损失函数为对数损失 (binary_crossentropy), 在训练模型时设置优化器为 adam, $batch_size = 32$, $dropout = 0.8$. 从图中的数据可以看出,当 $epochs = 20$ 时,嵌入用户评分的模型准确率达到最高点 91.8%。随着迭代次数的进一步增加,模型的准确率在 90% 以上起伏震荡,这表明模型性能较为平稳,且取得了较好的分类准确率。

4.3 推荐预测实验

4.3.1 对比算法

在本文的实验中,选取了 6 种较为经典的推荐算法,其中前 4 种推荐算法^[41]主要评测 MSE 性能指标,作为本文提出的 HRSM 算法的对比方法;后 2 种推荐算法主要基于 TopN 推荐,用来评估本文 HRSM 算法在推荐方面的性能。这 6 种对比推荐算法详细描述如下:

(1) ItemKNN 算法^[42]. 该方法将聚类物品分成若干类,并使用每个类别的平均评分作为算法的预测评分;

(2) MF 算法 (Matrix Factorization)^[43]. 该方法是一种特征分解算法,它对“用户-物品”评分矩阵进行分解,得到一个用户隐向量矩阵和一个物品隐向量矩阵,从而预测原评分矩阵中缺失的评分;

(3) SVD+十算法^[44]. 该方法是一种改进的奇异值分解 (SVD) 技术,在 SVD 的基础上引入隐式反馈,使用用户的历史浏览数据、用户历史评分数据等作为新的参数;

(4) HFT 算法 (Hidden Factors and Hidden Topics)^[40]. 它通过对用户评分和用户评论进行建模,在用户评论的主题分布和用户或物品的潜在因素间建立联系;

(5) DMF 算法 (Deep Matrix Factorization)^[45]. 利用评分和隐式反馈构建用户评分矩阵,将其作为输入,然后利用深度神经网络将用户和物品映射到共同低维空间;

(6) NCF 算法 (Neural Collaborative Filtering)^[46]. 一个基于协同过滤的用于模拟用户和物品潜在特征的神经网络框架,利用多层感知机制来学习用户-物品间的交互。

4.3.2 两种情感分析方法的 MSE 性能对比

为了实现用户评分与用户评论信息情感的融合,本文设计了两种方法:基于观点预过滤的方法和基于用户评分嵌入的方法。在观点预过滤的方法中,设置 $\alpha = 0.7$ 时推荐模型取得了最佳效果。利用本文提出的推荐模型,在 Amazon product data 的 10 个数据集上得到的预测评分结果如表 3 所示。

表 3 基于两种情感综合方法的评分预测结果 (MSE 值)

| 数据集 | 观点预过滤方法 | 评分嵌入方法 |
|---------------------|---------|--------|
| Automotive | 1.313 | 1.305 |
| Baby | 1.435 | 1.514 |
| Beauty | 1.373 | 1.341 |
| Electronics | 1.287 | 1.357 |
| Home and Kitchen | 1.132 | 1.156 |
| Kindle Store | 1.513 | 1.536 |
| Musical Instruments | 1.272 | 1.157 |
| Office Products | 1.312 | 1.336 |
| Pet Supplies | 1.598 | 1.672 |
| Sports and Outdoors | 0.966 | 1.015 |
| Average | 1.320 | 1.339 |

从表 3 中的实验数据来看,将上述两种情感融合方法引入到本文提出的 HRSM 算法中得到的实验结果比较接近,两者互有胜负。从平均成绩来看,基于观点预过滤的方法要好于评分嵌入方法,故本文后面的实验,在情感分析上采用的是基于观点预过滤的方法。

4.3.3 模型的评分预测结果

在推荐算法的实验环节,使用模型的预测评分与真实评分的平方误差 MSE 作为评价指标来衡量各种推荐算法的实验效果。在相同的 Amazon product

data 上,分别利用 ItemKNN 算法、MF 算法、SVD++ 算法、HFT 算法以及本文提出的 HRSM 算法进行比较实验. 在本文提出的 HRSM 算法中,设置取得最佳结果时的参数 $\alpha=0.6$, $K=60$. 此外在本节的 HRSM 算法中,采用观点预过滤方法的综合评分结果. 对于其它每一种推荐算法,参数亦设置为获得最佳结果时的参数. 本文使用 MyMediaLite 推荐系统库^[47]来获得 ItemKNN 算法、MF 算法、SVD++ 算法的实验结果,并使用论文中提供的源代码获得

HFT 算法的结果. 在 MF 算法中,设置 $num_factors=20$, $bias_reg=0.25$, $reg_u=0.4$, $reg_i=1.2$, $frequency_regularization=true$, $learn_rate=0.03$, $num_iter=80$, $bold_driver=true$. 在 SVD++ 算法中,设置 $num_factors=10$, $regularization=0.1$, $bias_reg=0.005$, $bias_learn_rate=0.007$, $learn_rate=0.01$, $num_iter=50$. 在测试集上得到的 MSE 指标如表 4 所示.

