(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 110162961 A
(43)申请公布日 2019.08.23

(21)申请号 201910392107.8

(22)申请日 2019.05.13

(71)申请人 华东师范大学
地址 200241 上海市闵行区东川路500号
申请人 杭州电子科技大学

(72)发明人 何道敬 周贝贝 杨肖 徐向华

(74)专利代理机构 上海蓝迪专利商标事务所（普通合伙）31215
代理人 徐筱梅 张翔

(51)Int.Cl.
G06F 21/46(2013.01)
G06F 16/215(2019.01)

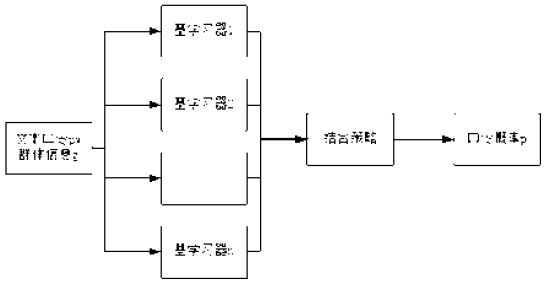
权利要求书1页 说明书3页 附图1页

(54)发明名称

基于集成学习的群体口令强度评价方法

(57)摘要

本发明公开了一种基于集成学习的群体口令强度评价方法，首先按照年龄、地区等不同属性划分群体，然后使用对应群体用户真实使用的口令数据集训练n个口令口令强度评价器作为子模型；然后将n个经过训练的子模型作为基学习器，每个基学习器得到一个口令评价结果，最后根据结合策略选出最终的口令强度评价结果。本发明集合各基学习器的优点，实现准确性高的通用口令评价方法。集成后的口令强度评价结果结合了多个基学习器的优势，在保证准确性的情况下更具通用性，可以广泛应用于电商网站、军事治安等。



CN 110162961 A

1.一种基于集成学习的群体口令强度评价方法，其特征在于，该方法包括以下具体步骤：

步骤1：口令集预处理

根据属性特征确定群体集合成员，构建群体集合，收集群体集合成员口令集并对用户真实使用的口令集数据进行清洗，得到预处理后的口令训练集；

步骤2：基学习器训练

使用预处理后的口令训练集训练n个基学习器，每个基学习器分别进行口令强度评价，并给出各自的即n个口令强度评分；

步骤3：群体口令强度评价

将步骤2所得的n个口令强度评分通过结合策略进行集成，得出最终的口令强度评分；其中，n为至少为2的整数。

2.根据权利要求1所述的群体口令强度评价方法，其特征在于，所述属性特征包括但不限于职业、地区或性别。

3.根据权利要求1所述的口令强度评价方法，其特征在于，所述步骤1具体为：

步骤A1：确定群体属性特征；

步骤A2：根据属性特征确定群体成员，构建群体集合；

步骤A3：根据集合成员收集群体口令集；

步骤A4：对收集到的口令集数据进行清洗；

步骤A5：对清洗过后的口令数据集进行预处理；其中：

所述步骤A4对口令集数据清洗包括但不限于对非法字符无效口令进行剔除；

所述步骤A5对口令数据集预处理包括但不限于对口令集数据进行one-hot编码。

4.根据权利要求1所述的群体口令强度评价方法，其特征在于，步骤3中所述结合策略包含但不局限于最弱项投票法、平均法或学习法。

基于集成学习的群体口令强度评价方法

[0001]

技术领域

[0002] 本发明属于信息安全技术领域，特别涉及一种基于集成学习的群体口令强度评价方法。

背景技术

[0003] 随着互联网技术不断发展，生活环境信息化、数字化不断推进，网络技术已经渗透到人们生活的方方面面。由于网络安全问题的频发，导致人们增强对网络环境安全情况重视，身份认证技术已经成为保护用户个人信息安全最基本也是最重要手段。在身份认证技术中，由于文本口令其易部署性、低成本性及实现方式简单性，文本口令在短时间甚至更远的将来都仍作为身份认证技术的重要手段。但是用户为了便于记忆，用户使用的口令大多具有关联性，导致口令强度降低，从而导致口令容易遭受猜测攻击。为了提高用户口令的安全强度，一般互联网服务商都会强行执行口令强度评测(Password Strength Metirc，简称PSM)向用户反馈用户口令的强弱程度，并根据评测结果向用户提出构建高强度口令建议。但是目前各个互联网服务商并未统一口令评价方法，导致相同口令在不同环境下得出不同的反馈结果，从而给用户带来困惑，导致用户口令有安全隐患。

