



(12) 发明专利

(10) 授权公告号 CN 110059498 B

(45) 授权公告日 2021.07.23

(21) 申请号 201910216242.7

G06Q 50/00 (2012.01)

(22) 申请日 2019.03.21

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 105574434 A, 2016.05.11

申请公布号 CN 110059498 A

CN 101655895 A, 2010.02.24

CN 103268454 A, 2013.08.28

(43) 申请公布日 2019.07.26

CN 104981816 A, 2015.10.14

(73) 专利权人 中国科学院计算技术研究所

审查员 谢永坚

地址 100080 北京市海淀区中关村科学院

南路6号

(72) 发明人 曹娟 郭俊波 谢添 刘浩远

吕博 王蕊

(74) 专利代理机构 北京律诚同业知识产权代理

有限公司 11006

代理人 祁建国 梁挥

(51) Int. Cl.

G06F 21/62 (2013.01)

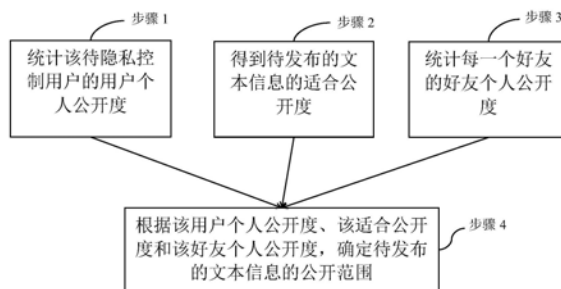
权利要求书1页 说明书6页 附图1页

(54) 发明名称

一种用于社交网络的隐私控制自动设置方法及系统

(57) 摘要

本发明提出一种用于社交网络的隐私控制自动设置方法及系统,包括:根据待隐私控制用户的个人档案中各用户属性的可见性,统计待隐私控制用户的用户个人公开度;获取由多种隐私属性组成的隐私属性集合,以及获取待隐私控制用户待发布的文本信息,将文本信息输入至多个分类器,每个分类器对应一种隐私属性,分类器输出文本信息属于对应隐私属性取值上的概率分布,归一化概率分布的熵,得到文本信息的适合公开度;根据待隐私控制用户的好友的个人档案和发布内容,得到待隐私控制用户的每一个好友的好友个人公开度;根据用户个人公开度、适合公开度和好友个人公开度,度量文本信息对于各好友的隐私敏感度,并根据隐私敏感度确定文本信息的公开范围。



1. 一种用于社交网络的隐私控制自动设置方法,其特征在于,包括:

步骤1、根据待隐私控制用户的个人档案中各用户属性的可见性,统计该待隐私控制用户的用户个人公开度;

步骤2、获取由多种隐私属性组成的隐私属性集合,以及获取该待隐私控制用户待发布的文本信息,将该文本信息输入至多个分类器,每个该分类器对应一种该隐私属性,分类器输出该文本信息属于对应隐私属性取值上的概率分布,归一化该概率分布的熵,得到该文本信息的适合公开度;

步骤3、根据该待隐私控制用户的好友的个人档案和发布内容,得到该待隐私控制用户的每一个好友的好友个人公开度;

步骤4、根据该用户个人公开度、该适合公开度和该好友个人公开度,度量该文本信息对于各好友的隐私敏感度,并根据该隐私敏感度确定该文本信息的公开范围。

2. 如权利要求1所述的用于社交网络的隐私控制自动设置方法,其特征在于,步骤1中该用户个人公开度的具体统计方式为:

$$U = \sum_i w_{ui} v_i$$

其中U为该用户个人公开度, $w_{ui}$ 表示该用户属性i对应的权值,且 $\sum_i w_{ui} = 1$ , $v_i$ 表示第i种属性的可见性,当第i种属性对所有人可见时 $v_i = 1$ ,分组可见时 $v_i = 0.5$ ,仅自己可见时 $v_i = 0$ 。

3. 一种用于社交网络的隐私控制自动设置系统,其特征在于,包括:

模块1、根据待隐私控制用户的个人档案中各用户属性的可见性,统计该待隐私控制用户的用户个人公开度;

模块2、获取由多种隐私属性组成的隐私属性集合,以及获取该待隐私控制用户待发布的文本信息,将该文本信息输入至多个分类器,每个该分类器对应一种该隐私属性,分类器输出该文本信息属于对应隐私属性取值上的概率分布,归一化该概率分布的熵,得到该文本信息的适合公开度;

模块3、根据该待隐私控制用户的好友的个人档案和发布内容,得到该待隐私控制用户的每一个好友的好友个人公开度;

模块4、根据该用户个人公开度、该适合公开度和该好友个人公开度,度量该文本信息对于各好友的隐私敏感度,并根据该隐私敏感度确定该文本信息的公开范围。

4. 如权利要求3所述的用于社交网络的隐私控制自动设置系统,其特征在于,模块1中该用户个人公开度的具体统计方式为:

$$U = \sum_i w_{ui} v_i$$

其中U为该用户个人公开度, $w_{ui}$ 表示该用户属性i对应的权值,且 $\sum_i w_{ui} = 1$ , $v_i$ 表示第i种属性的可见性,当第i种属性对所有人可见时 $v_i = 1$ ,分组可见时 $v_i = 0.5$ ,仅自己可见时 $v_i = 0$ 。

## 一种用于社交网络的隐私控制自动设置方法及系统

### 技术领域

[0001] 本发明涉及社交网络的隐私保护,特别涉及一种根据用户信息自动设置其隐私控制的方法和系统。

### 背景技术

[0002] 近年来在线社交网络(OSN)的快速发展促进了人与人的沟通,加快了信息的传播。与此同时,用户间的交互增多,不恰当的隐私设置会使个人信息扩散到无法预料的范围,使得个人的隐私受到威胁,产生很多隐私相关的问题。

[0003] 虽然大多数OSN,如微博、Twitter、各类论坛等,都提供了个人档案中属性的隐私设置,如所有人可见、分组可见、仅自己可见等,但这类设置通常默认为所有人可见,且大部分隐私设置的位置并不明显。此外,当用户发布信息时,默认的隐私设置也是所有人可见,而不会根据用户发布内容自动调整,导致用户会无意识地泄露自己的隐私。

### 发明内容

[0004] 针对现有技术的不足,本发明提出用于社交网络的隐私控制自动设置方法,包括用户个人档案中属性的隐私保护程度、用户发布内容包含隐私相关信息程度、用户好友隐私保护程度的度量方法,结合三者的度量值,给出了用户发布内容的隐私控制的自动设置方法。

[0005] 具体来说,本发明提供了一种用于社交网络的隐私控制自动设置方法,其中包括:

[0006] 步骤1、根据待隐私控制用户的个人档案中各用户属性的可见性,统计该待隐私控制用户的用户个人公开度;

[0007] 步骤2、获取由多种隐私属性组成的隐私属性集合,以及获取该待隐私控制用户待发布的文本信息,将该文本信息输入至多个分类器,每个该分类器对应一种该隐私属性,分类器输出该文本信息属于对应隐私属性取值上的概率分布,归一化该概率分布的熵,得到该文本信息的适合公开度;

[0008] 步骤3、根据该待隐私控制用户的好友的个人档案和发布内容,得到该待隐私控制用户的每一个好友的好友个人公开度;

[0009] 步骤4、根据该用户个人公开度、该适合公开度和该好友个人公开度,度量该文本信息对于各好友的隐私敏感度,并根据该隐私敏感度确定该文本信息的公开范围。

[0010] 所述的用于社交网络的隐私控制自动设置方法,其中步骤1中该用户个人公开度的具体统计方式为:

$$[0011] \quad U = \sum_i w_{ui} v_i$$

[0012] 其中U为该用户个人公开度, $w_{ui}$ 表示该用户属性i对应的权值,且 $\sum_i w_{ui} = 1$ , $v_i$ 表示该属性i的可见性。