表 4 推荐模型的评分预测结果(MSE 值)

| 数据集 | ItemKNN 算法 | MF 算法 | SVD++ 算法 | HFT 算法 | HRSM 算法 |
|---------------------|------------|-------|----------|--------|---------|
| Automotive | 1.548 | 1.601 | 1.572 | 1.419 | 1.313 |
| Baby | 1.605 | 1.654 | 1.627 | 1.549 | 1.435 |
| Beauty | 1.612 | 1.687 | 1.616 | 1.373 | 1.407 |
| Electronics | 1.644 | 1.688 | 1.661 | 1.659 | 1.487 |
| Home and Kitchen | 1.635 | 1.705 | 1.673 | 1.533 | 1.432 |
| Kindle Store | 1.482 | 1.502 | 1.436 | 1.341 | 1.373 |
| Musical Instruments | 1.404 | 1.454 | 1.455 | 1.412 | 1.272 |
| Office Products | 1.676 | 1.787 | 1.754 | 1.617 | 1.412 |
| Pet Supplies | 1.679 | 1.705 | 1.671 | 1.682 | 1.699 |
| Sports and Outdoors | 1.321 | 1.357 | 1.331 | 1.118 | 0.966 |

表 4 显示了本文提出的算法和 ItemKNN 算法、MF 算法、SVD++ 算法、HFT 算法等 4 种经典算法在 Amazon product data 上实验结果. 总体来看,本文提出的 HRSM 算法在 MSE 评价指标上优于其它 4 种经典的推荐算法,其中 MF 算法总体表现最差,这表明一个简单的评分矩阵分解算法不太适合于在 Amazon 实验数据集上构建推荐模型. ItemKNN 算法的总体性能与 SVD++ 算法较相近, SVD++ 算法的性能较 MF 算法有较明显的提高,这表明在模型中增加用户历史数据可以提高推荐模型的性能,同时也表明本文期望利用用户评论等辅助信息提高推荐效果具有可行性.

进一步对表 4 中的实验数据分析,发现在 Amazon product data 中的 10 个实验数据集上,本文提出的 HRSM 算法相较其它 4 种较为经典的算法在 Automotive、Baby 等 7 个数据集上获胜. 分析计算显示:在 Automotive、Baby 数据集上,HRSM 算法相比最好的算法(HFT)提升了 7.47% 和 7.36%;在 Electronics 和 Musical Instruments 数据集上,HRSM 算法相比最好的算法(ItemKNN)提升了 9.55% 和 9.40%;研究发现这两类物品有较丰富的用户评论信息,在模型中引入这些隐式参数能有效地降低系统评分的均方差. 在 Home and Kitchen、Office Products 和 Sports and Outdoors 数据集上,HRSM 算法相比最好的算法(HFT)提升了 6.59%、

12.68% 和 13.60%;研究表明在基于物品内容推荐视图中融入用户评论信息有助于改善系统性能. 在其它 3 个数据集 Beauty、Kindle Store 和 Pet Supplies 上,HRSM 算法相比最好的算法分别有 2.48%、2.39%、1.19% 的降低. 分析表明在 Beauty、Kindle Store 和 Pet Supplies 这三个数据集上,HRSM 算法的效果都要比 MF 算法要好,但相比 HFT 算法和 SVD++ 算法有略微的降低. 可能的原因是:购买儿童和女性用品的用户多为女性,其对物品的质量要求比较严格,其兴趣偏好波动也比较大,故评论信息和兴趣偏好并非总是正相关. 这表明通过多推荐视图的融合并不总是能提高模型的性能,如果模型中引入一些不可靠的推荐因素,反而会对系统的性能起反作用.

综合上述分析数据,本文提出的 HRSM 算法在 MSE 评价指标上较传统算法有明显的改善,表明推荐模型的预测准确率与真实的用户评分相关,也表明利用基于观点预过滤的方法来融合虚拟评分并得到用户的综合评分,能有效地改善用户的评分准确性,并最终影响推荐模型的评分预测准确率. 此外,冷启动问题是推荐场景中的一个备受关注的问题,然而很少有记录(包括评分和评论)被认为与“冷启动”相关. 本文提出的 HRSM 算法结合了基于物品的协同过滤推荐和基于物品内容的推荐,在推荐因素中融入了用户评论的情感倾向,以及物品

内容的自然语言描述的语义信息,从理论上分析,这些辅助信息在一定程度上将有助力解决推荐系统的冷启动问题.

4.4 推荐预测中参数的影响

4.4.1 平衡因子 α 的影响

在本文提出的 HRSM 算法中,有一个重要的参数 α ,它反映了基于观点预过滤(opinion pre-filtering)方法中原始用户评分和用户评论的情感分析虚拟评分的权重问题.对该物品的情感倾向进行综合评分时利用公式 $Score_c = \alpha Score_r + (1 - \alpha) Score_o$,其中 $Score_o$ 是用户对物品的原始评分值, $Score_r$ 是模型预测的用户对物品的虚拟评分. α 的值越大,则表示情感分类模型预测的虚拟评分在综合评分中的权重越大.在本实验中,分别设置 α 的值从 0 到 1.0,步长为 0.1.在 10 个 Amazon 数据集上得到的实验结果如图 8、图 9 所示.在图 8、图 9 中,当 $\alpha=0$ 时,表示只考虑用户的原始评分,当 $\alpha=1$ 时,表示只考虑用户的虚拟评分.

