(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 108509790 A
(43)申请公布日 2018.09.07

(21)申请号 201810207642.7

(22)申请日 2018.03.14

(71)申请人 华东师范大学
地址 200241 上海市闵行区东川路500号

(72)发明人 何道敬 杨肖 周贝贝

(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)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

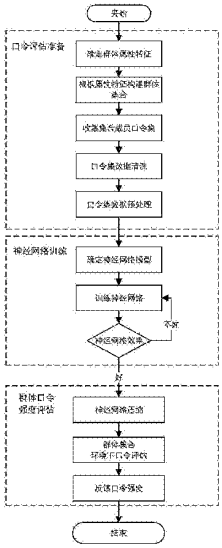
权利要求书1页 说明书4页 附图2页

(54)发明名称

一种基于群体的口令强度评估方法

(57)摘要

本发明公开了一种基于群体的口令强度评估方法，包括口令评估准备、神经网络训练、和群体口令强度评估步骤。其中，口令评估准备步骤：确定群体属性特征，根据属性特征构建群体集合，收集口令集，对口令集数据进行清洗，完成口令集数据预处理；神经网络训练步骤：确定神经网络模型，集合预处理结果训练神经网络；群体口令强度评估步骤：通过向训练好神经网络中输入口令，评估口令在已训练群体中的强度。本发明利用神经网络对口令在群体中强度进行评估，将神经网络压缩后可达到即插即用的效果，及时反馈口令强度，一定程度消除群体的弱口令，增强了口令抵抗口令猜测攻击的能力，提高了口令的安全性。



CN 108509790 A

1.一种基于群体的口令强度评估方法，其特征在于，该方法包括以下具体步骤：

步骤1：口令评估准备

确定群体属性特征，根据群体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口令集数据进行清洗，完成口令数据集预处理；

步骤2：神经网络训练

确定神经网络模型，结合预处理后口令数据集训练神经网络模型；

步骤3：群体口令强度评估

对训练好的神经网络模型进行压缩，对口令在群体集合环境下进行强度评估，并及时反馈口令强度。

2.根据权利要求1所述的口令强度评估方法，其特征在于，步骤1中，所述群体为具有相同属性特征的一类人；

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所述的口令强度评估方法，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所述的口令强度评估方法，其特征在于，所述步骤2具体为：

步骤B1：确定神经网络模型；

步骤B2：训练神经网络模型；

步骤B3：通过试验效果对神经网络模型进行参数调整，进行再次训练，直到效果好为止；其中：

所述神经网络模型包括但不限于循环神经网络模型及注意力机制；

所述调整的参数包括但不限于神经网络学习率、步长和训练轮数。

5.根据权利要求1所述的口令强度评估方法，其特征在于，所述步骤3具体为：

步骤C1：对训练效果好的神经网络模型进行模型压缩；

步骤C2：将口令在已训练群体环境中进行口令强度计算及评估；

步骤C3：反馈口令强度；其中：

所述模型压缩包括但不限于：轻量化压缩、有损压缩或ZigZag编码；

所述口令强度计算及评估包括但不限于：在已训练群体环境情况中，利用神经网络学习到的口令概率分布，结合蒙特卡洛仿真抽样方法，计算出口令被成功猜测出所需猜测次数，通过将猜测次数与预先设置的强度阈值进行比较，从而得出该口令在已训练群体中强度。

一种基于群体的口令强度评估方法

技术领域

[0001] 本发明属于信息安全技术领域，特别涉及一种基于群体的口令强度评估方法。

背景技术

[0002] 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的生活方方面面。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估方法，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。

[0003] 由于便记忆性，用户的口令并非是完全随机或者是毫无意义的，通常会采用生日、姓名等有意义的字符，从而降低了口令安全性。此外，通过研究发现，不同群体用户可能由于其职业、地区、姓名等潜在属性，常常会选择具有群体特征的字符作为口令，从而存在潜在的口令安全问题。在信息量爆炸的时代，大数据与人工智能技术的飞速发展，在短时间内从大规划数据中查找潜在关联已经变得比较容易。

发明内容

[0004] 本发明的目的在于弥补现有口令强度评估技术的不足，将神经网络技术与口令强度评估相结合，提供一种基于群体的口令强度评估方法，在继承了传统口令强度评估方法的效率性和鲁棒性的同时，一定程度消除已训练群体的弱口令，增强了口令抵抗口令猜测攻击的能力，提高了口令的安全性。

