

融合评分矩阵与评论文本的商品推荐模型

李琳¹⁾ 刘锦行¹⁾ 孟祥福²⁾ 苏畅¹⁾ 李鑫³⁾ 钟璐¹⁾

¹⁾(武汉理工大学 计算机科学与技术学院, 武汉 430070)

²⁾(辽宁工程技术大学 电子与信息工程学院, 辽宁 葫芦岛 125105)

³⁾(科大讯飞大数据研究院, 合肥 230088)

摘要 随着电子商务的迅速发展和 Web 上信息的不断增长, 推荐系统成为有效帮助用户做出决策的重要智能软件之一。基于评分数据的矩阵分解推荐模型被广泛研究和应用, 但数据稀疏性问题影响了该模型的推荐质量。另一方面推荐系统中用户撰写的评论文本可以反映用户的兴趣偏好, 有研究工作提出了相应的文本分析及观点挖掘方法来缓解评分数据的稀疏性。评分矩阵分解模型与评论文本挖掘模型的融合有助于提高推荐质量, 因此本文首先提出了一种融合评分数值和评论文本的推荐模型 DTMF (Double Topics with Matrix Factorization), 将用户评论集和商品评论集各自的潜在主题向量分别与传统矩阵分解的用户潜在因子向量和商品潜在因子向量建立正向映射关系, 然后通过添加潜在主题为预测评分引导项进一步优化 DTMF 模型提出了 DTMF+模型。在两组公开数据集上, 以推荐结果的均方误差 (MSE) 为评估指标进行了实验验证。实验结果表明本文提出的 DTMF 和 DTMF+ 两个模型整体上优于仅融入商品评论集的 HFT(item) 模型, 在子类数据上预测误差最大分别降低了 3.68% 和 7.31%。最后本文探讨了有用性评论排序问题来增强推荐结果的可解释性。

关键词 评分矩阵; 评论文本; 矩阵分解; 主题模型; 均方误差

中图法分类号 TP18

Recommendation Models by Exploiting Rating Matrix and Review Text

LI Lin¹⁾ LIU Jin-Hang¹⁾ MENG Xiang-Fu²⁾ SU Chang¹⁾ LI Xin³⁾ ZHONG Luo¹⁾

¹⁾(School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China)

²⁾(School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China)

³⁾(iFLYTEK Big Data Research Institute, Hefei 230088 China)

Abstract The explosive growth of information available on the World Wide Web and the rapid introduction of new E-business services have frequently overwhelmed Internet users, leading them to make poor decisions without recommendation systems. With the successful application of recommendation systems, more and more users participate actively in their online life. How to utilize user-generated information has become one way to improve the quality of recommendation systems. Usually rating data is used as a direct feedback and matrix decomposition is widely used to predict ratings in recommendation systems. However the sparsity problem of rating data has largely affected recommendation performance. Making full use of user-generated information like review text has become one way to enhance the recommendation performance in terms of accuracy. As we know, the reviews written by users can represent users' preferences. Recently researchers have shown achievements in improving recommendation quality by utilizing review text analysis and opinion mining

methods. Therefore, it is interesting to combine rating data based matrix decomposition model and review data based text mining model for improving recommendation quality. This paper studies how to exploit rating matrix and review text simultaneously, and an improved recommendation model, called DTMF(Double Topics with Matrix Factorization) is firstly proposed. DTMF model learns latent rating factors obtained from rating matrix decomposition with latent review topics inferred from a topic model like LDA at the same time and build a monotonic transform between them. By doing so, the two latent factors are combined together. The information inside review texts is richer than a single rating score. Therefore, the proposed DTMF model can better alleviate the data sparsity than those only consider rating scores. Moreover, its enhancement model, i.e., DTMF+, adds latent topics as user rating preferences for further improvement. DTMF+ thinks that latent topics from the review collection of users reflect users' preferences, while latent topics from the review collection of items reflect merchant items' characteristics. The predicted rating score for a user should also consider this two kinds of learned topics. Therefore, the objective function of DTMF+ considers them to accurately model users' ratings. Experiments are conducted on two public data sets. Both of them can be downloaded from Stanford Network Analysis Project. One is Fine food reviews data with the size of 299M, and the other is the review and rating of Amazon.com with the size of 3.3G. Experimental results show that our DTMF and DTMF+ models reduce the recommendation errors of the popular HPT (item) model in terms of mean square error in average, and their best improvements on some category data are up to 3.68% and 7.31%, respectively. In terms of running time, the model training time of the proposed DTMF and DTMF+ is about 2-3 times more than the HPT (item). In online recommendation, the latent factors got from the trained model can be directly used for dot product operation, so the proposed models are applicable in real world applications. In addition, given an item we discuss how to sort its reviews and some preliminary experimental results are reported.

Keywords: rating matrix; review text; matrix decomposition; topic model; mean square error

1 引言

伴随信息技术的发展, 互联网上的数据呈指数级增长, 如何有效地将相关信息提供给需要的用户面临巨大挑战。推荐系统能够满足用户的个性化信息需求, 得到了广泛的研究和推广应用。近年来, 随着用户参与度不断的提高, 产生了大量的用户生成信息(User generated content, i.e., UGC), 如何利用 UGC 信息来优化推荐模型和提高推荐质量成为研究的热点问题之一。

主流的协同过滤推荐方法是基于显性评分等用户生成信息进行建模, 虽然具有较好的推荐质量, 但是存在评分数据的稀疏性问题。研究者讨论了和分析了各种 UGC 信息对推荐模型的影响, 例如用户的社交信息^[1,2]、隐性用户反馈^[3]、评论文本的情感^[4]以及用户对实体评论的文本内容^[5], 使得推荐模型不断地被改进和优化。

如前所述的 UGC 信息中, 作为反映用户喜好和意见的评论文本受到了广泛的关注。与之相关的

文本分析与观点挖掘方法被提出, 并传统基于显性评分的协同过滤推荐方法结合^[6]。评论文本的加入能够有效改善评分数据的稀疏性问题, 从而提高推荐系统的推荐质量^[7-10]。因此, 研究融合评分数值与评论文本的推荐模型具有理论意义和应用价值, 然而如何挖掘评论文本主题并与传统的评分矩阵分解方法有效融合是该类推荐模型的难点问题之一^[11-14]。

另一方面, 目前研究者对推荐系统的评估主要考虑“准确度”, 根据预测值的类型分为两类任务。一类任务的预测评分为实数值, 以均方误差、均方根误差和平均绝对误差等评价指标为代表。另一类是非 0 即 1 的二值预测, 以正确率和召回率等为主流的评价指标。此外还有点击率和购买率等^[15], 但是对于推荐系统“可解释性”方面的评价关注度不足。用户在得到推荐列表的同时也希望了解推荐系统为何提供这样的结果。因此, 需要增强推荐系统结果的解释性, 帮助用户做出正确的决策^[16]。

产品： 电脑	
5个评价方面： 价格 外观 材质 硬件配置 系统软件	
用户 1	打分评级 评论内容
用户 ID: audrey535	star 5 感觉还是值这个钱的,建议大家还是不要重装系统,我看了一般的游戏都能玩,但也有确实玩不了的,比如lol三国无双7什么的,但是网上都有win8.1的解决办法。照着弄下就可以了。整个机器很霸气.....就是有点重,加电源12斤。
购买日期: 2014-06-01	
评论日期: 2014-06-03	
用户 2	打分评级 评论内容
用户 ID: jd_山东小熊	star 5 性能没的说!!!功能强大,很强大!!基于性能,重量的缺点就不算缺点了,只是电源适配器到电脑的电源接口略松!!不知道是个别情况还是都这样?移动的台式电脑还有,非常非常非常的漂亮大气!!!除了电源口略松,其他全非常满意!!!
购买日期: 2014-05-31	
评论日期: 2014-06-02	

图 1: 某电子商务平台上对电脑商品的评价数据

图 1 中的例子来自于中国某电子商务平台,包含了用户对电脑产品的评分和评论数据。相比于评分数值(图 1 中的“star5”),评论文本表达了用户对商品不同方面的关注和偏好。用户 1 和用户 2 从价格、外观、材质、硬件配置和系统软件等 5 个方面进行了评论,总体比较满意给了 5 分的打分。从这个例子观察到评分数值和评论文本都可以帮助推荐系统了解用户。传统基于评分矩阵隐因子分解的推荐模型对于新用户和新商品缺乏足够的评分数据来准确学习其多个潜在特征(因子),而且评分数据为数值类型,信息单一。相比之下评论文本天然具有丰富的信息,对 Amazon 每条评论的单词个数进行统计的结果表明,有 70% 以上评论的单词个数大于 20。以关键词为主体的评论文本有利于潜在特征(主题)的学习。用户评论文本集反映了用户的打分偏好和习惯,可以通过挖掘评论文本的潜在主题(图 1 中的 5 个评价方面)来弥补单一评分数值信息量不足的缺点。本文优化融合评分数值与评论文本的推荐模型,通过潜在主题和潜在因子的同步学习,将高维稀疏的特征空间映射到相对稠密的低维语义特征空间,缓解了原始数据稀疏的问题,进而提高预测评分的准确度。同时建模学到的低维语义特征可以看作影响用户评分的不同方面和角度。