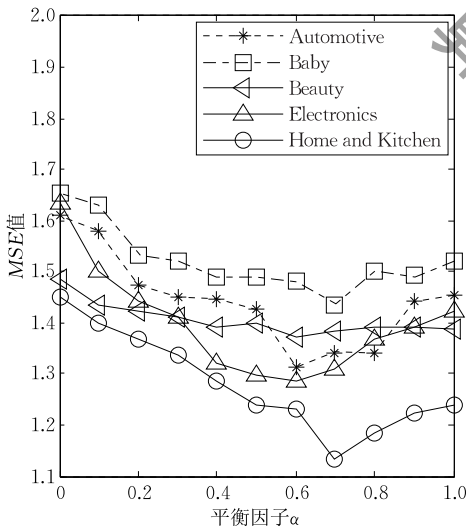


图 8 随着 α 的取值变化,模型在 Automotive 等 5 个数据集上的 MSE 值

从图 8 的数据可以看出, $\alpha=0.6$ 时,在 Automotive 等 5 个评论数据集中,有 4 个数据集的 MSE 值取得最小值;而 $\alpha=0.7$ 时,另一数据集 Home and Kitchen 的 MSE 值取得最小值.在基于观点预过滤(opinion pre-filtering)方法中, α 的值代表虚拟评分在综合评分中的权重.故图 8 中的实验数据表明:相较于用户的原始评分,虚拟评分在推荐预测评分模型中起着更重要的作用.图 9 表示模型在 Kindle Store 等 5 个数据集上随着 α 的取值变化而变化的 MSE 值.对图中的数据进行分析,我们几乎可以得出与图 8 相似的结论,即 $\alpha=0.6$ 或 $\alpha=0.7$ 时,模型

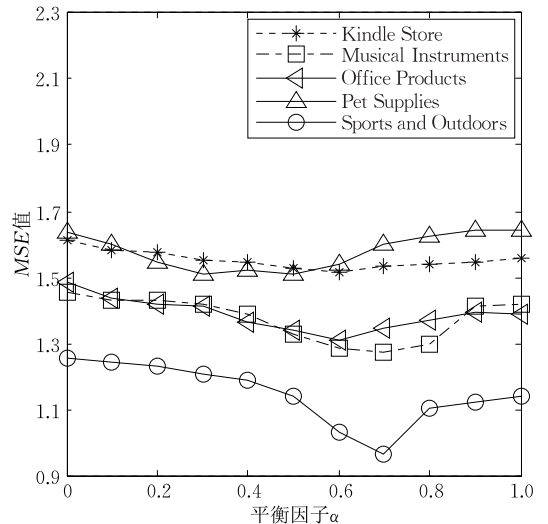


图 9 随着 α 的取值变化,模型在 Kindle Store 等 5 个数据集上的 MSE 值

取得了最小的 MSE 值.相较于当 $\alpha=0$ 时,模型只考虑用户的原始评分,当 $\alpha \neq 0$ 时(即考虑虚拟评分在综合评分中的作用),模型的 MSE 值得都有不同程度的降低,表明 α 对推荐系统的预测评分模型有着重要的影响.

对图 8 和图 9 中的实验数据进行分析,可以看出利用情感分类模型计算出来的虚拟评分对推荐预测评分模型的准确性有重要影响.这也在一定程度上验证了本文提出的设想,即在用户的原始评分较随意、且评分等级分布极度不均衡的情况下,用户的评论信息更能反映用户的真实兴趣偏好.采用基于观点预过滤方法对原始用户评分和用户评论的情感虚拟评分进行加权综合,可以在一定程度上解决用户原始评分与真实兴趣偏好存在较大偏差的问题,从而减少噪声数据对评分预测的估计偏差.

4.4.2 最近邻物品个数 K 的影响

本文利用协同训练策略对用户评分数据和物品内容描述信息进行融合,构建混合推荐系统.在基于物品内容推荐模型中,利用 KNN 算法来计算物品内容描述的距离,同时利用余弦相似度度量物品内容描述的相似度,以期利用 K 个最近邻物品的评分来更新或填充用户评分和缺省值.实验表明选取适当的 K 值对最终的推荐结果有重要影响.在本实验中,分别设 K 的值为 $\{5, 10, 20, 40, 60, 80, 100\}$,在 10 个 Amazon product data 上得到的实验结果如图 10、图 11 所示.

从图 10 的数据可以看出,在 $K=60$ 时,模型在 Automotive、Baby、Beauty、Electronics 等 4 个数据集的 MSE 值取得最小值;当 $K=80$ 时,模型的

