智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

大语言模型驱动的口令管理系统优化与实践

刘志勇, 何道敬, 成嘉轩, 陈志雄, 梁承东, 彭世强

引用本文:
刘志勇, 何道敬, 成嘉轩, 等. 大语言模型驱动的口令管理系统优化与实践[J]. 智能系统学报, 2026, 21(1): 257-271.
LIU Zhiyong, HE Daojing, CHENG Jiaxuan, et al. Optimization and practice of password management system driven by large language models[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(1): 257-271.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202504017

DOI: 10.11992/tis.202504017

大语言模型驱动的口令管理系统优化与实践

刘志勇[1,2], 何道敬[2], 成嘉轩[1], 陈志雄[3], 梁承东[1], 彭世强[1]
(1. 广州竞远安全技术股份有限公司, 广东 广州 510641; 2. 哈尔滨工业大学 (深圳) 计算机科学与技术学院, 广东 深圳 518055; 3. 香港城市大学 电气工程学院, 香港 999077)

摘 要: 随着互联网服务的增多, 口令管理成为一大挑战。尽管口令管理系统是安全的解决方案, 但其可用性受到口令强度评估器和非随机口令生成器设计缺陷的制约, 导致口令评估不准确、生成口令强度不足且难以记忆。为解决这些问题, 提出了一种基于大语言模型的口令管理系统优化方案。该方案结合微调技术与检索增强生成技术, 设计了专门针对口令安全的大语言模型, 能够有效识别脆弱口令并提取深层语义特征。同时, 创新的非随机口令生成器框架提升了生成口令的强度和易记忆性。通过改进的 Zxcvbn 算法和口令猜测模型, 优化了口令强度评估器的准确性。该方案显著提高了口令管理系统的可用性, 促进了其在实际应用中的普及。

关键词: 网络安全; 密码管理; 身份认证; 人工智能; 大语言模型; 口令管理系统; 口令破解; 口令强度计; 口令生成器

中图分类号: TP304 **文献标志码:** A **文章编号:** 1673−4785(2026)01−0257−15

中文引用格式: 刘志勇, 何道敬, 成嘉轩, 等. 大语言模型驱动的口令管理系统优化与实践 [J]. 智能系统学报, 2026, 21(1): 257–271.

英文引用格式: LIU Zhiyong, HE Daojing, CHENG Jiaxuan, et al. Optimization and practice of password management system driven by large language models[J]. CAAI transactions on intelligent systems, 2026, 21(1): 257–271.

Optimization and practice of password management system driven by large language models

LIU Zhiyong[1,2], HE Daojing[2], CHENG Jiaxuan[1], CHEN Zhixiong[3], LIANG Chengdong[1], PENG Shiqiang[1]
(1. Guangzhou Jingyuan Security Technology Co., Ltd., Guangzhou 510641, China; 2. Department of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China; 3. Department of Electrical Engineering, City University of Hong Kong, Hong Kong 999077, China)

Abstract: As the number of internet services continues to grow, password management has become a significant challenge. Although password management system (PMS) provide secure solutions, their usability is limited by design flaws in password strength meters (PSM) and non-random password generators (NRPG), leading to inaccurate password assessments, insufficient password strength, and poor memorability. To address these issues, this paper proposes an optimization scheme for PMS based on large language model (LLM). The proposed approach combines fine-tuning techniques with retrieval-augmented generation, creating a specialized LLM model for password security that can effectively identify weak passwords and extract deep semantic features. Meanwhile, an innovative NRPG framework enhances both password strength and memorability. The accuracy of the PSM is optimized through an improved Zxcvbn algorithm and password guessing model. This solution significantly enhances the usability of PMS and promotes its widespread adoption in practical applications.

Keywords: network security; password management; authentication; artificial intelligence; large language model; password management system; password cracking; password strength meter; password generator

随着互联网服务的增加, 用户需要管理多个口令, 口令管理系统 (password management system, PMS) 已成为解决这一问题的主流方案[1]。PMS 通过自动生成复杂且安全的口令来平衡安全性与可记忆性。然而, 尽管 PMS 在口令管理中具有潜

收稿日期: 2025−04−23.
基金项目: 国家自然科学基金项目 (62376074); 国家重点研发计划项目 (2024YFE0215300); 深圳市科技计划项目 (KJZD20240903100505007, SGDX20230116091244004, JSGGKQTD20221101115655027).
通信作者: 何道敬. E-mail: hedaojinghit@163.com.

力，因其安全性、成本和可用性问题，广泛应用仍受限[2]。PMS 生成的高随机性口令虽提高了安全性，但也增加了记忆难度，尤其在跨设备验证时，口令的可用性显著降低，影响了用户体验[3-4]。已有研究集中在提高 PMS 的可用性。例如，Glory 等[5] 提出了基于说服性文本的口令生成系统，Oesch 等[6] 改进了传统非随机口令生成器 (non-random password generator, NRPG)，但仍存在脆弱口令识别和安全性保障不足的问题。因此，PMS 中的口令强度评估器 (password strength meter, PSM) 与 NRPG 的设计缺陷亟需改进[7]。

本研究旨在通过专用大语言模型 (large language model, LLM) 优化 PMS 的可用性，重点提升 PSM 和 NRPG 的设计。具体贡献包括：1) 通过构建专用 LLM 模型提升口令分析和猜测能力，并验证其有效性；2) 基于 LLM 设计新型 NRPG 框架，满足个性化口令创建需求；3) 提出重构口令强度评估算法的新方法，优化 PSM，并提高识别脆弱口令的精度。

1　相关工作

1.1　LLM 赋能口令安全研究

近年来，LLM 在口令安全领域的应用取得了显著进展[8]。2019 年，Li 等[9] 首次基于 Transformer 架构构建了双向掩码语言模型，并基于该模型开发了一种新型口令猜测模型。此后，Pal 等[10] 利用 14 亿条泄露的邮箱账户数据训练了口令相似度模型，并提出了针对定向攻击的防御策略。2021 年，Pasquini 等[11] 设计了一种基于 Transformer 架构的攻击者行为模拟方法，随后提出了新型的动态口令猜测策略，该策略能够根据具体攻击目标的上下文调整猜测方法，从而提高基于规则的字典攻击效率。

随着大规模预训练模型的迅速发展，学者们逐步将 LLM 的生成能力融入口令猜测和强度评估任务中。2023 年，Xu 等[12] 提出了基于双向 Transformer 的 PassBERT(bidirectional encoder representations from Transformers) 框架。该框架结合了预训练与微调机制，充分利用预训练模型捕获的口令分布特征，针对不同攻击场景提出了微调策略，增强了模型在实际攻击环境中的适应性。此外，生成模型也开始应用于口令生成领域。PassGAN[13] 是首个基于生成对抗网络 (generative adversarial network, GAN) 构建的口令生成模型，改进后的 Wasserstein GAN 学习了 RockYou 泄露口令数据集的分布规律，使生成的口令能够很好拟合该数据集中用户的常见口令创建习惯。Rando 等[14] 后续提出了 PassGPT 模型，该模型通过不断训练泄露的口令数据集来提高生成口令的猜测命中率，实验证明其在口令猜测性能上优于传统的 GAN 模型。

2025 年，Gewida 等[15] 提出了一种基于机器学习的工具，用于预测智能家居物联网设备中的密码漏洞，并利用 LLM 提供动态推荐。该方法通过提供数据驱动的模型，预测潜在漏洞并实时生成个性化的安全建议，从而增强物联网安全性。研究表明，检索增强生成 (retrieval-augmented generation, RAG) 架构相比基线方法，在召回率和 F_1 分数方面表现显著提升，展示了 RAG 在应对物联网设备口令安全漏洞中的潜力。

总体而言，通用 LLM 在提取口令模式和优化口令猜测策略方面已经展示出一定的潜力，但仍存在一些明显的局限性[16]。通用 LLM 通常是为处理广泛领域的自然语言任务而设计的，并未专门微调以满足口令安全领域的特殊需求[17]。因此，通用 LLM 在识别口令创建模式时，往往忽略了潜在的语义联系，导致其在口令猜测、PSM 和 NRPG 等赋能场景中的表现未能达到预期[18]。例如，基于通用 LLM 的 NRPG 未能充分考虑生成口令的安全性特征，如复杂度、不可预测性以及抵抗字典攻击的能力。此外，基于 LLM 优化 PSM 和 NRPG 任务通常需要模型深入理解、识别和分析口令文本中隐含的口令创建模式和对口令文本间相似性的敏锐察觉力，然而这些特性是通用 LLM 所欠缺的，这也是现有的通用 LLM 在口令安全领域无法充分发挥潜力的主要原因之一。因此，开发一种专门面向口令安全研究的 LLM，不仅能够提高模型在口令安全任务中的赋能表现，还能根本上解决通用 LLM 在专用任务中的局限性[19]。本研究旨在填补这一空白，通过设计和训练专门的 LLM 基础模型，提升 PMS 系统中 PSM 与 NRPG 的性能，从而提高 PMS 的可用性。

1.2　PMS 的优化方法

目前，关于 PMS 优化的研究主要集中在系统可用性与系统安全性两个维度。研究的核心目标是消除用户的使用顾虑并尽可能提高用户体验，从而推动 PMS 的普及和使用[1,3]。Collins 等[20] 在分析 PMS 使用意向数据后发现，调查对象普遍认为 PMS 产品提供商应严格控制系统的安全性，并定期向用户反馈 PMS 的安全状况，以消除用户对 PMS 安全性的担忧。此外，调查结果还表明，用户普遍期望 PMS 能够提供更加准确的口令强

