마스킹 기반 계층적 임베딩을 통한 네트워크 트래픽 분류 모델의 일반화 성능 향상

김주성, 유경민, 장윤성, 남승우, 박재원, 최태상°, 김명섭*
고려대학교, °㈜모비젠, *고려대학교 세종캠퍼스
{jsung0514, rudals2710, brave1094, nam131119, 2018270614}@korea.ac.kr,
°choits@mobigen.com, *tmskim@korea.ac.kr

Masking-based Hierarchical Embedding for Generalized Network Traffic Classification

Ju-Sung Kim, Gyeong-Min Yu, Yoon-Seong Jang, Seung-Woo Nam, Jae-Won Park,
Taesang Choi°, Myung-Sup Kim*
Korea Univ., °Mobigen Co., LTD., *Korea Univ

요 약

기존 네트워크 트래픽 분류 모델은 데이터셋의 수집 시기나 환경 차이로 인해 최신 트래픽 분포를 충분히 반영하지 못하고, 도메인 간 일반화 성능이 저하되는 한계가 있다. 이를 해결하기 위해 사전학습 기반 모델의 구조를 유지하면서 토큰 마스킹과 계층별 마스킹을 결합한 커스텀 임베딩 구조를 제안하였다. 공개 및 사설 데이터셋을 활용한 실험 결과, 다양한 환경에서 일관된 성능을 보였으며 서로 다른 트래픽 분포 간에서도 안정적인 적응력을 나타냈다.

Ⅰ. 서 론

네트워크 트래픽 분석은 복잡한 계층 구조와 빠르게 변화하는 통신 환경으로 인해, 안정적이고 일반화 가능한 모델을 구축하는 것이 중요한 과제로 인식되고 있다. 특히 데이터의 생성 시점과 환경이 상이할 경우, 모델이 학습 과정에서 습득한 표현이 실제 최신 트래픽 패턴을 충분히 반영하지 못할 가능성이 존재한다. 이러한 이유로 최근 연구들은 다양한 네트워크 상황을 포괄할 수 있는 데이터 구성과 구조적 표현 학습의 중요성을 강조하고 있다.

한편, 사전학습과 미세조정을 결합한 접근은 복잡한 트래픽 구조를 효율적으로 학습하기 위한 주요 전략으로 자리 잡고 있다. 그러나 기존의 데이터셋들은 시간이 지남에 따라 최신 환경을 충분히 반영하지 못하거나, 모델의 일반화 성능을 체계적으로 검증하기 위한 교차검증 절차가 제한적으로 적용된 사례가 많았다. 이러한 점은 모델이 실제 운용 환경에서 일관된 성능을 보장하기 어렵게 만드는 요인으로 작용할 수 있다. 이에 본 논문은 효율적인 사전학습 구조를 기반으로, 토큰 마스킹을 이용하여 불필요한 입력을 제거하고, 계층별 마스킹을 통해 세션, 패킷, 계층, 필드 단위의 독립적인 특징을 학습하는 구조와 최신 데이터 환경을 반영하여 새로 구축한 데이터셋을 제안한다.

Ⅱ. 본 론

본 연구에서는 사전학습을 통해 네트워크 트래픽의 계층적 표현을 학습하는 모델[1]을 기반으로 하며 구조는 그림 1과 같다. 해당 모델은 세션(Session), 패킷(Packet), 계층(Layer), 필드(Field) 수준의 다양한 정보를 통합적으로 반영하도록 설계되었으며 다중 과제 학습을 통해 각 계층 간의 관계를 동시에 학습한다. 이러한 사전학습 전략은 비라벨 데이터로부터 일반화된 표현을 얻는 데 효과적이며, 이후의 다운스트림 과제에 활용될 수 있는 기반 표현을 제공한다.

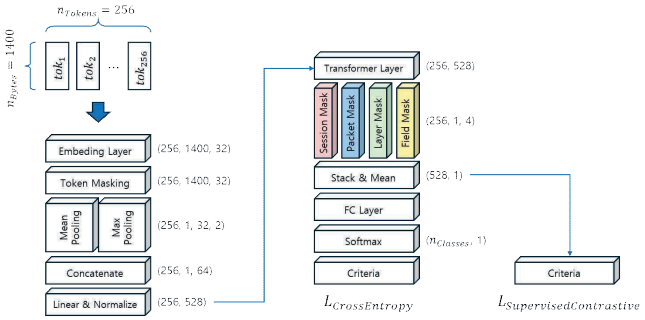

그림 2. 트래픽 계층 구조 반영을 위한 커스텀 임베딩 및 미세조정 과정

본 연구의 핵심은 입력 데이터를 모델 입력에 맞게 변환하는 커스텀 임베딩 구조에 있으며 이는 그림 2와 같다. 이 구조는 토큰 단위의 바이트 임베딩을 생성한 후, 토큰 마스킹을 적용하여 CLS, SEP, PAD와 같이 유효하지 않은 데이터를 0으로 변환함으로써 불필요한 정보를 제거한다. 마스킹된 입력은 mean pooling과 max pooling을 각각 수행하여 평균적 특성과 극값 기반 특성을 동시에 추출하고, 두 특성을 결합 및 정규화를 통해 안정적인 표현을 형성한다.

