빈출 키워드 및 문맥 기반 보이스피싱 수법 분류 기술

임정민[1,2], 오지연[1,3], 박현호[1,4]

[1]한국전자통신연구원, [2]한림대학교, [3]명지대학교, [4]과학기술연합대학원대학교

limjm1617@gmail.com, yeonaf6e@mju.ac.kr, hyunhopark@etri.re.kr

Voice Phishing Scam Method Classification Using Keyword-Based and Context-Based Approaches

Jeongmin Lim[1,2], Jiyeon Oh[1,3], Hyunho Park[1,4]

[1]Electronics and Telecommunications Research Institute, [2]Hallym University, [3]Myongji University, [4]University of Science and Technology

요 약

본 연구에서는 통화 내 빈출 키워드 분석과 문맥 기반 분석을 활용하여 보이스피싱 수법(예: 대출 사기형, 수사기관 사칭형)을 식별하는 분류 기술을 제안한다. 빈출 키워드 기반 분류 방식에서는 대출 사기형과 수사기관 사칭형 보이스피싱에서 자주 등장하는 핵심 키워드를 추출하고, 텍스트 내 수법별 키워드 출현 여부를 기준으로 보이스피싱 수법을 분류하였다. 문맥 기반 분류 방식에서는 SBERT(Sentence-BERT)를 활용하여 보이스피싱 텍스트의 문맥 정보를 학습하고, 학습된 임베딩을 기반으로 보이스피싱 수법을 분류하였다. 제안한 기법의 성능 평가를 위해 금융감독원에서 공개한 보이스피싱 음성 데이터를 텍스트로 변환하여 사용하였다. 실험 결과, 빈출 키워드 기반 분류 방식은 93.00%의 수법 분류 정확도를 달성하였으며, 문맥 기반 분류 방식은 100%의 수법 분류 정확도를 보였다. 본 연구에서 제안한 보이스피싱 수법 분류 기술은 보이스피싱 수사 과정에서 수사 인력 배치 및 수사 계획 수립을 지원하는 데 활용될 수 있을 것으로 기대된다.

Ⅰ. 서 론

보이스피싱 범죄가 증가함에 따라 분석이 필요한 통화 데이터의 규모가 확대되고 있으며, 이를 수사 인력이 직접 청취하여 판단하는 방식은 시간과 인력 소요 측면에서 한계를 보이고 있다[1]. 이에 따라 기존 학계 및 산업계에서는 보이스피싱 피해 예방을 목적으로 실시간 통화가 보이스피싱인지 여부를 판단하는 기술(예: LG유플러스 ixi)을 중심으로 연구가 이루어져 왔다[2, 3]. 그러나 보이스피싱 범죄에 대한 효과적인 수사를 위해서는 단순한 탐지 여부를 넘어, 범죄 수법을 식별하고 분류하는 과정이 중요하다.

범죄 수법에 따라 수사 인력 배치와 수사 계획은 달라질 수 있다. 예를 들어, 저금리 대출이나 대환대출을 미끼로 피해자의 금융 정보를 유도하는 “대출 사기형” 보이스피싱은 금융 거래 흐름 분석과 계좌 추적을 중심으로 한 수사 인력 배치가 요구된다. 반면, 검찰·경찰 등 수사기관을 사칭하여 심리적 압박을 가하는 “수사기관 사칭형” 보이스피싱은 통화 패턴 분석과 통신 경로 추적, 그리고 조직형 범죄 여부 판단을 수행할 수 있는 수사 인력의 집중적인 투입이 필요하다.

이에 본 연구에서는 “대출 사기형”과 “수사기관 사칭형” 보이스피싱 수법을 대상으로, 빈출 키워드 분석과 문맥 기반 분석을 활용한 보이스피싱 수법 분류 기술을 제안한다. 빈출 키워드 기반 분류 방식은 각 수법에서 반복적으로 등장하는 핵심 단어의 출현 빈도를 기준으로 통화 내용을 분석하는 접근법이다. 반면, 문맥 기반 분류 방식은 SBERT(Sentence-BERT)[4]를 활용하여 통화 문장 전체의 의미와 흐름을 학습함으로써, 동일한 단어가 사용되더라도 문맥에 내포된 범죄 의도를 고려하여 수법을 구분한다. 본 연구는 두 분류 기법의 성능을 비교·분석하고, 이를 통해 실제 보이스피싱 수사 현장에서 범죄 수법을 신속하고 정확하게 식별하는 데 기여하고자 한다.

