计算机科学
COMPUTER SCIENCE

一种基于强化学习的口令猜解模型

李小玲, 吴昊天, 周涛, 鲁辉

引用本文

李小玲, 吴昊天, 周涛, 鲁辉.一种基于强化学习的口令猜解模型[J]. 计算机科学, 2023, 50(1): 334-341.

LI Xiaoling, WU Haotian, ZHOU Tao, LU Hui. Password Guessing Model Based on Reinforcement Learning [J]. Computer Science, 2023, 50(1): 334-341.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.211100001

一种基于强化学习的口令猜解模型

李小玲[1] 吴昊天[1] 周 涛[1] 鲁 辉[2]
1 华南理工大学计算机科学与工程学院 广州 510006
2 广州大学网络空间先进技术研究院 广州 510006
(202021044839@mail.scut.edu.cn)

摘 要 口令猜解是口令安全研究的重要方向之一。基于生成式对抗网络(Generative Adversarial Network,GAN)的口令猜解是近几年提出的一种新方法,其通过判别器对生成口令的评判结果来指导生成器的更新,进而生成口令猜测集。然而由于判别器对生成器的指导不足,现有的基于GAN的口令猜解模型的猜解效率较低。针对这个问题,提出了一种基于强化学习Actor-Critic算法改进的GAN口令猜解模型AC-Pass。AC-Pass模型通过Critic网络和判别器输出的奖赏共同指导Actor网络每一时间步生成策略的更新,实现了对口令序列生成过程的强化指导。将AC-Pass模型应用到RockYou,LinkedIn和CSDN口令集进行实验,并与PCFG模型、已有基于GAN的口令猜解模型PassGAN和seqGAN进行比较。实验结果表明,无论是同源测试集还是异源测试集,AC-Pass模型在9×10^8猜测集上的口令破解率均高于PassGAN和seqGAN;且当测试集与训练集之间的口令空间分布差异较大时,AC-Pass表现出了优于PCFG的口令猜解性能;另外,AC-Pass模型有较大的口令输出空间,其破解率随着口令猜测集的增大而提高。

关键词:口令猜解;深度学习;强化学习;Actor-Critic算法;生成式对抗网络
中图法分类号 TP309

Password Guessing Model Based on Reinforcement Learning

LI Xiaoling[1],WU Haotian[1],ZHOU Tao[1] and LU Hui[2]
1 School of Computer Science and Engineering,South China University of Technology,Guangzhou 510006,China
2 Cyberspace Institute of Advanced Technology,Guangzhou University,Guangzhou 510006,China

Abstract Password guessing is an important research direction in password security. Password guessing based on generative adversarial network(GAN) is a new method proposed in recent years,which guides the update of the generator according to evaluation results on passwords generated by the discriminator. Consequently,password guessing sets can be generated with trained GANs. However,the existing GAN-based password guessing models have low efficiency due to inadequate guidance of the discriminator to the generator. To solve this problem,an improved GAN password guessing model AC-Pass based on reinforcement learning Actor-Critic algorithm is proposed. The AC-Pass model guides the update of the generation strategy of the Actor network at each time step through the output rewards of the discriminator and the Critic network,and realizes the reinforce guidance of password sequence generation process. The proposed AC-Pass model is implemented on RockYou,LinkedIn and CSDN data sets and compared with PCFG model and the existing GANs-based password guessing models such as PassGAN and seqGAN. Results on homologous testing sets and heterologous testing sets indicate that password cracking rate of AC-Pass model on the guessing set is higher than that of PassGAN and seqGAN. Moreover,AC-Pass shows better guessing performance than PCFG when the password spatial distribution between the testing set and the training set is significant. In addition,the AC-Pass model has a large password output space. As the size of password guessing set increases,the cracking rate continues to rise.

Keywords Password guessing,Deep learning,Reinforcement learning,Actor-Critic algorithm,Generative adversarial network

1 引言

在目前乃至未来可预见的时间范围内,口令认证都是应用系统最主要的身份认证方式之一[1-2]。然而,我们从多个泄露口令集中发现,用户为了方便记忆,往往偏向于设置结构简单、与个人信息相关、基于语义等的弱口令[3]。攻击者喜欢

到稿日期:2021-11-01 返修日期:2022-03-27
基金项目:广东省重点领域研发计划(2019B010137004);广东省自然科学基金面上项目(2021A1515011798)
This work was supported by the R&D Project in Key Areas of Guangdong Province,China(2019B010137004) and Natural Science Foundation of Guangdong Province,China(2021A1515011798).
通信作者:吴昊天(wuht@scut.edu.cn)

针对弱口令发起攻击，一旦攻击成功，将会导致用户和系统的大量隐私信息被泄露，带来不可预计的后果。因此，通过研究口令猜解方法来了解用户设置口令的规律和攻击者的攻击手段，对保护系统及用户信息安全具有重要意义。

传统的口令猜解方法有基于字典的口令猜解方法和基于统计概率的口令猜解方法。基于字典的猜解方法是将口令中常用的字符和字符串构成字典，对字典中的词汇进行多规则变换以生成口令猜测集，但是由于变换规则需要人为制定，因此其猜解范围有限。基于统计概率的猜解方法则是对口令进行统计建模，生成带概率的口令猜测集。与基于字典的方法相比，基于统计概率的方法理论性较强，在口令猜解上也取得了一定的进展，但该方法需要一定的先验知识或结构假设，口令猜解空间有限。

基于深度学习的口令猜解方法是随着深度学习技术发展起来的，与传统口令猜解方法相比，它有较大的口令输出空间，而且可以产生训练集中没有的新的口令模式。根据网络模型的不同可以将其分为基于循环神经网络(Recurrent Neural Networks，RNNs)的口令猜解模型和基于生成式对抗网络(Generative Adversarial Networks，GANs)的口令猜解模型。基于 RNNs 的口令猜解模型的口令猜解过程本质是马尔可夫过程，其通过循环神经网络学习口令中前后字符的关系，之后根据已生成的口令字符依次预测下一个口令字符的概率，其破解率在大猜测集下高于传统的口令猜解方法。基于 GANs 的口令猜解模型通过判别器指导生成器自主学习口令空间的分布规律，使得生成器能生成具有相似分布的口令猜测集。对于相同的破解率，现有基于 GANs 的口令猜解模型需要生成比其他两类方法更多的口令才能达到，因此其口令猜解效率还有待进一步提高。

