KIPS Trans. Comp. and Comm. Sys.
Vol.12, No.2 pp.53~60
ISSN: 2287-5891 (Print), ISSN 2734-049X (Online)

https://doi.org/10.3745/KTCCS.2023.12.2.53

Real-Time Streaming Traffic Prediction Using Deep Learning Models Based on Recurrent Neural Network

Jinho Kim† · Donghyeok An††

ABSTRACT

Recently, the demand and traffic volume for various multimedia contents are rapidly increasing through real-time streaming platforms. In this paper, we predict real-time streaming traffic to improve the quality of service (QoS). Statistical models have been used to predict network traffic. However, since real-time streaming traffic changes dynamically, we used recurrent neural network-based deep learning models rather than a statistical model. Therefore, after the collection and preprocessing for real-time streaming data, we exploit vanilla RNN, LSTM, GRU, Bi-LSTM, and Bi-GRU models to predict real-time streaming traffic. In evaluation, the training time and accuracy of each model are measured and compared.

Keywords : Real-Time Streaming Service, Traffic Prediction, Recurrent Neural Network, Deep-Learning

순환 신경망 기반 딥러닝 모델들을 활용한 실시간 스트리밍 트래픽 예측

김 진 호† · 안 동 혁††

요 약

최근 실시간 스트리밍 플랫폼을 기반으로 한 다양한 멀티미디어 컨텐츠의 수요량과 트래픽 양이 급격히 증가하고 있는 추세이다. 본 논문에서는 실시간 스트리밍 서비스의 품질을 향상시키기 위해서 실시간 스트리밍 트래픽을 예측한다. 네트워크 트래픽을 예측하기 위해 통계적 모형을 활용하였으나, 실시간 스트리밍 트래픽은 매우 동적으로 변화함에 따라 통계적 모형보다는 순환 신경망 기반 딥러닝 모델이 적합하다. 따라서, 실시간 스트리밍 트래픽을 수집, 정제 후 Vanilla RNN, LSTM, GRU, Bi-LSTM, Bi-GRU 모델을 활용하여 예측하며, 각 모델의 학습 시간, 정확도를 측정하여 비교한다.

키워드 : 실시간 스트리밍 서비스, 트래픽 예측, 순환 신경망, 딥러닝

1. 서 론

현대 사회에서 무선 통신 시장이 급속도로 성장함에 따라, 사용자들은 LTE, 5G와 같은 고품질 네트워크를 통해 향상된 품질의 컨텐츠를 자유롭게 이용할 수 있게 되었다. 이와 동시에 코로나-19 팬데믹으로 인하여 1인 미디어, 실시간 스트리밍과 같이 사용자들이 이용하는 멀티미디어 컨텐츠의 수요가 변화하였다. 이러한 변화에 따라, 최근 실시간 스트리밍 플랫폼을 기반으로 한 서비스 트래픽이 폭발적으로 증가하고 있는 추세이다. 실시간 스트리밍 서비스 트래픽은 서비스를 제공받는 호스트의 환경, 서비스를 제공하는 서버의 과부하, 네트워크 상의 혼잡과 같은 다양한 네트워크 환경 요인에 영향을 받으며, 이와 같은 이유로 인하여 트래픽 패턴이 동적으로 변화한다. 변화하는 트래픽을 정확하게 예측할 수 있다면 네트워크 관리 측면에서 보다 효율적이고 사전 예방적인 대응이 가능하며, 플랫폼에서 제공하는 서비스 품질(QoS, Quality of Service)과 사용자가 서비스를 경험하는 품질(QoE, Quality of Experience)을 향상시킬 수 있다.

시간과 환경에 따라 변하는 네트워크 트래픽과 같은 시퀀스 데이터는 ARIMA(Auto Regressive Integrated Moving Average), SARIMA(Seasonal ARIMA)와 같은 통계적 모형 또는 ANN(Artificial Neural Network)과 같은 신경망 모델을 기반으로 예측이 가능하다[1, 2]. 하지만 통계적 모형은

※ 이 논문은 2021~2022년도 창원대학교 자율연구과제 연구비 지원으로 수행된 연구결과임.
※ 본 논문은 한국컴퓨터종합학술대회에서 발표된 학부생 논문을 기반으로 확장 및 추가 연구됨.
† 비 회 원 : 창원대학교 컴퓨터공학과 학사과정
†† 종신회원 : 창원대학교 컴퓨터공학과 부교수
Manuscript Received : October 4, 2022
First Revision : November 2, 2022
Accepted : November 14, 2022
* Corresponding Author : Donghyeok An(donghyeokan@changwon.ac.kr)

데이터의 크기와 패턴에 따라 예측 결과가 불안정한 단점이 존재한다. 실시간 스트리밍 서비스 트래픽 변화량은 매우 동적이기 때문에 통계적 모형 기반 예측은 적합하지 않다. 본 논문에서는 실시간 스트리밍 서비스 트래픽을 예측하여 서비스를 제공하는 공급자와 수요자에게 더 좋은 품질의 스트리밍 서비스를 이용할 수 있도록 순환 신경망을 기반으로 한 Vanilla RNN, LSTM (Long Short-Term Memory)[3], GRU (Gated Recurrent Unit)[4], Bi-LTSM (bidirectional LSTM)[14], Bi-GRU (bidirectional GRU)[14] 모델을 활용한다. 스트리밍 서비스 종류로는 수요가 많은 드라마, 뉴스, 음악을 대상으로 진행한다. 성능 평가 결과에서 Vanilla RNN 모델보다 LSTM 모델과 GRU 모델의 트래픽 예측 정확도 비교적으로 높았으며 학습 시간 측면에서는 GRU 모델이 우수한 성능을 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구와 관련된 연구들과 배경 지식을 기술한다. 3장에서는 실시간 스트리밍 서비스 트래픽 데이터를 수집하기 위한 패킷 필터링 과정과 모델 학습을 위한 전처리(Pre-Processing) 과정에 대해 기술하며, 4장에서 순환 신경망을 기반으로 한 Vanilla RNN, LSTM, GRU 각각의 모델 아키텍처와 학습에 사용된 하이퍼 파라미터를 기술한다. 5장에서는 모델 학습 시간과 회귀 모델에서 사용되는 평가 지표를 통해 실험 결과를 기술한다. 마지막으로 6장에서 결론과 향후 연구에 대해 기술하며 마무리한다.

