

一种融合项目特征和移动用户信任关系的推荐算法*

胡 勋^{1,2}, 孟祥武^{1,2}, 张玉洁^{1,2}, 史艳翠^{1,2}

¹(智能通信软件与多媒体北京市重点实验室(北京邮电大学),北京 100876)

²(北京邮电大学 计算机学院,北京 100876)

通讯作者: 胡勋, E-mail: hu_xun@163.com, http://www.bupt.edu.cn

摘 要: 协同过滤推荐系统中普遍存在评分数据稀疏问题.传统的协同过滤推荐系统中的余弦、Pearson 等方法都是基于共同评分项目来计算用户间的相似度;而在稀疏的评分数据中,用户间共同评分的项目所占比重较小,不能准确地找到偏好相似的用户,从而影响协同过滤推荐的准确度.为了改变基于共同评分项目的用户相似度计算,使用推土机距离(earth mover's distance,简称EMD)实现跨项目的移动用户相似度计算,提出了一种融合项目特征和移动用户信任关系的协同过滤推荐算法.实验结果表明:与余弦、Pearson 方法相比,融合项目特征的用户相似度计算方法能够缓解评分数据稀疏对协同过滤算法的影响.所提出的推荐算法能够提高移动推荐的准确度.

关键词: 推土机距离;推荐系统;协同过滤;混合推荐

中图法分类号: TP311

中文引用格式: 胡勋,孟祥武,张玉洁,史艳翠.一种融合项目特征和移动用户信任关系的推荐算法.软件学报,2014,25(8): 1817-1830. http://www.jos.org.cn/1000-9825/4491.htm

英文引用格式: Hu X, Meng XW, Zhang YJ, Shi YC. Recommendation algorithm combing item features and trust relationship of mobile users. Ruan Jian Xue Bao/Journal of Software, 2014,24(8):1817-1830 (in Chinese). http://www.jos.org.cn/1000-9825/4491.htm

Recommendation Algorithm Combing Item Features and Trust Relationship of Mobile Users

HU Xun^{1,2}, MENG Xiang-Wu^{1,2}, ZHANG Yu-Jie^{1,2}, SHI Yan-Cui^{1,2}

¹(Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia (Beijing University of Posts and Telecommunications), Beijing 100876, China)

²(School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Corresponding author: HU Xun, E-mail: hu_xun@163.com, http://www.bupt.edu.cn

Abstract: The sparsity of user-item ratings is a common problem in collaborative filtering recommender systems. In traditional collaborative filtering recommender systems, similarity of users is often calculated with cosine and Pearson methods based on common ratings. When user-item ratings are sparse, the ratio of common rated items is less, and the accuracy of recommendations will be influenced because users with similar preferences can't be found accurately. To change calculation method of user similarity based on the same rated items, this paper applies EMD (earth mover's distance) to implement cross-item similarity calculation of mobile user and proposes a collaborative filtering recommendation method combining item features and trust relationship of mobile users. The experimental results show that, comparing with cosine and Pearson, user similarity calculation method combining item features can relieve influence of the sparsity of user-item ratings on collaborative filtering recommender systems. And the proposed recommender method can improve accuracy of mobile recommendations.

Key words: earth mover's distance (EMD); recommender system; collaborative filtering; hybrid recommendation

* 基金项目: 国家自然科学基金(60872051); 北京市教育委员会共建项目

收稿时间: 2012-07-04; 修改时间: 2013-03-29; 定稿时间: 2013-09-02; jos 在线出版时间: 2014-05-16

CNKI 网络优先出版: 2014-05-16 11:21, http://www.cnki.net/kcms/doi/10.13328/j.cnki.jos.004491.html

互联网的迅速发展引发了信息过载问题.推荐系统作为解决信息过载问题的重要手段,在电子商务等领域^[1-4]被广泛应用,主要通过发现用户的潜在需求并主动向用户推荐满足其需求的项目,将不相关的项目过滤掉以缓解用户的信息负担.目前,协同过滤算法是推荐系统中应用最广泛的推荐算法,协同过滤算法主要分为两类:基于内存的方法和基于模型的方法^[5].但协同过滤推荐系统中普遍存在评分数据稀疏问题,即用户评价或查看的项目远远小于用户未评价或未查看的项目.由于评分数据的稀疏不能准确地获取用户偏好,影响了协同过滤推荐算法的性能,是协同过滤推荐的一大挑战^[5,6].随着移动通信技术和移动互联网的迅速发展,移动网络服务和信息内容的迅速增长已超出人们所能接受的范围,导致了移动信息的过载.推荐系统作为信息过滤的有效手段,在移动领域中也得到了广泛的关注.由于移动设备的限制(如屏幕小、输入能力差等),移动用户的评价反馈行为变得更不方便,评分数据稀疏,在移动推荐系统中依然是影响移动推荐性能的重要因素.

基于用户的协同过滤推荐算法的基本思想是:通过计算用户偏好的相似性,找到与目标用户相似的其他用户,然后利用相似用户的偏好对目标用户的偏好进行预测并推荐.传统的基于用户的协同过滤推荐算法利用用户对项目的评分,通过余弦、Pearson 等方法来计算用户间的相似度,把不同的项目视为不同的维度,通过用户在相同维度上的评分来计算其相似度.当用户评分数据稀疏时,用户之间的共同评分较少;当根据用户之间的共同评分数据来计算用户之间的相似度时,不能准确地衡量用户之间的相似度,从而影响推荐系统的性能.余弦、Pearson 等方法在计算用户相似度时都忽略了用户在不同项目上的评分对用户间相似度的影响,如某演员主演的动作电影 *A* 与电影 *B*,用户 u_1 对电影 *A* 进行了评分,用户 u_2 对电影 *B* 进行了评分,由于用户 u_1 与 u_2 之间没有共同评分,余弦或 Pearson 等方法不能计算用户之间的相似度,从而认为用户 u_1 与 u_2 之间不相似.但从人的认知来说,这两个用户有一定的相似性,因为他们对相似的两部电影(主演相同而且都是动作片)进行了评分.为了充分利用移动用户现有的、在不同项目上的评分来计算移动用户之间的相似度,本文将图像检索中常用的距离计算方法 EMD^[7,8]引入到移动推荐中,将项目特征融入到用户的相似度计算中,改变传统方法中一对一(只考虑移动用户在相同项目的评分)的相似度计算,实现多对多(考虑移动用户在不同项目上的评分)的相似度计算,在一定程度上缓解评分稀疏对移动用户相似度计算的影响.为避免评分数据稀疏对计算用户相似度的影响,研究者将用户社会关系引入到推荐系统中^[9,10],利用用户社会关系进行推荐.本文将通过移动用户的移动通信数据挖掘移动用户间的信任关系,并与融合项目特征的移动用户相似关系进行混合推荐.

本文第 1 节介绍协同过滤推荐算法的相关工作.第 2 节介绍基于 EMD 的移动推荐算法.第 3 节介绍实验并分析实验结果.第 4 节总结全文并给出下一步的研究方向.

