

기계학습 기반의 단기 모바일 트래픽 예측기법

윤 문 섭*, 권 혜 연**, 박 승 근**, 장 한 승*

ung172726@gmail.com, hykwon@etri.re.kr, seungkp@etri.re.kr, hsjang@jnu.ac.kr

Machine Learning Based Short-term Mobile Traffic Estimation Scheme

Munseop Yun, Hye Yeon Kwon, Seug Keun Park, Han Seung Jang

*전남대학교 전기·전자통신·컴퓨터공학부

**한국전자통신연구원

요 약

본 논문에서는 기계학습을 이용한 표본셀 단기 트래픽 예측 방법을 고안하였다. 구체적으로는 양방향 장단기 메모리(Bidirectional LSTM) 학습 모델을 적용하여 국내 이동통신 3사의 트래픽 데이터를 바탕으로 6개월치 단기 표본셀 트래픽 예측을 진행하였다. 외부 데이터 없이 트래픽 값만을 단기 예측기술의 입력 데이터로 사용했다. 입력 시퀀스 길이에 따라서 1개월 전에서부터 6개월 전까지의 데이터를 이용해 향후 6개월의 이동 3사 및 총 트래픽을 예측하는 모델과 모델을 구성하는 기술을 설명한다. 연구에서는 대표적으로 이동 통신사 3사 데이터를 합친 총 트래픽의 6개월치 예측 성능을 보여주며, 예측 결과는 개월수가 늘어날수록 점차 증가하는 경향을 보였다.

I. 서 론

네트워크 트래픽의 사용량은 셀룰러 통신 시스템 구축 및 확장에 가장 중요한 요소로 작용한다. 트래픽 밀집 지역에는 원활한 서비스를 위해 기지국 밀도를 향상시키거나, 주파수 대역폭 등의 기지국 용량을 증대할 필요가 있다[1]. 그러나 네트워크 트래픽은 공간적으로만 영향을 받는 것이 아니라, 시간적으로도 변화 추이가 뚜렷한 시계열 특성을 보인다. 인구 밀집 지역이라도 한산한 시간대에는 트래픽이 크게 집중되지 않으며, 유동 인구가 많지 않은 지역도 출퇴근 및 업무 시간과 같은 특정 시간대에는 트래픽이 밀집될 수 있다. 하지만, 시계열 특성을 보이는 트래픽 데이터를 정확히 예측하는 것은 매우 고난도의 문제로 기존의 통계적 예측 기술로는 해결하기가 쉽지 않다. 이를 위해 네트워크 시스템에서 수집되는 빅데이터를 기반으로 기지국의 트래픽 변화 추이를 추정하는 기계학습 기반 데이터 전처리 및 딥러닝 기술 연구가 필요하다[2]. 따라서 본 논문에서는 국내 이동 통신사의 최번시(busy hour) 트래픽 데이터를 기반으로 6개월치 단기 표본셀 트래픽 데이터 예측을 진행하였다. 구체적으로 양방향 장단기 메모리(Bidirectional LSTM: B-LSTM) 학습 모델을 적용하였고 입력 시퀀스 길이를 1개월에서부터 6개월까지 변화해 가며 향후 6개월의 트래픽량을 예측하는 모델을 제시하고 평가하였다.

II. 본론

본 논문에서는 Artificial Neural Network (ANN)에서 파생된 순환신경망(Recurrent Neural Network, RNN)[3]의 일종인 B-LSTM을 이용한 시계열 데이터 분석 모델의 구현을 통해, 보다 정확하고 효율적인 방식으로 표본셀 트래픽에 대해 단기예측을 수행하는 모델을 구성하여 향후 6개월의 트래픽을 예측하는 모델을 만들었다. 모델 학습을 위한 데이터는 이동 3사로부터 17년 3월부터 20년 6월까지 수집된 RB 사용률 기준 상위 10% 표본 셀 트래픽 데이터를 사용하였다. 여기에 추가적으로 3사 트래픽 데이터를 합하여 매달 총 4개의 데이터를 사용하였으며, 총 40×4 의 40개월의 데이터 중 34×4 데이터는 학습(training)과 6×4 데이터는 시험

