计算机科学
COMPUTER SCIENCE

基于Apriori算法填充数据及改进相似度的推荐算法

董云薪，林耿，张清伟，陈颖婷

引用本文

董云薪，林耿，张清伟，陈颖婷.基于Apriori算法填充数据及改进相似度的推荐算法[J]．计算机科学，2022，49(11A)：211000005-5.

DONG Yun-xin，LIN Geng，ZHANG Qing-wei，CHEN Ying-ting. Recommendation Algorithm Based on Apriori Algorithm and Improved Similarity [J]. Computer Science，2022，49(11A)：211000005-5.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.211000005

基于 Apriori 算法填充数据及改进相似度的推荐算法

董云薪[1]　林　耿[2]　张清伟[1]　陈颖婷[1]
1 福建农林大学计算机与信息学院　福州 350028
2 闽江学院数学与数据科学学院　福州 350108
(765150341@qq.com)

摘　要　针对协同过滤算法中存在的数据稀疏和算法精确度不高的问题,提出了一种融合关联规则的协同过滤算法。首先,利用关联规则 Apriori 算法挖掘出用户间潜在的联系,该潜在联系采用用户间的关联规则的置信度来表示,紧接着进一步构建用户置信度矩阵,用于填充用户评分矩阵。其次,利用置信度矩阵来改进传统的相似度计算公式,构建一个用户间的综合相似度计算公式。最后,利用填充过后的用户评分矩阵和用户间的综合相似度为用户进行推荐。所提算法相比传统算法具有更高的算法精度。此外,与其他算法相比,所提算法还能有效缓解推荐系统的长尾问题,从而进一步提高推荐系统的推荐质量。

关键词:协同过滤;关联规则;推荐算法;数据稀疏;相似度改进
中图法分类号　TP301

Recommendation Algorithm Based on Apriori Algorithm and Improved Similarity

DONG Yun-xin[1],LIN Geng[2],ZHANG Qing-wei[1] and CHEN Ying-ting[1]
1 School of Computer and Information,Fujian Agriculture and Forestry University,Fuzhou 350028,China
2 School of Mathematics and Data Science,Minjiang University,Fuzhou 350108,China

Abstract　In order to alleviate the data sparse problem and improve the accuracy of collaborative filtering algorithm,a recommendation algorithm based on Apriori algorithm and improved similarity is presented. Firstly,it uses Apriori algorithm to mine the potential connections between users,and uses the confidence of the association rules between users to represent the potential connections between users,then constructs a user confidence matrix to fill the user rating matrix. Secondly,the algorithm uses the confidence matrix to improve the traditional similarity calculation formula and build a comprehensive similarity calculation formula between users. Finally,the algorithm uses the filled user rating matrix and the comprehensive similarity between users to make recommendations for users. The proposed algorithm has higher algorithm accuracy than traditional algorithms. Compared with other algorithms,the proposed algorithm can effectively alleviate the long tail problem of the recommendation system,so as to further improve the recommendation quality of the recommendation system.

Keywords　Collaborative filtering,Association rules,Recommendation algorithm,Data sparse,Similarity improvement

1　引言

随着信息技术与互联网的快速发展,人们在生活中所接收到的信息量与日俱增,导致了“信息过载”问题的出现。“信息过载”指人们难以从海量的信息库中找到自己真正所需的有效信息[1]。推荐系统可以帮助用户在海量的信息中提炼出有效信息。推荐系统作为解决大数据时代存在的“信息过载”问题和满足用户购买商品的一种个性化需求手段,受到越来越多学者的关注和研究。推荐系统的核心是推荐算法,协同过滤算法是目前最为流行的推荐算法之一。协同过滤算法主要有基于用户的协同过滤算法(User-CF)和基于项目的协同过滤算法(Item-CF)[2]。

