DOI:10.13878/j.cnki.jnuist.2019.05.013

汪祥舜[1,2]　郑孝遥[1,2]　朱德义[1,2]　章玥[3]　孙丽萍[1,2]

基于偏好融合的群组推荐研究

摘要

传统的推荐系统主要针对单个用户,但随着社会和电子商务的快速发展,人们越来越多地以多个用户的形式一起参与活动,而群组推荐旨在为多个用户组成的群组提供服务,已成为当前研究的热点之一.针对目前群组推荐准确率低,群组成员之间偏好冲突难以融合的问题,本文提出了一种新的共识模型策略,融合了群组领袖影响因子和项目热度影响因子,基于K近邻为目标群组寻找邻居群组,借鉴邻居群组的偏好,设计了基于偏好融合的群组推荐算法.在MovieLens数据集上的实验结果表明,本文所提的融合策略较传统的偏好融合策略有着更优越的表现,推荐准确率(nDCG)的总体平均性能约提高13%,推荐列表多样性指标的总体平均性能约提高10%.

关键词

群组推荐;推荐系统;偏好融合;协同过滤;数据挖掘;偏好预测

中图分类号 TP311

文献标志码 A

收稿日期 2019-09-07

资助项目 国家自然科学基金(61772034,61602009);安徽省自然科学基金(1808085MF172,1908085MF190);高校优秀青年人才支持计划重点项目(gxyqZD2019010);国家重点研发计划重点专项(2018YFB2101301);上海市高可信计算重点实验室开放课题(07dz22304201607)

作者简介

汪祥舜,男,硕士生,研究方向为推荐系统和隐私保护.xiangswdream@126.com

郑孝遥(通信作者),男,博士,副教授,研究方向为个性化推荐及信息安全.zxiaoyao_2000@163.com

1 安徽师范大学 计算机与信息学院,芜湖,241002

2 网络与信息安全安徽省重点实验室(安徽师范大学),芜湖,241002

3 华东师范大学 上海市高可信计算重点实验室,上海,200062

0 引言

随着社会和经济的飞速发展,尤其是近年互联网的广泛应用,人类进入了大数据时代.面对纷繁复杂的商品信息,仅依靠搜索引擎很难准确获得满足自身需求和个性化的信息资源,用户往往难以发现最需要或最适合的项目,越来越多的在线服务不可避免地受到“信息过载”问题的困扰,这使得用户难以发现他们所需要的信息,推荐系统为解决这一问题应运而生.

传统的推荐系统,如上下文感知推荐系统、移动推荐系统、实时推荐系统等,都是向单个用户提供个性化推荐,但是随着社会的发展和人们联系的紧密,人们经常以多个用户所组成的群组形式参与某项活动,例如人们会组团出去度假、进餐、看电影等,因此进行群组推荐的研究有着十分重要的意义.群组推荐已经成为近些年来推荐系统领域一个研究热点.

群组推荐系统得到了国内外研究人员和开发工作者的广泛重视.文献[1-2]结合模型融合和推荐融合的优点,提出了一种改进的偏好融合方法,提高了组推荐的整体推荐性能.Yuan等[3]提出了一个称为COM的共识模型,考虑用户在群组中的行为变化,依赖于主题的影响和内容信息,有效地提高了组推荐的准确率.Tang等[4]提出了一种基于用户交互行为的偏好融合策略,但该策略未考虑群组内已消费项目的影响因素.Sabrine等[5]用基于用户的协同过滤算法为群组成员生成推荐列表,用Borda规则来生成群组推荐,但是该方法的融合策略过于简单,导致推荐准确率相对较低.

群组推荐系统与很多领域(如数据挖掘、机器学习等)都有着紧密的联系.Chen等[6]将遗算法应用于组推荐中,用来预测群组成员之间存在的相互作用,提高了群组成员的满意度.文献[7-10]将数据挖掘的很多方法与组推荐相结合,提升了群组推荐系统在多方面的作用效益.机器学习和深度学习的很多理论和方法也被广泛应用到组推荐系统[11-14]中,Wang等[15]提出了一种用于组推荐的双向张量分解模型,采用贝叶斯个性化排序技术来学习所提出的BTF-GR模型的参数.群组内的成员之间包含丰富的行为作用信息,挖掘用户之间的相互作用规律可大幅度提高推荐质量,文献[16-17]通过用户之间的信任关系和社交信息来提高组推荐的准确率.尽管国内外研究在群组推荐领域做出了很多有益的探索,但是现有方法的偏好融合策略都相

对较简单,群组成员之间的分歧度较高,推荐准确率仍有待提高.因此本文针对群组偏好建模问题,提出一种新的共识模型策略,融合群组领袖影响因子和项目热度影响因子,并基于 K 近邻设计了一种新的基于偏好融合的群组推荐算法,实验表明,本文提出的改进算法效能较传统的策略推荐性能有着更优越的表现.