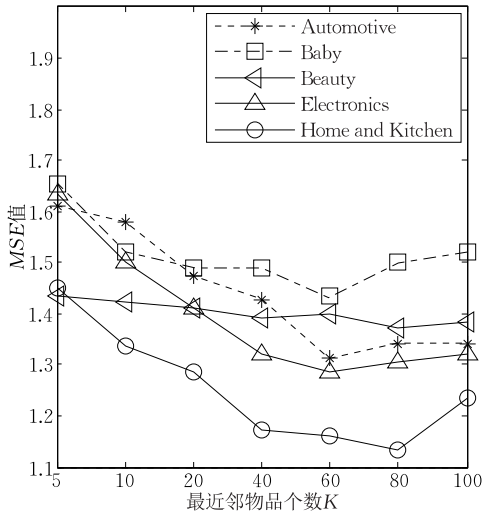


图 10 随着最近邻居数 K 的取值变化,模型在 Automotive 等 5 个数据集上的 MSE 值

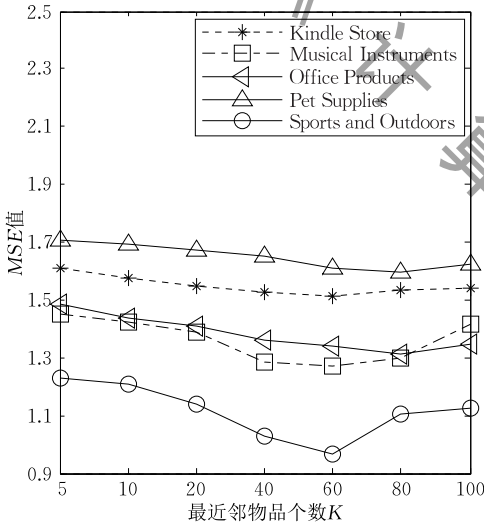


图 11 随着最近邻居数 K 的取值变化,模型在 Kindle Store 等 5 个数据集上的 MSE 值

MSE 值在 Home and Kitchen 数据集上取得最小.在上述 5 个数据集上,随着 K 值的继续增加,模型的 MSE 值也在增加,表明推荐模型的效果变差.图 11 是模型在 Kindle Store 等 5 个数据集上随着最近邻居数 K 的取值变化而变化的 MSE 值,从图中的数据可以得出,当 $K=60$ 时,模型在 5 个数据集上取得了最小的 MSE 值,而随着 K 值的继续增加,模型的 MSE 值也在增加,表明推荐模型的效果随着 K 值的增加反而变差.

从图 10 和图 11 中的数据可以看出,模型在 10 个数据集上的推荐效果与选取的最近邻居数 K 的取值有较大的关系.当然,在一些数据集上,推荐模型的 MSE 精度对 K 值也不是特别敏感,只在选择一

定的区域内,就可以获得比较好的 MSE 精度.例如,在本实验中,选择 K 的范围在 $[50, 70]$ 之间都是比较好的取值.故本文在利用 KNN 算法计算相似物品的内容描述时,即在表 5 中的对比实验结果中选取 $K=60$ 作为 HRSM 算法的参数.

4.5 推荐结果分析

为了评估推荐模型的性能,本文采用了大多数文献中广泛采用的留一法^[46] (leave-one-out). 本文使用命中率 HR (Hit Radio) 和标准化折扣累积增益 $NDCG$ (Normalized Discounted Cumulative Gain)^[26, 46] 来评估模型的性能.

$$HR@N = \frac{\text{Number of Hits}@N}{|GT|}, \text{其中分母表示}$$

候选测试集的数量,分子表示每个用户前 N 个被推荐物品中属于测试集物品列表的个数总和.

$NDCG_N = \frac{DCG_N}{idealDCG_N}$, 其中分子 DCG_N 表示按模型排名的计算结果,分母 $idealDCG_N$ 表示按实际等级排名的计算结果. $DCG_N = \sum_{i=1}^N \frac{2^{rel_i} - 1}{\log_2(i+1)}$, 其中 rel_i 表示第 i 个位置的“等级关联性”, i 表示推荐结果的位置.

在本节实验的 TopN 推荐中,选取 100 个用户未评定过且与用户喜欢的物品最相似的物品作为候选物品,在前 4 种推荐算法及本文的 HRSM 算法中,利用评分预测模型中得到的用户评分 $Score_{ui}$, 以及候选推荐物品与目标用户 u 喜欢物品的相似度 Sim_{ij} 的乘积作为 TopN 推荐排序的主要依据,同时也考虑推荐物品的时效性.在 DMF 算法中,使用交互矩阵 \mathbf{Y} 作为输入,构建一个 2 层的模型,设置顶层的因子数为 64.在 NCF 算法中,深度神经网络使用一个 4 层的全连接层,设置为 $layers = [64, 32, 16, 8]$, 正负实例比为 1:4,使用高斯分布(均值为 0, 标准差为 0.01)随机初始化模型参数,并用 mini-batch Adam 优化模型.得到的不同推荐算法的 $HR@N$ 和 $NDCG@N$ 性能比较如表 5.

表 5 显示了不同的推荐算法在 TopN 推荐中的最佳实验结果.从实验结果可以看出,本文提出的 HRSM 算法在 $HR@N$ 评价指标上取得了较其它算法更好的性能.相较于当前较流行的基于深度学习的 DMF 算法和 NCF 算法,由于本文提出的算法挖掘了用户评论信息,以及在协同训练模型中利用了物品内容的描述文本信息,所以可以较好地克服推荐系统的冷启动问题,因此在反映推荐系统的召

