The Transactions of the Korea Information Processing Society
TKIPS Letter Paper Vol.14, No.6, pp.397~399

eISSN: 3022-7011
https://doi.org/10.3745/TKIPS.2025.14.6.397

Compressed LLM-Based Secure Coding Support

Chan Woo Lee† · Junyoung Heo††

경량화 LLM 기반 시큐어 코딩 지원

이 찬 우† · 허 준 영††

ABSTRACT

As the importance of security in the IT industry continues to grow, secure coding has become essential. However, small-scale enterprises often struggle to fully implement secure coding guidelines due to a lack of specialized security personnel. This study proposes a compressed Large Language Model (LLM)-based approach to support secure coding, enabling cost-effective adoption of secure coding guidelines. The proposed method leverages an open-source model fine-tuned on a security dataset for domain-specific learning and applies the Low-Rank Adaptation (LoRA) technique to optimize training efficiency in a single GPU environment. Additionally, it incorporates BF16 transformation and 8-bit GGUF quantization to reduce model size, ensuring operability in standard computing environments. Through experiments, we evaluated the proposed model's ability to detect security vulnerabilities and generate improved code suggestions. The results demonstrate superior performance over the baseline model in key evaluation metrics, including F1 score and BLEU score.

Keywords : Secure Coding, Model Compression, Large Language Model(LLM)

요 약

IT 산업에서 보안의 중요성이 증가함에 따라, 시큐어 코딩은 필수적이다. 하지만 소규모 기업에서는 전문 보안 인력 부족으로 시큐어 코딩 가이드를 충실히 반영하지 못할 수 있다. 본 연구는 경량 LLM(Large Language Model) 기반의 시큐어 코딩 지원 기법을 제안하여 저비용으로 시큐어 코딩 가이드를 반영할 수 있도록 한다. 제안 기법은 오픈소스 모델을 활용하여 보안 데이터 세트로 도메인 특화 학습을 수행하고, LoRA(Low-Rank Adaptation) 기법을 적용하여 단일 GPU 환경에서도 효율적인 학습이 가능하도록 최적화하였다. 또한, BF16 변환 및 8bit GGUF 양자화를 통해 경량화하여 일반적인 컴퓨팅 환경에서도 구동할 수 있도록 설계하였다. 실험을 통해 제안 모델이 보안 취약점을 검출하고 개선 코드를 제공하는 기능을 수행하는지 평가하였으며, F1 스코어, BLEU 스코어 등의 평가지표에서 기반 모델보다 우수함을 보였다.

키워드 : 시큐어 코딩, 모델 경량화, 대규모 언어 모델(LLM)

1. 서 론

최근 다양한 산업에서 정보기술(IT)의 중요성이 커지고, 컴퓨팅 인프라와 생성형 인공지능 기술의 발전이 이런 중요성을 더 가속화하고 있다. 이에 따라 많은 기업이 IT 부서를 운영하고 있으며, SW 보안 약점 제거를 목표로 2012년 SW 개발 보안 제도가 도입되었다[1]. 그러나 신생 기업이나 보안 전문 인력이 부족한 조직에서는 서비스 출시 속도를 우선시하며 SW 보안 가이드 지침을 충분히 반영하지 못해 사용자 정보 유출과 같은 보안 사고가 발생할 위험이 크다. 특히 개발 초기 단계에서 보안을 고려하지 않으면 수정 비용이 기하급수적으로 증가해 대규모 보안 사고로 이어질 가능성이 높다.

이를 해결하기 위해 AI 기반의 시큐어 코딩 지원 방법론이 제안되었으며, 생성형 AI 기술의 거대 언어 모델(LLM)을 활용하여 보안 취약점을 사전에 식별하고 보완책을 제시할 수 있다[2]. 그러나 LLM의 높은 비용과 기업 내부 기밀 유출 위험성 등의 문제를 극복하기 위해 경량화 LLM(Compressed LLM)[3]을 활용하는 방안이 논의되고 있다. 경량화 LLM은 경량화된 구조로 일반적인 컴퓨팅 환경에서 구동 가능하며, 공신력 있는 자료를 학습함으로써 시큐어 코딩에 특화된 모델을 구축할 수 있다. 본 연구는 이를 통해 로컬 환경에서 운영 가능한 경량화 LLM 기반 시큐어 코딩 모델을 제안하고, 저비용으로 소프트웨어 보안을 강화할 수 있도록 한다.

