

用户非对称信任关系的推荐算法

张紫茵 张恒汝 徐媛媛 秦 琴

(西南石油大学计算机科学学院 成都 610500)

摘 要 数据稀疏性是目前协同过滤面临的主要挑战之一。用户间的信任关系为推荐系统提供了有用的附加信息。已有工作主要采用直接信任关系作为附加信息,对间接信任关系考虑得较少。针对这一情况,提出一种融合直接和间接的用户非对称信任关系的推荐算法(ATRec)。首先,构建一种信任传递机制,并利用该机制获得用户间的间接非对称信任关系。其次,根据直接和间接非对称信任关系获得每个用户的信任集合。最后,利用信任集合、最近邻的评分和好评阈值计算出商品的受欢迎程度,进而获得对用户的 top-N 推荐列表。在真实数据集上的实验结果表明,该算法比主流的推荐算法在 top-N 推荐性能上有更好的表现。

关键词 非对称信任关系,推荐系统,个性化推荐,信任传递机制

中图分类号 TP181 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.10.007

Recommendation Algorithm Combining User's Asymmetric Trust Relationships

ZHANG Zi-yin ZHANG Heng-ru XU Yuan-yuan QIN Qin

(School of Computer Science, Southwest Petroleum University, Chengdu 610500, China)

Abstract Data sparsity is one of the major challenges faced by collaborative filtering. Trust relationships between users provide useful additional information for the recommender system. In the existing studies, the direct trust relationships are mainly used as additional information, while the indirect trust relationships are less considered. This paper proposed a recommendation algorithm(ATRec) that combines the direct and indirect asymmetric trust relationships. First, a trust transfer mechanism is constructed and used to obtain asymmetric indirect trust relationships between users. Second, each user's trust set is obtained by the direct and indirect asymmetric trust relationship. At last, the popularity of the item is computed according to the rating information of the trust set or the k-nearest neighbors and the favorable threshold, thus generating user's top-N recommendation list by the recommended threshold. The experimental results show that this algorithm has better performance than the state-of-the-art recommendation algorithms in top-N recommendation.

Keywords Asymmetric trust relationship, Recommender system, Personalized recommendation, Trust transfer mechanism

1 引言

数据稀疏性^[1]是推荐系统面临的主要挑战之一。电子商务网站^[2]提供了数以万计的商品信息,单个用户浏览的商品仅占其中很少的一部分,售后评价的占比更小。例如,Epinions.com 在线服务网站提供的数据集^[3]中,有 4.9 万个用户和 13.9 万个商品,但实际评分次数只有 66 万,密度仅为 0.009%。最近邻算法^[4]是推荐系统中应用最广泛的技术之一,它主要采用评分信息/购买记录来寻找用户最近邻,并利用最近邻为用户推荐商品。这种技术的核心是计算用户间的相似度。在用户评分信息比较完善的情况下,该相似度能够较准确地反映用户购买习惯的异同^[5]。但当数据稀疏时,这

种异同不能准确反映出来,从而使得推荐质量急剧下降^[6]。

用户间的信任关系为推荐系统提供了有用的附加信息。Sinha 等^[7]的研究表明,朋友推荐的结果比原始推荐系统的推荐结果更新颖,即朋友提出的建议更有用。John 等^[5]认为在传统的推荐中主要重视用户间的相似性,信任因素可能在推荐系统中起着重要作用。Paolo 等^[8]认为信任关系可以在推荐系统中被提取和使用,且信任关系是一个缓解寻找相似用户不准确问题的突破口。这些工作主要采用直接信任关系作为附加信息来提高推荐性能,而对间接信任关系考虑得较少。

本文提出一种融合直接和间接的用户非对称信任关系的推荐算法(ATRec)。这种信任关系与生活中的社交关系不同^[9]:社交关系是指周围的关系,如亲戚、同事和同学等,此种

收稿日期:2018-04-16 返修日期:2018-05-16 本文受国家自然科学基金项目(41604114,61379089)资助。

张紫茵(1995-),女,硕士,CCF 会员,主要研究方向为推荐系统;张恒汝(1975-),男,副教授,CCF 会员,主要研究方向为数据挖掘、推荐系统和粒计算,E-mail:zhanghrswpu@163.com(通信作者);徐媛媛(1982-),女,硕士,CCF 会员,主要研究方向为信号处理和推荐系统;秦 琴(1997-),女,硕士,主要研究方向为推荐系统。

关系通常是对称的;而本文所采用的信任关系是非对称的,即A信任B,但B不一定信任A。本算法的输入包括信任关系矩阵和用户-商品的评分矩阵,输出商品的推荐列表,优化目标是 top-N 推荐列表的性能。首先,本文建立一种信任传递机制来获取用户的非对称间接信任关系。其次,获得用户的直接和间接信任集合。最后,利用信任集合、最近邻的评分和好评阈值计算得出商品的受欢迎程度,进而得到用户的 top-N 个商品的推荐列表。