McAuley J 等人^[11]提出的 HFT(Hidden Factors as Topics)模型是目前理论和实验效果上都比较出色的融合模型。该模型采用文本分析中的主题模型挖掘用户评论文本(或者商品评论文本)的潜在主题,对评分数值采用矩阵分解模型得到其潜在因

子,将这两类潜在特征一起进行参数学习达到模型融合的目的。本文对 HFT 模型进行了改进,主要包括以下三个方面的工作:(1)提出了优化的融合推荐模型: Double Topics with Matrix Factorization (DTMF),该模型同时将用户评论文本和商品评论文本的潜在主题与矩阵分解的潜在因子进行映射,并给出参数求解算法。(2)进一步将同步求解得到的潜在主题作为偏好引导项添加到预测评分中,提出了 DTMF+模型。(3)在两组公开的数据集上测试了模型预测评分的均方误差,DTMF 和 DTMF+均提高了推荐的准确度,在数据子类上将 HFT(item)模型的预测误差最大分别降低了 3.68% 和 7.31%。此外本文讨论和实验分析了如何对商品的评论按照主题相关进行排序,从而增强推荐结果的可解释性。

2 相关工作

2.1 基于评分矩阵的推荐方法

用户的评分数据是用户生成信息 UGC 之一,在基于评分数据的推荐方法中,研究者主要提出了基于邻域和基于模型的两类推荐方法。其中基于用户和基于物品的协同过滤方法是有代表性的基于邻域的推荐方法。基于用户的协同过滤方法通过计算用户与用户之间的相似度,找到与目标用户最为相似的用户来预测评分。基于物品的协同过滤方法则是依据物品间的相似度来预测,较早的应用是 Linden Greg 等人研究的亚马逊商品推荐系统^[17]。基于邻域的推荐方法通常关注如何改进传统的欧氏距离和余弦距离等相似度计算方法来提高推荐质量^[18-21]。

基于模型的推荐方法是通过建立数学模型,用用户评分数据来估计模型中的参数(模型训练阶段),然后用训练好的模型来预测用户对物品的评分(模型使用阶段)。目前最流行的方法之一是 Koren Y 等人^[22-24]研究的潜在因子矩阵分解模型(Latent factor model, LFM),其中基于奇异值分解的 SVD 和 SVD++等方法得到了广泛的应用研究。此外,Salakhutdinov R 等人^[25]提出了基于概率的矩阵分解方法 PMF,在 SVD 矩阵分解上添加了特征的高斯概率分布,以及进一步加入社交信息的其它改进方法^[26]。实验结果表明矩阵分解模型的推荐质量优于传统基于邻域的推荐方法。

目前基于评分矩阵的推荐方法在提高推荐系

统质量方面有着较好的效果。然而评分数据所具有的稀疏性、推荐的冷启动和用户喜好的转移等问题依然存在。因此,近年来研究者通过挖掘用户的关系、用户的评论和商品的标签等信息来进一步提高推荐方法的质量。本文研究如何挖掘评论文本来优化推荐模型,接下来主要介绍近年来融入评论文本的推荐方法。

2.2 融入评论文本的推荐方法

基于文本内容的挖掘方法,在文本的分类、聚类、语义分析等相关自然语言处理任务中有着广泛的研究基础。2003年 David M. Blei 等人^[27]提出了潜在狄利克雷概率模型 (Latent Dirichlet Allocation, LDA) 用于发现文本中的主题分布,成为最具影响力的文本主题挖掘模型。研究者在早期已经意识到评论文本在推荐系统中的价值, Basilico J 等人^[28]将回归分析应用于文本内容分析和协同过滤; Ganu G 等人^[29]发现评论文本中包含不同层面的信息,并探索用户和商品在这些层面上是如何互相权衡从而产生评分; Xu J 等人^[30]通过评论文本主题分析得到用户的偏好计算方法,依据用户的偏好相似度实现协同过滤推荐方法。

近年来,评分与评论结合的方法通过发现评论文本的用户偏好来提高评分预测的质量。Huang J 等人^[7]采用主题模型发现 Yelp 评论文本的子话题后再进行预测评分。Mukherjee S 等人在实验中发现评论文本的情感分析也有助于提高预测精度^[8]。文献^[9]关注将评论文本的潜在主题 (aspect) 用于回归模型实现评分预测。Li X 等人^[10]通过手动建立部分主题词将评论文本主题与评分矩阵分解模型融合。Fan 等人^[31]利用回归模型直接将评论文本和评分数据结合实现打分预测。

近年来表现突出的 HFT 模型^[11]在参数学习和拟合阶段对评分矩阵和评论文本进行融合。它的主要思想是将用户评论集和商品评论集之一的文本主题分布映射到一个由评分矩阵分解得到的潜在因子向量。在此基础上 TopicMF 模型^[12]采用非负矩阵分解挖掘单条评论 (某用户对某商品) 的主题分布,认为该主题分布反应了用户偏好和商品特性,并与用户潜在因子和商品潜在因子建立映射关系。不同于 HFT 和 TopicMF 将文本主题分布映射到矩阵分解的潜在因子向量, R-Linear 和 R-Logistic 方法^[13]将用户偏好与商品特性融合,通过主题分布与评分之间的线性回归和 logistic 回归算法,构建两者的关系。JMARS 方法^[14]先挖掘评论文本中的情

感主题,再建立与评分的关系。Wang C 等人^[32]提出的 CTR 模型将潜在因子矩阵与主题分布结合用于科学文章的推荐。RMR^[33]模型则将主题挖掘应用于基于评分数值的混合高斯推荐模型中。

与本文最为相近的研究是 HFT 模型^[11],该模型只将文本挖掘得到的用户偏好主题和商品特性主题之一融合到矩阵分解模型,分别得到 HFT (user) 和 HFT (item) 两个模型。然而用户评论集和商品评论集可以得出不同的潜在主题分布,仅仅将其中之一融合进推荐模型中,没有充分考虑两者对推荐质量的综合影响。本文提出的 DTMF 模型同时将用户评论集 (某个用户对多个商品的评论集合) 和商品评论集 (某个商品被多个用户评论的集合) 的潜在主题向量与传统矩阵分解的潜在因子向量进行映射,并且进一步提出的 DTMF+模型⁶将用户和商品两个潜在主题分布作为评分预测的引导项。实验结果表明本文提出两个模型均提高了推荐质量。

3 融合评分和评论的推荐模型

本节首先简要介绍融合评分和评论的 HFT 模型^[11],然后详细讨论本文所提出的 DTMF (Double Topics with Matrix Factorization) 模型和 DTMF+模型,以及参数求解算法。本文涉及到的主要符号及其含义如表 1 所示。

表 1 主要符号及其含义

符号	含义
$\tilde{r}_{i,j}$	用户 i 对商品 j 的预测评分
μ	全部评分的平均值
b_i, b_j	用户 i 的偏置, 商品 j 的偏置
$p_i \cdot q_j$	用户 i 的 K 维潜在因子向量, 物品 j 的 K 维潜在因子向量
τ	评分矩阵和评论文本的集合
ϕ_k	主题-词分布
θ_m	文档-主题分布
$z_{m,j}$	文档中每一个词的主题
N_m	m 文档中的词数量
K	潜在主题 (潜在因子) 数量
d	用户评论集 D 中的某个用户的评论
d'	商品评论集 D' 中的某个商品的评论
$z_{d,h}$	用户评论集中词的主题
$z'_{d',h}$	商品评论集中词的主题
$\theta_{d,z_{d,h}}$	用户评论集的文档-主题分布

$\theta_{d',z,d,h}$ 商品评论集的文档-主题分布

3.1 HFT 融合模型

融合评分矩阵和评论文本的 HFT 推荐模型^[11]根据潜在主题发现模型中输入的评论文档集不同,分为融合用户评论集的 HFT(user)模型和融合商品评论集的 HFT(item)模型。例如 HFT(item)模型将某商品的所有评论文本组合构成一篇文档,所有商品的评论集合作为主题发现模型的文档集,从而得到反映商品特性的主题分布。文献[11]假设评分矩阵分解的潜在因子数目 K 和评论文本的潜在主题数目 K 相同,且潜在因子具有相同的权重,构建的潜在因子与潜在主题间的转换关系参见公式(1)。

$$\theta_{j,k} = \frac{\exp(\kappa q_{j,k})}{\sum_k \exp(\kappa q_{j,k})} \quad (1)$$

公式(1)中 $q_{j,k}$ 表示商品 j 在潜在特征 k 上的主题, p_i 表示商品 j 的潜在因子向量在特征 k 上的值, κ 表示控制转换的权重。

HFT(item)模型利用公式(1)的转换函数,将评分的潜在因子与评论的潜在主题相结合,融合模型无需同时拟合求解参数集合 θ 和 q 。将评分矩阵和评论文本的集合用 τ 表示, HFT(item)模型的优化目标函数参见公式(2)。

$$f(\tau|\Theta, \Phi, \kappa, z) = \sum_{r_{i,j} \in \tau} (r_{i,j} - \tilde{r}_{i,j})^2 - \lambda \log L(\tau|\theta, \phi, z) \quad (2)$$

