

효율적인 전용회선 자원 사용량 예측을 위한 통계적 기법과 기계학습 모델 비교 연구

이인규*, 송미화*

*세명대학교 정보통신학부

db999@chol.com, mhsong@semyung.ac.kr

A Comparative Study of Statistical Techniques and Machine Learning Models for Efficient Leased Line Resource Usage Prediction

In-Gyu Lee*, Mi-Hwa Song*

*School of Information and Communication Science, Semyung University

요 약

전용회선은 두 지역을 독점적으로 사용하는 구조이기 때문에 안정된 품질수준과 보안성이 확보되어 교환 회선의 급격한 증가에도 불구하고 지속적으로 많이 사용하는 회선 방식이다. 하지만 비용이 상대적으로 고가이기 때문에 네트워크 전용회선의 자원을 적절히 배치하고 활용하여 최적의 상태를 유지하는 것이 중요한 요소이다. 이에 본 연구에서는 기업 네트워크에서 사용하는 전용회선의 실제 사용률 데이터를 기반으로 다양한 시계열 데이터 예측 모델을 적용하고 성능을 평가하였다. 일반적으로 통계적인 방법으로 많이 사용하는 평활화 모형 및 ARIMA 모형과 요즘 많은 연구가 되고 있는 인공지능망에 기반한 딥러닝의 대표적인 모델들을 적용하여 각각의 예측에 대한 성능을 측정하고 비교하였다.

1. 서론

데이터 전송을 위한 회선 방식은 전용회선 방식과 교환회선 방식으로 분류된다. 전용회선 방식은 점대점(point-to-point) 방식으로 두 지점을 배타적으로 연결하고 독점적으로 자원을 사용하는 방식으로 회선 사용에 대한 독립성이 확보되며 안정성과 보안성이 우수하다는 장점이 있다.

하지만 자원을 공유하여 사용하는 교환회선 방식과 비교하여 비교적 고가이므로 회선 용량 계획을 통하여 불필요한 비용의 낭비를 막고 부족한 자원에 대해서는 추가 자원을 투입하여 서비스 제공에 문제가 없도록 하는 일련의 네트워크 관리가 필수적인 요소이다. 이러한 관리를 위해서는 시계열 특성을 갖는 전용회선 사용량에 대하여 과거의 트래픽량을 기준으로 향후의 트래픽량을 예측하는 방법이 필요한데, 전통적으로 통계 기반 모형을 이용하는 방법과 데이터의 학습에 기반한 딥러닝 기법으로 분류할 수 있다.

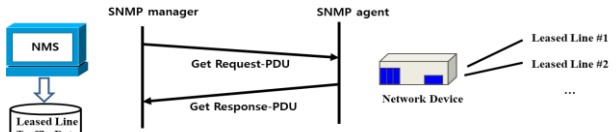
이러한 전용회선의 자원사용량 예측과 관련된 과거 연구들은 주로 통계적 방법을 사용하고 있으며 인공지능에 기반한 딥러닝 기법을 같이 적용 후 효과를 비교 분석한 사례는 찾기가 어렵다. 따라서, 본 연구

에서는 네트워크 관리 프로토콜인 SNMP 및 MIB-2를 이용하여 전용회선 사용량의 데이터를 수집하고 과거 트래픽량을 기준으로 향후의 트래픽량을 예측하는 시계열 분석 모형을 전통적 통계 기반 모형과 데이터 학습에 기반한 딥러닝 기법으로 분석한다. 통계적인 관점에서 과거 데이터의 평균을 이용하여 예측하는 방법인 평활법과 비정상적 요소를 제거하고 자기 상관을 이용하는 ARIMA 모형을 이용한다. 또한 요즘 인공지능 관점에서 많이 연구되고 있는 딥러닝 알고리즘에서는 시계열 데이터 처리에 적합한 RNN 과 LSTM 알고리즘을 적용하여 분석한다. 각 모델을 대상으로 예측한 결과에 대하여 정확도 및 성능을 평가하기 위해 많이 사용하는 MSE 를 사용하여 각각을 비교, 평가한다.

2. 관련 연구

네트워크 장비에 연동된 전용회선의 트래픽 데이터를 수집하기 위해서는 TCP/IP 기반의 SNMP 를 사용한다[1][2][3]. 데이터 수집을 위하여 사용되는 전용회선은 국내 한 신용카드사의 승인처리를 위하여 사용하는 실제 데이터를 대상으로 하며, 전용회선이 연결

된 네트워크 장비에 SNMP 프로토콜을 이용하여 전용회선 트래픽의 MIB-2 관련 객체 정보를 폴링 형태로 수집한다.



