

移动通信网中基于用户社会化关系挖掘的协同过滤算法

黄武汉* 孟祥武 王立才

(北京邮电大学智能通信软件与多媒体北京市重点实验室 北京 100876)

(北京邮电大学计算机学院 北京 100876)

摘要: 该文面对移动通信网中个性化服务推荐问题, 结合社会化网络分析方法提出一种基于移动用户社会化关系挖掘的协同过滤算法。利用移动通信网中所形成社会化网络, 预测潜在的社会化网络关系, 并按关系紧密程度找到相似用户; 然后结合基于用户评分相似度计算发现的最近邻用户, 找到最相似的用户集合, 进行移动用户偏好预测和推荐, 有效地缓解数据稀疏性。仿真数据集和公开数据集实验表明了该算法在预测移动用户偏好和提高推荐精确度方面的可行性和有效性。

关键词: 移动通信网; 移动社会化网络; 偏好预测; 协同过滤

中图分类号: TN915

文献标识码: A

文章编号: 1009-5896(2011)12-3002-06

DOI: 10.3724/SP.J.1146.2011.00364

A Collaborative Filtering Algorithm Based on Users' Social Relationship Mining in Mobile Communication Network

Huang Wu-han Meng Xiang-wu Wang Li-cai

(Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia,

Beijing University of Posts and Telecommunications, Beijing 100876, China)

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

Abstract: To solve the personalized information content and services recommended issues in mobile communication network, a collaborative filtering algorithm based on users' social relationship mining in mobile communication network is proposed with the social network analysis method. It effectively mitigates data sparsely for preferred forecasting and recommendations by using the information of social networks obtained from predicting the potential relationship between social networks and the most similar set of users according to the degree of similar relationship between the users. The algorithm is proved more accurate and feasible in the experiments by using the public data sets and the simulated data sets.

Key words: Mobile communication network; Mobile social network; Preference prediction; Collaborative filtering

1 引言

随着移动通信技术的飞速发展, 移动设备逐渐成为人们获取信息的主要平台之一。与传统互联网终端相比, 移动设备拥有移动性强、上下文感知能力强、可携带性好、入网方便等特点, 使得人们可以随时随地获取移动网络服务和信息内容^[1]。然而, 随着移动多媒体技术和移动信息承载、传输能力的提升, 尤其是移动社交网络的兴起, 大量丰富多彩的移动服务和信息内容日益涌现, 为人们带来严重的移动信息过载问题。此外, 移动网络资源有限,

移动设备也存在输入输出能力弱、电池续航能力弱等缺点, 如何从浩瀚的移动信息海洋中发现用户真正感兴趣的内容, 以提升移动用户的个性化服务体验, 成为面向移动通信网络的个性化服务领域亟待解决的技术难题。传统推荐系统在解决互联网信息过载问题方面取得较大进展, 成为辅助人们信息过滤和决策的主要工具之一。由于移动通信网络与传统桌面互联网相比, 有自身的特点, 而且移动信息传递也颇受上下文信息、移动社会化网络的影响, 所以传统推荐系统的各种模型和算法并不能完全适用于移动通信网络中的数据建模与推荐。目前, 国内外研究人员也开始研究如何将传统推荐系统的模型和算法应用于移动通信网络环境, 并取得一定进展^[2]。

2011-04-17 收到, 2011-07-20 改回

国家自然科学基金(60872051), 中央高校基本科研业务费(2009RC0203)和北京市教育委员会共建项目资助课题

*通信作者: 黄武汉 wuhan_05406@163.com

本文面向移动通信网络中的信息过载问题, 利用移动通信行为构建初步的移动社会化网络, 并深入挖掘潜在的用户社会化关系, 然后将其应用于协同过滤算法, 帮助移动用户发现与其兴趣最相似的近邻用户, 并进行移动用户偏好预测, 最终提出一种移动通信网中基于用户社会化关系挖掘的改进协同过滤算法。本文第 2 节介绍了协同过滤算法、相似性计算方法、移动推荐系统等方面的工作; 第 3 节提出移动通信网中基于用户社会化关系挖掘的协同过滤算法; 第 4 节给出仿真实验分析和利用公开数据集展开的实验, 结果表明了该方法的有效性和合理性; 最后给出结论。

