

Towards Carbon-Neutralization: Deep Learning-Based Server Management Method for Efficient Energy Operation in Data Centers

Sang-Gyun Ma[†] · Jaehyun Park^{††} · Yeong-Seok Seo^{†††}

ABSTRACT

As data utilization is becoming more important recently, the importance of data centers is also increasing. However, the data center is a problem in terms of environment and economy because it is a massive power-consuming facility that runs 24 hours a day. Recently, studies using deep learning techniques to reduce power used in data centers or servers or predict traffic have been conducted from various perspectives. However, the amount of traffic data processed by the server is anomalous, which makes it difficult to manage the server. In addition, many studies on dynamic server management techniques are still required. Therefore, in this paper, we propose a dynamic server management technique based on Long-Term Short Memory (LSTM), which is robust to time series data prediction. The proposed model allows servers to be managed more reliably and efficiently in the field environment than before, and reduces power used by servers more effectively. For verification of the proposed model, we collect transmission and reception traffic data from six of Wikipedia's data centers, and then analyze and experiment with statistical-based analysis on the relationship of each traffic data. Experimental results show that the proposed model is helpful for reliably and efficiently running servers.

Keywords : Data Center, Deep Learning, LSTM, Quality of Service, Server Management, Traffic Prediction

탄소중립을 향하여: 데이터 센터에서의 효율적인 에너지 운영을 위한 딥러닝 기반 서버 관리 방안

마 상 균[†] · 박 재 현^{††} · 서 영 석^{†††}

요 약

최근 데이터 활용이 중요해짐에 따라 데이터 센터의 중요도도 함께 높아지고 있다. 하지만 데이터 센터는 막대한 전력을 소모함과 동시에 24시간 가동되는 시설이기 때문에 환경적, 경제적 측면에서 문제가 되고 있다. 최근 딥러닝 기법들을 사용하여 트래픽을 예측하거나, 데이터 센터나 서버에서 사용되는 전력을 줄이는 연구들이 다양한 관점에서 이루어지고 있다. 그러나 서버에서 처리되는 트래픽 데이터양은 변칙적이며 이는 서버를 관리하기 어렵게 만든다. 또한, 서버 상황에 따라 서버를 가변적으로 관리하는 기법에 대한 연구들이 여전히 많이 요구되고 있다. 따라서 본 논문에서는 이러한 문제점을 해결하기 위해 시계열 데이터 예측에 강세를 보이는 장단기 기억 신경망 (Long-Term Short Memory, LSTM)을 기반으로 한 가변적인 서버 관리 기법을 제안한다. 제안된 모델을 통해 서버에서 사용되는 전력을 보다 효과적으로 줄일 수 있게 되며, 현업환경에서 이전보다 안정적이고 효율적으로 서버를 관리할 수 있게 된다. 제안된 모델의 검증을 위해 위키피디아 (Wikipedia)의 데이터 센터 중 6개의 데이터 센터의 전송 및 수신 트래픽 데이터를 수집한 뒤 통계기반 분석을 통해 각 트래픽 데이터의 관계를 분석 및 실험을 수행하였다. 실험 결과 본 논문에서 제안된 모델의 유의미한 성능을 통계적으로 검증하였으며 서버 관리를 안정적이고 효율적으로 수행할 수 있음을 보여주었다.

키워드 : 데이터 센터, 딥러닝, LSTM, Quality of Service, 서버 관리, 트래픽 예측

1. 서 론

최근 데이터 사용량이 급증함에 따라 대규모 데이터 센터가 많아지고 있다. 하지만 데이터 센터 특성상 수많은 서버, 저장장치, 냉방장치 등을 사용하여 엄청난 양의 전력을 사용하고 있다. 또한 데이터 센터는 24시간 동안 계속 작동하고 있으므로 전력 사용량이 많다[1]. 데이터 센터에서 전기 소비가 많아짐에 따라 온실가스도 급격하게 증가되고 있다[2]. 현재 전 세계적으로 지구온난화를 막기 위해 노력하고 있으며,

※ 이 논문은 2022년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No. NRF-2020R111A3073313).

※ 이 논문은 2022년 한국정보처리학회 ASK 2022의 "탄소중립적 데이터 센터 운영을 위한 LSTM기반 서버 관리 모델"의 제목으로 발표된 논문을 확장한 것임.

