doi:10.11959/j.issn.1000-436x.2015322

数据隐私保护的社会化推荐协议

刘曙曙[1,2]，刘安[1,2]，赵雷[1,2]，刘冠峰[1,2]，李直旭[1,2]，郑凯[1,2]，周晓方[1,2]
(1. 苏州大学 计算机科学与技术学院，江苏 苏州 215000；2. 江苏省软件新技术与产业化协同创新中心，江苏 南京 211102)

摘 要：基于邻域的社会化推荐需要同时依赖用户的历史行为数据和完善的社交网络拓扑图，但通常这些数据分别属于不同平台，如推荐系统服务提供商和社交网络服务提供商。出于维护自身数据价值及保护用户隐私的考虑，他们并不愿意将数据信息提供给其他方。针对这一现象，提出了 2 种数据隐私保护的社会化推荐协议，可以在保护推荐系统服务提供商和社交网络服务提供商的数据隐私的同时，为用户提供精准的推荐服务。其中，基于不经意传输的社会化推荐，计算代价较小，适用于对推荐效率要求较高的应用；基于同态加密的社会化推荐，安全程度更高，适用于对数据隐私要求较高的应用。在 4 组真实数据集上的实验表明，提出的 2 种方案切实可行，用户可以根据自身需求选择合适的方案。

关键词：推荐系统；不经意传输；同态加密；Yao's 协议
中图分类号：TP302 **文献标识码**：A

Preserving data privacy in social recommendation

LIU Shu-shu[1,2], LIU An[1,2], ZHAO Lei[1,2], LIU Guan-feng[1,2], LI Zhi-xu[1,2], ZHENG Kai[1,2], ZHOU Xiao-fang[1,2]
(1. School of Computer Science and Technology, Soochow University, Suzhou 215000, China;
2. Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing 211102, China)

Abstract: Social recommendation is a method which requires the participants of both user's historical behavior data and social network, which generally belong to different parties, such as recommendation system service provider and social network service provider. Considering the fact that in order to maintain the value of their own data interests and user's privacy, none of them will provide data information to the other, two privacy preserving protocols are proposed for efficient computation of social recommendation which needs the cooperation of two parties (recommendation system service provider and social network service provider). Both protocols enable two parties to compute the social recommendation without revealing their private data to each other. The protocol based on the well-known oblivious transfer multiplication has a low cost, and is suitable for the application of high efficiency requirements. And the one based on homomorphic cryptosystem has a better privacy preserving, and is more suitable for the application of higher data privacy requirements. Experimental results on the four real datasets show those two protocols are efficient and practical. Users are suggested to choose the appropriate protocol according to their own need.

Key words: recommendation system; oblivious transfer; homomorphic encryption; Yao's protocol

1 引言

推荐系统是一系列通过对用户或其购买行为进行分析，从而自动为用户推荐其可能感兴趣的信息和商品的算法集合。协同过滤算法出现以来就受到了广泛关注，但随之而来的“冷启动”问题却制约了该算法的进一步使用。为解决这一问题，研究学者们提出了基于邻域的社会化推荐方法[1,2]，该方法指出，将基于社交网络拓扑图得到的用户关系引入到推荐系统中，可以有效解决协同过滤计算过程中由于新用户缺少历史行为数据而带来的“冷启动”问题。基于邻域的社会化推荐算法得到普遍关注，同时，大量最新的研究也表明，该算法明显优于传统推荐算法。

收稿日期：2015-10-21；**修回日期**：2015-12-12
基金项目：国家自然科学基金资助项目（61572336, 61572335, 61303019, 61402313）
Foundation Item: The National Natural Science Foundation of China(61572336, 61572335, 61303019, 61402313)

尽管推荐系统能够帮助人们在海量数据中高效快速过滤掉大量无关信息，但是，这一过程带来的信息泄露却值得人们担忧。在传统的推荐系统中，用户的历史行为数据，如购物记录、对于电影或物品的评分记录等，都必须作为数据上传到推荐系统服务器。一旦这些信息发生泄漏，攻击者便可以基于以上信息，迅速推测出用户的年龄、性别、身体状况等个人隐私信息，由此而引发的一些问题可能会对用户的生命财产造成威胁。