2 相关工作

推荐系统^[3]作为缓解“信息过载”的重要手段, 已得到广泛研究。从信息过滤的角度, 可以将其分为: 协同过滤、基于内容的过滤和混合式过滤^[3]。其中, 协同过滤源于“集体智慧”的思想, 首先根据部分用户对项目的偏好信息或者评分信息计算用户之间或者项目之间的相似性, 然后利用相似用户或者相似项目的用户偏好来预测未知的潜在用户偏好, 最后按照偏好排序结果生成推荐。

协同过滤中的相似性计算方法主要有 3 种。第 1 种为余弦相似度公式^[4], 定义如下:

$$\text{sim}(x, y) = \begin{cases} \frac{\sum_{s \in S_{xy}} r_{x,s} \cdot r_{y,s}}{\sqrt{\sum_{s \in S_{xy}} r_{x,s}^2 \cdot \sum_{s \in S_{xy}} r_{y,s}^2}}, & |S_{xy}| \geq 2 \\ 0, & |S_{xy}| < 2 \end{cases} \quad (1)$$

其中 $\text{sim}(x, y)$ 表示用户 x 与 y 的相似度, $r_{x,s}$ 表示用户 x 对项目 s 的偏好度, S_{xy} 表示 x 与 y 共同评分的项目集合。

第 2 种为 Pearson 相关系数计算公式^[5], \bar{r}_x 表示用户 x 对已评分项目的平均评分值, 定义如下:

$$\text{sim}(x, y) = \begin{cases} \frac{\sum_{s \in S_{xy}} (r_{x,s} - \bar{r}_x)(r_{y,s} - \bar{r}_y)}{\sqrt{\sum_{s \in S_{xy}} (r_{x,s} - \bar{r}_x)^2 \times \sum_{s \in S_{xy}} (r_{y,s} - \bar{r}_y)^2}}, & |S_{xy}| > 2 \text{ 且 } H \neq 0 \\ 0, & |S_{xy}| \leq 2 \text{ 或 } H = 0 \end{cases} \quad (2)$$

其中 $H = \sqrt{\sum_{s \in S_{xy}} (r_{x,s} - \bar{r}_x)^2 \times \sum_{s \in S_{xy}} (r_{y,s} - \bar{r}_y)^2}$ 。

第 3 种算法为修正的 Pearson 相关系数计算公式^[5], r_{mid} 表示修正中值, 定义如下:

$$\text{sim}(x, y) = \begin{cases} \frac{\sum_{s \in S_{xy}} (r_{x,s} - r_{\text{mid}})(r_{y,s} - r_{\text{mid}})}{\sqrt{\sum_{s \in S_{xy}} (r_{x,s} - r_{\text{mid}})^2 \times \sum_{s \in S_{xy}} (r_{y,s} - r_{\text{mid}})^2}}, & |S_{xy}| > 2 \text{ 且 } H_{\text{mid}} \neq 0 \\ 0, & |S_{xy}| \leq 2 \text{ 或 } H_{\text{mid}} = 0 \end{cases} \quad (3)$$

其中 $H_{\text{mid}} = \sqrt{\sum_{s \in S_{xy}} (r_{x,s} - r_{\text{mid}})^2 \times \sum_{s \in S_{xy}} (r_{y,s} - r_{\text{mid}})^2}$
 $= 0$ 。

