软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2013,24(11):2721−2733 [doi: 10.3724/SP.J.1001.2013.04478]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于时序行为的协同过滤推荐算法*

孙光福, 吴 乐, 刘 淇, 朱 琛, 陈恩红

(中国科学技术大学 计算机科学与技术学院,安徽 合肥 230027)

通讯作者: 陈恩红, E-mail: cheneh@ustc.edu.cn

摘 要: 协同过滤直接根据用户的行为记录去预测其可能喜欢的产品,是现今最为成功、应用最广泛的推荐方法.概率矩阵分解算法是一类重要的协同过滤方式.它通过学习低维的近似矩阵进行推荐,能够有效处理海量数据.然而,传统的概率矩阵分解方法往往忽略了用户(产品)之间的结构关系,影响推荐算法的效果.通过衡量用户(产品)之间的关系寻找相似的邻居用户(产品),可以更准确地识别用户的个人兴趣,从而有效提高协同过滤推荐精度.为此,提出一种对用户(产品)间的时序行为建模的方法.基于该方法,可以发现对当前用户(产品)影响最大的邻居集合.进一步地,将该邻居集合成功融合到基于概率矩阵分解的协同过滤推荐算法中.在两个真实数据集上的验证结果表明,所提出的 SequentialMF 推荐算法与传统的使用社交网络信息与标签信息的推荐算法相比,能够更有效地预测用户实际评分,提升推荐精度.

关键词: 协同过滤;时序行为;概率矩阵分解
中图法分类号: TP183 文献标识码: A

中文引用格式: 孙光福,吴乐,刘淇,朱琛,陈恩红.基于时序行为的协同过滤推荐算法.软件学报,2013,24(11):2721−2733. http://www.jos.org.cn/1000-9825/4478.htm

英文引用格式: Sun GF, Wu L, Liu Q, Zhu C, Chen EH. Recommendations based on collaborative filtering by exploiting sequential behaviors. Ruan Jian Xue Bao/Journal of Software, 2013,24(11):2721−2733 (in Chinese). http://www.jos.org.cn/1000-9825/4478.htm

Recommendations Based on Collaborative Filtering by Exploiting Sequential Behaviors

SUN Guang-Fu, WU Le, LIU Qi, ZHU Chen, CHEN En-Hong

(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027,China)

Corresponding author: CHEN En-Hong, E-mail: cheneh@ustc.edu.cn

Abstract: Collaborative filtering, which makes personalized predictions by learning the historical behaviors of users, is widely used in recommender systems. The key to enhance the performance of collaborative filtering is to precisely learn the interests of the active users by exploiting the relationships among users and items. Though various works have targeted on this goal, few have noticed the sequential correlations among users and items. In this paper, a method is proposed to capture the sequential behaviors of users and items, which can help find the set of neighbors that are most influential to the given users (items). Furthermore, those influential neighbors are successfully applied into the recommendation process based on probabilistic matrix factorization. The extensive experiments on two real-world data sets demonstrate that the proposed SequentialMF algorithm can achieve more accurate rating predictions than the conventional methods using either social relations or tagging information.

Key words: collaborative filtering; sequential behavior; probabilistic matrix factorization

伴随着互联网的迅速发展,网络上记录的数据量急剧增长,用户逐渐陷入信息的汪洋大海之中,快速而高效地从如此浩瀚的数据海洋中获取我们所需要的信息变得越来越紧迫.尽管传统搜索引擎可以在一定程度上解

* 基金项目: 国家自然科学基金(61073110); 国家科技支撑计划(2012BAH17B03)
收稿时间: 2013-04-30; 修改时间: 2013-07-17; 定稿时间: 2013-08-27

决用户的信息检索需求,然而它们只能呈现给所有的用户同样的排序结果,无法针对不同用户的兴趣爱好主动提供个性化的服务.在此背景下,推荐系统应运而生.具体而言,推荐系统通过收集和分析用户的各种数据来学习用户的兴趣和行为模式,从而为用户推荐它所需要的信息和服务.由于推荐系统可以有效地解决信息过载问题,因而受到来自学术界和工业界的广泛关注.

在众多推荐方法中,协同过滤是目前应用最多的算法.协同过滤的核心思想是:根据用户行为记录分析用户兴趣,在用户群中找到与目标用户(兴趣)相似的邻居用户,综合这些邻居用户对某一信息的评价,形成系统对该目标用户在产品喜好程度方面的预测,系统再根据这些喜好程度进行相应的推荐.由于协同过滤对被推荐的产品没有特殊要求,能够处理音乐、电影等难以进行文本描述的对象,因而被广泛应用于推荐系统中并取得了显著的效果,尤其是给电子商务领域带来了巨大的商业利益.例如,据 VentureBeat 统计,Amazon 的推荐系统为其提供了 35%的商品销售额[1].

虽然协同过滤推荐算法取得了巨大的成功,但其仍然存在着诸多问题,其中很重要的一点是传统的协同过滤算法经常忽略用户(产品)之间的结构关系.有效利用用户(产品)之间的联系可以丰富单个用户(产品)的信息,从而更准确地识别用户的个人兴趣.许多学者[2−8]以此为出发点,提出了基于用户(产品)关系挖掘的协同过滤算法,并取得了良好的效果.如何获取用户(产品)之间的关系并量化用户(产品)的相似程度,将对这些算法的效果产生重要的影响.现有的解决方案主要包含两种:一种是利用显式的社交网络关系[2−6];另一种则是通过隐式的标签信息来计算用户(产品)之间的相似度[7,8],得到用户(产品)之间的关系.然而在实际的数据中,获取足够的社会关系或标签信息都是比较困难的.另外,上述方法中,一般模糊地假设用户(产品)之间的相互影响关系是无向的,然而在现实生活中,这种假设并不合理.例如,A 是 B 的粉丝,B 的行为对 A 影响很大,A 对 B 的影响则微乎其微.

本文基于以上问题提出了一种基于时序消费行为的最近邻建模方法,通过构建基于时间序列的消费网络,获取用户(产品)的相互影响关系.由于该建模方法只需要用户的消费时间,不需要用户标签、社交关系等复杂信息,并且其计算的影响力是有向的,能够更精确地发现用户(产品)之间相互影响的关系,因此可以准确识别对当前用户(产品)影响最大的邻居集合,并且可应用的领域也非常广泛.在此基础上,设计了名为 SequentialMF 的推荐算法,将该邻居集合成功应用到基于概率矩阵分解的协同过滤推荐算法中.最后,本文在真实的豆瓣推荐数据集上进行了相应的实验,实验结果表明,SequentialMF 比传统使用社交网络信息与标签信息的推荐算法能够更有效地预测用户实际评分,从而提升推荐精度.

本文第 1 节回顾相关工作.第 2 节给出本文针对的研究问题、标识符号和当前流行的一个基础模型.第 3 节详细介绍和分析 SequentialMF 算法.第 4 节说明 SequentialMF 推荐框架.第 5 节对 SequentialMF 算法与其他方法进行对比实验,并对实验结果进行分析.第 6 节对全文进行总结.

