JDFR Journal of Data Forensics Research
데이터포렌식연구, 2권 2호, 2025년 12월, 175-195면
Journal of Data Forensics Research, Volume 2, Issue 2, December 2025, Pages 175-195.
https://doi.org/10.12972/JDFR.2025.2.2.4
Print ISSN 3058-9460
Online ISSN 3058-9479

원저

키워드 기반 지식그래프와 GNN을 활용한 사이버 범죄 분류

정희훈[1], 이다은[1], 김선호[2], 윤수식[3], 김동주[4]

[1]고려대학교 뇌공학과 Brainlab 연구실 연구원
[2]고려대학교 컴퓨터학과 DAIS 연구실 연구원
[3]고려대학교 컴퓨터학과 교수
[4]고려대학교 뇌공학과 교수

교신저자: 윤수식, susik@korea.ac.kr; 김동주, dongjookim@korea.ac.kr

요약

디지털 금융거래와 온라인 커뮤니케이션 채널의 확산으로 텔레그램·웹사이트를 매개로 한 사이버 범죄가 지속적으로 증가하고 있다. 그러나 방대한 텍스트 기반 오픈 소스 인텔리전스(Open-Source Intelligence, OSINT) 데이터를 수작업으로 분석·분류하는 데에는 명확한 한계가 있으며, 이에 따라 자동화된 분류 체계에 대한 필요성이 커지고 있다. 본 연구는 이러한 문제를 해결하기 위해 텔레그램 채널 메시지와 웹사이트 HTML을 키워드 기반 지식그래프(Knowledge Graph)로 구조화하고, 그래프 신경망(Graph Neural Network, GNN)을 적용하여 범죄 여부 및 유형을 자동 분류하는 통합 프레임워크를 제안한다. 구조화된 지식그래프는 텔레그램 채널의 다중 클래스 범죄 유형 분류와 웹사이트의 이진 범죄/정상 분류를 위한 그래프 합성곱 신경망(Graph Convolutional Networks, GCN) 학습에 활용된다. 실험 결과, GCN 모델은 로지스틱 회귀, 다층 퍼셉트론, XGBoost 등 비교 모델 대비 전반적으로 우수한 성능을 보였으며, 텔레그램 범죄 유형 분류에서 약 0.80, 웹사이트 범죄 탐지에서 약 0.92의 F1 점수를 달성하였다. 종합하면, 본 연구는 텍스트 기반 OSINT를 지식그래프–GNN 워크플로로 연결함으로써 사이버 범죄 탐지·분류의 자동화를 실현하고, 향후 멀티모달 및 이질 그래프 기반 위협 인텔리전스 시스템으로 확장할 수 있는 기반을 제시한다.

주제어

지식그래프, 그래프 신경망(GNN), 오픈소스 인텔리전스(OSINT), 사이버 범죄 탐지, 단어 임베딩

Open Access

Received: December 11, 2025
Revised: December 17, 2025
Accepted: December 27, 2025
Published: December 31, 2025

Original Article

Cybercrime classification using keyword-based knowledge graphs and graph neural networks

Heehun Jeong[1], Daeun Lee[1], Sunho Kim[2], Susik Yoon[3], Dong-joo Kim[4]

[1]Researcher, Brainlab Laboratory, Department of Brain Engineering, Korea University, Republic of Korea
[2]Researcher, DAIS Laboratory, Department of Computer Science, Korea University, Republic of Korea
[3]Professor, Department of Computer Science, Korea University, Republic of Korea
[4]Professor Department of Brain Engineering, Korea University, Republic of Korea

Corresponding Author: Susik Yoon, susik@korea.ac.kr; Dong-joo Kim, dongjookim@korea.ac.kr

ABSTRACT

With the spread of digital financial transactions and online communication channels, cybercrime involving Telegramand websites has been constantly increasing. However, clear limitations exist in manually analyzing and classifying vast text-based open-source intelligence (OSINT) dataaccordingly, the need for an automated classification system is growing. To solve this problem, this study proposes an integrated framework for automatically classifying crime conditions and types by structuring Telegram channel messages and website HTML into keyword-based knowledge graphs and applying graph neural networks (GNNs). Structured knowledge graphs are used to learn graph convolutional networks (GCNs) for multiclass crime type classification of Telegram channels and binary crime/normal classification of websites. As a result of the experiment, the GCN model showed overall superior performance compared with comparative models such as logistic regression, multilayer perceptron (MLP), and XGBoost and achieved an F1 score of approximately 0.80 in Telegram crime type classification and approximately 0.92 in website crime detection. Taken together, this study proposes the basis for automating cybercrime detection and classification by connecting text-based OSINT to a knowledge graph–GNN workflow and expanding it to multimodal and heterogeneous graph-based threat intelligence systems in the future.

KEYWORDS

knowledge graphs, graph neural networks (GNNs), open-source intelligence (OSINT), cybercrime detection, word embedding

I. 서론

1.1. 사이버 범죄 확산과 위협 양상의 변화

최근 수년간 랜섬웨어, 피싱, 계정 탈취와 같은 사이버 범죄는 빈도와 피해 규모 모두에서 꾸준한 증가세를 보이고 있다. 예를 들어 미국 FBI 인터넷범죄신고센터(IC3)에 접수된 2024년 인터넷 범죄 피해액은 약 166억 달러로 전년 대비 30% 이상 증가했으며, 피싱·스푸핑, 개인 정보 침해, 투자 사기 등이 상위를 차지한다[1]. 글로벌 보고서들 역시 2024년 이후 랜섬웨어와 사이버전, 소셜 엔지니어링 기반 공격이 전 세계 조직과 일반 사용자에게 구조적인 위협이 되고 있음을 반복해서 지적하고 있다[2].

국내 또한 경찰청 통계에 따르면 2016~2020년 사이 연평균 약 16만 9천 건의 사이버 범죄가 신고되었고, 특히 인터넷 사기 건수는 2017년 9만 2천여 건에서 2020년 17만 4천여 건으로 급증한 뒤 2022년에도 15만 건 이상 수준을 유지하고 있다[3]. 현재 텔레그램과 같은 메신저 채널, 위장 웹사이트, 암호화폐 기반 결제 수단을 활용한 신종 범죄 수법은 짧은 주기로 등장하고 변형되고 있으며, 수사기관과 보안 담당자가 모든 위협을 수작업으로 추적·분류하는 것은 점점 더 어려워지고 있다.

1.2. 사이버 범죄 탐지·분류를 위한 기존 연구 동향

최근 다양한 온라인 공간에서 발생하는 이러한 사이버 범죄를 자동으로 탐지하고 분류하기 위한 기술적 연구가 활발하게 이루어지고 있다. 대표적으로 온라인 성적 그루밍과 같은 성범죄 탐지 분야에서는 실제 대화 데이터를 기반으로 한 머신러닝·딥러닝 기반 분류 기법이 주요 접근으로 자리 잡아왔다. Borj et al.은 온라인 채팅 대화의 문장 임베딩을 대조 학습 방식으로 학습하는 SimCSE 기반 모델을 제안함을 통해 문장 간 긍정·부정 쌍을 대비시키며 표현 공간을 구조화하고, 이후 서포트 벡터 머신(Support Vector Machine, SVM) 분류기를 결합하여 그루밍 대화와 비(非)그루밍 대화를 효과적으로 구분할 수 있음을 보여주었다[4]. 후속 연구에서는 역번역 기반 데이터 증강을 적용해 그루밍 탐지 정확도를 향상시키는 방법이 제시되었으며[5], 또한 한국어 SNS 환경을 대상으로 텍스트와 스크린샷 이미지를 함께 처리하는 딥러닝 모델이 개발되는 등, 실제 서비스 환경을 반영한 언어·매체 융합형 탐지 기술도 등장하고 있다[6].

금융사기 탐지 영역에서도 다양한 형태의 자동 분류 연구가 보고되고 있다. 신용카드 거래 데이터를 활용한 사기 트랜잭션 탐지는 전통적 머신러닝과 딥러닝 기법을 비교·평가하는 연구가 꾸준히 축적되어 왔으며[7], 최근에는 딥웹·다크웹, 텔레그램, Reddit 등 여러 플랫폼에서 발생하는 불법 거래 문서를 준지도 학습으로 자동 분류하는 연구가 제안되었고, 마약·무기·자격 증명 등 다양한 범죄 범주를 높은 정확도로 구분할 수 있음이 보고되었다[8].

한편, 메신저·소셜 네트워크에서 유포되는 사기성 메시지나 악성 콘텐츠를 자동으로 분류하려는 시도도 꾸준히 이어지고 있다. 초기 연구에서는 메시지의 의미적 특징을 기반으로 한 텍스트 중심 사기 탐지 모델이 제안되었으며[9], 이러한 접근은 이후 다양한 플랫폼에서의 사기 행위를 자동 분석·탐지하기 위한 기초적인 방법론으로 활용되고 있다. 최근에는 텔레그램에서 운영되는 대규모 사이버범죄 채널을 분석하여, 게시물을 범주별로 자동 분류하고 신규 악성 채널을 탐지하는 실증 연구가 수행되었는데, 해당 연구는 BERT 기반 분류기를 통해 90% 이상의 탐지 성능을 달성하고 실제 차단 사례로 이어졌음이 보고되었다[10].

이처럼 기존 연구는 주로 텍스트·트랜잭션 데이터를 독립적인 입력으로 간주하여 분류하는 모델 개발에 집중되어 있으며, 플랫폼 특성에 따라 다양한 탐지 기법이 제시되어 왔다. 그러나

오픈 소스 인텔리전스(Open-Source Intelligence, OSINT) 데이터 기반 텔레그램 메시지와 한국어 웹사이트 콘텐츠처럼 이질적이고 구조화되지 않은 데이터를 지식그래프(Knowledge Graph) 형태로 재구성한 뒤, 이를 그래프 신경망(Graph Nueral Network, GNN)을 통해 통합적으로 분석하는 연구는 아직 제한적으로만 보고되고 있다. 이러한 점에서, 본 연구가 제안하는 키워드 기반 지식그래프-GNN 결합 프레임워크는 기존 텍스트 중심 접근을 보완하고, 복잡하게 변화하는 사이버 사기 유형을 보다 구조적으로 탐지·분류할 수 있는 새로운 방향성을 제시한다.