有学者分析，基于 GANs 的口令猜解模型猜解效率较低的原因是 GANs 模型大多是用于图像生成任务，对图像而言，其结果并不需要那么精细，即使存在一些偏差，也不会明显影响最终结果，但是对于口令猜解任务来说，一个字符的偏差也会导致口令无法和测试集匹配[4]。经过分析，我们认为基于 GANs 的猜解模型的偏差来源于判别器对生成器的指导不足，现有的 GANs 猜解模型只能通过判别器对完整口令的评判结果来对生成器进行简单指导，不能直接指导生成器生成口令的过程，从而导致 GANs 猜解模型生成的口令偏差较大，口令猜解效率较低。另外，GANs 猜解模型对生成器的指导存在滞后问题，即使生成器生成的部分口令已经显示出偏差，还是需要等待生成完整口令后才能指导生成器更新。Yu 等[5]提出了一种序列生成模型 seqGAN，seqGAN 通过 rollout 策略和判别器对完整状态序列中每一时间步的动作进行奖赏，之后根据奖赏的等权平均值更新生成策略。它虽然考虑到了每一时间步动作的奖赏，但是依然需要等待生成器生成完整序列后再进行更新，因此 seqGAN 中判别器对生成器生成过程的指导是一种简单的间接指导。

针对上述问题，本文提出了一种基于强化学习的口令猜解模型——AC-Pass。本文的主要贡献如下：

(1)首次在口令猜解任务中引入了强化学习框架，通过强化学习 Actor-Critic 算法共同训练 Actor 网络、Critic 网络和判别器，以判别器和 Critic 网络输出的奖赏不断优化 Actor 网络的生成策略。强化学习的引入量化了口令序列生成过程中每一时间步动作和状态的价值，实现了对 Actor 网络的强化指导。

(2)构建了价值网络 Critic。Critic 网络用于估计口令序列生成过程中每一序列状态的未来回报，与判别器输出的动作奖赏一起指导 Actor 网络进行单步更新，实现对口令序列生成过程的强化指导，从而减小了模型生成口令的偏差。

(3)在 RockYou，LinkedIn，CSDN 口令集上对 AC-Pass 模型进行充分的实验和分析，实验结果表明，无论是同源测试集还是异源测试集，AC-Pass 模型在 9×10^8 猜测集上的口令破解率均高于现有的基于 GAN 的口令猜解模型。另外，当测试集与训练集之间的口令空间分布差异显著时，AC-Pass 模型的猜解性能优于 PCFG 模型。

2 相关工作

2.1 口令猜解方法

在深度学习发展之前，口令猜解方法主要分为两类。第一类是基于字典的方法，该方法将口令常用字符或字符串构成特定字典，通过对字典中的词汇进行多规则变换得到口令猜测集，对应的口令猜解工具有 HashCat 和 John the Ripper。第二类是基于统计概率的方法，典型代表有 Markov 模型[6]和概率上下文无关文法(Probabilistic Context-Free Grammar，PCFG)模型[7]。Markov 模型是对口令前后字符的依赖关系进行建模，而 PCEG 模型是基于口令的结构组成进行建模。之后许多研究者陆续提出 Markov 模型和 PCFG 模型的改进方法，如 Tansey[8]提出分层 Markov 模型方法，将一阶 Markov 模型扩展到 n 层模型上，解决了一阶 Markov 模型中部分口令生成模型信息丢失的问题；Dürmuth 等[9]提出了 OMEN 方法，在 Markov 模型的基础上按概率递减顺序枚举生成的口令，提高了口令猜解的速度和效率；Houshmand 等[10]在详细分析了键盘规则在口令生成中的重要性之后，提出在 PCFG 的基础上加入键盘词模式。此外，为了增强口令猜解方法的针对性，研究者们还提出了多种融合用户个人信息的口令定向攻击算法，如 Targeted-Markov[11]，Personal-PCFG[12]和 TarGuess 猜测框架[13]。

随着深度学习技术的发展，基于深度学习的口令猜解方法已成为第三类口令猜解方法，可根据网络模型的不同将其分为基于 RNNs 的口令猜解模型和基于 GANs 的口令猜解模型。Melicher 等[14]首次将深度神经网络引入口令猜解，提出了基于 RNN 的口令猜解模型 FLA。FLA 模型在口令猜解任务中取得了较好的成果，证明了深度神经网络在口令猜解任务上的有效性，其不足之处在于它限制了口令格式，因此生成的口令不具有广泛性。LSTM(Long Short-Term Memory)相比 RNN 能捕捉更长序列间的相关性，之后，Xu 等[15]提出了基于 LSTM 的口令猜解模型，其破解率在 3.35×10^9 猜测数下高于 Markov 模型和 PCFG 模型。一些研究者还将 RNNs 和 PCFG 模型相结合提出混合攻击模型，如 PL(PCFG+LSTM)[16]和 PR/PR+(PCFG+RNN)[4]，与 PCFG，Markov 和 RNNs 模型相比，混合攻击模型的破解率有一定的提高。

Hitaj 等[17] 于 2019 年首次提出基于生成式对抗网络(Generative Adversarial Network, GAN)的口令猜解模型——PassGAN,其本质是带策略梯度的 Wasserstein GAN[18]。其中的生成器致力于生成使判别器判断为真的虚假口令,而判别器则希望能够准确判断口令是来源于生成器还是真实口令集,两个网络以博弈的方式进行训练,直到判别器难以判断生成器生成的口令是虚假口令还是真实口令。当生成 5×10^{10} 个猜测口令时,PassGAN 在同源测试集上的破解率为 34.192%,在异源测试集上的破解率为 21.9%。之后,Nam 等[19]提出改进 passGAN 的方法:一是使用在序列问题上表现良好的 RNN 代替 passGAN 中的卷积神经网络(CNN),实验结果表明,RNN 实现的 passGAN 口令猜解效率有所提高,生成的口令与泄露口令集的空间分布有更高的相似性;二是在第一种改进方法的基础上采用双鉴别器的 GAN 结构,使模型训练更加稳定,实验结果显示,改进的 PassGAN 在大猜测集($\geqslant2.5\times10^{10}$)下的破解率比 PassGAN 提高了 20%左右,而且在与训练集空间分布相差较大的测试集上的破解性能优于 PCFG 模型。Pasquini 等[20]同样对 PassGAN 提出了改进方法:在口令字符的独热编码中添加噪声,再对加噪后的编码进行归一化,此方法改善了 GAN 模型训练时的模式坍塌问题,使得模型在大猜测集下(10^{10} 量级)的破解率得到了提高,但在小猜测集下的破解率提升较小。这两种改进方法虽然都提高了 PassGAN 模型的口令破解率,但是仍然没有从根本上解决基于 GANs 的口令猜解模型生成的口令存在较大偏差的问题。