Ⅱ. 빈출 키워드 기반 분석과 문맥 기반의 보이스피싱 수법 분류 기술

본 연구에서는 제안하는 수법 분류 기술의 성능 평가를 위해 실제 보이스피싱 범죄 양상을 분석하고자 금융감독원에 공개된 보이스피싱 음성 데이터 200건(대출 사기형 100건, 수사기관 사칭형 100건)을 대상으로 수집하였다. 해당 데이터는 주변 소음과 통화 품질 저하 등 비정형적 특성이 강하므로, 고성능의 텍스트 추출을 위해 OpenAI의 Whisper large-v3 모델을 활용하였다[5]. Whisper 엔진을 통해 전사된 데이터는 JSON 형식으로 구조화되어, 단순 텍스트뿐만 아니라 발화 시점의 메타데이터를 포함한 데이터셋으로 확보되었다.

보이스피싱 수법 분류 성능을 향상시키기 위해 입력 데이터의 언어적 특징을 정제하는 전처리 과정을 수행하였다. 구어체 발화에서 빈번히 발생하는 형태소 파편화 문제를 완화하기 위해 정규표현식 기반의 ‘핵심어 통합 매핑 알고리즘’을 적용하였다. 해당 알고리즘은 형태가 일부 상이한 변이형 어휘(예: ‘계좌가’, ‘계좌로’)를 통일된 어근 형태(예: ‘계좌’)로 변환함으로써 단어 출현 빈도의 일관성을 확보한다. 또한 ‘그리고’, ‘네’와 같이 분석 유의성을 저하시킬 수 있는 불용어를 제거하여, 범죄 수사 관점에서 의미 있는 키워드를 중심으로 한 데이터셋을 구축하였다.

1. 빈출 키워드 기반 보이스피싱 수법 분류

본 단계에서는 수법별 언어적 특징을 정량화하기 위해 유형별 상위 10개 빈출 키워드를 도출하였다. 또한, 분석의 객관성을 위해 전체 데이터셋을 7:3 비율로 무작위 분할하여 실험을 진행하였다. 데이터 분석을 통해 각 수법의 빈출 키워드를 정의한 결과, 대출 사기형과 수사기관 사칭형 수법 모두 “통장”, “계좌”가 빈출 단어에 포함됨을 확인하였다. 이 외에 대출 사기형은 “대출”, “은행”, “상환”, “금리” 등 경제적 어휘가 주로 도출되었고, 수사기관 사칭형은 “사건”, “조사”, “명의”, “수사” 등 법적 어휘가 주로 도출되었다.

위와 같이 수법별 빈출 키워드가 실제 데이터 속에 얼마나 포함되어 있

는지를 기준으로 보이스피싱 수법 분류를 진행하였다. 테스트 데이터의 각 문장에서 수사기관 사칭형 키워드와 대출 사기형 키워드가 각각 몇 번 나타나는지 합산한 뒤, 더 많이 발견된 쪽으로 수법을 결정하는 방식을 채택하였다. 실험 결과, 그림 1과 같이 단어 출현 횟수만을 기준으로 분류했을 때 약 93%의 정확도를 기록하였다. 그러나 '계좌', '확인', '본인' 등 두 수법에서 공통으로 발견되는 단어들이 한 문장에 섞여 나올 경우, 단순히 단어의 개수를 세는 것만으로는 범죄자가 어떤 의도를 가지고 말하는지 판별하기 어려워 수법을 오분류하는 한계가 관찰되었다.

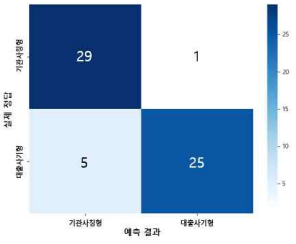

그림1. 빈출 키워드 기반 수법 분류 성능

2. 문맥 기반 보이스피싱 수법 분류

본 단계에서는 SBERT(Sentence-BERT) 모델을 활용하여 데이터 전체의 맥락을 학습시키는 고도화 과정을 수행하였다. 본 방식은 빈출 키워드 기반 접근과 달리 단어의 출현 여부에 의존하지 않으며, SBERT를 활용해 텍스트를 고차원 임베딩 공간으로 변환함으로써 대화의 흐름과 수법별 언어적 패턴을 학습한다. 예를 들어, 동일한 "계좌"라는 단어가 포함되더라도 그것이 기관 사칭형의 "수사 목적의 자금 동결" 맥락인지, 대출 사기형의 "대출 실행을 위한 입금" 맥락인지를 전체적인 흐름을 통해 식별한다.

앞서와 동일한 데이터를 7:3의 비율로 무작위 분할하여, 문맥 기반 보이스피싱 수법 분류 실험을 진행하였다. 학습 과정에서는 JSON 파일을 SBERT 모델에 입력하여 각 문장의 의미를 벡터으로 변환하였으며 문맥을 학습을 진행했다. 이후 입력된 문장이 학습된 두 가지 수법 중 어느 쪽과 통계적으로 더 유사한지를 계산하여 유형을 판별하도록 설계하여 테스트를 진행하였다. 실험 결과, 그림 2와 같이 문맥을 학습하도록 한 SBERT 모델은 테스트 데이터 전체에 대해 100%의 분류 정확도를 달성하였으며, 단순 단어 빈도 분석보다 문장 간의 의미론적 관계를 파악하는 방식이 지능화된 보이스피싱 수법을 식별하는 데 훨씬 효과적이라는 사실을 입증하였다.

