基于协同过滤的美食推荐算法

**摘 要：**当今中国正在面临经济快速发展，人们的生活节奏不断加快，如何推荐出更符合用户喜好的物品，成为当下人们较为关心的话题。本文通过对美食领域的初步调查，对标本数据的初步剖析。针对如何提供推荐个性化美食服务，提高推荐准确率的问题，分析基于用户的推荐算法、基于餐厅的推荐算法、基于的协调过滤算法以及流行度推荐算法这四种推荐算法，解决推荐餐厅与用户喜好适配度问题；针对算法衡量问题，采用准确率和召回率以及这三种指标比较，提高算法的准确性。

**关键词：**协同过滤；美食推荐系统；混合算法

**Food recommendation algorithm based on collaborative filtering**

**【Abstract】**Today's China is facing rapid economic development, people's pace of life is accelerating, how to recommend items that are more in line with user preferences, has become a topic that people are more concerned about. This paper provides a preliminary analysis of the specimen data through a preliminary survey of the gastronomic field. Aiming at the problem of how to recommend personalized food services and improve the accuracy of recommendation, four recommendation algorithms are analyzed: user-based recommendation algorithm, restaurant-based recommendation algorithm, coordinated filtering algorithm based and popularity recommendation algorithm, which solves the adaptability of recommended restaurants with user preferences. For the algorithm measurement problem, the accuracy rate and recall rate and the comparison of these three indicators are used to improve the accuracy of the algorithm.

**【Key words】**Collaborative filtration; Gourmet recommendation system; Hybrid algorithms

随着我国经济的快速发展和人民生活水平的不断提高,人们对饮食的要求从“吃不饱”到“吃饱”再到“吃好”,逐步走上了追求吃特色、健康、营养的新阶段。越来越多的人倾向于在互联网上寻找美食,但是人们普遍识别真假信息的能力有限,面对网络数据的爆炸增长,使得用户难以在海量的数据中快速的找到适合自己的美食，具有一定的盲目性。通过对现有美食的调查发现，美食数据排行过于笼统，不能有效的解决用户的个人喜好等问题。针对以上问题，对于现阶段的市场提供个性化美食推荐服务、提高检索效率、优化用户体验成为了广大用户的诉求。面对这一诉求，根据数据集的用户和餐厅特征，提出基于协同过滤[1-3]的美食推荐系统算法，该算法采用基于用户的推荐算法、基于餐厅的推荐算法、基于的协调过滤算法以及流行度推荐算法这四种推荐算法结合了准确率和召回率以及这三种指标比较。

**1相关工作**

协同过滤算法[1-3]在推荐系统领域被广泛应用，该算法通过对物品之间的相似度或者具有相同兴趣爱好的群众进行物品推荐。根据上述思想可分为四大类：一是基于用户的协调过滤；二是基于物品的协调过滤[4]。基于用户的协同过滤算法根据相似用户的喜好产生对目标用户的推荐; 基于项目的协同过滤推荐算法则根据用户对相似项目的评分数据预测目标项目的评分[5]；三是基于SVD的协同过滤[6]；四是基于流行度的协同过滤。

本文首先对数据集进行预处理，通过对不同算法的准确率和召回率以及F1指标的比较，优化推荐算法，提出混合算法，提高推荐算法的准确率，最后将其推荐算法应用于美食推荐系统领域。

**2 基于协同过滤的混合推荐算法**

本文在对基于用户、基于物品、基于的协同过滤算法实验测试后，又引进基于流行度的推荐算法，通过对这四种算法的三种指标进行比较，确定参数。为了进一步提高推荐算法的准确率，实验又将四种算法两两混合，这里着重介绍基于流行度与基于协同过滤的混合推荐算法。实验分为以下三步骤：

