



(12) 发明专利

(10) 授权公告号 CN 109559245 B

(45) 授权公告日 2022. 02. 25

(21) 申请号 201710882135.9
 (22) 申请日 2017.09.26
 (65) 同一申请的已公布的文献号
 申请公布号 CN 109559245 A
 (43) 申请公布日 2019.04.02
 (73) 专利权人 北京国双科技有限公司
 地址 100080 北京市海淀区北四环中路229号海泰大厦4层南401号
 (72) 发明人 郑立颖
 (74) 专利代理机构 北京集佳知识产权代理有限公司 11227
 代理人 王宝筠
 (51) Int. Cl.
 G06Q 50/00 (2012.01)

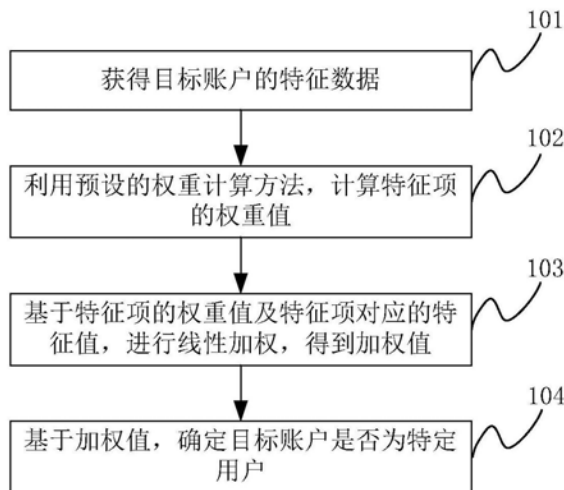
(56) 对比文件
 CN 106940732 A, 2017.07.11
 CN 102855440 A, 2013.01.02
 CN 103198161 A, 2013.07.10
 CN 103617235 A, 2014.03.05
 CN 105611084 A, 2016.05.25
 CN 106657062 A, 2017.05.10
 CN 106557663 A, 2017.04.05
 审查员 杨怡睿

权利要求书2页 说明书9页 附图3页

(54) 发明名称
 一种识别特定用户的方法及装置

(57) 摘要

本发明公开了一种识别特定用户的方法及装置,方法包括:获得目标账户的特征数据,所述特征数据包括至少一个特征项及每个所述特征项对应的特征值;利用预设的权重计算方法,计算所述特征项的权重值;基于所述特征项的权重值及所述特征项对应的特征值,进行线性加权,得到加权值;基于所述加权值,确定所述目标账户是否为特定用户。本申请利用特征权重计算的方案来识别特定用户,如网络水军等,而区别于现有技术中人工标注的方案,能够节省识别的时间,提高识别的准确率,达到有效识别特定用户的目的。



1. 一种识别特定用户的方法,其特征在于,包括:

获得目标账户的特征数据,所述特征数据包括至少一个特征项及每个所述特征项对应的特征值;

计算所述特征项在不同账户上所对应的特征值的标准差;

计算不同的特征项所对应的特征值之间的关联度;

基于所述标准差与所述关联度,计算所述特征项的权重值;

基于所述特征项的权重值及所述特征项对应的特征值,进行线性加权,得到加权值;

基于所述加权值,确定所述目标账户是否为特定用户;

其中,所述获得目标账户的特征数据,包括:

采集目标账户对其他账户进行操作的特征项及每个特征项对应的特征属性;

对所述特征项对应的特征属性进行数值化处理,得到所述特征项对应的特征值;

基于所述加权值,确定所述目标账户是否为特定用户,包括:

判断所述加权值是否小于或等于预设的第一阈值,如果所述加权值小于或等于所述第一阈值,确定所述目标账户为特定用户;

或者,

将各目标账户的加权值按照从小到大进行排序,再将加权值排名比例在前预设的第二阈值的目标账户确定为特定用户。

2. 根据权利要求1所述的方法,其特征在于,基于所述特征项的权重值及所述特征项对应的特征值,进行线性加权,得到加权值,包括:

将所述特征项的权重值乘以该特征项所对应的特征值,再将所得到的乘积相加,得到加权值。

3. 一种识别特定用户的装置,其特征在于,包括:

数据获得单元,用于获得目标账户的特征数据,所述特征数据包括至少一个特征项及每个所述特征项对应的特征值;

权重计算单元,用于利用预设的权重计算方法,计算所述特征项的权重值;

特征加权单元,基于所述特征项的权重值及所述特征项对应的特征值,进行线性加权,得到加权值;

用户确定单元,用于基于所述加权值,确定所述目标账户是否为特定用户;

其中,所述权重计算单元,具体用于:计算所述特征项在不同账户上所对应的特征值的标准差,并计算不同的特征项所对应的特征值之间的关联度,基于所述标准差与所述关联度,计算所述特征项的权重值;