2.2 强化学习

强化学习是一种机器学习算法,用于描述和解决智能体通过与环境交互不断学习策略以达成回报最大化的问题。强化学习算法可以分为基于值函数和基于策略两大类,基于值函数的强化学习算法定义了状态或动作的价值函数,智能体倾向于根据输出的价值选择价值最高的状态或动作,常见算法有 DQN(DeepQ-Network)[21];基于策略的强化学习算法定义了策略函数,该函数能输出下一个动作的概率,智能体基于当前的状态和概率选择相应动作,也就是说,智能体不一定会选择概率最高的动作,常见算法有 DPG(Deterministic Policy Gradient)[22]。此外,还有将基于值函数和基于策略相结合的 Actor-Critic 算法[23],Actor-Critic 算法中的 Actor 对应策略函数,负责选择合适的动作,Critic 对应价值函数,负责评估 Actor 的表现,并指导 Actor 下一阶段的动作,常见算法有 DDPG(Deep Deterministic Policy Gradient)[24]和 A3C(Asynchronous Advantage Actor-Critic)[25]。

强化学习也越来越多地应用在复杂场景、多任务场景中[26]。在复杂应用场景下,算法样本效率低是强化学习模型的主要缺陷。为了解决这个问题,对于无模型类方法使用离线方法(Off-policy)学习,使用行动策略产生样本,存入经验池,之后通过重放不同策略的采样经验来优化目标策略。该方法不仅提高了样本效率,也降低了样本复杂度。对于 model-based 则进行策略学习,从采样数据中对环境进行建模,之后通过模拟仿真自动生成大量的样本数据,使用规划的手段快速进行策略学习。在多任务场景下,强化学习算法泛化性能较差。针对这个问题,一个解决思路是多任务强化学习,其核心思想是在不同但相关的源任务和目标任务之间迁移知识,以提升用于学习目标任务的机器学习算法的性能;另一个解决思路是元强化学习,通过学习与相似任务匹配的内部表示,为模型提供一种使用少量样本快速适应新任务的方法[26]。

目前,强化学习在自然语言处理领域的应用已非常广泛。如 Yu 等[5]提出的 SeqGAN 模型基于策略梯度算法实现了文本序列的自动生成;Lin 等[27]提出的基于策略梯度的对抗学习框架 RankGAN 可产生语义连贯的语言描述;Fedus 等[28]提出的 MaskGAN 模型通过上下文语义填充缺失部分的场景,基于 Actor-Critic 算法训练生成器,同时采用极大似然估计(Maximum Likelihood Estimate, MLE)和随机梯度下降算法训练判别器,使得生成器能够生成语义连贯的文本。

3 基于 Actor-Critic 的口令猜解模型

为了对生成网络生成口令过程进行强化指导以减小基于 GANs 的猜解模型生成口令的偏差,本文受 seqGAN[5] 的启发,提出了一种基于强化学习的口令猜解模型——AC-Pass,本节首先介绍该模型的总体架构,之后对各个部分进行详细说明。

首先将口令序列生成过程形式化为马尔可夫决策过程,其中的四大要素状态(s_t)、动作(a_{t+1})、策略($p(a_{t+1}|s_t)$)和奖励($R(s_t,a_{t+1})$)定义如下。

s_t 表示口令生成过程中 t 时间步的状态,包括 t 时间步之前生成的所有口令字符,可表示为 $s_t=(y_1,y_2,\cdots,y_i,\cdots,y_t)$,$y_i$ 来源于候选口令字符表 Y。

a_{t+1} 表示在 t 时间步基于当前状态 s_t 生成的口令字符(即动作)。注意,在口令序列生成问题中,动作 a_{t+1} 一旦被选定为候选口令字符表 Y 中的某个字符,下一个状态 s_{t+1} 就确定了,即状态转移概率 $P(s_{t+1}|s_t=(y_1,y_2,\cdots,y_i,\cdots,y_t),a_t=y_{t+1})=1$。

$p(a_{t+1}|s_t)$ 表示在 t 时间步基于当前状态 s_t 选择动作 a_{t+1} 的概率。

$R(s_t,a_{t+1})$ 表示在 t 时间步基于当前状态 s_t 选定动作为 a_{t+1} 后,最终可获得的累积折扣奖赏。

3.1 模型概述

本文提出的 AC-Pass 模型包括 Actor 网络、Critic 网络和判别器,其中 Actor 网络用于学习真实口令的空间分布规律并生成口令猜测集;Critic 网络是一个状态值函数,用于估计 Actor 网络生成口令序列过程中每一状态的未来回报;判别器则用于对 Actor 网络每一时间步的动作进行评价。图 1 为 AC-Pass 模型一个时间步的学习过程图。Actor 网络根据已生成的口令字符(即状态)进行决策,生成下一个口令字符(即动作),判别器和 Critic 网络分别对动作、获得的下一个状态进行评价,获得对应的单步奖赏和状态回报,随后 Actor 网络会根据该单步奖赏和状态回报优化下一个时间步的生成策略。

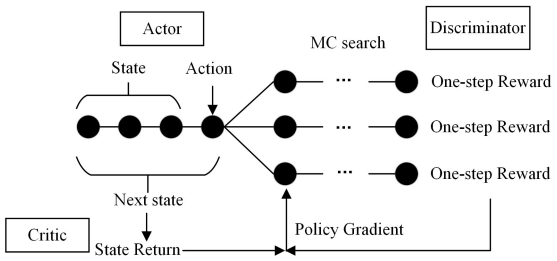


图 1 AC-Pass 模型一个时间步的学习过程

Fig. 1 One time step learning process of AC-Pass mode

3.2 Actor 网络

在 t 时间步时，Actor 网络（记为 G_θ）基于当前状态 s_t 和概率 $G_\theta(a_{t+1}\mid s_t)$ 选择下一个口令字符 a_{t+1}，得到下一个状态 $s_{t+1}=(s_t,a_{t+1})$。Actor 网络的目标是最大化每个时间步动作的期望回报 J_θ。

$$J_\theta(s_t,a_{t+1})=E_{a_{h+1}\sim G_\theta}\sum_{t+1}^{T}\gamma^{h-t}r(s_h,a_{h+1}) \tag{1}$$