[0004] 口令安全性研究的难点在于，口令是人产生的，与人的行为直接相关，而每个人行为因内在或者外在的环境而千差万别，所以口令之间具有很大的差异。在口令评价方面，基于对猜测攻击方法和用户脆弱口令行为的深入理解，常用的方法是使用通用口令列表来评价用户输入的口令，如：用户输入口令是否在通用口令列表里，来判断口令是否可接受。这种方法具有很大的局限性，其准确程度取决于黑名单口令列表的大小，并且影响用户体验。目前，根据美国国家标准技术研究所 (National Institute of Standards andTechnology, NIST)的建议而衍生的启发式口令强度估计也颇受欢迎，它是基于大小写字母、数字和特殊字符 (counts of lower and uppercase letters, digits and symbol,LUDS)数量来计算信息熵的，信息熵越大，口令强度就越强。然而，相关文献表明基于信息熵的口令强度评价方法，只能提供一个粗略的评价结果。

[0005] 口令强度评价是口令安全研究的重要组成部分，但是通过调研发现目前的口令强度评价方法均忽略了口令的区域性特征，未考虑区域性群体对口令生成的影响；同时，主流的基于启发式或概率模型的口令强度评价方法通常只能对特定类型口令进行准确强度评价，同一个口令在不同口令强度评价器中的评价结果截然不同，缺乏通用性。

发明内容

[0006] 本发明的目的是针对现有技术的不足而提供的一种基于集成学习的群体口令强度评价方法，该方法集合各基学习器的优点，实现准确性高的通用口令评价方法。集成后的强度评价模型结合了多个基学习器的优势，在保证准确性的情况下更具通用性，可以广泛

应用于电商网站、军事治安等。

[0007] 实现本发明目的的具体技术方案是：

一种基于集成学习的群体口令强度评价方法，该方法包括以下具体步骤：

步骤1：口令集预处理

根据属性特征确定群体集合成员，构建群体集合，收集群体集合成员口令集并对用户真实使用的口令集数据进行清洗，得到预处理后的口令训练集；

步骤2：基学习器训练

使用预处理后的口令训练集训练n个基学习器，每个基学习器分别进行口令强度评价，并给出各自的即n个口令强度评分；

步骤3：群体口令强度评价

将步骤2所得的n个口令强度评分通过结合策略进行集成，得出最终的口令强度评分；其中，n为至少为2的整数。

[0008] 所述属性特征包括但不限于职业、地区或性别。

[0009] 所述步骤1具体为：

步骤A1：确定群体属性特征；

步骤A2：根据属性特征确定群体成员，构建群体集合；

步骤A3：根据集合成员收集群体口令集；

步骤A4：对收集到的口令集数据进行清洗；

步骤A5：对清洗过后的口令数据集进行预处理；其中：

所述步骤A4对口令集数据清洗包括但不限于对非法字符无效口令进行剔除；

所述步骤A5对口令数据集预处理包括但不限于对口令集数据进行one-hot编码。

[0010] 步骤3中所述结合策略包含但不局限于最弱项投票法、平均法或学习法。

[0011] 本发明针对不同群体使用多个基学习器进行训练，结合多个口令强度评价基模型的优点，有效提高了口令强度评价的准确性和通用性，使得评价结果可以跨平台使用在电商网站、军事安全等各方面。

附图说明

[0012] 图1为本发明流程图；

图2为本发明实施例流程图。

具体实施方式

[0013] 结合以下具体实施例和附图，对本发明作进一步的详细说明。实施本发明的过程、条件、实验方法等，除以下专门提及的内容之外，均为本领域的普遍知识和公知常识，本发明没有特别限制内容。

实施例

[0014] 参阅图2，本实施例包括如下步骤：

对某个在CSDN网站要注册的口令“password123”。

[0015] 步骤1：口令集预处理

按照“地区”作为群体分类标准，它属于中文用户群体，因此使用CSDN、12306等群体口令数据集并对用户真实使用的口令数据集进行清洗，得到预处理后的口令训练集。

[0016] 步骤2：基学习器训练

使用步骤1中预处理后的口令集分别训练基于流行口令的口令强度评价基学习器、基于模式检测的口令强度评价基学习器、基于语义的口令强度评价基学习器和基于注意力机制的神经网络群体口令强度评价基学习器，每个基学习器分别进行口令强度评价，并给出各自出现此口令的概率分别为0.06、0.32、0.05和0.41；

