

## ISP의 OTT 트래픽 품질모니터링과 예측에 관한 연구

남창섭\*

### A Study on the Quality Monitoring and Prediction of OTT Traffic in ISP

Chang-Sup Nam\*

**요약** 본 논문은 급증하는 인터넷 트래픽예측을 위해 빅데이터와 인공지능기술을 이용하였다. 기존에 트래픽 예측에 관해 다양한 연구가 있었지만 최근 스마트폰이나 스트리밍 등 거대한 인터넷 트래픽을 유발하는 증가 요소를 반영하지는 못했다. 더불어 대용량 인기 게임 출시나 OTT(Over the Top)사업자의 신규 콘텐츠 제공과 같은 이벤트성 요소는 사전 예측이 더욱 어렵다. 이러한 특성으로 기존 방법으로는 ISP(Internet Service Provider)가 실시간적 서비스 품질관리나 트래픽 예측치를 네트워크 사업환경에 반영하기가 불가능하였다. 따라서 본 연구에서는 이러한 문제점을 해결하고자 기존 NMS와는 별개로 트래픽 데이터를 실시간적으로 탐색, 판별하여 수집하는 인터넷 트래픽 수집시스템을 구축하였다. 이를 통해 수집대상의 데이터를 자동등록할 수 있는 유연성과 탄력성을 확보하였으며 실시간 네트워크 품질모니터링을 가능하게 하였다. 또한 시스템에서 수집된 대량의 트래픽 데이터를 머신러닝(AI)으로 분석하여 OTT 사업자의 미래 트래픽을 예측하였다. 이를 통해 보다 과학적이고 체계적인 예측이 가능해졌으며 더불어 ISP 사업자 간의 연동 최적화와 대형 OTT 서비스의 품질확보가 가능할 수 있게 되었다.

**Abstract** This paper used big data and artificial intelligence technology to predict the rapidly increasing internet traffic. There have been various studies on traffic prediction in the past, but they have not been able to reflect the increasing factors that induce huge Internet traffic such as smartphones and streaming in recent years. In addition, event-like factors such as the release of large-capacity popular games or the provision of new contents by OTT (Over the Top) operators are more difficult to predict in advance. Due to these characteristics, it was impossible for an ISP (Internet Service Provider) to reflect real-time service quality management or traffic forecasts in the network business environment with the existing method. Therefore, in this study, in order to solve this problem, an Internet traffic collection system was constructed that searches, discriminates and collects traffic data in real time, separate from the existing NMS. Through this, the flexibility and elasticity to automatically register the data of the collection target are secured, and real-time network quality monitoring is possible. In addition, a large amount of traffic data collected from the system was analyzed by machine learning (AI) to predict future traffic of OTT operators. Through this, more scientific and systematic prediction was possible, and in addition, it was possible to optimize the interworking between ISP operators and to secure the quality of large-scale OTT services.

**Key Words** : artificial intelligence, big data, internet quality monitoring, internet traffic, Internet traffic prediction, Machine learning, network traffic,

### 1. 서론

코로나로 인해 전 세계인들의 생활이 온라인으로 변화되면서 인터넷 사용량은 전례 없이 증가하고 있다. 재택근무와 온라인 학습 그리고 방대한 규모의 커머스와 엔터테인먼트가 온라인으로 일어나고 있다. 이와 같은 사회적 현상에 따라 인터넷 사용량이 급증하면서 인터넷이 과연 지속적으로 증가하는 트래픽을 처리할 수 있을지에 대한 관심이 높아지고 있다. 더불어 모바일도 일상생활이 되면서 다양한 응용서비스가 저변 확대되고 네트워크 트래픽이 급증하는 또 하나의 원인이 되고 있다. 이로 인해 네트워크 트래픽에 대한 예측이 ISP(Internet Service Provider)나 네트워크 관리자에게 매우 중요한 이슈로 떠오르고 있다. 또한 최근에는 Youtube, Netflix 등 OTT (Over-the-top) 사업자의 증가로 인한 트래픽 폭등이 전체 네트워크에 영향을 미치게 되면서 트래픽 예측이 네트워크관리의 중요한 이슈가 되었다. 본 논문은 이러한 트래픽 현황을 효율적으로 모니터링하고 OTT 트래픽의 미래를 정확히 예측하기 위해 인터넷 트래픽 수집시스템을 구축하고 방대한 데이터를 기반으로 OTT 미래 트래픽을 예측하였다.

