

• 复杂性科学 •

基于四阶奇异值分解的推荐算法研究

郭强¹, 岳强¹, 李仁德¹, 刘建国^{2*}

(1. 上海理工大学复杂系统科学研究中心 上海 杨浦区 200093; 2. 上海财经大学会计与财务研究院 上海 杨浦区 200433)

【摘要】三阶奇异值分解推荐算法可以综合考虑用户、物品标签和物品三部分信息,挖掘三者之间的潜在关系进行推荐,然而该方法并没有引入其他方面的有效信息,如用户情感。为了考虑更多维度的信息,本文在三阶奇异值分解推荐算法的基础上,提出了一种加入用户情感信息的四阶奇异值分解推荐算法。该方法基于从评论中的emoji表情提炼出的用户情感偏好,再引入四阶张量模型,存储用户、用户情感、物品标签和物品四元组数据,应用四阶奇异值分解,从而进行个性化推荐。在某在线互联网教育的实证数据集上的实验结果表明,该方法比三阶奇异值分解推荐算法以及传统推荐算法在准确率和召回率性能指标上都有明显提升,其中进行Top-1推荐时,准确率和召回率可以达到0.513和0.339。本文的工作为移动通信端的个性化推荐提供了借鉴。

关键词 四阶; 多维信息; 推荐算法; 奇异值分解

中图分类号 TP301.6 文献标志码 A doi:10.3969/j.issn.1001-0548.2019.04.017

Recommendation Algorithm Based on 4th-order Singular Value Decomposition

GUO Qiang¹, YUE Qiang¹, LI Ren-de¹, and LIU Jian-guo^{2*}

(1. Complex Systems Science Research Center, University of Shanghai for Science and Technology Yangpu Shanghai 200093;

2. Institute of Accounting and Finance, Shanghai University of Finance and Economics Yangpu Shanghai 200433)

Abstract The 3rd-order singular value decomposition recommendation algorithm can comprehensively consider the three parts of information of user, tag and item, and explore the potential relationship between the three to make recommendations. However, this method does not introduce any other effective information, such as the user's emotion. Considering more dimension information, in this paper we propose a 4th-order singular value decomposition recommendation algorithm based on the third-order one. The method extracts user's emotional preference from the emoji expression in the commentary, introduces a 4th-order tensor model to store user, user emotion, tag, and item quad data, and applies the 4th-order singular value decomposition to make personalized recommendations. The experimental results on an empirical dataset of online internet education shows that the proposed method has a significant improvement in accuracy and recall performance than the third-order singular value decomposition recommendation algorithm and the traditional recommendation algorithms. In the Top-1 recommendation, the accuracy rate and recall rate of proposed method can reach 0.513 and 0.339. The work of this paper provides a reference for personalized recommendation of mobile.

Key words 4-th order; multidimensional information; recommendation algorithm; singular value decomposition

推荐算法是电子商务中最重要的技术之一^[1]。其中,协同过滤(collaboration filtering)推荐是目前应用最广泛的个性化推荐算法。它的核心思想是,利用目标用户的相似用户对特定物品的评价来产生该用户对此物品的评价预测^[2-3]。协同过滤算法最大的优点是对推荐对象没有特殊的要求,能处理电影、

音乐等难以进行文本结构化表示的对象^[4]。但协同过滤算法存在评分数据稀疏性等问题并且只考虑了用户的评分数据却忽略了物品和用户本身的诸多特征,如:电影类型、发布时间等,用户性别、年龄等^[5]。如何充分、合理地利用这些特征,获得更好的推荐效果,是基于内容(content-based)的推荐算法

收稿日期: 2018-04-25; 修回日期: 2019-02-12

基金项目: 国家自然科学基金(71771152)

作者简介: 郭强(1975-),女,教授,博导,主要从事复杂网络、数据挖掘、科学知识图谱分析方面的研究。

通信作者: 刘建国,教授, E-mail: liu.jianguo@sufe.edu.cn

所解决的主要问题。基于内容的推荐算法通过分析用户已经评过分的物品的特征来得到对用户兴趣的描述, 通过比较用户兴趣描述与物品内容信息之间的相似性来实现推荐功能^[6]。但基于内容的推荐算法不可避免地收到信息获取技术的约束, 如自动提取图形、视频等内容特征具有技术上的困难^[4]。文献[7]将SVD(singular value decomposition)引入协同过滤算法中, 利用SVD方法将用户对物品的评分分解为用户、物品特征向量矩阵, 并利用用户和物品之间的潜在关系, 利用评分矩阵的奇异值提取一些本质特征, 从而用这些特征做出推荐。基于SVD的协同过滤推荐算法不仅可以不受信息挖掘技术的制约, 而且可以解决协同过滤推荐算法中的评分稀疏性。随着Web2.0时代的来临, 社会标签系统开始兴起^[8]。社会标签系统主要包含用户、标签和物品三元素, 而上述全部推荐算法仅考虑了用户-物品二维信息, 因而传统推荐算法不能直接应用于社会标签系统。为解决这个问题, 文献[9]首先提出将能完整地表示高维数据并且能维持高维空间数据地本征结构信息的张量(tensor)应用于社会标签系统中, 并且利用HOSVD(high order singular value decomposition)分解算法进行标签预测。