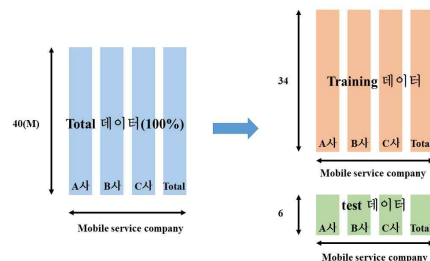


그림 1. 모델의 학습 및 시험 데이터 구성

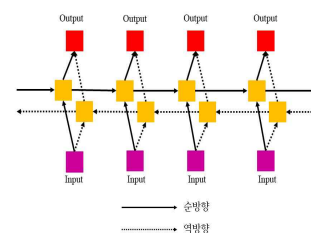


그림 2. 양방향 장단기 메모리(Bidirectional LSTM)의 구조

(test)을 위해 그림 1과 같이 구성하였다[4]. 데이터 양이 충분치 않아 입력 데이터의 구성에서 검증용 데이터는 따로 두지 않았으나, 시험용 데이터를 검증처럼 활용해 검증 손실 값이 가장 적었을 때의 가중치를 불러와 나머지 향후 6개월을 예측하는데 사용하였다.

예측 모델에는 순환신경망 모델을 이용한다. 순환신경망 모델은 사슬처럼 이어지는 성질을 가지며 기존 인공신경망과는 다르게 은닉층(hidden layer)을 통해 과거부터 현재까지의 데이터를 기억해낼 수 있으며, 이전 신경 노드에서 전달받은 가중치를 이용하여, 새로운 입력에 대해서 예측을 하고 네트워크에서 기억을 조금씩 수정한다. 본 연구에서는 순환신경망 모델 중에서도 양방향 장단기 메모리 모델을 사용한 예측 모델을 제안한다. 기존의 순환신경망 모델의 구조는 순방향으로의 순서만 고려했다면, 그림 2에서처럼 양방향 장단기 메모리 모델은 역방향도 고려하는 방식을 사용한다[5].

다음으로, 훈련을 위해 아래와 같은 평균제곱오차(Mean Squared Error, MSE) 손실값(loss) 모델을 사용한다.

$$\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2,$$

여기서 \hat{y}_i 및 y_i 는 i 번째 데이터에 대한 예측값과 실제값을 의미하고 n 은 학습용 데이터의 총 개수를 의미한다. 학습은 손실 값을 최소화하는 방식으로 학습이 진행된다. 그리고 모델의 입력 데이터의 길이는 1개월에서부터 6개월까지 시퀀스 길이(개월 수)를 변경해가는 방법을 사용했다. 따라서, x 개월 시퀀스 길이를 사용할 경우 입력 데이터의 크기는 $x \times 4$ 가 된다. 학습에서는 시퀀스 길이를 변화하여 더 정확한 예측 결과를 찾을 수 있도록 하였다. 예를 들어 시퀀스 길이 x 가 2일 경우에는 과거 2개월치 데이터를 이용하여 향후 트래픽 값을 예측하는 것을 의미한다. 모델에 사용한 파라미터들은 시퀀스 별로 동일하게 epochs = 2500, learning rate = 0.01, unit = 5로 모델을 구성했다. 모델을 훈련 시키는 과정에서는 모델의 가중치 값들을 저장시킨 후, 가장 손실 값이 적었던 가중치 값을 가져와 모델에 적용시키는 방식으로 예측을 진행했다.

III. 실험 결과

예측에 대한 평가는 아래 식과 같은 절대 평균 오차율을 통해서 검증했다.

$$\frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100,$$

여기서 \hat{y}_i 및 y_i 는 i 번째 데이터에 대한 예측값과 실제값을 의미하고 N 은 시험용 데이터의 총 개수를 의미한다. 향후 6개월 예측의 정확도를 높이기 위해 검증의 손실 값이 낮았던 지점의 가중치와 시퀀스 길이를 사용하였다. 표 1과 2는 제안 방식과 완전 연결 인공신경망(Fully-Connected Neural Network, FNN)방식의 6개월 예측에 관한 절대 평균 오차율 성능을 각각 보여준다. 결과적으로 제안한 모델의 경우 이동 3사 총 트래픽 예측은 2개월(시퀀스 길이 2)의 데이터를 이용해 향후 6개월 데이터를 예측했을 때, 가장 오차율이 낮음을 확인할 수 있다. 예측 절대 평균 오차율은 3.48%로 6개월 예측 구간 안에서 10% 미만의 절대 평균 오차율 성능을 보였다. 하지만, FNN은 6개월 예측 평균으로는 5.52%의 성능을 보이지만 특정 개월에서는 10%를 초과하는 오차 성능을 보여주는 지점이 발생했다. 그림 3은 각 기술의 단기 트래픽 예측 결과를 보여준다. 34개월부터 40개월에 해당하는 지점은 검증 데이터 학습을 통해 나온 예측 결과이며, 본 연구에서 제안한 모델이 FNN 모델보다 실제값과 더욱 유사한 예측 결과를 보여준다. 다음으로, 41개월부터 46개월에 해당하는 지점은 마지막 데이터 수집 시점으로부터 향후 6개월의 예측 결과를 나타내었다. 예측 결과를 바탕으로 향후 3개월까지는 데이터 트래픽 증가가 있으며 이후 트래픽 사용량이 줄어든 것으로 예상된다.