1 相关工作

基于内存的协同过滤推荐算法主要关注用户之间或项目之间的相似关系,基于用户的协同过滤推荐算法首先根据用户的历史数据计算用户之间的相似度,如用户评分数据、浏览数据等.其中,评分数据是推荐系统中常用的用户历史数据.通常,基于评分数据使用余弦、Pearson 等方法来计算用户之间的相似度^[2,9-11],再根据用户相似度选择最相似的用户作为邻居,利用相似邻居的偏好进行预测.余弦方法使用向量空间模型来表示用户的偏好,每个项目作为向量的一个维度,用户对该项目的评分表示用户对项目的偏好权重,用户的相似度根据两个用户在相同项上的评分来计算,如果没有共同评分则无法计算用户相似度.余弦方法没有考虑用户自身的评分尺度,不同用户有不同的评分尺度,用户对同样喜欢的电影评分不同,如用户 u_1 评分为 3 表示喜欢该电影,而用户 u_2 评分为 4 才表示喜欢.为了更准确地计算用户之间的相似度,Pearson、改进余弦等方法考虑了用户自身的评分尺度.推荐系统中普遍存在评分数据稀疏问题^[5],用户评分的项目远小于没有评分的项目.余弦、Pearson 等方法都是基于共同评分项目计算用户相似度,由于评分数据稀疏,用户之间的共同评分项目数量小,不能准确计算用户之间的相似度.文献[12]为准确度量用户之间的相似性,将余弦、Pearson 等方法与 Jaccard 系数(两个用户评分项目交集与两个用户评分项目并集的比)融合,如两个用户对 400 个项目进行评分,只有 2 个共同评分项目且评分都为 4,按余弦等方法,其相似度为 1,但实际上,在极小量项目上的相似度不能真正体现用户之间的相似性,通过 Jaccard 系数能避免用户之间只有小量共同评分造成用户之间的高相似度问题.将共同评分数量作为用

户相似度计算中的因素,能够更准确地度量用户之间的相似性,但上述研究还是基于用户间共同评分项目来计算用户间的相似度.

为了缓解评分稀疏对协同过滤推荐算法的影响,研究者引入评分数据外的其他信息来缓解评分稀疏对用户间相似度计算的影响,但有时这些信息并不容易获取.文献[13]引入用户的人口统计学信息,如年龄、性别、受教育程度等,对用户评分稀疏的用户使用人口统计学信息来计算与其他用户的相似度,避免使用评分数据来计算用户相似度,能够在一定程度上缓解评分数据稀疏的影响.为了避免评分数据稀疏对用户相似度计算的影响,研究者将用户社会关系引入到推荐系统中^[9,10,14-16],利用用户社会关系进行推荐.文献[10]将用户的信任关系与协同过滤推荐结合,利用用户显式给定的信任关系和信任关系的传递性计算用户间的信任度,基于信任关系的强度选择最近邻居,避免了传统协同过滤推荐中的用户相似度计算,从而缓解了用户评分稀疏对推荐性能的影响.文献[14,15]利用信任用户对用户偏好的影响,将矩阵分解模型与用户信任关系结合,利用信任关系的影响特性或传递性将用户和项目映射到相同特征空间,在此特征空间上进行偏好预测.文献[2]在移动服务推荐中,基于用户的移动通信数据分析移动用户之间的信任关系,提出一种基于启发式的基于社会化网络和上下文感知的移动服务推荐方法,将移动社会化网络信息和上下文信息同时引入移动服务推荐系统.上述研究考虑了用户社会化关系对推荐的影响,但没有考虑用户相似关系和项目相似性对推荐的影响.本文考虑项目相似性和移动用户社会化关系对移动推荐的影响,提出一种融合项目特征和移动用户信任关系的推荐算法,利用EMD方法实现融合项目特征的移动用户相似度计算,基于移动用户的相似网络和信任网络构建邻居集合并进行推荐.

2 基于 EMD 的移动推荐算法

2.1 基本定义

定义 1(移动用户-项目评分矩阵). $R=\{r_{u,j}\}$, R 表示移动用户对项目的评分矩阵, $r_{u,j}$ 表示移动用户 u 对项目 j 的评分,评分值通常为 1~5 的整数.

定义 2(项目集合). $I=\{i|i \in L\}$, I 表示项目集合,其中, L 为项目 ID 集合.

定义 3(移动用户集合). $U=\{u_k|k \in M\}$, U 表示移动用户集合,其中, M 表示移动用户的 ID 集合.

定义 4(项目相似度矩阵). $S_I=\{s_{i,j}|i,j \in I\}$, S_I 表示项目之间的相似度矩阵, $s_{i,j} \in [0,1]$ 表示项目 i 与项目 j 的相似度,值越大,表示项目之间相似度越高,0 表示项目之间不相似,1 表示完全相似.

定义 5(项目距离矩阵). $D_I=\{d_{i,j}=1-s_{i,j}|i,j \in I\}$, D_I 表示项目之间的距离矩阵, $d_{i,j}$ 表示为项目 i 与 j 间的距离,取值为 $[0,1]$,值越大,表示项目之间距离大(不相似程度越高),其中,0 表示项目之间完全相似,1 表示项目之间完全不相似.

定义 6(移动用户相似度矩阵). $S_U=\{sim(u_k,u_o)|u_k,u_o \in U\}$, S_U 表示移动用户之间的相似度矩阵, $sim(u_k,u_o)$ 表示移动用户 u_k 与 u_o 的相似度,值越大,表示移动用户之间相似度越高.

2.2 基于EMD的移动用户相似度计算

用户相似度计算是基于用户协同过滤推荐算法的关键,余弦、Pearson 等用户相似度计算方法基于用户在相同评分项目上的偏好进行计算,忽略了项目之间的关联,当用户没有共同评分项目时,无法计算用户之间的相似度.在推荐系统中,项目之间存在一定的相似度,用户在相似项目上的评分也能反映用户之间的相似性.本节介绍将EMD方法与移动推荐系统结合,改变传统协同过滤推荐系统中针对相同评分项目进行用户相似度计算.

定义 7(移动用户的 EMD 距离)^[7]. 设移动用户的项目评分向量为 $V_{u_1} = \{(1, r_{u_1,1}), (2, r_{u_1,2}), \dots, (i, r_{u_1,i})\}$, $V_{u_2} = \{(1, r_{u_2,1}), (2, r_{u_2,2}), \dots, (j, r_{u_2,j})\}$, $D = \{d_{i,j}|i,j \in I\}$ 表示项目 i 与 j 之间的距离, $w_{u_1} = \sum_{i \in RI_{u_1}} r_{u_1,i}$, $w_{u_2} = \sum_{j \in RI_{u_2}} r_{u_2,j}$, 偏好流量 $F = \{f_{i,j}\}$, $f_{i,j}$ 表示移动用户 u_1 在项目 i 和移动用户 u_2 在项目 j 之间的偏好转移量,计算时需要满足公式(1)~公式(4)的约束:

$$f_{i,j} \geq 0, i, j \in I \tag{1}$$

$$\sum_{j \in RI_{u_2}} f_{i,j} \leq w_{u_1,i}, i \in RI_{u_1} \tag{2}$$

$$\sum_{i \in RI_{u_1}} f_{i,j} \leq w_{u_2,j}, j \in RI_{u_2} \tag{3}$$

$$\sum_{i \in RI_{u_1}} \sum_{j \in RI_{u_2}} f_{i,j} = \min(w_{u_1}, w_{u_2}) \tag{4}$$

$$EMD(u_1, u_2) = \frac{1}{\sum_{i \in RI_{u_1}} \sum_{j \in RI_{u_2}} f_{i,j}} \sum_{i \in RI_{u_1}} \sum_{j \in RI_{u_2}} d_{i,j} \times f_{i,j} \tag{5}$$

其中, V_{u_1}, V_{u_2} 表示移动用户对项目的评分向量, $(i, r_{u_k,i})$ 表示移动用户 u_k 对项目 i 的评分为 $r_{u_k,i}$ 即移动用户 u_k 对项目 i 的偏好, RI_{u_1}, RI_{u_2} 表示用户 u_1, u_2 评分的项目集合, w_{u_1}, w_{u_2} 表示移动用户对项目偏好的总和. 公式(1)表示在计算 EMD 距离时, 移动用户可以选择性地转移偏好. 公式(2)表示用户 u_1 在项目 i 上能提供的偏好转移总量不能超过用户 u_1 对项目 i 的偏好. 公式(3)表示移动用户 u_2 在项目 j 上能接收的偏好转移总量不能超过用户 u_2 对项目 j 的偏好. 公式(4)表示移动用户 u_1 与 u_2 在所有项目上的偏好转移总量为 $\sum_{i \in RI_{u_1}} \sum_{j \in RI_{u_2}} f_{i,j}$, 由两个用户中已知偏好总量小的用户决定. 公式(5)表示移动用户间的 EMD 距离, 根据偏好转移总量和转移代价计算, 并将距离归一化.