1.3. 그래프 기술의 적용 및 특화 인공지능의 필요성

최근 사이버 범죄의 상당수가 텔레그램·카카오톡과 같은 SNS 채널과 각종 웹사이트를 통해 이루어지면서, 이들 플랫폼에 남는 메시지·게시물·HTML 콘텐츠가 사이버 수사에서 중요한 단서이자 핵심 데이터 자원으로 부상하고 있다. 그러나 이러한 텍스트 기반 OSINT는 통상 문서 단위의 키워드 빈도나 단순 통계로만 요약되는 경우가 많아, 어떤 용어들이 어떤 맥락에서 함께 등장하는지, 특정 범죄 수법을 둘러싼 의미적·맥락적 구조가 어떠한지를 체계적으로 파악하기 어렵다.

텍스트를 기반으로 한 지식그래프는 이러한 한계를 완화하기 위한 현실적인 출발점이 될 수 있다. 예를 들어 텔레그램 채널 내 메시지나 웹사이트 본문에서 핵심 키워드를 추출한 뒤, 이들 키워드의 공출현(Co-occurrence) 관계를 노드-엣지로 표현하면, 개별 문서를 넘어 “어떤 키워드들이 함께 뭉쳐서 하나의 범죄 수법·주제를 이룰 때가 많은지”, “어떤 용어가 특정 범죄 유형에서만 자주 등장하는지”와 같은 구조적 패턴을 그래프 형태로 분석할 수 있다. 이는 단순 빈도 기반 통계에서는 잘 드러나지 않는 키워드 군집, 하위 토픽, 수법별 특유의 어휘 조합을 보다 명시적으로 드러나게 해 준다.

또한, 키워드 지식그래프는 시간·채널·도메인 별로 발생하는 텍스트 패턴의 차이를 비교·분석하는 데에도 유용하다. 예를 들어 범죄 도메인 집단에서 구축한 키워드 공출현 그래프와 정상 도메인 집단에서 구축한 그래프를 비교하면, 두 집단에서 유사하게 등장하는 일반적 용어와, 범죄 집단에서만 응집적으로 나타나는 특이 키워드 집합을 구조적으로 구분해낼 수 있다. 텔레그램 채널에 대해서도 마찬가지로, 서로 다른 채널 간에 공유되는 키워드 패턴이나, 특정 범죄 유형에서만 나타나는 국지적 구조(Local structure)를 분석함으로써, 표면적으로는 비슷한 문장·레이아웃을 사용하는 “위장 정상” 채널과 실제 정상 채널을 구분하는 단서 역시 얻을 수 있다.

이처럼 그래프 구조로 표현된 사이버 위협 데이터를 효과적으로 학습하기 위해 GNN 기반 인공지능 모델이 활용될 수 있다. GNN은 노드와 엣지로 구성된 정형 그래프 구조를 직접 입력으로 받아, 인접 노드의 정보를 반복적으로 집계·전파하는 방식으로 표현을 학습한다. 최근에는 네트워크 트래픽 흐름을 그래프로 모델링해 침입 여부를 분류하거나, 공격 그래프 상에서 이상 행위를 탐지하는 등 다양한 사이버 보안 문제에 GNN을 적용한 연구들이 제안되었으며, 여러 데이터셋에서 기존 CNN·RNN 기반 모델과 동등하거나 더 우수한 성능이 보고되고 있다[11].

1.4. 본 연구의 목표

본 연구의 궁극적인 목표는 텔레그램 채널 메시지와 웹사이트 HTML과 같은 텍스트 기반 OSINT 데이터를 지식그래프-GNN 기반으로 자동 분석하여, 범죄 관련성을 빠르고 정확하게

판별할 수 있는 분류 프레임워크를 제시하는 데 있다. 이를 위해 텍스트에서 핵심 키워드를 추출하고, 키워드 공출현 관계를 그래프로 구조화한 뒤, 해당 그래프를 입력으로 하는 GNN 모델을 통해 웹사이트 도메인의 범죄 관련 여부 및 텔레그램 채널의 범죄 유형을 다중 분류하는 것을 기술적 목표로 설정한다.

II. 관련 연구

2.1. 지식그래프 구축 기술과 응용 사례

지식그래프는 현실 세계의 개체와 그 관계를 그래프 형태로 표현한 지식베이스로, 최근 인공지능·검색·추천 시스템을 지탱하는 핵심 인프라로 자리 잡았다. Google이 2012년 “things, not strings”라는 슬로건과 함께 검색 품질 향상을 위해 지식 그래프를 도입한 이후, 웹 상의 인물·장소·사물·조직 등에 대한 사실 정보를 그래프 형태로 통합하여 질의응답·자동 완성·지식 패널 등에 활용하는 것이 대표적인 성공 사례로 널리 알려져 있다[12]. 이후 다양한 연구들은 지식그래프를 단순한 데이터 저장소를 넘어, 표현 학습, 추론, 시맨틱 검색, 추천, 설명 가능한 AI 등을 위한 도구 등으로 활발히 활용하고 있다.

지식그래프 구축 기술은 크게 (1) 구조화 데이터 기반 통합, (2) 비정형 텍스트 기반 정보 추출, (3) 대형 언어모델(LLM)을 활용한 하이브리드 방식으로 발전해 왔다. 먼저, 기존의 관계형 데이터베이스(Relational Database, RDB), 온톨로지, 도메인별 표준 코드 체계인 CVE(Common Vulnerabilities and Exposures), ICD(International Classification of Diseases) 등에 존재하는 구조화·반구조화 데이터를 RDF(Resource Description Framework) 트리플이나 속성 그래프로 매핑하는 방식은 비교적 안정적인 품질의 지식그래프를 빠르게 구축할 수 있어 초기부터 널리 사용되었다. 한편, 대규모 문헌·웹 문서에서 개체와 관계를 자동으로 추출하는 정보 추출(Information Extraction, IE), 관계 추출(Relation Extraction), 오픈 정보 추출(Open IE) 기술을 활용하여, 지질학·과학 문헌·뉴스 기사 등 비정형 텍스트로부터 직접 지식그래프를 구축하려는 연구도 활발히 진행되어 왔다[13]. 최근에는 2022년 이후의 연구를 정리한 서베이를 중심으로, 이러한 전통적인 규칙 기반·통계적 방법 위에 LLM을 이용한 지식 추출·보완을 결합하여 보다 유연하고 범용적인 지식그래프 구축 파이프라인을 설계하려는 시도가 주목받고 있다[14].

이처럼 선행 연구들은 다양한 데이터원을 그래프로 구조화하고, 그 위에서 표현 학습·추론·응용 서비스까지 연결하는 전체 파이프라인을 중심으로 기술을 고도화해 왔다. 본 연구에서 활용하는 텍스트 기반 키워드 지식그래프 역시 이러한 지식그래프 구축 기술의 세부 분야로 볼 수 있으며, 기존의 대규모 일반 도메인·전문 도메인 지식그래프와 달리, 특정 OSINT 텍스트 코퍼스를 대상으로 한 지식그래프 표현에 초점을 둔다는 점에서 차별성을 갖는다.

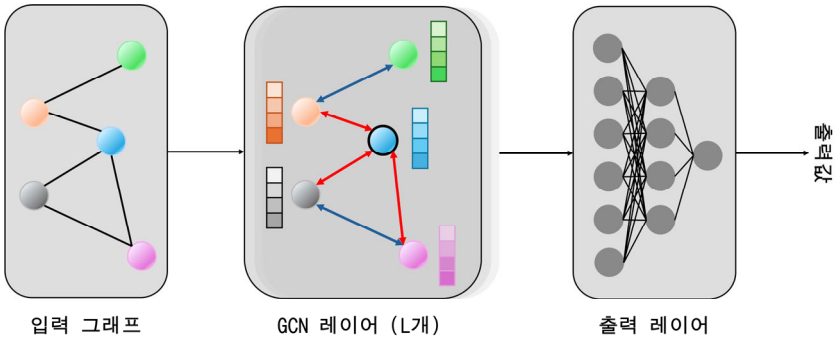
2.2. GNN 기술의 발전 동향

그래프 신경망은 노드와 엣지로 구성된 그래프 구조를 직접 입력으로 받아, 인접 노드 정보를 반복적으로 집계·전파하는 방식으로 표현을 학습하는 딥러닝 패러다임이다. 초기에는 순환 신경망을 그래프에 확장한 형태나 스펙트럴 도메인(spectral domain)에서의 그래프 컨볼루션 이론을 바탕으로 한 모델들이 제안되었고, 이후 Kipf & Welling의 Graph Convolutional Network(GCN) 등 반정형 그래프에서의 반지도 학습(semi-supervised learning)을 위한 단순·효율적인 구조가 등장하면서, GNN은 비유클리드(non-Euclidean) 데이터에 대한 대표적

딥러닝 기법으로 빠르게 자리 잡았다. 이러한 발전 과정을 체계적으로 정리한 여러 기존 연구에서는, GNN이 그래프 상의 메시지 패싱(message passing)을 통해 구조적 의존성을 모델링하는 공통된 틀을 공유하면서도, 구현 방식에 따라 다양한 아키텍처로 분화되어 왔음을 보여준다[15].

모델 구조 관점에서, GNN 연구는 주로 스펙트럴 기반 GNN과 스패셜(spatial) 기반 GNN이라는 두 축을 중심으로 전개되어 왔다. 스펙트럴 접근은 그래프 라플라시안의 고유분해를 이용해 그래프 푸리에 변환을 정의하고, 주파수 도메인에서 컨볼루션을 수행하는 이론적 토대를 제공하는 반면, 스패셜 접근은 각 노드의 이웃 정보를 직접 집계하는 형태로 설계되어 확장성과 구현의 단순성을 갖춘다. 이후 그래프 컨볼루션에 주의(attention) 메커니즘을 결합한 GAT 계열, 샘플링 전략을 통해 대규모 그래프에서 효율적으로 학습하는 GraphSAGE 계열, 그래프 풀링·읽기(readout) 연산을 도입한 그래프 수준 분류 모델 등 다양한 변형이 제안되었다. 최근 서베이들은 이러한 모델들을 스펙트럴/스패셜, 메시지 패싱 구조, 학습 목적 등에 따라 분류하고, 이론적 표현력과 한계(oversmoothing, oversquashing 등)에 대한 분석을 제공한다[16].

