**珠海科技学院**

**大学生创新创业训练计划**

**创新训练项目**

**申请书**

|  |  |  |  |
| --- | --- | --- | --- |
| 立项年份： | 2023年 | | |
| 项目编号： |  | | |
| 项目名称： | 基于机器学习实现涉诈网址自动分类识别 | | |
| 项目级别： | □国家级 ☑省级 □校级 | | |
| 项目负责人： | 钟耀荣 | 学号： | 20200619 |
| 学院: | 阿里云大数据应用学院 | | |
| 专业班级: | 数据科学与大数据技术2006班 | | |
| 联系电话： | 13822478129 | | |
| 第一指导教师： | 肖悦 | 电话： | 15876666808 |
| 项目起止年月： | 2023年 3月 至 2024 年 3 月 | | |

填写须知

1.填写申请书前请认真阅读《珠海科技学院大学生创新创业训练计划项目管理办法（试行）》。

2.本申请书所列各项内容均须实事求是，认真填写，表达明确严谨，简明扼要。

3.申请人可以是个人，也可为创新团队，首页只填负责人。“项目编号”一栏不填。

4.项目所属学科门类应是哲学、经济学、大学、教育学、文学、历史学、理学、工学、农户学、医学、管理学和艺术学等一级学科门类中的一种。所属专业大类及代码参考《专业大类对照表》。

5.导师类型分为：第一指导教师、指导教师、企业导师等。（创业实践项目必须有企业导师参与指导）。导师职称参考《职称对照表》。

6.本申报表中各项内容用“小四”号仿宋体填写，格式、内容、大小均须与表格样式一致；填写完后用A4纸张双面打印，（经费预算和意见页单独一页）左侧装订成册。

**一、基本情况**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 项目名称 | | 基于机器学习实现涉诈网址自动分类识别 | | | | | | |
| 所属学科门类 | | 工学 | | 所属专业大类/代码 | | | 0809 | |
| 实施时间 | | 2023 年 3月至 2024年 3 月 | | 研究周期 | | | ☑1年 □2年 | |
| 项目负责人及团队成员信息 | 序号 | 姓 名 | 年级 | 学号 | 学院专业 | | 联系电话 | 项目中的分工 |
| 负责人 | 钟耀荣 | 20级 | 20200619 | 数据科学与大数据技术 | | 13822478129 | 项目负责人 |
| 成员 | 苏洇如 | 20级 | 13200608 | 数据科学与大数据技术 | | 13670263808 | 技术负责人 |
| 梁家杰 | 20级 | 20200611 | 数据科学与大数据技术 | | 15016485509 | 研发负责人 |
| 伍思名 | 20级 | 13200606 | 物流工程 | | 17275493268 | 市场负责人 |
| 陈心卉 | 20级 | 13200309 | 电子商务 | | 18320414522 | 运营负责人 |
| 负责人曾经参与科研的情况 | 参与了基于机器学习的电商用户群体分类。 | | | | | | |
| 项目指导教师 | 1 | 姓名 | 学院/单位 | 职称 | 联系电话 | 邮箱 | | 导师类型 |
| 肖悦 | 阿里云大数据应用学院 | 助教 | 15876666808 | xiaoyue810@zcst.edu.cn | | 第一指导老师 |
| 指导教师承担科研课题情况：  NLP技术在医疗领域文本生成和智能问答系统的研究  发表国际会议EI 2篇，1篇Sci，参与多项省级项目，主要研究领域机器学习和自然语言处理 | | | | | | |
| 2 | 姓名 | 学院/单位 | 职称 | 联系电话 | 邮箱 | | 导师类型 |
| 陈倩 | 物流管理与工程学院 | 副教授 | 13555936422 | 1033670790@qq.com | | 第二指导老师 |
| 指导教师承担科研课题情况：  发表国际会议EI 2篇，1篇Sci，参与多项省级项目，主要研究领域机器学习和自然语言处理；NLP技术在医疗领域文本生成和智能问答系统的研究。 | | | | | | |
| 指导教师、  企业导师支持情况 | |  | | | | | | |
| 项目简介  (200字以内) | | 这个项目基于机器学习，也是运用了机器学习中的模型和训练方法进行训练；从互联网的现状得知，万维网逐渐从一个提供对静态信息的访问的系统发展演变成一个成熟的分布式基础设施。因此人们的生产、生活越来越多地依赖于网络的支持，离不开网络技术的支撑。然而网络犯罪行为也日益严重，不法分子通过构建恶意网站，利用网络漏洞牟利，对互联网用户的信息，财产安全构成巨大威胁。基于现在的网络大时代，网址存在着诈骗、恶意等非法行为已经屡见不鲜，严重危害了社会的网络治安导致电脑环境的损伤和民众财产的损失。所以我们以涉诈网址为立项之本，运用机器学习的模型应用来为网络安全出一份力，在网络时代，网络安全尤为重要；我们以机器学习的分类算法识别恶意、非法、有诈骗嫌疑的网站为核心展开建设,我们运用globalpointer模型可以无差别识别非嵌套实体（flat-ner）和嵌套实体（nested-ner）速度比CRF要更快。利用高效模型来快速识别网址的恶意性，使得网络环境得以改善，网络安全得到保障。 | | | | | | |