近年来, 研究人员针对移动通信网络环境和移动用户的特点, 提出“移动推荐系统”的概念。文献[6]提出可以通过分析移动通信网下的服务内容, 并结合用户使用的反馈信息等, 为用户推荐其更喜欢的服务。文献[7]则认为要结合移动终端的可移动性特点, 利用一些便携式的传感器设备, 捕获用户在日常服务体验中的直接情感表达信息, 从而为用户提供其所喜爱的服务。相对互联网而言, 针对移动通信网中以社会化推荐方式进行用户评分预测的研究相对较少。文献[8]提出了一种基于社会化网络的混合推荐算法, 该算法将传统的协同过滤与社会化网络中的人际关系结合, 但并未陈述如何衡量社会化网络中的信任关系, 而且并没有考虑如何利用移动通信行为构建社会化网络关系及其进一步挖掘。文献[9]直接以互联网下的邮件系统为例子, 介绍了如何利用用户日常的邮件交互行为建立一个互联网下的社会化网络, 以此来进行偏好预测, 同样未考虑移动通信网络环境下的移动社会化网络与推荐系统之间的结合。

本文针对移动通信网中的个性化推荐问题, 提出了一种基于移动用户社会化关系挖掘的协同过滤算法。首先对余弦相似度公式进行了改进, 使得共同评分项目较少的用户间也能进行相似度计算, 扩大推荐数据源; 其次, 通过挖掘移动社会化网络中存在的直接及其潜在关系, 获得与用户喜好密切相关的用户数据, 将这些重要的用户偏好信息作为填充初始推荐数据源的数据之一进行评分预测, 在一定程度上缓解了数据稀疏性问题。

3 算法设计

3.1 建立原始数据模型

移动通信网中存在着大量的用户交互数据, 主要包括电话、短信等通信记录及用户使用移动通信网服务的一些评价信息。这些信息对于分析移动网络用户需求和构建移动社会化网络具有重要价值。本文建立的数据模型包括 4 个数据集:

(1)用户集合,即所有移动通信网用户集合,使用 U 表示。

(2)移动网络服务集合,即需要处理的所有移动网络服务集合,使用 S 表示。

(3)用户社会化关系矩阵,即集合 U 中的用户间的交互行为,这种行为是双向的,本文使用 \mathbf{F} 来表示最初的通信关系矩阵,对这个矩阵中的数据本文将按一定的规则进行挖掘,计算出好友关系矩阵 \mathbf{F} , \mathbf{F} 是一个不对称矩阵,即用户 x 认为用户 y 是可以信任的好友并不代表 y 也认为 x 是可以信任的,在交互行为中,表现为 x 联系了 y ,那就表示 x 信任 y 。

(4)用户对服务的评分矩阵,即集合 U 中用户对集合 S 中服务的评分矩阵,这些数据来源于用户对所使用业务的反馈、评价等记录。如用户连续订阅了某个业务,则认为该用户喜欢这个业务;在一段时间内订阅,则表明用户对这个业务感兴趣等,结合用户的使用反馈及网上评分来修正这些数值,就可以挖掘出原始的评分矩阵,使用 \mathbf{R} 表示。

在上述两个矩阵中,用 $f_{x \rightarrow y}$ 表示 \mathbf{F} 中存在用户 x 信任 y 的记录, $r_{x,s}$ 表示用户 x 对于服务 s 的评分值。

3.2 好友关系挖掘

将 \mathbf{F} 中的短信与通话数据进行如下处理,来确定 \mathbf{F} 。

步骤 1 将数据中的通信主体抽取出来,判断是否是所需要预测的目标,如果不是,则删除数据,形成矩阵 \mathbf{F}_1 ,这一步是为了删除一些广告及网外用户号码的无用数据。

