



(12) 发明专利

(10) 授权公告号 CN 103118043 B

(45) 授权公告日 2015. 12. 02

(21) 申请号 201110363026. 9

CN 102200987 A, 2011. 09. 28, 全文.

(22) 申请日 2011. 11. 16

审查员 蔡宜飞

(73) 专利权人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼岛资本大厦一座
四层 847 号邮箱

(72) 发明人 张立柱 张正威 杨廷辉

(74) 专利代理机构 北京同达信恒知识产权代理
有限公司 11291

代理人 郭润湘

(51) Int. Cl.

H04L 29/08(2006. 01)

G06F 17/30(2006. 01)

(56) 对比文件

US 7725421 B1, 2010. 05. 25, 说明书第 5 栏
35 行到第 6 页第 52 行, 第 7 栏第 45 行到第 10 栏
第 36 行、附图 5.

US 2007192122 A1, 2007. 08. 16, 全文.

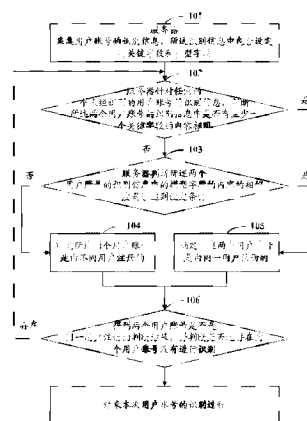
权利要求书2页 说明书9页 附图2页

(54) 发明名称

一种用户账号的识别方法及设备

(57) 摘要

本申请公开了一种用户账号的识别方法及设备, 通过采集各用户账号所对应的识别信息, 所述识别信息中包含设定的关键字段和模型字段, 当任意两个用户账号的识别信息中包含有至少一个相同的关键字段内容时, 则确定该两个用户账号为同一用户所注册, 否则, 需要判断该两个用户账号的识别信息中模型字段的内容相似度, 并根据所述模型字段的内容相似度评级来判断该两个用户账号是否为同一用户所注册, 从而解决了无法有效地从大量的用户账号中识别出同一用户注册的用户账号的问题。



1. 一种用户账号的识别方法,其特征在于,包括以下步骤:

服务器采集用户账号的识别信息,所述识别信息中包含设定的关键字段和模型字段;

针对任意两个用户账号的识别信息,所述服务器判断所述两个用户账号的识别信息中是否有至少一个关键字段的内容相同;

若是,则确定该两个用户账号是由同一用户注册的;

否则,所述服务器确定识别信息中的模型字段的内容相似度,判断所述两个用户账号的识别信息中的模型字段的内容相似度是否达到设定条件,若是,则确定该两个用户账号是由同一用户注册的,否则,确定该两个用户账号是由不同用户注册的;

所述确定识别信息中的模型字段的内容相似度包括:

确定各模型字段的权重值;具体地,采用层次分析模型的方式,通过建立层次结构模型、构造成对比较阵、计算权向量并做一致性检验的步骤来得到各模型字段的权重值;

分别将两个用户账号的识别信息中名称相同的模型字段的内容进行相似度运算,并将各运算结果与对应的模型字段的权重值进行加权求和,得到模型字段的内容相似度;

其中,所述识别信息包括以下至少一种信息:

用户账号注册时的注册信息、从第三方获取的信息和用户账号运行时产生的信息;

以及,针对同一用户账号的相同名称的字段出现不同内容的情况,确定不同内容所属来源的权重,保留权重较高的来源提供的内容,或者,将不同来源的针对同一字段不同的内容,都作为该字段的内容保留。

2. 如权利要求1所述的用户账号的识别方法,其特征在于,通过以下方式确定两个用户账号的识别信息中的模型字段的内容相似度:

确定识别信息中的注册信息、从第三方获取的信息和用户账号运行时产生的信息的权重值;

分别将两个用户账号的识别信息中名称相同的模型字段的内容进行相似度运算,并确定进行相似度运算的内容所属信息的权重值;

将各运算结果与进行相似度运算的内容所属信息的权重值进行加权求和,得到模型字段的内容相似度。

3. 一种服务器,其特征在于,包括:

数据采集模块,用于采集用户账号的识别信息,所述识别信息中包含设定的关键字段和模型字段;

识别判断模块,用于接收所述数据采集模块发送的用户账号的识别信息,针对任意两个用户账号的识别信息,判断所述两个用户账号的识别信息中是否有至少一个关键字段的内容相同,若是,则确定该两个用户账号是由同一用户注册的,否则,判断所述两个用户账号的识别信息中的模型字段的内容相似度是否达到设定条件,若是,则确定该两个用户账号是由同一用户注册的,否则,确定该两个用户账号是由不同用户注册的;

第一相似度运算模块,用于确定各模型字段的权重值,并分别将两个用户账号的识别信息中名称相同的模型字段的内容进行相似度运算,将各运算结果与对应的模型字段的权重值进行加权求和,得到模型字段的内容相似度并发送至识别判断模块;具体地,采用层次分析模型的方式,通过建立层次结构模型、构造成对比较阵、计算权向量并做一致性检验的步骤来得到各模型字段的权重值;