**二、立项依据**

|  |
| --- |
| （一）研究目的  伴随着社会信息化和全球一体化的高速发展，计算机网络迅速走进人们的生活，当今世界已经越来越依赖于信息网络的正常运行。以此同时，计算机、网络相关犯罪也出现在社会各个领域并迅速蔓延尤其是伴随着网络经济的发展，传统的诈骗方式正不断向网络渗透。这之中，网络诈骗犯罪利用恶意网站进行诈骗愈演愈烈。较之于传统诈骗犯罪，网络诈骗充分利用网络本身所具有的匿名性、跨国性、实时性等特点，使其欺骗性和隐蔽性更强，它所造成的经济损失之巨、受害人之广、社会危害性之大，远远超过了传统诈骗犯罪，已成为世界各国共同关注的犯罪问题。  回顾这几年，我国的诈骗分子利用人工智能技术实施电信诈骗活动，特别是基于人工智能技术的个人信息挖掘、深度伪造等各种诈骗手段，使得电信网络诈骗活动治理难度加剧，社会危害性巨大。   1. 在政府政策方面：   近年来，网站诈骗呈高发趋势，诈骗手法也不断翻新，造成了巨大的财产损失，严重影响了人民群众的正常生活。为应对当前网络安全问题，政府部门坚决贯彻党中央决策部署，持续开展电信网络安全领域治理工作。工信部于 2020 年 8 月印发《关于运用大数据推进防范治理电信网络诈骗长效机制建设工作方案》，提出要打造信息通信行业反诈大数据技术手段，持续提升大数据技术管控水平。**因此，我们可以承运国家政策之便，运用机器学习的模型应用改善网络安全问题。**   1. 在人工智能方面：   随着近年来我国网站诈骗犯罪日益与科技手段融合，尤其是在信息获取、诈骗模式设计等方面，利用大数据技术和深度学习等技术手段，大大加大了网络诈骗活动的治理难度。**在这样的大环境下，加强对电信、网络欺诈、建恶意网址等行为的识别分类就显得尤为重要。**   1. 在社会安全方面：   据相关公安部统计，2021 年 1 月至 9 月，全国共破获电信网络诈骗案件 26.2万起，抓获犯罪嫌疑人 37.3 万名，同比分别上升 41.1% 和 116.4%；共紧急止付涉案资金 2770 亿元 。由此可见，电信网址诈骗案件在全国公安机关受理案件中占比较大，对社会安全已经造成较大的影响。因此，如何利用机器学习实现涉诈网址自动分类识别是目前社会面临的大挑战之一。**在此方面下，我们将利用相关模型，提高对恶意网址的识别速度，助力情报机构高效获取研判数据。**  现阶段，涉诈网址呈现变种数量多、传播速度快、影响范围广的特点。尤其需要注意的  是，涉诈网址常针对新型漏洞进行设计，而传统的涉诈网址检测方法，例如维护一个黑名单数据库来简单粗暴地将URL进行识别分类，该方法需要耗费大量的人力、物力，并且很难有效抵御新的未知涉诈网址。**因此，为了精准、高效地规避潜在的诈骗风险和财务损失，满足互联网用户安全上网的需求期望，我们运用globalpointer 模型可以无差别识别非嵌套实体 (flat-ner) 和嵌套实体(nested-ner) ，借助机器学习能力更强的模型结构提高检测的准确度，同时保持实时目标检测所需要的检测速度。**通过以上方法，我们将实现高效率检测已经出现、正在出现以及将要出现的已知以及未知的恶意网址，第一时间保护互联网用户的信息安全，降低识别和监管工作的难度。 |
| 1. 研究内容   **研究内容我们主要侧重于对网络安全中的涉诈网址的自动分类识别与模型优化，此次的研究是通过机器学习的Global Pointer模型来创新识别具有恶意的、非法的、涉嫌诈骗的网址，通过globalpointer的设计思想，可以精准的定位出关键信息从而进行分类识别，利用全局归一化的思路进行命名实体识别（NER），无差别地识别嵌套实体技术为其核心，使其发挥在识别网址上并改进其由数据变化的判断精确度。研究内容我们会分为以下四点进行开发优化：**   1. 机器学习global pointer模型能应用的场景   现代的网络技术不断发展，网络的安全尤为重要，因此，我们研发的基于机器学习的涉诈网址分类识别是利用了global pointer技术和过往数据的支撑所实现的。而基于机器学习也是运用了其模型的可行性、可用性以及高效性，而我们研究的此内容正式借助了global pointer的无差别地识别嵌套实体技术，而GlobalPointer的设计思想就比CRF更合理，GlobalPointer是一种基于span分类的解码方法。而对于任意句子，GlobalPointer构造一个上三角矩阵来遍历所有有效的span，因此超越CRF的性能可以使得落地使用的场景便更为广泛，利用庞大的数据训练和高效率的模型实训，让我们开发的项目可以不仅分类出涉诈网站，还可以分类出其他带有恶意的，如博彩、色情以及会侵害系统安全性的网站,极大可能的发现网址的欺诈性，面对网址的繁多与杂乱，仍能精确的识别出网址是否携带恶意，是守护网络安全的一把利刃。   1. 