步骤 2 以用户 x 为例,将 \mathbf{F}_1 中由 x 充当发送方的短信数据提取出,统计短信总数 c 及各个充当接受方的用户 i 和对应短信数 c_i ;将 \mathbf{F}_1 中由 x 充当主叫方的通话数据提取出,统计用户 x 通话总时间 t ,通话总次数 q 及各个充当被叫方的用户 i 与用户 x 通话的总时间 t_i ,通话总次数 q_i 。定义 p 为 x 对用户 i 的信任程度,则

$$p = \frac{c_i}{c} + \frac{t_i}{t} + \frac{q_i}{q} \quad (4)$$

本文认为,在 \mathbf{F}_1 中的所有通信组合中, x 仅与其通信对象中的80%用户是好友关系,所以取 p 最高的80%用户,将这些用户定义为 x 信任的对象即好友,并将信任程度 p 作为数值加入矩阵 \mathbf{F}_2 中。

步骤 3 对于 U 中的每一个用户,都执行步骤2,并把好友关系及信任程度 p 写入矩阵 \mathbf{F}_2 中,则 \mathbf{F}_2 就是从通信记录中挖掘出的直接好友关系矩阵。

由通信记录确定的直接好友关系比较稀疏,不利于推荐,而且现实中,也不是所有的好友间都有

通信行为,因而从通信记录中无法全面地获取用户间的好友关系,所以需要通过矩阵 \mathbf{F}_2 挖掘更多有用的信息。对于任意两个好友而言,他们的好友集合或多或少会存在交集。用 f_x 表示 x 的好友集合,如果 x 与 y 是好友关系,则 f_x 与 f_y 将存在交集,基于此本文认为:当 $f_x \cap f_y$ 中的元素在 f_x 中的比例达到某个阈值 α (如50%)时, x 信任 y ,即 y 是 x 的好友。所以进一步做如下处理:

步骤 4 以用户 x 为例,将 \mathbf{F}_2 中的 f_x 取出,与 U 中 x 的非好友关系用户在 \mathbf{F}_2 中的好友集合进行比较,用 f_y 表示 y 的好友集合,当满足条件式(5)时,将 y 定义为 x 的好友。

$$a = \frac{|f_x \cap f_y|}{|f_x|} \geq \alpha \quad (5)$$

步骤 5 由于集合 $f_x \cap f_y$ 确定了 y 与 x 的好友关系,因此 x 对 y 的信任程度也应由此集合确定,用 \bar{p} 表示集合 x 对 $f_x \cap f_y$ 所有用户信任程度的平均值,即

$$\bar{p} = \frac{\sum_{i \in (f_x \cap f_y)} p_i}{|f_x \cap f_y|} \quad (6)$$

其中 p_i 表示 x 对用户 i 的信任程度。使用式(7)计算 p_y :

$$p_y = \bar{p} \cdot a \quad (7)$$

将好友关系及对应的信任程度 p 写入矩阵 \mathbf{F}_3 中。显然,这里的信任程度只是对以单个用户为中心的社交网络关系的描述,在全局网络中不具有可比性。

步骤 6 对 \mathbf{F}_2 中每一个有好友的用户都重复步骤4,步骤5,将这种间接好友关系及信任程度写入矩阵 \mathbf{F}_3 。

步骤 7 由 \mathbf{F}_2 与 \mathbf{F}_3 最终确定好友关系矩阵 \mathbf{F} ,如式(8)所示,其中 \mathbf{F}_2 代表直接好友关系矩阵, \mathbf{F}_3 代表间接好友关系矩阵。

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_2 & 0 \\ 0 & \mathbf{F}_3 \end{bmatrix} \quad (8)$$

3.3 基于评分数据的用户相似度计算

对于 U 中的任意用户组合 x 与 y ,用其共同评分项目来计算他们之间基于业务评分的相似度,从而得到相似度集合SM。

$$SM = \{\text{sim}(x, y) \mid x \in U \text{ and } y \in U \text{ and } x \neq y\} \quad (9)$$

