텍스트 유사도 기반 RAG와 sLLM을 활용한 소프트웨어 보안 버그 리포트 템플릿 생성 및 예측 기법

손영준[O] (한경국립대학교 컴퓨터응용수학부), 양근석* (한경국립대학교 컴퓨터응용수학부 (컴퓨터시스템연구소))

[O]Speaker , *Corresponding Author

요약

소프트웨어 유지보수 과정에서 보안 버그 리포트는 보안 취약점을 식별하고 수정하는 데 중요한 역할을 한다. 그러나 버그 보고자의 보안 지식수준에 따라 보안 버그 리포트의 설명이 보안 위험을 충분히 전달하지 못할 수 있으며, 설명의 누락이나 불명확한 정보가 포함될 가능성도 있다. 이러한 문제를 해결하기 위해 본 논문에서는 텍스트 유사도 기반 RAG와 sLLM 모델을 활용한 소프트웨어 보안 버그 리포트 템플릿 생성 및 보안 버그 리포트 예측 기법을 제안한다. 구체적으로는, 먼저 텍스트 유사도를 활용하여 보안 버그 리포트와 관련된 컴포넌트의 유사한 보안 버그 리포트를 분류하고, 분류된 유사 버그 리포트의 정보를 RAG를 통해 필요한 정보를 추출한다. 이후, sLLM 모델을 사용해 보안 정보를 보완하고 재구성하여 보안 버그 리포트 템플릿을 생성하며, 생성한 버그 리포트 템플릿이 보안과 관련된 것인지 아닌지를 예측한다. 제안된 방법의 성능은 오픈 소스 프로젝트인 Ambari, Wicket, Derby, Camel 데이터셋을 활용하여 평가되었으며, 기존 베이스라인 대비 높은 정확도를 보였다.

Ⅰ. 서론

소프트웨어 유지보수에서 보안 버그 리포트(Security Bug Report, SBR)는 소프트웨어의 안정성과 신뢰성을 유지하는 데 중요한 역할[1]을 한다. 그러나 여전히 다음과 같은 문제가 발생한다.

- 첫째, 작성자의 주관적인 해석으로 인해 버그 리포트의 품질이 일관되지 않음. 프로젝트 매니저, 개발자 등 작성자별로 보안 결함에 대한 해석이 달라지는 경향이 있음
- 둘째, 설명 누락과 불명확한 정보가 포함되어 취약점을 재현하거나 해결하는 데 필요한 정보를 충분히 제공하지 못함
- 셋째, 표준화된 템플릿과 자동화 도구의 부재로 인해 작성 및 분석 과정이 비효율적이며, 중요한 정보가 누락되는 경우도 발생함

이러한 문제를 해결하기 위해 본 연구는 텍스트 유사도 기반 Retrieval-Augmented. Generation (RAG)[2]과 경량화된 거대 언어 모델(sLLM)[2]을 활용한 자동화된 보안 버그 리포트 템플릿 생성 및 예측 기법을 제안한다. 제안된 기법은 유사 버그 리포트를 분류하고, RAG와 sLLM을 사용하여 구조화된 템플릿을 생성함으로써 버그 리포트의 일관성과 정확성을 향상한다.