所述数据获得单元具体用于:采集目标账户对其他账户进行操作的特征项及每个特征项对应的特征属性,对所述特征项对应的特征属性进行数值化处理,得到所述特征项对应的特征值;

所述用户确定单元,具体用于:基于所述加权值,确定所述目标账户是否为特定用户,包括:

判断所述加权值是否小于或等于预设的第一阈值,如果所述加权值小于或等于所述第一阈值,确定所述目标账户为特定用户;

或者

将各目标账户的加权值按照从小到大进行排序,再将加权值排名比例在前预设的第二阈值的账户确定为特定用户。

4. 根据权利要求3所述的装置,其特征在于,所述权重计算单元具体用于:计算所述特征项在不同账户上所对应的特征值的标准差,并计算不同的特征项所对应的特征值之间的关联度,基于所述标准差与所述关联度,计算所述特征项的权重值。

5. 一种存储介质,其特征在于,所述存储介质包括存储的程序,其中,所述程序执行权利要求1至2中任意一项所述的识别特定用户的方法。

6. 一种处理器,其特征在于,所述处理器用于运行程序,其中,所述程序运行时执行权利要求1至2中任意一项所述的识别特定用户的方法。

一种识别特定用户的方法及装置

技术领域

[0001] 本发明涉及数据处理技术领域,特别涉及一种识别特定用户的方法及装置。

背景技术

[0002] 随着社交媒体的迅速发展,越来越多的用户开始使用微博、论坛等媒体来进行沟通。在用户量急剧增长的情况下,各大品牌广告也将广告从传统的搜索引擎转向了越来越热的社交媒体,如通过各种知名社交应用账号来传播商品信息,传播效果则是靠社交应用账号的粉丝点赞数、评论数或转发数量等来衡量,因此,社交应用账号可能会为了更高的广告费,而使用网路水军为账号进行虚假的点赞、转发及评论来扩大传播效果。

[0003] 但是,对于广告品牌来说,需要获得真实用户的社交互动量,则需要对特定用户进行识别,如真爱粉、黑粉、水军等账号进行识别,再真实衡量广告效果及社交应用账号的影响力。

[0004] 因此,目前亟需一种能够有效识别特定用户的方案。

发明内容

[0005] 鉴于上述问题,提出了本发明以便提供一种识别特定用户的方法及装置,用以解决现有技术中无法有效识别特定用户的技术问题。

[0006] 本申请提供了一种识别特定用户的方法,包括:

[0007] 获得目标账户的特征数据,所述特征数据包括至少一个特征项及每个所述特征项对应的特征值;

[0008] 利用预设的权重计算方法,计算所述特征项的权重值;

[0009] 基于所述特征项的权重值及所述特征项对应的特征值,进行线性加权,得到加权值;

[0010] 基于所述加权值,确定所述目标账户是否为特定用户。

[0011] 上述方法,优选的,获得目标账户的特征数据,包括:

[0012] 采集目标账户对其他账户进行操作的特征项及每个特征项对应的特征属性;

[0013] 对所述特征项对应的特征属性进行数值化处理,得到所述特征项对应的特征值。

[0014] 上述方法,优选的,利用预设的权重计算方法,计算所述特征项的权重值,包括:

[0015] 计算所述特征项在不同账户上所对应的特征值的标准差;

[0016] 计算不同的特征项所对应的特征值之间的关联度;

[0017] 基于所述标准差与所述关联度,计算所述特征项的权重值。

[0018] 上述方法,优选的,基于所述特征项的权重值及所述特征项对应的特征值,进行线性加权,得到加权值,包括:

[0019] 将所述特征项的权重值乘以该特征项所对应的特征值,再将所得到的乘积相加,得到加权值。

[0020] 上述方法,优选的,基于所述加权值,确定所述目标账户是否为特定用户,包括:

[0021] 判断所述加权值是否小于或等于预设的第一阈值,如果所述加权值小于或等于所述第一阈值,确定所述目标账户为特定用户;

[0022] 或者

[0023] 将各目标账户的加权值按照从小到大进行排序,再将加权值排名比例在前预设的第二阈值的账户确定为特定用户。

[0024] 本申请还提供了一种识别特定用户的装置,包括:

[0025] 数据获得单元,用于获得目标账户的特征数据,所述特征数据包括至少一个特征项及每个所述特征项对应的特征值;

[0026] 权重计算单元,用于利用预设的权重计算方法,计算所述特征项的权重值;

[0027] 特征加权单元,基于所述特征项的权重值及所述特征项对应的特征值,进行线性加权,得到加权值;

[0028] 用户确定单元,用于基于所述加权值,确定所述目标账户是否为特定用户。

[0029] 上述装置,优选的,所述数据获得单元具体用于:采集目标账户对其他账户进行操作的特征项及每个特征项对应的特征属性,对所述特征项对应的特征属性进行数值化处理,得到所述特征项对应的特征值。