[0013] 所述的用于社交网络的隐私控制自动设置方法,其中步骤2中该适合公开度的确

定方式为：

$$[0014] \quad E_i = - \sum_j p_{ij} \log p_{ij} / \log n_i$$

$$[0015] \quad C = \sum_i w_{ci} E_i$$

[0016] 其中，C为该适合公开度， $p_{ij}$ 为隐私属性 $S_i$ 取值上的概率分布， $n_i$ 表示隐私属性 $S_i$ 取值的个数， $w_{ci}$ 表示隐私属性 $S_i$ 的权重，且满足 $\sum_i w_{ci} = 1$ 。

[0017] 所述的用于社交网络的隐私控制自动设置方法，其中步骤3中好友i的该好友个人公开度的确定方式为：

$$[0018] \quad F_i = (1 - U_i) \sum_j C_j / n_c$$

[0019] 其中， $U_i$ 为该好友i的用户个人公开度， $\sum_j C_j / n_c$ 用来衡量好友i发布过的内容中泄露隐私敏感信息的程度， $n_c$ 为好友i发布内容的总数， $C_j$ 表示好友发布的第j个文本的隐私敏感程度。

[0020] 所述的用于社交网络的隐私控制自动设置方法，其中步骤4中好友i的隐私敏感度度量方式为：

$$[0021] \quad M_i = U * C * F_i$$

[0022] 其中， $M_i$ 为好友i的隐私敏感度。

[0023] 本发明还提供了一种用于社交网络的隐私控制自动设置系统，其中包括：

[0024] 模块1、根据待隐私控制用户的个人档案中各用户属性的可见性，统计该待隐私控制用户的用户个人公开度；

[0025] 模块2、获取由多种隐私属性组成的隐私属性集合，以及获取该待隐私控制用户待发布的文本信息，将该文本信息输入至多个分类器，每个该分类器对应一种该隐私属性，分类器输出该文本信息属于对应隐私属性取值上的概率分布，归一化该概率分布的熵，得到该文本信息的适合公开度；

[0026] 模块3、根据该待隐私控制用户的好友的个人档案和发布内容，得到该待隐私控制用户的每一个好友的好友个人公开度；

[0027] 模块4、根据该用户个人公开度、该适合公开度和该好友个人公开度，度量该文本信息对于各好友的隐私敏感度，并根据该隐私敏感度确定该文本信息的公开范围。

[0028] 所述的用于社交网络的隐私控制自动设置系统，其中模块1中该用户个人公开度的具体统计方式为：

$$[0029] \quad U = \sum_i w_{ui} v_i$$

[0030] 其中U为该用户个人公开度， $w_{ui}$ 表示该用户属性i对应的权值，且 $\sum_i w_{ui} = 1$ ， $v_i$ 表示该属性i的可见性。

[0031] 所述的用于社交网络的隐私控制自动设置系统，其中模块2中该适合公开度的确定方式为：

$$[0032] \quad E_i = - \sum_j p_{ij} \log p_{ij} / \log n_i$$

$$[0033] \quad C = \sum_i w_{ci} E_i$$

[0034] 其中,  $C$  为该适合公开度,  $p_{ij}$  为隐私属性  $S_i$  取值上的概率分布,  $n_i$  表示隐私属性  $S_i$  取值的个数,  $w_{ci}$  表示隐私属性  $S_i$  的权重, 且满足  $\sum_i w_{ci} = 1$ 。

[0035] 所述的用于社交网络的隐私控制自动设置系统, 其中模块3中好友  $i$  的该好友个人公开度的确定方式为:

$$[0036] \quad F_i = (1 - U_i) \sum_j C_j / n_c$$

[0037] 其中,  $U_i$  为该好友  $i$  的用户个人公开度,  $\sum_j C_j / n_c$  用来衡量好友  $i$  发布过的内容中泄露隐私敏感信息的程度,  $n_c$  为好友  $i$  发布内容的总数,  $C_j$  表示好友发布的第  $j$  个文本的隐私敏感程度。