度反馈, 并支持生成既安全又易于记忆的非随机口令功能, 以确保生成口令的安全性和便捷性。本文主要聚焦于 PMS 的可用性维度, 因此本节将概述与本研究相关的研究现状。

在优化 PMS 的 NRPG 方面, Oesch 等[6] 开发了一种基于说服性文本的口令生成系统, 通过引导用户的口令创建行为间接提高生成口令的可用性。Murmu 等[21] 基于双向生成对抗网络 (Bi-GAN) 提出了一种个性化的口令生成方法, 能够在更短时间内生成更具代表性的口令样本。然而, 这类方法未能有效保证生成口令的安全强度, 从而可能使用户使用弱口令, 带来安全隐患。为此, Houshmand 等[22] 提出了一种结合口令脆弱性识别、分析与纠正的综合解决方案, 显著降低了口令使用的安全风险。栗会峰等[23] 通过口令猜测对比实验发现, 目前 PMS 主流的非随机口令生成器在强化脆弱口令方面存在显著不足, 且难以兼顾生成口令的便捷记忆性。因此, 平衡口令安全性与易记性之间的矛盾, 依然是优化 PMS 中的 NRPG 的重要研究方向。

在优化 PMS 的 PSM 方面, Bonneau 等[24] 指出, 传统基于 Shannon 熵和猜测熵的评估指标无法准确模拟真实攻击者的行为, 导致口令强度评估结果不可靠。随着基于规则工具 (如 Zxcvbn[25]) 的出现, 口令强度评估的准确性有所提高, 但这些方法在处理多语言类型口令时的适应性和评估精度仍然存在局限, 尤其对于非英语口令 (例如汉语口令) 准确性较低。为此, 部分研究者致力于将 Zxcvbn 适配到非英语口令强度评估中。例如, Doucek 等[26] 成功将 Zxcvbn 适配到西斯拉夫语类口令数据集, 但其新版本的 Zxcvbn 仍难以推广到其他语言。

近年来, 基于神经网络的口令强度评估方法逐渐兴起。Melicher 等[27] 设计了一种能够估算口令破解所需猜测次数的神经网络生成器, 显著提高了口令强度反馈的准确性。Pereira 等[28] 提出了一种专门用于口令强度评估的轻量级深度学习框架, 实现了可解释的、概率化的口令强度评估。

尽管如此, 现有主流的口令强度评估器仍面临诸多挑战: 一方面, 传统的熵基方法过度强调随机性, 忽视了社会工程学因素和已知口令模式, 导致评估结果不够准确; 另一方面, 基于多类型口令猜测模型的评估方法, 因硬件配置和猜测性能的差异, 常导致强度评估结果的不一致, 进而影响用户对口令强度评估器的信任。因此, 如何设计一种能够适应汉语口令的 Zxcvbn 模型, 或是改进现有的口令猜测模型以更好评估汉语口令强度, 仍是优化 PMS 中的 PSM 的关键研究方向。

2　主流 PMS 的脆弱性调研

PMS 脆弱性调研对优化当前 PMS 至关重要[1], 因为它们直接影响优化工作质量、效率和优化后 PMS 的性能提升效果。本节将从 PMS 信息收集、PMS 性能对比分析两个方面展开, 系统地分析当前主流 PMS 在口令强度评估和生成方面的不足之处, 为后续研究提供理论依据。

2.1　主流 PMS 的数据采集

本研究的数据采集工作主要通过安全网站、学术平台、学术期刊以及专业论坛等多个渠道进行, 旨在获取主流 PMS 的基础信息。具体采集内容包括: PMS 的基本功能、面向客户的主要语言支持、用户规模、系统功能属性及其在实际应用中的显著缺陷。例如, 系统在口令强度评估准确性、脆弱口令识别能力、生成口令的随机性和易记性等方面存在的不足。

数据采集的具体步骤如下:

1) 基于安全网站、学术搜索引擎、期刊文章、会议论文及专业论坛, 收集并核实公开的 PMS 相关信息, 确保信息的真实性与可靠性, 制作如表 1 所示的 PMS 基本信息表。

表 1　国内外主流 PMS 基本信息

Table 1　Basic information of mainstream PMS at home and abroad

地区	PMS名称	系统基本信息	功能插件
国内PMS	腾讯密码管家	面向QQ/微信个人用户; 主动使用口令管理用户量未知; 免费	PSM、随机口令生成器(random password generator, RPG)、NRPG(支持自定义生成)
	芯盾时代口令管家	面向企业客户; 2000+家企业客户; 按企业用户数定制收费	PSM、RPG、NRPG(支持自定义生成)
	安恒信息密码保险箱	面向企业客户; 客户量未公开; 按企业需求定制	PSM、RPG、NRPG(无自定义生成)
	夏冰加密软件	面向个人用户; 100万; 基础版(免费)、升级版(30元/月)	PSM、RPG、NRPG(无自定义生成)

续表 1

地区	PMS名称	系统基本信息	功能插件
国际PMS	1Password	面向个人/企业用户；1 500万+；个人版(2.99美元/月)、家庭版(4.99美元/月)	PSM、RPG(可定制长度、符号类型)、NRPG(支持自定义生成)
	LastPass	面向个人/企业用户；2 500万+；免费版(功能受限)、升级版(3美元/月)	PSM、RPG、NRPG(无自定义生成)
	Dashlane	面向个人/企业用户；1 500万+；高级版(4.99美元/月)、家庭版(7.49美元/月)	PSM、RPG(可定制长度、符号类型)、NRPG(支持自定义生成)
	Keeper Security	面向个人/企业用户；1 000万+；个人版(2.91美元/月)、家庭版(6.24美元/月)	PSM、RPG、NRPG(支持自定义生成)

注：PSM、RPG、NRPG、“支持自定义生成”NRPG可根据用户的个性化设置调整非随机口令生成策略。

2) 通过注册、体验、测试不同 PMS 产品的主功能，获取一手的功能缺陷数据，包括脆弱口令识别准确率、口令强度评估精度以及生成口令的易记性等实测结果。

3) 将所有收集到的数据经过严格清洗与整合，制作 PMS 缺陷汇总表，为后续的 PMS 脆弱性分析奠定了数据基础，以便清晰呈现不同 PMS 系统的性能差异。

2.2　PMS 的对比分析

本研究共收集了 8 个 PMS 信息，其中国内常用 PMS、国际常用 PMS 各 4 个，关于它们的详细信息如表 1 所示。分析表 1 中的数据后，可以清晰发现，国内外 PMS 的发展存在显著差异。

国内市场分析：国内 PMS 主要集中在企业用户领域，且面向个人用户的产品较为稀缺。特别是，国内许多个人用户依赖于系统内置的口令管理服务 (如腾讯密码管家的口令管理功能、浏览器的口令管理插件等)。这些内置化的 PMS 虽然方便，但通常存在数据泄露的隐患，导致许多中国用户对其信任度较低。出于对隐私保护的担忧以及对国外 PMS 付费服务的顾虑，国内用户常选择较为传统的、不安全的口令管理方式 (如本地存储等)。然而，这种做法不仅安全性差，而且缺乏便捷性，极大限制了口令安全研究和技术的进一步发展[6]。

国际市场分析：相对而言，国际 PMS 如 1Password、LastPass、Dashlane、Keeper Security 等，在欧美等地得到了广泛应用。这些产品不仅具备良好的用户基础，且逐渐采用合理的付费模式，深受国际用户信任。数据显示，国外互联网用户对 PMS 的需求明显高于国内用户，这一现象也间接反映出国际市场在口令安全方面的关注度较高。

从表 1 的数据来看，国内外 PMS 产品普遍集成了 PSM、RPG 和 NRPG 等基本功能。然而，尽管这些功能普遍存在，实测结果显示，国内外 PMS 的 PSM 和 NRPG 都普遍存在一些关键性问题。具体而言，这些问题包括脆弱口令识别能力差、口令强度反馈不准确、生成的非随机口令难以记忆、支持的自定义生成功能受限，以及生成策略灵活性不足等。(详见表 2)

表 2　PMS 功能属性缺陷汇总

Table 2　PMS function attribute defects summary

地区	PMS名称	PSM缺陷	NRPG缺陷
国内PMS	腾讯密码管家	无法准确反馈脆弱口令强度、反馈的强度提升建议不具针对性	生成的非随机口安全性不足且难以记忆、自定义生成功能受限、灵活性差
	芯盾时代口令管家	无法准确反馈脆弱口令强度、强度提升建议不具针对性	生成的非随机口令安全性不足且难以记忆、自定义生成功能受限、灵活性差
	安恒信息密码保险箱	反馈的强度提升建议不具针对性	生成的非随机口令安全性不足且难以记忆
	夏冰加密软件	无法准确反馈脆弱口令强度、强度提升建议不具针对性	生成的非随机口令安全性不足且难以记忆
国际PMS	1Password	强度提升建议不具针对性	生成的非随机口安全性不足、自定义生成功能受限、灵活性差
	LastPass	强度提升建议不具针对性	长度区间不足、生成的非随机口安全性不足且难以记忆
	Dashlane	强度提升建议不具针对性	生成的非随机口安全性不足、自定义生成功能受限、灵活性差
	Keeper Security	强度提升建议不具针对性	生成的非随机口安全性不足、自定义生成功能受限、灵活性差

