基于机器学习的身份认证专利技术综述

□段玥　国家知识产权局专利局专利审查协作江苏中心

【摘要】　随着云计算、物联网、大数据等新兴技术的迅猛发展，给网络空间安全带来了巨大的困难和挑战，传统的安全问题解决方案面对海量数据变得效率低下，尤其是在数据信息安全领域，身份认证面临着巨大的挑战，而机器学习以其强大的自适应性、自学习能力为安全领域提供了一系列有效的分析决策工具，近年来引起了学术界与工业界的广泛关注和深入研究。基于机器学习的身份认证技术有效保证用户的身份安全，对提高用户数据和设备的安全性起着十分重要的作用。本文围绕基于机器学习的身份认证技术进行研究，总结了与基于机器学习的身份认证技术相关的国内和国外专利的申请发展趋势，其中对主要申请人分布以及重点技术的发展路线进行了脉络梳理。

【关键词】　机器学习　身份认证　识别　生物特征　行为模式

引言

随着大数据时代的到来，机器学习与深度学习技术已经成为当前人工智能领域的一个研究热点，与此同时国内外各大公司，高校，研究所开始着重进行机器学习方面的学习和研究；同时，用户现在可以在他们的智能手机上使用和存储更为私密和敏感的信息，所以识别用户身份的真实性以保护用户隐私在现下已经势在必行，基于机器学习的身份认证是指从信号中提取反映身份的特征，利用机器学习算法对身份特征进行训练学习，建立相应模型，通过综合分析后得到是否来自合法用户的判断，有效保证用户的身份安全，对提高用户数据和设备的安全性起着十分重要的作用。

一、专利申请现状分析

本节主要对全球的专利申请状况的趋势以及重要申请人进行分析，如图 1 所示，从专利技术的国家 / 地区来源来看，美国申请的专利数量最多，接近全部专利的二分之一，其次是中国，其专利数量也相当可观，占全部专利的 39%，分别位列世界第一位和第二位，韩国和日分别本占全部专利的 7% 和 4%，其他国家 / 地区的申请总量和为 11%，前四位专利申请量国家 / 地区专利申请总量超过全部专利的 90%。

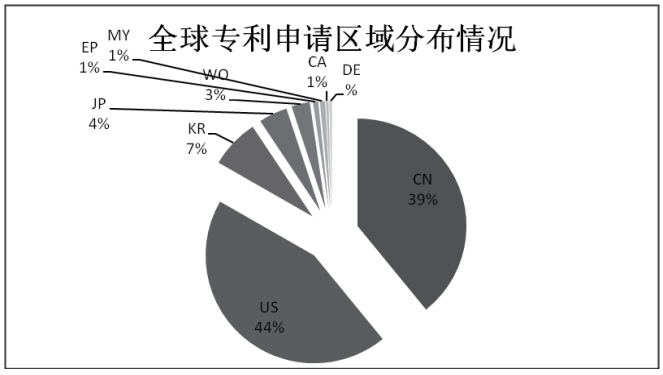

图 1　全球专利申请区域分布

从本领域重要申请人方面做进一步分析，主要考虑申请人历年的申请总量：

图 2 是专利申请数量排名前 15 的国际申请人，从本领域重要申请人方面做进一步分析得到前 15 位申请人，其中，不乏美国的 Google、英特尔，韩国的三星，以及中国的阿里巴巴、百度等国际型大公司，且这些申请人在申请数量以及质量方面都从始至终占据较为重要的地位，部分公司一直属于领域的领头羊。

