

融合可信度和时效标签的商品推荐算法

黄 兰¹, 钱育蓉¹, 于 炯^{1,2}, 郑 洁¹, 马婉贞¹

(1 新疆大学 软件学院, 新疆 乌鲁木齐 830008; 2 新疆大学 信息科学与工程学院, 新疆 乌鲁木齐 830046)

摘 要: 为了提高电子商务推荐系统的性能, 提出了考虑可信度的基于时效性的用户-项目-属性-标签四部图模型, 并针对该模型提出了一种新的推荐算法——融合可信度和时效标签的商品推荐算法. 该模型算法改善新项目的冷启动问题; 提出了基于评分信息的、基于项目属性的和基于用户时效标签的 3 种个性化预测评分方法, 将这 3 种评估方法融合, 通过调节参数 α , β 和 γ 平衡因子, 判断这 3 种评估方法的影响因素权值, 从而提高推荐的准确度. 结果表明该推荐算法既有较高的准确度, 也有较高的新颖度, 并且在某种程度上有效地处理了新项目推荐的冷启动问题.

关键词: 关键字: 推荐系统; 冷启动; 可信度; 时效标签; 四部图

中图分类号: TP181

文献标识码: A

文章编号: 1000-7180(2017)06-0078-06

Goods Recommendation Fusing Credibility and Timeliness-tag

HUANG Lan¹, QIAN Yu-rong¹, YU Jiong^{1,2}, ZHENG Jie¹, MA Wan-zhen¹

(1 School of Software, Xinjiang University, Urumqi 830008, China;

2 School of Information Science and Engineering, Xinjiang University, Urumqi 830046, China)

Abstract: To solve improve the performance of e-commerce recommendation system, considering four graph model of credibility based on the user - projects - property - label of timeliness is proposed, And this paper proposes also a new recommendation algorithm-Goods Recommendation Fusing Credibility and Timeliness-tag(CRFCT) for the model. The model algorithm using reliability model judges user type to improve the cold start of the new project, 3 kinds of personalized prediction scoring methods are proposed based on rating information and the attributes of the project and user timeliness method, incorporating the three evaluation methods, by adjusting the parameters of alpha, beta, and gamma balance factor to determine the weights of influence factors of the three evaluation methods, so as to improve the accuracy of recommendation. Experimenting show that has higher accuracy and also has higher novelty, and to some extent effective to deal with the cold start problem of the new project recommendation compared with the traditional recommendation algorithm.

Key words: recommendation system; cold start; credibility; timeliness-tag ; four graph

1 引言

随着互联网的发展, 人们逐渐进入一个信息过载的时代. 为了解决信息过载问题, 电商平台均使用推荐系统来提高销售量, 因此推荐系统^[1]已成为个性化服务的重要研究热点.

推荐算法是推荐系统的核心, 在当前的研究中有 3 种主流的推荐方法: 基于内容的推荐、协同过滤

推荐、组合推荐方法^[2-5]. 为提高推荐的准确率, 研究者们提出了各种改进和优化方法^[6-8], 解决以上问题. 这些方法在一定程度上解决了协同过滤算法的现存问题, 提高推荐性能, 但是这些方法考虑用户和项目属性时不够全面, 在预测分值时存在较大误差, 影响推荐质量.

基于上述问题, 本文提出基于多属性的商品推荐算法(Goods Recommender Combining Multi-at-

tribute,CRFCT-C).

2 相关基本概念

2.1 用户-项目评分信息

本文使用 $m \times n$ 阶的用户-项目评分矩阵来表示用户对项目的评价信息. 其中, m 个用户集合用集合 $U = \{user_1, user_2, \dots, user_m\}$ 表示, n 个项目集合用集合 $I = \{item_1, item_2, \dots, item_n\}$ 表示, 用户 i 对项目 j 的评分值用 r_{ij} 表示(r_{ij} 值越大, 表示用户 i 对项目 j 越喜欢). 关于 $m \times n$ 阶的用户-项目评分矩阵如表 1 所示.