[0030] 上述装置,优选的,所述权重计算单元具体用于:计算所述特征项在不同账户上所对应的特征值的标准差,并计算不同的特征项所对应的特征值之间的关联度,基于所述标准差与所述关联度,计算所述特征项的权重值。

[0031] 上述装置,优选的,所述特征加权单元具体用于:将所述特征项的权重值乘以该特征项所对应的特征值,再将所得到的乘积相加,得到加权值。

[0032] 上述装置,优选的,所述水军确定单元具体用于:判断所述加权值是否小于或等于预设的第一阈值,如果所述加权值小于或等于所述第一阈值,确定所述目标账户为水军;

[0033] 或者,所述水军确定单元具体用于:将各目标账户的加权值按照从小到大进行排序,再将加权值排名比例在前预设的第二阈值的账户确定为水军。

[0034] 本申请还提供了一种存储介质,所述存储介质包括存储的程序,其中,所述程序执行上述任意一项所述的识别特定用户的方法。

[0035] 本申请还提供了一种处理器,所述处理器用于运行程序,其中,所述程序运行时执行上述任意一项所述的识别特定用户的方法。

[0036] 借由上述技术方案,本发明提供了一种识别特定用户的方法及装置,通过在获得目标账户的各种特征项及特征值等特征数据之后,计算特征项的权重值,再对这些权重值进行特征加权得到加权值,最终基于该加权值来判断出目标账户是否为特定用户,从而实现了对特定用户的识别。本申请利用特征权重计算的方案来识别特定用户,而区别于现有技术中人工标注的方案,能够节省识别的时间,提高识别的准确率,达到有效识别特定用户的目的。

[0037] 上述说明仅是本发明技术方案的概述,为了能够更清楚了解本发明的技术手段,而可依照说明书的内容予以实施,并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂,以下特举本发明的具体实施方式。

附图说明

[0038] 通过阅读下文优选实施方式的详细描述,各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的,而并不认为是对本发明的限制。而且在整个附图中,用相同的参考符号表示相同的部件。在附图中:

[0039] 图1示出了本申请实施例提供的一种识别特定用户的方法的流程图;

[0040] 图2示出了本申请实施例的应用示例图;

[0041] 图3及图4分别示出了本申请实施例提供的一种识别特定用户的方法的部分流程图;

[0042] 图5示出了本申请实施例的另一应用示例图;

[0043] 图6示出了本申请实施例提供的一种识别特定用户的装置的结构示意图;

[0044] 图7示出了本申请实施例提供的一种存储介质的结构示意图;

[0045] 图8示出了本申请实施例提供的一种处理器的结构示意图;

[0046] 图9示出了本申请实施例提供的一种设备的结构示意图。

具体实施方式

[0047] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例,然而应当理解,可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反,提供这些实施例是为了能够更透彻地理解本公开,并且能够将本公开的范围完整的传达给本领域的技术人员。

[0048] 参考图1,为本申请实施例提供的一种识别特定用户的方法的实现流程图,适用于对各种社交应用账号的特定用户进行有效识别,例如,对社交应用账户的水军点赞、水军评论、水军转发等进行识别。

[0049] 具体实现中,本实施例的方法可以包括以下步骤:

[0050] 步骤101:获得目标账户的特征数据。

[0051] 其中,该目标账户是指需要识别是否为特定用户的社交应用账户,如微博账户等,这些账户可以通过移动客户端或者非移动客户端推送状态,另外,这些目标账户可以对其他社交应用账户进行关注成为账户粉丝,以及对其他社交应用账户推送的状态进行点赞、评论、转发等。

[0052] 而目标账户的特征数据包括至少一个特征项及每个特征项对应的特征值,这里的特征项可以为:目标账户的被关注项、目标账户的关注项、目标账户推送的状态项、目标账户推送状态的客户端属性项及目标账户被社交应用官方认证的状态等,相应的特征值是指:目标账户的被关注数、目标账户的关注数、目标账户推送的状态数量、目标账户使用移动客户端推送状态的次数及目标账户是否被社交应用官方认证等。

[0053] 步骤102:利用预设的权重计算方法,计算特征项的权重值。

[0054] 其中,权重值可以理解为:特征项在表明目标账户为特定用户如水军的可能性上所占的分量,例如:权重值越高的特征项的值越高,则说明目标账户为特定用户的可能性越大。

[0055] 步骤103:基于特征项的权重值及特征项对应的特征值,进行线性加权,得到加权值。

[0056] 在一种实现方式中,可以将特征项的权重值乘以该特征项所对应的特征值,得到多个乘积,再将这些乘积相加,得到加权值。如,特征项 $X_1 \sim X_n$, n 为大于2的正整数,而相应的特征值分别为 $y_1 \sim y_n$,本实施例中,将 X_1 乘以 y_1 、 X_2 乘以 y_2 、 \dots 、 X_n 乘以 y_n ,如图2中所示,加权值 Z 则为这些乘积之和。