트랜스포머 인코더는 커스텀 임베딩의 결과를 입력으로 받아 트래픽 시퀀스를 처리한다. 이때, 모델은 계층별 마스킹을 적용하여 세션, 패킷, 계층, 필드 수준의 각 구조별 토큰을 통해 각 계층의 표현을 독립적으로 집약하며, 이렇게 얻어진 구조별 표현 벡터는 각각 네트워크 트래픽의 세부적 의미를 반영한다. 모델은 세션 - 패킷 - 계층 - 필드 수준의 표현을 개별적으로 계산한 후, 이들을 하나의 특징 벡터로 병합(stacking)하여 통합 특징 벡터를 구성한다.

미세조정 단계의 학습 목표는 크로스 엔트로피 손실(Cross-Entropy Loss)을 통해 클래스 라벨 정합도를 학습하고, 감독형 대조 손실(Supervised Contrastive Loss)을 추가하여 동일 클래스 간 표현은 가깝게, 서로 다른 클래스 간 표현은 멀게 학습되도록 유도한다. 이 두 가지 손실항은 비율 α로 가중 결합되어, 모델이 구조별 특징의 분리도와 표현 안정성을 동시에 학습하도록 한다.

본 논문은 과기정통부·정보통신기획평가원의 정보통신방송표준개발지원(R&D,정보화)사업으로 수행한 결과이며(No. RS-2025-02219319, 양자컴퓨터 공격에도 안전한 양자암호 기반 제로트러스트 보안 네트워크/서비스 및 제어/관리 기술 표준개발), 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(00235509, ICT융합 공공 서비스●인프라의 암호화 사이버위협에 대한 네트워크 행위기반 보안관제 기술 개발)과 2024년 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원(P0028063, 2024년 산업기술국제협력사업)을 받아 수행된 연구임.

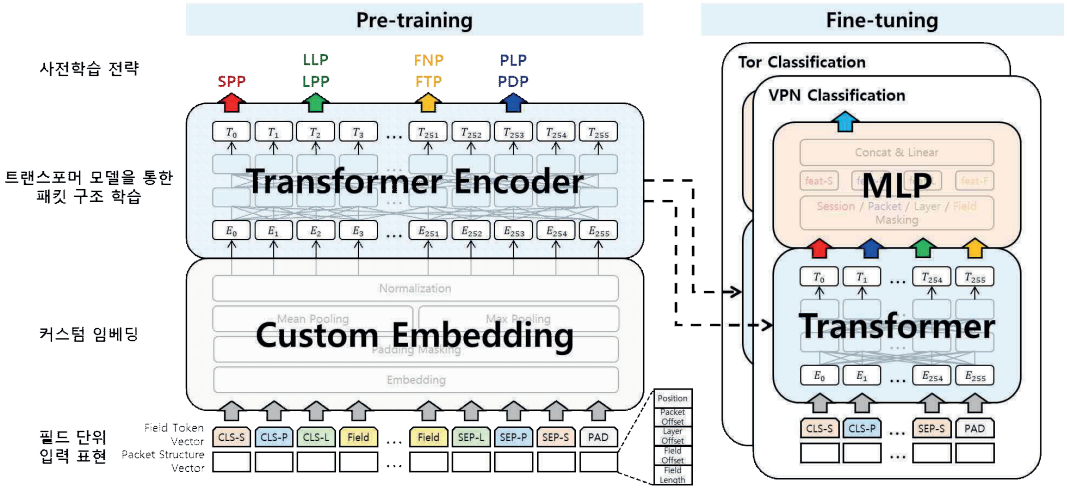

그림 1 트래픽 계층 구조 사전학습 및 미세조정 모델 개요

Ⅲ. 실 험

데이터셋	클래스 개수	세션 개수	
		전처리 전	전처리 후
ISCX VPN 2016	16	310,189	4,802
ISCX Tor 2016	14	31,367	26,165
CSTNET-TLS1.3	120	46,372	46,372
KU 2023	48	71,918	50,166
KU 2024	48	70,380	32,108
KU 2025	300	399,722	351,536