步骤3：群体口令强度评价

通过最弱项评分法，即选择上述四个口令评价基学习器的评分最低(即口令出现概率最大)的结果0.41作为最终的口令强度得分，并反馈口令强度。

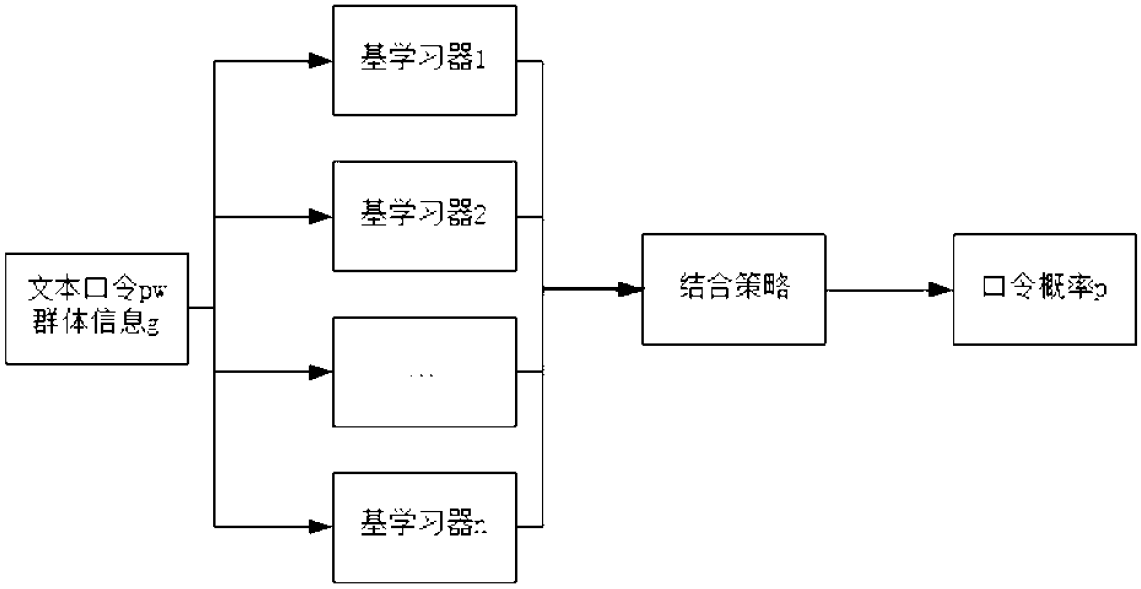

基学习器1
基学习器2
...
基学习器x
文本口令pw
群体信息g
结合策略
口令概率p

图1

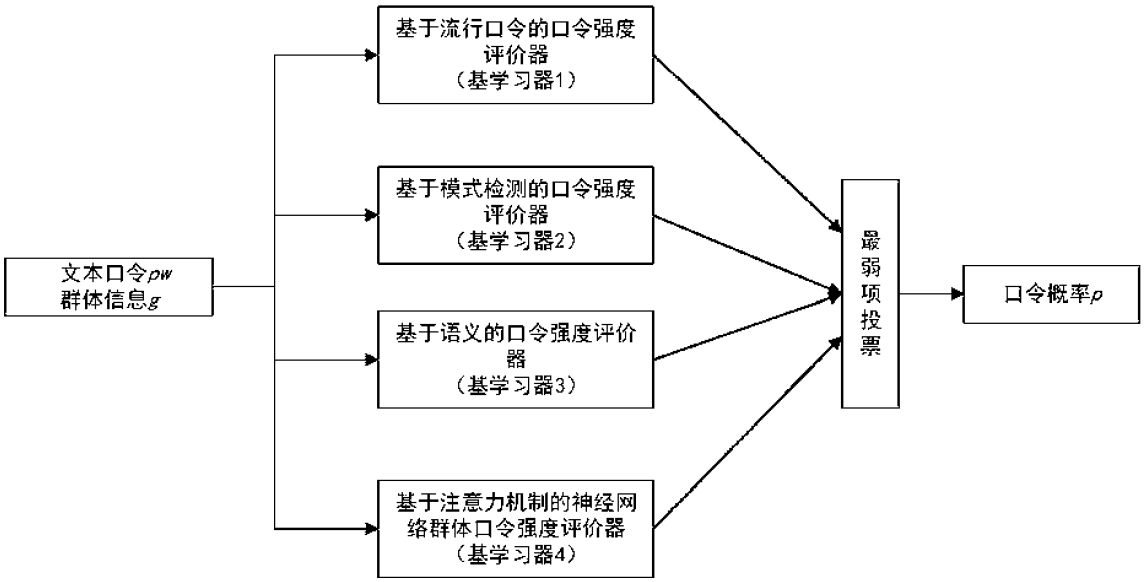

基于流行口令的口令强度评价器（基学习器1）
基于模式检测的口令强度评价器（基学习器2）
基于语义的口令强度评价器（基学习器3）
基于注意力机制的神经网络群体口令强度评价器（基学习器4）
文本口令pw
群体信息g
最弱项投票
口令概率p

图2