OMECNN:一种基于有序马尔可夫枚举器和判别神经网络的口令生成模型

杨龙龙，杨　频，刘　亮，张　磊
(四川大学网络空间安全学院，成都 610065)

摘　要：基于口令的身份鉴别是目前最流行的鉴别方式之一，利用口令生成技术进行大规模口令集的生成，进而检测现有用户口令保护机制的缺陷、评估口令猜测算法效率等，是研究口令安全性的重要手段. 本文提出一种基于有序马尔可夫枚举器和判别神经网络的口令生成模型 OMECNN，使用有序马尔可夫口令枚举器按照口令组合概率的高低生成组合口令，同时基于判别神经网络进行打分筛选口令，选出得分高于阈值的口令组成最终口令集. 采用本文提出方法生成的口令集具有按照口令组合概率高低排序的特点，以及符合真实训练口令集的口令分布的特点. 实验结果表明，在生成 10^7 条口令时，OMECNN 模型生成的口令集在 Rockyou 测试集上的匹配条目比 OMEN 模型高出 16.60%，比 PassGAN 模型高出 220.02%.

关键词：口令生成；马尔科夫链；判别神经网络；生成对抗网络
中图分类号：TP391.1　　**文献标识码**：A　　DOI：10.19907/j.0490-6756.2021.042004

OMECNN: a password-generation model based on ordered markov enumerator and critic neural network

YANG Long-Long, *YANG Pin*, *LIU Liang*, *ZHANG Lei*
(College of Cybersecurity, Sichuan University, Chengdu 610065, China)

Abstract: Password identification is one of the most popular way of identification. Generating a large-scale password set based on password-generation techniques is a principal method to research password security, which can be applied to evaluate the efficiency of password-generation algorithm and detect the defects of existing user-password protective mechanisms. In this paper, we propose a password-generation model based on an ordered Markov enumerator and critical neural network (OMECNN). The OMECNN model combines both Markov chain and neural network techniques. OMECNN utilizes the ordered Markov passwords enumerator to generate the passwords according to the probability of combinations, and then uses the critic neural network to score those passwords, and selects the passwords whose score is higher than the threshold to form the final password set. The generated password set has the characteristics of sorting according to the combination probability of passwords and the distribution of passwords in accordance with the real training password set. The experimental results show that when 10^7 passwords are generated, the hits of OMECNN model on Rockyou test set is 16.60% higher than that of OMEN model and 220.02% higher than that of Pass GAN model.

Keywords: Password generation; Markov; Critic neural network; Generative adversarial network

收稿日期：2020-11-03
基金项目：四川省科技计划(2020YFG0076)
作者简介：杨龙龙(1996—)，男，湖南衡阳人，硕士研究生，主要研究领域为软件应用安全.
通讯作者：张磊. E-mail：zhanglei2018@scu.edu.cn

1 引　言

当前,基于口令的身份鉴别是目前最流行的鉴别方式之一,也是一个重要研究领域[1],然而口令的使用不总是安全的,人们习惯将口令设置成让自己容易记住的短的或者有规律的字符序列[2-3],像“123456”这样的弱口令. 对此,许多研究学者提出了口令设置策略[4-8]来帮助用户设置易于记忆且不易被猜解的口令. 然而,由于禀赋效应[9]的存在,即使有很高安全意识的人,也还是会使用弱口令. 研究口令策略和口令安全性如检测现有用户口令保护机制的缺陷等都需要有大规模口令明文样本. 获取真实的大规模口令明文是困难的,利用口令生成技术来生成大规模口令集是目前广泛使用的一个手段.

在口令生成技术研究方面,前人已经做了大量的研究[10-15],其中比较具有代表性的是 OMEN 模型[10]和 PassGAN 模型[11],这两个模型分别是基于传统马尔可夫链的口令生成方法和基于生成对抗神经网络的口令生成方法.

OMEN 模型的出发点是高概率出现的口令排在口令集前面对口令破解有加速作用. OMEN 实验数据表明,OMEN 模型生成的口令集在测试集上的命中率要高于 JtR-Markov 以及其改进版 JtR-Inc 模型[16]. 然而,OMEN 模型在设计上没有考虑到口令集的分布是否与测试集口令分布一致的问题.

PassGAN 模型使用了生成对抗神经网络(Generative Adversarial Network,GAN)[17],生成的口令具有符合原口令集分布的特点. 然而,PassGAN 生成的口令集中具有大量重复的口令,其次,PassGAN 生成的口令不会按照一定次序排序,并且存在口令重复率高的问题.