数据来源的可靠性与训练结构优化   我们在数据集上先采用定位较为泛性的数据进行训练再与用较为精确定位的数据做出的训练结果进行对比，来比较两个结果的精确率，从而决定数据定位的大走向，我们的采用的网站数据集来自于Kaggle，基于Global pointer的无差别识别算法，取得的数据集是符合显示网络的具有真实性，使得训练出来的模型才具有生命力。而我们在于模型上的调整确实会导致准确率下降，因此训练结构优化也是我们重要内容的一环，而我们已有的损失函数也得根据实际的网络情况进行针对性的调整，并且发挥了Global pointer的特点优势。     1. 模型精度调整并在数据中进行测试   在模型精度上我们受限于损失函数的算法，因此调整也要从损失函数开始做起，如果将softmax+交叉熵推广到多标签分类问题将会使得整个模型的精确度得到极大的提升，而能否在多标签上也运用softmax，softmax又称归一化函数，是二分类函数sigmoid在多分类上的推广。多分类的交叉熵损失：    类比得出，每个目标类得分都不小于每个非目标类的得分，如果固定类别，则：    回到多标签上，类别不确定需要一个阈值，希望目标类的分数都大于s0，非目标类的分数都小于s0，应该是：    阈值如果取为0，既s0=0时：    损失函数的资料搜寻后需要不断进行调试知道找出适合自身需求的条件。Global pointer追求的是将首尾看成一个整体，调试完后将需要训练数据集进行精确度反馈，通过不断反复调整找出最为适合方向。突破现有的算法难点是Global pointer比一般CRF更为突出的地方，希望从深度学习中演变出一个智能化模型，计划运用在各大主机系统，让其网络浏览的安全性大为提高。我们进行反复的演算、推论、实验、实际网址识别、恶意拦截等追踪到许多仍需改进的通点。   1. 开发后以适配于实际网络情况、识别算法的优化完善   开发完成后需要进行最后的环境调试，因为所有主机系统的兼容性并不一定可以完全容纳这一模型，需要根据实际的嵌入来发现不适配的问题所在，提高其用户体验感，让其精密的算法模型可以更为简单的部署在其主机，让在因为环境导致的识别失误的可能性尽量减到最小，抽象的模型想要兼容市面上的次流环境应该是仍需改进，面对网络繁杂、拦截失效等也是我们不能忽视的问题。基于Global pointer开发的模型所需部署环境也是要从严苛逐渐向大众转移。我们识别算法之所以不用CRF是因为CRF进行了标签约束，这种序列标注采取BILOU标注框架，每一个token只能属于一种，不能解决重叠实体问题，如：    这里出现了两个实体《叶圣陶》、《叶圣陶散文选集》，分别代表了作者和作品两个实体，传统的token只能属于一种Tag，无法解决此类问题，因此我们选用了Global pointer来取代CRF，难点自然也就少了些许，并且识别算法使得数据更为精确，研究其识别算法使其效率更快。 |
| 1. 国、内外研究现状和发展动态   1.国外研究现状  Szegedy等人在生成对抗样本时首次提出了梯度方法，通过盒约束有限内存BFGS(limited-memoryBroyden-Fletcher-Goldfarb-Shanno,L-BFGS)优化算法最小化原始图像和对抗样本之间的欧几里得距离，向原始图像中添加噪声来生成图像对抗样本，并攻击深度神经网络[13]。这种方法能够有效地生成对抗样本，但寻找最优解所需的计算开销也非常大。Goodfellow等人提出一种快速梯度下降方法(FastGradientSignMethod，FGSM)，通过计算损失函数关于输入的梯度，在每个特征上仅执行一次沿着梯度符号方向上的梯度更新，最小化添加到原始样本中的最大扰动强度，导致模型产生错误的分类结果。Grosse等人基于FGSM的思想，对Android恶意软件提取出相关API及硬件申请等信息，构建0/1特征向量，最后对特征向量进行特定修改后指导生成了Android恶意软件对抗样本。FGSM方法在有限的计算开销内取得了相对较好的效果。  Garlini和Wagner将攻击表示为优化问题，基于L-BFGS创建了一种新的攻击方式，称为Carlini&Wagner攻击(C&W)，通过改变损失函数，取得了比L-BFGS方法更优越的效果，并且能够抵御防御蒸馏和对抗训练。  梯度方法通常假设目标攻击模型的内部结构和参数已知，以便获取梯度信息进行反向传播训练，这样的攻击方法也称为白盒攻击。但实际上攻击者无法获知这些权重参数，只能进行黑盒攻击。Xu等人使用遗传编程方法在ContagioPDF恶意软件数据集上执行黑盒攻击，规避恶意软件分类器。该策略通过插入、移除或替换PDF文件树结构中的对象来执行变异操作，直至变异样本成功逃过分类器检测。在每次迭代过程中，仅留下根据oracle指示保留了恶意行为的样本。最后针对PDFrate和Hidost两种分类器进行了测试，实现了100%的逃避率。Calleia等人采用了类似的基因编程方法修改恶意Android应用程序，以将其错误地标记为DREBIN数据集中的其他恶意软件家族，从而攻击RevealDroid分类器[22]。