

基于聚类 and 随机森林的协同过滤推荐算法

杨兴雨, 李华平, 张宇波

YANG Xingyu, LI Huaping, ZHANG Yubo

广东工业大学 管理学院, 广州 510520

School of Management, Guangdong University of Technology, Guangzhou 510520, China

YANG Xingyu, LI Huaping, ZHANG Yubo. Collaborative filtering algorithm based on clustering and random forests. Computer Engineering and Applications, 2018, 54(16): 152-157.

Abstract: To handle the inefficiency problem of online recommendation of neighborhood-based collaborative filtering algorithms, this paper proposes a method to train a rating prediction model offline. The method firstly reduces the dimensions of the user vectors and the item vectors in the user-item rating matrix, and transforms this matrix so as to use supervised learning models. A random forest model is then trained by using the transformed data, and the online rating prediction is made by the previous trained model without the search of the nearest neighborhoods. The experiment results show that the method performs much better than neighborhood-based collaborative filtering algorithms in term of online recommendation efficiency without decreasing the precision of rating prediction.

Key words: collaborative filtering; recommendation algorithm; clustering; random forests

摘 要: 针对基于邻近关系的协同过滤算法在线推荐效率低的问题, 提出了一种可离线训练评分预测模型的算法。通过聚类算法降低用户-项目评分矩阵中用户向量和项目向量的维数, 并对数据进行转换使其适用于监督模型; 利用转换后的数据离线训练随机森林模型, 在线推荐时只需根据随机森林模型的规则进行评分预测, 无需查找最邻近用户或项目。实验结果表明, 该算法在不降低评分预测精度的情况下, 在线推荐效率远高于基于邻近关系的协同过滤算法。

关键词: 协同过滤; 推荐算法; 聚类; 随机森林

文献标志码: A **中图分类号:** TP399 **doi:** 10.3778/j.issn.1002-8331.1712-0089

1 引言

互联网的快速发展使其成为信息传递和商品流通日益重要的平台。急剧膨胀的网络资源在为用户提供丰富多样的信息的同时, 也对用户搜索信息的能力和精力提出了挑战, 用户不得不投入更多时间和精力搜索其所需要的信息。推荐系统(Recommendation Systems, RS)由此而生, 它利用用户的偏好信息自动向用户推荐符合其兴趣特点的项目^[1], 项目泛指商品、信息等被推荐给用户的对象。

推荐系统分析的数据主要有两种类型: 一种是通过矩阵表示的数值型数据, 比如用户对项目的评分、购买、浏览等, 用户是否购买商品或是否浏览网页可用数字 1 和 0 表示。另一种类型的数据是文本数据, 比如用

户对项目的评论、网页的内容、电影的简介等。根据所分析数据的类型, 推荐技术可分为协同过滤(Collaborative Filtering, CF)推荐技术、基于内容(Content-based)的推荐技术和混合(Hybrid)推荐技术。协同过滤推荐算法分析数值矩阵; 基于内容的推荐算法分析文本数据; 而混合推荐算法则综合前两者的优势。

协同过滤算法根据用户-项目评分矩阵中已有的评分, 利用用户间或者项目间的相关关系预测评分矩阵中缺失的评分, 然后根据评分的高低衡量用户对项目的偏好程度。协同过滤算法可分为基于邻近关系(Neighborhood-based)的推荐算法和基于模型(Model-based)的推荐算法。前者根据与目标用户相似的用户对某项目的评分, 预测目标用户对该项目的评分; 或者在目标用户

基金项目: 国家自然科学基金(No.71301029)。

作者简介: 杨兴雨(1981—), 男, 博士, 副教授, 研究领域为机器学习、不确定性决策, E-mail: yangxy@gdut.edu.cn; 李华平(1992—), 男, 硕士研究生, 研究领域为机器学习、推荐算法。