표 1. 데이터셋 주요 구성 정보

평가를 위해 총 6종의 네트워크 트래픽 데이터셋을 사용하였으며 이 중 공개 데이터셋은 ISCX VPN 2016[2], ISCX Tor 2016[3], CSTNET-TLS1.3[4]의 3가지이다. 이들 공개 데이터셋은 다양한 네트워크 응용과 트래픽 유형을 포괄하고 있어 모델의 기본 학습에 적합하지만, 수집 시점이 오래되고 네트워크 인프라와 암호화 프로토콜이 현재와 상이하다는 한계가 존재한다. 이러한 문제를 보완하기 위해 사설 환경에서 수집한 3종의 데이터셋을 추가로 사용하였다. 이 중 두 개의 데이터셋은 교차검증에 활용되었으며, 나머지 한 개의 데이터셋은 최신 네트워크 트래픽을 기반으로 신규 제작된 데이터셋이다.

전처리 단계에서는 Ethernet 헤더를 제거하고, IPv4 TCP/UDP 세션만을 사용하였다. 그리고 3-way 핸드셰이크가 불완전한 세션, 응용과 관계가 없는 세션들을 제거하고 사전학습에 사용하였으며, 이를 다시 9:1로 분할 후 각각 미세조정 단계의 학습 및 테스트에 사용하였다. 표 1은 실험에 사용한 데이터셋들의 주요 구성 정보를 요약한 것이다.

데이터셋	분류 정확도	
	Proposed	ET-BERT[4]
ISCX VPN 2016	95.22	85.19
ISCX Tor 2016	98.01	83.11
CSTNET-TLS1.3	97.22	95.10
KU 2023	95.14	91.72
KU 2024	96.45	89.66
KU 2025	95.72	90.32

표 2. 기존 연구와의 정확도 비교 실험 결과

표 2는 동일한 데이터셋을 대상으로 분류 모델을 통해 비교 실험을 진행한 결과를 정리한 것이다. 실험 결과, 기존 ET-BERT 대비 제안하는 방법의 모델이 6가지 데이터셋에서 모두 더 높은 성능을 보였다. 교차검증은 클래스 개수 및 종류가 동일한 사설 데이터셋 2개를 대상으로 하였으며, 6개 데이터셋 중 하나를 검증 대상으로 제외하고 나머지 5개 데이터셋으로 사전학습 후, 제외된 데이터셋을 이용해 미세조정과 검증을 수행하는 방식으로 진행하였다. 실험 결과, 제안한 미세조정 전략을 적용한 모델은 KU 2023에서 94.31, KU 2024에서 94.49의 분류 정확도를 나타내었고 이는 두 데이터셋 모두에서 안정적인 성능을 보였으며 미세조정 후 성능 편차가 크지 않았음을 알 수 있다.

Ⅳ. 결 론

본 논문은 사전학습 기반 네트워크 트래픽 분류 모델의 표현력과 일반화 성능을 향상시키기 위해, 마스킹 기반 커스텀 임베딩 구조를 제안하였다. 제안한 구조는 토큰 마스킹을 적용하여 유효하지 않은 입력을 제거하고, 계층별 마스킹을 통해 계층별 특징을 통합하여 하나의 표현으로 구성한다. 또한, 크로스 엔트로피 손실과 감독형 대조 손실을 결합하여 계층 간 표현 분리도와 학습 안정성을 동시에 확보하였다. 다양한 데이터셋을 활용한 실험 결과, 제안한 미세조정 전략은 기존 방식 대비 일관된 성능 향상을 보였으며, 특히 데이터셋 단위 교차검증을 통해 일반화 성능을 실증하였다. 향후 연구에서는 실시간 트래픽 환경에 적용 가능한 경량화 모델 구조에 관하여 연구할 계획이다.

참 고 문 헌

[1] 김주성, 장윤성, 김지민, 백의준, 김명섭, "패킷 구조 학습을 통한 트랜스포머 기반 응용 트래픽 분류", 2025년 한국통신학회 하계종합학술발표회 (KICS 2025), 신화월드, 제주, Jun. 18-20, 2025, pp.1-2.

[2] Gil, et al. "Characterization of encrypted and VPN traffic using time-related features." Proceedings of the 2nd international conference on information systems security and privacy (ICISSP 2016). Setúbal, Portugal: SciTePress, 2016.

[3] Arash Habibi Lashkari, Gerard Draper-Gil, Mohammad Saiful Islam Mamun and Ali A. Ghorbani, "Characterization of Tor Traffic Using Time Based Features", In the proceeding of the 3rd International Conference on Information System Security and Privacy, SCITEPRESS, Porto, Portugal, 2017.

[4] Lin, et al. "Et-bert: A contextualized datagram representation with pre-training transformers for encrypted traffic classification." Proceedings of the ACM Web Conference 2022. 2022.