HOSVD推荐算法具有良好的性能和效果, 并且在实际应用中也有着不错的表现^[10], 然而经典的HOSVD方法仅考虑了用户-标签-物品三部分信息, 并没有考虑更高维度的信息。与此同时, 随着网络的普及与发展, 网络信息量日益增长, “信息过载”与“资源迷航”现象愈发严重, 其中信息过载是指用户对过多的信息难以及时地消化、吸收; 资源迷航是指用户难以确切表达对信息资源的需求, 以及难以准确有效地寻找资源^[11]。文献[12]认为在传统推荐算法中加入多维信息是一种解决信息过载与资源迷航现象的有效手段。文献[13]首次提出在信息推荐中考虑情景信息, 建立了一个多维推荐模型。所谓情景信息即对于同一个用户而言, 即使对同一个物品, 当用户处于不同的情景之下对该物品的爱好(兴趣)可能都是不一样的。文献[14]利用用户的上下文信息, 提出了一种基于情景的多维协同过滤方法, 针对多维信息对信息推荐效果的重要性以及如何在信息推荐中纳入多维信息进行了分析。文献[15]使用多维兴趣向量刻画用户的兴趣, 采用多智能体模型模拟, 并引入用户和新闻的质量, 分析了用户网络的结构特征以及质量因素对新闻推荐和传播的影响。文献[16]指出传统推荐算法未能充分利用视频社

会化网站中的多维信息, 导致冷启动和数据稀疏的问题; 并提出一种社交网络服务中的多维空间视频推荐算法, 综合分析构成视频社会化网络的多维信息源要素。在经典的用户物品评分矩阵中, 考虑到在线用户的评分行为受到各种类似情绪、时间、流派等潜在因素的影响, 文献[17]提出了一种多线性融合矩阵分解方法。以电影评分为例, 分别单独刻画用户和其他场景信息之间的相互作用, 然后通过线性融合构成最终的用户评分。由于标签包含关于个性化偏好和物品内容的丰富信息, 文献[18]提出了一种基于用户-物品-标签三部分图集成扩散的推荐算法, 使用Del.icio.us、MovieLens和BibSonomy三个基准数据集对该算法进行了评价。实验结果表明, 标签信息的使用可以显著提高推荐的准确性、多样性和新颖性。文献[19]在引入标签信息的过程中, 认为有些相似商品之间的标签其实是可以互补的, 利用这种协同关系, 可以解决商品的信息不对称问题。为此, 提出了基于商品协同网络的张量分解模型, 采用了Pairwise排序学习的方式对模型进行优化, 并在lastfm、Yelp数据集上验证了该方法在处理稀疏物品排序预测挑战时的准确性和灵活性。社交网络随时间不断演化, 用户兴趣也是不断变化的^[20]。并发现在用户兴趣点推荐中, 时间扮演着重要的角色, 文献[21]提出了时间感知兴趣点推荐, 建立了结合时态信息的协同推荐模型, 以达到在一天中的指定时间为给定用户推荐相应兴趣点。考虑到用户购买兴趣的变化, 文献[22]提出了一种时间权重方法, 该方法使用聚类来区分不同类型的物品, 对于每个物品集群, 跟踪每个用户的购买兴趣变化, 并根据用户自己的购买行为引入时间衰减因子。除了推荐领域, 考虑多维信息在其他领域应用的也十分广泛。文献[23]综合考虑影响信任关系的多种可能要素, 提出了一个新的基于多维决策属性的信任关系量化模型, 引入直接信任、风险函数、反馈信任、激励函数和实体活跃度等多个决策属性, 从多个角度推理和评估信任关系的复杂性和不确定性, 用来解决传统量化模型对环境的动态变化适应能力不足的问题。文献[24]指出贫困的概念经历了由一元向多维度的扩展, 传统单维度测量方法的局限性使得其无法解释贫困的现象, 而多维方法有助于理解贫困的本质特性, 包括复杂性、模糊性和渗透性等, 并提出了多维贫困的识别、加总和分解方法。鉴于诱发交通事故的原因复杂, 不仅和驾驶员的行为有关, 而且和道路环境条件、天气条件、交通状况等有一定的联

系,文献[25]提出采用多维分析技术分析交通事故数据,建立相应的信息组织模型,深层次地挖掘诱发事故的因素之间的内在关系。随着网络信息资源的数量呈爆炸式增长,将传统的二维或三维推荐算法扩展为多维推荐算法显得尤为重要。为此,本文尝试在三阶奇异值分解推荐算法的基础上,提出了一种加入用户情感信息的四阶奇异值分解推荐算法。通过提取评论中的emoji表情,依据情感分成积极、中立和消极三类,分别给每类情感赋予不同的权重,之后计算不同类emoji表情数量的加权和来表征用户的情感,再利用四阶张量模型,存储用户、用户情感、物品标签和物品四元组数据,应用四阶奇异值分解,进而进行个性化推荐。最后在某在线互联网教育的实证数据集上的实验结果表明,在TOP1~TOP5推荐中,基于四阶奇异值分解推荐算法在准确率、召回率等指标上相较于三阶或传统算法,表现优异。