† 준 회 원 : 영남대학교 컴퓨터공학과 학사과정

†† 준 회 원 : 영남대학교 컴퓨터공학과 석사과정

††† 종신회원 : 영남대학교 컴퓨터공학과 교수

Manuscript Received : August 1, 2022

First Revision : August 29, 2022

Accepted : August 31, 2022

* Corresponding Author : Yeong-Seok Seo(ysseo@yu.ac.kr)

탄소중립을 실현하기 위해 온실가스를 줄이는 방안을 마련하고 있다[3]. 탄소중립은 인간의 활동에 의한 온실가스 배출을 최대한 줄이고, 남은 온실가스는 산림으로 탄소를 흡수, 제거한 뒤 산소를 재생산하여 실질적인 배출량을 0(Zero)으로 만든다는 개념이다. 데이터 센터의 전기사용량을 줄이는 것은 탄소중립을 실현하기 위해서 필수적이다[4]. 데이터 센터의 전기사용량 감축을 위해 서버에서 사용되는 전기를 줄이려는 다양한 연구들이 진행되었다. 시스템의 부하 수준에 따라서 일부 서버들에게 부하를 집중시킨 후 나머지 서버는 꺼버리거나 소비전력의 추이를 분석하여 소비전력이 가장 작은 서버를 선택하여 끄는 등 데이터 센터의 에너지를 절감하기 위한 다양한 관점에서의 연구들이 존재한다[5-7]. 그러나 서버에서 처리되는 트래픽 데이터양은 변칙적이며, 이는 서버 관리하기 어렵게 만든다. 또한, 가변적으로 서버를 관리하는 기법에 대한 연구들이 여전히 많이 요구되고 있다.

따라서 본 논문에서는 데이터 센터에서 사용되는 에너지 절약 및 서버 전원 제어를 위한 딥러닝 기반 트래픽 예측 모델을 제안하고자 한다. 제안된 모델은 시계열 데이터 예측에 강세를 보이는 LSTM 기반 학습 모델을 사용한다[8]. 또한, 학습 및 성능 평가를 위해 Wikipedia의 데이터 센터 중 6개의 데이터 센터들의 전송 및 수신 트래픽 데이터들을 수집하였으며 데이터 간의 상관관계를 상관분석으로 검토하였다. 제안된 모델을 통해 정확하게 트래픽 데이터양을 예측하게 되면 트래픽 데이터양이 작은 서버의 작업을 남는 서버로 옮기고 해당 서버의 전원을 끌 수 있게 된다. 이렇게 제안된 모델로 서버 전원을 제어하게 된다면 보다 효과적으로 서버의 전력을 제어할 수 있게 되고 서버에서 사용되는 에너지를 절감하여 탄소중립에 한 걸음 더 나아갈 수 있을 것이다.

본 논문의 기여는 다음과 같다.

- 서버 사용 에너지 절감을 위해 LSTM 모델을 활용하여 트래픽 데이터양이 감소하는 시점과 트래픽 데이터양이 낮은 시점을 예측하기 위한 새로운 모델을 제안하였다.
- 6개의 데이터 센터에서 제공하는 실제 트래픽 데이터들을 수집하여 서버 여유자원 및 보정치 분석을 통해 제안된 모델을 실질적으로 검증하였다.
- CPU 사용률과 네트워크 통신 서비스 품질의 보장 범위를 고려하여 서버 여유자원 및 에너지 관리를 현실적으로 수행할 수 있는 기반을 제공하였다.

이하, 1장 서론에 이어 본 논문의 구성은 다음과 같다. 2장에서는 본 논문과 관련된 연구를 기술한다. 3장에서는 본 논문에 대한 전반적인 접근을 설명한다. 다음 4장에서는 실험 설계 및 실험 결과에 관해 서술한다. 마지막으로 5장에서는 본 논문의 결과를 요약하며 결론 및 향후 연구를 기술한다.

2. 관련 연구

본 장에서는 LSTM을 통한 시계열 예측, 에너지 절약을 위

한 서버 제어, 데이터 센터의 전기 사용량 감축 관련 선행 연구들에 대해 소개한다. LSTM을 통한 이상 트래픽 탐지, 서버에서 데이터 배치 및 전원관리, 동적 서버 전원 모드 제어 등의 연구들이 제안되었다[5-6, 9-12].