또한, 이종 그래프(heterogeneous graph), 동적 그래프(dynamic graph), 시공간 그래프(spatio-temporal grpah) 등 보다 복잡한 구조로 GNN을 확장하는 연구도 활발하다. 다양한 타입의 노드와 관계를 가진 이종 그래프에 대해서는, 타입별 변환과 관계별 집계 함수를 설계한 Heterogeneous GNN(HGNN) 계열 모델과 이를 종합적으로 비교 정리한 연구가 다수 수행되었으며, 이러한 모델들이 추천, 소셜 네트워크, 바이오 네트워크 등에서 동종 그래프 기반 GNN보다 더 강력한 표현력을 보인다는 결과가 보고되고 있다[17].



<Figure 1> Example of a graph convolutional network as a basic GNN model.

2.3. OSINT 데이터 기반 사이버범죄 탐지 및 분류에 대한 동향

OSINT는 웹·소셜미디어·포럼·코드 저장소·도메인/WHOIS 정보 등 공개 데이터를 수집·분석하여 위협 정보를 도출하는 접근으로, 최근 사이버 위협 인텔리전스(CTI)와 침해사고 대응에서 핵심 축으로 자리 잡고 있다. 여러 연구와 산업 보고서는 OSINT가 다양한 공개 소스를 연계 분석함으로써 공격자의 전술·기법·절차(TTPs)를 파악하고, 위협 탐지와 대응 속도를 향상시키는 데 기여한다고 평가한다[18]. 동시에, 데이터 규모·신뢰도·법·윤리 이슈 등 OSINT 특유의 한계도 함께 지적되고 있어, 자동화된 수집·정제·분석 기술과 AI 기반 분류·탐지 기법에 대한 수요가 지속적으로 증가하고 있다.

소셜미디어·메신저·다크웹 등 비전통적 플랫폼을 대상으로 한 OSINT 기반 사이버범죄 연구도 활발하다. 최근 관련 연구에서는 트위터, 페이스북, 텔레그램, 디스코드, 다크웹 포럼 등이 악성코드 유통, 피싱 키트 판매, 데이터 유출 공유, 금융 사기 지침 공유 등 다양한 범죄 활동의

허브로 사용되고 있음을 보고하며, 이러한 채널에서의 텍스트·링크·이미지·사용자 행태를 분석하는 AI·머신러닝 기반 포렌식 연구 동향을 정리하였다[19]. 텔레그램에 특화된 연구로는 12만 개 이상 채널과 4억 개 이상의 메시지를 수집한 TGDataset과 같이, 대규모 채널 생태계를 포괄적으로 수집·특성화하여 스팸·선전·불법 거래·극단주의 등 다양한 주제를 분석할 수 있도록 한 벤치마크 데이터셋 구축 연구가 제안되었다[20].

웹사이트와 도메인 수준에서는 OSINT를 활용한 악성 도메인·피싱 사이트 탐지가 오래전부터 연구되어 왔다. 초기 연구들은 URL 패턴, 도메인 길이, 문자 구성, 호스팅 정보 등 정적 특징에 기반한 머신러닝 분류기를 구축했으나, 최근에는 DNS 로그, WHOIS·BGP·지리 정보, 검색엔진·트래픽 메트릭, HTML 구조·콘텐츠, 심지어 시각적 스크린샷까지 통합하는 멀티모달·OSINT 확장 특징을 활용하는 방향으로 발전하고 있다[21].

III. OSINT 데이터 기반 지식그래프 및 학습데이터 구축

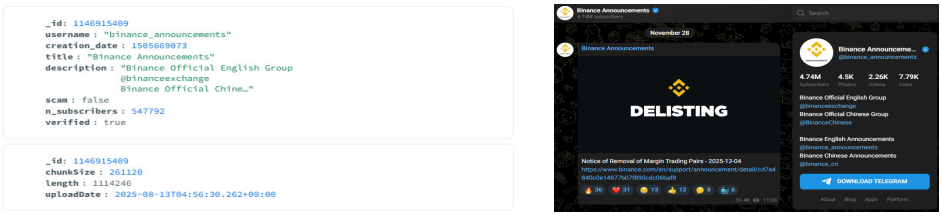
3.1. OSINT 데이터 수집

본 연구에서는 지식그래프-GNN 기반 범죄 분류 모델을 학습하기 위해, 텍스트 기반 OSINT 자료로부터 텔레그램 채널 데이터셋과 웹사이트 도메인/HTML 데이터셋 총 두 종류의 데이터셋을 구축하였다. 텔레그램 데이터셋의 경우, 공개 벤치마크 데이터셋인 TGDataset을 활용하여 범죄와 연관된 채널과 비교 대상이 되는 일반 채널을 선별하고, 각 채널의 메시지 이력을 수집하였으며, 사이버범죄 분류체계 기준을 참고하여 범죄 유무 및 범죄 유형 라벨을 부여하였다. 웹사이트 데이터셋은 범죄 연관 사이트와 정상 사이트로 구분하여 구축하였다. 범죄 연관 웹사이트의 도메인 후보는 피해 신고 플랫폼인 더치트(thecheat)에 게시된 신고 목록에서 도메인 주소를 추출하여 구성하였으며, 정상 웹사이트의 경우 Tranco 패키지를 통해 수집한 상위 트래픽 도메인 목록을 사용하였다.

3.1.1. TGDataset 기반 데이터셋 구축

TGDataset은 공개 텔레그램 채널의 스냅샷을 대규모로 수집한 공개 데이터셋으로, 120,979개의 채널과 4억 건 이상의 메시지로 구성된, 현재까지 알려진 것 중 가장 큰 규모의 텔레그램 채널 데이터셋이다. 각 채널에 대해 채널 ID, 생성 시각, 사용자명, 제목과 설명, 가입자 수와 같은 메타데이터뿐 아니라, 텔레그램이 부여한 사기 의심 채널 여부를 뜻하는 scam 플래그와 공식 계정 여부를 뜻하는 verified 플래그 정보, 그리고 채널에 게시된 메시지 이력을 함께 제공한다. 이러한 구조 덕분에 TGDataset은 텔레그램 생태계 전반에서 나타나는 정상·악성 채널의 분포와 행태를 동시에 분석할 수 있는 기반을 제공한다.

본 연구에서는 TGDataset에 포함된 scam 및 verified 메타 정보를 활용하여, 범죄와 연관된 텔레그램 채널과 정상적인 공식 채널을 자동으로 구분하였다. 구체적으로는, ‘scam = true’로 표시된 채널을 범죄 관련 채널로, ‘verified = true’로 표시된 채널을 정상 공식 채널로 간주하여 비교 집단을 구성하였다. 각 채널에 대해 TGDataset이 제공하는 메시지 본문을 모두 수집한 뒤, 채널 단위로 메시지를 통합하여 이후 텍스트 전처리 및 키워드 추출, 지식그래프 구축에 활용하였다.



<Figure 2> TGDataset channel metadata (left) and the corresponding Telegram channel (right).

TGDataset에서 제공하는 verified 및 scam 플래그가 모두 false인 일반 채널의 경우, 범죄 연관성이 불명확하고 정상·악성의 기준이 모호하다고 판단하여 양질의 학습 환경을 확보하기 위해 데이터셋에서 제외하였다. 또한 채널 내 메시지 개수가 10개 미만인 채널은 충분한 텍스트 정보가 존재하지 않아 통계적 특성과 언어 패턴을 안정적으로 학습하기 어렵다고 보아 분석 대상에서 제거하였다.

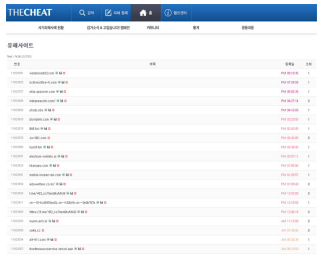
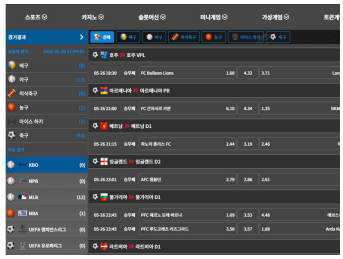
3.1.2. 범죄 관련 웹사이트 데이터셋 수집

범죄 연관 웹사이트 데이터셋은 국내 사기 피해 정보 공유 플랫폼인 더치트를 기반으로 구축하였다. 더치트는 2006년부터 운영되고 있는 금융·인터넷 사기 피해 정보 공유 서비스로, 피해자가 직접 사기 거래 사례와 함께 계좌번호, 전화번호, 도메인 등 가해자 식별 정보를 신고·등록하면, 다른 이용자가 거래 전에 해당 정보를 조회하여 사기 이력을 확인할 수 있도록 지원하는 플랫폼이다[22]. 특히 중고거래·온라인 거래 사기를 중심으로 국내에서 가장 큰 규모의 사기 정보 데이터베이스를 유지하고 있으며, 금융사기 예방 및 2차 피해 방지를 위한 대표적인 민간 서비스로 활용되고 있다.

본 연구에서는 더치트에 신고·공유된 피해 사례 중 불법·사기 웹사이트와 연관된 신고 목록을 대상으로, 게시글 본문과 메타데이터에서 도메인 주소를 추출하여 범죄 연관 웹사이트 도메인 후보군을 구성하였다. 이후 각 도메인에 대해 크롤러를 이용해 웹사이트의 HTML 문서를 수집하였다. 수집된 HTML 문서는 이후 텍스트 기반 OSINT로 활용하기 위해 별도의 전처리 파이프라인을 통해 평문 텍스트 코퍼스로 변환하였다. 먼저, 각 HTML 파일은 원시 바이너리 형태로 읽은 뒤, <meta charset="…"> 태그를 우선적으로 탐색하여 명시된 문자 인코딩을 추출하고, 해당 정보가 존재하지 않는 경우에는 자동 인코딩 추정 라이브러리 chardet을 이용하여 인코딩을 추정한 후 이를 기반으로 문자열로 디코딩하였다. 디코딩 과정에서 오류가 발생하는 경우에는 UTF-8을 기본값으로 사용하여 손상된 문자를 대체하는 방식으로 처리함으로써, 다양한 인코딩을 사용하는 웹페이지에서도 가능한 한 많은 텍스트 정보를 보존하도록 하였다.