1 相关工作

1.1 群组推荐系统的形式化定义

到目前为止,群组推荐系统仍没有统一的定义形式.文献[18-19]详细阐述了群组推荐的语义和效率,并提出了一种使用共识分数的方法来定义群组推荐系统,该定义主要由共识函数和分歧度组成,具体的形式化定义如下:

1)群组预测评级.群组 G 对项目 i 的预测评级 $R_G(G,i)$ 是对每个群组成员的预测评级的融合结果.使用的融合策略不一样,得到的群组的预测评级也不一样,如均值策略的群组预测分数的计算公式为

$$R_G(G,i)=\frac{1}{|G|}\sum_{u\in G}R(u,i)\ . \tag{1}$$

2)群组分歧度.群组 G 对项目 i 的分歧,表示为 $\mathrm{dis}(G,i)$,反映了群组成员之间预测评级的分歧程度.文献[1-2]考虑了以下两种主要的计算分歧度的方法:

① 平均成对分歧:

$$\mathrm{dis}(G,i)=\frac{2}{|G|(|G|-1)}\sum(|R(u,i)-R(v,i)|)^2, \tag{2}$$

其中 $u\neq v$,且 $u,v\in G$.

② 分歧差异:

$$\mathrm{dis}(G,i)=\frac{1}{|G|}\sum_{u\in G}(R(u,i)-R_m(G,i))^2, \tag{3}$$

其中 $R_m(G,i)$ 是群组内所有成员对项目 i 的预测评级的平均值.

3) 共识函数.共识函数 $F(G,i)$ 通过群组预测评级和群组分歧度聚合计算得到:

$$F(G,i)=w_1\times R_G(G,i)+w_2\times(1-\mathrm{dis}(G,i))\ , \tag{4}$$

其中 $w_1+w_2=1$,w_1 和 w_2 分别代表了预测评分和群组分歧度在总体推荐评分中所占的重要程度.该加权模型的共识函数比较简单,不能有效对群组推荐做出精准的预测,适应性较差.

1.2 群组推荐偏好融合策略

群组推荐系统有多种不同的偏好融合策略,文献[20]对常见的融合策略做了详细阐述.到目前为止,在群组推荐系统的研究中,常用的策略有最小痛苦策略、均值策略和最开心策略.具体定义如下:

1)最小痛苦策略(Least Misery Strategy,LM):以群组成员最低的个人评级作为该群组的评级.设 $R_{G_i,j}$ 表示群组 G_i 对项目 j 的评分,$R_{U_l,j}$ 为群组中成员 u_l 对项目 j 的评分,则可用下述式(5)表示:

$$R_{G_i,j}=\min_{l\in G}\{R_{U_l,j}\}. \tag{5}$$

2)均值策略(Average Strategy,AVG):以群组成员的评级的平均值作为该群组的评级.若用 avg 表示求平均值运算,则可用下述式(6)表示:

$$R_{G_i,j}=\operatorname{avg}_{l\in G}\{R_{U_l,j}\}. \tag{6}$$

3)最开心策略(Most Pleasure Strategy,MP):以群组成员最高的个人评级作为该群组的评级.若用 max 表示求最大值运算,则可用下述式(7)表示:

$$R_{G_i,j}=\max_{l\in G}\{R_{U_l,j}\}. \tag{7}$$

2 问题定义和模型构建

2.1 相关定义

为了更好的描述本文对群组推荐系统所做的研究,将本文所做的研究工作定义如下:

定义 1 群组领袖:一个群组由多个用户组成,群组内的成员各有差异,消费的项目及其数量各不相同,对推荐的决策产生的影响也有所不同.因此,本文根据群组成员的作用大小为其分配不同的权重,将群组内对推荐决策起到关键作用的用户定义为该组的群组领袖.群组领袖可以是一个,也可以是多个,群组领袖的排名按照影响大小逐步递减的次序依次确定.群组领袖往往是该群组内活跃度最大、消费项目数量最多的用户.

定义 2 群组领袖影响因子:为了精确衡量群组内成员作用程度的大小,本文定义群组领袖影响因子为 α_i.设用户 u_i 消费的项目总数记为 N_i,项目种类数为 C_i,群组 G 消费的项目总数和种类数分别为 N_G 和 C_G,则 α_i 定义如下:

$$\alpha_i=\frac{2\dfrac{N_i}{N_G}\dfrac{C_i}{C_G}}{\dfrac{N_i}{N_G}+\dfrac{C_i}{C_G}}=\frac{2N_iC_i}{N_iC_G+N_GC_i}. \tag{8}$$

本文将 f_{leader} 定义为群组影响影响因子 α_i 组成的集合,表示为 $f_{\mathrm{leader}}=\{\alpha_1,\alpha_2,\alpha_3,\cdots,\alpha_k\}$,其中 k 表

示群组 G 中群组领袖的个数,本文按照群组领袖影响因子值的大小将其降序排列.群组领袖的个数可以根据需要合理选择,若群组内只选择一个群组领袖,则 $f_{\text{leader}}=\alpha_1$,此时 α_1 表示该群组内影响力最大的领袖因子,依次类推.