[0057] 步骤104:基于加权值,确定目标账户是否为特定用户。

[0058] 在一种实现方式中,本实施例中可以根据预设的加权判定方法对加权值进行判断,从而判断出目标账户是否为特定用户。例如,本实施例首先判断加权值是否小于或等于预设的第一阈值,如果加权值小于或等于第一阈值,则可以确定目标账户为特定用户。

[0059] 其中,第一阈值可以根据需求设置。

[0060] 或者,本实施例中先将各个目标账户的加权值按照从小到大进行排序,再将加权值排名比例在前预设的第二阈值的账户确定为特定用户。例如,设置第二阈值为20%,将加权值最低的20%的账户确定为特定用户。

[0061] 由上述技术方案,本申请实施例提供的一种识别特定用户的方法,通过在获得目标账户的各种特征项及特征值等特征数据之后,计算特征项的权重值,再对这些权重值进行特征加权得到加权值,最终基于该加权值来判断出目标账户是否为特定用户,从而实现特定用户如水军的识别。本实施例利用特征权重计算的方案来识别特定用户,而区别于现有技术中人工标注的方案,能够节省识别的时间,提高识别的准确率,达到有效识别特定用户的目的。

[0062] 在一种实现方式中,本实施例中在获得目标账户的特征数据时,可以通过以下步骤实现,如图3中所示:

[0063] 步骤301:采集目标账户对其他账户进行操作的特征项及每个特征项对应的特征属性。

[0064] 本实施例中,可以利用网页爬虫获得目标账户的特征项及各自对应的特征属性,或者可以在社交应用的后台抽取目标账户的特征项及各特征项对应的特征属性。例如,抽取社交应用的各个用户的特征数据,如粉丝数、关注数、推送数、移动推送比例、是否认证等。

[0065] 步骤302:对特征项对应的特征属性进行数值化处理,得到特征项对应的特征值。

[0066] 例如,本实施例中对特征属性进行数据清洗、特征归一化等处理,例如,将被官方认证的特征属性转换为1的值,将未被官方认证的特征属性转换为0的值,最终得到特征值。

[0067] 其中,数据清洗可以采用min-max标准化的方法,也成为离差标准化,是对原始数据的线性变换,使得结果值映射到[0-1]之间,计算公式(1)如下:

$$[0068] \quad x^* = \frac{x - \min}{\max - \min}$$

[0069] 其中, x 即为需要进行数据清洗的数据。

[0070] 在一种实现方式中,本实施例中在利用预设的权重计算方法,计算特征项的权重值时,可以采用CRITIC方法(Criteria Importance Though Intercriteria Correlation)计算权重,具体可以通过以下步骤实现,如图4中所示:

[0071] 步骤401:计算特征项在不同账户上所对应的特征值的标准差。

[0072] 例如,计算目标账户的点赞在不同账户上所对应的点赞数的标准差。

[0073] 步骤402:计算不同特征项所对应的特征值之间的关联度。

[0074] 例如,计算点赞与转发所对应的点赞数与转发数之间的关联度。

[0075] 步骤403:基于标准差与关联度,计算特征项的权重值。

[0076] 其中,在CRITIC方法中,其指标(特征项)的客观权重以两个基本概念为主,一是对比强度,它表示了同一指标值之间取值差距的大小,以标准差的形式来表现,标准差越大各方案的取值差距越大,应赋予较高权重;二是指标之间的冲突性,以指标之间的相关性为基础,如果两个指标之间具有较强的正相关,说明两个指标冲突性较低,则指标反应的信息量相似,应赋予较小的权重。

[0077] CRITIC方法的具体计算步骤如下:

[0078] Step1:计算指标标准差:

[0079] 假设有一组数值 x_1 、 x_2 、 x_3 ... x_n (皆为实数),其平均值(算术平均值)为 μ ,公式如下:

$$[0080] \quad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

[0081] Step2:计算指标冲突性:

[0082] 指标之间的冲突性是以指标之间的相关性为基础,如两个指标之间具有较强的正相关,说明两个指标冲突性较低。

[0083] 第 j 个指标与其他指标的冲突性计算公式为:

$$[0084] \quad \sum_{i=1}^n (1 - r_{ij}), \text{其中 } r_{ij} \text{ 是指标 } i \text{ 和 } j \text{ 的相关系数}$$

[0085] Step3:求指标的critic系数,并归一化后获得指标权重

$$[0086] \quad C_j = \sigma_j \sum_{i=1}^n (1 - r_{ij}), \text{其中 } \sigma_j \text{ 是指标 } j \text{ 的标准差}$$