在真实数据集上的实验结果表明,ATRec 算法在不同数据集上,直接和间接信任用户的权重对推荐结果的影响不同;在密度更大的数据集中,直接朋友的权重越大,推荐质量越好。实验得出,信任用户的偏好的影响力略大于邻居的偏好的影响力。ATRec 的精确率 Precision 在大多数实验中高于相关算法,而综合指标 F_1 -measure 在部分实验中略低于相关算法。

2 相关工作

推荐系统从大数据中为用户推荐其感兴趣的商品,是信息过滤中的一种重要技术。主要的推荐算法包括以下几类:基于内容推荐、基于知识推荐、基于组合推荐和基于协同过滤推荐^[10]。最近邻算法是基于协同过滤推荐算法中较为流行的一种算法。基于用户的最近邻算法的核心是用户间的相似度,根据相似度找出当前用户的最近邻,利用最近邻预测当前用户对商品的偏好。该算法考虑了全局的评价信息,增加了推荐可利用的信息量,提高了推荐的质量和效率。

数据的稀疏性影响了最近邻的准确性,从而影响了推荐的精确率。一些学者将信任以及其他因素相融合来解决这个问题,并结合协同过滤来获得更高的推荐精确率。Sinha 等^[7]对 19 个拥有不同职业的测试者进行了问卷调查,且对基于 6 个推荐系统得出的建议与其朋友的建议进行比较后得出后者的建议更有用。Paolo 等^[8]提出了一种推荐系统的信任感知方法,将协同过滤和信任传播结合起来,与协同过滤的不同之处在于 Paolo 提出的方法是将评分信息和信任关系加权得出最近邻。上述研究均缓解了数据稀疏性带来的问题,同时提高了推荐精确率,保证了推荐的多样性。根据结论得出,用户间的信任关系可以缓解协同过滤中数据稀疏带来的问题,从而提高推荐的精确率。

3 数据模型

本文所使用的数据集包括用户对商品的评分以及用户间的信任数据。

3.1 用户-商品评分矩阵

数据集中有 n 个用户及 m 个商品。本文中 $U = \{u_0, u_1, \dots, u_{n-1}\}$ 表示用户的集合, $E = \{e_0, e_1, \dots, e_{m-1}\}$ 表示商品的集合。用户对商品的评分可以构成一个评分矩阵,表示为:

$$R: U \times E \rightarrow r \quad (1)$$

其中, R 是一个 $n \times m$ 阶的矩阵。矩阵 R 中的元素 r_{ij} 是用户 u_i 对商品 e_j 的评分,表示为 $r_{ij} = R(u_i, e_j)$,其中 $i \in [0, n-1]$

且 $j \in [0, m-1]$ 。如表 1 所列,用户对商品的评分为离散值,即 $\{1, 2, 3, 4, 5\}$ 。

表 1 用户-商品的评分矩阵

Table 1 User-item rating matrix

	e_0	e_1	e_2	e_3	e_4
u_0	0	?	4	5	0
u_1	3	4	?	0	3
u_2	3	0	4	0	0
u_3	0	2	?	1	0
u_4	2	5	0	1	1
u_5	0	0	2	3	5

注:0 表示用户对商品没有评分;? 表示需要预测用户对商品的偏好

3.2 非对称信任关系矩阵

信任是指直接联系的用户间存在的信任。图 1 给出一个信任关系的示例,其中每个节点代表一个用户,节点间存在的有向边代表用户间存在信任关系。

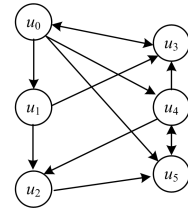


图 1 用户间的信任关系

Fig. 1 Trust relations among users

从图 1 可得,用户 u_0 和用户 u_1 间只有一条有向边,表示 u_0 信任 u_1 ,但并不代表 u_1 也信任 u_0 ;而 u_1 与 u_3 互相信任,因为这两个节点之间双向连接。表 2 列出图 1 所对应的信任关系。矩阵中,数值 1 或 0 反映了用户间是否存在直接信任关系。

表 2 非对称信任关系矩阵

Table 2 Asymmetric trust relationship matrix

	u_0	u_1	u_2	u_3	u_4	u_5
u_0	0	1	0	1	1	1
u_1	0	0	1	1	0	0
u_2	0	0	0	0	0	1
u_3	1	0	0	0	0	0
u_4	0	0	1	1	0	1
u_5	0	0	0	0	1	0