1 相关工作

1.1 传统的协同过滤算法

协同过滤(collaborative filtering,简称 CF)利用与目标用户相似的用户行为(评分、点击次数等)推断目标用户对特定产品的喜好程度,然后根据这种喜好程度进行相应推荐[9].目前,协同过滤推荐算法主要包括基于近邻和基于模型两类.

基于近邻的协同过滤算法首先是根据用户的历史信息计算用户(产品)之间的相似性,然后利用与目标用户(产品)相似性较高的邻居对其他产品的评价来预测用户对特定产品的喜好程度,系统根据这一喜好程度对目标用户进行推荐.目前,基于近邻的协同过滤算法主要包括基于用户的[10]和基于产品的[11]两类.基于用户的协同过滤算法的核心在于找相似的用户,基于产品的算法主要是找相似的产品.

与基于近邻的算法不同,基于模型的协同过滤算法主要通过用户对产品的评分信息训练出相应的模型,利用此模型预测未知的数据[12].目前,基于模型的算法主要包括聚类模型[13]、概率相关模型[14]、潜在因子模型[15]、贝叶斯层次模型[16]等.最近,由于处理大数据的需要,Salakhutdinov 等人[17]提出了利用低维近似矩阵分解模型进行推荐的概率矩阵分解算法(probabilistic matrix factorization,简称 PMF),它一般假设每个用户的兴趣只受到少

数几个因素的影响,然后将用户(产品)映射到低维的特征空间中,通过用户(产品)的评分信息来学习用户(产品)的特征向量,从而重构评分矩阵,利用重构的低维矩阵预测用户对产品的评分,进行相应的推荐.由于用户和产品的特征向量维数比较低,因而可以通过梯度下降的方法高效地求解.文献[17]中的实验结果表明,基于矩阵分解的算法可以有效地处理大数据并能取得比较理想的精度.为了减少 PMF 当中参数设定对算法的影响,Salakhutdinov 等人[18]进一步提出了贝叶斯概率矩阵分解算法(Bayesian probabilistic matrix factorization,简称BPMF).BPMF 采用马尔可夫链蒙特卡洛算法进行参数估计,其推荐效果与 PMF 相比有了一定的提高.文献[19]证明了 PMF 与概率主成分分析在理论上一致,在此基础上提出了 NPMF(non-linear PMF),NPMF 通过高斯过程对 PMF 进行非线性扩展,进一步提高了算法效果.

传统的协同过滤算法虽然能够进行相应的推荐,但其往往仅采用评分信息,而与之密切相关的时序信息和关系信息则被忽略,有效利用这些信息可以进一步提高推荐算法的精度.许多学者以此为出发点,提出了基于时序信息和关系信息的推荐算法.

1.2 基于时序信息的推荐算法

基于时序信息的推荐算法通过将时序信息加入到现有的推荐模型中,使得模型能够学习到数据的动态变化,从而优化推荐效果.Koren 等人[20]提出的 TimeSVD++算法将时间信息加入到用户(产品)的特征向量中,有效解决了兴趣漂移问题,取得了较好的结果.Liang 等人[21]将时间信息作为第 3 个维度,然后利用张量分解的方式模型化动态变化.Koshneshin 等人[22]根据演化联合聚类(evolutionary co-clustering)的方式将用户(产品)动态的分配给不用的聚类,从而做进一步的推荐.Li 等人[23]认为,每个用户在特定时间段的兴趣只会集中在 1 个或几个方面,在此基础上提出了跨域协同过滤算法框架(cross-domain CF framework).实验结果表明,该算法不仅能够有效地进行推荐,还可以追踪用户的兴趣漂移.Ren 等人[24]认为,现有的推荐系统中,用户的偏好模式(preference pattern)和偏好的动态效应(preference dynamic effect)被忽略.以此为出发点,他们将用户的偏好模式规则化为一个稀疏矩阵,进而采用子空间来逐步模型化个性化和全局的偏好模式.

与上述相关工作不同,本文通过时序信息来构建用户(产品)之间的结构关系,在此基础上进行相似度计算;进而,本文将此相似度集成到概率矩阵分解算法当中,提出了一个全新的推荐框架.

1.3 基于关系挖掘的协同过滤算法

传统的协同过滤算法一般假设用户(产品)是独立的,因而忽略了用户(产品)之间的结构联系.基于用户(产品)关系挖掘的推荐算法将用户(产品)的关系融入到现有的协同过滤算法中,以丰富单个用户(产品)的信息,提高算法的精度.

基于关系挖掘的协同过滤算法核心步骤之一是获取用户(产品)的关系,目前,获取用户关系的方式主要分显式和隐式两类.Ma[2,3],Guo[4],Yang[25,26]等人提出通过显式的社交网络关系来获取用户之间的联系,在原有的评分矩阵基础上增加用户社交关系矩阵,极大地提高了算法的效果.Jamali 等人[5]提出了一种利用用户信任关系网络的随机游走模型.它通过在社交网络上进行随机游走寻找相似的产品,从而增加对目标产品的预测评分来源,减少数据稀疏性造成的影响.由于在实际的系统中获取足够的网络关系比较有难度,Zhou 等人[7]提出采用标签信息获取隐式的关系矩阵,然后通过文献[3]的方法进行相应的推荐.Wu 等人[8]利用标签信息计算近邻,并且假设近邻会直接影响用户(产品)的特征向量,为用户(产品)的特征向量增加基于近邻关系的先验;在此基础上,通过梯度下降方法学习特征向量,进一步完善了基于近邻关系的矩阵分解模型.

考虑到以上两类关系建模方法经常会受到关系数据不易大量收集的影响,本文主要是利用用户消费的时间先后信息来挖掘用户(产品)的隐式影响关系,并将用户(产品)关系融入到矩阵分解模型中,设计了基于时序行为的协同过滤推荐算法(SequentialMF),以提高推荐算法的精度.

2 问题定义和概率矩阵分解

本文的推荐算法主要是基于用户-产品评分矩阵进行相应的计算,目的是对推荐系统中用户的评分行为进

行预测.具体而言,假设在推荐系统中存在 N 个用户,其构成的集合为 $U=\{u_1,\ldots,u_N\}$,存在 M 个产品,其构成的集合为 $I=\{i_1,\ldots,i_M\}$,用户-产品评分矩阵为 $R=[R_{u,i}]_{N\times M}$,在这个评分矩阵中,$R_{u,i}$ 表示用户 u 对产品 i 的评分(比如 1~5 分).协同过滤算法利用概率矩阵分解模型学习用户(产品)的特征向量,然后基于此特征向量预测未知的评分.