该分类器由决策树组成，根据应用程序是否具有各种静态功能来进行决策。该攻击策略通过执行功能添加(包括向现有应用程序添加新功能，例如新的API调用和新权限)实现了攻击目的。另外，考虑到程序的完整性可能会受到损害，该策略避免了功能移除。  Papernot等人训练了一个本地模型替代目标深度神经网络(DeepNeuralNetworkDNN)模型，通过观察目标DNN对特定输入的模型输出，以此指导本地替代模型生成恶意软件对抗样本，最终证明了这些对抗样本确实能被目标DNN模型错误分类。Goodfellow等人探索了对抗性攻击的属性，即模型之间的可转移性。当对抗样本成功地被模型错误分类时，常常也会被其他模型错误分类，即使其他模型具有不同的结构或拥有不同的训练数据集，并且这些模型经常在多类别场景的预测类别上达成共识。当执行黑盒攻击时，这种属性能够发挥强大的作用，攻击者可以创建一个在遵循相同分布的数据上进行训练的替代模型，最终能够生成在目标模型上被错误分类的样本。  根据这种攻击可转移性，胡玮玮和谭颖基于GAN提出一种新模型MalGAN，在目标神经网络上生成了恶意软件对抗样本。GAN最初是由Goodfellow等人提出，是一个由两个神经网络进行互训练的机器学习系统。建中一个作为生成器，目标神经网络上生成了恶意软件对抗样本。GAN最初是由Goodfellow等人提出，是一个由两个神经网络进行互训练的机器学习系统，其中一个作为生成器，另一个作为鉴别器，生成器试图生成能够被鉴别器错误分类的样本，鉴别器试图区分真实样本和生成样本。MalGAN的生成器使用128个恶意软件API调用作为输入特征，为其构建0/1特征向量，通过向输入特征向量中加入噪声矢量来生成对抗样本。Rosenberg等人利用自适应进化算法(EvolutionaryAlgorithm,EA)向离散的API调用序列中添加大量无效操作的API，通过对数回溯方法迭代减少了插入的冗余API的数量。  2.国内研究现状  针对信息和通信安全问题，PKI证书是国内外常用来保障服务可靠性、保密  性和完善性的工具，例如OpenADR2.0协议就常借助于RSA和ECC证书，这些证书满足NIST网络安全方面的需求。此外该协议借助于公钥认证机制和签名保障通信安全。根据DR(DemandResponse，DR）系统的安全和隐私目标，有研究者在物联网设备信息安全的前期工作基础上，定义了一个合适的、现实的对手模型来分析需求响应系统了。针对需求侧设备的网络环境所涉及的问题，许多学者集中在动态完整性测量模型、可信计算技术和匿名远程证明协议等方面进行研究。与上述模式不同的是，有一种安全匿名智能电表发送电能计量数据的方法(如每隔几分钟发送一次)。这种方法为难以与具体智能电表或容户关联的真实匿名电表读数提供了第三方托管机制门。有学者对基于匿名网络的智能电网隐私增强协议(PEPs）进行了分析和评价。同时对基于居名网络的PEPs的安全性和效率进行了评价，展示了另一种保护物联网设备免受潜在攻击的模式。除了国内的可信密码模块(即国家电网公司的安全芯片〕外，有研究者对智能电表原有实时传输用电的协议进行升级，增加了用户敏感信息保护的部分，利用环形签名新技术匿名化用户身份，不仅如此智能抄表平台的配置也被平台属任证书所隐藏。出现了一种新型的隐私保护智能测量系统，它给智能电表提供了远离攻击的机会，电表只需要在数据中加入噪声，井使用有效的流密码器对结果进行加密。新的智能测量系统具有众多的优点，但其安全性和隐私性的保障成为新的问题。现有的方法主要是通过隐私保护来解决这个问题，但是都存在一定的局限性。随机化算法在隐私保护方面得到充分应用，有研宄将其运用于功率数据防护、敏感属性保护等。优于上文所提PEPs，新的用电传输协议可以在定义的仪表群上计算总的仪表测量值，支持攻击检测方法，并对网络管理以及仪表测量值进行进一步的统计处理，而不显示单个仪表读数的任何附加信息。  然而以上提到的物联网设备安全方案主要着眼于信息安全和客户隐私保护，对于DDoS这种可同时采用多种方法的复杂攻击类型则关注较少。随者海量运行数据的积累和物联网技术的快速发展，攻击检测方法的应用也在研究之中，以改善需求侧物联设备的安全问题川。现己存在一些DDoS攻击检测方法，比如通过流特征提取采用基于聚类的算法进行柠测。还有方法基于流信息闵值来判定流量正常与否，优点是能够检测多种入侵，不足之处在于需要耗费巨大成本预处理收集的流量。与这些方法相比，机器学习检测方法具有更高的检测准确率和成功率，关键在于如何选择和优化算法来提升检测准确率。  对于DDoS攻击检测最关键的在于检测方法的设计。灵活的检测技术能够确保快速、准确地对DDoS攻击进行有效识别，从而为下一步的防御提供有力保障。  目前DDoS检测相关研究中普遍采用误用检测(misusedetection）和异常检测(anomalydetection）两种策略。