其中,所述数据采集模块采集到的用户账号的识别信息具体包括以下至少一种:

用户账号注册时的注册信息、从第三方获取的信息和用户账号运行时产生的信息;

以及,针对同一用户账号的相同名称的字段出现不同内容的情况,确定不同内容所属来源的权重,保留权重较高的来源提供的内容,或者,将不同来源的针对同一字段不同的内容,都作为该字段的内容保留。

4. 如权利要求 3 所述的服务器,其特征在于,还包括:

第二相似度运算模块,用于确定识别信息中的注册信息、从第三方获取的信息和用户账号运行时产生的信息的权重值,分别将两个用户账号的识别信息中名称相同的模型字段的内容进行相似度运算,并确定进行相似度运算的内容所属信息的权重值,并将各运算结果与进行相似度运算的内容所属信息的权重值进行加权求和,得到模型字段的内容相似度;

所述识别判断模块,具体用于根据所述第二相似度运算模块得到的内容相似度,判断所述两个用户账号的识别信息中的模型字段的内容相似度是否达到设定条件。

5. 如权利要求 3-4 任一所述的服务器,其特征在于,还包括:

数据存储模块,用于存储数据采集模块所采集到的各用户账号的识别信息;

人工识别模块,用于提供操作界面来接收人工干预信息,利用接收到的人工干预信息对识别判断模块所得到的识别结果再次进行判定。

6. 如权利要求 5 所述的服务器,其特征在于,

所述数据存储模块,还用于将由同一用户注册的多个用户账号进行绑定,并存储绑定信息。

一种用户账号的识别方法及设备

技术领域

[0001] 本申请涉及计算机网络技术领域,尤其涉及一种用户账号的识别方法及设备。

背景技术

[0002] 随着互联网技术的不断发展,互联网业务也得到飞速发展,多种多样的互联网网站为不同的用户提供了种类繁多的业务服务,如:用户可以利用互联网网站进行商品的买卖、发表时政见解以及进行交友活动等,从而将现实生活中需要面对面进行的业务交易或来往,扩展到了虚拟网络领域,丰富了用户的业务交易方式。

[0003] 在利用上述互联网的方式进行相应的业务时,用户需要通过在互联网网站注册用户账号的方式来登录该网站,从而使用该网站为用户提供的业务服务信息,完成用户所需要的业务。

[0004] 由于目前网络身份证和网络实名制等身份认证制度尚未实施,绝大多数网站只需要用户提供自命名的用户名和其他身份信息(如合法的邮箱信息、手机号码)等,就可以注册一个新的用户账号,因此,同一用户可以利用不同的用户名和身份信息重复注册多个用户账号。

[0005] 网站服务器在对本地注册的用户账号进行管理时,是无法区分已注册的用户账号是否是同一用户注册的,网站服务器只能不加区别地对各个已注册的用户账号分别进行管理,如向用户账号推送信息、对该用户账号在网站服务器中登录后进行的操作管理等。

[0006] 而实际上,对于同一用户注册的多个用户账号之间很可能具有一定的相关性,用户使用各用户账号登录网站服务器后的操作也具有一定的相关性,因此,如果能找到一种从大量的用户账号中识别出同一用户注册的用户账号的方案,则可以将同一用户注册的用户账号集中管理,有效提高对用户账号的管理效率,进一步地,还可以有效地分析、跟踪用户行为,向用户注册的用户账号精准地推送有用信息。

发明内容

[0007] 本申请的目的在于:提供一种用户账号的识别方法及设备,用以解决现有互联网网站中无法有效地从大量的已注册用户账号中识别出同一用户注册的用户账号的问题。

[0008] 一种用户账号的识别方法,包括以下步骤:

[0009] 服务器采集用户账号的识别信息,所述识别信息中包含设定的关键字段和模型字段;

[0010] 针对任意两个用户账号的识别信息,所述服务器判断所述两个用户账号的识别信息中是否有至少一个关键字段的内容相同;

[0011] 若是,则确定该两个用户账号是由同一用户注册的;

[0012] 否则,所述服务器确定识别信息中的模型字段的内容相似度,判断所述两个用户账号的识别信息中的模型字段的内容相似度是否达到设定条件,若是,则确定该两个用户账号是由同一用户注册的,否则,确定该两个用户账号是由不同用户注册的;

[0013] 所述确定识别信息中的模型字段的内容相似度包括：

[0014] 确定各模型字段的权重值；

[0015] 分别将两个用户账号的识别信息中名称相同的模型字段的内容进行相似度运算，并将各运算结果与对应的模型字段的权重值进行加权求和，得到模型字段的内容相似度。