其中，T 为完整口令序列的长度，$t+1\in[1,T)$；$\gamma\in[0,1]$ 为衰减系数，表明了未来奖赏相对于当前奖赏的重要程度，$\gamma=0$ 表示不考虑未来奖赏，只考虑当前奖赏，$\gamma=1$ 则表示未来奖赏和当前奖赏一样重要，本文设定 $\gamma=0.9$；$r(s_h,a_{h+1})$ 表示在 h 时间步基于状态 s_h 选择动作 a_{h+1} 所获得的单步奖赏；$R(s_t,a_{t+1})=\sum_{t+1\leqslant h+1\leqslant T}\gamma^{h-t}r(s_h,a_{h+1})$ 表示从 t 时间步开始一直到最后一个时间步所有单步奖赏的带衰减总和；$J_\theta(s_t,a_{t+1})$ 为 t 时间步开始遵循生成策略 G_θ 获得的期望回报。

由式（1）可知，求解 J_θ 的关键在于获得单步奖赏 $r(s_h,a_{h+1})_{1\leqslant h+1\leqslant T}$。当 $h+1=T$ 时，$s_T=(s_{T-1},a_T)$ 是完整的口令序列，可直接输入判别器获得对应的单步奖赏 $r(s_{T-1},a_T)$；当 $1\leqslant h+1<T$ 时，$s_{h+1}=(s_h,a_{h+1})$ 不是完整的口令序列，不能直接输入判别器得到对应单步奖赏。鉴于此，我们考虑是否可以通过采样填充不完整口令序列，再输入判别器得到单步奖赏。但是在实际应用时，考虑到时间因素，我们无法进行足够数量的交互采样，这会导致 $r(s_h,a_{h+1})$ 存在较大方差，进而导致 $R(s_t,a_{t+1})$ 也存在较大方差。受 DQN[21] 算法的启发，本文使用动作值函数 Q 近似 $R(s_t,a_{t+1})$：

$$Q(s_t,a_{t+1})=R(s_t,a_{t+1}) \tag{2}$$

则目标函数 $J_\theta(s_t,a_{t+1})$ 的梯度为：

$$\nabla_\theta J_\theta=E_{a_{t+1}\sim G_\theta}[\nabla_\theta\log G_\theta(s_t,a_{t+1})\cdot Q(s_t,a_{t+1})] \tag{3}$$

如果策略权重 $Q(s_t,a_{t+1})$ 恒为正值，那么生成策略 G_θ 会掉入正数陷阱，所有被采样到的口令字符下一次被采样的概率都会增加，还可能出现部分“好”的口令字符没有被采样到的情况。为了解决这个问题，本文选取 $Q(s_t,a_{t+1})$ 的期望作为基准值，$Q(s_t,a_{t+1})$ 与基准值的差值作为策略权重。又 $E[Q(s_t,a_{t+1})]=V(s_t)$，根据马尔可夫决策过程和状态转移概率 $P(s'\mid s,a)=1$ 可得新的策略权重为：

$$Q(s_t,a_{t+1})-V(s_t)=r(s_t,a_{t+1})+\gamma V(s_{t+1})-V(s_t) \tag{4}$$

其中，$V(s_t)$ 是状态值函数，表示从状态 s 开始并采用同样的生成策略 G_θ 获得的期望回报，$r(s_t,a_{t+1})$ 为单步奖赏，γ 为衰减系数，$\delta_t=r(s_t,a_{t+1})+\gamma V(s_{t+1})-V(s_t)$ 被称作时间差分误差（TD-error）。对于 $V(s)$ 和 $r(s,a)$ 的计算，我们将在下面的小节进行详细说明。由此得到 Actor 网络的梯度和参数更新式为：

$$\nabla_\theta J_\theta(s_t,a_{t+1})=E_{G_\theta}[\nabla_\theta\log G_\theta(s_t,a_{t+1})\cdot\delta_t] \tag{5}$$

$$\theta_{t+1}\leftarrow\theta_t+\alpha_t\cdot\nabla_\theta J_\theta \tag{6}$$

其中，δ_t 是 t 时间步的 TD-error，α_t 为 t 时间步的学习速率。循环神经网络是处理时间序列问题的常用网络，在口令猜解等任务上有较好的表现。因此，在本文提出的 AC-Pass 模型中，采用 LSTM 作为 Actor 网络。

3.3 Critic 网络

Critic 网络是用于评价口令序列状态优劣的价值网络，记作 V_μ，其输入是口令序列状态 s，输出是从状态 s 开始预期获得的回报 $V_\mu(s)$。本文用 $V_\mu(s)$ 近似估计式（5）中的真实期望回报 $V(s)$，两者的差值越小越好，因此将 Critic 网络的损失函数定义为：

$$\min_\mu L(\mu)=\frac{1}{2}(V(s)-V_\mu(s)) \tag{7}$$

但如何获得真实期望回报 $V(s)$？根据马尔可夫决策过程和状态转移概率 $P(s'\mid s,a)=1$，可得：

$$V(s)\approx r(s,a)+\gamma V_\mu(s') \tag{8}$$

其中，$r(s,a)$ 为单步奖赏，$s'=(s,a)$ 是对应的下一个状态。在 $r(s,a)$ 较为准确的情况下，$r(s,a)+\gamma V_\mu(s')$ 总会比 $V_\mu(s)$ 更接近 $V(s)$ 的真实值。因此 Critic 网络的梯度和参数更新公式如下：

$$\begin{aligned}\nabla_\mu L(\mu)&=(V(s)-V_\mu(s))\nabla_\mu\log V_\mu(s)\\&\approx[r(s,a)+\gamma V_\mu(s')-V_\mu(s)]\cdot\nabla_\mu\log V_\mu(s)\\&\approx\delta\cdot\nabla_\mu\log V_\mu(s)\end{aligned} \tag{9}$$

$$\mu_{t+1}\leftarrow\mu_t+w_t\cdot\nabla_\mu L(\mu) \tag{10}$$

其中，w_t 是 t 时间步的学习速率。口令序列状态由 Actor 网络已生成的口令字符组成。为了使 Critic 网络能更准确地估计从状态 s 开始获得的回报，同时适应输入状态的可变化性，本文选择 LSTM 作为 Critic 网络。