2. 배경 지식 및 관련 연구

해당 장에서는 기본적인 순환 신경망의 원리와 이를 기반으로 한 Vanilla RNN, LSTM, GRU 모델의 내부 구조와 학습 과정을 나타낸다. 이후 관련 연구들과 본 연구와의 차별성을 서술한다.

2.1 Vanilla RNN

순환 신경망은 Fig. 1과 같이 은닉층(Hidden Layer)이 순환하는 구조를 가지고 있으며 각 시점 t에서의 은닉층들은 모두 같다. 이러한 구조를 기반으로 순환 신경망은 시간에 따라 변화하는 시계열 데이터의 학습에 특화되어 있는 특징이 있다. 순환 신경망을 기반으로 동작하는 Vanilla RNN, LSTM, GRU 모델은 은닉 층 종류에 따라 구분되며 Vanilla RNN의 구조는 Fig. 2와 같다.

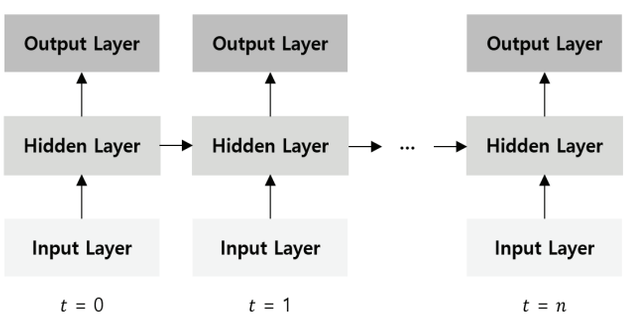

Fig. 1. Recurrent Neural Network

Vanilla RNN은 현재 시점 t에서의 입력 값 행렬 X_t와 이전 시점 t-1에서의 출력 값 행렬 Y_{t-1}을 입력 받는다. X_t와 Y_{t-1}은 각각 가중치 행렬 W_1, W_2와 원소 간의 곱 연산(MatMul)을 진행하며 편향 행렬 B와 원소 간의 합 연산(Add)을 진행하여 행렬 U_t를 도출한다. U_t는 활성화 함수 tanh를 거친 뒤 현재 시점 t에서의 출력 값 행렬 Y_t를 도출한다. 마지막으로, Y_t는 다음 시점 t+1과 다음 계층으로 각각 전파된다.

$$U_t = X_t^* W_1 + Y_{t-1}^* W_2 + B \tag{1}$$

$$Y_t = \tanh(U_t) \tag{2}$$

하지만 Vanilla RNN은 학습을 진행할 때 항상 같은 가중치 행렬을 사용하기 때문에 장기 의존성 문제[5]가 발생한다는 단점이 존재한다.

2.2 LSTM

LSTM은 Fig. 3과 같이 망각 게이트(Forget gate), 입력 게이트(Input gate), 출력 게이트(Output gate)와 1개의 기억 셀(Memory cell)로 이루어져 있으며, 해당 구조를 통해 과거 정보를 반영할 비율을 조정하여 장기 기억을 학습할 수 있는 특징이 존재한다.

망각 게이트는 현재 시점의 입력 값 행렬 Xt와 이전 시점의 출력 값 행렬 Y_{t-1}을 입력 받아 각각 가중치 행렬 W_{f1}, W_{f2}와 원소 간의 곱 연산을 진행한 후 편향 행렬 B_f와 원소 간의 합 연산을 진행한다. 해당 값은 활성화 함수 sigmoid를 거친 뒤 출력 값 행렬 A_{forget}을 도출하는 과정을 통해 과거에 학습했던 기억을 남길 비율을 조정하는 역할을 한다. 또한, 입력 게이트는 이전 시점의 출력을 기억 셀에 반영할 비율을 조정하는 역할을 하며, 출력 게이트는 기억 셀이 출