EMD 方法的计算如图 1 所示.

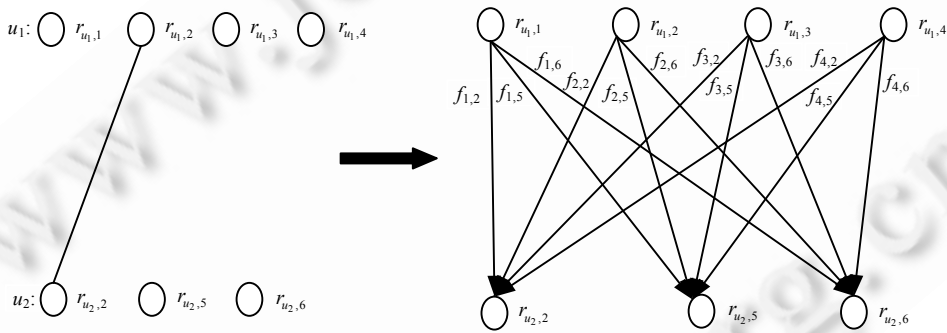


Fig.1 An illustration of EMD computing
图 1 EMD 计算示意图

在推荐系统中, 移动用户偏好可以通过用户对项目的评分向量来表示. 图 1 中, 移动用户 u_1 的评分向量为 $\{(1,4), (2,3), (3,3), (4,5)\}$, 移动用户 u_2 的评分向量为 $\{(2,4), (5,4), (6,3)\}$. 如果使用余弦或 Pearson 等方法来计算图 1 中移动用户之间的相似度, 由于移动用户间只有共同评分项目 2, 相似度的计算只是基于移动用户对项目 2 的评分, 共同评分项目很少, 不能有效地反映用户之间的偏好, 使用余弦或 Pearson 方法得到用户间的相似度不能准确地衡量用户之间的相似关系. 而 EMD 方法的跨项目特性, 能够将用户共同评分以外的其他项目评分融入用户相似度的计算. 它利用项目之间的相似性, 用两个用户的所有项目评分信息来计算两个用户之间的距离. EMD 的计算被视为运输问题来求解, 直观上, EMD 距离的计算可以认为是求解将一堆土填到一系列坑中的最优转移路径^[7]. 在示意图中, 移动用户 u_1 对不同项目的偏好视为不同的堆, 用户 u_2 对不同项目的偏好视为不同的坑. 根据定义, 移动用户 u_1 对项目 1 的偏好是 4, 他能提供的最大转移偏好量是 4, 用户 u_2 对项目 6 的偏好是 3, 他能接收的最大偏好量是 3. 在满足公式(1)~公式(4)的情况下, 移动用户之间的 EMD 距离就是用户之间偏好转移量的最小转移代价, 总的偏好转移量以用户 u_2 的偏好总量来计算, 转移代价是偏好转移代价乘以偏好转移量. 如上例中, 假设项目 1 与项目 5 之间的距离是 0.3, 用户 u_1 和 u_2 在项目 1 和项目 5 之间的偏好转移量是 4, 则其转移代价是 1.2.

最优转移路径决定了最终的转移代价. 在运输问题中, 最优转移路径可以使用不同的方法来求解, 本文使用线性规划的单纯形方法来求解^[7]. 求解运输的最优转移路径时, 通常需要知道源和目的, 资源的流向是从源到目

的.示意图中的EMD计算,其偏好的转移是单向的,是从源用户到目的用户.但是在计算两个用户之间的EMD距离时,如果项目之间距离是对称的,即项目*i*和项目*j*之间的距离与项目*j*和项目*i*之间的距离是相等的,那么EMD距离也是对称的,即移动用户*i*和移动用户*j*的EMD距离与移动用户*j*和移动用户*i*的EMD距离是相等的,源用户和目的用户的选择不会影响两个用户之间的EMD值^[7].本文基于项目的特征向量,使用余弦方法计算项目之间的距离,向量之间的距离是对称的,因此项目之间的距离是对称的.由于项目之间的距离是对称的,源到目的的最优转移路径也是目的用户到源用户的最优转移路径.因此,本文EMD距离计算时,源用户和目的用户的选择不会影响两个用户之间的EMD距离.在求解最优转移路径时,可能存在多条最优转移路径,即它们的转移代价相等而且是最优的,EMD距离计算的是最小转移代价,因此只需要找到一条最优转移路径并计算其最终的转移代价.

在移动推荐系统中,移动用户的评分项目数量不一定相等,移动用户对项目的偏好也不尽相同,移动用户在推荐系统中的评分总量也不一定相等.针对移动推荐系统中移动用户的评分特点,为了将EMD方法有效地应用于移动推荐中,对移动用户的评分进行了预处理.针对评分项目数量不相等的情况,为了减少评分项目数量对EMD计算效率的影响,以评分项目数量小的用户为基准,从另一个用户中选择等量的评分项目,以此计算两个用户的EMD距离.在选择等量的评分项目时,两个用户间的共同评分项目都会被选择,对于共同评分项目外的项目,以基准用户的项目为目标,从另一用户的评分项目中选择与之最相似的项目,以此来构建等量的评分项目集合.针对用户间评分总量不同的情况,如用户 u_1 的项目评分 $V_{u_1} = \{(1,2), (2,3), (3,2)\}$,用户 u_2 对项目的评分 $V_{u_2} = \{(1,4), (2,3), (3,5)\}$,直接使用EMD方法计算用户 u_1 与 u_2 的距离,其值为0.因为项目与自身的距离为0,而最优运输路径将用户 u_1 对项目1~项目3偏好转移到用户 u_2 中对应的项目1~项目3,其转移总量为7,用户 u_1 没有可转移的偏好,因此算法结束.但在相同项目上的转移代价为0,根据公式(5)知两个移动用户间的距离为0,即相似度为1.但实际上,移动用户在不同项目的评分是不同的,其相似度不为1.为了解决此问题,我们在两个用户中构建一个虚拟项目^[8],在评分量小的用户端,该用户对其评分为两个用户间的评分总量差值.在评分量大的用户端,其评分为0.因为是虚拟项目,我们假设虚拟项目与系统其他项目的相似度为0,即与系统其他项目的距离为1.经过预处理后的两个移动用户评分,他们的评分项目数量和评分总量相等.

在计算EMD距离时,需要依赖项目之间的距离即用户偏好在不同项目间转移所需的代价.由定义5可知,项目之间的距离由项目之间的相似度来计算,使用不同相似度方法来计算项目之间的相似度会影响EMD方法中最优路径的选择和用户间的EMD距离.因此,需要准确地度量项目之间的相似度.在推荐系统中,通常只知道用户对项目的评分矩阵,并不知道项目的显式特征(如电影类型、电影导演等),基于用户对项目的评分可以启发式地计算项目之间的相似度^[17],或使用矩阵分解技术挖掘评分矩阵中项目的潜在特征向量,利用项目的潜在特征向量计算项目之间的相似度^[18].启发式方法计算简单、容易理解,但易受评分稀疏的影响,当项目只被很少的用户评价时,由于评分稀疏不能准确而有效地计算与其他项目的相似度.矩阵分解方法对用户和项目进行降维处理挖掘其潜在特征,其方法不易受评分稀疏的影响^[18],能够有效地计算项目之间的相似度.为了确保准确度量项目之间的相似度,本文选择使用矩阵分解方法计算项目的潜在特征向量,并根据项目的潜在特征向量,使用余弦方法计算项目之间的相似度.由于用户评分矩阵只是用户对部分项目的评分,在求解潜在特征向量时常用随机梯度下降方法或交替最小二乘方法来最小化公式(6)得到用户和项目的特征向量^[18]:

$$\min \sum_{(u,i) \in K} (r_{u,i} - q_i^T p_u)^2 + \lambda (\|q_i\|^2 + \|p_u\|^2) \quad (6)$$

其中, $r_{u,i}$ 表示用户*u*对项目*i*的评分, K 表示用户项目对集合, q_i 表示项目*i*的潜在特征向量, p_u 表示用户*u*的潜在特征向量.

