

基于移动用户上下文相似度的协同过滤推荐算法

徐风苓* 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: 该文面向移动通信网络领域的个性化服务推荐问题, 通过将移动用户上下文信息引入协同过滤推荐过程, 提出一种基于移动用户上下文相似度的改进协同过滤推荐算法。该算法首先计算基于移动用户的上下文相似度, 以构造目标用户当前上下文的相似上下文集合, 然后采用上下文预过滤推荐方法对“移动用户-移动服务-上下文”3维模型进行降维得到“移动用户-移动服务”2维模型, 最后结合传统2维协同过滤算法进行偏好预测和推荐。仿真数据集和公开数据集实验表明, 该算法能够用于移动网络服务环境下的用户偏好预测, 并且与传统协同过滤相比具有更高的推荐精确度。

关键词: 移动网络; 用户上下文; 相似度计算; 协同过滤

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2011)11-2785-05

DOI: 10.3724/SP.J.1146.2011.00384

A Collaborative Filtering Recommendation Algorithm Based on Context Similarity for Mobile Users

Xu Feng-ling 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: Towards the problem of personalized services recommendation in mobile telecommunication network, a collaborative filtering algorithm based on context similarity for mobile users is proposed by incorporating mobile users' context information into collaborative filtering recommendation process. The algorithm calculates firstly user-based context similarities to construct a set of similar contexts related to the current context of the active user. Then it reduces the “mobile user-mobile service-context” 3D model to the “mobile user-mobile service” 2D model by using context pre-filtering recommendation method. Finally it predicts the unknown user preferences and generates recommendations based on the traditional 2D Collaborative Filtering (CF) algorithm. Experimental results indicate that this algorithm can be applied to predict user preferences in mobile network service environment and achieve better recommendation accuracy than the traditional CF algorithm.

Key words: Mobile network; User context; Similarity measure; Collaborative Filtering (CF)

1 引言

随着移动通信技术的快速发展, 越来越多的用户通过移动设备获取移动网络服务和信息内容。由于移动设备具有处理能力低、电池续航能力差、输入输出能力有限等缺点, 移动用户获取信息的能力受到严重影响。而且, 移动用户所处的上下文环境动态多变, 使其在不同上下文条件下具有不同的移动信息需求。例如, 有些用户喜欢“在乘坐地铁时”

通过移动设备浏览新闻, 而“在办公室”“休息”时则喜欢玩游戏。因此, 如何帮助移动用户发现其真正感兴趣的、符合其上下文条件约束(如时间、地点、活动状态、网络条件等)的移动网络服务及其信息内容, 显得尤为重要。

推荐系统^[1,2]通过识别和预测用户偏好, 提供个性化服务, 是缓解“信息过载”问题的重要手段之一。传统的推荐系统较少考虑移动环境下影响用户偏好的上下文因素, 因而不适用于个性化移动网络服务领域。上下文感知计算是指系统能发现并有效利用上下文信息进行计算的一种计算模式, 已经广泛应用于许多领域^[3,4]。上下文感知推荐系统通过将

2011-04-20 收到, 2011-08-22 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: 徐风苓 xfl198712@gmail.com

上下文信息引入推荐系统^[5,6], 成为一个刚刚兴起的领域, 还有许多问题等待解决, 如上下文用户偏好提取、高维数据稀疏性问题等^[6-8], 在移动领域的应用实践也才刚刚开始。文献[9]针对移动领域的推荐系统进行介绍, 分析了几种典型的模型和算法, 并指出当前存在的重要挑战。

本文在现有上下文推荐系统和移动推荐系统的研究基础上, 针对个性化移动网络服务要求, 基于“移动用户-移动服务-上下文”3维模型, 提出了基于用户的上下文相似度计算方法, 并结合上下文预过滤推荐算法思想, 改进了3维模型的降维方法, 最终提出一种基于移动用户上下文相似度的改进协同过滤推荐算法, 并通过实验表明了该算法的可行性和有效性。本文第2节介绍了协同过滤、上下文预过滤等方面的相关工作, 第3节论述了基于移动用户上下文相似度的协同过滤推荐算法, 第4节展开实验分析, 第5节总结全文。

