

O-RAN 구조에서의 dApp 기반 네트워크 슬라이스의 초저지연 분류에 관한 연구

박시연, 최다영, 박형곤

이화여자대학교 공과대학 전자전기공학과

{siyoun0116, dayoung.choi, hyunggon.park}@ewha.ac.kr

A Study on Ultra-Low Latency Network Slice Classification based on O-RAN dApp

Siyoun Park, Dayoung Choi, Hyunggon Park

Department of Electronic and Electrical Engineering, Ewha Womans University

요약

본 논문은 Open Radio Access Network (O-RAN) 구조에서 슬라이스 분류를 초저지연으로 실현하기 위해, 특정 선택 기법이 통합된 dApp 기반 프레임워크를 제안하였다. 5G 네트워크의 대표 트래픽 유형인 초고신뢰·저지연 통신은 10ms 이하의 초저지연, 99% 이상의 초고신뢰 서비스를 요구한다. 제안한 dApp 기반 프레임워크는 10ms 이하의 실시간 분류를 가능하게 하여 rApp과 xApp의 지원 한계를 보완하고 초저지연 처리를 실현한다. 또한, 상호 정보량을 이용한 특정 선택을 적용하여 입력 데이터의 차원을 축소함으로써, 경량화된 트래픽 분류기를 설계할 수 있다. 실제 5G 트래픽 데이터를 활용한 실험을 통해 소수의 주요 특징만으로도 0.99 이상의 정확도와 2.6ms 이하의 추론 지연을 보여 저지연성과 고신뢰성을 동시에 만족함을 확인하였다.

I. 서 론

5G와 6G 네트워크에서는 데이터 기반의 적응적 인공지능을 활용한 지능형 자동화가 필수적 요소로 부상함에 따라, RAN 구조에 인공지능을 통합하려는 Artificial Intelligence-Radio Access Network (AI-RAN)이 등장하였다. AI-RAN을 실현하기 위해 개방적이고 상호운용 가능한 Open RAN (O-RAN)이 제시되었다[1]. 5G는 초고속, 초연결, 초저지연을 목표로 하며, 이에 따라 향상된 광대역 이동통신(enhanced mobile broadband, eMBB), 대규모 사물통신(massive machine type communications, mMTC), 초고신뢰·저지연 통신(ultra-reliable and low-latency communications, URLLC)으로 네트워크 슬라이스 트래픽 유형이 구분된다. 이러한 유형들은 제한된 지연 시간 내에서 신속한 자원 스케줄링과 처리가 요구된다. 특히 URLLC와 일부 mMTC 서비스는 10ms 이하의 초저지연이 보장되어야 하며, 99% 이상의 높은 신뢰도를 확보하는 것이 필수적이다.

기존 O-RAN의 비실시간/근실시간 무선 접속망 지능형 컨트롤러(non/near-real-time RAN Intelligent Controller, non/near-RT RIC)의 폐루프 지연 범위는 각각 1s 이상과 10-1000ms이다. 따라서, 이러한 구조에서는 10ms 이하의 초저지연 요구를 충족하기 어렵다. 이를 보완하기 위해, RAN의 중앙부(Central Unit, CU), 분배부(Distributed Unit, DU)에 적용 가능한 10ms 이하의 지연 범위를 가지는 제3자 어플리케이션 dApp이 제안된 바 있다[2]. 트래픽별 스케줄링을 위해서는 신속하고 정확한 트래픽 분류가 선행되어야 하며[3], 이를 위한 dApp 개발이 필요하다. 특히, DU는 RIC에 비해 더욱 한정된 자원에서 동작하기 때문에, 데이터 분석을 기반으로 자원 효율성과 경량성을 고려할 수 있는 설계가 요구된다[3].

본 논문에서는 dApp 기반 슬라이스 트래픽 분류 프레임워크에 특정 선택(feature selection) 기법을 통합하여, 핵심적인 입력 변수만을 활용한 경량화된 모델을 설계한다. 이를 통해 머신러닝 기반 슬라이스 분류 실험에서 빠른 추론 속도와 높은 정확도를 동시에 달성함을 확인한다.

