



(12) 发明专利申请

(10) 申请公布号 CN 112182654 A

(43) 申请公布日 2021.01.05

(21) 申请号 202011048516.5

(22) 申请日 2020.09.29

(71) 申请人 浙江鸿程计算机系统有限公司

地址 310023 浙江省杭州市余杭区五常街
道溪沁街10号

(72) 发明人 吴勇 严伟强 钟宏泽 王凯
李纺 邱龙 应悦

(74) 专利代理机构 杭州之江专利事务所(普通
合伙) 33216

代理人 张慧英

(51) Int.Cl.

G06F 21/62 (2013.01)

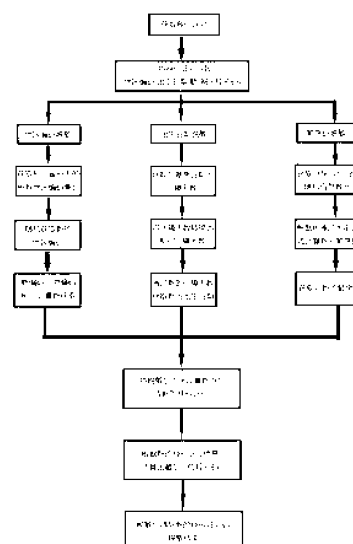
权利要求书2页 说明书4页 附图1页

(54) 发明名称

一种保留统计级特征属性的身份证号脱敏方法

(57) 摘要

本发明涉及数据处理领域,尤其涉及一种保留统计级特征属性的身份证号脱敏方法,包括以下步骤:身份证号进行分段;分别对地区编码、出生日期、顺序码进行脱敏;获取相同省份下的所有地区编码集和,从集合中获取新的地址编码;获取基准日期与待脱敏出生日期的间隔天数,间隔天数通过转换得到新的间隔天数,新的间隔天数进制转换后与一个随机数进行模运算,得到最终的间隔天数,根据最终的间隔天数得到新的出生日期;通过与一随机整数进行计算得到新的顺序码;根据固定算法得到新的校验码;组合形成新的身份证号。本发明的有益效果在于:采用分段处理机制,可以灵活选择要脱敏的分段信息,同时保持身份证号本身的结构特征和分析价值。



1. 一种保留统计级特征属性的身份证号脱敏方法,其特征在于包括以下步骤:

1) 提取待脱敏的身份证号,将身份证号分为待脱敏地区编码、待脱敏出生日期、待脱敏顺序码和校验码;

2) 对待脱敏地区编码、待脱敏出生日期、待脱敏顺序码和待脱敏校验码进行脱敏,

(1) 待脱敏地区编码进行脱敏的步骤:

① 加载全国所有地区编码信息(港澳台除外),取地区编码前2位作为Key值,相同的Key值放置到对应的List集合,若干List集和构成Map集合;

② 获取待脱敏地区编码的前2位,在Map集合中根据Key值,找到对应的List集合,并对List集合中的地区编码进行随机置乱;

③ 获取List集合中首个地区编码,作为新的地区编码,若新的地区编码和待脱敏地区编码相同,则重复上面的步骤②,直到获取到新的地区编码;

(2) 待脱敏出生日期进行脱敏的步骤:获取基准日期与待脱敏出生日期的间隔天数,间隔天数通过转换得到新的间隔天数,新的间隔天数进制转换后与一个随机数进行模运算,得到最终的间隔天数,根据最终的间隔天数得到新的出生日期;

(3) 待脱敏顺序码进行脱敏的步骤:获取待脱敏顺序码sequence,根据公式计算新的顺序码newSequence,如果顺序码超过3位,则去掉最高位,所述n为随机整数:

$$\text{newSequence} = \text{sequence} + 2 * n;$$

(4) 根据身份证号的固定算法得到新的校验码;

3) 新的地区编码、新的出生日期、新的顺序码和新的校验码依次组合成新的身份证号。

2. 根据权利要求1所述的一种保留统计级特征属性的身份证号脱敏方法,其特征在于,所述待脱敏地区编码为待脱敏身份证的第1-6位、待脱敏出生日期为待脱敏身份证的第7-14位、待脱敏顺序码为待脱敏身份证的第15-17位,校验码为待脱敏身份证的第18位。

