

# 基于协同过滤的移动电子商务个性化推荐系统若干研究

吴吉义<sup>1</sup>, 林志洁<sup>2</sup>, 龚祥国<sup>1</sup>

(1. 浙江广播电视大学 远程开放实验室, 浙江 杭州 310012;

2. 浙江科技学院 信息与电子工程学院, 浙江 杭州 310018)

**摘要:** 在简要介绍移动电子商务个性化推荐系统概念的基础上, 给出了移动电子商务推荐系统 EMC-PRS 的模块结构。重点分析比较了基于最近邻居的协同过滤算法和基于项目评分预测的协同过滤算法。经测试发现, 基于项目评分预测的协同过滤推荐算法可以显著提高个性化推荐系统的推荐质量。

**关键词:** 协同过滤 移动电子商务 个性化推荐系统 算法分析

随着移动通信技术的发展和电子商务市场环境的不断成熟, 移动电子商务 MEC(Mobile Electronic Commerce) 正成为电子商务发展新的增长点。移动电子商务是指通过手机、个人数字助理(PDA)和掌上电脑等手持移动终端进行的商务活动, 是移动通信技术与 Internet 技术相结合的产物。我国移动通信用户基数巨大, 市场增长迅速, 截至 2006 年 10 月已达到 3.8 亿户, 这无疑是中

国最有消费能力的群体。如果能够吸引 5%, 也就是 1900 万手机用户参与移动电子商务, 年人均支付金额为 100 元, 那么通过移动电子商务的年市场交易总额将达到 20 亿元左右。IBM 无线电子商务副总裁 James Keegan 预估, “全球无线电子商务市场在 2008 年时, 总值将达 2000 亿美元”。

与传统商务方式不同, 移动电子商务环境中交易双方是不谋面的,

商家不能直观地了解客户, 能获得的只是大量的相关数据(如用户注册信息、历史购买记录等)。因此通过对数据的分析尽可能地揣摩客户需求, 在恰当的时间向相应的客户个性化地推荐商品或服务就显得尤为重要。因此, 个性化推荐系统 PRS(Personalized Recommendation System) 在移动电子商务领域具有极高的研究价值和应用前景。

## 移动电子商务个性化推荐系统 EMC-PRS 简介

在移动电子商务的虚拟环境中, 商家能够在网上提供的商品种类和数量非常多, 但用户不可能通过一个狭小的移动终端屏幕很方便地发现自己感兴趣的商品。用户既不愿意花太多时间在漫无边际的电

子商务平台上寻找商品,也不可能像在物理环境下那样检查商品的质量。因此,用户需要移动电子商务系统具有一种类似采购助手的功能来帮助其选购商品,并能根据用户的兴趣爱好个性化地给每个用户推荐他可能感兴趣且满意的商品。

在这种背景下,移动电子商务个性化推荐系统 EMC-PRS (Mobile Electronic Commerce-PRS)应运而生。它能在移动电子商务环境中根据客户的兴趣偏好,向客户提供商品信息和建议,推荐符合用户兴趣爱好的商品对象,模拟销售人员帮助客户完成购买过程。EMC-PRS 可以根据其他客户的信息或是当前客户的信息,根据该顾客以往的购买行为预测未来的购买行为,模拟销售人员帮助客户完成购买过程,为客户提供个性化服务。

移动电子商务平台中,个性化推荐系统的重要职能包括:

(1)将浏览者转变为购买者:有时客户只是浏览商务平台的内容而没有购买的意向,推荐系统可以利用该机会帮他们找到感兴趣的、愿意购买的商品。

(2)改善交叉销售能力:基于用户已经购买的商品,为他们推荐一些相关的商品。

(3)建立客户忠诚度:客户往往愿意到那些最能满足自己需求的网站购物。

## 移动电子商务个性化推荐系统 EMC-PRS 模块结构

一般地,移动电子商务个性化推荐系统 EMC-PRS 包括输入功能模块、推荐方法模块和输出功能模块。图1所示为移动电子商务个性化推荐系统 EMC-PRS 的模块结构。