II. dApp 기반 트래픽 분류를 위한 구조 및 데이터 분석

A. O-RAN 구조에서의 dApp 동작 프레임워크

제안하는 dApp 기반 슬라이스 트래픽 분류가 적용된 O-RAN 구조는 그림 1과 같다. 사용자 단말(User Equipment, UE)에서 발생한

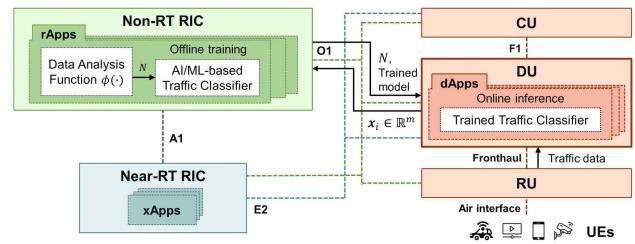


그림 1. 슬라이스 트래픽 분류를 수행하는 dApp이 적용된 O-RAN 구조

트래픽 흐름은 무선부(Radio Unit, RU)에서 수집되어 프론트홀(Front haul)을 통해 DU로 전송된다. 전달되는 데이터 $\mathbf{D} \in \mathbb{R}^{T \times m}$ 는 T 개의 시점 동안 수집된 m 개의 특징 $\mathbf{f}_j \in \mathbb{R}^T$ ($j = 1, 2, \dots, m$)으로 구성되며, 각 시점의 인스턴스는 $\mathbf{x}_i \in \mathbb{R}^m$ ($i = 1, 2, \dots, T$)으로 표현된다. 수집된 데이터는 O1 인터페이스를 통해 non-RT RIC으로 전달된다. 본 논문에서는 rApp 내 데이터 분석 함수 $\phi(\cdot)$ 를 적용함으로써, 주요 특징 n ($\leq m$) 개에 대한 인덱스셋 N 을 도출한다. 이렇게 선별된 특징을 이용하여 \mathbf{x}_i 로부터 차원 축소된 데이터 $\mathbf{x}'_i \in \mathbb{R}^n$ 을 매핑할 수 있다. 머신러닝 기반 트래픽 분류기는 오프라인(offline)으로 학습되며, 학습된 모델과 N 은 O1 인터페이스를 통해 DU로 배포된다. 따라서, dApp은 학습된 분류기와 N 을 통해 차원이 축소된 온라인(online) 데이터 \mathbf{x}'_i 의 트래픽 유형을 실시간으로 분류할 수 있다.

B. 데이터 분석 기반 중요 특징 선택

본 연구에서는 성능 저하 없이 입력 데이터의 차원을 축소하기 위해 상호 정보량(mutual information)[4]을 활용한다. 상호 정보량은 한 확률 변수에 대한 정보가 다른 확률 변수의 불확실성을 감소시키는 정도를 통계적으로 나타내는 지표로, 각 특징 \mathbf{f}_j 가 트래픽 유형 $\mathbf{y} \in \mathbb{R}^T$ 의 구분에 미치는 기여도 MI_j 를 다음과 같이 정량화할 수 있다.

$$MI_j = \sum_{y \in \mathcal{Y}} \sum_{f_j \in \mathcal{F}_j} P(f_j, y) \log \left(\frac{P(f_j, y)}{P(f_j)P(y)} \right) \quad (1)$$

$P(\cdot, \cdot)$ 은 두 변수 사이의 결합 확률 질량 함수(joint probability mass function, joint pmf), $P(\cdot)$ 은 한 변수의 주변 확률 질량 함수 (marginal

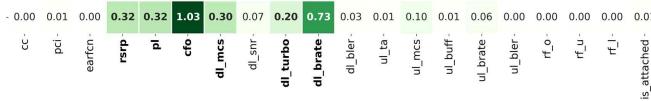


그림 2. 특징별 트래픽 유형과의 상호 정보량

pmf_j 의 의미하고, 특징의 확률변수 f_i 는 확률변수 값 집합 \mathcal{F}_j 내의 값을, 라벨의 확률 변수 y 는 집합 \mathcal{Y} 내의 값을 가진다. 기여도 측정 이후, 계산된 MI_j 값에 니들(Kneedle) 알고리즘[5]을 적용하여, 트래픽 분류에 유의미한 상관성을 가지는 상위 n 개의 특징들을 선별한다.