EMD方法具有跨项目特性,能够合理利用用户间共同评分项目外的评分项目,实现项目之间多对多的用户相似度计算,充分利用了用户的现有评分.EMD方法最终得到的是用户之间的距离,由定义4和公式(5)可知,用户之间的相似度为

$$\text{sim}(u_1, u_2) = 1 - \text{EMD}(u_1, u_2) \quad (7)$$

EMD 通过寻找源用户与目的用户之间的最优偏好转移路径来计算用户间的距离,但最优转移路径的计算时间复杂度比较高,当计算两个 N 维评分用户的距离时,使用线性规划的单纯形方法来计算时间复杂度是 $O(N^3 \log M)^{[7]}$. 本文主要是将 EMD 方法与推荐系统进行融合,文中使用单纯形方法来计算最优转移路径,没有对 EMD 方法进行优化. 时间复杂度高会影响 EMD 方法的应用,可以通过得到近似 EMD 值来解决时间复杂度高的问题. 文献[19,20]通过小波变换将 EMD 方法的时间复杂度降为线性,以得到近似的 EMD 值.

2.3 移动用户的信任关系

由于信任或朋友等社会关系的相互影响,利用信任或朋友关系来选择最近邻居,从而避免计算移动用户之间的评分相似度,可以缓解评分数据稀疏问题对协同过滤推荐算法的影响^[10]. 移动用户的通信数据在一定程度上反映了移动用户之间的信任关系,在人们日常的通信行为中,通话时间长的用户之间的信任度一般会大于通话时间短的用户之间的信任度,通信频率高的用户之间的信任度一般会大于通信频率低的用户之间的信任度. 通话时长和通信频率都能体现用户之间的信任关系^[21],因此使用公式(8),从通话时长和通信次数(包括通话次数和短信次数)两个方面通过线性方法来融合计算移动用户之间的信任度. 在融合时,一些用户之间的信任度可能受通话时长的影响更大,一些用户之间的信任度可能受通信频率的影响更大,一些用户之间的信任度可能受两者的影响相同,但是在不知道其他信息的情况下,本文假定这两个方面的信息对用户之间的信任关系的影响同等重要,这样的假定是合理的,在一定程度上能够正确反映用户现实生活中通信行为对用户之间信任关系的影响.

$$trust(u_1, u_2) = \alpha \times \frac{time_{u_1, u_2}}{\max_{u \in C_{u_1}} time_{u_1, u}} + (1 - \alpha) \times \frac{f_{u_1, u_2}}{\max_{u \in C_{u_1}} f_{u_1, u}} \quad (8)$$

其中, $trust(u_1, u_2)$ 表示移动用户 u_1 对 u_2 的信任程度; C_{u_1} 表示与移动用户 u_1 有通信行为的移动用户集合; $time_{u_1, u_2}$ 表示移动用户 u_1 与 u_2 的通话时长; $\max_{u \in C_{u_1}} time_{u_1, u}$ 表示移动用户 u_1 与其他移动用户的最大通话时长; f_{u_1, u_2} 表示移动用户 u_1 与 u_2 的通信次数,包括移动用户间的通话次数和短信次数; $\max_{u \in C_{u_1}} f_{u_1, u}$ 表示移动用户 u_1 与其他移动用户的最大通信次数; $\alpha, 1-\alpha$ 表示通话时长和通信次数在信任度计算中的权重,根据上面的分析,在实验中 α 取值为 0.5.

为了进一步缓解评分数据稀疏性对移动协同过滤算法的影响,将移动用户之间的信任关系与移动用户的相似关系进行融合,将在第 2.4.2 节中加以介绍.

2.4 评分预测

2.4.1 基于相似项目的评分预测

基于用户的协同过滤算法根据相似用户的偏好进行预测,通常使用阈值、top- N 等策略选择最相似的用户作为最近邻居集合,根据最近邻居集合,使用预测公式(9)^[5]进行评分预测:

$$r_{u,i}'' = \bar{r}_u + \frac{1}{\sum_{n \in N} sim(u, n)} \sum_{n \in N} sim(u, n) \times (r_{n,i} - \bar{r}_n) \quad (9)$$

其中, N 表示最近邻居集合中对项目 i 进行了评分的邻居用户,基于相似用户对相同项目评分相似的基本思想,不同邻居与移动用户 u 的相似度不同,对移动用户 u 的评分预测影响也不同,通过邻居用户对该项目的评分加权来预测移动用户 u 对项目 i 的评分; $sim(u, n)$ 表示用户 u 与 n 之间的相似度; \bar{r}_n 表示用户 n 评分的均值; $r_{u,i}''$ 表示用户 u 对项目 i 的预测评分. 传统的预测公式根据邻居用户对待预测项目的评分来进行预测. 如果邻居用户中都没有对待预测项目进行评分,在这种情况下,传统预测公式将失去意义,只能根据移动用户 u 的平均评分进行预测. 如预测移动用户 u_1 对项目 i 的评分,移动用户 u_1 的最近邻居是移动用户 u_2 和 u_3 ,移动用户 u_2 和 u_3 都没有对项目 i 进行评分,移动用户 u_3 与 u_1 的相似度为 0.85,而移动用户 u_3 对项目 j 进行了评分,且项目 i 与 j 很相似,其相似度为 0.95,由于最近邻居没有对项目 i 进行评分,将使用移动用户 u_1 的平均评分来预测对项目 i 的评分. 实际上,移动用户 u_3 与 u_1 相似度很高,项目 i 与 j 的相似度也很高,基于相似用户对相似项目的评分相似的思想,

移动用户 u_3 对项目 j 的评分信息对预测移动用户 u_1 对项目 i 的评分是有帮助的.因此,本文提出一种基于相似项目的评分预测公式(10):

$$r_{u,i}'' = \bar{r}_u + \frac{1}{\sum_{s \in S} sim(u,s)} \sum_{s \in S} sim(u,s) \times (r'_{s,i} - \bar{r}_s) \tag{10}$$

公式(10)与公式(9)基本相似,区别是使用不同的策略来处理待预测用户最近邻居用户没有对项目 i 进行评分的情况:公式(9)直接忽略没有评价项目 i 的最近邻居用户,而公式(10)利用这些最近邻居在项目 i 相似项目上的评分来进行预测.在使用公式(10)进行预测时,需要确定与项目 i 相似的项目,本文利用公式(6)中获取的项目潜在特征向量,并使用余弦方法来计算项目之间的相似度.为了确保选择的相似项目与待预测项目 i 的相关性,利用阈值方法来过滤相似度低的相似项目.公式(10)中的 $r'_{s,i}$ 可以通过公式(11)来计算:

$$r'_{s,i} = \begin{cases} \frac{1}{\sum_{j \in S_{s,i}} s_{i,j}} \sum_{j \in S_{s,i}} r_{m,j} \times s_{i,j}, & r_{s,i} = 0 \\ r_{s,i}, & r_{s,i} \neq 0 \end{cases} \tag{11}$$

$r'_{s,i}$ 表示移动用户 s 对与项目 i 相似项目的评分,在不同情况下,根据不同方式来获取.当移动用户 s 没有评价项目 i 时,根据移动用户 s 评价项目中与项目 i 的相似项目集合来计算移动用户 s 对项目 i 的近似评分,其中, $s_{i,j}$ 表示项目 i 与 j 的相似度, $S_{s,i}$ 表示移动用户 s 进行了评分且与项目 i 相似的项目集合.相似项目集合根据设定的项目相似度阈值来选择,将项目相似度大于阈值的项目作为相似项目.对最近邻居用户来说,有时并不能找到满足条件的相似项目,这时的策略将和公式(9)一样将忽略这样的最近邻居.如上述实例,如果将项目相似度阈值设为 0.98,那么用户 u_3 将没有满足条件的相似项目,那么将按公式(9)的策略来进行评分预测.