EMC-PRS 输入功能模块输入的信息包括目标客户输入和社团输入。其中目标客户输入可以分为:①客户的浏览行为作为推荐系统的输入,但客户并不知情,称为隐式浏览输入;②客户的浏览行为是有目的地向推荐系统提供自己的喜好,

称为显式浏览输入;③客户输入关键词或项目的有关属性以得到推荐系统有价值的推荐,称为关键词和项目属性输入;④用户过去的购买记录,称为用户购买历史。社团输入可以分为:①社团对商品风格和类别的集体评判,称为项目属性;②社团过去的购买记录,称为社团购买历史;③其他客户对商品的文本评价,计算机并不知道评价是好是坏,称为文本评价;④其他客户对商品的评分,计算机可以对评分进行处理,称为等级评分。

推荐方法模块中可以采用的推荐技术包括协同过滤推荐、基于内容的推荐、基于用户统计信息的推荐、基于效用的推荐、基于知识的推荐和基于关联规则的推荐等。组合推荐也是可以采用的推荐方式,研究和应用最多的是协同过滤推荐和内容推荐的组合。最简单的做法是分别用基于内容的方法和协同推荐方法,产生一个推荐预测结果,然后用某方法组合其结果。

EMC-PRS 输出功能模块输出的信息包括:①建议:包括单个建议、未排序建议列表和排序建议列表等;②预言:系统对给定项目的总体评分;③个体评分:输出其他客户对商品的个体评分;④评论:输出其他客户对商品的文本评价。

## 移动电子商务个性化推荐系统 EMC-PRS 中的协同过滤算法分析

### 基于最近邻居的协同过滤算法

最近邻居协同过滤推荐是当前

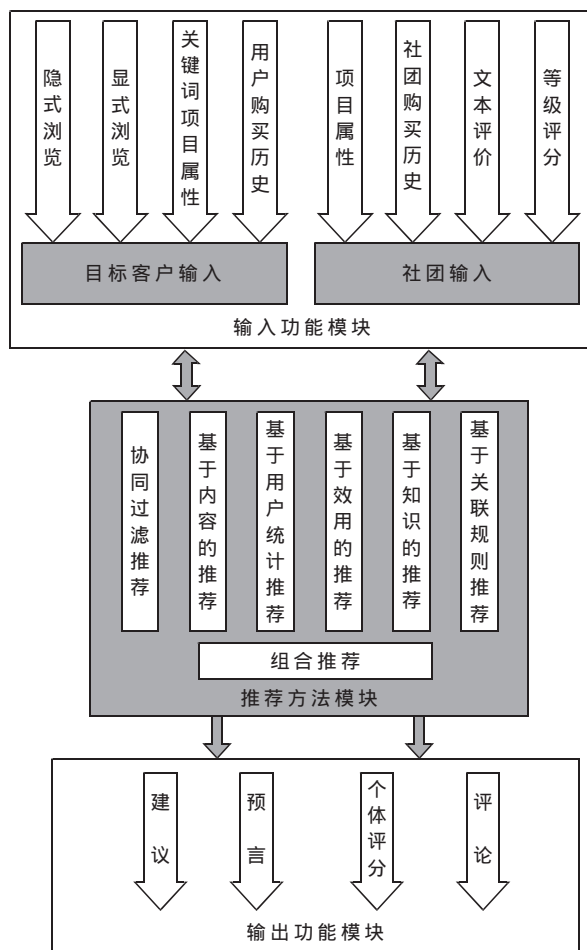


图1 移动电子商务个性化推荐系统 EMC-PRS 模块结构

最成功的推荐技术之一,其基本思想就是基于评分相似的最近邻居的评分数据向目标用户产生推荐。由于最近邻居对项目(移动电子商务中的商品,如书籍、电影、音乐等)的评分与目标用户非常相似,因此目标用户对未评分项目的评分可以通过最近邻居对该项目评分的加权平均值逼近。该算法一般分为表示、邻居形成和推荐产生三个主要阶段。