[0005] 实现本发明目的的具体技术方案是：

一种基于群体的口令强度评估方法，该方法包括以下具体步骤：

步骤1：口令评估准备

确定群体属性特征，根据群体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口令集数据进行清洗，完成口令数据集预处理；

步骤2：神经网络训练

确定神经网络模型，结合预处理后口令数据集训练神经网络模型；

步骤3：群体口令强度评估

对训练好的神经网络模型进行压缩，对口令在群体集合环境下进行强度评估，并及时反馈口令强度。

[0006] 本发明步骤1中，所述群体为具有相同属性特征的一类人；

所述属性特征包括但不限于职业、地区或性别。

[0007] 本发明所述步骤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；

所述口令数据集预处理方法包括但不限于对口令集数据进行one-hot编码。

[0008] 本发明所述步骤2具体为：

步骤B1：确定神经网络模型；

步骤B2：训练神经网络模型；

步骤B3：通过试验效果对神经网络模型进行参数调整，进行再次训练，直到效果好为止；其中：

所述神经网络模型包括但不限于循环神经网络模型(Recurrent Neural Networks，RNN)及注意力机制(Attention)；

所述调整的参数包括但不限于神经网络学习率、步长、训练轮数。

[0009] 本发明所述步骤3具体为：

步骤C1：对训练效果好的神经网络模型进行模型压缩；

步骤C2：将口令在已训练群体环境中进行口令强度计算及评估；

步骤C3：反馈口令强度；其中：

所述模型压缩方法包括但不限于：轻量化压缩、有损压缩或ZigZag编码；

所述口令强度计算及评估包括但不限于：在已训练群体环境情况中，利用神经网络学习到的口令概率分布，结合蒙特卡洛仿真抽样方法，计算出口令被成功猜测出所需猜测次数，通过将猜测次数与预先设置的强度阈值进行比较，从而得出该口令在已训练群体中强度。

[0010] 本发明在继承了传统口令强度评估方法的效率性和鲁棒性的同时，一定程度消除已训练群体的弱口令，增强了口令抵抗口令猜测攻击的能力，提高了口令的安全性，保障了用户的信息安全及财产安全。同时，本发明将训练好的神经网络进行压缩，达到了即插即用的效果，能够快速反馈口令强度，帮助用户构建强度较高的口令。此外，本发明利用机器学习方法从海量真实口令数据中找出不同群体使用口令的内在联系，提高强度评估结果的准确性和真实性。

附图说明

[0011] 图1是本发明的流程图；

图2是网络模型图。

具体实施方式

[0012] 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。

[0013] 本发明中有关的技术术语代表的含义如下：

RNN-循环神经网络(Recurrent neural networks)

LSTM-长短期记忆网络(Long Short-Term Memory)

Attention-注意力机制

群体属性特征:A

以属性A作为群体分类的表示:S = {s1, s2 … sn | A}

口令集:P(S) = {P(s1), P(s2) … P(sn)}

Lr-学习效率(learning rate)

Step-步长

Ep-训练轮数

Count-猜测次数

M-神经网络模型

如图1所示，本发明一种基于群体的口令强度评估方法包括如下三个步骤：

步骤1:口令评估准备:确定群体属性特征A，根据群体属性特征确定群体集合成员s，构建群体集合S，收集群体集合成员口令集P(S)并对口令集数据进行清洗，完成口令数据集预处理；

步骤2:神经网络训练:确定神经网络模型M，结合预处理后口令数据集训练神经网络模型；

步骤3:群体口令强度评估:对训练好的神经网络模型进行压缩，对口令在群体集合环境下进行强度评估，并及时反馈口令强度。

[0014] 其中，步骤1中群体S为具有相同属性特征的一类人；

属性特征A包括但不限于职业、地区或性别。

[0015] 其中，步骤1:口令评估准备具体步骤如下：

A1、确定群体属性特征A；

A2、根据属性特征A确定群体成员s，构建群体集合S；

A3、根据评估范围收集群体相关口令集P(S)；

A4、对收集到的口令集数据进行清洗；

A5、对清洗过后的口令数据集进行预处理。

[0016] 其中，步骤A4中对口令集数据P(S)进行清洗的方法包括但不限于对非法字符无效口令进行剔除；

口令数据集预处理方法包括但不限于对口令集数据进行one-hot编码；

其中，步骤2:神经网络训练具体步骤如下：

B1、确定神经网络模型M，如图2所示，将口令数据集预处理送入1一个LSTM进行训练，再将LSTM输出到一个Attention，经过Attentoin计算过后将输出输入另外一个LSTM重复下去，图2模型中共有3层LSTM，2个Attention作为训练编码口令集，2层LSTM作为解码输出口令；