定义 3 热度项目:推荐系统更倾向于向用户推荐消费次数多的项目,项目被消费的次数越多,在一定程度上表明其受欢迎的程度越高.本文定义消费数量排名靠前的 k 个项目为热度项目.

定义 4 项目热度影响因子:为了量化项目热度的高低,本文定义项目热度影响因子为 β_j,设 m 为消费该项目的用户个数,M 为群组 G 内的用户总数,则 β_j 定义如下:

$$\beta_j=\frac{\log_2(1+m)}{\log_2(1+M)}. \tag{9}$$

本文将 f_{hot} 定义为群组影响影响因子 β_j 组成的集合,表示为 $f_{\text{hot}}=\{\beta_1,\beta_2,\beta_3,\cdots,\beta_s\}$,其中 s 表示项目热度影响因子的个数,并且其大小依次递减.与群组领袖影响因子类似,项目热度影响因子的个数也可根据实际情况合理选取.

定义 5 邻居群组:群组之间也具有相似性,例如学生群组、球迷群组、旅游群组等,群组之间彼此都有着紧密的联系或相同的兴趣爱好等.相似的群组在很大程度上会有着相同的偏好,可以根据相似群组的偏好为目标群组提供推荐参考,本文将这样相似的群组定义为邻居群组.邻居群组的寻找,可首先计算该群组与目标群组之间的相似度,根据群组之间相似度的大小为其合理归类到相应的目标群组.一个目标群组可找到多个邻居群组,这些邻居群组可构成一个邻居群组集合.

2.2 共识模型

为进一步提高群组推荐的准确率,本文融合了群组领袖影响因子和项目热度影响因子,为目标群组寻找近邻组,并借鉴邻居群组的偏好.本文用 $F(G_i,I_j)$ 表示群组 G_i 对项目 j 的预测评分,α_i 表示群组领袖影响因子,β_j 表示项目热度影响因子,I 表示群组的项目集,j 表示需要预测的项目,$R_{\text{CF}}(u_i,I_j)$ 表示单个用户之间协同过滤产生的预测评级,$R_{\text{NG}}(G_i,I_j)$ 表示基于群组间的 K 近邻推荐产生的预测评级,则本文的共识模型定义如下:

$$F(G_i,I_j)=a\left\{\frac{1}{2n+1}\left[\sum_{i=1}^{n}\alpha_i x_i+\sum_{j=1}^{n}\beta_j y_j+R_{CF}(u_i,I_j)\right]\right\}+bR_{\text{NG}}(G_i,I_j), \tag{10}$$

其中 $a+b=1$,$a,b\in[0,1]$;x_i 表示该领袖已经消费项目的评分的平均值,表示为 $x_i=\frac{1}{k}\sum_{i=1}^{k}R_{U_i,j}$,其中 $R_{Ui,j}$ 表示领袖用户 U_i 对项目 j 的评级;y_j 表示该热度项目获得评分的平均值,表示为 $y_j=\frac{1}{s}\sum_{j=1}^{s}R_{Ij}$,其中 R_{Ij} 表示热度项目 j 获得的评级.

3 算法设计与分析

本文的算法在训练集上对共识模型进行训练,在测试集上验证推荐结果,本文算法的主要思想流程如图 1 所示.

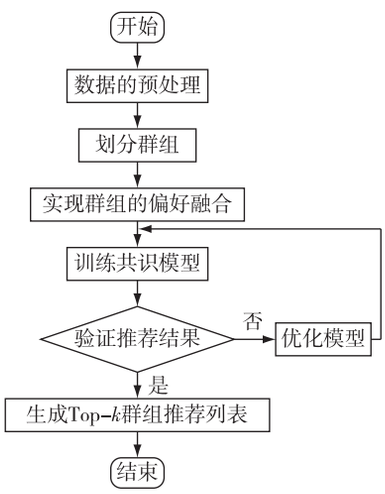


图 1 算法思想流程

Fig. 1 Algorithm flow

3.1 划分群组

为了确定两个用户之间的关联程度,可以通过计算两个用户之间相似度的方法度量两个用户之间的亲密程度.设 u 和 v 表示任意两个用户,$\boldsymbol{u}$ 和 $\boldsymbol{v}$ 分别为由用户 u 和用户 v 评分信息所组成的评分向量,用 $s_{\cos}(u,v)$ 表示两个用户之间的余弦相似度,则计算式为

$$s_{\cos}(u,v)=\frac{\boldsymbol{uv}}{|\boldsymbol{u}||\boldsymbol{v}|}. \tag{11}$$

为合理划分群组,本文首先通过计算任意 2 个用户之间的余弦相似度,将相似度高的用户构造成一个群组,并且总是优先选择有相同或相似偏好的用户加入该群组,以减少群组成员之间的分歧度.