其中研究最多的是基千异常的检测，即通过发现网络中存在的异常流量并进行归类来识别潜在风险。该基于异常的检测方法可进一步细化为监督、半监督及无监督三类。针对有监督的学习方法，必须学习带标签的大量数据来生成模型，再用于检测无标签数据。针对半监督的分类模型，只需比对正常数据与待检测数据来发现异常行为，但不能得到确切异常类型且结果取决于正常参数。针对无监督的学习方法，不需正常数据的参与而只对未知数据直接处理，通过发现整个数据集中的差异性数据来判别异常。具体来讲，有监督的学习方法可划分为基于分类、基于回归两类，半监督的学习方法又包括聚类和分类两种，无监督的学习方法分类较为繁玉，主流方法涉及基于距商、基于分布、基于密度等几种。  具体来讲，监督学习分类检测方法预先生成模型，并根据模型做出高效判断。  可细化为基于分类和基于回归的监督。  监督学习网络异常检测分类算法从非异常数据中筛选特征参量，并用这些参量训练模型对等待检测数据划分类别。基于分类的监督包括两个环节，在第一个环节中把正常数据各类别区分开来，在第二个环节中用己经区分类别的数据生成训练模型环节中把正常数据各类别区分开来，在第二个环节中用己经区分类别的数据生成训练模型，用于识别需要检测的数据。KNN、SVM等都是研究较多的主流分类方法。回归监督则不同于分类监督，通过对一系列数据点拟合分析揭示变量输入输出间作用关系，以此作出预测。逻年回归、KNN等在回归范畴之内。  半监督异常检测学习模型在监督学习基础上加入了待检测数据，即有标签数据和未知数据共存，经过不断学习这些混杂数据生成检測模型，可细化为面向聚  类和面向分类两种方式。  在面向聚类的半监督异常检测方法中，并不刻意区分数据是否带有标记，统一求取样本数据相似度并将数值较高的聚成一族。整个聚类过程在有类别标记的数据监督下进行，聚类效果远胜于完全无标记样本。面向分类的半监督学习模型同样借助于己知数据，不同之处在于直接用这些已知数据生成性能较好的分类系统，再用于未知数据的异常检测。这种方法本质上与监督学习类似，都是基于己有标签的数据对未知数据辦别，极易发生过拟合的情况。  无监督异常检测学习模型面向的是没有任何标签的数据集，通过发掘数据本身的潜在特性来学习，并应用于新类型的判别。细化之下有距高判别、密度判别  等几种。  其中，距离判别方法建立在大量的距离计算工作之上，依据距离判别样本数据状态，如果是大数据样本或存在样本不均衡现象就不太灵敏;基于分类的方法在己知统计意义上的数据分布情况下，通过大量运算利用待测无标记数据的模型来检测无标记数据，检测结果极为准确，类似于基于距离的无监督学习，基于密度的学习方法需不断对相近数据的密度运算以作出精准判断，同样建立在较商运算代价上因此对大容量数据样本并不实用。  己有研究提出对KNN分类检测性能的提升方法，基于界标的谐聚类算法对大容量样本分筷后在族内查找k近邻，最终分类结果过于依赖聚类算法的表现，但对于大容量样本的处理十分高效。另有文献将主成分分析PCA(principalComponentAnalysis）算法与KNN结合进行网络流量降维检测，且组合成族的方法不需数据逐一判别，但降维处理中极易丢失关键数据特征使模型性能下降。同样是对高维数据检测，苏州大学的学者提出一种考虑角度方差划分数据集网格的思路，降低了算法时间复杂度。华中科技大学的研究者设计了特征提取基础上的LSTM人工神经网络异常检测系统，对用户行为模式进行预测，但存在多线程系统搜索的性能限街120。陈瑜等人提出将遗传算法和决策树组合算法用于攻击流特征的优选，并集成梯度提升和决策树对特征集检测，但存在计算复杂耗时久的弊端。贾斌等学者设计了一种基于RotationForest集成策胳的混合异构多分类器，优点在于泛化能力强、效果稳定，缺点在于与同构分类器检测性能相当，且实验方法并未在真实网络环境中得到验证。经海飞等人则设计了软硬件协同的电力监控系统DDoS攻击入侵检测系统，将支持向量机模型与实时采集报文的防火墙装置封装一起，实现了实时在线检测。除此之外，还将数据统计分析的思想融合于DDoS攻击检测研究，这类型方法建立在大量统计基础上，运算代价大且计算过程复杂。另有文献设计了两级攻击检测策略，第一级序贯概率比检验快速查找攻击端口，第二级轻量级梯度提升机判别攻击类型，能够快速准确响应攻击。此外还有国外学者提出一种修正的多图表累计和算法来自适应检测攻击。还有方法针对工业物联网的DDoS攻击，提出了多级DDoS防御框架，涵盖云计算等多种级别。DDoS攻击检测从根本上讲是对每个数据包的分类问题，关键在于检测模型的构建。而基于分类的监督学习充分学习了待标记的网络特征流量本质特性及相关性，构造的分类检测模型能够更加合理预测分类结果。其中经典分类算法KNN由于简单学习、可解释性强的优点，常用于多分类情况。基于此本文从基于分类的监督学习算法KNN出发，探索适用于需求侧物联设备的攻击检测方法。  