(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 110110218 A
(43)申请公布日 2019.08.09

(21)申请号 201810105358.9

(22)申请日 2018.02.01

(71)申请人 重庆邮电大学
地址 400065 重庆市南岸区崇文路2号重庆邮电大学

(72)发明人 陈龙 李葱

(74)专利代理机构 广州三环专利商标代理有限公司 44202
代理人 郝传鑫 熊永强

(51)Int.Cl.
G06F 16/9535(2019.01)
G06Q 50/00(2012.01)

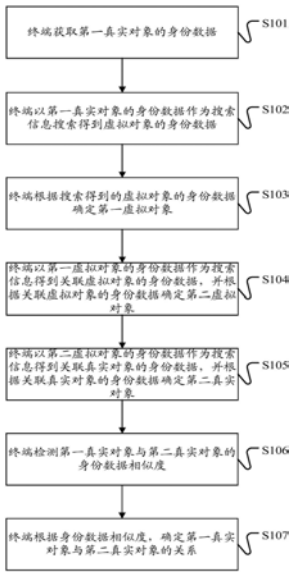
权利要求书2页 说明书18页 附图4页

(54)发明名称

一种身份关联方法及终端

(57)摘要

本发明实施例公开了一种身份关联方法和终端，其中方法包括，获取第一真实对象的身份数据，找到第一真实对象在社交网络中对应的第一虚拟对象，以第一虚拟对象的身份数据为搜索信息确定第二虚拟对象，找到第二虚拟对象对应的第二真实对象，并检测第一真实对象和第二真实对象的相似度，进而确定第一真实对象和第二真实对象的关系。本发明实施例还公开了一种身份数据相似度计算方法，具体包括获取真实对象和虚拟对象的身份数据，构建身份数据模型，计算身份数据相似度，进而判断真实对象与虚拟对象之间，不同虚拟对象之间，不同真实对象之间的身份关联性。通过实施上述方法，可以自动地计算对象之间的身份数据相似度并判断对象之间的身份关联性。



CN 110110218 A

1.一种身份关联方法,其特征在于,包括:

获取第一真实对象的身份数据;

以所述第一真实对象的身份数据作为搜索信息搜索得到虚拟对象的身份数据,并根据所述搜索得到的虚拟对象的身份数据确定第一虚拟对象;

以所述第一虚拟对象的身份数据作为搜索信息得到关联虚拟对象的身份数据,并根据所述关联虚拟对象的身份数据确定第二虚拟对象。

2.根据权利要求1所述的方法,其特征在于,所述根据所述关联虚拟对象的身份数据确定第二虚拟对象之后,还包括:

以所述第二虚拟对象的身份数据作为搜索信息得到关联真实对象的身份数据,并根据所述关联真实对象的身份数据确定第二真实对象。

3.根据权利要求1所述的方法,其特征在于,所述根据所述关联真实对象的身份数据确定第二真实对象之后,还包括:

检测所述第一真实对象与所述第二真实对象的身份数据相似度;

根据所述身份数据相似度,确定所述第一真实对象与所述第二真实对象的关系。

4.根据权利要求1任一项所述的方法,其特征在于,所述根据所述搜索得到的虚拟对象的身份数据确定第一虚拟对象,包括:

分别计算各个虚拟对象的背景数据与所述第一真实对象的背景数据相似度;

将所述虚拟对象按所述背景数据相似度从高到低的顺序降序排列;

计算排序为前n位的虚拟对象与所述第一真实对象的兴趣数据相似度;

计算排序为前n位的虚拟对象与所述第一真实对象的社交数据相似度;

将所述排序为前n位的虚拟对象与所述第一真实对象的背景数据相似度、兴趣数据相似度和社交数据相似度进行加权汇总,得到所述排序为前n位的虚拟对象与所述第一真实对象的身份数据相似度;

将所述排序为前n位的虚拟对象按所述身份数据相似度从高到低的顺序降序排列;

将排序为第一位的虚拟对象确定为所述第一虚拟对象,其中,n为大于1的整数。

5.根据权利要求4所述的方法,其特征在于,所述分别计算各个虚拟对象的背景数据与所述第一真实对象的背景数据相似度,包括:

提取背景数据中的字符串数据和数字数据,所述字符串数据包括姓名数据和地址数据,所述数字数据包括生日数据和性别数据;

分别计算所述第一真实对象和虚拟对象的姓名数据相似度和地址数据相似度;

将所述姓名数据相似度和所述地址数据相似度进行加权汇总得到所述第一真实对象和所述虚拟对象的字符串数据相似度;

分别计算所述第一真实对象和所述虚拟对象的生日数据相似度和性别数据相似度;

将所述生日数据相似度和所述性别数据相似度进行加权汇总得到所述第一真实对象和所述虚拟对象的数字数据相似度;

将所述字符串数据相似度和所述数字数据相似度进行加权汇总得到所述第一真实对象与所述虚拟对象的背景数据相似度。

6.根据权利要求4所述的方法,其特征在于,所述计算排序为前n位的虚拟对象与所述第一真实对象的兴趣数据相似度,包括:

提取所述第一真实对象和虚拟对象的兴趣数据中的兴趣关键词；

获取所述兴趣关键词的在所述兴趣数据中的权重；

将所述兴趣数据采用空间向量模型表示为兴趣向量，其第k个向量的值为第k个关键词对应的权重，其中，k为大于或等于1的整数；

计算所述第一真实对象的兴趣向量与所述虚拟对象的兴趣向量的余弦值；

将所述余弦值作为所述第一真实对象和所述虚拟对象的兴趣数据相似度。

7.一种身份关联方法，其特征在于，包括：

获取第一虚拟对象的身份数据；

以所述第一虚拟对象的身份数据作为搜索信息搜索得到关联虚拟对象的身份数据，并根据所述关联虚拟对象的身份数据确定第二虚拟对象；

以所述第二虚拟对象的身份数据作为搜索信息得到关联真实对象的身份数据，并根据所述关联真实对象的身份数据确定第二真实对象。

8.根据权利要求4所述的方法，其特征在于，所述根据所述关联真实对象的身份数据确定第二真实对象之后，还包括：

获取第一虚拟对象对应的第一真实对象的身份数据；

检测所述第一真实对象与所述第二真实对象的身份数据相似度；

根据所述身份数据相似度，确定所述第一真实对象与所述第二真实对象的关系。

9.一种终端，其特征在于，包括处理器、输入设备、输出设备和存储器，所述处理器、输入设备、输出设备和存储器相互连接，其中，所述存储器用于存储计算机程序，所述计算机程序包括程序指令，所述处理器被配置用于调用所述程序指令，执行如权利要求1-8任一项所述的方法。

10.一种计算机可读存储介质，其特征在于，所述计算机存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处理器执行时使所述处理器执行如权利要求1-8任一项所述的方法。

一种身份关联方法及终端

技术领域

[0001] 本发明涉及计算机领域，尤其涉及一种身份关联方法及终端。

背景技术

[0002] 随着计算机技术和网络技术的发展，社交网络已经融入了人们的日常生活，几乎每个人在社交网络中都有自己的虚拟身份，人们通过社交网络上的虚拟身 份进行沟通，极大的方便了人们日常的交流，在社交网络上也会留下大量的可 以在一定程度上表明用户身份的身份数据。

[0003] 目前在对某些违法违纪的嫌疑人进行追踪时，除了通过办案人员实地考察 追踪以外，还可以借助社交网络等网络上的虚拟身份数据，来查找识别嫌疑人。在网络上查找嫌疑人的过程中，主要是通过人工查找虚拟身份数据并进行分析，现有的查找确定方式费时费力，效率低下。

发明内容

[0004] 本发明实施例提供了一种身份关联方法和终端，可以自动地计算对象之间 的身份数据相似度并判断对象之间的身份关联性。

[0005] 为了解决上述技术问题，本发明实施例第一方面公开了一种身份关联方法，所述方法包括：

[0006] 获取第一真实对象的身份数据；

[0007] 以所述第一真实对象的身份数据作为搜索信息搜索得到虚拟对象的身份数 据，并根据所述搜索得到的虚拟对象的身份数据确定第一虚拟对象；

[0008] 以所述第一虚拟对象的身份数据作为搜索信息得到关联虚拟对象的身份数 据，并根据所述关联虚拟对象的身份数据确定第二虚拟对象。

[0009] 本发明实施例第二方面公开了一种终端，所述终端包括：

[0010] 获取模块，用于获取第一真实对象的身份数据；

[0011] 搜索模块，用于以所述第一真实对象的身份数据作为搜索信息搜索得到虚 拟对象的身份数据，并根据所述搜索得到的虚拟对象的身份数据确定第一虚拟 对象；

[0012] 所述搜索模块，还用于以所述第一虚拟对象的身份数据作为搜索信息得到 关联虚拟对象的身份数据，并根据所述关联虚拟对象的身份数据确定第二虚拟 对象。

[0013] 本发明实施例第三方面公开了一种终端，所述终端包括处理器、输入设备、输出设备和存储器，所述处理器、输入设备、输出设备和存储器相互连接，其 中，所述存储器用于存储计算机程序，所述计算机程序包括程序指令，所述处 理器被配置用于调用所述程序指令，执行所述身份关联的方法。

[0014] 本发明实施例第四方面公开了一种计算机可读存储介质，所述计算机存储 介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处 理器执行时使所述处理器执行所述身份关联的方法。

[0015] 本发明实施例中，终端获取第一真实对象的身份数据，终端以第一真实对 象的身份数据为搜索信息搜索得到虚拟对象的身份数据，并根据搜索得到的虚 拟对象的身份数据确定第一虚拟对象；终端以第一虚拟对象的身份数据为搜索 信息得到关联虚拟对象的身份数据，并确定第二虚拟对象；终端以第二虚拟对 象的身份数据为搜索信息得到关联真实对象的身份数据，并确定第二真实对象； 终端检测第一真实对象与第二真实对象的身份数据相似度；根据身份数据相似 度大小确定第一真实对象与第二真实对象的关系。通过实施上述方法，可以找 出真实对象在社交网络中对应的虚拟对象，以及与真实对象具有亲密关系的其 他真实对象。