其中 $\text{sim}(x, y)$ 表示 x 与 y 的相似度,用 r_{\max} 表示 \mathbf{R} 中所定义的最大评分值, r_{\min} 表示最小评分值,相似度计算公式如式(10)所示。

$$\text{sim}(x, y) = \begin{cases} \frac{\sum_{s \in S_{xy}} r_{x,s} \cdot r_{y,s}}{\sqrt{\sum_{s \in S_{xy}} r_{x,s}^2 \cdot \sum_{s \in S_{xy}} r_{y,s}^2}}, & |S_{xy}| > 1 \\ \frac{|r_x - r_y|}{r_{\max} - r_{\min}}, & |S_{xy}| = 1 \\ 0, & |S_{xy}| = 0 \end{cases} \quad (10)$$

3.4 用户偏好预测

用 SM_x 表示 x 与集合 U 中其他用户的相似度集合, m_x 表示 x 与好友的相似度集合, 用 n_x 表示 x 与非好友的相似度集合, 其中 m_x 中的各个用户与 x 还存在着通过社会化关系而得到的信任程度关系, 显然,

$$SM_x = m_x \cup n_x \quad (11)$$

数据选取方式可以根据 3 种不同方法进行。

方法 1: 选取 SM_x 中相似度大小为前 50 的用户数据。

方法 2: 选取 m_x 中相似度大小在前 $t\%$ 的用户数据。

方法 3: 选取 n_x 中相似度大小为前 50 的用户数据与 m_x 中相似度大小在前 $t\%$ 的用户数据。

第 1 种是传统的协同过滤算法中的数据选取方式(Traditional Data Selection, TDS); 第 2 种是建立在好友关系的数据选择方式(Data Selection Based on Friends, DSBF); 第 3 种是本文所提出的基于用户社会化关系挖掘的混合数据选择方式(Data Selection Based on the Combination of Tradition and Friends, DSBCTF)。

将通过社会化关系挖掘出的好友关系及其信任程度与传统的基于项目评分的相似度结合, 进行评分预测, 具体如式(12)。

$$r_{x,s} = \begin{cases} \bar{r}_x + k \sum_{y \in C} ((\text{sim}(x, y) + p_y) \times (r_{y,s} - \bar{r}_y)), & \bar{r}_x \neq 0 \\ k \sum_{y \in C} ((\text{sim}(x, y) + p_y) \times r_{y,s}), & \bar{r}_x = 0 \end{cases} \quad (12)$$

$$\text{其中 } k = \frac{1}{\left(\sum_{y \in C} \text{sim}(x, y) + \sum_{y \in f_x} p_y \right)}。$$

与传统协同过滤算法相比, 本文算法增加了通过社会化关系挖掘出的好友关系及其信任程度对推荐结果的影响, 而且当预测用户无评分项目即 $\bar{r}_x = 0$ 时, 采用与预测用户及其相关用户的其他间接评分无关的预测公式, 使得推荐结果更加符合实际。公式中 C 表示上述提到的 3 种不同的用户集合,

\bar{r}_x 表示 x 已知评分的平均值, p_y 表示 x 对 y 的信任程度, 当 y 不是 x 好友时: $p_y = 0$ 。对于 3 种数据选取方式, 本文将用实验的方法确定最优解; 其中 $t\%$ 的取值, 将在 4.2 节中确定。通过计算, 便可以得到 x 对未知评分业务的评分值。若无法预测, 将用 0 代替。

4 实验

本文实验环境为: 2 GB 内存, 2.93 GHz 双核 CPU, Windows XP2 操作系统, Java1.6 开发语言, Eclipse3.5 集成环境, Mysql5.0 数据库。实验将分成两个部分: (1)使用仿真数据集展开实验表明本算法的可行性和有效性; (2)使用公开数据集来研究相关参数的选择方法以及与其他基准算法的性能比较。