디코딩된 HTML 문자열은 Python 기반 HTML 파서 BeautifulSoup을 이용해 파싱하였으며, 이 과정에서 <script>, <style>, <noscript>와 같이 실제 페이지 표시에는 관여하지 않거나 분석에 불필요한 태그들은 제거하였다. 이후 DOM 구조에서 눈에 보이는 텍스트 노드만 추출하고, 문단·블록 단위 구분을 유지하기 위해 구분자(|)를 사이에 두고 하나의 문자열로 합쳐 텍스트를 구성하였다. 최종적으로 구축된 HTML 텍스트 코퍼스는 토큰화·키워드 추출 및 키워드 공출현 그래프 생성의 입력으로 사용되며, 범죄 사이트에 해당하는 더치트 기반 도메인 집합과 정상 사이트에 해당하는 Tranco 기반 도메인 집합을 병합하여 웹사이트 도메인의 범죄/정상 이진 분류 모델을 학습하기 위한 학습 데이터로 활용하였다.

<Table 1> Example of data within HTML text corpus creation workflow

더치트 신고목록	범죄 관련 사이트(사이버도박)	HTML 텍스트 코퍼스
		원벳원 > 베팅 > 라이브벳 1 페이지|한업광고|×|오늘하루 열지 않기|닫기|한업광고|×|오늘하루 열지 않기|닫기|한업광고|×|오늘하루 열지 않기|닫기|POUP|로그인|×|USERN AME|PASSWORD|LOGIN|SIGN UP|TELEGRAM|CLOSE|OUR SERVICES|CLOSE|HEADER|SPORTS|>|크로스|파셜|실시간|라이브|스포츠|월한스|B) CREATOR|E-SPORTS|CASINO|SLOT|MINI GAME|>|보글보글|1분게임|2분 게임|3분 게임|슈퍼마리오|1분게임|2분게임|3분게임|룰렛|1분게임|2분게임|3분 게임|바카라게임|로투스바카라|로투스식보|로투스홀짝|MGM바카라|스카이파크|바카라|홀짝젝|스피드바카라|드래곤타이거|룰렛|BTC파워볼|1분게임|2분게임|3분게임|4분게임|5분게임|BET GAME|>|BET 365 가상 스포츠|축구|농구|미식축구|테니스|크리켓|다트|BET 365 가상 레이스|개경주|말경주|자동차경주|오토바이 경주|싸이클경주|마차경주 (GRAPH|MORE)|출석|쿠폰|룰렛|고객센터|배팅내 역|지 인추천이벤트|이용규정|입금하기|출금하기|충계티비|도메인|트위터 (MESSENGER|LOGIN|OR|SIGN UP|Main Visual|LIVE SPORTS|라이브 영상|홀짝|라이브 중영상 준비중입니다.|ALL|SOCCER|BASEBALL|BASKETBALL|VOLLEYBALL|HOCKEY|UFC|LUGBY|TENNIS|ESPORTS|BOOKMARK|<|>|시 간순|>|시 간순|리그순 |OUR SERVICES|VIEW ALL|+|LIVE CASINO|VIEW ALL|+|SLOT GAMES|VIEW ALL|+|FOOTER

3.1.3. 정상 웹사이트 데이터셋 수집

정상 웹사이트 데이터셋은 범죄 연관 웹사이트와 대비되는 범죄와 관련 없는 정상 도메인 집합을 구성하기 위해, 웹 트래픽 기반 도메인 랭킹 리스트인 Tranco를 활용하여 구축하였다. Tranco는 여러 상이한 소스의 트래픽 데이터를 기반으로 도메인 인기 순위를 산출하는 연구용 리스트로, 단기적인 변동이나 인위적인 조작에 덜 민감하고, 보안·프라이버시 연구에서 재현 가능한 도메인 샘플링 기준을 제공한다는 점에서 널리 사용된다[23]. 본 연구에서는 Python용 Tranco 패키지를 이용하여 최신 Tranco 리스트로부터 상위 트래픽 도메인들을 자동으로 수집하였으며, 이를 일반적인 정상 웹사이트의 대표 표본으로 간주하였다. 수집된 각 도메인에 대해 3.1.2절의 범죄 연관 웹사이트와 동일한 방식으로 원본 HTML 문서를 수집한 뒤, 동일한 전처리 파이프라인(인코딩 판별 및 디코딩, HTML 파싱, 불필요 태그 제거, 가시 텍스트 추출)을 적용하여 텍스트 코퍼스로 변환하였다.

이와 같이 Tranco 상위 도메인 목록에서 수집·전처리된 HTML 텍스트는 범죄와 직접적인 관련성이 없는 정상인 웹사이트 집합으로 간주하였다. 최종적으로, 더치트 기반의 범죄 연관 도메인과 Tranco 기반의 정상 도메인을 통합하여 웹사이트 도메인 단위의 범죄 또는 정상 이진 분류를 위한 학습 데이터셋을 구성하였으며, 이후 지식그래프-GNN 기반 분류 모델의 범죄 클래스와 반대되는 정상 클래스로 활용하였다.

3.2. 범죄 유무 및 범죄 유형 라벨링 기준

3.2.1. TGDataset 기반 텔레그램 채널 범죄 유형 라벨링

본 연구에서는 텔레그램 채널의 범죄 유무 및 범죄 유형 라벨을 영어 채널을 대상으로 LDA(Latent Dirichlet Allocation) 기반 토픽 모델링을 수행하여 각 채널에 대해 주요 토픽을 추정하여 얻어진 TGDataset 내 채널-토픽 매핑 결과를 바탕으로 정의하였다. 먼저 3.1.1절과 같이, TGDataset에서 scam = true로 표시된 채널은 범죄 연관 채널로, verified = true로 표시된 채널은 정상 공식 채널로 간주하여 이진 수준의 범죄 유무 라벨을 부여하였다. 이후 범죄 연관 채널에 대해서는 LDA 기반 토픽 정보를 ECRM(European Cybercrime Research Manual)의 사이버범죄 분류 체계와 의미적으로 가장 가까운 범주에 매핑함으로써 채널 단위의 범죄 유형 라벨을 정의하였다. 사이버범죄 도메인 지식이 제한적인 상황에서 수천 개 채널을 개별적으로 직접 열람·판단하는 것은 연구 인력과 시간 측면에서 비현실적이었기 때문에, 본 연구에서는 TGDataset의 토픽 라벨과 ECRM 사이버범죄 분류 체계를 결합한 이러한 준자동 라벨링 방식을 채택하였다.

구체적으로, 카드 정보 거래나 금융사기와 직접적으로 연관된 'Carding' 토픽과, 사기성 광

고·투자 사기와 연관된 'US news', 'Crypto', 'Indian edu' 토픽은 모두「사이버 사기」범주로 통합하였으며, 악성코드 유포, 크랙·패치 공유, 비인가 소프트웨어 변조와 연관된 'videogame modding' 및 'Software' 토픽의 경우「악성프로그램」범주로 통합하였다. 음란물·불법 성 콘텐츠 유통과 관련된 'Porn' 토픽은 「사이버성범죄」 범주로 라벨링하였다. 이와 같이 TGDataset의 토픽 레이블을 기준으로, 최종적으로는「사이버 사기」-「악성프로그램」-「사이버성범죄」의 세 가지 주요 범죄 유형 클래스를 구성하였다.

3.2.2. 웹사이트 데이터셋 범죄 유무 라벨링

웹사이트 데이터셋은 3.1.2절의 더치트 기반 도메인 집합과 3.1.3절의 Tranco 기반 도메인 집합을 통합하여 구성하였으며, 라벨링은 도메인 단위의 범죄 연관 여부 이진 분류를 기준으로 수행하였다. 구체적으로, 더치트에 사기 피해 사례와 함께 신고·게시된 도메인들은 실제 피해 신고를 통해 수집된다는 점에서 범죄 연관성이 높은 사이트(범죄 사이트)로 간주하여 "crime" 라벨을 부여하였다. 반대로, Tranco 패키지를 통해 수집한 상위 트래픽 도메인들은 일반적인 웹 이용 환경에서 자주 방문되는 정상 사이트로 보고 "normal" 라벨을 부여하였다.

웹사이트의 경우에도 각 도메인별 HTML 스냅샷과 텍스트 내용만으로는 구체적인 범죄 수법이나 목적(예: 금융사기, 악성코드 유포, 사이버성범죄 등)을 일관되게 판별하기 어렵고, 사이버범죄 도메인 지식의 부족과 대규모 사이트를 일일이 검토할 인력·시간의 제약이 존재하였다. 따라서 본 연구에서는 웹사이트 데이터셋에 대해 세부 범죄 유형까지 라벨링하기보다는, 범죄 관련 사이트 여부만을 구분하는 이진 수준의 라벨링으로 범위를 제한하였다.

3.3. 지식그래프 스키마 설계 및 그래프 생성

본 연구의 지식그래프는 텍스트에서 추출한 키워드 중심 노드로 설계하였다. 텔레그램 채널과 웹사이트라는 두 데이터 소스에 대해 공통적으로, (1) 형태소/토큰 단위로 텍스트를 분해한 뒤, (2) 노이즈가 많은 단어를 제거하고, (3) 빈도와 공출현·유사도 정보를 기준으로 상위 핵심 키워드만 노드로 전략을 사용하였다. 이러한 노드 선택 기준을 통해 그래프 규모를 불필요하게 키우지 않으면서도, 각 채널/사이트의 범죄적 특성을 잘 드러내는 단어들만 남기도록 하였다.

3.3.1. 노드 구성 정의

TGDataset 기반 텔레그램 그래프에서는 각 채널(uid)을 하나의 그래프 단위로 보고, 그 안에 포함된 메시지에서 영어 토큰을 추출하여 노드로 사용하였다. 이를 위해 먼저 메시지 문자열을 리스트 형태로 정규화하고, RegexpTokenizer를 이용해 알파벳·숫자 조합만 남기도록 토큰화한 뒤, 불용어와 한 글자 토큰, 숫자·기호 위주의 토큰을 제거하였다. 이후 WordNet 기반 표제어 추출을 적용해, 파생형이 많은 단어들도 하나의 어근으로 묶이도록 하여 노드 수를 줄이면서 의미적 일관성을 높였다.

한편 웹사이트 그래프에서는 각 도메인의 단위로 HTML에서 추출된 한국어 텍스트를 대상으로 명사 중심 키워드 노드를 구성하였다. 먼저 더치트·Tranco에서 수집한 각 도메인의 텍스트를 | 구분자를 기준으로 문장/구를 나눈 뒤, 숫자만으로 구성된 문자열과 순수 기호 문자열은 제외하였다. 각 구문에 대해 Kkma 형태소 분석기를 사용하여 명사만 추출한 뒤, 불용어 및 의미 없는 토큰을 제거하였다. 더 나아가, 어떤 단어가 다른 단어에 완전히 포함되는 경우(예: "게임", "온라인게임") 짧은 쪽을 제거하여, 보다 구체적인 복합 명사가 남도록 필터링하였다.

