



DOI:10.13364/j.issn.1672-6510.20190230

一种基于标签的电影组推荐方法

熊聪聪, 林颖, 史艳翠, 冯阔
(天津科技大学人工智能学院, 天津 300457)

摘要: 现有的电影推荐系统大多是分析用户评分矩阵, 没有考虑电影中的标签信息对推荐结果的影响, 造成标签信息利用率较低. 针对该问题提出一种基于标签的电影组推荐方法. 通过改进的 TF-IDF 方法得到用户-标签矩阵以及结合时间因素得到归一化的用户评分矩阵, 分别计算用户相似度并融合, 通过融合相似度进行群组划分. 在此基础上计算组成员中电影标签的好评率, 根据好评率设置张量初始总权重. 根据近似张量值进行群组偏好融合, 得到组推荐列表. 在 MovieLens 数据集上进行实验对比, 结果表明本方法在准确率上有明显提高, 召回率与 F 值有较大提高.

关键词: 组推荐; 标签推荐; 张量分解

中图分类号: TP391 文献标志码: A 文章编号: 1672-6510(2020)05-0074-07

A Movie Group Recommendation Method Based on Tag

XIONG Congcong, LIN Ying, SHI Yancui, FENG Kuo

(College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China)

Abstract: Most of the existing movie recommendation systems analyze the user rating matrix without considering the influence of tag information in movies on recommendation results, resulting in low utilization rate of tag information. This research proposes a new film group recommendation method based on tag. The improved TF-IDF method is used to obtain a user-tag matrix and a normalized user scoring matrix is also obtained by combining time factors. User similarities are calculated and fused, and then the users are divided into groups according to fusing their similarities. On this basis, the rating of movie tags in groups is calculated, and the tensor initial total weight is set according to the rating. Finally, group preference fusion is carried out according to the approximate tensor value to obtain a group recommendation list. The experimental comparison on MovieLens data set shows that the accuracy rate is obviously improved by this method, and both the recall rate and F-Measure are greatly improved as well.

Key words: group recommendation; tag recommendation; tensor decomposition

当今互联网发展迅速, 同时互联网信息量以爆炸式的速度增长, 而这种信息量对用户个人搜索而言异常困难, 互联网用户需要花费大量时间去筛选符合条件的信息. 推荐系统成为如何从众多信息中筛选出符合用户需要信息的方法之一. 传统推荐系统通过分析个人行为推荐, 而现在社交网络中, 用户不仅不是孤立的而且极有可能还与其他用户存在大量的关系. 在影视作品中, 用户会自主地根据观看内容为电影标记符合内容的标签. 标签作为电影最重要附属信息, 同时也是用户对电影最直观的感受, 因此标签

可用于提取用户行为首选项. 大多数用户有时会与朋友和家人一起观看电影, 需要考虑群组成员观看的电影类型.

为了解决上述推荐问题, 对群组推荐问题提出相关研究. 目前, 部分组推荐在用户评分预测中直接使用聚类方法对用户进行分类, 这样用户的相似度仅依赖用户评分矩阵, 而部分组推荐则是挖掘分组之前用户间的相似度信息^[2]. 这样对于标签的信息量挖掘不足, 标签只是作为评分矩阵的附加信息, 起到较小的作用. 例如, 郑剑等^[3]将标签信息加入到项目相似度

收稿日期: 2019-09-03; 修回日期: 2020-02-07

基金项目: 天津市教委科研计划资助项目(2018KJ105); 国家自然科学基金资助项目(61402331, 61702367)

作者简介: 熊聪聪(1961—), 女, 四川泸州人, 教授; 通信作者: 史艳翠, 副教授, syc@tust.edu.cn

的计算过程中,随后融入到矩阵分解. 研究人员可以通过构建用户、项目、标签的三维张量,挖掘数据之间的潜在关系. 例如,杨林等^[4]利用标签惩罚机制和用户评分构建张量;曾辉等^[5]则根据用户使用产品的活跃性以及用户之间相似性对用户聚类,构建用户、项目、标签之间的权重. 这样能将标签与用户的联系建立起来,而不是作为权值添加进评分矩阵.

鉴于以上研究,本文提出了一种基于标签的电影组推荐方法. 首先对标签信息与结合时间因素的评分矩阵分别计算用户相似度,融合二者相似度后,对用户聚类;然后在群组中找出高评分率的标签,建立用户、电影、标签的初始张量,运用梯度下降法^[6]迭代求解近似张量;最后融合群组偏好,给出推荐列表.

1 相关工作

1.1 TF-IDF

TF(Term Frequency)表示词条 N 在某一文本词条总数中的频率, IDF(Inverse Document Frequency)取包含 N 词条的文件数在总文件中占比的对数. TF-IDF(词频-逆向文件频率)计算的是文本中所有词条的重要度权重.