2.1 시계열 예측 모델 기반 네트워크 관리

Input feature들의 특징을 추출하는 convolution neural network (CNN)와 시계열 데이터에서 높은 성능을 보이는 LSTM을 결합해 각 모델의 장점을 모두 살린 하이브리드 모델이 제안되었다[10]. 그는 Yahoo에서 제공하는 시계열 트래픽 데이터를 제안된 모델을 통해 분석했다. 제안된 모델은 Denial of Service, Probe, User to Root, Remote to User와 같은 비정상적인 트래픽을 탐지할 수 있었다. 또한, 기존 머신러닝 모델에 비해 더 향상된 성능을 보였다. 테스트 데이터에 대해 98.6%의 accuracy와 89.7%의 recall을 보여줬다. 그리고 기존 ARIMA 모델에서 계절적 성분을 추가한 Seasonal AutoRegressive Integrated Moving Average (SARIMA) 모델, LSTM 모델, 그리고 CNN 모델을 사용한 네트워크 트래픽 예측 모델 또한 제안되었다[9]. 제안된 모델은 기존 모델보다 향상된 네트워크 트래픽 예측 성능을 보였으며, 효율적인 네트워크 트래픽 관리에 기여했다. 또한, LSTM과 CNN-LSTM 하이브리드 모델이 SARIMA 모델보다 더 높은 precision을 보였으며, 11% 감소된 오차율을 보인다는 것을 증명했다.

2.2 에너지 절약을 위한 서버 제어

현재 시스템의 부하 수준에 따라서 특정 서버들에게 부하를 집중시킨 후 서버를 즉시 종료함으로써 에너지 소모량을 줄이는 방법이 제안되었다[5]. 기존 방식은 멀티미디어 서버에서 시스템 성능을 향상시키기 위해 저장장치 간 부하 균등을 지원하기 때문에 높은 에너지 소모량을 보였다. 제안된 연구는 기존 방식과 비교했을 때 가동되는 서버의 수를 줄였으며, 이로 인해 에너지 소모량을 크게 감소시킨다는 것을 증명했다. 그리고 서버 소비 전력 이력을 통해 가까운 장래에 서버 소비 전력이 증가할 것인지를 예측한 뒤, 서버의 가동 여부를 제어하여 에너지 소모량을 줄이는 방법 또한 제안되었다[6]. 그는 일정 시간 동안 각 서버에서 소비 전력 데이터를 수집한 뒤 이를 통해 총 소비 전력의 평균과 과거 소비 전력의 평균을 비교하여 이전보다 증가했으면 이후에 증가할 것이라고 예측했다. 제안된 연구는 동일한 성능을 유지하면서 에너지 소모량이 기존 대비 29% 줄어든 것을 증명했다.

2.3 에너지 절약을 위한 냉방 제어

데이터 센터의 효율적 냉각 시스템 운영을 위한 Artificial Neural Network(ANN) 기반 데이터 센터의 최적 냉수 유량 제어알고리즘이 제안되었다[11]. ANN 제어알고리즘은 기존 기법 대비 정확성, 안정성, 에너지 절약성 측면에서 성능 향

상을 보여줬다. 그리고 데이터 센터의 냉방 에너지 사용량을 감축하기 위해 외기의 이용을 극대화할 수 있는 멀티 외기 냉방 시스템 또한 제안되었다[12]. 에너지 시뮬레이션을 통해 검토했을 때 멀티 외기 냉방 시스템은 총 냉방 에너지의 26.7%를 절약할 수 있음을 보여줬다.

관련연구를 통해 서버에서 사용되는 에너지 소모량을 감소시키기 위해 다양한 연구들이 진행되고 있는 것을 알 수 있다. 하지만 가변적으로 서버를 관리하는 기법에 대한 연구는 여전히 도전적인 이슈이다. 따라서 본 논문에서는 트래픽 데이터와 딥러닝 기반 예측 모델을 통해 서버 가동 여부를 제어하며, 서버 상황에 따라 가변적으로 서버를 관리하는 방식에 초점을 맞춘다. 제안된 논문을 통해 트래픽 데이터양이 급증할 때 또는 트래픽 데이터양이 급감할 때 서버를 이전보다 안정적, 효율적으로 관리할 수 있으며 궁극적으로 에너지 소모량을 보다 효율적으로 감소시킬 수 있을 것이다.

3. 접근 방법

Fig. 1은 제안된 논문에 대한 접근법을 나타낸 그림이다. 먼저, 딥러닝 학습을 위해 Wikipedia의 데이터 센터 중 6개의 데이터 센터 (Eqiad, Codfw, Esams, Ulsfo, Eqsin, Drmrs)들의 전송 및 수신 트래픽 데이터들을 수집한 뒤 데이터 간의 상관관계를 상관 분석으로 검토하였다. 데이터 센터에 관한 설명은 Table 1과 같다. 서버 관리를 위해서는 하나의 데이터 센터 트래픽을 예측하면 된다. 따라서 본 논문에서는 Esams 데이터 센터의 전송 트래픽을 예측하였다. 다음, Esams 데이터 센터의 전송 트래픽과 관련된 데이터를 통해 LSTM 모델을 학습하며, 학습된 모델로 미래의 전송 트래픽 데이터양을 예측한다. 일반적으로 Quality of Service (QoS)는 CPU 사용률이 약 60~70% 사이일 때 보장된다 [6]. 따라서 가동되는 서버의 수를 줄이기 위해 QoS 보장과 함께 CPU 사용률을 높이고자 예측된 값에 보정치 1.5를 곱한다. 보정치를 곱한 값에 실제 트래픽 데이터양을 뺀 값을