表 1 用户-项目评分矩阵				
用户	项目			
	item ₁	item ₂	...	item _n
user ₁	$r_{1,1}$	$r_{1,2}$	...	$r_{1,n}$
user ₂	$r_{2,1}$	$r_{2,1}$	...	$r_{2,n}$
...	...	...	...	...
user _m	$r_{m,1}$	$r_{m,2}$	...	$r_{m,n}$

2.2 项目-属性信息

本文使用 $n \times p$ 阶的项目-属性矩阵来表示项目属性信息. 其中, p 个属性集合用集合 $A = \{attribute_1, attribute_2, \dots, attribute_p\}$ 表示, 项目 i 对是否有属性 j 用 a_{ij} 表示($a_{ij} = 0$ 时, 表示项目 i 没有属性 j ; 反之, $a_{ij} = 1$ 时, 表示项目 i 具有属性 j). 关于 $n \times p$ 阶的项目-属性矩阵如表 2 所示.

表 2 项目-属性矩阵				
项目	属性			
	attribute ₁	attribute ₂	...	i attribute _p
item ₁	$a_{1,1}$	a_{12}	...	$a_{1,p}$
item ₂	$a_{2,1}$	$a_{2,1}$	...	$a_{2,p}$
...	...	...	...	...
item _n	$a_{n,1}$	$a_{n,2}$	...	$a_{n,p}$

定义 1 (基于属性的项目相似性) 基于属性的项目相似性计算公式 $sim(i, j)$ 如下:

$$sim(i, j) = \frac{AC_i \cap AC_j}{AC_i \cup AC_j} \tag{1}$$

式(1)中, AC_i 表示项目 i 所具有的属性, AC_j 表示项目 j 所具有的属性, $AC_i \cap AC_j$ 表示项目 i 和项目 j 的共同属性个数, $AC_i \cup AC_j$ 表示项目 i 和项目 j 的所有属性个数. 其中, $AC_i \cap AC_j$ 值越大, $sim(i, j)$ 越大, 说明两个项目的共同属性越多, 因此两个项目越相似.

2.3 属性-标签信息

本文使用 $p \times q$ 阶的标签-属性评分矩阵来表示

标签属性信息. 其中, p 个属性集合用集合 $A = \{attribute_1, attribute_2, \dots, attribute_p\}$ 表示, q 个标签集合用集合 $L = \{label_1, label_2, \dots, label_q\}$ 表示, 标签 i 是否有属性 j 用 l_{ij} 表示($l_{ij} = 0$ 时, 表示标签 i 没有属性 j ; 反之, $l_{ij} = 1$ 时, 表示标签 i 具有属性 j). 用户和项目都可具有多个标签. 关于 $p \times q$ 阶的属性-标签矩阵如表 3 所示.

表 3 属性-标签矩阵				
属性	标签			
	label ₁	label ₂	...	label _q
attribute ₁	$l_{1,1}$	$l_{1,2}$	...	$l_{1,q}$
attribute ₂	$l_{2,1}$	$l_{2,1}$	...	$l_{2,q}$
...	...	...	...	...
attribute _p	$l_{p,1}$	$l_{p,2}$	...	$l_{p,q}$

2.4 用户-时效标签信息

本文使用 $m \times s$ 阶的用户-时效标签矩阵来表示用户-时效标签信息. 其中, m 个用户集合用集合 $U = \{user_1, user_2, \dots, user_m\}$ 表示, s 个时间值集合用集合 $T = \{time_1, time_2, \dots, time_s\}$ 表示(时间值是所有用户购买商品的时间, 按从小到大的顺序排列), 用户 i 在时间 j 标注的标签用 t_{ij} 表示($t_{ij} \in L = \{label_1, label_2, \dots, label_q\}$). 关于 $m \times s$ 阶的用户-时效标签矩阵如表 4 所示.

表 4 用户-时效标签矩阵				
用户	时间			
	time ₁	time ₂	...	time _s
user ₁	$t_{1,1}$	$t_{1,2}$	...	$t_{1,s}$
user ₂	$t_{2,1}$	$t_{2,1}$	...	$t_{2,s}$
...	...	...	...	...
user _m	$t_{m,1}$	$t_{m,2}$	...	$t_{m,s}$

定义 2 (基于标签的用户相似性) 基于标签的用户相似性计算公式 $sim(u, v)$ 如下:

$$sim(u, v) = \frac{ALC_u \cap ALC_v}{ALC_u \cup ALC_v} \tag{2}$$

式(2)中, ALC_u 表示用户 u 所有标签的属性个数, ALC_v 表示用户 v 所有标签的属性个数, $ALC_u \cap ALC_v$ 表示用户 u 和用户 v 的共同属性个数, $ALC_u \cup ALC_v$ 表示项用户 u 和用户 v 的所有属性个数. 其中, $ALC_u \cap ALC_v$ 值越大, $sim(u, v)$ 越大, 说明两个用户标签的共同属性越多, 因此两个用户越相似.