传统的协同过滤算法中主要存在着数据稀疏和算法精确度不高这两大问题[3-4]。针对这两大问题,学者们提出了许多方法用以缓解数据稀疏问题和提高算法精度。针对数据稀疏问题,学者们提出通过挖掘用户评论、项目属性、项目标签、用户行为等间接数据中的隐藏评分来计算项目预测评分,填充稀疏的用户评分矩阵[5-8]。Hong 等[5]利用 TF-IDF 算法提取项目标签的词向量,并计算出其在不同项目中的评分,填充用户评分矩阵。Chen 等[6]提出一种标签扩展的协同过滤算法,通过用户行为及标签语义计算标签相似度,用标签相似度来拓展标签向量和填充项目评分矩阵。Guo 等[7]利用显式用户信息(用户项目评分)和隐式用户信息(用户信任行为)进行用户项目评分矩阵的填充。Liu[8]提出一种基于电子学习群体行为影响集的协同过滤算法,电子学习是通过用户在互联网使用中产生的行为,从中学习用户的个人行为及兴趣,通过学习到的这些用户行为及兴趣对用户评分矩阵进行填充。

针对算法精确度不高的问题,主要采用改进相似度计算方法来提高推荐算法的精确度。学者们提出引入数学模型以及挖掘数据中用户的隐藏信息,来改进传统的相似度计算

基金项目:福建省自然科学基金(2020J01843)
This work was supported by the Natural Science Foundation of Fujian Province,China(2020J01843).
通信作者:林耿(lingeng413@163.com)

公式，达到提高算法精确度的目的[9-13]。Zhang 等[9]提出概率矩阵来预测用户对未评分项目的偏好值，并结合时间函数构建用户-项目-时间的三维矩阵模型以改进用户相似度计算。Wang 等[10]提出把用户兴趣融入相似度计算中，将从用户兴趣中提取出来的用户项目评分倾向的离散系数融入用户间的相似度计算。Han 等[11]提出用户的人口统计学信息(年龄、性别、职业和邮编)以及项目的种类信息，结合用户的项目评分，来改进用户间的相似度计算。Wu 等[12]提出采用用户属性、用户项目评分数、类别属性、用户评分时间，结合用户评分得到用户的显隐兴趣相似度，以改进用户相似度计算。Li 等[13]提出将用户共同评分项目数与用户评分平均数作为修正因子，融入到用户相似度改进计算中。

以上文献针对协同过滤算法中存在的数据稀疏及算法精确度不高的问题进行了多方面的研究。学者们大多从挖掘用户评论、标签等文本类信息中所蕴涵的隐藏信息和改进传统相似度计算着手研究，其中对于用户间的潜在联系挖掘还不够充分，忽略了用户间的潜在联系对用户间相似度计算以及项目预测评分的影响。

针对以上问题，本文提出了一种融合关联规则的协同过滤算法(AR-CF)。本文主要通过研究用户观影行为对推荐结果的影响，通过对用户间的共同观影行为进行关联规则挖掘，得到用户间的关联规则置信度矩阵，将置信度矩阵融入预测项目评分计算及用户间相似度的改进计算，从而提高推荐算法的推荐质量。

2 融合关联规则的协同过滤算法

2.1 关联规则

2.1.1 关联规则基本概念

设含有 n 个项的集合为 $I=\{i_1,i_2\cdots i_n\}$。设与事务相关的数据 D 是数据库事务的集合，D 称为事务集。其中每个事务 Q 是一个非空项集，使得 $Q\subseteq I$。设 Z 是一个项集，事务 Q 包含项集 Z，当且仅当 $Z\subseteq Q$。

关联规则如 $R:X\rightarrow Y$ 这般的蕴含式，其中 $X\subseteq I,Y\subseteq I$，$X\neq\varnothing,Y\neq\varnothing$，并且 $X\cap Y\neq\varnothing$。关联规则 R 在事务集 D 中成立，具有支持度 S，其支持度 S 是事务集 D 中集合 X 和 Y 的百分比。规则 R 在事务集 D 中具有置信度 C，其置信度 C 表示同时包含 X 和 Y 的事务在事务集 D 中所占的百分比，关联规则 R 的置信度公式如下：

$$C(X\rightarrow Y)=P(X|Y)=\frac{S(X\cup Y)}{S(X)} \tag{1}$$

其中 $S(X\cup Y)$ 表示 X 事务与 Y 事务在事务集 D 中共同发生的百分比，$S(X)$ 表示 X 事务在事务集 D 中发生的百分比。

2.1.2 关联规则挖掘基本原理

(1)找出所有的频繁项集。根据要求，设定最小支持度 S，当项集在事务集中所占百分比满足最小支持度时，则称该项集为频繁项集。要找出强关联规则，首先要找出满足最小支持度 S 的频繁项集。