附图说明

[0016] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施 例中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是 本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的 前提下，还可以根据这些附图获得其他的附图。

[0017] 图1为本发明实施例提供的一种身份关联方法的流程示意图；

[0018] 图2为本发明实施例提供的另一种身份关联方法的流程示意图；

[0019] 图3为本发明实施例提供的一种身份数据相似度检测方法的流程示意图；

[0020] 图4为本发明实施例提供的一种终端的结构示意图；

[0021] 图5为本发明实施例提供的另一种终端的结构示意图。

具体实施方式

[0022] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清 楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是 全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造 性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0023] 请参见图1，为本发明实施例提供的一种身份关联方法的流程示意图，该方 法可包括：

[0024] S101、终端获取第一真实对象的身份数据。

[0025] 本发明实施例中，该身份数据包括背景数据、兴趣数据和社交数据，其中， 背景数据为记录对象的身份背景的数据；兴趣数据为记录对象的兴趣爱好的数 据；社交数据为记录对象的人际关系的数据。背景数据可以为对象的姓名、性 别、地址、出生日期、身份证号、电话号、教育/工作经历等，兴趣数据可以为 对象喜爱的体育、军事、动漫、明星、宠物、美食等，社交数据可以为对象的 情侣/配偶、父母/子女、兄弟、朋友等。对象可以为现实中一个实际的人物，也 可以是社交网络上的一个虚拟身份，如各类应用软件的账号等，第一真实对象 可以为现实中一个实际的人物。

[0026] S102、终端以第一真实对象的身份数据作为搜索信息搜索得到虚拟对象的 身份数据。

[0027] 本发明实施例中，终端获取到该第一真实对象的身份数据之后，会以第一 真实对象的身份数据作为搜索信息在社交网络上搜索得到虚拟对象的身份数据。搜索得到的虚

拟对象与该第一真实对象有一定的共同点。

[0028] 例如,该共同点为相同的名字,该第一真实对象的名字为小明,则可以在 社交网络上寻找用户名为小明的虚拟对象,或者好友中包含小明的虚拟对象; 或者,该共同点为相同的爱好,该第一真实对象的爱好为足球,则可以在社交 网络上寻找兴趣爱好为足球的虚拟对象,或者,该共同点为出生地址、所在学 校、工作单位等,共同点可以是一个,也可以是多个,对于具体的共同点,本 发明实施例不做限定。

[0029] S103、终端根据搜索得到的虚拟对象的身份数据确定第一虚拟对象。

[0030] 本发明实施例中,终端以第一真实对象的身份数据为搜索信息搜索得到虚 拟对象的身份数据之后,可以计算第一真实对象与虚拟对象之间的身份数据相 似度,并确定该第一真实对象在社交网络上对应的第一虚拟对象。

[0031] 可选的,该第一虚拟对象为与该第一真实对象身份数据相似度最高的虚拟 对象,为了减少计算第一真实对象与虚拟对象的相似度时的运算量,本发明实 施例首先计算社交网络上的多个虚拟对象的背景数据与第一真实对象的背景数 据相似度,将各个虚拟对象按背景数据相似度从高到低的顺序降序排列。

[0032] 可选的,在计算背景数据相似度时,具体计算的数据可包括背景数据中的 字符串数据和数字数据,其中,字符串数据包括对象的姓名和地址,数字数据 包括对象的生日和性别。第一真实对象和虚拟对象的背景数据相似度的计算过 程可参见步骤S303。

[0033] 可选的,在计算第一真实对象和虚拟对象的背景数据相似度之后,将该多 个虚拟对象按背景数据相似度从高到低的顺序降序排列;然后从中提取出排名 前n位的虚拟对象,并计算排序为前n位的虚拟对象与所述第一真实对象的兴 趣数据和社交数据相似度。其中,n为大于等于1的整数,其具体可由研发人员 设定,本发明实施例不做限定。

[0034] 可选的,在计算第一真实对象与虚拟对象的兴趣数据相似度时,根据预设 的关键词与权重的对应关系确定第一真实对象的兴趣数据中的兴趣关键词对应 的第一权重;获取虚拟对象在社交网路上发布的文本数据;检测兴趣关键词在 文本数据中出现的频率;根据预设的频率与权重的对应关系确定虚拟对象的兴 趣关键词对应的第二权重。将兴趣数据采用空间向量模型表示为兴趣向量,其 第k个向量的值为第k个关键词对应的权重,计算第一真实对象的兴趣向量与 虚拟对象的兴趣向量的余弦值,将得到的余弦值作为所述第一真实对象和所述 虚拟对象的兴趣数据相似度,其中,k为大于等于1的整数。可选的,第一真实 对象和虚拟对象的兴趣数据相似度的计算过程可参见步骤S304。

[0035] 可选的,在计算第一真实对象与虚拟对象的社交数据相似度时,获取与第 一真实对象有关联的第三真实对象集合和与虚拟对象有关联的第三虚拟对象集 合;检测第三真实对象集合与第三虚拟对象集合的交集的人数,将交集的人数 与所述第三虚拟对象集合中的人数的比值作为虚拟对象与第一真实对象的社交 数据相似度。其中,第三真实对象集合包括第一真实对象的亲人、朋友、同事 等;第三虚拟对象集合包括第一虚拟对象在社交网络中的好友。例如,第三真 实对象集合中包括真实对象A、真实对象B和真实对象C,第三虚拟对象集合 中包括真实对象A创建的虚拟身份A、真实对象B创建的虚拟身份B和真实对象D创建的虚拟身份D,则第三真实对象集合与第三虚拟对象集合的交集的人 数为2个,第三虚拟对象集合中人数为3个,则第一真实对象与虚拟对象的社 交数据相似度为2/3,可选的,第一真实对象与虚拟对象的社交数据相似度计算 过程可参见步骤S305。

[0036] 在获取到第一真实对象和虚拟对象的背景数据相似度，兴趣数据相似度和 社交数据相似度之后，将排序为前n位的虚拟对象与第一真实对象的背景数据 相似度、兴趣数据相似度和社交数据相似度进行加权汇总，得到排序为前n位 的虚拟对象与第一真实对象的身份数据相似度；将排序为前n位的虚拟对象按 身份数据相似度从高到低的顺序降序排列，并将排序为第一位的虚拟对象确定 为第一虚拟对象(即该第一真实对象在社交网络上的虚拟身份)。可选的，也可 以再次提取排名前m位的虚拟对象，m为小于等于n的正整数，再对该前m位 的虚拟对象进行其他数据分析对比，找到该第一真实对象对应的第一虚拟对象， 其他数据可以是对象的字符串风格、情感倾向、定位数据、设备持有情况、文 化水平和计算机操作能力等。

[0037] S104、终端以第一虚拟对象的身份数据作为搜索信息得到关联虚拟对象的 身份数据，并根据关联虚拟对象的身份数据确定第二虚拟对象。

[0038] 本发明实施例中，终端确定第一真实对象在社交网络上的虚拟身份(即第 一虚拟对象)之后，分析第一虚拟对象的身份数据，以第一虚拟对象的身份数 据作为搜索信息得到关联虚拟对象的身份数据，并根据关联虚拟对象的身份数 据确定第二虚拟对象，可选的，第二虚拟对象为与该第一虚拟对象相似度最高 的虚拟对象。其中，第二虚拟对象与第一虚拟对象的相似度计算方法也是通过 计算虚拟对象之间的背景数据相似度、兴趣数据相似度和社交数据相似度得到， 其具体过程与S103类似，本发明实施例不在赘述。

[0039] S105、终端以第二虚拟对象的身份数据作为搜索信息得到关联真实对象的 身份数据，并根据关联真实对象的身份数据确定第二真实对象。

[0040] 本发明实施例中，终端在社交网络上找到第二虚拟对象的身份数据之后， 终端以第二虚拟对象的身份数据作为搜索信息得到关联真实对象的身份数据， 并根据关联真实对象的身份数据确定第二真实对象，可选的，第二真实对象为 与该第二虚拟对象相似度最高的虚拟对象，或者，第二真实对象为与第二虚拟 对象相似度大于预设阈值的真实对象，其相似度计算的具体过程与S103类似， 本发明实施例不在赘述。

[0041] S106、终端检测第一真实对象与第二真实对象的身份数据相似度；

[0042] 本发明实施例中，终端确定第一真实对象和第二真实对象之后，可以继续 计算第一真实对象和第二真实对象之间的相似度。其相似度计算的具体过程与 S103类似，本发明实施例不在赘述。

[0043] S107、终端根据身份数据相似度，确定第一真实对象与第二真实对象的关系。

[0044] 本发明实施例中，若第一真实对象与第二真实对象的相似度大于第一预设 阈值，则确定第一真实对象与第二真实对象为同一个人；其中第一预设阈值可 以为95％、99％等，具体可由研发人员具体设定，本发明实施例不做限定。

[0045] 若第一真实对象与第二真实对象的相似度介于第一预设阈值和第二预设阈 值之间，则确定所述第二真实对象与所述第一真实对象为亲密关系；其中，亲 密关系可以是父母、兄弟等。若第一真实对象为犯罪嫌疑人，则该第一真实对 象的亲密关系也可以为共犯等。其中，第一预设阈值大于第二预设阈值，第二 预设阈值可以为75％、80％等，本发明实施例不做限定。

[0046] 当第一真实对象与第二真实对象的相似度大小小于第二预设阈值时，则可 以将第一真实对象与第二真实对象的关系设置为待定关系，并通过其他数据(如 对象的字符串