为了防止传统推荐系统中由于用户历史行为数据泄露而带来的担忧，研究学者们提出了一系列解决方案。Canny[5]最初提出了针对推荐系统的隐私保护方案，通过借助同态加密和一个端对端的协议，他能够为多种基于协同过滤的推荐系统提供隐私服务。该方案同样在文献[4, 6]中得到了应用。借助随机扰动技术，Polat 和 Du[7]将用户的历史行为数据进行一定程度的干扰后再将其发送给推荐服务提供商，从而保证了用户的数据隐私安全。文献[8]使用了相同的原理来实现这一目标，他们都能够保证用户的历史行为数据在推荐服务提供商端的安全。Jorgensen[3]等提出利用差分隐私技术可以保证目标用户无法从推荐结果中推测任何和其他用户相关的信息，从而保证了其他用户个人隐私的安全性。不同于上述传统推荐系统中的问题模型，在社会化推荐系统中，以推荐系统服务提供商及社交网络服务提供商为主要参与方。在现实生活中，推荐系统服务提供商通常对应在线电子商务平台，他们持有用户的历史行为数据，却没有完善的社交网络拓扑图；社交网络拓扑图通常来自于第三方的在线社交网络服务提供商，如 FaceBook 或者 Twitter 等，他们都拥有用户信赖及用户数据信息。但是，出于对社交网络拓扑图拥有重要的利益价值和维护用户隐私的考虑，在线社交网络服务提供商并不愿意将数据信息提供给推荐系统服务提供商。同样，推荐系统服务提供商愿意为用户提供高效的推荐服务，但是并不愿意透漏用户的历史行为数据。

出于对以上实际情况的考虑，本文认为在社会化推荐系统中，保证两主要参与方的数据隐私安全是完成社会化推荐的前提。上述方案虽然能够有效解决传统推荐系统中用户隐私保护问题，但是这些方案并不适用于本文的问题模型。如何能够在保护双方数据隐私的前提下，实现两参与方协同计算的问题是本文的重点，将在后面详细讲解。

2 相关背景及问题定义

在基于邻域的社会化推荐系统中，通常有 2 个参与方，分别称为 Alice 和 Bob。Alice 代表持有用户历史行为数据的推荐系统服务提供商（如 ebay、淘宝等电子商务平台），Bob 是拥有社交网络拓扑图的第三方社交网站（如 Facebook、Twitter 等）。本文分别对以上 2 个数据模型给出形式化的定义。

定义 1 用户历史行为二分图。如图 1（a）所示，用户历史行为数据图 $G_t=(U, I, E_t)$是一个单向二分图，其中，$U(|U|=M)$是所有用户的集合，$I(|I|=N)$是物品集合，E_t 是由用户 U 指向物品 I 的单向边集合，每一条边都附有权重值 $w(u, i)\geqslant 0$，$w(u,i)>0$ 时，表示用户 u 对于物品 i 的相应评分或购买频次，当用户 u 未曾购买过物品 i, $w(u, i)=0$。

定义 2 用户社交网络拓扑图。如图 1（b）所示，社交网络拓扑图 $G_s=(U, E_s)$是一个由用户集合 U 和用户关系边 E_s 构成的无向图，其中，E_s (u, v) 表示用户 u 和 v 之间存在联系。

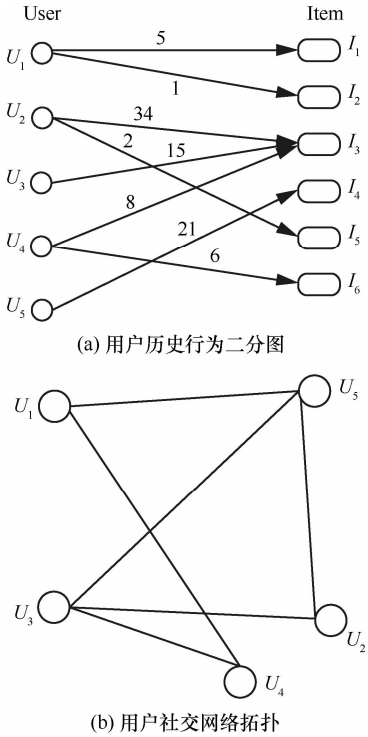


图 1 用户历史行为二分图和用户社交网络拓扑

上述假设模型可以方便地扩展到其他应用领域，如在 Last.fm 中，边 $E_t(u,i)$表示用户 u 听过作者 i 的歌曲，在 Brightkite.com 中，边 $E_t(u,i)$表示用户

u 曾经访问过位置 i，在上述假设中，权重值 $w(u, i)$分别代表了用户 u 听过作者 i 的歌曲的次数，以及用户 u 曾经访问过位置 i 的次数。同样，各种类型的在线社交网站均可映射到定义 2 的模型中，如关系边（u, v）既可表示 Facebook 中用户 u 和 v 之间的好友关系，也可以代表 Twitter 中用户之间的“Following”关系。