3.2 基于偏好融合的 K 近邻组推荐

基于偏好融合的 K 近邻(K Nearest Nighborhood,KNN)组推荐,其基本思想是根据邻居

群组的偏好为目标群组提供推荐,如图 2 所示.设 Group_m 为需要推荐的目标群组,Group_n 和 Group_s 是和目标群组相似的 2 个群组,可借鉴近邻组的偏好,为目标群组生成 Top-k 推荐列表.

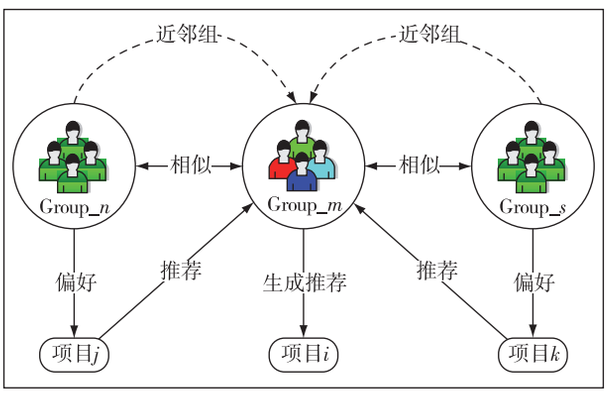


图 2　基于偏好融合的 K 近邻组推荐示意图

Fig. 2　Schematic diagram of K-Nearest Neighbor group recommendation based on preference aggregation

基于偏好融合的 K 近邻组推荐,首先计算任意 2 个群组之间的相似度,然后挑选合适的邻居,最后完成推荐过程.挑选合理的邻居集合规模是基于 K 近邻推荐最重要的一步.

3.2.1　领域的选择

领域数量和选择领域的规则会对推荐结果产生重要的影响,传统基于 K 近邻的推荐可用如图 3 左图所示.在选定目标群组以后,首先选定 K 个邻居,然后根据欧式距离或者计算 2 个群组之间的相似度来挑选邻居,然而由于 K 的值是人为事先选定,如果挑选的不合理,加入了关联程度较差的群组,这样反而会大大降低推荐的精度.例如图 3 左图所示,若事先选定 6 个邻居,从图中可以看出,群组 G8 和目标群组 G1 的距离较远,并不适合挑选出来作为目标群组 G1 的邻居,如果把 G8 挑选出来当做邻居反而降低了推荐的准确度,此时的推荐质量不如挑选 G2、G3、G5、G7、G9 这 5 个邻居更合适.

与选择固定数量的邻居数目的做法不同,本文首先设定一个阈值过滤,保留相似度权重数大于设定阈值的近邻,在算法执行时,过滤掉相似度低的群组,如下方右图所示,阈值的过滤效果相当于在目标群组 G1 的周边画了一个圆,可以合理舍弃掉周边相似度低的离散的点,这样便可以去掉 G8 这个偏离中心的点.

阈值的合理选取会影响预测的精度,如果阈值

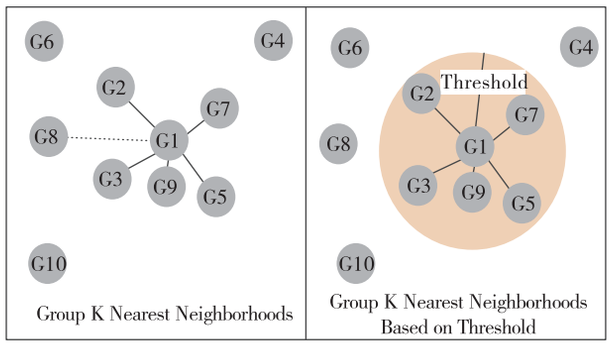


图 3　K 近邻组和阈值 K 近邻组的邻域选择

Fig. 3　Neighborhood selection of group KNN and group KNN based on threshold

设置过高,会降低可供选择的近邻的数量,这会降低很多群组的覆盖率;相反,如果阈值设置太低,邻居数量会很多,如果邻居群组中有很多关联度低的群组,也会对推荐结果产生消极的影响.当所需要的邻居数量较少的时候,2 种方法产生的结果几乎一致,但当可供挑的邻居数量较多时,第二种方案对推荐结果更能起到更好的作用.