(2)由频繁项集产生强关联规则。当找出频繁项集后，设定最小置信度 C，当频繁项集里的关联规则满足最小置信度时，则称该关联规则为强关联规则。

2.1.3 关联规则算法

对于关联规则的挖掘，目前有许多相关算法，其中应用最为广泛与经典的是 Apriori 算法。Apriori 算法是第一个用于挖掘数据集中关联规则的算法[14]。

Apriori 算法主要步骤如下：

设共有 K 个项集，每个项集 I 均不为空集，项集 $I=\{i_1,i_2,\cdots,i_n\}$。

(1)计算出数据集中每个最小项集 1-项集$\{i_1\}$，$\{i_2\}$，…，$\{i_n\}$的支持度，找出满足最小支持度 S 的 1-项集，生成一维最大项目集。

(2)根据第一步生成的一维最大项目集，产生 2-项集$\{i_a,i_b\}$，$a\neq b$，找出满足最小支持度的 2-项集，生成二维最大项目集。

(3)重复以上步骤，在第 L 步中，根据第 $L-1$ 步生成的 $(L-1)$维最大项目集，产生 L 维侯选项目集，找出满足最小支持度 S 的 L-项集，生成 L 维最大项目集。

(4)找出频繁项集后，设定最小置信度 C，计算各频繁项集中关联规则的置信度，保留满足最小置信度的关联规则，称为强关联规则。

2.2 数据填充

目前推荐系统中存在数据稀疏这一主要问题。数据稀疏是用户评分矩阵中未评分数据占比多而导致的。在传统协同过滤算法的计算过程中，有海量的用户和项目信息，但用户对项目未评分数目的增多，导致用户项目评分矩阵高度稀疏，任意两个向量之间是近似正交的，计算用户间的相似度通常情况下为零，导致推荐结果不理想。数据集的稀疏性采用稀疏度作为参考指标。稀疏度指数据集中用户未评分项目占用户评分总项目的百分比。稀疏度越大，对传统协同过滤算法的影响越大，进而导致算法精确度下降，推荐结果不够理想。

为了缓解数据稀疏性，本文提出使用关联规则算法挖掘用户间的潜在联系，通过 Apriori 算法挖掘出用户间的关联规则，将关联规则的置信度用于填充用户评分矩阵，达到缓解数据稀疏的目的，提高算法的精确度。

首先，对电影数据集进行处理，处理为以下形式：

$$I_k:\{u_n\cdots u_m\},n\neq m,n<m \tag{2}$$

其中，I_k 表示第 K 部电影，u_n 表示第 n 名用户，用户集合中的用户是对第 k 部电影进行过评分的用户。

其次，使用 Apriori 算法挖掘用户间的关联规则。先将电影数据集处理成式(2)，再利用 Apriori 算法挖掘用户间满足最小支持度的 2-项集$\{u_i,u_j\}$，$i\neq j$，再通过式(1)计算出用户间的置信度，接着构建置信度矩阵，置信度矩阵 $\boldsymbol{Con}$ 具体如下：

$$\boldsymbol{Con}=\begin{pmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & c_{mi} & \vdots \\ c_{m1} & \cdots & c_{mm} \end{pmatrix} \tag{3}$$

其中 c_{mi} 表示用户 m 和用户 n 之间的置信度。

然后，再对电影数据集进行处理，得到原始用户评分矩阵，用户评分矩阵 $\boldsymbol{R}$ 如下：

$$\boldsymbol{R}=\begin{pmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mm} \end{pmatrix} \tag{4}$$

其中，r_{mn} 表示用户 m 对项目 n 的评分。

最后，将置信度矩阵和原始用户评分矩阵相结合，来计算未评分项目的预测评分，预测评分 p' 的计算式如下：

$$P'_{mn}=\sum_{i\in I}c_{mi}*r_{in}/\sum_{i\in I}c_{mi} \tag{5}$$

其中,p'_{mn}表示用户 m 对项目 n 的预测评分,I 表示所有对项目 n 评分过的用户集合,c_{mi}表示用户 m 与用户 i 之间的置信度,r_{in}表示用户 i 对项目 n 的评分。

利用计算得到的预测评分 p'填充原始的用户评分矩阵 $\boldsymbol{R}$,填充过后的用户评分矩阵 $\boldsymbol{R}'$如下:

$$\boldsymbol{R}'=\begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p'_{mn} \end{pmatrix} \tag{6}$$