3. 根据权利要求2所述的一种保留统计级特征属性的身份证号脱敏方法,其特征在于,所述新的间隔天数newGapDay的计算公式如下:

$$\text{newGapDay} = (\text{baseDate} - \text{birthDate}) - (32760 * \text{flag})$$

$$\text{flag} = (\text{baseDate} - \text{birthDate}) / 32760$$

其中baseDate为基准日期,birthDate为待脱敏出生日期,flag取其商值并取整。

4. 根据权利要求3所述的一种保留统计级特征属性的身份证号脱敏方法,其特征在于,所述新的间隔天数转换成5位八进制数 $G_1G_2G_3G_4G_5$,高位不够用0补齐;所述随机数为5位八进制数,记为 $A_1A_2A_3A_4A_5$ 。

5. 根据权利要求4所述的一种保留统计级特征属性的身份证号脱敏方法,其特征在于,所述模运算得到5位八进制数 $D_1D_2D_3D_4D_5$,计算公式如下:

$$\text{计算 } D_5 = (G_5 + A_5) \% 8$$

$$\text{计算 } D_4 = (((G_4 + A_4) \% 8) + D_5) \% 8$$

$$\text{计算 } D_3 = (((G_3 + A_3) \% 8) + D_4) \% 8$$

$$\text{计算 } D_2 = (((G_2 + A_2) \% 8) + D_3) \% 8$$

$$\text{计算 } D_1 = (((G_1 + A_1) \% 8) + D_2) \% 8。$$

6. 根据权利要求5所述的一种保留统计级特征属性的身份证号脱敏方法,其特征在于,所述 $D_1D_2D_3D_4D_5$ 转换为十进制数tempDay

$\text{tempDay} = D_1 * 8^4 + D_2 * 8^3 + D_3 * 8^2 + D_4 * 8 + D_5$

计算最后的间隔天数ResultDay

$\text{ResultDay} = \text{tempDay} + (32760 * \text{flag})$

得到新的出生日期 $\text{newBirthDay} = \text{baseDate} - \text{ResultDay}$ 。

7. 根据权利要求6所述的一种保留统计级特征属性的身份证号脱敏方法, 其特征在于, 所述n为[0, 50) 以内的随机整数。

一种保留统计级特征属性的身份证号脱敏方法

技术领域

[0001] 本发明涉及数据处理领域,尤其涉及一种保留统计级特征属性的身份证号脱敏方法。

背景技术

[0002] 随着互联网的发展,信息传递越来越方便。数据信息已经开始慢慢成为一个企业的核心竞争力,已经成为了一种新型的战略资源。未来,数据信息将是企业的核心资产,对企业的重要性不言而喻。数据即价值,这已是公认的产业趋势。正因为数据蕴含着巨大的商业价值,关于数据盗窃的案件越来越多,犯罪手法多种多样。所以加强敏感信息的保护也就变得越来越重要,尤其是一些关于个人隐私的敏感信息,如:姓名,身份证号,手机号,银行卡号等。

[0003] 现阶段对于敏感信息的主要保护手段是对敏感信息进行脱敏处理。数据的脱敏,又叫数据漂白或数据去隐私化,指我们对某些敏感信息通过一定的脱敏规则和脱敏手段,对敏感数据进行覆盖隐藏或变形等方法,实现对敏感数据的替换,从而起到保护敏感数据的作用。

[0004] 身份证号可以说是一种很重要的敏感信息,因为里面包含了用户很多隐私信息,所以关于身份证号的脱敏方法也有很多。现有技术中,对于身份证号进行脱敏时,常用的手段是覆盖和替换。

[0005] 覆盖即将身份证号的某一段数字或全部数字用特殊符号覆盖,以达到保护信息的效果。替换又分为随机替换和固定替换,固定替换即将身份证号中的地址编号和出生日期用一个固定的地址编号和日期替换,这样所有的身份证号的地址编号和出生日期都是相同的,这样也可以达到保护信息的效果。随机替换即将身份证号中的地址编号和出生日期进行随机的替换,改变身份证号信息以达到保护信息的效果。但是这些脱敏方法在对身份证号进行脱敏后,身份证号也就失去了它本身的特征,同时也意味着数据失去了分析价值。