3.3.2. 엣지 구성 정의

본 연구에서 엣지는 키워드 노드 간 관계 구조를 표현하는 핵심 요소로, 공출현 기반 엣지와 코사인 유사도(cosine similarity) 기반 엣지 두 가지로 구성하였다. 데이터 소스의 특성에 따라 (1) 텔레그램 채널 그래프에는 공출현 엣지와 의미 유사도 엣지를 각각 적용하였고, (2) 웹사이트 그래프에는 의미 유사도 엣지만 적용하였다. 이는 HTML 구조 특성상 웹페이지 텍스트에서 신뢰할 만한 문장·메시지 단위를 정의하기 어렵기 때문에, 웹사이트 데이터에서는 공출현 관계가 실제 의미적인 "문맥 공존"이라기보다 레이아웃·템플릿 구조를 반영하는 경우가 많다는 점을 고려하였다.

공출현 빈도 엣지 : 텔레그램 채널 지식그래프에서는 각 채널 내의 메시지를 기준으로 하여, 메시지 안에서 함께 등장하는 키워드 쌍을 공출현 엣지로 정의하였다. 먼저 채널별로 메시지 리스트를 정규화한 뒤, 토큰화·정제 과정을 거쳐 얻은 토큰 시퀀스에 대해, 한 메시지 안에서 등장하는 모든 키워드쌍을 생성하였다. 이때 동일한 메시지 안에서 같은 쌍이 여러 번 등장하더라도 한 번만 카운트하도록 하여, 스팸성 반복보다는 "같은 메시지 안에 함께 출현했는가"라는 공존 여부를 강조하였다. 이에 따라 텔레그램 그래프의 엣지는 "같은 채널의 메시지 안에서 자주 함께 등장하는 키워드들 사이의 연결"을 나타내며, 그 강도는 공출현 빈도와 정규화 가중치로 표현되어 엣지 가중치(edge weight)로 사용되었다.

언어 유사도 엣지 : 의미 유사도 엣지는 텔레그램과 웹사이트 그래프 모두에 공통으로 적용되는 관계 구조로, 사전학습 언어모델을 통해 산출된 키워드 임베딩 벡터 간 코사인 유사도를 이용하여 키워드 간의 의미적 근접성을 표현한다. 먼저 각 노드 후보로 선정된 채널 및 도메인별 고빈도 명사 키워드에 대해 한국어·영어 문맥을 반영한 사전학습 언어모델을 적용하여 임베딩 벡터를 계산한다. 이후 모든 키워드 쌍에 대해 코사인 유사도를 산출하여 엣지를 생성하였으며, 이렇게 생성된 엣지에는 두 키워드 임베딩 간 코사인 유사도 값을 엣지 가중치 속성으로 부여하여, 값이 클수록 두 키워드의 의미적 연관성이 더 강함을 나타내도록 하였다.

3.4. 학습데이터 통계 및 특성

최종 텔레그램 학습 데이터셋은 총 223개 채널로 구성되었으며, 포함된 메시지 수는 857,623개이고 채널당 평균 메시지 수는 약 3,846개였다. 이 중 정상 채널의 수는 115개이며, 범죄 관련 채널 수는 유형별로 각각 사이버 사기 88개, 악성프로그램 15개, 사이버 성범죄 5개로 구분되었으며, 평균 그래프 노드 수는 18.12개로 형성되었다. 한편 최종 웹사이트 학습 데이터셋은 총 1120개의 도메인으로 구성되어 있으며, 범죄와 연관되지 않은 도메인 677개와 범죄와 연관된 도메인 443개로 구분되었다. 또한 평균 그래프의 노드 수는 19.33개로 형성되었다.

IV. GCN 기반 사이버범죄 분류 모델 설계

4.1. 노드·엣지 특성 정의 및 입력 표현

앞서 텔레그램 기반 TGDataset 그래프와 웹사이트 그래프의 노드 및 공출현·유사도 기반 엣지 구성 방식을 정의하였다. GNN은 정수 형태의 노드·엣지 정의만으로는 동작할 수 없기 때문에, 각 노드를 의미적으로 표현하는 노드 피처(feature)를 위한 연속형 벡터와 유사도 기반 엣지의 관계 강도를 계산하기 위한 임베딩 기반 특성이 필요하다. 반면 공출현 엣지는 메시지 내 공동 등장 여부만으로 정의되므로 별도의 임베딩 모델을 요구하지 않는다. 따라서 본 연구에서

는 구축된 지식그래프를 GNN 학습이 가능한 입력 텐서로 변환하기 위해 임베딩 모델 선택 → 노드·엣지 특성 계산 → Cypher-to-Tensor 변환 절차를 적용하였다. 또한 텔레그램과 웹사이트 데이터는 언어적 특성과 단어 분포가 상이하므로, 두 데이터셋에 대해 서로 다른 사전학습 언어모델을 구분하여 적용하였다.

4.1.1. TGDataset 기반 텔레그램 그래프의 임베딩 모델 적용 전략

텔레그램 데이터는 영어 기반의 짧고 파편적인 메시지로 구성되어 있으며, 단어 단위 토큰 구조가 명확하다. 이에 단어 간 의미적 유사도를 안정적으로 반영하기 위해 MEN Test Collection(Bruni et al., 2012) 을 활용하여 영어 임베딩 모델을 평가하였다. MEN은 총 3,000개의 단어쌍과 인간 유사도 점수로 구성되며, 모델이 산출한 코사인 유사도와 인간 점수 간 Spearman 상관계수(ρ) 로 임베딩 품질을 비교할 수 있다.

FastText, Sentence-BERT, Dense Retrieval 기반 모델 등 총 3종을 비교한 결과, FastText가 가장 높은 성능(ρ = 0.8427) 을 기록하였다(Sentence-BERT: 0.7643, Dense Model: 0.6811). 이에 따라 텔레그램 그래프의 유사도 기반 엣지 가중치는 FastText 임베딩 cosine similarity로 정의하였다. 한편 노드 임베딩은 단어 수준 의미 유사도보다 더 풍부한 문맥·형태 정보를 반영할 필요가 있어, WordPiece 기반 subword 표현이 가능한 bert-base-uncased 모델을 사용하여 생성하였다.

4.1.2. 웹사이트 그래프의 임베딩 모델 적용 전략

웹사이트 데이터는 한국어 기반 명사 중심의 키워드로 구성되므로, 한국어 명사 간 의미적 유사도를 안정적으로 표현할 수 있는 임베딩 모델 선정이 필요하다. 본 연구에서는 MEN Dataset이 영어 기반이라는 점을 고려하여 이를 한국어로 직접 번역한 MEN-KR 벤치마크를 구축하였고, 원본 MEN 점수(0-50)를 0-1 구간으로 Min-max 정규화하여 평가에 활용하였다.

이후 monologg/kobert, skt/kobert-base-v1, klue/bert-base, monologg/koelectra-base-v3-discriminator, klue/roberta-base, tunib/electra-ko-base, kakaobank/kf-deberta-base 등 총 7종의 한국어 사전학습 언어모델을 MEN-KR에서 비교하였다. 실험 결과, kakaobank/kf-deberta-base가 가장 높은 평균 Spearman 상관계수(Mean ρ = 0.573) 를 기록하였으며, 이는 한국어 명사 간 의미적 연관성을 가장 안정적으로 반영함을 의미한다. 따라서 본 연구에서는 웹사이트 그래프의 유사도 기반 엣지 가중치를 kf-deberta-base 임베딩을 바탕으로 정의하였으며, 노드 임베딩 또한 동일한 모델의 토크나이저 및 embedding layer를 사용하여 일관된 의미 공간에서 표현하였다.

4.1.3. 지식그래프 텐서화

본 연구에서 구축된 텔레그램 및 웹사이트 기반 지식그래프는 Neo4j Cypher 형식으로 저장되며, 이를 GNN의 학습에 활용하기 위해 PyTorch Geometric(PYG)이 요구하는 텐서 구조로 변환하였다. PYG의 단일 그래프 입력은 (1) 노드 임베딩 행렬, (2) 엣지 연결 구조, (3) 엣지 가중치, (4) 그래프 단위 레이블로 구성된다. 본 절에서는 Cypher 스키마로부터 해당 텐서들을 생성한 절차를 기술한다.

공출현 엣지 텐서화 : 공출현 기반 엣지는 텔레그램 그래프에서 메시지 단위의 문맥적 공존 관계를 반영하기 위해 사용되었다. Cypher 파일에는 각 메시지에서 함께 등장한 단어 쌍(s, t)

과 채널 전체에서의 누적 공출현 횟수가 weight 속성으로 기록되어 있으며, 본 연구에서는 이를 기반으로 텐서 형태의 엣지 정보를 구성하였다. 먼저 노드명을 정수 인덱스로 매핑한 뒤, 공출현 관계의 무방향성을 반영하기 위해 각 엣지를 양방향(s→t, t→s)으로 확장하여 edge_index 텐서를 생성하였다. 또한 Cypher에 저장된 공출현 가중치 값을 실수형 텐서로 변환하여 edge_attr에 저장함으로써, 엣지의 강도를 표현하였다.

유사도 엣지 : 텍스트 기반 의미 연관성을 반영하기 위해 두 데이터셋 모두 코사인 유사도 기반의 유사도 엣지를 포함하였으나, 데이터 구조의 차이로 인해 유사도 계산 대상 단어쌍 선정 방식은 서로 다르게 설계하였다. 텔레그램 메시지는 짧고 비정형적이며 노이즈가 많아 모든 단어쌍의 유사도를 계산하는 것은 비효율적이므로, 본 연구에서는 메시지 내에서 생성된 후보 단어쌍 중 유사도 상위 약 40개 쌍만을 선별하여 유사도 엣지로 포함하였다. 선별된 단어쌍은 Cypher 스키마에 코사인 유사도와 함께 저장되며, 텐서 변환 단계에서는 이를 그대로 edge_attr로 매핑하였다. 반면 웹사이트 텍스트는 도메인 단위의 비교적 안정적인 한국어 명사 중심 구조를 가지므로, 텔레그램과 달리 메시지 단위 후보 추출이 어렵다. 이에 각 도메인에서 빈도가 높은 약 20개 내외의 핵심 명사를 노드로 선정한 뒤, 이들 노드 간 코사인 유사도를 계산하여 미리 설정한 임계값을 만족하는 단어쌍을 유사도 엣지로 구성하였다. 해당 유사도는 kf-deberta-base 임베딩을 기반으로 산출되며, Cypher 스키마에 기록된 뒤 텐서 변환 시 edge_attr로 반영된다.