[0038] 所述的用于社交网络的隐私控制自动设置系统, 其中模块4中好友  $i$  的隐私敏感度度量方式为:

$$[0039] \quad M_i = U * C * F_i$$

[0040] 其中,  $M_i$  为好友  $i$  的隐私敏感度。

[0041] 由以上方案可知, 本发明包括: 基于对用户的分析, 设置用户个人档案页面中各项属性的默认可见范围和基于用户发布的内容, 结合用户的隐私偏好、好友隐私保护程度, 设置欲发布内容的默认可见范围。本发明的优点在于: 在用户发布内容时, 提供一种自动的隐私控制设置方法, 根据其个人隐私偏好、发布的内容及其好友的隐私保护程度, 为这次内容的发布提供保护用户隐私的分享策略。相比于现有技术, 保护隐私的策略更加灵活, 无需用户手动为每次分享做隐私设置。

## 附图说明

[0042] 图1为本发明流程图。

## 具体实施方式

[0043] 本发明的目的是提出一种自动设置隐私控制的策略, 缓解信息分享导致的隐私泄露问题。

[0044] 本发明将用户的隐私控制分为两部分: 用户个人档案中属性的隐私控制和用户发布内容的隐私控制。本发明根据用户信息, 给这两类隐私控制提供更能保护隐私的默认设置。

[0045] 一、用户个人档案中属性的隐私控制。

[0046] 针对具体应用的平台, 根据其活跃用户数量, 采取两类统计方法。若该平台已有足够的活跃用户, 则直接对其用户进行统计分析; 否则, 借助其他大型OSN(如微博, Twitter等)的用户数据进行统计分析。活跃用户的定义视平台不同有所区别, 例如对于微博, 可以

定义微博总数大于30条、账户注册时间多于一个月、原创微博与转发微博比例在一定范围内的用户为活跃用户。

[0047] 分析用户对用户档案中各类属性隐私敏感程度的普遍观点,对每种属性(姓名,年龄,住址,婚姻状况等)分别进行统计分析。若该属性有超过预设数值或比例的用户(例如半数用户)设为他人不可见,则该属性的默认隐私设置为仅自己可见。

[0048] 用户可自行修改其个人档案中属性的可见性,定义U为用户个人档案中信息的公开程度,用于后续用户发布内容的隐私设置。计算如下:

$$[0049] \quad U = \sum_i w_{ui} v_i$$

[0050] 其中, $w_{ui}$ 表示用户属性i对应的权值,且满足 $\sum_i w_{ui} = 1$ , $v_i$ 表示该属性i的可见性,所有人可见时 $v_i = 1$ ,分组可见时 $v_i = 0.5$ ,仅自己可见时 $v_i = 0$ 。不失一般性,此处各属性权重 $w_{ui}$ 平均分配。

[0051] 二、如图1所示,用户发布内容的隐私控制。

[0052] 1、计算用户发布内容包含隐私相关信息的程度。

[0053] 本发明只考虑文本信息的隐私问题。定义所有隐私敏感的属性类别为集合S,集合的设定可参考各大公司的隐私条款,例如:健康状况;政治、宗教信仰;年龄;性取向;性别;残疾;经济状况等。

[0054] 对于文本内容,基于机器学习的方法训练多组分类器,每个分类器对应一种隐私敏感的属性 $S_i$ ,输出向量 $p_i$ ,该向量的每一维表示一段文本在属性 $S_i$ 各个可能取值j上的概率分布 $p_{ij}$ 。计算 $p_i$ 的熵,归一化得到 $E_i$ 。例如:敏感属性 $S_i$ ,如 $S_i =$ '年龄段',可能的取值有<20;20-40;>40三种, $n_i = 3$ ,向量 $p_i = [0.3, 0.4, 0.3]$ ,分别对应三个类别的概率,最后计算向量 $p_i$ 的熵, $n_i$ 表示属性 $S_i$ 所有可能取值的个数,则 $E_i$ 计算如下:

$$[0055] \quad E_i = - \sum_j p_{ij} \log p_{ij} / \log n_i$$

[0056]  $E_i$ 越小,表明这段文本在属性 $S_i$ 上不确定性越低,越容易泄露该属性的隐私,反之表示与该属性关系较弱。因此使用 $E_i$ 度量这段文本对于敏感属性 $S_i$ 适合公开的程度。