(1) 直接信任用户

若用户 u_i 和用户 u_j 之间存在直接信任关系,即他们的信任关系为:

$$T(u_i, u_j) = \begin{cases} 1, & u_i \text{ trust } u_j \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

若 $T(u_i, u_j) = 1$, 则 u_j 是 u_i 的直接信任用户。例如, $T(u_0, u_1) = 1$ 表示 u_1 是 u_0 的直接信任用户,而 $T(u_0, u_2) = 0$ 表示 u_2 不是 u_0 的直接信任用户。因此,可以得出直接信任用户的集合,记为:

$$C_i = \{u_k \mid T(u_i, u_j) = 1, k \in [0, n-1]\} \quad (3)$$

在 C_i 中的每一个用户都是 u_i 的直接信任用户。如图 1 中的 $C_0 = \{u_1, u_3, u_4, u_5\}$ 。

(2) 间接信任用户

根据信任关系的传递性,当直接信任用户无法预测偏好时,可以寻找间接信任用户的预测偏好。 C_i 中用户的直接信

任用户就是 u_i 的间接信任用户,记为:

$$C_i' = \{u_k | T(u_x, u_k) = 1, u_x \in C_i, k \in [0, n-1]\} \quad (4)$$

其中, u_x 为 u_i 的间接信任用户,组成的间接信任用户集合为 C_i' 。如图 1 中的 $C_0' = \{u_2, u_3\}$ 。

然而,不是每一个信任的用户都可以影响预测结果,只有其中对预测商品 e_j 有评分的用户才可以预测结果。该用户集合称为可预测用户集合,记为:

$$U_{ij} = \{u_k | \forall r_{ij} > 0, u_k \in C_i \cup C_i'\} \quad (5)$$

U_{ij} 表示信任用户中对商品 e_j 有评分的用户。根据信任用户的评分,可以得出其中评分大于 t 的次数,记为:

$$h(t) = |U'_{ij}|, U'_{ij} = \{u_k | r_{ij} > t, u_k \in U_{ij}\} \quad (6)$$

4 融合用户非对称信任关系的推荐算法

本文提出了一种融合用户非对称信任关系的推荐算法(ATRec)。首先根据 T 获得目标用户的信任用户集合,然后利用该集合中用户对预测商品的评分预测目标用户的偏好,最后得出 top-N 个商品的推荐列表。该算法的具体步骤如算法 1 所示。

算法 1 ATRec 算法

输入:用户 u_i , 评分矩阵 R , 信任矩阵 T , 好评阈值 α , 信任权重 β , 邻居权重 γ , 邻居数量 k , 推荐个数 N

输出:top-N 个商品的推荐列表 L_i

1. for (each e_j in E) do
2. $C_i \cup C_i' \leftarrow T$;
3. $U_{ij} \leftarrow R$;
4. if ($U_{ij} = \emptyset$) then
5. $U_{ij} \leftarrow kNN(u_i, e_j, k)$;
6. end if
7. get the ratings' distributions in U_{ij} by R ;
8. Compute the $P(j, \alpha) = h(\alpha) / h(1)$; /* Based on β to get the item's popularity, and record it */
9. end for
10. orderItemPopularity(); /* Sorting the popularity of each item by γ , and return the item sequence. */
11. get the top-N recommended list L_i ;
12. return L_i ;

步骤 2—步骤 6 主要用于寻找商品的可预测用户集合 U_{ij} 。首先利用信任关系矩阵获得 u_i 的直接或间接信任用户,然后检查是否能从中得到 U_{ij} 。若不能得到 U_{ij} ,则利用最近邻算法寻找 k 个邻居作为可预测用户集合 U_{ij} 。 $kNN(u_i, e_j, k)$ 是最近邻算法,返回 u_i 的 k 个对 e_j 有评分的邻居。以预测用户 u_0 对商品 e_1 的偏好为例,在图 1 中可得 $C_0 = \{u_1, u_4, u_5\}$, $C_0' = \{u_2, u_3\}$,则根据表 1 得出对 e_1 有评分的用户集合 $U_{01} = \{u_1, u_3, u_4\}$ 。

在步骤 7 中,获取可预测用户集合 U_{ij} 对 e_j 的评分分布(评分分布是用户对商品的不同评分次数)。图 2 描述了根据 U_{01} 得到评分分布的过程。

用户 u_0 通过信任关系获得信任用户集合 U_{01} 后,根据 U_{01} 对 e_1 的评分得到评分分布(评分分布是用户对商品的不同评分次数)。图 2 表示根据其余用户对 e_1 的评分情况,得出评分分布。可以看出:对于 e_1 ,在 U_{01} 中,有 1 次 2 分评分,1 次 4 分评分,1 次 5 分评分。