2.5 评价-购买时间信息

本文使用 $n \times m$ 阶的评价-购买时间矩阵来表示评价-购买时间信息. 其中, m 个用户集合用集合

$U = \{user_1, user_2, \dots, user_m\}$ 表示, n 个项目集合用集合 $I = \{item_1, item_2, \dots, item_n\}$ 表示, 用户 i 对项目 j 的评价时间用 Pd_{ij} 表示, 用户 i 对项目 j 的购买时间用 Bd_{ij} 表示. 关于 $m \times n$ 阶的评价-购买时间矩阵如表 5 所示.

表 5 评价-购买时间矩阵

Item	user ₁		user ₂		...		user _m	
	rt	bt	rt	bt	rt	bt	rt	bt
item ₁	rt_{11}	bt_{11}	rt_{21}	bt_{21}	...	...	rt_{m1}	bt_{m1}
item ₂	rt_{12}	bt_{12}	rt_{22}	bt_{22}	...	...	rt_{m2}	bt_{m2}
...	...	...	...	...	...	...	...	...
item _n	rt_{1n}	bt_{1n}	rt_{2n}	bt_{2n}	...	...	rt_{mn}	bt_{mn}

3 融合可信度和时效标签的商品推荐算法

3.1 用户评分可信度

在推荐系统中, 由于用户对项目评分的主观性及用户兴趣的变化性影响了评分的真实性, 从而降低了推荐的准确率. 本文从评分准确度和评分时间权重两方面来衡量用户评分可信度. 评分可信度高的称为可信用用户; 评分可信度低的称为不可信用用户. 对于可信用用户而言, 该用户对项目 i 的评分与该项目平均评分值差较小, 且该用户的评论时间和购买时间一般也会较长一些; 对于不可信用用户, 该用户对项目 i 的评分与该项目平均评分值差较大, 且该用户的评论时间和购买时间一般也会较短一些. 在商品推荐过程中, 这两种用户都存在, 本文将新项目优先推荐给可信用用户.

定义 3 (评分准确度) 评分准确度计算公式 $Accu(u)$ 如下:

$$Accu(u) = \frac{1}{N} \sum_{i=1}^N \left[1 - \frac{r_{u,i} - \bar{r}_i}{\bar{r}_i} \right] \quad (3)$$

式中, $Accu(u)$ 表示用户 u 的平均评分准确度, $r_{u,i}$ 表示用户 u 对项目 i 的评分, $\bar{r}_i$ 表示项目 i 的平均评分值, N 是用户 u 评价过所有项目的集合. 其中, $(r_{u,i} - \bar{r}_i)$ 值越小, $Accu(u)$ 越大, 说明用户 u 对项目 i 的评分误差越小, 因此用户的评分准确度越高, 即用户评分越可信.

定义 4 (评分时间权重) 评分时间权重计算公式 $Time(u)$ 如下:

$$Time(u) = \frac{1}{N} \sum_{i=1}^N \left[1 - \frac{1}{rt_{u,i} - bt_{u,i}} \right] \quad (4)$$

式中, $Time(u)$ 表示用户 u 评论所有项目的平均评分时间权重, $rt_{u,i}$ 表示用户 u 对用户 i 的评论时间, $bt_{u,i}$ 表示用户 u 对用户 i 的购买时间, N 是用户 u 评

价过所有项目的集合. 其中, $(rt_{u,i} - bt_{u,i})$ 值越大, $Time(u)$ 越大, 说明用户评分时间和购买时间差值越大, 用户评分可信度越高.

