

6G 네트워크 슬라이스 시뮬레이션을 위한 합성 트래픽 데이터셋 생성 방법

이현호, 최원석, 최성곤*
충북대학교

hhl9438@cbnu.ac.kr, choiws@cbnu.ac.kr, *choisg@cbnu.ac.kr

Synthetic Traffic Dataset Generation Method for 6G Network Slice Simulation

Lee Hyeon Ho, Choi Won Seok, Choi Seong Gon*
Chungbuk National Univ.

요약

본 논문은 6G 네트워크 슬라이싱 연구를 위한 합성 트래픽 데이터셋 생성 방법을 제안한다. AI 기반 네트워크 관리 연구에는 학습용 트래픽 데이터셋이 필수적이나, 실제 6G 트래픽 데이터 획득은 현실적으로 어렵다. 본 논문에서는 Poisson 기반 세션 도착, Heavy-tail ON/OFF 분포, Superimposed Periodicity, Event-driven Abrupt Changes 등 실제 셀룰러 트래픽의 핵심 특성을 반영하고, C-V2X, eMBB, IoT 슬라이스별로 차별화된 패턴을 적용한 합성 데이터셋 생성 방법을 제시한다. LSTM 기반 트래픽 예측 모델 학습을 통해 생성된 데이터셋의 활용 가능성을 검증하였다.

I. 서론

6G 네트워크는 C-V2X(Cellular Vehicle-to-Everything), eMBB(enhanced Mobile Broadband), mMTC(massive Machine Type Communications) 등 이질적인 서비스 요구사항을 네트워크 슬라이싱 기술을 통해 단일 물리 인프라에서 동시에 지원한다[1].

각 슬라이스는 서비스 특성에 따라 상이한 트래픽 패턴을 보이며, 이러한 동적 트래픽 변화에 대응하기 위해 AI 기반 트래픽 예측 및 자원 관리 연구가 활발히 진행되고 있다[2].

AI 기반 네트워크 관리 연구에는 대규모 학습 데이터셋이 필수적이다. 그러나 실제 6G 네트워크 트래픽 데이터는 상용 서비스 전 단계로 인해 획득이 어렵고, 기존 LTE/5G 데이터는 보안 및 프라이버시 문제로 공개가 제한적이다. 기존 합성 트래픽 생성 연구는 단순 정규분포 기반 패턴을 사용하거나 슬라이스별 서비스 특성을 충분히 반영하지 못하는 한계가 있다[3].

따라서, 본 논문에서는 실제 셀룰러 트래픽의 핵심 통계적 특성을 반영하고 슬라이스별로 차별화된 패턴을 적용한 합성 트래픽 데이터셋 생성 방법을 제안한다. 생성된 데이터셋을 LSTM 기반 예측 모델 학습에 활용하여 AI 연구용 데이터셋으로서의 활용 가능성을 검증한다.

II. 본론

그림 1은 제안하는 합성 트래픽 데이터셋 생성 파이프라인을 보여준다. 파이프라인은 트래픽 특성 모델링, 슬라이스별 패턴 적용, 노이즈 추가 단계로 구성된다.

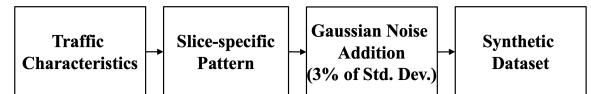


그림 1. 합성 트래픽 데이터셋 생성 파이프라인

생성된 트래픽 데이터는 실제 셀룰러 트래픽의 네 가지 핵심 특성을 반영한다. 첫째, Poisson 기반 세션 도착 패턴을 적용하여 모바일 사용자 세션의 무작위 발생을 모사한다[3]. 둘째, Heavy-tail ON/OFF 분포를 적용하여 세션 지속시간의 Burstiness 와 Long-range dependence 를 반영한다[4]. 셋째, 24 시간 일정 패턴, 교통신호 기반 단기 주기 등 여러 시간 주기가 중첩된 Superimposed Periodicity 패턴을 적용한다[5]. 넷째, 날씨, 사고, 지역 이벤트 등으로 인한 급증 및 급감 패턴을 Event-driven Abrupt Changes 로 모델링한다[5].

각 슬라이스별로 서비스 특성에 맞는 트래픽 패턴을 차별화하여 생성한다. C-V2X 슬라이스는 Rush-hour 집중 패턴과 이동성 기반 변동, 신호등 및 주행 경로에 따른 단기 주기적 특성을 반영한다. eMBB 슬라이스는 스트리밍 및 대용량 전송에 해당하는 고대역폭 기반 트래픽으로, 단기 burst 와 함께 평균 throughput 이 높은 패턴을 적용한다. IoT 슬라이스는 주기적이고 저변동량의 대규모 소량 패킷 패턴을 반영한다.