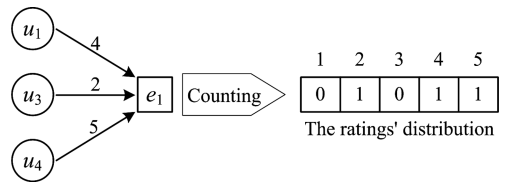


图 2 评分分布的获取过程

Fig. 2 Process of getting ratings' distribution

在步骤 8 中,根据评分分布预测该用户对目标商品的偏好,在此设定好评阈值 α 和信任权重 β 。如果用户对商品的评分大于好评阈值 α ,则认为用户喜欢该商品。在 U_{ij} 中, β 表示间接信任用户的偏好对受欢迎程度的影响因子,即直接和间接信任用户的影响比为 $1:\beta$ 。如果在 U_{ij} 中有 q 个间接用户,则基于 β 可以得出 U_{ij} 对 e_j 的欢迎程度为:

$$P(j, \alpha) = \frac{h(\alpha) - q + q \times \beta}{h(0) - q + q \times \beta} \quad (7)$$

其中, $P(j, \alpha)$ 表示引入 β 后, U_{ij} 中的用户对 e_j 的评分大于 α 的概率。在步骤 2 中得知: U_{01} 中只有 u_3 一个间接信任用户,假设 $\alpha=3, \beta=0.5$,则上述例子中 $P(j, \alpha) = P(1, 3) = [h(3) - 1 + 1 \times 0.5] / [h(0) - 1 + 1 \times 0.5] = [2 - 1 + 1 \times 0.5] / [3 - 1 + 1 \times 0.5] = 1.5 / 2.5 = 0.6$ 。

重复步骤 2—步骤 8,可以得到所有商品的受欢迎程度。

在步骤 10 和步骤 11 中,对商品的受欢迎程度进行排序,可以得出用户 u_i 的 top-N 推荐列表。orderItemPopularity() 表示根据受欢迎程度对商品进行排序,返回的是排序后的所有商品列表。由于每件商品的受欢迎程度可能来自于信任用户或邻居,因此本文在排序时设定邻居权重 γ ,即邻居的偏好对推荐的影响因子,表明信任用户和邻居的影响比为 $1:\gamma$ 。假如信任用户和邻居推荐的商品分别为 e_1 和 e_2 ,其受欢迎程度均为 0.6,那么在排序时 e_1 的受欢迎程度为 0.6, e_2 的受欢迎程度更新为 $0.6 \times \gamma$ 。由更新后的受欢迎程度对商品进行排序,最后选取排序后的 top-N 个商品作为每个用户的推荐列表。

5 实验与结果

5.1 数据集

本文使用 FilmTrust 以及 Epinions 真实数据集对本文算法进行实验。这两个数据集都来源于不同的在线服务网站。用户可以在这些网站上对商品或电影进行评分,也可以将其余用户加入到自己的信任列表中。数据集信息如表 3 所列。

实验过程中为了避免偶然性,会将数据集随机划分成测试集和训练集。在测试集中,根据用户的评分可以得到用户喜欢的商品列表,记为:

$$L_i' = \{e_j | r_{ij} > \alpha, j \in [0, m-1]\} \quad (8)$$

其中, α 是好评阈值, r_{ij} 是用户 u_i 对商品 e_j 的评分,即评分大于 α 时,就将商品放入 L_i' 中。在训练集中,根据测试集中的评分信息可以得出商品的推荐列表 L_i 。如果商品 $e_j \in L_i \cap L_i'$,则本文视此情况为推荐正确,反之为推荐错误。最后根据评价指标得出算法的推荐质量。

表3 数据集信息

Table 3 Information of data sets

Datasets	Users	Items	Ratings	Relations	Rating density/%	Trust density/%
FilmTrust	1642	2071	35497	1853	1.044	0.069
Epinions	49290	139738	664824	487183	0.001	0.020

5.2 评价指标

本文用精确率 Precision 和综合指标 F_1 -measure 作为推荐质量的评价指标。这两个指标被广泛用于信息检索和统计学分类领域来评价结果的质量。

Precision 是推荐正确的商品数与推荐的商品总数的比率,用于衡量检索系统的查准率。定义为:

$$Precision = \frac{|L_i \cap L_i'|}{|L_i|} \quad (9)$$

其中, L_i 表示向用户 u_i 推荐的商品列表, L_i' 表示用户 u_i 喜欢的商品列表。

F_1 -measure 是 Precision 和 Recall 的调和平均值。Recall 是指推荐正确的商品数和喜欢的商品总数的比率,衡量的是检索系统的查全率。定义为:

$$F_1\text{-measure} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

5.3 实验设计

在实验中,首先设置训练集和测试集。从每个用户的评分信息中随机选择 g 个评分信息,并把选出的所有信息作为训练集,剩下的作为测试集。为了保证每个用户的推荐列表中至少有 10 个商品,需要过滤掉评分次数少于 $(g+10)$ 的用户。考虑到实验的可比性,本文采用与文献[11]相同的设置,即 $g=10, 20, 50$, 即要过滤掉评分次数少于 20, 30, 60 的用户。