本文针对上述的 OMEN 模型与 PassGAN 模型在生成口令过程中存在的问题,提出 OMECNN 模型,通过引入判别神经网络打分筛选口令的方法解决上述问题. OMECNN 模型使用有序马尔可夫口令枚举器按照口令组合概率的高低生成组合口令,然后利用判别神经网络进行打分筛选口令,选出得分高于阈值的口令组成最终口令集. 本文主要的工作和贡献如下.

1) 本文提出一种基于有序马尔可夫枚举器和判别神经网络的口令生成模型. 首先由有序马尔可夫枚举器进行口令生成,随即使用判别神经网络进行打分筛选,得分高于阈值的口令组成最终口令集. 通过该模型生成的口令具有按照口令组合概率高低排序的特点和符合真实训练口令集的口令分布的特点.

2) 与 PassGAN 相比,OMECNN 解决了重复率的问题且生成的口令是按照口令组合概率高低排序的. 其次,OMECNN 模型是首个利用生成对抗网络的判别器和有序马尔可夫技术相结合进行口令生成的技术. 本文发现随着 GAN 的训练轮次的增加,GAN 的生成器的性能没有增加反而更差,而口令集重复率却增加了. 相反,GAN 的判别器的性能随着 GAN 的训练轮次的增加而得到增加.

3) 实验结果表明,OMECNN 模型的命中率较 OMEN 模型有所提升,较 PassGAN 模型有大幅的提升. 在 Rockyou 数据集上,在生成 10^7 条口令的条件下,OMECNN 模型比 OMEN 模型多命中 16.60%的口令,比 PassGAN 模型多命中 220.02%的口令.

2 相关工作及背景

2.1 生成对抗网络

2014 年,研究学者提出生成对抗网络[18]. 生成器网络尽可能生成逼真样本,判别器网络则尽可能去判别样本是来自真实样本,还是来自生成器生成的伪样本,示意图如图 1 所示.

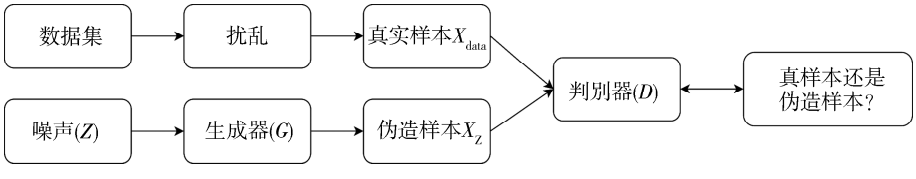

图 1　生成对抗网络框架图

Fig. 1　GAN framework

GAN 的目标函数如下：

$$\min_G \max_D V(G,D) = \min_G \max_D E_{x\sim p_{\text{data}}}[\log D(x)] + E_{z\sim p_z}[\log(1-D(G(z)))] \quad (1)$$

对于判别器 D，其目标函数为 $\max V(G,D)$，当生成器 G 固定时，对 $V(G,D)$ 求导，得到最优判别器 $D^*(x)$，如式(2).

$$D^*(x) = \frac{p_z(x)}{p_z(x)+p_{\text{data}}(x)} \quad (2)$$

将最优判别器代入上述目标函数式(1)中，可以求出在最优判别器下，生成器的目标函数，其等价于优化 $p_{\text{data}}(x)$ 与 $p_z(x)$ 的 JS 散度. 当 $p_{\text{data}}(x)$ 与 $p_z(x)$ 相等时，对于判别器而言，任一样本被预测为真实样本的概率均为 0.5，达到了难以区分真实样本与伪造样本的地步. 实际训练时，生成器和判别器采取交替训练的方式，先训练 k 次 D，然后训练 1 次 G，不断往复. 在 PassGAN 模型中 k 取 10.

原始 GAN 存在训练梯度不稳定问题，使 GAN 的训练变得困难或失败. 有进一步的工作[17-18]改善了训练时梯度的稳定性的问题. 本文介绍的 PassGAN 和判别神经网络均是用由 IW-GAN 模型[18]实现对抗神经网络模型的.

2.2　口令生成技术

早期的口令生成技术的研究是基于 Markov 模型，后经过改进得到 JtR-Markov 模型[16]. 在该模型中，用户口令被解析成一串字符序列，字符之间存在相互的联系，且字符之间的关系符合统计规律. 通过统计字符串子序列之间的关系，得到字符序列间的统计规律，随后基于这些规律进行口令字典的生成.