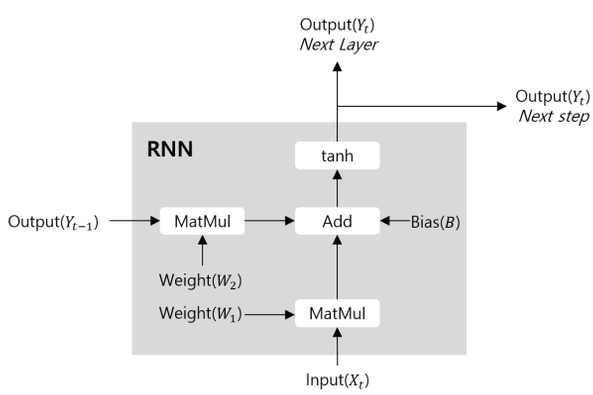

Fig. 2. Vanilla RNN

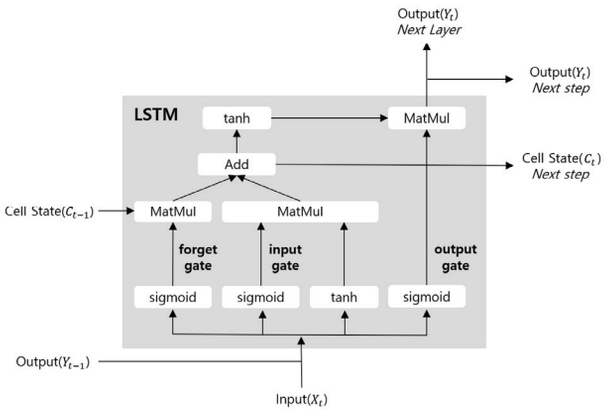

Fig. 3. LSTM

력에 반영되는 비율을 조정하는 역할을 한다. 입력 게이트와 출력 게이트는 각각 가중치 행렬 W_{i1}, W_{i2}와 W_{o1}, W_{o2}를 사용하여 해당 역할을 수행한다. A_{forget}, A_{input}, A_{output} 계산식은 아래와 같다.

$$A_{forget} = sigmoid(X_t * W_{f1} + Y_{t-1} * W_{f2} + B_f) \quad (3)$$

$$A_{input} = sigmoid(X_t * W_{i1} + Y_{t-1} * W_{i2} + B_i) \quad (4)$$

$$A_{output} = sigmoid(X_t * W_{o1} + Y_{t-1} * W_{o2} + B_o) \quad (5)$$

기억 셀은 셀의 상태(Cell state)를 저장하고 전파하는 역할을 하며, 이전 시점의 셀의 상태 정보 C_{t-1}과 여러 게이트들의 출력 값 행렬들을 반영하여 현재 시점의 셀의 상태 정보 C_t를 다음 시점으로 전파하며, 활성화 함수 tanh를 거친 뒤 A_{output}과 원소 간 곱 연산을 진행한 후 다음 시점 t+1과 계층으로 각각 전파된다.

$$C_t = A_{forget} * C_{t-1} + (A_{input} * \tanh(X_t * Y_{t-1})) \quad (6)$$

$$Y_t = A_{output} * \tanh(C_t) \quad (7)$$

2.3 GRU

GRU는 LSTM보다 구조적으로 단순하며, 작은 연산 횟수를 통해 빠른 학습 시간을 제공하는 모델이다. 업데이트 게이트(Update gate)와 리셋 게이트(Reset gate)로 이루어져 있다. 또한, 기억 셀이 별도로 존재하지 않으며, 해당 기능은 이전 시점의 출력 Y_{t-1}을 통해 반영하게 되는 특징이 있다.

업데이트 게이트는 LSTM의 입력 게이트와 망각 게이트의 기능을 합한 역할을 하며, 현재 시점의 입력 값 행렬 X_t와 이전 시점의 출력 값 행렬 Y_{t-1}을 입력 받아 각각 가중치 행렬 W_{u1}, W_{u2}와 원소 간의 곱 연산을 진행한 후 편향 행렬 B_u와 원소 간의 합 연산을 진행한다. 해당 값은 활성화 함수 sigmoid를 거친 뒤 현재 시점의 새로운 기억과 이전 시점의 과거 기억의 비율을 조정하는 역할을 한다. 리셋 게이트는 LSTM의 출력 게이트를 대신하여, 각각 가중치 행렬 W_{r1}, W_{r2}를 사용하여 과거에서 이어받은 기억을 선별하는 역할을 한다. 업데이트 게이트의 출력 값 행렬 A_{update}와 리셋 게이트의 출력 값 행렬 A_{reset}을 통해 현재 시점 t 에서의 출력 값 행렬 Y_t를 도출하며, Y_t는 다음 시점 t+1과 다음 계층으로 각각 전파된다.