本文将与以下主流的推荐方法进行比较。在实验中采用与对比算法相同的参数,取推荐列表中 top- N 个推荐商品。实验分为参数调整和对比实验两部分。在实验结果中“@ N ”表示取推荐列表的前 N 个推荐商品,本文设定 $N=5, 10$ 。

参数调整实验中,主要针对算法中的主要参数 β 和 γ 进行讨论。在不同数据集的不同划分方式下进行实验并分析实验结果。在实验中,“ $g@N$ ”表示在数据集划分为 g 的情况下推荐前 N 个商品,“10@5”表示在 $g=10$ 的数据集划分情况下推荐前 5 个商品。

对比实验中,主要将 ATRec 算法在 top- N 上的推荐性能与目前主流的算法进行比较。对比实验分为 2 种不同的类别。

(1) 基于评分的推荐方法:WRMF^[12],带有隐式反馈数据的加权矩阵分解推荐算法;BPRMF^[13],基于贝叶斯的矩阵分解推荐算法;ListRank^[14],基于列表式学习排序的推荐算法。

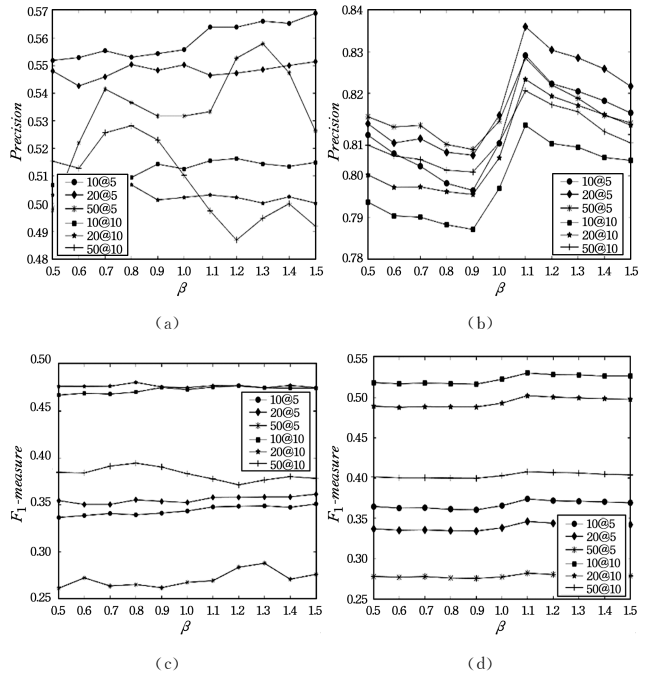
(2) 基于附加信息的推荐方法:SBPR^[15],基于社会信息的 BPRMF 扩展推荐算法;SoRank^[16],基于社会信息的列表式排名学习的推荐算法;TRecSo^[11],基于排序的推荐方法;ATRec,本文算法。

本文选择基于评分的推荐方法是为了验证信任关系为推荐提供了有效信息,由于附加信息的推荐方法与本文算法使用了相同的附加信息,以此可知 ATRec 的推荐精确率更高。

5.4 参数调整实验

ATRec 算法中的参数有 α, β, γ 和 k ,主要的参数为 β 和 γ 。其中,参数 α 为好评阈值,若评分高于 α ,则表示用户喜欢该商品,实验中设定 $\alpha=3$,表示用户对喜欢的商品的评分大于 3;参数 k 为最近邻数,由于本文研究的重点不在于最近邻,因此实验中设定 $k=10$ 。

参数 β 为信任权重,表示间接信任用户对受欢迎程度的影响大小。实验选取 β 为 $[0.5, 1.5]$ 区间的离散值,暂定 $\gamma=1$ 。图 3 表示在 FilmTrust 数据集和 Epinions 数据集上推荐性能随 β 变化的结果,横坐标为 β 。图 3 给出了 β 对在不同数据集的不同划分方式下的推荐性能的影响,其中,图 3(a)和图 3(b)分别给出不同数据集的 Precision,图 3(c)和图 3(d)分别给出不同数据集的 F_1 -measure。