2. 제안 기법

제안 기법은 EXAONE 3.0 7.8B Instruct 모델[4]을 기반으

※ 본 연구는 한성대학교 교내학술연구비 지원과제임.
† 준 회 원 : 한성대학교 컴퓨터공학과 석사과정, temp2temp@daum.net, https://orcid.org/0009-0003-3346-2500
†† 정 회 원 : 한성대학교 컴퓨터공학부 교수, jyheo@hansung.ac.kr, https://orcid.org/0000-0001-6407-6678
Manuscript Received : February 6, 2025
Accepted : February 15, 2025
* Corresponding Author : Junyoung Heo(jyheo@hansung.ac.kr)

로 시큐어 코딩을 지원할 수 있도록 설계되었으며, 일반 데스크톱 환경에서도 구동 가능하도록 경량화를 수행하였다. 기존 LLM의 높은 하드웨어 요구사항과 데이터 유출 문제를 극복하고자 3B~70B 사이 크기의 오픈소스 모델을 기반으로 한국인터넷진흥원의 언어별 시큐어코딩 가이드와 SEI CERT Coding Standards 등에서 수집한 약 600개의 취약한 코드와 안전한 코드의 데이터세트를 활용한 도메인 특화 학습을 진행하였다. 데이터 전처리, 모델 학습, 모델 병합의 과정을 거쳐 구현된 제안 모델은 시큐어 코딩 가이드라인에 기반해 보안 약점을 검출하고 개선 코드를 생성하는 데 특화된 모델이다. Table 1은 수집한 데이터세트 상세 정보를 요약한 것이다.

Fig. 1은 제안 모델 생성 과정을 보여준다. 제안 모델은 공신력 있는 시큐어 코딩 가이드 데이터를 기반으로 데이터 수집 및 전처리를 거쳐 비준수 코드와 준수 코드를 매핑하고 불필요한 정보를 제거하여 학습 데이터를 준비한다. 모델 학습 단계에서는 하드웨어 자원과 시간을 절약하기 위해 일부 파라미터만 학습시키는 PEFT(Parameter-Efficient Fine-Tuning) 기법[5], 특히 LoRA(Low-Rank Adaptation)[6]를 활용해 단일 GPU 환경에서도 효율적으로 학습을 진행하며, 특수 토큰을 통해 문맥 이해 능력을 강화하였다. 마지막으로, 학습된 어댑터 파일을 기존 EXAONE 3.0 모델과 병합하고, 8bit GGUF 양자화[7]를 통해 일반적인 컴퓨팅 환경에서도 구동 가능한 최적화된 모델을 구현하였다.

3. 실험 및 성능 평가

LoRA 기법을 사용하더라도 모델의 실행과 학습에는 여전히 많은 자원이 요구된다. 이를 해결하기 위해 본 연구는 클라우드 환경에서 학습과 추론을 지원하는 Google Colab의 유료 서비스인 Colab Pro+의 NVIDIA A100 (40GB) GPU로 진행되었다.

학습에 필요한 모든 조건을 설정한 뒤, Trainer API를 사용하여 학습을 시작하였다. 초기 10회 에포크 학습 결과, Training Loss가 1.3x로 나타나 추가 학습이 필요함을 확인하였다. 이에 따라, 10회 학습된 어댑터 파일을 기본 모델에 병합한 후, 미세 조정된 모델을 바탕으로 다시 10회 에포크 학습을 진행하였다. 학습 과정에서는 모델이 잘 학습되고 있는지, 과적합이 발생하지 않는지를 확인하기 위해 데이터세트를 학습용과 검증용으로 8:2 비율로 나누어 실험을 수행하였다. Fig. 2는 Training Loss와 Validation Loss의 감소 추세를 시각화한 것이다. 학습 결과, Training Loss와 Validation Loss 모두 학습이 진행될수록 안정적으로 감소하는 경향을 보였다. 특히, 두 손실 값 간의 차이가 점진적으로 줄어들며, 과적합 없이 모델이 효과적으로 학습되었음을 확인할 수 있다.

완성된 모델을 연구 목적에 맞는 추론이 가능한지 평가를 진행하였다. 본 연구는 비 준수 코드를 입력하면 해당 코드의

Table 1. Details of the collected dataset

Data name	Language	Key Contents
Python Secure Coding Guide(2023), KISA	Python	An introduction to 46 of the 49 implementation phase security weakness remediation criteria items.
JavaScript Secure Coding Guide(2023), KISA	Javascript	An introduction to 42 of the 49 implementation phase remediation criteria items.
Software Development Security Guide(2021), KISA	C, C#, JAVA	Implementation phase Describes criteria for eliminating security weaknesses and introduces security measures and examples of secure coding in JAVA, C, and C#.
SEI CERT Oracle Coding Standard for Java, Software Engineering Institute(Carnegie Mellon University)	JAVA	Breaking down rules into topics and organizing the tools available to inspect them.
OpenSSF, Secure Coding One Stop Shop for Python	Python	Common Weakness Enumeration (CWE) Mapping.