假设 $U\in\boldsymbol{R}^{K\times M}$ 和 $V\in\boldsymbol{R}^{K\times N}$ 代表用户和产品的特征矩阵,其中,U_u 和 V_i 代表某个特定用户 u 和产品 i 的 K 维特征向量.概率矩阵分解模型的推荐算法的核心步骤就是学习用户和产品的特征向量.根据以上的定义,已有评分数据的条件概率定义如下:

$$p(R\mid U,V,\sigma_R^2)=\prod_{u=1}^{N}\prod_{i=1}^{M}[\boldsymbol{N}(R_{u,i}\mid g(U_u^TV_i),\sigma_R^2)]^{I_{u,i}^R} \tag{1}$$

其中,

- $\mathrm{N}(x|\mu,\sigma^2)$表示平均值为$\mu$、方差为$\sigma^2$的正态分布.
- $I_{u,i}^R$ 是一个 0-1 函数,如果 u 对 i 有评分,其值为 1;否则为 0.
- $g(x)$将 $U_u^TV_i$ 的值映射到[0,1]区间内,本文中,$g(x)=1/(1+\mathrm{e}^{-x})$.

同时,为了防止过拟合,用户和产品的特征向量均假设服从平均值为 0 的高斯先验:

$$p(U\mid\sigma_U^2)=\prod_{u=1}^{N}\boldsymbol{N}(U_u\mid 0,\sigma_U^2\boldsymbol{I}),\ p(V\mid\sigma_V^2)=\prod_{i=1}^{M}\boldsymbol{N}(V_i\mid 0,\sigma_V^2\boldsymbol{I}) \tag{2}$$

根据以上的表述,通过贝叶斯推断,特征向量 U 和 V 的后验概率如下:

$$p(U,V\mid R,\sigma_R^2,\sigma_U^2,\sigma_V^2)\propto p(R\mid U,V,\sigma_R^2)p(U\mid\sigma_U^2)p(V\mid\sigma_V^2)=\prod_{u=1}^{N}\prod_{i=1}^{M}[\boldsymbol{N}(R_{u,i}\mid g(U_u^TV_i),\sigma_R^2)]^{I_{u,i}^R}\times \prod_{u=1}^{N}\boldsymbol{N}(U_u\mid 0,\sigma_U^2\boldsymbol{I})\times\prod_{i=1}^{M}\boldsymbol{N}(V_i\mid 0,\sigma_V^2\boldsymbol{I}) \tag{3}$$

图 1 介绍了该方法[17]的图模型,根据公式(3),我们只需用户-产品评分矩阵,就可以学习出相应的特征向量.

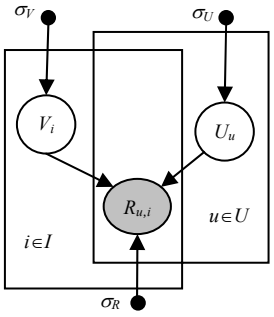

Fig.1 Graph model for probabilistic matrix factorization

图 1 概率矩阵分解图模型

3 SequentialMF 推荐算法描述

如前所述,第 2 节介绍的基础协同过滤算法忽略了用户(产品)之间的联系,因此,本节将详细描述本文所设计的 SequentialMF 推荐算法.首先介绍如何利用用户消费的时间先后信息挖掘用户(产品)间的相互影响关系,从而确定对用户(产品)影响最大的邻居集合;然后介绍如何将该邻居集合成功融合到基于概率矩阵分解的协同过滤算法中;此后,简要分析 SequentialMF 的时间复杂度;最后对算法的特性进行分析.

3.1 基于时序行为建模的最近邻选择

在基于关系的矩阵分解模型中,核心步骤之一是用户(产品)关系的获取.传统的协同过滤算法均忽略了用户或产品的消费时间信息,然而用户消费产品的时序信息可能会隐藏着一部分规律,利用这些规律可以在一定程度上挖掘用户或产品之间的联系.例如,用户 A 看了一部电影,用户 B 在 A 之后较短时间内看了同样一部电影,

如果这种情况多次出现,那么很可能 A 对 B 存在潜在的影响关系.为了发现这种潜在的关系,我们首先引入基于时序的用户消费网络图,如图 2 所示.

在这个消费网络图 $G=\{U,E\}$中,U 为用户的集合,E 为边的集合,W 表示边的权重,“()”内表示用户消费的产品数.如果在设定的时间段(例如 1 天)内 U_i 和 U_j 先后消费了同一个产品,则其边 $E_{i\to j}$ 的权重 $W_{i\to j}$ 增加 1.遍历所有的产品,符合这样的产品数目即是 U_i 到 U_j 的有向边权重 $W_{i\to j}$.根据此网络图,我们定义用户的影响关系权重如下:

$$T_{i\to j}=\frac{W_{i\to j}}{f(U_i,U_j)} \tag{4}$$

其中,$f(U_i,U_j)$为用户 U_i 和 U_j 消费产品的并集,$T_{i\to j}$ 为 i 对 j 的影响力.例如在图 2 中,假设用户 U_1 和用户 U_2 消费产品的并集数目为 100,那么 U_1 对 U_2 的影响力 $T_{1\to 2}$=5/100=0.05,而 U_2 对 U_1 的影响力 $T_{2\to 1}$=25/100=0.25.由此可以看出,两个用户之间的影响是有向的,这与现有的无向计算方法相比更具合理性.

同理,我们可以建立基于产品的消费网络图(如图 3 所示),基于产品的消费网路图与图 2 类似,其节点为产品,“()”内的数字变为消费该产品的用户数,而边权重表示有多少用户先后消费了端点的两个产品.建立相应的网络图之后,就可以利用公式(5)进行相应的影响关系计算.

$$S_{i\to j}=\frac{W_{i\to j}}{f(V_i,V_j)} \tag{5}$$

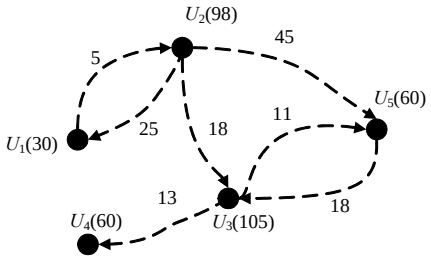

Fig.2 User consumption network

图 2 用户消费网络

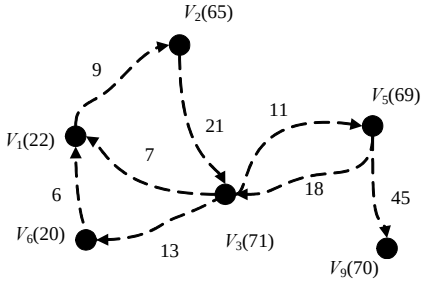

Fig.3 Item consumption network

图 3 产品消费网络

实验结果表明,本文提出的影响力计算方法简单、有效,关于其他影响力的计算方式可以参见文献[27,28].利用此影响关系的强弱,可以为特定的用户(产品)寻找对其影响最大的邻居(比如 Top-20)集合,进一步地将这些最近邻居融合到基于矩阵分解的协同过滤中,从而提高推荐系统的精度.