### 2. 선행연구

인터넷 및 모바일 트래픽 예측 연구에 많은 연구자들은 다양한 통계적 방법을 사용하여 시물레이션 형태로 연구가 수행되고 있다[1]. 기존의 트래픽 발생량 예측 연구 중 통계학 분야에서 시계열 모형 예측 방법으로 널리 쓰이는 ARIMA 모형을 사용한 연구 결과도 있지만 시계열 분석은 과거 시계열 자료의 패턴이 미래에도 지속적으로 유지된다는 가정하에서 예측을 수행하는 것이므로 가정을 만족할 경우에 한해서 정확한 예측을 수행할 수 있다는 제약이 있다[2][3]. 최근에는 머신러닝 알고리즘 중 순환신경망(RNN) 알고리즘으로 트래픽 발생량 예측 모델을 제안하는 연구가 국내, 외에서 진행되고 있다[4][5][6][13].

데이터 흐름 예측기법과 예측모형의 적합성 검증과 같은 이론적인 연구와 실제 통신망 관리나 인터넷 트래픽 등 네트워크를 타겟팅한 연구도 최근까지 지속되고 있다[7][10]. 상기 내용 중 일부 연구내용은 학내망 등에서 발생하는 소규모 트래픽을 대상으로 하였거나 수학적 모델을 비교, 연구하여 정확도나 처리시간 측면에서의 성능향상을 시켰음을 입증하는 내용으로 국한된다[9][11][12]. 이러한 다양한 트래픽 예측 연구 환경과는 달리 최근 실제 네트워크에서는 인터넷 트래픽 증감이 매우 다양한 요소에 의해 결정된다. 과거에는 스마트폰이나 스트리밍 같은 거대한 트래픽을 유발하는 요소가 없었다는 점과 대용량 인기 게임 출시나 OTT 사업자의 신규 콘텐츠 제공과 같은 사전파악이 어려운 이벤트성 요소가 없었다. 이러한 점을 생각하면 현재 트래픽 예측 연구에 필요한 요소는 더욱 다양하게 반영되어야 하고 이러한 예측 결과는 신뢰도를 유지할 수가 있을 것이다.

### 3. 트래픽 수집환경 구축

본 연구는 예측을 위해 우선 실제 ISP망에서 인터넷 트래픽 데이터를 실시간적으로 수집하는 인터넷 트래픽 수집 서버를 구축하였다. 그림1과 같이 수집 서버 구축을 통해 트래픽 모니터 예측을

