

모바일 트래픽 예측 성능 향상을 위한 시스템 연구

방석영, 홍인기
경희대학교

qkdtjr97@khu.ac.kr, ekhong@khu.ac.kr

A study on the system for improving mobile traffic prediction performance

Seok-Yeong Bang, Een-Kee Hong

Kyunghee University

요 약

폭발적인 트래픽 증가로 인해 모바일 네트워크의 정체가 빈번하게 일어나고, 이를 수용하기 위한 시스템 용량 증가 및 효율적인 트래픽 운영 방식에 대한 필요성이 증가하고 있다. 이러한 문제를 해결하기 위해 트래픽 예측을 통한 대처 및 이를 이용한 무선 자원 할당 방안이 제안되었다. 본 논문은 트래픽 예측 성능을 향상시키기 위하여 기존의 원본 데이터를 활용한 예측이 아닌 원본 데이터를 전처리 하여 예측을 진행하는 예측 시스템을 제안하였다. 제안하는 예측 시스템은 전처리부와 예측부로 나뉘는데, 전처리부에서는 STL(Seasonal Trend Loess) 분해법을 통해 원본 데이터를 데이터 특성에 따라 분해하여 그중 트렌드(Trend) 데이터와 계절성(Seasonal) 데이터를 예측부의 입력 데이터로 사용하고, 예측부에서는 각각의 입력 데이터에 맞게 딥러닝 모델을 최적화하여 원본 데이터를 예측한다. 시뮬레이션 결과 제안하는 예측 시스템의 예측 성능이 기존의 원본 데이터에 전처리를 하지 않았을 때에 비해 우수한 성능을 보였다. 제안하는 예측 시스템은 무선 자원 할당 등 네트워킹 장비 운영에 유용하게 사용될 수 있을 것으로 기대된다.

I. 서 론

모바일 기술과 셀룰러 네트워크는 날마다 발전하고 있으며, 이에 따라 스마트폰, 태블릿, 웨어러블 기기과 같은 모바일 장비의 수와 휴대 전화 가입자의 수는 빠르게 증가하고 있다. CISCO에서 발표된 보고서에 의하면 전 세계 휴대 전화 가입자 수는 2015년 51억 명에서 2023년까지 57억 명으로 세계 인구수의 71%가 될 것이며, 네트워크 장치는 293억 개로 전 세계 인구수의 3배 이상에 달하는 수가 될 것이다[1]. 또한 전 세계 모바일 트래픽 데이터는 2022년까지 한 달에 77엑사바이트(Exabyte)를 생성할 것이며, 연간 트래픽 생성량은 1제타바이트(Zettabyte)에 가까울 것이다[2].

폭발적인 트래픽 증가로 인하여 모바일 네트워크의 정체는 빈번하게 일어나고, 이를 수용하기 위한 시스템 용량 증가 및 효율적인 트래픽 운영 방식에 대한 필요성이 증가하고 있다. 네트워크는 계속 증가하는 트래픽 요구에 대처해야 하며, 상당한 수의 휴대 전화 가입자들에게 양질의 서비스를 제공해야 한다. 특히, 모바일 트래픽은 항상 일정하게 발생하기보다는 특정 시간, 특정 장소에서 폭발적으로 증가하는 경향이 있기 때문에 대부분의 모바일 네트워크는 데이터가 폭증할 때도 원활히 운영될 수 있도록 용량을 확보하고 있다. 따라서 대부분의 시간이나 장소에서 용량을 남지만 트래픽이 폭증할 때 적절한 용량이 확보되어 있지 않다면 문제가 발생할 수 있다. 이러한 문제를 해결하기 위해 만약 특정 시간대에 발생할 트래픽에 대한 예측이 가능하다면 이에 대하여 적절히 대처할 수 있고, 무선 자원 할당 등 네트워크 장비의 운용을 효과적으로 수행할 수 있다.

최근 많은 연구들이 트래픽 예측을 위해 딥러닝 기반 예측 모델을 사용하여 모바일 트래픽을 예측하는 문제를 다루고 있다. 하지만 이런 연구들의 대부분은 예측 모델 자체를 고도화하거나 위치 정보를 이용한 클러스터링을 활용하는 형태를 보이고 있다. 이런 연구의 문제점은 데이터에 민감하다는 것인데, 모바일 트래픽은 변칙적인 확률이 높기 때문에 단순히 예측 모델만의 연구로는 한계가 있다. 이를 해결하는 방법으로 데이터 전처리 과정을 통해 데이터를 가공한 이후에 딥러닝 예측 모델의 입력으로 사용하는 시스템을 설계하는 방법이 있다. [3]은 데이터 전처리 방법으로 시계열 분해를 사용해서 계절성(Seasonal) 데이터만을 ARIMA를 통해 예측했다. 모바일 트래픽에서는 계절성 데이터 뿐만 아니라 트렌드

(Trend) 데이터도 전체 트래픽의 큰 부분을 차지하고 있기 때문에 해당 방법은 모바일 트래픽을 예측할 때 사용할 수 없다. 따라서 본 논문에서는 딥러닝을 통해 예측하기 전에 전처리 과정으로 STL(Seasonal Trend Loess) 분해법을 통해 특성에 따라 모바일 트래픽을 분해한 후, 그중 트렌드, 계절성 데이터를 딥러닝 예측 모델의 입력으로 사용하여 예측을 진행하는 시스템을 제안한다.