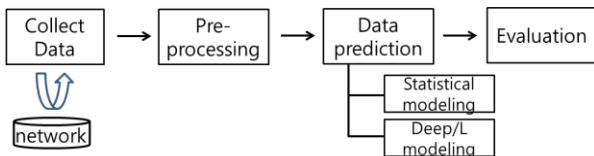
(그림 1) SNMP 를 이용한 전용회선 트래픽의 수집

수집된 정보는 시간의 순으로 정리된 시계열 데이터이며 패턴을 파악하고 미래의 데이터를 예측하기 위한 다양한 기법과 모델이 적용 가능하다. 우선, 시간에 따른 패턴을 파악하기 위해 많이 사용하는 방법 중 하나는 평활 기법이며 이동평균법과 지수평활법이 있다. 이동 평균법은 이전 데이터들의 평균으로 예측하는 개념이며 지수평활법은 최근의 데이터에 가중치를 더 두는 개념이다. ARIMA 모델은 자기회귀모델과 이동평균법을 합친 형태로 차분 단계를 거쳐 비정상성을 제거한 방법이다[4][5].

인공지능 분야의 딥러닝은 많은 데이터의 학습을 통하여 미래의 데이터를 예측하는 방법으로 RNN 과 RNN 모델의 구조를 개선하고 기울기 소실 문제등을 개선한 LSTM 모델이 있다[6].

3. 방법론

전용회선 사용량의 수집과 예측 방법들을 분석하기 위한 과정은 (그림 2)와 같다.



(그림 2) 전용회선 사용량 예측방법

데이터 수집을 위하여 SNMP 프로토콜을 이용하여 MIB-2 의 IfInOctet 와 IfOutOctet 값을 수집한다. 폴링 주기는 1 분으로 하고 1 개월 동안 40,000 개 이상의 시계열 데이터를 수집한다.

수집된 데이터는 전처리 과정을 거친다. 수집된 누적 데이터를 폴링 주기로 나누어 실시간 형태의 데이터로 바꾸고 이를 회선 속도로 나누어 사용률로 변환한다. 모델들의 성능평가를 위하여 통계적 모델링과 딥러닝 모델링을 나눈다. 본 논문에서는 이동평균법, 지수평활법, ARIMA, RNN, LSTM 에 대해서 평가를 진행한다.

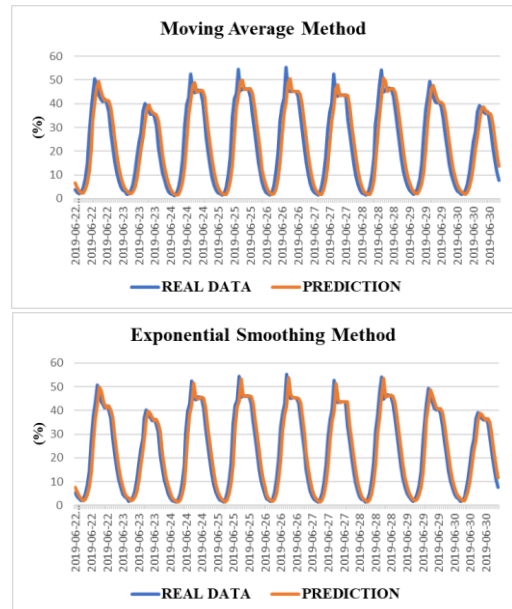
모델의 성능 평가를 위하여 수집된 데이터 중, 70%

는 학습데이터로, 30%는 테스트 데이터로 나누고 예측된 데이터와 테스트 데이터의 평균 제곱 오차를 산출하여 각 모델의 성능을 평가한다.

4. 실험 결과

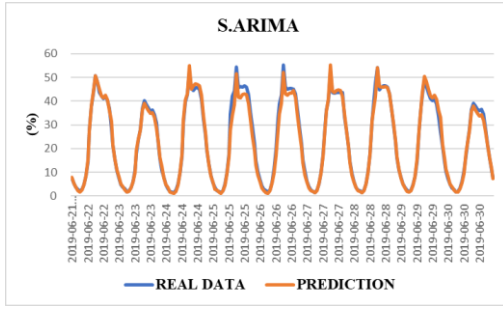
실험환경은 구글에서 제공하는 Colab pro 를 이용하였고 프로그램은 Python 3.6.9 및 TensorFlow 2.4.0 을 사용하였다.

평활화 방법 중 이동평균법은 전반적으로 데이터의 예측이 무난하지만 과거 데이터 평균을 적용하기 때문에 트래픽 피크치의 예측이 어렵고 과거의 데이터의 평균이 한 스텝 늦게 적용되는 형태를 보인다. 과거 데이터에 일정한 가중치를 부여하는 지수 평활법은 이동평균법보다 전체적으로 양호한 결과를 보인다. (그림 3)은 테스트 데이터에 대한 평활화 기법의 예측치의 결과이다.