1.2 张量基本知识

张量是对多维数组的线性分析总称,如一阶张量代表向量,二阶张量代表矩阵,高阶张量是多维数组的统称^[7]. 本文是由用户、电影、标签 3 种属性构建的张量,即 $X \in \mathbf{R}^{I_U \times I_M \times I_T}$,张量 \mathbf{R} 的维度大小由用户数 I_U 、电影数 I_M 、标签数 I_T 组成.

不同推荐系统在构建张量方式上也不同. “0/1”表示形式即用户-产品-标签的三维张量值为 0 或 1,用户使用一个产品并属于一个标签,则对应的张量值为 1;否则,用户没有使用该产品也不属于该标签,则张量值为 0^[8]. 多元权值张量通过构建用户-电影-标签之间的关系权值建立初始张量,而对没有看过的电影及标签张量值为 0.

1.3 张量分解

本文采用的分解算法为 tucker 分解,该算法分解后得到张量各维度下的矩阵和 1 个核张量,3 个矩阵包含各维度上的潜在特征^[9]. 表示为

$$X \approx A \times U \times M \times T \quad (1)$$

其中: $X \in \mathbf{R}^{I_U \times I_M \times I_T}$ 为初始张量; $A \in \mathbf{R}^{r_1 \times r_2 \times r_3}$ 是核心张量; $U \in \mathbf{R}^{I_U \times r_1}$ 、 $M \in \mathbf{R}^{I_M \times r_2}$ 、 $T \in \mathbf{R}^{I_T \times r_3}$ 分别为各维度下的特征矩阵. 张量上的任一元素估计值为

$$x_{ijk} \approx \sum_{x=1}^{r_1} \sum_{n=1}^{r_2} \sum_{l=1}^{r_3} (a_{xnl} \times u_{ix} \times m_{jn} \times t_{kl}) \quad (2)$$

其中: a_{xnl} 表示核张量 A 的元素; u_{ix} 、 m_{jn} 、 t_{kl} 分别表示矩阵 U 、 M 、 T 元素. 如果为得到近似张量,即简单地使总的误差最小,那么可以将张量分解的逼近问题转化为一个无约束的优化问题,即梯度下降方法

$$\min J = \frac{1}{2} \sum_{(i,j,k) \in S} e_{ijk}^2 = \frac{1}{2} \sum_{(i,j,k) \in S} (x_{ijk} - \sum_{x=1}^{r_1} \sum_{n=1}^{r_2} \sum_{l=1}^{r_3} (a_{xnl} \times u_{ix} \times m_{jn} \times t_{kl}))^2 \quad (3)$$

其中 S 表示所有非零元素集合. 对目标函数 J 对 u_{im} 、 m_{jn} 、 t_{kl} 和 a_{mnl} 求偏导,得到

$$\begin{aligned} \frac{\partial J}{\partial u_{ix}} &= - \sum_{j,k:(i,j,k) \in S} e_{ijk} \times (\sum_{n=1}^{r_2} \sum_{l=1}^{r_3} (a_{xnl} \times m_{jn} \times t_{kl})) \\ \frac{\partial J}{\partial m_{jn}} &= - \sum_{i,k:(i,j,k) \in S} e_{ijk} \times (\sum_{x=1}^{r_1} \sum_{l=1}^{r_3} (a_{xnl} \times u_{ix} \times t_{kl})) \\ \frac{\partial J}{\partial t_{kl}} &= - \sum_{i,j:(i,j,k) \in S} e_{ijk} \times (\sum_{x=1}^{r_1} \sum_{n=1}^{r_2} (a_{xnl} \times u_{ix} \times m_{jn})) \\ \frac{\partial J}{\partial a_{xnl}} &= - \sum_{(i,j,k) \in S} e_{ijk} \times u_{ix} \times m_{jn} \times t_{kl} \end{aligned} \quad (4)$$

再根据梯度下降方法, u_{ix} 、 m_{jn} 、 t_{kl} 和 a_{xnl} 在每次迭代过程中的更新公式为

$$\begin{aligned} u_{ix} &= u_{ix} + \alpha \sum_{j,k:(i,j,k) \in S} e_{ijk} \times (\sum_{n=1}^{r_2} \sum_{l=1}^{r_3} (a_{xnl} \times m_{jn} \times t_{kl})) \\ m_{jn} &= m_{jn} + \alpha \sum_{i,k:(i,j,k) \in S} e_{ijk} \times (\sum_{x=1}^{r_1} \sum_{l=1}^{r_3} (a_{xnl} \times u_{ix} \times t_{kl})) \\ t_{kl} &= t_{kl} + \alpha \sum_{i,j:(i,j,k) \in S} e_{ijk} \times (\sum_{x=1}^{r_1} \sum_{n=1}^{r_2} (a_{xnl} \times u_{ix} \times m_{jn})) \\ a_{xnl} &= a_{xnl} + \alpha \sum_{(i,j,k) \in S} e_{ijk} \times u_{ix} \times m_{jn} \times t_{kl} \end{aligned} \quad (5)$$