3　本文方法

为有效解决 PMS 共性问题并增强用户群体对系统安全性和可靠性的信任, 本研究构建专门面向口令安全研究的 LLM, 并基于该模型提出可行的改进方案, 以弥补现有 PMS 的不足之处, 如优化生成策略与强度评估算法等。同时, 本研究设计了一个验证框架, 用以评估所提改进方案的有效性与实际表现。研究的总体框架如图 1 所示。

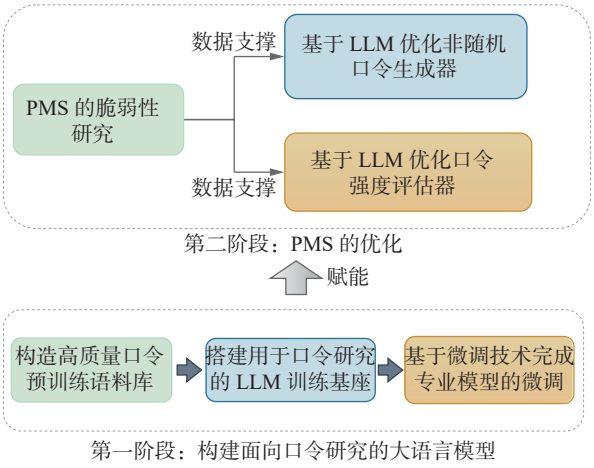

图 1　基于 LLM 优化 PMS 的总体流程

Fig. 1　Overall flowchart of optimizing PMS based on LLM

主要包括以下两个核心阶段:

1) 口令安全研究专用 LLM 的构建。为了提升口令分析的准确性, 本研究开发了一个安全导向型的高质量 LLM。该模型通过整合相关的预训练口令安全语料库, 并采用无监督自回归技术, 增强了模型的泛化能力, 从而能够有效捕获口令安全领域的相关知识。本阶段的详细实施过程将在第 4 章中进一步阐述。

2) PMS 的优化。在 PMS 优化阶段, 本研究主要聚焦于改进非随机口令生成与强度评估功能。通过对 LLM 的微调, 本研究设计了个性化的非随机口令生成策略, 确保生成的口令不仅符合用户偏好, 而且能够维持较高的安全强度。此外, 为了进一步提升强度评估模块的效果, 本研究结合 LLM 进行口令特征分析, 重构了评估算法, 以提高口令强度判定的准确性与可靠性。这些改进旨在显著增强 PMS 生成口令的安全性与可用性, 从而推动 PMS 的广泛应用, 并为口令安全技术的持续演进提供有力支撑。优化阶段的具体工作流程如图 2 所示, 相关的详细细节将在第 5 章和第 6 章中分别展开论述。

通过将上述方法紧密融合, 本研究所构建的框架实现了对 PMS 的全面优化, 充分展现了 LLM 在口令安全领域的广泛应用潜力。

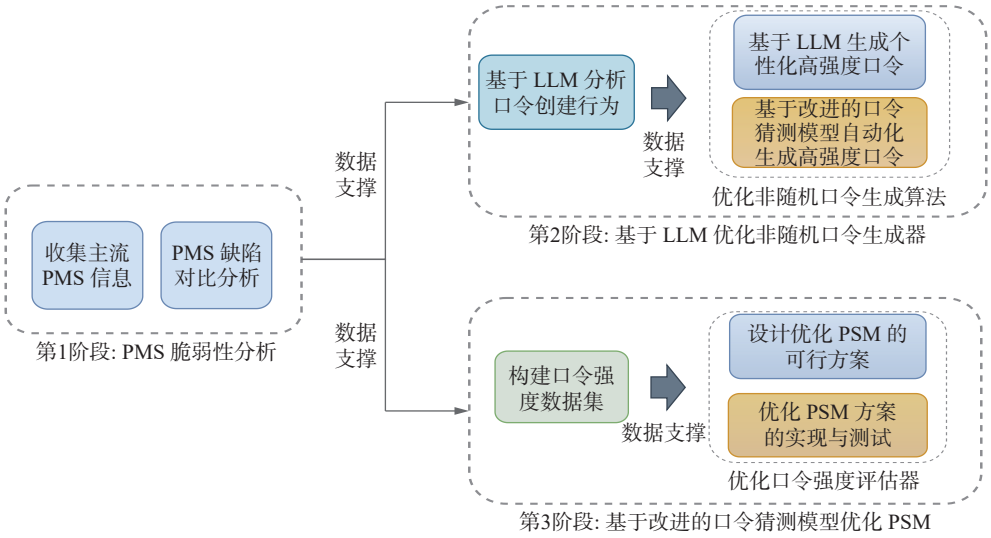

图 2　优化 PMS 的具体工作流程

Fig. 2　Specific workflow of optimizing PMS

4　口令安全研究专用 LLM 的构建

随着口令安全威胁的不断演变, 口令猜测和生成面临越来越复杂的挑战。传统的口令分析方法受限于静态规则和浅显模型, 无法有效处理动态变化的口令创建模式和复杂的攻击策略。因此, 针对口令安全研究的专用 LLM 的构建, 要求深度结合口令结构的独特性和生成规律, 同时结合高效的优化技术以应对大规模数据和多样化攻击。

4.1　面向口令安全研究的专用 LLM

通用型 LLM 通常缺乏针对口令生成和分析

的特异性[17]。在此背景下，本研究构建面向口令安全研究的 LLM 框架，如图 3 所示，该框架主要通过以下几个技术创新点来解决这一问题。

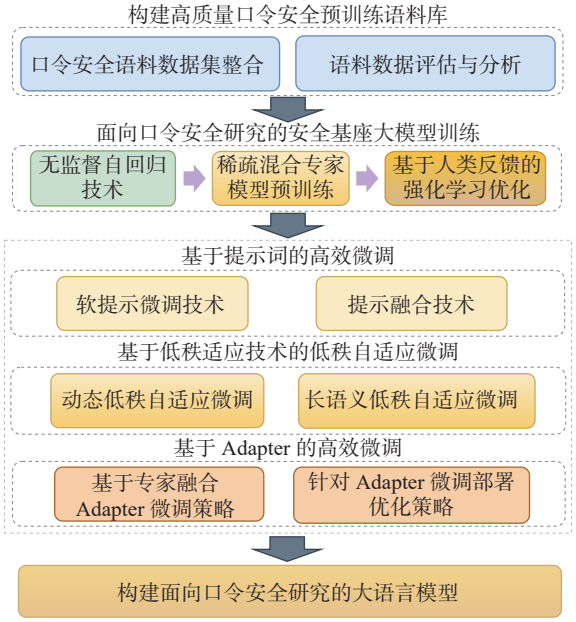

图 3　面向口令安全研究的大语言模型的整体框架

Fig. 3　Framework of LLM for command security research

4.1.1　定制化语料库的构建与处理

口令的创建具有独特的结构特征，通常在语法和语义上都遵循某些非随机的规律。例如，日期类数字串是最常见的数字串类型，特殊字符主要出现在字符串的尾部，“a”“e”“o”等在自然语言学上常见的元音在口令串中依然常见。

为了有效捕捉这些规律，我们构建了一个面向口令安全的定制化语料库，专门包括社会工程数据、常见口令组合、攻击模式 (如字典攻击、暴力破解) 及其动态演化。语料库构建过程包括多源数据收集、去重清洗、字符规范化及异常样本剔除，并结合社会工程特征与攻击模式生成样本，以保证数据的多样性与真实性[29]。

通过对口令生成过程的详细建模，定义口令生成的概率分布：

$$P(x)=\prod_{i=1}^{n}p(x_i|x_{<i}) \tag{1}$$

式中：x 表示一个口令序列，$x<i$ 表示当前口令前 i 个字符的上下文。式 (1) 基于自回归语言建模方法，通过对每个字符的条件概率建模，可以捕捉口令生成过程中的潜在规律和变化趋势。

4.1.2　自回归模型与上下文信息的深度融合

为了更好地捕捉口令生成过程中的结构性和变化性，本研究采用自回归模型 (autoregressive model)，并对模型进行了优化，使其在训练时能够高效学习口令的字符结构[30]。具体而言，我们在训练过程中通过最小化损失函数来优化模型参数：

$$L(\theta)=-\sum_{i=1}^{n}\log p(x_i|x_{<i};\theta) \tag{2}$$

式中：θ 为模型的参数，$P(x_i \mid x_{<i};\theta)$ 为在给定上下文条件下生成字符 x_i 的条件概率。该训练目标能使模型在每次生成字符时，考虑到先前字符的上下文信息，从而更好模拟实际口令生成的复杂性。

4.1.3　稀疏混合专家模型的引入

针对口令分析中存在的计算瓶颈问题，特别是在大规模数据处理与模型扩展方面，提出稀疏混合专家 (sparse mixture of experts, SMoE) 模型的应用[31]。SMoE 模型通过选择性地激活部分专家网络来提升计算效率，并在处理高维度口令数据时保持良好的表现。SMoE 的核心思想是将整个模型划分为多个子模型 (专家)，每个子模型在处理某些特定口令模式时表现最佳。在推理过程中，仅激活与当前输入最相关的部分专家，从而显著减少计算量。