[0016] 一种服务器，包括：

[0017] 数据采集模块，用于采集用户账号的识别信息，所述识别信息中包含设定的关键字段和模型字段；

[0018] 识别判断模块，用于接收所述数据采集模块发送的用户账号的识别信息，针对任意两个用户账号的识别信息，判断所述两个用户账号的识别信息中是否有至少一个关键字段的内容相同，若是，则确定该两个用户账号是由同一用户注册的，否则，根据第一相似度运算模块的运算结果判断所述两个用户账号的识别信息中的模型字段的内容相似度是否达到设定条件，若是，则确定该两个用户账号是由同一用户注册的，否则，确定该两个用户账号是由不同用户注册的；

[0019] 第一相似度运算模块，用于确定各模型字段的权重值，并分别将两个用户账号的识别信息中名称相同的模型字段的内容进行相似度运算，将各运算结果与对应的模型字段的权重值进行加权求和，得到模型字段的内容相似度并发送至识别判断模块。

[0020] 本申请的有益效果如下：

[0021] 本申请实施例提供了一种用户账号的识别方法及设备，通过采集各用户账号所对应的识别信息，所述识别信息中包含设定的关键字段和模型字段，当任意两个用户账号的识别信息中包含有至少一个相同的关键字段内容时，则确定该两个用户账号为同一用户所注册，否则，需要判断该两个用户账号的识别信息中模型字段的内容相似度，并根据所述模型字段的内容相似度评级来判断该两个用户账号是否为同一用户所注册，从而解决了无法有效地从大量的用户账号中识别出同一用户注册的用户账号的问题。

附图说明

[0022] 图 1 所示为本申请实施例一中用户账号的识别方法流程示意图；

[0023] 图 2 所示为本申请实施例二中服务器的结构示意图。

具体实施方式

[0024] 本申请实施例提供了一种用户账号的识别方法及设备，通过采集各用户账号所对应的识别信息，所述识别信息中包含设定的关键字段和模型字段，当任意两个用户账号的识别信息中包含有至少一个相同的关键字段内容时，则确定该两个用户账号为同一用户所注册，否则，需要判断该两个用户账号的识别信息中模型字段的内容相似度，并根据所述模型字段的内容相似度评级来判断该两个用户账号是否为同一用户所注册，从而解决了无法有效地从大量的用户账号中识别出同一用户注册的用户账号的问题。

[0025] 通过本申请的技术方案，可以实现对同一用户注册的各用户账号的集中管理、有效提高对用户账号的管理效率，进一步地，还可以有效地分析、跟踪用户行为，向用户注册的用户账号精准地推送有用信息。

[0026] 实施例一：

[0027] 如图 1 所示,为本申请实施例一中用户账号的识别方法流程示意图,所述识别方法包括以下步骤:

[0028] 步骤 101:服务器采集用户账号的识别信息,所述识别信息中包含设定的关键字段和模型字段。

[0029] 所述用户账号是指用户通过互联网网站注册的用于在注册网站进行相应网上业务行为的虚拟账号。所述用户可以是注册用户账号的个人、公司或者组织。

[0030] 所述用户账号的识别信息包括以下至少一种信息:

[0031] (1)、用户账号注册时的注册信息。

[0032] 所述注册信息可以是用户在注册用户账号时填写的信息,例如,针对个人用户,所述注册信息可以包括:姓名、身份证号码、性别、电子邮箱、教育水平等信息;针对企业用户,所述注册信息可以包括:企业名称、组织机构代码、企业类型、联系地址等信息。

[0033] (2)、从第三方获取的信息。

[0034] 所述第三方可以是指注册网站外的其他网站或是工商单位、民事单位、司法单位等政府、民间机构等,从所述第三方获取的信息可以是依据注册信息中的内容,从所述第三方获取与该用户相关的其他信息,例如,根据所述注册信息确定企业用户的企业名称后,可以从工商单位中获取该企业用户的注册资本、所属行业等信息。

[0035] (3)、用户账号运行时产生的信息,如用户账号登录时所使用的 IP 地址或者 MAC 信息等。

[0036] 采集的识别信息中的内容按其性质可以划分为两类,一类为关键字段中的内容,另一类为模型字段中的内容。

[0037] 所述关键字段表示能够唯一或是接近唯一地识别用户账号信息注册者的身份的识别信息,例如,针对个人用户,关键字段的名称可以为身份证号码、护照号码等,针对身份证号码这一关键字段,其内容可以为用户在注册用户账号时填写的身份证号码。

[0038] 所述模型字段为识别信息中的非关键字段,例如,针对个人用户,模型字段的名称可以为性别、出生日期、国籍、电子邮箱、教育水平,针对出生日期这一模型字段,其内容可以为用户在注册用户账号时填写的出生日期。

[0039] 需要说明的是,采集的识别信息中并不一定包含预先设定的关键字段或模型字段,也就是说,若将预设的关键字段或模型字段以数据表的形式设置,针对每次采集的识别信息,将识别信息中的内容填充至表 1 所示的数据表中时,可能存在某些关键字段或模型字段有内容,某些关键字段或模型字段的内容为空的情况。