Weir 等提出上下文概率无关语法技术，后经 Durmuth 等拓展得到 OMEN 模型[10]. OMEN 按照 N-Gram 的发生概率高低来生成口令，概率越高的口令输出的可能性就越靠前. 具体做法是引入用于存储具有不同概率的 N-Gram 序列的桶的概念. 随后将所有的 N-Gram 序列按照统计概率的高低分类到不同的桶中，OMEN 模型按照不同桶的概率由高到低进行取 N-Gram 序列组成口令，直到所有桶被遍历完或者生成的口令的条目达到目标条目. 在可选字符集 $\sum$ 中，OMEN 模型生成的口令序列可以表述为

$$P(x_1x_2\ldots x_n) = P(x_1)P(x_2|x_1)\ldots P(x_n|x_1x_2\ldots x_{n-1}),\ x_i \in \sum \quad (3)$$

OMEN 模型生成的口令集 D 可以表述为

$$D_{V,\vartheta,\ell,\eta,\vartheta',\ell',\eta'} = \{\beta : \alpha\beta \in D_{V,\vartheta,\ell,\eta}\} \quad (4)$$

以及

$$D_{V,\vartheta,\ell,\eta} = \{\alpha : |\alpha| = \ell \wedge \prod_{X\in a} V(X) > \vartheta \wedge V(X) > \eta\} \quad (5)$$

在式(4)和式(5)中，l 为字符序列的长度；$V(X)$ 是字符的条件概率函数；ϑ 为 OMEN 模型中的阈值，低于该阈值的口令组合将被丢弃；η 为某个 N-Gram 序列桶的最小概率. OMEN 的实验表明，OMEN 性能超过 JtR-Markov 以及 JtR-Inc 模型.

随着机器学习的快速发展，运用机器学习的方法进行安全领域课题的研究也逐渐流行起来，出现了一些新的研究方向[13,19-20]. 在口令猜解中，Hitaj 等基于生成对抗神经网络提出 PassGAN[11] 口令猜解模型，使用生成对抗神经网络来进行学习口令集的分布，学习到分布后然后再使用生成器网络来生成口令. 本文实验部分将讨论 PassGAN 的重复率问题.

本文提出的模型结合马尔科夫枚举器和判别神经网络方法，使用有序马尔可夫口令枚举器按照口令组合概率的高低生成组合口令，同时使用判别神经网络进行打分筛选口令，选出得分高于阈值的口令组成最终口令集. 生成的口令集具有按照口令组合概率高低排序的特点和符合真实训练口令集的口令分布的特点，并且不存在重复的口令的情况.

3　OMECNN 模型实现

3.1　OMECNN 模型总括

OMECNN 模型包括两个子模块，一个是口令生成模块，另一个是口令筛选模块，整体模型训练和口令生成框架图如图 2 所示.

如图 2 所示，使用数据集训练 GAN 和有序马尔可夫口令枚举器，训练完成后得到有序马尔可夫口令枚举器与 GAN 神经网络两个模块. OMECNN 模型中，取 GAN 的判别器作为判别神经网络子模块，GAN 生成器网络被丢弃. 随用使用有序马尔可夫口令枚举器进行口令生成，将判别神经网络应用于有序马尔可夫口令枚举器生成的口令，输出相应的分值，OMECNN 模型选择那些分值高于或者等于阈值常数 λ 的口令.

3.2　口令选择模块

OMECNN 模型的口令选择模块需要满足以下 3 点特性：(1)所选的判别神经网络模块有能力学习给定口令集的分布特征；(2)所选的判别神经网络模块以口令字符序列作为输入；(3)所选的判

别神经网络模块输出一个分值. 要求:当分值大于或者等于一个阈值常数 λ,能在一个高概率情况下得出该口令符合原分布的结论,反之,当分值低于该阈值常数 λ 时,得出该口令不符合原口令集分布的结论.

图 3~图 5 给出了本文 OMECNN 模型的生成对抗网络结构图.