4.1 基于模拟数据集的实验

4.1.1 模拟数据集介绍 实验所需要生成的模拟数据包括移动用户列表, 移动业务列表, 移动用户的关系列表和移动用户对移动业务的原始评分列表, 其中移动用户的关系列表只是直接的好友关系, 实验中, 将根据这个列表进一步挖掘潜在的社会化关系。为了使实验更加简便, 直接规定: 对单个用户而言, 他对所有好友的信任程度相等, 并且等于直接参与评分预测的好友数的倒数。

这里用实际的 900 个移动手机号码代表用户, 将用户分成 3 个集合, 假定每个集合中的用户在业务评分上相似, 不同集合中的用户业务评分差异较大; 移动业务数为 90, 这些业务分为 3 个集合, 分别代表电话套餐、短信套餐和 GPRS 套餐业务。另外随机确定一些好友关系组合, 其中每个用户平均有 10-30 个好友。这些好友分布在不同的 3 个集合中, 本集合成员(自己除外)占有 80%的比例, 其余两个集合成员占有 20%的比例, 形成最初的用户关系矩阵。

假定 3 个集合的用户对 3 种不同的套餐业务评分差异很大并预先规定了这些评分矩阵的不同取值范围, 具体如表 1 所示, 表中 U_S1 表示用户对电话套餐业务集合的评分值范围, U_S2 表示用户对短信套餐业务集合的评分值范围, U_S3 表示用户对 GPRS 套餐业务集合的评分值范围。根据这个规则和设定的评分覆盖率(这里为 30%), 形成原始评分矩阵。

4.1.2 评价方法 本文使用两种度量标准来衡量实验结果: 首先借鉴信息检索领域广泛使用的查准率^[10], 查准率(Precise), 又称“准确率”, 用来衡量系统的查准率。定义如下:

表1 评分范围限定

用户组	U_S1	U_S2	U_S3
1	[1,4]	[4,7]	[7,10]
2	[4,7]	[7,10]	[1,4]
3	[7,10]	[1,4]	[4,7]

$$P = T_h / T \quad (13)$$

其中 T_h 表示可预测结果中,符合表1所述规则的数据量, T 表示可预测结果的总数据量。

然后,使用 MAE^[10]公式的另一种形式的绝对误差率 $|\bar{E}|$,来评价实验结果,定义如下:

$$|\bar{E}| = \frac{\sum_{i=1}^N |t_i - r_i|}{N(r_{\max} - r_{\min})} \quad (14)$$

其中 N 表示可预测数据量, r_i 表示预测结果中不符合表1的评分值, t_i 表示与 r_i 对应的在表1中应该所处的范围的临界值。 r_{\max} 与 r_{\min} 表示数据集中的最大与最小评分值,本实验中 $r_{\max} = 10$, $r_{\min} = 0$ 。

4.1.3 实验描述及结果分析 首先我们从模拟生成的初始好友矩阵中挖掘出更多的社会化关系,其中间接好友关系挖掘阈值 α 定为 50%。而后将数据作为输入,代入本文所提的 DSBCTF 数据选择算法中进行评分预测,其中 $t\%$ 定为 70%,在预测评分阶段,数据集中的好友信任程度,由参与计算的好友数的倒数确定。

经过计算,将得到了一个预测的评分表。利用 4.1.2 节中的评价标准衡量实验结果,结果显示:本文算法的准确率 $P = 80.74\%$,绝对误差率 $|\bar{E}| = 5.56\%$ 。

处于同一个社交网络的用户,由于在生活中频繁的交流,使得他们的业务喜好相近。通过周围好友的偏好数据,能够更加准确且合理的预测出用户偏好。试验结果表明,本文所提出的移动网络用户社会化关系挖掘算法在实践中是可行的,理论上通过对用户社会化关系的挖掘可以使推荐结果更加准确。

4.2 基于公开数据集的实验

4.2.1 公开数据集介绍 公开数据集使用 Filmtipset^[11], Filmtipset 是瑞士最大的电影推荐社区,它拥有超过 90000 位注册用户和两千万以上的电影评分数据。这个数据集包含:用于 Weekly Recommendation 的训练集、用于圣诞周推荐任务测试集和评测集、用于奥斯卡周推荐任务的测试集和评测集;用于社会化推荐任务的训练集、测试集