II. 본론

본 논문에서 제안하는 예측 시스템은 데이터를 특성에 따라 분해하는 전처리부와 분해된 데이터를 입력 데이터로 받아 딥러닝 모델 최적화 및 예측을 진행하는 예측부로 나누어져 있다. 전처리부에서 사용된 시계열 분해 방법은 STL 분해법인데, STL 분해법은 고전적인 분해, SEATS 분해, X11 분해를 뛰어넘는 몇 가지 장점을 가지고 있다[4]. STL 분해법은 계절 성분의 변화율을 사용자가 조절할 수 있어 계절적인 성분이 시간에 따라 변해도 분해가 용이하고, 가끔 있는 이상값이 추세-주기와 계절 성분에 영향을 주지 않게 이상값을 잔여(Residual) 성분으로 분해한다. 이상값이 잔여 성분으로 분해되기 때문에 트렌드 데이터는 장기적인 트래픽의 상승 및 하락하는 패턴으로 모바일 트래픽의 전반적인 변화를 보여주고, 계절성 데이터는 반복하는 시간 패턴으로 모바일 트래픽에의 하루 단위 변화율을 보여준다. 따라서 전처리 과정의 데이터 분해 과정을 통해 데이터의 특성에 따라 분해된 데이터 중 트렌드와 계절성 데이터를 딥러닝 모델의 입력 데이터로 선택한다.

제안하는 예측 시스템의 예측부에서는 딥러닝 데이터의 입력 데이터로 선택된 트렌드 데이터와 계절성 데이터 각각에 맞게 딥러닝 모델을 선택한 후, 모델 파라미터 최적화를 진행하였다. 전처리 과정 이후 최적화된 모델을 통해 트렌드 데이터와 계절성 데이터 예측을 진행하는데, 트렌드 데이터와 계절성 데이터를 예측한 결과를 합하여 원본 데이터에 대한 예측을 진행한다.

본 논문에서 제안하는 예측 시스템의 성능을 비교하기 위해 SKT에서 제공한 강남역에 위치한 기지국의 2020년 8월 한달 간 1시간 단위로 집계된 모바일 트래픽 데이터 중 임의로 하나를 선택하여 예측을 진행하였다. 정보 보호를 위해 선택된 모바일 트래픽 데이터에 가중치를 곱하여 데이

터양을 변경하였다.

전처리 과정을 거치지 않은 원본 데이터를 예측하는 모델로는 [5]에서 제안한 것과 같이 LSTM을 통해 예측을 진행하였고, 모델 파라미터는 Hidden State = 64, Hidden Layers = 2로 최적화하여 예측을 진행하였다. 본 논문에서 제안하는 예측 시스템도 트렌드 데이터와 계절성 데이터의 딥러닝 모델 최적화를 진행하였는데, 딥러닝 모델로는 LSTM과 CNN을 사용하였다.

표 1. 딥러닝 모델 최적화를 위한 성능 비교

데이터	CNN	LSTM
트렌드 데이터	1493	1217
계절성 데이터	21086	23901

표 1은 딥러닝 모델의 입력 데이터로 선택된 트렌드 데이터와 계절성 데이터의 예측 모델을 선택하기 위해 LSTM과 CNN으로 예측을 진행한 결과를 MAE(Mean Absolute Error)로 나타낸 결과이다. 예측 결과는 총 30번 시행을 통해 얻은 결과를 평균 낸 결과로, 모바일 트래픽의 트렌드 데이터는 단기적인 변화보다 장기적으로 변하는 패턴을 보이므로 LSTM에서 더 좋은 예측 성능을 보이고, 계절성 데이터는 하루 단위로 비슷한 양상을 보이지만 분산이 변하므로 지역적인 특징을 추출하는 CNN에서 더 좋은 예측 성능을 보인 것을 확인할 수 있다. 예측 결과에 따라 트렌드 데이터는 LSTM, 계절성 데이터는 CNN으로 딥러닝 모델을 결정하였고, 최종적으로 제안하는 시스템의 구조는 그림1과 같다.