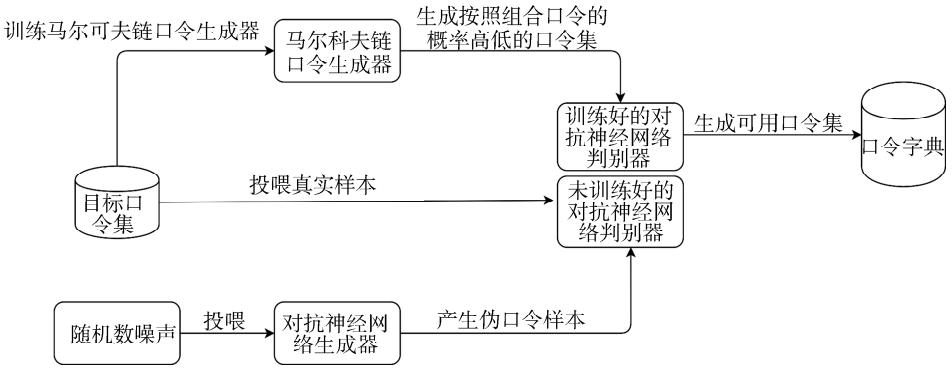

图 2　OMECNN 训练和口令生成框架图

Fig. 2　OMECNN and password generation framework

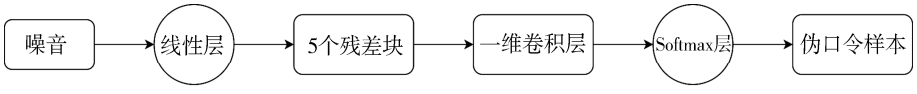

图 3　生成器网络结构图

Fig. 3　Generator structure

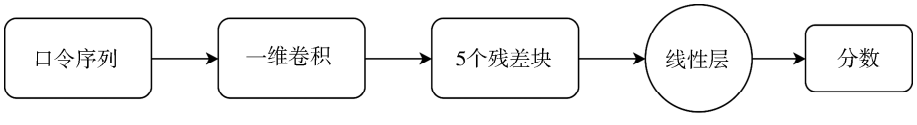

图 4　判别器网络结构图

Fig. 4　Discriminator structure

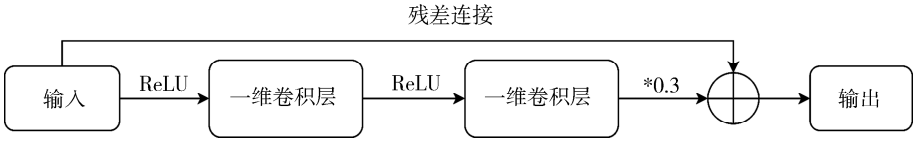

图 5　残差块结构图

Fig. 5　Residual block structure

训练完成后,OMECNN 模型将 GAN 中的判别器网络作为判别神经网络口令选择模块,并将其应用到由口令生成模块所生成的口令集上,对口令进行打分筛选,这一过程如下式所示.

$$score=C(X) \tag{6}$$

3.3　口令生成模块

OMECNN 模型的口令生成模块需要具备以下 3 个特征:(1) 所选的模型有能力学习给定口令集的口令相关特征;(2) 所选的模型能生成不重复的口令;(3) 所选模型输出的口令有一个特征,即先出现的口令是真实口令的概率比后出现的高.

有序马尔可夫枚举器符合上述 3 个特征. 用数学语言描述由有序马尔可夫枚举器生成的口令集合 D 可以被描述为如下式所示.

$$\begin{aligned} D_{V,\vartheta,\eta,\ell}=\{&x_1x_2\cdots x_\ell : V(x_1)\prod_{i=1}^{\ell-1}V(x_{i+1}\mid x_ix_{i-1}\ldots x_1)>\vartheta \wedge \\ &\forall n_j\in\eta, V(x_1)=n_1, V(x_{i+1}\mid x_ix_{i-1}\ldots x_1)>n_{i+1}\} \end{aligned} \tag{7}$$

式(7)中,$V(x_1)$是口令起始序列的概率函数;$V(x_{i+1}\mid x_ix_{i-1}\ldots x_1)$是第 $i+1$ 个字符的条件概率函数,这两个函数均由 N-Gram 马尔可夫模型学得.

结合式(6)和式(7)可得 OMECNN 模型生成的口令集 D 可以表示为

$$D^{*}_{OMECNN}=\{X:X\subseteq D_{V,\vartheta,\eta,\ell}\wedge\forall x_i\in X\Rightarrow C(x_i)\geqslant\lambda\} \tag{8}$$

综合两个模块，OMECNN 模型可由算法 1 表述.