3.4 判别器

口令由一个个口令字符组成，每一个口令字符都会影响完整口令的匹配结果。本文使用判别器评估 Actor 网络基于当前状态 s 选择口令字符 a（即动作）的优劣，将完整口令序列输入判别器，输出判断其为真实口令的概率，并将该概率作为对应口令字符 a 的单步奖赏 $r(s,a)$。当 $t+1=T$ 时，$s_T=(s_{T-1},a_T)$ 是完整的口令序列，直接由判别器得到 $r(s_{T-1},a_T)$；当 $1\leqslant t+1<T$ 时，$s_{t+1}=(s_t,a_{t+1})$ 不是完整的口令序列，因此使用 rollout 算法和蒙特卡罗搜索方法计算单步奖赏。具体来说，构建一个 Actor 网络的完全副本 G_β，使用 G_β 生成 $(s_t,a_{t+1})_{1\leqslant t+1<T}$ 余下的 $T-(t+1)$ 个口令字符，其蒙特卡罗搜索 N 次采样的结果表示如下：

$$\{Y_{1:T}^1,Y_{1:T}^2,\cdots,Y_{1:T}^n,\cdots,Y_{1:T}^N\}=MC_{G_\beta}(Y_{1:t+1};N) \tag{11}$$

其中，$Y_{1:t+1}^n=(y_1,y_2,\cdots,y_{t+1})$，$Y_{t+2:T}^n$ 是 G_β 对 $Y_{1:t+1}^n$ 剩余口令字符的一次采样结果。将 $(s_t,a_{t+1})_{1\leqslant t+1<T}$ 的 N 次采样输入判别器得到 N 个单步奖赏，将这 N 个单步奖赏的平均值作为 $r(s_t,a_{t+1})_{1\leqslant t+1<T}$。因此单步奖赏 $r(s,a)$ 的计算式如下：

$$r(s_t=Y_{1:t},a_{t+1}=y_{t+1})=\begin{cases}\dfrac{1}{N}D_\varphi(Y_{1:T}^n),Y_{1:T}^n\in MC_{G_\beta}(Y_{1:t+1};N), & t+1<T\\ D_\varphi(Y_{1:T}), & t+1=T\end{cases} \tag{12}$$

其中，D_φ 为判别器。

判别器D_φ通过真实口令和 Actor 网络生成的虚假口令进行训练。对于真实口令序列，我们希望判别器的输出概率趋近于1；对于虚假口令序列，则希望判别器的输出概率趋近于 0，因此判别器的损失函数如式(13)所示。

$$DLoss_\varphi = -E_{Y\sim R_{data}}[\log D_\varphi(Y)] - E_{Y'\sim G_\theta}[\log(1-D_\varphi(Y'))] \tag{13}$$

其中，R_{data}是口令集的真实分布，G_θ表示 Actor 网络的生成策略，指代其学习到的口令集分布。

判别器的关键在于特征提取，只有从输入口令序列中提取到准确有效的特征，才能保证输出概率的准确性，进而保证$r(s,a)$的准确性。已知卷积神经网络在文本分类中有较为突出的表现[29]，因此选取卷积神经网络作为特征提取器；同时，为了从输入口令序列中提取到尽可能多的不同粒度的特征，在判别器中设置了多个平行的卷积核大小不同的卷积层。图 2 给出了判别器网络结构图。输入完整口令序列后，首先进行词嵌入处理，之后通过多个平行卷积层和池化层提取特征，然后汇总提取到的特征，依次输入残差层和 dropout 层中，最终得到该完整口令序列为真实口令的概率。引入残差层和 dropout 层是为了避免出现网络退化和过拟合问题。

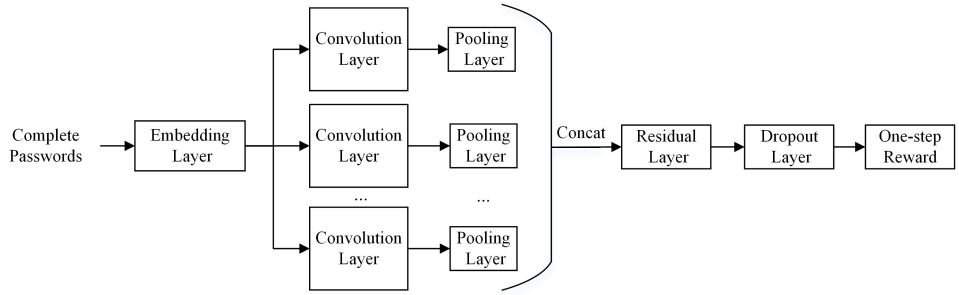


图 2 判别器网络结构

Fig. 2 Network structure of discriminator

3.5 算法流程

AC-Pass 模型的训练流程如算法 1 所示。首先采用极大似然估计对 Actor 网络进行预训练，然后使用预训练后的 Actor 网络生成部分虚假口令，将其与等量的真实口令一起输入判别器进行预训练。预训练完成后，需要对 Actor 网络、Critic 网络和判别器交替进行训练。Actor 网络每生成一个口令字符，就基于对应的动作奖赏和状态奖赏依次更新 Actor 网络和 Critic 网络的参数；然后定期利用 Actor 网络生成的虚假口令和真实口令训练判别器，从而提升判别器的判别能力，使其与 Actor 网络和 Critic 网络保持良好的同步，直到 Actor 网络获得较优的生成策略，能够生成与真实口令相似的口令。

算法 1 AC-Pass 模型训练流程

输入：(G_θ, G_β, V_μ, D_φ, R_{data}, T, γ, ac-iters, d-iters)

1. 初始化G_θ, V_μ, D_φ的网络参数；
2. 在R_{data}上利用 MLE 对G_θ进行预训练
3. $\beta \leftarrow \theta$
4. 基于G_θ生成的虚假口令和R_{data}对D_φ进行预训练
5. for epoch←1 to total-epoch do
6. for iteration←1 to ac-iters do
7. Initialize s_0(first state)
8. for t←1 to T do
9. Generate a password character $a_t \sim G_\theta(\cdot|s_{t-1})$, observe s_t
10. Compute $r(s_{t-1}, a_t)$ by Eq. (12), compute $V_\mu(s_{t-1})$, $V_\mu(s_t)$ by V_μ
11. $\delta_t \leftarrow r(s_{t-1}, a_t) + \gamma V_\mu(s_{t-1}) - V_\mu(s_t)$
12. Update G_θ via policy gradient Eq. (6)
13. Update V_μ via Eq. (10)
14. end for
15. end for
16. $\beta \leftarrow \theta$
17. for iteration←1 to d-iters do
18. Use current G_θ to generate fake passwords and combine with given real passwords R_{data}
19. Train discriminator D_φ by Eq. (13)
20. end for
21. end for