表 5 不同算法的 $HR@N$ 和 $NDCG@N$ 比较

| 评测指标 | ItemKNN 算法 | MF 算法 | SVD++ 算法 | HFT 算法 | DMF 算法 | NCF 算法 | HRSM 算法 |
|-----------|------------|--------|----------|--------|--------|--------|---------|
| $HR@5$ | 0.1897 | 0.3027 | 0.3082 | 0.3035 | 0.2693 | 0.3117 | 0.3156 |
| $NDCG@5$ | 0.1279 | 0.2068 | 0.2063 | 0.1917 | 0.1848 | 0.2141 | 0.2163 |
| $HR@10$ | 0.3126 | 0.4278 | 0.3225 | 0.3206 | 0.3715 | 0.4309 | 0.4493 |
| $NDCG@10$ | 0.1672 | 0.2471 | 0.2483 | 0.2319 | 0.2179 | 0.2524 | 0.2545 |
| $HR@15$ | 0.3901 | 0.5054 | 0.4672 | 0.4409 | 0.4328 | 0.5258 | 0.5487 |
| $NDCG@15$ | 0.1877 | 0.2676 | 0.2701 | 0.2685 | 0.2332 | 0.2774 | 0.2689 |
| $HR@20$ | 0.4431 | 0.5680 | 0.5692 | 0.5610 | 0.4850 | 0.5897 | 0.5932 |
| $NDCG@20$ | 0.2002 | 0.2824 | 0.2749 | 0.2606 | 0.2458 | 0.2925 | 0.2906 |

回率性能($HR@N$)上取得了较好的成绩. 在 $NDCG@N$ 评价指标上, HRSM 算法和 NCF 算法各取得了两项最佳成绩. NCF 利用深度神经网络构建用户与物品之间的交互, 在学习二者交互的潜在因子方面表现出了较强的能力, 在 $NDCG@15$ 和 $NDCG@20$ 指标上获得了较 HRSM 算法更好的性能. 同时, 在挖掘用户偏好的时效性方面, 考虑到本文实验数据的涵盖了 1995 年至 2013 年, 时间跨度非常大. 本文在利用观点预过滤的方法计算用户对物品的综合评分时, 加入了对用户评论的时间维度, 遵循时间越近, 权重因子越大, 时间越远, 权重因子越小的线性函数; 同时在进行 TopN 推荐排序时, 也考虑了推荐物品的时效性因素. 实验表明, 在推荐过程中考虑时间跨度对最终的推荐结果有较重要的影响.

对表 5 的实验数据进一步分析, 通过在推荐模型中融入用户评论信息、物品的内容描述信息等构建混合推荐系统, 较传统的基于物品的协同过滤算法及利用主题分布信息等构建模型在各方面性能上均有较大的提升, 这也验证本文提出的通过融合多个推荐视图构建混合推荐系统的设想的可行性. 在反映推荐系统的召回率性能($HR@N$)上, HRSM 算法取得了最佳的成绩, 表明 HRSM 算法可在一定程度上解决推荐系统的冷启动问题.

4.6 推荐算法的时间和内存消耗比较

在推荐系统时间消耗方面, 大多数推荐算法是基于 RMSE 值最小化的目标, 因此评分矩阵的预测与循环赋值是一个重要的时间优化目标. 本文实验使用的计算机平台是 64 位操作系统, Intel i7-8700 CPU, 8GB 内存, NVIDIA GTX1080 8G. 实验中随机选择了 Automotive 和 Electronics 两个数据集, 在 DMF 和 NCF 模型中, 设置 $epochs=100$, 将深度学习模型的训练时间作为算法的时间消耗, 得到的几种推荐算法的时间消耗比较如表 6 所示.

表 6 推荐算法的时间消耗比较 (单位: min)

| 数据集 | Automotive | Electronics |
|------------|------------|-------------|
| ItemKNN 算法 | 31.5 | 206.8 |
| MF 算法 | 11.9 | 80.7 |
| SVD++ 算法 | 12.6 | 82.8 |
| HFT 算法 | 20.8 | 220.2 |
| DMF 算法 | 26.7 | 185.6 |
| NCF 算法 | 23.6 | 156.1 |
| HRSM 算法 | 33.6 | 180.3 |

从表 6 中的数据可以看出, MF 和 SVD++ 算法速度最快, 因为他们使用了梯度下降的方法进行近似计算, 大大减少了时间消耗; DMF 和 NCF 等深度学习算法利用 GPU 加速, 其训练速度也比 HFT 算法略慢; 本文提出的 HRSM 算法与 ItemKNN 算法的处理时间相当. 综合比较而言, 在 Automotive 数据集上, HRSM 算法训练时间最长, 但在 Electronics 数据集上 (Electronics 数据集是 Automotive 数据集的 9 倍), HRSM 算法训练时长要小于 ItemKNN、HFT 和 DMF 算法. 这表明随着数据集的加大, HRSM 算法的时间消耗下降速度更快. 考虑到用户的评分矩阵预测和赋值在整个推荐过程中是离线进行的, 故不影响 TopN 推荐最终的实时性.

在内存消耗方面, 由于 MF 和 SVD++ 算法需要将一个用户评分矩阵分解为三个大的矩阵, 这三个矩阵都需存储在计算机内存中, 所以需要消耗较多的内存; ItemKNN 算法只需存储物品的内容信息, HRSM 算法也只需要存储用户评分矩阵和物品内容信息, 且这里的物品内容信息用低维向量表示, 内存空间消耗远小于 MF 和 SVD++ 等基于矩阵因式分解的方法. DMF 和 NCF 等深度学习算法在训练过程中由于参数众多、需要进行批量训练, 故内存消耗是最大的.