Table 1. Descriptions of Data Centers

Data Center	Description
Eqiad	Data center located in Ashburn, USA
Codfw	Data Center in Carrollton, Texas, USA
Esams	Data center located in Amsterdam, The Netherlands
Ulsfo	Data Center located in San Francisco
Eqsin	Data Center located in Singapore
Drmrs	Data Center located in Marseille, France.

Equation (1)과 같이 서버 여유 자원이라고 하며, 최솟값, 최댓값, 평균값을 통해 서버 관리 모델을 검증한다. 예를 들어 음수 상태의 서버 여유 자원 값은 트래픽 처리의 실패 및 서버 다운의 발생으로 간주한다. 제안된 모델이 정확하게 트래픽을 예측한다면 LSTM을 통해 예측한 트래픽 값에 보정치를 곱한 값을 통해 서버를 관리하게 된다. 따라서 트래픽을 실제 값보다 다소 낮게 예측하여도 보정치로 인해 서버를 원활하게 관리할 수 있게 된다. 제안된 모델은 실제 값과 예측된 값의 차이를 비교해야하므로 R-squared(R^2) Score를 평가 지표로 사용한다. 또한, QoS를 보장하기 위해 예측된 트래픽 데이터양에서 보정치를 1.5($1/1.5=66.7%$)으로 곱한 방식뿐만 아니라 서버 상황에 따라 보정치를 1.4($1/1.4=71.4%$)~1.8($1/1.8=55.6%$) 사이의 값을 가변적으로 변화시켜가며 검증하였다[6].

$$S_{free}(T) = c * Predicted(T) - Real(T) \quad (1)$$

S: 서버 여유 자원, c: 보정치, T: 트래픽 데이터양

3.1 서버 상황에 따른 가변적 제어

본 장에서는 서버 상황에 따라 보정치를 가변적으로 제어하는 방식을 제안한다. 보정치의 초깃값은 1.5로 설정한다. 현재까지의 서버 여유 자원의 평균을 기준으로 서버 여유 자원이 기준보다 크거나 같으면 보정치에서 0.001을 빼며, 이후

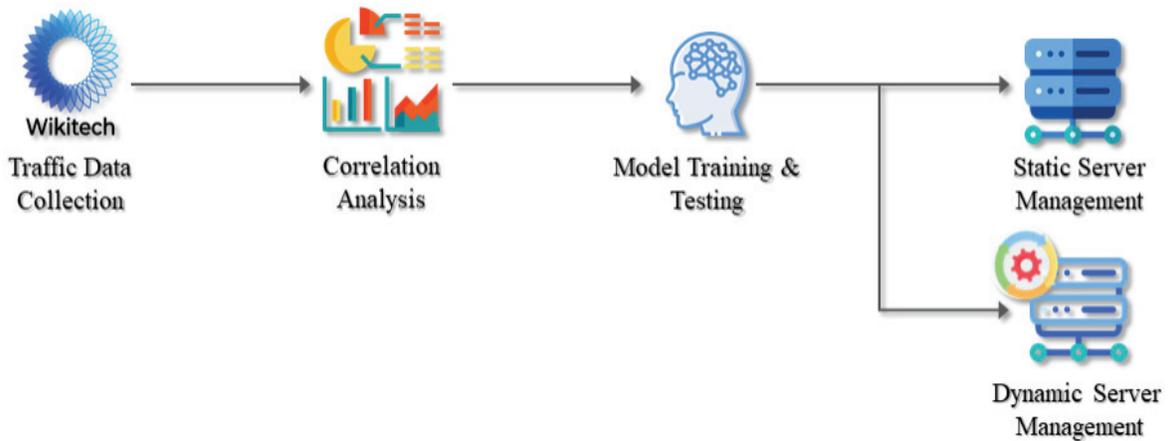


Fig. 1. Overall Approach of the Proposed Method

연속으로 서버 여유 자원이 기준보다 크거나 같으면 0.001에 2를 반복해서 곱한 후 뺀다. 반대로 서버 여유 자원이 기준보다 작으면 보정치에서 0.001을 더하며, 이후 연속으로 서버 여유 자원이 기준보다 작으면 0.001에 2를 반복해서 곱한 후 더한다. 해당 알고리즘의 Pseudo Code는 Table 2와 같다.