$$A_{update} = sigmoid(X_t * W_{u1} + Y_{t-1} * W_{u2} + B_u) \quad (8)$$

$$A_{reset} = sigmoid(X_t * W_{r1} + Y_{t-1} * W_{r2} + B_r) \quad (9)$$

$$U_t = \tanh(X_t * W_1 + (A_{reset} * Y_{t-1}) * W_2 + B) \quad (10)$$

$$Y_t = (1 - A_{update}) * Y_{t-1} + A_{update} * U_t \quad (11)$$

2.4 관련 연구

네트워크 트래픽을 예측하기 위한 다양한 연구들이 진행되었다. [1, 2]의 연구는 네트워크 트래픽 예측을 위해 과거 정보와 추세를 반영해 특정 시점의 미래 정보를 예측하는 통계적 모형과 일반적인 인공 신경망 모델인 ANN과 같은 접근 방법을 제시한다. 하지만 실시간 스트리밍 서비스 트래픽은 컨텐츠 종류에 따른 트래픽 패턴이 다양하므로 정확한 통계적 모델 생성이 어려우며, 단순 인공 신경망이 아닌 순환 신경망을 기반으로 예측을 수행한 점에서 본 연구와 차별화된다. [6]의 연구는 MLP(Multi-Layer Perceptron), RNN와 같은 전통적인 신경망 모델과 SAE(Stacked Auto Encoder)와 같은 데이터를 효율적으로 표현할 수 있는 신경망 모델을 사용하였다. [7]에서는 5G 네트워크 내 가상화 기반 네트워크 슬라이싱을 위해 RNN을 사용해 트래픽을 예측하였다. 하지만, 본 연구는 RNN 외에도 LSTM, GRU와 같은 순환 신경망 계열의 모델을 기반으로 트래픽 예측을 진행하였다. [8]의 연구는 LSTM 모델을 사용하여 무선 네트워크 상에서의 정확한 트래픽 예측과 빠른 수렴 속도에 대한 연구를 진행하였지만, 실시간 스트리밍 트래픽은 전체 경로 상태에 영향을 받기 때문에 무선 네트워크 내 트래픽 예측과는 다르다. [9]에서는 전용회선 구축을 위한 트래픽 예측을 위해 RNN과 LSTM을 사용했다는 점에서 본 연구와 유사하다. 하지만 전용회선 내 트래픽은 네트워크 내 백그라운드 트래픽이 없기 때문에 트래픽 변화량이 크지 않다. 통신 경로 상 주변 트래픽에 영향을 받는 실시간 스트리밍 트래픽 예측과는 차별화된다. [10]에서는 HTTP, DNS 등 웹 트래픽을 예측하기 위해서 LSTM과 CNN-LSTM을 활용하였으나 웹과 DNS 트래픽들은 필요에 의해 순간적으로 발생한다는 점에서 스트리밍 트래픽 예측과는 다르다. [11]에서는 모바일 앱 트래픽 예측을 위해 마르코프(Markov) 모델링을 활용하였다.

3. 데이터 셋

실시간 스트리밍 서비스 트래픽 예측 모델을 위한 학습 및 테스트에 활용하기 위해 데이터를 수집하였다. 실시간 스트리밍 서비스 플랫폼으로 Youtube를 사용하였으며 서비스 종류는 국내외에서 사용자 수요가 가장 많은 뉴스, 드라마, 음악을 선정하였다. 각 컨텐츠 별로 4시간(14400초) 분량의 패킷 데이터를 Wireshark[12]를 활용해 수집하였다. Wireshark는 지정한 NIC가 연결된 네트워크 내 모든 패킷들을 수집하기 때문에 수집된 패킷들 중 실시간 스트리밍 서비스와 상관없는 패킷들이 존재한다. 불필요한 패킷들을 제거하기 위해 아래 알고리즘과 같은 패킷 필터링 과정을 수행하였다.

packet filtering

```
for packet do
  if packet_dst_ip != host_ip or packet_ip not in
     platform_ip_list
     then invalid packet's ip address, drop packet
  else
    if packet_protocol != udp or packet_protocol
     != quic
       then invalid packet's protocol, drop packet
    else
       then append packet to the packet list
    end if
  end if
end for
```

먼저 목적지 IP가 호스트 IP와의 동일 여부를 확인한다. 만약 목적지 IP가 호스트 IP와 상이할 경우 네트워크에 연결된 다른 호스트를 위한 패킷이므로 패킷을 삭제한다. 목적지 IP가 호스트 IP와 동일하지만 송신자 IP가 스트리밍 플랫폼 IP가 아니면 패킷을 삭제한다. 스트리밍 플랫폼은 원활한 스트리밍 서비스를 제공하기 위해서 CDN(Contents Delivery Network)을 활용하고 있다. 이로 인해 시간 및 네트워크 환경에 따라 빈번하게 서버의 IP 주소가 변경된다. 해당 플랫폼 서버의 IP 주소에 서브넷 마스크(Subnet Mask)를 적용해 플랫폼 IP 리스트를 작성하였다. 플랫폼 IP 리스트를 기반으로 플랫폼 IP 여부를 결정한다. 목적지 IP가 호스트 IP와 동일하고 송신자 IP가 스트리밍 플랫폼 IP라면 UDP 또는 QUIC 패킷 여부를 판단한다. Youtube는 QUIC 프로토콜을 활용하고 QUIC 프로토콜은 UDP 기반으로 동작하기 때문에 UDP 패킷 여부를 먼저 판별한다. UDP 패킷이 아닌 경우 패킷을 삭제한다. 또한 UDP 패킷이지만 QUIC 패킷이 아니라면 Youtube 트래픽이 아니라고 판단해 패킷을 삭제한다. UDP 패킷이고 QUIC 패킷인 경우 Youtube 스트리밍 패킷으로 판단해 수신 패킷을 저장한다.

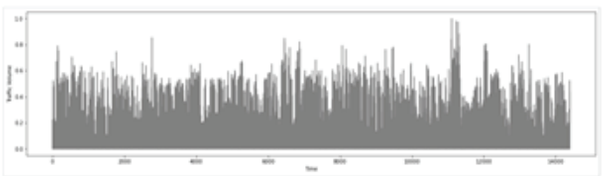

Fig. 4. Real-Time Drama Traffic

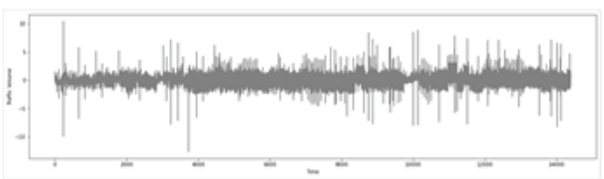

Fig. 5. Real-Time Music Traffic

패킷 필터링 후 실시간 스트리밍 트래픽에서 획득할 수 있는 여러 정보 중 시간과 패킷 길이를 활용해 특징 추출(Feature selection)을 진행하였다. 수집한 트래픽들의 수신 시간을 기반으로 1초, 5초, 10초 간격으로 트래픽을 그룹핑 하였으며 동일 그룹에 속한 패킷을 대상으로 패킷 길이의 합을 계산하였다.