5 总结与展望

推荐系统作为解决信息过载的最有效工具, 其

在当前的学术界和工业界都备受关注. 一直以来协同过滤推荐和基于内容的推荐是两种最重要的推荐方法, 但单一的协同过滤推荐模型或多或少地受评分数据稀疏、冷启动等制约; 而基于内容的推荐, 其推荐因素通常不够充分, 特别是当目标用户可用的历史数据较少时. 融合多视图的兴趣偏好信息构建混合推荐模型是个性化推荐研究发展的趋势. 鉴于用户评论、物品内容描述等短文本的情感及语义难以分析, 单一视图的推荐模型易导致对用户画像的建模粗放等问题. 本文在对用户行为偏好进行分析的基础上, 将研究聚焦于文本信息的情感挖掘和深层语义分析; 同时挖掘物品内容的自然语言描述信息, 并结合半监督学习中的协同训练的策略实现基于物品的协同过滤推荐视图和基于内容推荐视图的融合, 来构建混合推荐系统. 该方法一定程度上解决了基于模型推荐方法缺乏足够有标签数据的问题. 在 Amazon product data 上的对比实验表明本文提出的综合情感评分的有效性, 有助于解决推荐系统中用户原始评分与真实兴趣偏好存在偏差、且评分等级分布极度不均衡的问题. 在推荐效果方面, 将本文提出的 HRSM 算法与 4 种较为经典的推荐算法 ItemKNN 算法、MF 算法、SVD++ 算法、HFT 算法进行对比, 结果表明本文提出的 HRSM 算法在均方误差评价指标 (MSE) 上有显著的改善. 在 TopN 推荐上, 通过与上述 4 种推荐算法和当前较流行的基于深度学习的 DMF 算法和 NCF 算法进行对比实验, 结果显示本文提出的 HRSM 算法在 $HR@N$ 和 $NDCG@N$ 均取得了较好的成绩. 这表明在推荐模型中融入物品的内容描述视图, 及对辅助信息建模, 可以提高推荐的准确性, 一定程度上解决推荐系统面临的冷启动问题.

本文提出了基于用户评论的深度情感分析与多源推荐视图协同融合的混合推荐方法, 从上下文层面对用户评论的短文本进行语义分析和情感计算. 通过利用段落向量对物品内容描述的短文本进行相似度计算, 并基于协同训练的策略来解决多推荐视图的融合问题. 由于本文提出的 HRSM 推荐算法兼顾了物品的内容信息, 故有效地解决了推荐系统的冷启动问题, 同时提高了推荐系统的推荐召回率. 下一步的工作, 我们拟从互联网电商网站上收集并整理推荐系统数据集, 对本文提出的推荐算法进行评测、改进和参数调优设置, 设计合理的推荐过滤和排序算法, 进一步用本文提出的 HRSM 推荐算法在推荐准确率和召回率等方面进行实验.

参 考 文 献

- [1] Cremonesi P, Koren Y, Turrin R. Performance of recommender algorithms on top-N recommendation tasks//Proceedings of the 4th ACM Conference on Recommender Systems. Barcelona, Spain, 2010; 39-46
- [2] Wang H, Wang N, Yeung D Y. Collaborative deep learning for recommender systems//Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney, Australia, 2015; 1235-1244
- [3] Wei J, He J, Chen K, et al. Collaborative filtering and deep learning based recommendation system for cold start items. Expert Systems with Applications, 2017, 69: 29-39
- [4] Bobadilla J, Ortega F, Hernando A, et al. Recommender systems survey. Knowledge-Based Systems, 2013, 46: 109-132
- [5] Hu Zhong-Kai, Zheng Xiao-Lin, Wu Ya-Feng, et al. Product recommendation algorithm based on user's reviews mining. Journal of Zhejiang University (Engineering Science), 2013, 47(8): 1475-1485(in Chinese)
(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. 浙江大学学报(工学版), 2013, 47(8): 1475-1485)
- [6] Wang Zhi-Sheng, Li Qi, Wang Jing, et al. Real-time personalized recommendation based on implicit user feedback data stream. Chinese Journal of Computers, 2016, 39(1): 52-64(in Chinese)
(王智圣, 李琪, 汪静等. 基于隐式用户反馈数据流的实时个性化推荐. 计算机学报, 2016, 39(1): 52-64)
- [7] Zhang F, Yuan N J, Lian D, et al. Collaborative knowledge base embedding for recommender systems//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, 2016; 353-362
- [8] Huang Zhen-Hua, Zhang Jia-Wen, Tian Chun-Qi, et al. Survey on learning-to-rank based recommendation algorithm. Journal of Software, 2016, 27(3): 691-713(in Chinese)
(黄震华, 张佳雯, 田春岐等. 基于排序学习的推荐算法研究综述. 软件学报, 2016, 27(3): 691-713)
- [9] Karatzoglou A, Baltrunas L, Shi Y. Learning to rank for recommender systems//Proceedings of the 7th ACM Conference on Recommender Systems. Hong Kong, China, 2013; 493-494
- [10] Zhang Y, Zhang H, Zhang M, et al. Do users rate or review?: Boost phrase-level sentiment labeling with review-level sentiment classification//Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval. Gold Coast, Australia, 2014; 1027-1030
- [11] Chen Long, Guan Zi-Yu, He Jin-Hong, et al. A survey on sentiment classification. Journal of Computer Research and Development, 2017, 54(6): 1150-1170(in Chinese)
(陈龙, 管子玉, 何金红等. 情感分类研究进展. 计算机研究与发展, 2017, 54(6): 1150-1170)
- [12] Zhang W, Wang J. A collective Bayesian Poisson factorization model for cold-start local event recommendation//Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney, Australia, 2015; 1455-1464
- [13] Huang Li-Wei, Jiang Bi-Tao, Lv Shou-Ye, et al. Survey on