图3 信任权重 β 对推荐性能的影响Fig. 3 Influence of β on recommendation performance

可以看出:在图 3 中, β 对同一个数据集的不同划分方式下的 Precision 和 F_1 -measure 的影响均不一样,但 F_1 -measure 的变化和 Precision 的变化基本一致。图 3(a)和图 3(c)给出了在 FilmTrust 数据集上 β 对推荐性能的影响,其 Precision 的变化幅度在 5% 以内,而 F_1 -measure 的变化幅度在 3% 以内。图 3(b)和图 3(d)为在 Epinions 数据集上 β 对推荐性能的影响,其 Precision 的变化幅度在 3% 以内,而 F_1 -measure 的变化幅度在 2% 以内。从整体来看,在 FilmTrust 数据集上,当 $\beta=0.8$,即直接和间接信任用户偏好的影响为 1:0.8 时, Precision 最好;在 Epinions 数据集上,当 $\beta=1.1$,即直接和间接信任用户偏好的影响为 1:1.1 时, Precision 最好。

参数 γ 为邻居权重,表示邻居的偏好对推荐的影响大小。

实验选取 γ 为 $[0.5, 1.5]$ 中的离散值, 设定 $\beta = 1.1$ 。图 4 表示在 FilmTrust 数据集和 Epinions 数据集上推荐性能随 γ 变化的结果, 即在不同数据集的不同划分方式下 γ 对数据集的推荐性能的影响, 横坐标为 γ 。其中, 图 4(a) 和图 4(b) 分别给出不同数据集的 Precision, 图 4(c) 和图 4(d) 分别给出不同数据集的 F_1 -measure。

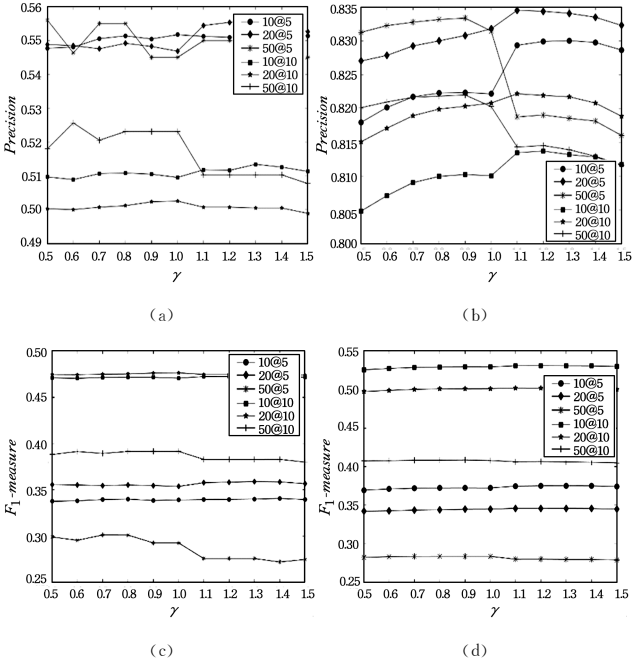


图 4 邻居权重 γ 对推荐性能的影响

Fig. 4 Influence of γ on recommendation performance

从图 4 中可以看出: 在同一条件下, 当 γ 逐渐增大时, Precision 呈现明显“逐渐升高”或“逐渐降低”的趋势, 但 F_1 -measure 变化较小; 当 γ 从 1 逐渐增大到 1.1 时, 大多条件下 Precision 有明显升高或降低的趋势, F_1 -measure 也有小幅度升高或降低的趋势, 但总的来说降低幅度更大。图 4(a) 和图 4(b) 给出在不同数据集上 γ 对 Precision 的影响情况, 其变化幅度在 2% 以内; 图 4(c) 和图 4(d) 给出在不同数据集上 γ 对 F_1 -measure 的影响情况, 其变化幅度在 2% 以内。整体来看, 当 $\gamma = 0.9$, 即邻居偏好所占权重稍小时, Precision 的效果最好。

可以看出: β 和 γ 对推荐质量的影响并不是特别突出, 这是因为本算法倾向于对热门产品的推荐, 所以无论是间接信任用户还是邻居偏好, 对 top-N 推荐列表的 Precision 的影响均不大; 而 F_1 -measure 是 Precision 和 Recall 的调和平均值, 因此其受影响程度也不大。

5.5 对比实验

按照 5.3 节的对比实验设计进行实验。根据参数调整实验结果, 在 FilmTrust 数据集上设置 $\beta = 0.8, \gamma = 0.9$; 在 Epinions 数据集上设置 $\beta = 1.1, \gamma = 0.9$ 。图 5 和图 6 为 FilmTrust 数据集、Epinions 数据集上的 Precision 和 F_1 -measure 的 top-N 比较实验。在此实验结果中, “Precision@N” 表示推荐前 N 个商品时的精确率, 而 “F1-measure@N” 表示推荐前 N 个商品时的综合评价结果。