[0040] 例如:如表 1 所示,为设置的数据表形式的用户账号 A 的识别信息,所述识别信息包括设定的关键字段和模型字段,若本次针对用户账号 A 采集的识别信息中包含了用户的姓名、身份证号码、性别、国籍,则表 1 中的名称为身份证号码的关键字段有内容,名称为护照号码的关键字段中内容为空;名称为姓名、性别、国籍的模型字段有内容,名称为出生日期、电子邮箱的模型字段内容为空。

[0041]

识别信息		用户账号标识	用户账号 A
关键字段	身份证号码		1101001980
	护照号码		
模型字段	姓名		张三
	国籍		中国
	电子邮箱		
	性别		男
	出生日期		

[0042] 表 1

[0043] 需要说明的是,由于识别信息是可以包含以上三种信息的(用户账号注册时的注册信息、从第三方获取的信息和用户账号运行时产生的信息),而这三种信息又可以从不同的途径采集获得,可能出现针对同一用户账号的相同名称的字段出现不同内容的情况。仍以表 1 所示的情况为例,若从用户账号 A 的注册信息中确定国籍这一模型字段的内容为中国,再从用户账号 A 登录时所使用的 IP 地址确定国籍这一模型字段的内容为日本,则针对同一用户账号的相同名称的字段出现不同内容的情况,本实施例包括但不限于以下两种做法:

[0044] 第一种做法:根据识别信息采集的来源不同,为不同来源的识别信息分配权重,也就是说,可以为注册信息、从第三方获取的信息和用户账号运行时产生的信息分配权重,当针对同一用户账号的相同名称的字段出现不同内容的情况时,确定不同内容所属来源的权重,保留权重较高的来源提供的内容。例如:若注册信息的权重最高,从第三方获取的信息的权重其次,用户账号运行时产生的信息的权重最低,则若从用户账号 A 的注册信息中确定国籍这一模型字段的内容为中国,再从用户账号 A 登录时所使用的 IP 地址确定国籍这一模型字段的内容为日本,可保留注册信息中确定的国籍的内容,丢弃用户账号 A 登录时所使用的 IP 地址确定的国籍的内容。

[0045] 第二种做法:将不同来源的针对同一字段不同的内容,都作为该字段的内容保留,例如:若从用户账号 A 的注册信息中确定国籍这一模型字段的内容为中国,再从用户账号 A 登录时所使用的 IP 地址确定国籍这一模型字段的内容为日本,可在表 1 国籍这一模型字段的内容中填写中国和日本。

[0046] 在采集到各用户账号的识别信息后,可将该识别信息以表 1 所示的数据表的形式存储至服务器的数据库中,较优地,可为存储的识别信息分配标识,通过该分配的标识不仅可以从数据库中查询出各用户账号的识别信息,还可以识别出该用户账号是个人用户注册的还是企业用户注册的。例如:针对表 1 所示的个人用户账号 A 的识别信息分配的标识可为 A_001,其中,A 表示对应的识别信息是个人用户账号的识别信息,001 为个人用户账号中的识别信息的序列号。再例如,针对某一企业用户账号的识别信息分配的标识可为 B_001,

其中, B 表示对应的识别信息是企业用户账号的识别信息, 001 为企业用户账号中的识别信息的序列号。

[0047] 另外, 为了使本申请中采集的识别信息能够方便地进行数据库存储并进行分析, 不论识别信息是注册信息、从第三方获取的信息还是用户账号运行时产生的信息, 采集的信息中关键字段和模型字段的指代需一致。例如: 针对表 1 中名称为国籍的模型字段, 不论从何处采集的识别信息, 针对国籍这一模型字段, 其内容都应当表示用户的国籍信息。

[0048] 步骤 102: 服务器针对任意两个未经识别的用户账号的识别信息, 判断所述两个用户账号的识别信息中是否有至少一个关键字段的内容相同, 若是, 则执行步骤 105; 否则, 执行步骤 103。

[0049] 较优地, 若在步骤 101 中, 用户账号的识别信息以数据表的形式存储在服务器的数据库中, 且为存储的识别信息分配了可区分个人用户账号和企业用户账号的标识, 则在本步骤 102 中, 可根据标识, 将同为个人用户账号的识别信息执行本步骤 102, 并将同为企业用户账号的识别信息执行本步骤 102。

[0050] 由于识别信息中关键字段的内容是表示能够唯一或是接近唯一地识别用户账号信息注册者的身份的内容, 因此, 只要两个识别信息中, 有至少一个关键字段的内容相同, 则可认为这两个识别信息所对应的用户账号是同一用户注册的。

[0051] 假设, 用户账号 A 的识别信息中, 身份证号码这一关键字段的内容与用户账号 B 的识别信息中的这一关键字段的内容相同, 则不论这两个识别信息的其他关键字段或是模型字段的内容有无相关性, 都可认为用户账号 A 和用户账号 B 是同一用户注册的。