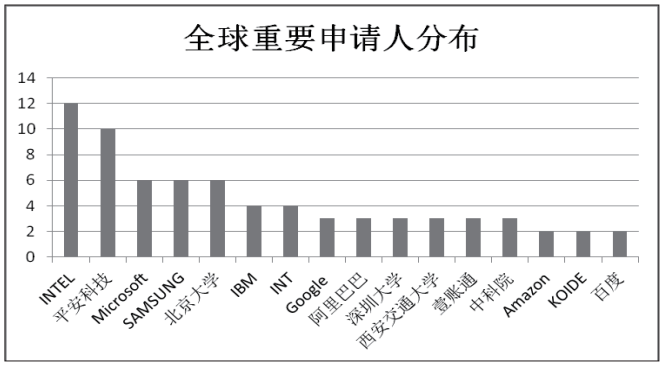

图 2　全球重要申请人分布

二、技术分支分析

基于机器学习的身份认证技术通过机器学习建立相应模型，针对用户的多个身份因素进行识别认证，其涉及多种生物采集识别器、传感器、重力感应器等硬件设备，实现对身份因素数据的采集、交互、处理，因此涉及多项基础技术。申请人将多项基础技术进行结合，从技术应用的角度可将其分为：用户行为模式、用户生物特征、设备状态数据、系统使数据用主要技术分支：

（1）用户行为模式：用户在使用相关设备时，每个用户都用其独特的输入特征，通过采集这些数据，基于机器学习建立相应的用户行为模型，基于模型识别合法用户，并且能够对用户进行持续认证。

基于数据来源，通常分为两类，第一类是用户在触摸屏上的相关操作构成具有用户个人信息的触摸行为数据，包括：点击模式、按压程度、触摸区域、触摸时间。滑动轨迹等，由于用户使用设备的习惯通常也是各具特点，上述数据通常也是唯一性的，其可用于识别用户的合法性，由于用户行为一般没有标签，通过采用无监督模式的机器学习对采集到的数据进行分析训练，如西京学院提出的专利申请（CN201710032948，申请日为 20170118），以习惯性手势先后接触触摸屏，系统记录手指之间的距离和触屏时间差，对数据进行模糊聚类分析处理，判断是否与合法用户数据匹配。

第二类是通过用户的运动和姿态模式形成的动态数据，如拿起设备、解锁设备，放下设备等中产生的相关用户行为的数据，数据出采集来源包括运行和姿态传感器，如加速计、重力感应器、陀螺仪、定位器等，尤其是可穿戴设备的普及，更是将数据的采集工作变得更加方便准确，如中国科学院深圳先进技术研究院提出的专利申请（CN201810822151，申请日为 20180724），身份认证方法，包括：获取设置于用户身上的加速度传感器采集的步态加速度信号；对步态加速度信号进行预处理，以降低步态加速度信号中的噪声；对步态加速度信号进行步态周期检测，得到多个步态周期；将多个步态周期分割为多个步态分割 段，并对每个步态分割段均进行模式提取，得到每个步态分割段对应的步态模式，其中，以四个步态周期为一个步态分割段，相邻的两个步态分割段之间设置 (50)% 的重叠，将多个步态周期分割为多个步态分割段；利用 MFCC 算法对每个步态模式均进行特征向量提取，得到每个步态模式的步态特征；依据每个步态模式的步态特征，构建身份识别模型，以确认用户身份。

（2）用户生理特征: 用户的生理特征是唯一且不可变的，通过采集用户的生理特征，尤其是针对传统认证方式中不容易识别的心率、脑电波等生物特征进行处理，识别合法用户。

生物特征识别技术是利用人体固有的生理特征来进行个人身份认证的技术，指纹识别技术在生物特征识别技术领域中起步较早，在技术上发展较为成熟，已投产应用最多，所占的市场份额最大的一类生物特征识别技术，并基于机器学习技术对指纹识别技术进行了新的研究和发展，如阿里巴巴集团控股有限公司提出的专利申请：一种基于网络指纹的身份识别方法和装置（CN201811109943，申请日为 20180921），基于用户之间相互认识的关系构建出用户关系网，然后通过网络指纹指标进行初步的去重筛选，最后通过聚类处理根据用户之间相互认识关系的程度，分析得到其中的真实用户，为后续处理提供数据基础。同时，对非接触式的生物特征也得到了广泛的发展，如人脸识别、语音识别、虹膜识别等，对此，多家企业和高校科研机构均提出了相关的专利申请，北京工业大学提出的专利申请（CN 201510420256，申请日为 20150716），提供一种人脸图像性别识别方法及系统，组合特征更具有泛化能力，决策结果更稳定，识别准确率更高；无锡中科奥森科技有限公司提出的专利申请（CN201210059454，申请日为 20120308），提供一种双验证人脸防伪方法及装置，以提高识别精确度，更加安全可靠，并且避免用户负担重、人机交互时间长。

