[文章编号] 1007－7405(2019)03－0234－07 DOI:10.19715/j.jmuzr.2019.03.13

基于深度卷积神经网络的SQL注入攻击检测

叶永辉，谢加良，李青岩

（集美大学理学院，福建 厦门 361021）

[摘要] 结合自然语言处理技术，采用卷积神经网络算法训练SQL注入检测模型，主要包括文本分词处理、提取文本向量和训练检测模型三个部分。实验结果与BP神经网络算法结果对比，发现基于卷积神经网络的SQL注入检测模型仅需提取用户输入的信息，就可以对攻击行为进行检测，具有很强的预测能力，同时针对变异SQL注入攻击具有良好的识别能力。

[关键词] SQL注入；检测；CNN；自然语言处理

[中图分类号] TP 309

SQL Injection Detection Method Based on Deep Convolutional Neural Network

YE Yonghui, XIE Jialiang, LI Qingyan

(College of Science, Jimei University, Xiamen 361021, China)

Abstract: This paper combines natural language processing technology and uses convolution neural network algorithm to train SQL injection detection model. It includes three parts: text segmentation processing, extracting text vectors and training detection models. By comparing the BP neural network algorithm, the experimental results show that the SQL injection detection model based on the convolution neural network only needs to extract the information from the user input, and can detect the attack behavior, which has a strong prediction ability and is good for the variant SQL injection attack with clockwise. At the same time, it has good recognition ability against variant SQL injection attacks.

Keywords: SQL injection; detection; CNN; natural language processing

0 引言

随着计算机技术和互联网的迅速发展，Web应用迅速崛起，网络安全问题备受关注[1]。SQL注入漏洞攻击[2]是目前网上最流行最热门的黑客脚本攻击方法之一[3]。SQL注入具有攻击危害大、类型多、变异快、攻击隐蔽等特点。因此，SQL注入漏洞的检测和防御一直是Web安全领域关注的重点。

SQL注入[4]是指攻击者通过恶意查询语句来获取服务器数据库信息的攻击行为，可通过预编译、转义、过滤关键字、部署WAF等方式防御攻击。SQL注入方式有多种，基于攻击方式的不同可分为联合查询注入、报错注入、重言式攻击等[4]。基于此，入侵检测技术保障网络安全的重要性也与日俱增。入侵检测技术主要分为基于分类、基于聚类、基于统计和基于信息理论4大类算法[5-6]。李红

[收稿日期] 2018－07－05

[基金项目] 国家自然科学基金资助项目（11371130）；福建省自然科学基金资助项目（2017J01558）；福建省中青年教师教育科研项目（JAT160696，JA15265）

[作者简介] 叶永辉（1995—），男，从事机器学习、信息安全方面的研究。通信作者：谢加良（1981—），男，副教授，博士，主要从事不确定信息处理、信息安全方面的研究。E-mail:xiejialiang@jmu.edu.cn

灵等[7]在解决检测方法的适用性和提高注入检测准确率方面提出了 SVM 算法训练注入检测模型；杨连群等[8]提出结合隐马尔科夫模型来降低 SQL 注入检测误报率；张志超等[9]提出了 BP 神经网络训练检测模型，其特点是快速高效。卷积神经网络[10]是一种前馈神经网络，相较于其他机器学习算法[11-12]，它通过卷积层和池化层的优化降低了网络参数个数，使卷积神经网络的计算量大大降低。卷积神经网络成功应用于图像处理[13]、视觉领域[14]以及围棋人工智能程序[15]等。为了进一步提高对变异攻击的识别率，降低人为因素对检测模型的影响，本文提出使用卷积神经网络（convolutional neural network，CNN）算法[10]训练注入检测模型。

1　基于卷积神经网络的 SQL 注入攻击检测方法

SQL 注入检测模型的主要原理是通过拦截客户端与 Web 服务器的通信数据，利用 SQL 注入模型对数据内容进行检测分析。若存在攻击行为，则不对数据包进行转发处理，否则向服务器转发数据包。