风格、情感倾向、定位数据等)综合考量第一真实对象与第二真 实对象的关系。

[0047] 需要说明的是,本发明实施例只是举例,在其他可能的情况中,终端可以 直接获取社交网络中第一虚拟对象的身份数据,然后通过图3所示的相似度计 算方法找到该第一虚拟对象对应的第一真实对象,进一步的,终端获取到第一 真实对象之后,可以找到与该第一真实对象有关联的第二真实对象。并通过图3 所示的相似度计算方法找到该第二真实对象对应的第二虚拟对象。并判断第一 虚拟对象和第二虚拟对象在社交网络中的关系。

[0048] 举例说明,将第一真实对象表示为A,第一虚拟对象表示为A',第二虚拟 对象表示为B',第二真实对象表示为A。可选的,终端获取A的身份数据,分 析找出相似度最高的A'。可选的,终端获取A'的身份数据,在社交网络中分析 找出相似度最高的B'。可选的,终端获取B'的身份数据,分析找出与其相似度 最高的B。可选的,终端检测A和B之间的相似度,判断A和B之间的关系。

[0049] 在一种可能实现的方式中,终端获取A的身份数据,分析找出相似度最高 的A',并分析A'的身份数据,找出相似度最高的B'。

[0050] 在一种可能实现的方式中,终端获取A的身份数据,分析找出相似度最高 的A',并分析A'的身份数据,找出相似度最高的B'之后,分析B'的身份数据, 找出相似度最高的B。

[0051] 在一种可能实现的方式中,终端获取A的身份数据,分析找出相似度最高 的A',并分析A'的身份数据,找出相似度最高的B',分析B'的身份数据,找出 相似度最高的B之后,检测A和B的相似度,并根据相似度的大小判定A和B 之间的关系。可选的,若A和B的相似度大于第一预设阈值,则确定A和B为 同一个人;若A和B的相似度介于第一预设阈值和第二预设阈值之间,则确定 A和B为亲密关系;可选的,若A和B的相似度小于第二预设阈值,则交由人 工判断A和B之间的关系;其中,所述第一预设阈值大于所述第二预设阈值。

[0052] 在一种可能的实现方式中,终端获取A'的身份数据,分析A'的身份数据, 找出相似度最高的B',分析B'的身份数据,找出相似度最高的B。

[0053] 在一种可能的实现方式中,终端获取A'的身份数据,分析A'的身份数据, 找出相似度最高的B',分析B'的身份数据,找出相似度最高的B之后,获取A' 对应的A的身份信息,检测A和B的相似度,并根据相似度的大小判定A和B 之间的关系。

[0054] 在一种可能实现的方式中,终端获取B'的身份数据,分析B'的身份数据, 找出相似度最高的B之后,获取A的身份数据,检测A和B的相似度,并根据 相似度的大小判定A和B之间的关系。

[0055] 本发明实施例中,在已知第一真实对象身份信息的情况下,在社交网络上 通过相似度检测算法分析找出与第一真实对象相似度最高的第一虚拟对象,在 取证过程中,可以直接通过控制犯罪嫌疑人(即第一真实对象)的设备来获取 到第一真实对象在社交网络上的虚拟身份(即第一虚拟对象),对第一虚拟对象 的身份数据进行分析,找出与第一虚拟对象相似度最高的第二虚拟对象,在取 证过程中,有可能推断出第二虚拟对象是第一虚拟对象的共犯,或者第二虚拟 对象和第一虚拟对象对应同一个真实对象。然后对第二虚拟对象的身份数据进 行分析,找出与第二虚拟对象相似度最高的第二真实对象,则第二真实对象可 能与第一真实对象为同一个人,或者第二真实对象与第一真实对象为亲密关系, 在取证过程中,若已经获得了目标嫌疑人的虚拟身份信息,和若干嫌疑人,则 可以通过分析判断出谁是真正的嫌疑人。本发明实施例可以运用于以下两种场 景,如已知目标嫌疑人的真

实身份信息，从大量虚拟身份数据中，找到目标嫌 疑人的虚拟身份，或者，已经控制目标嫌疑人，从其设备获取到其虚拟身份信 息，从而找出该目标嫌疑人再社交网络中可能存在的其他虚拟身份和现实中的 共犯。

[0056] 请参见图2，为本发明实施例提供的另一种身份关联方法的流程示意图，该 方法可包括：

[0057] S201、终端获取真实对象和虚拟对象的身份数据。

[0058] 本发明实施例中，真实对象和虚拟对象的身份数据包括背景数据、兴趣数 据和社交数据。

[0059] 可选的，终端以显式的方式与对象进行交互获取对象的身份数据，例如， 终端首先提出一些关于身份数据的初始问题，根据对象的回答确定对象的身份 数据。可选的，根据对象答案的不同，终端提供给对象的问题也不同，例如， 当问到对象的年龄时，根据对象输入的结果确定对象年龄对应的问题组，以更 为准确的针对不同人群获取更精细的身份数据。

[0060] 可选的，终端以隐式方式获取对象的身份数据，具体的，终端在社交网络 上跟踪、分析、挖掘一些对象的身份数据。具体的，通过对象的网络使用数据 进行挖掘。或者，根据对象点击流数据进行分析、挖掘，对于点击流数据的分 析，将对象、查询及点击的网页作为一组数据来考虑，并对该组数据进行潜在 语义分析和概率潜在语义分析。或者，通过对对象查询历史或浏览历史进行分 析处理，通过反馈建立对象兴趣评价，获取用户的身份数据。

[0061] S202、终端根据身份数据的类别与相似度检测算法的对应关系，从预设的 多种相似度检测算法中筛选出与所述真实对象和虚拟对象的身份数据相对应的 目标相似度检测算法。

[0062] 本发明实施例中，身份数据的类别包括背景数据、兴趣数据和社交数据， 终端根据预设的相似度检测算法计算真实对象和虚拟对象的身份数据相似度， 其中，身份数据中的背景数据对应第一相似度检测算法，兴趣数据对应第二相 似度检测算法，社交数据对应第三相似度检测算法。

[0063] S203、终端根据目标相似度检测算法检测真实对象和虚拟对象的相似度。

[0064] 本发明实施例中，终端根据第一相似度检测算法检测真实对象和虚拟对象 的背景数据的相似度，并得到第一相似值；具体的，分别提取所述真实对象和 所述虚拟对象的背景数据中的字符串数据和数字数据；根据预设的字符串数据 相似度检测算法检测字符串数据的相似度，得到字符串相似度值；根据预设的 数字数据相似度检测算法检测数字数据的相似度，得到数字相似度值，对计算 得到的字符串数据相似度值和数字数据相似度值进行汇总得到背景数据相似度 对应的第一相似值。其详细步骤可参见步骤S303。

[0065] 终端根据第二相似度检测算法检测真实对象和虚拟对象的兴趣数据的相似 度，并得到第二相似值。具体的，终端提取兴趣数据中的关键词；记录所述关 键词出现的次数，并根据次数与权值的对应关系得到所述关键字对应的权值； 根据所述权值对所述兴趣数据进行向量表示，计算兴趣数据的向量积，得到真 实对象和虚拟对象的兴趣数据相似度对应的第二相似值。其详细步骤可参见步 骤S304。

[0066] 终端根据第三相似度检测算法检测真实对象和虚拟对象的兴趣数据相似度， 并

得到第三相似值。具体的，终端找到真实对象的社交关系圈和虚拟对象的社 交关系圈，检测两者社交关系圈中相同对象的个数，并将相同人数与总数的比 值作为第三相似值，以此得到真实对象和虚拟对象的社交数据相似度。其详细 步骤可参见步骤S305。

[0067] 最终汇总真实对象和虚拟对象的背景数据相似度、兴趣数据相似度和社交 数据相似度即可得到真实对象和虚拟对象的身份数据相似度，详细步骤可参见 步骤S306。

[0068] 本发明实施例中，终端获取真实对象和虚拟对象的身份数据，身份数据包 括背景数据、兴趣数据和社交数据，终端根据身份数据的类别与相似度检测算 法的对应关系，从预设的多种相似度检测算法中筛选出与真实对象和虚拟对象 的身份数据相对应的目标相似度检测算法，终端根据目标相似度检测算法检测 真实对象和虚拟对象的相似度。通过本发明实施例，可以判断真实对象与虚拟 对象的相似度，进而找到真实对象在社交网络上的虚拟身份，或者，通过获取 社交网络上的虚拟身份，找到该虚拟身份对应的真实对象。

[0069] 请参见图3，为本发明实施例提供的一种身份数据相似度检测方法的流程示 意图，该方法可包括：

[0070] S301、终端获取真实对象和虚拟对象的身份数据。

[0071] 本发明实施例中，真实对象和虚拟对象的身份数据包括背景数据、兴趣数 据和社交数据。

[0072] S302、终端根据获取到的身份数据构建身份数据模型。

[0073] 本发明实施例中，终端获取到真实对象和虚拟对象的身份数据之后，会根 据获取到的身份数据构建身份数据模型。

[0074] 具体的，给定一个真实对象u，其身份数据包括3种属性数据(背景数据， 兴趣数据和社交数据)，身份数据Profile(u)的具体表达式为：

[0075] Profile(u)＝{Background(u)，Interest(u)，Relative(u)}

[0076] 其中，Background(u)表示真实对象u的背景数据，Interest(u)表示u的 兴趣数据，Relative(u)表示u的社交数据。

[0077] 给定一个虚拟对象v，其身份数据也包括3种属性数据(背景数据，兴趣数 据和社交数据)，身份数据Profile(v)的具体表达式为：

[0078] Profile(v)＝{Node(v)，Tweet(u)，Relation(v)}

[0079] 其中，Node(v)表示虚拟对象v的背景数据，Tweet(v)表示v的兴趣数据， Relation(v)表示v的社交数据。

[0080] S303、终端根据第一相似度检测算法检测真实对象和虚拟对象的背景数据 的相似度，并得到第一相似值。

[0081] 本发明实施例中，终端构建了真实对象u和虚拟对象v的身份数据模型之 后，可以进一步构建u和v的背景数据模型，并根据第一相似度检测算法检测 真实对象u和虚拟对象v的背景数据的相似度，得到第一相似值。