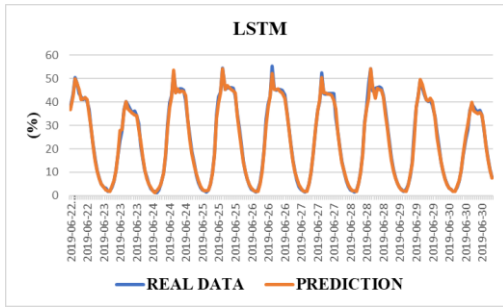
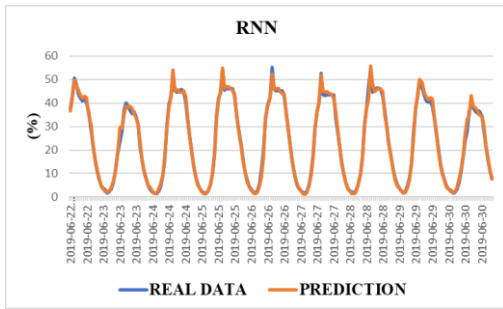
(그림 3) 평활화기법의 성능 평가 (상: 이동평균법, 하: 지수평활법)

평활화 모형의 단점은 전용회선 트래픽이 급격히 변화할 경우 적절한 예측이 부족하다는 문제가 있다. 평활화는 기본적으로 과거의 데이터를 적절히 평균화하여 반영하기 때문인데, 실제 전산 운영 환경에서 급격한 변화에 대한 최대값을 적절히 예측하지 못할 경우 전용회선 자원 부족으로 적절한 서비스가 어려운 문제가 생길 수 있다. 이에 반하여 과거 데이터에 대한 회귀 분석적인 방법을 이용하는 ARIMA 는 평활화 기법에 대하여 훨씬 개선된 결과를 나타낸다. (그림 4)는 테스트 데이터에 대한 계절성을 고려한 ARIMA 의 예측치 결과이다.



(그림 4) 계절 ARIMA 기법의 성능 평가

인공신경망 기반의 딥러닝 방법은 정상성 데이터로 변환할 필요가 없이 비정상성 구조의 데이터도 학습을 통하여 예측하는 방법을 이용하며 통계학적 모델의 성능보다 좋은 결과를 보인다. 또한 RNN의 경우 보다는 LSTM 방법이 더 좋은 결과를 나타내는 것으로 확인되어 RNN을 개선시킨 모델임을 확인하였다. (그림 5)는 테스트 데이터에 대한 딥러닝 방법의 예측치 결과이다.



(그림 5) 딥러닝 기법의 성능 평가
(상: RNN, 하: LSTM)

상기 실험 결과에 대하여 각 모델에 대한 MSE를 이용한 성능 평가의 결과를 요약하면 <표 1>과 같다.

〈표 1〉 성능평가 결과		
	모델, 방법	MSE
통계적 방법	평균이동법	54.32
	단순지수평활법	32.80
	ARIMA	2.87
딥러닝 방법	RNN	1.62
	LSTM	1.28

5. 결론

본 연구에서는 실제 기업의 운영환경에서 사용중인 전용회선의 데이터를 이용하여 시계열 데이터의 예측이 가능한 모델에 대해서 각각의 성능을 측정하였다. 데이터의 예측을 위한 방법은 시계열 패턴을 분석하고 상관관계, 추세 등의 특징을 파악하는 해석적 측면의 통계학적 방법과, 학습을 통하여 미래 예측데이터의 정확성을 높이는 딥러닝 방법으로 나눌 수 있으며, 각각에 대하여 통계학적 방법으로는 전통적인 평활법과 ARIMA 모델을 검토하였으며 딥러닝 방법에서는 RNN과 LSTM 모델의 성능을 검토하였다.

평활법에서는 지수평활법이 단순히 과거의 데이터의 균일한 평균을 사용하는 평균이동법 보다 좋은 성능을 나타냈으나 전반적으로 과거데이터에 대하여 가중치를 둔 평균값이라는 한계성이 있다. 반면 ARIMA 모델은 평활법 보다 좋은 예측 성능을 보였다. 수집된 데이터의 추세 및 계절성에 대하여 차분을 이용하여 정상적인 데이터를 산출하고 일정한 주기로 반복되는 데이터 패턴을 분해 기법을 통하여 확인 후 적절한 파라미터를 적용한 후에 안정된 성능 결과를 확인하였다. 인공신경망의 딥러닝 방법에서는 통계적 방법의 결과보다는 훨씬 더 나은 성능 결과를 확인할 수 있었다. 또한 추세 및 계절성 같은 비정상성 데이터를 별도로 계산하지 않아도 스스로 학습을 통하여 정확한 예측치를 찾아내는 편리함을 제공하였다. 그리고 RNN 보다는 개선된 LSTM이 보다 좋은 성능을 보여줌을 확인하였다.

감사의 글

본 연구 진행에 도움을 주신 세명대학교 김식 교수님께 감사드립니다.

참고문헌

- [1] J.D. Case & M. Fedor. (1990). Simple Network Management Protocol (SNMP), RFC1157
- [2] K. McCloghrie & M. Rose (1991). Management Information Base for Network Management of TCP/IP-based internets: MIB-II, RFC1213
- [3] William Stallings. (1998). SNMP and SNMPv2: The Infrastructure for Network Management. IEEE Communications Magazine. 37-43
- [4] Box G. E. P, Jenkins G. M. (1976). Time Series Analysis: Forecasting and Control. Holden-Day
- [5] R. J. Hyndman, G. Athanasopoulos. (2018). Forecasting: principles and practice. Available from: <https://otexts.com/fpp2>
- [6] Colah's blog. (2015) Understanding LSTM Networks. Available from: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>