机器学习安全性问题的暴露激起了大批学者研究兴趣，为了对这个问题进行系统、统一的处理，研究者们从理论着手，分析传统的机器学习算法，针对其不足设计一种新型的算法，总体而言存在以下三方面需解决的问题。  (a）分析分类算法中存在的漏洞及对应的攻击手段。在不同攻击场景、不同敌手棪型下机器学习算法通常存在不同威胁，诸生研究者通过设计-攻击模型证明了机器学习中的安全风险。Fredikson等将病人的人口统计信息作为轴助信息，预测药物剂量的线性回归模型作为目标西数，通过反复请求目标模型恢复出训练集中用户的敏感数据，证明目标模型存在安全性问题，其基本思路是根据输出结果找到对应可能性最大的一组输入。TraTEr等提出了一和等式求解模型参数的方法，根据模型己知的结构信息和标签信,息，反复请求模型，根据等式求出目标模型的参数，进而实现对模型的攻击。  (b）开发新的评价机器学习安全性的指标。在机器学习中，传统性能估方法的有效性是建立在测试集和训练集符合同一概率分布的前提条件下，因而忽略了一些与安全性相关的问题。为了解決这个问题，研究者提出使用一种特性曲线来评估算法的安全性，其中较为出名的便是ROC曲线，用来衡量算法的检测率与虚报率，该曲线的优势在于当正货样本的分布出现变化时，其形状能够基本保特不变。在此理论基础上，Laskovd等人提出了一个用于对机器学习算法进行定量安全性分析的框架，该框架的核心功能有部署机器学习模型、对攻击者进行行为约束、计算最优攻击方式和攻击最大北响上限。他们将此框架应用于在线质心异常检测的场景中，通过实验证明理论方法的有效性。  (c)在对抗环境中设计相应的抵抗措施保证模型的安全性。有学者针对这个问题提出了两种解决方法，分别为最大-最小方法和博弈论方法。最大-最小方法假设模型设计者和攻击者的损失函数是对立的，此时会产生一个相对简单的优化问题，可通过求解最优值得到解决方案。博弈论方法适用于非对抗性损失，例如垃圾邮件过滤者希望准确识别合法电子邮件，而垃圾邮件发送者则试图提高垃圾邮件漏检率。在一定条件下，这类问题可用纳什均衡法(Nashequilibriumapproach)解决。两种方法都为各自的安全问题提出了一个解决方案，但现实中的攻击却过于复杂和多面，无法融入到现有的博弈论方法中，比如躲避攻击，旨在通过操纵恶意测试样本避免检测。Biggio等对躲避攻击进行深入研究，并且提出了一种简单有效的基于梯度的方法，对几种广泛使用的分类算法的安全性进行系统的评估。通过增加攻击者对系统的了解和处理攻击样本的能力来模拟不同风险等级下的攻击场景，使机器学习模型设计者能够更好地了解躲避攻击下分类器的性能，提醒他们选择更好的模型或调整该模型下的参数设置。基于此研究，Demontis等提出了分类器敏感性的概念，从多方面证明了线性模型中权重的均匀分布将有效提高躲避攻击下模型的安全性。Demontis等人还重新设计了线性SVM算法，在求解最优参数时使用了随机梯度(StochasticGradientDescent)算法。为提高模型的安全性，他们对线性SVM算法进行改进，将每次求解的最优参数向量的权重值限定在特定区间内。实验结果表明，改进后的SVM算法在几乎不影响准确率的情况下明显提高了模型的安全性。但该方法仅应用在安卓恶意软件测试，当检测其他恶意攻击时，若将特征向量均限定在特定区间。可能会出现丢失范围外其他最优值的情况。  3.Web攻击检测的研究现状  为减少Web攻击带来的安全隐患问题,国内外的学者分别从推理树验证法、规则集匹配法。污点追踪技术和机器学习四个方面进行的广泛而深入的研究。Su等提出的SQLCheck方法和Buehrer等提出的SOLGuard方法均是基于推理树验证的理论，这两种方法的理论思想相似，且都是在攻击语句操作数据库之前进行攻击检测，在使用SQLGuard方法进行攻击检测时，Buehrer等并未给出准确率等相关衡量指标的结果。Reisl4)等人提出BrowerShield方法和Yu等人提出的CoreScript方法主要依赖于规则集匹配理论。BrowerShield方法主要用于检测和防御XSS攻击，通过在服务器端建立一个代理，依据服务器的反应解析HTML脚本，若发现解析出的脚本是恶意脚本，则置换成规则库中功能相同、结构不同的安全脚本。CoreScript方法的核心思想是将不受信任的JavaScript代码重写，并在适当时提示用户以决定如何运行。污点追踪技术的检测方法关键在于如何标记及跟踪污点数据，Haldar在Java虚拟机上事先写入一些方法，将用户输入的数据标记为污点数据，当污点数据用作链接该方法的参数时便会引发异常。为提高检测的准确性，Xu等提出应对污点数据做进一步的检查。他们在内存字节级别跟踪污染信息，为了确定污点数据的合法性或恶意性，他们另外使用正则表达式的方法查询污点数据中是否含有可疑字符。  机器学习的检测模型思想最早来源于Denning,他提出了一个实时的入侵检测系统模型。该系统能够成功检测攻击者的入役、渗透和其他形式的计算机滥用操作。