定义 3　基于邻域的社会化推荐。假设两参与方 Alice 和 Bob，Alice 拥有用户历史行为数据，Bob 拥有社交网络拓扑图（以及用户之间相似度度量算法 sim），Alice 和 Bob 通过合作计算，为目标用户 $u\in U$，推荐 K 个评分最高的物品 $R_u\in I$。

对于 Alice 中的任一目标用户 u，Alice 将会为 u 推荐 K 个得分最高的物品，记为集合 R_u。其中，$s(u,i)$，即对于用户 u 而言物品 i 的得分

$$s(u,i)=\sum sim(u,v)w(v,i)\ (v\in U, v\neq u) \qquad (1)$$

其中，$sim(u,v)$是用户 u 和 v 之间的相似度值，在非社会化推荐系统中，如图 2(a)所示，$sim(u,v)$的计算主要基于用户的历史交易记录向量，常用方法有皮尔逊相关系数法、余弦相似度以及基于概率的相似度计算方法等，其中前 2 种方法使用最为广泛。在社会化推荐系统中，如图 2(b)所示，相似度的计算通常基于社交网络拓扑图（即图中用户之间的邻接矩阵），常用方法有共同邻居（common neighbor），Katz 以及随机游走（random walk with restart）等。$w(v,i)$是用户 v 与物品 i 的连接边的权重值。

	I_1	I_2	I_3	I_4	I_5	I_6
U_1	5	1	?	?	?	?
U_2	?	?	34	?	2	?
U_3	?	?	15	?	?	?
U_4	?	?	8	?	?	6
U_5	?	?	?	21	?	?

(a) 传统推荐系统

	I_1	I_2	I_3	I_4	I_5	I_6
U_1	5	1	?	?	?	?
U_2	?	?	34	?	2	?
U_3	?	?	15	?	?	?
U_4	?	?	8	?	?	6
U_5	?	?	?	21	?	?

	U_1	U_2	U_3	U_4	U_5
U_1	0	0	0	1	1
U_2	0	0	1	0	1
U_3	0	1	0	1	1
U_4	1	0	1	0	0
U_5	1	1	1	0	0

(b) 社会化推荐系统

图 2　传统和社会化推荐系统

在社会化推荐中，$sim(u,v)$的计算依赖于社交网络拓扑图 G_s。为了能够完成社会化推荐，Alice 即电子商务平台必须利用社交网络拓扑。出于商业利益或维护用户隐私的权益，Bob 即社交网络拓扑图的持有者，并不愿意将私有数据，社交网络拓扑图直接提供给 Alice 使用，保证两参与方的数据隐私安全是完成社会化推荐的前提。下面将给出保护隐私的社会化推荐的定义。

定义 4　保护隐私的社会化推荐。假设两参与方 Alice 和 Bob，Alice 拥有用户历史行为数据，Bob 拥有社交网络拓扑图以及用户之间相似度度量算法 sim，Alice 和 Bob 通过合作计算，为目标用户 $u\in U$，推荐 K 个评分最高的物品 $R_u\in I$，计算过程中，双方私有信息（如 G_t 和 G_s）不能暴露给对方。

假设前提：定义中的推荐结果都是基于某一时刻的静态图谱 G_t 和 G_s 展开的。对于图谱 G_t 和 G_s 的动态更新，需要重新运行协议，更新推荐结果。下面将给出该问题的具体解决方案。

3　保护隐私的社会化推荐系统

针对社会化推荐系统需要在两方隐私数据上进行协同计算，本文提出了 2 种数据隐私保护的社会化推荐协议，可以同时保护推荐系统服务提供商和社交网络服务提供商的数据隐私。其中，基于不经意传输的社会化推荐，计算代价较小，适用于对推荐效率要求较高的应用。基于同态加密的社会化推荐，安全程度较高，适用于对数据隐私要求较高的应用。下面是 2 个协议的详细介绍。

3.1　基于不经意传输的社会化推荐

不经意传输(OT，oblivious transfer)是安全计算领域的重要工具，在众多问题中得到了广泛应用。借助不经意传输协议，可以高效地实现两方安全乘法计算[13]。计算过程中，两参与方私有数据信息安全完成后，结果由两方以和形式秘密共享，即参与方 A 持有数据 a，参与方 B 持有数据 b，利用不经意传输乘法协议后，A 将持有结果 x，B 持有结果 y，并且满足公式 $x+y=ab$。相关细节参见文献[13]。