III. 실험

생성된 데이터셋은 1 달 분량으로 구성되며, 1 초 단위 리샘플링을 통해 총 2,592,000 개의 샘플을 확보하였다. 현실적인 미세 변동을 포함시키기 위해 각 슬라이스별로 신호의 표준편차 대비 3%에 해당하는 가우시안 노이즈를 추가하였다.

그림 2는 제안한 방법으로 생성된 7일간의 슬라이스별 트래픽 패턴을 보여준다. C-V2X 슬라이스는 출퇴근 시간대(Rush-hour)에 명확한 피크를 보이며, eMBB 슬라이스는 저녁 시간대에 스트리밍 트래픽 증가로 인한 피크가 나타난다. IoT 슬라이스는 낮은 변동성과 안정적인 패턴을 유지하여 각 슬라이스의 서비스 특성이 반영되었음을 확인할 수 있다.

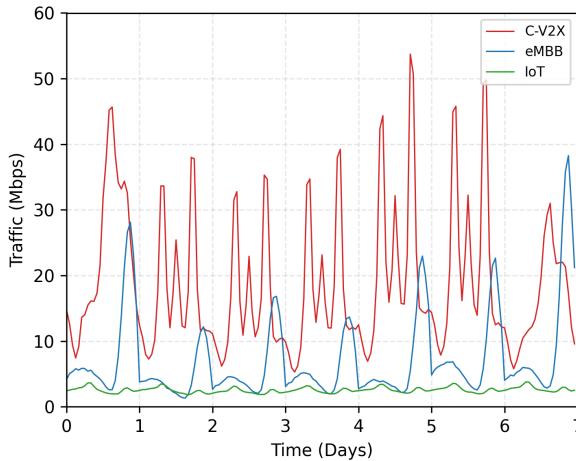


그림 2. 생성된 슬라이스별 합성 트래픽 (7일)

표 1. 실험 환경

Component	Specification
CPU	Intel i7-12700K
RAM	DDR4 32GB
Python	3.10
TensorFlow	2.20.0

표 1은 생성된 데이터셋의 활용 가능성을 검증하기 위해 사용한 하드웨어와 소프트웨어의 스펙을 보여준다. 트래픽 예측 모델은 LSTM 기반으로 학습 하였다. 모델은 2-layer LSTM(128→64 units)으로 구성하였으며, 지난 60 초 데이터를 입력으로 미래 20 초 트래픽을 예측하는 구조이다. 데이터셋은 Train 70%, Validation 15%, Test 15%로 분할하였다.

Slice	Accuracy
IoT	97.31%
eMBB	91.84%
C-V2X	90.46%

표 2는 슬라이스별 예측 성능을 보여준다. IoT 슬라이스는 규칙적인 트래픽 패턴으로 인해 97.31%의 가장 높은 예측 정확도를 달성하였다. eMBB 슬라이스는 중간 수준의 변동성을 가지며 91.84%의 안정적인 예측 정확도를 보였다. C-V2X 슬라이스는 높은 변동성과 노이즈로 인해 상대적으로 낮은 90.46%의 정확도를 보였으나,

추세 판단 및 자원 증설 의사결정에는 충분한 예측 품질을 확보하였다.

IV. 결론

본 논문에서는 6G 네트워크 슬라이싱 연구를 위한 합성 트래픽 데이터셋 생성 방법을 제안하였다. 실제 셀룰러 트래픽의 통계적 특성을 반영하고 슬라이스별 차별화된 패턴을 적용하여 현실적인 데이터셋을 생성하였으며, LSTM 기반 예측 모델 학습을 통해 AI 연구용 데이터셋으로서의 활용 가능성을 확인하였다.

향후 연구에서는 실제 네트워크 트래픽 데이터와의 통계적 유사성 비교 분석 및 다양한 AI 모델에 대한 범용성 검증을 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초 연구사업임(No. RS-2020-NR049604). 또한, 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. RS-2024-00397979, 6G 수십 테라급 초정밀 전달망 시스템 기술 개발).

*교신저자: 최성곤(choisg@cbnu.ac.kr)

참고 문헌

- [1] 5G Americas, "Artificial Intelligence and Cellular Networks," White Paper, Dec. 2024.
- [2] S. T. Arzo, C. Naiga, F. Granelli, R. Bassoli, M. Devetsikiotis, and F. H. P. Fitzek, "A Theoretical Discussion and Survey of Network Automation for IoT: Challenges and Opportunity," IEEE Internet of Things Journal, vol. 8, no. 15, pp. 12021–12045, 2021.
- [3] R. K. Polaganga and Q. Liang, "Self-Similarity and Modeling of LTE/LTE-A Data Traffic," Measurement, vol. 75, pp. 218–229, 2015.
- [4] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, "Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level," IEEE/ACM Trans. Networking, vol. 5, no. 1, pp. 71–86, 1997.
- [5] F. Xu, Y. Li, H. Wang, P. Zhang and D. Jin, "Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment," in IEEE/ACM Transactions on Networking, vol. 25, no. 2, pp. 1147–1161, April 2017.