B2、训练神经网络模型；

B3、通过试验效果对神经网络模型进行参数调整，进行再次训练，直到效果较好为止。

[0017] 其中，神经网络训练阶段步骤B1神经网络模型包括但不限于RNN，LSTM及Attention；步骤B3中参数包括但不限于神经网络学习率lr，步长step和训练轮数ep。

[0018] 其中，步骤3：群体口令强度评估具体步骤如下：

C1、对训练效果好的神经网络模型M进行模型压缩；

C2、将口令在已训练的群体环境中进行口令强度计算及评估；

C3、反馈口令强度；其中：

步骤C1中神经网络模型压缩的方法包括但不限于：轻量化压缩、有损压缩或ZigZag编码；

其中，口令强度计算及评估方法包括但不限于：在已训练的群体环境中，利用神经网络学习到的口令概率分布，结合蒙特卡洛仿真抽样方法，计算出该户口令被猜测出所需猜测次数count，通过将猜测次数与预先设置的强度阈值{1e3，1e6，1e9，1e12}进行比较，从而得出该口令在已训练群体中强度。

[0019] 通过向用户反馈口令在已训练群体中口令强度，帮助用户构建强度较高口令，在一定程度上消除了属于该群体中的弱口令，同时也提高了用户口令防止猜测攻击的能力，提高了口令的安全性，保障了用户的信息安全及财产安全。同时，将训练好的神经网络进行压缩，达到了即插即用的效果，能够快速反馈口令强度，帮助用户构建强度较高的口令。此外，利用机器学习方法从海量真实口令数据中找出不同群体使用口令的内在联系，提高强度评估结果的准确性和真实性。

[0020] 本发明的保护内容不局限于以上实施例。在不背离发明构思的精神和范围下，本领域技术人员能够想到的变化和优点都被包括在本发明中，并且以所附的权利要求书为保护范围。