Table 2. The Proposed Pseudo Code for Dynamic Server Management

```

1. Num = 1
2. Increase_Decrease = 1 // If the variable is 1, the correction
   range increases and if it is -1, the correction range decreases
3. While True
4.   // Calculate server free resources and average of
   server free resources after predicting network traffic
5.   Traffic_Predict = Result of Traffic Prediction
6.   Free_Resources = Traffic_Predict * Correction_Ranges
   - Real_Traffic
7.   Free_Resources_Average = Result of Server Free
   Resource Average Calculation
8.   // Calculating Correction Ranges
9.   if Free_Resources>Free_Resources_Average then
10.    // If the server resources are sufficient,
    the correction range is reduced.
11.    if Increase_Decrease == 1 then
12.      Num = 1
13.      Increase_Decrease = -1
14.    // If the server resources are sufficient,
    the correction range is further reduced.
15.    else
16.      Num*=2
17.    Endif
18.    Correction_Ranges -= Num*0.001
19.  else
20.    // If the server resources are insufficient,
    the correction range is increased
21.    if Increase_Decrease==-1 then
22.      Num = 1
23.      Increase_Decrease = 1
24.    // If the server resources are insufficient,
    the correction range is further increased
25.    else
26.      Num*=2
27.    Endif
28.    Correction_Ranges += Num*0.001
29.  Endif
30.  // Prevention of exceeding the correction ranges
31.  if Correction_Ranges<=1.4 then
32.    Correction_Ranges = 1.4
33.  Endif
34.  if Correction_Ranges>=1.8 then
35.    Correction_Ranges = 1.8
36.  Endif
37. Endwhile

```

보정치를 서버 상황에 따라 1.4~1.8 사이의 값으로 가변적으로 조절하면 고정적인 값인 1.5로 정하는 것 보다 트래픽 데이터양이 급증할 때 더 안정적으로 서버를 관리할 수 있다. 또한 트래픽 데이터양이 급감할 때 더 효율적으로 서버를 관리하여 에너지 소모량 감소를 기대할 수 있다.

3.2 서버 여유 자원의 최소, 최대, 평균을 통한 모델 검증

본 장에서는 예측 모델을 통해 나온 값을 가공한 결과인 서버 여유 자원의 최소, 최대, 평균에 대해 서술한다. 이 결과는 본 실험의 평가지표로 사용된다.

1) 서버 여유 자원의 최솟값

서버 여유 자원의 최솟값이 높을수록 모든 트래픽 데이터 양을 안정적으로 처리할 만큼 충분한 용량을 가졌으며, 서버 다운에 떨어지기 때문에 서버를 안정적으로 관리했다고 판단한다. 또한 이 값이 낮을수록 안정적으로 관리하지 않았다고 판단한다. 예를 들어 서버 여유 자원의 최솟값이 0.1GB라고 가정한다. 이는 실제 트래픽 데이터양이 조금만 더 높았다면 서버다운이 되었기에 안정적으로 관리되지 않았음을 의미한다. 이 값들은 실숫값을 가지며 그 값이 클수록 더 높은 서버 관리 성능을 의미한다.

2) 서버 여유 자원들의 최댓값, 평균값

서버 여유 자원의 최댓값과 평균값이 낮을수록 용량의 낭비가 적었으며, 서버를 효율적으로 관리했다고 판단한다. 또한 이 값들이 높을수록 비효율적으로 관리했다고 판단한다. 예를 들어 안정적으로 동작하는 서버의 서버 여유 자원의 평균값이 1GB이고 모델에서 서버 여유 자원의 최댓값과 평균이 3GB, 2GB라고 가정한다. 이는 그 차이만큼 서버가 관리되는 동안 낭비가 되고 있었기 때문에 비효율적으로 관리했다고 본다. 서버 여유 자원의 최솟값과 동일하게 실숫값을 가지며 그 값이 작을수록 더 높은 서버 관리 성능을 의미한다. 최종적으로 서버 여유 자원의 최솟값은 높을수록, 서버 여유 자원의 최댓값과 평균이 낮을수록 서버 관리 성능이 높은 것으로 판단한다.

3.3 가변적인 보정치 범위

Table 3에서 볼 수 있듯이 보정치의 최솟값을 1.5으로 설정했을 때는 서버 여유 자원의 최솟값이 높지만 서버 여유 자원의 평균값과 최댓값도 같이 높아져서 비효율적으로 서버관리를 하게 된다. 또한 1.3로 설정했을 때는 서버 여유 자원의 최솟값이 매우 낮아 안정적으로 관리한다고 볼 수 없다. 따라서 실험에서 사용할 보정치의 최솟값을 1.3과 1.5의 평균인 1.4으로 설정했다.