根据定义 4，Alice 持有数据 G_t，Bob 持有数据 G_s。对于目标用户 u，Bob 可以根据事先确定好的相似度计算方法计算出 $sim(u,v)$ (v 是除 u 外

的所有其他用户)。以物品 i 为例，Bob 端持有相似度向量 $SIM=\{sim(u,u_1),sim(u,u_2),\cdots,sim(u,u_m)\}$，Alice 持有物品 i 的评分向量 $W_i=\{w(u_1,i),w(u_2,i),\cdots,w(u_m,i)\}$，根据式（1）可知，对于目标用户 u 而言，物品 i 的推荐得分 $s(u,i)$为对应位积之和。具体算法如下。

算法 1 基于 OT 乘法的推荐得分算法

输入：Alice 端 $G_t=(U,I,E_t)$

Bob 端 $G_s=(U,E_s)$， 相似度计算函数 $Fsim$

目标用户 $u\in U$

输出：所有物品 I 的推荐得分

Bob

1) $sim(u,v)=Fsim(G_s)$ ($v\in U$， $v\neq u$)

Alice + Bob

2) for $i\rightarrow N$

3) for $j\rightarrow M$

4) $s(u,i)=s(u,i)+sim(u,u_j)w(u_j,i)$

5) end for

6) end for

在计算过程中，使用 OT 乘法直接算出所有用户和物品之间的乘积，其复杂度为 MN。但是由于用户历史行为记录是一个及其稀疏的矩阵，通常为 $M+N$，直接对所有元素进行 OT 乘法操作，会产生大量不必要的计算。通用的解决方案是，Alice 将用户历史行为记录矩阵的数据分布情况（0 为用户未曾够买过物品；1 为用户购买过物品）共享给 Bob，同样 Bob 也需要将自己数据分布情况共享给 Alice，两端仅需计算 $sim(u,u_j)\neq 0$ 和 $w(u_j,i)\neq 0$ 的项，从而减少不必要的计算开销。在这一共享过程中，两方共享的仅为数据分布情况，并未涉及两方的数据值信息，在对安全要求不是很高的情况下，该方法是安全可信的。

利用 OT 乘法协议完成物品的推荐得分计算后，所有物品的推荐得分由 Alice 和 Bob 以加法和形式秘密共享，即 Alice 端持有 s_1，Bob 持有 s_2，同时 $s_1+s_2=s$。

接下来，需要从所有候选物品中，挑选出 K 个推荐得分最高的物品推荐给用户。因为最终只需要将 K 个推荐得分最高的物品推荐给目标用户即可，K 个物品之间的排列顺序并不影响推荐，所以采用线性时间复杂度的随机选择算法[14]来实现 Top$K$ 选择。

随机选择算法的基本思想是：随机选择枢纽元，将数据分为 2 个独立的部分，其中一部分的所有数据都比枢纽元小，另外一部分的数据都比枢纽元大，然后再按此方法继续对其中某一部分数据进行划分，直到找到的枢纽元在整个序列中处于 K 的位置。具体算法如下。

算法 2 安全的 Top$K$ 选择算法

输入：推荐得分向量 $S=\{s_1,s_2,\cdots,s_n\}$

输出：K 个最高推荐得分物品集合 R_u

Alice + Bob

1) $l\leftarrow 1,h\leftarrow N,A[i]\leftarrow i(1\leqslant i\leqslant K)$

2) loop

3) $k\leftarrow$RANDOM(l,h)

4) $k\leftarrow$PARTITION(l,h,k)

5) case

6) $k=K$: return ($i_{A[1]},i_{A[2]},\cdots,i_{A[K]}$)

7) $k<K$: $l\leftarrow k+1$

8) $k>K$:$h\leftarrow k-1$

9) endcase

10) end loop

Procedure PARTITION(l,h,k)

11) $p\leftarrow l,q\leftarrow h,m\leftarrow k$

12) loop

13) while COMPARE(s_p,s_m) = 1 do

14) $p\leftarrow p+1$

15) end while

16) while COMPARE(s_p,s_m) = 1 do

17) $q\leftarrow q-1$

18) end while

19) if $p<q$ then

20) exchange(s_p,s_m)