其中 p'_{mn}表示用户 m 对项目 n 的预测评分。

2.3 融合用户间置信度的相似度

推荐系统中还存在传统算法精确度不高这一问题。推荐系统的好坏主要依赖于目标用户(项目)相邻用户的选择,也就是依赖于用户(项目)间相似度的计算,因此相似度计算的准确性在一定程度上会影响推荐质量。传统的相似度计算是通过用户间对共同项目的评分来计算的,但由于数据稀疏性的影响,用户间对项目的共同评分不足,通过传统相似度计算方式得到的推荐结果质量较差。因此改进相似度计算成为了缓解数据稀疏和提高推荐质量的方法之一。

本文通过提高挖掘用户间的相关关系来改进传统相似度计算,首先利用 Apriori 算法挖掘出用户间置信度,再与传统用户间相似度相融合,构建综合用户间相似度,最后利用综合相似度来对用户进行推荐,提高推荐质量。

本文设置参数 α,综合考虑用户间的置信度与相似度,并将两者相结合,构建用户间的综合相似度。用户间的综合相似度表达式如下:

$$new\_sim(u,v)=\alpha\times c(u,v)+(1-\alpha)\times sim(u,v) \tag{7}$$

其中,$new_sim(u,v)$表示用户 u 和用户 v 间的综合相似度,$c(u,v)$表示用户 u 和用户 v 间的置信度,$sim(u,v)$表示用户 u 和用户 v 之间的相似度。

3 算法描述

3.1 算法具体描述

本文提出的融合关联规则的协同过滤算法(ARCF)主要分为两个部分,第一部分是数据填充,第二部分是相似度改进。数据填充部分,主要考虑用户间的潜在关系,通过关联规则算法挖掘用户间的潜在关系,该联系用关联规则的置信度表示,将其同原始用户评分矩阵相结合,计算未被用户评分的项目的预测评分,达到缓解数据稀疏的目的。相似度改进部分,考虑到传统相似度计算会随着数据量的增多而使得算法推荐精度下降,因此对传统相似度计算进行改进。传统的用户间相似度是通过用户对项目的共同评分得到的,而用户间的置信度是通过用户间的共同评分行为挖掘得到的,两者之间具有一定的共通性,因而将两者相结合,构建用户间的综合相似度公式,通过挖掘用户间的隐式联系来改进传统的相似度计算。

3.2 算法具体步骤

融合关联规则的协同过滤算法步骤如算法 1 所示。

算法 1 融合关联规则的协同过滤算法

输入:用户项目评分矩阵

输出:用户推荐列表

步骤 1 将原始数据集处理为式(2)所示的电影用户集合。

步骤 2 使用 Apriori 算法对处理过后的集合进行关联规则挖掘,得到用户间的关联规则,使用式(1)来计算每个关联规则的置信度。

步骤 3 利用第二步中得到的用户间的置信度构建用户置信度矩阵 **Con**,如式(3)所示。

步骤 4 将原始数据集处理为用户评分矩阵 **R**,如式(4)所示。

步骤 5 将用户置信度矩阵 **Con** 通过式(5)来填充用户评分矩阵 **R**,得到填充过后的用户评分矩阵 **R'**,如式(6)所示。

步骤 6 使用式(7)改进传统的相似度计算。

步骤 7 将用户评分矩阵 **R'**和综合相似度 new_sim(u,v)用于基于用户的协同过滤算法(User-CF)。

步骤 8 根据上一步所得推荐结果为用户进行推荐。

4 实验结果及分析

4.1 实验数据集

本次实验采用 MovieLens-latest-small 数据集(以下简称 MLS 数据集)和 MovieLens100K 数据集(以下简称 ML100K 数据集)。MLS 数据集共有 610 名用户对 9 742 部电影进行了 100 836 条评分,其中稀疏度为 98.30%。ML100K 数据集共有 943 名用户对 1 682 部电影进行了 100 000 条评分,其中稀疏度为 93.70%。电影评分采用 5 分制,分数从 1~5 分不等,取整数。分数越高表示用户对该电影越喜爱。电影种类按年代分为 90 年代、00 年代、10 年代,按体裁分为浪漫、动作、犯罪等 19 种影视体裁。本次实验将 80%的数据集作为训练集,将 20%的数据集作为测试集。