其中: 更新公式中的求和项下标 $j, k: (i, j, k) \in S$ 、 $i, k: (i, j, k) \in S$ 、 $i, j: (i, j, k) \in S$ 分别表示矩阵 $X(i, :, :)$ 、 $X(:, j, :)$ 、 $X(:, :, k)$ 所有非零元素索引.

2 算法设计

本文首先对标签信息与结合时间因素的评分矩阵分别计算用户相似度,融合二者相似度后,对用户聚类;然后挖掘群组成员中的观影信息,统计群组用户观看的高评分率电影,计算电影标签的好评率,建立“用户-电影-标签”的关系张量,并对其分解,得到近似张量;最后对群组偏好融合,得到群组推荐列表.

2.1 计算个人总体偏好

2.1.1 基于标签的用户相似度

电影标签是对每部电影的总结概述或是描述该电影所拥有的特点,以用户对电影标签的偏好代表用户对不同电影特征属性的偏好.通过各个用户观看过的标签在总标签的重要程度,以及用户对标签的观看占比计算用户对每个标签的偏好权重值,本文在TF-IDF基础上进行改进,TF-IDF作为一种统计方法,用于计算标签对于所有标签的是否有代表性^[10].用户对标签的偏好表示为

$$p_{u_i,t_j} = \frac{f_{u_i,t_j}}{\sum_{k=1}^n f_{u_i,t_k}} \log \frac{m_t}{m_{t_j}} \quad (6)$$

其中: f_{t_j,u_i} 表示用户 u_i 观看的标签 t_j 个数; m_t 表示所有包含标签的电影数; m_{t_j} 表示包含标签 t_j 的电影数.在所有标签中,用户观看的标签出现频率较低,可以更好地区别用户之间偏好,相较于高频率标签的出现更具有代表性.而同时标签出现在电影中的次数越少,对于电影的区分度则越高,标签更能区分用户的偏好.标签的重要性为

$$w_{t_j} = \log \left(\frac{t}{t_j} \times \frac{n}{n_{t_j}} \right) \quad (7)$$

其中: t 表示所有电影的标签出现次数; t_j 表示标签 t_j 在总标签中出现次数; n 表示用户数; n_{t_j} 表示观看过标签 t_j 的用户数.标签之间比率表示标签重要性,用户之间比率表示标签独特性.用户对标签的偏好权重为

$$p'_{u_i,t_j} = p_{u_i,t_j} \times w_{t_j} \quad (8)$$

用户基于标签信息的相似度公式为

$$sim_{tag}(u_i, u_j) = \frac{\sum_{k \in t_{ij}} (p'_{u_i,t_k} - \overline{p'_{u_i}})(p'_{u_j,t_k} - \overline{p'_{u_j}})}{\sqrt{\sum_{k \in t_{ij}} (p'_{u_i,t_k} - \overline{p'_{u_i}})^2 \sum_{k \in t_{ij}} (p'_{u_j,t_k} - \overline{p'_{u_j}})^2}} \quad (9)$$

其中: t_{ij} 表示用户 u_i 和用户 u_j 均观看过的标签, $\overline{p'_{u_i}}$ 和 $\overline{p'_{u_j}}$ 分别表示用户对于标签偏好权重的均值.

2.1.2 基于评分矩阵的用户相似度

用户兴趣随时间的推移可能发生变化,在不同时间段会有不同需求.将以前的用户历史信息与现在的信息作为同等价值的信息处理,会使用户行为分析有偏差.用户最近时段的偏好,能够更准确地推断用户偏好.通过指数函数模拟遗忘函数,以此作为时间

对用户偏好影响^[11].

$$r'_{u_i,m_k} = r_{u_i,m_k} \times e^{-T_0 \Delta t} \quad (10)$$

其中: Δt 表示当前时间与评分时间差,时间计算单位为周; T_0 是半衰期常量; r_{u_i,m_k} 表示用户 u_i 对电影 m_k 的评分.基于评分矩阵的用户相似度公式为

$$sim_r(u_i, u_j) = \frac{\sum_{k \in m_{ij}} (r'_{u_i,m_k} - \overline{r'_{u_i}})(r'_{u_j,m_k} - \overline{r'_{u_j}})}{\sqrt{\sum_{k \in m_{ij}} (r'_{u_i,m_k} - \overline{r'_{u_i}})^2 \sum_{k \in m_{ij}} (r'_{u_j,m_k} - \overline{r'_{u_j}})^2}} \quad (11)$$