노드 임베딩: 두 데이터셋의 노드 피처는 동일한 절차로 생성하였다. 각 단어(영어·한국어 명사)에 대해 해당 언어의 사전학습 모델 토크나이저를 적용하여 WordPiece 기반 서브토큰(subword token)으로 분해하였다. 서브토큰은 의미적·형태적 정보를 세분화한 단위로, 예를 들어: 영어 단어 “international”은 inter, nation, al 등의 서브토큰으로, 한국어 단어 “금융범죄”는 금, 융, 범, 죄와 같이 여러 부분 단위로 분해된다. 분해된 서브토큰에 대해 embedding layer의 표현을 추출한 뒤, 모든 서브토큰 임베딩을 평균(pooling)하여 최종 768차원 단어 임베딩 벡터를 생성하였다. 이 방식은 단일 단어 벡터보다 더 세밀한 형태·문맥 정보를 반영하여 노드 피처의 표현력을 강화한다. 최종적으로 구축된 노드 피처 텐서는 $X \in R^{N \times 768}$와 같다.

4.2. 사이버범죄 분류를 위한 GCN 설계

본 연구에서는 4.1절에서 정의한 텔레그램·웹사이트 지식그래프를 입력으로, 그래프 합성곱 신경망(Graph Convolutional Network, GCN)을 이용해 그래프 단위 사이버범죄 분류를 수행하였다. 제안 모델은 Kipf & Welling의 GCN 구조를 기본 골격으로 하되, 다층 GCN 기반 주 경로(main path)와 입력 특성을 직접 반영하는 보조 경로(auxiliary path)를 병렬로 구성한 뒤, 두 경로의 출력을 결합하여 그래프-level 표현을 얻는 경량 멀티 패스 구조를 사용하였다.

각 그래프 G_i는 노드 특성 행렬 $X^{(0)} \in \mathrm{R}^{\mathrm{N_i} \times \mathrm{d}}$, 정규화된 인접 행렬 $\tilde{\mathrm{A}}$, 엣지 가중치(공출현 빈도 또는 임베딩 기반 유사도)를 포함하는 형태로 표현된다. 단일 GCN 층은 아래 메시지 패싱 형태로 정의되며, $H^{(0)} = X^{(0)}$, $W^{(l)}$는 학습 가능한 가중치 행렬이며, $\sigma(\cdot)$는 비선형 활성화 함수이다. 구현은 PyTorch Geometric 라이브러리의 GCNConv 모듈을 활용하였고, 엣지 가중치는 edge_attr로 전달하여 공출현·유사도 기반 엣지 강도를 반영하였다.

$$H^{(l+1)} = \sigma(\widetilde{\mathrm{AH}}^{(1)}\mathrm{W}^{(1)})$$

<Table 2> Layer configuration of the GCN-based cybercrime classification model

단계	경로	연산/레이어	입력 차원	출력 차원	활성화/정규화 /드롭아웃
0	공통(입력)	입력 노드 특성 : X(0)	N i × d	N i × d	-
1	메인	GCNConv_1	N i × d	N i × h	GELU, GraphNorm, Dropout(0.5)
2	메인	GCNConv_2	N i × h	N i × h	GELU
3	메인	GCNConv_3	N i × h	N i × (h/2)	GELU
4	보조	GCNConv_aux	N i × d	N i × (h/2)	GELU, Dropout(0.5)
5	결합	Concatenate (Hmain, Haux)	N i × (h/2), N i × (h/2)	N i × h	-
6	공통	Global Mean Pooling	N i × h	1 × h	-
7	공통(출력)	Linear(h → C)	1 × h	1 × C	Softmax (출력 시)

도식적으로는, 입력 노드 임베딩이 상단 경로에서 3개의 GCN 블록을 통과하며 점진적으로 차원이 d → h → h/2로 축소되고, 하단 경로에서는 입력에서 바로 d → h/2로 투영된 뒤, 두 표현을 연결하여 그래프 수준 표현을 얻는 구조이다. 상단 경로는 고차 이웃 구조와 엣지 가중치를 반영한 고수준 특성을 학습하고, 하단 경로는 원본 임베딩에 더 가까운 저차 표현을 보존함으로써, 깊은 층에서 발생할 수 있는 정보 소실을 완화하는 일종의 스킵 경로 역할을 수행한다. 또한 첫 번째 GCN 층 출력에 GraphNorm을 적용하여 배치 간 통계 차이로 인한 학습 불안정을 줄였으며, 각 경로에 Dropout(0.5)을 적용해 소규모 그래프 데이터셋에서의 과적합을 완화하였다. 활성화 함수는 ReLU 대비 부드러운 비선형성을 제공하는 GELU를 사용하였다.

모델 구조 탐색 과정에서 GCN 층 수(2~4층), 은닉 차원 크기(h∈32,64,128), 활성화 함수(ReLU, LeakyReLU, GELU) 등을 바꿔가며 최적의 성능을 보이는 모델 구조를 찾는 최적화 실험을 수행하였다. 그러나 약 220개 및 1,100개 규모의 텔레그램 및 웹사이트 데이터 규모에서는 복잡한 구조로 갈수록 통계적으로 유의한 성능 향상이 나타나지 않거나 오히려 변동성이 커지는 경향이 관찰되었다. 이에 본 연구에서는 성능과 안정성, 해석 가능성 간 균형을 고려하여 위와 같은 경량 멀티 패스 GCN 구조를 최종 모델로 채택하였다.

4.3. 비교 모델 및 베이스라인

본 연구에서는 제안한 GCN 기반 그래프 분류기의 성능을 검증하기 위해, 그래프 구조를 활용하지 않는 전통적 머신러닝 및 딥러닝 모델을 비교 대상으로 설정하였다. 모든 비교 모델은 동일한 사전학습 언어모델에서 생성된 임베딩을 입력으로 사용하며, 엣지와 같은 그래프 구조를 제거한 상태에서 문서 단위의 텔레그램 메시지 또는 웹페이지 텍스트의 의미적 표현만을 활용하여 분류를 수행한다. 이를 통해 그래프 기반 구조 정보가 범죄 유형 분류 성능 향상에 기여하는 정도를 정량적으로 평가하고자 하였다.

4.3.1. 머신러닝 베이스라인

머신러닝 기반 모델에서는 각 문서를 구성하는 768차원의 토큰 임베딩을 평균(pooling)하여 문서별 고정 크기의 벡터(document embedding)를 생성한 후 이를 분류기에 입력하였다. 이러한 설정은 그래프 구조 없이 의미 기반 벡터 표현만을 사용했을 때의 분류 성능을 평가하기 위한 것이다.

로지스틱 회귀 (Logistic Regression): 선형 결정경계 기반의 다중분류 모델로, 최소한의 비선형성을 가정하는 전통적 분류 방식이다. 이를 통해 문서 임베딩이 클래스 간 선형적으로 분리 가능한지를 확인하는 베이스라인 역할을 수행한다.

XGBoost: 트리 기반 gradient boosting 모델로, 문서 수준 임베딩만을 활용하여 비선형적 패턴을 어느 정도까지 학습할 수 있는지 평가하기 위한 강한 머신러닝 baseline으로 설정하였다.

4.3.2. 딥러닝 베이스라인

딥러닝 비교 실험에서는 그래프 구조 정보를 완전히 제거한 상태에서 순수 신경망 기반 학습 성능을 측정하였다. 이를 위해 다층 퍼셉트론(Multi-layer Perceptron, MLP) 기반 분류기를 구축하였다.

다층 퍼셉트론: 입력 임베딩을 256차원과 128차원의 두 개의 Fully Connected 층을 통해 축소하며, 각 층 사이에 Dropout을 적용하여 과적합을 방지하였다. 다층 퍼셉트론은 그래프 구조가 없는 조건에서 신경망 모델이 학습할 수 있는 표현력의 한계를 평가하기 위한 것으로, 주어진 임베딩의 분포적 특징만을 기반으로 분류를 수행한다.

V. 실험 설계 및 결과

이 절에서는 실제로 구축된 텔레그램·웹사이트 OSINT 데이터를 지식그래프로 표현한 뒤 GNN에 입력하는 방식이, 단순히 노드 임베딩만을 사용하여 베이스라인 머신러닝 모델을 학습하는 경우보다 분류 성능 향상에 얼마나 기여하는지를 검증한다. 이를 위해 각 채널 및 도메인에 대해 노드 임베딩을 집계하여 고정 길이 벡터를 구성하고, 로지스틱 회귀, 다층 퍼셉트론(MLP), XGBoost 등 그래프 구조를 사용하지 않는 비교 모델을 학습한 뒤, 제안한 GCN 기반 분류 모델과의 정량적 성능 차이를 분석하였다.

5.1. 실험 환경 및 설정

본 연구의 모든 실험은 NVIDIA RTX 6000 Ada Generation GPU 2개(각각 약 49 GB VRAM, 총 98 GB VRAM)와 251 GB 시스템 메모리를 탑재한 워크스테이션에서 수행하였다. 소프트웨어 환경은 Python 3.11, PyTorch 2.6, PyTorch Geometric 2.7, scikit-learn 1.7.2를 기반으로 구축하였으며, 노드 임베딩은 4.1 절에서 선별된 노드 임베딩 모델 및 텐서화 전략을 활용하였다.

텔레그램 채널 데이터는 학습/테스트 데이터셋을 각각 8:2 비율로 계층적(stratified) 분할하였고, 웹사이트 도메인 데이터 또한 범죄/정상 비율을 유지하도록 하여 학습/테스트 데이터셋을 각각 8:2 비율로 분할하였다. GCN 기반 제안 모델과 로지스틱 회귀, 다층 퍼셉트론, XGBoost 등 비교 모델은 동일한 학습·테스트 분할을 공유하도록 하여, 모델 구조 외의 요인으로 인한 성능 차이를 최소화하였다. 분할의 우연성을 줄이기 위해 난수 시드 1~100에 대해 각각 독립적으로 데이터를 분할하고 학습을 반복한 뒤, 각 실험에서 얻은 성능 지표의 평균을 보고하였다.