定义 5 (用户评分可信度) 用户评分可信度计算公式 $Cret(u)$ 如下:

$$Cret(u) = \varphi Accu(u) + (1 - \varphi) Time(u) \quad (5)$$

式中, $Cret(u)$ 表示用户 u 的评分可信度. 用户评分可信度由评分准确度和评分时间权重共同确定. φ 是用户评分可信度的平衡因子 ($\varphi \in [0, 1]$), φ 越大, 说明评分准确度对用户可信度影响越大; 反之, φ 越小, 说明评分时间权重对用户可信度影响越大.

3.2 用户个性化预测评分方法

如图 1 所示, 该图为用户-项目-属性-标签四部图, R 表示用户对项目的评分, A 表示项目具有的属性, T 表示标签具有的属性.

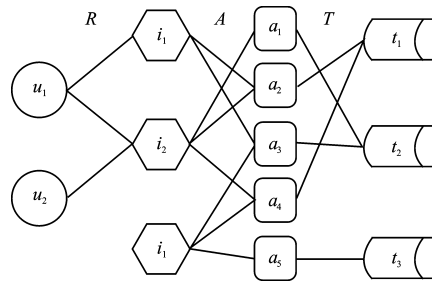


图 1 用户-项目-属性-标签四部图

定义 6 (基于评分信息的预测评分值) 基于评分信息的预测评分值计算公式 $f_u(u, i)$ 如下:

$$f_u(u, i) = \sum_{v=1}^{UC_i} \frac{sim(u, v) \times r_{vi}}{UC_i} \quad (6)$$

式中, $f_u(u, i)$ 表示基于评分信息的用户 u 对项目 i 的预测评分值, $sim(u, v)$ 表示用户 u 和用户 v 的相似度, r_{vi} 表示用户 v 对项目 i 的评分, UC_i 表示共同评价项目 j 的用户数.

定义 7 (基于项目属性的预测评分值) 基于项目属性的预测评分值计算公式 $f_a(u, i)$ 如下:

$$f_a(u, i) = \sum_{j=1}^{IC_U} \frac{sim(i, j)}{IC_U} \quad (7)$$

式中, $f_a(u, i)$ 表示基于项目属性的用户 u 对项目 i 的预测评分值, $sim(i, j)$ 表示项目 i 和项目 j 的相似度, IC_U 表示用户 U 评价过的项目数.

定义 8 (基于用户时效标签的预测评分值) 基于用户时效标签的预测评分值计算公式 $f_t(u, i)$ 如下:

$$f_t(u, i) = \sum_{j=1}^J \frac{1}{t_x - t_j} \times \frac{ALC_j}{ALC_U} \quad (8)$$

式中, $f_t(u, i)$ 表示基于时效标签的用户 u 对项目 i 的预测评分值, J 表示项目 i 的标签数集合, ALC_j 表示项目 i 标签 j 的属性个数, ALC_U 表示用户 U 总标签属性个数. $t_x - t_j$ 表示用户标签 j 的时间与现在时间的间隔, 时间间隔越小, 用户标签对用户兴趣影响越大.

3.3 融合可信度和时效标签的商品推荐算法

根据前文描述, 本节结合用户、项目属性、标签、时间等提出一种基于多属性的商品推荐算法. 实验结果表明, 本文提出的算法不仅提高了推荐的准确度, 还有效解决了新项目的冷启动问题. 基于多属性的商品推荐算法的预测评分值计算公式如下:

$$P(u, i) = \begin{cases} \alpha \times f_u(u, i) + \beta \times f_a(u, i) + \\ \gamma \times f_t(u, i), \\ \alpha + \beta + \gamma = 1, \alpha \times \beta \times \gamma > 0 \\ Cret(u) \times \frac{AC_i}{AIC_u} \end{cases} \quad (9)$$

(10)

在上述公式中, $P(u, i)$ 表示用户 u 对项目 i 的预测评分, 其中式(9)表示项目 i 不是新项目, 式(10)中的项目 i 是新项目. 式(9)是不考虑可信度模型的商品推荐方法(简称 GRFCT), 式(9)、(10)是考虑可信度模型的商品推荐方法(简称 GRFCT-C).

3.4 融合可信度和时效标签的商品推荐算法描述

输入: 目标用户 u , 目标用户 i , 用户-项目评分矩阵 $\mathbf{R}_{m \times n}$, 项目-属性矩阵 $\mathbf{A}_{n \times p}$, 属性-标签矩阵 $\mathbf{L}_{p \times q}$, 用户时效标签矩阵 $\mathbf{T}_{m \times s}$, 用户时间评价矩阵 TIME.