3.2 矩阵分解模型

挖掘出相应用户(产品)间的影响关系并找到最近邻居集合之后,将其应用到矩阵分解模型中.此时,用户(产品)的特征向量应当受到其最近邻居用户(产品)的影响,即相似的用户(产品)应当有相似的特征向量.

$$\tilde{U}_u=\sum_{v\in N_u}T_{v\to u}U_v,\ \tilde{V}_i=\sum_{j\in N_i}S_{j\to i}V_j \tag{6}$$

其中,$\tilde{U},\tilde{V}$ 表示近似的特征向量,N_u,N_i 分别表示用户 u 和产品 i 的邻居集合.与文献[6,8]不同,本文在学习特征向量时综合考虑了用户(产品)自身特征与关系用户(产品)特征的双重影响,每个用户(产品)的特征向量不仅服从平均值为 0 的高斯先验以防止过拟合,而且要与关系用户(产品)的特征向量相似.此外,由于我们考虑了时间序列信息,其影响关系更加明确,更符合实际情形:

$$p(U\mid T,\sigma_U^2,\sigma_T^2)\propto p(U\mid\sigma_U^2)\times p(U\mid T,\sigma_T^2)=\prod_{u=1}^{N}\boldsymbol{N}(U_u\mid 0,\sigma_U^2\boldsymbol{I})\times\prod_{u=1}^{N}\boldsymbol{N}\left(U_u\middle|\sum_{v\in N_u}T_{v\to u}U_v,\sigma_T^2\boldsymbol{I}\right) \tag{7}$$

$$p(V\mid S,\sigma_V^2,\sigma_S^2)\propto p(V\mid\sigma_V^2)\times p(V\mid S,\sigma_S^2)=\prod_{i=1}^{M}\boldsymbol{N}(V_i\mid 0,\sigma_V^2\boldsymbol{I})\times\prod_{i=1}^{M}\boldsymbol{N}\left(V_i\middle|\sum_{j\in N_v}S_{j\to i}V_j,\sigma_S^2\boldsymbol{I}\right) \tag{8}$$

与公式(3)相似,通过贝叶斯推断,其后验概率如下:

$$\begin{aligned}p(U,V\mid R,T,S,\sigma_R^2,\sigma_U^2,\sigma_V^2)\propto{}& p(R\mid U,V,\sigma_R^2)p(U\mid T,\sigma_U^2,\sigma_T^2)p(V\mid S,\sigma_V^2,\sigma_S^2)=\\ &\prod_{u=1}^{N}\prod_{i=1}^{M}[\boldsymbol{N}(R_{u,i}\mid g(U_u^TV_i),\sigma_R^2)]^{I_{u,i}^R}\times\prod_{u=1}^{N}\boldsymbol{N}\left(U_u\middle|\sum_{v\in N_u}T_{v\to u}U_v,\sigma_T^2\boldsymbol{I}\right)\times\\ &\prod_{i=1}^{M}\boldsymbol{N}\left(V_i\middle|\sum_{j\in N_v}S_{j\to i}V_j,\sigma_S^2\boldsymbol{I}\right)\times\prod_{u=1}^{N}\boldsymbol{N}(U_u\mid 0,\sigma_U^2\boldsymbol{I})\times\prod_{i=1}^{M}\boldsymbol{N}(V_i\mid 0,\sigma_V^2\boldsymbol{I})\end{aligned} \tag{9}$$

该模型的概率图模型如图 4 所示.

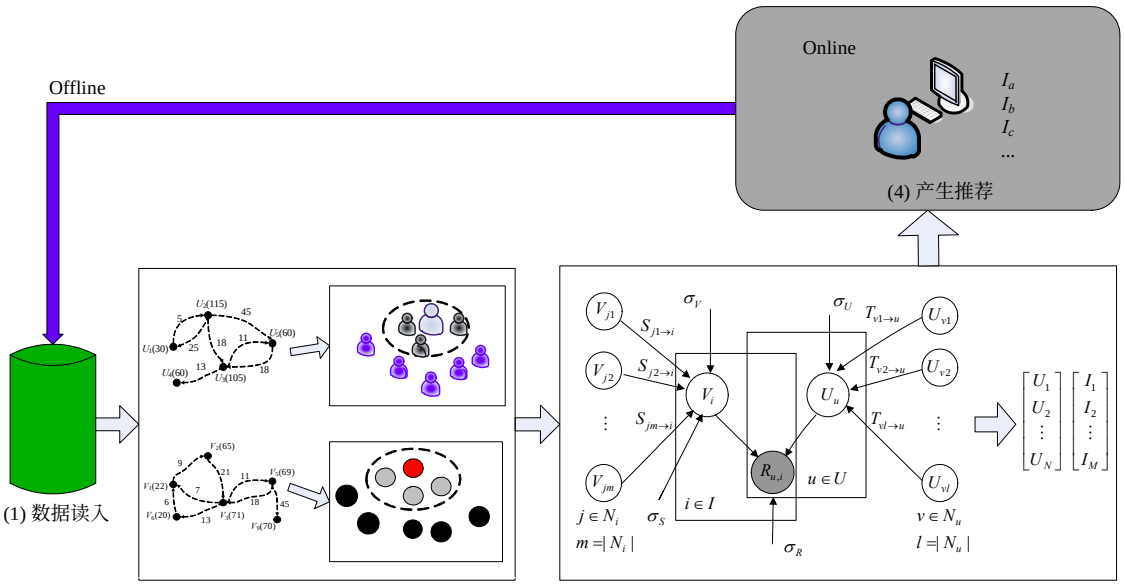

Fig.4 Framework of SequentialMF recommendation

图 4 基于时序行为的协同过滤推荐算法框架

为了便于求解,对公式(9)得到的后验概率进行对数处理如下:

$$\begin{aligned}\ln p(U,V\mid R,T,S,\sigma_R^2,\sigma_T^2,\sigma_S^2,\sigma_U^2,\sigma_V^2)=&-\frac{1}{2\sigma_R^2}\sum_{u=1}^{N}\sum_{i=1}^{M}I_{u,i}^R(R_{u,i}-g(U_u^TV_i))^2-\frac{1}{2\sigma_U^2}\sum_{u=1}^{N}U_u^TU_u-\frac{1}{2\sigma_V^2}\sum_{i=1}^{M}V_i^TV_i-\\ &\frac{1}{2\sigma_T^2}\sum_{u=1}^{N}\left(\left(U_u-\sum_{v\in N_u}T_{v\to u}U_v\right)^T\left(U_u-\sum_{v\in N_u}T_{v\to u}U_v\right)\right)-\\ &\frac{1}{2\sigma_S^2}\sum_{i=1}^{M}\left(\left(V_i-\sum_{j\in N_i}S_{j\to i}V_j\right)^T\left(V_i-\sum_{j\in N_i}S_{j\to i}V_j\right)\right)-\\ &\frac{1}{2}\left(\sum_{u=1}^{N}\sum_{i=1}^{M}I_{u,i}^R\right)\ln\sigma_R^2-\\ &\frac{1}{2}((N\times K)\ln\sigma_U^2+(M\times K)\ln\sigma_V^2+(N\times K)\ln\sigma_T^2+(M\times K)\ln\sigma_S^2)+C\end{aligned} \tag{10}$$