1、**基于流行度的协同过滤算法**

该混合算法的第一步研究基于流行度的推荐算法，该基础算法是根据用户对餐厅点击的热度将当下最热门的内容推荐给用户，由此对求解基于流行度的推荐算法的算法思路基本如下：

首先对去过同一家餐厅的总人数记作,流行度记作。

其次，为了让用户之前的相似度更能说明用户之前的兴趣喜爱程度是接近的，我们把两个用户对流行度相对较低的餐厅进行权重计算，如下：

对于低于平均流行度的餐厅我们进行 对其权重赋值大1，对于高于平均流行度的餐厅对其权重赋值小于1。基于此，计算相似度，如下：

考虑到冷门项目的兴趣度，引入权重因子 ,如下：

根据上述引入的权重因子,前个为用户的推荐餐厅。

2、**基于SVD的协同过滤算法**

对基于SVD的协同过滤算法的研究，首先对矩阵空白值进行相关行列的平均值进行填充。下一步对矩阵进行分解，其中 为的矩阵，即用户和餐厅的关系。为的正交矩阵，即用户给出的评分与匹配的另一位用户给出的评分的总和；为的对角矩阵，即矩阵的奇异值；矩阵为的正交矩阵，即该餐厅收到的评分与匹配的另一家餐厅收到的评分总和，得奇异值分解，如下：

对矩阵进行维化简得到远远小于，计算用户对餐厅的预测评分推荐，如下：

利用 SVD 来逼近矩阵并从中获得主要的特征。通过保留矩阵的 80%~90%的能量，就可以得到重用的特征并去除噪声。

3**、偏好程度**

通过对以上两步骤基础算法的研究，本文引入用户对于餐厅的一个偏好程度[7]，目的就是根据用户的目标用户评分来学习用户对不同餐厅偏好，从而给出用户感兴趣的餐厅分布[8],为了提高推荐算法的准确程度，提出了基于与流行度的协同过滤推荐算法，衡量与流行度的两者重要性引入权重因子，提出新偏好程度算法，如下：

通过对权重因子的测试，获得本推荐算法的相似邻居集，从而产生推荐列表。

**3实验及评估结果**

1. **实验数据集**

本文数据来自于和鲸社区美食数据集①,共4万家餐厅，54万用户，440万条评论数据。首先对数据集进行数据预处理，删除低于100条的数据记录数，通过整理最终参与实验一共4417家餐厅，3291用户，76208条数据。由于用户评价较为主观性，我们采用归一化处理评价数据。

1. **评估指标**

随着人工智能和大数据时代的到来,推荐系统在各行各业中的应用也是越来越多,但是从目前的发展来说,如何在不同的领域中判断推荐系统性能的优劣也是个尤为重要的课题[9]。考虑到实际需求,应用推荐系统的目的就在于挖掘更多的用户需求,为用户推荐最感兴趣的物品,以此来促进物品的销售,一般会用一些指标来度不同用户量评估推荐系统的质量。

设为符合用户B喜爱的推荐餐厅, 为用户B总喜爱的推荐餐厅，为用户B总推荐餐厅[8]。

第一方面：精确率。表示符合用户喜爱的推荐物品数在用户总推荐物品数的比例，主要考虑用户对推荐物品感兴趣的可能性，定义如下：

第二方面：召回率。表示符合用户喜爱的推荐物品数与用户总喜爱的推荐物品数的比例，主要考虑用户喜好物品被推荐的可能性，定于如下：

第三方面：指标，代表精确率和召回率的一个综合考量，两者一起使用才能评价推荐系统的好坏，定义如下：

上述三种指标表示推荐系统的准确度。

1. **参数确定**

在实验开始前，将数据集分为训练集和测试集，通过误差计算，发现不同参数对应的误差值都不同。这些参数包括邻居个数，训练，测试集的划分比例。为测试其灵敏性，分别对训练集、测试集进行数据测试。