对 Actor 网络和判别器进行预训练是为了避免模型训练出现训练困难、难收敛等问题。由式(6)、式(10)以及式(13)可知，Actor 网络生成策略的优化依赖于 Critic 网络的输出$V_\mu(s)$和与判别器输出相关联的单步奖赏$r(s,a)$，当$V_\mu(s)$和$r(s,a)$值越准确(即越近似对应的真实值)，Actor 网络的生成策略训练就越好；同样地，Critic 网络的训练也依赖于判别器；而判别器的训练又依赖于 Actor 网络的生成策略G_θ。可见，Actor 网络、Critic 网络和判别器三者相互依赖和制约，这导致 AC-Pass 在训练过程中容易出现训练困难、难收敛等问题。针对这些问题，同时为了保证 Actor 网络、Critic 网络训练时单步奖赏$r(s,a)$的准确性，因此对 Actor 网络、判别器进行预训练。

4 实验结果与分析

根据多个泄露口令集发现，长度在 6～10 之间的口令在各口令集占比达 60%以上[3]，这说明用户在设置口令时偏向于设置长度在 6～10 之间的口令，因此我们重点研究如何猜解长度在 6～10 之间的口令。

4.1 数据集

本文使用了国内外 3 个大规模的真实泄露口令集：RockYou，LinkedIn 和 CSDN。其中 RockYou 口令集与 LinkedIn 口令集来源于文化背景相同但领域不同的网站，口令空间分布差异较小；CSDN 口令集来源于与前两个口令集文化背景

和领域都不相同的网站，口令空间分布差异较大。我们在RockYou口令集中随机选择600万个长度在6～10之间的口令，选取80%作为训练集，将剩下的20%中未出现在训练集的不重复口令作为测试集；另外，将LinkedIn和CSDN口令集中所有长度在6～10之间且未出现在RockYou训练集的不重复口令也作为测试集。表1详细列出了所用数据集的统计信息。

表1 数据集的统计信息

Table 1 Statistics of datasets

Dataset	Training Set	Testing Set
RockYou	4800000	1197868
LinkedIn	0	42068370
CSDN	0	2709189

4.2 对比模型

为了评价本文提出的口令猜解模型，采用文献中已有的一些口令猜解模型进行性能对比，包括基于统计概率的口令猜解模型PCFG[7]、基于GAN的口令猜解模型PassGAN[17]和seqGAN[5]。

PCFG：对口令的结构组成进行建模。首先统计口令集的口令结构信息并用上下文无关文法表示，然后利用优先队列以概率降序排列的方式产生预口令结构，最后使用字符组件对预口令结构进行填充，得到以概率降序排列的口令猜测集。

PassGAN：首个基于GAN的口令猜解模型。在生成器生成完整口令后，首先将完整口令输入判别器中判断其是否为真，然后根据判别器的判断指导生成器进行参数更新，最后用训练好的生成器生成口令猜测集。

seqGAN：首个基于强化学习和GAN的离散序列生成模型。将离散序列生成过程形式化为马尔可夫决策过程，生成器生成完整序列后，通过rollout策略和判别器对序列中每一时间步的动作给出奖赏，之后根据奖赏的等权平均值进行策略梯度更新，最后得到较优的生成策略以生成离散序列。

4.3 参数设置与评估

Actor网络与Critic网络中LSTM的隐藏层维度均设置为128，判别器的词嵌入维度（emb_dim）为64，卷积层共5个，卷积核大小分别为（6，emb_dim）（7，emb_dim）（8，emb_dim）（9，emb_dim）（10，emb_dim），对应的卷积核个数为100，200，200，200，200。所有网络统一采用Adam优化器。Actor网络的完全副本G_β不参与梯度下降过程，直接根据G_θ更新自身参数。为了增加长序列强化学习的稳定性，G_β的参数更新稍滞后于G_θ，滞后因子设置为0.8。

本文选择口令破解率作为模型猜解性能的评估指标，包括与训练集来源相同的同源测试集上的破解率以及与训练集来源不同的异源测试集上的破解率。口令猜解模型的目标就是产生高概率为“真”的口令，因此，模型口令破解率越高，生成的未匹配口令为用户真实口令的概率就越大。

口令破解率的计算方法为：

$$CrackingRate=\frac{guess\_unique\_nums}{testing\_set\_nums}\times 100\% \tag{14}$$

其中，$guess_unique_nums$为生成的口令猜测集与测试集不重复的口令个数，$testing_set_nums$为测试集包含的口令总数。

4.4 实验结果与分析

本文使用RockYou训练集训练PCFG，PassGAN，seqGAN和AC-Pass模型，为了保证实验结果的可靠性，seqGAN和AC-Pass的预训练设置（包括用于预训练的口令集数量、训练迭代次数等）基本相同，之后各模型分别生成9×10^8大小的猜测集，并在RockYou测试集进行测试，结果如表2所列。可以看出，本文提出的AC-Pass模型的口令破解率略低于PCFG模型。当猜测集大小为1.3×10^9时，AC-Pass模型在RockYou测试集上的破解率可达20.002%，虽然所提模型比PCFG模型在同一破解率下多生成了大约4×10^8个口令，但因为算力的增长和存储成本的降低，这多余的生成成本在可接受范围内。此外，不同于PCFG模型是基于口令的统计概率生成口令，需要结构假设，生成口令总量有限，AC-Pass模型可以完全自主学习口令集的空间分布规律，生成大量与口令集分布相近的不重复口令。

表2 不同模型的口令破解率对比（RockYou测试集）

Table 2 Comparison of password cracking rate of different models (RockYou testing set)

Model	Guessing set	Cracking rate/%
PCFG[7]	9×10^8	19.955
PassGAN[17]	9×10^8	8.692
seqGAN[5]	9×10^8	16.563
AC-Pass(ours)	9×10^8	18.426

由表2还可得，seqGAN模型的破解率较PassGAN提高了大约90%，AC-Pass模型的破解率较PassGAN提高了大约112%，说明AC-Pass模型和seqGAN模型对口令序列生成过程的指导有助于减小GANs模型生成口令的偏差。与seqGAN模型中判别器对口令序列生成过程简单的间接指导相比，AC-Pass模型增加了价值网络Critic，和判别器一起强化对口令序列生成过程的直接指导，破解率比seqGAN提高了大约11%，这表明AC-Pass模型对口令序列生成过程的强化指导进一步减小了基于GANs的猜解模型生成口令的偏差，提高了口令猜解效率。随后又在不同大小的猜测集上测试AC-Pass，PassGAN，seqGAN模型的口令破解率，结果如图3所示，随着口令猜测集的增大，破解率都呈现先快速增长再缓慢增长的趋势，AC-Pass与PassGAN和seqGAN之间口令破解率的差距也在缓慢增大。又由表3可知，AC-Pass模型有较大的口令输出空间。可以预计，随着猜测集的增大，AC-Pass模型的破解率会进一步增长，且会高于PassGAN和seqGAN的破解率。