3.2.2　算法实现

基于偏好融合的 K 近邻组推荐算法,首先需要确定相似的邻居群组集,本文首先采用 Pearson 相关系数来计算 2 个群组之间的相似度,用 $s_{\text{Pearson}}(G,G')$ 表示群组 G 和群组 G' 之间的 Pearson 相似度,计算式如下:

$$s_{\text{Pearson}}(G,G')=\frac{\sum_{i\in I}(r_{G,i}-\overline{r_G})(r_{G',i}-\overline{r_{G'}})}{\sqrt{\sum_{i\in I}(r_{G,i}-\overline{r_G})^2}\sqrt{\sum_{i\in I}(r_{G',i}-\overline{r_{G'}})^2}},\tag{12}$$

其中,I 表示群组的项目集,i 表示项目集 I 的某一个具体的项目,$\overline{r_G}$ 表示群组 G 中的项目集 I 评分的平均值,$\overline{r_{G'}}$ 表示群组 G' 中的项目集 I 评分的平均值.

在确定好近邻组的集合以后,对于 N 个最相似的邻居群组,设 $R_{G,i}$ 值表示群组 G 后对项目 i 的预测评分,则目标群组 G 对项目 i 的预测评分可用下式表示:

$$R_{G,i}=\overline{r_G}+\frac{\sum_{G'\in N}s(G,G')(r_{G',i}-\overline{r_{G'}})}{\sum_{G'\in N}|s(G,G')|}.\tag{13}$$

具体的算法描述如下所示.

算法 1　基于偏好融合的 K 近邻群组推荐算法

输入:需要预测的群组编号 id 和项目编号 id
输出:该群组对该项目的预测评分
1.divide Groups
2.calculate GroupJaccardSim
3.**while**(Current Nearest Neighbor Groups! =required Neighbor Groups) **do**
4.**for each** k **in** GroupNum//k 表示计数器
5.　　**if** (k! =G&&GroupJaccardSim>max) **then**
//G 为群组 id,max 用来临时保存 GroupJaccardSim 所得的最大值
6.　　**for each** j< Current NearestNeighborGroups
7.　　**if**(k==flaG[j]) **then**
//flaG[ ]表示已存在的与群组 G 相似度最大的群组 id 的数组
8.　　Break;
9.　　**end if**
10.　　if(j==Current Nearest Neighbor Groups)
11.　　max←GroupJaccardSim;
12.　　flaG[$i2$]←k
13.　　**end if**
14.　　**end for**
15.　　**end if**
16.　**end for**
17.　utill find required K Neighbor Groups
18.**end while**
19.caculate GroupPearsonSim
20.return G on Item i Predicted Rating Based on KNN Groups

4　实验结果与分析

4.1　数据集和实验设置

本文实验采用的是 MoviesLens 100K 数据集,MovieLens 数据集是由明尼苏达大学的 GroupLens 研究项目收集,该数据集包括 943 名用户对 1 682 部电影的 100 000 点评分,评分范围是 1~5 分,评分越大表示用户越喜欢该电影,其中每位用户至少点评了 20 部电影.该数据集被广泛应用于推荐系统相关的实验或科研项目中.为了有效对推荐结果和算法进行衡量,本文将 MoviesLens 数据集随机选择 80%作为训练集,剩余 20%作为测试集,算法在训练集中进行测试,然后将推荐结果与测试集中的数据进行验证,对比分析论证实验结果.

4.2　评价指标

1)归一化折损累计增益

归一化折损累计增益指标(normalized Discounted Cumulative Gain,nDCG)是基于排序的评价方法,是衡量组推荐准确率的一种重要指标.本文使用 nDCG 来作为推荐准确率的评价标准.由于群组推荐列表是由群组内成员若干偏好得到,要反应该群组成员对推荐结果的满意程度,可对比验证测试集中的数据,度量预测的排序与实际排序的准确程度.

设 $p_1,p_2,\cdots,p_k$ 为算法产生的有序推荐结果列表,u 表示某一个用户,r_{upi} 表示用户 u 对项目 p_i 的真实评分,i 为列表中的位置.DCG(Discounted Cumulative Gain)的计算公式为

$$G^u_{\mathrm{DC},k}=r_{up1}+\sum_{i=2}^{k}\frac{r_{upi}}{\log_2(i)},\tag{14}$$

nDCG 的计算公式为

$$G^u_{\mathrm{nDC},k}=\frac{G^u_{\mathrm{DC},k}}{G^u_{\mathrm{IDC},k}},\tag{15}$$

其中,$G^u_{\mathrm{IDC},k}$表示用户 u 的最大 DCG 值,可根据用户真实评分高低,将 k 个项目按照最理想情况进行排序计算出 IDCG 的值.nDCG 等于真实 DCG 值与 IDCG 的商,nDCG 的数值越大,表明推荐准确率越高.本文的 nDCG 是群组成员的平均 nDCG 值.