[0087] Step4:求得归一化后的权重

$$[0088] \quad W_j = \frac{C_j}{\sum_{j=1}^n C_j}$$

[0089] 本实施例中利用CRITIC方法来计算目标账户各特征项的权重值,如图5中所示,以对水军识别为例,首先抽取社交应用的用户特征数据,例如粉丝数、关注数、推送数、移动推送占比、是否认证等,在对特征数据进行数据清洗、特征构建如特征数据归一化之后,使用critic方法针对以上特征计算出权重,各特征数据针对权重进行线性加权得出最终用户得分即加权值,得分最低的20%可以判别为水军。以上方案中,本实施例利用无监督方法给社交应用账号打分,打分越低的则越可能为水军,省去了有监督方法中的大量人工标注过程。而相比于用规则单一特征筛选水军账号的方法,更全面的考虑水军的各方面特征,综合各个特征值来判断水军账号的得分。

[0090] 参考图6,为本申请实施例提供的一种识别特定用户的装置的结构示意图,适用于对各种社交应用账号的特定用户进行有效识别,例如,对社交应用账户的水军点赞、水军评论、水军转发等进行识别。

[0091] 具体实现中,本实施例的装置可以包括以下结构:

[0092] 数据获得单元601,用于获得目标账户的特征数据。

[0093] 其中,该目标账户是指需要识别是否为特定用户的社交应用账户,如微博账户等,这些账户可以通过移动客户端或者非移动客户端推送状态,另外,这些目标账户可以对其他社交应用账户进行关注成为账户粉丝,以及对其他社交应用账户推送的状态进行点赞、评论、转发等。

[0094] 而目标账户的特征数据包括至少一个特征项及每个特征项对应的特征值,这里的特征项可以为:目标账户的被关注项、目标账户的关注项、目标账户推送的状态项、目标账户推送状态的客户端属性项及目标账户被社交应用官方认证的状态等,相应的特征值是指:目标账户的被关注数、目标账户的关注数、目标账户推送的状态数量、目标账户使用移动客户端推送状态的次数及目标账户是否被社交应用官方认证等。

[0095] 在一种实现方式中,数据获得单元601具体用于:采集目标账户对其他账户进行操作的特征项及每个特征项对应的特征属性,对所述特征项对应的特征属性进行数值化处理,得到所述特征项对应的特征值。

[0096] 本实施例中,可以利用网页爬虫获得目标账户的特征项及各自对应的特征属性,或者可以在社交应用的后台抽取目标账户的特征项及各特征项对应的特征属性。例如,抽取社交应用的各个用户的特征数据,如粉丝数、关注数、推送数、移动推送比例、是否认证等。例如,本实施例中对特征属性进行数据清洗、特征归一化等处理,例如,将被官方认证的特征属性转换为1的值,将未被官方认证的特征属性转换为0的值,最终得到特征值。

[0097] 权重计算单元602,用于利用预设的权重计算方法,计算所述特征项的权重值。

[0098] 其中,权重值可以理解为:特征项在表明目标账户为特定用户如水军的可能性上所占的分量,例如:权重值越高的特征项的值越高,则说明目标账户为特定用户的可能性越大。

[0099] 在一种实现方式中,权重计算单元602具体用于:计算所述特征项在不同账户上所对应的特征值的标准差,并计算不同的特征项所对应的特征值之间的关联度,基于所述标准差与所述关联度,计算所述特征项的权重值。

[0100] 例如,计算目标账户的点赞在不同账户上所对应的点赞数的标准差,再计算点赞与转发所对应的点赞数与转发数之间的关联度。

[0101] 其中,在CRITIC方法中,其指标(特征项)的客观权重以两个基本概念为主,一是对比强度,它表示了同一指标值之间取值差距的大小,以标准差的形式来表现,标准差越大各方案的取值差距越大,应赋予较高权重;二是指标之间的冲突性,以指标之间的相关性为基础,如果两个指标之间具有较强的正相关,说明两个指标冲突性较低,则指标反应的信息量相似,应赋予较小的权重。

[0102] 特征加权单元603,基于所述特征项的权重值及所述特征项对应的特征值,进行线性加权,得到加权值。

[0103] 在一种实现方式中,特征加权单元603可以将特征项的权重值乘以该特征项所对应的特征值,得到多个乘积,再将这些乘积相加,得到加权值。如,特征项 $X_1 \sim X_n$, n 为大于2的正整数,而相应的特征值分别为 $y_1 \sim y_n$,本实施例中,将 X_1 乘以 y_1 、 X_2 乘以 y_2 、 \dots 、 X_n 乘以 y_n ,如图2中所示,加权值 Z 则为这些乘积之和。