[0082] 具体的，终端构建真实对象u的背景数据模型Backrgound(u)。

[0083] Background(u)＝{String(u)，Number(u)}

[0084] 其中，本发明实施例将背景数据中的字符串和数字分开表示，String(u)表 示真实对象u的背景数据中的字符串集合，由背景数据中的字符串组成， Number(u)表示u的背景数据中的数字集合，由背景数据中的数字组成。

[0085] 具体的,String(u)可以具体表示为:

[0086] String(u)＝{Name(u),lace(u),Describe(u)}

[0087] 其中,Name(u)表示对象u的名称数据,包括现用名、曾用名、英文名、学校名称和公司名称等。Place(u)表示对象u的地址数据,包括生源地、工作地、旅游地等,Describe(u)表示对对象u的描述数据,包括星座、生肖、教育经历、工作经历等。

[0088] 具体的,Number(u)可以具体表示为:

[0089] Number(u)＝{Date(u),Figure(u)}

[0090] 其中,Date(u)表示对象u的日期数据,包括生日、纪念日、节日等,Figure(u) 表示与对象u有关的数字数据,包括车牌号、门牌号、幸运数字、手机号和身 份证号等。

[0091] 终端构建虚拟对象v的背景数据模型Node(v)。

[0092] Node(v)＝{String(v),Number(v)}

[0093] 其中,String(v)表示虚拟对象v的背景数据中的字符串集合,由v的背景 数据中的字符串组成,Number(v)表示v的背景数据中的数字集合,由v的背 景数据中的数字组成。

[0094] 具体的,String(v)可以具体表示为:

[0095] String(v)＝{UName(v),Address(v),Tag(v)}

[0096] 其中,UName(v)表示虚拟对象v的用户名数据。Address(v)表示虚拟对象 v的注册时填写的地址数据,Tag(v)表示对虚拟对象v的标签数据,如星座、生 肖等。

[0097] Number(v)可以具体表示为:

[0098] Number(v)＝{Birth(v),Sex(v),Other(v)}

[0099] 其中:Birth(v)表示虚拟对象v注册时填写的生日;Sex(v)中,男性Male 用1表示,女性Female用0表示。Other(v)用于存储用户名中包含的数字和用 户发布文本中出现的日期或数字。

[0100] 终端创建真实对象u和虚拟对象v的背景数据模型之后,则会根据第一相 似度检测算法计算真实对象u和虚拟对象v的背景数据相似度,得到第一相似 值。

[0101] 具体的,采用第一相似度检测算法计算真实对象u和虚拟对象v的背景数 据相似度Sim_1(Background(u),Node(v)),具体计算公式为:

[0102] Sim_1(Background(u),Node(v))

[0103] ＝$\omega_1 Sim_{11}$(String(u),String(v))

[0104] ＋$\omega_2 Sim_{12}$(Number(v),Number(v))

[0105] 对于String(u)和String(v),相似度可以根据他们的姓名相似度、地址相似 度、个人描述相似度来度量。其中,姓名相似度和地址相似度更能揭示(u,v) 之间潜在的身份相似性,而身份的描述数据在真实对象对自己的个人描述和虚 拟对象对自己选择的标签中,可能会存在描述范围太广而只有极少部分的重叠, 甚至是严重的数据缺失等情况。因此本发明实施例在对字符串相似性进行度量 的时候,只考虑姓名相似度和地址相似度。

[0106] 具体的,String(u)和String(v)的相似度计算公式为:

[0107] Sim_{11}(String(u),String(v))

[0108] ＝$\omega_{11} Sim_{111}$(Name(u),UName(v))

[0109] ＋$\omega_{12} Sim_{112}$(Place(v),Address(v))

[0110] 其中,ω_i为各个属性相似度的权值,$\omega_1+\omega_2=1$,$\omega_{11}+\omega_{12}=1$,对于$\omega_i$的 具体数

值,本发明实施例不做限定。

[0111] 对于真实对象u和虚拟对象v的姓名相似度Sim_{111}(Name(u),UName(v))的 计算之前,将字符串数据中的汉字转化为拼音,二是对用户名字符串进行处理, 只保留字母,如果原用户名中含有数字,则将数字另存入Number(v)数据集中 的Other(v)集中。

[0112] 真实对象u和虚拟对象v的姓名相似度Sim_{111}(Name(u),UName(v))的具体 计算算法如下:

[0113] 输入:两个名字字符串Name(u)和UName(v)记做N_u和N_v

[0114] 输出:N_u和N_v的相似度

[0115] 1.cn←0//cn为对比次数

[0116] 2.while(N_u和N_v中存在相同字符)DO

[0117] 3.lcs_i←N_u和N_v中最长公共子字符串长度

[0118] 4.cn++

[0119] 5.删除检测到的N_u和N_v中的最长公共字符串

[0120] 6.end while

[0121] 7.if(cn==0)//当不存在相同字符时进行参数调整

[0122] 8.cn←1

[0123] 9.end if

[0124] 10.return

[0125] $$Sim_{111}=\frac{\sum_{i=1}^{cn} lcs_i - cn + 1}{|N_u| + |N_v| + \sum_{i=1}^{cn} lcs_i}$$

[0126] 其中,$|N_u|$和$|N_v|$为最终删除所有最长公共子字符串后字符串N_u和N_v的长度。

[0127] 举例说明,若N_u为abcde,N_v为abcdf,则$|N_u|=1$,$|N_v|=1$, $\sum_{i=1}^{cn} lcs_i = 4$,cn=1,则求出最终的$Sim_{111}$=0.67,若$N_u$为abcd,$N_v$为abcd, $|N_u|=0$,$|N_v|=0$,$\sum_{i=1}^{cn} lcs_i = 4$,cn=1,求出最终的$Sim_{111}$=1。

[0128] 对于真实对象u和虚拟对象v的地址相似度Sim_{112}(Place(v),Address(v)), 首先采用国家-省份-地市三段数据结构来表示,通过分层比较计算转换次数的方 式来计算其相似度,Sim_{112}(Place(v),Address(v))的具体计算公式为:

[0129] $$Sim_{112}\big(Place(v), Address(v)\big) = \sum_{i=1}^{n} \omega_{ai} * \left(1 - \frac{T(Place(u), Address(v))}{3}\right)$$

[0130] 其中,ω_{ai}表示地理位置的权值,所有ω_{ai}相加之和为1,对于ω_{ai}的具体数 值,本发明实施例不做限定。T(Place(u),Address(v))表示两个地理位置属性的 转换次数,即分别比较真实对象u和虚拟对象v的地址中的国家、省份和地市 是否相同,如果不同,则转换次数加1。若国家、省份和地市都相同,则转换次 数为0,若国家、省份和地市都不同,则转换次数为3。

[0131] 例如,终端获取到真实对象u和虚拟对象v之间需要对比的地址有3个, 分别为当前所在地,户籍地,工作地址。则可以为当前所在地分配权值ω_{a1}=0.5, 户籍地权值ω_{a2}=0.3,工作地址ω_{a3}=0.2。且三个地址的国家和省份都相同,地市 都不同,即T(Place(u),

Address(v))＝1,则Sim_{112}(Place(v),Address(v))最终 的计算结果为0.67。

[0132] 对于真实对象u和虚拟对象v之间的数字数据Number(v)和Number(v)， 相似度可以根据他们的生日相似度和性别相似度来度量。

[0133] Sin_{12}(Number(v),umber(v))

[0134] ＝$\omega_{21}Sim_{121}$(Birth(u),Birth(v))+$\omega_{22}Sim_{122}$(Sex(u),Sex(v))

[0135] 其中,ω21+ω_{22}＝1,Birth(u)和Birth(v)分别表示u和v的生日数据， Sex(v)和Sex(v)分别表示u和v的性别数据。

[0136] 对于对象的生日数据,按年-月-日(YYYY-MM-DD)的格式记录8位数字,如 1995-05-26表示对象的生日是1995年5月26日。对于生日的相似度计算,本发 明实施例将分为两步计算,第一步完成对年份的相似度计算,第二步完成对月 和日的相似度计算。

[0137] Sim_{121}(Birth(u),Birth(v))

[0138] ＝$\omega_{23}Sim_{123}$(Y(u),(v))+$\omega_{24}Sim_{124}$(MD(u),MD(v))

[0139] 第一步:因为不同年龄层的对象往往拥有不同的阅历和关注点,本发明实 施例通过生日中的年份直接得出年龄。一般而言,年龄差越小,用户的相似度 越高,但仅用年龄差不能准确描述年龄相似度,年龄差对年龄值的比也是重要 的计算因素,则关于年份的相似度计算公式:

[0140] $$Sim_{123}\big(Y(u),Y(v)\big)=1-\frac{|Y(u)-Y(v)|}{MAX(\big(m-Y(u)\big),\big(m-Y(v)\big))}$$

[0141] 其中:m表示当年年份,如2018,Y(u)表示真实对象u的生日的年份,Y(v)表 示虚拟用户v的生日的年份,MAX((m-Y(u)),(m-Y(v)))表示u和v之中年 龄较大的年龄值。

[0142] 对于月和日的部分(4位),本发明实施例采用编辑距离方法来计算相似度， 编辑距离用于评价两个字符串间的相似度。编辑距离反映了两个字符串的绝对 差异,而相似度以一个[0,1]之间的数值反应两个字符串的相似程度,数值越大表 示相似程度越高。生日中月日的相似度的计算公式:

[0143] $$Sim_{124}\big(MD(u),MD(v)\big)=1-\frac{T(MD(u),MD(v))}{4}$$

[0144] 其中:MD(u)表示真实对象的生日的月日部分,MD(v)表示虚拟用户的生 日的月日部分,T(MD(u),MD(v))表示转换次数。

[0145] 基于生日的月日部分的相似度Sim_{124}(MD(u),MD(v))的计算,本发明实施 例提出的算法如下:

[0146] 输入:两个生日月日部分的数字MD(u)和MD(v)记做M_u,M_v;

[0147] 输出:相似度;