输出: 目标用户 u 对项目 i 的预测评分值 Pre_{ui} .

步骤1 根据用户-项目评分矩阵 $\mathbf{R}_{m \times n}$, 使用式(6)计算基于项目评分信息的预测评分值 $f_u(u, i)$.

步骤2 根据用户-项目评分矩阵 $\mathbf{R}_{m \times n}$ 和项目-属性矩阵 $\mathbf{A}_{n \times p}$, 使用式(7)计算基于项目属性的预测评分值 $f_a(u, i)$.

步骤3 根据属性-标签矩阵 $\mathbf{L}_{p \times q}$ 和用户时效标签矩阵 $\mathbf{T}_{m \times s}$, 使用式(8)计算基于用户时效标签的预测评分值.

步骤4 根据用户-项目评分矩阵 $\mathbf{R}_{m \times n}$, 使用式(3)计算用户评分准确度; 再根据用户时间评价矩阵 TIME 使用式(4)计算用户评分时间权重; 最后, 使用式(5)计算用户评分可信度权值.

步骤5 结合前四个步骤计算出的结果使用式(9)、(10), 计算目标用户 u 对项目 j 的预测评分值

Pre_{ui} .

4 实验结果与分析

本文设计3组实验验证所提出方法的有效性. 首先, 针对2.1节提出的用户可信度对推荐准确度和新颖度的影响进行了实验; 然后, 针对2.3节提出的商品各属性对推荐准确度的影响进行对比了实验; 最后, 将本文所提出的方法与其他方法进行对比实验.

实验1 用户可信度对推荐准确度和新颖度的影响.

首先, 针对本文所提出的用户可信度算法进行平衡因子 Φ 的调整; 然后, 计算出用户可信度对推荐准确度和新颖度的影响.

由图2可知, 当 Φ 在取值 0.4~0.6 时, 用户的可信度最高, 这说明评分准确度和评分权重对用户可信度的影响一样重要. 用户兴趣建模一般根据用户用户浏览与购买评价信息进行建模, 当用户评价越客观时, 对用户的兴趣建模越准确; 用户兴趣还与评论的时间有关系, 当用户评论时间与购买时间较长时, 说明用户经过自身体验才进行了评价, 这样的评价相对客观, 对用户的兴趣建模越准确.

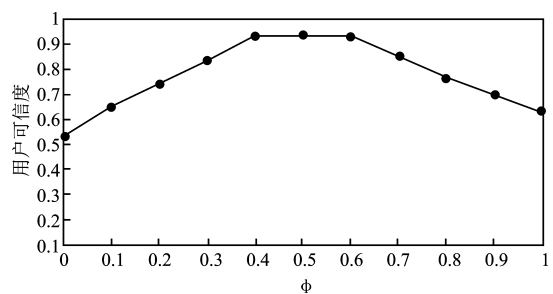


图2 参数 Φ 对用户可信度的影响

由图3可知, 随着用户可信度增大, 其准确度和新颖度也就越高. 用户可信度越高, 说明该用户购买商品行为较多, 且用户评分相对客观, 因此可信度越大, 其推荐的准确度和新颖度越高. 当可信度大于0.7时, 准确度和新颖度变化就相对平稳, 这是因为当用户可信度较高时, 他们的购买商品行为都较多, 以及他们评分也都相对客观, 所以对这些用户的兴趣建模误差不会太大, 变化会相对平稳.

结论1: 用户可信度与推荐准确度和新颖度呈正相关关系.

实验2: 不同算法的准确度对比.

将本文所提出的考虑可信度模型商品推荐方法 CRFCT-C、不考虑可信度模型的商品推荐方法

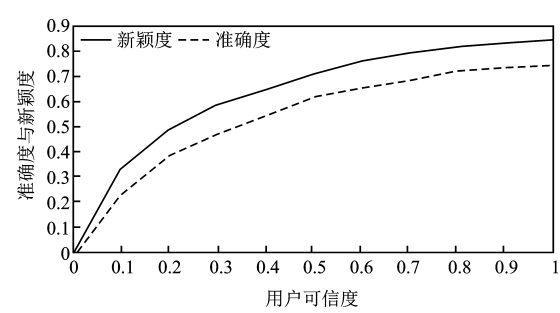


图3 用户可信度对准确度和新颖度的影响

CRFCT 和文献[9]的推荐算法 CUT 进行准确度的对比。在这些算法中,由于 α 、 β 和 γ 不同的取值共有 36 种组合,本文只给出了最优组取值参数。本文针对这 36 种组合进行了对比实验,实验结果表明:当 γ 取值越大时,对推荐的准确度影响越大。由于用户时效标签记录了用户在某个时段的兴趣爱好,它很好地展现了用户兴趣变化的情况,因此 γ 取值变化对推荐结果影响很大。