[0104] 用户确定单元604,用于基于所述加权值,确定所述目标账户是否为特定用户。

[0105] 在一种实现方式中,用户确定单元604可以根据预设的加权判定方法对加权值进行判断,从而判断出目标账户是否为特定用户。例如,本实施例首先判断加权值是否小于或等于预设的第一阈值,如果加权值小于或等于第一阈值,则可以确定目标账户为特定用户如水军。

[0106] 其中,第一阈值可以根据需求设置。

[0107] 或者,用户确定单元604可以先将各个目标账户的加权值按照从小到大进行排序,再将加权值排名比例在前预设的第二阈值的账户确定为特定用户。例如,设置第二阈值为20%,将加权值最低的20%的账户确定为特定用户。

[0108] 本实施例中利用CRITIC方法来计算目标账户各特征项的权重值,如图5中所示,以对水军识别为例,首先抽取社交应用的用户特征数据,例如粉丝数、关注数、推送数、移动推送占比、是否认证等,在对特征数据进行数据清洗、特征构建如特征数据归一化之后,使用critic方法针对以上特征计算出权重,各特征数据针对权重进行线性加权得出最终用户得分即加权值,得分最低的20%可以判别为水军。以上方案中,本实施例利用无监督方法给社交应用账号打分,打分越低的则越可能为水军,省去了有监督方法中的大量人工标注过程。而相比于用规则单一特征筛选水军账号的方法,更全面的考虑水军的各方面特征,综合各个特征值来判断水军账号的得分。

[0109] 在本实施例中,该识别装置可以包括有处理器及存储器,处理器及存储器为服务器等承载以上本实施例的设备上的元器件,上述的数据获得单元601、权重计算单元602、特征加权单元603及用户确定单元604等均作为程序单元存储在存储器中,由处理器执行存储在存储器中的上述程序单元来实现相应的功能。

[0110] 例如,上述各个程序单元以安装包或处理类的形式存储在存储器中,同时存储器中还存储有预先设置的配置文件,处理器通过调用安装包处理类,来执行以上各个程序单元,实现相应的功能。

[0111] 具体的,处理器中包含内核,由内核去存储器中调取相应的程序单元,内核可以设置一个或多个,通过调整内核参数来获得目标账户的特征数据,再利用预设的权重计算方法,计算特征数据中特征项的权重值,基于所述特征项的权重值及所述特征项对应的特征值,进行线性加权,得到加权值之后,基于加权值,确定目标账户是否为特定用户。

[0112] 其中,存储器可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM),存储器包括至少一个存储芯片。

[0113] 由上述技术方案,本申请实施例提供的一种识别特定用户的装置,通过在获得目标账户的各种特征项及特征值等特征数据之后,计算特征项的权重值,再对这些权重值进行特征加权得到加权值,最终基于该加权值来判断出目标账户是否为特定用户,从而实现特定用户如水军的识别。本实施例利用特征权重计算的方案来识别特定用户,而区别于现有技术中人工标注的方案,能够节省识别的时间,提高识别的准确率,达到有效识别特定用户的目的。

[0114] 本发明实施例提供了一种存储介质,如图7中所示,该存储介质处于设备上,而该存储介质上存储有程序,该程序运行时控制该存储介质所在的设备执行图1至图5中所描述

或展示的识别特定用户的方法。

[0115] 本发明实施例提供了一种处理器,如图8中所示,该处理器用于运行程序,该程序可以为图7所示的存储介质中存储的程序,由此,处理器通过运行程序执行图1至图5中所描述或展示的识别特定用户的方法。

[0116] 本发明实施例提供了一种设备,如图9中设备的结构示意图所示,该设备中可以包括处理器、存储器及存储在存储器上并可在处理器上运行的应用程序,处理器执行应用程序时实现以下步骤:获得目标账户的特征数据,所述特征数据包括至少一个特征项及每个所述特征项对应的特征值;利用预设的权重计算方法,计算所述特征项的权重值;基于所述特征项的权重值及所述特征项对应的特征值,进行线性加权,得到加权值;基于所述加权值,确定所述目标账户是否为特定用户,如水军等。

[0117] 其中,本文中的设备可以是服务器、PC、PAD、手机等。

[0118] 本申请还提供了一种计算机程序产品,当在数据处理设备上执行时,适于执行初始化有如下方法步骤的程序代码:获得目标账户的特征数据,所述特征数据包括至少一个特征项及每个所述特征项对应的特征值;利用预设的权重计算方法,计算所述特征项的权重值;基于所述特征项的权重值及所述特征项对应的特征值,进行线性加权,得到加权值;基于所述加权值,确定所述目标账户是否为特定用户,如水军等。

[0119] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0120] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0121] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0122] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0123] 在一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0124] 存储器可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。存储器是计算机可读介

质的示例。

[0125] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器 (CD-ROM)、数字多功能光盘 (DVD) 或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括暂存电脑可读媒体 (transitory media),如调制的数据信号和载波。