[0148] 1.定义

[0149] m＝M_u的长度＝4,n＝M_v的长度＝4,

[0150] d[m+1][n+1]//矩阵

[0151] temp//记录相同字符,在某个矩阵位置值的增量,非0即1;

[0152] 整型变量i,j;字符型变量ch1,ch2;

[0153] 2.d[i][0]＝i d[0][j]＝j//初始化第一行和第一列;

[0154] 3.遍历M_u去匹配M_v；

[0155] if(ch1==ch2)temp=0；

[0156] else temp=1;//ch1记录M_u的字符,ch2记录M_v的字符；

[0157] 4.d[i][j]=min(d[i-1][j]+1,d[i][j-1]+1,d[i-1][j-1]+temp)

[0158] //矩阵上边+1,左边+1,左上+temp取最小；

[0159] 5.T=d[m][n]//d[m][n]即为Mu转换为M_v需要编辑的次数；

[0160] 6.return

[0161] $$Sim_{124}(M_u, M_v) = 1 - \frac{T}{4}$$

[0162] 对于对象的性别数据,当真实对象u和虚拟用户v的性别相同时,在性别 这一维度的相似度为1,反之相似度为0(本发明实施例不考虑将虚拟用户的性别 故意设置为与本人真实性别相反的情况)。性别相似度的计算公式为：

[0163] $$Sim_{122}\big(Sex(u), Sex(v)\big) = \begin{cases} 0, Sex(u) \neq Sex(v) \\ 1, Sex(u) = Sex(v) \end{cases}$$

[0164] 根据上述算法计算出背景数据中的各个子相似度值之后,将计算出的各个 子相似度值带入背景数据相似度计算公式Sim_1(Background(u),Node(v))中即 可计算真实对象u与虚拟对象v的背景数据相似度,得到第一相似值S_1。

[0165] S304、终端根据第二相似度检测算法检测真实对象和虚拟对象的兴趣数据 的相似度,并得到第二相似值。

[0166] 本发明实施例中,终端构建了真实对象u和虚拟对象v的身份数据模型之 后,可以进一步构建u和v的兴趣数据模型,并根据第二相似度检测算法检测 真实对象u和虚拟对象v的兴趣数据的相似度,得到第二相似值。

[0167] 具体的,终端构建真实对象u的兴趣数据模型Interest(u),本发明实施例 在记录真实对象的兴趣数据时,获取到的兴趣数据的文档中可能包含是字、词、 句、章等,因此采用空间向量模型VSM的表示法,将兴趣数据Interest(u)表示 为：

[0168] Interest(u)=(ω_{u1},ω_{u2},…ω_{un})

[0169] 其中,i可以对应记录的对象感兴趣的特征词,ω_{ui}为i对应的特征词的权 重。

[0170] $$\omega_{ui} = tf_i(Interest(u)) \times log\left(\frac{N}{n_i} + 0.01\right)$$

[0171] 其中,tf_i(Interest(u))表示i对应的特征词在文档中的频率,log(N/n_i+ 0.01)表示为i对应的特征词的逆文档频率。N表示全部训练集的文本数,n_i表示 训练文本中出现i对应的特征词的文本频数。

[0172] 终端构建虚拟对象v的兴趣数据模型Tweet(v),具体的,Tweet(v)表示虚 拟对象v在社交网络中发布的各个文本数据组成的长文本(其文本内容可能包含 兴趣词、情感词、事件时间词、数字等)。本发明实施例将其表示为一个文本向 量。过程如下：

[0173] 第一步:文本预处理:对Tweet(v)进行过滤噪音数据、分词、词性标注、 去除停用词等处理；

[0174] 第二步:数字处理:把文本中出现的日期和数字存入Other(v)中；

[0175] 第三步:特征提取:采用数据增益的特征选择算法提取Tweet(v)的特征词，对文本进行降维处理;

[0176] 第四步:权重计算:Tweet(v)中的每个特征词t_{vi}的权重ω_{vi}。

[0177] $$\omega_{vi} = tf_i(Tweet(v)) \times log\left(\frac{N}{n_i} + 0.01\right)$$

[0178] 其中,tf_i(Tweet(v))表示i对应的特征词在文档中的频率,log(N/n_i+0.01) 表示为i对应的特征词的逆文档频率。N表示全部训练集的文本数,n_i表示训练 文本中出现i对应的特征词的文本频数,取对数是为了平衡,避免N/n_i值过大而 占据主要作用,0.01的作用是为了避免当N＝n_i时对数为0。

[0179] 第五步:向量表示:Tweet(v)＝(ω_{v1},ω_{v2},…ω_{vn}),其中ω_{vi}为虚拟对象v 在社交网络中发布的各个文本数据中某个i对应的特征词的权重。

[0180] 终端构建真实对象u和虚拟对象v的兴趣数据模型之后,将根据第二相似 度检测算法检测真实对象和虚拟对象的兴趣数据的相似度,得到第二相似值。

[0181] 具体的,对于真实对象u的兴趣文本可以表示为文本特征向量:

[0182] Interest(u)＝(ω_{u1},ω_{u2},…ω_{un})

[0183] 对于虚拟用户v的兴趣文本可表示为文本特征向量:

[0184] Tweet(v)＝(ω_{v1},ω_{v2},…ω_{vn})

[0185] 则真实对象u和虚拟对象v之间的兴趣数据相似度对应的第二相似度检测 算法Sim_2(Interest(u),Tweet(v))的计算公式为:

[0186] $$Sim_2\big(Interest(u), Tweet(v)\big) = \frac{\sum_{i=1}^{n}(\omega_{ui} * \omega_{vi})}{\sqrt{\sum_{i=1}^{n}\omega_{ui}^2}\sqrt{\sum_{i=0}^{n}\omega_{vi}^2}}$$

[0187] 其对应的具体算法如下:

[0188] 输入:两个兴趣向量Interest(u)和Tweet(v)记做arrayNum1[],arrayNum2[];

[0189] 输出:真实对象u和虚拟对象v之间的兴趣数据相似度;

[0190] 1.定义

[0191] arrayNum1[],arrayNum2[]//1,2数组分别存放Interest(u),Tweet(v)

[0192] arrayKey[]//存放关键词合并后的数据

[0193] 2.计算两个向量的点积

[0194] x＝0 i＝0

[0195] while

[0196] i<arrayKey[]的长度

[0197] x＝x+arrayNum1[i]*arrayNum2[i]

[0198] i++

[0199] printx

[0200] 3.计算两个向量的模

[0201] sq1＝0 i＝0

[0202] while

[0203] i<arrayKey[]的长度

[0204] sq1＝sq1+pow(arrayNum1[i],2)//pow(a,2)＝a*a

[0205] i++

[0206] sq2＝0 j＝0

[0207] while

[0208] j<arrayKey[]的长度

[0209] sq2＝sq2+pow(arrayNum2[j],2)

[0210] j++

[0211] 4.return

[0212] $$Sim_2\big(Interest(u),Tweet(v)\big)=\frac{x}{math.sqrt(sq1)*math.sqrt(sq2)}$$

[0213] 终端将计算出的i对应的特征词的权重ω_{ui}和ω_{v1}带入上述相似度计算公式Sim_1(Background(u),Node(v))中，即可计算真实对象u与虚拟对象v的兴趣 数据相似度，得到第二相似值S_2。

[0214] S305、终端根据第三相似度检测算法检测所述真实对象和虚拟对象的社交 数据的相似度，并得到第三相似值。

[0215] 本发明实施例中，终端构建了真实对象u和虚拟对象v的身份数据模型之 后，可以进一步构建u和v的社交数据模型，并根据第三相似度检测算法检测 真实对象和虚拟对象的兴趣数据的相似度，得到第三相似值。

[0216] 具体的，构建真实对象u的社交数据模型，Relative(u)：表示u的社交数 据，用树型结构来表示，对象u为根节点，其余对象为子节点，按与对象u的 亲疏关系依次往下排列，其中每条边的权值为1，从该子节点出发到根节点的距 离越远，数值越大，则两人关系越疏远，反之，若距离为1，则说明与对象u关 系密切，一般为伴侣，父母，子女等。

[0217] 构建虚拟对象v的社交数据模型，Relation(v)，表示v的社交数据，包括2 种属性(链接数据、互动数据)，其中链接数据包括关注数据和粉丝数据，互动数 据包括转发数据、评论数据和@数据。本发明实施例将它们分别表示为五个向 量：关注向量Followee(v)、粉丝向量Follower(v)、转发向量Repost(v)、评论 向量Comment(v)，@(v)，则Relation(v)可以表示为：

[0218] Relation(v)＝{Followee(v),Follower(v),Repost(v),Comment(v),@(v)}

[0219] 终端构建虚拟对象v的社交数据模型之后，会在社交网络上寻找与虚拟对 象相关联的其他虚拟对象v’，并计算根据v与v’的相似度。其中v与v’的相 似度Relation(v,v′)的计算公式为：

[0220] Relation(v,v’)

[0221] ＝ω_4Sim4(Link(v),Link(v’))

[0222] +ω_5Sim5(Interactuon(v),Interaction(v’))

[0223] 其中，Sim_4(Link(v),Link(v’))表示v与v’链接数据相似度，v与v’的互动 数据相似度表示为Sim_5(Interaction(v),Interaction(v’))，$\omega_4+\omega_5=1$。对于 ω_4和ω_5的具体数值，本发明实施例不做限定。

[0224] 可选的，对于虚拟对象的链接数据相似度，本发明实施例提供了如下分析 方法，

用户的链接数据包含2种属性数据(关注数据和粉丝数据),表示为:

[0225] Link(v)＝{Followee(v),Follower(v)}

[0226] 其对象链接数据相似度,可以根据(v,v')之间的各种属性相似度而计算,链 接数据相似度Sim_4(Link(v),Link(v'))的计算公式为:

[0227] Sim_4(Link(v),Link(v'))

[0228] ＝$\omega_6 Sim_6$(Followee(v),Followee(v'))