在口令安全研究中，SMoE 模型能够在不同的口令结构 (如数字、字母、符号混合的口令) 下表现出较强的适应性。具体而言，SMoE 的训练目标可以描述为

$$\hat{y}=\sum_{k=1}^{k}\alpha_k(x)f_k(x) \tag{3}$$

式中：$\hat{y}$ 为模型的预测输出，k 为专家的数量，$\alpha_k(x)$ 为输入 x 与第 k 个专家相关的权重系数，$f_k(x)$ 为第 k 个专家的输出函数。通过稀疏激活机制，SMoE 模型能够动态选择适应性强的专家进行计算，从而在口令分析任务中实现高效处理。

4.2　口令安全研究专用 LLM 的微调技术

4.2.1　提示词微调

提示词微调是一种通过设计特定的提示词，使得 LLM 能够生成更符合口令生成任务要求的输出的方法。在这一过程中，我们通过优化提示词的权重，以提高模型在口令生成和安全评估任务中的表现[32]。具体而言，模型的输入可以表示为

$$z=\text{Prompt}+x \tag{4}$$

式中：x 为模型输入的口令数据，Prompt 为特定设计的提示词。通过训练模型来优化 Prompt 和 x 的结合方式，模型可以更好地生成符合实际需求的口令。

4.2.2　低秩适应

为了提升微调的计算效率，本研究引入了低秩适应 (low-rank adaptation, LoRA) 技术。低秩适

应通过在模型内部引入低秩矩阵来减少模型参数的更新量, 从而在不牺牲性能的前提下大幅度减少计算资源的消耗[33]。具体的数学表述为

$$\Delta W = \boldsymbol{A}\boldsymbol{B} \tag{5}$$

式中: ΔW 为模型参数的更新; $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别为低秩矩阵, 通常 $\boldsymbol{A} \in \mathbf{R}^{d\times r}$, $\boldsymbol{B} \in \mathbf{R}^{r\times d}$, 其中 r 是低秩矩阵的秩。通过这种方式, LoRA 能够在减少参数更新的同时保留良好的泛化能力和适应性。

4.2.3　适配器微调

适配器微调技术通过在模型中插入轻量级的适配器模块来快速调整模型的任务特定行为[34]。这种方法不需要修改预训练的主模型架构, 从而减少了微调的复杂性。适配器模块的数学表示可以描述为

$$h_{\text{out}} = h + \boldsymbol{W}_{\alpha} \cdot \text{Activation}(h) \tag{6}$$

式中: h 为输入特征, $\boldsymbol{W}_{\alpha}$ 为适配器的权重矩阵, Activation(h)为激活函数。通过这一方法, 适配器能够灵活地适应不同口令生成和评估任务, 并在面对新型安全威胁时迅速做出调整。值得注意的是, 尽管适配器微调主要关注参数效率和任务适应性, 但在实际部署中, 为了进一步提升模型输出的质量与对齐人类偏好, 常结合基于人类反馈的强化学习 (RLHF) 进行优化[35]。RLHF 通过人类标注的偏好数据训练奖励模型, 并利用该模型指导策略优化, 使生成内容更符合安全、有用和无害的标准。

4.3　专用 LLM 赋能口令猜测

基于微调后的专用 LLM 基础上, 本研究构建了用于口令猜测的多维口令特征库。口令猜测模型可结合专用 LLM 深入分析目标口令的创建规则、提取目标口令所包含的社会工程学信息, 再利用构建的多维度口令特征库来提升口令猜测模型的破解准确率, 特别是在破解复杂口令模式和多变口令时, 专用 LLM 提供了强大的赋能支持。

4.3.1　专用 LLM 在口令猜测模型中的应用

构建的专用 LLM 能够从大规模的口令数据集中提取多维口令创建特征, 并将这些特征以加载特征库的形式映射至现有的口令猜测模型中, 如 PCFG(概率上下文无关文法)[36] 与 PassGAN[13] 等模型。这一集成不仅增强了模型对脆弱口令的识别能力, 还能根据用户兴趣与常用术语对社会工程信息进行建模, 提高了口令猜测的精准度。

1) 口令猜测模型特征解析增强。LLM 框架提取的多维特征经过标准化后, 输入至 PCFG 与 PassGAN 模型进行训练和优化。提取的特征包括口令的字符分布、生成规律、用户行为特征等, 这些信息能够显著提升模型对复杂口令模式的预测能力, 尤其是在应对具有特殊字符和多变模式的口令时。

2) LLM 微调增强口令猜测模型。将口令猜测任务视为文本生成问题, 我们采用 GPT 等先进的语言模型生成候选口令, 并通过交叉验证微调模型的参数。该方法能够实时适应口令生成的趋势, 并有效提升模型在长期使用过程中的准确性和稳定性。

4.3.2　性能优化与实验验证

本节将上述优化策略应用于 PCFGv4 和 PassGAN 模型, 重点解决现有模型在破解含特殊字符口令方面的不足。为验证优化效果, 选择两个数据集进行实验: 7k7k 数据集 (2011 年泄露的中国某游戏平台数据集, 包含 19 121 128 条口令, 其中 295422 条含有特殊字符) 与 MySpace 数据集 (2016 年泄露的国际化社交平台数据集, 包含 13 128 817 条口令, 其中 11.44% 的口令含有特殊字符)。实验数据集按 4:1 的比例划分为训练集和测试集。

如图 4 所示, 优化后的 PCFGv4 与 PassGAN 模型在这两个数据集上的表现有所提升, 且当口令猜测规模达到 4×10^7 时, 性能提升效果尤为明显。这表明, 优化后的模型在大规模口令猜测场景中表现出显著的性能改进。

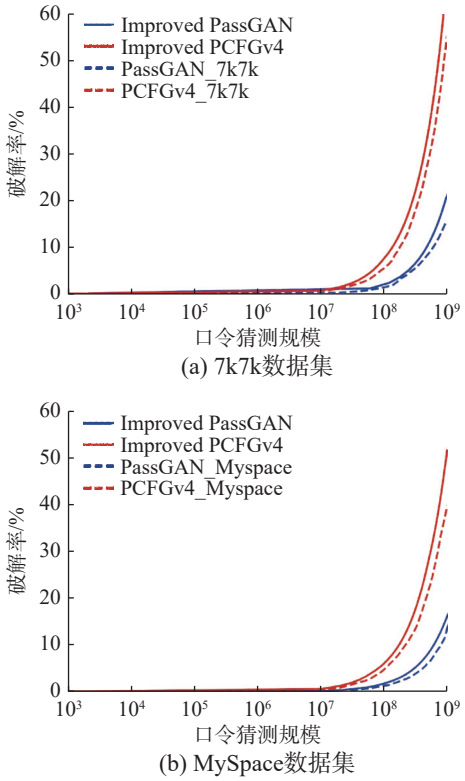

图 4　可靠性验证实验

Fig. 4　Reliability verification experiment

然而，在小规模猜测的情况下，性能提升相对有限，仍有进一步优化的空间。这也反映出优化策略在不同规模攻击场景下的效果存在差异。因此，后续研究将进一步探索针对小规模在线口令猜测场景的优化方法，以提高模型在实际应用中的适应性和效率。

5　基于专用 LLM 优化 NRPG

尽管现有 NRPG 在平衡口令安全性与可用性方面已有显著进展，但在生成高强度非随机口令时仍面临技术瓶颈[7]。尤其在面对用户个性化需求时，现有 NRPG 无法兼顾生成口令的安全性与便捷记忆性，导致用户对 PMS 的信任度下降。本文提出一种基于专用 LLM 优化 NRPG 的方法，通过 3 点核心特性克服上述挑战：1) 基于专用 LLM 实现深层次的口令创建行为感知；2) 基于专用 LLM 实现灵活动态调整口令生成策略、定制化口令生成；3) 基于 LLM 及改进的口令猜测模型实现高强度口令自动生成。优化 NRPG 的整体流程如图 5 所示。

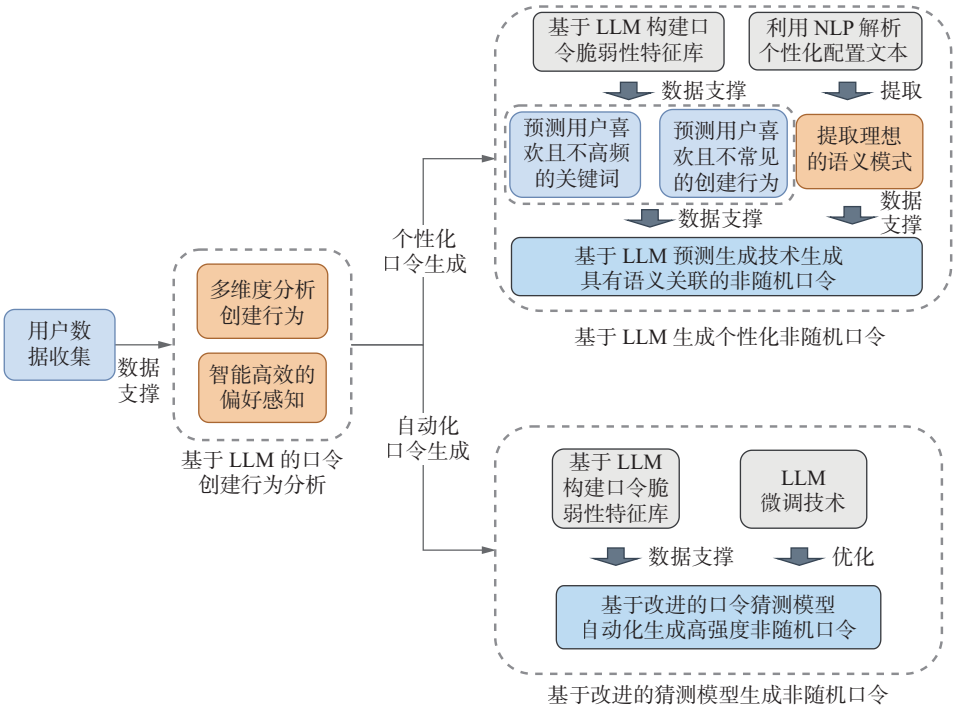

图 5　优化 NRPG 的整体流程

Fig. 5　Overall process of optimizing NRPG

5.1　基于 LLM 的用户行为分析

本研究利用专用 LLM 对用户口令创建行为进行深入分析，通过对历史数据的建模与学习，为个性化或自动化口令生成提供关键数据支撑。具体的行为分析过程如下：

1) 多维度行为分析。通过使用专用 LLM，分析用户历史口令数据，提取多维度口令创建特征，包括用户输入的习惯、口令的频率分布及常见的口令模式。为此，采用了基于自注意力机制的特征抽取方法 (Transformer 架构[37]) 对这些数据进行建模，并对口令的生成规律进行量化：