公式(2)中 $\Theta, \Phi = \{\theta, \phi\}$ 分别是基于评分矩阵分解模型的参数集合和基于评论 LDA 主题发现模型的参数集合,详细内容请参看文献[11]。 κ, z 分别表示转换函数的权重控制参数和词的主题参数。公式(2)的右式第一部分是预测评分的误差,第二部分是评论文本主题模型的对数似然函数。 λ 是控制两个部分的权重。公式(2)中 $\tilde{r}_{i,j}$ 为用户 i 对商品 j 的预测评分,可以通过公式(3)得到。

$$\tilde{r}_{ij} = \mu + b_i + b_j + p_i \cdot q_j \quad (3)$$

公式(3)中 μ 表示全局偏置,即全部评分数据的平均值,反映不同数据集对用户评分的影响。 b_i 表示用户偏置, b_j 为商品偏置,两个偏置反映不同用户和不同商品对评分的影响。 p_i 表示用户 i 的 K 维潜在主题向量, q_j 表示物品 j 的 K 维潜在主题向量。

本质上 HFT 模型所建立的目标优化函数公式(2)是用评论文本主题模型的对数似然函数取代了传统矩阵分解模型中由潜在因子组成的正则化项。

3.2 DTMF 融合模型

HFT 模型是将用户评论的潜在主题和商品评论的潜在主题之一与评分矩阵的潜在因子进行融合。然而用户评论集中隐性包含了用户的偏好,商品评论集则代表了商品的特性,本文提出的 DTMF 模型同时将这两者与评分矩阵进行融合。模型的参数逻辑关系图见图 2 所示。

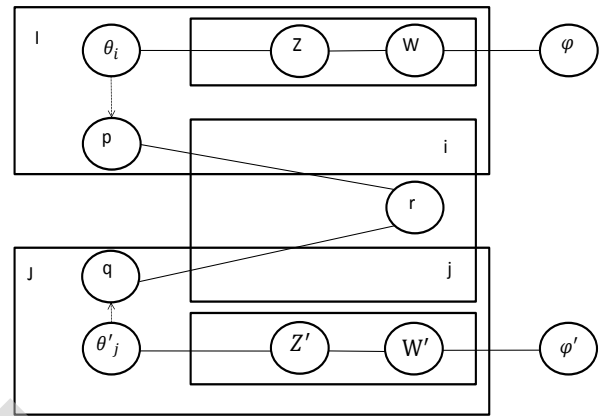


图 2 DTMF 模型的参数逻辑关系图

DTMF模型将从用户 i 的评论文本中挖掘其主题分布 θ_i 并与评分矩阵分解的用户潜在因子向量 p_i 建立映射关系。同时从商品 j 的评论文本中挖掘主题分布 q_j 与评分矩阵分解的商品潜在因子向量 q_j 建立映射关系。并进一步将这两个潜在主题分布概率作为正则化项融入到预测评分模型中得到预测评分 $\tilde{r}_{i,j}$ 。

图 2 表明本文提出 DTMF 模型在用户评论集与商品评论集上同时使用了两次主题发现模型。换言之,DTMF 模型中商品在 K 个主题上的评分要与用户在这 K 个主题上的评分相对应,从而使 K 个主题能够对 K 个潜在因子进行解释。而 HFT 模型只取一个潜在主题分布与 p_i 或 q_j 映射,例如 HFT(item)模型将商品评论集的主题分布 θ'_j 与 q_j 映射。与 HFT 模型的主要区别参见表 2。

下面给出商品评论集和用户评论集的 LDA 主题概率分布。公式(4)中 d 表示用户评论集 $D = \{d_i\}$ 中某个用户的全部评论文本, d' 表示商品评论集 $D' = \{d'_i\}$ 中某个商品的全部评论文本。 $\theta_{d',z,d,h}$ 和 $\theta_{d,z,d,h}$ 表示两个评论集的文档-主题分布。

$$L_{review}(user) = \prod_{d \in D} \prod_{h=1}^{N_d} \theta_{d,z_{d,h}} \phi_{z_{d,h},w_{d,h}} \quad (4)$$

$$L_{review}(item) = \prod_{d' \in D'} \prod_{h=1}^{N_{d'}} \theta'_{d',z'_{d',h}} \phi'_{z'_{d',h},w'_{d',h}} \quad (5)$$

评论文本的潜在主题与矩阵分解的潜在因子的关系借鉴 HFT 模型中的映射函数,得到的用户评论集的潜在因子向量与主题分布的关系参见公式(6),与商品评论集的映射关系参见公式(7)。

$$\theta_k = \frac{\exp(\kappa p_{i,k})}{\sum_{k=1}^K \exp(\kappa p_{i,k})} \quad (6)$$

$$\theta'_k = \frac{\exp(\kappa q_{j,k})}{\sum_{k=1}^K \exp(\kappa q_{j,k})} \quad (7)$$

公式(6)和(7)假设潜在因子数量和潜在主题数量相同,然后构建一个单调函数实现,使 θ_k 与 $p_{i,k}$, θ'_k 和 $q_{j,k}$ 分别成正比映射关系。图1的例子,用户2的评论文本表现出对“硬件配置”主题的偏好,则该主题将对应评分矩阵分解中权重最大的潜在因子。在模型求解过程中,不再分别去学习潜在主题和潜在因子,只需要求解其中一个,再利用映射关系计算其对应的另外一个。为了便于理解, θ 和 θ' 表示潜在主题, $p_{i,k}$ 和 $q_{j,k}$ 表示潜在因子。

DTMF模型将公式(6)和公式(7)融合进LDA^[27]的最大似然函数中,并将它们作为正则化项实现模型的融合,优化的目标函数参见公式(8)所示。

$$O = \sum_{r_{i,j} \in \tau} (r_{i,j} - \tilde{r}_{i,j})^2 - \mu_1 \log L(D|\theta, \phi, z) - \mu_2 \log L(D'| \theta', \phi', z') \quad (8)$$

公式(8)中, μ_1 和 μ_2 用于控制主题似然函数的正则权重参数。 $\log L(D|\theta, \phi, z)$ 和 $\log L(D'| \theta', \phi', z')$ 分别表示用户评论集LDA对数似然函数和商品评论集LDA对数似然函数。将两个评论集的LDA概率分布公式(4)和(5)带入公式(8),得到如公式(9)所示的最小化目标函数。

$$\begin{aligned} \arg \min_{\Theta, \Phi, \kappa, \kappa'} \sum_{r_{i,j} \in \tau} (r_{i,j} - \tilde{r}_{i,j})^2 - \mu_1 \log \prod_{d \in D} \prod_{h=1}^{N_d} \theta_{d,z_{d,h}} \phi_{z_{d,h},w_{d,h}} \\ - \mu_2 \log \prod_{d' \in D'} \prod_{h=1}^{N_{d'}} \theta'_{d',z'_{d',h}} \phi'_{z'_{d',h},w'_{d',h}} \end{aligned} \quad (9)$$

设定模型的参数集合 $\Theta = \{\mu, b_{i,d}, b_j, p_i, q_j\}$ 和 $\Phi = \{\theta, \phi, \theta', \phi'\}$,已经知道 θ 和 θ' 与 p 和 q 存在映射关系,不能独立训练 Θ 和 Φ ^[11],具体求解过程参见3.3节的推导部分和算法1。

3.3 DTMF模型的拟合求解

如前面所述,DTMF算法模型需要拟合求解参数,由于潜在因子和潜在主题参数间已经存在映射关系,因此它们构建的模型无法独立的对参数 Θ 和 Φ 进行求解,可以通过迭代公式(10)和公式(11)来实现公式(9)的求解。

公式(10)固定词的主题 $z_{d,h}$ 和 $z'_{d',h}$ 后,可采用梯度下降法。因此在一次迭代中公式(10)使用梯度下降法对其进行求解参数 Θ, Φ, κ 。主题模型使用Gibbs采样方法,迭代公式(10)和(11)求解参数。

$$\begin{aligned} \arg \min_{\Theta^{(t)}, \Phi^{(t)}, \kappa^{(t)}, \kappa'^{(t)}} \sum_{r_{i,j} \in \tau} (r_{i,j} - \tilde{r}_{i,j})^2 \\ - \mu_1 \log \prod_{d \in D} \prod_{h=1}^{N_d} \theta_{d,z_{d,h}^{(t-1)}} \phi_{z_{d,h}^{(t-1)},w_{d,h}} \\ - \mu_2 \log \prod_{d' \in D'} \prod_{h=1}^{N_{d'}} \theta'_{d',z'_{d',h}^{(t-1)}} \phi'_{z'_{d',h}^{(t-1)},w'_{d',h}} \end{aligned} \quad (10)$$

$$\begin{aligned} sample \rightarrow z_{d,h}^{(t)}, z'_{d',h}^{(t)}, \quad p(z_{d,h} = k) = \theta_{d,k} \phi_{k,w_{d,h}}^{(t)}, \\ p(z'_{d',h} = k) = \theta'_{d',k} \phi_{k,w'_{d',h}}^{(t)} \quad (11) \end{aligned}$$