算法 1　OMECNN 模型生成口令算法

Count=0

WHILE Count<targetCount：

FOR each vector $(n)_{2\leqslant i\leqslant\eta}$ with $\sum{}_i n_i=\eta$

FOR each $x_1x_2x_3x_4\in\sum{}^4$ with $V(x_1x_2x_3x_4)=n_2$

WHILE length of $X<\ell$：

FOR each $x_{j+4}\in\sum$ with $V(x_{j+4}|x_jx_{j+1}x_{j+2}x_{j+3})>n_{j+4}$

END WHILE

IF $C(X)\geqslant\lambda$：

Output password X

Count=Count+1

END IF

END FOR

END FOR

END WHILE

4　OMECNN 实验及分析

4.1　口令数据集描述

实验中使用到的是 Rockyou 数据集[21]，将数据集中的重复口令进行删除操作；同时，口令中包含非 ASCII 的字符也被删除，由于 PassGAN 中的口令序列最长为 10 个字符，故将口令长度大于 10 的口令也进行删除操作. 对口令集进行扰乱操作后，口令集被按照 4∶1 的比例分成训练集和测试集. 最终训练集包含 7 909 309 条口令，测试集包含 1 977 328 条口令. 该数据集包含的字符序列如下.

*abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ*0123456789!. * @−_$ # <?

4.2　实验环境

实验软件与硬件环境如表 1 所示.

表 1　实验软硬件环境

Tab. 1　Experience environment

操作系统	Microsoft Windows 10. 0. 18363. 1082
CPU	Intel® Core™ i7-9700 CPU @ 3. 00 GHz 3. 00 GHz
GPU	2 张 NVIDIA GeForce RTX 2070 SUPER
Tensorflow 版本	2. 3. 0
CUDA 版本	release 10. 1，V10. 1. 243
储存	NVMe GALAX TA1H0240N 240 G

4.3　实验评估方法

本文共对 3 个模型进行评估，其中评估的指标为口令的命中率(precision)，计算统计方法如下.

$$precision=\frac{Hit}{testing\text{-}count} \tag{9}$$

Hit 为命中测试集中的口令的数目；$testing\text{-}count$ 为测试集口令数量. 对于 PassGAN 模型而言，其重复率(repetition)也被评估，计算统计方法如下式.

$$repetition=\frac{Hit-Hit_{\text{unique}}}{\text{Maxtry}} \tag{10}$$

Hit_{unique} 为命中测试集中的已删除重复的口令的条目. OMECNN 模型将被按照算法 1 来进行口令生成，按照算法 1，有序马尔可夫枚举器生成的口令，得分低于阈值常数的口令将被丢弃，故 OMECNN 模型的口令的命中率(precision)的计算方法如下

$$precision=\frac{Hit_{C(X)\geqslant\lambda}}{testing\text{-}count_{C(X)\geqslant\lambda}} \tag{11}$$

4.4　对 OMEN 模型的评估

将 Rockyou 训练集口令给 OMEN 模型投喂学习，待其统计完 N-Gram 序列概率后，使用 OMEN 模型生成口令，其中，N-Gram 序列中的 N 取值分别为 2，3，4，5. 将训练完成的 OMEN 模型进行口令生成，将生成的口令集进行逐一取口令在测试集中搜索统计，得到的命中率如图 6 所示.

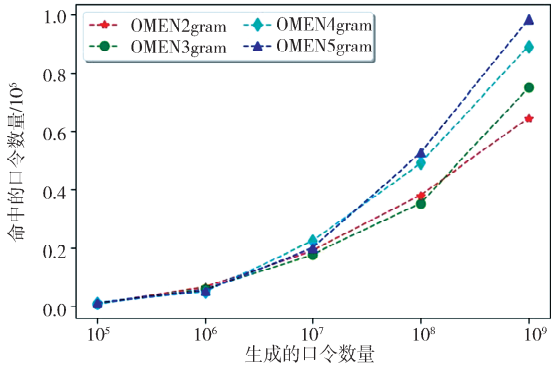

图 6　OMEN 模型命中情况

Fig. 6　The hit rate of OMEN model

图 6 中,OMEN2gram 表示 OMEN 模型中,N-Gram 取为 2,其他图例亦类似. 实验结果显示,当 OMEN 模型中的 N-Gram 模型为 5-Gram 时,OMEN 性能最好,其命中条目数如表 2 所示.