1 数据清洗与提取

本文使用某在线互联网教育中用户评论的数据集来评估所提出算法的性能。该数据集包含2017年1月~2017年3月间共1 324 501条用户评论记录,其中用户评论是指用户在选择过的某一社团刊物(简称社刊)上发表评论,一个用户可以对某一社刊进行多次评论。本文定义上述实证数据为原始数据。为了最终实验的进行,需要在原始数据上进行数据清洗、用户情感计算以及社刊标签提取3个步骤。

1.1 数据清洗

首先删除不含emoji表情的评论记录。其次,为了获得更稠密的数据,本文构造了如图1所示的投影网络^[26],即把某一用户评论过某一社刊视作一个新单元节点,若与其他单元节点存在相同的用户或者社刊则产生连边。

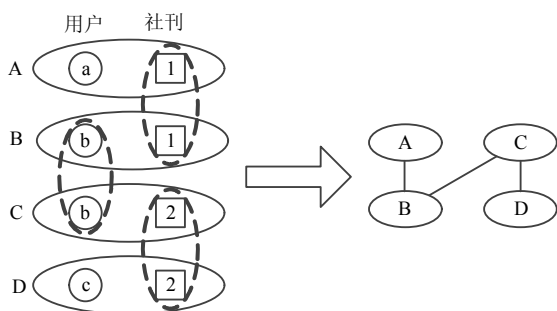


图1 投影网络构建示意图

如图1所示,用户a在社刊1上评论过,将这条记录看作一个单元节点,用A表示,即A中既包含用户

a也包含社刊1的信息。相应的,将用户b和社刊1看作一个单元节点B。因为单元节点A和B中包含同样的社刊1,则关系A与关系B之间存在一条连边。以此类推,单元节点之间依靠是否拥有同样的用户或者社刊信息而建立连边,这样便构成了本文使用的投影网络。最后,在投影网络上应用k核分解^[27]。实验中k取最大值3 442,便可得到最核心的单元节点。

从上述k值最大所对应的单元节点中,保留用户和社刊信息,并定义这部分用户与社刊为核心用户与社刊。再筛选出包含这部分核心用户与社刊的评论数据即筛选后的评论数据是由核心用户对核心社刊所发表的包含emoji表情的评论数据,本文定义这部分筛选后的数据为实验数据。至此,数据清洗工作全部结束。

1.2 用户情感计算

根据用户在相应社刊评论中的emoji表情数据提炼出用户情感。详细用户情感计算过程如下。

经过数据清洗,得到15 370条记录的数据。在这些记录中一共提取出126种emoji表情,并由7名志愿者主观的对这126种emoji表情分成3类,分别是:积极、中立和消极。为了检验志愿者主观分类的结果的一致性,避免出现某个志愿者胡乱分类的情况,本文对分类结果进行了Krippendorff's alpha检验^[28]。文中实验的Alpha值为0.706 9,大于基准值0.6^[28],则表明不同志愿者的分类结果之间的差异是可以被接受的。之后,对7名志愿者的分类结果取众数即获得相应emoji表情的所属分类。部分emoji表情分类结果如表1所示。

表1 emoji表情分类部分结果

| 积极表情 | 中立 | 消极表情 |
|------|-----|------|
| 送花 | 天然呆 | 晕 |
| 得意 | 扭扭 | 尴尬 |
| ⋮ | ⋮ | ⋮ |
| 拍地大笑 | 滚床 | 委屈 |
| 超人 | 饿 | 汗 |

表情分类结束后,本文根据式(1)来计算某条记录表达对社刊的情感:

$$e_{um} = \frac{\sum_{i=1}^{n_2} \frac{p_a - n_a}{n_1}}{n_2} \quad (1)$$

式中, e_{um} 表示在记录x上用户u对社刊m的情感; p_a 表示积极表情的数量; n_a 表示消极表情的数量; n_1 表示一条记录上emoji表情的总个数; n_2 表示用户u对社刊m的所有评论记录数。由式(1)可知,用户u对社

2.3 SVD分解

在 n -模展开矩阵 A_1 、 A_2 、 A_3 和 A_4 上分别进行SVD分解:

$$\begin{aligned} A_1 &= Y^{(1)} S_1 V_1^T \\ A_2 &= U^{(2)} S_2 V_2^T \\ A_3 &= U^{(3)} S_3 V_3^T \\ A_4 &= U^{(4)} S_4 V_4^T \end{aligned} \quad (2)$$

式中, U 表示SVD分解后的左奇异矩阵; S 是由奇异值组成的对角矩阵; V 表示右奇异矩阵。

SVD分解过程中需要对 S 矩阵进行调整, 以便于过滤掉原始矩阵中的噪声信息。本文保留原始矩阵中信息量, 定义为信息量阈值 σ , 即信息量阈值不大于对角矩阵 S 中前 k 大奇异值的和与所有奇异值之和的比例。设对角矩阵前 k 大奇异值的和为 o_k , 所有奇异值之和为 o , 则信息量阈值满足:

$$\sigma \leq \frac{o_k}{o} \quad (3)$$

为了取得最佳的实验结果, 令左奇异矩阵 S_1 、 S_2 、 S_3 和 S_4 对应的信息量阈值分别为 σ_1 、 σ_2 、 σ_3 和 σ_4 。而 σ_1 、 σ_2 、 σ_3 和 σ_4 的可能取值为0.5、0.6、0.7、0.8和0.9。确定 σ_1 、 σ_2 、 σ_3 和 σ_4 的取值实验过程如下: 遍历 σ_1 、 σ_2 、 σ_3 和 σ_4 的全部取值组合, 以F1值的Top1~Top5的平均值为因变量。当因变量为最大时, σ_1 、 σ_2 、 σ_3 和 σ_4 的取值即作为本文使用的最优取值组合。以HOSVD(uetm)算法为例, 其中部分实验结果如表5所示。当F1值的Top1~Top5的平均值取最大值0.292时, 此时的 σ_1 、 σ_2 、 σ_3 和 σ_4 分别为0.5、0.9、0.7和0.5。

表5 信息量阈值 σ 实验结果表

| σ_1 | σ_2 | σ_3 | σ_4 | F1均值 |
|------------|------------|------------|------------|--------------|
| 0.5 | 0.5 | 0.5 | 0.5 | 0.285 |
| 0.5 | 0.5 | 0.5 | 0.6 | 0.285 |
| 0.5 | 0.5 | 0.5 | 0.7 | 0.286 |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| 0.5 | 0.9 | 0.7 | 0.5 | 0.292 |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| 0.9 | 0.9 | 0.9 | 0.7 | 0.282 |
| 0.9 | 0.9 | 0.9 | 0.8 | 0.284 |
| 0.9 | 0.9 | 0.9 | 0.9 | 0.278 |

信息量阈值确定后就可以求 k 的值。 k 是满足式(3)的最小整数。当 k 值确定后, 从对角矩阵 S 中取出前 k 大奇异值组成新的对角矩阵 S_k ; 从 U 和 V 中选取相应的前 k 个奇异向量, 分别组成新的 U_k 和 V_k , 则降噪后的 A_1 、 A_2 、 A_3 和 A_4 分别为:

$$\begin{aligned} A_1 &\approx U_k^{(1)} S_{1k} V_{1k}^T \\ A_2 &\approx U_k^{(2)} S_{2k} V_{2k}^T \\ A_3 &\approx U_k^{(3)} S_{3k} V_{3k}^T \\ A_4 &\approx U_k^{(4)} S_{4k} V_{4k}^T \end{aligned} \quad (4)$$

实例中, 在信息量阈值 σ 等于0.5, 0.9, 0.7, 0.5的情况下, 矩阵 A_1 、 A_2 、 A_3 和 A_4 对应的 k 值分别是: 1, 2, 2, 2。进而降噪后的 $U_k^{(1)}$ 、 $U_k^{(2)}$ 、 $U_k^{(3)}$ 和 $U_k^{(4)}$ 分别表示如下:

$$\begin{aligned} U_k^{(1)} &= \begin{bmatrix} -0.894 & 4 \\ -0.447 & 2 \end{bmatrix} \\ U_k^{(2)} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ U_k^{(3)} &= \begin{bmatrix} -0.708 & 5 & 0 \\ -0.559 & 2 & -0.447 & 2 \\ -0.279 & 6 & 0.894 & 4 \end{bmatrix} \\ U_k^{(4)} &= \begin{bmatrix} 1 & 0 \\ 0 & -0.850 & 7 \\ 0 & -0.525 & 7 \end{bmatrix} \end{aligned}$$

2.4 构造近似核心张量

近似核心张量 \hat{C} 是由初始张量 A 和降噪后的 U_k 矩阵构造的, 具体计算方式为:

$$\hat{C} = A \times_1 U_k^{(1)T} \times_2 U_k^{(2)T} \times_3 U_k^{(3)T} \times_4 U_k^{(4)T} \quad (5)$$

式中, 符号 \times_n ($n=1, 2, 3, 4$)表示矩阵 $U_k^{(n)T}$ ($n=1, 2, 3, 4$)与张量 A 的 n -模乘积。

实例中, 计算出的近似核心张量 \hat{C} 如表6所示。近似核心张量 $\hat{C} \in R^{2 \times 3 \times 2}$, 结合式(5)易知, 矩阵 A_1 、 A_2 、 A_3 和 A_4 对应的 k 值决定了张量 \hat{C} 的维度。

表6 通过实例数据构造的近似核心张量

| 用户 | 用户情感 | 社刊标签 | 社刊 | 关联权重 |
|-------|-------|-------|-------|----------|
| u_1 | e_1 | t_1 | m_1 | 1.797 3 |
| u_1 | e_1 | t_2 | m_1 | 0.600 0 |
| u_1 | e_2 | t_1 | m_2 | -0.938 0 |
| u_1 | e_2 | t_2 | m_2 | 1.101 1 |

2.5 构造近似张量

由Tucker分解法可知, 初始张量 A 的近似张量 \hat{A} 的计算方式为:

$$\hat{A} = \hat{C} \times_1 U_k^{(1)} \times_2 U_k^{(2)} \times_3 U_k^{(3)} \times_4 U_k^{(4)} \quad (6)$$