21) else

22) return p

23) end if

24) end loop

Yao 协议[9, 10]允许 2 个半诚实参与方分别输入 x 和 y 作为一个任意函数 $f(x,y)$的输入，协议能够保证两参与方私有信息安全的前提下，准确计算函数值，没有任何关于输入或者中间值的相关信息泄露。关于 Yao 协议的定理证明可以参见文献[10]。基于 Yao 协议实现的两方安全计算框架 FGC[11]近年来凭借其高效的性能得到普遍使用。本文将基于 FGC 中的 2-ADD 和 2-CMP 这 2 个基本模块，实现 Top$K$ 选择的比较模块。其中，2-ADD 可以实现任

意 2 个 L 位整数之间的加法，2-CMP 可以实现任意 2 个 L 位整数之间的比较，输出结果为 0 或 1。2-ADD 可以以密文形式还原推荐得分，随后使用 2-CMP 完成得分的比较即可。

3.2　基于同态加密的社会化推荐

在上述协议中，为了减少不必要的开销，实现高效率计算，Alice 和 Bob 需要将本身的数据分布情况共享给对方，尽管这一共享并没有泄露两参与方实际的数据值信息，但对于高安全级别的系统而言仍然存在一定程度的安全隐患。为了实现更高层次的安全保护，防止任何形式的信息泄露，提出了基于同态加密的完全隐私保护的社会化推荐协议。

Paillier 同态加密系统[8]是 Paillier 于 1999 年发明的用于公钥加密的概率非对称算法。在本文中，用 $E(x)$表示对明文 x 的 Paillier 加密函数，$D(x)$表示对密文 x 的 Paillier 解密函数，证明参见文献[8]。该加密系统具有加法同态性质，即 2 个密文乘积的解密值，与两密文对应明文的和相等；密文的 k 次幂解密值，与 k 和对应明文的乘积相等；Paillier 加密系统的语义安全特性保证了攻击者无法由给定密文导出任何相关明文信息。基于同态加密的推荐得分算法如下。

算法 3　基于 Paillier 同态机密的推荐得分算法

输入：Alice 端 $G_t = (U, I, E_t)$

Bob 端 $G_s = (U, E_s)$，相似度计算函数 $Fsim$

目标用户 $u \in U$

输出：所有物品 I 的推荐得分

Bob

1) $sim(u,v) = Fsim(G_s)$ $(v \in U, v \neq u)$
2) $En(sim(u, v))$ $(v \in U, v \neq u)$
3) send $En(sim(u, v))$

Alice

4) for $i \rightarrow N$
5)　for $j \rightarrow M$
6)　if $w(u_j,i) \neq 0$
7)　$En(s(u,i)) = En(s(u,i))En(sim(u,v))w(v,i)$
8)　end if
9)　end for
10) end for

为了保证数据信息的安全，Bob 端的相似度值需要用 Paillier 加密函数 En 加密后方可发送给 Alice 端，Paillier 的语义安全特性保证了 Alice 无法从密文获取与 Bob 相关的任何信息。同时，鉴于 Paillier 加密的同态性质，Alice 端可以单独计算出所有物品的推荐结果得分，得分以密文形式存在。Alice 端推荐得分计算公式如下

$$En(s(u,i)) = \prod En(sim(u,v))^{w(v,i)} (v \in U, v \neq u) \quad (2)$$

由于 $w(v,i)$是 Alice 端的数据，所以 Alice 可以自动过滤掉 $w(u_j,i) \neq 0$ 的项，减少不必要的计算开销。基于 Paillier 完成推荐得分计算后，结果以密文形式由 Alice 持有，密钥由 Bob 端持有。

因为 Pailler 并不支持基于密文的比较操作，Alice 无法单独实现安全的 Top$K$ 选择。加法秘密共享能够保证推荐得分对于两参与方的保密，同时，加法和秘密共享形式能保证安全的 Top$K$ 选择顺利进行。为了实现推荐得分的秘密共享，首先，Alice 生成 N 个足够大的随机数 r，并结合 Paillier 加密算法计算 $En(s-r)$，密文结果发送给 Bob；Bob 借助其本身持有的密钥可以解密数据，从而得到 $s'=s-r$。至此，Alice 端持有随机数 r，Bob 持有 s'，同时 $r+(s')=s$。尽管 Alice 持有随机数 r 和推荐得分的密文，但是 Alice 并没有密钥，所以无法得知任何与中间结果 s 相关的任何信息。Bob 持有密钥和 s'，但是 $s'=s-r$，由于随机数的干扰，Bob 端无法推测出 s 的相关信息。