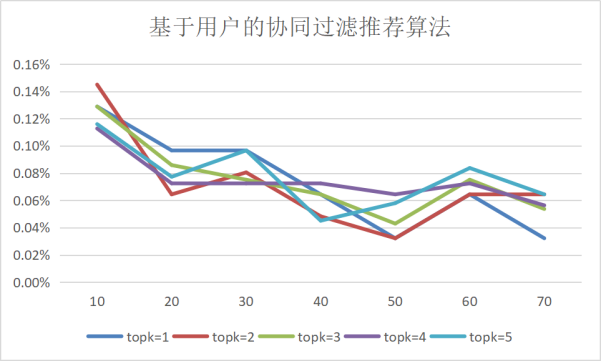


图1基于用户的协同过滤推荐算法

在测试基于用户的推荐算法中，我们对不同推荐个数以及邻居个数进行研究，结果如上图1所示。根据图表中数据显示，邻居个数的增加，基于用户的推荐相似度总体趋势呈下降状态。并通过对数值反复的模拟计算，前期topk=5的曲线趋势值要处于相对于中间趋势，但后期随着邻居个数的增加，取topk=5的效果要好于其他。

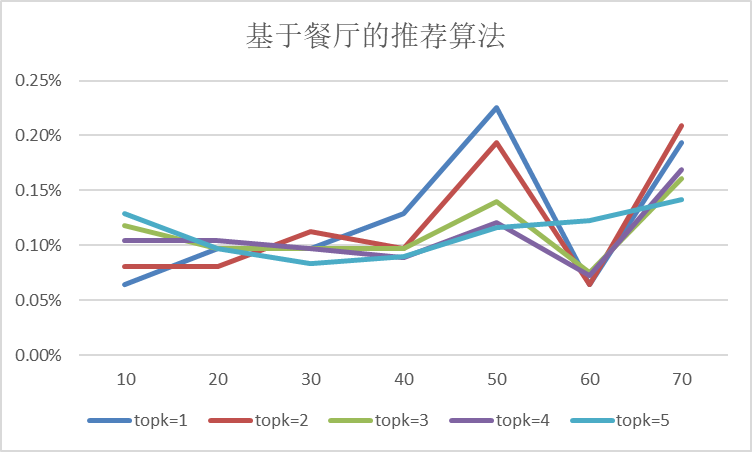


图2基于餐厅的推荐算法

为了测试基于餐厅的推荐算法对推荐程度的影响，改变topk和邻居个数的值通过控制变量法，测得数据如上图。通过图2我们可以看出在邻居个数位于50左右，基于餐厅推荐算法的相似度达到一个相对峰值，为相对优值。

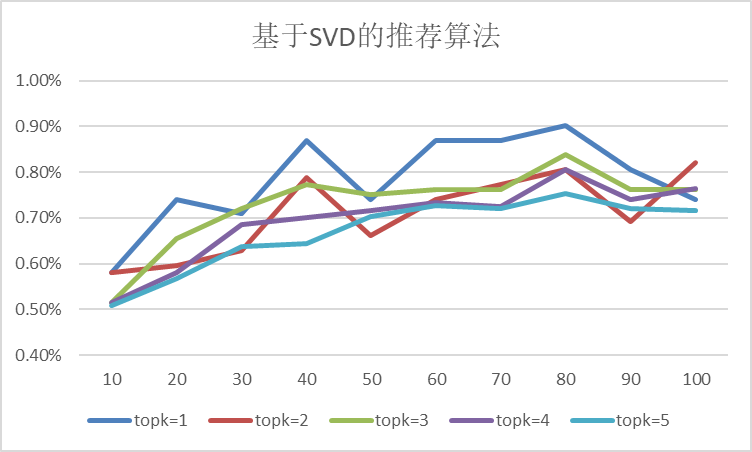


图3基于SVD的推荐算法

基于上述两种推荐算法引入第三种基于SVD的推荐算法对实验的影响程度，根据数据得出上述图3。我们发现基于SVD 推荐总体数据区间位于0.50%—0.90%，整体相似度相对提高，呈上升趋势，数据波动较小，在邻居个数处于80左右得到此算法在同一topk下的相对峰值。