监控系统在训练阶段可以学习所有合法的代码语句。在实际检測阶段，只有那些被识别的语句才可被调用执行。基于这样的思想，更多的研究者一方面针对不同情景使用不同的机器学习算法，并将多种机器学习算法的检测结果进行对比，另一方面改进特征提取方法以提高检测的效果。Kamtuo等2提出了一种如何提取数据集重要元煮并将其标记为输入属性的方法,该方法总共提取出20个属性值，将这些输入屆性发送到机器学习模型，最终输出sQL注入攻击的预测。该方染选择了生种机器学习模型进行測试，最终结果农明，决策树(DecisionTree,DT）花费的时间开销最小，检测的准确度最高。决策树能够根据一系列的屆性值做出决策，但它必须要定期建立，一旦攻击特征库发生改变，所有的树结构必须要重新建立，使用决策树检测攻击手段时可能出现不稳定的现象。Su等I50提出了基于执行路径控制流的轻量级在线模型和基于特定应用程序的离线模型，离线模型使用了随机森林算法（RandomForest，RF)，对SQL注入攻击和XSS攻击同时进行检测测试。随机森林是由多个决策树组成，相较于决策树，随机森林不仅更加稳定，同时也能提高检测精度，但会增加时间开销。昊少华等(提出基于SVM的Web攻击异常检测方法，利用人工挑选和数据统计概括出6个特征，将原始样本集转化成固定维数的特征向量。这种方法能够有效检测出SQL注入攻击和XSS攻击，但未过它们攻击的多样性与变异性进行深入研究，且并末明确给出该方法的漏报率。陈君新!2对变形的XSS攻击进行了深入研究，首先分析样本总结XSS的普通攻击方式和变形攻击方式，然后使用词集模型对XSS攻击样本进行特征提取，最后对比隐马尔可夫模型（HiddenMarkovModel,HIMM)和支特向量机(SupportVectorMacline，SVM))两种检测校型的检测效果，结果花明HIMM模型在检測变形的XSS攻击时漏报率、误报率等方面更胜一筹。胡松峰等提出基于朴素贝叶斯算法(NaiveBayesianModel,NBM)的SQL注入攻击检测方法，改进特征提取的方法，对提取出的特征加入词法分析过程，根据词法分析的结果设计一个闪值进行去噪，从而提高SQL注入攻击检测的准确率。然而在提取特征时，特征项设置过多，这将影响机器学习算法的学习的效率。Komiya等使用TF-IDF方法计算每个特征向量的权重，将处理好的特征向量放入支持向量机、朴素贝叶斯及上近邻(k-NearestNoighbor,KNN)三种机器学习模型中训练、测试并进行结果分析，实验结果表明，三种机器学习模型对SQL注入攻击和XSS攻击检测效果均表现较好，但测试时使用的数据集规模较小。 |
| （四）创新点与项目特色   1. 基于机器学习实现对涉嫌诈骗的网址自动分类识别   近年来随着互联网的发展，网络中存在越来越多的威胁。虽然网络安全问题层出不穷，但大部分攻击技术都是通过传播被破坏的URLs去实现的，亦或者这些URLs的传播是攻击操作的关键部分。同时，URLs 作为万维网上文件和资源的全球地址，针对 URLS 特征进行检测系统的设计与研究实现更具有通用性，其中基于 URLs 特征提取的方法被重点关注，相关研究也日益增加。  采用基于特征提取的恶意网站的检测方法中，机器学习算法的应用是必不可少的，适合的方法十分有利于检测准确率的提升，因此机器学习的方法选择也十分重要。通过对大量学者研究的分析总结，发现分类算法中主要运用传统的机器学习算法，包括决策树、贝叶斯或朴素贝叶斯、支持向量机、最大熵模型、逻辑回归、多层感知器、KNN等。然而对于利用以上多种算法的恶意URLs检测方法都未能成功进入市场完成实际应用这一现象，其原因除了分类算法的不足外，还因未能发现广谱性的特征空间，目前基于特征提取的恶意 URLs 检测均依赖于人工规则的设计，但此类规则设计要求高，更新慢，无法实时适应多变的URLs 数据，新的特征提取方式也有待研究。所以在此基础上，我们大胆猜想，能不能通过运用GlobalPointer去做分类识别算法，来实时适应多变的URLs数据，以提高精度。  2、运用GlobalPointer做分类识别算法，提高精度，减少人力物力  对于涉诈、恶意网址检测方面，较为流行的是直接维护一个黑名单，当下次浏览器在访问某个网址的时候，如果该域名已经存在于blacklist中，那么直接阻止访问或弹出警告提醒即可，这种方式最大的一个缺点在于不能及时更新，同时抗干扰能力也比较差，如果人家改了一个域名、或者域名是算法自动生成的，那这种方法就失效了。基于黑名单及人工规则的检测方法需要基于专家设计特征提取规则，需要深层的专家知识，消耗大量人工，因此高质量人工提取规则的设计越来越困难，特征的提取与更新时效性也越来越差，使得越来越多新兴涉诈网址有机可乘。  为了改善以上不足，我们基于机器学习提出了一种更为新颖的GlobalPointer，采取了全局归一化的思路来进行命名实体识别（NER），可以无差别地识别嵌套实体和非嵌套实体，在非嵌套（Flat NER）的情形下它能取得媲美CRF的效果，而在嵌套（Nested NER）情形它也有不错的效果。