和评测集。本文选择其中有关 1000 个用户与 500 部电影的数据集合,并选取了这 1000 个用户的关系矩阵和 1000 个用户对 500 部电影的评分集。对于这些评分数据按照不同的训练集和测试集比划分,形成 15 组测试数据,每组测试数据包含 1 个训练集和 1 个测试集。

4.2.2 评价方法 本实验使用 4.1.2 节中介绍的绝对误差率 $|\bar{E}|$ 来评价试验结果,对于式(17)中的参数, N 表示预测结果数据量, r_i 表示由训练集中数据经本文算法预测得到的评分值, t_i 表示与 r_i 对应的测试集评分值, r_{\max} 与 r_{\min} 表示数据集中的最大与最小评分值,本实验中 $r_{\max} = 10$, $r_{\min} = 0$ 。

4.2.3 实验描述及结果分析

(1)基于参数 $t\%$ 的交叉实验 使用本文提出的 DSBCTF 数据选取方式,将 4.2.2 节中的 15 组训练集数据随机抽取 10 组分别代入本文的算法进行计算,使用改进的余弦相似度计算方法,间接好友关系挖掘阈值 α 定为 50%,当好友选取比例分别为 20%, 50%和 70%时, $|\bar{E}|$ 值比较如图 1 所示。

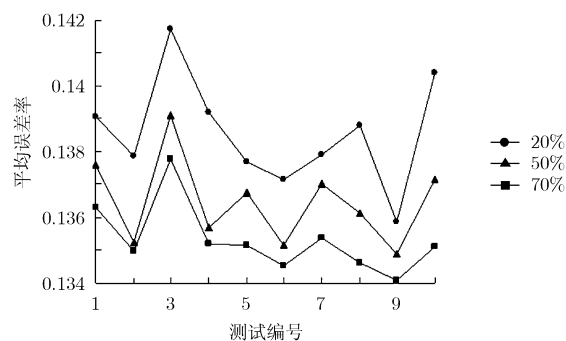


图1 不同好友数据选取比例的平均误差率比较

对于挖掘出的好友关系,仅表明他们之间的信任程度很高,但是他们的业务喜好也有可能差异很大,所以不能将这些关系全部用于预测业务评分。从这些好友关系中通过他们之间基于业务评分的相似性大小来选取部分数据,将使得推荐更加合理。选取比例决定了选取的好友数量,当参与预测的数量过少时,有可能忽略掉某些关键数据的影响,通过试验发现:当好友选取比例为 70%时,平均误差率比较小,所以本文将参数 $t\%$ 定为 70%。

(2)TDS、DSBF、DSBCTF 对比实验 将 4.2.2 节中的 15 组训练集数据随机抽取 10 组代入本文算法进行偏好预测,其中 $t\%$ 的取值定为 70%,间接好友关系挖掘阈值 α 定为 50%,当选用 3 种不同的数据选取方式时, $|\bar{E}|$ 值比较如图 2 所示。