上式(6)中, 等号右边部分在上述步骤中都已求解, 故易得实例中的近似张量 \hat{A} 如表7所示。

表7 标签元数据构建的近似张量 \hat{A}

| 用户 | 用户情感 | 社刊标签 | 社刊 | 关联权重 |
|-------|-------|-------|-------|----------|
| u_1 | e_1 | t_1 | m_1 | 1.254 7 |
| u_1 | e_1 | t_2 | m_1 | 1.139 0 |
| u_1 | e_1 | t_3 | m_1 | -0.030 5 |
| u_1 | e_2 | t_1 | m_2 | 0.557 0 |
| u_1 | e_2 | t_2 | m_2 | 0.024 4 |
| u_1 | e_2 | t_3 | m_2 | 0.948 9 |
| u_1 | e_2 | t_1 | m_3 | 0.344 2 |
| u_1 | e_2 | t_2 | m_3 | 0.015 1 |
| u_1 | e_2 | t_3 | m_3 | 0.568 4 |
| u_2 | e_1 | t_1 | m_1 | 0.627 3 |
| u_2 | e_1 | t_2 | m_1 | 0.569 5 |
| u_2 | e_1 | t_3 | m_1 | -0.015 3 |
| u_2 | e_2 | t_1 | m_2 | 0.278 5 |
| u_2 | e_2 | t_2 | m_2 | 0.052 2 |
| u_2 | e_2 | t_3 | m_2 | 0.474 4 |
| u_2 | e_2 | t_1 | m_3 | 0.172 1 |
| u_2 | e_2 | t_2 | m_3 | 0.007 6 |
| u_2 | e_2 | t_3 | m_3 | 0.293 2 |

2.6 生成推荐列表

近似张量 \hat{A} 反映用户、用户情感、社刊标签和社刊之间的关联, 其中近似张量 \hat{A} 中的每一个元素表示用户、用户情感、社刊标签以及社刊之间的关联权重。对比初始张量 A , 不难发现近似张量 \hat{A} 中的元素值发生了变化, 不再是1或者0; 并且会增加一些新元素即出现了用户选择社刊的新纪录。如果要向用户推荐 k 个社刊, 只需选择该用户在近似张量 \hat{A} 中出现的新纪录中关联权重最高的前 k 条记录即可, 即Top- k 推荐。

在实例中, 近似张量 \hat{A} 中出现的新纪录是用户 u_2 选择了社刊 m_2 和 m_3 , 并且 u_2 选择社刊 m_2 的最大关联权重为0.474 4, 而 u_2 选择社刊 m_3 的最大关联权重为0.293 2。因0.474 4大于0.293 2, 故进行Top-1推荐即给每个用户推荐一种社刊, 则在这个实例中, 应给用户 u_2 推荐社刊 m_2 。

3 实验设置及结果

3.1 实验设置

按照本文1.1节清洗数据后, 以用户与社刊为关键来划分训练集和测试集^[29], 即用户选择某一社刊的全部记录只会同时存在于训练集中或者测试集中, 而训练集与测试集的比例采用了5:5、6:4、7:3、8:2和9:1, 共5种组合。本文采用的对比算法主要是5种, 分别是基于物品的协同过滤算法(简称itemCF算

法)、PersonalRank算法、采用用户-用户情感-社刊三元组信息的HOSVD算法(下文简称HOSVD(uem)算法)、采用用户-社刊标签-社刊三元组信息的HOSVD算法(下文简称HOSVD(utm)算法)以及采用用户-社刊热门度-社刊标签-社刊四元组信息的HOSVD算法(下文简称HOSVD(ustm)算法)。其中PersonalRank算法^[9]是基于PageRank的推荐算法, 在用户、情感和物品三部图上计算各顶点之间的相关性, 根据相关性大小从而进行推荐; HOSVD(uem)算法输入的是用户、用户情感以及社刊三部分信息; HOSVD(utm)算法输入的是用户、社刊标签以及社刊三部分信息; HOSVD(ustm)算法中的社刊热门度的定义为:

$$g(m_i) = \frac{h_{m_i}}{h} \tag{7}$$

式中, $g(m_i)$ 为社刊 m_i 的热门度; h_{m_i} 为选择社刊 m_i 的用户数量; h 表示全部用户数量。

实验中 itemCF 算法中的最近邻 $k=10$, PersonalRank算法中标度常数 $\alpha=0.85$ 。HOSVD(ustm)算法中的信息量阈值 σ 分别是: 0.5, 0.7, 0.7, 0.5; HOSVD(uem)算法中的信息量阈值 σ 分别是: 0.5, 0.9, 0.5; HOSVD(utm)算法中的信息量阈值 σ 分别是: 0.5, 0.7, 0.5。评价指标采用推荐算法中常见的准确率(Precision)、召回率(Recall)以及F1值^[29]。其中对于目标用户 $u_i (i=1,2,\dots,n)$ 来说, 进行Top- k 推荐, 准确率被定义为测试集中记录同时出现在推荐列表中的个数与推荐列表长度 k 的比值, 计算所有用户的准确率的平均值 $P(k)$:

$$P(k) = \frac{1}{n} \sum_i \frac{d_i(k)}{k} \tag{8}$$

式中, $d_i(k)$ 表示用户 u_i 在测试集中记录同时出现在推荐列表中前 k 个位置的记录个数; 召回率是出现在在测试集中记录的同时也位于推荐列表前 k 个位置的记录个数与测试集中所有记录个数 C_i 的比值。平均所有用户的召回率, 得到数据集上所有用户的平均召回率, 即为:

$$R(k) = \frac{1}{n} \sum_i \frac{d_i(k)}{C_i} \tag{9}$$

F1值是一种统一准确率和召回率的系统性能评估标准, 定义如下:

$$F1 = \frac{2 \times P(k) \times R(k)}{P(k) + R(k)} \tag{10}$$

3.2 实验结果

由表8可知, 在不同训练集和测试集比例下, 本

文提出的HOSVD(uetm)算法, 相较于其他算法, 以F1值的Top1~Top5的平均值而言一直是最优的, 其中, 比同为四阶的HOSVD(ustm)算法在各比例下分别在数值上提高了0.031、0.033、0.039、0.048和0.035。表3的结果, 也从侧面验证了训练集和测试集的比例并不影响各算法结果的优劣排名。当训练集和测试集比例为6:4时, 如图2所示, 横坐标表示召回率, 纵坐标表示准确率。图2中每个算法出现5个点由左至右依次表示Top1~Top5推荐, 以HOSVD(uetm)算法为例, 第一个点表示HOSVD(uetm)算法在进行Top1推荐时准确率和召回率分别是0.513和0.339, 依次类推。为了更直观的看出本文提出的引入用户情感的HOSVD(uetm)算法的推荐效果, 下文进行各个算法的F1值对比分析。如图3所示, 横坐标分别表示Top1~Top5推荐, 纵坐标表示F1值。以HOSVD(uetm)算法为例, 第一个点表示HOSVD(uetm)算法在进行Top1推荐时F1值是0.408。

由图2和图3可知, HOSVD(uetm)算法在准确率、召回率及F1值性能指标上, 实验结果相不仅优于低阶方法(三阶或传统算法), 同时也比考虑社刊热门度的四阶HOSVD(ustm)算法优异。其中结合表9可知, HOSVD(uetm)算法的F1值的Top1~Top5的平均值, 在数值上比HOSVD(ustm)算法提高了0.033, 比HOSVD(uem)算法提高了0.055, 比HOSVD(utm)算法提高了0.122, 比PersonalRank算法提高了0.234, 比itemCF算法提高了0.278。除此之外, 相较于Top2~Top5推荐, HOSVD(uetm)算法在Top1推荐时的表现最为优异, 准确率、召回率和F1值分别是: 0.513、0.339和0.408。

表8 不同训练集和测试集比例下各算法F1值的Top1~Top5的平均值

| 比例 | HOSVD (uetm) | HOSVD (ustm) | HOSVD (uem) | HOSVD (utm) | Personal Rank | Item CF |
|-----|--------------|--------------|-------------|-------------|---------------|---------|
| 5:5 | 0.355 | 0.324 | 0.291 | 0.197 | 0.071 | 0.018 |
| 6:4 | 0.292 | 0.259 | 0.237 | 0.170 | 0.058 | 0.014 |
| 7:3 | 0.247 | 0.198 | 0.183 | 0.130 | 0.042 | 0.012 |
| 8:2 | 0.189 | 0.141 | 0.117 | 0.099 | 0.027 | 0.008 |
| 9:1 | 0.137 | 0.102 | 0.086 | 0.042 | 0.014 | 0.003 |

表9 各算法F1值的Top1~Top5的平均值

| | HOSVD (uetm) | HOSVD (ustm) | HOSVD (uem) | HOSVD (utm) | Personal Rank | Item CF |
|------|--------------|--------------|-------------|-------------|---------------|---------|
| F1均值 | 0.292 | 0.259 | 0.237 | 0.170 | 0.058 | 0.014 |