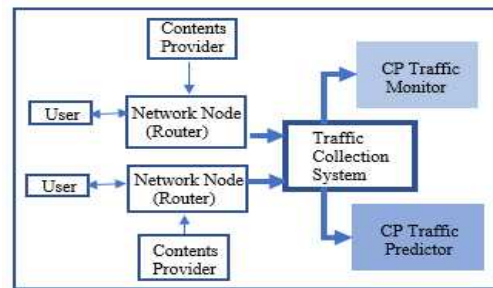


그림 1. 트래픽 수집 서버의 역할  
Fig. 1. Role of traffic collection server

모두 가능하도록 했다. 트래픽 수집 서버는 Azure 클라우드 플랫폼으로 구축하는 사례도 있지만 본

연구에서는 기존 망관리 시스템과는 별개로 그림 2와 같이 클라우드 기반의 가상서버로 구축하였고 운용시스템은 사용자가 쉽게 설치하고 운용이 가능하도록 설계된 Ubuntu OS를 사용하였다[8]. 또한 가상서버는 분산 환경을 쉽게 구축할 수 있는 도커와 실행에 필요한 시스템 환경을 모아서 관리하는 컨테이너로 구성하였다. 컨테이너는 ZABBIX, Apache server, NGINX server, Mysql server, SNMP server로 각각 이루어져 있다. ZABBIX는 네트워크 하드웨어를 감시하고 추적하여 장애 발생을 신속히 알리기 위해 만들어진 네트워크 관리 시스템이다. 그 외는 망관리 프로토콜(SNMP), 데이터베이스 관리 시스템(MySQL), 웹서버(Apache)와 웹서버 소프트웨어(NGINX) 이다. 이렇게 구성된 트래픽 데이터 수집 서버는 클라우드로 구축되었기 때문에 상황에 맞게 자유롭고 쉽게 확장, 축소가 가능하도록 했다. 또한 가상서버의 장애 시, 추가 백업 서버도 구동이 가능하도록 하였다. 이를 통해 유연성과 탄력성을 갖춘 트래픽 빅데이터 수집 환경을 확보할 수 있게 되었다.

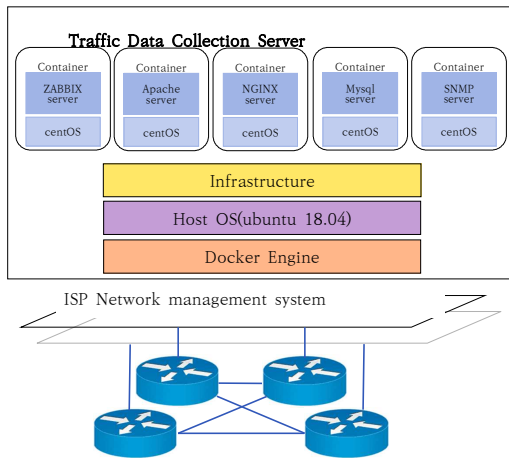


그림 2. 트래픽 수집 가상서버 구조  
Fig. 2. traffic collection virtual server structure

#### 4. 트래픽 모니터링

트래픽 모니터링을 위해서는 모든 라우터로부

터 수집할 대상 데이터를 자동으로 탐색 및 판별한 다음 실시간으로 수집할 수 있도록 했다. 기존 NMS에서는 트래픽 수집 오류와 Data 손실이 수시로 발생하는데 이러한 이슈를 수집 서버를 통해 해결했고 Data 수집대상의 수동 등록 절차도 자동으로 할 수 있도록 했다. 연구 결과의 보다 큰 성과는 불가능했던 네트워크 품질모니터링을 가능하도록 했다는 것이다. 대상 라우터로부터 7,000여 개의 빅데이터를 탐색하여 1,500여 개의 품질모니터링 그래프가 자동으로 생성되도록 하였다. 그래프가 웹브라우저를 통해 실시간으로 생성되면서 인터넷 품질을 직관적으로 관측할 수 있도록 했다. 이러한 그래프의 일부를 그림3에 도시하였다. 적색은 해외사업자 입력 트래픽, 청색은 출력 트래픽이다. 인터넷 트래픽의 관측은 네트워크 트래픽의 특성과 상태를 파악하는 데 매우 중요하다. 인터넷 서비스를 제공하는 ISP가 네트워크 사용 트래픽의 종류와 양을 정량적으로 정확히 파악한다면 효율적인 네트워크 서비스와 과금 정책을 구사할 수가 있다. 즉, 현재 발생하는 트래픽을 최적으로 서비스할 수 있는 라우팅 정책과 스케줄링 기법 그리고



그림 3. 라우터 품질모니터링 그래프  
Fig. 3. router quality monitoring graph

라우터 버퍼관리기법 등을 선택할 수 있게 된다. 또한 네트워크 중단 호스트들 간의 인터넷 데이터 패킷이 어떻게 목적지로 이동하고 있는지를

파악하여 네트워크 상태를 진단할 수 있게 한다. 이와 같은 중요한 역할을 본 연구를 통해 구축된 트래픽 데이터 수집 서버를 통해 가능하도록 하였다.

## 5. OTT 트래픽 예측

### 5.1 머신러닝(AI) 트래픽 예측모델

OTT 미래 트래픽예측은 트래픽 데이터 수집서버로부터 과거 특정날짜의 트래픽 데이터값을 가져온다. 그 데이터값과 예측에 필요한 변수를 반영하여 머신러닝을 하게되고 인공지능 기반의 머신러닝(AI) 트래픽 예측모델이 만들어진다. 이러한 과정을 통해 생성된 머신러닝(AI) 트래픽 예측모델을 통해 OTT 미래트래픽을 예측하였다. 인공지능기반 머신러닝(AI) 트래픽 예측모델은 그림 4와 같다.