(1)表示阶段:对客户已经购买的商品进行建模,采用  $m \times n$  阶客户-商品矩阵表示,如图2。其中如果  $R_{ij}=1$ ,则第  $i$  个客户购买了第  $j$  件商品;如果  $R_{ij}=0$ ,则第  $i$  个客户没有购买第  $j$  件商品。

上述表示称为原始表示。这种表示的主要问题有:①稀疏性:大部分客户购买的商品不到全部商品的1%,从而使得推荐精度很低;②适应性:计算代价随着客户数目和商

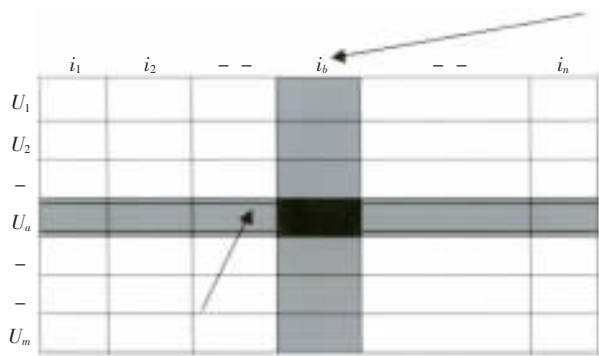


图2  $m \times n$  阶客户-商品矩阵

品数目的增加而增加,很难满足实时性要求;③同义词问题:同一类商品的名字不一样。通过奇异值分解将  $m \times n$  阶客户-商品矩阵变换为  $m \times k$  阶矩阵,这种表示称为降维表示,可以部分解决原始表示存在的稀疏性、适应性和同义词问题。

(2)邻居形成阶段:寻找目标客户所对应的邻居。该阶段的关键在于计算客户之间的相似性,目标是对于每个客户  $i$ ,找到它的  $l$  个邻居  $j = \{j_1, j_2, \dots, j_l\}$ ,使得  $\text{sim}(i, j_1)$  最大,  $\text{sim}(i, j_2)$  次之……。可以采用的相似性度量方法:

①相关相似性:设经用户  $i$  和用户  $j$  共同评分的项目集合用  $I_{ij}$  表示,则用户  $i$  和用户  $j$  之间的相似性  $\text{sim}(i, j)$  通过 Pearson 相关系数度量:

$$\text{sim}(i, j) = \frac{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_{ij}} (R_{j,c} - \bar{R}_j)^2}}$$

$R_{i,c}$  表示用户  $i$  对项目  $c$  的评分,  $\bar{R}_i$  和  $\bar{R}_j$  分别表示用户  $i$  和用户  $j$  对项目的平均评分。

②余弦相似性:用户评分被看做是  $n$  维项目空间上的向量。如果用户对项目没有进行评分,则将用户对该项目的评分设为 0,用户间的相似性通过向量间的余弦夹角度量。设用户  $i$  和用户  $j$  在  $n$  维项目空间上的评分分别表示为向量  $\vec{i}$ 、 $\vec{j}$ ,则用户  $i$  和用户  $j$  之间的相似性  $\text{sim}(i, j)$  为:

$$\text{sim}(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\| \|\vec{j}\|}$$

分子为两个用户评分向量的内积,分母为两个用户向量模的乘积。

③修正的余弦相似性:在余弦相似性度量方法中没有考虑不同用户的评分尺度问题,修正的余弦相似性度量方法通过减去用户对项目的平均评分来改

善上述缺陷,设经用户  $i$  和用户  $j$  共同评分的项目集合用  $I_{ij}$  表示,  $I_i$  和  $I_j$  分别表示经用户  $i$  和用户  $j$  评分的项目集合,则用户  $i$  与用户  $j$  之间的相似性  $\text{sim}(i, j)$  为:

$$\text{sim}(i, j) = \frac{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_{ij}} (R_{j,c} - \bar{R}_j)^2}}$$