안정적인 딥러닝 학습을 위해 데이터 스케일링을 진행하였다. 스케일링을 위해 Min-Max scaling과 Standard scaling을 사용하였다. Min-Max scaling는 데이터의 분포를 유지하면서 값의 범위를 0 ~ 1 사이로 조정하는 기법이며 시간 별 트래픽 변화가 큰 실시간 드라마 또는 실시간 뉴스 데이터에 적용하였다. Standard scaling은 데이터 평균값이 0, 분산이 1인 표준 정규 분포를 따르도록 조정하는 기법으로 시간 별 트래픽 변화가 적은 실시간 음악 데이터에 적용하였다. Fig. 4와 Fig. 5는 수집한 실시간 트래픽에 Min-Max scaling과 Standard scaling을 적용한 결과이다.

마지막으로, 학습과 테스트를 위해 데이터 분할을 진행하였다. 학습 데이터와 테스트 데이터의 비율을 7:3으로 분할하여 학습 데이터는 약 3시간(10080초), 테스트 데이터는 약 1시간(4320초)의 플레이 시간을 가진다.

4. 모델 아키텍처

해당 장에서는 실시간 스트리밍 트래픽 예측을 빠르게 수행하기 위한 Vanilla RNN, LSTM과 GRU 기반 경량화 모델 아키텍처는 Fig. 6과 같다. 시간 간격에 따른 그룹핑 과정과 전처리 과정을 진행한 데이터는 시점 t와 입력 데이터의 길이 시퀀스 길이 (sequence length)에 따라 순차적으로 모델에 입력된다. 또한, 시점 t에서의 입력 값 행렬 X_t와 이전 시점에서의 출력 값 행렬 Y_{t-1}은 모델의 은닉층으로 입력된다. 은닉층의 종류로는 Vanilla RNN, LSTM, GRU의 셀들이 해당된다. 각 은닉층을 통해 학습된 데이터들은 현재 시점 t의 출력 값 행렬 Y_t로 출력되며, 해당 데이터는 다음 시점의 입력 데

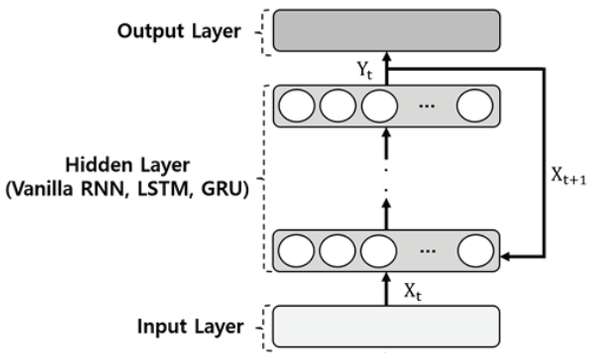

Fig. 6. Model Architecture

이터로 사용되는 순환하는 과정을 통해 학습이 진행된다. 모델에 따라 사용된 파라미터는 학습 횟수 (epoch), 학습 률 (learning rate), 시퀀스 길이, 오차 함수 (criterion), 최적화 함수 (optimizer), 은닉층 개수 (number of hidden layer), 은닉층 뉴런 수(hidden size)가 존재하며, 5.2 최적 파라미터 학습에서 파라미터 설정 기준을 자세히 서술한다.

5. 성능 평가

5.1 실험 환경 및 평가 지표

제안한 LSTM 및 GRU 모델을 구현하기 위해 PyTorch와 CUDA를 사용하였다. 실험은 3.8GHz CPU, 32GB RAM, RTX 3060 12GB GPU가 장착된 컴퓨터에서 시간 간격 별로 LSTM 및 GRU의 예측 성능과 학습 시간을 측정하였다.

실험은 실시간 드라마, 뉴스, 음악 서비스를 위한 트래픽 별 1초, 5초, 10초 동안 측정한 트래픽 양을 계산한다. 시간 별 트래픽 양을 기반으로 Vanilla RNN, LSTM, GRU, Bi-LSTM, Bi-GRU 기반 모델의 예측 정확도와 학습 시간을 비교한다. 실시간 트래픽에 대한 예측을 목표로 하기 때문에 학습 시간도 중요한 요소이기 때문이다. 예측 정확도는 NRMSE 평가 지표를 활용한다[13]. NRMSE는 회귀 모델에서 대표적으로 사용되는 평가 지표이며, 값의 범위가 다른 데이터에 대해 일관된 평가 지표를 보여줄 수 있으며, 0에 가까울수록 좋은 성능을 의미한다.

NRMSE를 구하기 위해서는 Equation (12)와 같이 RMSE 값에 데이터의 최대값 y_{max}와 최소값 y_{min}의 차이를 나누어 계산한다. RMSE는 Equation (13)와 같이 데이터의 실제 값 y_n과 해당 값에 대한 예측 값 $predict_n$의 차이를 통하여 오차를 구하며, 음수 값을 처리하기 위해 제곱 연산을 진행한다. 해당 과정은 전체 데이터의 길이 L에 대해 수행하며, 수행된 전체 오차의 합을 L로 나누어 평균 오차 값을 구한다. 원래의 데이터 범위에 맞추기 위해 제곱근 연산을 수행하여 결과를 도출한다.