本文提出的基于卷积神经网络算法的 SQL 注入检测系统由以下三个部分组成：

1）文本分词处理　针对文本中的链接、数字进行规范化处理，从而降低分词数量，减小无关变量对系统模型的影响。

2）提取文本向量　使用 Word2Vec 工具中的 CBOW 算法，对已经过分词处理的训练样本进行词汇模型训练，进而将文本数据转化为文本向量。

3）训练检测模型　设计卷积神经网络结构，选取卷积层、池化层以及激活函数等参数，进行模型训练。

1.1　文本分词处理

1）URL 解码。浏览器向服务器发送数据时，客户端将用户输入的参数进行打包编码后发送到服务端，采集的训练样本往往都进行过编码处理，在训练前则需进行解码。针对存在多重 URL 编码的样本需采用递归 URL 解码[16]进行解析，以保证数据编码的一致性。

2）对已解码的数据进行规范化处理。进行的操作主要有：将 URL 中数字替换为“0”；替换超链接为“http://u”的形式；对于数字、超链接等无关因素进行统一规范处理，以降低分词后的分词数量。

3）进行分词处理。分词处理模块以特殊符号（‘@ #/空格等）为分隔符对样本进行分割处理，将文本字符串转化为文本序列的形式，既保留原本文本信息，也方便进行文本特征提取。

1.2　提取文本向量

Word2Vec[17]是 Google 在 2013 年开源的一款将自然语言转化为计算机可以理解特征向量的工具。Word2Vec 主要有 CBOW 和 Skip-Gram 两种。本文采取 CBOW 模型进行训练，该模型由输入层、映射层和输出层组成，其神经网络结构示意图如图 1 所示。

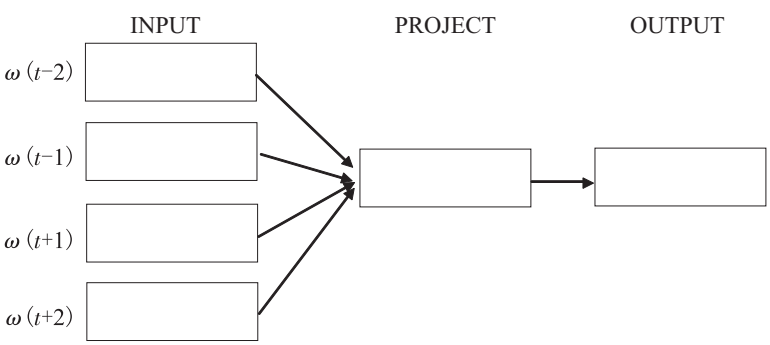


图 1　CBOW 的神经网络结构

Fig. 1　CBOW neural network structure

输入层为单词 X 周围 $n-1$ 个单词的词向量。例如，$n=5$，则词 X（记为 $w(t)$）前两个和后两个的单词为 $w(t-2)$，$w(t-1)$，$w(t+1)$，$w(t+2)$。相应地，4 个单词的词向量记为 $v(w(t-2))$，$v(w(t-1))$，$v(w(t+1))$，$v(w(t+2))$。

从输入层到映射层（$Pro(t)$）是将 $n-1$ 个词向量相加：

$$Pro(t)=v(w(t-2))+v(w(t-1))+v(w(t+1))+v(w(t+2))。\tag{1}$$

映射层到输出层需构造 Huffman 树（依照各词汇出现频率构造 Huffman 树），Huffman 树构造过程如下：

1）将 $w(t-2), w(t-1), w(t), w(t+1), w(t+2)$ 看作 n 棵树，每棵树仅有一个节点。

2）选取权值最小的两棵树进行合并，得到一棵新树，这两棵树作为新树的左右节点，新树的根节点权值等于左右子树的权重之和。

3）重复步骤 2）直到 n 棵树都已合并为止。

然后从根节点开始，映射层的值需要沿着 Huffman 树不断地进行 logistic 分类，并且不断地更正各词向量。

词向量模型训练完成后，以字典的形式保存，从而完成单词到向量的映射。

1.3 训练检测模型

本文采取卷积神经网络对文本进行分类，在数据集较大的情况下，也可以自动提取特征。

其输入层由文本向量组成，文本向量的长度即为输入参数的个数。通过若干个一维卷积层进行特征提取；最大池化层将提取的特征再次压缩，并提取主要特征且简化网络计算复杂度；dropout 连接所有特征并将计算结果通过 softmax 分类器进行输出。