2 相关工作

2.1 协同过滤

协同过滤算法在推荐系统领域应用最广, 其基本思想是通过相似性计算找到邻居用户, 目标用户对对象的偏好可以通过邻居用户对该对象偏好的加权平均值逼近^[1]。其主要实现过程包括: 相似性计算和用户偏好预测。相似性的度量方法主要包括3种^[2]: 余弦相似性、Pearson相关系数和修正的余弦相似性。其中Pearson^[1]相关系数使用最多, 计算方法如下:

$$\text{sim}_u(x, y) = \frac{\sum_{s \in S_{xy}} (r_{(x,s)} - \bar{r}_x)(r_{(y,s)} - \bar{r}_y)}{\sqrt{\sum_{s \in S_{xy}} (r_{(x,s)} - \bar{r}_x)^2 \times \sum_{s \in S_{xy}} (r_{(y,s)} - \bar{r}_y)^2}} \quad (1)$$

其中 $S_{xy} = S_x \cap S_y$, 表示用户 x 和用户 y 共同评过分的对象集合; $r_{(x,s)}$ 表示用户 x 对对象 s 的评分; \bar{r}_x 表示用户 x 对对象的平均评分。

偏好预测则普遍采用加权平均预测公式^[1]:

$$R(u, s) = k \sum_{u' \in N} [\text{sim}_u(u, u') \times R(u', s)] \quad (2)$$

其中集合 N 为用户 u 的邻居集合。 k 为一个标准化因子, $k = \frac{1}{\sum_{u' \in N} |\text{sim}_u(u, u')|}$ 。

2.2 上下文预过滤推荐

文献[5]基于上下文信息将2维推荐空间延伸到多维, 并将OLAP处理多维信息的能力应用到推荐机制上, 提出多维推荐模型。Chen^[10]面向普适计算环境, 提出一种上下文感知的协同过滤系统, 其主要贡献是设计了基于项目的上下文相似度的计算公

式和含上下文信息的用户评分预测计算公式。文献[6]根据上下文信息在推荐过程中的不同位置, 总结了上下文感知推荐系统的3种范式: 上下文预过滤、上下文后过滤和上下文模型。其中, 上下文预过滤推荐范式是指, 在生成推荐结果之前, 利用当前上下文信息过滤掉无关的用户偏好数据, 从而构建当前上下文信息相关的推荐数据集; 然后利用传统推荐方法处理这些筛选后的数据以进行评分预测, 并生成满足当前上下文条件的推荐结果。其核心思想是将多维推荐问题转化为传统2维推荐问题求解。例如, 假定一个用户想在周末看电影, 可以首先根据“周末”这一条上下文具体实例将“非周末”的用户偏好数据过滤掉, 然后利用传统推荐算法为其生成一个推荐列表。目前, 上下文具体实例的分类或者粒度, 是影响上下文预过滤范式的重要因素, 关系到数据稀疏性问题。

3 基于移动用户上下文相似度的协同过滤推荐算法

3.1 数据模型

定义1 使用向量模型^[3]表示由 n 个上下文属性(如时间、位置、活动状态等)组成的上下文 $\text{Context} = \{C_1, C_2, \dots, C_i, \dots, C_n\}$; 一个上下文实例表示为 $C = (c_1, c_2, \dots, c_n)$, 其中 $c_i (1 \leq i \leq n)$ 为上下文属性 C_i 的一个具体属性值。若 C_i 为时间上下文, c_i 可能取值有上午、下午、晚上。

定义2 “移动用户-移动服务-上下文”3维模型记作U-S-C模型, 它是一个3维的向量空间 $\{\text{User}, \text{Service}, \text{Context}\}$, 3个维度分别是: 移动用户、移动服务、上下文, 每个维度分别用各自属性值组成的向量来表示。其中, 移动用户、移动服务、上下文、偏好值所组成的一条记录, 称为一条偏好记录。偏好记录的集合叫做偏好数据集。

3.2 算法设计思想

本文借鉴上下文预过滤推荐的思想, 用移动用户当前上下文 C 的相似上下文集合 $N(c)$ 代替 C 作为过滤条件, 对“移动用户-移动服务-上下文”3维模型进行降维, 以期有效缓解数据稀疏性问题, 并利用传统2维推荐算法进行用户偏好预测。因此, 需要设计一种基于用户的上下文相似度计算来构造 $N(c)$ 。

公式(3)对文献[10]提出的基于对象的上下文相似度计算公式进行改进, 是基于用户的上下文相似度计算公式, 用来度量同一用户的不同上下文的相似度, 根据该用户在不同上下文情况下对相同对象