$R_{i,c}$  表示用户  $i$  对项目  $c$  的评分,  $\bar{R}_i$  和  $\bar{R}_j$  分别表示用户  $i$  和用户  $j$  对项目的平均评分。

(3)推荐产生阶段:从目标客户的邻居中产生  $N$  项产品推荐,可以采用如下两种不同方法产生推荐:①最频繁项目推荐:扫描目标客户每一个邻居的购买数据,对其购买的商品进行计数,选择出现频率最高且目标客户没有购买的前  $N$  件商品最佳推荐结果。②基于关联的推荐:类似于基于关联规则的推荐算法,只是将目标客户的邻居作为算法的输入。

### 基于项目评分

#### 预测的协同过滤算法

基于项目评分预测的协同过滤算法基本思路是根据目标客户已经评价过的项目与目标项目的相似性,选择  $k$  个最相似的项目  $\{i_1, i_2, \dots, i_k\}$ ,同时得到  $k$  个最相似的项目与目标项目的相似度,记为  $\{s_{i1}, s_{i2},$

$\dots, s_{ik}\}$ ,然后将目标客户对这  $k$  个最相似的项目的评分及这  $k$  个最相似的项目与目标项目的相似度的加权平均值作为对目标项目的评分。该算法一般分为项目相似性计算和推荐产生两个主要阶段:

(1)项目相似性计算阶段:从所有的用户中分离出同时对项目  $i$  和项目  $j$  进行评价的用户,根据上面得到的数据计算项目  $i$  和项目  $j$  的相似性,如图 3 所示。

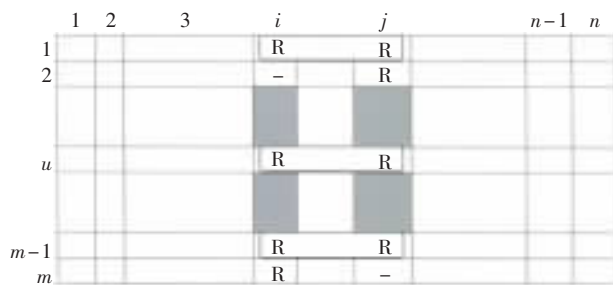


图3 计算项目  $i$  和项目  $j$  的相似性

相似性度量与基于最近邻技术的协同推荐系统类似:可以采用基于相关性的相似性和基于余弦的相似性。

(2)推荐产生阶段:方法很简单,根据目标客户对最相似项目的评分及最相似项目与目标项目的相似度产生推荐。

$$P_{i,p} = \frac{\sum_{n \in M_p} \text{sim}_{p,n} \times R_{i,n}}{\sum_{n \in M_p} (|\text{sim}_{p,n}|)}$$

利用相似性度量方法得到目标用户的最近邻居,下一步需要产生相应的推荐。设用户  $u$  的最近邻居集合用  $NBS_u$  表示,则用户  $u$  对项目  $i$  的预测评分  $P_{u,i}$  可以通过用户  $u$  对最近邻居集合  $NBS_u$  中项目的评分得到,计算方法如下:

$$P_{u,i} = \bar{R}_M + \frac{\sum_{n \in NBS_u} \text{sim}(u, n) \times (R_{n,i} - \bar{R}_n)}{\sum_{n \in NBS_u} (|\text{sim}(u, n)|)}$$

$\text{sim}(u, n)$  表示用户  $u$  与用户  $n$  之间的相似性,  $R_{n,i}$  表示用户  $n$  对项目  $i$  的评分,  $\bar{R}_n$  分别表示用户  $u$  和用户  $n$  对项目的平均评分。