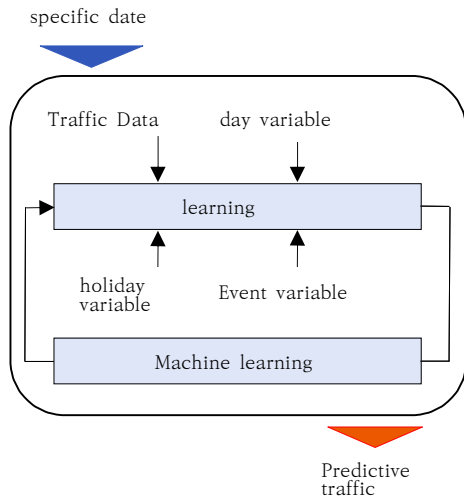


그림 4. AI 트래픽예측 모델  
Fig. 4. AI traffic prediction model

### 5.2 트래픽 예측기능 구현방법

본 연구에서는 Python 언어로 프로그래밍하여 데이터를 작업했으며 Interactive 형식의 라이브 코드 제공과 시각화가 가능한 Jupyter Notebook에서 예측기능을 구현했다. IDE/웹브라우저 안에서 필요한 코드 부분을 실행하고 다양한 그래프를 통해 예측 데이터를 시각화하였다. 예측기능 구현은 크게 두 가지 부분으로 진

행하였다. 인터넷 트래픽 수집 서버로부터 예측을 위해 필요한 Dataframe을 생성하는 부분과 이를 통해 필요한 파라미터 값을 세팅하고 예측 모델을 통해 예측을 하는 부분이다. 구현된 기능과 그 예측절차를 함께 설명하면 다음과 같다. 먼저 Dataframe을 생성하기 위해서는 SQL등 필요한 모듈을 설치한다. 다음은 수집 서버와 연결하고 운용자가 등록한 host로부터 host name과 Key값인 hostid 정보를 가져온다. 이를 실행하면 hosts라는 dictionary 타입의 변수에

```
hosts = host_id 1 : host_name1,
        host_id 2 : host_name2, ..
```

와 같은 구조가 저장이 된다. 다음은 예측할 때, 상한값을 제한하기 위해 값을 가져오면 {ds : cap}과 같은 형식으로 저장이 되고 마지막으로 OTT 트래픽 데이터를 가져온다. 구조형식의 Dataframe이 생성되는 과정을 도식화하면 그림 5와 같으며, 생성된 Dataframe은 표 1과 같다.

| item                                    | action                         |
|---|--------------------------------|
| prophet<br>workalendar<br>pymysql       | installation                   |
| traffic data<br>collection<br>DB Server | connection                     |
| host                                    | resistration                   |
| host name<br>host id                    | query                          |
| upper value                             | restriction(key,value)         |
| OTT traffic data                        | query{ds:cap},<br>{ds:traffic} |

Dataframe Creation

그림 5. 데이터프레임 생성 절차  
Fig. 5. Dataframe creation procedure

표 1. 데이터프레임  
table 1. Dataframe

|     | ds         | cap              | y                |
|-----|------------|------------------|------------------|
| 0   | 2018-XX-01 | C <sub>0</sub>   | Y <sub>0</sub>   |
| 1   | 2018-XX-02 | C <sub>1</sub>   | Y <sub>1</sub>   |
| ... | ...        | ...              | ...              |
| 724 | 2020-XX-24 | C <sub>724</sub> | Y <sub>724</sub> |

두 번째, 예측을 위해 파라미터 값을 세팅하고 예측모델을 통해 예측하는 기능과 절차이다. 시계열 예측은 과거값의 미래 확률 패턴 반영과 오차의 최소화가 관건이다. 그리고 파라미터의 직관성과 튜닝의 용이성 그리고 반복을 통한 정확성 향상이다. 따라서 본 논문에서 이러한 점을 고려하여 시계열 분석/예측 라이브러리 Prophet을 활용하였다. 연구에 적용한 모델은 Trend, Seasonality, Holidays 3개의 주요 요소로 구성되어 있으며 식은 아래와 같다.