Ⅱ. 제안한 방법론 및 실험

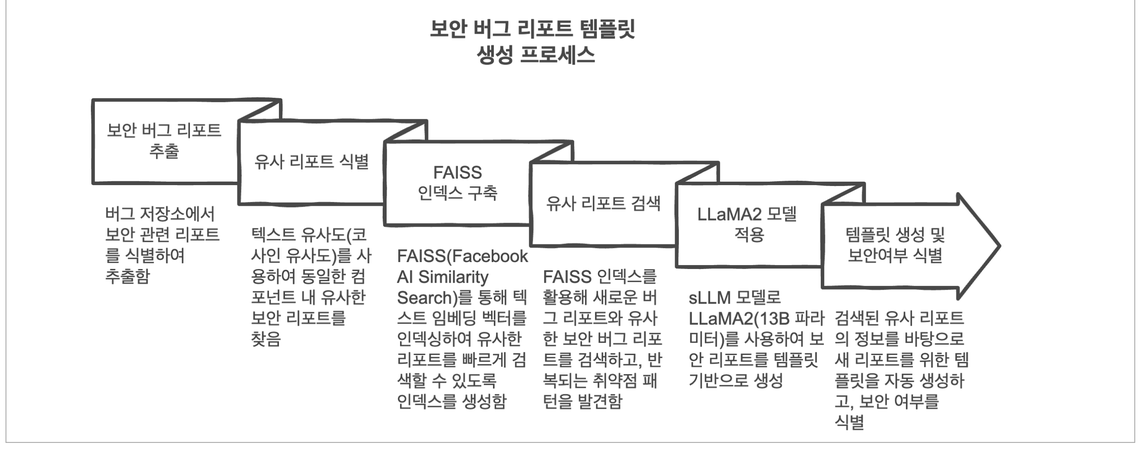


Figure1. 제안한 방법론 도식도

(1) 제안한 방법론

본 연구는 텍스트 유사도 기반 RAG와 LLaMA2[2]를 활용하여 보안 버그 리포트를 템플릿 기반으로 생성하고, 보안 여부를 효과적으로 식별하는 방법을 제안한다. 주요 과정은 다음과 같다.

- **보안 버그 리포트 추출:** 버그 저장소에서 보안 관련 리포트를 식별하고, 요약과 설명을 중심으로 데이터를 수집하여 분석에 필요한 정보를 추출함. 이를 통해 보안 버그와 관련된 주요 정보를 효율적으로 확보함
- **유사 리포트 식별:** 코사인 유사도를 활용하여 동일한 컴포넌트 내에서 유사한 리포트를 그룹화하고, 반복적으로 나타나는 패턴을 분석함 이를 통해 유사한 보안 버그 리포트를 신속하게 분류함
- **FAISS 인덱스 구축:** FAISS(Facebook AI Similarity Search)를 사용하여 텍스트 임베딩 벡터를 인덱싱 하고, 유사 리포트를 빠르게 검색할 수 있는 환경을 제공함. 이를 통해 검색 속도를 향상시키고 대규모 데이터에서도 효율적인 처리가 가능함
- **LLaMA2 모델 활용:** 유사 리포트를 기반으로 LLaMA2(13B 파라미터)를 활용해 템플릿을 자동 생성하고, 보안 여부를 신뢰성 있게 판별함. 이를 통해 보안 관련 문제를 체계적으로 분석하고 대응 방안을 제시함

(2) 데이터셋 및 평가척도

본 연구는 Ambari, Camel, Derby, Wicket 오픈 소스 프로젝트 데이터[1]를 활용하며, Precision, Recall, F1-Score를 기준으로 모델의 성능[1]을 평가한다. 이를 통해 보안 리포트(1)와 비보안 리포트(0)를 이진 분류[3]로 구분하는 성능을 측정한다.

(3) 모델 구성

FAISS 기반 텍스트 유사도 검색과 LLaMA2 모델을 활용한 템플릿 생성 및 보안 여부 판별 과정을 통해, 자동화된 보안 리포트 분류 및 구조화를 효과적으로 수행한다.

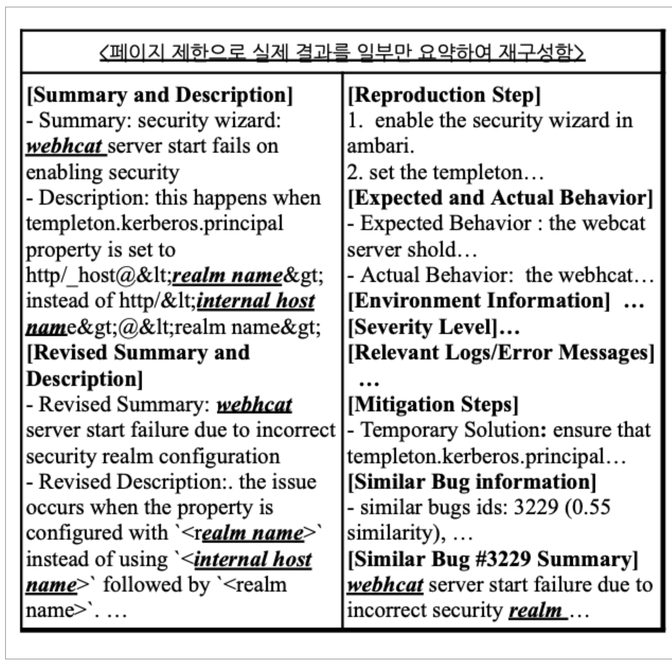
Ⅲ. 실험 결과

제안한 방법(RAG + LLaMA2)의 성능을 기존 베이스라인 모델인 Ji et al.[1]과 Wu et al.[1]의 결과와 비교하였으며, Precision, Recall, F1-Score를 주요 성능 지표로 사용하여 다음 표 1과 같이 요약한다.