接下来，调用算法 2 提出的安全 Top$K$ 选择算法，即可从所有候选物品中，挑选出 K 个推荐得分最高的物品推荐给用户。

4　理论分析

攻击模型：本文假设 Alice 和 Bob 都是半诚实的，两方将严格的执行协议，但是计算过程中两方也会尽可能地根据中间信息推测出更多的额外信息。针对恶意攻击模型的安全协议虽然存在，但是计算代价过大，在实际中并不实用。而针对半诚实模型的安全协议不但能够实现高效的计算，而且对恶意攻击模型下的安全协议研究具有重要参考价值。

4.1　安全性分析

如上所述，社会化推荐方法包括物品的推荐得分计算和 Top$K$ 选择 2 个过程。此处将就这 2 个过程逐一分析。

本文就不经意传输和同态加密提出了 2 种不同的隐私保护方法来计算推荐得分，在不经传输乘法协议中，两参与方私有数据信息安全，计算完成后，结果由两方以和形式秘密共享，即参与方 A 持有数据 a，参与方 B 持有数据 b，利用不经意传输乘法

协议后，A 将持有结果 x，B 持有结果 y，并且满足 $x+y=ab$（相关安全性证明参见文献[13]）。在不经意传输乘法协议可信的前提下，基于不经意传输的推荐得分计算不会泄露任何一方的私有信息。

在基于同态加密实现的推荐得分计算中，为了保证数据信息的安全，Bob 端将相似度值加密后发送给 Alice 端，Paillier 的语义安全特性保证了 Alice 无法从密文获取与 Bob 相关的任何信息。同时，基于密文，Alice 端可以单独计算出所有物品的推荐结果得分，得分以密文形式存在。为了实现加法秘密共享从而保证 Garbled Circuit 的调用，首先，Alice 生成 N 个足够大的随机数 r，并结合 Paillier 加密算法计算 $En(s-r)$，密文结果发送给 Bob；Bob 借助其本身持有的密钥可以解密数据，从而得到 $s'=s-r$。至此，Alice 端持有随机数 r，Bob 持有 s'，同时 $r+(s')=s$。在此过程中，尽管 Alice 持有随机数 r 和推荐得分的密文，但是 Alice 并没有密钥，所以无法得知任何与中间结果 s 相关的任何信息。Bob 持有密钥和 s'，但是 $s'=s-r$，由于随机数的干扰，Bob 端无法推测出 s 的相关信息。

完成推荐得分计算后，所有得分由 Alice 和 Bob 以加法和形式秘密共享，即 Alice 端持有 s_1，Bob 持有 s_2，同时 $s_1+s_2=s$。结合这一特点，可以基于 Garbled Circuit 提供的 2-ADD 和 2-CMP 这 2 个基本模块完成 Top$K$ 推荐。文献[10]指出在半诚实模型下，Garbled Circuit 允许 2 个参与方分别输入 x 和 y 作为一个任意函数 $f(x,y)$的输入，协议能够保证两参与方私有信息安全的前提下，准确计算函数值，没有任何关于输入或者中间值的相关信息泄露。这一性质保证了在 Top$K$ 推荐过程中，任何与两方相关的输入或者中间信息都不会泄露，该 Top$K$ 选择过程安全可靠。

综上，基于不经意传输的社会化推荐方法和基于同态加密的社会化推荐方法，都能在保证两方（推荐系统服务提供商和社交网络服务提供商）数据隐私的前提下，为目标用户提供精确的推荐。

4.2 复杂度分析

如表 1 所示为两方案的复杂度分析，其中，$|U|$ 表示用户个数，$|I|$表示物品个数，$|E_t|$表示用户历史购买记录条数，t 为推荐得分的二进制比特数。在基于不经意传输的社会化推荐中，步骤 1 对应 OT 乘法协议计算推荐得分，步骤 2 是安全的 Top$K$ 选择。在基于同态加密的社会化推荐方法中，步骤 1 至步骤 3 分别对应了基于 Paillier 同态加密的推荐得分，加法秘密共享及安全的 Top$K$ 选择。

表 1 复杂度分析

步骤	基于不经意传输的推荐	基于同态加密的社会化推荐
步骤 1	$\|E_t\|$不经意传输乘法	$\|U\|$加密，$\|I\|$指数计算
步骤 2	$(3t+1)\|I\|$非异或门	$\|I\|$加密，$\|I\|$解密，$\|I\|$指数计算
步骤 3	—	$(3t+1)\|I\|$非异或门

通过算法 1，可以明显看出，在不经意传输协议的步骤 1 中，双方共需调用不经意传输乘法协议 $|E_t|$次。同时，由上文可知，在 Top$K$ 选择中基本单元包含 2 个 2-ADD 和 1 个 2-CMP 模块，对于 t 位的电路输入，共包含非异或门 $3t+1$ 个。因为采用了线性时间复杂度的随机选择算法[14]来实现 Top$K$ 选择，其平均比较次数为 $|I|$，所以共需非异或门 $(3t+1)|I|$个。