2.4.2 融合移动用户信任关系的评分预测

社会化网络分析的相关研究表明,社会化网络中相互影响的用户会变得相似^[22,23].在移动用户通信网络中,移动用户通话时间越长、通信越频繁,他们之间相互影响的可能性就越大,在第 2.3 节,通过计算用户之间的信任关系来体现他们之间相互影响的关系.因此,用户之间信任度越高,他们之间相互影响的可能性越大,他们之间的相似度可能越大.移动通信网中,信任关系体现了用户之间的相似度,它是用户相似性的综合体现,并不针对具体项目的偏好.由于评分数据的稀疏性,通过评分数据并不能准确地度量用户之间相似度,而信任关系在一定程度上也体现了用户之间的相似性,本文将移动用户偏好的相似关系与移动用户的信任关系进行融合偏好预测,以缓解由于评分稀疏性造成不能准确计算移动用户之间的相似度,从而不能准确预测用户偏好的问题.在启发式地融合用户相似关系和用户信任关系时,文献[16]比较了几种融合策略,基于此,本文选择混合方法来融合用户相似度和信任度,根据邻居选择策略,分别从相似用户和信任用户中分别选择相应的相似用户和信任用户.由于移动用户之间的相似度和信任度使用不同的计算方法,他们之间的值并不能直接进行比较,如移动用户之间的相似度为 0.8,移动用户之间的信任度为 0.1,在预测评分时并不能表示相似用户的影响比信任用户的影响大,因此,本文将区分考虑两类邻居用户对预测评分的影响,其计算公式如下:

$$r_{u,i}'' = \begin{cases} \bar{r}_u + \beta \times \frac{\sum_{s \in S} sim(u,s) \times (r'_{s,i} - \bar{r}_s)}{\sum_{s \in S} sim(u,s)} + (1 - \beta) \frac{\sum_{t \in T} trust(u,t) \times (r_{t,i} - \bar{r}_t)}{\sum_{t \in T} trust(u,t)}, & T! = \emptyset \text{ 且 } S! = \emptyset \\ \bar{r}_u + \frac{\sum_{s \in S} sim(u,s) \times (r'_{s,i} - \bar{r}_s)}{\sum_{s \in S} sim(u,s)}, & S! = \emptyset, T = \emptyset \\ \bar{r}_u + \frac{\sum_{t \in T} trust(u,t) \times (r_{t,i} - \bar{r}_t)}{\sum_{t \in T} trust(u,t)}, & T! = \emptyset, S = \emptyset \end{cases} \tag{12}$$

其中, S 和 T 分别表示相似用户集合和信任用户集合, $sim(u,s)$ 和 $trust(u,t)$ 分别表示移动用户之间的相似度和信任

度, β 和 $1-\beta$ 表示相似用户和信任用户在预测中所占的权重.参数 β 调整信任关系在用户评分预测中的权重, β 的取值通过交叉验证实验来设定.

2.5 推荐过程描述

输入:移动用户-项目的评分矩阵,移动用户的通信数据.

输出:移动用户的推荐项目列表.

步骤 1. 计算项目的潜在特征向量,使用随机梯度下降算法计算满足公式(6)的项目潜在特征向量.

步骤 2. 计算项目之间的相似度,根据项目潜在的特征向量,使用余弦方法计算项目之间的相似度,得到项目相似度矩阵 S_i ,并根据定义 5 计算项目距离矩阵 D_i .

步骤 3. 计算移动用户之间的 EMD 距离,根据移动用户-项目的评分矩阵和项目距离矩阵 D_i ,根据定义 7 的 EMD 方法,使用线性规划的单纯性方法计算移动用户之间的 EMD 距离,根据公式(7)得到移动用户相似度矩阵 S_u .

步骤 4. 计算移动用户之间的信任权重,根据移动用户的通话时长、通话次数和短信次数,使用公式(8)计算移动用户之间的信任权重.

步骤 5. 为目标用户选择最近邻,根据 top-N 最近邻选择策略,从相似用户和信任用户中选择相应的用户作为最近邻.

步骤 6. 预测目标用户对未评分项目的偏好,根据目标用户的最近邻,使用评分预测公式(12)预测移动用户对未评价项目的偏好.

步骤 7. 构造推荐列表,根据移动用户对未评价项目的预测偏好,选择预测值前 top-N 的项目作为推荐项目推荐给移动用户.

3 实验及结果分析

3.1 数据集及实验环境

目前,没有同时包括移动用户通信数据和移动用户项目评分数据的公开可用数据集,为了构建移动通信网中移动用户的项目评分,本文将使用两个公开的数据集:MIT 的 the reality mining 移动用户通信数据集^[21]和 MovieLens 的电影评分数据集,根据特定规则将这两个数据集进行匹配,得到需要的数据集.

MovieLens 电影评分数据集包括 6 040 个用户对 3 952 部电影的约 100 万条电影评分信息,评分值为 1~5 的整数,其中,电影包含电影类别、电影名称和上映时间,该数据集评分稀疏度为 95.53%.表 1 统计了数据集中不同评分量的用户分布,其中,评分数量都是小于 100 的用户占了 51.2%.MIT 数据集收集了 94 个用户从 2004 年 9 月~2005 年 7 月的通话、短信、蓝牙交互等移动用户通信数据,与该 94 个用户有通信行为的有 9 950 个用户.本文根据 MIT 和 MovieLens 数据集中用户的活跃度进行匹配.在 MIT 中,用户的活跃度指用户主动拨打和发送短信的次数,用户拨打和发送短息的次数越多,表明他越活跃,而与 94 个用户有通信关系的其他用户并没有全部的通信数据,只根据其部分通信记录来统计其活跃度.MovieLens 数据集中,用户的活跃度指用户评分的数量,评分数量越多,表明用户越活跃.由于 MovieLens 中的用户数量少,按活跃度从高到低,从 MIT 数据集中选择与 MovieLens 数据集中等量的用户(6 040 个移动用户),并将两个数据集中的用户按活跃度高低进行匹配,从而得到包含移动用户通信数据和偏好数据的合成数据集.在合成数据集中,只是利用了 MovieLens 中的项目 ID 和用户评分偏好信息.从合成数据集的角度来看,项目 ID 可以视为不同的事物,如移动服务、电影、音乐或餐馆.

为了验证推荐算法的准确性,将合成数据分为训练集和测试集,训练集用来学习或训练推荐方法中的相关参数,测试集用来验证推荐的准确性.为了保证移动用户在训练集和测试集中都有评分数据,按一定的比例将每个移动用户的评分数据随机的分为训练集和测试集.本文实验中按 80:20 的比例将移动用户的评分数据随机地分为训练集和测试集.

Table 1 Distribution of users with different number of ratings in MovieLens dataset

表 1 MovieLens 数据集中不同评分量的用户分布

评分量	10~50	50~100	100~150	150~200	200~250	250~300	≥300
MovieLens 用户数	1 743	1 352	849	507	364	260	758

在 EMD 方法中,需要使用矩阵分解技术获取项目的潜在特征向量,并计算项目间的相似度.本文使用随机梯度下降方法来实现公式(6)对评分矩阵的分解^[14],以求解项目潜在的特征向量.其中,特征向量维度和 λ 分别设置为 20 和 0.06.实验参数根据交叉验证实验获取,即在训练集中进行模型参数学习,并在测试集中进行模型准确度验证,选择在测试集中准确度最好的参数作为实验中的参数设定.通过实验设定参数,在不同的数据集中取值可能不同.