[0057] 对所有敏感属性的 $E_i$ 加权,得到C:

$$[0058] \quad C = \sum_i w_{ci} E_i$$

[0059] 其中 $w_{ci}$ 表示各个属性的权重,且满足 $\sum_i w_{ci} = 1$ 。C表示综合考虑各类敏感属性,这段文本适合公开的程度。C越接近1,内容与隐私敏感的属性越无关,从隐私保护的角度而言更适合公开。不失一般性,此处各类别权重平均分配。

[0060] 2、度量用户好友的隐私保护程度。

[0061] 用户好友的隐私保护程度由好友的用户个人档案和发布内容共同决定,用以度量该用户的好友i的隐私意识,也即向该好友公开要发布的内容的可靠程度,用 $F_i$ 表示,计算如下:

$$[0062] \quad F_i = (1 - U_i) \sum_j C_j / n_c$$

[0063] 其中,  $U_i$  为该好友的U值, 表示其个人档案的可见程度,  $\sum_j C_j / n_c$  用来衡量其发布过的内容中泄露隐私敏感信息的程度,  $n_c$  为其发布过内容的总数,  $C_j$  表示好友发布的第j个文本的隐私敏感程度, 加和后取平均, 衡量好友所有发布内容的平均敏感程度。

[0064]  $F_i$  越高, 表明该好友隐私意识越强, 发布内容对该好友可见的隐私风险较低。

[0065] 3、结合用户的过往偏好, 设置该信息的默认可见范围。

[0066] 结合上述度量值, 计算要发布的内容对好友i的度量值  $M_i$

$$[0067] \quad M_i = U * C * F_i$$

[0068] 设定阈值T, 当  $M_i > T$  时, 要发布的内容设置为对该好友可见。即当用户隐私意识较弱 (U较大)、文本内容与隐私敏感属性关系较小 (C较大)、好友隐私意识较强 ( $F_i$  较大) 时, 用户要发布的内容更可能向好友i公开。

[0069] 以下为与上述方法实施例对应的系统实施例, 本实施方式可与上述实施方式互相配合实施。上述实施方式中提到的相关技术细节在本实施方式中依然有效, 为了减少重复, 这里不再赘述。相应地, 本实施方式中提到的相关技术细节也可应用在上述实施方式中。

[0070] 本发明还提供了一种用于社交网络的隐私控制自动设置系统, 其中包括:

[0071] 模块1、根据待隐私控制用户的个人档案中各用户属性的可见性, 统计该待隐私控制用户的用户个人公开度;

[0072] 模块2、获取由多种隐私属性组成的隐私属性集合, 以及获取该待隐私控制用户待发布的文本信息, 将该文本信息输入至多个分类器, 每个该分类器对应一种该隐私属性, 分类器输出该文本信息属于对应隐私属性取值上的概率分布, 归一化该概率分布的熵, 得到该文本信息的适合公开度;

[0073] 模块3、根据该待隐私控制用户的好友的个人档案和发布内容, 得到该待隐私控制用户的每一个好友的好友个人公开度;

[0074] 模块4、根据该用户个人公开度、该适合公开度和该好友个人公开度, 度量该文本信息对于各好友的隐私敏感度, 并根据该隐私敏感度确定该文本信息的公开范围。