Table 1. Experiment Environment

CPU	R7 5800X ~ 3.8GHz
RAM	32GB
GPU	RTX 3060 12GB
OS	Windows 10
IDE / Language	PyCharm / Python 3.7

$$NRMSE = \frac{RMSE}{y_{\max} - y_{\min}} \tag{12}$$

$$RMSE = \sqrt{\frac{1}{L}\sum_{n=1}^{L}(y_n - predict_n)^2} \tag{13}$$

5.2 파라미터 설정

모델에 따른 성능을 극대화하기 위하여 파라미터 튜닝 과정을 진행하였다. 최적의 파라미터를 선택하기 위해 10초의 시간 간격으로 그룹핑한 트래픽 데이터셋을 기준으로 다양한 파라미터 값을 입력으로 넣는 실험 과정을 수행하였다. 실험에 대상이 되는 파라미터의 종류로는 순환 신경망 기반 모델에서 가장 중요한 파라미터 중 하나인 입력 데이터의 길이인 시퀀스 길이(sequence length)와 은닉층 개수(number of hidden layer), 은닉층 뉴런 수(hidden size)가 존재한다. 파라미터 별 성능을 평가하기 위해 NRMSE 값을 사용했으며 실제 모델에서 사용한 파라미터 값은 NRMSE 값이 가장 낮은 파라미터를 선택하였다. Fig. 7, Fig. 8, Fig. 9는 각각 시퀀스 길이, 은닉층 개수, 은닉층 뉴런 수에 대한 결과이다.

Vanilla RNN, LSTM 모델은 시퀀스 길이가 40인 경우 가장 성능이 좋았고, GRU 모델은 30일 때 예측 정확도 성능이 가장 좋았다. 은닉층의 개수는 Vanilla RNN 모델은 4개, LSTM 모델과 GRU 모델에서는 3개를 사용하였을 때 가장 성능이 좋은 것을 확인할 수 있다. 마지막으로 은닉층 당 뉴런의 개수 파라미터는 Vanilla RNN 모델은 32개, LSTM, GRU 모델은 24개를 사용하였을 때 높은 성능을 보여주었다. 모델들이 공통적으로 사용한 파라미터 설정은 다음과 같다. 학습 횟수는 600, 학습률은 0.001로 설정하였다. 최적화 함수 및 오차 함수로 Adam과 MSELoss를 사용하였다.

5.3 실시간 스트리밍 트래픽

실시간 스트리밍 트래픽 예측 성능을 평가하기 위해서 예측 정확도와 학습 시간을 측정하였다. 실시간 스트리밍으로 드라마, 뉴스, 음악 서비스를 평가하였다. Fig. 10은 실시간 드라마 트래픽 예측 정확도를 나타낸다. 측정 시간이 짧을수록 트래픽 변화량이 크지 않기 때문에 예측 정확도가 높게 나타나며 측정 시간이 늘어날수록 트래픽 변화량이 크기 때문에 예측 정확도가 떨어진다. 성능 평가 결과에서 알 수 있듯이 측정 시간 별로 대부분의 모델들이 유사한 정확도를 나타

낸다. Fig. 11은 모델 별 학습 시간을 나타내고 있다. 측정 시간이 클수록 학습에 사용되는 데이터가 감소하기 때문에 학습시간이 감소한다. 모델들 중 GRU 모델이 모든 측정 시간에서 학습 시간이 가장 작게 측정된다.

Fig. 12와 Fig. 13은 실시간 뉴스 트래픽 예측 정확도와 학습 시간을 측정한 결과이다. 뉴스는 드라마에 비해 영상 변화뿐만 아니라 트래픽 변화량도 크지 않다. 이로 인해 모든 모델에서 비슷한 예측 정확도를 확인할 수 있다. Bi-LSTM 모델의 예측 정확도가 타 모델 대비 약간 좋은 것으로 나타나지만 학습시간이 가장 길게 측정되었다. 실시간 뉴스 트래픽 예측을 위한 모델 학습은 GRU 모델이 가장 빠르게 진행된다.