一维卷积网络降低维度的原理如下：

1）假设输入数据维度（in）为 8，过滤器维度（filter）为 5，则卷积操作后数据维度（out）为 $8-5+1=4$，即：$\text{out}=\text{in}-\text{filter}+1$。

2）如果过滤器数量仍为 1，而输入数据的 channel 数量变为 16，则表示输入的数据有 8 个单词而每个单词的词向量维度为 16。此时，过滤器的维度则变为 5×16，输出维度仍为 4。

3）如果过滤器数量为 n，那么输出的数据维度就变为 $4n$，即：$\text{out}=n\times(\text{in}-\text{filter}+1)$。

卷积神经网络属于有监督学习，训练前需对样本进行标签化处理。在这个模型中，将正常样本标记为 0，SQL 注入样本标记为 1。训练过程中，优化器根据预测结果与实际结果的偏差（由损失函数计算所得）不断进行反向传递优化参数，最后得到检测模型。

2 SQL 注入检测实验

本文数据来自互联网的 SQL 注入实例。训练时使用了三种数据集：第一部分为训练集，主要用于训练检测模型；第二部分为验证集，验证当前训练模型的准确率；第三部分为测试集，测试已训练完成的模型对样本的识别率。训练集中正常样本 24 500 条，SQL 注入攻击样本 25 527 条，XSS 攻击样本 25 112 条；验证集中正常样本 10 000 条，SQL 注入攻击样本 10 000 条，XSS 攻击样本 10 000 条；测试集共 4 组，每组正常样本 2000 条，SQL 注入攻击样本 2000 条，XSS 攻击样本 2000 条。

正常样本数据格式如下：

code% 3Dzs _ 000001% 2Czs _ 399001% 2Czs _ 399006% 26cb% 3Dfortune _ hq _ cn% 26 _% 3D1498591852632；

SQL 注入样本数据格式如下：

-9500%22%20WHERE%206669%3D6669%20OR%20NOT%20%284237%3D6337%29；

XSS 注入样本数据格式如下：

site_id%3Dmedicare%22%3E%3Cscript%3Ealert%281337%29%3C/script%3E%2Casdf。

2.1 文本分词处理

URL 解码后，对一些无关变量进行替换，以一些特殊符号（如@ * # $ （）/空格等）为分割符，对语句进行分词处理。对数据进行规范化处理、分词处理，是非常关键的一个步骤，不仅能最大化地保留文本信息，还减少了一些噪声的影响。提取结果如下：

未分词处理样本 1））)%252bAND%252b8941%25253d8941%252bAND；

分词处理后样本 ['0',')',')',')','and','0=','0','and']。

2.2　提取文本向量

通过 Python 的 gensim 模块来使用 Word2Vec 工具，将分词处理后的语句作为输入数据。经过 Word2Vec 训练后，可得到一个以字典形式保存的词向量模型。通过此模型对每个单词进行向量化，再由词向量组成语句向量作为检测模型的输入。提取词向量结果如下：

样本单词　And；

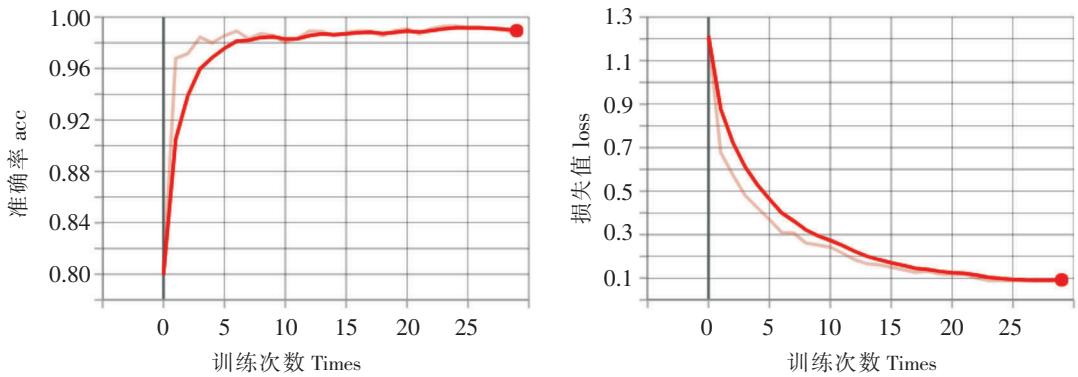
样本词向量　[-4.609 003 54　2.700 308 80　-0.033 447 10　0.946 626 66　0.517 221 75
1.236 753 94　-2.057 605 03　-2.369 857 31　4.133 568 29　2.375 314 47
-5.805 147 65　-1.499 013 90　-3.302 577 02　-2.151 923 66　0.870 841 15
-1.487 621 19]。

