

一种身份识别方法和装置

技术领域

本发明涉及网络技术，特别涉及一种身份识别方法和装置。

背景技术

5 随着中国国内互联网实名制的趋势发展，越来越多的互联网场景要求实名认证，金融和电商等行业首当其冲。在这种趋势下，欺诈作弊分子为了藏匿自己的真实身份，常常通过互联网泄露或者批量购买等方式获取到大量的他人身份证件信息，冒用他人身份证号码、姓名，并利用自己掌控的手机号，进行互联网场景下的账户注册、认证，用以进行信用卡申请或贷款等信贷申请欺诈行为，给商户和金融机构带来损失。

而现有的身份识别方式，主要是基于网络层面或设备层面来认定，比如，可以根据欺诈作弊分子使用的设备的 IP 地址、MAC 地址，或者设备的 IMEI 等标识，通过一定的识别模型来识别身份冒用。但是，有很多欺诈作弊分子是职业黑客，具有很强的网络技术，能够通过执行一定的策略绕过现有的身份识别模型，使得身份识别较为困难。

发明内容

有鉴于此，本发明提供一种身份识别方法和装置，以实现对于身份冒用行为的识别。

具体地，本发明是通过如下技术方案实现的：

20 第一方面，提供一种身份识别方法，所述方法包括：

采集通讯录大数据，所述通讯录大数据包括：多个用户拥有的通讯录，每个通讯录中记录有多个身份信息对，所述身份信息对包括姓名和手机号码；

将待识别的身份信息对与所述通讯录大数据进行比对，获取信息比对结果，所述待识别的身份信息对包括：待识别的用户使用的姓名和手机号码；

若所述信息比对结果满足风险条件，则确定所述用户是风险用户。

第二方面，提供一种身份识别装置，所述装置包括：

数据采集模块，用于采集通讯录大数据，所述通讯录大数据包括：多个用户拥有的通讯录，每个通讯录中记录有多个身份信息对，所述身份信息对包括

5 姓名和手机号码；

信息比较模块，用于将待识别的身份信息对与所述通讯录大数据进行比对，获取信息比对结果，所述待识别的身份信息对包括：待识别的用户使用的姓名和手机号码；

风险判断模块，用于在所述信息比对结果满足风险条件时，则确定待识别
10 的所述用户是风险用户。

本发明实施例的身份识别方法和装置，通过采集通讯录大数据，建立了身份信息数据库，并将待识别的身份信息对与该身份信息数据库比对，就可以判断出某一个姓名和手机号码的身份信息对是否是真实的，从而识别到用户的身份是否是冒用，实现了对身份冒用行为的识别。

15 附图说明

图 1 为本发明实施例提供的一种身份识别方法的流程图；

图 2 为本发明实施例提供的用户的通讯录的大数据示意图；

图 3 为本发明实施例提供的另一种身份识别方法的流程图；

图 4 为本发明实施例提供的一种身份识别装置的结构示意图；

20 图 5 为本发明实施例提供的另一种身份识别装置的结构示意图。

具体实施方式

本申请实施例提供了一种身份识别方法，该方法可以用于识别身份冒用行为，比如，欺诈作弊分子冒用他人身份证号码、姓名，并利用自己掌控的手机
25 号，进行互联网场景下的账户注册、认证，用以进行信用卡申请或贷款等信贷

申请欺诈行为。为了使得即使在欺诈作弊分子绕过网络设备层面的识别模型时，也能将其识别出来，本申请考虑了一种“判断用户使用的手机号码是否是其声称的这个人正常使用的电话号码”的识别方案。

上述识别方案的基本思想是：要对客户进行身份识别的身份识别方，在获
5 取到足够多用户的通讯录后，其对所有潜在客户手机号码基本全部覆盖，组成了一个通讯录数据库，如果后续验证的客户不在这个通讯录数据库里出现，或者在这个数据库里出现时权重非常低，则是非本人操作的可能性非常大，即所验证的客户更倾向于属于身份冒用。