2)多样性

推荐项目的多样性是衡量组推荐系统的推荐性能高低的一个重要评价指标,本文用 Divers 衡量推荐项目的多样性.首先采用 Jaccard 相关系数来计算两个项目之间的相似度,然后按照公式(17)计算推荐列表项目的多样性.

计算两个项目之间的 Jaccard 相似度用下式表示:

$$s_{\mathrm{Jac}}(I_i,I_j)=\frac{I_i\cap I_j}{I_i\cup I_j}.\tag{16}$$

推荐列表多样性 Divers 可用下式计算:

$$I_{\mathrm{Divers}}=1-\frac{2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}s_{\mathrm{Jac}}(I_i,I_j)}{k(k-1)},\tag{17}$$

其中 $s_{\mathrm{Jac}}(I_i,I_j)$ 表示项目 i 和项目 j 之间的 Jaccard 相关系数,k 表示推荐列表的项目的个数.Divers 的数值越大,表明推荐结果的多样性越高.

3)均方根误差

均方根误差被广泛应用于推荐系统的实验评估指标,可衡量预测结果的精确度,本文所用的群组推荐的均方根误差公式 RMSE 如下:

$$I_{\mathrm{RMSE}}=\sqrt{\frac{\sum_{i=1}^{T}(R_{G,i}-\hat{R}_{G,i})^2}{|T|}}.\tag{18}$$

设 $R_{G,i}$ 表示群组 G 对项目 i 的预测评分,$\hat{R}_{G,i}$ 表

示群组 G 对项目 i 的真实评分,T 表示群组 G 中的参与预测的项目个数,I_{RMSE} 的值越小表明推荐的准确度越高.

4.3 敏感性分析

4.3.1 参数权重对推荐性能的影响

为达到最佳的推荐效果,选取最佳的参数,通过实验对参数 a 和 b 进行灵敏度分析.从图 4 中可以看出,当 $a=0.6$,$b=0.4$ 时,RMSE 取得最小值,此时推荐准确度最好,这也从侧面说明在推荐过程中起决定性因素的是群组内的用户偏好.

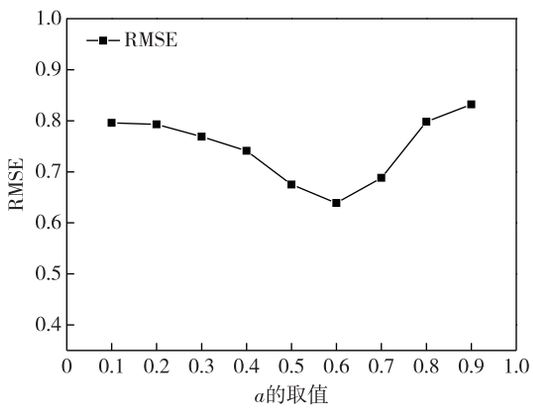


图 4 参数对群组推荐性能的影响

Fig. 4 Effect of parameters on group recommendation performance

4.3.2 近邻组个数对推荐精度的影响

本文在 MoiveLens 数据集上,改变近邻组的个数,采用本文所提的共识模型,完成对预测精度的测试实验,实验结果如图 5 所示.

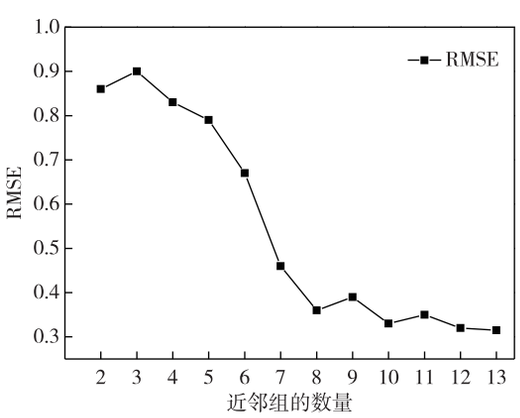


图 5 近邻组的数量对群组推荐精度的影响

Fig. 5 Influence of the number of neighbor groups on the accuracy of group recommendation

从实验结果可以看出:预测的准确度受近邻组数量的影响,RMSE 随着近邻组的数量的增加总体呈下降的趋势,当近邻组达到一定的数量的时候,RMSE 的值开始上下波动、维持动态平衡.实验结果表明预测的精度随着近邻组的个数的增加总体呈上升的趋势;当近邻组的个数比较少的时候,预测的准确度不是很高;当近邻组达到一定数量的时候,我们会发现预测的精确度总体趋势几乎保持不变.我们更换不同的群组,多次进行重复实验,发现当近邻组的个数在 9 左右的时候,群组推荐的预测准确度受近邻组的个数敏感较小,预测的准确度可稳定在一定的范围内.实验发现并不是近邻组越多,预测的准确度就越高,而是到达一定数量的时候,预测的准确度基本维持不变.