Table1. 제안한 방법의 성능평가

	Precision	Recall	F1-Score
Ji[1]	0.8375	0.8725	0.8375
Wu[1]	0.775	0.4375	0.5425
Our Model (LLaMA2)	0.9134	0.9067	0.9100
Our Model (RAG+LLaMA2)	0.9756	0.9758	0.9757

RAG + LLaMA2 모델은 모든 성능 지표에서 기존 방법보다 뛰어난 결과를 보였으며, 특히 F1-Score는 0.9757로 Ji et al.[1]의 0.8375와 Wu et al.[1]의 0.5425 보다 크게 향상하였다. 또한, RAG를 활용함으로써, 이전 버그 리포트에서 관련 패턴을 효과적으로 추출하여 전체적으로 성능이 크게 향상하였다.



〈페이지 제한으로 실제 결과를 일부만 요약하여 재구성함〉

[Summary and Description]
- Summary: security wizard: ***webhcat*** server start fails on enabling security
- Description: this happens when templeton.kerberos.principal property is set to http/_host@<***realm name***> instead of http/<***internal host name***>@<realm name>

[Revised Summary and Description]
- Revised Summary: ***webhcat*** server start failure due to incorrect security realm configuration
- Revised Description:. the issue occurs when the property is configured with `<***realm name***>` instead of using `<***internal host name***>` followed by `<realm name>`. …

[Reproduction Step]
1. enable the security wizard in ambari.
2. set the templeton…

[Expected and Actual Behavior]
- Expected Behavior : the webcat server shold…
- Actual Behavior: the webhcat…

[Environment Information] …

[Severity Level]…

[Relevant Logs/Error Messages] …

[Mitigation Steps]
- Temporary Solution: ensure that templeton.kerberos.principal…

[Similar Bug information]
- similar bugs ids: 3229 (0.55 similarity), …

[Similar Bug #3229 Summary]
webhcat server start failure due to incorrect security ***realm***…

제안한 템플릿 생성 기법은 보안 버그 리포트를 체계적이고 명확하게 구조화하였다.

- 주요 문제를 신속하게 식별하여 대응 시간을 단축함
- 문제의 근본 원인과 예상 동작을 명확히 이해하도록 지원함
- 효과적인 해결책 실행이 가능하도록 구조화된 정보를 제공함
- 유사 버그 정보를 활용하여 문제 해결의 정확성과 효율성을 향상함

Figure2. 출력 템플릿 예시

RAG와 LLaMA2를 결합함으로써 명확하고 간결하며 보안 중심의 버그 리포트 템플릿을 생성하는 능력이 크게 향상되었다. 제안된 방법은 보안 버그 식별 및 수정에서 기존 방법보다 뛰어난 정확도와 실용성을 보였다.

Ⅳ. 결론

본 연구는 텍스트 유사도 기반 RAG와 sLLM 모델을 활용하여 기존 보안 버그 리포트의 주관성과 일관성 문제를 해결하고, 소프트웨어 보안 버그 리포트 템플릿 생성 기법 및 보안 버그 리포트 예측 기법을 제안하였다. 오픈 소스 프로젝트를 활용한 성능 평가 결과, 제안된 모델은 기존 베이스라인 대비 더 높은 정확도와 일관성을 보여주었다. 향후 연구에서는 보안 패턴 확장, 다양한 평가 지표를 활용한 검증, 그리고 생성된 보안 버그 리포트가 원래 리포트의 정보와 의미를 온전히 반영하는지 확인하기 위해, 원본 리포트와 생성된 리포트 간의 정보 손실 여부와 일치율을 정밀히 검증할 계획이다.

참고문헌

1. J. Ji, G. Yang, “Enhancing Security Bug Report Prediction with Severity-Based Feature Selection through LSTM Algorithm,” in Proc. of the KCC, 2024.
2. Y. Gao, "Improving the Capabilities of Large Language Model Based Marketing Analytics Copilots With Semantic Search And Fine-Tuning," International Journal on Cybernetics & Informatics, 2024.
3. R. Shu, T. Xia, L. Williams, and T. Menzies, “Better Security Bug Report Classification via Hyperparameter Optimization,” arXiv preprint arXiv:1905.06872, 2019.