最大化这个后验概率,相当于最小化以下的目标函数:

$$L(R,T,S,U,V)=\frac{1}{2}\sum_{u=1}^{N}\sum_{i=1}^{M}I_{u,i}^{R}(R_{u,i}-g(U_u^TV_i))^2+\frac{\lambda_U}{2}\sum_{u=1}^{N}U_u^TU_u+\frac{\lambda_V}{2}\sum_{i=1}^{M}V_i^TV_i+\frac{\lambda_T}{2}\sum_{u=1}^{N}\left(\left(U_u-\sum_{v\in N_u}T_{v\to u}U_v\right)^T\left(U_u-\sum_{v\in N_u}T_{v\to u}U_v\right)\right)+\frac{\lambda_S}{2}\sum_{i=1}^{M}\left(\left(V_i-\sum_{j\in N_i}S_{j\to i}V_j\right)^T\left(V_i-\sum_{j\in N_i}S_{j\to i}V_j\right)\right) \tag{11}$$

在上述公式中, $\lambda_U=\sigma_R^2/\sigma_U^2,\lambda_V=\sigma_R^2/\sigma_V^2,\lambda_T=\sigma_R^2/\sigma_T^2,\lambda_S=\sigma_R^2/\sigma_S^2$. 通过梯度下降的方法,可以得到每个用户(产品)的特征向量,其梯度计算方法如下:

$$\frac{\partial L}{\partial U_u}=\sum_{i=1}^{M}I_{u,i}^{R}V_ig'(U_u^TV_i)(g(U_u^TV_i)-R_{u,i})+\lambda_UU_u+\lambda_T\left(U_u-\sum_{v\in N_u}T_{v\to u}U_v\right)-\lambda_T\sum_{\{v|u\in N_v\}}T_{u\to v}\left(U_v-\sum_{w\in N_v}T_{w\to v}U_w\right) \tag{12}$$

$$\frac{\partial L}{\partial V_i}=\sum_{u=1}^{N}I_{u,i}^{R}U_ug'(U_u^TV_i)(g(U_u^TV_i)-R_{u,i})+\lambda_VV_i+\lambda_S\left(V_i-\sum_{j\in N_i}S_{j\to i}V_j\right)-\lambda_S\sum_{\{j|i\in N_j\}}T_{i\to j}\left(V_j-\sum_{k\in N_j}S_{k\to j}V_k\right) \tag{13}$$

其中,$g'(x)$为 $g(x)$的导数.本文中, $g'(x)=\mathrm{e}^{-x}/(1+\mathrm{e}^{-x})^2$.

3.3 SequentialMF时间复杂度分析

在基于时序信息的关系挖掘推荐模型中,其计算过程主要包括两个步骤:第 1 步建立消费网络图挖掘出相应的关系,第 2 步是利用获取的影响关系进行推荐.

对于消费网络图的建立,以基于用户的网络图为例,假设平均每个产品被 $\hat{r}$ 个用户消费,对每个评分信息按照用户的评分时间进行排序,然后判断涉及这些用户的边权重是否增加.根据该过程,对每个产品建立用户的消费图的复杂度为 $O(\hat{r}^2)$, 因此建立所有用户的消费网路图的时间复杂度为 $O(N\hat{r}^2)$; 同理,假设平均每个用户消费产品数目为 $\hat{t}$, 那么构建产品的消费网络图的时间复杂度为 $O(M\hat{t}^2)$. 因此,构建消费网络图的时间复杂度为 $O(N\hat{r}^2+M\hat{t}^2)$. 建立此网络图之后,利用该网络图进行影响关系挖掘,由于每个用户平均消费 $\hat{t}$ 个产品,因此平均每个用户在网络图中有 $\hat{t}$ 条出边,依次计算每个用户对其他用户的影响力,并进行从小到大排序的复杂度为 $O(N\hat{t}^2)$; 同理,在基于产品的消费网络图中,此时间复杂度为 $O(M\hat{r}^2)$. 根据以上的分析,第 1 步总的时间复杂度为 $O((N+M)\times(\hat{r}^2+\hat{t}^2))$.

对于算法的第 2 步,根据文献[6]的分析,其计算梯度时间复杂度为 $O(N\hat{t}K+Nl^2K+M\hat{r}K+Ml^2K)$, 其中,$l$ 表示影响用户的邻居数目.由于网络图的建立只需要依次遍历评分信息即可,并不需要迭代.此外, $\hat{r}$ 和 $\hat{t}$ 往往比较小,所以该模型总的时间复杂度并不高,可以有效地对大数据进行处理.

3.4 算法讨论

SequntialMF 利用评分数据携带的时序信息构建用户(产品)之间的结构关系,在此基础上计算用户(产品)的影响关系程度,并将此影响关系集成到概率矩阵分解的推荐算法中.该算法是一种基于模型的引入关系挖掘的协同过滤算法,通过时序信息分析用户(产品)的隐形关系,所采用的推荐模型利用了群组组合的方式,从而可以更精确地预测用户行为.算法为影响关系程度计算方面提供了一个思路.同时,算法也为如何扩展基于概率矩阵分解的推荐算法提供了一定的借鉴意义.通过对算法进行复杂度分析,SequentialMF 能够有效处理数据,应用范围比较广.

4 推荐框架

第 3.1 节和第 3.2 节分别介绍了最近邻构建方式和求解优化方法,在本节中将给出本文提出的基于时序行为的推荐算法具体的推荐框架.该推荐算法一共分 4 个步骤:

(1) 读入数据,其中数据的信息应包括用户的评分信息和评分的时间信息;

(2) 系统将根据读入的信息构建用户的消费网络图和产品的消费网络图,并根据此图计算影响力最大的近邻集(第 3.1 节);

(3) 算法将该近邻集合运用到概率矩阵分解模型当中,利用概率矩阵分解模型学习出用户的特征向量和项目的特征向量(第 3.2 节);

(4) 根据该特征向量预测重构评分矩阵,利用该评分矩阵对用户形成相应的推荐.

图 4 介绍了该算法的具体过程,图中对每个步骤都进行了相应的介绍.从图中可以看出,该算法中主要的计算复杂度在步骤(2)和步骤(3)中,第 3.3 节已经对这两个步骤的时间复杂度进行了具体的分析.值得注意的是,在实际的推荐系统中,往往将计算复杂度较高的步骤(1)~步骤(3)放在线下处理,通过已有的数据,在线下学习到用户和产品的特征向量,而在线上则通过学习到的特征向量进行推荐,以提高系统的推荐速度;在线上推荐结束后,系统将用户在线上的行为数据再传输给存储系统,然后将根据新的数据更新相应的推荐模型.

从图 4 中可以更直观地看出,SequentialMF 推荐框架利用更易获取的时序信息获取关系信息,并且算法采用有向的影响力来衡量关系程度,从而进一步提高推荐精度.最后算法将计算的关系信息融入到概率矩阵分解模型当中,为如何扩展矩阵分解提供了一条思路.

5 实验结果及分析

本节首先介绍实验所用数据集,然后说明评价标准及对比算法,最后给出 SequentialMF 模型与其他方法的对比实验结果,并对实验结果进行了相应的分析.