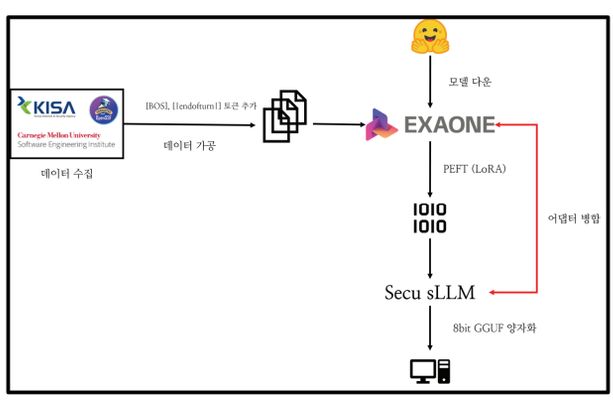

Fig. 1. Proposed Model Generation

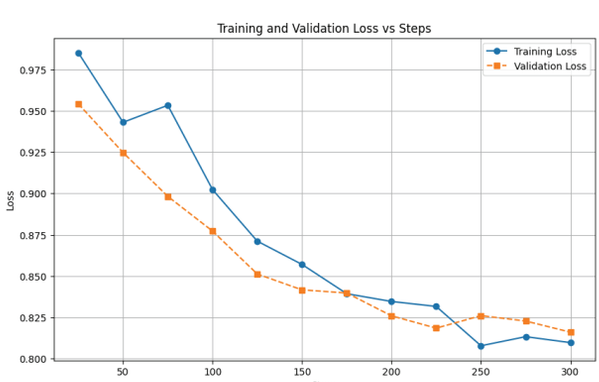

Fig. 2. Training Loss and Validation Loss

Table 2. Performance Comparison

	BLEU Score	F1 Score	Perplexity
Proposed Model(BF16)	0.112	0.257	3.617
Base Model(EXAONE 3.0)	0.088	0.221	5.829

보안 위협을 알려주며 준수 코드를 알려주는 모델이므로 관련 평가를 진행할 수 있는 평가 데이터세트를 찾아보았지만, 특정 언어만 검사하거나, 모델이 평가 데이터세트에 지원을 안 하는 등의 제약조건이 있었다. 그래서 기존에 만든 데이터세트를 활용해 무작위로 30개를 추출해 평가 데이터세트로 제작하여 LLM을 평가할 때 주로 사용하는 평가 방법인 F1 Score, BLEU(Bilingual Evaluation Understudy) Score, Perplexity를 사용했다.

Table 2는 세 가지 평가지표를 활용하여 비교 분석한 결과를 시각화한 것이다. 그래프에서 확인할 수 있듯이, 제작된 모델의 F1 Score와 BLEU가 기본 모델보다 높게 나타났으며, 혼란도를 측정하는 Perplexity 역시 낮은 값을 기록하였다. 이는 완벽한 응답에 더 가까워졌음을 의미한다.

또한 추가 실험에서 FP32(32비트 부동소수점) 타입보다 BF16(16비트 부동소수점)[8]으로 최적화된 제안 모델이 더 나은 성능이 보임을 알 수 있었다.

4. 결 론

본 연구는 보안 중심의 소프트웨어 개발을 지원하기 위해 시큐어 코딩에 특화된 제안 모델을 설계하고 구현하였다. 기존 LLM 모델의 한계를 극복하고자 LoRA 기법과 BF16 최적화를 적용하여, 경량화와 효율성을 동시에 추구하였다. 이를 통해 제안 모델은 소프트웨어 보안 약점 탐지와 보안 코드 생성을 효과적으로 수행할 수 있는 모델임을 확인하였다.

BLEU 점수와 F1 Score에서 기반 모델인 EXAONE 3.0 7.8B 모델 대비 우수한 성과를 보였으며, Perplexity에서도 낮은 값을 기록하여 신뢰할 수 있음을 확인하였다. 이러한 결과를 통해 제안 모델이 일반 수준의 컴퓨터 환경에서도 시큐어 코딩 가이드 LLM으로 활용 가능성이 있음을 보여준다.

제안 모델에 대한 지속적인 발전을 위해 사용하고 검증할 수 있도록 허깅페이스 Hub에 학습된 모델[1]과 데이터세트[2]를 공개하였다.

1) 학습된 모델 공개: https://huggingface.co/LINKER98/secusLLM-Fine-tuning
2) 학습 데이터 공개: https://huggingface.co/datasets/LINKER98/secusLLM-Dataset

References

[1] Korea Internet & Security Agency(KISA), “Software Development Security Guide,” 2021.

[2] F. N. Motlagh, M. Hajizadeh, M. Majd, P. Najafi, F. Cheng and C. Meinel, “Large language models in cybersecurity: State-of-the-art,” *arXiv preprint arXiv:2402.00891*, 2024.

[3] X. Zhu, J. Li, Y. Liu, C. Ma and W Wang, “A survey on model compression for large language models,” *Transactions of the Association for Computational Linguistics*, Vol.12, pp.1556-1577, 2024.

[4] LG AI Research, “EXAONE 3.0 7.8B Instruction Tuned Language Model,” *arXiv preprint arXiv:2408.03541*, 2024.

[5] H. Liu et al., “Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning,” *Advances in Neural Information Processing Systems*, Vol.35, pp.1950-1965, 2022.

[6] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang and W. Chen, “Lora: Low-rank adaptation of large language models,” *arXiv preprint arXiv:2106.09685*, 2021.

[7] A. Chavan, R. Magazine, S. Kushwaha, M. Debbah and D. Gupta, “Faster and Lighter LLMs: A Survey on Current Challenges and Way Forward,” *arXiv preprint arXiv:2402.01799*, 2024.

[8] J. Osorio, A. Armejach, E. Petit G. Henry and M. A. Casas, “BF16 FMA is all you need for DNN training,” *IEEE Transactions on Emerging Topics in Computing*, Vol.10, No.3, pp.1302-1314, 2022.