其中公式(10)的目标函数经过推导可以得到公式(12)所示。

$$\begin{aligned}
& \arg \min_{\Theta^{(t)}, \Phi^{(t)}, \kappa^{(t)}, \kappa'^{(t)}} \sum_{r_{i,j} \in \tau} (r_{i,j} - \tilde{r}_{i,j})^2 - \mu_1 \sum_{d=1}^D \sum_{k=1}^K n_{d,k} (\kappa \cdot p_{i,k} \\
& \quad - \log(\sum_{k'=1}^K \exp(\kappa \cdot p_{i,k'}))) \\
& \quad - \mu_2 \sum_{d'=1}^{D'} \sum_{k=1}^K n_{d',k} (\kappa' \cdot q_{j,k} - \log(\sum_{k'=1}^K \exp(\kappa' \cdot q_{j,k'}))) \\
& \quad - \mu_1 \sum_{d=1}^D \sum_{h=1}^{N_d} \log(\phi_{z_{d,h}, w_{d,h}}) - \mu_2 \sum_{d'=1}^{D'} \sum_{h=1}^{N_{d'}} \log(\phi'_{z_{d',h}, w_{d',h}})
\end{aligned} \tag{12}$$

公式(12)中 $n_{d,k}$ 和 $n_{d',k}$ 表示为文档 d 或文档 d' 中主题 k 出现的次数。使用拟牛顿法还要求解各个参数的偏导数, 得到 p 、 q 、 κ 、 κ' 参数的部分重要偏导数公式, 参见公式(13)、(14)、(15)、(16)所示。

$$\begin{aligned}
\frac{\partial O}{\partial p_i} &= \frac{\partial O}{\partial p_i} - \mu_1 \sum_{j=1}^J \sum_{d=1}^D \sum_{k=1}^K (n_{d,k} \cdot \kappa \\
& \quad - n_d \cdot \kappa \cdot \exp(\kappa \cdot p_{i,k}) / \sum_{k'=1}^K \exp(\kappa \cdot p_{i,k'}))
\end{aligned} \tag{13}$$

$$\begin{aligned}
\frac{\partial O}{\partial q_j} &= \frac{\partial O}{\partial q_j} - \mu_2 \sum_{i=1}^I \sum_{d'=1}^{D'} \sum_{k=1}^K (n_{d',k} \cdot \kappa' \\
& \quad - n_{d'} \cdot \kappa' \cdot \exp(\kappa' \cdot q_{j,k}) / \sum_{k'=1}^K \exp(\kappa' \cdot q_{j,k'}))
\end{aligned} \tag{14}$$

$$\begin{aligned}
\frac{\partial O}{\partial \kappa} &= \frac{\partial O}{\partial \kappa} - \mu_1 \sum_{i=1}^I \sum_{d=1}^D \sum_{k=1}^K p_{i,k} \cdot (n_{d,k} \\
& \quad - n_d \cdot \exp(\kappa \cdot p_{i,k}) / \sum_{k'=1}^K \exp(\kappa \cdot p_{i,k'}))
\end{aligned} \tag{15}$$

$$\begin{aligned}
\frac{\partial O}{\partial \kappa'} &= \frac{\partial O}{\partial \kappa'} - \mu_2 \sum_{j=1}^J \sum_{d'=1}^{D'} \sum_{k=1}^K q_{j,k} \cdot (n_{d',k} \\
& \quad - n_{d'} \cdot \exp(\kappa' \cdot q_{j,k}) / \sum_{k'=1}^K \exp(\kappa' \cdot q_{j,k'}))
\end{aligned} \tag{16}$$

通过对公式(10)和(11)迭代直到输出的参数不再变化或者达到一定的迭代次数, 算法收敛。虽然目标函数公式(9)是非凸的, 产生局部最优解, 但是本文前期的实验结果表明不同的初始条件仍然得到类似的潜在主题分布。

算法 1. DTMF 参数求解算法¹

输入: 用户评分集合 $\{r_{i,j}\}$, 用户评论集 $D = \{d_i\}$,
商品评论集 $D' = \{d'_j\}$

输出: 矩阵分解参数集合 $\Theta = \{\mu, b_i, b_j, p_i, q_j\}$,
主题发现参数集合 $\Phi = \{\theta, \phi, \theta', \phi'\}$

步骤:

1. $\{r_{i,j}\}, \{d_{i,j}\} \leftarrow$ 数据预处理。
2. $\{d_i\}, \{d'_j\} \leftarrow$ 用户/商品评论集。
3. $n_{d,k}, n_{d'} \leftarrow \{n_{d',k}\} n_{d'} \leftarrow D$; 统计文档中的词和主题数目, 用于参数求解的偏导数计算。
4. $f(\tau | \Theta, \Phi, \kappa, z, \kappa', z') \leftarrow \text{lsq}(W[NW]);$ // lsq()是构建的最小目标函数, 见公式(12), $W[]$ 是参数数组, NW 是参数个数。
5. WHILE 迭代次数 < 式(10)和(11)的最大迭代次数 Do:
6. WHILE 迭代次数 < 梯度法迭代次数 Do:
7. Lib-BGFS-Operation(f); 牛顿法求解
8. END WHILE;
9. $z, z' \leftarrow$ sample from Gibbs; 主题采样
10. END WHILE;
11. RETURN $W[NW]$; //返回求解得到的参数数组

3.6 DTMF+模型

如公式(9)所示, DTMF 模型将评论文本的潜在主题作为正则化项融入到矩阵分解模型中, μ_1 和 μ_2 分别表示用户评论集主题正则项和商品评论集主题正则项的权重。本文进一步提出 DTMF+模型, 该模型将从用户评论集和商品评论集中得到的两个主题分布作为偏好引导项也融入到预测评分中, 参数的逻辑关系如图 3 所示。

用户对商品的评分受到两个部分的影响(引导项), 一个是用户自身兴趣偏好, 另一个是大众对商品的主体特性描述, 也就是从商品评论集中提取的商品特性。例如, 一个用户对某一件商品感兴趣的话题是商品外观, 然而在所有的商品评论中对商

$n_{d',k}$

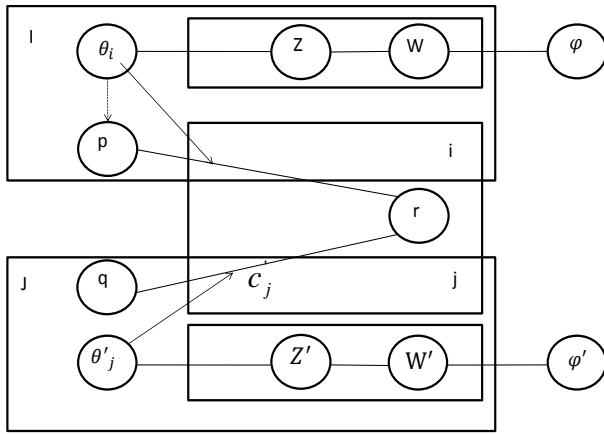


图3 DTMF+模型的参数逻辑关系图

品的质量表示堪忧,这种商品特性的评论会影响用户的偏好,使得用户转而开始关注该商品的质量。DTMF+模型考虑添加这两个引导项,用 y_i' 和 c_j' 分别表示用户个人偏好和商品特性,得到评分预测如

¹DTMF 算法源代码下载地址: <https://github.com/WUT-IDEA/DLMF>
公式(17)所示。

$$\tilde{r}_{i,j} = \mu + b_i + b_j + (\alpha \cdot y_i' + \beta \cdot c_j' + p_i) \cdot q_j \quad \text{公式 (17)}$$

其中 α 和 β 为控制引导项的权重,本文中采用历史评价数量和总评论数的比例作为权重分配机制。当用户历史评论数越多时,用户评论引导项赋予越大的权重。 y_i' 和 c_j' 两个潜在因子与主题分布 θ 和 θ' 的映射,参见公式(6)和(7)。通过上述引导项加入,可以得到目标函数如公式(18)所示。

$$\min_{P,Q} \sum_{\tilde{r}_{i,j} \in \tau} (r_{i,j} - \tilde{r}_{i,j})^2 - \mu_1 \log L(D | \theta, \phi, z) - \mu_2 \log L(D' | \theta', \phi', z') \quad \text{公式(18)}$$

其中 μ_1 和 μ_2 等参数同公式(9),为了拟合求解公式(18),可以通过梯度下降法求解参数,过程和DTMF模型的参数求解算法1一致。

3.7 模型对比

本节讨论 HFT 模型与本文所提出的 DTMF 和 DTMF+模型的主要区别,如表 2 所示。HFT 模型根据主题模型中评论文本的组成不同分为 HFT (user) 和 HFT (item),然后用不同的主题分布

作为正则化项。DTMF 和 DTMF+同时考虑用户的评论文本集和商品的评论文本集,并将两个主题分布作为正则化项。DTMF+则进一步将用户潜在因子和商品潜在因子作为预测评分的引导项。

表2 融合评论和评分的推荐模型对比

模型	评论文本	正则化项	引导项
HFT (user)	某用户的所有评论	用户评论主题分布	无
HFT (item)	某商品的所有评论	商品评论主题分布	无
DTMF	某用户的所有评论 某商品的所有评论	用户评论主题分布 商品评论主题分布	无
DTMF+	某用户的所有评论 某商品的所有评论	用户评论主题分布 商品评论主题分布	用户潜在因子 商品潜在因子

4 数据集与评估方法

4.1 数据来源及格式

本文使用两组数据集均来自 Stanford Network Analysis Project²。一组是大小为 299M 的 Fine food reviews data,另一组是大小为 3.3G 的 Amazon.com 电商网站全部 28 个子类的评论数据。数据集的主要内容是用户对该网站商品的打分、评论及评论的用户有用性反馈,统计结果见表 3。