$$P(x_t|x_{t-1},x_{t-2},\cdots)=\mathrm{Softmax}(\boldsymbol{W}_h\cdot\mathrm{Attention}(x_{t-1})) \quad (7)$$

式中：x_t 表示时间步 t 的输入特征，$\boldsymbol{W}_h$ 是模型的权重矩阵，Attention() 是自注意力机制。

2) 偏好预测。利用 LLM 进行用户的语言偏好预测，分析其常用短语、个人兴趣及安全偏好，从而定制化生成具有高安全性且符合用户偏好的口令。

为了增强预测的准确性，我们引入了基于用户上下文信息的语言模型微调方法，通过最大化用户行为数据的对数似然估计进行微调：

$$L(\theta)=\sum_{i=1}^{N}\log P_\theta(y_i|x_i) \quad (8)$$

式中：θ 表示模型参数，y_i 是第 i 个用户行为数据的输出，x_i 是输入的特征信息。

5.2　基于 LLM 的定制化口令生成

在定制化口令生成阶段，发挥 LLM 在自然语言处理和生成方面的优势，生成满足用户偏好与安全要求的非随机口令：

1) 关键词预测。使用 LLM 模型分析用户历

史行为, 特别是低频但用户偏好的口令元素。通过挖掘这些潜在的关键词, 结合语义模型 (双向变换器模型), 生成适合用户个性化需求的口令。

2) 语义分析。通过使用自然语言处理技术, 解析用户输入的文本或关键提示, 提取目标语义模式, 并将这些语义模式映射到口令生成空间中。例如, 使用命名实体识别技术来从文本中提取重要的实体 (如地名、日期等), 并将其嵌入到口令中, 保证口令的高语义相关性。

3) 安全口令生成。结合 LLM 模型的预测能力, 生成强度高且语义连贯的非随机口令。通过对生成的口令进行随机性调整和复杂度增强 (如混合字母大小写、添加符号等), 确保口令具有高强度安全性。

5.3　利用改进猜测模型生成高强度口令

为了提高生成口令的强度, 本文对猜测模型进行改进, 并与 LLM 微调结合使用, 以实现以下优化目标:

1) 关键特征提取。基于 LLM 提取影响口令强度的关键特征, 例如字符分布[38]、口令长度[39]、字典与非字典词汇比例等[40]。这些特征被用于对口令的复杂性进行建模。

2) 性能优化。引入改进后的门控循环单元 (gated recurrent unit, GRU) 模型, 通过调整隐藏层参数、优化梯度更新方法来提升口令生成模型的预测效率[41]。具体来说, 采用长短时记忆网络和 GRU 的结合, 能够提高模型对时序信息的捕捉能力:

$$h_t = \mathrm{GRU}(h_{t-1}, x_t) \tag{9}$$

式中: h_t 表示时刻 t 的隐藏状态, x_t 是输入特征。

3) 模块优化。对 LLM 进行微调, 调整生成策略, 提升口令生成的准确性和多样性。使用基于 Adam 优化器的自适应学习率调整, 以确保生成策略更符合实际需求[42]。

4) 候选池筛选。在生成的口令集合中, 根据预设的安全标准 (如抗破解次数、复杂性评分) 筛选出最具安全性的口令, 作为最终用户部署的候选项。此过程利用了多目标优化算法 (Pareto 优化) 来平衡安全性与可用性[43]。

上述优化流程能够确保 NRPG 生成的口令既能满足高强度要求, 又能适配用户个性化偏好, 很好解决了当前 PMS 集成的 NRPG 无法提供高强度易记忆非随机口令的难题, 从而提升用户对 PMS 的使用体验。图 6 给出了改进后的 NRPG 的工作流程。

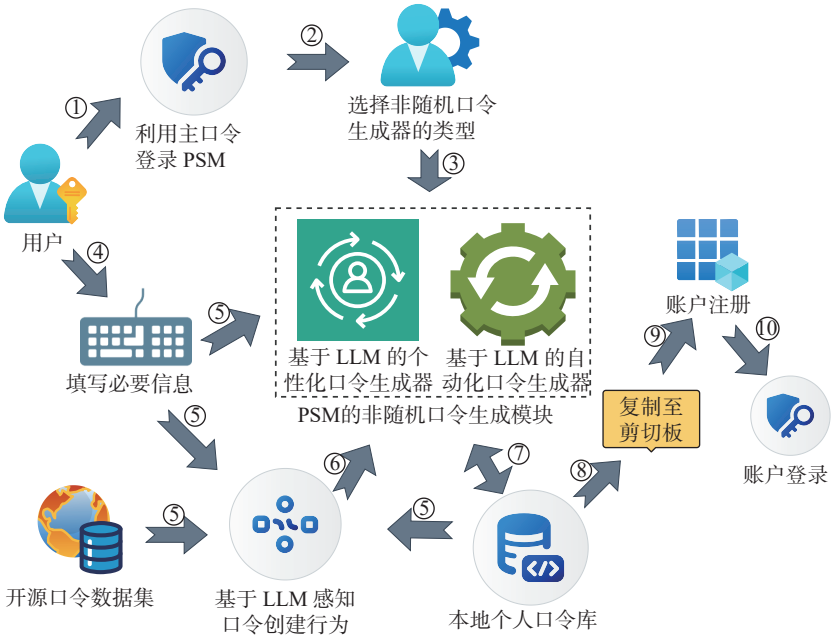

图 6　改进后的 NRPG 工作流程

Fig. 6　Workflow of improved NRPG

5.4　可用性与有效性验证

为全面评估优化模块的性能, 本文通过多维度验证测试, 确保口令生成模块的实际应用效果。

1) 记忆性分析。通过问卷收集了 100 名用户对生成口令的记忆性情况, 再对用户记忆生成口令的程度进行量化:

$$S = 1 - \frac{N_{\mathrm{N}}}{N} \tag{10}$$

式中: N_{N} 是错误记忆数, N 是总记忆数。基于

S 将生成口令的易记忆性等级划分为高 ($0.66 \leqslant S \leqslant 1$)、中 ($0.33 \leqslant S < 0.6$)、低 ($0 \leqslant S < 0.33$)3 个等级，以此量化生成口令可记忆性的评价指标。通过对比个性化与自动化生成口令的效果，验证了个性化生成的口令具有更高的记忆性。

2) 安全性评估。采用 TarGuess-I、PCFG 和 PassGAN 等模型对生成口令的抗破解能力进行评估。通过对口令进行多次攻击模拟，评估口令抵御高强度攻击的能力。安全性评分系统 (如 Zxcvbn[25]) 进一步对口令的强度进行等级划分。

通过对比个性化与自动化生成组的数据样本 (表 3)，发现优化后的模块显著提升了口令的安全性与便捷记忆性。具体表现在：生成的口令能够抵抗高达 10^9 次的离线攻击，并且个性化生成的口令在用户记忆性方面评分明显高于自动化生成的口令。研究结果表明，优化的 NRPG 有效克服了现有 PMS 集成的 NRPG 的不足，在保证生成的非随机口令强度的同时显著提升了生成口令的便捷记忆性，实现了预期的性能改进目标。

表 3　优化的 NRPG 模型性能测试结果

Table 3　Performance test results of optimized NRPG module

	示例场景	Top-5生成的口令	易记忆性等级	强度等级			强度分数
				TarGuess-I	PCFG	PassGAN	
示例场景1	生成类型：个性化生成非随机口令；用户输入的内容：Zhang San和Li Si的结婚周年纪念日是8月13日，口令长度不超过16，特殊字符的数量不超过4。	[Zhang&Li]Wa0813	高	高	高	高	4
		Wa0813(Zhang&Li)	中	高	高	高	4
		Zhang&Li(Wa813)	高	中	高	高	4
		San&Si(Wa0813)	高	高	高	高	4
		[San&Si]Wa0813	中	高	高	高	4
示例场景2	生成类型：自动化生成非随机口令；用户基本信息{姓名：Wu Xiaoming；生日：20001205；邮件：wuxiaoming@163.com；号码：12365504321。}	100%Wxm 1205	中	中	高	高	4
		Wxm1205@163.coM	中	中	高	高	4
		Ming 2000.1205	高	中	高	中	4
		WuXM@1205.coM	中	高	高	高	4
		Xiaoming1205_Wu	中	中	高	高	4