实验将本文提出的融合关联规则的协同过滤算法(ARCF)在稀疏度不同的 MLS 数据集和 ML100K 数据集上进行横向比较。

4.2 实验评价指标

本实验的推荐系统的评价指标采用准确率(Precision)和覆盖率(Coverage),以及综合指标 F[15-16]。

准确率(Precision)可以用于衡量推荐算法的精确度,从而衡量推荐系统的推荐质量。准确率(Precision)又称查准率,是表示推荐给用户的项目中用户感兴趣的概率,准确率(Precision)公式如下:

$$precision=\frac{\sum\limits_{u\in U}|R_u\cap T_u|}{\sum\limits_{u\in U}|R_u|} \tag{8}$$

覆盖率(Coverage)用于衡量推荐系统推荐产品的范围是否具有广泛性,尤其在应用协同过滤算法的推荐系统下,随着数据量的增加,一些高质量的推荐系统往往只针对具有大量数据的小部分物品,而忽略了其余大部分的物品,产生长尾问题。覆盖率越高的推荐系统,代表该系统推荐给用户的物品范围越广泛和多元。覆盖率(Coverage)表示推荐给用户的商品种类占总的商品种类的百分比,覆盖率(Coverage)公式如下:

$$Coverage=\frac{|U_{u\in U}R_u|}{n} \tag{9}$$

由于准确率(Precision)会随着推荐数目的增加而降低,而覆盖率(Coverage)会随着推荐数目的增加而增加,因此单看两个指标无法准确表示算法的好坏。将准确率(Precision)和覆盖率(Coverage)结合起来,构建一个综合指标 F,以更为全面地表示算法的性能好坏。综合指标 F 值公式如下:

$$F=\frac{2\times Coverage\times Precision}{Coverage+Precision} \tag{10}$$

4.3 实验结果分析

实验将本文提出的基于融合关联规则的协同过滤算法(AR-CF)同传统的基于用户的协同过滤算法(User-CF)。在同一电影数据集上进行比较,算法实现语言为 python3 版本。

4.3.1 参数分析

将基于融合关联规则的协同过滤算法(AR-CF)应用于MLS数据集上进行调参。当推荐系统中目标用户邻居数 $L=20$、推荐电影数目 $n=10$ 时,不同的 α 值下所对应的算法准确率(Precision)的变化,α 值每次变化以 0.1 递增。如图 1 所示,准确率在 0.7～0.8 区间逐渐上升,在接近 0.8 处得到准确率的最大值。因此针对 0.7～0.8 区间的 α 值,每次变化以 0.01 递增。如图 2 所示,准确率在 0.73～0.74 最高,故最后 α 值取两者中间值 0.735。

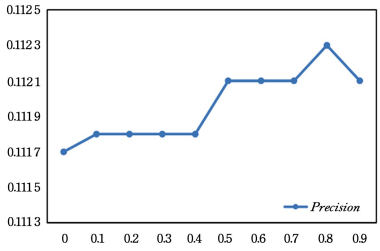


图 1 α 值对准确率的影响

Fig. 1 Effect of α on precision

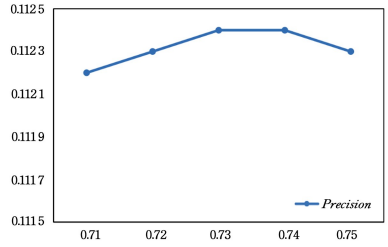


图 2 α 值对准确率的影响

Fig. 2 Effect of α on precision

4.3.2 结果分析

(1)推荐算法精确性

在推荐系统中,设目标用户邻居数 $L=20$,推荐电影数目 n 以 5 为单位逐步递增,在不同的电影推荐数目 n 下,将基于用户的协同过滤算法(User-CF)和基于项目的协同过滤算法(Item-CF)同本文算法(AR-CF)的准确率(Precision)、覆盖率(Coverage)及两者综合指标 F 值(F-Measure)在 ML100K 数据集上进行比对。

