|  |  |  |  |
| --- | --- | --- | --- |
| **论文开题报告进展表** | | | |
| **姓 名** | 胡希文 | 填表时间 | 2109年8月28日 |
| **论文题目** | 大致方向为机器学习用于网络安全 | | |
| 已读文献1：基于 Python 的机器学习入侵检测的研究  文献来源：网络安全技术与应用期刊  文献主要研究内容：该文献主要是基于流量是否恶意来进行入侵检测。首先是用流量探针来捕获流量，然后将捕获的数据报文经过特征提取，之后组成一个样本，然后将样本用于训练模型和模型判断。该文献直接使用KDD99比赛中提供的五百万现成的样本，作为训练模型和模型评估的数据集合。由于入侵检测的样本判断可以看做是分类，故该文献中使用的是分类器的相关算法，它选中的是Logistic Regression算法来做分类器，之后直接使用KDD99比赛的数据集，用python做了实现和准确度测试。  已读文献2：网络入侵检测中的机器学习方法与应用  文献来源：2019年北邮硕士论文  文献主要研究内容：  该文献研究了用机器学习的方法来做网络入侵检测。它首先是介绍了入侵检测和机器学习的相关内容，然后将整个研究分为三个阶段，分别是预处理阶段、入侵检测阶段、误报消除阶段。  预处理阶段先是将数据格式进行转换，转换成数值类型，然后将得到数值进行标准化，之后再将数据归一化。之后对特征选择方法进行了探讨，特征选择方法可以分为过滤式、包裹式和嵌入式。然后选了4种过滤式、1种包裹式、1种嵌入式在测试环境下，通过KDD99数据集进行了实验。  入侵检测阶段首先是从理论上评估和探讨了朴素贝叶斯、KNN算法、决策树、SVM向量机、人工神经网络这几个分类器的性能。之后是选取KNN、SVM、朴素贝叶斯三种算法进行实验。得到结论是，从运行时间的角度出发，数据集规模不大的情况下，三种算法在计算时间方面差距不大；从检测率的角度出发，递归特征消除法子集更适用于KNN，树形特征子集更适用于朴素贝叶斯，SVM对两个子集的特征都不是很敏感；误报率方面，朴素贝叶斯对于递归特征子集可以保持较低的误报率，SVM对于树形特征子集误报率较高切波动较大。  误报消除阶段研究使用的是DARPA2000数据集，首先是从理论上评估了几种聚类算法的性能，分别是K-means聚类、层次化聚类、基于密度的DBSCAN聚类、基于模糊的FCM聚类，根据实际需求出发，选择了K-means、DBSCAN和FCM算法进行实验性能探测。实验结果表明，三种算法的消除率都很高，其中K-means最高，但波动最大，FCM最稳定。  已读文献3：基于深度卷积神经网络的入侵检测研究  文献来源：2019-08-12计算机科学  文献主要研究内容：  该论文研究了将深度卷积神经网络应用到入侵检测中去。首先是提出了一种将网络数据转换为图像的方法，然后再针对图像，使用深度卷积神经网络对其进行处理。  该研究的流程是数据集准备-->数据预处理-->图像化处理  数据集采用KDD99数据集进行实验，抽取10%的KDD数据集作为训练集，Corrected数据集作为测试集。  数据预处理流程与上一篇论文相似，先是数值化，然后归一化得到标准数据。预处理之后的取值在[0,1]之间，其中每条数据的维度为122维，数据集的维度很大。  数据图像化之前，先是删除其中一维在训练集和测试集中特征值全为0的数据，将122为数据变为121维数据，然后将121维数据转变为11\*11的图像。由文献给出的转化后的图像看到同种类型的数据图像的相似性很高，证明这个思路是可行的。  之后是CNN网络。该论文使用的CNN网络分为8层，第1层为数据输入层，第2层和第4层为卷积层，第3层和第5层为池化层；第6层和第7层为全连接层；最后一层为CNN结构的输出层。  实验采用准确率(Accuracy,AC)、误报率(False Alarm Rate,FA)和召回率(Recall,RE)作为实验效果判断标准。该实验设置了3种不同结构的CNN模型，一组卷积和池化层结构、两组卷积和池化层结构和三组卷积和池化层结构。实验结果是具有两组卷积和池化层结构的CNN网络在三个判断指标上明显优于其他两个结构。  最后，对比了目前常用的入侵检测算法，KPCA-SVM、BPNN、RNN、和DBN算法，从准确率、误报率、检测速率、测试时间四个方面来对比，发现该论文的算法在这四个方面都优于上述几个算法。  （若有更多，按此格式填写） | | | |
| **开题报告进展情况**： | | | |
| （本周关于开题报告所进行的相关工作）  1、重新确定方向为基于机器学习做入侵检测  2、机器学习的相关学习  3、python的相关学习  （下周计划）  1、挑重点学习机器学习或者是深度学习  2、继续看入侵检测方面的论文  3、尝试实际使用模型 | | | |