注：强度分数指经优化后Zxcvbn评估的口令强度分数，4分对应强度最高等级。

6　优化 PSM

现有 PMS 的漏洞分析表明，当前缺乏面向汉语用户的有效 PSM。现有的基于数学模型或机器学习理论的 PSM 主要面向国际用户 (母语为英语用户为主) 所设计，难以直接适用于汉语用户场景。为弥补这一空白，本研究提出一种基于改进的口令猜测模型和优化 Zxcvbn 算法的综合评估方法，旨在通过两者的结合提升口令强度评估的准确性。

6.1　新型 PSM 框架

为更准确评估口令强度，本研究提出了一种综合的口令强度评估框架，结合了改进的口令猜测模型与优化 Zxcvbn 算法。两者相辅相成，共同解决了传统 PSM 的口令强度评估中的准确性问题。

1) 改进的口令猜测模型。该模型基于专用 LLM 的口令特征感知与分析能力，分析了大规模口令数据集中口令生成规律与常见结构，能够更精确地估算破解口令所需的猜测次数，尤其对复杂口令的评估具有重要价值。它为破解复杂口令提供了一个动态的、数据驱动的评估工具。

2) 优化 Zxcvbn 算法。Zxcvbn 是一个广泛应用于口令强度评估的工具，但其在汉语口令和复杂口令的评估上存在一定的局限性。本研究在 Zxcvbn 算法的基础上，进一步增强了对汉语口令的支持，改进了字典匹配的能力，并提升了复杂度分析的精确性。

两者的结合能够充分利用各自的优势，确保口令强度评估的全面性与精准性。改进后的整体评估框架如图 7 所示。

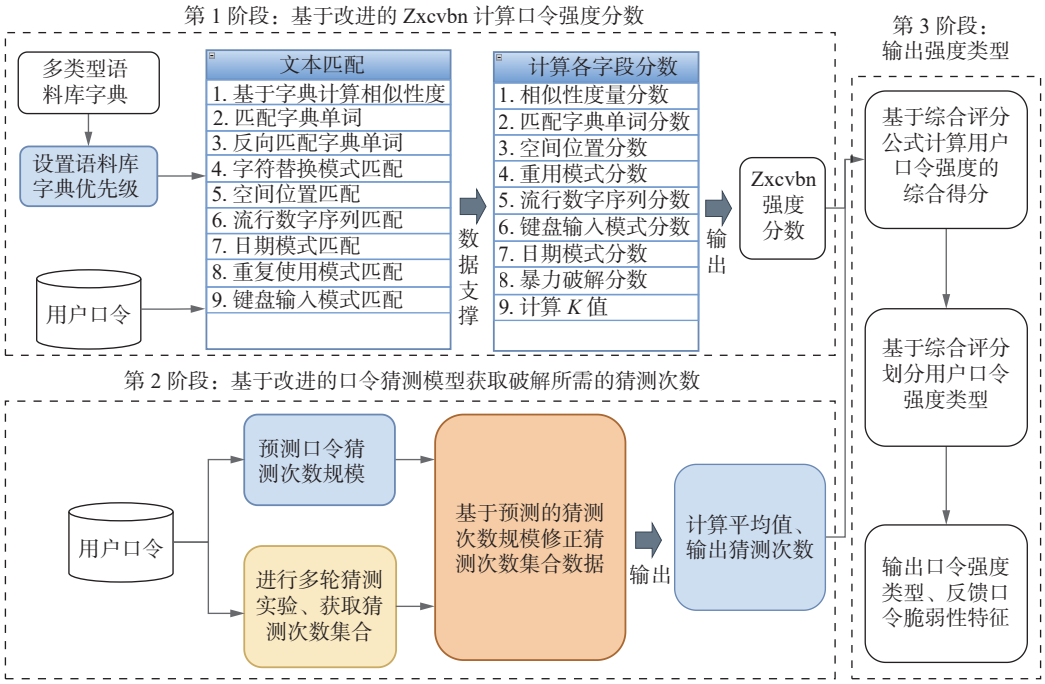

图 7　改进后的 PSM 整体评估框架

Fig. 7　Overall evaluation framework of improved PSM

6.2　口令强度数据集构建

为了优化口令强度评估方法, 首先需要构建一个适应汉语用户口令的高质量口令强度数据集。数据集构建方法如图 8 所示。

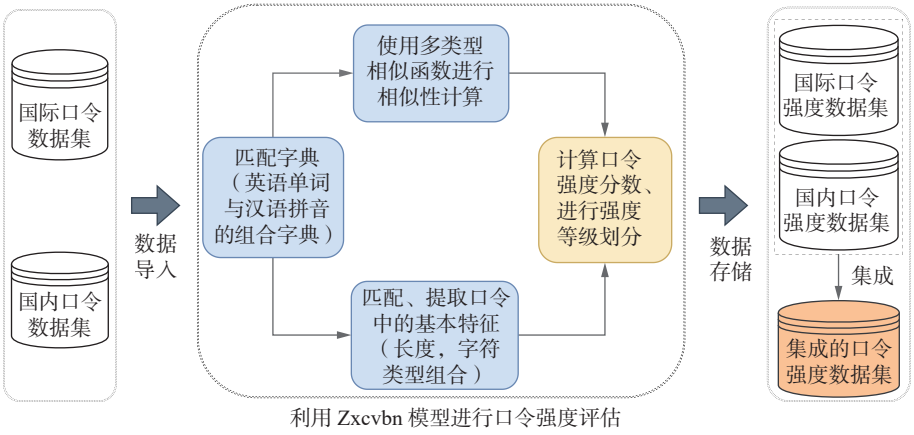

图 8　口令强度数据集构建方法

Fig. 8　Password Strength Dataset Construction Method

1) 数据集清理与筛选。为确保数据质量, 从国际和国内两类数据集中筛除掉不包含英文或汉语拼音的口令, 避免噪声数据对训练过程造成干扰[38]。所有口令数据均通过人工审查与数据预处理, 去除拼写错误、无效字符或极其简化的组合。

2) 数据集平衡。为避免类别不平衡问题, 在国内和国际口令数据集中随机抽取等量的样本, 确保数据集的代表性, 且符合多样性需求。

3) Zxcvbn 标记。首先利用 Zxcvbn 模型对国际数据集进行初始的强度标记, 并针对国内数据集, 利用融合了汉语词典的 Zxcvbn 模型进行标注。这里需要特别指出, Zxcvbn 模型的强度标签通过机器学习算法生成, 原始标签 (弱/中/强) 提供了初步的强度评估。

4) 数据集融合。结合来自国内外的口令强度数据集, 确保训练数据的多样性与广泛性。最终, 合并后的数据集作为模型训练和评估的基础。

6.3　优化方法

6.3.1　优化原理

优化方法的核心原理是通过引入先进的口令猜测模型和相似度计算方法, 对 Zxcvbn 算法进行改进, 并基于 GRU 训练的口令猜测模型进一步提升口令强度的评估能力。优化过程为:

1) 基于 Zxcvbn 的优化。改进 Zxcvbn 算法,

通过引入中文词典、拼音词典和多种相似度计算方法 (如 Levenshtein、Fuzz、Jaro-Winkler 等[44]), 提升模型在中文口令上的评估能力。Zxcvbn 模型的基础功能是根据字典匹配和字符组合复杂度来评估口令的强度。

2) 基于口令猜测模型的优化。除了改进 Zxcvbn 模型外, 还引入了基于改进的口令猜测模型的技术, 通过对口令序列的学习和预测, 提高口令强度的识别能力。口令猜测模型可以通过分析用户行为数据、历史口令创建模式等信息, 捕捉到复杂口令结构中的潜在弱点。

3) 多层次评估。将 Zxcvbn 算法和口令猜测模型结合, 建立一个多层次的强度评估方法。该方法不仅能够通过字典匹配进行初步评估, 还能通过口令猜测模型捕捉到更为复杂的口令模式, 从而提供更加精确的强度评估。

6.3.2　优化 Zxcvbn 算法

优化流程的关键在于本优化方法引入了更为细致的强度标记机制, 将口令强度细化为 5 个等级 (0~4 分)。原始 Zxcvbn 模型仅使用简单的弱、中、强标签, 这种粗糙的标记方法无法处理复杂的中文口令。具体优化方法流程如图 9 所示。

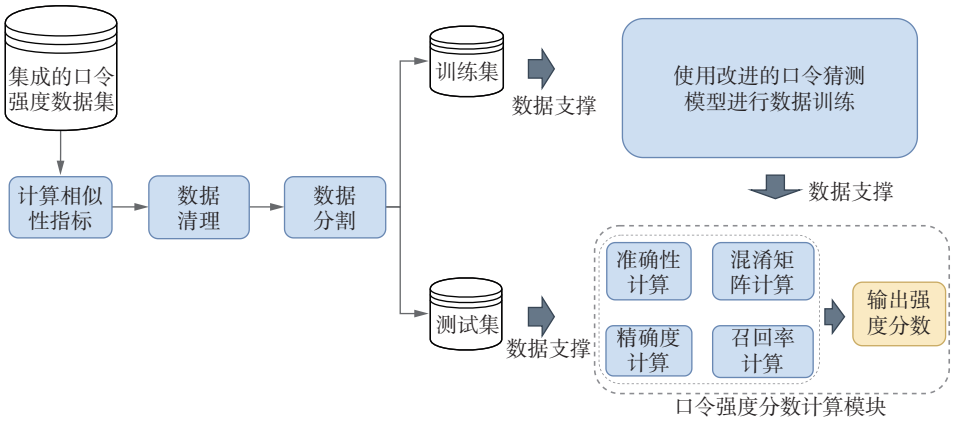

图 9　优化 Zxcvbn 流程

Fig. 9　Optimized Zxcvbn flowchart

1) 词汇匹配与相似度计算。在相似度计算中, 本方法使用了多种距离度量方法, 包括 Fuzz、Levenshtein、Jaro 和 Jaro-Winkler 算法[44]。每种方法依据其不同的匹配策略, 从字面到语义, 分别计算口令与字典条目之间的相似度。该方法特别重视汉语词汇的语义相似性, 而不仅仅是字符匹配。