收稿日期: 2017-12-07 **修回日期:** 2018-03-12 **文章编号:** 1002-8331(2018)16-0152-06

已评分的项目中搜索与目标项目最相似的 k 个项目,根据目标用户对这些项目的评分,预测其对目标项目的评分。基于邻近关系的推荐算法易于理解和实现,但每次进行推荐时都要查找目标用户(项目)的 k 最邻近用户(项目)。当推荐系统面临数以百万甚至千万级别的用户和项目时,计算开销非常庞大,算法的实时性将很难保证,相应的推荐系统将面临可扩展性问题^[2]。针对此问题,学者提出的解决方法主要分为两类:一类是采用并行存储和计算技术^[3-5],提高推荐系统的可扩展性;另一类方法则是采用基于模型的推荐技术,此类技术可离线根据用户-项目矩阵构建模型,在线根据模型产生推荐,提高在线推荐效率。基于模型的推荐技术常用的模型包括聚类模型^[6]、贝叶斯网络^[7]、矩阵分解^[8]等。

随着推荐系统中用户量和项目量的增长,在数据库中查找相似用户或项目会消耗大量的计算资源和时间。为此,Herlocker 等人^[9]提出一种基于项目聚类的协同过滤算法,该算法先对项目进行聚类,然后在各个簇中独立应用基于邻近关系的推荐算法,缩小了最邻近用户(项目)的搜索范围,提高在线推荐的效率;Wei 等人^[10]提出的算法先对项目进行聚类,然后在每个簇中采用基于用户的协同过滤算法进行推荐;Liao 等人^[11]在聚类后的每个项目簇中独立应用随机游走过程对项目进行排名;魏慧娟等人^[12]先对用户进行聚类,在线推荐时在目标用户所属簇的若干最邻近簇中查找最邻近用户。Sarwar 等人^[13]提出基于二分 k -means 用户聚类的协同过滤算法,该算法可以相对均匀地对用户进行聚类;Koohi 等人^[14]提出基于模糊 C-均值聚类模型的协同过滤算法;Tsai 等人^[15]提出基于聚类集成的协同过滤算法。在大数据背景下,部分学者^[16,16]提出基于 Hadoop 分布式平台的聚类协同过滤推荐算法,在云环境中具有良好的可扩展性。除了用户-项目评分矩阵数据,郭弘毅等人^[17]提出一种融合用户社交网络信息和兴趣聚类的推荐算法;王媛媛等人^[18]则结合用户人口统计学信息和评分矩阵计算用户之间的相似性,并使用分层近邻传播聚类算法对用户进行聚类;Najafabadi 等人^[19]提出基于聚类和关联规则挖掘的推荐算法,并结合项目属性数据提高推荐效果。

以上基于聚类的协同过滤算法主要通过缩小最邻近用户或项目的搜索空间从而提高推荐的效率。本文提出一种基于聚类和随机森林的协同过滤推荐算法(CRF)。首先通过聚类算法将用户-项目评分矩阵转换成适用于监督学习模型的数据;然后利用这些数据训练随机森林回归模型,在线推荐时只需根据预先构建的随机森林模型进行评分预测,无需查找最邻近用户或项目,提高了推荐的效率。

2 协同过滤算法

假设推荐系统中有 n 个用户 $U = \{u_1, u_2, \dots, u_n\}$ 和 m

个项目 $I = \{i_1, i_2, \dots, i_m\}$,用户对项目的历史评分可由矩阵 $R_{n \times m}$ 表示(见表1)。评分矩阵中空缺的元素表示对应的用户没有给对应的项目评分。协同过滤推荐算法预测评分矩阵的空缺值,并向用户推荐评分预测值较高的项目。

表1 评分矩阵的简单例子

	i_1	i_2	i_3	i_4
u_1	4	3	5	
u_2	4	2	1	
u_3	3		2	4
u_4	4	4		
u_5	2	1	3	5

2.1 基于邻近关系的协同过滤算法

基于邻近关系的协同过滤推荐算法可分为基于用户的协同过滤(User-based CF,UCF)算法和基于项目的协同过滤(Item-based CF,ICF)算法。前者的基本思想是,相似的用户有共同的偏好,因此在向目标用户推荐项目时,UCF算法在用户集中查找与目标用户最相似的若干用户,并向目标用户推荐最相似用户所偏好的项目。而后者假设用户会喜欢与其过去喜欢的项目相似的项目,因此ICF向用户推荐与其过去喜欢的项目相似的项目。