3 种算法中,从直接参与评分预测的数据量上

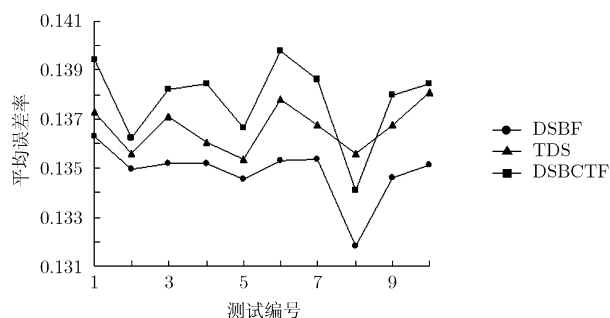


图 2 3 种数据选取方法的性能比较

看, DSBCTF 方法所选取的数据量最多, 它包含了传统基于项目评分相似性及社会化关系挖掘的用户信息, TDS, DSBF 则仅有部分数据; 从数据的推荐意义上看, TDS 缺少了来自社会化关系挖掘的好友数据, 而 DSBF 缺少了来自传统基于项目评分相似性的用户数据, 与 DSBCTF 相比, 这两种方法都不能较全面地预测用户项目评分。实验结果也表明: 本文提出的基于用户社会化关系挖掘的混合数据选取算法(DSBCTF)性能最好, 误差率最小。

5 结论

基于移动社交网络服务现状, 多数用户更相信来自真实社交网络中朋友而非系统的推荐信息, 且交往密切的用户在兴趣上较为相似。本文面向移动通信网, 提出一种基于移动用户社会化关系挖掘的协同过滤算法, 通过对移动通信网下的社会化潜在关系进行挖掘, 将好友数据融入传统的协同过滤算法中。在实验部分, 首先通过一个仿真数据实验验证了该方案的可行性和有效性, 然后通过公开的 Filmtipset 数据集实验确定了合理的参数设置, 并与 TDS, DSBF 数据选取方法进行了性能比较。实验结果表明, 本文算法不仅缓解了评分矩阵稀疏性问题, 而且有效地提高了偏好预测和推荐结果的准确度。

参考文献

- [1] Jonna H, Albrecht S, Jani M, *et al.* Context-aware mobile media and social networks. Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, Bonn, Germany, 2009: 1-3.
- [2] Ricci F. Mobile recommender systems. *International Journal of Information Technology and Tourism*, 2011, 12(3): 205-231.
- [3] Wang L C, Meng X W, Zhang Y J, *et al.* New approaches to mood-based hybrid collaborative filtering. Proceedings of the Workshop on Context-Aware Movie Recommendation at the 4th ACM Conference on Recommender Systems (ACM Recsys'10), Barcelona, Spain, 2010: 28-33.
- [4] Gediminas A and Alexander T. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(6): 152-162.
- [5] 方娟, 梁文灿. 一种基于协同过滤的网格门户推荐模型. 电子与信息学报, 2010, 32(7): 1585-1590.
- [6] Fang Juan and Liang Wen-can. A grid portal recommendation model based on collaborative filtering. *Journal of Electronics & Information Technology*, 2010, 32(7): 1585-1590.
- [7] Matthias B and Gernot B. Improving the recommendation of mobile services by interpreting the user's icon arrangement. Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, Bonn, Germany, 2009: 15-18.
- [8] Gupta A, Kalra A, Boston D, *et al.* MobiSoC: a middleware for mobile social computing applications. *Mobile Networks and Applications*, 2009, 14(10): 35-52.
- [9] Arazy O, Kumar N, and Shapira B. Improving social recommender systems. *IT Professional*, 2009, 11(4): 38-44.
- [10] Dijiang H and Vetri A. Email-based social network trust. IEEE International Conference on Social Computing /IEEE International Conference on Privacy, Security, Risk and Trust, Boston, USA, 2010: 363-370.
- [11] Herlocker J, Konstan J, Terveen L, *et al.* Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 2004, 22(1): 20-21.
- [12] Fernando D and Pedro G C. Movie recommendations based in explicit and implicit features extracted from the Filmtipset dataset. Proceedings of the Workshop on Context-Aware Movie Recommendation at the 4th ACM Conference on Recommender Systems (ACM Recsys'10), Barcelona, Spain, 2010: 45-52.

黄武汉: 男, 1987 年生, 硕士生, 研究方向为社会化网络分析方法与推荐系统。

孟祥武: 男, 1966 年生, 教授, 博士生导师, 研究领域为网络服务、智能信息处理、通信软件。

王立才: 男, 1984 年生, 博士生, 研究方向为推荐系统、智能信息处理。