表 2　OMEN_5gram 模型命中条目情况表

Tab. 2　Hits of OMEN_5gram

生成的口令条目	命中的条目(命中率/%)
10^5	10 045(0.51)
10^6	53 257(2.69)
10^7	199 507(10.09)
10^8	526 296(26.62)
10^9	986 322(49.88)

4.5　对 PassGAN 模型的评估

将 Rockyou 训练集对 PassGAN 模型进行训练,训练模型的参数与文献中的一致,实验评估中,将更改最大字符序列长度(max sequence length)与训练的轮数(iteration)进行数据对比. 得到的命中情况如图 7 所示.

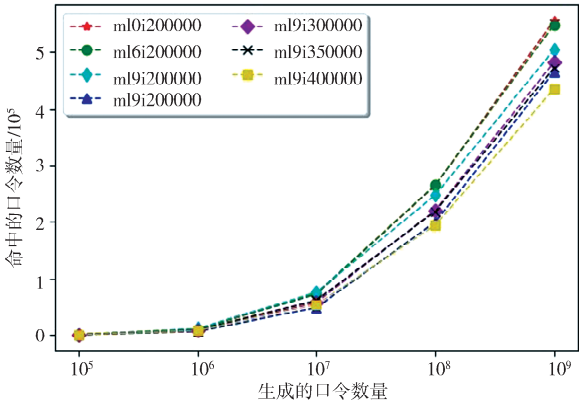

图 7　PassGAN 模型命中情况

Fig. 7　The hit rate of PassGAN model

图 7 中,m10i200000 表示在 PassGAN 模型中,最大序列长度取 10,最大训练轮数取 2×10^5,其他图例亦类似. 结果得出,PassGAN 模型中,当最大序列长度取 10,最大训练轮数取 2×10^5 时,PassGAN 模型命中性能最好,命中条目情况如表 3 所示.

上述结果表明,PassGAN 的命中率比 OMEN 模型差,原因之一是因为 PassGAN 模型会产生重复的口令,这是去掉重复口令的结果,按照式(10)对 PassGAN 模型重复率进行统计,得出结果如图 8 所示.

表 3　PassGAN 模型命中条目情况表

Tab. 3　The hits of PassGAN

生成的口令条目	命中的条目(命中率/%)
10^5	1 146(0.06)
10^6	10 539(0.53)
10^7	72 691(3.68)
10^8	264 978(13.40)
10^9	555 724(28.10)

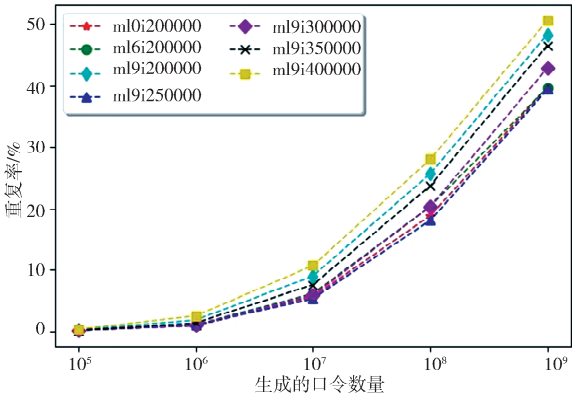

图 8　PassGAN 模型重复率

Fig. 8　Repetition of PassGAN

结合图 7 和图 8 可以分析出以下结论,当训练轮数参数为 4×10^5,最大字符序列长度为 19 时,在生成 10^9 条口令时,口令重复率达到了 50.60%. 其次,随着训练轮数和最大字符序列长度的增加,PassGAN 的生成器的性能并没有增加反而更差,而重复率会随着训练轮数的增加而增加.

4.6　对 OMECNN 模型的评估

将 Rockyou 训练集对 OMECNN 模型的有序马尔可夫枚举器和判别神经网络进行训练,其中,有序马尔可夫枚举器中的 N-Gram 设置为 5-Gram,判别神经网络中的训练轮数分别设置为 2×10^5, 3×10^5, 4×10^5,最大字符序列长度参数值设为 10,将 score 阈值分别设置为 −1.2,−1.3 和 −1.4. 按照算法 1 生成口令集,随后按照式(11)与测试集进行比对统计,结果如图 9 所示.

图 9 中,OMECNNn5i200000s13 表示为 OMECNN 模型中,有序马尔可夫枚举器的 N-Gram 为 5-Gram,训练轮数为 2×10^5,score 阈值常数取 −1.3,如无特殊情况,后文不再复述. 结果得出,当训练轮数为 4×10^5,score 阈值常数设置为 −1.3 时,在生成相同条目的口令情况下 OMECNN 命中条目最高,此时,命中条目具体情况如表 4 所示.