其中: m_{ij} 表示用户 u_i 和用户 u_j 共同观看过电影, $\overline{r'_{u_i}}$ 和 $\overline{r'_{u_j}}$ 分别表示用户对于电影评分的均值.用户观看电影的数量差会对用户偏好相似预测产生偏差,需要修正因数量级造成的相似度变差.为此,加入用户参数 λ 修正数量级差造成的偏好相似.

$$\lambda = m_{u_{ij}} / |m_{u_i} \cup m_{u_j}| \quad (12)$$

其中: $m_{u_{ij}}$ 表示用户 u_i 与用户 u_j 观看相同的电影数; m_{u_i} 表示用户 u_i 观看的电影数; m_{u_j} 表示用户 u_j 观看的电影数.修正后的用户相似度为

$$sim_{rating}(u_i, u_j) = \lambda \times sim(u_i, u_j) \quad (13)$$

2.1.3 计算用户总体偏好相似度

基于标签信息计算出的用户相似度,与基于评分矩阵计算出的用户相似度存在差异.标签传输的信息是对电影所属信息的总结,而电影能够传达的信息要远远多于标签信息所能概括的信息.用户在选择观看的电影时不仅仅会选择熟悉的电影类型,有时会挑选拥有熟悉标签的冷门电影类型.比如,用户在看完漫威电影后,需要看看其他类型的电影,也会选择主演们演绎的其他电影.因此,将标签信息计算出的相似度结合评分矩阵计算出的相似度作为用户总相似度作为用户之间的相似度,其计算式为

$$sim(u_i, u_j) = (1 - \alpha) sim_{rating}(u_i, u_j) + \alpha sim_{tag}(u_i, u_j) \quad (14)$$

其中 α 用来权衡用户标签相似度影响大还是评分相似度影响大.

在计算出总体相似度后,通过用户之间的相似度划分群组,将拥有高相似偏好的用户集合到一起进行推荐.对于如何将用户分组,可以通过分析用户对其他用户的综合相似度,从而采用 K-means 聚类方法,将拥有高相似度的用户聚类,得到用户群组^[12].

2.2 电影好评率计算

用户分组后的群组中,用户之间偏好相似,直接

建立群组的“0/1”初始张量,用户对电影和标签的喜好没有任何表现.通过加入权值的方式,考虑用户对各个维度下的元素喜好程度,同时也可以反映各维度元素之间的关系.直接将用户评分作为初始张量的权重,能够反映用户对电影的偏好.由于标签是对电影的某一方面总结,通过计算用户对电影的评分,以此隐式反映含有该标签的电影是否好看,从而作为标签的好评率.

由于在群组中,用户偏好相似,在张量创建初始可以为高评分率的电影添加一个相对较高的权重值,使得求解近似张量时好电影能够得到更好的推荐.群组评分权值

$$gw_{t_k} = \frac{\left(\sum_{i=1}^n \sum_{j=1}^a r_{u_i, m_j, t_k} \times f_{r_{u_i, m_j, t_k}} + 1 \right) \times a}{\sum_{i=1}^n \sum_{j=1}^a r_{u_i, m_j, t_k} \times |G_M|} \quad (15)$$

其中: r_{u_i, m_j, t_k} 表示用户 u_i 对标签 t_k 的评分; $f_{r_{u_i, m_j, t_k}}$ 表示用户判断评分是否达到高评分标准; n 表示看过包含标签 t_k 的用户数; a 表示包含标签 t_k 的电影数; $|G_M|$ 表示该群组中用户看过的电影数.

群组对于一个标签评价高低表明大多数观看人对标签的喜好程度,并以此给没有观看过该标签的用户进行推荐.而对于用户看过拥有该标签的电影,通过计算用户对标签的好评率,降低群组对标签的好评率权重,以此作为用户的权重,避免出现用户不喜欢的电影类型.用户标签好评率权值为

$$w_{u_i, t_k} = \frac{\sum_{j=1}^n r_{u_i, m_j, t_k} \times f_{r_{u_i, m_j, t_k}} + 1}{\sum_{j=1}^n r_{u_i, m_j, t_k}} \quad (16)$$

其中: r_{u_i, m_j, t_k} 表示用户 u_i 对含有标签 t_k 的电影评分, $f_{r_{u_i, m_j, t_k}}$ 表示用户评分是否达到高评分标准.