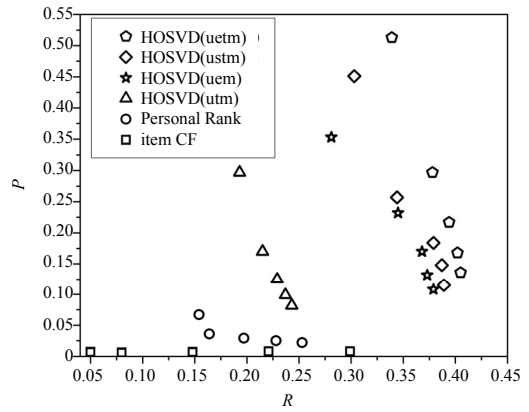


图2 各算法准确率、召回率结果对比图

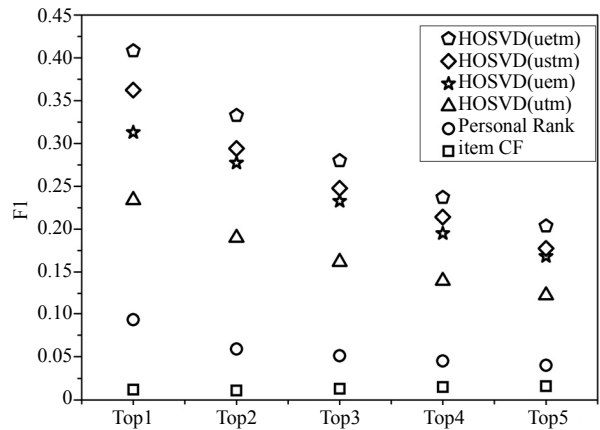


图3 各算法F1值结果对比图

4 结束语

本文在3阶奇异值分解推荐算法的基础上, 提出了一种加入用户情感信息的四阶奇异值分解推荐算法。该方法通过提取评论emoji表情, 依据情感分成积极、中立和消极三类, 分别给每类情感赋予不同的权重, 之后计算不同类emoji表情数量的加权和来表征用户的情感; 再利用四阶张量模型, 存储用户、用户情感、物品标签和物品四元组数据, 应用四阶奇异值分解, 进而进行个性化推荐。实证数据的结果中, 引入emoji表情的四阶HOSVD(uetm)算法不仅比三阶或传统算法结果优异, 还超越了引入社刊热门度的四阶HOSVD(ustm)算法。其中进行Top1推荐时, 本文提出的方法的准确率和召回率分别达到0.513和0.339。Top1推荐结果的优异也表明了本文的方法更适合应用在移动通信端的个性化推荐中。

国内外现有研究成果表明目前大部分推荐方法都是二维或三维的推荐方法, 忽略了时间、空间、设备、用户行为等其他维度信息对推荐结果的影响。而利用多维信息符合个性化推荐服务的发展趋势, 并在多个领域被广泛应用, 具有较强的理论价值和

实践意义。此外, emoji表情已经成为互联网时代必不可少的元素之一, 是人们表达情感的重要载体。通过分析海量的emoji表情信息, 可以了解用户的情绪状况、对某个社会的现象的观点、对某款物品的喜好等, 从而不仅在人机交互、市场营销等领域产生重要影响, 具有十分可观的商业价值; 还可以被用来分析用户心理、态度等, 对维护社会稳定有所帮助, 意义非凡。基于四阶奇异值分解推荐算法在推荐效果上的表现优异, 但因涉及到四阶张量, 算法的时间复杂度较高, 不适用于大规模网络。在未来的工作中, 可在实验中通过用户和物品之间的选择关系以及物品自身的标签信息可先将用户进行聚类, 形成兴趣圈。在这个过程中, 一个用户可同时属于多个兴趣圈, 且一个兴趣圈的用户规模可通过阈值进行限制。之后在每个兴趣圈内应用本文方法, 综合全部兴趣圈的推荐结果, 形成最终推荐。通过这种方式, 可有效减少计算复杂度, 提高本文方法在真实系统中实施的可行性。

参 考 文 献

- [1] 张亮, 柏林森, 周涛. 基于跨电商行为的交叉推荐算法[J]. 电子科技大学学报, 2013, 42(1): 154-160.
ZHANG Liang, BAI Lin-sen, ZHOU Tao. Crossing recommendation based on multi-B2C behavior[J]. Journal of University of Electronic Science & Technology of China, 2013, 42(1): 154-160.
- [2] 张海霞, 吕振, 张传亭, 等. 一种引入加权异构信息的改进协同过滤推荐算法[J]. 电子科技大学学报, 2018, 47(1): 112-116.
ZHANG Hai-xia, LÜ Zhen, ZHANG Chuan-ting, et al. An improved collaborative filtering recommendation algorithm with weighted heterogeneous information[J]. Journal of University of Electronic Science & Technology of China, 2018, 47(1): 112-116.
- [3] 丁哲, 秦臻, 郑文韬, 等. 基于移动用户浏览行为的推荐模型[J]. 电子科技大学学报, 2017, 46(6): 907-912.
DING Zhe, QIN Zhen, ZHENG Wen-tao, et al. A recommendation model based on browsing behaviors of mobile users[J]. Journal of University of Electronic Science & Technology of China, 2017, 46(6): 907-912.
- [4] 刘建国, 周涛, 郭强, 等. 个性化推荐系统评价方法综述[J]. 复杂系统与复杂性科学, 2009, 6(3): 1-10.
LIU Jian-guo, ZHOU Tao, GUO Qiang, et al. Overview of the evaluated algorithms for the personal recommendation systems[J]. Complex Systems & Complexity Science, 2009, 6(3): 1-10.
- [5] 刘贵松, 解修蕊, 黄海波, 等. 基于最短路径信任关系的推荐项目计算方法[J]. 电子科技大学学报, 2014, 43(2): 162-166.
LIU Gui-song, XIE Xiu-rui, HUANG Hai-bo, et al. Fast computing method for items recommendation based on shortest-path trust relationship[J]. Journal of University of Electronic Science & Technology of China, 2014, 43(2): 162-166.
- [6] ZHOU T, REN J, MEDO M, et al. Bipartite network projection and personal recommendation[J]. Physical Review E, 2007, 76(4): 046115.
- [7] SARWAR B M, KARYPIS G, KONSTAN J A, et al. Application of dimensionality reduction in recommender system-a case study[C]//ACM Web kdd Workshop. [S.l.]: ACM, 2000.
- [8] HEYMANN P, KOUTRIKA G, GARCIA-MOLINA H. Can social bookmarking improve web search?[C]//Proceedings of the 2008 International Conference on Web Search and Data Mining. [S.l.]: ACM, 2008: 195-206.
- [9] SYMEONIDIS P, NANOPOULOS A, MANOLOPOULOS Y. Tag recommendations based on tensor dimensionality reduction[C]//Proceedings of the 2008 ACM Conference on Recommender Systems. [S.l.]: ACM, 2008: 43-50.
- [10] SYMEONIDIS P. Clusthosvd: Item recommendation by combining semantically enhanced tag clustering with tensor HOSVD[J]. IEEE Transactions on Systems Man & Cybernetics Systems, 2016, 46(9): 1240-1251.
- [11] 杨君, 汪会玲, 艾丹祥. 一种基于输出情景化的多维信息推荐新方法研究[J]. 情报科学, 2014, 32(11): 126-132.
YANG Jun, WANG Hui-ling, AI Dan-xiang. A new method of multi-dimensional information recommendation based on contextual post-filtering[J]. Information Science, 2014, 32(11): 126-132.
- [12] SAKAGUCHI T, AKAHO Y, OKADA K, et al. Recommendation system with multi-dimensional and parallel-case four-term analogy[C]//IEEE International Conference on Systems, Man, and Cybernetics. [S.l.]: IEEE, 2011: 3137-3143.
- [13] ADOMAVICIUS G, SANKARANARAYANAN R, SEN S, et al. Incorporating contextual information in recommender systems using a multidimensional approach[J]. ACM Transactions on Information Systems, 2005, 23(1): 103-145.
- [14] CHO S, LEE M, JANG C, et al. Multidimensional filtering approach based on contextual information[C]//International Conference on Hybrid Information Technology. [S.l.]: IEEE, 2006: 497-504.
- [15] 王冠楠, 陈端兵, 傅彦. 新闻推荐的多维兴趣模型与传播分析[J]. 计算机科学, 2013, 40(11): 126-130.
WANG Guan-nan, CHEN Duan-bing, FU Yan, et al. Analysis of news diffusion in recommender systems based on multidimensional tastes[J]. Computer Science, 2013, 40(11): 126-130.
- [16] 周娇, 霍欢. 社交网络服务中的多维空间视频推荐算法[J]. 计算机工程, 2015, 41(1): 245-250.
ZHOU Jiao, HUO Huan. Multi-dimensional space video recommendation algorithm in social network service[J]. Computer Engineering, 2015, 41(1): 245-250.
- [17] YU L, LIU C, ZHANG Z K. Multi-linear interactive matrix factorization[J]. Knowledge-Based Systems, 2015, 85: 307-315.
- [18] ZHANG Z K, ZHOU T, ZHANG Y C. Personalized