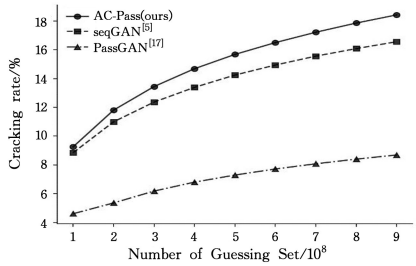


图3 不同模型在不同大小猜测集上的破解率

Fig. 3 Cracking rate of different models on guessing sets with differernt sizes

表 3 采用 AC-Pass 模型生成的不重复口令数量

Table 3 Number of unique passwords generated with AC-Pass model

Number of generated passwords	Number of unique passwords
9×10^8	218589897
1.1×10^9	253806510
1.3×10^9	287258773
1.5×10^9	319280991
1.7×10^9	350100980

为进一步探究 AC-Pass 模型的口令猜解能力，本文在与训练集口令空间分布差异较小的异源测试集——LinkedIn 测试集上测试各猜解模型的猜解性能，结果如表 4 所列。可以看出，PCFG 的口令破解率最高，AC-Pass 次之。与同源测试集(即 RockYou 测试集)相比，PCFG 模型在 LinkedIn 测试集上的口令破解率有小幅下降，AC-Pass，seqGAN，PassGAN 下降幅度较大，但本文提出的 AC-Pass 模型的破解率相比 PassGAN 提高了约 117%，相比 seqGAN 提高了约 11%。

表 4 不同模型的口令破解率对比(LinkedIn 测试集)

Table 4 Comparison of password cracking rates of different models (LinkedIn testing set)

Model	Guessing set	Cracking rate/%
PCFG[7]	9×10^8	15.703
PassGAN[17]	9×10^8	4.425
seqGAN[5]	9×10^8	8.601
AC-Pass(ours)	9×10^8	9.592

在与训练集口令空间分布差异较大的异源测试集——CSDN 测试集上对各模型进行测试，结果如表 5 所列。可以看出，AC-Pass 模型的口令破解率最高，猜解性能最好。其中 AC-Pass 模型的破解率相比 PCFG 模型提高了约 21%，相比 PassGAN 和 seqGAN 模型分别提高了 56%和 121%。然而 PCFG 模型在 RockYou 测试集和 LinkedIn 测试集上的破解率都高于 AC-Pass 模型，在 CSDN 测试集上的破解率却低于 AC-Pass 模型。当将猜测集增大到 1.1×10^9 时，AC-Pass 与 PCFG 在 CSDN 测试集上的口令破解率分别为 4.461%和 3.839%，AC-Pass 的口令破解率依然高于 PCFG。为了排除 CSDN 口令集的特殊性，我们增加了两个国内泄露口令集 Weibo 和 7K7K[2]，选取 Weibo 和 7K7K 口令集中长度在 6～10 之间的不重复口令作为测试集测试 PCFG 和 AC-Pass 模型的猜解性能，结果如表 6 所列。由表可知，AC-Pass 模型在 Weibo 和 7K7K 口令集上的破解率远高于 PCFG 模型。之后又测试了在 3.5×10^9 猜测集上 PCFG 模型对 Weibo 和 7K7K 口令集的猜解性能，结果如表 7 所列。对比表 6、表 7 可得，对于 Weibo 和 7K7K 测试集，AC-Pass 模型在 1.7×10^9 猜测集上的破解率高于 PCFG 模型在 3.5×10^9 猜测集上的破解率。这充分说明，当测试集与训练集口令空间分布相差较大时，本文提出的 AC-Pass 模型表现出了比 PCFG 更好的猜解性能。换言之，AC-Pass 模型在口令空间分布差异较大的猜解任务上具有一定的优势。

表 5 不同模型的口令破解率对比(CSDN 测试集)

Table 5 Comparison of password cracking rates of different models (CSDN testing set)

Model	Guessing Set	Cracking Rate/%
PCFG[7]	9×10^8	3.441
PassGAN[17]	9×10^8	2.673
seqGAN[5]	9×10^8	1.887
AC-Pass(ours)	9×10^8	4.175

表 6 AC-Pass 与 PCFG 模型的口令破解率对比

Table 6 Comparison of password cracking rates of AC-Pass and PCFG models

Password Source	Testing Set	Guessing Set	PCFG[7]	AC-Pass (ours)
Weibo	1519929	1.7×10^9	8.709	27.523
7K7K	1361661	1.7×10^9	7.861	26.288

表 7 PCFG 模型的口令破解率(Weibo，7K7K 测试集)

Table 7 Passwordc racking rates of PCFG model(Weibo, 7K7K testing set)

Password Source	Testing Set	Guessing Set	Cracking Rate/%
Weibo	1519929	3.5×10^9	13.471
7K7K	1361661	3.5×10^9	11.721

无论是在 LinkedIn 测试集还是 CSDN 测试集，AC-Pass 模型的破解率都高于 PassGAN 和 seqGAN 模型，这充分说明了 AC-Pass 模型对口令序列生成过程的强化指导能够有效减小基于 GANs 模型生成口令的偏差，提高口令猜解效率。我们进一步对比表 2、表 4 和表 5 发现，与 RockYou 测试集上的破解率相比，PassGAN，seqGAN，AC-Pass 模型在 LinkedIn 测试集上的破解率下降程度基本相等，但在 CSDN 测试集上的破解率下降程度相差较大，其中 seqGAN 最大，AC-Pass 次之，PassGAN 最小。这可能是因为，不同于 PassGAN 根据已生成的完整口令指导生成网络，seqGAN 和 AC-Pass 是对生成网络生成口令序列的过程进行指导，可以学习到更细粒度的口令空间分布规律，因此 seqGAN 和 AC-Pass 模型在 CSDN 测试集上口令破解率下降程度较大。