[0052] 本步骤 102 也不限于对数据库中已存储的识别信息两两进行比较。

[0053] 步骤 103: 服务器判断所述两个用户账号的识别信息中的模型字段的内容相似度是否达到设定条件, 若是, 则执行步骤 105; 否则, 执行步骤 104。

[0054] 具体地, 服务器可以通过以下步骤来确定该两个用户账号的识别信息中的模型字段的内容相似度:

[0055] 步骤一: 首先需要确定所述两个用户账号的识别信息中各模型字段的权重值。

[0056] 模型字段的权重值可以根据该模型字段表示用户身份的重要性来确定, 对于表示用户身份的重要性越高的模型字段, 其分配的权重也就越高。

[0057] 较优地, 本实施例中可以采用设置模型算法的方式来确定识别信息中各模型字段的权重值, 具体可以采用层次分析模型的方式, 通过建立层次结构模型、构造成对比较阵、计算权向量并做一致性检验的步骤来得到各模型字段的权重值。采用构造层次分析模型的方式来确定各模型字段的权重值的方式, 可以较大程度地避免主观设定模型字段的权重值所带来的结果上的误差, 使得到的权重值的结果更为准确。

[0058] 步骤二: 分别将两个用户账号的识别信息中名称相同的模型字段的内容进行相似度运算, 并将各运算结果与对应的模型字段的权重值进行加权求和, 得到模型字段的内容相似度。

[0059] 本实施例中所涉及的相似度运算算法不限于目前可用的任一算法, 例如: 两个模型字段内容相同, 则确定其相似度为 1, 若两个模型字段内容不相同, 则确定其相似度为 0, 再例如, 还可以使用海明距离算法等。

[0060] 在确定多个名称相同的模型字段的内容之间的相似度后, 可将每个运算结果与对

应的模型字段的权重值进行加权求和,例如:针对国籍、电子邮箱、性别和出生日期这4个模型字段进行内容相似度判断,假设其权重分别为a、b、c和d,两个用户账号的识别信息中针对这4个模型字段的内容进行相似度运算后,其运算结果分别为X1、X2、X3和X4,则进行加权求和为: $aX1+bX2+cX3+dX4$,该加权求和的结果即为两个用户账号的识别信息中模型字段整体的内容相似度。

[0061] 所述设定条件可以根据对相似度判断的精度要求来设置,若对相似度判断的精度要求较高,则可将所述设定条件设置为较大的门限值,若对相似度判断的精度要求较低,则可将所述设定条件设置为较小的门限值。

[0062] 较优地,在本步骤103中计算两个用户账号的识别信息中模型字段整体的内容相似度时,还可以考虑识别信息来源的可靠性程度,对于来源比较可靠的信息,可为其分配较高的权重,利用识别信息来源的权重值计算模型字段整体的内容相似度的方式为:

[0063] 分别将两个用户账号的识别信息中名称相同的模型字段的内容进行相似度运算,并确定进行相似度运算的内容所属信息的权重值,然后将各运算结果与进行相似度运算的内容所属信息的权重值进行加权求和,得到模型字段的内容相似度。

[0064] 例如:针对国籍、电子邮箱、性别和出生日期这4个模型字段进行内容相似度判断,国籍信息、性别和出生日期是通过注册信息获取的,电子邮箱是通过第三方获取的,假设注册信息的权重值为A,从第三方获取信息的权重值为B,两个用户账号的识别信息中针对这4个模型字段的内容进行相似度运算后,其运算结果分别为X1、X2、X3和X4,则进行加权求和为: $AX1+BX2+AX3+AX4$,该加权求和的结果即为两个用户账号的识别信息中模型字段整体的内容相似度。

[0065] 若同时考虑表示用户身份的重要性来为模型字段分配的权重和信息来源的权重,则可将相同名称的模型字段进行内容相似度运算后,将运算结果分别与对应的模型字段的权重、信息来源的权重进行加权求和,得到模型字段的内容相似度。

[0066] 仍以针对国籍、电子邮箱、性别和出生日期这4个模型字段进行内容相似度判断为例,假设模型字段的权重分别为a、b、c和d,国籍信息、性别和出生日期是通过注册信息获取的,电子邮箱是通过第三方获取的,假设注册信息的权重值为A,从第三方获取信息的权重值为B,两个用户账号的识别信息中针对这4个模型字段的内容进行相似度运算后,其运算结果分别为X1、X2、X3和X4,则进行加权求和为: $AaX1+BbX2+AcX3+AdX4$,该加权求和的结果即为两个用户账号的识别信息中模型字段整体的内容相似度。

[0067] 步骤104:确定所述两个用户账号是由不同用户注册的,并跳转至步骤106。

[0068] 较优地,为了保证判断结果的可靠性,还可以对步骤104的结果进一步采用人工干预的方式再次进行判定,本实施例的方案也不限于:若人工干预的方式再次判定的结果与步骤104的确定结果不同时,以人工干预的结果为准。

