문장 길이에 따른 딥러닝 기반 보이스 피싱 탐지 기법 성능 분석

최상현, 김홍국*

광주과학기술원, *광주과학기술원

shchoiga@gm.gist.ac.kr, *hongkook@gist.ac.kr

Performance Analysis of Deep Learning-Based Voice Phishing Detection According to Sentence Length

Sang Hyun Choi, Hong Kook Kim*

Gwangju Institute of Science and Technology (GIST)

요 약

보이스 피싱은 인간의 통신 취약점을 악용하여 개인 보안과 사생활에 상당한 위협이 되고 있다. 본 논문에서는 보이스 피싱 탐지 정확도에서 문장 길이에 대한 중요성을 알아본다. 이를 위해, 보이스 피싱을 탐지하기 위한 딥러닝 기법은 KoBERT 모델을 활용한다. 금감원에 공개된 279개의 보이스 피싱 음성파일을 비롯하여 유튜브에서 얻은 121개의 보이스 피싱 음성파일로부터 KoBERT를 활용한 보이스 피싱 탐지 기법의 정확도와 문장 길이와의 유의미한 상관관계를 확인할 수 있었다. 본 연구의 결과가 시사하는 바를 보이스 피싱 탐지 기술의 성능 개선에 활용한다면 보이스 피싱 탐지의 속도와 정확도를 크게 향상할 수 있을 것으로 기대된다.

Ⅰ. 서 론

피싱 또는 보이스 피싱이라고 불리는 범죄는 대한민국을 비롯한 전 세계에서 지속해서 발생하고 있다. 이러한 보이스 피싱은 인간의 의사소통에 내재하여 있는 취약점을 공격하는 방향으로 중대한 위협이 되고 있다. 보이스 피싱 범죄는 신뢰와 기만의 미묘한 차이로 무차별적으로 일반 시민들을 대상으로 공략하고 있으며 보이스 피싱 범죄는 시간이 지날수록 수법이 다양하고 전문화가 되어가고 있다.

하지만 보이스 피싱 범죄는 한 가지 공통점이 있다. 그건 바로 모든 보이스 피싱 범죄는 금전 요구와 개인정보를 요구한다는 점이다. 이 특징을 이용하여 딥러닝 기법을 활용하여 효과적으로 보이스 피싱을 탐지하는 방법들이 연구되어 왔다[1]-[3]. 하지만, 이러한 탐지 기법의 단점은 보이스 피싱 범죄를 빠르고 정확하게 보이스 피싱 여부를 판단하지 못한다는 것이다.

본 논문에서 제시하는 연구는 이러한 필요성에 의해 출발한다. 보이스 피싱 시도의 언어적 특성, 특히 문장 길이에 초점을 맞춰 조사하고자 한다. 즉, 문장 길이의 차이에 따른 기존 딥러닝 기법의 정확성에 미치는 영향을 연구하고자 한다. 이를 통해 효율성을 가장 크게 향상할 수 있는 최적의 문장 길이를 도출해 내고자 한다. 본 연구의 결과는 문장 길이가 탐지 정확도에 미치는 영향에 대한 유의미한 분석을 제공하여 단순히 학술적 발전이 아니라 사이버 보안을 개선하는 데 실질적인 시사점을 제공하는 데 있다. 본 연구의 결과를 활용하면 보이스 피싱 탐지의 정확도와 속도를 향상시켜 그 위험과 영향을 효과적으로 줄일 수 있을 것이다.

Ⅱ. 본론

본 연구에서 학습 데이터는 이전의 연구[1],[2]에서와 같이 KorCCVi_v2의 데이터셋을 활용하였다. 이 데이터셋에서 보이스 피싱과 상관없는 무해한 파일 데이터는 일상 대화 말뭉치 2020 (NIKL_DIALOGUE_2020_v1.2) 가 활용되었으며, 유해한 파일은 금융감독원 사이트에 게시된 파일이 활용되었다. 음성으로 추출한 스크립트는 다듬는 과정을 거쳐 한 파일 안에 묶였는데 무해한 파일은 2,230개의 말뭉치로 이루어져 있고, 유해한 파일은 695개로 이루어져 있다.