基于上述思想，本申请实施例提供的身份识别方法，可以参见图 1 所示的
10 流程，该方法可以包括：

在步骤 101 中，采集通讯录大数据，所述通讯录大数据包括：多个用户拥有的通讯录，每个通讯录中记录有多个身份信息对，所述身份信息对包括姓名和手机号码。

例如，通讯录大数据可以包括很多用户的通讯录数据，图 2 示例了用户 1、
15 用户 2、用户 3 直至用户 y 的通讯录数据，其中包括的通讯录的数量足够多，尽可能多的覆盖所有可能的潜在业务客户，这样才能在后续步骤用于对业务客户的身份验证。每个通讯录中都包括多个身份信息对，所述身份信息对包括姓名和手机号码。以用户 1 的通讯录为例，“姓名 N11-号码 P11”是一个身份信息对，表示姓名 N11 所代表的人使用的手机号码是 P11；“姓名 N12-号码 P12”是另一
20 个身份信息对，表示姓名 N12 所代表的人使用的手机号码是 P12。

本步骤中，通讯录数据的采集方式可以有多种，例如，可以通过运行在用户手机上的客户端软件采集用户手机上的通讯录数据。

在步骤 102 中，将待识别的身份信息对与通讯录大数据进行比对，获取信息
比对结果，待识别的身份信息对包括：待识别的用户使用的姓名和手机号码。

25 本步骤中的信息对比结果，例如，可以是通讯录大数据中是否存在与待识别的身份信息对相同的身份信息对，或者还可以是通讯录大数据中包括所述待识别的身份信息对的数量是多少，等。

在步骤 103 中，若信息比对结果满足风险条件，则确定用户是风险用户。

例如，该风险条件的设定也可以包括多种，比如，可以设定为若通讯录大数据中不存在与待识别的身份信息对相同的身份信息对时，认为该待识别的用户是风险用户；或者，若通讯录大数据中即使存在待识别的身份信息对，但是
5 包含的数量较少时，认为该待识别的用户是风险用户。

本例子的身份识别方法，通过采集通讯录大数据，建立身份信息数据库，可以根据该大数据判断出某一个姓名和手机号码的身份信息对是否是真实的，从而识别到用户的身份是否是冒用，实现了对身份冒用行为的识别。

在一个例子中，还可以按照图 3 所示的方法执行身份识别，该图 3 中的方法，将根据通讯录大数据，构建一个信息权重表，该信息权重表可以用于后续
10 对用户身份的验证。如图 3 所示，该流程可以包括：

在步骤 301 中，采集通讯录大数据。在步骤 302 中，对通讯录大数据中的身份信息对进行统计，得到每个身份信息对分别对应的信息权重，生成信息权重表。

15 本步骤中所述的信息权重，可以用于表示身份信息对的可信程度，比如，如果一个身份信息对“姓名 N11-号码 P11”在很多用户的通讯录中出现，则很大程度上该身份信息对的信息是真实的，是被很多人所认可的；反之，表示该身份信息对的可信程度较低，可能信息是伪造的。

而信息权重的计算方法可以有多种，只要能根据各个身份信息对在通讯录
20 中不同的出现情况，体现出不同身份信息对的权重差异。

例如，可以统计包括所述身份信息对的通讯录的个数，作为所述身份信息对的信息权重，假设身份信息对“姓名 N11-号码 P11”在五个用户的通讯录中出现，则对应的信息权重可以为 5，假设身份信息对“姓名 N12-号码 P12”在八个用户的通讯录中出现，则对应的信息权重可以为 8。

25 又例如，还可以根据 pagerank 方法计算每一个身份信息对的 pagerank 值，将该 pagerank 值作为身份信息对的信息权重。其中，在构建 pagerank 方法使用的 Web 图模型时，可以将每一个身份信息对分别作为一个页面节点（相当于

pagerank 中的页面节点), 该节点的出链指向所述身份信息对所属用户的通讯录中包括的另一个身份信息对, 例如, “姓名 N11-号码 P11” 这个节点所属的用户即姓名 N11 的用户, 该用户的通讯录中还包括身份信息对“姓名 N12-号码 P12”, 则 “姓名 N11-号码 P11” 节点的出链将指向节点 “姓名 N12-号码 P12”。一个页面节点的入链来自包含所述身份信息对的通讯录用户的身份信息对, 同样以上述例子来说, 对于节点 “姓名 N12-号码 P12” 来说, 其入链来自 “姓名 N11-号码 P11”, 而该 “姓名 N11-号码 P11” 节点用户的通讯录中包含 “姓名 N12-号码 P12”。在构造 Web 图模型后, 可以根据 pagerank 方法计算每一个身份信息对的 pagerank 值, 将所述 pagerank 值作为所述身份信息对的信息权重。

10 其中, 在通过 pagerank 方法计算时, 可以基于如下两个假设:

数量假设: 在 Web 图模型中, 如果一个页面节点接收到的其他网页指向的入链数量越多, 那么这个页面越重要。在本申请的例子中, 即对于某一个身份信息对来说, 如果越多的通讯录中包含该身份信息对, 则表明该身份信息对的可信程度越高。

15 质量假设: 指向页面的入链质量不同, 质量高的页面会通过链接向其他页面传递更多的权重。所以越是质量高的页面指向页面, 则页面越重要。在本申请的例子中, 即考虑了出现身份信息对的通讯录所属的用户的影响力, 身份信息对出现在一个广为熟知的公众人物的通讯录中, 相比于出现在一个不知名的普通人物的通讯录中, 其信息的真实可信程度应该有所区别。

20 经过本步骤的计算, 可以得到如下表 1 所示的信息权重表。其中需要说明的是, 在本申请的方案中, 所生成的信息权重表主要包括身份信息对与对应的信息权重, 在存储方式上可以不局限于以表的形式存储。

表 1 信息权重表

身份信息对		
姓名	号码	信息权重
N11	P11	t1

N12	P12	t2
.....		

此外，在通讯录中记录的身份信息对，可能会出现记录不规范的情况。比如，一个用户的真实姓名叫“王晓月”，结果该用户的朋友在记录其姓名和手机号码时，不小心记录成了“王小月”，即写成了错别字，本实施例对于这种情况，

5 可以进行一致性纠偏处理，即原本应该是相同的身份信息对，可能在记录时在不同的通讯录中出现了记录误差。在一个例子中，这种情况可以按照如下方式处理：在对通讯录大数据的身份信息进行信息权重的统计计算之前，在信息权重表中记录身份信息对时，“王晓月—号码 H”和“王小月—号码 H”，均统一记录为“wangxiaoyue—号码 H”，即还是将“王小月”和“王晓月”，两者作为

10 同一身份信息对，对应的信息权重可以是 2（即“wangxiaoyue—号码 H”出现两次）。在后续的将待识别的身份信息对与信息权重表比对时，根据待识别的身份信息，先根据号码找到匹配的号码 H，再将姓名转换为拼音，看是否有匹配的拼音姓名。经过上述处理，可以使得对信息权重的计算更加准确，不过，具体对于何种情况的误差才能进行一致性纠偏处理，可以根据实际业务情况或者

15 实验确定具体的策略。

此外，也可以有其他的实现方式，比如，对于上述例子中出现相同拼音不同字，且相同号码的情况，可以在信息权重表中通过记录拼音字符串的方式进行统一处理；而对于不包含错别字的情况，也还可以在信息权重表中采用汉字记录姓名，那么对于待识别的身份信息对，可以先根据号码找到匹配的号码 H，

20 再进行姓名的匹配，可以先判断是否有匹配的汉字，没有时再将姓名转换为拼音，看是否有匹配的拼音姓名，姓名和号码都匹配时，则找到了匹配的身份信息对，获取对应的信息权重即可。

在又一个例子中，在寻找匹配的身份信息对时，也可以采取容许一定范围误差的匹配方式。比如，信息权重表中记录的是“xiaoyue—号码 H”（即少了姓），

25 待识别的身份信息对是“王晓月—号码 H”，在匹配时将发现，这两个身份信息

对的号码 H 能够匹配, 是相同的, 而姓名字段中, “xiaoyue” 与 “王晓月” (具体是与王晓月转换的拼音 wangxiaoyue) 特别相似, 例如按照一定算法计算得到的相似度达到 70% 以上, 那么也可以认为 “xiaoyue” 与 “王晓月” 是匹配的。这种情况可以设定一相似度阈值, 高于该阈值时, 则尽管不完全相同, 也可以认为两者匹配。而诸如 “xiaoyue” 与 “王家惠” 这两个, 明显是不相似的, 一般会低于相似度阈值, 判定为不匹配。

在生成信息权重表的基础上, 如下的步骤中将利用该信息权重表, 用于身份信息识别。可以将待识别的身份信息对与预先生成的信息权重表进行比对, 获取信息比对结果, 该待识别的身份信息对包括: 待进行身份识别的用户使用的姓名和手机号码。若信息比对结果满足风险条件, 则确定用户是风险用户。

在步骤 303 中, 获取待识别用户的身份信息对。

例如, 可以获取正在注册用户的一些身份信息, 以用于识别该用户是否是冒用别人身份的非法分子。该身份信息可以包括: 身份证号、姓名、手机号码、地址等联系信息, 其中的姓名和手机号码, 本例子中可以称为身份信息对。