[0229] +$\omega_7 Sim_7$(Follower (v),Follower(v'))

[0230] 其中,

[0231] $$Sim_6\big(Followee(v),Followee(v')\big)=\frac{Followee(v)\cdot Followee(v')}{\|Followee(v)\|\|Followee(v')\|}$$

[0232] $$Sim_7\big(Followee(v),Followee(v')\big)=\frac{Followee(v)\cdot Follower(v')}{\|Follower(v)\|\|Follower(v')\|}$$

[0233] 可选的,对于虚拟对象的互动数据相似度,本发明实施例提供了如下分析 方法,用户的互动数据包含3种属性数据(转发数据、评论数据、@数据),表 示为:

[0234] Interaction(v)＝{Repost(v),Comment(v),@(v)}

[0235] 其对象互动数据相似度,可以根据(v,v')之间的各种属性相似度而计算,链 接数据相似度Sim_5(Interaction(v),Interaction(v'))的计算公式为:

[0236] Sim_5(Interaction(v),Interaction(v'))

[0237] ＝$\omega_8 Sim_8$(Repost(v),epost(v'))

[0238] +$\omega_9 Sim_9$(Comment(v),Comment(v'))+$\omega_{10} Sim_{10}$(@(v),@(v'))

[0239] 其中,$\omega_8+\omega_9+\omega_{10}$＝1,对于$\omega_8$、$\omega_9$和$\omega_{10}$的具体数值,本发明实施例 不做限定。

[0240] 具体的,

[0241] $$Sim_8\big(Repost(v),Repost(v')\big)=\frac{Repost\_num(v->v')+Repost\_num(v'->v)}{2}$$

[0242] 其中,Repost_num(v->v')表示虚拟对象v是否转发对象v'在社交网络上 发布的数据文本;Repost_num(v'->v)表示虚拟对象v'是否转发对象v在社交 网络上发布的数据文本,若是,则为1,若否,则为0。

[0243] $$Sim_9\big(Comment(v),Comment(v')\big)=\frac{Comment\_num(v->v')+Comment\_num(v'->v)}{2}$$

[0244] 其中,Comment_num(v->v')表示虚拟对象v是否评论对象v'在社交网 络上发布的数据文本,Comment_num(v'->v)表示虚拟对象v'是否评论对象v 在社交网络上发布的数据文本。若是,则为1,若否,则为0。

[0245] $$Sim_{10}\big(@(v),@(v')\big)=\frac{@\_num(v->v')+@\_num(v'->v)}{2}$$

[0246] 其中,@_num(v->v')表示虚拟对象v是否@对象v',@_num(v'->v)表 示虚拟对象v'是否@对象v。若是,则为1,若否,则为0。

[0247] 最终将上述公式计算得出的相似度带入Relation(v,v′)的计算公式中,即可 得到虚拟对象v和v'的相似度。选取相似度排名前k位的虚拟对象v'即为与虚 拟对象v相关联的虚拟对象,构成虚拟对象v的社交关系。

[0248] 终端构建虚拟对象v的社交关系之后，可以根据第三相似度检测算法检测 真实对象和虚拟对象的社交数据的相似度$Sim_3(Relative(u),Relation(v))$，并得 到第三相似值$S_3$，第三检测算法的具体公式为：

[0249] $$Sim_3\left(Relative(u),Relation(v)\right)=\frac{\|Relative(u)\cap Relation_k(v)\|}{\|Relation_k(v)\|}$$

[0250] 其中，Relative(u)表示真实对象u的社交关系，包括多个与对象u有关联 的真实对象，$Relation_k(v)$表示与虚拟对象v相似度排名前k位的虚拟对象v'。

[0251] 举例说明，若k个虚拟对象v'都与真实对象u的社交数据中的对象对应， 则第三相似值$S_3=1$，若k个虚拟对象v'都不与真实对象u的社交数据中的对象 对应，则第三相似值$S_3=0$。若有m个虚拟对象v'与真实对象u的社交数据中 的对象对应，则$S_3=m/k$，其中，$m\leqslant k$。

[0252] S306、终端根据预设的加权规则对第一相似值、第二相似值和第三相似值 进行加权处理。

[0253] 本发明实施例中，根据相似度检测算法计算出第一相似值S_1、第二相似值S_2和第三相似值S_3之后，会给计算出的相似值赋予一个加权系数，得到真实对象u 和虚拟对象v的身份数据相似度

[0254] $Sim(Profile(u),Profile(v))=\beta_1Sim_1(Background(u),Node(v))$

[0255] $+\beta_2Sim_2(Interest(u),Tweet(v))$

[0256] $+\beta_3Sim_3(Relative(u),Relation(v))$

[0257] 其中，β_1、β_2和β_3为加权系数，$\beta_1+\beta_2+\beta_3=1$，对于$\beta_1$、$\beta_2$和$\beta_3$的具体 数值，本发明实施例不做限定。

[0258] S307、终端将加权处理后的各个相似值进行汇总得到所述真实对象和虚拟 对象的身份数据的相似度对应的相似值。

[0259] 本发明实施例中，终端根据S306获取到各个相似值的加权系数之后，对各 个相似值进行汇总处理即可得到真实对象u和虚拟对象v的身份数据相似度对 应的相似值S。

[0260] $S=\beta_1S_1+\beta_2S_2+\beta_3S_3$

[0261] 其中，S_1为第一相似值，S_2为第二相似值，S_3为第三相似值。

[0262] 可选的，当知道一个真实对象u，需要在社交网络上找到该真实对象u的虚 拟身份时，可以根据上述相似度检测算法找到身份数据相似度排名前n位的虚 拟对象v，根据数值大小对S(u)中的虚拟用户再进行降序排列，得到一个新的身 份相似序列S'(u)，最后根据真实对象u的其它数据(例如：文化水平、计算机操 作能力、性格、气质、行为等)对比虚拟身份v的字符串风格、情感倾向、时间 属性、设备持有情况、定位数据等。选择S'(u)中排名靠前的k位用户，来综合 考虑虚拟用户v是不是真实对象u在社交网络上的虚拟身份。

[0263] 本发明实施例中，终端获取真实对象和虚拟对象的身份数据之后，会根据 身份数据构建背景数据模型、兴趣数据模型和社交数据模型，并计算出真实对 象和虚拟对象的背景数据相似度、兴趣数据相似度和社交数据相似度，最后汇 总得到真实对象和虚拟对象的身份数据相似度。通过本发明实施例，可以判断 真实对象与虚拟对象的相似度，进而找到真实对象在社交网络上的虚拟身份， 或者，通过知道社交网络上的虚拟身份，知道该虚拟身份对应的真实对象。

[0264] 下面将结合附图4对本发明实施例提供的终端进行详细介绍。需要说明的 是，附

图4所示的终端，用于执行本发明图1-图3所示实施例的方法，为了便 于说明，仅示出了与本发明实施例相关的部分，具体技术细节未揭示的，经参 照本发明图1-图3所示的实施例。

[0265] 请参见图4，为本发明提供的一种终端的结构示意图；该终端40可包括：获取模块401、搜索模块402、检测模块403、确定模块404、计算模块405和 排序模块406。

[0266] 获取模块401，用于获取第一真实对象的身份数据；

[0267] 搜索模块402，用于以所述第一真实对象的身份数据作为搜索信息搜索得到 虚拟对象的身份数据，并根据所述搜索得到的虚拟对象的身份数据确定第一虚 拟对象；

[0268] 所述搜索模块402，还用于以所述第一虚拟对象的身份数据作为搜索信息得 到关联虚拟对象的身份数据，并根据所述关联虚拟对象的身份数据确定第二虚 拟对象；

[0269] 所述搜索模块402，还用于以所述第二虚拟对象的身份数据作为搜索信息得 到关联真实对象的身份数据，并根据所述关联真实对象的身份数据确定第二真 实对象；

[0270] 检测模块403，用于检测所述第一真实对象与所述第二真实对象的身份数据 相似度；

[0271] 确定模块404，根据所述身份数据相似度，确定所述第一真实对象与所述第 二真实对象的关系。

[0272] 可选的，本发明实施例所述的终端，还包括：

[0273] 计算模块405，用于分别计算各个虚拟对象的背景数据与所述第一真实对象 的背景数据相似度；

[0274] 排序模块406，用于将所述虚拟对象按所述背景数据相似度从高到低的顺序 降序排列；

[0275] 所述计算模块405，还用于计算排序为前n位的虚拟对象与所述第一真实对 象的兴趣数据相似度；

[0276] 所述计算模块405，还用于计算排序为前n位的虚拟对象与所述第一真实对 象的社交数据相似度；

[0277] 所述计算模块405，还用于将所述排序为前n位的虚拟对象与所述第一真实 对象的背景数据相似度、兴趣数据相似度和社交数据相似度进行加权汇总，得 到所述排序为前n为的虚拟对象与所述第一真实对象的身份数据相似度；

[0278] 所述排序模块406，还用于将所述排序为前n位的虚拟对象按所述身份数据 相似度从高到低的顺序降序排列；

[0279] 所述确定模块404，还用于将排序为第一位的虚拟对象确定为所述第一虚拟 对象，其中，n为大于1的整数。

[0280] 可选的，本发明实施例所述的终端，还包括：

[0281] 所述获取模块401，还用于提取背景数据中的字符串数据和数字数据，所述 字符串数据包括姓名数据和地址数据，所述数字数据包括生日数据和性别数据；

[0282] 所述计算模块405，还用于分别计算所述第一真实对象和虚拟对象的姓名数 据相似度和地址数据相似度；

[0283] 所述计算模块405，还用于将所述姓名数据相似度和所述地址数据相似度进 行加权汇总得到所述第一真实对象和所述虚拟对象的字符串数据相似度；

[0284] 所述计算模块405，还用于分别计算所述第一真实对象和所述虚拟对象的生 日数

据相似度和性别数据相似度；

[0285] 所述计算模块405，还用于将所述生日数据相似度和所述性别数据相似度进 行加权汇总得到所述第一真实对象和所述虚拟对象的数字数据相似度；

[0286] 所述计算模块405，还用于将所述字符串数据相似度和所述数字数据相似度 进行加权汇总得到所述第一真实对象与所述虚拟对象的背景数据相似度。

[0287] 可选的，本发明实施例所述的终端，还包括：

[0288] 所述获取模块401，还用于提取所述第一真实对象和虚拟对象的兴趣数据中 的兴趣关键词；

[0289] 所述获取模块401，还用于获取所述兴趣关键词的在所述兴趣数据中的权重；

[0290] 所述计算模块405，还用于将所述兴趣数据采用空间向量模型表示为兴趣向 量，其第k个向量的值为第k个关键词对应的权重，其中，k为大于或等于1的 整数；