高评分的设定标准通过对比国内外各电影评分网站的评分标准,以及观众通过评分判定电影好坏的程度.国外的影评网烂番茄对于一部电影正面评价超过 60%,会给出新鲜的评价,网站 80% 以上影评人给出超过 75% 的评分值,电影达到坚定新鲜标准.同样是国外影评网的 IMDb,作为老牌影评网,十分制下 6~7 分的电影属于正常范围,达到 8 分基本达到 top250 水平.与 IMDb 相似的国内豆瓣影评网评分同样采取十分制,用户对于电影评分判定分水岭为 6 分,达到该分数线的电影进入可观看行列,豆瓣 top250 分值也在 8 分左右.基于以上各大知名网站

的比较,用户对于好电影的评价都在分值 6 分以上,佳片的评分均达到 8 分左右.但由于能达到 8 分的电影在各网站相对于电影片库的数量均是占比较少,用户分组后能达到电影推荐条件的数量又需要减少.为提高推荐质量和数量取 6 分电影与 8 分电影的中位数,将 7 分电影定为高评分的电影评价标准,即为五分制的 3.5 分.

$$f_{r_{u_i, m_j, t_k}} = \begin{cases} 1 & r_{u_i, m_j, t_k} \geq 3.5 \\ 0 & r_{u_i, m_j, t_k} < 3.5 \end{cases} \quad (17)$$

在对用户、电影、标签张量权重中综合群体对标签电影的评分高低,结合各个用户对该电影标签的偏好,综合作为总权重,即

$$w_{u_i, m_j, t_k} = r_{u_i, m_j, t_k} \times gw_{t_k} \times w_{u_i, t_k} \quad (18)$$

张量的 3 个维度构建出用户 U 、电影 M 、标签 T 的相互关系,由计算得到的综合权值确定对应张量维度上的元素值并构建初始三维张量.通过分解初始张量得到核张量与 3 个因子矩阵,之后采用梯度下降更新获得新初始核张量与因子矩阵,乘积后相似张量 X ,其中张量中的值为 x'_{u_i, m_j, t_k} .对于整体的基于标签的电影组推荐算法设计流程如下:

输入:用户、电影、评分、时间数据的四元组 (U, M, R, Ti) ,用户、电影、标签三元组 (U, M, T) .

输出:近似张量 X .

- (1) (U, M, T) 按公式(6)~(9)计算基于标签的用户相似度 sim_{tag} ;
 - (2) (U, M, R, Ti) 按公式(10)~(13)计算基于评分矩阵的用户相似度 sim_{rating} ;
 - (3) 根据公式(14)融合 sim_{tag} 和相似度 sim_{rating} 的相似度得到融合相似度 $sim(u_i, u_j)$;
 - (4) 根据融合相似度用 K-means 分组用户,得到群组用户;
 - (5) 计算群组用户中标签好评率,按公式(15)~(17)计算;
 - (6) 按公式(18)建立初始张量权重值.
- 初始张量按公式(4)和(5)更新迭代张量,得到近似张量 X .

2.3 偏好融合

通过近似张量可以得到用户对标签的选择概率,通过这个概率计算用户对每部电影的选择概率.确定群组成员对每部电影的选择概率后,为平衡成员对电影选择概率不同,需要融合用户对电影的偏好,以此满足组内成员的喜好.目前均值策略是组推荐应

用较广泛的策略,均值策略采用组内成员对每部电影的
选择概率均值作为组选择概率,而后为组内成员选
取 top_N 作为推荐列表.

3 实验

3.1 实验数据

本文使用 MovieLens 数据集,包含用户对视频的
评分,以及用户给视频标注的标签数据.通过对数
据处理,得到 100 004 条数据,这些数据集中包含
694 个用户、9 743 个电影、1 239 个标签,每个用户对
电影的评分范围为 0.5 ~ 5,用户的频分等级划分为
十,评分区间间隔为 0.5.

3.2 评价标准

本文使用 F 值 (F -Measure)、准确率 ($Precision$)
和召回率 ($Recall$) 作为算法的评价标准.在推荐系统
中, F 值作为评价推荐结果, F 值越高推荐结果越好.

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (19)$$

其中: $Precision$ 表示在推荐列表中得到的推荐结果
与测试集中实际情况相同的物品数与所有的推荐物
品数的比值, $Recall$ 指的是推荐列表中准确的结果占
测试样本的比例.准确率和召回率的计算公式为

$$Precision = \frac{|test \cap top_N|}{N} \quad (20)$$

$$Recall = \frac{|test \cap top_N|}{|test|} \quad (21)$$

其中: $test$ 表示测试集中群组用户观看电影合集;
 top_N 表示推荐的结果; N 表示推荐的数目.在实验
过程中,处理后的数据集将被划分成两部分:训练集
和测试集.其中,训练集占 80%,测试集占 20%.