由图 4 可知,这 3 种算法都是随着推荐项目个数越多,推荐准确度就随之下降,同时,CRFCT-C 推荐准确度上略优于 CRFCT 算法和 CUT 算法。其中 CRFCT-C 算法平均推荐准确度为 0.63,CRFCT 算法平均推荐准确度为 0.61,CUT 算法平均推荐准确度为 0.58。由于 CRFCT-C 算法从评分准确度和评论时间权重两方面考虑了用户的可信度,然而 CRFCT 算法和 CUT 算法未考虑,因此 CRFCT-C 算法预测用户的兴趣越接近用户真实兴趣,推荐比其他两种方法的准确度高。

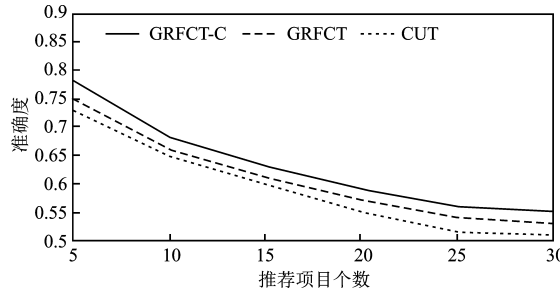


图4 不同算法的准确度对比

结论 2:通过 CRFCT-C、CRFCT 和 CUT 3 种算法的实验对比可知,CRFCT-C 算法比 CRFCT 算法多考虑了用户可信度,其推荐的准确度高了 2%,CRFCT-C 算法比 CUT 算法多考虑了项目属性,其推荐的准确度高了 3%,通过实验结果可知考虑用户和项目的属性越多,其推荐准确率越好。

实验 3:不同算法的新颖度对比。

将本文所提出的考虑可信度模型商品推荐方法

CRFCT-C、不考虑可信度模型商品推荐方法 CRFCT 和文献[9]的推荐算法 CUT 进行新颖度的对比。

图 5 可知,这 3 种算法都是随着推荐项目个数越多,推荐新颖准确度就随之下降,同时,CRFCT-C 推荐新颖准确度上很明显高于 CRFCT 算法和 CUT 算法。其中 CRFCT-C 算法平均推荐新颖度为 0.72,CRFCT 算法平均推荐新颖度为 0.18,CUT 算法虽然提高了推荐的准确度,但并未解决新项目推荐问题,因此 CUT 算法推荐新颖度为 0。通过 3 种算法的比较,说明本文提出的用户可信度模型可以很好地解决新项目冷启动问题。

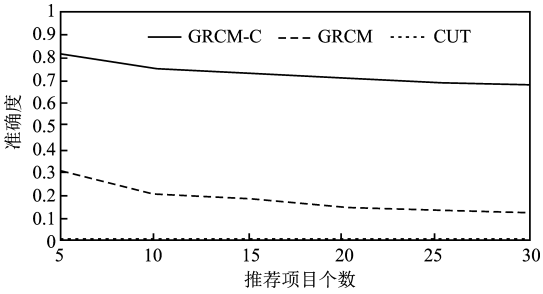


图5 不同算法的新颖度对比

结论 3:通过 CRFCT-C、CRFCT 和 CUT 3 种算法的实验对比,CRFCT-C 算法比 CRFCT 算法提高了约 54% 的新颖度,CRFCT 算法比 CUT 算法提高了约 18% 的新颖度,通过实验结果可知本文提出的用户可信度模型对新颖度影响较大,CRFCT-C 算法有效地解决了新项目推荐的冷启动问题。

5 结束语

本研究结果发现,本文提出的 CRFCT-C 算法不仅提高了推荐的准确度,还有效解决了新项目的冷启动问题。首先,本文从用户评分、项目属性、用户标签等多方面进行分析并定义了基于多属性的商品推荐算法计算公式,然后进行实验和分析,最后与其他类似方法进行实验对比,实验结果表明,本文所提出的 CRFCT-C 方法是有效的。由于本研究数据集是购物网站的数据集,因此比较适合电子商务购物网站。

参考文献:

[1] Isinkaye F O, Folajimi Y O, Ojokoh B A, et al. , Recommendation systems : Principles, methods and evaluation[J]. Egyptian Informatics Journal, 2015, 16(3): 261-273.