보정치의 최댓값을 2에서 1.8로 줄였을 경우 서버 여유 자원의 평균값이 줄어드는 추세이며, 서버 여유 자원의 최솟값에는 차이가 없다. 그러므로 서버를 효율적으로 관리하기 위해 1.8을 실험에서 사용할 보정치의 최댓값으로 설정했다.

Table 3. Results of Dynamic Correction Ranges

		Server Free Resources		
		Average	Maximum	Minimum
Correction Ranges	1.3~2	0.6315GB	1.7029GB	0.0489GB
	1.4~2	0.8023GB	2.0407GB	0.2662GB
	1.5~2	0.9665GB	2.3785GB	0.2667GB
	1.3~1.9	0.6229GB	1.7029GB	0.0489GB
	1.4~1.9	0.7915GB	2.0407GB	0.2662GB
	1.5~1.9	0.9508GB	2.3785GB	0.2667GB
	1.3~1.8	0.6146GB	1.7029GB	0.0489GB
	1.4~1.8	0.7768GB	2.0407GB	0.2662GB
	1.5~1.8	0.9352GB	2.3785GB	0.2667GB

Table 3의 결과를 통해 얻은 가변적인 보정치 범위의 최소 CPU 사용률은 55.6%(1/1.8)고, 최대 CPU 사용률은 71.4%(1/1.4)이다. QoS는 CPU 사용률이 60~70% 사이일 때 보장되므로 QoS 보장과 함께 서버를 안정적이고 효율적으로 관리할 수 있는 최적의 서버 상황 보정치를 알 수 있다[6].

4. 실험

본 장에서는 모델의 성능 평가에 사용된 평가지표와 실험을 하기 위해 선행된 설계과정에 대해 서술한다. 또한 앞에서 제안된 방법에 대해 실험한 결과를 서술한다.

4.1 평가 지표

1) R² Score

3.2에서 서술된 평가지표 외에 시계열 예측 모델에 적합한 평가지표인 R² Score를 추가적으로 사용하였다. R² Score는 오차제곱의 합에 편차제곱의 합을 나눈 값을 1에 뺀 값이며, 선형 회귀 모델에 대한 적합도 측정값이다. 0에서 1 사이의 값을 가지며 그 값이 클수록 더 높은 예측 성능을 의미한다[13].

2) 상관분석

상관분석은 두 입력 변수 간에 선형적 또는 비선형적 관계가 있는지 분석하는 기법이다[14]. 본 실험에서는 피어슨 상관계수를 사용한다. -1에서 1 사이의 값을 가지며, 두 변수에 대해 동일한 양의 선형관계일수록 1에 가까워지고, 동일한 음의 선형관계를 가지면 -1에 가까워진다. 또한 두 변수에 대해 관계가 다를수록 0에 가까워진다. 값의 범위에 대한 설명은 Table 4과 같다[15].

4.2 실험설계

1) 데이터 수집

본 논문은 실험을 위해 Wikitech에서 Wikipedia의 데이

Table 4. Values of the Pearson Correlation Coefficient

Range	Description
-1.0 ~ -0.7	Strong negative linear relation
-0.7 ~ -0.3	Distinct negative linear relation
-0.3 ~ -0.1	Weak negative linear relation
-0.1 ~ +0.1	Negligible linear relationship
+0.1 ~ +0.3	Weak positive linear relation
+0.3 ~ +0.7	Distinct positive linear relation
+0.7 ~ +1.0	Strong positive linear relation

Table 5. Traffic Data

Time	Transmission data in the Esams Data Center (byte/s)
2022-08-01 0:00	876,219,108
2022-08-01 0:04	914,283,844
2022-08-01 0:08	936,286,381
2022-08-01 0:12	878,264,774
2022-08-01 0:16	826,663,298

터 센터 중 6개의 데이터 센터들의 전송 및 수신 트래픽 데이터를 수집한다[16]. Train set과 test set은 Wikitech에서 제공하는 트래픽 데이터 중 입력변수로 사용할 수 있는 데이터들의 최대 개월 수 차이로 구성하였다. 2022년 4월 1일부터 2022년 4월 10일까지 4분 간격의 10일간 데이터인 3,600개 데이터를 train set으로 사용하였으며, 이 데이터의 4개월 후인 2022년 8월 1일부터 2022년 8월 5일까지 4분 간격의 5일간 데이터인 1,800개 데이터 test set으로 사용하였다.