在步骤 304 中, 校验用户的身份证号和手机号码的使用权是否通过验证。

本步骤中, 可以通过公安网实名校验身份证号和姓名, 或者通过人脸与身份证关联的公安网照片做人脸比对, 或者是其他形式的校验。还可以校验用户的手机号, 以保证用户此时拥有该手机号码的使用权。

如果本步骤通过验证, 在继续执行步骤 305; 否则, 转至步骤 309。

在步骤 305 中, 查询身份信息对是否出现在信息权重表中。

如果在信息权重表中, 则继续执行步骤 306; 否则, 若所述信息权重表不包括所述待识别的身份信息对, 则转至步骤 309。

在步骤 306 中, 由信息权重表中获取对应的信息权重。

例如, 可以由预先建立的信息权重表中, 获取与步骤 303 中的身份信息对对应的信息权重。

在步骤 307 中, 判断信息权重是否大于或等于权重阈值。

假设权重阈值是 t_0 , 该权重阈值的设置, 可以依据建立信息权重表时采集

的大数据数量占全部潜在客户中的覆盖度，以及使用本身份识别方法的业务方对身份冒用风险的管控尺度等因素确定。比如，假设业务方需要严格把控用户身份，则可以将权重阈值设置的较高，保证较高的信息真实可靠度；又比如，如果采集的大数据数量占全部潜在客户中的覆盖度稍低，可以将权重阈值设置的较高，以提高信息真实可靠度。

如果本步骤的判断结果为是，则继续步骤 308；否则，执行步骤 309。

在步骤 308 中，确定待识别的用户通过验证，为合法用户。

在步骤 309 中，确定待识别的用户是风险用户。

在确定用户是风险用户后，相应的，可以定位到该用户对应的欺诈操作。

本例子的身份识别方法，通过根据通讯录大数据建立信息权重表，预先确定了各个身份信息对的可信度，并结合权重阈值，可以判断出某一个姓名和手机号码的身份信息对是否是真实的，从而识别到用户的身份是否是冒用，实现了对身份冒用行为的识别。

为了实现上述的方法，本申请实施例提供了一种身份识别装置，如图 4 所示，该装置可以包括：数据采集模块 41、信息比较模块 42 和风险判断模块 43。

数据采集模块 41，用于采集通讯录大数据，所述通讯录大数据包括：多个用户拥有的通讯录，每个通讯录中记录有多个身份信息对，所述身份信息对包括姓名和手机号码；

信息比较模块 42，用于将待识别的身份信息对与所述通讯录大数据进行比对，获取信息比对结果，所述待识别的身份信息对包括：待识别的用户使用的姓名和手机号码；

风险判断模块 43，用于在所述信息比对结果满足风险条件时，则确定待识别的所述用户是风险用户。

在一个例子中，如图 5 所示，该装置中的信息比较模块 42，可以包括：

权重统计单元 421，用于对所述通讯录大数据中的身份信息对进行统计，得到每个身份信息对分别对应的信息权重，所述信息权重用于表示身份信息对的可信程度；

权重获取单元 422，用于由统计结果中，获取对应所述待识别的身份信息对的信息权重。

在一个例子中，风险判断模块 43，具体用于：若统计结果中不存在对应所述待识别的身份信息对的信息权重；或者，若对应所述待识别的身份信息对的信息权重，低于预定的权重阈值，则确定所述待识别的用户是风险用户。

在一个例子中，权重统计单元 421，具体用于：以包括所述身份信息对的通讯录的数量，作为所述身份信息对的信息权重；或者，根据 pagerank 方法计算每一个身份信息对的 pagerank 值，将所述 pagerank 值作为所述身份信息对的信息权重。

10 在一个例子中，权重统计单元 421，还用于在对通讯录大数据中的身份信息对进行统计之前，对不同通讯录出现的同一身份信息对，进行一致性纠偏处理。

本例子的身份识别装置，通过根据通讯录大数据建立信息权重表，预先确定了各个身份信息对的可信度，并结合权重阈值，可以判断出某一个姓名和手机号码的身份信息对是否是真实的，从而识别到用户的身份是否是冒用，实现了对身份冒用行为的识别。

15 以上所述仅为本发明的较佳实施例而已，并不用以限制本发明，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明保护的范围之内。