表3 数据集统计信息

	评论数	用户数	商品数	平均 词数	时间跨度
Fine food	568454	256059	74258	56	1999.10-2012.10
Amazon	34686770	6643669	2441053	82	1995.06-2013.03

从表 3 中可以看出数据的评分与评论的时间跨度都在 10 年以上,两组数据格式一致。每一条样本中包含有如下 10 字段:商品标识、商品名称、商品价格、用户标识、署名、有用性反馈、评分、时间、评论标题文本、评论正文文本。

本文实验目的是验证 DTMF 和 DTMF+融合推荐模型优于目前具有代表性的 HFT 模型。为了全面评估本文提出的模型,首先在 Fine food reviews 数据集上,与传统的协同过滤方法进行了实验对比,然后在 Amazon 数据集上重点对比了传统的全局偏置 Offset、LFM、SVD++、HFT 等四类方法,并且对参数选择、数据稀疏性和推荐结果可解释性进行了实验分析和讨论。

4.2 评估指标

当推荐结果的输出为实数值时，通常用于评估的指标有均方误差（MSE）、均方根误差（RMSE）和平均绝对值误差（MAE）。这三个指标都是计算预测值和真实值之间的误差，其中 MSE 的定义见公式（19）。

$$MSE = \frac{\sum_{u,i \in TestSet} (r_{u,i} - \tilde{r}_{u,i})^2}{|TestSet|} \quad (19)$$

其中， $TestSet$ 表示测试样本集， $|TestSet|$ 表示样本集数目。MSE 值越小，表示系统的推荐质量越好。同时 RMSE 是 MSE 开根号后的值，而 MSE 为 MAE 的平方值，这两个评价指标与 MSE 评价的结论保持一致。为了同参考文献[11]中的模型在同一评估方法下进行对比，本文采用 MSE 作为评价指标。此外若预测评分为 0 和 1 二值，可以采用准确率（precision）和召回率（recall）等，而本文的预测评分为实数值，因此用 MSE 作为评估指标。

²实验数据：<http://snap.stanford.edu/index.html>

5 实验与结果分析

本节将详细介绍实验的过程，对比不同推荐模型和方法在两组公开数据集上的 MSE 误差值，并讨论如何为推荐的商品找到有用性的评论。

5.1 实验设置

(1) 数据预处理

本文数据集包含两种数据，一种是关于用户的评分数据（1-5），另一种是关于用户的评论文本内容。本文对评论文本进行了预处理，包括分词（本文使用 IKAnalyzer）、去停用词，去噪等。

(2) 输入数据划分

本文使用“留出法”将数据集随机划分为训练集和测试集，其中训练集占整个数据集比例的 80%。进行了 20 次的随机划分后，用 20 次测试结果的 MSE 平均值来评估模型的预测质量。

(3) 实验平台配置

本文使用的 Amazon 数据集包含 28 个子类，对超过 1GB 的数据子集进行随机抽取。本文的实验在单机平台下完成，系统为 CentOS6.5，4 核 intel CORE i5 CPU，8G 内存，使用 C/C++ 语言实现。

5.2 Fine food reviews 数据实验结果与分析

DTMF 和 DTMF+ 模型对比传统协同过滤方法和

HFT 模型的结果如表 4 所示。协同过滤方法 (Item based CF) 采用的是欧氏距离，本文对皮尔逊和 Jaccard 相关性度量方法进行了实验，但它们的推荐误差大于欧氏距离。实验结果表明传统的协同过滤方法仅根据评分数据进行推荐，速度快，但是推荐误差远大于融入评论文本的 HFT、DTMF 和 DTMF+，其中 DTMF+ 将协同过滤方法的推荐误差减少了 41.38%。虽然基于评论文本的融合模型训练时间明显增长，但是大幅度提高了推荐质量。接下来将在 Amazon 数据上重点对比不同融合模型。

表 4 Fine food reviews 数据实验结果

	Item-based CF	HFT(item)	DTMF	DTMF+
MSE	0.8491	0.543073	0.519123	0.497725
Imp.	---	36.04%	38.86%	41.38%
Run time	1442ms	5205.67s	8726.84s	9599.52s

5.3 Amazon 数据实验结果与分析

5.3.1 Amazon 评分数据的稀疏性分析

电商平台中用户和商品总体数量巨大，但是每一个用户购买的商品总数相对来说只占很少一部分，本文对 Amazon 3.3G 数据集的统计结果如表 5 所示。

表 5 Amazon 数据集中商品评分统计信息

	<=20	<=10	<=5
占总商品数比例	0.921	0.849	0.753

总体来看，平均 92.1% 的商品评分数量少于 20 个，平均 84.9% 的商品评分数量少于 10 个，平均 75.3% 的商品评分少于 5 个，表明该数据集的评分数据是稀疏的，将影响矩阵分解模型对预测评分的准确度。在以往的研究中使用数据初始化填充、基于内容的推荐等来缓解该问题。本文研究将评论文本的数据融入预测评分，因此对商品评论文本中单词进行了统计，结果如表 6 所示。

表 6 Amazon 数据集中商品评论文本单词个数统计信息

	>=20	>=50
占总商品数比例	0.719	0.472

总体来看，统计后得到 71.99% 的商品评论中单词数大于 20，47.33% 的商品评论中单词数大于 50。结果表明大部分商品具有一定长度的评论文本，通过对用户评论集的主题发现可以缓解单一评分数

据的稀疏性问题。因此可以将评论文本内容融合到评分数据中,根据用户评论所隐藏的用户打分偏好和商品评论所隐藏的商品特性来帮助对未知商品的预测评分。

5.3.2 对比模型

本文提出的优化融合模型与以下传统方法和融合模型进行对比。

(1) Offset 方法是一种基于全局偏置的协同过滤,将商品的平均值作为该商品的预测值,即某用户对某商品的预测打分为该商品所有打分的平均分。本文保留此方法是和文献[11]保持一致,作为对比的基准方法。

(2) LFM(Latent Factor Model), 隐语义模型^[23]。通过矩阵分解(SVD)对未知商品预测评分(公式 2)。此模型中没有应用到用户的评论文本信息。

(3) SVD++模型,本文在 SVD 模型中添加邻域商品的信息得到 SVD++模型,领域商品信息是指用户历史评论商品所具有的潜在因子的累加。

(4) HFT(user)和 HFT(item)模型^[11],融合评分与评论文本的推荐模型(公式 2)。将评论划分为用户评论集或商品评论集,取其一作为主题模型的输入,得到的主题与矩阵分解中的潜在因子之一进行融合。

(5) DTMF 模型。本文提出的融合模型,使用 LDA 方法同时得到用户评论集主题和商品评论集主题,并分别和矩阵分解潜在因子特征融合,构建融合推荐模型(公式 8)。

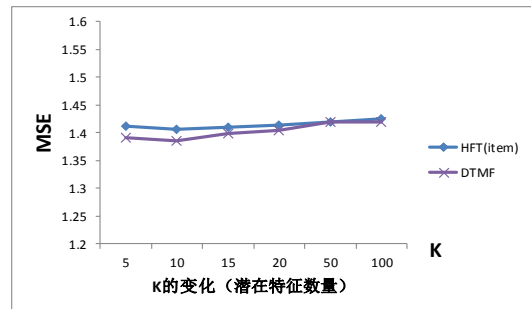
(6) DTMF+模型。在 DTMF 模型上添加了用户主题分布和商品特性主题分布作为两个引导项,进一步优化推荐质量(公式 18)。

5.3.3 参数选取

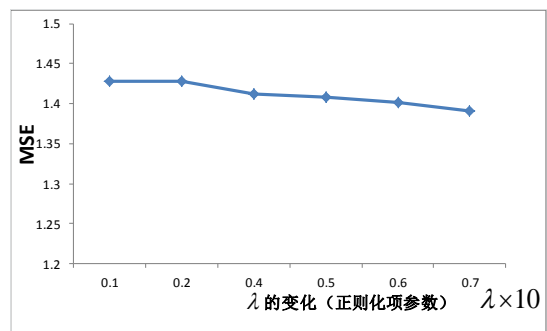
由于对比模型中均有需要调节的参数,首先通过多次实验选取多个特征维数 K ,观察不同特征维数 K 对 MSE 值结果的影响,如图 4 中的(a) 所示(以 Musical_Instruments 数据子集为例)。发现本文提出的 DTMF 算法随着 K 的增加一直优于 HFT,DTMF 和 HFT 均在 $K=5$ 或 10 时性能达到最佳,这与文献[11]的实验结果一致。并且 $K=10$ 时的 MSE 比 $K=5$ 时只是略有降低(DTMF: 1.3918 vs 1.3859, HFT: 1.4117 vs 1.4064)。

HFT 模型中正则化参数 λ (公式 2) 用于控制主题偏好正则化的权重,正则化项是机器学习中避免模型过拟合的有效手段之一。图 4 中(b)的实验结果表明 λ 在 0.5 附近 MSE 值趋于平稳。本文实验中

潜在特征维度 K 设为 5,包括潜在因子数和潜在主题数,迭代次数为 50,正则化参数 λ 为 0.5,DTMF 和 DTMF+中两个正则化权重 μ_1 为 0.5, μ_2 为 0.1。