5.1 实验数据集

为了可以比较不同信息(关系)对推荐结果的影响,实验数据集应当含有评分信息、标签信息以及用户社会关系.为此,本文使用从豆瓣网站抓取的数据作为实验数据集.豆瓣网站是一个针对电影、书籍及音乐的评价讨论网站,网站为用户提供了评分、讨论以及推荐服务.它拥有目前中国最大的中文类书籍、音乐以及电影数据库,并且是中国最大的网上社区之一.在该网站中,每个用户可以为书籍、音乐或电影做出[1,5]范围内的评分.另外,豆瓣中也提供了类似于 Facebook 的社交关系服务,它允许用户通过 E-mail 发现自己的朋友.综上所述,豆瓣数据集比较适合本文中的实验研究.

本文从豆瓣网站中抓取了两组数据集:一组数据为用户对书籍的评分信息、用户社交关系以及标签信息,另一组数据为用户对电影的评分信息及对应的其他信息.数据情况见表 1.

Table 1 Douban data set

表 1 豆瓣数据集

信息	产品类型	
	书籍	电影
用户数量	23 944	9 601
产品数量	219 725	44 779
标签数量	74 095	50 530
评分记录	1 642 111	1 960 682
社交关系	588 269	91 945

5.2 评价标准

本文实验采用 RMSE 作为评价标准.RMSE 通过计算预测的用户评分与实际的用户评分之间的偏差来度量预测的准确性.RMSE 为推荐质量提供了直观的方法,是最常用的一种推荐质量度量方法.推荐算法整体 RMSE 越小,意味着推荐的质量越高.假设算法对 C 个产品预测的评分向量表示为$\{p_1,p_2,\ldots,p_C\}$,对应的实际评分向量为$\{r_1,r_2,\ldots,r_C\}$,则算法的 RMSE 表示为

$$RMSE=\sqrt{\frac{\sum_{i=1}^{C}(p_i-r_i)^2}{C}} \tag{14}$$

5.3 比较算法及参数设定

本文选取了 4 种方法作为对比算法:

- 概率矩阵分解方法(PMF):文献[17]提出的概率矩阵分解方法.它没有考虑用户(产品)的关系.
- 引入时序信息的张量分解算法(BPTF):文献[21]提出的一种引入时序信息的推荐算法.该算法将时间作为一个张量,在此基础上,利用时间信息提高推荐效果.
- 基于社会关系推荐算法(SocialMF):文献[6]提出的模型,将用户的社交网络关系考虑到推荐模型中.此方法并没有考虑产品之间的联系.
- 基于标签信息的推荐方法(TagMF):文献[8]利用标签信息计算用户(产品)的关系,利用此关系提高矩阵分解算法的精度.此方法没有考虑影响关系的有向性.

在实验中,为了降低模型计算的复杂度,本文设定$\lambda_U=\lambda_I=0.001$,并且假设$\lambda_T=\lambda_S=\lambda$.由于 SocialMF 没有产品的关系,因此设定$\lambda_S=0$,其中,特征向量 U 以及 V 的初始值通过均值为 0 的正态分布随机抽取获得.此后,在每一代的运算中,特征向量 U 和 V 根据其前一代的值进行迭代更新,直到其收敛为止.对于 TagMF 和 SequentialMF,本文均选取 Top-20 最相似或影响力最大的用户(产品)作为目标用户(产品)的邻居,对于 SocialMF 的用户的邻居数目,根据社交关系的数据获取,而其产品的邻居数目为 0.

5.4 实验结果与分析

实验 1. 不同特征向量维度下的算法结果.

实验首先比较了各种算法在不同的特征向量维度下的结果.在实验中,我们分别设定了特征向量维度 K=5, 10,20.在实验中,算法的其他参数均设置为使各算法最优时的相应值.表 2 给出了算法在不同特征向量维度下,RMSE 的比较结果.从实验结果可以看出:

(1) 随着 K 的增大,算法精度都有一定的提高.但是需要指出的是,K 的增大在一定程度上会增加模型的时间复杂度.

(2) BPTF,TagMF,SocialMF,SequentialMF 与 PMF 相比有了较大的提高,进一步说明时序信息和用户(产品)之间的关系信息对传统的协同过滤算法精度的提高起着较大的作用;

(3) BPTF 比 SocialMF 精度有所提高,这主要是 SocialMF 中显示的社交关系比较稀疏造成的.与 TagMF 和 SequentialMF 相比,BPTF 的结果则稍差些.

(4) SocialMF 并没有 TagMF 和 SequentialMF 的精度高,这主要是由于 SocialMF 并没有考虑到产品之间的关系,另外,SocialMF 并没有考虑好友之间影响关系的有向性.

(5) SequentialMF 比 TagMF 有了进一步的提高,说明本文提出的影响关系可以有效地提高算法的精度.另外,由于 SequentialMF 只需要简单的时间信息,从信息的获取以及可应用范围来讲,SequentialMF 也具有更大的优势.

Table 2 RMSE comparisons for different setting of dimensionality K

表 2 不同维度 K 下的 RMSE 比较结果

模型	K=5		K=10		K=20	
	书籍	电影	书籍	电影	书籍	电影
PMF	0.751 1	0.735 8	0.746 5	0.732 7	0.743 5	0.731 1
BPTF	0.731 7	0.727 9	0.726 7	0.723 1	0.724 2	0.722 8
SocialMF	0.733 9	0.730 9	0.730 7	0.726 9	0.728 9	0.724 5
TagMF	0.729 8	0.726 7	0.724 0	0.723 5	0.721 9	0.721 8
SequentialMF	**0.729 4**	**0.725 1**	**0.723 8**	**0.722 9**	**0.721 7**	**0.720 8**

实验结果表明,SequentialMF 与传统的 PMF 相比有了较大的提高,充分说明了本文提出的影响关系的合理性和有效性.相对于 SocialMF,TagMF 以及 BPTF,本文的算法虽然在精度上没有十分显著的提高,然而在实际的推荐系统中,SocialMF 和 TagMF 所需要的社会网络关系或者是标签信息往往较难获取或者极其稀疏,而 SequentialMF 则只需要比较容易获取的时间信息,因此算法的实际应用场景更为广泛.而对于 BPTF,由于 BPTF

需要进行马尔可夫蒙特卡洛算法进行参数估计,其时间效率与 SequentialMF 相比大为降低.此外,SequentialMF 也为关系行为的获取以及如何扩展概率矩阵分解模型提供了新的思路.

实验 2. 算法运行时间.

在第 3.3 节,本文分析了 SequentialMF 的时间复杂度.在此实验中,我们比较了各种算法每更新一代具体的运行时间.该实验的运行环境为:Intel Core i3 CPU,2.67GHZ 主频,Windows7 系统,2G 内存.实验中设定特征向量维数 K=5,实验结果见表 3.从实验结果可以看出,算法的运行时间满足:

$$BPTF>SequentialMF\approx TagMF>SocialMF>PMF.$$

BPTF 的运行时间要远远高于另外几种算法.这主要是由于 BPTF 进行马尔可夫蒙特卡洛训练消耗过多时间,因此该算法的时间效率不高.另外,表 3 也可以说明:考虑的关系越多,其时间复杂度就越高.但是总体来看,SequentialMF 的运行时间是可以接受的.同时,从实验中可以看出:对于 PMF 和 BPTF,书籍要比电影数据的运行速度快,而其余的算法则相反.这主要是由于 PMF 和 BPTF 没有考虑邻居关系,因此其时间复杂度只与训练集的评分数据量有关系.书籍的评分数据要比电影评分数据稀少(见表 1),因而其需要的运行时间较少;但是其余算法均考虑了邻居关系,由于书籍包含的用户和产品都要比电影数据集多,因而其复杂度更高.