由于训练前的样本经过分类、标签处理，因此需对样本的训练顺序进行随机处理。通过随机处理可以有效避免实验结果的偶然性。本次实验对训练时用到的训练集和验证集的样本数据进行随机处理，使用 Python 的随机函数产生随机序列，并由该随机序列的顺序决定样本读取顺序。

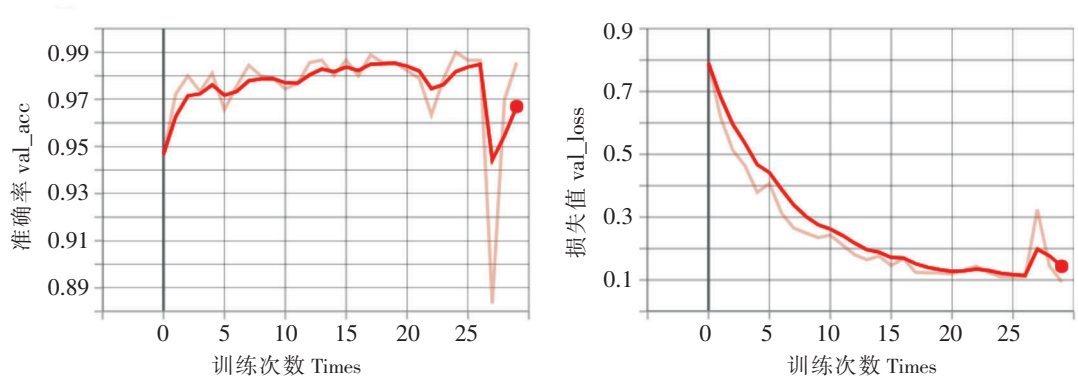
2.3　训练检测模型

本文采用的卷积神经网络结构参考文献［18］，具体由三个卷积层、三个池化层组成，最后连接全连接层。由于 CNN 只能接受固定长度的向量输入，在训练前需要将样本填充数据，使其输入长度固定。

训练集的准确率(acc)、损失值(loss)与验证集的准确率(val_acc)、损失值(val_loss)如图 2 所示。



a) 训练集 Train set



b) 验证集 Validation set

图 2　准确率与损失值

Fig.2　Acc and loss

2.3.1　CNN 中的参数选取

该卷积神经网络中使用了两种激活函数，卷积层与全连接层使用了 ReLu 函数作为激活函数，而输出层使用的是 Softmax 函数作为激活函数。

1）ReLU 函数是个分段线性函数，输入为负值时输出为 0，而正值输出不变，这种操作被称为单侧抑制。

$$\mathrm{ReLU}(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leqslant 0 \end{cases} \tag{2}$$

2）Softmax 函数用于多分类过程时，将多个神经元通过 Softmax 函数映射到（0,1）区间内，可以将其看成概率来理解，从而进行多分类。假设有一个数组 V，V_i 表示 V 中的第 i 个元素，那么这个元素的 Softmax 值 $S_i = e^{v_i} / \sum_{j=1}^{j} e^{v_j}$。

输入的参数经过 Softmax 函数，映射成为（0,1）的值，而这些值的累和为 1（满足概率的性质）。因此，在最后选取输出结点的时候，选取概率最大（也就是值对应最大）的结点作为预测目标。

2.3.2 代理模块

客户端在与 Web 服务器进行数据包交互前需进行 TCP 三次握手，经过 TCP 连接后传输 HTTP 报文。代理服务可作为中间者，由代理服务器完成与服务器的 TCP 连接、传输 HTTP 报文，而客户端实质上是与代理服务器进行报文交互，由代理服务器决定是否转发数据。

恶意用户可通过修改 HTTP 数据里的参数，构造攻击载荷，对数据库进行攻击。代理模块拦截客户端与服务端的数据包，通过注入检测模型检测当前数据包是否存在攻击，若不存在攻击则由代理服务器向 Web 服务器转发数据，否则丢弃数据包。