[0006] 但是在某些场景下,我们需要对身份证号信息进行统计分析,但是我们又不能把真实的身份证号信息给客户分析,这样很容易造成信息泄露。但是对身份证号脱敏后,数据又会失去分析价值。

[0007] 比如固定替换后,所有身份证号基本相同,随机替换后有可能出生日期是未来某一天或者几百年前的一天,这就让我们的数据失去了分析价值。所以现阶段希望有一种脱敏方法,既可以对身份证号进行脱敏,又可以保持身份证号本身的结构特征和分析价值。

发明内容

[0008] 本发明为克服上述的不足之处,目的在于提供一种保留统计级特征属性的身份证号脱敏方法,既可以对身份证号进行脱敏,又可以保持身份证号本身的结构特征和分析价值。

[0009] 本发明是通过以下脱敏方案达到上述目的:1、一种保留统计级特征属性的身份证

号脱敏方法,包括以下步骤:

[0010] 一、提取待脱敏的身份证号,将身份证号分为待脱敏地区编码、待脱敏出生日期、待脱敏顺序码和待脱敏校验码;

[0011] 二、对待脱敏地区编码、待脱敏出生日期、待脱敏顺序码和校验码进行脱敏,

[0012] (1) 待脱敏地区编码进行脱敏的步骤:

[0013] ①加载全国所有地区编码信息(港澳台除外),取地区编码前2位作为Key值,相同的Key值放置到对应的List集合,若干 List集合构成Map集合;

[0014] ②获取待脱敏地区编码的前2位,在Map集合中根据Key 值,找到对应的List集合,并对List集合中的地区编码进行随机置乱;

[0015] ③获取List集合中首个地区编码,作为新的地区编码,若新的地区编码和待脱敏地区编码相同,则重复上面的步骤②,直到获取到新的地区编码;

[0016] (2) 待脱敏出生日期进行脱敏的步骤:获取基准日期与待脱敏出生日期的间隔天数,间隔天数通过转换得到新的间隔天数,新的间隔天数进制转换后与一个随机数进行模运算,得到最终的间隔天数,根据最终的间隔天数得到新的出生日期;

[0017] (3) 待脱敏顺序码进行脱敏的步骤:获取待脱敏顺序码 sequence,根据公式计算新的顺序码newSequence,如果顺序码超过 3位,则去掉最高位,所述n为随机整数:

[0018] $newSequence = sequence + 2 * n$;

[0019] (4) 根据身份证号的固定算法得到新的校验码;

[0020] (三) 新的地区编码、新的出生日期、新的顺序码和新的校验码依次组合成新的身份证号。

[0021] 作为优选,所述待脱敏地区编码为待脱敏身份证的第1-6位、待脱敏出生日期为待脱敏身份证的第7-14位、待脱敏顺序码为待脱敏身份证的第15-17位,校验码为待脱敏身份证的第18位。

[0022] 作为优选,所述新的间隔天数newGapDay的计算公式如下:

[0023] $newGapDay = (baseDate - birthDate - (32760 * flag))$

[0024] $flag = (baseDate - birthDate) / 32760$

[0025] 其中baseDate为基准日期,birthDate为待脱敏出生日期,flag取其商值并取整。

[0026] 作为优选,所述新的间隔天数转换成5位八进制数 $G_1G_2G_3G_4G_5$,高位不够用0补齐;所述随机数为5位八进制数,记为 $A_1A_2A_3A_4A_5$ 。

[0027] 作为优选,所述模运算得到5位八进制数 $D_1D_2D_3D_4D_5$,计算公式如下:

[0028] 计算 $D_5 = (G_5 + A_5) \% 8$

[0029] 计算 $D_4 = (((G_4 + A_4) \% 8) + D_5) \% 8$

[0030] 计算 $D_3 = (((G_3 + A_3) \% 8) + D_4) \% 8$

[0031] 计算 $D_2 = (((G_2 + A_2) \% 8) + D_3) \% 8$

[0032] 计算 $D_1 = (((G_1 + A_1) \% 8) + D_2) \% 8$

[0033] 作为优选,所述 $D_1D_2D_3D_4D_5$ 转换为十进制数tempDay

[0034] $tempDay = D_1 * 8^4 + D_2 * 8^3 + D_3 * 8^2 + D_4 * 8 + D_5$

[0035] 计算最后的间隔天数ResultDay

[0036] $ResultDay = tempDay + (32760 * flag)$

[0037] 得到新的出生日期 $\text{newBirthDay} = \text{baseDate} - \text{ResultDay}$ 。

[0038] 作为优选,所述 n 为 $[0, 50)$ 以内的随机整数。

[0039] 本发明的有益效果在于:一、本发明采用分段处理机制,可以灵活选择要脱敏的分段信息;二、脱敏后的身份证地区编码所在省份保持不变,满足做人员地域分布等相关方面的分析;三、脱敏后的各年龄段所占比例基本保持不变,满足做人员年龄分布等方面的分析;四、脱敏后的顺序码奇偶性保持不变、信息性别比例保持不变,满足做人员性别分布等方面的分析;五、通过身份证校验码算法获取校验码,身份证号信息结构特征保持不变。

附图说明

[0040] 图1是本发明方法的流程图。

具体实施方式

[0041] 下面结合具体实施实例对本发明进行进一步描述,但本发明的保护范围并不仅限于此:

[0042] 实施例:对身份证号642225182502231271(本身份证号为虚构号码)进行脱敏。

[0043] 一、对身份证号进行分段,将身份证号分为地区编码:642225,出生日期:18250223,顺序码:127,校验码1。

[0044] 二、对身份证号进行分段脱敏:

[0045] (1)对地区编码642225进行脱敏:

[0046] ①加载全国所有地区编码信息(港澳台除外),将其放置到 Map集合中;

[0047] ②在Map集合中获取Key值为64的集合,并对List集合中的地区编码进行随机置乱;

[0048] ③获取List集合中首个地区编码,获取到的新的地区编码为642127;

[0049] (2)对出生日期18250223进行脱敏:

[0050] ①定义基准日期为20201029:

[0051] ②计算间隔天数:基准日期-出生日期

[0052] 间隔天数为:71471天;

[0053] ③使用间隔天数整除32760,获取其商值并取整。

[0054] $\text{flage} = 71471 / 32760 = 2$

[0055] ④获取新的间隔天数为:5951天;

[0056] $5951 = 71471 - (32760 * 2)$

[0057] ⑤将新的间隔天数转换为5位八进制数为:13477。

[0058] ⑥获取一个随机的5位八进制数06426。

[0059] ⑦通过模运算,获取新的5位八进制数07665。

[0060] 计算 $5 = (7+6) \% 8$

[0061] 计算 $6 = (((7+2) \% 8) + 5) \% 8$

[0062] 计算 $6 = (((4+4) \% 8) + 6) \% 8$

[0063] 计算 $7 = (((3+6) \% 8) + 6) \% 8$

[0064] 计算 $0 = (((1+0) \% 8) + 7) \% 8$

- [0065] ⑧将新的5位八进制数转换为十进制数(tempDay)。
- [0066] $\text{tempDay} = 0 \times 8^4 + 7 \times 8^3 + 6 \times 8^2 + 6 \times 8 + 5 = 4021$ 天
- [0067] ⑨计算最后的间隔天数(ResultDay)为:69541天
- [0068] $69541 = 4021 + (32760 \times 2)$
- [0069] ⑩计算脱敏后的出生日期(newBirthDay)为:18300607
- [0070] $\text{newBirthDay} = \text{baseDate}(\text{基准日期}) - \text{ResultDay}$
- [0071] (3)对顺序码进行脱敏
- [0072] ①获取一个[0,50)以内的随机整数23,原顺序码为:127。
- [0073] ②计算新的顺序码为:173= $127 + 2 \times 23$
- [0074] 三、将脱敏后的分段信息重新组合成新的身份证号为:
- [0075] 64212718300607173,还少最后一位的校验码。
- [0076] 四、根据新的身份证号信息,通过身份证号的固定算法,获取新的身份证号校验码为:3。
- [0077] 五、最后将校验码补充完整,即可获取到脱敏后的身份证号码为:642127183006071733。
- [0078] 以上的所述乃是本发明的具体实施例及所运用的技术原理,若依本发明的构想所作的改变,其所产生的功能作用仍未超出说明书及附图所涵盖的精神时,仍应属本发明的保护范围。