的不同评分值计算而得。

$$\text{sim}_u(u, x, y) = \frac{\sum_{s \in S} (r_{(u,s,x)} - \bar{r}_{u,x})(r_{(u,s,y)} - \bar{r}_{u,y})}{\sqrt{\sum_{s \in S} (r_{(u,s,x)} - \bar{r}_{u,x})^2 \times \sum_{s \in S} (r_{(u,s,y)} - \bar{r}_{u,y})^2}} \quad (3)$$

其中 S 表示用户 u 在上下文 x 和上下文 y 下都存在评分值的移动服务集合; $r_{(u,s,x)}$ 表示用户 u 对服务 s 在上下文 x 下的评分值; $\bar{r}_{u,x}$ 表示用户 u 在上下文 x 下的平均评分值。

实际情况中, 如果数据集中的移动用户在不同的上下文下共同评分的移动服务数很少或者没有, 将很难得到目标用户当前上下文 C 的相似上下文集合。为了解决这一问题, 可以利用上下文特征相似度计算公式代替式(3)。这里, 使用余弦相似性公式计算上下文实例 x 和 y 的特征相似度:

$$\text{sim}_c(x, y) = \frac{x \cdot y}{\|x\| \|y\|} \quad (4)$$

构造当前上下文的相似上下文集合后, 便可以按照上下文预过滤思想进行偏好预测。

3.3 算法过程描述

本节在上述数据模型和设计思想的基础上, 对算法的输入、输出、详细流程展开描述。

输入: “移动用户-移动服务-上下文” 3 维模型; 目标用户 u 及当前上下文 C ; 待推荐的移动服务集合 $\text{Service}(R)$ 。

输出: 用户 u 在当前上下文 C 下, 对 $\text{Service}(R)$ 中偏好最大的 TOP- N 个移动服务。

算法的具体过程, 分为以下几步:

(1) 取出目标用户 u 的所有偏好数据, 构造上下文-移动服务 2 维偏好矩阵。

(2) 在第 1 步矩阵基础上, 利用 Pearson 相关相似性公式计算当前上下文 C 与其余上下文间的相似度, 构造当前上下文的相似上下文集合 $N(c)$, 由于该矩阵是目标用户的上下文-移动服务 2 维偏好矩阵, 所以生成结果基于目标用户当前上下文的相似上下文集合 $N(c)$ 。

(3) 判断第 2 步是否由于数据稀疏性难以构造相似上下文集合, 如果不存在该问题, 执行第(4)步, 否则, 计算当前上下文 C 与其余上下文间的特征相似度, 用与当前上下文 C 特征相似度很大的邻居上下文来填充第 2 步的集合 $N(c)$ 。

(4) 将“移动用户-移动服务-上下文” 3 维模型降维生成 U-S 2 维模型。降维方法如下: 在得到的基于目标用户当前上下文的相似上下文集合 $N(c)$ 基础上, 从“移动用户-移动服务-上下文”模型中取出上下文属于集合 $N(c)$ 的所有偏好记录, 若某一用户在

不同上下文下对同一移动服务存在不同的偏好值, 则仅保留其上下文与目标用户当前上下文 C 相似度最大的那条偏好记录。例如, 对于偏好记录 (U_1, C_1, S, R_1) 和 (U_1, C_2, S, R_2) , 若 C_1, C_2 都在集合 $N(c)$ 中, 但 C_2 与目标用户当前上下文 C 有更高的相似度, 降维时只保留偏好记录 (U_1, C_2, S, R_2) 。这样保证了用户对一个服务有唯一的偏好值, 过滤上下文信息后得到 2 维 U-S 模型。

(5) 对 2 维 U-S 模型, 采用传统的基于用户的协同过滤算法进行偏好预测。用协同过滤算法确定目标用户 u 的 M 个邻居用户, 预测用户 u 对待推荐移动服务集合 $\text{Service}(R)$ 中各移动服务的偏好, 将偏好最大的 TOP- N 个移动服务推荐给目标用户 u 。

4 实验结果及分析

本文采用的实验平台为 PC(Intel(R), CPU 2.93 GHz, RAM 2 GB) 和 Windows XP SP3 操作系统, 开发工具使用 Java SE6, Eclipse 3.2 和 MySQL5.0, 算法使用 Java 语言编写。