2)GRU 模型的引入。为适应汉语口令的特点, GRU 模型被引入进行序列建模, 处理口令中的上下文信息。与传统的循环神经网络相比, GRU 在处理长序列时能够避免梯度消失问题, 因此能有效识别复杂的口令组合。

3) 多层次评分机制。针对汉语口令的特殊性, 在 Zxcvbn 的基础上, 设计了多层次评分机制, 细化口令强度的评估。口令的评分从 0~4 进行细分, 0 代表极易猜测, 4 代表非常安全。同时考虑口令的形态复杂度 (字符的多样性)、字典匹配度和口令生成的随机性等多个因素。

4) K 值与破解时间估算。本文模型进一步将 K 值 (破解所需猜测次数) 作为关键指标。K 值由多个因素共同决定, 包括口令的词汇形态多样性 (R)、与字典的匹配片段长度 (L) 以及相似度得分 (S)。K 值的计算公式为

$$K = R \times L \times S \tag{11}$$

式中: R 表示口令与字典的词汇差异性, L 是匹配片段的长度, S 是相似度分数。通过公式 (11), 系统能够估算口令的破解时间并给出强度分级。

5) 结果评估。最终, 模型评估结果基于估算的 K 值, 按照原始 Zxcvbn 模型的安全标准进行评分, 并提供比原始模型更精细的安全评估分数。

6.3.3　利用口令猜测模型计算猜测次数

首先, 利用改进的口令猜测模型快捷预测成功破解目标口令所需要的猜测次数规模范围; 然后, 经过多轮猜测实验, 进一步精确定位目标口令对应的“破解所需的猜测次数”参数值的区间范围; 紧接着, 基于预测的猜测次数规模大小修正目标口令对应的“破解所需的猜测次数”参数值集合; 最后, 计算“破解所需的猜测次数”参数值集合的平均值作为破解目标口令所需的猜测次数。

6.3.4　综合评分计算方法

核心思想是结合 Zxcvbn 算法的初步评分与口令猜测模型的猜测次数, 形成综合评分。破解所需的猜测次数作为关键指标, 结合 Zxcvbn 的字典匹配得分, 给出口令的最终强度评分。综合评分体系不仅考虑口令的表面复杂度, 还考虑了潜

在的猜测路径与破解难度。

6.4 可行性分析

6.4.1 实践可行性

为验证改进算法的有效性, 本研究基于 He 等[39]的测试用例构建对照实验。对比原始 Zxcvbn 与改进的 PSM 在同一数据集上的表现。通过引入更细化的评分机制和增强的字典匹配, 改进的 PSM 在多个案例中表现出更精确的强度评估。强度评估准确性变化如表 4 所示。从表 4 可以看出, 口令“Ch@rles@bcd123”在 Zxcvbn 中被评为“中”, 而在改进后的 PSM 中被评为“弱”, 更准确地反映了其真实的弱口令强度。此外, 改进后的 PSM 能够避免原模型对复杂口令的高估, 提供了更为精准的弱口令识别能力。例如, 口令“(Char!es)@123.coM”在 Zxcvbn 中被评为“强”, 而在改进后的 PSM 中被评为“中”。

表 4　改进后 PSM 的强度评估准确性变化

Table 4　Change in strength evaluation accuracy of improved PSM

脆弱口令示例	PSM类型	
	Zxcvbn	改进的PSM
Charlesabcd123	中	弱
Ch@rles@bcd123	强	弱
Char!es123!@#$	强	弱
Char!es@123.com	强	弱
(Char!es)@123.coM	强	中

为定量评估准确性提升, 引入强度评估准确性提升率, 并通过多个脆弱口令测试集 (数据规模 10^4 以上) 进行计算。强度评估准确性提升率指标 (RImprove) 指改进的 PSM 评估的正确性相对于 Zxcvbn 评估准确性的提高程度。计算公式为

$$R_{\text{Improve}} = \frac{A_{\text{Improved PSM}} - A_{\text{Zxcvbn}}}{A_{\text{Zxcvbn}}} * 100\% \qquad (12)$$

式中: $A_{\text{Improved PSM}}$ 指改进的 PSM 的口令强度评估准确性, A_{Zxcvbn} 指 Zxcvbn 的口令强度评估准确性。

基于收集的改进的 PSM 和 Zxcvbn 对多组测试集的强度评估结果, 利用式 (12) 可以计算出改进的 PSM 在多个测试集上 R_{Improve} 的取值范围为 62.5%~85.1%。这一结果表明, 在处理具有复杂字符和语义结构的汉语用户口令时, 改进的 PSM 的评估准确性提升最为显著。

6.4.2 理论可行性

理论上, 改进的口令猜测模型通过深度学习方法能够有效预测口令破解所需的猜测次数, 具有良好的适应性与灵活性。而优化后的 Zxcvbn 算法, 则通过增强字典与复杂度分析, 提高了口令评估的全面性。两者结合, 能够为口令强度评估提供更可靠的理论支持。

综上所述, 本章提出的基于改进的口令猜测模型与优化 Zxcvbn 算法的优化 PSM 方案, 有效解决了传统 PSM 在口令强度评估中的不足。通过综合利用两者的优势, 形成了一个更加精准、全面的口令强度评估框架。

7 结束语

本文通过针对 PMS 可用性问题, 提出了一种基于 LLM 的优化方案, 解决了 PSM 和 NRPG 中的关键技术瓶颈。实验结果表明, 基于专用 LLM 的 PMS 在多种实际应用场景中有效提升了生成口令的安全性及易记忆性和口令强度评估的准确性, 特别是在提高系统可靠性和用户体验方面具有显著优势。与传统方法相比, 本文的创新之处在于结合 LLM 与现有的口令生成与评估技术, 突破了口令安全性和记忆性之间的矛盾。

尽管取得了良好的结果, 仍存在一些挑战, 如在面对更复杂的攻击方式时, 新型 PMS 的适应性和鲁棒性仍需进一步加强。此外, 本研究的优化方案与先前研究相比, 具有较好的普遍性和应用前景, 但尚未完全解决 PMS 中的安全性问题。

理论上, 本文提出的优化框架为 PMS 的改进提供了新的思路, 并在口令强度评估和生成策略上做出了更为精确的贡献; 实用性上, 本方案能够有效提升 PMS 的可用性, 具有广泛的应用潜力。未来的研究应进一步结合深度学习与安全技术, 增强系统的鲁棒性, 并探索与用户行为分析的结合, 以应对更复杂的安全挑战。

参考文献:

[1] PEARMAN S, ZHANG Shikun, BAUER L, et al. Why people (don’t) use password managers effectively[C]//Fifteenth symposium on usable privacy and security. Santa Clara: [s. n.], 2019: 319−338.

[2] FÁBREGA A, NAMAVARI A, AGARWAL R, et al. Exploiting leakage in password managers via injection attacks [EB/OL]. (2024−08−13)[2025−04−23]. https://arxiv.org/abs/2408.07054.

[3] 娄峰, 韩超. 数据中心用户管理系统的设计与实现[J]. 铁路计算机应用, 2022, 31(4): 65−69.
LOU Feng, HAN Chao. Design and implementation of data center user management system[J]. Railway computer application, 2022, 31(4): 65−69.

[4] 孙亮, 陈小春, 鲍天明, 等. 面向多处理器的内网计算机 BIOS 口令集中管理机制研究[J]. 信息安全与通信保密,

2021, 19(2): 63−74.
SUN Liang, CHEN Xiaochun, BAO Tianming, et al. Research on centralized management mechanism of BIOS password of intranet computer for multiprocessor[J]. Information security and communications privacy, 2021, 19(2): 63−74.

[5] GLORY F Z, UL AFTAB A, TREMBLAY-SAVARD O, et al. Strong password generation based on user inputs[C]// 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference. Piscataway: IEEE, 2019: 416−423.

[6] OESCH S, RUOTI S. That was then, this is now: a security evaluation of password generation, storage, and autofill in thirteen password managers[EB/OL]. (2019−08−09)[2025−04−23]. https://arxiv.org/abs/1908.03296.

[7] GOLLA M, DÜRMUTH M. On the accuracy of password strength meters[C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2018: 1567−1582.

[8] PATEL S, MADISETTI V K. PhishGuard: integrating fine-tuned large language models (LLMs) into password management[J]. Journal of information security, 2024, 15(4): 474−493.

[9] LI Hang, CHEN Mengqi, YAN Shengbo, et al. Password guessing via neural language modeling[M]//Machine Learning for Cyber Security. Cham: Springer International Publishing, 2019: 78−93.