本文实验环境为:Windows XP3 操作系统,2GB 内存,Intel Core(TM) 2 Duo CPU 2.66GHz,实验程序使用 Java 1.6 语言开发,数据库为 Mysql5.0.

3.2 评价指标

为了验证推荐的性能,使用 MAE 和 $P@N$ 评价指标:

- MAE 主要根据用户的预测评分和实际评分衡量评分预测的准确性,其定义如下:

$$MAE = \frac{\left(\sum_{i=1}^N |r_i - p_i| \right)}{N_{test}} \tag{13}$$

其中, r_i 表示用户对项目 i 的实际评分, p_i 表示用户对项目 i 的预测评分, N_{test} 表示测试集中的评分数量.

- $P@N$ 指标对预测项目按预测评分的大小排序,将 top- N 的项目推荐给用户,并与测试集中用户评分前 top- N 的项目比较,计算其准确度,定义如下:

$$P@N = \frac{\text{移动用户top-}N\text{推荐集合中包含测试集中top-}N\text{项目的数量}}{N} \tag{14}$$

3.3 实验结果与分析

3.3.1 与传统相似度方法的比较

该实验主要比较余弦、Pearson 等方法与 EMD 方法在协同过滤推荐中的评分预测准确度,使用公式(9)作为预测公式.实验结果如图 2 和图 3 所示.

由于每个移动用户的评分预测准确度并不相同,实验结果是所有移动用户评价指标的均值.由图 2 可知:在不同数量的最近邻居集合中,EMD 方法的预测准确度在几种方法中是最好的,与余弦和 Pearson 方法相比,最好时准确率分别改善了 5.4%和 5.7%($K=10$).这是因为 EMD 方法在计算用户相似度时考虑了项目间的相似度,通过跨项目的特性,充分利用用户的现有评分,能够更准确地计算用户间的相似度,提高预测的准确度.由图 3 可知:在不同数量的最近邻居集合中,EMD 方法在 $P@5$ 和 $P@10$ 的准确度也是最高的,最高时比余弦和 Pearson 方法的准确度分别提高了 1.2%和 1.1%.由以上分析可知:与余弦和 Pearson 方法相比,EMD 方法能够获得更好的推荐结果.

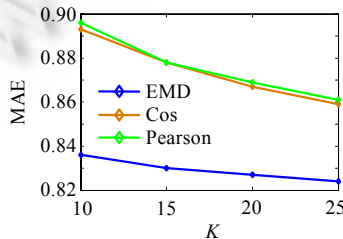


Fig.2 MAE comparison of different similarity computing methods

图 2 各种相似度计算方法的 MAE 比较

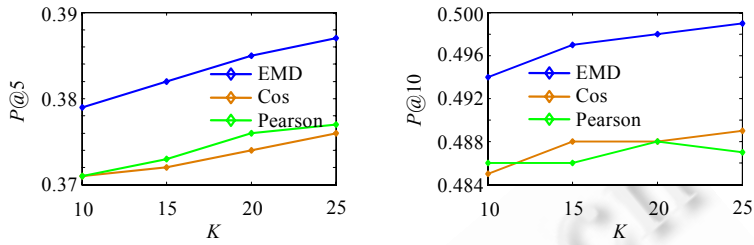


Fig.3 P@N comparison of different similarity computing methods

图 3 各种相似度计算方法的 P@N 比较

3.3.2 基于相似项目的评分预测

该实验主要验证基于相似项目的评分预测公式与传统评分预测公式的预测准确度,在选择相似项目时,为确保选择的相似项目与待预测项目的相关性,我们将从邻居用户的评价项目中选择与待预测项目最相似的并且与其相似度不低于 0.9 的项目作为相似项目,其中,项目相似度阈值 0.9 通过交叉实验来选择.实验结果如图 4 和图 5 所示.

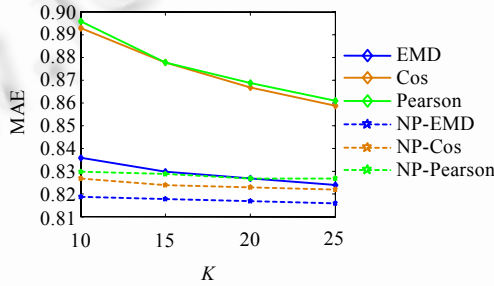


Fig.4 MAE comparison of two different rating prediction formulas

图 4 两种评分预测公式的 MAE 比较

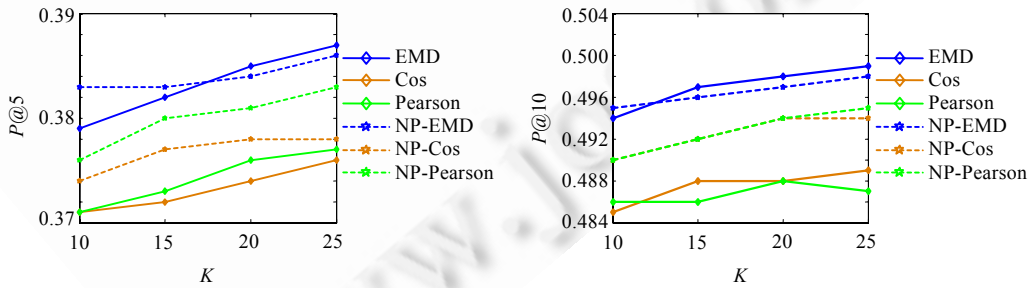


Fig.5 P@N comparison of two different rating prediction formulas

图 5 两种评分预测公式的 P@N 比较

如图 4 所示,在不同数量的最近邻居集合中,余弦、Pearson 和 EMD 方法在融合相似项目的评分预测方法 NP-Cos, NP-Pearson 和 NP-EMD 中, MAE 值低于传统的评分预测公式,与传统评分预测公式的预测准确度相比,实验结果最好时分别提高了 6.6%, 6.5% 和 1.9% (K=10). 在 P@5 和 P@10 指标中,余弦和 Pearson 方法在基于相似项目的评分预测比传统预测公式的准确度高,对 EMD 方法效果并不明显.传统评分预测公式(9)在评分预测时没有考虑相似项目对评分预测的影响,当最近邻居中没有对待预测的项目进行评分时,则认为该邻居用户对预测没有影响.基于相似项目的评分预测公式基于用户对相似项目的评分相似,充分利用最近邻居用户对评分

预测的影响,因此,基于相似项目的评分预测能够有效地提高预测准确度.

3.3.3 稀疏性验证

用户评分稀疏导致用户之间共同评分的电影减少,从而影响了余弦和 Pearson 方法的推荐准确性.第 3.3.1 节的实验在 Movielens 数据集上进行比较,由表 1 可知,其中评分数量都是小于 100 的用户占了 51.2%.为了比较在不同稀疏度下算法的性能,本实验从合成数据集中选择评分量小于 60 的 2 012 个移动用户进行实验,将选择移动用户的评分电影按 80:20 的比例随机生成训练集和测试集.实验结果如图 6 和图 7 所示.