在基于同态加密的社会化推荐方法中，由算法 3 可知，为了计算物品得分，Bob 共需$|U|$次加密操作，Alice 共需$|I|$次指数操作。在步骤 2，即加法秘密共享中，Alice 共需$|I|$次加密和指数操作，Bob 需要$|I|$次解密操作。由于两方案使用同一 Top$K$ 协议，所以复杂度仍为$(3t+1)|I|$。下面将通过实验对两方案的性能做进一步的比较。

5 实验部分

本文提出了 2 种方案，都能够在保证两参与方私有信息（G_t 和 G_s）不泄露的前提下，为目标用户提供精确的推荐。在这一部分，将使用 4 个公开数据集测试所提的方法，数据集相关统计信息如表 2 所示。

表 2 数据集统计信息

数据	Last.fm	Flixster	Brightkite	Gowalla
$\|U\|$	1 892	786 936	58 228	196 591
$\|E_s\|$	12 717	7 058 819	214 078	950 327
$\|I\|$	17 632	48 796	314 417	1 280 969
$\|E_t\|$	92 198	8 196 077	4 491 143	6 442 890

其中，$|U|$表示用户个数，$|I|$表示物品个数，$|E_s|$是用户之间的关联边数，$|E_t|$表示用户历史购买记录条数，从表中数据可以看出，用户历史记录矩阵相当稀疏。本实验主要包含了 4 个数据集，Last.fm[1]是

一个相对较小的数据集，主要包含了不同用户对音乐家的收听习惯。Flixter.com[2]是一个电影评分网站。Brightkite.com[3]和 Gowalla.com[4]则是基于位置信息的信息分享网站，主要记录不同用户在不同地点有多次的登录行为。

实验在 2.6 GHz CPU，1TB RAM 的服务器上执行，软件环境为 Centos Linux release 7.1，JDK7。

使用 Java 实现了 Paillier 同态加密系统，实验中，密钥空间设为 1 024 bit（1 024 bit 相当于对称密钥方案中 80 bit 的安全级别，在这个设置下，可以忽略由于密码被攻破而带来的信息泄露）。基于 FGC 框架，实现了安全的 Top$K$ 选择协议。默认情况下，推荐数 K 设置为 10。

实验结果如表 3 和表 4 所示，表 3 和表 4 分别是方法 1 和方法 2 对应的时间统计。表 3 中，步骤 1 对应 OT 乘法协议计算推荐得分，步骤 2 是安全的 Top$K$ 选择。表 4 中包含了 3 步操作，基于 Paillier 同态加密的推荐得分以及得分的秘密共享及安全的 Top$K$ 选择。

表 3　基于 OT 乘法的社会化推荐系统时间统计

数据集	步骤 1/min	步骤 2/min	总时间/min
Last.fm	2.39	1.22	3.61
Flixster	161.54	3.39	164.93
Brightkite	10.25	23.27	33.52
Gowalla	162.52	88.55	251.07

表 4　基于同态加密的社会化推荐系统时间统计

数据集	步骤 1/min	步骤 2/min	步骤 3/min	总时间/min
Last.fm	4.14	7.11	1.22	12.48
Flixster	102.95	19.40	3.39	125.74
Brightkite	70.37	134.23	23.27	227.87
Gowalla	249.92	467.85	88.55	806.32

结合算法分析可知，由于矩阵的稀疏度影响，表 3 中步骤 1 的时间主要与非零项相关。步骤 2 中时间和物品数$|I|$成正比。由于表 4 中步骤 1 推荐得分计算与$|E_t|$成正比，步骤 2 加法秘密共享与$|I|$成正比。

从表中可以看到，除了 Flixster 外的其他 3 个数据集上，基于 OT 乘法的社会化推荐协议比基于同态加密的社会化推荐方案更为高效。而 Flixster 由于矩阵的非零项记录$|E_t|$远大于$|I|$，所以基于 OT 乘法的协议并没有因为省略加法秘密共享操作而获得足够优势。在方法选择过程中，用户可以根据自己的数据集特性及要求选择合适的方案。

6　结束语

本文提出了 2 种数据隐私保护的社会化推荐协议。两协议都能够在保证不泄露两参与方私有数据信息的前提下，完成社会化推荐。其中，基于不经意传输的社会化推荐，计算代价较小，适用于对推荐效率要求较高的应用。基于同态加密的社会化推荐，安全程度较高，适用于对数据隐私要求较高的应用。4 组真实数据集实验表明，本文提出的方案切实可行，用户可以根据自身需求选择合适的方案。

参考文献：

[1] KONSTAS I, STATHOPOULOS V, JOSE J M. On social networks and collaborative recommendation[A]. 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval[C]. ACM, 2009. 195-202.