4.4 实验结果

在本节中我们改变群组规模来验证本文所提共识模型的推荐准确率和推荐列表多样性.

1)推荐准确率

实验结果如下图所示,其中横轴表示群组规模,群组规模为 5 表示该群组成员的个数为 5,依次类推;纵轴表示 nDCG.GROUP_CF 表示群组间协同过滤算法策略;PROPOSED 表示本文所提的共识模型策略.从图 6 中可以看出,在相同的群组规模下,本文所提的模型策略在推荐准确率上有所提高;当群组规模为 15 和 25 时,本文所提共识模型有更着较为优良的表现;实验也表明在其他的分组规模下,本文所提的融合模型策略也明显优于另外常见的 4 种策略,在推荐准确率上总体的平均性能约提高 13%.

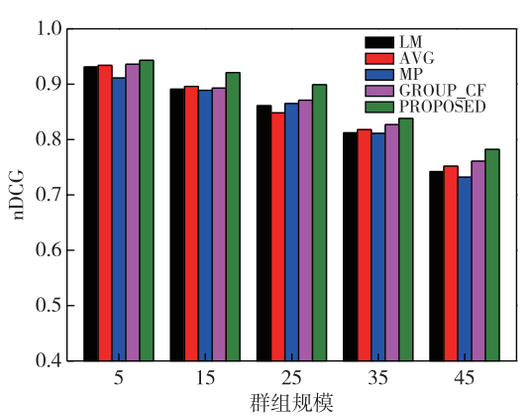


图 6 推荐准确率比较

Fig. 6 Comparison of recommendation accuracy

2)推荐项目的多样性

实验结果如图 7 所示,其中纵轴 Divers 表示群组推荐项目的多样性.实验结果表明,当群组成员的个数

较少的时候,推荐列表的多样性相对较低;随着群组成员个数的增加,群组推荐列表的多样性有一定的增加.同时实验结果也验证了本文所提的共识模型的多样性优于前4种融合策略,本文所提的共识模型在多样性指标上总体平均性能上约提高10%.

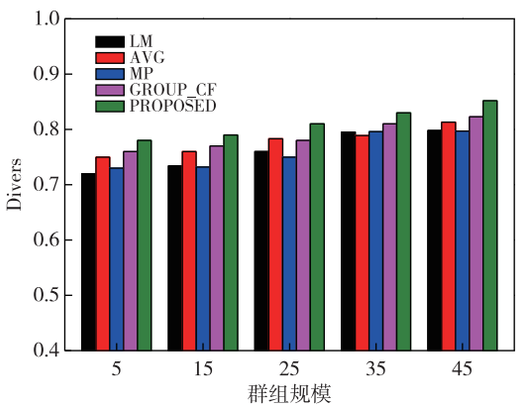


图7 推荐项目多样性比较

Fig. 7 Comparison of diversity of recommended items

5 总结和展望

现有的大多数推荐系统主要是针对单个用户进行推荐,随着人们联系的紧密和社会的发展,群组推荐已成为研究热点.如何为群组提供精准的推荐仍然是一个开放性的问题,也是一个挑战.群组推荐系统研究的难点在于是找出一个合理的偏好融合策略,尽可能的降低群组成员的之间的分歧度,使推荐结果能够满足群组内大多数成员的偏好.本文提出了一种新的共识模型,提出了群组领袖影响因子和项目热度影响因子等概念,融合群组内和群组间的多种因素,提高了群组推荐的准确率;并设计了基于偏好融合的K近邻群组推荐算法,为目标群组寻找邻居群组,并借鉴邻居群组的偏好,在MoviesLens数据集上也验证了本文所提的共识模型的优越性,相较几种常见的偏好融合策略,本文所提的共识模型策略在推荐准确率nDCG上总体的平均性能约提高13%,在多样性上总体的平均性能约提高10%.接下来的工作拟打算从群组成员之间的信任关系和社交关系等展开进一步的研究.