[0291] 所述计算模块405，还用于计算所述第一真实对象的兴趣向量与所述虚拟对 象的兴趣向量的余弦值；

[0292] 所述计算模块405，还用于将所述余弦值作为所述第一真实对象和所述虚拟 对象的兴趣数据相似度。

[0293] 所述获取模块401，还用于获取与所述第一真实对象有关联的第三真实对象 集合和与虚拟对象有关联的第三虚拟对象集合；

[0294] 所述检测模块403，还用于检测所述第三真实对象集合与所述第三虚拟对象 集合的交集的人数；

[0295] 所述计算模块405，还用于将所述交集的人数与所述第三虚拟对象集合中的 人数的比值作为虚拟对象与所述第一真实对象的社交数据相似度。

[0296] 若所述第一真实对象与所述第二真实对象的相似度大于第一预设阈值，则 所述确定模块404确定所述第一真实对象与所述第二真实对象为同一个人；

[0297] 若所述第一真实对象与所述第二真实对象的相似度介于第一预设阈值和第 二预设阈值之间，则所述确定模块404确定所述第二真实对象与所述第一真实 对象为亲密关系；

[0298] 本发明实施例中，通过对真实对象与真实对象之间的相似度计算、真实对 象与虚拟对象之间的相似度计算，可自动地进行虚拟身份数据的查找以及对象 之间相似度的识别确定.

[0299] 请参见图5，为本发明实施例提供了另一种终端的结构示意图。如图5所示， 该终端包括：至少一个处理器501，输入设备503，输出设备504，存储器505， 至少一个通信总线502。其中，通信总线502用于实现这些组件之间的连接通信。 其中，输入设备503可以是控制面板或者麦克风等，输出设备504可以是显示 屏等。其中，存储器505可以是高速RAM存储器，也可以是非不稳定的存储器 (non-volatile memory)，例如至少一个磁盘存储器。存储器505可选的还可以是 至少一个位于远离前述处理器501的存储装置。其中处理器501可以结合图4 所描述的终端，存储器505中存储一组程序代码，且处理器501，输入设备503， 输出设备504调用存储器505中存储的程序代码，用于执行以下操作：

[0300] 输入设备503获取第一真实对象的身份数据；

[0301] 处理器501以所述第一真实对象的身份数据作为搜索信息搜索得到虚拟对 象的

身份数据,输出设备504根据所述搜索得到的虚拟对象的身份数据确定第 一虚拟对象

[0302] 处理器501以所述第一虚拟对象的身份数据作为搜索信息得到关联虚拟对 象的身份数据,输出设备504根据所述关联虚拟对象的身份数据确定第二虚拟 对象;

[0303] 若处理器501以所述第二虚拟对象的身份数据作为搜索信息得到关联真实 对象的身份数据,输出设备504根据所述关联真实对象的身份数据确定第二真 实对象;

[0304] 若处理器501检测所述第一真实对象与所述第二真实对象的身份数据相似 度;

[0305] 处理器501根据所述身份数据相似度,确定所述第一真实对象与所述第二 真实对象的关系;

[0306] 本发明实施例中,通过对真实对象与真实对象之间的相似度计算、真实对 象与虚拟对象之间的相似度计算,可自动地进行虚拟身份数据的查找以及对象 之间相似度的识别确定。

[0307] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程, 是可以通过计算机程序来指令相关的硬件来完成,所述的程序可存储于计算机 存储介质中,该程序在执行时,可包括如上述各方法的实施例的流程。其中, 所述的计算机存储介质可为磁碟、光盘、只读存储记忆体(Read-Only Memory, ROM)或随机存储记忆体(Random Access Memory,RAM)等。

[0308] 以上所揭露的仅为本发明较佳实施例而已,当然不能以此来限定本发明之 权利范围,因此依本发明权利要求所作的等同变化,仍属本发明所涵盖的范围。

[0309] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程, 是可以通过计算机程序来指令相关的硬件来完成,所述的程序可存储于计算机 存储介质中,该程序在执行时,可包括如上述各方法的实施例的流程。其中, 所述的计算机存储介质可为磁碟、光盘、只读存储记忆体(Read-Only Memory, ROM)或随机存储记忆体(Random Access Memory,RAM)等。

[0310] 以上所揭露的仅为本发明较佳实施例而已,当然不能以此来限定本发明之 权利范围,因此依本发明权利要求所作的等同变化,仍属本发明所涵盖的范围。

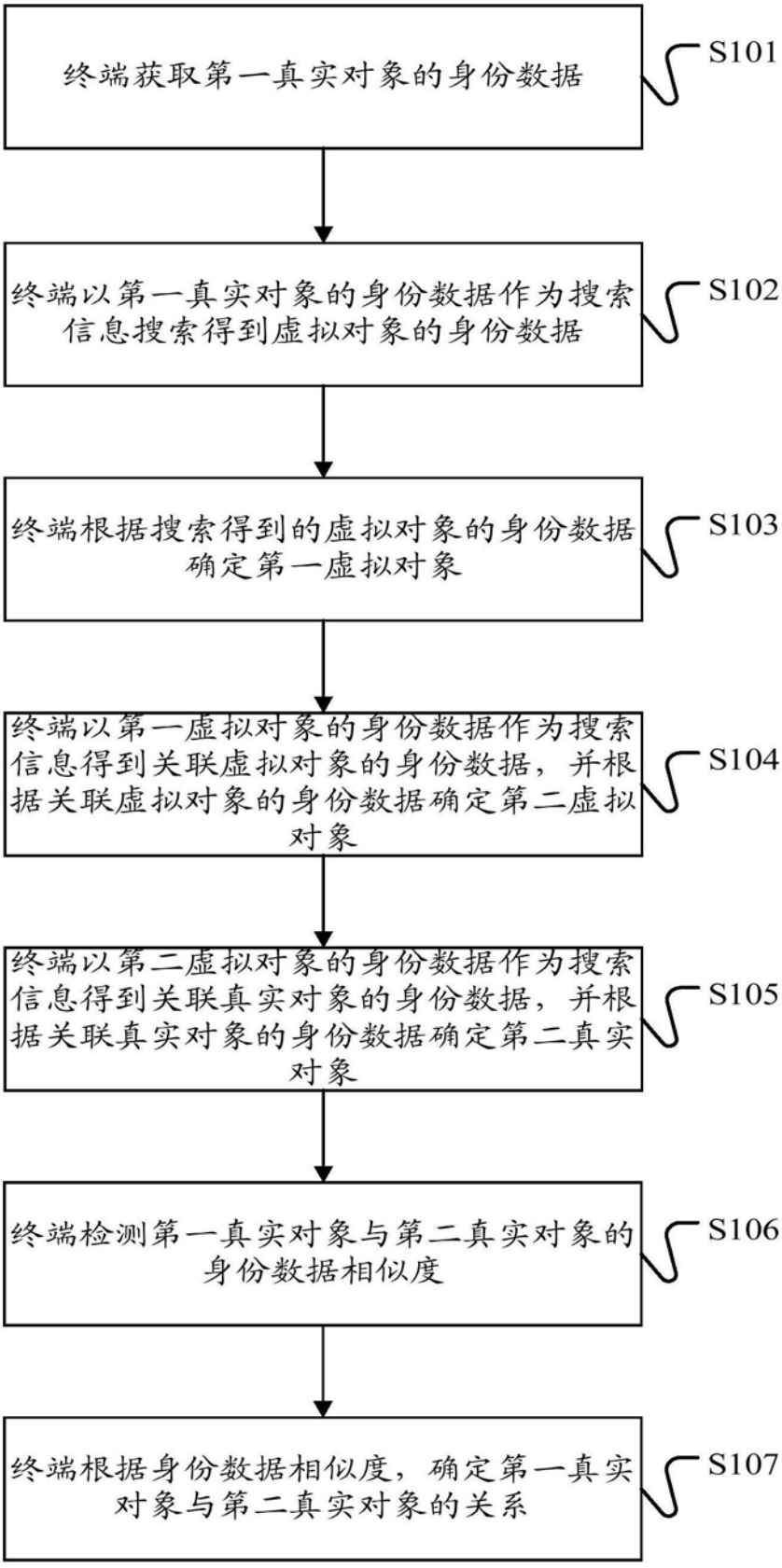

终端获取第一真实对象的身份数据
S101
终端以第一真实对象的身份数据作为搜索信息搜索得到虚拟对象的身份数据
S102
终端根据搜索得到的虚拟对象的身份数据确定第一虚拟对象
S103
终端以第一虚拟对象的身份数据作为搜索信息得到关联虚拟对象的身份数据，并根据关联虚拟对象的身份数据确定第二虚拟对象
S104
终端以第二虚拟对象的身份数据作为搜索信息得到关联真实对象的身份数据，并根据关联真实对象的身份数据确定第二真实对象
S105
终端检测第一真实对象与第二真实对象的身份数据相似度
S106
终端根据身份数据相似度，确定第一真实对象与第二真实对象的关系
S107