[2] YUAN Q, ZHAO S, CHEN L, et al. Augmenting collaborative recommender by fusing explicit social relationships[A]. Workshop on Recommender Systems and the Social Web[C]. 2009. 2009.

[3] JORGENSEN Z, YU T. A privacy-preserving framework for personalized, social recommendations[A]. EDBT[C]. 2014. 571-582.

[4] HAN S, NG W K, YU P S. Privacy-preserving singular value decomposition[A]. ICDE[C]. 2009.

[5] CANNY J. Collaborative filtering with privacy[A]. Security and Privacy, Proceedings of 2002 IEEE Symposium[C]. IEEE, 2002. 45-57.

[6] MILLER B N, KONSTAN J A, RIEDL J. PocketLens: toward a personal recommender system[J]. ACM Transactions on Information Systems (TOIS), 2004, 22(3): 437-476.

[7] POLAT H, DU W. Privacy-preserving collaborative filtering[J]. International Journal of Electronic Commerce, 2005, 9(4): 9-35.

[8] BERKOVSKY S, EYTANI Y, KUFLIK T, et al. Enhancing privacy and preserving accuracy of a distributed collaborative filtering[A]. 2007 ACM Conference on Recommender Systems[C]. ACM, 2007. 9-16.

[9] YAO A. How to generate and exchange secrets[A]. Foundations of Computer Science 27th Annual Symposium on[C]. 1986. 162-167.

[10] LINDELL Y, PINKAS B. A proof of security of Yao's protocol for two-party computation[J]. Journal of Cryptology, 2009, 22(2): 161-188.

[11] HUANG Y, EVANS D, KATZ J, et al. Faster secure two-party computation using garbled circuits[A]. USENIX Security Symposium[C]. 2011, 201(1).

[12] NAOR M, PINKAS B. Efficient oblivious transfer protocols[A].

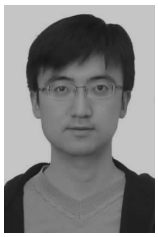
Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics[C]. 2001. 448-457.

[13] GILBOA N. Two party RSA key generation[A]. Advances in Cryptology—CRYPTO'99[C]. Springer Berlin Heidelberg, 1999. 116-129.

[14] CORMEN T H. Introduction to Algorithms[M]. MIT Press, 2009.

作者简介：

刘曙曙（1992-），女，江苏南通人，苏州大学硕士生，主要研究方向为数据安全与隐私和推荐系统。



刘安（1981-），男，安徽泾县人，博士，苏州大学副教授、硕士生导师，主要研究方向为数据安全与隐私、时空数据库、云计算与服务计算以及图数据库等。

赵雷（1972-），男，江苏无锡人，博士，苏州大学教授，主要研究方向为数据库及数据仓库、图数据库、空间数据库、数据挖掘、并行及分布式系统、高性能计算等。

刘冠峰（1982-），男，山东青岛人，博士，苏州大学副教授、硕士生导师，主要研究方向为可信计算、社交网络信息挖掘、图数据库等。

李直旭（1983-），男，安徽泾县人，博士，苏州大学副教授、硕士生导师，主要研究方向为数据库、机器学习与数据挖掘、大数据应用、信息检索与信息抽取、数据质量以及移动计算等。

郑凯（1983-），男，山东淄博人，博士，苏州大学特聘教授、博士生导师，主要研究方向为大数据管理、社交媒体数据分析、时空数据库、不确定数据库、内存数据库、数据挖掘等。

周晓方（1963-），男，江苏无锡人，博士，苏州大学特聘教授，主要研究方向为空间数据库、多媒体数据库、数据质量、高性能数据处理及网络信息系统，以及这些技术在生物信息学、地理信息系统、移动对象管理、水文信息系统、医疗卫生系统、Web查询及视频数据检索等方面的应用。

勘误声明

本刊2015年第6期刊出的《卫星通信的近期发展与前景展望》一文的作者为易克初、李怡、孙晨华、南春国。作者简介中孙晨华（男）应为孙晨华（女），特此更正，并向作者致歉。

《通信学报》编辑部