Table 3 Runtime comparison of a single iteration in training (s)

表 3 每代运行时间比较结果 (秒)

模型	书籍	电影
PMF	1.7	1.9
BPTF	15.2	18.6
SocialMF	2.9	2.7
TagMF	4.2	3.9
SequentialMF	4.4	4.0

实验 3. 参数λ对算法的影响.

在 SequentialMF 中,λ可以衡量用户(产品)受到其关系信息影响的程度,λ越大,表明用户(产品)影响关系对算法的作用越大.实验中为了降低复杂度,本文设定$\lambda_S=\lambda_T=\lambda$,$\lambda$的设定值分别为 0.1,0.5,1,5,10,20;此外,设定 K=5.图 5 表明了参数λ对算法的影响.根据结果可以看出:参数λ对算法有较大的影响,随着λ的增加,算法的精度不断提高.这充分说明了本文通过时序信息获取的关系的可靠性,也说明了关系的引入对算法的有效性.同时我们发现,当λ增大到 20 的时候,算法的效果开始下降.这主要是由于λ过大导致了算法的过拟合,从而导致精度的降低.

实验 4. 不同评分稀疏度下的实验结果比较.

为了比较不同稀疏度下的算法效果,本文以用户的评分数量作为划分依据,将训练集中的用户分为 4 组,评分数量分别为[0:10],[10:100],[100,500],500 以上.图 6 说明了两组数据中每组用户所占的比例.

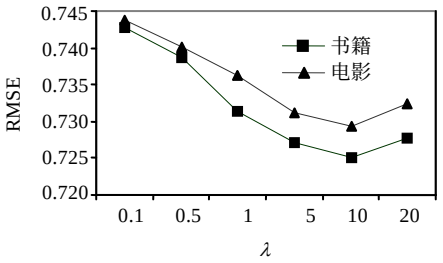

Fig.5 Impact of λ

图 5 λ的影响

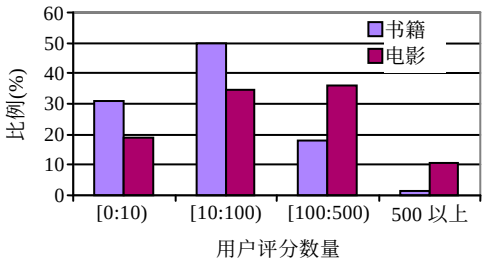

Fig.6 Distribution of different types of users

图 6 不同类型用户的分布

对数据进行相应的划分之后,本文利用训练集学习相应的模型,然后在测试集上对 4 组用户分别计算 RMSE,结果如图 7、图 8 所示.从两个数据集的实验中可以看出:

- 在数据极度稀疏的情况下(评分数量小于 10),SequentialMF 的改进效果并不明显,比引入额外信息的

SocialMF 和 TagMF 要差.这主要是因为数据稀疏会导致 SequentialMF 建立的网络图极其稀疏,从而影响其精度;而 BPTF 引入了参数训练方法,因此比 SequentialMF 效果要好.不过,SequentialMF 仍然比传统的 PMF 要好.

- 当用户评分数据增加以后,SequentialMF 比 TagMF,SocialMF 和 BPTF 都要好,从而进一步证明了引入本文中的影响关系能够有效提高推荐系统的精度.

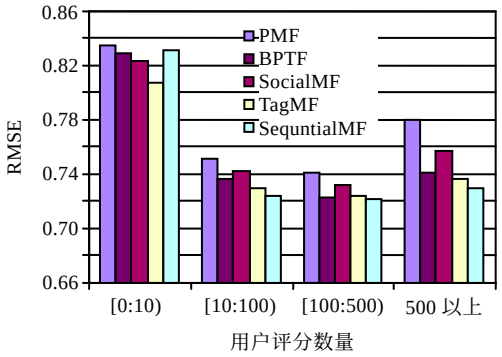

Fig.7 RMSE comparison of different types of users (book)

图 7 不同类型用户 RMSE 比较结果(书籍)

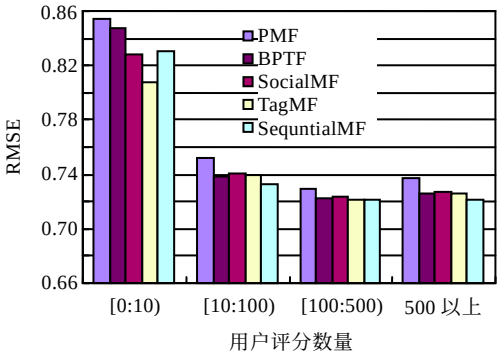

Fig.8 RMSE comparison of different types of users (movie)

图 8 不同类型用户 RMSE 比较结果(电影)

由图 7、图 8 还可以看出:无论是本文提出的 SequentialMF 还是其余算法,其 RMSE 结果并没有随着评分数量的递增而持续变小,当评分数量增大到 500 之后,算法的效果并没有持续变优而是有所下降.这主要是因为当评分数量较小时,数据的稀疏影响了算法的效果,此时的模型容易出现过拟合[8,25]现象,即其在训练样本的精度较好,而在测试样本上的精度很差;随着评分数量的增加,数据逐渐变得稠密,从而缓解了数据稀疏造成的影响,算法的效果逐渐变好;当评分的训练样本过多之后,用户的兴趣将会发散,这样将会造成模型无法学习到用户的特征喜好,从而影响模型精度.因而,控制好样本数量对于模型的预测效果具有一定的影响.从图 7、图 8 可以看出:对于本文的数据集,当用户的样本数量在[100:500)时,算法的效果最优.

6 总 结

本文利用用户消费的时序信息建立用户(产品)消费网络图,通过该网络图挖掘用户(产品)潜在的相互影响关系并寻找最近邻居集合,然后将其融入基于矩阵分解的协同过滤推荐算法中,从而提高评分预测的精度.由于与社交网络信息、标签信息相比消费时间信息更容易获取,因此,基于时序行为的协同过滤推荐算法可应用的范围更加广泛.在真实的豆瓣推荐数据集上的实验表明,该方法与传统的推荐算法相比有一定的效果提高.在今后的工作中,我们将进一步研究解决消费网络图稀疏、用户和产品的冷启动问题等给本文方法带来的挑战.此外,我们提出的关系获取方法并不仅仅局限于概率矩阵分解算法,也可以考虑把该技术应用到其他矩阵分解方法中.在接下来的工作中,我们也将研究利用该方式进一步提高推荐效果.

致谢 在此,我们向对本文的工作给予支持和建议的老师和同学表示衷心的感谢.

References:

[1] Liu JG, Zhou T, Wang BH. Research progress of personalized recommendation system. Progress in Natural Science, 2009,19(1): 1−15 (in Chinese with English abstract).