Table 5는 test set 데이터의 일부이며, Fig. 2는 test set으로 사용된 6개의 데이터 센터의 전송 및 수신 트래픽 그래프를 나타낸 것이다. 각 그래프에 대해 가로축은 시간, 세로축은 해당 시간의 트래픽 데이터양을 의미한다.

2) 데이터 센터 간의 상관분석 결과

본 장에서는 6개의 데이터 센터들의 전송 및 수신 트래픽들이 Esams 데이터 센터의 전송 트래픽에 미치는 영향력을 확인하기 위한 통계기반의 실험을 수행한다. 트래픽 데이터에 대해 피어슨 상관계수를 기반으로 상관분석을 수행하여 각 트래픽 데이터 간의 관계성을 실험 및 검증한다.

Fig. 3는 6개의 데이터 센터들의 전송 및 수신 트래픽들에 대한 상관계수를 보여준다. Drmrs 데이터 센터의 전송 및 수신 트래픽과 Esams 데이터 센터의 수신 트래픽은 Esams 데이터 센터의 전송 트래픽에 대해 값의 편차가 크지 않고 거의 완벽한 선형적 관계를 보인다. 이로 인해 0.97에서 0.99 사이의 값을 가지고 있으며, 이를 통해 매우 강한 상관관계를

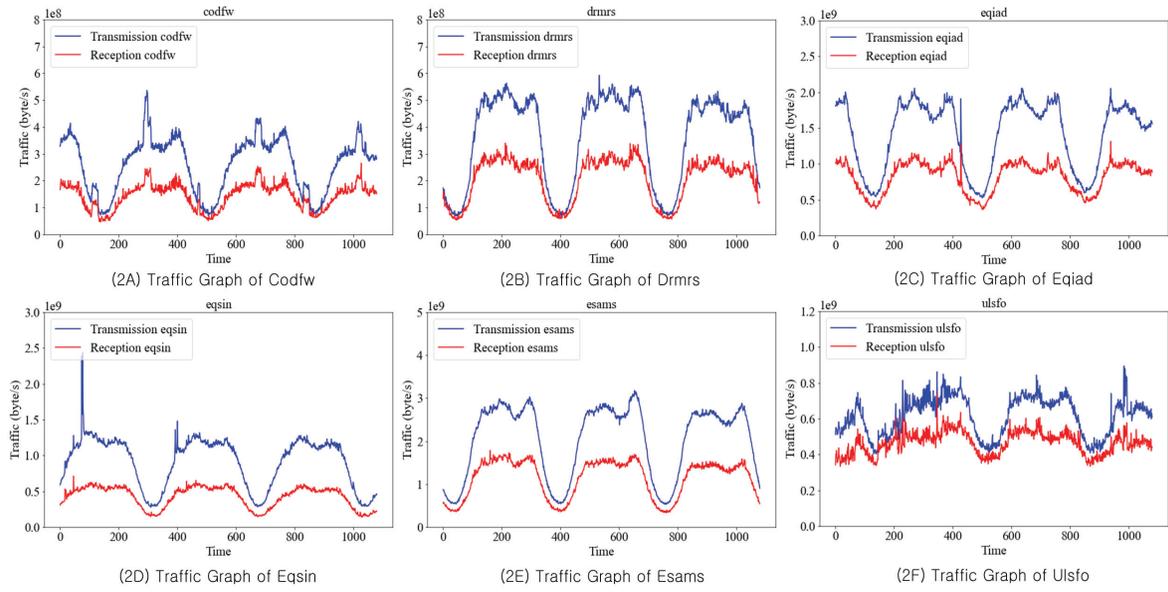


Fig. 2. Traffic Graphs of Data Centers

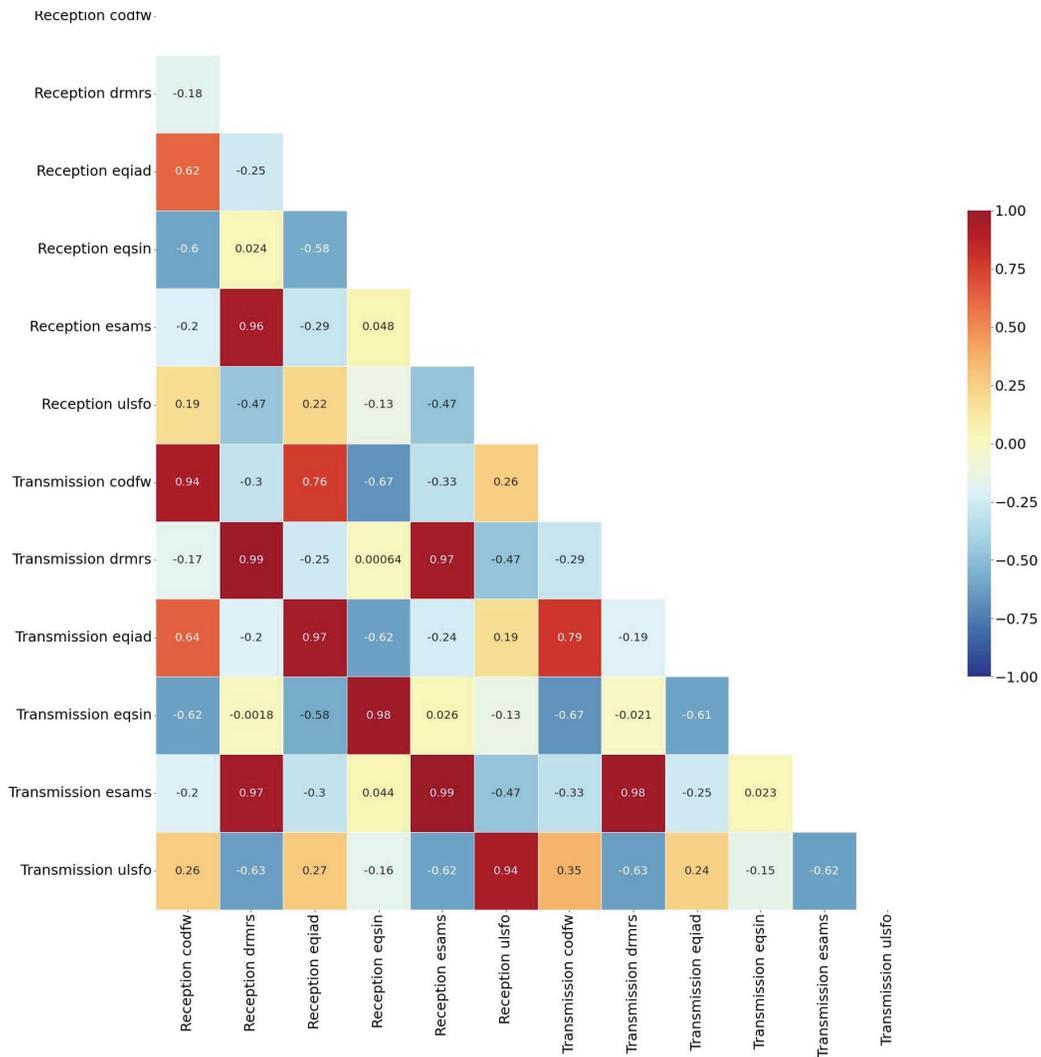


Fig. 3. Results of the Correlation Analysis