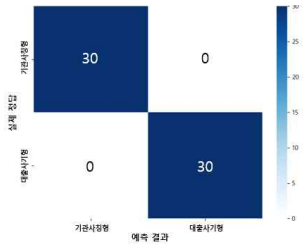

그림2. 문맥 기반 수법 분류 성능

3. 빈출 키워드 및 문맥 기반 수법 분류 성능 결과 비교 및 분석

수법별 빈출 키워드 방식과 SBERT 기반 문맥 학습 방식의 성능을 비교 분석한 결과, 문맥에 따라 의미가 달라지는 언어의 특성을 어떻게 처리하고 반영하느냐에 따라 달라진 것으로 판단된다. 키워드 방식은 특정 단어가 나타나면 기계적으로 점수를 합산하기 때문에, 수사기관을 사칭하면서 대출 관련 용어를 섞어 쓰거나, 반대로 대출에 관해 설명하면서 "사건", "조사"와 같은 단어를 사용한다면 이에 대응하기 어렵다. 반면, SBERT는 단순한 단어의 일치 유무나 단어의 나열이 아닌, 대화 전체에 내포된 수법별 흐름을 파악한다. 이러한 특징은 실제 수사 현장에서 신종 변종 수법이 등장하더라도 핵심적인 범죄 의도를 놓치지 않고 분류해낼 수 있다는 점에서 높은 실무적 효용성을 지닌다.

Ⅲ. 결론

본 연구에서는 지능화되는 보이스피싱 범죄 수법에 효과적으로 대응하기 위해 빈출 키워드 기반 수법 분류와 문맥 기반 수법 분류를 결합한 지능형 분류 기술을 제안하고, 실제 범죄 데이터를 통해 그 실효성을 검증하였다. 특히 전국 단위에서 발생하는 대량의 보이스피싱 사건을 대상으로 범죄 수법을 신속하게 식별하여 수사의 초기 방향성과 인력 배치 결정을 지원하는 데 연구의 초점을 두었다.

실험 결과, 빈출 키워드 기반 수법 분류는 93%의 정확도를 기록하여 대규모 사건에 대한 신속한 1차 분류에 유용함을 확인하였으나, 여러 수법에서 공통적으로 사용되는 어휘가 혼재된 경우 범죄 의도를 충분히 반영하지 못하는 한계가 관찰되었다. 반면, SBERT 기반 문맥 학습을 적용한 문맥 기반 수법 분류는 테스트 데이터 전체에 대해 100%의 분류 정확도를 달성하였으며, 통화 문장 전체의 의미 흐름과 발화 맥락을 고려함으로써 수법별 언어적 패턴을 효과적으로 식별하였다. 이러한 문맥 기반 분류 방식은 사건 간 연관성 분석과 조직형 범죄 패턴 추적에 활용 가능한 정량적 단서를 제공한다는 점에서 높은 실무적 가치를 지닌다.

다만 본 연구는 제한된 규모의 수사 데이터를 활용하였다는 한계를 가지며, 향후 보다 방대하고 다양한 데이터셋을 확보하여 모델의 일반화 성능과 신뢰성을 추가로 검증할 필요가 있다. 나아가 제안한 문맥 기반 분석 기술은 보이스피싱 뿐만 아니라, 스미싱과 같은 텍스트 기반 사이버 범죄 분류에도 확장 적용이 가능할 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 26년도 정부(경찰청)*의 재원으로 과학치안진흥센터 사이버 범죄 수사단서 통합분석 및 추론시스템 개발 사업의 지원을 받아 수행된 연구임(No. RS-2025-02218280)

참 고 문 헌

[1] 김상민, 노승민, "딥러닝 기반 NLP 및 지식증류기법을 활용한 보이스피싱 의심 발언 탐지," 한국전자거래학회지, vol. 29, no. 4, pp. 139-148, 2024.

[2] 이치훈, 윤철희, "생성형 AI를 활용한 범죄 예방을 위한 머신러닝 기반 딥보이스 판별에 관한 연구," 한국정보기술학회 2024년도 하계종합학술대회 및 대학생논문경진대회, pp. 25-29, 2024.

[3] LG유플러스, "ixi (익시): 온디바이스 AI 기반 보이스피싱 실시간 탐지 및 차단 기술," LG U+ 기술 백서, 2024.

[4] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in Proc. 2019 Conf. Empirical Methods Natural Language Processing 9th Int. Joint Conf. Natural Language Processing (EMNLP-IJCNLP), pp. 3982-3992, 2019.

[5] A. Radford et al., "Robust Speech Recognition via Large-Scale Weak Supervision," in Proc. 40th Int. Conf. Machine Learning (ICML), PMLR 202, pp. 28,368-28,387, 2023.