3.3 实验对比方法

本文实验选择的对比方法如下:

(1) method1,对原始评分矩阵聚类分组后的张
量分解方法 OR (Original Rating).

(2) method2,评分矩阵与标签信息相似聚类的
张量分解方法 TIR (Tag Information and Rating).即
本文提出用户总体偏好方法,通过标签相似聚类得
到的用户群与通过评分矩阵相似聚类得到的用户群,
同时通过调整 α 值,计算各不同比重下聚类得
到的用户群推荐结果的差异.

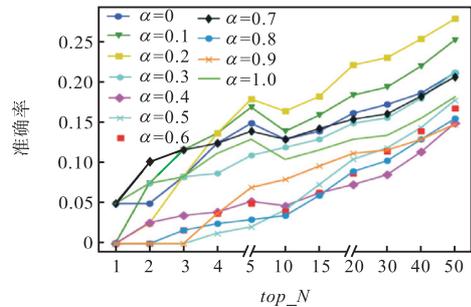
(3) method3,在 method2 方法上加入文献[4]中
的 PMUS (Penalty Mechanism and User Score) 方法进
行张量分解 TIR_PMUS.

(4) method4,在 method2 方法上加入本文提出
的电影好评率 FR (Favorable Rate) 后的张量分解方法
TIR_FR.

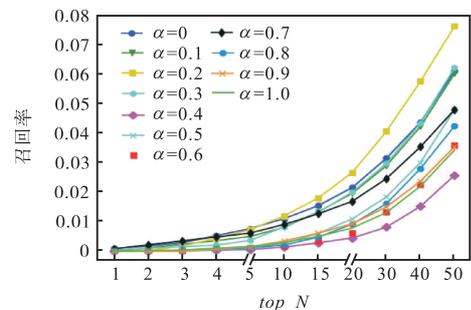
3.4 实验结果

3.4.1 相似度融合中参数 α 对推荐结果的影响

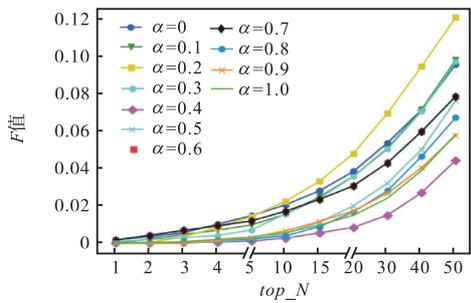
当参数 α 取值不同时,TIR 推荐结果如图1所示.



(a) TIR 推荐结果的准确率比较



(b) TIR 推荐结果的召回率比较



(c) TIR 推荐结果的 F 值比较

图 1 参数 α 对推荐结果的影响

Fig. 1 Effect of parameter α on recommendation results

从图 1 中得到只有标签信息计算的用户相似度,
与通过评分矩阵计算的用户相似度得到的群组推荐
结果.单独计算两种信息相似度后的推荐结果与融
合相似度得到的群组推荐结果对比显示,评分矩阵
得到的结果显示在所有比重中表现较好,排到所有比
重的第三,而标签信息得到的结果推荐比评分矩阵
结果差,排到第五.从图 1 可看出,准确率、召回
率和 F 值均随推荐数增长而增加,推荐数目在 50
时准确率、召回率和 F 值达到最高点.综合 3 种
评价标准,

$\alpha = 0.2$ 的推荐结果表现比其他比重推荐结果表现好. 准确率在推荐数目为 4 时居于所有比重的最高点, 并且推荐数目增加一直是最佳推荐结果. 召回率与 F 值的表现同其他比重差值在 0.02 中变化, 在推荐数达到 5 后居于所有比重最高点, 并随推荐数增长, 召回率与 F 值增加最快.

3.4.2 加入好评率后相似度融合中参数 α 对推荐结果的影响

method4 在 $\alpha = 0.3$ 时, 准确率、召回率与 F 值的结果均表现比其他比重优秀(图 2). 由于推荐列表是与群组用户的合集进行对比, 推荐列表相较测试集数量过于稀疏, 使得准确率很高, 同时召回率和 F 值整体很小.