학습은 미니배치 크기 32으로 진행하였으며, 손실 함수로 다중 클래스 교차 엔트로피(Cross-Entropy)를 사용하였다. 최적화는 Adam 옵티마이저를 사용하고 초기 학습률

(learning rate)은 0.0005, weight decay는 0.0001로 설정하였다. 추가로, 과적합을 완화하기 위하여 모델 가중치의 계수 0.001의 L2 정규화 기법을 적용하여 학습하였다. 최대 1,000 epoch까지 학습을 수행하면서 매 epoch마다 검증 세트에 대한 손실과 정확도를 계산하고, 검증 손실이 최소가 되는 시점의 가중치를 최종 모델로 채택하였다. 이후 절에서는 각 난수 시드에 대해 산출된 정확도, 정밀도, 재현율, F1-score를 모두 macro-averaging 기준으로 평균값을 계산하여 정량적 성능 지표로 제시한다.

5.2. 정량적 성능비교

<Table 3과 4>는 텔레그램 채널 데이터셋 기반 다중 클래스 분류와 웹사이트 도메인 데이터셋 기반 이진 분류에 대해 제안한 GCN 기반 모델과 비교 모델들의 정량적 성능을 요약한 것이다. 평가 지표는 Precision, Recall, F1-score, Accuracy를 macro 기준으로 산출하였다.

<Table 3> Performance comparison of GCN and baseline models on the Telegram multiclass dataset

모델	Precision	Recall	F1-score	Accuracy
GCN (Co-occur)	0.8490	0.7787	0.7968	0.9167
GCN (Similarity)	0.6527	0.5136	0.5407	0.8182
Logistic Regression	0.2917	0.2652	0.2101	0.5318
XGBoost	0.4410	0.4587	0.4474	0.7984
MLP	0.5938	0.5932	0.5917	0.8682

<Table 4> Performance comparison of GCN and baseline models on the website binary dataset

모델	Precision	Recall	F1-score	Accuracy
GCN (Similarity)	0.9195	0.9196	0.9195	0.9196
Logistic Regression	0.8788	0.8750	0.8758	0.8750
XGBoost	0.8800	0.8795	0.8795	0.8795
MLP	0.8634	0.8616	0.8593	0.8616

먼저 텔레그램 채널 범죄 유형 분류 결과를 보면, GCN(Co-occur) 모델이 모든 지표에서 가장 우수한 성능을 보였다. 동일한 GCN 구조이지만 노드 간 유사도를 중심으로 엣지를 구성한 GCN(Similarity) 는 성능이 다소 떨어져, 공출현 스키마가 텔레그램 메시지의 범죄 유형을 반영하는 데 더 적합함을 확인할 수 있었다. 반면, 노드 임베딩만을 입력으로 사용하는 로지스틱 회귀, XGBoost, 다층 퍼셉트론 등 비그래프 기반 모델은 전반적으로 낮은 F1-score와 정확도를 보였으며, 특히 Logistic Regression의 경우 다중 클래스 구조를 충분히 표현하지 못해 성능이 크게 저하되는 경향이 나타났다. 이는 텔레그램 채널과 같이 키워드·주제 간 관계 구조가 중요한 데이터에서는 단순 벡터화보다 키워드 공출현 관계를 반영한 그래프 표현과 GCN이 유의미한 이득을 제공함을 시사한다.

웹사이트 도메인의 범죄 여부 이진 분류에서는 전반적인 성능 수준이 전 모델에서 높게 나타났다. GCN(Similarity) 가 정확도와 F1-score 약 0.92로 가장 높은 지표를 기록하였으나, 로지스틱 회귀, XGBoost, 다층 퍼셉트론 역시 0.86~0.88 수준의 유사한 성능을 보이며 GCN과의 격차가 크지 않았지만, GCN 모델의 경우 두 번째로 높은 XGBoost(Accuracy 0.8795)와 비교했을 때 약 4%p 높은 정확도와 더 높은 F1-score를 달성하여, 동일한 OSINT 텍스트를 사용하

더라도 그래프 구조 정보를 함께 활용하는 것이 추가적인 성능 향상을 제공함을 확인할 수 있었다.

종합하면, 다중 범주를 갖는 텔레그램 채널 분류 문제에서는 그래프 구조와 공출현 기반 스키마를 활용한 GCN이 비그래프 모델 대비 큰 폭의 성능 향상을 보였으며, 이진 분류인 웹사이트 도메인 문제에서도 전통적인 머신러닝 모델들이 높은 기본 성능을 보이는 상황에서 GCN이 가장 높은 정확도와 F1-score로 일관된 우위를 유지하는 것을 확인하였다. 이는 OSINT 데이터의 특성과 과업 난이도에 따라 효과의 크기는 달라지지만, 지식그래프 기반 GNN 접근법이 전통적인 벡터 기반 분류기에 비해 구조 정보를 활용한 추가적인 성능 개선을 제공한다는 점을 시사한다.

5.3. 사례 분석 및 오류 사례 검토

본 절에서는 제안한 모델의 실제 예측 결과를 바탕으로, 대표적인 성공 사례와 오분류 사례를 비교·분석함으로써 모델의 동작 특성과 한계를 구체적으로 검토하였다. 텔레그램 채널 메시지와 웹사이트 HTML을 입력으로 범죄 유무를 분류하는 실험에서, 일반적인 사기 계정이나 피싱 사이트에 대해서는 사기 관련 키워드, 비정상적인 문장 패턴, 특정 도메인 구성 요소 등이 일관되게 나타나 비교적 안정적인 분류 성능을 확인할 수 있었다. 그러나 보다 고도화된 사칭 유형에 대해서는 현 모델의 구조적 한계가 뚜렷하게 드러났다. 먼저 텔레그램 채널의 경우, 공식 계정의 공지 문구와 게시 내용을 거의 그대로 복제한 사칭 채널이 존재하는데, 이들 채널은 메시지의 내용·형식·작성 패턴이 실제 공식 채널과 매우 유사하였다. 제안한 모델은 주로 메시지 텍스트 기반 통계 및 임베딩 정보를 활용하기 때문에, 이러한 사칭 채널을 정상 공식 채널과 구별하지 못하고 '정상'으로 오분류하는 사례가 반복적으로 관찰되었다. 이는 텍스트 내용만으로는 운영 주체·채널 생성 맥락·메타데이터와 같은 추가 정보를 반영하기 어렵다는 한계를 시사한다.

웹사이트 도메인 분류 실험에서도 유사한 문제 양상이 확인되었다. 정상 웹사이트와 철자만 약간 다른 도메인명을 사용하고, HTML 구조와 CSS 디자인, 메뉴 구성 등을 거의 동일하게 복제한 사칭 사이트의 경우, 현재 모델이 활용하는 텍스트·DOM 구조 기반 특징만으로는 정상 사이트와의 차이를 충분히 포착하기 어려웠다. 그 결과 일부 사칭 사이트가 '정상'으로 분류되는 오류가 발생하였다. 이러한 사례 분석 결과는, 단순한 정적 콘텐츠와 구조 정보만을 사용하는 현재 OSINT 기반 분류 모델로는 고도화된 사칭·위장형 범죄를 완전히 탐지하기 어렵고, 도메인 등록 정보, 호스팅 인프라, 접속 이력, 외부 평판 정보 등 추가적인 맥락·행위 기반 특징을 지식그래프에 통합하는 후속 연구가 필요함을 보여준다.

<Figure 3> Example of similar HTML structures in a legitimate and spoofed shopping site

Ⅵ. 논의 및 결론

6.1. 연구 결과 요약 및 시사점

본 연구는 텔레그램 채널 메시지와 웹사이트 HTML과 같은 텍스트 기반 OSINT를 키워드 중심 지식그래프 형태로 구조화한 뒤, 이를 입력으로 하는 GCN 기반 그래프 분류 모델을 통해 범죄 유무 및 범죄 유형을 자동 분류하는 프레임워크를 제안하였다. 텔레그램의 경우 채널 단위로 키워드 노드와 공출현·의미 유사도 엣지를 구성하고, 웹사이트는 도메인 단위 한국어 명사 키워드를 중심으로 의미 유사도 그래프를 구성하여, Neo4j-PyTorch Geometric 간 자동 텐서 변환 파이프라인까지 포함한 “수집-전처리-그래프화-학습” 전 과정을 일관된 지식그래프-GNN 워크플로로 제시했다는 점에 의의가 있다.

정량적 실험 결과, 텔레그램 범죄 유형 다중 분류에서는 공출현 기반 엣지를 활용한 GCN(Co-occur) 모델이 F1-score 약 0.80, Accuracy 약 0.92 수준의 성능을 보이며, 동일 임베딩을 사용하되 그래프 구조를 사용하지 않는 로지스틱 회귀, 다층 퍼셉트론, XGBoost 대비 뚜렷한 우위를 보였다. 웹사이트 범죄/정상 이진 분류에서도 GCN(Similarity)이 F1-score·Accuracy 약 0.92로 가장 높은 성능을 기록하였고, 다른 비교 모델들이 0.86~0.88 수준에서 수렴한 것과 비교해 일관된 성능 향상을 확인할 수 있었다. 이는 동일한 텍스트 임베딩이라 하더라도, 키워드 간 공출현 및 의미 유사도를 그래프 구조로 반영하고 GNN으로 학습할 경우, 특히 다중 범주 분류와 같이 구조적 맥락이 중요한 문제에서 의미 있는 성능 개선을 달성할 수 있음을 시사한다.

사례 분석에서도 키워드 지식그래프-GCN 조합이 전형적인 사기 계정·피싱 사이트에 대해서는 안정적인 탐지 성능을 보이는 반면, 실제 공식 계정과 문구·메시지 패턴이 거의 동일한 텔레그램 사칭 채널이나, 정상 사이트의 레이아웃과 도메인 구조를 정교하게 모방한 웹사이트 사칭 사례에서는 여전히 혼동이 발생함을 확인하였다. 이는 텍스트·정적 구조 정보만 반영한 1차 그래프 표현만으로는 고도화된 위장형 범죄를 완전히 구분하기 어렵고, 향후 도메인 등록 정보, 접속 행태, 외부 평판 지표 등 추가적인 OSINT·행위 기반 신호를 지식그래프에 통합할 필요가 있음을 보여준다.