(a) 潜在特征 K 取值不同对 MSE 的影响



(b) λ 取值不同对 MSE 的影响 ($\lambda \times 10$, 公式 2)

图 4 潜在特征 K 和 λ 取值对 MSE 的影响

5.3.4 实验结果与分析

实验1. 推荐准确度

图 5 直方图给出了 7 个推荐模型各自的 MSE 平均值 (Amazon 数据集的 28 个子类)。分析得出如下结论: (1)潜在因子矩阵分解给推荐质量带来了极大的提高, LFM 将基于全局偏置的 Offset 方法误差减少了 14.6%, SVD++ 进一步将误差减少了 17.1%。(2)基于商品评论集输入的融合模型 HFT(item)推荐质量优于基于用户评论集输入的融合算法 HFT(user),且 HFT 的两个模型都优于基于评分数据的 Offset 方法, LFM 和 SVD++, HFT(item)比 SVD++提升了 7.6%。(3)DTMF 和 DTMF+模型均整体优于 HFT 模型,将 HFT(item)模型的推荐误差分别减少了 0.36%和 1.68%,并且将 HFT(user)模型的推荐误差分别减少了 1.6%和 2.94%。

本文对比方法分为三类,第一类是非矩阵分解方法 (Offset),第二类是矩阵分解模型 LFM 和 SVD++,第三类是融入了评论文本的潜在因子矩阵分解模型 HFT(user)、HFT(item)、DTMF 和 DTMF+。

属于第一类的传统协同过滤方法在 5.2 节中给出了实验结果，见表 4。

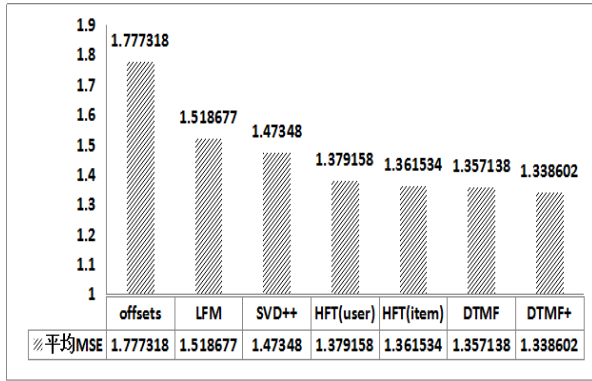


图 5 对比算法 MSE 结果直方图

第一类方法没有使用矩阵分解，和第二类使用矩阵分解的方法对比来看，矩阵分解方法显著提高了推荐质量。理论上矩阵分解中得到高层的潜在因子，可以代表用户历史信息的偏好，缓解了数据稀疏性，对于预测评分起到了积极作用。但是评分数据单一且特征信息不如评论文本丰富，第三类模型考虑融合了评论文本的信息，来弥补评分数据的不足，它们之间的差别在于融合的方法以及融合程度。

HFT 模型以用户评论集或商品评论集为一篇

文档，将得到的主题分布与矩阵分解得到的其中一个潜在因子向量进行映射。而 DTMF 模型则在划分的两个评论集上同时进行主题发现，并作为正则化项融合进基于矩阵分解的预测评分模型中。DTMF+模型进一步将两个评论集上的潜在主题作为预测评分引导项。相比 HFT 模型，DTMF 和 DTMF+模型的优势在于同时考虑了用户偏好和商品特性。

综上所述，DTMF 和 DTMF+模型减少了预测误差 (MSE)，Amazon 28 个子类数据上的 HFT(user)、HFT(item)、DTMF 和 DTMF+四个模型的 MSE 值，参见表 7 所示。其中在每一个数据子集下方给出了 DTMF 和 DTMF+算模型相比于 HFT(user)、HFT(item)提升的百分比。可以看到 DTMF+模型在 27 个子类上优于 HFT(user)模型，同时 23 个子类的结果优于 HFT(item)模型，在子类上将 HFT(item)误差最大减少了 7.31%，而 DTMF 在子类上比 HFT(item)误差最大减少了 3.68%。综合所有数据子集来看 DTMF 和 DTMF+模型的平均推荐质量优于 HFT 模型。

表 7 在 Amazon 28 组数据下的实验结果(K=5)

Method	Video	Arts	Automotive	Baby	Beauty	Books	Phones	Clothing	Electronics	Foods
a: HFT(user)	1.403331	1.403269	1.423605	1.465442	1.392671	1.26835	2.107897	0.375229	1.816131	1.457064
b: HFT(item)	1.237879	1.398137	1.426024	1.449722	1.336476	1.250076	2.112647	0.349138	1.819961	1.460673
c: DTMF	1.281167	1.395728	1.432097	1.456733	1.341773	1.245885	2.118866	0.353376	1.753016	1.459025
d: DTMF+	1.239675	1.388343	1.415873	1.451042	1.310354	1.220125	2.094322	0.343894	1.741132	1.448238
Imp. c vs. a	8.71%	0.54%	-0.60%	0.59%	3.65%	1.77%	-0.52%	5.82%	3.48%	-0.13%
Imp. c vs. b	-3.50%	0.17%	-0.43%	-0.48%	-0.40%	0.34%	-0.29%	-1.21%	3.68%	0.11%
Imp. d vs. a	11.66%	1.07%	0.54%	0.98%	5.91%	3.80%	0.64%	8.35%	4.13%	0.61%
Imp. d vs. b	-0.15%	0.70%	0.71%	-0.09%	1.95%	2.40%	0.87%	1.50%	4.33%	0.85%
Method	Health	Kitchen	Industrial	Jewelry	Kindle	Movies	Music	Musical	Office	Patio
a: HFT(user)	1.554054	1.664797	0.34432	1.197654	1.451514	1.382538	1.031931	1.387048	1.66677	1.708102
b: HFT(item)	1.50557	1.530157	0.344535	1.202037	1.431611	1.376415	1.030399	1.389323	1.676403	1.715282
c: DTMF	1.508202	1.548969	0.373955	1.19741	1.41824	1.36914	1.044047	1.391368	1.642782	1.704495
d: DTMF+	1.499384	1.534892	0.337923	1.186237	1.420342	1.340986	1.029334	1.350072	1.639354	1.681479
Imp. c vs. a	2.95%	6.96%	-8.61%	0.02%	2.29%	0.97%	-1.17%	-0.31%	1.44%	0.21%

Imp. c vs. b	-0.17%	-1.23%	-8.54%	0.38%	0.93%	0.53%	-1.32%	-0.15%	2.01%	0.63%
Imp. d vs. a	3.52%	7.80%	1.86%	0.95%	2.15%	3.01%	0.25%	2.67%	1.64%	1.56%
Imp. d vs. b	0.41%	-0.31%	1.92%	1.31%	1.31%	2.57%	0.10%	0.10%	2.21%	1.97%
Method	Pet	Shoes	Software	Sports	Tools	Toys	Games	Watches	AVG.	
a: HFT(user)	1.555542	0.235528	2.245577	1.177441	1.498481	1.390398	1.515444	1.496303	1.379158	
b: HFT(item)	1.549682	0.218553	2.270249	1.147139	1.499123	1.378072	1.515565	1.502108	1.361534	
c: DTMF	1.553574	0.221303	2.202432	1.151549	1.485249	1.356396	1.48787	1.505208	1.357137	
d: DTMF+	1.548234	0.219845	2.104321	1.149234	1.475873	1.341986	1.470916	1.497437	1.338602	
Imp. c vs. a	0.13%	6.04%	1.92%	2.20%	0.88%	2.45%	1.82%	-0.60%	1.60%	
Imp. c vs. b	-0.25%	-1.26%	2.99%	-0.38%	0.93%	1.57%	1.83%	-0.21%	0.32%	
Imp. d vs. a	0.47%	6.66%	6.29%	1.59%	1.51%	3.48%	2.94%	-0.08%	2.94%	
Imp. d vs. b	1.97%	-0.59%	7.31%	-0.18%	1.55%	2.62%	2.95%	0.31%	1.68%	

实验2. 模型训练时间

本文随机选择了 Arts 和 Watch 子类数据作为模型的输入, 得到了 DTMF 和 DTMF+ 模型与 HFT(item)模型在模型训练时间上的实验结果, 如表 8 所示, 其中 K 表示潜在因子数量。

随着 K 的增加, 训练时间都明显增长, 而前面的参数选择实验中发现 K 达到 5 和 10 时, 推荐准确度变化不大, 所以最终的实验选择 $K=5$ 。此外 DTMF 和 DTMF+ 融合模型在模型训练上需要花费比 HFT(item)更多的时间, 但是从实验 1 的结果可知, 它们的推荐准确度高于 HFT(item)。模型训练属于数据处理的离线过程, 而在线预测时只需要将训练得到的用户潜在因子向量和商品潜在因子向量进行点积运算, 运算时间取决于潜在因子的个数 (即 K 的大小)。因此本文提出的 DTMF 和 DTMF+ 模型可以满足实际应用系统的需求。

表 8 模型训练时间对比 (单位: s)

数据集	Arts	Watch
HFT(item)	63.77	155.34
K=5 DTMF	138.96	322.15
DTMF+	152.86	383.03
HFT(item)	136.57	323.62
K=10 DTMF	276.13	664.17
DTMF+	314.78	796.94
HFT(item)	297.421	720.13
K=20 DTMF	615.47	1140.74
DTMF+	676.96	1287.90