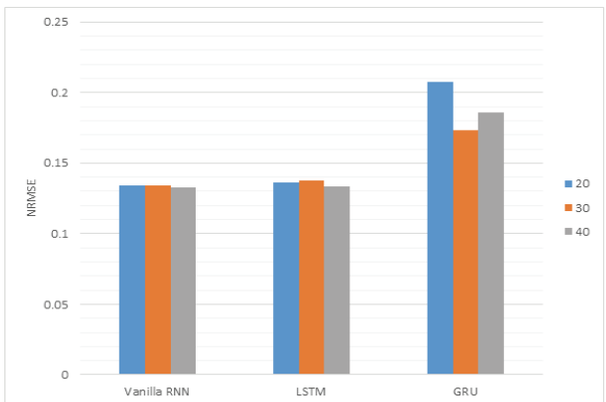

Fig. 7. Sequence Length Comparison

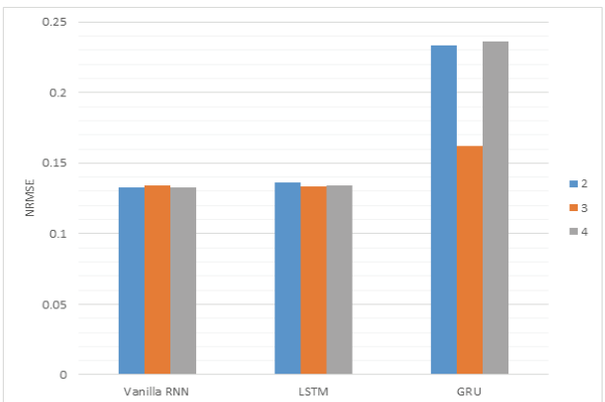

Fig. 8. Comparison of Number of Hidden Layer

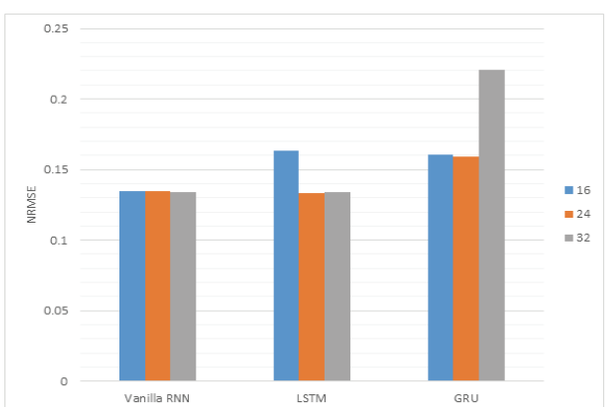

Fig. 9. Comparison of Hidden Size

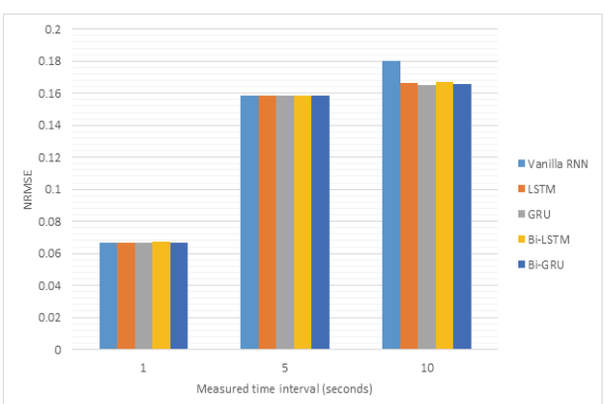

Fig. 10. Prediction Accuracy of Real-Time Drama Traffic

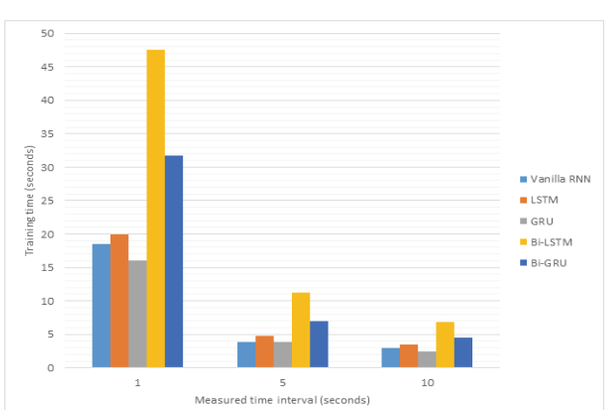

Fig. 11. Training Time of Real-Time Drama Traffic Prediction Model

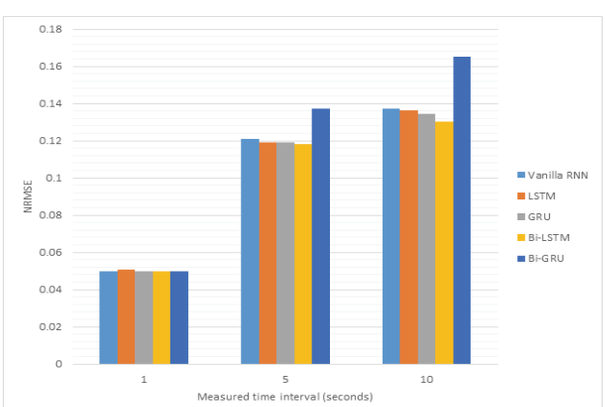

Fig. 12. Prediction Accuracy Real-Time News Traffic

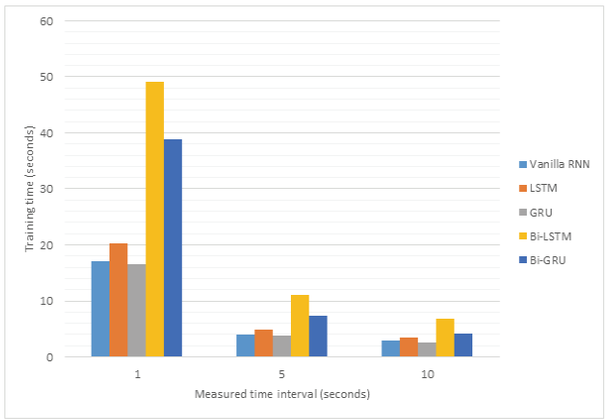

Fig. 13. Training Time of Real-Time News Traffic Prediction Model

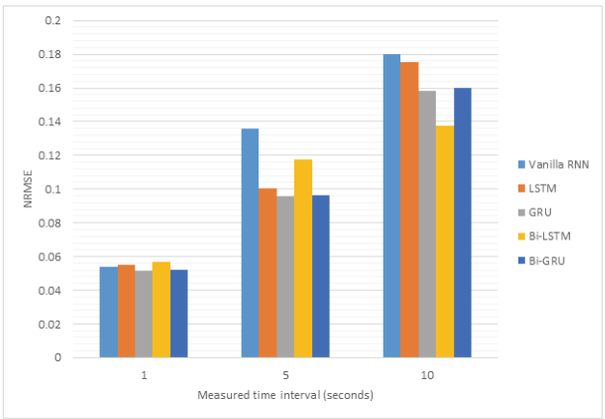

Fig. 14. Prediction Accuracy of Real-Time Music Traffic

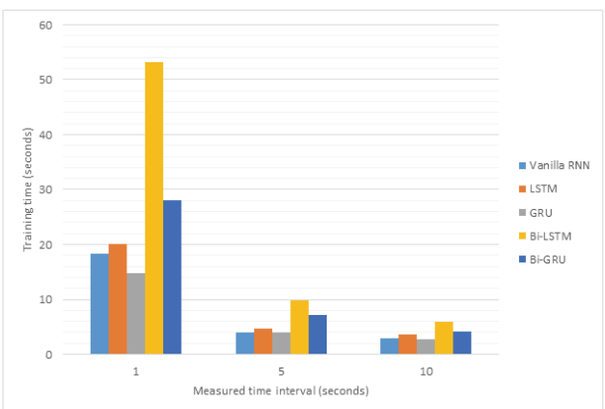

Fig. 15. Training Time of Real-Time Music Traffic Prediction Model

실시간 음악 스트리밍 트래픽 성능 평가 결과는 Fig. 14과 Fig. 15에서 나타낸다. 측정 시간이 5초 일 때는 GRU와 Bi-GRU가 타 모델 대비 정확도가 조금 높으며, 측정 시간이 10초일 때는 Bi-LSTM 정확도가 우수하다. 모델 학습 시간은 GRU가 타 모델 대비 우수한 성능을 보인다. 실시간 스트리밍 종류와 측정 시간에 따라 가장 우수한 예측 정확도를 보이는 모델들이 상이하지만 대부분 모델들의 성능이 비슷하다. 하지만 학습 시간 측면에서도 GRU 모델이 타 모델 대비 우수한 성능을 보인다.

6. 결론 및 향후 연구

본 논문에서는 실시간 스트리밍 서비스 트래픽을 예측하기 위해 순환 신경망 기반 Vanilla RNN, LSTM, GRU의 모델을 활용하여 실시간 드라마, 뉴스, 음악 스트리밍 서비스 트래픽을 예측하였다. LSTM 모델과 GRU 모델의 예측 정확도가 RNN 대비 비교적 높았다. 두 모델 중 GRU 모델은 LSTM 모델과 비교해 경량화된 모델 구조를 통해 보다 빠른 학습 속도를 보여주었다. 향후 연구로서 AR/VR, Holographic과 같은 차세대 네트워크 장비를 통해 전송되는 데이터 트래픽들의 예측 가능성에 대한 연구를 진행할 예정이다.

References

[1] H. Feng and Y. Shu, "Study on network traffic prediction techniques," *Proceedings 2005 International Conference on Wireless Communications, Networking and Mobile Computing*, pp.1041-1044, 2005.

[2] S. J. Jung, Y. K. Chung, and C. G. Kim, "Network routing by traffic prediction on time series models," *Journal of KISS: Information Networking*, Vol.32, No.4, pp.433-442, 2005.

[3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, Vol.9, No.8, pp.1735-1780, 1997.

[4] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," CoRR, abs/1412.3555, 2014.

[5] A. M. Schaefer, S. Udluft, and H. G, Zimmermann, "Learning long-term dependencies with recurrent neural networks," *Neurocomputing*, Vol.71, No.13-15, pp.2481-2488, 2008.