在UCF算法中,目标用户 a 对项目 i 的评分为:

$$\hat{r}_{ai} = \bar{r}_a + \frac{\sum_{u \in N(a)} (r_{ui} - \bar{r}_u) \times \text{sim}(a, u)}{\sum_{u \in N(a)} |\text{sim}(a, u)|} \quad (1)$$

其中, $N(a)$ 表示已经对项目 i 评分的用户中,与目标用户最相似的 k 个用户; r_{ui} 表示用户 u 对项目 i 的评分; \bar{r}_a 和 \bar{r}_u 表示用户 a 和 u 历史评分的均值; $\text{sim}(a, u)$ 表示两用户间的相似度。在UCF算法中,常用的相似度计算方式是 Pearson 相关系数(式(2))和余弦相似度(式(3)), C 表示用户 a 和 u 共同评分的项目集。

$$\text{sim}^{\text{PCC}}(a, u) = \frac{\sum_{i \in C} (r_{ai} - \bar{r}_a)(r_{ui} - \bar{r}_u)}{\sqrt{\sum_{i \in C} (r_{ai} - \bar{r}_a)^2} \sqrt{\sum_{i \in C} (r_{ui} - \bar{r}_u)^2}} \quad (2)$$

$$\text{sim}^{\text{COS}}(a, u) = \frac{\sum_{i \in I} r_{ai} \times r_{ui}}{\sqrt{\sum_{i \in I} r_{ai}^2} \sqrt{\sum_{i \in I} r_{ui}^2}} \quad (3)$$

ICF算法通过目标用户 a 过去对其他项目的评分预测该用户对项目 i 的评分,即:

$$\hat{r}_{ai} = \frac{\sum_{j \in N(i)} r_{aj} \times \text{sim}(i, j)}{\sum_{j \in N(i)} |\text{sim}(i, j)|} \quad (4)$$

其中, $N(i)$ 表示用户 a 已评分的项目中与项目 i 最相似的项目。

2.2 基于聚类的协同过滤算法

由于传统基于邻近关系的协同算法在进行推荐时,在所有用户中查找与目标用户最相似的 k 个用户,推荐算法的效率随着推荐系统中用户数量的增加而降低。Sarwar等人^[13]提出一种基于用户聚类的协同过滤算法,算法首先将推荐系统的用户划分为多个不相交的用户簇(集合) $\{U_i | U_i \cup U_j = U; \cap U_i = \emptyset; i=1,2,\dots,l\}$,使得每个簇内用户的相似度尽可能高,而簇间用户的相似度尽可能低。在进行推荐时,仅需在目标用户所属的簇中应用UCF算法进行推荐即可,这缩小了最相似用户的查找范围,提高了推荐系统的效率。然而,搜索空间的缩小会导致评分预测精度的下降。

3 基于聚类 and 随机森林的协同过滤算法

已有基于聚类的协同过滤算法主要通过缩小最邻近用户(或项目)的搜索空间提高推荐的效率。虽然聚类过程可离线完成,但最邻近搜索仍然是一个在线的过程。本文对这一点进行了改进,对用户和项目进行聚类后,计算每个用户(项目)对每个簇的隶属度,根据隶属度和评分矩阵构造监督学习模型的训练数据,并离线训练随机森林模型。在线推荐时根据预先构建的随机森林模型进行评分预测,这个过程所需的时间远小于最邻近搜索过程。

3.1 聚类算法

聚类算法对数据中的对象进行分组,使得分组后同一组内的对象相似度尽可能高,而组间对象的相似度尽可能低^[20]。聚类算法可分为三类:划分聚类、基于密度的聚类、层次聚类^[21]。这三类算法中最常用的分别为 k -means^[22]、DBSCAN^[23]和BIRCH^[24]。

本文采用二分 k -means^[25]聚类算法对用户和项目进行聚类,该算法先将所有样本划分为两簇,再对所有簇中样本数最多的簇进行再次二分,重复此步骤直到簇数达到预先设定的正整数 k 为止。其具体流程如下:

- (1)在样本集中随机选取两个样本作为初始簇中心;
- (2)计算所有样本与簇中心的相似度,并将这些样本划分到与其相似度最高的簇中;
- (3)计算各簇中所有样本的均值,作为新的簇中心;
- (4)重复步骤(2)和(3),直到所有簇中心都不再改变;
- (5)从划分后的簇中选取样本数最多的簇,对其进行步骤(1)至(4)的划分;
- (6)重复步骤(5)直到簇数达到 k 为止。

二分 k -means算法的聚类效果会受到参数 k 的影响。实际应用中,最优簇数 k_{opt} 需要经过多次试验获得。为了缩小搜索范围,多数学者采用的经验规则为:在区间 $[2, \sqrt{n}]$ 内搜索 k_{opt} , n 为样本的数量。

3.2 随机森林

监督学习是机器学习任务中根据带标记的训练数

据推断模型参数的过程^[26]。训练数据由训练样本组成,每个训练样本包含输入变量(通常是一个向量)和输出变量(数值或类别标签)。通过训练数据训练得到的模型可用于预测新样本(已知输入变量)的输出变量。

随机森林是一种常用的监督学习模型,可以用于分类和预测,它利用Bootstrap重抽样方法从原始样本中有放回地抽取多个样本集,利用每个样本集进行决策树建模。进行预测时,随机森林取所有决策树预测结果的均值作为最终预测结果。大量的理论和实证研究证明了随机森林具有很高的预测准确率,对异常值和噪声具有很好的容忍度,且不容易出现过拟合^[27]。

3.3 数据转换及评分预测

为了离线构建随机森林模型,需要将用户-项目评分矩阵转换成适用于监督学习模型的数据。具体地,通过聚类算法对用户进行聚类,聚类结束后,计算各个用户对各个用户簇的隶属度,见表2。相应地,对项目向量进行同样的转换,得到表3所示的隶属度矩阵。本文采用向量的余弦值表示样本(用户或项目)与簇中心的相似度,并以此作为样本对簇的隶属度,簇中心为该簇中所有样本的均值。

表2 用户隶属度矩阵

	U_1	U_2	U_3
u_1	0.8	0.1	0.2
u_2	0.4	0.7	0.2
u_3	0.1	0.8	0.3
u_4	0.4	0.2	0.6
u_5	0.7	0.1	0.3

表3 项目隶属度矩阵

	I_1	I_2	I_3
i_1	0.7	0.2	0.3
i_2	0.3	0.6	0.3
i_3	0.3	0.8	0.2
i_4	0.1	0.2	0.9

通过聚类得到的数据更适用于监督学习模型。一方面,用户向量的维数得到大幅度降低。用户-项目评分矩阵中用户向量的维数等于项目的数量,而表2中用户向量的维数等于用户簇的数量,通常用户簇的数量远小于项目的数量。另一方面,经过数据转换后的用户向量不存在缺失值。同样,转换后的项目数据也具有以上两个优点。

根据表1~表3的数据构造监督学习模型的输入变量和输出变量,得到如表4所示的数据。表1中每个元素对应表4中的一个样本(一行),因此在用户数和项目数各为 n 和 m 的推荐系统中,总共包含 $n \times m$ 个样本。每个样本的输入变量隐含着该样本对应的用户和项目的信息,来自隶属度矩阵;样本的输出变量对应着表1中的评分。

表4 监督学习模型的数据形式

用户-项目	输入变量						输出变量
u_1i_1	0.8	0.1	0.2	0.7	0.2	0.3	4
u_2i_2	0.4	0.7	0.2	0.3	0.6	0.3	2
u_3i_3	0.1	0.8	0.3	0.3	0.8	0.2	2
u_4i_4	0.4	0.2	0.6	0.1	0.2	0.9	?