[0069] 步骤105:确定所述两个用户账号是由同一用户注册的,并跳转至步骤106。

[0070] 与步骤104类似的,还可以对本步骤105的确定结果采用人工干预的方式再次进行判定。

[0071] 在本发明实施例的方案中,若在步骤104或步骤105后还进行过人工干预操作,则可以通过日志信息的方式记录人工干预的操作过程以及操作结果。

[0072] 步骤106:得到两个用户账号是否是同一用户注册的判断结果,并判断是否还存

在两个用户账号没有进行识别,若是,则跳转至步骤 102,否则,结束本次用户账号的识别过程。

[0073] 通过本发明实施例一的方案,可以识别出在同一网站或不同网站内的多个用户账号是否是由同一用户注册的,对于由同一用户注册的多个用户账号,可将该多个用户账号进行绑定,并向所述多个用户账号所在的网站服务器提供通知机制和结果查询机制,以便于网站服务器对同一用户注册的用户账号集中管理,有效提高对用户账号的管理效率。

[0074] 较佳地,通过对同一用户所注册多个用户账号的分析,可有效地跟踪该用户的业务行为、综合分析该用户的业务发展趋势,进而可以向该用户所注册的多个用户账号的任一用户账号精准地推送有用信息,不仅降低了网站服务器向该用户所注册的多个用户账号都推送信息所占用的网络资源和管理成本,而且使得信息的推送更有目的性和针对性,便于网站服务一方进行准确营销。

[0075] 较佳地,在识别出多个用户账号为同一用户所注册时,还可以将所述多个用户账号划分为主、次用户账号(如可以根据用户登录用户账号的次数来判定,登录次数比较频繁的为主用户账号、相对比较少登录的为次用户账号),并将该用户的次用户账号按照离线时间长短进行计时,当计时达到设定值时,注销该用户的次用户账号,使得网站服务器的存储空间得到一定的降低,从而一定程度地降低相应互联网网站的运行成本。

[0076] 较佳地,在识别出多个用户账号为同一用户所注册时,若确定其中有一个用户账号具有一定风险性时,则可将其他绑定的由同一用户注册的用户账号都识别为具有风险性的用户账号;另外,若识别出多个用户账号为同一用户短时期内在同一网站注册时,且所述多个用户账号的数量较大,可以认为是用户在网站的恶意注册。因此,通过本发明实施例的方案,还可以避免用户账号的安全风险,降低其所带来的网络安全隐患,提高互联网网站的安全性。

[0077] 实施例二:

[0078] 如图 2 所示,为与本发明实施例一属于同一发明构思下的服务器的结构示意图,所述服务器包括数据采集模块 11、识别判断模块 12 和第一相似度运算模块 13。

[0079] 所述数据采集模块 11 用于采集用户账号的识别信息,所述识别信息中包含设定的关键字段和模型字段。

[0080] 具体地,所述数据采集模块 11 采集到的用户账号的识别信息包括以下至少一种信息:用户账号注册时的注册信息、从第三方获取的信息、用户账号运行时产生的信息;所述数据采集模块 11 采集的识别信息中的关键字段表示能够唯一或是接近唯一地识别用户账号信息注册者的身份的识别信息,所述模型字段则为识别信息中的非关键字段。

[0081] 需要说明的是,为了使数据采集模块 11 所采集的识别信息能够方便地进行数据库存储并进行分析,不论识别信息是注册信息、从第三方获取的信息还是用户账号运行时产生的信息,采集的信息中关键字段和模型字段的指代需一致。

[0082] 所述识别判断模块 12 用于接收所述数据采集模块 11 发送的用户账号的识别信息,针对任意两个用户账号的识别信息,判断所述两个用户账号的识别信息中是否有至少一个关键字段的内容相同,若是,则确定该两个用户账号是由同一用户注册的,否则,判断所述两个用户账号的识别信息中的模型字段的内容相似度是否达到设定条件,若是,则确定该两个用户账号是由同一用户注册的,否则,确定该两个用户账号是由不同用户注册的。

[0083] 所述第一相似度运算模块 13 具体用于确定各模型字段的权重值,并分别将两个用户账号的识别信息中名称相同的模型字段的内容进行相似度运算,将各运算结果与对应的模型字段的权重值进行加权求和,得到模型字段的内容相似度并发送至识别判断模块 12。

[0084] 具体地,所述服务器还可以包括第二相似度运算模块 14。

[0085] 所述第二相似度运算模块 14 用于确定识别信息中的注册信息、从第三方获取的信息和用户账号运行时产生的信息的权重值,分别将两个用户账号的识别信息中名称相同的模型字段的内容进行相似度运算,并确定进行相似度运算的内容所属信息的权重值,并将各运算结果与进行相似度运算的内容所属信息的权重值进行加权求和,得到模型字段的内容相似度并发送至识别判断模块 12。

[0086] 所述识别判断模块 12 具体还用于根据所述第一相似度运算模块 13 或第二相似度运算模块 14 得到的内容相似度,判断所述两个用户账号的识别信息中的模型字段的内容相似度是否达到设定条件。