$$y(t) = g(t) + s(t) + h(t) + \text{error}$$

위 식에서 Trend를 의미하는 g(t)는 비주기적인 변화를 파악할 수 있다. 그리고 Seasonality를 나타내는 s(t)는 주별, 월별 등 주기적으로 나타나는 패턴 예측을 가능하게 한다. Holidays를 나타내는 h(t)는 주기성을 가지지는 않지만, 공휴일이나 이벤트와 같이 추이에 영향을 미치는 요소이다. error는 정규분포라고 가정할 오차이다. 이러한 직관적인 파라미터를 가지고 예측 모델에 적용하였다.

먼저 Prophet의 Seasonality 요소를 모델에 반영하기 위해 python의 카렌다를 가져와 휴일이나 이벤트 정보를 모델에 반영하여 주,월별 패턴 예측이 가능하도록 하였다. 그리고 시기별로 예측 상한값과 하한값을 추가 적용하고 90일간의 예측 기간을 정하였다. 다음은 생성한 미래 데이터의 Dataframe을 가지고 90일간의 데이터를 예측했다. 그리고 다양한 컴포넌트(trend, weekly, yearly)로 예측 결과를 시각화할 수 있게 하였다. 트래픽 예측절차와 시각화를 도식화하면 그림 6과 같다.

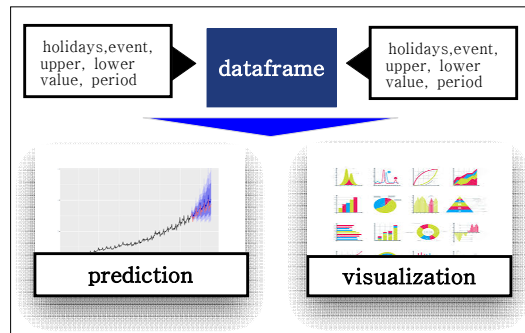


그림 6. 트래픽예측과 시각화  
Fig. 6. prediction & visualization

### 5.3 트래픽 예측 결과 분석

과거 트래픽 데이터를 머신러닝(AI)으로 분석한 트래픽 예측모델과 예측을 위한 Dataframe 그리고 Prophet 라이브러리로 그림 7과 같이 미래 90일의 OTT 트래픽 데이터를 예측하였다. 진한 파란색 선이 모델이 예측한 결과값이고, 연한 파란색은 오차 범위, 점선은 상한값과 하한값을 나타낸다. 검정색 점들이 실제 데이터를 의미하고 있다. 결과 그래프를 보면, 마지막으로 실제 값(검은 점)이 수집된 이후 90일 동안의 예측 그래프가 출력된 것을 볼 수 있다. 2020년 3/4분기 후반부와 4/4분기 전반부의 OTT 트래픽은 전반적으로 증가됨을 예측하고 있다. 90일간의 예측 데이터뿐만 아니라, 컴포넌트별로 시각화하여, trend로 “추세”를 예측하였으며 weekly와 yearly를 통해 “주계절성”과 “연계절성”을 분석하였다. 아래 그림 8을 보면 예측한 OTT 트래픽의 경우, 과거부터 지금까지 트래픽이 증가하는 추세를 보이고 있고, 휴일 등 특정 event가 있는 날에는 트래픽 사용량이 증가하는 모습을 보이고 있다. 그리고 요일별로 보면 금요일 저녁부터 월요일까지 트래픽이 증가추세이다가 화요일부터 트래픽이 감소하여 목요일까지 이어진다. 월별로는 6월과 10월이 가장 트래픽이 낮은 것으로 예측되었는데 날씨로 인한 바깥 활동이 원인인 것으로 추정된다.