표 1. 제안 방식의 3사 총 트래픽 절대 평균 오차율(%)

	1개월	2개월	3개월	4개월	5개월	6개월	평균
seq=1	5.72	5.49	2.77	4.13	4.74	2.77	4.28
seq=2	6.19	4.61	1.12	4.19	3.88	0.88	3.48
seq=3	6.13	9.35	6.22	0.10	2.26	6.57	5.11
seq=4	12.31	13.16	0.45	0.45	0.74	5.57	6.88
seq=5	7.27	8.77	7.74	7.74	1.88	4.57	6.78
seq=6	7.79	5.47	3.94	3.94	4.12	7.08	5.48

표 2. FNN 방식의 3사 총 트래픽 절대 평균 오차율(%)

	1개월	2개월	3개월	4개월	5개월	6개월	평균
seq=1	8.87	10.86	15.40	1.36	8.21	8.01	8.79
seq=2	8.47	11.82	15.66	0.11	8.11	7.55	8.62
seq=3	4.27	8.47	11.54	5.40	1.84	1.61	5.52
seq=4	0.15	2.33	4.71	14.75	6.92	8.04	6.15
seq=5	4.71	4.24	6.79	13.95	6.87	8.72	7.55
seq=6	11.41	11.61	15.36	4.14	0.79	1.17	7.41

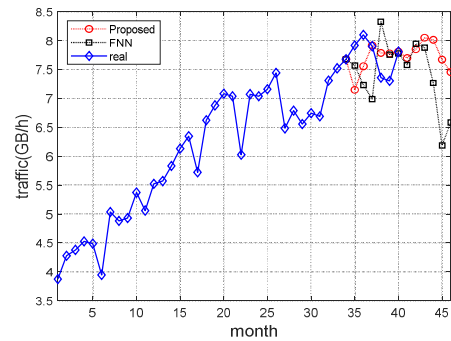


그림 3. 단기 트래픽 예측 결과

IV. 결론

본 논문에서는 양방향 장단기 메모리(Bidirectional LSTM) 모델을 활용하여 단기 트래픽 예측을 수행하였다. 입력 데이터 길이를 1개월부터 6개월까지 시퀀스 길이를 변경해가며 학습을 통해 더욱 정확한 예측 결과를 찾았다. 제안된 모델과 FNN 모델의 6개월 동안의 절대 평균 오차율을 비교하였을 때, 제안된 방식의 오차율이 더 낮은 것을 확인할 수 있었다. 본 연구에서 제안된 단기 트래픽 예측기술은 이동 통신사들의 셀 구축 및 확장에 중요한 정보로 사용될 것으로 예상된다.

ACKNOWLEDGMENT

이 논문은 2021년 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00092, 주파수 확보 및 공급 기반 기술개발)

참고 문헌

- [1] 국내 이동통신 트래픽 추세 분석", 한국전자통신연구원, 2018.06
- [2] Le Nguyen, Phi, and Yusheng Ji. "Deep Convolutional LSTM Network-Based Traffic Matrix Prediction with Partial Information," In Proc. 2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), 2019.
- [3] Schuster, Mike, and Kuldip K. Paliwal. "Bidirectional Recurrent Neural Networks." IEEE Transactions on Signal Processing vol. 45, no. 11, pp. 2673-2681, Nov. 1997.
- [4] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. Vol. 1. No. 2. Cambridge: MIT press, 2016.
- [5] Cheng, Hongju, et al. "Data Prediction Model in Wireless Sensor Networks Based on Bidirectional LSTM." EURASIP Journal on Wireless Communications and Networking 2019.1 (2019): 1-12.