III. dApp 기반 트래픽 분류 실험

본 연구에서 사용한 데이터셋은 Colosseum O-RAN COMMAG[6]이며, 대규모 무선 애뮬레이터인 Colosseum을 통해 수집되었다. 이는 물리, MAC 계층의 경로 손실(pathloss), 참조 신호 수신 전력(Reference Signal Received Power), 하향 링크 변조 코딩 구성(downlink Modulation Coding Scheme) 등이 포함된 $m=20$ 의 특징으로 구성된 사용자 데이터이며, eMBB, mMTC, URLLC 총 3개의 슬라이스 트래픽 유형을 가진다. 제안한 dApp의 유효성을 검증하기 위해 eXtreme Gradient Boosting (XGBoost)[7], Support Vector Machine (SVM)[8], k -Nearest Neighbors (k -NN)[9]을 이용해 트래픽 분류를 수행하였다. 각각 early stopping, radial basis function (RBF) 커널, $k=5$ 가 적용되었다. 성능 지표는 정밀도와 재현율을 모두 고려하기 위해 F1-score가 사용되었다.

그림 2는 각 특징에 대한 상호 정보량 값을 나타낸 것이다. 높은 상관성을 보이는 특징들은 전체 중 일부에 불과하며, 상호 정보량이 0인 특징을 제외하면 오직 12개의 특징만 의미 있는 정보를 가진다. 또한, 상대적으로 높은 상관성을 가진 특징과 그렇지 않은 특징 간의 대비가 뚜렷한 것은 트래픽 구분에 실질적으로 기여하는 주요 특징이 일부에 집중됨을 시사한다. 계산된 상호 정보량 값들을 통해 니들 알고리즘 적용 결과, 상위 6개 ($n=6$)의 특징이 가장 높은 기여도를 가진 것으로 선별되었다.

그림 3은 선택된 특징의 개수 n 에 따른 모델별 트래픽 분류 성능으로, 선별된 6개의 특징만 사용한 경우, 모든 모델이 전반적으로 높은 성능을 보이는 것을 확인할 수 있다. 그중에서도 XGBoost, k -NN은 $n=3$ 부터 $n=20$ 까지의 모든 부분집합에서 0.98 이상의 안정적인 성능을 나타낸다. 특히, XGBoost는 일관적으로 0.99 이상의 정확도를 달성함으로써, 트래픽 서비스의 신뢰성 보장 기준을 만족한다. 반면, SVM은 비교적 낮은 성능을 기록하여, 고신뢰 환경에서의 실시간 분류에는 다소 한계가 있음을 확인할 수 있다. 또한, 모든 모델에 걸쳐 $n=3$ 까지 특징 개수가 감소하여도 0.90 이상의 높은 성능이 유지되어, 복잡한 고차원 입력 없이도 소수의 핵심적 특징만으로 효과적인 트래픽 분류를 실현할 수 있음을 보여준다.

표 1은 $n \in \{3, 6, 20\}$ 에 따른 모델별 추론 지연 시간을 측정한 결과이며, 테스트 데이터셋 내의 각 인스턴스의 평균 추론 시간으로 계산되었다. 모든 n 에서 추론 지연이 10ms 이하로 측정되어, 트래픽 서비스 및 dApp의 지연 요구를 충족함으로써 저지연성 유효성이 확인되었다. 특히, 최고 성능 모델인 XGBoost는 최대 2.60ms의 지연 시간으로, 실시간 처리에 충분히 빠른 추론 속도를 보였다. 이러한 결과는 dApp이 트래픽 분류를 신속히 수행함으로써, 이후 스케줄링 및 자원 제어의 전체 지연을 단축하고 O-RAN 제어 프로세스의 응답 효율을 향상시키는 데 기여할 수 있다.