3 实验结果分析

利用 4 组测试数据集对前述 SQL 注入检测模型进行验证。每组测试集对三种样本的识别率和误报率如表 1、表 2 所示。

表 1 样本检测结果

Tab. 1 Results of Sample test

数据组别 Group	数据样本 Sample	正常数据识别率 Normal data recognition rate/%	XSS 攻击识别率 XSS attack reconition rate/%	SQL 攻击识别率 SQL attack recognition rate/%	误报率 False positive rate /%
第一组 First group	正常数据 Normal data	98.85	0.60	0.55	1.15
	XSS 攻击 XSS attack	1.45	98.50	0.05	1.50
	SQL 注入攻击 SQL injection attack	1.50	2.80	95.70	4.30
第二组 Second group	正常数据 Normal data	98.75	1.00	0.25	1.25
	XSS 攻击 XSS attack	1.70	97.90	0.40	2.10
	SQL 注入攻击 SQL injection attack	0.20	1.45	98.35	1.65
第三组 Third group	正常数据 Normal data	97.8	1.95	0.25	2.20
	XSS 攻击 XSS attack	1.60	98.20	0.20	1.80
	SQL 注入攻击 SQL injection attack	0.20	0.10	99.70	0.30
第四组 Fouth group	正常数据 Normal data	98.10	1.75	0.15	1.90
	XSS 攻击样本 XSS attack	5.05	90.80	4.15	9.20
	SQL 注入攻击样本 SQL injection attack	2.20	0.10	97.70	2.30

表 2　实验结果数据统计

Tab. 2　Data statistics of experient

数据样本 Samples	第一组 First group /个	第二组 Second group /个	第三组 Third group /个	第四组 Fouth group /个	平均识别率 Average recognition /%	平均误报率 Average positive /%
正常数据 Normal data	1977	1975	1956	1962	98.375	1.625
XSS 攻击 XSS attack	1970	1958	1964	1816	96.350	3.650
SQL 注入攻击 SQL injection attack	1914	1967	1994	1954	97.863	2.135

测试结果表明经过卷积神经网络（CNN）训练的模型具有极强的预测能力，对样本识别的准确率基本都在 97% 左右并且误报率极低。此方法不仅仅能够识别单一 SQL 注入或者 XSS 攻击，只要拥有足够并且典型的训练样本，不需设计更多的算法，就可训练出能识别多种网络攻击的检测模型，且模型的准确性和检测攻击的多样性完全取决于样本数据本身。当然，卷积神经网络对设备的性能也有所要求，随着数据量的增加，设备的计算量也会随之增加，服务器的响应速度也会有所影响。

4　CNN 与 BP 神经网络算法对比

本文对基于 CNN 算法与 BP 神经网络算法的 SQL 注入检测模型进行测试，两种算法选择同样的训练集。

变异 SQL 注入攻击样本由特殊字符进行混淆，且未曾在训练样本中使用。变异 SQL 注入攻击数据样本格式为：/? id = -1/ * ! UnIon/ * / * %0n%0y * / / * / * %0n%0y * /seLeCT * /1,'test',3。

实验结果为：1）BP 神经模型对 4 组包含 SQL 注入的攻击样本进行测试，平均识别率 97%，误报率 3%；2）CNN 模型对 1500 条变异 SQL 注入攻击样本测试，显示正常样本 1 条，SQL 注入 1499 条，准确率 99%，误报率 1%；3）BP 神经网络模型对 1500 条变异 SQL 注入攻击样本测试，显示正常样本 369 条，SQL 注入 1131 条，准确率 75%，误报率 25%。

通过实验分析，BP 神经网络模型与 CNN 模型对普通 SQL 注入攻击的识别能力相仿，其准确率均为 97% 左右，但是，对 SQL 注入变异攻击的识别能力，CNN 模型远远优于 BP 神经网络，CNN 准确率为 99%，而 BP 神经网络仅为 75%。面对 SQL 变异快的特点，基于 CNN 训练的 SQL 注入检测模型具有更强的适应性。