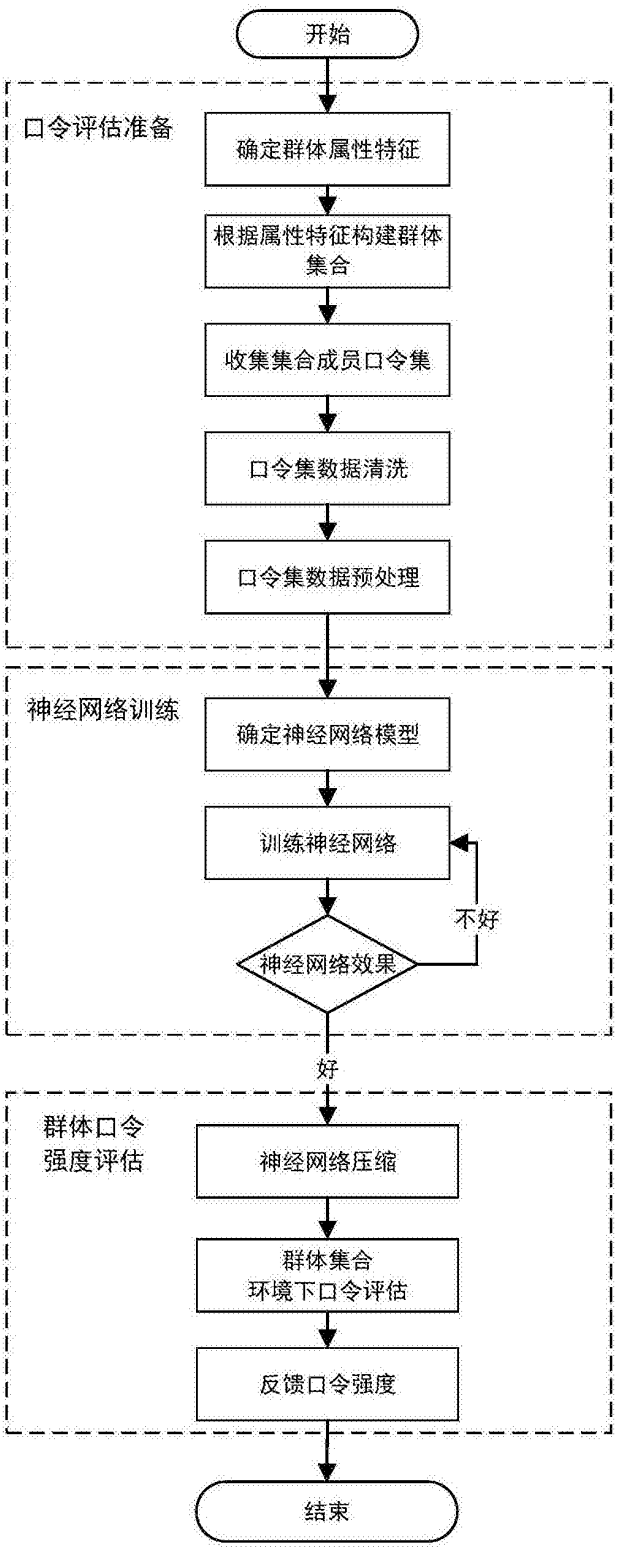

开始
口令评估准备
确定群体属性特征
根据属性特征构建群体集合
收集集合成员口令集
口令集数据清洗
口令集数据预处理
神经网络训练
确定神经网络模型
训练神经网络
不好
神经网络效果
好
群体口令强度评估
神经网络压缩
群体集合环境下口令评估
反馈口令强度
结束

图1

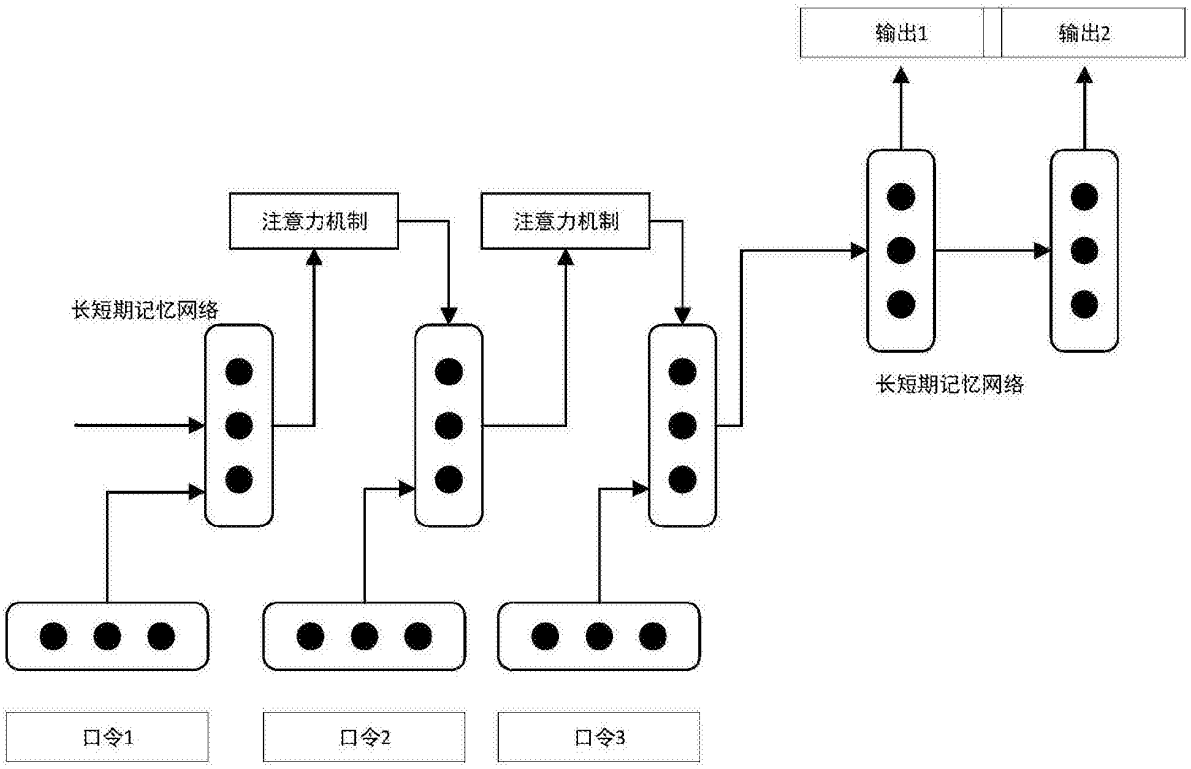


图2