본 연구에 테스트 데이터는 최대한 전화환경과 최대한 비슷하게 설정하였다. 금감원에 공개된 279개의 보이스 피싱 음성파일을 비롯하여 유튜브에서 얻은 121개의 보이스 피싱 음성파일을 추출하였다. 이렇게 추출된 음성파일을 네이버 클로바노트에서 제공하는 Speech-to-Text 기능을 활용하여 스크립트를 얻었다. 무해한 파일은 일상 대화 말뭉치 2020 (NIKL_DIALOGUE_2020_v1.2)에서 608개를 추출하였다. 추출한 데이터로 문장 단위로 끊어서 테스트를 수행하였다. 이때 마침표, 느낌표, 물음표 등 문장 기호를 기준으로 문장을 나누었다. 하지만, 본 연구에서는 추출한 데이터가 7문장 이상인 경우가 드물어 2문장에서 7문장까지로 데이터를 나누었고 이 경우들만을 고려하였다.

기존의 보이스 피싱 탐지에 관한 연구는 여러 딥러닝 모델을 활용하여 연구가 되고 있다. 이들 딥러닝 모델 가운데 KoBERT라는 구조를 활용한 모델이 가장 뛰어난 성능을 보였다[4]. KoBERT란 한국어의 성능한계를 극복하기 위해 기존의 BERT 구조를 발전시킨 것이다. KoBERT를 활용한 binary classification의 주요한 parameter는 표 1과 같다.

표 1. KoBERT를 활용한 Binary Classification의 주요한 parameter

Parameters	Specific Value
Batch size	3
Learning rate	32
Max grad norm	5e-5
Log interval	200
Dr rate	0.4

총 10번의 epoch을 통해 KorCCVi_v2의 데이터셋으로 학습하였고 training accuracy는 99.84%를 얻을 수 있었다. Binary classification의 구조를 블록다이어그램으로 나타내면 그림 1과 같다.