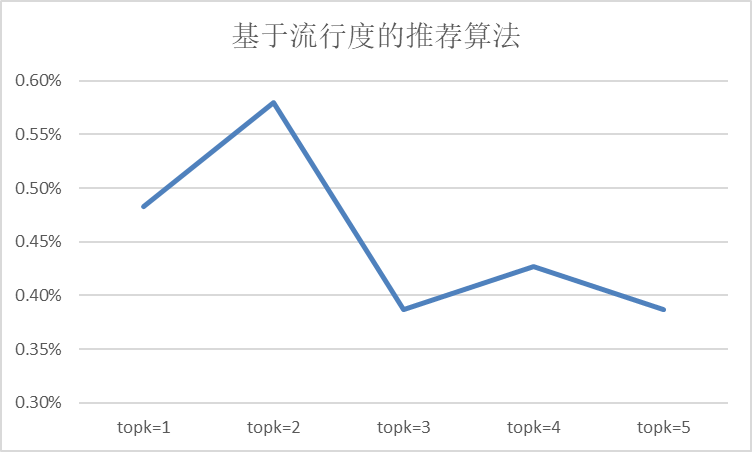


图4基于流行度的推荐算法

在对流行对进行数据分析，数据波动于0.4%—0.6%这一区间内，在topk=2时处于该算法的峰值0.58%。

4 评估情况

4.1 **基础算法的分析**

通过对上述图表数据结论分析以及相关文献查阅，当topk=5时，为本文的最优推荐个数，选取三种指标，对四种推荐算法进行对比择优。

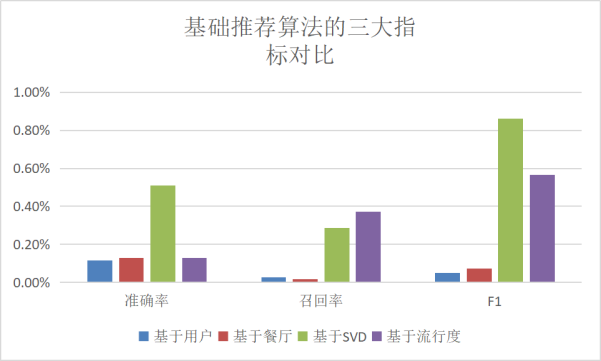


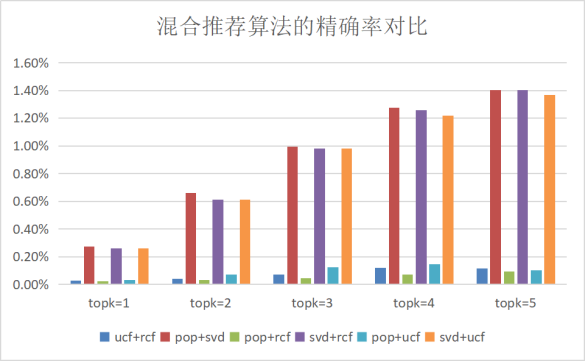
图5 基础算法的三大指标对比

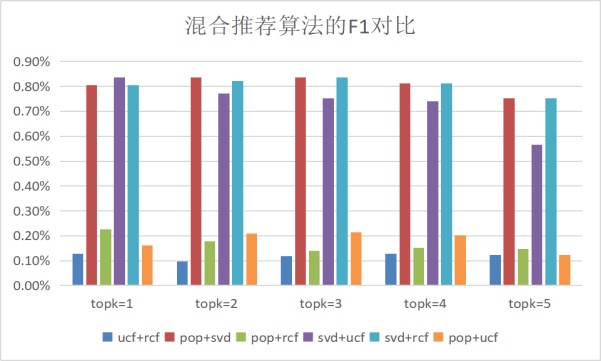
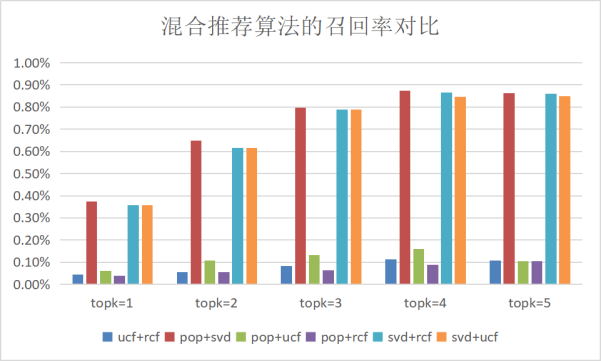
实验数据表明，对精确率的分析，发现在该指标下，基于SVD推荐算法的性能是最好的，在推荐餐厅个数中餐厅用户喜爱餐厅的匹配度较高。对召回率的分析，发现基于流行度和基于SVD的推荐算法的召回率值要高于基于用户和基于餐厅，这进一步说明基于流行度和基于SVD的推荐算法的找到餐厅比较全面。F1指标是对精确度和召回率的调和平均，说明在考虑算法结果既要准确又要全面的平衡点时基于SVD和基于流行度的推荐算法性能好。