- recommendation via integrated diffusion on user-item-tag tripartite graphs[J]. *Physica A Statistical Mechanics & Its Applications*, 2009, 389(1): 179-186.
- [19] YU L, HUANG J, ZHOU G, et al. TIIREC: A tensor approach for tag-driven item recommendation with sparse user generated content[J]. *Information Sciences*, 2017, 411: 122-135.
- [20] LÜ L, MEDO M, CHI H Y, et al. Recommender systems[J]. *Physics Reports*, 2012, 519(1): 1-49.
- [21] YUAN Q, CONG G, MA Z, et al. Time-aware point-of-interest recommendation[C]//*International ACM SIGIR Conference on Research and Development in Information Retrieval*. [S.l.]: ACM, 2013: 363-372.
- [22] DING Y, LI X. Time weight collaborative filtering[C]//*ACM International Conference on Information and Knowledge Management*. [S.l.]: ACM, 2005: 485-492.
- [23] 李小勇, 桂小林. 可信网络中基于多维决策属性的信任量化模型[J]. *计算机学报*, 2009, 32(3): 405-416.
LI Xiao-yong, GUI Xiao-ling. Trust quantitative model with multiple decision factors in trusted network[J]. *Chinese Journal of Computers*, 2009, 32(3): 405-416.
- [24] ALKIRE S, FOSTER J. Counting and multidimensional poverty measurement[J]. *Journal of Public Economics*, 2012, 95(7): 476-487.
- [25] 袁炜毅, 杨东援. 基于多维分析技术的公路交通事故分析方法研究[J]. *交通信息与安全*, 2000, 18(4): 4-9.
QIU Wei-yi, YANG Dong-yuan. Analysis of road traffic accident based on multi-dimensional analysis[J]. *Computer & Communications*, 2000, 18(4): 4-9.
- [26] JÄSCHKE R, MARINHO L, HOTHO A, et al. Tag recommendations in folksonomies[M]//*Knowledge Discovery in Databases: PKDD 2007*. Berlin Heidelberg: Springer, 2007: 506-514.
- [27] BATAGELJ V, ZAVERŠNIK M. Generalized cores [EB/OL]. [2018-01-07]. <http://arXiv preprint cs/0202039>, 2002.
- [28] KRIPPENDORFF K. Agreement and information in the reliability of coding[J]. *Communication Methods and Measures*, 2011, 5(2): 93-112.
- [29] 项亮. 推荐系统实践[M]. 北京: 人民邮电出版社, 2012.
XIANG Liang. *Practical application of recommendation* [M]. Beijing: Posts & Telecom Press, 2012.

编辑 蒋晓