[0087] 较优地,所述服务器还可以包括数据存储模块 15 和人工识别模块 16。

[0088] 所述数据存储模块 15 用于存储数据采集模块 11 所采集到的各用户账号的识别信息;所述人工识别模块 16 用于提供操作界面来接收人工干预信息,利用接收到的人工干预信息对识别判断模块 12 所得到的识别结果再次进行判定,若人工干预的方式再次判定的结果与识别判断模块 12 的确定结果不一致时,以人工干预的结果为准。

[0089] 需要说明的是,由于数据采集模块 11 所采集的识别信息中并不一定包含预先设定的关键字段或模型字段,也就是说,若将预设的关键字段和模型字段以数据表的形式或其他形式存储在数据存储模块 15 中时,针对每次采集的识别信息,可能存在某些关键字段或模型字段有内容,某些关键字段或模型字段的内容为空的情况;另外,需要说明的是,数据存储模块 15 中以数据表的形式或其他形式存储的所预设的关键字段和模型字段的数目也应根据每次采集到的用户账号的识别信息而动态变化,较优地,可以根据采集到的识别信息增加所预设的关键字段数目或是模型字段的数目。

[0090] 需要说明的是,由于数据采集模块 11 所采集的识别信息是可以包含三种信息的(用户账号注册时的注册信息、从第三方获取的信息和用户账号运行时产生的信息),而这三种信息又可以从不同的途径采集获得,可能出现针对同一用户账号的相同名称的字段出现不同内容的情况。则针对同一用户账号的相同名称的字段出现不同内容的情况,数据存储模块 15 存储该用户账号的识别信息的方式可以包括但不限于以下两种做法:

[0091] 第一种做法:数据存储模块 15 根据识别信息采集的来源不同,为不同来源的识别信息分配权重,当针对同一用户账号的相同名称的字段出现不同内容的情况时,确定不同内容所属来源的权重,保留权重较高的来源提供的内容,并丢弃其他权重较低的来源所提供的内容。

[0092] 第二种做法:数据存储模块 15 将不同来源的针对同一字段不同的内容,都作为该字段的内容保留。

[0093] 较优地,所述数据存储模块 15 还可以为存储的识别信息分配唯一标识,通过该分配的标识不仅可以从数据库中查询出各用户账号的识别信息,还可以识别出该用户账号是个人用户注册的还是企业用户注册的。

[0094] 较优地,所述数据存储模块 15 还用于将由同一用户注册的多个用户账号进行绑定,并存储绑定信息,还可以采用日志信息的方式记录人工识别模块 16 对本地存储的识别信息判断结果进行的相应操作结果和操作过程。

[0095] 本申请实施例提供了一种用户账号的识别方法及设备,通过采集多个用户账号的识别信息,所述识别信息中包含设定的关键字段和模型字段,当任意两个用户账号的识别信息中包含有至少一个相同的关键字段内容时,则确定该两个用户账号为同一用户所注册,否则,计算所述两个用户账号的模型字段的内容相似度并根据得到的内容相似度结果来判断该两个用户账号是否为同一用户所注册,从而解决了无法有效地识别出多个用户账号是否为同一用户注册的问题,进一步地,利用本申请的技术方案可以实现对同一用户注册的各用户账号的集中管理、有效提高对用户账号的管理效率,还可以有效地分析、跟踪用户行为,向用户注册的用户账号精准地推送有用信息。

[0096] 显然,本领域的技术人员可以对本申请进行各种改动和变型而不脱离本申请的精神和范围。这样,倘若本申请的这些修改和变型属于本申请权利要求及其等同技术的范围之内,则本申请也意图包含这些改动和变型在内。