（3）设备状态数据：设备状态数据通常与设备的环境风险等级相关，通过设备上的传感器判断当前环境的风险等级，并基于风险等级采取相应的认证方式，如当设备位于家中时是低风险级别，可以不采用任何显式认证机制，当设备位于办公室时是中等风险级别，可以采用中等显式认证机制，当设备处于户外或者陌生环境时认为是高风险级别，需要采用强显式认证机制。

设备的环境状态，如位置信息，则对用户认证是十分重要的，尤其是针对隐式身份认证，在用户不易察觉的情况下，判断设备所处的风险等级，并基于风险等级身份认证的相关策略，以提高设备的安全性，如通过 CPS 设备定位信息、通过用户的日历计划表，获得用户预期的地点或者基于设备接入的基站信息、IP 地址等定位用户的当前位置和移动路线，如阿里巴巴集团控股有限公司提出的专利申请（CN201810845003，申请日为 20180727），身份验证方法，包括：在接收到账户信息变更请求之后，获取当前操作者的个人信息和网络 环境信息；分别对个人信息和网络环境信息进行风险识别，对应获取第一风险识别结果和第二风险识别结果，其中，第一风险识别结果用于表征当前操作者与账户所有者不一致的风险，第二风险识别结果用于表征当前操作者所处网络环境存在安全隐患的风险；根据第一风险识别结果和第二风险识别结果选择身份验证方式，并根据身份验证方式对当前操作者进行身份验证，能够提高识别准确率，降低账户被盗用的风险，提高实时性和安全性，减少用户不必要的操作。

（4）系统使用数据：用户对手机系统的使用数据可以识别出非法用户，系统的使用数据具体包括：通话模式、短消息、网站访问记录、软件使用情况、当时使用软件列表等。当相关数据发生显著改变时，则说明设备可能被非法使用。

因此可利用系统使用数据对用户的身份进行认证，如 OPPO 广东移动通信有限公司提出的专利申请（CN201710526399，申请日为 20170630），基于用户习惯自动启动应用，根据用户使用移动设备的习惯自动解锁并进行预设的应用中。

三、总结

本文在对国内外基于机器学习的身份认证领域专利申请定量统计分析的基础上，对基于机器学习的身份认证领域中的技术进行了分类，并且分析了每类的关键技术点，有助于本领域技术人员快速了解当前基于机器学习的身份认证领域的发展现状。

参 考 文 献

[1] 张帆 . 机器学习与深度学习相关研究综述 [C].2018 成都人工智能发展论坛 ,20190128:58-62.
[2] 李卫 . 深度学习在图像识别中的研究及应用 [D]. 武汉理工大学 ,2014:1-64.
[3] 孙权森 , 曾生根 , 王平安 , 等 . 典型相关分析的理论及其在特征融合中的应用 [J]. 计算机学报 ,2005,(9).
[4] 毋泽南 . 基于用户动态行为的安全生产数据的访问控制技术研究与实现 [D]. 华北科技学院 ,2018.
[5] 章敏敏 , 徐和平 , 王晓洁 , 等 . 谷歌 TensorFlow 机器学习框架及应用 [J]. 微型机与应用 ,2017,(10).

段玥（1989—），女，汉族，江苏省苏州市，国家知识产权局专利局专利审查协作江苏中心，研究方向为信息安全领域的专利审查。