数据转换后,将输出变量已存在的样本作为随机森林模型的训练样本,用于学习随机森林的预测规则,这个训练过程可离线完成。在线推荐时,系统根据随机森林的预测规则进行评分预测,无需查找最邻近用户或项目,可提高实时推荐的效率。

4 实验与分析

本文的实验采用网站 MovieLens(<https://grouplens.org/datasets/>)提供的数据集 ml-100k 和 Book-Crossing。其中 ml-100k 包含了来自 943 个用户对 1 682 部电影的 100 000 次评分;Book-Crossing 包含了 278 858 个用户对 271 379 本书的 1 149 780 次评分,由于该数据集数据量较大,本文从中剔除评分次数少于 20 次的用户和书籍,最后得到的实验数据包含了 23 225 次评分。对于每个实验数据集,本文采用监督学习实验中常用的方式,随机抽取 10% 的样本作为测试集,剩下的 90% 作为训练集。训练集用于训练随机森林模型,而测试集用于测试模型的预测精度。

本实验将本文提出的算法 CRF 与 UCF^[28]、ICF^[29]、基于聚类的推荐算法 Clust^[13]进行对比。以平均绝对误差(Mean Absolute Error,MAE)和均方根误差(Root Mean Square Error, RMSE)作为评分预测精度的评价标准:

$$MAE = \frac{\sum_{r \in T} |r - \hat{r}|}{\#T}$$

(5)

$$RMSE = \sqrt{\frac{\sum_{r \in T} (r - \hat{r})^2}{\#T}}$$

(6)

其中, $\#T$ 表示测试集中样本的数量; r 和 \hat{r} 分别表示评分的实际值和预测值。

图 1 展示了各算法在两个数据集上的表现(MAE 和 RMSE 随着 k 的变化情况, k 为最邻近用户或项目的数量), MAE 和 RMSE 越小说明评分预测的精度越高。可以发现,除了在 Book-Crossing 数据集上 MAE 值比 ICF 的大之外,本文提出的 CRF 在多数情况下表现最好。为了便于对比,图 1 中 Clust 和 CRF 算法在聚类时,簇的数量设为 10,后文会单独分析簇的数量对推荐质量的影响。

图 2 展示了 Clust 和 CRF 算法在两个数据集上随着簇数量的变化预测精度的变化情况。随着簇数的增长, Clust 的 MAE 和 RMSE 逐渐增长,而 CRF 的逐渐下降。不难解释,随着簇数的增长,每个簇中的样本会减少, Clust 算法在目标用户所属簇内查找得到的 k 个最邻近用户并非全局最近的 k 个用户,因此预测精度会下降。而簇数一定程度增加反而有利于提升 CRF 算法的预测精度,因为这增加了随机森林训练数据的输入变量的维度,可以训练得到更加精确的随机森林模型。

除了比较评分预测的精度之外,本文也对算法的在线推荐效率进行比较。图 3 展示了各算法在线推荐所需时间随着聚类簇数的变化情况。UCF 和 ICF 算法没有聚类过程,因此在线推荐时长不受簇数变化的影响。随着簇数的增长, Clust 的在线推荐时长下降,因为最邻近用户的搜索范围缩小。CRF 的在线推荐时长几乎不随簇数的变化而变化,因为 CRF 在线推荐时无搜索最邻近用户(或项目)的过程,只需根据随机森林模型预