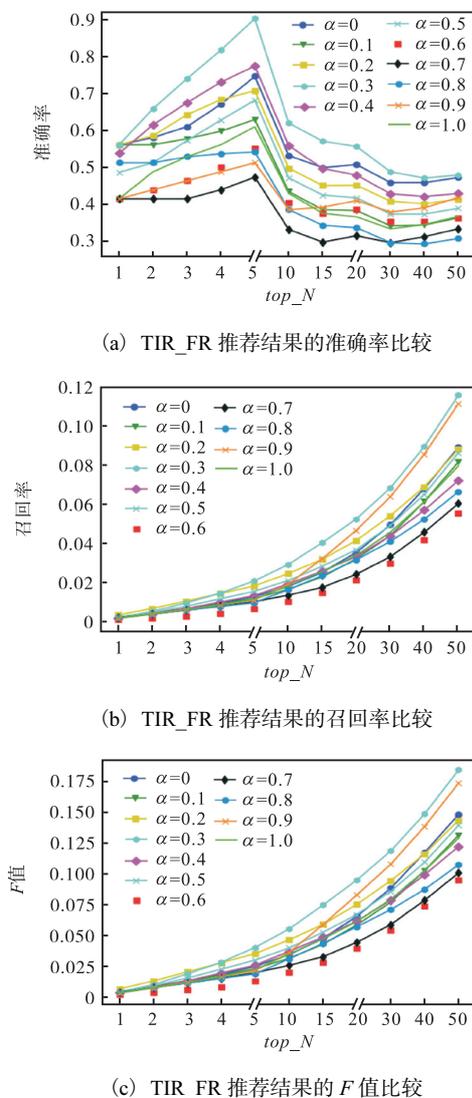


图 2 加入好评率后参数 α 对推荐结果的影响

Fig. 2 Effect of α changes on results after adding favorable rate

由于好评率计算中以评分作为选择概率, 并在此

基础上计算组内成员对标签的好评率, 将评分与标签好评率结合作为最终权重. 这使得分组中评分信息与标签信息对推荐结果影响程度改变, 造成 TIR 和 TIR_FR 在不同 α 值下取得最佳推荐结果.

3.4.3 不同方法推荐结果的比较

图 3 表示 4 种方法在准确率、召回率和 F 值上的比较.

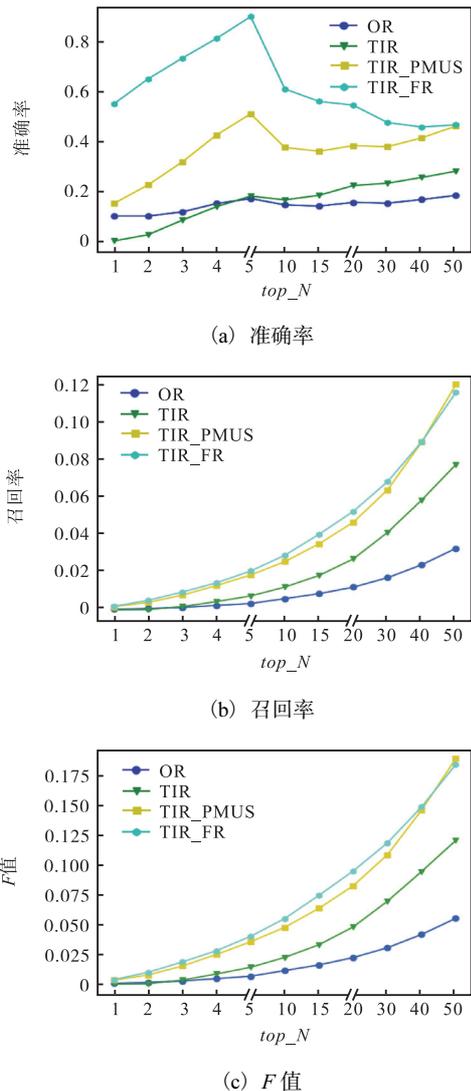


图 3 不同方法推荐结果的比较

Fig. 3 Comparison of results by different recommendation methods

OR 的准确率、召回率和 F 值增长速度相较于其他 3 种较低, 准确率基本持平. TIR 在推荐数目小于 4 时, 准确率低于 OR, 但是在召回率和 F 值的表现上要优于 OR. TIR_PMUS 与 TIR_FR 在各方面均优于其他两种方法, 准确率在推荐数为 5 时达到峰值, 随后均在推荐数到 10 时急速下降. 之后 TIR_PMUS 准确率能够在下降后缓慢增长, 接近 0.5 左右, 但

TIR_FR 准确率一直在缓慢下降. TIR_FR 在推荐数达到 50 时, 准确率与 TIR_PMUS 表现结果相同, 均达到 0.5. TIR_FR 的召回率与 F 值表现在推荐数达到 40 后被 TIR_PMUS 超过. 加入标签后的推荐相对于只通过用户评分预测用户偏好电影, 更相似度用户之间推荐效果更佳突出, 通过标签信息构建的张量在比只在用户评分中构建张量能够得到更佳的结果.