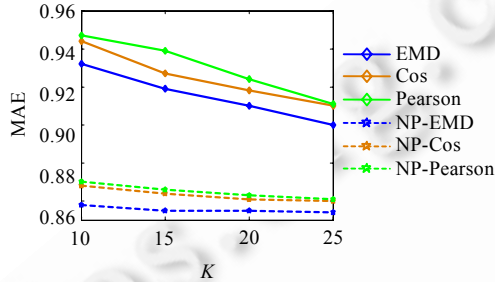


Fig.6 MAE comparison of different similarity methods in sparse rating data

图 6 评分数据稀疏情况下各种相似度方法的 MAE 比较

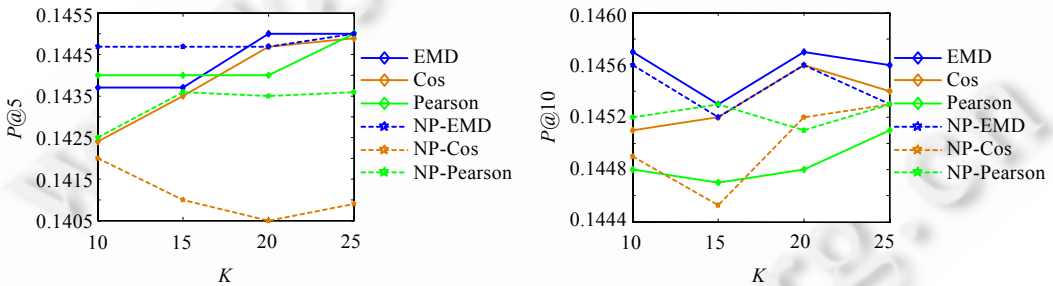


Fig.7 P@N comparison of different similarity methods in sparse rating data

图 7 评分数据稀疏情况下各种相似度方法的 P@N 比较

由图 6 可知:在稀疏数据下,不同数量的最近邻居集合中,EMD 方法的预测更准确,比余弦和 Pearson 方法的预测准确度高 1%左右.使用基于相似项目的评分预测公式比使用传统预测公式的预测准确度高,EMD、余弦和 Pearson 方法的 MAE 值最好时分别降低了 7.1%,7%和 7.5%.在评分稀疏的情况下,基于相似项目的评分预测方法更有效.由图 7 可知:在 P@5 和 P@10 指标中,EMD 方法能达到更好的准确度,但只是微小的改善.在 P@N 性能比较中,基于相似项目的评分预测未能有效地提高准确度.从实验结果可知,EMD 方法在评分数据稀疏性时能够较好地地进行预测推荐.这是因为移动用户评分稀疏时,移动用户间共同评分量减少,跨项目特性使 EMD 方法能够准确地计算移动用户间的相似度,从而提高预测的准确度.

3.3.4 融合移动用户信任关系的推荐

该实验用于验证融合基于移动通信数据的移动用户信任关系和移动用户相似关系的推荐准确性.由于 MIT 数据集中只收集 94 个移动用户的通信数据,无法完全计算其他用户的信任关系,因此推荐准确度只针对这 94 个移动用户的推荐结果进行计算.为了验证融合移动用户信任关系和相似关系对推荐的影响,实验中,我们使用 top-N 邻居选择策略来选择用户的最近邻居数量,其中,最近邻居数量设为 30.从信任用户和相似用户中分别选择最信任的 10 个用户和最相似的 20 个用户、从信任用户中选择 30 个最相似用户和从相似用户中选择最相似的 30 个用户组成不同的邻居集合,基于不同的邻居集合进行实验验证推荐准确度.公式(12)中的 β 参数通过交叉验证实验其值设为 0.8.实验结果如图 8 和图 9 所示.

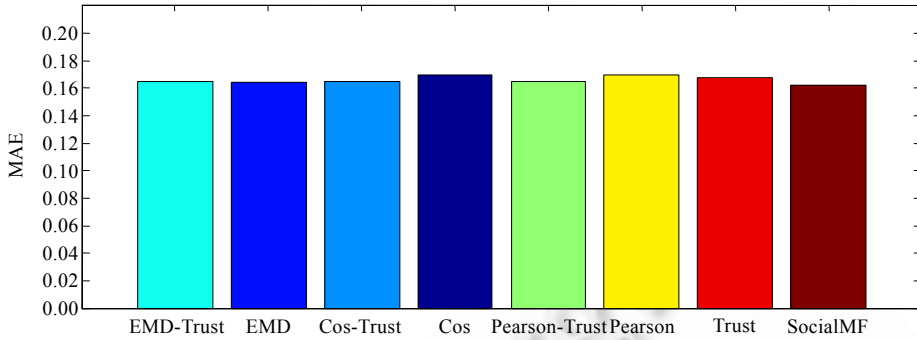
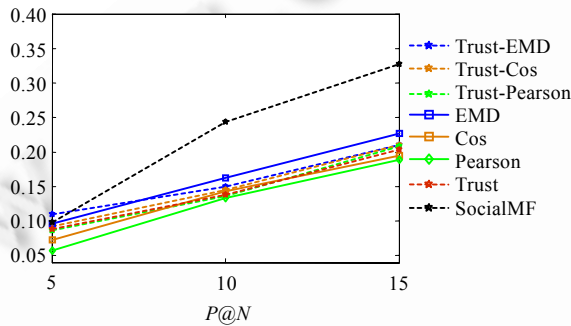


Fig.8 MAE comparison of methods combining trust relationship of mobile users

图 8 融合移动用户信任关系的 MAE 比较

Fig.9 $P@N$ comparison of methods combining trust relationship of mobile users图 9 融合移动用户信任关系的 $P@N$ 比较

由图 8 和图 9 可知:将移动用户的信任关系与余弦相似关系结合(Trust-Cos)进行预测推荐,以及与 Pearson 相似关系结合(Trust-Pearson)进行预测推荐,与完全利用相似关系进行预测推荐相比,MAE 值分别降低了 0.6% 和 0.5%, $P@N$ 值分别绝对提高了 0.3% 左右;与完全利用移动用户的信任关系进行预测推荐相比,MAE 值分别降低了 0.5% 和 0.4%, $P@N$ 值最好时分别提高了 1.4% 和 1.3% ($N=15$)。移动用户信任关系缓解了余弦和 Pearson 相似度计算方法由于评分稀疏对推荐性能的影响。在实验时,我们预先指定了移动用户相似邻居和移动用户信任邻居的数量,可能会影响实验的结果。但在预先指定的情况下,从实验结果可知,本文相似关系和信任关系的融合预测方法提高了移动推荐系统的准确性,表明了本文融合信任关系评分预测方法的有效性,同时也表明可以通过合理融合相似关系和信任关系来提高移动推荐系统的准确性。由于信任用户之间的相互影响,随着时间的推移,信任用户之间的相似度和信任度都会增加。在推荐系统中,由于没有足够的评分信息来计算信任用户之间的相似度,因此,将移动用户的信任关系与相似度进行融合能够提高推荐的准确度。但随着移动用户提供的评分信息越来越多,或使用更有效的方法来计算移动用户之间的偏好相似度,移动用户之间的相似度能够更真实地体现用户之间的相似关系,能够更有效地根据相似关系进行偏好预测。由第 3.3.1 节的实验结果可知,EMD 方法比余弦和 Pearson 能够更准确地计算移动用户之间的相似度。由图 8 和图 9 的实验结果可知:融合信任关系的评分预测方法将信任关系与基于 EMD 的移动用户相似度进行融合预测,并不能有效地提高 EMD 方法的推荐准确度。

实验还比较了基于内存方法和基于模型方法融合移动用户信任关系的推荐准确度。SocialMF^[14]利用矩阵分解技术构建融合社会关系的推荐模型,实验中,模型参数 λ_U, λ_V 和 λ_T 值为 0.1, 0.1 和 5, 特征维度为 5。由于 SocialMF 将预测评分值限定在 $[0, 1]$, 在实验中,使用函数 $f(x)=x/R_{\max}$ 将评分从 $[1\sim 5]$ 映射到 $[0\sim 1]$ 。由图 8 和图 9 可

知,基于模型的 SocialMF 方法比基于内存的方法在 MAE 和 $P@N$ 指标上的准确度更高.主要原因是基于内存的方法只是利用移动用户的局部信息来计算移动用户的信任度和相似度,而基于模型的方法在模型学习时从全局来考虑移动用户评分和信任关系对模型的影响.