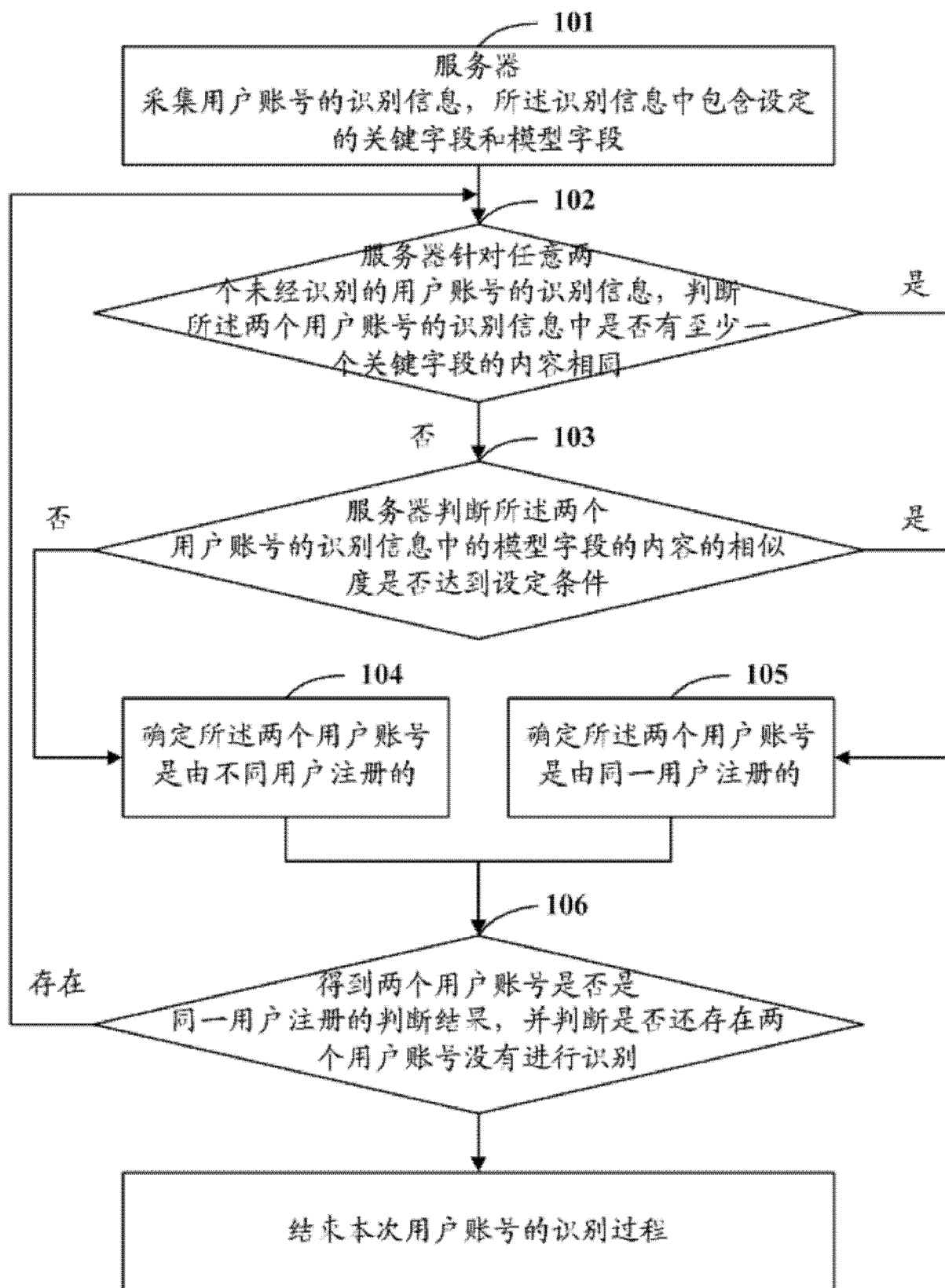


图 1

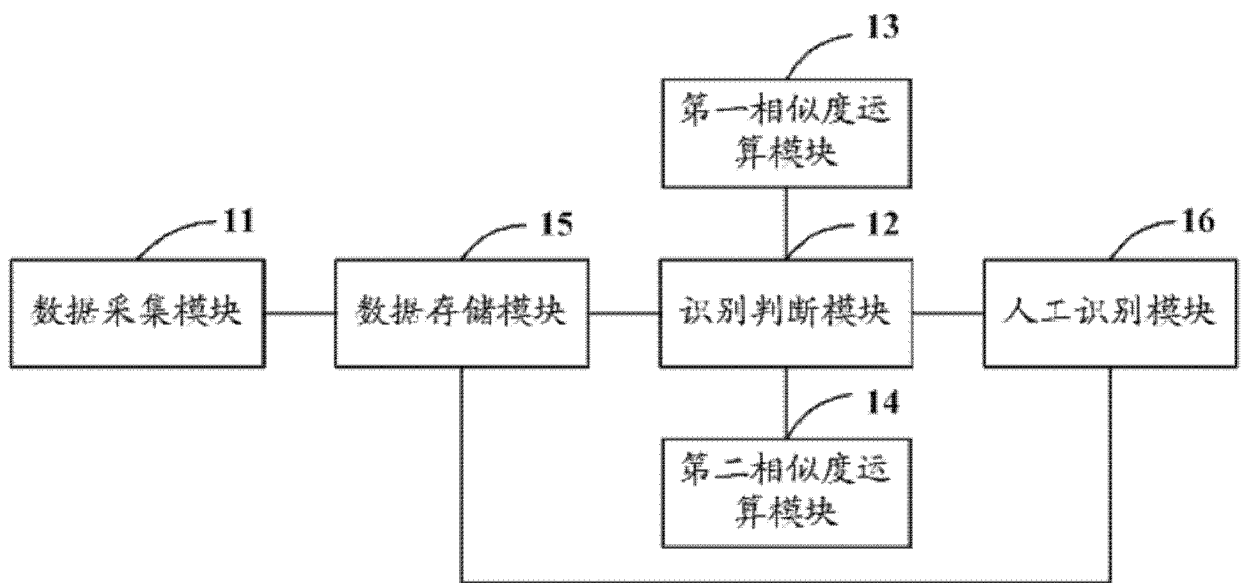


图 2