IV. 결론

본 논문에서는 O-RAN 구조에서의 실시간 트래픽 처리를 위한 dApp 기반 초기지연 트래픽 분류 프레임워크를 제안하였다. 제안한 방법은 O-RAN 아키텍처를 준수하는 dApp의 적용 구조와 데이터 흐름을 설계하고, 데이터

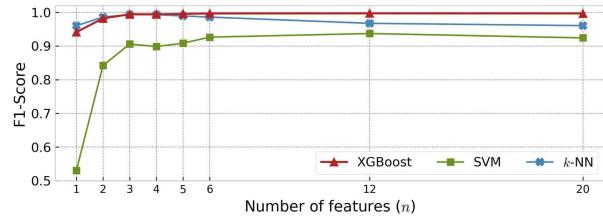


그림 3. 특징 개수(n)에 따른 모델별 트래픽 분류 성능

	XGBoost	SVM	k -NN
$n=3$	0.84 ± 0.50	0.64 ± 0.01	0.60 ± 0.01
$n=6$	1.15 ± 0.02	0.41 ± 0.00	0.60 ± 0.01
$n=20$	2.60 ± 0.46	0.53 ± 0.01	0.39 ± 0.45

표 2. 특징 개수(n)에 따른 모델별 추론 지연 시간 (ms)

분석 기반의 특징 선택을 통해 효율적으로 높은 분류 성능을 유지하도록 하였다. 실제 5G O-RAN 트래픽 데이터를 활용한 실험 결과, 머신러닝 기반 트래픽 분류기들은 전반적으로 높은 성능과 낮은 지연을 달성하였으며, 특히 XGBoost는 0.99 이상의 정확도와 2.6ms 이하의 추론 지연을 기록하며 dApp과 URLLC의 요구 조건을 모두 충족하였다. 이는 머신러닝 기반 dApp이 near-RT RIC보다 낮은 지연 범위에서도 안정적이고 효율적인 트래픽 분류가 가능함을 보여준다. 향후, 제안한 트래픽 분류 dApp과 연계된 스케줄링 dApp 구조를 논의함으로써, 보다 통합적이고 구체적인 dApp 기반 트래픽 제어 및 자원 관리 프레임워크로 확장할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획 평가원의 지원(No. 2021-0-00739, 분산/협력AI 기반 5G+ 네트워크 데이터 분석 기능 및 제어 기술 개발)과 한국연구재단의 지원(No.RS-2025-16066913, AI-RAN 내 분산형 자율 제어에 관한 연구 및 디지털 트윈 기반 검증)을 받아 수행된 연구임.

참 고 문 현

- [1] O-RAN Alliance, “Towards an Open and Smart RAN,” O-RAN Alliance, White Paper, Oct. 2018.
- [2] O-RAN nGRG, “dApps for Real-Time RAN Control: Use Cases and Requirements,” Research Report. RR-2024-10, Oct. 2024.
- [3] Haque, M. E., et al., “A Survey of Scheduling in 5G URLLC and Outlook for Emerging 6G Systems,” *IEEE Access*, vol. 11, pp. 34372–34396, 2023.
- [4] Shannon, C. E. “A Mathematical Theory of Communication,” *The Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
- [5] Satopaa, V., et al., “Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior,” in *Proc. ICDCSW 2011*, 2011, pp. 166–171.
- [6] Bonati, L., et al., “Intelligence and Learning in O-RAN for Data-driven NextG Cellular Networks,” *IEEE Commun. Mag.*, vol. 59, no. 10, pp. 21–27, Oct. 2021.
- [7] Chen, T., and Guestrin, C. “XGBoost: A Scalable Tree Boosting System,” in *Proc. SIGKDD 2016*, 2016, pp. 785–794.
- [8] Cortes, C., and Vapnik, V. “Support-Vector Network,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [9] Silverman, B. W., and Jones, M. C., “E. fix and j.l. hedges (1951): An important contribution to nonparametric discriminant analysis and density estimation,” *Int. Stat. Rev.*, vol. 57, no. 3, pp. 233–238, 1989.