不同算法在邻居数 $L=20$、不同的推荐数目 n 下准确率的变化如图 3 所示。

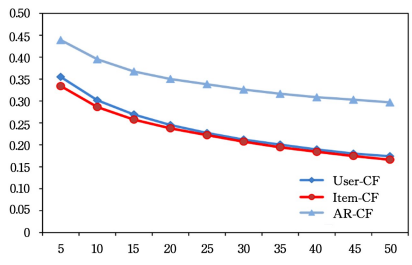


图 3 不同算法下推荐数量 n 对准确率的影响

Fig. 3 Effect of recommendation number n on precision

不同算法在邻居数 $L=20$、不同的推荐数目 n 下 F 值的变化如图 4 所示。

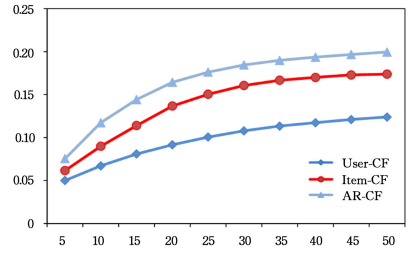


图 4 不同算法下推荐数量 n 对 F 值的影响

Fig. 4 Effect of recommendation number n on F values

如图 4 和图 5 所示,本文所提算法较协同过滤算法中两大经典算法的精确度(precision)及算法性能 F 值都有较大提升。说明本文所提算法具有较高的推荐精度,在一定程度上提高了推荐算法的精确度。

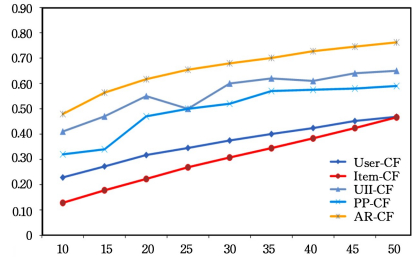


图 5 不同算法下推荐数量 n 对覆盖率的影响

Fig. 5 Effect of recommendation number n on coverage

(2)推荐算法长尾性

在提升算法精确度的基础上,本文进一步将所提算法(AR-CF)同文献[10]提出的结合用户兴趣和改进的协同过滤推荐算法(UII-CF)和文献[13]提出的基于改进相似度的协同过滤算法(PP-CF),以及 User-CF 和 Item-CF 几种主流推荐算法在算法覆盖率上进行对比,在确保提升算法准确率的同时,观察改进过后的算法对缓解推荐系统所存在的长尾问题的能力否有所提高。

长尾问题指推荐系统在为用户进行推荐时,为了提高准确率,往往会忽略尾部产品,只为用户推荐头部产品,会造成推荐商品缺乏多样性[3]。

不同算法在邻居数 $L=20$、不同的推荐数目 n 下覆盖率的变化对比结果如图 5 所示。由图 5 可知,本文所提 AR-CF 算法的覆盖率一直大于 User-CF, Item-CF, UII-CF, PP-CF 这 4 种推荐算法,说明本文所提算法具有较好的推荐能力,可以有效解决推荐系统存在的长尾问题,在一定程度上提高了推荐系统的推荐质量。

(3)推荐算法适用数据集

在推荐系统中,设目标用户邻居数 $L=20$,推荐电影数目 n 以 5 为单位逐步递增,不同的电影推荐数目 n 下,本文算法(AR-CF)在基于用户的协同过滤算法上所提高的准确率(Precision)百分比,在 ML100K 数据集和 MLS 数据集上进行比对。

不同数据集上在邻居数 $L=20$、不同的推荐数目 n 下 AR-CF 算法较 User-CF 算法所提高的准确率百分比的具体变化如图 6 所示。从图 6 可以看出,对比 AR-CF 算法应用于 ML100K 数据集和 MLS 数据集,AR-CF 算法在 ML100K 上的准确率提高的效果更为明显,且随着推荐数目的增加准确

率提高越明显。说明在评分数目及稀疏度相似的情况下,用户数越多,AR-CF 算法的效果越好。

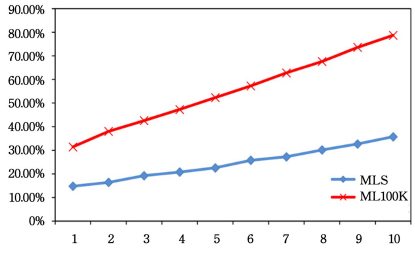


图 6　不同数据集上推荐数量 n 对提高准确率百分比的影响

Fig. 6　Effect of recommendation number n on improving accuracy percentage on different data sets