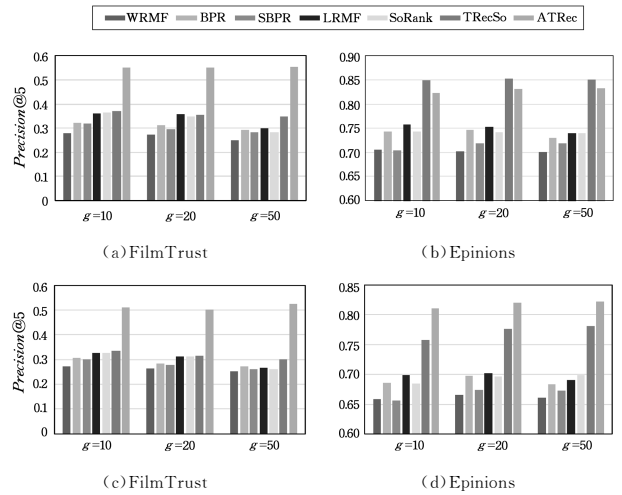


图 5 Precision 的 top-N 比较实验

Fig. 5 Top-N comparison of Precision

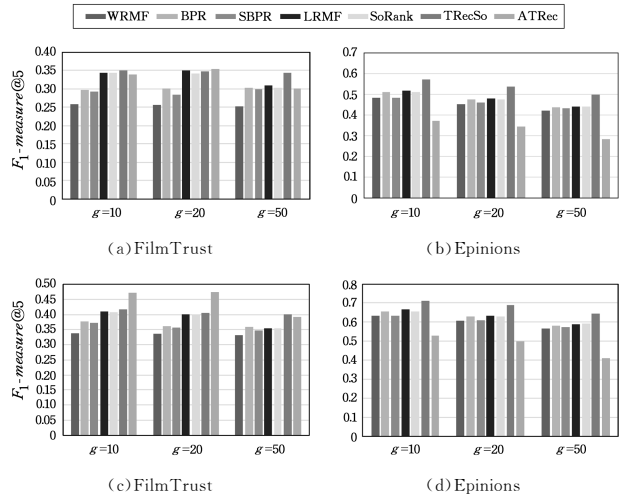


图 6 F_1 -measure 的 top-N 比较实验

Fig. 6 Top-N comparison of F_1 -measure

从图 5 可以看出, ATRec 在不同数据集上的 Precision 高于大多数算法。在 FilmTrust 数据集中, ATRec 比 TRecSo 在 Precision 上平均提高 19.4%。在 Epinions 数据集中, ATRec 比 TRecSo 在 Precision 上平均提高 1.2%。从图 6 可以看出: ATRec 在不同数据集上的 F_1 -measure 略低于 TRecSo。在 FilmTrust 数据集中, ATRec 比 TRecSo 在 F_1 -measure 上平均降低 1.2%; 在 Epinions 数据集中, ATRec 比 TRecSo 在 F_1 -measure 上平均降低 20%。

同时, 从以上实验中也可以看出:

(1) 随着滤除用户的增多, Precision 和 F_1 -measure 均越来越小, 原因可能在于少数信任用户存在不准确推荐。因为滤除的用户越多, 利用信任关系找到的用户集合可能越小。根据信任关系得到的受欢迎程度可能只从很少一部分或单个信任用户的偏好而来, 由此可能得到不准确的推荐。

(2) 在 Epinions 数据集中推荐 top-5 个商品时, ATRec 的效果均略差于 TRecSo, 其原因在于数据集评分密度低使得寻找邻居不准确, 从而导致推荐质量不理想。Epinions 数据集本身的评分密度就比较稀疏, 在数据集划分时还过滤掉了一部分用户, 不仅存在少数信任用户推荐不准确的情况, 还可能

使得在寻找 k 个相似邻居时存在不精准的情况,导致推荐质量不理想。

结束语 针对数据的稀疏性,本文引入了非对称信任关系,建立了基于用户信任关系的推荐算法(ATRec)。ATRec首先通过建立的信任传递机制,得到用户的非对称间接信任关系;其次,融合直接和间接信任关系,获取用户的信任集合;最后,利用信任集合、最近邻的信息和好评阈值计算商品的受欢迎程度,进而获取 top-N 推荐列表。在 FilmTrust 和 Epinions 数据集中的实验表明,ATRec 中直接与间接信任用户的影响因子差别不大;而且与目前的主流算法 TRecSo 相比,ATRec 在推荐 top-N 商品的 Precision 上最高可提升 22.1%,而 F_1 -measure 略低。本文方法可推广到对称信任关系,并且可以轻松应用于现实场景中,能取得较好的推荐结果。但在推荐本质中,ATRec 倾向于将热门商品推荐给用户,下一步研究工作可能会考虑到冷门商品的推荐,实现推荐结果的多样性。