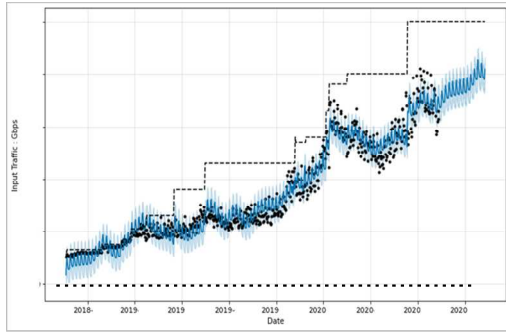
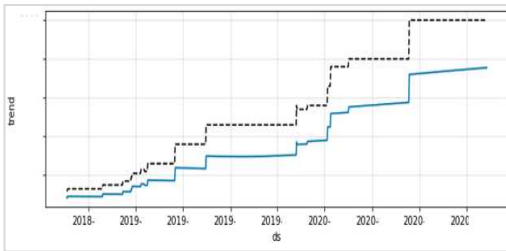


그림 7. 트래픽 예측 결과 (90일간)  
Fig. 7. traffic forecast for 90 days providing network

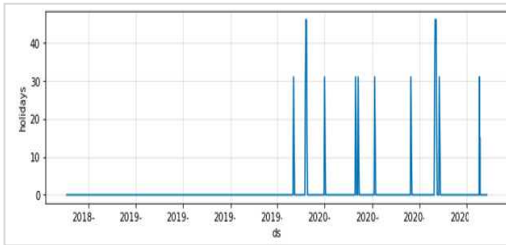
표 2. 트래픽 증가율 (예측기준일로부터 90일간)

table 2. Rate of traffic increase (90 days from the prediction base date)

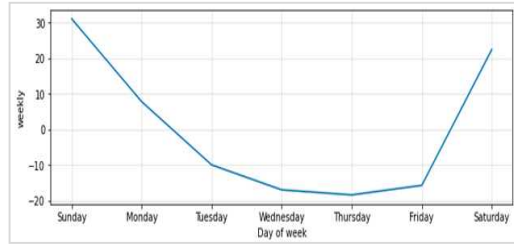
| 일차 | 증가율   | 일차 | 증가율   | 일차 | 증가율   |
|----|-------|----|-------|----|-------|
| 3  | 100.6 | 33 | 106.6 | 63 | 110.2 |
| 6  | 100.7 | 36 | 106.9 | 66 | 111.3 |
| 9  | 101.4 | 39 | 107.3 | 69 | 112.5 |
| 12 | 101.7 | 42 | 107.4 | 72 | 113.2 |
| 15 | 102.0 | 45 | 107.5 | 75 | 113.9 |
| 18 | 102.9 | 48 | 107.5 | 78 | 114.7 |
| 21 | 103.6 | 51 | 107.6 | 81 | 116.2 |
| 24 | 104.4 | 54 | 108.2 | 84 | 117.6 |
| 27 | 105.3 | 57 | 108.5 | 87 | 119.1 |
| 30 | 105.8 | 60 | 108.8 | 90 | 120.5 |