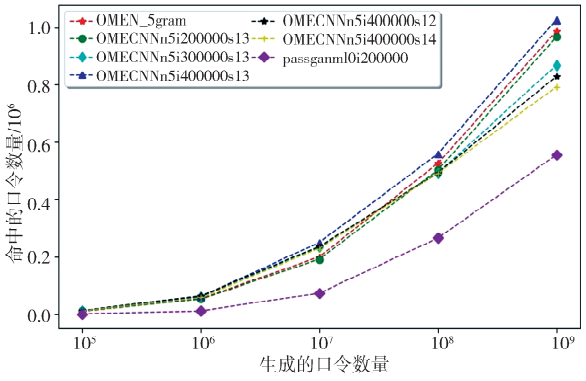

图 9　三个模型不同参数下命中情况

Fig. 9　The hits of 3 models under different parameters under OMECNN model

表 4　OMECNN 模型命中条目情况表

Tab. 4　The hits of OMECNN

生成的口令条目	命中的条目(命中率/%)
10^5	10 362(0.52)
10^6	60 101(3.04)
10^7	232 628(11.76)
10^8	542 640(27.44)
10^9	1 014 720(51.32)

比较表 2～表 4 可以得出，在生成 10^7 条口令时，OMECNN 模型命中的条目比 OMEN 模型命中的条目高出 16.60%，比 PassGAN 模型高出 220.02%.

4.7　实验小结

通过上述实验可以发现 OMEN 模型生成的口令对比测试集有较好的命中率，在生成 10^9 条口令时，OMEN 模型能命中测试集的 49.88% 的口令. 而 PassGAN 则只能命中 28.10%，造成 PassGAN 命中率低的原因之一是因为 PassGAN 模型生成的口令集存在大量重复的口令，本文还对 PassGAN 模型生成口令的重复率进行了实验，实验发现，当训练轮数参数为 4×10^5，最大口令字符序列长度为 19 时，在生成 10^9 条口令的情况下，口令重复率达到 50.60%. 其次，随着训练轮数和最大口令字符序列长度的增加，PassGAN 的生成器的性能并没有增加反而更差，而重复率会随着训练轮数的增加而增加. 对比 OMECNN 模型的判别神经网络模块，还可以发现，GAN 的判别器的性能会随着训练轮数的增加而增加. 在命中率上，OMECNN 在生成 10^9 条口令的情况下，命中率能达到 51.32%，高出 OMEN 模型 1.44%，高出 PassGAN 模型 23.22%. 在生成 10^7 条口令的情况下，OMECNN 模型生成的口令命中条目高出 OMEN 模型生成的口令命中条目的 16.60%，高出 PassGAN 模型生成的口令命中条目的 220.02%. OMECNN 模型比 OMEN 模型和 PassGAN 模型均要更优，是一个性能更好的口令生成模型.

5　结　论

利用口令生成技术进行大规模口令集的生成，进而检测现有用户口令保护机制的缺陷、评估口令猜测算法效率等，是研究口令安全性的重要手段. 针对口令生成技术研究，本文提出一种基于有序马尔可夫枚举器和判别神经网络的口令生成模型 OMECNN，OMECNN 模型在 Rockyou 公开口令数据集上，在生成 10^9 条口令的情况下，命中率能达到 51.32%，高出 OMEN 模型 1.44%，高出 PassGAN 模型 23.22%. 在生成 10^7 条口令的情况下，OMECNN 模型生成的口令命中条目高出 OMEN 模型生成的口令命中条目的 16.60%，高出 PassGAN 模型生成的口令命中条目的 220.02%. OMECNN 模型比 OMEN 模型和 PassGAN 模型均要更优，是一个性能更好的口令生成模型.

今后的工作可以关注判别神经网络模块输出的分数的分布的研究，研究判别神经网络模块的输出可以发现判别神经网络内在的工作模式，以及发现什么性质结构的口令的得分值能超过阈值，这对改善口令构造策略的研究有积极作用.

参考文献：

[1]　赵宗渠，黄鹂娟，叶青，等. 基于格的两方口令认证密钥交换协议[J]. 重庆邮电大学学报：自然科学版，2019，31：833.