综合以上实验结果,可以得出本文提出的 AR-CF 算法可以一定程度上提高算法精确度,且所提算法可以较好地缓解推荐算法的长尾问题,具有较高的推荐质量。

结束语　传统的协同过滤算法存在着数据稀疏和算法精度不高的问题,导致推荐结果质量不高。因此本文提出一种融入关联规则的协同过滤算法(AR-CF),通过关联规则挖掘用户间潜在的相关关系,用于用户评分数据的填充和传统相似度的改进,达到缓解数据稀疏和提高算法精度的目的。由对比实验可知,相较于传统的协同过滤算法,本文算法具有更高的算法精度,并且在提高算法精确度的基础上,与近年来的推荐算法相比,本文算法具有较好的覆盖率,可以有效解决推荐系统的长尾问题,在一定程度上提高了推荐系统的推荐质量。在未来的研究中,将考虑在本文算法的基础上融合更多的隐式因子,如用户评论、电影类别、评分时间等,挖掘更多隐式信息来填充用户评分矩阵以及改进相似度计算。

参 考 文 献

[1] XU H L, WU X, LI X D, et al. Comparison Study of Inter Recommendation System[J]. Journal of Software, 2009, 20(2): 350-362.

[2] LIU J G, ZHOU T, WANG B H. Research Progress of Personalized Recommendation System [J]. Progress in Natural Science, 2009, 19(1): 1-15.

[3] KIM B D, KIM S O. A new recommender system to combine content-based and collaborative filtering systems [J]. Journal of Database Marketing & Customer Strategy Management, 2001, 8: 244-252.

[4] KIM B M, LI Q, PARK C S, et al. A new approach for combining content-based and collaborative filters[J]. Journal of Intelligent Information Systems, 2006, 27: 79-91.

[5] HONG B, YU M. A collaborative filtering algorithm based on correlation coefficient [J]. Neural Computing and Applications, 2019, 31: 8317-8326.

[6] CHEN H, YAN W, SUN H, et al. Tag-Extended Collaborative Filtering Recommendation Algorithm[J]. SN Computer Science, 2020, 1(5): 302.

[7] GUO L M, LIANG J K, ZHU Y, et al. Collaborative filtering recommendation based on trust and emotion [J]. Journal of Intelligent Information Systems, 2019, 53: 113-135.

[8] LIU X J. A collaborative filtering recommendation algorithm based on the influence sets of elearning group's behavior [J]. Cluster Computing, 2019, 22: 2823-2833.

[9] ZHANG P, ZHANG Z, TIAN T, et al. Co-llaborative filtering recommendation algorithm integrating time windows and rating predictions [J]. Applied Intelligence, 2019, 49: 3146-3157.

[10] WANG Y, ZHANG J, XU H L. Combining User Interests with Improved Collaborative Filtering Recommendation Algorithm [J]. Journal of Chinese Computer Systems, 2020, 41(8): 1665-1669.

[11] HAN S B, YI H W, LI X H, et al. Cold Start Recommendation Algorithm Based on Fusion Similarity and Hierarchical Clustering[J/OL]. Journal of Chinese Computer Systems: 1-8. http://kns.cnki.net/kcms/detail/21.1106.TP.2021517.1243.006.html.

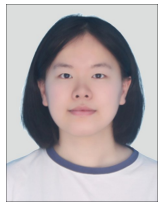
[12] WU J X, ZHANG Z H. Collaborative Filtering Recommendation Algorithm Based on User Rating and Similarity of Explicit and Implicit Interest[J]. Computer Science, 2021, 48(5): 147-154.

[13] LI R, LI M G, GUO W Q. Research on Collaborative Filtering Algorithm Based on Improved Similarity [J]. Computer Science, 2016, 43(12): 206-208, 240.

[14] LU L N, CHEN Y P, WEI H Y, et al. Research on Apriori Algorithm in Mining Association Rules [J]. Journal of Chinese Computer Systems, 2000(9): 940-943.

[15] ZHU Y X, LU L Y. Evaluation Metrics for Recommender Systems [J]. Journal of University of Electronic Science and Technology of China, 2012, 41(2): 163-175.

[16] PAZZANI M J, BILLSUS D. Learning and revising user profiles: the identification of interesting Web sites [J]. Machine Learning, 1997, 27(3): 313-331.



DONG Yun-xin, born in 1997, postgraduate. Her main research interests include recommendation system and data mining.

LIN Geng, born in 1981, Ph.D, professor, is a member of China Computer Federation. His main research interests include combinatorial optimization and artificial intelligence.