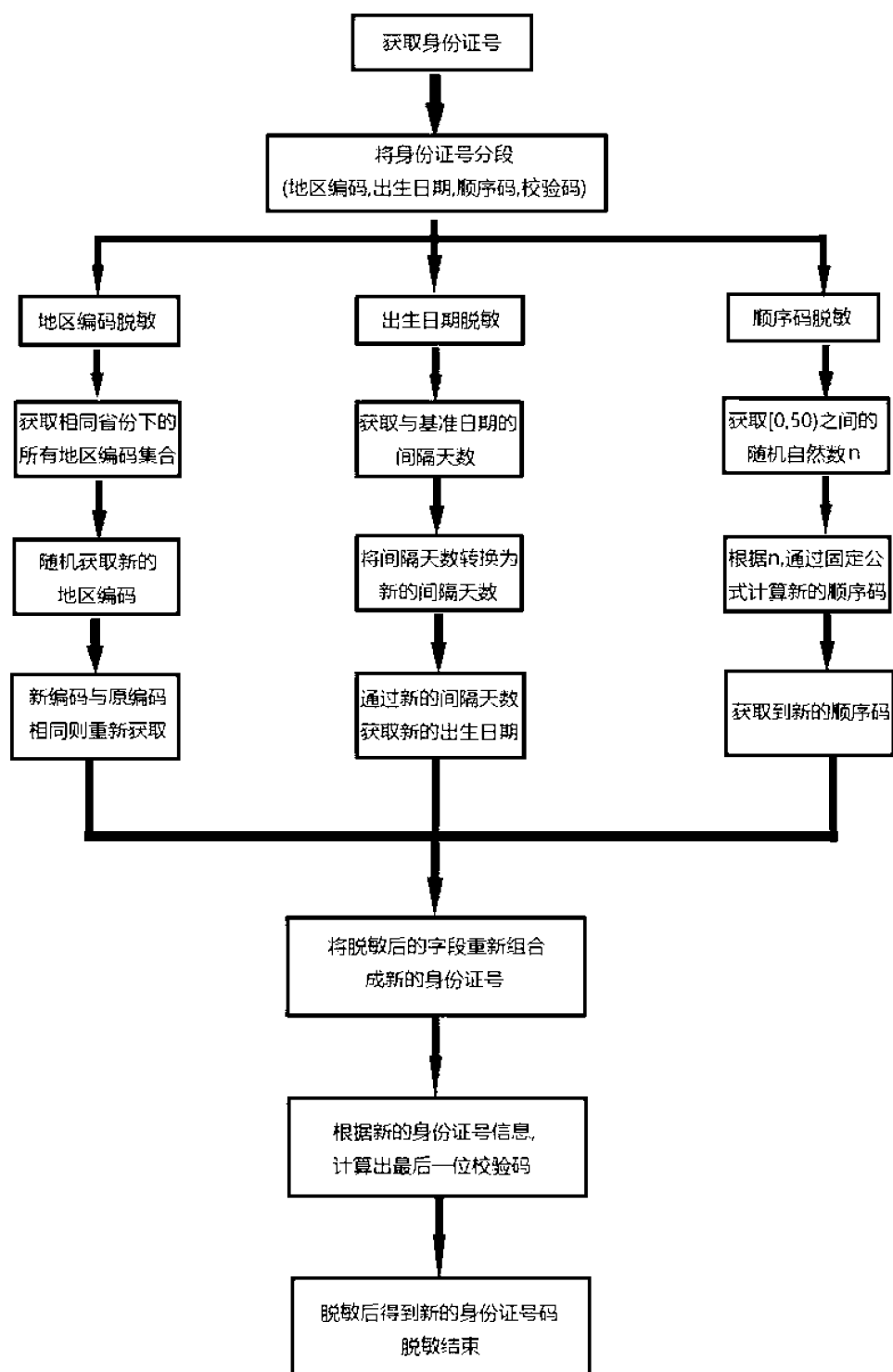


图1