4 结 语

本文针对电影推荐中标签信息的利用率低, 提出一种基于标签的电影组推荐方法. 针对现有组推荐技术大多是在随机分组或原始信息上聚类分组后推荐, 现提出分组时尽可能提高组内成员相似度, 通过评分矩阵与标签信息进行融合相似度, 以此提高组内相似度. 通过实验验证本方法对推荐准确率有所提高. 本实验在使用张量进行用户-电影-标签信息隐形特征分析时, 张量分解运行时间长, 需要提高推荐速度.

参考文献:

- [1] 唐福喜, 刘克剑, 冯玲, 等. 基于用户交互行为的群组推荐偏好融合策略[J]. 西华大学学报: 自然科学版, 2016, 35(3): 51-56.
- [2] 王海艳, 董茂伟. 基于动态卷积概率矩阵分解的潜在

群组推荐[J]. 计算机研究与发展, 2017, 54(8): 1853-1863.

- [3] 郑剑, 王啸乾. 融合标签相似度的差分隐私矩阵分解推荐算法[J/OL]. 计算机应用研究: 1-6[2020-02-06]. <https://doi.org/10.19734/j.issn.1001-3695.2018.08.0654>.
- [4] 杨林, 顾军华, 官磊, 等. PMUS-HOSGD 张量分解方法及其在标签推荐中的应用[J]. 计算机工程, 2018, 44(11): 300-305.
- [5] 曾辉, 胡强, 淦修修. 基于多元关系的张量分解标签推荐方法[J]. 计算机应用研究, 2019, 36(10): 2907-2910.
- [6] 肖亦康. 上下文感知的群组服务推荐研究[D]. 南京: 南京邮电大学, 2017.
- [7] Kolda T G, Bader B W. Tensor decompositions and applications[J]. SIAM Review, 2009, 51(3): 455-500.
- [8] 杨秋勇. 应用张量分解法学习最优评分的推荐系统研究[D]. 广州: 华南理工大学, 2012.
- [9] Golub G, Charles V L. Matrix Computation[M]. 4th ed. Beijing: Posts & Telecom Press, 2014: 731-739.
- [10] 叶柏龙, 徐静静, 严笋. 基于评分和项目特征的群组推荐方法[J]. 计算机应用研究, 2017, 34(4): 1032-1035.
- [11] 史艳翠, 戴浩男, 石和平, 等. 一种基于时间戳的新闻推荐模型[J]. 计算机应用与软件, 2016, 33(6): 40-43.
- [12] 金涛. 社会群组快速生成及其群组推荐研究[D]. 上海: 华东师范大学, 2016.

责任编辑: 郎婧

(上接第 35 页)

- [3] 万丽, 葛静华, 茹成章, 等. 有源电力滤波器的应用分析[J]. 机械设计, 2018, 35(S1): 279-283.
- [4] EI-Hahronk M, Darwish M K, Mehta P. Active power filters: A review[J]. IEE Proceedings-Electric Power Applications, 2000, 147(5): 403-413.
- [5] 武健, 刘瑜超, 李晓萌, 等. 基于瞬时值比较的电流型有源滤波器谐振抑制方法[J]. 电工技术学报, 2013, 28(19): 73-78.
- [6] 全生财, 郭文勇, 李大飞. 三相电流型 PWM 整流器改进型模型预测控制[J]. 电工电能新技术, 2017, 36(8): 31-35.
- [7] 李卫东, 李达义, 杨国庆, 等. 一种基于复合控制的通用型有源电力滤波器[J]. 电工电能新技术, 2016, 35(2): 14-17.
- [8] 熊宇, 高潮. 一种新型双频控制电流型有源电力滤波

器[J]. 电力电子技术, 2013, 47(8): 58-60.

- [9] 墨柯. 超导储能技术及产业发展简介[J]. 新材料产业, 2013(9): 61-65.
- [10] 郭文勇, 蔡富裕, 赵闯, 等. 超导储能技术在可再生能源中的应用与展望[J]. 电力系统自动化, 2019, 43(8): 2-19.
- [11] 李海滨, 李雪, 胡富静. 超导储能技术在现代配电网中的应用研究[J]. 技术与市场, 2012, 19(9): 5-8.
- [12] 黄晶晶, 孙元岗, 同向前. 基于滑模变结构方法的 APF 电流控制策略[J]. 电气传动, 2016, 46(5): 44-48.
- [13] 白圣建, 黄新生. 变结构控制的抖振问题研究[J]. 计算机仿真, 2006, 23(6): 155-158.
- [14] 陈志梅, 王贞艳, 张井岗. 滑模变结构控制理论及应用[M]. 北京: 电子工业出版社, 2012.

责任编辑: 郎婧