参考文献

References

[1] 张玉洁,杜雨露,孟祥武.组推荐系统及其应用研究[J].计算机学报,2016,39(4):745-764
ZHANG Yujie, DU Yulu, MENG Xiangwu. Research on group recommender systems and their applications[J]. Chinese Journal of Computers,2016,39(4):745-764

[2] 胡川,孟祥武,张玉洁,等.一种改进的偏好融合组推荐方法[J].软件学报,2018,29(10):3164-3183
HU Chuan, MENG Xiangwu, ZHANG Yujie, et al. Enhanced group recommendation method based on preference aggregation[J]. Journal of Software, 2018, 29(10):3164-3183

[3] Yuan Q, Cong G, Lin C Y. COM: a generative model for group recommendation[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014:163-172

[4] Tang F, Liu K, Feng L, et al. Research on the integration strategy of group recommendation based on user's interactive behaviors[C]// 2016 IEEE International Conference on Cloud Computing and Big Data Analysis, 2016: 367-372

[5] Abdrabbah S B, Ayadi M, Ayachi R, et al. Aggregating top-K lists in group recommendation using Borda rule [M]// Advances in Artificial Intelligence: From Theory to Practice. Cham: Springer International Publishing, 2017:325-334

[6] Chen Y L, Cheng L C, Chuang C N. A group recommendation system with consideration of interactions among group members[J]. Expert Systems with Applications, 2008,34(3):2082-2090

[7] Guo Z W, Tang C W, Tang H, et al. A novel group recommendation mechanism from the perspective of preference distribution[J]. IEEE Access, 2018, 6:5865-5878

[8] Seo Y D, Kim Y G, Lee E, et al. An enhanced aggregation method considering deviations for a group recommendation[J]. Expert Systems with Applications, 2018,93:299-312

[9] Feng S S, Cao J, Wang J, et al. Group recommendations based on comprehensive latent relationship discovery[C]// 2016 IEEE International Conference on Web Services, 2016:9-16

[10] Boratto L, Carta S, Fenu G, et al. Influence of rating prediction on group recommendation's accuracy[J]. IEEE Intelligent Systems, 2016,31(6):22-27

[11] Guo Z W, Zeng W R, Wang H, et al. An enhanced group recommender system by exploiting preference relation [J]. IEEE Access, 2019, 7:24852-24864

[12] Yin H, Wang Q, Zheng K, et al. Social influence-based group representation learning for group recommendation [C]// 2019 IEEE 35th International Conference on Data Engineering, 2019:566-577

[13] Du Y L, Meng X W, Zhang Y J. CVTM: a content-venue-aware topic model for group event recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2019. DOI:10.1109/TKDE.2019.2904066

[14] Zhang W, Wang J, Feng W. Combining latent factor model with location features for event-based group recommendation[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013:910-918

[15] Wang J K, Jiang Y C, Sun J S, et al. Group recommendation based on a bidirectional tensor factorization model[J]. World Wide Web, 2018, 21(4):961-984

[16] Guo C, Tian X M. Flickr group recommendation using content interest and social information[C]//Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics, 2014:405-410
[17] Santos E B, Manzato M G, Goularte R. A conceptual architecture with trust consensus to enhance group recommendations[C]//2014 IEEE/ACIS 13th International Conference on Computer and Information Science, 2014: 137-142
[18] Amer-Yahia S, Roy S B, Chawlat A, et al. Group recommendation: semantics and efficiency[J]. Proceedings of the VLDB Endowment, 2009, 2(1):754-765
[19] Roy S B, Amer-Yahia S, Chawla A, et al. Space efficiency in group recommendation[J]. The VLDB Journal, 2010, 19(6):877-900
[20] Masthoff J. Group modeling: selecting a sequence of television items to suit a group of viewers[J]. User Modeling and User-Adapted Interaction, 2004, 14(1):37-85

Research on group recommendation based on preference aggregation

WANG Xiangshun[1,2] ZHENG Xiaoyao[1,2] ZHU Deyi[1,2] ZHANG Yue[3] SUN Liping[1,2]
1 School of Computer and Information, Anhui Normal University, Wuhu 241002
2 Anhui Provincial Key Laboratory of Network and Information Security, Anhui Normal University, Wuhu 241002
3 Shanghai Key Lab for Trustworthy Computing, East China Normal University, Shanghai 200062

Abstract The traditional recommendation system is mainly for a single user, but people are increasingly participating in activities in the form of multiple users. Group recommendation is intended to serve groups of multiple users and has already become one of the hotspots of current research. Aiming at the low accuracy of current group recommendation and the difficulty of integrating preference conflicts among group members, this paper proposes a new consensus model strategy, which combines group leader impact factor and project heat impact factor, and is based on K Nearest Neighbor recommendation. A group recommendation algorithm based on preference aggregation is designed to find the neighbor groups for the target group and draw on the preference of the neighbor groups. The experimental results on the MovieLens dataset show that the aggregation strategy proposed in this paper has better performance than traditional preference aggregation strategy. The overall average performance of the recommended accuracy (measured by normalized Discounted Cumulative Gain, nDCG) is increased by about 13%, and the overall average performance of the recommendation list diversity is improved by about 10 %.

Key words group recommendation; recommendation; preference aggregation; collaborative filtering; data mining; preference prediction