[0126] 还需要说明的是,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0127] 本领域技术人员应明白,本申请的实施例可提供为方法、系统或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0128] 以上仅为本申请的实施例而已,并不用于限制本申请。对于本领域技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等,均应包含在本申请的权利要求范围之内。

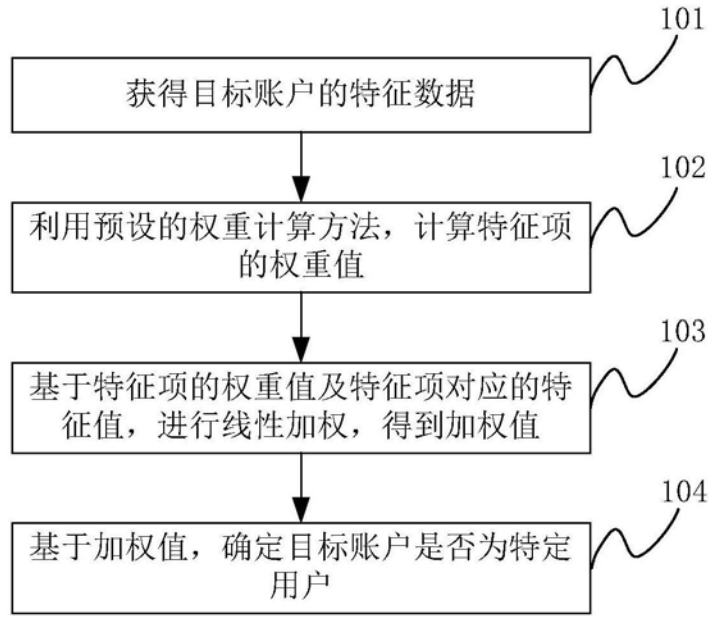


图1

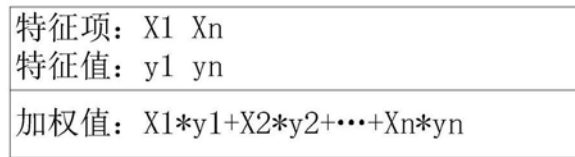