종합하면, 본 연구는 텍스트 기반 OSINT를 키워드 지식그래프로 구조화하고 GNN을 결합하는 접근이, 수작업 분석에 의존해 온 기존 사이버 범죄 탐지 업무를 자동화·고도화할 수 있는 실질적인 기술적 옵션이 될 수 있음을 실험적으로 검증하였다. 더 나아가, 제안 프레임워크는 향후 멀티모달 정보와 이질적인 관계를 추가로 수용할 수 있는 확장성을 가지므로, 사이버 위협 인텔리전스 및 수사 지원 시스템의 기반 인프라로 발전할 잠재력을 지닌다.

6.2. 한계점 및 향후 연구 방향

본 연구는 OSINT 기반 지식그래프와 GNN을 결합하여 사이버범죄 유무·유형을 자동 분류할 수 있는 가능성을 보여주었으나, 사이버범죄의 구체적인 지식과 결합하여 전문적인 라벨링이 되지 않았다는 점, OSINT 데이터 내의 이미지 정보가 들어가지 않았다는 점과 같은 부분에서 데이터 구성의 측면에서 여전히 개선의 여지가 남아 있다. 또한 본 연구에서 구축한 지식그래프는 텍스트 단일 모달리티에 기반한 초기 형태의 그래프로서, 향후 다양한 노드 타입을 추가하고 이에 적합한 R-GCN(Relational Graph Convolutional Network), GAT(Graph Attention Network), HGT(Heterogeneous Graph Transformer) 등의 고도화된 GNN 모델을 적용할 수 있는 구조적 확장성을 지닌다.

6.2.1. 도메인 지식 기반 라벨링 체계 고도화 및 범죄 데이터셋 규모/다양성 확대

현재 텔레그램 채널과 웹사이트 도메인에 대한 라벨은 공개 데이터셋의 플래그 정보나 단순한 범죄 유무 기준에 크게 의존하고 있으며, 세부 범죄유형을 일관되게 반영하기에는 한계가 있다. 향후에는 수사·보안 전문가와 협력하여 범죄유형 체계를 정교화하고, 유형별 정의·예시·경계 사례를 포함한 라벨링 가이드를 마련함으로써 보다 신뢰도 높은 학습 데이터셋을 구축할 필요가 있다. 또한 현재 데이터셋은 크기가 제한적이고, 특정 범죄유형에 표본이 집중되는 라벨 불균형 문제가 존재한다. 텔레그램 채널과 웹사이트 도메인을 지속적으로 수집·갱신하고, 신규 범죄 수법·플랫폼을 포함하는 데이터를 추가하여, 범죄유형·언어·서비스 종류 측면에서 데이터 분포를 넓혀가는 작업이 요구된다.

6.2.2. 지식그래프 스키마 고도화 및 관계·이질성 반영 GNN 확장

향후에는 동일한 키워드라도 등장 위치와 역할에 따라 다른 의미와 위험도를 가질 수 있다는 점을 반영하기 위해, 지식그래프의 노드·엣지 스키마를 세분화하여 구조적 표현력을 높이는 방향으로 확장이 필요하다. 예를 들어 웹사이트 HTML 데이터의 경우, 현재는 페이지 내 텍스트를 동일한 키워드 노드로 취급하고 있으나, 추후에는 <title>·헤더·네비게이션 메뉴·본문·푸터·링크 앵커 텍스트 등 HTML 구조와 역할에 따라 서로 다른 노드 타입 또는 속성으로 구분함으로써, 페이지 내에서 어떤 위치의 텍스트가 범죄 판별에 더 중요한지 학습할 수 있도록 설계할 수 있다. 텔레그램 메시지 역시 본문 텍스트, 첨부 링크의 미리보기(프리뷰) 텍스트, 해시태그, 고정 메시지 등 출처와 기능이 다른 텍스트 요소를 구분된 노드/속성으로 모델링함으로써 채널 운영 패턴과 위장 수법을 보다 세밀하게 반영할 수 있을 것이다.

엣지 관점에서도 단순 공출현·유사도 관계를 넘어, “동일 HTML 블록 내 포함”, “메시지 간 회신/전달 관계”, “동일 기간 내 동시 등장” 등 다양한 의미의 관계를 개별 엣지 타입으로 정의하는 방향을 고려할 수 있다. 이러한 다종의 노드와 관계를 도입하면 그래프가 자연스럽게 이질적 구조를 띠게 되므로, 모델 측면에서는 이에 상응하는 R-GCN, GAT, HGT 등 관계·이질성 정보를 직접 반영할 수 있는 GNN으로의 확장이 요구된다. 더 나아가, 그래프 자기지도 학습 기법을 결합하여, 라벨이 없는 대규모 OSINT 그래프에서 먼저 일반적인 표현을 학습한 뒤, 소규모 범죄 라벨 데이터로 미세 조정하는 전략을 적용한다면, 복잡한 지식그래프 구조를 보다 효율적으로 활용하면서도 일반화 성능을 향상시킬 수 있을 것으로 기대된다.

참고문헌 (References)

[1] Federal Bureau of Investigation. 2024. Internet crime report 2024. Internet Crime Complaint Center, Federal Bureau of Investigation, Washington DC, USA. Technical Report IC3-2024. https://www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf

[2] World Economic Forum. 2025. Global cybersecurity outlook 2025: Insight report. World Economic Forum, Geneva, Switzerland. Technical Report GCO-2025. https://reports.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2025.pdf

[3] Kwon D, Borrion H, Wortley R. 2024. Measuring cybercrime in calls for police service. Asian Journal of Criminology, 19, 329-351. https://doi.org/10.1007/s11417-024-09432-2

[4] Borj PR, Raja K, Bours P. 2023. Detecting online grooming by simple contrastive chat embeddings. Proceedings of the 9th ACM International Workshop on Security and Privacy Analytics, Charlotte, NC, USA, pp. 57-65. https://doi.org/10.1145/3579987.3586564

[5] Waezi H, Fani H. 2025. Enhancing online grooming detection via backtranslation augmentation. Proceedings of the 31st International Conference on Computational Linguistics, Abu Dhabi, UAE, pp. 2340-2350. https://aclanthology.org/2025.coling-main.160/

[6] Kim S, Lee B, Maqsood M, et al. 2025. Deep Learning-based natural language processing model and optical character recognition for detection of online grooming on social networking services. Computer Modeling in Engineering and Sciences, 143(2), 2079-2108. https://doi.org/10.32604/cmes.2025.061653

[7] Marazqah Btoush EAL, Zhou X, Gururajan R, et al. 2023. A systematic review of literature on credit card cyber fraud detection using machine and deep learning. PeerJ Computer Science, 9, e1278. https://doi.org/10.7717/peerj-cs.1278

[8] Yazdanjue N, Rakhshaninejad M, Yazdanjouei H, et al. 2025. Cyber threat management using semi-supervised ensemble learning and enhanced interior search algorithm: applications for illicit marketplace classification in deep/dark web and social platforms. Annals of Operations Research. https://doi.org/10.1007/s10479-025-06651-3

[9] Chen LC, Hsu CL, Lo NW, et al. 2017. Fraud analysis and detection for real-time messaging communications on social networks. IEICE Transactions on Information and Systems, E100.D(10), 2267-2274. https://doi.org/10.1587/transinf.2016INI0003

[10] Roy SS, Vafa EP, Khanmohammadi K, et al. 2025. DarkGram: A large-scale analysis of cybercriminal activity channels on telegram. Proceedings of the 34th USENIX Conference on Security Symposium, Seattle, WA, US, 249, pp. 4839-4858. https://dl.acm.org/doi/10.5555/3766078.3766327

[11] Alshehri SM, Sharaf SA, Molla RA. 2025. Systematic review of graph neural network for malicious attack detection. Information, 16, 470. https://doi.org/10.3390/info16060470

[12] Singhal A. 2012. Introducing the knowledge graph: Things, not strings. Google The Keyword. Available at: https://blog.google/products/search/introducing-knowledge-graph-things-not/ accessed on 2025. 11. 20.

[13] Wang C, Ma X, Chen J, et al. 2018. Information extraction and knowledge graph construction from geoscience literature. Computers & Geosciences 112, 112-120. https://doi.org/10.1016/j.cageo.2017.12.007

[14] Choi S, Jung Y. 2025. Knowledge graph construction: Extraction, learning, and evaluation. Applied Sciences, 15(7), 3727. https://doi.org/10.3390/app15073727

[15] Zhou J, Cui G, Hu S, et al. 2020. Graph neural networks: A review of methods and applications. AI Open, 1, 57-81. https://doi.org/10.1016/j.aiopen.2021.01.001

[16] Khemani B, Patil S, Kotecha K, et al. 2024. A review of graph neural networks: Concepts, architectures, techniques, challenges, datasets, applications, and future directions. Journal of Big Data, 11, 18. https://doi.org/10.1186/s40537-023-00876-4

[17] Bing R, Yuan G, Zhu M, et al. 2023. Heterogeneous graph neural networks analysis: A survey

of techniques, evaluations and applications. Artificial Intelligence Review, 56, 8003-8042. https://doi.org/10.1007/s10462-022-10375-2

[18] Rahman MS. 2025. The art of open source intelligence (OSINT): Addressing cybercrime, opportunities, and challenges. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5281845

[19] Bokolo BG, Liu Q. 2024. Artificial intelligence in social media forensics: A comprehensive survey and analysis. Electronics, 13(9), 1671. https://doi.org/10.3390/electronics13091671

[20] La Morgia M, Mei A, Mongardini AM. 2023. TGDataset: Collecting and exploring the largest telegram channels dataset. arXiv:2303.05345. https://doi.org/10.48550/arXiv.2303.05345

[21] Alshehri L, Alajmani S. 2025. Malicious domain name detection using ML algorithms. International Journal of Advanced Computer Science and Applications, 16(3), 483-494. https://doi.org/10.14569/IJACSA.2025.0160348

[22] The Cheat. [n.d.]. The Cheat: Fraud information sharing service [더치트: 사기피해 정보공유 서비스]. Available at: https://thecheat.co.kr/ accessed on 2025. 11. 20.

[23] Le Pochat V, Van Goethem T, Tajalizadehkhoob S, et al. 2019. TRANCO: A research-oriented top sites ranking hardened against manipulation. Network and Distributed Systems Security (NDSS) Symposium 2019, San Diego, CA, USA. https://doi.org/10.14722/ndss.2019.23386