- deep learning based recommender systems. *Chinese Journal of Computers*, 2018, 41(7): 1619-1647(in Chinese)
(黄立威, 江碧涛, 吕守业等. 基于深度学习的推荐系统研究综述. *计算机学报*, 2018, 41(7): 1619-1647)
- [14] Li Lin, Liu Jin-Hang, Meng Xiang-Fu, et al. Recommendation models by exploiting rating matrix and review text. *Chinese Journal of Computers*, 2018, 41(7): 1559-1573(in Chinese)
(李琳, 刘锦行, 孟祥福等. 融合评分矩阵与评论文本的商品推荐模型. *计算机学报*, 2018, 41(7): 1559-1573)
- [15] Xie Xin-Qiang, Yang Xiao-Chun, Wang Bin, et al. Multi-feature fused software developer recommendation. *Journal of Software*, 2018, 29(8): 2306-2321(in Chinese)
(谢新强, 杨晓春, 王斌等. 一种多特征融合的软件开发者推荐. *软件学报*, 2018, 29(8): 2306-2321)
- [16] Pan Yi-Teng, He Fa-Zhi, Yu Hai-Ping. Social recommendation algorithm using implicit similarity in trust. *Chinese Journal of Computers*, 2018, 41(1): 65-81(in Chinese)
(潘一腾, 何发智, 于海平. 一种基于信任关系隐含相似度的社会化推荐算法. *计算机学报*, 2018, 41(1): 65-81)
- [17] Kong Xin-Xin, Su Ben-Chang, Wang Hong-Zhi, et al. Research on the modeling and related algorithms of label-weight rating based recommendation system. *Chinese Journal of Computers*, 2017, 40(6): 1440-1452(in Chinese)
(孔欣欣, 苏本昌, 王宏志等. 基于标签权重评分的推荐模型及算法研究. *计算机学报*, 2017, 40(6): 1440-1452)
- [18] Tian Chao, Qin Zuo-Yan, Zhu Qing, et al. SuperRank: An intelligent recommendation system based on review analysis. *Journal of Computer Research and Development*, 2010, 47(1): 494-498(in Chinese)
(田超, 覃左言, 朱青等. SuperRank: 基于评论分析的智能推荐系统. *计算机研究与发展*, 2010, 47(1): 494-498)
- [19] Shmueli E, Kagian A, Koren Y, et al. Care to comment?: Recommendations for commenting on news stories//*Proceedings of the 21st International Conference on World Wide Web*. Lyon, France, 2012: 429-438
- [20] Zhang Y, Tan Y, Zhang M, et al. Catch the black sheep: Unified framework for shilling attack detection based on fraudulent action propagation//*Proceedings of the 24th International Joint Conference on Artificial Intelligence*. Buenos Aires, Argentina, 2015: 2408-2414
- [21] Wu Y, DuBois C, Zheng A X, et al. Collaborative denoising auto-encoders for top-N recommender systems//*Proceedings of the 9th ACM International Conference on Web Search and Data Mining*. San Francisco, USA, 2016: 153-162
- [22] Chen L, Chen G, Wang F. Recommender systems based on user reviews: the state of the art. *User Modeling and User-Adapted Interaction*, 2015, 25(2): 99-154
- [23] Zhang W, Yuan Q, Han J, et al. Collaborative multi-level embedding learning from reviews for rating prediction//*Proceedings of the 25th International Joint Conference on Artificial Intelligence*. New York, USA, 2016: 2986-2992
- [24] Zheng L, Noroozi V, Yu P S. Joint deep modeling of users and items using reviews for recommendation//*Proceedings of the 10th ACM International Conference on Web Search and Data Mining*. Cambridge, United Kingdom, 2017: 425-434
- [25] Chen C, Zhang M, Liu Y, et al. Neural attentional rating regression with review-level explanations//*Proceedings of the 2018 World Wide Web Conference on World Wide Web*. Lyon, France, 2018: 1583-1592
- [26] Han X, Shi C, Wang S, et al. Aspect-level deep collaborative filtering via heterogeneous information networks//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm, Sweden, 2018: 3393-3399
- [27] Liu Zhi-Yuan, Sun Mao-Song, Lin Yan-Kai, et al. Knowledge representation learning: A review. *Journal of Computer Research and Development*, 2016, 53(2): 1-16(in Chinese)
(刘知远, 孙茂松, 林衍凯等. 知识表示学习研究进展. *计算机研究与发展*, 2016, 53(2): 1-16)
- [28] Wang P, Xu J, Xu B, et al. Semantic clustering and convolutional neural network for short text categorization//*Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*. Beijing, China, 2015: 352-357
- [29] Wang Zhong-Yuan, Chen Jian-Peng, Wang Hai-Xun, Wen Ji-Rong. Short text understanding: A survey. *Journal of Computer Research and Development*, 2016, 53(2): 262-269(in Chinese)
(王仲远, 陈健鹏, 王海勋, 文继荣. 短文本理解研究. *计算机研究与发展*, 2016, 53(2): 262-269)
- [30] Ma W, Zhang M, Wang C, et al. Your tweets reveal what you like: Introducing cross-media content information into multi-domain recommendation//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm, Sweden, 2018: 3484-3490
- [31] Lian J, Zhang F, Xie X, et al. Towards better representation learning for personalized news recommendation: A multi-channel deep fusion approach//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm, Sweden, 2018: 3805-3811
- [32] Nguyen T T, Hui P M, Harper F M, et al. Exploring the filter bubble: The effect of using recommender systems on content diversity//*Proceedings of the 23rd International Conference on World Wide Web*. Seoul, Korea, 2014: 677-686
- [33] Wu Z, Wu J, Cao J, et al. HySAD: A semi-supervised hybrid shilling attack detector for trustworthy product recommendation//*Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Beijing, China, 2012: 985-993
- [34] Zhang M, Tang J, Zhang X, et al. Addressing cold start in recommender systems: A semi-supervised co-training algorithm//*Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*. Gold Coast, Australia, 2014: 73-82
- [35] Ding J, Yu G, He X, et al. Improving implicit recommender systems with view data//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm, Sweden, 2018: 3343-3349
- [36] Wang X, He X, Feng F, et al. TEM: Tree-enhanced embedding model for explainable recommendation//*Proceedings of the 2018 World Wide Web Conference on World Wide Web*. Lyon, France, 2018: 1543-1552
- [37] Pero Š, Horváth T. Opinion-driven matrix factorization for