### 协同过滤算法的

#### 度量标准与性能测试

评价推荐系统推荐质量的度量



标准主要包括统计精度度量方法和决策支持精度度量方法两类。统计精度度量方法中的平均绝对偏差 MAE(Mean Absolute Error)易于理解,可以直观地对推荐质量进行度量,是最常用的一种推荐质量度量方法,本文采用平均绝对偏差 MAE 作为度量标准。平均绝对偏差 MAE 通过计算预测的用户评分与实际用户评分之间的偏差度量预测准确性,MAE 越小,推荐质量越高。设预测的用户评分集合表示为  $\{p_1, p_2, \dots, p_N\}$ , 对应的实际用户评分集合为  $\{q_1, q_2, \dots, q_N\}$ , 则平均绝对偏差 MAE 定义为:

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N}$$

为了检验基于项目评分预测的协同过滤算法的有效性,以传统的协同过滤推荐算法作为基准。在传统的协同过滤推荐算法中,分别以余弦相似性和相关相似性作为相似性度量标准,计算其 MAE,邻居个数从 4 增加到 20,间隔为 4,然后与基于项目评分预测的协同过滤推荐算法作比较,实验结果如图 4 所示。

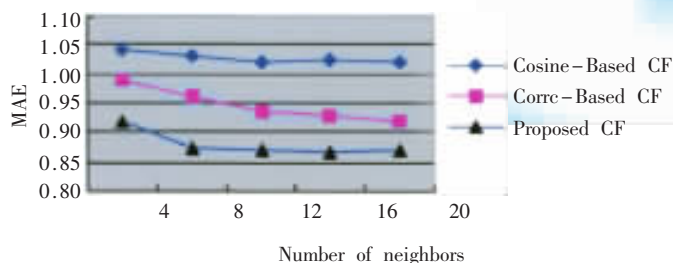


图 4 推荐算法推荐精度比较

由图 4 可知,在各种实验条件下,基于项目评分预测的协同过滤推荐算法均具有最小的 MAE。由此可知,与传统的协同过滤推荐算法相比,基于项目评分预测的协同过滤推荐算法可以显著地提高推荐系统的推荐质量。

此外,在基于最近邻居技术的协同过滤推荐算法中,邻居生成阶段,特别是客户相似性计算是算法提高性能的瓶颈,使得大规模移动电子商务网站的实时推荐不能实

现。该算法将邻居生成阶段与推荐产生阶段分离,使邻居生成阶段可以离线进行,从而使移动电子商务网站的实时推荐得以实现。

移动电子商务个性化推荐是一个全新的研究领域,尽管目前已取得了初步的研究成果,但尚存在许多问题,如数据获取主要依赖用户的显式评价,在自动获得用户的隐式信息方面做得不够;对推荐系统的开发与应用,尤其是与企业其他系统的集成应用研究不够。针对用户评分数据的极端稀疏性,采用基于项目评分预测的最近邻居协同过滤推荐算法,通过计算项目之间的相似性,初步预测用户对未评分项目的评分,然后采用一种新颖的相似性度量方法计算用户的最近邻居。当然,该方法尚存在某些不成熟的方面,需要在实际应用中不断改进与完善。

#### 参考文献

- [1] SARWAR B, KARYPIS G, KONSTAN J et al. Item-Based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International World Wide Web Conference. 2001: 285-295.
- [2] BREESE J, HECHERMAN D, KADIE C. Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence(UAI'98). 1998: 43-52.
- [3] SARWAR B M, KARYPIS G, KONSTAN J A. Analysis of recommendation algorithms for ecommerce. Proceedings of the ACM EC 00 Conference[C]. 2000: 158-167.
- [4] 黎星星, 黄小琴, 朱庆生. 电子商务推荐系统研究[J]. 计算机工程与科学, 2004, 26(5): 7-10.
- [5] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(9): 1621-1628.

(收稿日期: 2006-11-06)