4 结 论

本文提出使用跨项目的 EMD 方法计算移动用户间的相似度,改变了传统基于用户的协同过滤推荐方法中利用用户在相同项目上的评分来计算用户相似度.与余弦、Pearson 等相似度计算方法相比,EMD 方法推荐更准确.在评分预测阶段,为了充分利用移动用户的现有评分,将项目相似性融合到评分预测公式,利用最近邻居对相似项目的评分进行预测,能够有效地提高预测的准确度.为了进一步缓解评分数据稀疏性的影响,将相似度与移动用户通信信任关系融合.实验结果表明:有效地将移动用户通信信任关系与移动用户的相似度进行融合,能够提高推荐准确度.本文基于启发式的方法将项目相似性和移动用户信任关系与协同过滤算法融合,进行移动推荐.下一步的研究将对移动用户的相似关系和移动用户社会关系网络中的信任关系在推荐系统中的相互关系及作用进行更深入的研究,把项目的相似性、移动用户的相似性和移动用户的社会关系网络融合到推荐模型中,通过模型来准确地构建它们之间的关系以及对推荐的影响,从而提高移动推荐的准确度.

References:

- [1] Wang LC, Meng XW, Zhang YJ. Context-Aware recommender systems. Ruan Jian Xue Bao/Journal of Software, 2012,23(1):1–20 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4100.htm> [doi: 10.3724/SP.J.1001.2012.04100]
- [2] Wang LC, Meng XW, Zhang YJ. A heuristic approach to social network-based and context-aware mobile services recommendation. Journal of Convergence Information Technology, 2011,6(10):339–346. [doi: 10.4156/jcit.vol6.issue10.43]
- [3] Meng XW, Hu X, Wang LC, Zhang YJ. Mobile recommender systems and their applications. Ruan Jian Xue Bao/Journal of Software, 2013,24(1):91–108 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4292.htm> [doi: 10.3724/SP.J.1001.2013.04292]
- [4] Wang LC, Meng XW, Zhang YJ. A cognitive psychology-based approach to user preferences elicitation for mobile network services. Acta Electronica Sinica, 2011,39(11):2547–2553 (in Chinese with English abstract).
- [5] Adomavicius G, Tuzhilin A. Towards the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. on Knowledge and Data Engineering, 2005,17(6):734–749. [doi: 10.1109/TKDE.2005.99]
- [6] Xu HL, Wu X, Li XD, Yan BP. Comparison study of internet recommendation system. Ruan Jian Xue Bao/Journal of Software, 2009,20(2):350–362 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3388.htm> [doi: 10.3724/SP.J.1001.2009.03388]
- [7] Rubner Y, Tomasi C, Guibas LJ. The earth mover's distance as a metric for image retrieval. Int'l Journal of Computer Vision, 2000, 40(2):99–121. [doi: 10.1109/ICME.2005.1521516]
- [8] Pele O, Werman M. Fast and robust earth mover's distances. In: Proc. of the Int'l Conf. on Computer Vision. Piscataway: IEEE Computer Society, 2009. 460–467. [doi: 10.1109/ICCV.2009.5459199]
- [9] Massa P, Avesani P. Trust-Aware collaborative filtering for recommender systems. In: Proc. of Federated Int'l Conf. on the Move to Meaningful Internet: CoopIS, DOA, ODBASE. Berlin, Heidelberg: Springer-Verlag, 2004. 492–508. [doi: 10.1007/978-3-540-30468-5_31]
- [10] Massa P, Avesani P. Trust-Aware recommender systems. In: Proc. of the ACM Recommender Systems Conf. New York: ACM Press, 2007. 17–24. [doi: 10.1145/1297231.1297235]
- [11] Deshpande M, Karypis G. Item-Based top- N recommendation algorithms. ACM Trans. on Information Systems, 2004,22(1):143–177. [doi: 10.1145/963770.963776]
- [12] Candillier L, Meyer F, Fessant F. Designing specific weighted similarity measures to improve collaborative filtering systems. In: Proc. of ICDM Industrial Conf. on Advances in Data Mining: Medical Applications, E-Commerce, Marketing, and Theoretical Aspects. Berlin, Heidelberg: Springer-Verlag, 2008. 242–255. [doi: 10.1007/978-3-540-70720-2_19]

- [13] Pazzani M. A framework for collaborative, content-based, and demographic filtering. *Artificial Intelligence Review*, 1999,13(5-6): 393–408. [doi: 10.1023/A:1006544522159]
- [14] Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks. In: *Proc. of the ACM Recommender Systems Conf.* New York: ACM Press, 2010. 135–142. [doi: 10.1145/1864708.1864736]
- [15] Ma H, King I, Lyu MR. Learning to recommend with social trust ensemble. In: *Proc. of the SIGIR.* New York: ACM Press, 2009. 203–210. [doi: 10.1145/1571941.1571978]
- [16] Liu F, Lee HJ. Use of social network information to enhance collaborative filtering performance. *Expert Systems with Applications*, 2010,37(7):4772–4778. [doi: 10.1016/j.eswa.2009.12.061]
- [17] Sarwar B, Karypis G, Konstan J, Riedl J. Item-Based collaborative filtering recommendation algorithms. In: *Proc. of the 10th Int'l WWW Conf.* New York: ACM Press, 2001. 285–295. [doi:10.1145/371920.372071]
- [18] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *IEEE Computer*, 2009,42(8):30–37. [doi: 10.1109/MC.2009.263]
- [19] Shirdhonkar S, Jacobs DW. Approximate earth mover's distance in linear time. In: *Proc. of the Computer Vision and Pattern Recognition.* Piscataway: IEEE Computer Society, 2008. 1–8. [doi: 10.1109/CVPR.2008.4587662]
- [20] Applegate D, Dasu T, Krishnan S, Urbanek S. Unsupervised clustering of multidimensional distributions using earth mover distance. In: *Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining.* New York: ACM Press, 2011. 636–644. [doi: 10.1145/2020408.2020508]
- [21] Eagle N, Pentland A, Lazer D. Inferring friendship network structure by using mobile phone data. *Proc. of the National Academy of Sciences (PNAS)*, 2009,106(36):15274–15278. [doi: 10.1073/pnas.0900282106]
- [22] Marsden PV, Friedkin NE. Network studies of social influence. *Sociological Methods and Research*, 1993,22(1):127–151. [doi: 10.4135/9781452243528]
- [23] Crandall D, Cosley D, Huttenlocher D, Kleinberg J, Suri S. Feedback effects between similarity and social influence in online communities. In: *Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining.* New York: ACM Press, 2008. 160–168. [doi: 10.1145/1401890.1401914]

附中文参考文献:

- [1] 王立才,孟祥武,张玉洁.上下文感知推荐系统.软件学报,2012,23(1):1–20. <http://www.jos.org.cn/1000-9825/4100.htm> [doi: 10.3724/SP.J.1001.2012.04100]
- [3] 孟祥武,胡勋,王立才,张玉洁.移动推荐系统及其应用.软件学报,2013,24(1):91–108. <http://www.jos.org.cn/1000-9825/4292.htm> [doi: 10.3724/SP.J.1001.2013.04292]
- [4] 王立才,孟祥武,张玉洁.移动网络服务中基于认知心理学的用户偏好提取方法.电子学报,2011,39(11):2547–2553.
- [6] 许海玲,吴潇,李晓东,阎保平.互联网推荐系统比较研究.软件学报,2009,20(2):350–362. <http://www.jos.org.cn/1000-9825/3388.htm> [doi: 10.3724/SP.J.1001.2009.03388]



胡勋(1985—),男,广西桂林人,博士,主要研究领域为社会化网络分析,推荐系统.
E-mail: hu_xun@163.com



张玉洁(1969—),女,讲师,主要研究领域为智能信息处理,通信软件,网络服务.
E-mail: zhangyj@bupt.edu.cn



孟祥武(1966—),男,博士,教授,博士生导师,CCF高级会员,主要研究领域为网络服务,通信软件,人工智能.
E-mail: mengxw@bupt.edu.cn



史艳翠(1982—),女,博士,主要研究领域为用户偏好自适应,智能信息处理.
E-mail: shi_yancui@126.com