- rating prediction//Proceedings of the 21st International Conference on User Modeling, Adaptation, and Personalization, Rome, Italy, 2013: 1-13
- [38] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality//Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, USA, 2013: 3111-3119
- [39] Le Q, Mikolov T. Distributed representations of sentences and documents//Proceedings of the International Conference on Machine Learning, Beijing, China, 2014: 1188-1196
- [40] McAuley J, Leskovec J. Hidden factors and hidden topics: Understanding rating dimensions with review text//Proceedings of the 7th ACM Conference on Recommender Systems, Hong Kong, China, 2013: 165-172
- [41] Qiu L, Gao S, Cheng W, et al. Aspect-based latent factor model by integrating ratings and reviews for recommender system. Knowledge-Based Systems, 2016, 110: 233-243
- [42] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms//Proceedings of

the 10th International Conference on World Wide Web, Hong Kong, China, 2001: 285-295

- [43] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer, 2009, 42(8): 42-49
- [44] Koren Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model//Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, USA, 2008: 426-434
- [45] Xue H J, Dai X, Zhang J, et al. Deep matrix factorization models for recommender systems//Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 2017: 3203-3209
- [46] He X, Liao L, Zhang H, et al. Neural collaborative filtering //Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 2017: 173-182
- [47] Gantner Z, Rendle S, Freudenthaler C, et al. MyMediaLite: A free recommender system library//Proceedings of the 5th ACM Conference on Recommender Systems, Chicago, USA, 2011: 305-308



ZHANG Yi-Hao, Ph.D., associate professor. His research interests include recommender system, machine learning and natural language processing.

ZHU Xiao-Fei, Ph.D., professor. His research interests include big data search and recommendation, web mining, machine learning.

XU Chuan-Yun, Ph.D., associate professor. His research interests include machine learning and image processing.

DONG Shi-Du, Ph.D., associate professor. His research interests include machine learning and image processing.

Background

With the rapid development of E-commerce and social networks, information overload is becoming a serious problem. As the most effective tool for solving information overload, recommendation systems have attracted attention in both academia and industry. Currently, most of the various types of recommender techniques use user-provided ratings to infer user preferences. Collaborative filtering techniques perform well when there is sufficient rating information. However, their effectiveness is limited when the well-known rating sparsity problem occurs, due to the poor coverage of recommendation space, or the difficulty in letting users express their preferences as scalar ratings on items. Ours research on user's ratings also shows that there is a great deviation between the users' rating and real interest preference. Content-based recommender approaches have been developed that rely instead on the content representations of items to locate items that have similar content to items the target user liked. However, these methods are still inadequate, especially when the target user has little historical data. They are also of limited usefulness when the sparsity level of overall data is high. At present, it is a recent development trend to do personalized recommendation through fusing multi-view of interest preferences to build the hybrid recommendation model, as a single view of the recommended model lead to

user reach is extensive.

In this paper, we focus on the following issues; (1) study the distributed representation of the short text based on the paragraph vector, and realize the sentiment analysis of the user review on the context semantic level. The proposed method will solve the problem that there is a great deviation between the user's rating and real interest preference, and also solve the extreme imbalance problem of the user rating distribution; (2) study the similarity calculation method of short text based on word embedding and paragraph vector, and study the fusion method of recommendation view based on collaborative training. The proposed method will address the challenge of similarity calculation of item contents, and solve the problem of the lack of sufficient labeled data for modeling. We study recommendation on the basis of user's behavior analysis, which will focus on sentiment mining and deep semantic analysis of the text information, and will consequentially achieve the integration of multi-recommended views and the personalized needs of users.

This work is supported by the National Natural Science Foundation of China (No. 61702063), the Key Project of Chongqing Basic Science and Frontier Technology Research (No. cstc2017jcyjBX0059).