4.1 数据集

为了验证本文所提方法在移动网络服务环境下的有效性和可行性, 及其相对传统算法具有较优的推荐精确度, 分别使用两个数据集展开实验。一个是从中国移动应用商城(<http://mm.10086.cn/>)上采集移动用户购买行为并处理得到的仿真数据集; 一个是公开可用的面向上下文感知的电影推荐的评分数据集 Moviepilot^[11]。

(1) 个性化移动服务偏好数据集 该数据集是从移动应用商城采集并经处理得到的仿真数据集, 采集和处理过程如下:

采集移动应用商城上移动用户的购买记录, 从中选择出 50 个忠诚用户的购买记录; 然后将记录中的具体时间划归到(清晨、早上、上午、中午、下午、傍晚、晚上、凌晨)8 个时间段中; 再统计每个用户在每个时间段内对各类产品的购买次数; 最后根据每个用户在某时间段内对各类产品的购买次数统计生成用户基于时间的产品偏好(1-10)。最终得到 50 个用户在不同上下文具体实例下对 41 类产品的 2975 条偏好记录, 随机选出 100 条记录作为测试集, 剩余 2875 条记录作为训练集。

(2) Moviepilot_week 公开数据集 Moviepilot_week 数据集包含用户对电影评分的时间上下文信息, 所以可以用来作为基于上下文推荐算法的实验数据集(评分值是从 0 到 100 的整数)。本文从原始数据集中选出评分记录最多的用户和测试集中的用户作为用户集, 最终采用的训练集包括 2,320 个用

用户对 23,628 部电影的 1,628,955 条评分记录。

4.2 度量标准

本文采用平均绝对误差(Mean Absolute Error, MAE)^[12]和 $P(u)@N$ ^[11]作为度量标准。

MAE 通过计算预测值与实际值之间的偏差来度量预测的准确性,偏差越小,预测精度越高。设预测的评分集合表示为 $\{p_1, p_2, \dots, p_N\}$, 对应的实际评分集合为 $\{q_1, q_2, \dots, q_N\}$, 则

$$MAE = \left(\sum_{i=1}^N |p_i - q_i| \right) / N \quad (5)$$

$P(u)@N$ 表示为用户 u 生成的 TOP- N 推荐列表中符合用户需求的项目数与 N 的比值:

$$P(u)@N = \frac{\text{\#relevant items in top } n \text{ items for } u}{N} \quad (6)$$

4.3 实验结果分析

(1)基于个性化移动服务偏好数据集的对比实验 本实验分别采用传统的上下文预过滤方法^[5]和本文所提出的改进算法对测试数据集中的“用户-服务”序对进行偏好预测。其中,传统的预过滤方法使用目标用户所处的时间段(如上午)进行过滤,而改进算法则利用目标用户当前时间段的相似时间段集合进行预过滤。图1给出了两种方法比较结果。

由图1可知,在不同邻居数目的情况下,改进算法比传统上下文预过滤推荐方法均有较低的 MAE 值,表明改进算法在一定程度上比传统的预过滤方法推荐质量高,从而表明其在个性化移动网络服务领域的有效性和可行性。这是因为,利用相似上下文集合代替当前上下文有利于缓解数据稀疏

性,也可充分利用上下文信息在提高推荐精确度方面的作用。

(2)基于 Moviepilot_week 公开数据集的对比实验 本实验首先将训练集和测试集中的时间戳按月份归类。由于用户对每部电影仅有一个评分,所以采用上下文特征相似度代替基于用户的上下文相似度进行预过滤,从训练集中选出与测试集数据所在月份相近的月份的所有评分记录作为预过滤后的数据集进行偏好预测。

实验结果见图2、图3。从中可以看出,在选取不同邻居数目的情况下,本文的改进算法与传统协同过滤算法相比,均具有更低的 MAE 误差值,更高的 $P@10$ 准确率,表明改进算法在合理考虑上下文的情况下,比传统协同过滤算法有更高的推荐质量。

5 总结

本文在上下文推荐系统和移动推荐系统相关研究的基础上,针对移动网络环境特点,提出了一种基于移动用户上下文相似度的协同过滤推荐算法。基于“移动用户-移动服务-上下文”3维推荐模型,提出了基于用户的上下文相似度计算方法和上下文特征相似度计算方法,并结合上下文预过滤推荐算法思想,改进了3维偏好模型的降维方法。通过实验表明,该算法适用于移动网络中个性化服务推荐领域,既缓解了传统上下文预过滤方法的数据稀疏性问题,又能充分利用上下文信息,从而提高了推荐质量。