参考文献

- [1] MASSA P, AVESANI P. Trust-aware recommender systems [C]//2007 ACM Conference on Recommender Systems. ACM, 2007:17-24.
- [2] ZHAO W X, LI S, HE Y, et al. Connecting social media to e-commerce: Cold-start product recommendation using microblogging information[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(5):1147-1159.
- [3] ZHOU J, TANG M, TIAN Y, et al. Social network and tag sources based augmenting collaborative recommender system [J]. IEICE Transactions on Information and Systems, 2015, 98(4):902-910.
- [4] ZHANG H R, MIN F. Three-way recommender systems based on random forests[J]. Knowledge-Based Systems, 2016, 91(C): 275-286.
- [5] O'DONOVAN J, SMYTH B. Trust in recommender systems [C]//10th International Conference on Intelligent User Interfaces. ACM, 2005:167-174.
- [6] DENG A L, ZHU Y Y, SHI B L. A collaborative filtering recommendation algorithm based on item rating prediction[J]. Journal of Software, 2003, 14(9):1621-1628. (in Chinese)
- [7] SINHA R R, SWEARINGEN K. Comparing recommendations made by online systems and friends [C]//DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries. 2001.
- [8] MASSA P, BHATTACHARJEE B. Using trust in recommender systems: an experimental analysis[C]//International Conference on Trust Management. Berlin:Springer, 2004:221-235.
- [9] MA H, ZHOU D, LIU C, et al. Recommender systems with social regularization[C]//Forth International Conference on Web Search & Web Data Mining. DBLP, 2011:287-296.
- [10] HERLOCKER J L, KONSTAN J A, RIEDL J. Explaining collaborative filtering recommendations[C]//2000 ACM Conference on Computer Supported Cooperative Work. ACM, 2000: 241-250.
- [11] PARK C, KIM D, OH J, et al. Improving top-K recommendation with trustor and trustee relationship in user trust network[J]. Information Sciences An International Journal, 2016, 374(C): 100-114.
- [12] PAN R, ZHOU Y, CAO B, et al. One-class collaborative filtering [C]//Eighth IEEE International Conference on Data Mining, 2008(ICDM'08). IEEE, 2008:502-511.
- [13] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback [C]//Twenty-fifth Conference on Uncertainty in Artificial Intelligence. AUAI Press, 2009:452-461.
- [14] SHI Y, LARSON M, HANJALIC A. List-wise learning to rank with matrix factorization for collaborative filtering[C]//Fourth ACM Conference on Recommender Systems. ACM, 2010: 269-272.
- [15] ZHAO T, MCAULEY J, KING I. Leveraging social connections to improve personalized ranking for collaborative filtering[C]//23rd ACM International Conference on Information and Knowledge Management. ACM, 2014:261-270.
- [16] YAO W, HE J, HUANG G, et al. SoRank: incorporating social information into learning to rank models for recommendation [C]//23rd International Conference on World Wide Web. ACM, 2014:409-410.
- [17] LI J H, LV Y J. Attribute reduction and rules extraction in decision formal context based on concept lattice[J]. Mathematics in Practice and Theory, 2009, 39(7):182-188. (in Chinese)
李金海,吕跃进.基于概念格的决策形式背景属性约简及规则提取[J].数学的实践与认知,2009,39(7):182-188.
- [18] LI J H, LV Y J. Knowledge reduction in decision formal contexts[J]. Knowledge-Based Systems, 2011, 24(5):709-715.
- [19] SHAO M W, LEUNG Y, WU W Z. Rule acquisition and complexity reduction in formal decision contexts[J]. International Journal of Approximate Reasoning, 2014, 55(1):259-274.
- [20] LI J J, ZHANG Y L, WU W Z, et al. Attribute reduction for formal context and consistent decision formal context and concept lattice generation [J]. Chinese Journal of Computer, 2014, 37(8):1768-1774. (in Chinese)
李进金,张燕兰,吴伟志,等.形式背景与协调决策形式背景属性约简与概念格生成[J].计算机学报,2014,37(8):1768-1774.

(上接第 36 页)

- [14] WEI L, QI J J, ZHANG W X. Attribute reduction theory of concept lattice based on decision formal contexts[J]. Science in China F, 2008, 51(7):910-923.
- [15] LI J H, MEI C L, LV Y J. Knowledge reduction in decision formal contexts[J]. Knowledge-Based Systems, 2011, 24(5):709-715.
- [16] LI J H, MEI C L, LV Y J. Knowledge reduction in real decision formal contexts [J]. Information Sciences, 2012, 189(7):191-207.
- [17] LI J H, MEI M L, WANG J H, et al. Rule-preserved object compression in formal decision contexts using concept lattices[J]. Knowledge-Based Systems, 2014, 71:435-445.