[6] T. P. Oliveria, J. S. Barbar, and A. S. Soares, "Computer network traffic prediction: A comparison between traditional and deep learning neural networks," *International Journal of Big Data Intelligence*, Vol.3, No.1, pp.28-37, 2016.

[7] P. W. Lee, S. Y. Park, and Y. T. Shin, "Machine learning-based network slicing resource reservation scheme in 5G network," *Proceedings of the Korea Information Processing Society Conference*, Vol.27, No.1, pp.56-59, 2020.

[8] S. Jaffry and S. F. Hasan, "Cellular traffic prediction using recurrent neural networks," *2020 IEEE 5th International Symposium on Telecommunication Technologies (ISTT)*, pp.94-98, 2020.
[9] I. G. Lee and M. H. Song, "Leased line traffic prediction using a recurrent deep neural network model," *KIPS Transactions on Software and Data Engineering*, Vol.10, No.10, pp.391-398, 2021.
[10] Y. J. Jang, "Network prediction of traffic generation amount using time series prediction model," Master degree at Hanyang University, 2022.
[11] G. Aceto, G. Bovenzi, D. Ciuonzo, A. Montieri, V. Persico, and A. Pescapé, "Characterization and prediction of Mobile-App traffic using markov modeling," *IEEE Transactions On Network And Service Management*, Vol.18, No.1, pp.907-925, 2021.
[12] Wireshark [Internet], https://www.wireshark.org/.
[13] Q. Liu, J. Li, and Z. Lu, "ST-Tran: Spatial-temporal transformer for cellular traffic prediction," in *IEEE Communications Letters*, Vol.25, No.10, pp.3325-3329, 2021
[14] D. Aloraifan, I. Ahmad, and E. Alrashed, "Deep learning based network traffic matrix prediction," *International Journal of Intelligent Networks*, Vol.2, pp.46-56, 2021.

김 진 호

https://orcid.org/0000-0003-1477-0754
e-mail : 20173051@gs.cwnu..ac.kr
2017년 ~ 현 재 창원대학교 컴퓨터공학과 학사과정
관심분야 : Networking & Artificial Intelligence

안 동 혁

https://orcid.org/0000-0001-6703-9311
e-mail : donghyeokan@changwon.ac.kr
2006년 한동대학교 전산전자공학부(학사)
2013년 KAIST 전산학과(박사)
2013년 성균관대학교 박사후연구원
2014년 삼성전자 책임연구원
2015년 계명대학교 컴퓨터공학과 조교수
2017년 ~ 현 재 창원대학교 컴퓨터공학과 부교수
관심분야 : 저지연 통신, 시간 민감성 네트워킹, 유무선 네트워크, 사물 인터넷