[0075] 所述的用于社交网络的隐私控制自动设置系统, 其中模块1中该用户个人公开度的具体统计方式为:

$$[0076] \quad U = \sum_i w_{ui} v_i$$

[0077] 其中U为该用户个人公开度,  $w_{ui}$  表示该用户属性i对应的权值, 且  $\sum_i w_{ui} = 1$ ,  $v_i$  表示该属性i的可见性。

[0078] 所述的用于社交网络的隐私控制自动设置系统, 其中模块2中该适合公开度的确定方式为:

$$[0079] \quad E_i = - \sum_j p_{ij} \log p_{ij} / \log n_i$$

$$[0080] \quad C = \sum_i w_{ci} E_i$$

[0081] 其中,  $C$  为该适合公开度,  $p_{ij}$  为隐私属性  $S_i$  取值上的概率分布,  $n_i$  表示隐私属性  $S_i$  取值的个数,  $w_{ci}$  表示隐私属性  $S_i$  的权重, 且满足  $\sum_i w_{ci} = 1$ 。

[0082] 所述的用于社交网络的隐私控制自动设置系统, 其中模块3中好友  $i$  的该好友个人公开度的确定方式为:

$$[0083] \quad F_i = (1 - U_i) \sum_j C_j / n_c$$

[0084] 其中,  $U_i$  为该好友  $i$  的用户个人公开度,  $\sum_j C_j / n_c$  用来衡量好友  $i$  发布过的内容中泄露隐私敏感信息的程度,  $n_c$  为好友  $i$  发布内容的总数,  $C_j$  表示好友发布的第  $j$  个文本的隐私敏感程度。

[0085] 所述的用于社交网络的隐私控制自动设置系统, 其中模块4中好友  $i$  的隐私敏感度度量方式为:

$$[0086] \quad M_i = U * C * F_i$$

[0087] 其中,  $M_i$  为好友  $i$  的隐私敏感度。



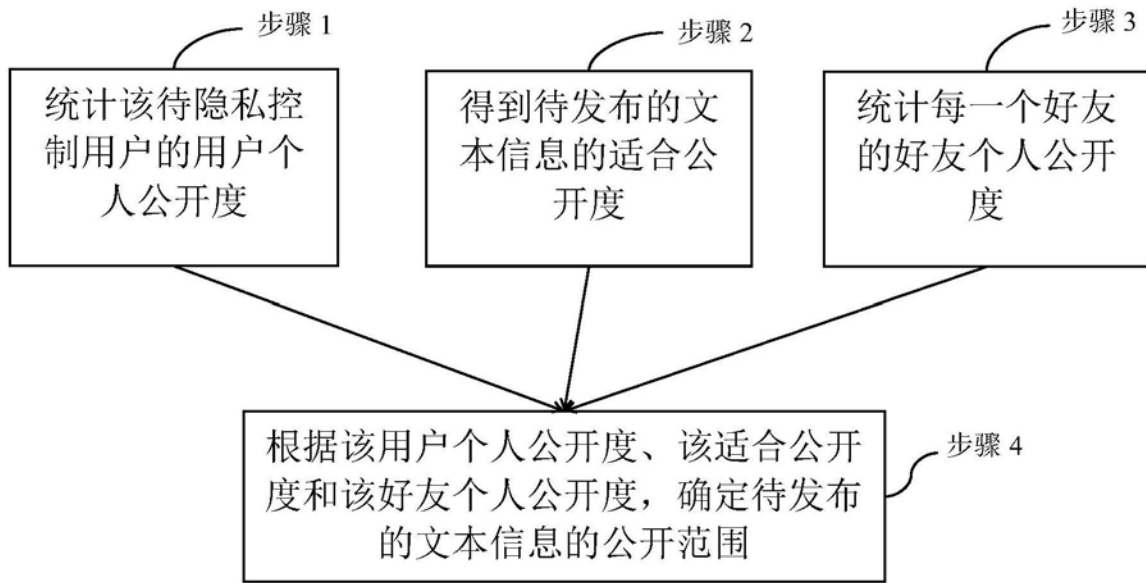


图1