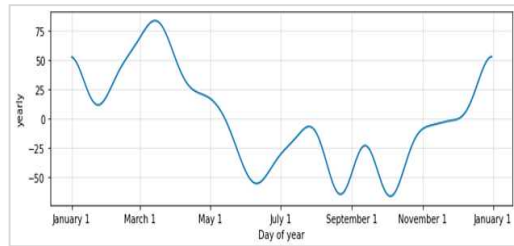
(a) Traffic trend by year



(b) Traffic trend by event day



(c) Traffic trend by day of the week



(d) Monthly traffic trend

그림 8. 컴포넌트별 트래픽 시각화  
Fig. 8. traffic visualization each component

## 6. 결론

본 연구에서는 인터넷 트래픽 예측을 위해 우선 실제 운용되고 있는 네트워크 환경에서 트래픽 데이터 수집했다는 데 의미가 있다. 이를 통해 트래픽 모니터와 트래픽 예측이 함께 가능할 수 있었다. 클라우드 기반의 모니터링 서버구축을 통해 기존 NMS의 트래픽수집오류 및 데이터 손실의 문제점을 개선하였다. 트래픽 예측을 위해서는 방대한 데이터를 머신러닝(AI)으로 분석하여 미래 트래픽 예측의 신뢰도를 한층 높였다. 따라서 과거 방식에서 벗어나 획기적인 모니터와 예측이 가능하도록 하였다. 또한 최근 코로나이슈로 증대된 트래픽과 사전예측 곤란한 이벤트성 트래픽을 실시간적으로 반영하였다. 이러한 현실적 데이터를 반영한 연구 결과는 시뮬레이션과는 또 다른 차원의 연구결과로 볼 수 있다. 본 연구 결과는 OTT 사업자와의 회선연동방식과 ISP의 백본 대역폭 구축 계획 등에 적극 활용이 가능하다. 아울러 예측이 어려운 인터넷 트래픽 폭등에 대비하고 OTT 서비스이용자의 체감 품질을

유지할 수 있게 되었다. 향후 이러한 연구 경험을 토대로 검색어 API를 이용한 독자적인 인터넷 트래픽 예측 모델을 개발할 계획이다.

## REFERENCES

- [1] Tikunov, D. and Nishimura, T. "Traffic Prediction for Mobile Network using Holt-Winter's Exponential Smoothing", Telecommunications and Computer Networks, 15th International Conference, 1-5. 2007
- [2] H. NIE et al. "Hybrid of ARIMA and SVMs for Short-Term Load Forecasting," International Conference on Future Energy, Environment, and Materials, pp.1455-1460, 2012
- [3] Shu, Y., Yu, M., Yang, O., Liu, J. and Feng, H. "Wireless traffic modeling and prediction using seasonal ARIMA models", IEICE-transactions on Communications, 10, 3992-3999. 2005
- [4] L.R. Medsker, L.C. Jain, "Recurrent Neural Network: Design and Applications", p.12-14, 2001
- [5] Luis G. B. R., Manuel P. C., Miguel, D. C. and Maria D. C. P. J., "An Application of Non-Linear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings," energies, 9(9), p.684, Aug. 2016.
- [6] L.R. Medsker, L.C. Jain, "Recurrent Neural Network: Design and Applications", p.12-14, 2001
- [7] DH Kim, MW Kim, BJ Lee, KT Kim, HY Yoon, "Data Flow Prediction Scheme using ARIMA Model", Proceedings of the Korean Society of Computer Information Conference 26(2), 2018.7, 141-142
- [8] JKLee, IP Cho, SY Lee "A study on data collection environment and analysis using virtual server hosting of Azure cloud platform, Proceedings of the Korean Society of Computer Information Conference, 2020.7, 329-330
- [9] MH Ha, HG Sona, S. Kim, "A Study on Performance Analysis of ShortTerm Internet Traffic Forecasting Models", Proceedings of KICS. 2012, 19(3B), 415-422
- [10] SJ. Jung, DJ Kim, YH Kwon "A Fitness Verification of time Series Models for Network Traffic Predictions", Proceedings of KICS. 2004 29(2B), 217-227
- [11] SH Ji, Huru Hasanova, KS Shim, MS Kim. "Prediction of Traffic Usage Using Machine Learning Algorithm For Efficient Network Management", Proceedings of Symposium of KICS, 2018.1, 824-825
- [12] Jin Sheng, MS Seok, GY Kim, "Internet Traffic Prediction By Neural Networks", Proceedings of Symposium of KICS, 019.6, 211-213
- [13] JS Won, SW Kim, "VNF traffic Prediction Using Recurrent Neural Network", Proceedings of Symposium of The Institute of Electronics and Information Engineers 2018.6, 351-353

---

## 저자약력

---

### 남 창 섭(Chang-Sup Nam)

[정회원]



- 1987년 2월: 숭실대학교 전자공학(공학사)
- 1989년 2월: 숭실대학교 전자공학(공학 석사)
- 2007년 2월: 충북대학교 정보통신공학과(공학 박사)
- 1989년 8월~2015년12월: KT 구매전략실 부장
- 2016년 2월~2019년 12월: KT commerce 본부장
- 2020년 3월~현재: 호서대학교 기계(CT공학과) 교수

〈관심분야〉 통신네트워크, 지능망, IT서비스, AI