[10] PAL B, DANIEL T, CHATTERJEE R, et al. Beyond credential stuffing: password similarity models using neural networks[C]//2019 IEEE Symposium on Security and Privacy. [S. l.]: IEEE, 2019: 417−434.

[11] PASQUINI D, CIANFRIGLIA M, ATENIESE G, et al. Reducing bias in modeling real-world password strength via deep learning and dynamic dictionaries[C]//30th USENIX Security Symposium. [S. l.]: USENIX, 2020.

[12] XU Ming, YU Jitao, ZHANG Xinyi, et al. Improving real-world password guessing attacks via bi-directional transformers[C]//32nd USENIX Security Symposium. [S. l.]: USENIX, 2023: 1001−1018.

[13] HITAJ B, GASTI P, ATENIESE G, et al. PassGAN: a deep learning approach for password guessing[M]//Applied Cryptography and Network Security. Cham: Springer International Publishing, 2019: 217−237.

[14] RANDO J, PEREZ-CRUZ F, HITAJ B. PassGPT: password modeling and (guided) generation with large language models[C]//Computer Security–ESORICS 2023. Cham: Springer, 2024: 164−183.

[15] GEWIDA M, QU Yanzhen. Enhancing IoT security: predicting password vulnerability and providing dynamic recommendations using machine learning and large language models[J]. European journal of electrical engineering and computer science, 2025, 9(1): 8−16.

[16] KAGNICI O, BISGIN H, ULUDAG S. Evaluating the efficacy of GPT-based password generation on real world data[C]//2025 1st International Conference on Secure IoT, Assured and Trusted Computing (SATC). Piscataway: IEEE, 2025: 1−5.

[17] GE Yingqiang, HUA Wenyue, MEI Kai, et al. OpenAGI: when LLM meets domain experts[J]. Advances in neural information processing systems, 2023, 36: 5539−5568.

[18] AN Shengnan, CHEN Weizhu, LIN Zeqi, et al. Make your LLM fully utilize the context[C]//Advances in Neural Information Processing Systems 37. New Orleans: NeurIPS, 2024: 62160−62188.

[19] XU Haoruo, ZHANG Ning, YIN Zhenyu, et al. GeoLLM: a specialized large language model framework for intelligent geotechnical design[J]. Computers and geotechnics, 2025, 177: 106849.

[20] MUNYENDO C W, MAYER P, AVIV A J. "I just stopped using one and started using the other": Motivations, Techniques, and Challenges When Switching Password Managers[C]//Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2023: 3123−3137.

[21] MURMU S, KASYAP H, TRIPATHY S. PassMon: a technique for password generation and strength estimation[J]. Journal of network and systems management, 2021, 30(1): 13.

[22] HOUSHMAND S, AGGARWAL S. Building better passwords using probabilistic techniques[C]//Proceedings of the 28th Annual Computer Security Applications Conference. New York: ACM, 2012: 109−118.

[23] 栗会峰, 李铁成, 姚启桂, 等. 基于形态学的电力系统弱口令深度学习检测方案[J]. 计算机应用与软件, 2024, 41(10): 379−385.
LI Huifeng, LI Tiecheng, YAO Qigui, et al. Weak password deep learning detection scheme for power system based on morphology[J]. Computer applications and software, 2024, 41(10): 379−385.

[24] BONNEAU J. The science of guessing: analyzing an anonymized corpus of 70 million passwords[C]//2012 IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2012: 538−552.

[25] WHEELER D L. Zxcvbn: Low-budget password strength Estimation[C]//USENIX Security Symposium. [S. l.]: USENIX, 2016: 157−173.

[26] DOUCEK P, PAVLÍČEK L, SEDLÁČEK J, et al. Adaptation of password strength estimators to a non-English environment: the Czech experience[J]. Computers & security, 2020, 95: 101757.

[27] MELICHER W, UR B, KOMANDURI S, et al. Fast, lean, and accurate: modeling password guessability using neural networks[C]//USENIX Annual Technical Conference. [S. l.]: USENIX, 2016.

[28] PEREIRA D, FERREIRA J F, MENDES A. Evaluating the accuracy of password strength meters using off-the-

shelf guessing attacks[C]//2020 IEEE International Symposium on Software Reliability Engineering Workshops. Piscataway: IEEE, 2021: 237−242.

[29] 岳增营, 叶霞, 刘睿珩. 基于语言模型的预训练技术研究综述[J]. 中文信息学报, 2021, 35(9): 15−29.
YUE Zengying, YE Xia, LIU Ruiheng. A review of pre-training technology based on language model[J]. Journal of Chinese information processing, 2021, 35(9): 15−29.

[30] 邱玉祥, 蔡艳, 陈霖, 等. 基于自回归神经网络的多维时间序列分析[J]. 吉林大学学报 (理学版), 2022, 60(5): 1143−1152.
QIU Yuxiang, CAI Yan, CHEN Lin, et al. Multidimensional time series analysis based on autoregressive neural network[J]. Journal of Jilin university (science edition), 2022, 60(5): 1143−1152.

[31] RIQUELME C, PUIGCERVER J, MUSTAFA B, et al. Scaling vision with sparse mixture of experts[EB/OL]. (2021−06−10)[2025−04−23]. https://arxiv.org/abs/2106.05974.

[32] OYMAK S, RAWAT A S, SOLTANOLKOTABI M, et al. On the role of attention in prompt-tuning[C]//International Conference on Machine Learning. Hangzhou: ACM, 2023.

[33] DEVALAL S, KARTHIKEYAN A. LoRa technology - an overview[C]//2018 Second International Conference on Electronics, Communication and Aerospace Technology. Piscataway: IEEE, 2018: 284−290.

[34] CHAUDHARI S, AGGARWAL P, MURAHARI V, et al. RLHF deciphered: a critical analysis of reinforcement learning from human feedback for LLMs[J]. ACM computing surveys, 2026, 58(2): 1−37.

[35] 张钦彤, 王昱超, 王鹤羲, 等. 大语言模型微调技术的研究综述[J]. 计算机工程与应用, 2024, 60(17): 17−33.
ZHANG Qintong, WANG Yuchao, WANG Hexi, et al. A survey of fine-tuning techniques for large language models[J]. Computer engineering and applications, 2024, 60(17): 17−33.

[36] MATSUZAKI T, MIYAO Y, TSUJII J. Probabilistic CFG with latent annotations[C]//Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics-ACL '05. Morristown: ACL, 2005: 75−82.

[37] 王辰成, 杨麟儿, 王莹莹, 等. 基于 Transformer 增强架构的中文语法纠错方法[J]. 中文信息学报, 2020, 34(6): 106−114.
WANG Chencheng, YANG Liner, WANG Yingying, et al. Chinese grammar error correction method based on transformer enhanced architecture[J]. Journal of Chinese information processing, 2020, 34(6): 106−114.

[38] HE Daojing, LIU Zhiyong, ZHU Shanshan, et al. Special characters usage and its effect on password security[J]. IEEE Internet of things journal, 2024, 11(11): 19440−19453.

[39] HE Daojing, LIU Zhiyong, ZHOU Beibei, et al. The impact of digit semantic patterns on password security[J]. IEEE transactions on dependable and secure computing, 2025, 3639170.

[40] HE Daojing, LIU Zhiyong, ZHU Shanshan, et al. Exploiting semantics of special characters to strengthen passwords[J]. IEEE transactions on dependable and secure computing, 2025, 3608956.

[41] DEY R, SALEM F M. Gate-variants of gated recurrent unit (GRU) neural networks[C]//2017 IEEE 60th International Midwest Symposium on Circuits and Systems. Piscataway: IEEE, 2017: 1597−1600.

[42] 杨观赐, 杨静, 李少波, 等. 基于 Dopout 与 ADAM 优化器的改进 CNN 算法[J]. 华中科技大学学报 (自然科学版), 2018, 46(7): 122−127.
YANG Guanci, YANG Jing, LI Shaobo, et al. Improved CNN algorithm based on dopout and ADAM optimizer [J]. Journal of Huazhong University of science and technology (nature science edition), 2018, 46(7): 122−127.

[43] 刘建昌, 李飞, 王洪海, 等. 进化高维多目标优化算法研究综述[J]. 控制与决策, 2018, 33(5): 879−887.
LIU Jianchang, LI Fei, WANG Honghai, et al. Review on evolutionary high-dimensional multi-objective optimization algorithms[J]. Control and decision, 2018, 33(5): 879−887.

[44] 王春柳, 杨永辉, 邓霏, 等. 文本相似度计算方法研究综述[J]. 情报科学, 2019, 37(3): 158−168.
WANG Chunliu, YANG Yonghui, DENG Fei, et al. Summary of research on text similarity calculation methods[J]. Information science, 2019, 37(3): 158−168.

作者简介:

刘志勇, 博士研究生, 主要研究方向为网络安全。发表学术论文 4 篇。E-mail: liuzhiyong0513@163.com。

何道敬, 教授、博士生导师, 哈尔滨工业大学 (深圳) 计算机学院副院长、哈尔滨工业大学 (深圳) 计算与智能研究院常务副院长。连续多年被评选为“爱思唯尔”中国高被引学者及全球前 2% 顶尖科学家。E-mail: hedaojinghit@163.com。

成嘉轩, 硕士, 主要研究方向为信息安全与 AI 安全, 并参与了多项网络安全标准的制定, 持有 CISSP 认证。发表学术论文 4 篇。E-mail: lssn1000@163.com。