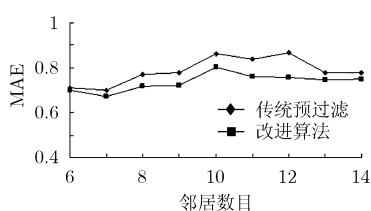


图1 改进算法与传统上下文预过滤推荐方法的 MAE 比较

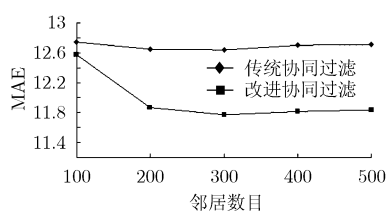


图2 改进算法与传统协同过滤算法的 MAE 比较

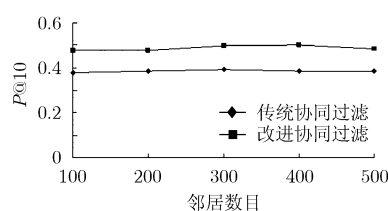


图3 改进算法与传统协同过滤的 P@10 比较

参考文献

- [1] Adomavicius G and Tuzhilin A. Towards the next generation of recommender systems: a survey of the state-of-the-art and possible extensions[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(6): 734-749.
- [2] 许海玲, 吴潇, 李晓东, 等. 互联网推荐系统比较研究[J]. 软件学报, 2009, 20(2): 350-362.
Xu Hai-ling, Wu Xiao, Li Xiao-dong, et al. Comparison study of internet recommendation system[J]. *Journal of Software*, 2009, 20(2): 350-362.
- [3] 李蕊, 李仁发. 上下文感知计算及系统框架综述[J]. 计算机研究与发展, 2007, 44(2): 269-276.
Li Rui and Li Ren-fa. A survey of context-aware computing and its system infrastructure[J]. *Journal of Computer Research and Development*, 2007, 44(2): 269-276.
- [4] 李艳娜, 乔秀全, 李晓峰. 基于证据理论的上下文本体建模以及不确定性推理方法[J]. 电子与信息学报, 2010, 32(8): 1806-1811.

- Li Yan-na, Qiao Xiu-quan, and Li Xiao-feng. An uncertain context ontology modeling and reasoning approach based on D-S theory[J]. *Journal of Electronics and Information Technology*, 2010, 32(8): 1806-1811.
- [5] Adomavicius G, Sankaranarayanan R, Sen S, *et al.* Incorporating contextual information in recommender systems using a multidimensional approach[J]. *ACM Transactions on Information Systems*, 2005, 23(1): 103-145.
- [6] Adomavicius G and Tuzhilin A. Context-aware Recommender Systems (Book Chapter)[M]. *Recommender Systems Hand-book*, New York, Dordrecht, Heidelberg, London, Springer Press, 2011: 217-253.
- [7] Zhang Yu-jie and Wang Li-cai. Some challenges for context-aware recommender systems[C]. In the 1st Workshop on Recommender System at The 5th IEEE International Conference on Computer Science & Education, Hefei, China, 2010: 362-365.
- [8] Wang Li-cai. Understanding and using contextual information in recommender systems[C]. In Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information, Beijing, China, 2011: 1329-1330.
- [9] Ricci F. Mobile recommender systems[J]. *Journal of Information Technology and Tourism*, 2011, 12(3): 205-231.
- [10] Chen Annie. Context-aware collaborative filtering system: predicting the user's preference in the ubiquitous computing environment[C]. The 1st International Workshop on Location-and Context-Awareness, Berlin, Heidelberg, Springer 2005, LNCS 3479: 244-253.
- [11] Wang Li-cai, Meng Xiang-wu, Zhang Yu-jie, *et al.* New approaches to mood-based hybrid collaborative filtering[C]. In Proceedings of the Workshop on Context-Aware Movie Recommendation at the 4th ACM Conference on Recommender System, Barcelona, Spain, 2010: 28-33.
- [12] Herlocker J L, Konstan J A, Terveen L G, *et al.* Evaluating collaborative filtering recommender systems[J]. *ACM Transactions on Information Systems*, 2004, 22(1): 5-53.
- 徐风苓：女，1985年生，硕士生，研究方向为上下文感知推荐系统。
- 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。