5　结束语

本文构建基于卷积神经网络算法的 SQL 注入检测模型，通过算法本身特性提取攻击样本的特征向量，从而降低漏洞检测的误报率；对多种类型攻击样本进行训练，实现多种攻击（SQL 注入与 XSS）的检测；利用卷积神经系统的非线性映射能力建立了多种攻击行为的映射关系。通过实验验证了此算法的通用性，其不仅可以应用在 SQL 注入检测，而且可以推广到多种入侵行为的检测。相比于 BP 神经网络模型，基于 CNN 的 SQL 注入检测模型识别能力更优，具有更强的适应性。当然，在数据量极为庞大的时候，检测模型的检测速度会有所下降甚至影响正常访问，完善模型的检测速率将是后续研究的重点。

[参 考 文 献]

[1] SATTER A, HOSSAIN B M M. Vulnerabilities assessment of emerging web-based services in developing countries [J]. International Journal of Information Engineering and Electronic Business, 2016, 8(5): 1-2.

[2] Open Web Application Security Project (OWASP): 2017 OWASP Top Ten[EB/OL][2017-10-25]. http://www.owasp.org.cn/owasp-project/OWASPTop102017v1.3.pdf.

[3] JUSTIN CLARKE. SQL 注入攻击与防御 [M]. 2 版. 北京：清华大学出版社，2013.
[4] ANTUNES J，NEVES N，CORREIA M，et al. Vulnerability discovery with attack injection [J]. IEEE Transactions on Software Engineering，2010，36(3)：357-370.
[5] AHMED M，MAHMOOD A N，HU J. A survey of network anomaly detection techniques [J]. Journal of Network & Computer Applications，2016，60：19-31.
[6] VALDES A，SKINNER K. Adaptive，model-based monitoring for cyber attack detection [C] //International Workshop on Recent Advances in Intrusion Detection. Berlin，Heidelberg：Springer，2000：80-93.
[7] 李红灵，邹建鑫. 基于 SVM 和文本特征向量提取的 SQL 注入检测研究 [J]. 信息网络安全，2017(12)：40-46.
[8] 杨连群，孟魁，王斌，等. 基于隐马尔可夫模型的新型 SQL 注入攻击检测方法 [J]. 信息网络安全，2017(9)：115-118.
[9] 张志超，王丹，赵文兵，等. 一种基于神经网络的 SQL 注入漏洞的检测模型 [J]. 计算机与现代化，2016(10)：67-71.
[10] 陈先昌. 基于卷积神经网络的深度学习算法与应用研究 [D]. 杭州：浙江工商大学，2014.
[11] 苑兆忠，姜华. Web 挖掘技术在信息检索中的应用研究 [J]. 聊城大学学报（自然科学版），2006，19(1)：74-77.
[12] 林晓佳. 基于改进 Adaboost M1 算法医学图像分类系统的研究 [J]. 聊城大学学报（自然科学版），2015，28(4)：29-32.
[13] KRIZHEVSKY A，SUTSKEVER I，HINTON G E. Imagenet classification with deep convloutional neural networks [C] //Advances in Neural Information Processing Systems. Massachusetts：The MIT Press，2012：1097-1105.
[14] SIMARD P，STEINKRAUS D，PLATT J C. Best practices for convolational neural networks applied to visual document analysis [C] //7th International Conference on Document Analysis and Recognition（ICDAR 2003）. Washington D C：IEEE Computer Society，2003(3)：958-962.
[15] CLARK C，STORKEY A，CLARK C，et al. Teaching deep convolutional neural networks to play go [J]. Eprint Arxiv，2014：1766-1774.
[16] 蒋磊. 基于机器学习的 SQL 注入检测技术研究 [D]. 南京：南京邮电大学，2017.
[17] MIKOLOV T，CHEN K，CORRADO G，et al. Efficient estimation of word representations in vector space [C/OL] //Proceedings of Workshop at ICLR，2013. [2018 - 06 - 03] http://www.researchgate.net/publication/234131319_Eiffi-cent_Estimation_of_word_Representations_in_Vector_Space.
[18] HU B，LU Z，LI H，et al. Convolutional neural network architectures for matching natural language sentences [C] //Advances in Neural Information Processing Systems. Massachusetts：The MIT Press，2014：2042-2050.

（责任编辑 朱雪莲 英文审校 黄振坤）