还有，在理论上，GlobalPointer的设计思想就比CRF更合理；而在实践上，它训练的时候不需要像CRF那样递归计算分母，预测的时候也不需要动态规划，是完全并行的，理想情况下时间复杂度是O(1)！  2377306125  图1  上图（图1）为GlobalPointer多头识别嵌套实体示意图。具体来说，假设要识别文本序列长度为n，简单起见先假定只有一种实体要识别，并且假定每个待识别实体是该序列的一个连续片段，长度不限，并且可以相互嵌套（两个实体之间有交集），那么该序列有多少个“候选实体”呢？不难得出，答案是n(n+1)/2个，即长度为n的序列有n(n+1)/2个不同的连续子序列，这些子序列包含了所有可能的实体，而我们要做的就是从这n(n+1)/2个“候选实体”里边挑出真正的实体，其实就是一个“n(n+1)/2选k”的多标签分类问题。如果有m种实体类型需要识别，那么就做成m个“n(n+1)/2选k”的多标签分类问题。这就是GlobalPointer的基本思想，以实体为基本单位进行判别，正如图1所示。 |
| （五）技术路线、拟解决的问题及预期成果  技术路线：1.通过文献查阅法对网络诈骗的案例信息进行分析整理，对国内外诈骗网址手段进行初步的了解。  2.通过问卷调查法，针对所做的项目以电子档纸质档结合的形式，就是否有必要使用机器学习的方式对涉诈网址进行识别分类对大众进行问卷调查，然后回收整理，统计和研究。  3.通过实地调查研究，切实了解是否有必要使用机器学习的方式对涉诈网址进行识别分类达到防诈效果，并获得分析调研数据，得出切实有保障的结论。  拟解决的问题：利用机器学习解决涉诈网址的识别分类问题。  预期成果：1.对已有成果的技术发表相关论文。 |
| 1. 项目研究进度安排  |  |  |  | | --- | --- | --- | | 第一阶段 | 2023.2-2023.4 | 通过中国知网查找文献研究如何解决涉诈网址识别分类问题——机器学习。 | | 第二阶段 | 2023.3-2023.5 | 利用珠海科技学院校内阿里云大数据实验室所提供的资源，构思出机器学习对涉诈网址识别分类的解决路线。利用校内实验室资源，研发算法模型。 | | 第三阶段 | 2023.4-2023.6 | 深入讨论我方团队项目系统细节 | | 第四阶段 | 2023.6-2023.9 | 团队成员经过与指导老师的探讨，确定系统所使用的算法。 | | 第五阶段 | 2023.10 | 准备材料拉赞助，创建公众号运营。 | | 第六阶段 | 2023.6-2024.2 | 继续完善项目系统，并预计与公司合作。 | |
| （七）已有基础：与本项目有关的研究积累和已取得的成绩  1.已获珠海科技学院“数聚珠科，智领未来”人工智能技术创新大赛二等奖  15334a5db6ffb327c6139210edcf5c6 |
| （八）已有基础：已具备的条件，尚缺少的条件及解决方法  已具备的条件：基本技术已大概掌握，思路框架清晰。  尚缺少的条件：1.大学生创业，缺少资金。  2.识别分类技术不够完善。  3.涉诈网站五花八门，种类繁多，难以面面俱到。  解决方案：1.团队优势，积极寻找资源客户，向与该项目相关的公司申请投资，并与其保持联系。  2.寻找技术指导、合作伙伴。  3.多向专业老师询问，多查询国内外相关文献。 |
| 项目预期成果：项目的预期成果种类、数量及完成时间  □申报校级预期成果  ☑拟申报省级、国家级预期成果 |

**三、经费预算**

|  |  |  |
| --- | --- | --- |
| **开支科目** | **预算经费（元）** | **主要用途** |
| 1. 业务费 |  |  |
| （1）计算、分析、测试费 |  |  |
| （2）能源动力费 |  |  |
| （3）会议、差旅费 |  |  |
| （4）文献检索费 |  |  |
| （5）论文出版费 | **2000** |  |
| 2. 仪器设备购置费 |  |  |
| 3. 实验装置试制费 |  |  |
| 4. 材料费 |  |  |
| 合计 |  |  |

注：如上述支出科目中一项或几项无预算，请在相应栏填写0，不得删减。

**四、指导教师意见**

|  |
| --- |
| 指导教师（签名）： .  年 月 日 |

**五、院系大学生创新创业训练计划意见**

|  |
| --- |
| 学院（签章）： .  年 月 日 |

**六、学校大学生创新创业训练计划专家组意见**

|  |
| --- |
| 专家组组长（签名）： .  年 月 日 |

**七、学校审批意见**

|  |
| --- |
| 学校（签章）： .  年 月 日 |