总的来说，本文提出的基于强化学习的口令猜解模型 AC-Pass 不仅在同源测试集和异源测试集上的破解率高于 PassGAN 和 seqGAN 模型，而且在与训练集空间分布相差较大的测试集上表现出比 PCFG 更好的猜解性能。另外，AC-Pass 模型有较大的口令输出空间，随着生成口令数目的增多，AC-Pass 模型的破解率持续增长。对于 AC-Pass 生成的未与测试集匹配的口令，它们有较大概率是未被泄露的真实口令，对后续的口令研究具有重要价值。

结束语 本文提出了一种基于强化学习的口令猜解模型 AC-Pass。AC-Pass 模型将口令序列生成过程形式化为马尔可夫决策过程，通过价值网络和判别器网络指导口令生成策略的更新，实现了对口令序列生成过程的强化指导。该模型提高了基于 GANs 的口令猜解模型的猜解效率，而且相比 PCFG 模型，所提模型在与训练集空间分布差异较大的口令集上表现出了更好的猜解性能。未来的工作考虑把传统的口令猜解方法或新的神经网络模型融合到 AC-Pass 模型中，以更好地学习口令空间的分布规律，进一步提升 AC-Pass 模型对口令(包括不同长度口令、同异源口令)的猜解能力。

参 考 文 献

[1] HAN W L, YUAN L, LI S S, et al. An Efficient Algorithm to Generate Password Sets Based on Samples[J]. Chinese Journal of Computers, 2017, 40(5): 1151-1167.

[2] LIU G S, QIU W D, MENG K, et al. Password Vulnerability Assessment and Recovery Based on Ruels Mined from Large-Scale Real Data[J]. Chinese Journal of Computers, 2016, 39(3): 454-467.

[3] XIE Z J, ZHANG M, LI Z H, et al. Analysis of Large-scale Real User Password Data Based on Cracking Algorithms[J]. Computer Science, 2020, 47(11): 48-54.

[4] WANG D, ZOU Y K, TAO Y, et al. Password Guessing Model Based on Recurrent Neural Networks and Generative Adversarial Networks[J]. Chinese Journal of Computers, 2021, 44(8): 1519-1534.

[5] YU L, ZHANG W, WANG J, et al. Seqgan: Sequence generative adversarial nets with policy gradient[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2017, 31(1), 2852-2858.

[6] NARAYANAN A, SHMATIKOV V. Fast dictionary attacks on passwords using time-space tradeoff[C]// Proceedings of the 12th ACM Conference on Computer and communications security. 2005: 364-372.

[7] WEIR M, AGGARWAL S, DE MEDEIROS B, et al. Password cracking using probabilistic context-free grammars[C]// 2009 30th IEEE Symposium on Security and Privacy. IEEE, 2009: 391-405.

[8] TANSEY W. Improved models for password guessing [EB/OL]. https://www. semanticscholar. org/paper/ImprovedModels-for-Password-Guessing-Tansey/3451ac7f102da12e1197c681b77d368ba3b19ac9.

[9] DÜRMUTH 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, 2015: 119-132.

[10] HOUSHMAND S, AGGARWAL S, FLOOD R. Next gen PCFG password cracking [J]. IEEE Transactions on Information Forensics and Security, 2015, 10(8): 1776-1791.

[11] WANG D, WANG P. The emperor's new password creation policies[C]// European Symposium on Research in Computer Security. Cham: Springer, 2015: 456-477.

[12] LI Y, WANG H, SUN K. A study of personal information in human-chosen passwords and its security implications[C]// IEEE INFOCOM 2016-the 35th Annual IEEE International Conference on Computer Communications. IEEE, 2016: 1-9.

[13] WANG D, ZHANG Z, WANG P, et al. Targeted online password guessing: An underestimated threat[C]// Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. 2016: 1242-1254.

[14] MELICHER W, UR B, SEGRETI S M, et al. Fast, lean, and accurate: Modeling password guessability using neural networks [C]// 25th {USENIX} Security Symposium({USENIX} Security 16). 2016: 175-191.

[15] XU L, GE C, QIU W, et al. Password guessing based on LSTM recurrent neural networks[C]// 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC). IEEE, 2017: 785-788.

[16] XIA Z Y, YI P, LIU Y, et al. GENPass: A multi-source deep learning model for password guessing[J]. IEEE Transactions on Multimedia, 2019, 22(5): 1323-1332.

[17] HITAJ B, GASTI P, ATENIESE G, et al. Passgan: A deep learning approach for password guessing [C]// International Conference on Applied Cryptography and Network Security. Cham: Springer, 2019: 217-237.

[18] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of wasserstein gans [J]. arXiv: 1704. 00028, 2017.

[19] NAM S, JEON S, KIM H, et al. Recurrent gans password cracker for iot password security enhancement [J]. Sensors, 2020, 20(11): 3106.

[20] PASQUINI D, GANGWAL A, ATENIESE G, et al. Improving password guessing via representation learning[C]// 2021 IEEE Symposium on Security and Privacy (SP). IEEE, 2021: 1382-1399.

[21] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518(7540): 529-533.

[22] SILVER D, LEVER G, HEESS N, et al. Deterministic policy gradient algorithms[C]// International Conference on Machine Learning. PMLR, 2014: 387-395.

[23] KONDA V R, TSITSIKLIS J N. Actor-critic algorithms[C]// Advances in Neural Information Processing Systems. 2000: 1008-1014.

[24] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning [J]. arXiv: 1509. 02971, 2015.

[25] MNIH V, BADIA A P, MIRZA M, et al. Asynchronous methods for deep reinforcement learning[C]// International Conference on Machine Learning. PMLR, 2016: 1928-1937.

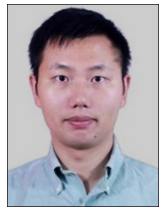
[26] YANG S M, SHAN Z, DING Y, et al. Survey of Research on Deep Reinforcement Learning[J]. Computer Engineering, 2021, 47(12): 19-29.

[27] LIN K, LI D, HE X, et al. Adversarial ranking for language generation [J]. arXiv: 1705. 11001, 2017.

[28] FEDUS W, GOODFELLOW I, DAI A M. Maskgan: better text generation via filling in the_ [J]. arXiv: 1801. 07736, 2018.

[29] ZHANG X, LECUN Y. Text understanding from scratch [J]. arXiv: 1502. 01710, 2015.

LI Xiaoling, born in 1998, postgraduate. Her main research interests include deep learning based password guessing and so on.



WU Haotian, born in 1980, Ph.D, associate professor. His main research interests include information hiding, privacy preservation, password guessing and blockchain.

(责任编辑:何杨)