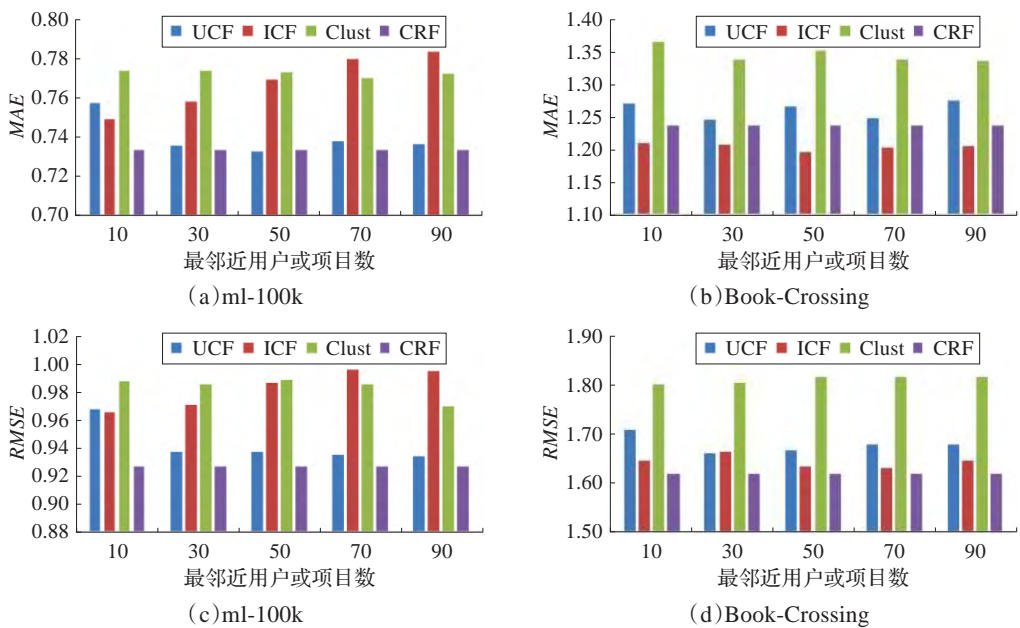


图1 评分预测精度比较

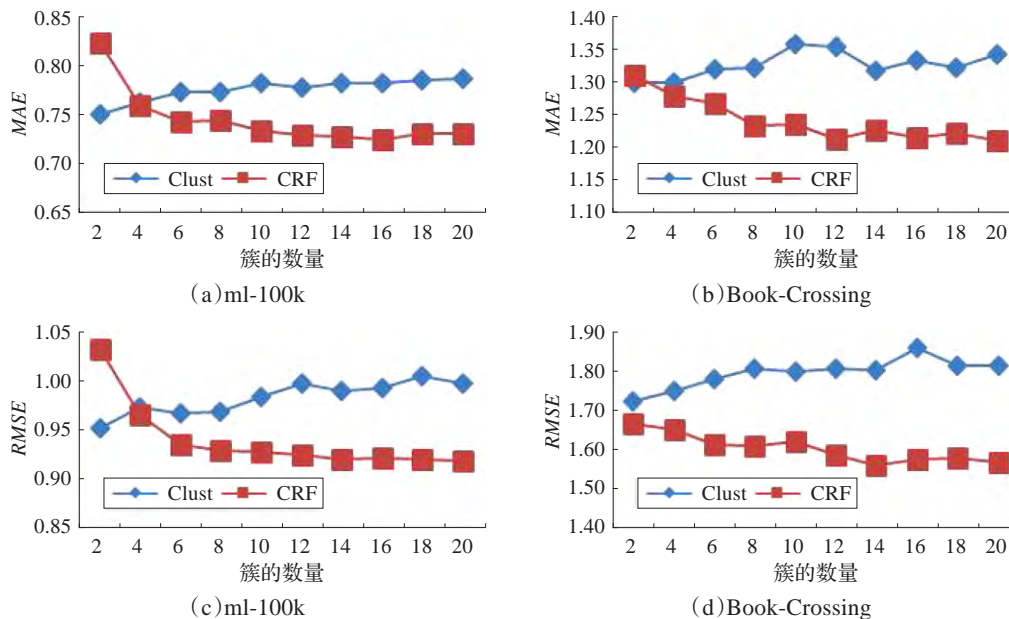


图2 评分预测精度随簇数的变化

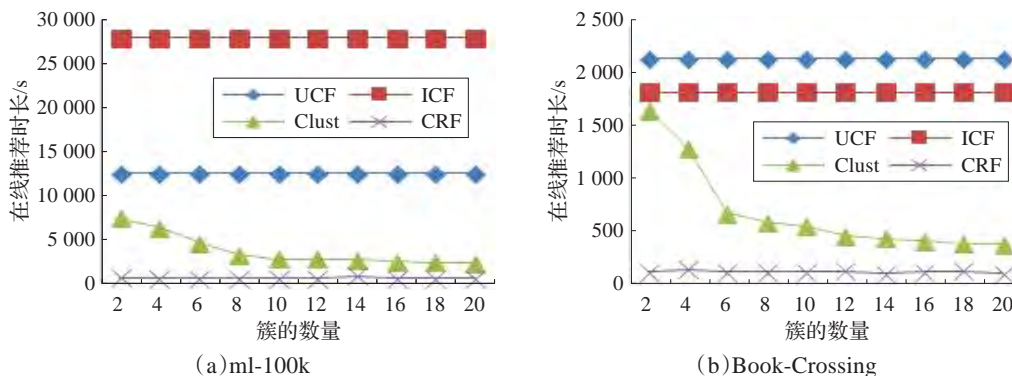


图3 在线推荐时长比较

先得到的规则进行推荐即可,所以在线推荐时长远小于其他模型。

5 结束语

针对基于邻近关系的推荐算法在线推荐效率低的问题,本文提出了一种基于聚类和随机森林的推荐算法。该算法通过聚类过程降低用户和项目向量的维数,同时将用户-项目评分矩阵转换成适用于监督学习模型的数据,并用于训练随机森林模型。在线推荐时,系统只需根据离线时构造的随机森林模型进行评分预测,不需要查找最邻近用户或项目,大幅提高了推荐的效率。此外,算法也保持了良好的预测精度,多数情况下比基于邻近关系的推荐算法优越。

参考文献:

- [1] Goldberg D, Nichols D, Oki B M, et al. Using collaborative filtering to weave an information tapestry[J]. Communications of the ACM, 1992, 35(12): 61-70.
- [2] 翁小兰, 王志坚. 协同过滤推荐算法研究进展[J]. 计算机工程与应用, 2018, 54(1): 25-31.
- [3] 程曦, 陈军. 基于 MapReduce 与项目分类的协同过滤算法[J]. 计算机工程, 2016, 42(7): 194-198.
- [4] 廖彬, 张陶, 国冰磊, 等. 基于 Spark 的 ItemBased 推荐算法性能优化[J]. 计算机应用, 2017, 37(7): 1900-1905.
- [5] Panigrahi S, Lenka R K, Stitipragyan A. A hybrid distributed collaborative filtering recommender engine using apache spark[J]. Procedia Computer Science, 2016, 83(1): 1000-1006.
- [6] 田保军, 胡培培, 杜晓娟, 等. Hadoop 下基于聚类协同过滤推荐算法优化的研究[J]. 计算机工程与科学, 2016, 38(8): 1615-1624.
- [7] 付永平, 邱玉辉. 一种基于贝叶斯网络的个性化协同过滤推荐方法研究[J]. 计算机科学, 2016, 43(9): 266-268.
- [8] 张航, 叶东毅. 一种基于多正则化参数的矩阵分解推荐算法[J]. 计算机工程与应用, 2017, 53(3): 74-79.
- [9] O'Conner M, Herlocker J. Clustering items for collaborative filtering[C]//ACM SIGIR Workshop on Recommender Systems, Berkeley, 1999.
- [10] Wei S, Ye N, Zhang S, et al. Collaborative filtering recommendation algorithm based on item clustering and global