实验3. 推荐结果可解释性

当推荐系统给用户推荐结果时, 用户经常会阅读其他用户的评论。亚马逊网站上根据用户对评论的人工反馈进行排序, 例如 7/10, 表示有 10 名用户反馈, 7 名用户认为评论有用, 3 名用户认为没有用。然而该简单的排序方法存在两个问题^[34]: 第一, 新商品的评论文本可能没有或是只有极少的用户进行了有用性反馈, 因此其打分是 0/0, 从而只能排在评论列表的后面。第二, 低销量的商品, 用户的评论偏少。若能在低销量商品中通过对评论文本内容的分析, 提高用户对冷门商品的认识, 引导用户的购买决策, 符合“长尾理论”。

针对上述问题, HFT 模型用评论主题分布与商品主题分布的相似程度进行评论排序, 为推荐的商品找到有用性的评论。但是在参考文献[11]中仅给出了方案, 并未用实验予以验证, 本文首先通过实验分析基于评论文本内容的排序与传统排序的差异, 并讨论了 DTMF 模型对此部分的改进。

为了能够观察 HFT 模型得到的有用性评论与原始 Amazon 评论排序的差异, 本文采用 Kendall tau distance 计算差异度, 如公式(20)。

$$KT_j = 1 - \frac{Num_{oo}}{N(N-1)/2} \quad (20)$$

公式(20)中 KT_j 表示数据子集中商品 j 的所有评论排序与用户反馈 Helpfulness 排序的 Kendall tau

distance 差异度。距离越小，两者的相似程度则越大，两者最不相似时，取为 1。 Num_{oo} 表示逆序对数的总数。排序对数量总和表示为 $N(N-1)/2$ 。

本文随机选取了 15 个数据子集进行计算差异度，表 9 的结果表明利用 HFT(item)模型产生的排序与用户反馈排序的差异度比较大，在实际系统中可以考虑将这两种排序机制综合起来。HFT(item)模型产生的排序将与某商品主题相近的评论排在前面，考虑的是评论文本自身的内容。

本文认为有用性评论是综合用户个人偏好和商品特性，因此在对每一条评论进行打分排序时应同时考虑这两个方面的内容。因此，本文给出了融合两种特性的评论排序方法，参见公式(21)。

$$\tilde{r}_{i,j,n} = \frac{sim(\theta_{j,n}, \theta_i) + sim(\theta_{j,n}, \theta_j)}{\sum_{n'=1}^{N_j} (sim(\theta_{j,n'}, \theta_i) + sim(\theta_{j,n'}, \theta_j))} \quad (21)$$

公式(21)中 $\tilde{r}_{i,j,n}$ 表示用户 i 对商品 j 第 n 条评论的评分， N_j 表示商品 j 的所有评论数量， sim 计算方法参考文献[11]。DTMF 模型中能够得到用户主题分布和商品主题分布，因此可以通过公式(21)

对商品的评论进行排序，推送给用户有用性评论，从而提升推荐结果的可解释性。

表 9 HFT(item)模型评论排序与 Amazon 排序差异度

数据子集	KT 差异度
Arts	0.7024
Automotive	0.7772
Baby	0.7570
Beauty	0.7315
Clothing_&_Accessories	0.7219
Gourmet_Foods	0.7171
Health	0.7085
Industrial_&_Scientific	0.7068
Jewelry	0.7119
Musical_Instruments	0.7110
Office_Products	0.7085
Pet_Supplies	0.7586
Shoes	0.7222
Software	0.7112
Watches	0.7127

表 10 有用性评论排序结果示例

排序	商品 ID	Helpfulness	相似度距离	评论文本
1	B000MVZ2E6	3/3	0.063111211	exquisite ink inky black supposed flows easily ...
2	B000MVZ2E6	1/2	0.080003211	bought replace omas black ink running lamy fountain ...
3	B000MVZ2E6	3/4	0.083149413	lamy safari starts paper ve including receipt paper ...
4	B000MVZ2E6	0/0	0.094894474	love ink reliable notes writing sketching watercolors
5	B000MVZ2E6	70/72	0.133192774	time buy fountain pen ink lamy safari converter ...
6	B000MVZ2E6	2/3	0.139970948	huge fan noodler ink company believes providing ...
7	B000MVZ2E6	6/6	0.140863446	noodler black bulletproof ink waterproof fraud ...
...
127	B000MVZ2E6	0/0	0.866689016	nice ink minimal bleeding crappy paper leaves...
128	B000MVZ2E6	7/17	0.880409567	ink pilot plumix converter brow pen flows dream ...
129	B000MVZ2E6	0/7	0.920992696	noodler bullet proof black saturated black ink ...
130	B000MVZ2E6	0/0	0.961184004	ink impressed hold glass bottle quality feels sturdy ...
131	B000MVZ2E6	0/3	0.977550082	haven ink love noodler ink waterproof black ...

然而由于两组公开的数据集中没有明确的给出单个用户对一件商品的所有评论的有效性打分，因此无法进行定量的实验验证。本文以商品 IDB000MVZ2E6 为例，在表 10 中给出了 131 条评论排序结果。

表 10 中的排序考虑了评论的主题分布与商品的主体分布的相似度，排在前面的评论多数是属于热门评论，同时有些评论虽然 Helpfulness 偏低，仍然排在前面，比如排序为 4 的评论。因此，本文的有用性评论排序方法可作为推荐系统的一种评论

推荐策略, 优化仅根据用户反馈次数的简单排序。

5.3.5 相关讨论

(1) 本文实验的两组数据的时间跨度都在 10 年以上, Fine food reviews 数据集包含了 1999.10-2012.10 的信息, Amazon 数据是从 1994.6 至 2013.3 约为 18 年。考虑到 Web 电商平台快速发展可能对推荐结果带来影响, 本文以 Amazon 的 video 子类数据为例, 以一年为时间段对推荐质量进行了实验分析, 如图 6 所示。结果表明, 本文提出的 DTMF 模型和 HFT 模型的推荐误差 (MSE) 没有随时间变化呈现一定上升或者下降规律, 近十年的 MSE 值基本上在 0.4-0.5 之间。1997 年到 2001 年由于每年的数据量不超过 10M, 且 1999 年仅 6.95K, MSE 变化幅度较大。

(2) 对于评论文本的主题发现, 本文将一个商品 (或者一个用户) 的所有评论合成一个文档作为

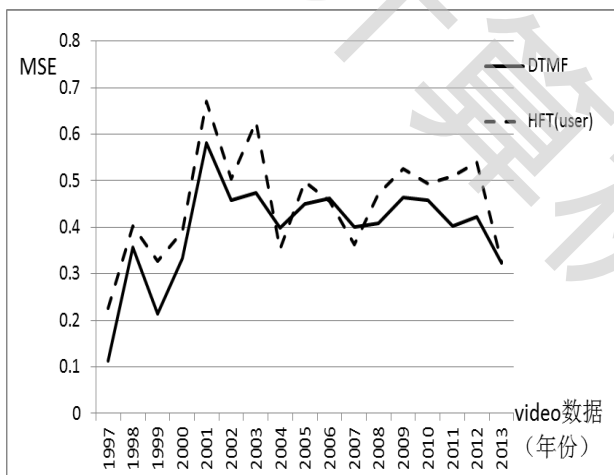


图 6 MSE 随时间变化曲线 (年份)

LDA 主题模型的输入, TopicMF 方法^[12]则采用单条评论文本为主题模型的输入。两种方法处理文本数据的粒度不一样。本文以 Amazon 数据为对象, 在 $K=5$ 时, TopicMF 的 MSE 为 1.3645, 推荐误差大于本文提出的 DTMF 和 DTMF+ (图 1 中分别为 1.3571 和 1.3386)。因此在本文的实验着重对比了以商品 (用户) 评论集为 LDA 输入的 HFT 模型。

(3) 本文以每一条评论文文档进行 LDA 主题发现, 将主题偏好作为用户潜在因子的引导项, 并将主题概率分布添加到正则约束中构建了基于单条评论文本的融合推荐模型。实验结果表明基于单条评论文本的推荐模型的 MSE 为 1.3520, 推荐质量优于 TopicMF、HFT(user) 和 HFT(item)。虽然推

荐误差比 DTMF 略低一点 (1.3571), 但是不如 DTMF+ (1.3386)。

6 总结

本文提出了融合评分矩阵和评论文本的优化推荐模型 DTMF 和 DTMF+。DTMF 模型将主题发现分为用户评论集的主题发现和商品评论集的主题发现, 并将两者同时与矩阵分解的潜在因子进行融合, DTMF+ 则再进一步将用户主题偏好和商品主题特性作为引导项加入到预测模型中。对比实验结果表明 DTMF 和 DTMF+ 模型在整体优于其它模型, 将 HFT(item) 的推荐误差分别减少了 0.36% 和 1.68%, 将 HFT(user) 的推荐误差分别减少了 1.6% 和 2.94%。此外针对推荐结果的可解释性问题, 对基于 HFT 模型的有效性评论排序问题进行实验验证, 并讨论基于 DTMF 模型的排序优化方法。