[2] Ma H, Yang HX, Lyu MR, King I. SoRec: Social recommendation using probabilistic matrix factorization. In: Proc. of the ACM Int'l Conf. on Information and Knowledge Management. ACM Press, 2008. 978−991. [doi: 10.1145/1458082.1458205]

[3] Ma H, King I, Lyu MR. Learning to recommend with social trust ensemble. In: Proc. of the Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. ACM Press, 2009. 203−210. [doi: 10.1145/1571941.1571978]

[4] Guo L, Ma J, Chen ZM, Jiang HR. Learning to recommend with social relation ensemble. In: Proc. of the ACM Int'l Conf. on Information and Knowledge Management. ACM Press, 2012. 2599−2602. [doi: 10.1145/2396761.2398701]

[5] Jamali M, Ester M. TrustWalker: A random walk model for combining trust-based and item-based recommendation. In: Proc. of the ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. ACM Press, 2009. 397−405. [doi: 10.1145/1557019.1557067]

[6] Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks. In: Proc. of the ACM Conf. on Recommender Systems. ACM Press, 2010. 135−142. [doi: 10.1145/1864708.1864736]

[7] Zhou TC, Ma H, King I, Lyu MR. UserRec: A user recommendation framework in social tagging systems. In: Proc. of the 24th AAAI Conf. on Artificial Intelligence. AAAI Press, 2010. 1486−1491.

[8] Wu L, Chen EH, Liu Q, Xu LL, Bao TF, Zhang L. Leveraging tagging for neighborhood-aware probabilistic matrix factorization. In: Proc. of the ACM Int'l Conf. on Information and Knowledge Management. ACM Press, 2012. 1854−1858. [doi: 10.1145/2396761.2398531]

[9] Liu Q, Chen EH, Xiong H, Ding CHQ, Chen J. Enhancing collaborative filtering by user interests expansion via personalized ranking. IEEE Trans. on Systems, Man and Cybernetics—B, 2012,42(1):218−233. [doi: 10.1109/TSMCB.2011.2163711]

[10] Adomavicius G, Tuzhilin A. Toward 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(16):734−749. [doi: 10.1109/TKDE.2005.99]

[11] Sarwar B, Karypis G, Konstan J, Riedl J. Item-Based collaborative filtering recommendation algorithms. In: Proc. of the 10th Int'l Conf. on World Wide Web. ACM Press, 2001. 285−295. [doi: 10.1145/371920.372071]

[12] 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]

[13] Ungar LH, Foster DP. Clustering methods for collaborative filtering. In: Proc. of the AAAI Workshop on Recommendation Systems. AAAI Press, 1998. 84−88.

[14] Getoor L, Sahami M. Using probabilistic relational models for collaborative filtering. In: Proc. of the Workshop Web Usage Analysis and User Profiling. Springer-Verlag, 2000. 83−96.

[15] Hofmann T. Collaborative filtering via gaussian probabilistic latent semantic analysis. In: Proc. of the Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. ACM Press, 2003. 259−266. [doi: 10.1145/860435.860483]

[16] Chen YH, George EI. A Bayesian model for collaborative filtering. In: Proc. of the Int'l Workshop on Artificial Intelligence and Statistics. 1999.

[17] Salakhutdinov R, Mnih A. Probabilistic matrix factorization. In: Proc. of the Annual Conf. on Neural Information Processing Systems. Curran Associates Press, 2008. 1257−1264.

[18] Salakhutdinov R, Mnih A. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In: Proc. of the 25th Int'l Conf. on Machine Learning. ACM Press, 2008. 880−887. [doi: 10.1145/1390156.1390267]

[19] Lawrence ND, Urtasun R. Non-Linear matrix factorization with Gaussian processes. In: Proc. of the 26th Annual Int'l Conf. on Machine Learning. ACM Press, 2009. 601−608. [doi: 10.1145/1553374.1553452]

[20] Koren Y. Collaborative filtering with temporal dynamics. In: Proc. of the ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. ACM Press, 2009. 89−97. [doi: 10.1145/1557019.1557072]

[21] Xiong L, Chen X, Huang TK, Schneider J, Carbonell JG. Temporal collaborative filtering with Bayesian probabilistic tensor factorization. In: Proc. of the SIAM Int'l Conf. on Data Mining. SIAM/Omnipress, 2010. 211−222.

[22] Khoshneshin M, Street WN. Incremental collaborative filtering via evolutionary co-clustering. In: Proc. of the ACM Conf. on Recommender Systems. ACM Press, 2010. 325−328. [doi: 10.1145/1864708.1864778]

[23] Li B, Zhu XQ, Li RJ, Zhang CQ, Xue XY, Wu XD. Cross-Domain collaborative filtering over time. In: Proc. of the 22nd Int'l Joint Conf. on Artificial Intelligence. IJCAI/AAAI Press, 2011. 2292−2298. [doi: 10.5591/978-1-57735-516-8/IJCAI11-382]

[24] Ren YL, Zhu TQ, Li G, Zhou WL. Top-*N* recommendations by learning user preference dynamics. In: Proc. of the Annual Conf. on Neural Information Processing Systems. Springer-Verlag, 2013. 390−401. [doi: 10.1007/978-3-642-37456-2_33]

[25] Yang XW, Steck H, Guo Y, Liu Y. On top-*K* recommendation using social networks. In: Proc. of the ACM Conf. on Recommender Systems. ACM Press, 2012. 431−438. [doi: 10.1145/2365952.2365969]

[26] Yang XW, Steck H, Liu Y. Circle-Based recommendation in online social networks. In: Proc. of the ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. ACM Press, 2012. 312−318. [doi: 10.1145/2339530.2339728]

[27] Liu Q, Xiang B, Chen EH, Ge Y, Xiong H, Bao TF, Zheng Y. Influential seed items recommendation. In: Proc. of the ACM Conf. on Recommender Systems. ACM Press, 2012. 245−248. [doi: 10.1145/2365952.2366005]

[28] Spertus E, Sahami M, Buyukkokten O. Evaluating similarity measures: A large-scale study intheorkut social network. In: Proc. of the ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. ACM Press, 2005. 678−684. [doi: 10.1145/1081870.1081956]

附中文参考文献:

[1] 刘建国,周涛,汪秉宏.个性化推荐系统的研究进展.自然科学研究进展,2009,19(1):1−15.

[12] 许海玲,吴潇,李晓东,阎宝平.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]

孙光福(1988－),男,山东济南人,硕士生,主要研究领域为个性化推荐系统,数据挖掘.
E-mail: sungf@mail.ustc.edu.cn

吴乐(1988－),女,博士生,主要研究领域为数据挖掘,推荐系统.
E-mail: wule@mail.ustc.edu.cn

刘淇(1986－),男,博士,副研究员,主要研究领域为数据挖掘,推荐系统,社交网络,情境感知的数据挖掘.
E-mail: qiliuql@ustc.edu.cn

朱琛(1990－),男,硕士生,主要研究领域为数据挖掘.
E-mail: zc3930155@gmail.com

陈恩红(1968－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为机器学习与数据挖掘,个性化推荐系统,社会网络.
E-mail: cheneh@ustc.edu.cn