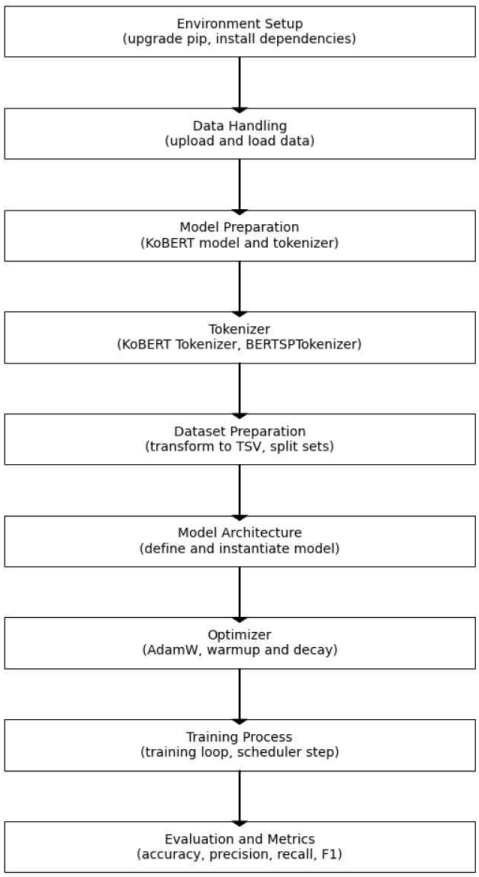

그림 1. Binary Classification의 블록다이어그램

표 2는 학습 이후 데스트 데이터를 적용하여 얻은 문장 길이에 따른 보이스 피싱 탐지 기법의 정확도를 보여 준다. 표에서 보인 바와 같이, 문장 변화가 없었을 때는 약 98.19% accuracy와 0.9716의 F1 score를 산출한다는 것을 알 수 있다. 문장을 6문장 입력으로 할 경우까지 이 accuracy와 F1 score가 유지가 되었다. 하지만 입력된 문장을 5문장 이하로 줄이면 성능이 점점 떨어져 2문장을 넣었을 때 accuracy는 약 95.93%, F1 score는 0.9463으로 accuracy의 경우 2.3%, F1 score는 2.6% 성능이 떨어지게 되었다. 표 2를 통해 5문장 이하에서는 빠른 속도를 위해 정확도를 희생해야 되지만 6문장의 경우 정확도는 최대로 유지하면서 최고로 빠르게 분류할 수 있다는 것을 확인할 수 있다.

표 2. 문장 길이에 따른 정확도

문장의 길이	Accuracy (%)	F1 score
2문장	95.93%	0.9463
3문장	96.17%	0.9564
4문장	96.49%	0.9463
5문장	97.67%	0.9644
6문장	98.19%	0.9716
7문장	98.19%	0.9716
문장 변화 X	98.19%	0.9716

Ⅲ. 결론

본 연구에서는 KoBERT를 활용한 보이스 피싱 탐지 기법의 정확도와 문장 길이와의 유의미한 상관관계를 확인할 수 있었다. 6문장보다 짧은 경우, 문장 길이가 감소함에 따라 정확도가 유의미하게 감소하는 경향을 보였다. 반면 문장 길이가 6일 때부터는 속도를 위해 정확도를 희생하지 않아도 된다는 것을 확인 할 수 있었다.

본 연구의 결과를 활용하여 보이스 피싱 방지 기법을 적용할 때 최적의 문장 길이인 6문장에 대하여 수행한다면 보이스 피싱에서 매우 중요한 빠른 속도와 높은 정확성을 만족시킬 수 있을 것이다. 본 연구의 결과는 사회적 경제적으로 매우 큰 문제를 야기하는 보이스 피싱 탐지에 활용되어 사이버 보안 기술의 성능 향상에 크게 도움을 줄 수 있을 것으로 예상이 된다. 보이스 피싱은 향후 고도화될 것으로 예상이 되므로 이에 따라 언어 모델과 데이터셋을 지속해서 개선하고 업데이트해야 한다. 특히 짧은 문장으로 빠르고 정확한 결과를 얻을 수 있는 모델이 요구되는데 본 연구가 시사하듯이 강력한 방어 메커니즘의 설계 및 구현에 있어 문장 길이가 중요한 매개변수로 역할을 하므로 향후 보이스 피싱 탐지 기술의 성능 개선에 문장 길이를 우선적으로 고려해야 할 것으로 판단된다.

ACKNOWLEDGMENT

이 논문은 광주과학기술원의 오디오지능연구실의 인턴으로 2024년도 광주과학기술원 GIST-MIT 공동연구사업의 지원으로 수행되었음.

참 고 문 헌

[1] Boussougou, M.K.M., An Artificial Intelligent Approach to Detect Voice Phishing Crime by Analyzing the Call Content: A Case Study on Voice Phishing Crime in South Korea, MS Thesis, Soongsil Univ., 2021.

[2] Boussougou, M.K.M, and Park, D.-J., “A real-time efficient detection technique of voice phishing with AI,” in Proc. of KCC 2021, pp. 768-770, July 2021.

[3] 김원웅, 강예준, 김현지, 양유진, 오유진, 이민우, 임세진, 서화정, “딥러닝과 감성 분석에 따른 보이스피싱 여부 판별,” ACK 2021 학술발표대회 논문집, 28권, 2호, pp. 811-814, 2021.

[4] Boussougou, M.K.M, and Park, D.-J., “Attention-based 1D CNN-BiLSTM hybrid model enhanced with FastText word embedding for Korean boice phishing detection,” Mathematics, vol. 11, 2023.