[2]　Wang D，Wang P，He D，*et al*. Birthday，name and Bifacial-Security：understanding passwords of Chinese web users [C]//Proceedings of the 28th USENIX Conference on Security Symposium. USA：USENIX Association，2019.

[3]　Zeng J，Duan J，Wu C. Empirical study on lexical sentiment in passwords from Chinese websites [J]. Comput Secur，2019，80：200.

[4]　Guo Y，Zhang Z，Guo Y. Optiwords：A new password policy for creating memorable and strong passwords [J]. Comput Secur，2019，85：423.

[5]　Siponen M，Puhakainen P，Vance A. Can individuals' neutralization techniques be overcome? A field experiment on password policy [J]. Comput Secur，

2020, 88: 101617.

[6] Ye B, Guo Y, Zhang L, *et al*. An empirical study of mnemonic password creation tips [J]. Comput Secur, 2019, 85: 41.

[7] Yildirim M, Mackie I. Encouraging users to improve password security and memorability [J]. Int J Inf Secur, 2019, 18: 741.

[8] Doucek P, Pavlíček L, Sedláček J, *et al*. Adaptation of password strength estimators to a non-English environment—the Czech experience [J]. Comput Secur, 2020, 95: 101757.

[9] Renaud K, Otondo R, Warkentin M. "This is the way 'I' create my passwords"... does the endowment effect deter people from changing the way they create their passwords? [J]. Comput Secur, 2019, 82: 241.

[10] Duermuth M, Angelstorf F, Castelluccia C, *et al*. OMEN: Faster password guessing using an ordered markov enumerator [C]//International Symposium on Engineering Secure Software and Systems. Cham: Springer International Publishing, 2015.

[11] Hitaj B, Gastip P, Ateniese G, *et al*. PassGAN: A deep learning approach for password guessing [C]// Deng R H, GAUTHIER-UMAÑA V, Ochoa M, *et al*. Applied cryptography and network security. Cham: Springer International Publishing, 2019.

[12] 徐凌智. 基于深度学习的密码字典生成技术研究[D]. 上海: 上海交通大学, 2018.

[13] Linghu Y, Li X, Zhang Z. Deep learning vs. traditional probabilistic models: case study on short inputs for password guessing [C]//Wen S, Zomaya A, Yang L T. Algorithms and architectures for parallel processing. Cham: Springer International Publishing, 2020.

[14] 刘功申, 王靖康, 罗宇韬, 等. 基于海量真实口令的挖掘分析及口令生成[J]. 通信学报, 2018, 39: 213.

[15] Fang Y, Liu K, Jing F, *et al*. Password guessing based on semantic analysis and neural networks [C]//Zhang H, Zhao B, Yan F. Trusted Computing and Information Security. Singapore: Springer Singapore, 2019.

[16] Openwall. John the Ripper [M/OL]. [2020-02-16]. https://www.openwall.com/john/.

[17] Cao Y, Ding G W, Lui K Y C, *et al*. Improving GAN training via Binarized Representation Entropy (BRE) regularization [C]//Proceedings of the 6th International Conference on Learning Representations. Vancouver, BC, Canada: [s. l.], 2018.

[18] Wei X, Gong B, Liu Z, *et al*. Improving the improved training of wasserstein GANs: a consistency term and its dual effect [C]// Proceedings of the 6th International Conference on Learning Representations. Vancouver, BC, Canada: [s. n.], 2018.

[19] 高杨晨, 方勇, 刘亮, 等. 基于卷积神经网络的 Android 恶意软件检测技术研究[J]. 四川大学学报: 自然科学版, 2020, 57: 673.

[20] 刘凯, 方勇, 张磊, 左政, 等. 基于图卷积网络的恶意代码聚类[J]. 四川大学学报: 自然科学版, 2019, 56: 654.

[21] Rockyou. Rockyou [EB/OL]. [2020-02-16]. https://www.kaggle.com/wjburns/common-password-list-Rockyoutxt.

引用本文格式:

中　文: 杨龙龙, 杨频, 刘亮, 等. OMECNN:一种基于有序马尔可夫枚举器和判别神经网络的口令生成模型[J]. 四川大学学报: 自然科学版, 2021, 58: 042004.

英　文: Yang L L, Yang P, Liu L, *et al*. OMECNN: a password-generation model based on ordered markov enumerator and critic neural network [J]. J Sichuan Univ: Nat Sci Ed, 2021, 58: 042004.