图1

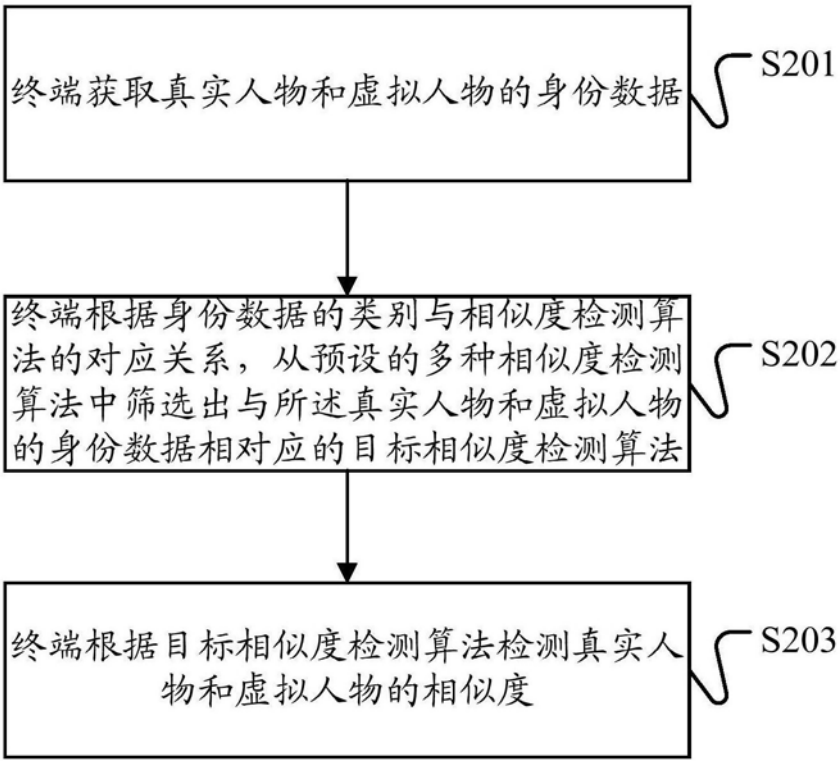

终端获取真实人物和虚拟人物的身份数据
S201
终端根据身份数据的类别与相似度检测算法的对应关系，从预设的多种相似度检测算法中筛选出与所述真实人物和虚拟人物的身份数据相对应的目标相似度检测算法
S202
终端根据目标相似度检测算法检测真实人物和虚拟人物的相似度
S203

图2

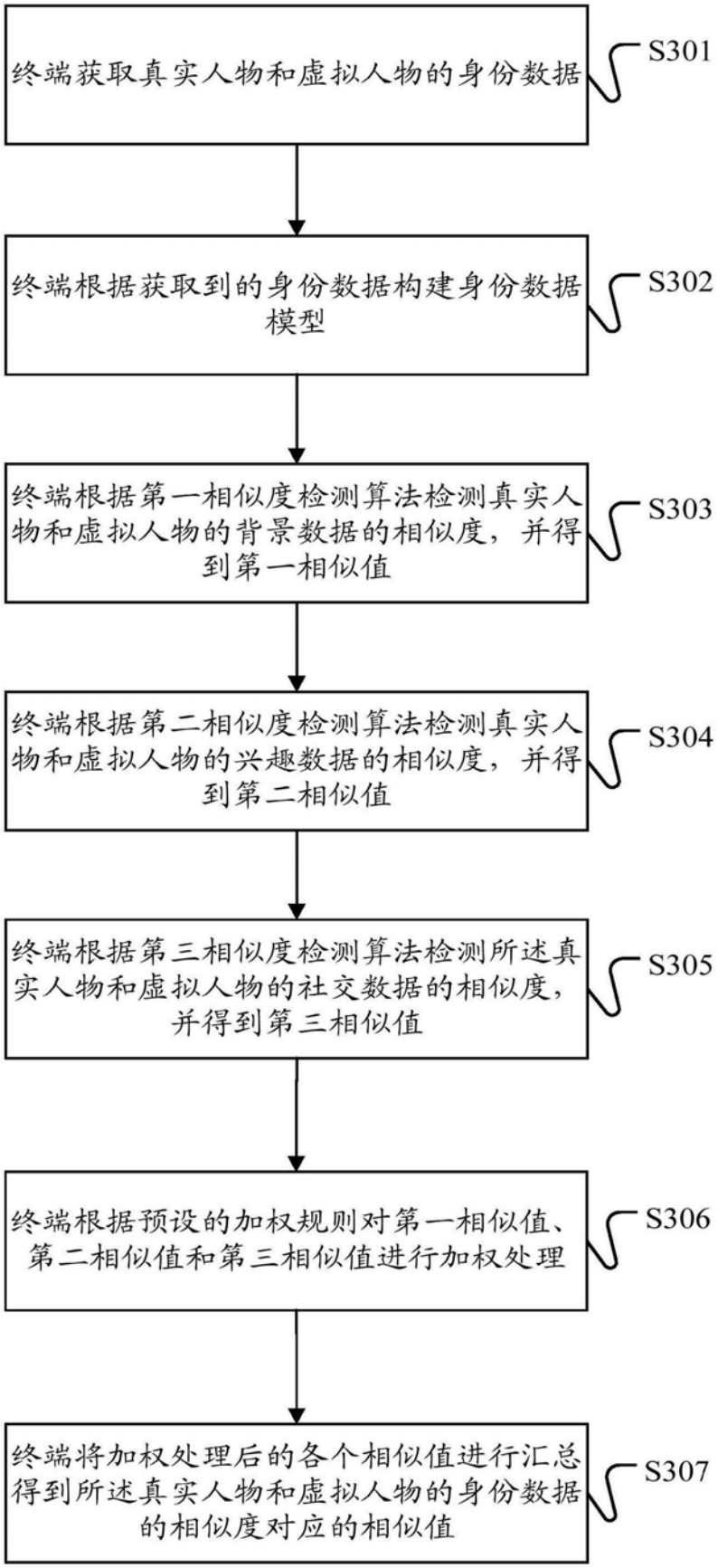

终端获取真实人物和虚拟人物的身份数据
S301
终端根据获取到的身份数据构建身份数据模型
S302
终端根据第一相似度检测算法检测真实人物和虚拟人物的背景数据的相似度，并得到第一相似值
S303
终端根据第二相似度检测算法检测真实人物和虚拟人物的兴趣数据的相似度，并得到第二相似值
S304
终端根据第三相似度检测算法检测所述真实人物和虚拟人物的社交数据的相似度，并得到第三相似值
S305
终端根据预设的加权规则对第一相似值、第二相似值和第三相似值进行加权处理
S306
终端将加权处理后的各个相似值进行汇总得到所述真实人物和虚拟人物的身份数据的相似度对应的相似值
S307

图3

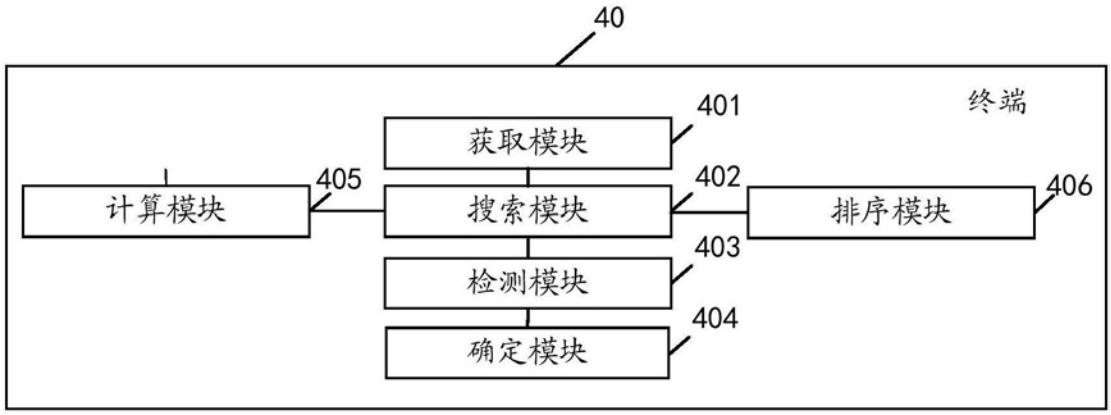

40
终端
401
获取模块
405
计算模块
402
搜索模块
406
排序模块
403
检测模块
404
确定模块

图4

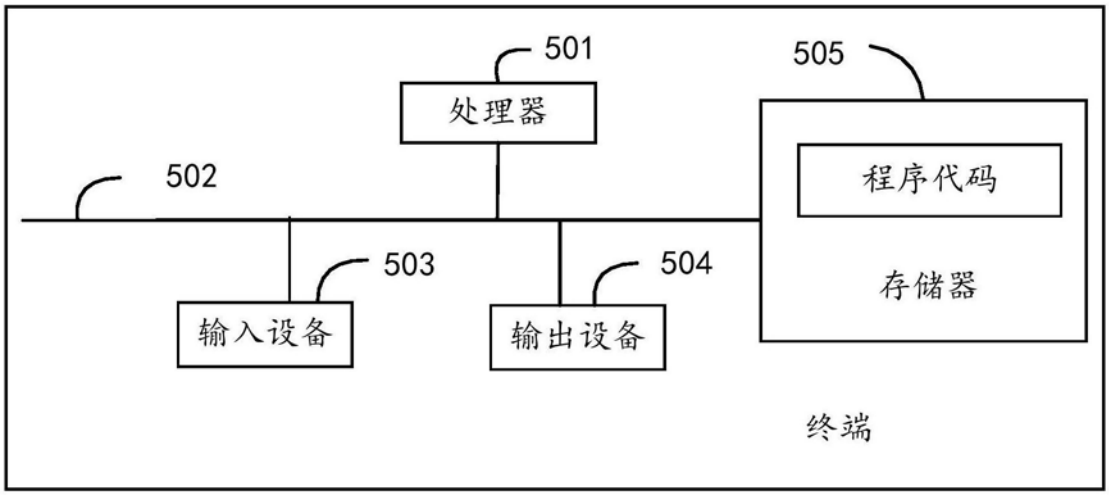

501
处理器
505
程序代码
502
503
504
输入设备
输出设备
存储器
终端

图5