- similarity[C]//Fifth International Conference on Business Intelligence and Financial Engineering,2012:69-72.
- [11] Liao C L, Lee S J. A clustering based approach to improving the efficiency of collaborative filtering recommendation[J]. Electronic Commerce Research & Applications, 2016, 18: 1-9.
- [12] 魏慧娟,戴壮红,宁勇余.基于最近邻居聚类的协同过滤推荐算法[J].中国科学技术大学学报,2016(9):736-742.
- [13] Sarwar B M, Karypis G, Konstan J, et al. Recommender systems for large-scale e-commerce: scalable neighborhood formation using clustering[C]//Conference on Computer and Information Technology, 2002.
- [14] Koohi H, Kiani K. User based collaborative filtering using fuzzy C-means[J]. Measurement, 2016, 91: 134-139.
- [15] Tsai C F, Hung C. Cluster ensembles in collaborative filtering recommendation[J]. Applied Soft Computing, 2012, 12(4): 1417-1425.
- [16] 孙天昊,黎安能,李明,等.基于Hadoop分布式改进聚类协同过滤推荐算法研究[J].计算机工程与应用,2015,51(15):124-128.
- [17] 郭弘毅,刘功申,苏波,等.融合社区结构和兴趣聚类的协同过滤推荐算法[J].计算机研究与发展,2016,53(8):1664-1672.
- [18] 王媛媛,李翔.基于人口统计学的改进聚类模型协同过滤算法[J].计算机科学,2017,44(3):63-69.
- [19] Najafabadi M K, Mahrin M N, Chuprat S, et al. Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data[J]. Computers in Human Behavior, 2017, 67: 113-128.
- [20] Jain A K, Murty M N, Flynn P J. Data clustering: a review[J]. ACM Computing Surveys, 1999, 31(3): 264-323.
- [21] Han J, Kamber M. Data mining: concepts and techniques[M]. San Francisco: Morgan Kaufmann, 2001.
- [22] Macqueen J. Some methods for classification and analysis of multivariate observations[C]//Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 1967: 281-297.
- [23] Ester M, Kriegel H P, Xu X. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise[C]//International Conference on Knowledge Discovery and Data Mining. AAAI Press, 1996: 226-231.
- [24] Zhang T, Ramakrishnan R, Livny M. BIRCH: an efficient data clustering method for very large databases[C]//ACM SIGMOD International Conference on Management of Data, 1996: 103-114.
- [25] Jain A K, Dubes R C. Algorithms for clustering data[J]. Technometrics, 1988, 32(2): 227-229.
- [26] Schapire R, Freund Y. Foundations of machine learning[M]. Cambridge: MIT Press, 2012.
- [27] 方匡南,吴见彬,朱建平,等.随机森林方法研究综述[J].统计与信息论坛,2011,26(3):32-38.
- [28] Herlocker J L, Konstan J A, Borchers A, et al. An algorithmic framework for performing collaborative filtering[C]//International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, 1999: 230-237.
- [29] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]//International Conference on World Wide Web. New York: ACM, 2001: 285-295.
- (上接100页)
- ence on Information Security and Assurance. Berlin, Heidelberg: Springer, 2009: 68-77.
- [6] Yang Y. Attribute-based data retrieval with semantic keyword search for e-health cloud[J]. Journal of Cloud Computing, 2015, 4(1): 10-16.
- [7] Yang L, Zheng Q, Fan X. RSPP: a reliable, searchable and privacy-preserving e-healthcare system for cloud-assisted body area networks[J]. arXiv: 1702.03467, 2017.
- [8] 陈兰香.一种基于同态Hash的数据持有性证明方法[J].电子与信息学报,2011,33(9):2199-2204.
- [9] 黄石,刘文卓,曹天杰.改进的基于同态哈希的云存储数据完整性验证方案[J].河海大学学报:自然科学版,2015,43(3):278-282.
- [10] Burton H B. Space/time trade-offs in hash coding with allowable errors[J]. Communications of the ACM, 1970, 13(7): 422-426.
- [11] 徐碧晗,郑东,任方.基于编码Hash同态性的数据持有性证明方案[J].计算机工程与应用,2016,52(9):34-44.
- [12] Bethencourt J, Sahai A, Waters B. Ciphertext-policy attribute-based encryption[C]//IEEE Symposium on Security and Privacy, 2007: 321-334.
- [13] 邵清,叶琨.基于编辑距离和相似度改进的汉字字符串匹配[J].电子科技,2016,29(9):7-11.
- [14] Augot D, Finiasz M, Sendrier N. A family of fast syndrome based cryptographic hash functions[C]//LNCS 3715: International Conference on Cryptology in Malaysia. Berlin, Heidelberg: Springer, 2005: 64-83.
- [15] Müller S, Katzenbeisser S. On multi-authority ciphertext-policy attribute-based encryption[J]. Bulletin of the Korean Mathematical Society, 2009, 46(4): 803-819.
- [16] Li J, Huang Q, Chen X, et al. Multi-authority ciphertext-policy attribute-based encryption with accountability[C]//ACM Symposium on Information, Computer and Communications Security, Hong Kong, China, March 2011, 2011: 386-390.