通过对上述指标的分析，更倾向于选着基于SVD和基于流行度的推荐算法为推荐系统的核心算法。但是只是简单的考虑到每种推荐算法的推荐准确程度，由于本文数据集过于庞大，结果相似度还是相对较低，基于本次实验进一步进行研究，结合多种算法进一步研究推荐准确程度更高的算法。

4.2 **混合算法的分析**

如下图给出了六种混合推荐算法各项指标下的实验结果，其中ucf为基于用户的协同过滤、rcf为基于餐馆的协同过滤、svd为基于svd的协同过滤、pop为基于流行度的协同过滤。



基于上述：

通过对基础算法的改进与融合，对精确率、召回率、F1的实验结果分析，相比于其他混合算法只有某一方面的提升，基于pop与svd的混合推荐算法无论是在美食餐厅推荐的准确程度、全面方面都有了明显的质量提高，在topk=5时达到了实验的峰值，推荐效果好。

4**结束语**

本文主要研究美食推荐算法。推荐是指由系统主导用户的浏览顺序 ,引导用户发现需要的结果[10]。在电子信息高速发展的21世纪，工作和生活如何达到平衡这一问题在当代社会日益突出。人们对生活要求越来越高，所以，能设计出一款准确率较高，符合用户喜好的美食推荐系统迫在眉睫。它既能分辨真假信息，又能高效解决用户的需求。本文除了研究基础推荐算法，还在此基础上提出了基于SVD和基于流行度(pop)的推荐算法，通过实验测试发现该算法有效的提高推荐的准确程度，后续将继续对该算法进入深入研究，将其推广能更好解决上述问题。

**参考文献**

[1] Bell R，Koren Y，Volinsky C.The bellkor 2008 solution to the Netflix prize[R].2007.

[2] Paterek A.Improving regularized singular value decomposition for collaborative filtering[C]//KDD-Cup and Workshop.[S.l.]：ACM Press，2007.

[3] Lee D D，Seung H S.Learning the parts of objects by non-negative matrix factorization[J].Nature，401：788-791.

[4] 张春丽 基于hadoop的协调过滤算法的研究与实现[D].东华大学，2015

[5] 熊聪聪,邓滢,史艳翠,陶鑫,陈亚瑞.基于协同过滤的美食推荐算法[J].计算机应用研究,2017

[6] 陈清浩. 基于SVD的协同过滤推荐算法研究[D].西南交通大学,2015.

[7]王立才，孟详武，张玉洁． 上下文感知推荐系统［J］． 软件学报，2012

[8]巨星海,周刚.一种基于用户偏好分析和论坛相似度计算的改进LFM推荐算法[J].信息工程大学学报,2021,22(04):433-437+449.

[9]张芳. 基于项目流行度与用户信任度的协同过滤推荐算法研究[D].山东科技大学,2018.DOI:10.27275/d.cnki.gsdku.2018.000964.

[10]许海玲,吴潇,李晓东,阎保平.互联网推荐系统比较研究[J].软件学报,2009,20(02):350-362