- [2] Silvia Puglisi, Javier Parra-Arnau, Jordi Forné. On content-based recommendation and user privacy in social-tagging systems[J]. Computer Standards & Interfaces, 2015, 41: 17-27.
- [3] Dongsheng Li, Qin Lv, Li Shang, et al. Item-based top-N recommendation resilient to aggregated information revelation[J]. Knowledge-Based Systems, 2014, 67: 290-304.
- [4] Alper Bilge, Lhsan Gunes, Huseyin Polat. Robustness analysis of privacy-preserving model-based recommendation schemes[J]. Expert Systems with Applications, 2014, 41(8): 3671-3681.
- [5] Jing Zhang, Peng Qinke, Sun Shiquan, et al. Collaborative filtering recommendation algorithm based on user preference derived from item domain features[J]. Physica A: Statistical Mechanics and its Applications, 2014, 396: 66-76.
- [6] 肖晓丽. 基于用户兴趣和社交信任的聚类推荐算法[J]. 计算机应用, 2016, 36(5): 1273-1278.
- [7] 韦素云. 结合项目类别和动态时间加权的协同过滤算法[J]. 计算机工程, 2014, 40(6): 206-210.
- [8] 丁少衡. 基于用户属性和评分的协同过滤推荐算法[J]. 计算机工程与设计, 2015, 36(2): 487-491.
- [9] Zhang Z K, Zhou T, Zhang Y C. Tag-Aware recommender systems: A state-of-the-art survey[J]. Journal of Computer Science and Technology, 2011, 26(5): 767-777. [doi:10.1007/s11390-011-0176-1]

作者简介:

黄 兰 女, (1988-), 硕士研究生. 研究方向为数据挖掘.

钱育蓉(通讯作者) 女(满族), (1980-), 博士, 副教授. 研究方向为网络计算和遥感图像处理. E-mail: qyr@xju.edu.cn.

于 炯 男, (1964-), 教授, 博士生导师. 研究方向为网络安全、网络与分布式计算等.

郑 洁 女, (1992-), 硕士研究生. 研究方向为数据挖掘.

马婉贞 女(回族), (1992-), 硕士研究生. 研究方向为高性能并行计算.

(上接第 77 页)

- [5] Fard A, Wang K, Yu P. Limiting link disclosure in social network analysis through subgraph-wise perturbation[C]//Proc of the 2012 EDBT International Conference on Extending DB Technology. New York: ACM, 2012: 109-119.
- [6] Ford A, Wang K. Neighborhood randomization for link privacy in social network analysis[J]. World Wide Web-internet & Web Information Systems, 2013, 18(1): 9-32.
- [7] Malewicz G, Austern M, Bik A, et al. Pregel: a system for large-scale graph processing[C]//Proc of the 2010 SIGMOD International Conference on Management of data. New York: ACM, 2010: 135-146.
- [8] Salihoglu S, Widom J. Optimizing graph algorithms on

pregel-like systems[J]. Proceedings of the Vldb Endowment, 2014, 7(7): 577-588.

作者简介:

张文超 女, (1989-), 硕士研究生. 研究方向为大规模数据下社会网络隐私保护. E-mail: 1028721352@qq.com.

张晓琳 女, (1966-), 博士, 教授. 研究方向为数据库理论与技术、大规模数据下社会网络隐私保护.

张 臣 男, (1990-), 硕士研究生. 研究方向为大规模数据下社会网络隐私保护.

刘立新 女, (1983-), 硕士, 讲师. 研究方向为数据库理论与技术.

何晓玉 男, (1991-), 硕士研究生. 研究方向为大规模数据下社会网络隐私保护.