本文关注对评论文本的主题挖掘, 今后的研究可以考虑用户偏好随时间变化、评论文本情感、潜在特征的权重、社交关系等对推荐结果的影响。此外可以将 DTMF 和 DTMF+ 模型应用到群组推荐、好友关系推荐等问题中。

参考文献

- [1] Lei G, Jun M A, Chen Z M. Incorporating item relations for social recommendation. Chinese Journal of Computers, 2014, 37:219-228.(in Chinese)
(郭磊, 马军, 陈竹敏, 等. 一种结合推荐对象间关联关系的社会化推荐算法. 计算机学报, 2014, 37(1):219-228.)
- [2] HE Jie-Yue MA Bei. Based on real-valued conditional restricted boltzmann machine and social network for collaborative filtering. Chinese Journal of Computers, 2016(1):183-195.(in Chinese)
(何洁月, 马贝. 利用社交关系的实值条件受限玻尔兹曼机协同过滤推荐算法. 计算机学报, 2016(1):183-195.)
- [3] Wang Z S, Li Q, Wang J, Yin J. Real-Time personalized recommendation based on implicit user feedback data stream. Chinese Journal of Computers, 2016(1):52-64.(in Chinese)
(王智圣, 李琪, 汪静, 印鉴. 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(1):52-64.)
- [4] Kevin Lerman, Sasha Blair-Goldensohn, Ryan T. McDonald. Sentiment summarization: evaluating and learning user preferences// Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Athens, Greece, 2009: 514-522.
- [5] Amit Sharma, Dan Cosley. Do social explanations work?: studying and modeling the effects of social explanations in recommender systems//

- Proceedings of the 22nd International World Wide Web Conference (WWW), Lyon, France, 2013: 1133-1144.
- [6] 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.
- [7] Huang J, Rogers S and Joo E. Improving restaurants by extracting subtopics from yelp reviews// *Proceedings of the iConference*, Berlin, Germany, 2014: 1-5.
- [8] Mukherjee S, Basu G, and Joshi S. Incorporating author preference in sentiment rating prediction of reviews// *Proceedings of the 22nd International World Wide Web Conference(WWW)*, Rio de Janeiro, Brazil, 2013: 47-48.
- [9] Wang H, Lu Y and Zhai C. Latent aspect rating analysis on review text data: A rating regression approach// *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, USA, 2010: 783-792.
- [10] Li X, Xu G, Chen E, Li L. Learning User Preferences across Multiple Aspects for Merchant Recommendation// *Proceedings of 2015 IEEE International Conference on Data Mining(ICDM)*, Atlantic City, USA, 2015: 865-870.
- [11] McAuley Julian, Leskovec Jure. Hidden factors and hidden topics: understanding rating dimensions with review text// *Proceedings of the 7th ACM conference on Recommender systems(RecSys)*. Hong Kong, China, 2013: 165-172.
- [12] Bao Y, Fang H, Zhang J. Topicmf: Simultaneously exploiting ratings and reviews for recommendation// *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. Quebec City, Canada, 2014: 2-8.
- [13] Gao YF, Yu WZ, Chao PF, Zheng ZL, Zhang R. Analyzing reviews for rating prediction and item recommendation. *Journal of East China Normal University(Natural Science)*, 2015(3): 80-90. (in Chinese)
(高祎璠, 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.)
- [14] Diao Q, Qiu M, Wu C Y, et al. Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS)// *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, USA, 2014: 193-202.
- [15] Cosley D, Lawrence S, Pennock D M. Referee: An open framework for practical testing of recommender systems using research index// *Proceedings of the 28th International Conference on Very Large Data Bases(VLDB)*, Hong Kong, China, 2002: 35-46.
- [16] Sinha R, Swearingen K. The role of transparency in recommender systems// *Proceedings of the Extended abstracts of the 2002 Conference on Human Factors in Computing Systems*. Minneapolis, Minnesota, 2002: 830-831.
- [17] Linden G, Smith B, York J. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 2003, 7(1): 76-80.
- [18] Choi K, Suh Y. A new similarity function for selecting neighbors for each target item in collaborative filtering. *Knowledge-Based Systems*, 2013, 37(1): 146-153.
- [19] Wu X, Cheng B, Chen J. Collaborative filtering service recommendation based on a novel similarity computation method. *IEEE Transactions on Services Computing*, 2017, 10 (3) :352-365.
- [20] Wen J H, Shu S. Improved collaborative filtering recommendation algorithm of similarity measure. *Computer Science*, 2014, 41(5): 68-71.
- [21] Huang X, Qin Z, Chen H. A new user similarity measurement based on a local item space in collaborative filtering recommendation. *Journal of Computational Information Systems*, 2015, 11(10): 3501-3508.
- [22] Hu Y, Koren Y, Volinsky C. Collaborative filtering for implicit feedback datasets// *Proceedings of the 8th IEEE International Conference on Data Mining(ICDM)*. Pisa, Italy, 2008: 263-272.
- [23] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *IEEE Computer*, 2009, 42(8): 30 - 37.
- [24] 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.
- [25] Salakhutdinov R, Mnih A. Probabilistic matrix factorization// *Proceedings of the Advances in Neural Information Processing Systems(NIPS)*. Vancouver, Canada 2007: 1257-1264.
- [26] Xiao R, Li Y, Chen H, et al. SRSP-PMF: A novel probabilistic matrix factorization recommendation algorithm using social reliable similarity propagation// *Proceedings of the International conference on Intelligent Computing Theories and Methodologies*. Fuzhou, China, 2015: 80-91.
- [27] Jordan M I, Blei D M, Ng A Y. Latent dirichlet allocation. *Journal of Machine Learning Research*, 2003(3): 993-1022.
- [28] Basilico J, Hofmann T. Unifying collaborative and content-based filtering// *Proceedings of the 21st International Conference on Machine learning(ICML)*. Banff, Canada, 2004: 65-72.
- [29] Ganu G, Elhadad N, Marian A. Beyond the stars: improving rating predictions using review text content// *Proceedings of the 12th International Workshop on the Web and Databases*. Providence, USA, 2009: 1-6.
- [30] Xu J, Zheng X, Ding W. Personalized recommendation based on reviews and ratings alleviating the sparsity problem of collaborative filtering// *Proceedings of the Ninth IEEE International Conference on e-Business Engineering*. Hangzhou, China, 2012: 9-16.
- [31] Fan M and Khademi M. Predicting a business star in yelp from its reviews text alone. *CoRR*, vol. abs/1401.0864, 2014.
- [32] Wang C, Blei D M. Collaborative topic modeling for recommending scientific articles// *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge discovery and data mining*. San Diego, USA, 2011: 448-456.
- [33] Guang L, Lyu Michael R., Irwin King. Ratings meet reviews, a combined approach to recommend// *Proceedings of the 8th ACM Conference on Recommender systems (RecSys)*. Foster City, USA, 2014: 105-112.

- [34] Kim S M, Pantel P, Chklovski T. Automatically assessing review helpfulness//Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics(EMNLP). Sydney, Australia, 2006: 423-430.

计算机学报



LI Lin, born in 1977, Ph.D., professor, Ph.D. supervisor. Her research interests include recommender system and information retrieval.

LIU Jin-Hang, born in 1986, Ph.D. candidate. His research interests include machine learning and natural language processing.

MENG Xiang-Fu, born in 1981, Ph.D., associate professor. His research interests include Web data querying and spatial database analysis.

SU Chang, born in 1992, master. His research interests include recommender system and machine learning.

LI Xin, born in 1989, Ph.D., research director. His research interests include recommender system and big education data.

ZHONG Luo, born in 1957, Ph.D., professor, Ph.D. Supervisor. His research interests include intelligent technology and intelligent system..

计算机学报

Background

Recommendation system exploits user-generated information (User generated content, i.e., UGC) has been widely studied in recent years. Traditional recommendation models make use of ratings, a kind of explicit UGC which pushes the successful application of recommendation system in our real life. A suite of popular recommendation approaches have been proposed through exploiting rating matrix, such as collaborative filtering, clustering, association rules, hidden factor model, and so forth. With the involvement of more and more users in recommender systems, the reviews written by them have become one of the most influential UGC. Through text analysis and opinion mining methods, we can not only know user preferences, but also guess or explain the reasons that they have bought. Therefore, recommendation models by exploiting rating matrix and review texts have attracted researchers' interests currently. However, how to model both of them simultaneously still needs research efforts for improving recommendation accuracy and interpretability.

This paper address two issues in building improved recommendation models based on ratings and reviews: (1)

combining rating matrix and review texts in latent factor analysis in different ways, and (2) finding 'informative' reviews, which text is good at explaining the factors that are contributed to a user's rating. By dividing review collections into user review collection and product reviews, the proposed models merge them into matrix factorization to improve recommendation accuracy. Moreover, based on understanding how hidden rating factors is related to hidden review factors, our models can sort all reviews belongs to an item based on user preferences. Some preliminary experimental results are reported, which improves recommendation interpretability.

This work is mainly supported by the National Social Science Foundation of China under Grant No. 15BGL048 and the National Natural Science Foundation of China under Grant No. 61772249. The two projects mainly aim to study the methods of massive and multimodal big data analysis and apply them in different applications. This paper discusses the problem of aggregating numerical rating data and textual review data in high-level dimensions, which is a effective study in improving the quality of recommender systems.