图2

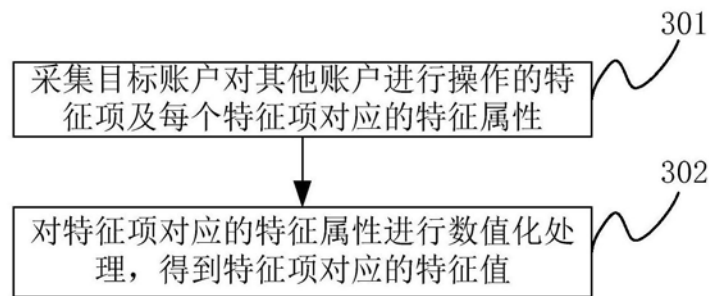


图3

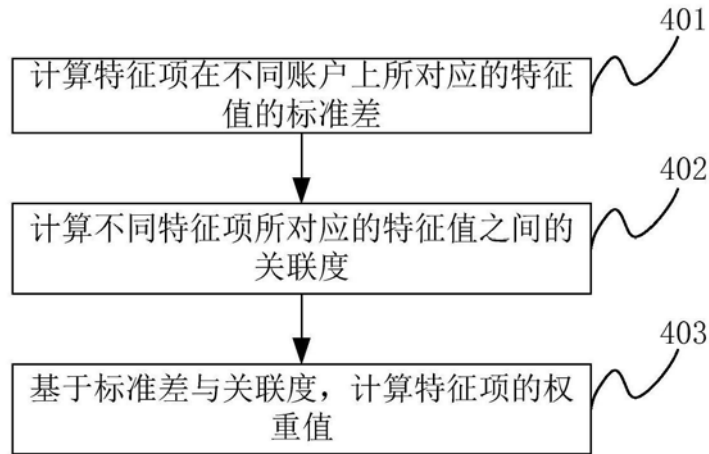


图4

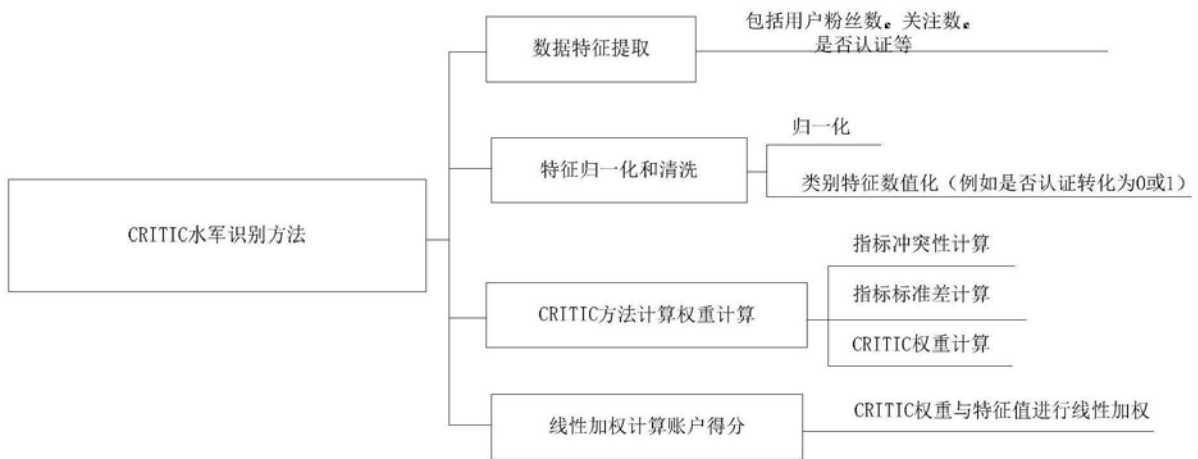


图5

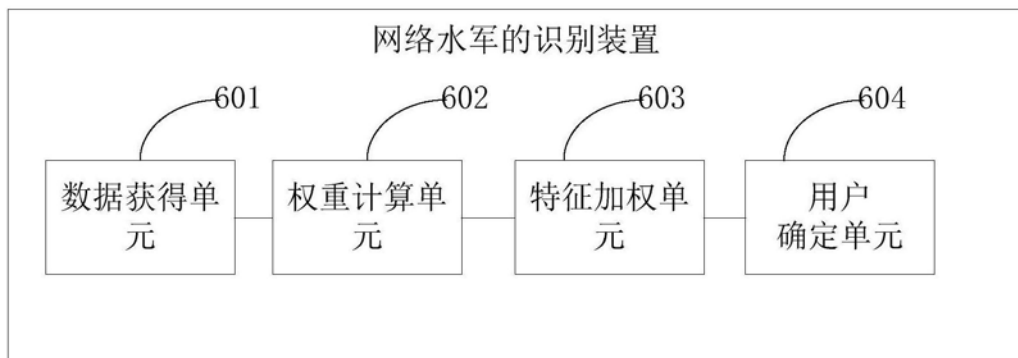


图6

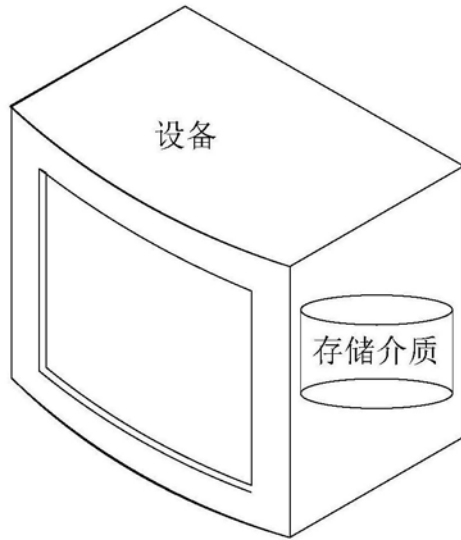


图7

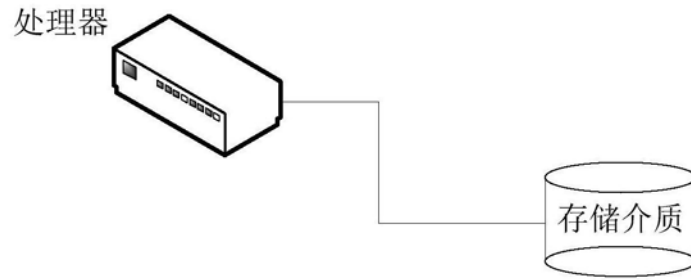


图8

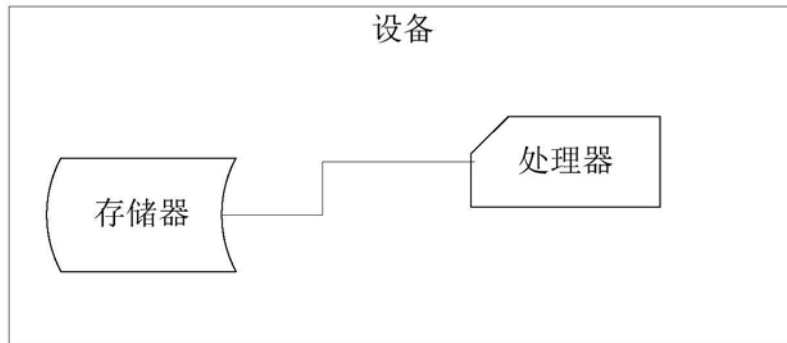


图9