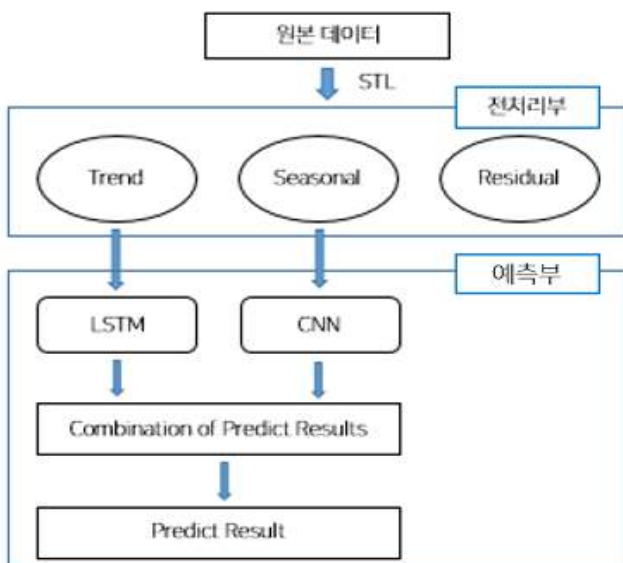


그림 1. 제안하는 시스템의 구조도

앞서 설명한 점들을 토대로 시뮬레이션을 진행하였고, 표 2에 구체적인 예측 오차의 수치 및 예측 시간을 제시하였다. 전처리를 하지 않고 원본 데이터를 예측한 경우 'Basic', 본 연구에서 제안하는 예측 시스템으로 예측한 경우 'Proposed'로 표기하였다. 예측 결과 본 연구에서 제안하는 예측 시스템을 통해 예측한 결과가 전처리 과정을 거치지 않은 원본 데이터를 예측한 결과에 비해 14.2% 예측 결과가 높은 것을 확인할 수 있고, 모델 학습에 걸린 시간은 27.7% 빨라졌음을 확인할 수 있었다. 그림 2의 그래프를 통해 예측 결과를 확인해보면 검은색 그래프가 원본 데이터, 초록색 그래프가 Basic, 파란색 그래프가 Proposed일 때, 피크치가 있는 경우를 제외한 모든 구간에서 Proposed가 Basic보다 원본 데이터를 더 잘 예측하는 것을 확인할 수 있었다.

표 2. 제안하는 예측 시스템의 예측 성능 비교

예측 방식	예측 성능(MAE)	걸린 시간(s)
Basic	26818	22
Proposed	23006	15.9

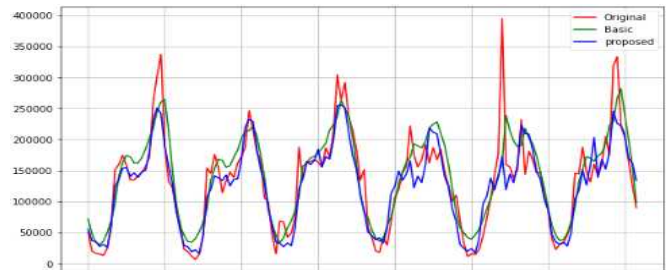


그림 2. 트래픽 예측 결과

III. 결론

본 논문에서는 딥러닝을 통해 예측하기 전에 전처리 과정으로 STL(Seasonal Trend Loess) 분해법을 통해 특성에 따라 모바일 트래픽을 분해한 후, 그중 트렌드, 계절성 데이터를 딥러닝 예측 모델의 입력으로 사용하여 예측을 진행하는 시스템을 제안하였다. 제안하는 시스템을 통해 원본 데이터를 특성에 따라 분해하여 데이터의 특성에 맞게 딥러닝 모델을 최적화 하였을 때, 모바일 트래픽을 더 빠르게 정확하고 예측할 수 있음을 확인하였다. 모바일 트래픽의 예측을 통해 무선 자원 할당 등 네트워크 장비 운영을 효율적으로 수행할 수 있는데, 본 논문에서 제안하는 예측 시스템을 활용해 무선 자원 관리의 성능을 향상시킬 수 있을 것으로 예상된다.

ACKNOWLEDGMENT

"본 연구는 한국전자통신연구원 NR 기반 지능형 오픈 스몰셀 기술 개발 과제의 위탁연구과제로 수행한 연구결과로 수행되었음" (ETRI-2018-0-0-01659)

참 고 문 헌

- [1] Cisco, U. (2020). Cisco annual internet report (2018-2023) white paper.
- [2] Forecast, G.M.D.T (2019). Cisco visual networking index: global mobile data traffic forecasting update, 2017-2022. Update, 2017, 2022.
- [3] Mingyan Zhang, Haohao Fu, Sheng Chen. (2019) "Understanding urban dynamics from massive mobile traffic data", in IEEE Transactions on Big Data, 5(2), 266-278
- [4] Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on <current date>
- [5] Hong Duy Trinh, Lorenza Giupponi, Paolo Dini. (2018) "Mobile Traffic Prediction from Raw Data Using LSTM Networks". in 2018 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)