가지고 있음을 알 수 있다. Ulsfo 데이터 센터의 전송 및 수신 트래픽과 Esams 데이터 센터의 전송 트래픽간의 상관관계수는 각각 -0.62와 -0.47으로 전반적으로 적지 않은 영향력을 미침을 알 수 있다. 마지막으로 Codfw, Eqiad, Eqsin 데이터 센터들의 전송 및 수신 트래픽과 Esams 데이터 센터의 전송 트래픽간의 상관관계수는 -0.33에서 0.044 사이의 값이므로 상관관계가 거의 없음을 알 수 있다.

상관분석을 통해 얻어진 결과를 통해 6개의 데이터 센터들의 전송 및 수신 트래픽 중 Esams 데이터 센터의 전송 트래픽과 매우 강한 상관관계를 가지는 Drmrs 데이터 센터의 전송 및 수신 트래픽과 Esams 데이터 센터의 전송 및 수신 트래픽을 입력 변수로 사용하였다.

3) 모델 구성

본 논문은 Tensorflow 2.7.0과 Keras 2.7.0 기반인 딥러닝 모델을 사용하였으며, 시계열 데이터 예측에 높은 성능을 보이는 LSTM을 사용하였다[8]. Drmrs 데이터 센터의 전송 및 수신 트래픽 데이터양과 Esams 데이터 센터의 전송 및 수

Table 6. Values of LSTM Hyper-parameters

Hyper-parameter	Value
Unit	50
Activation	Tanh
Optimizer	Adam
Learning_rate	0.001
Batch_size	32
Epoch	1000
Callbacks	Checkpoint, Earlystopping
Earlystopping	monitor="loss", patience=10
Window size	15

신 트래픽 데이터양을 입력 변수로 사용하며, 최종적으로 미래의 트래픽 데이터양이 output으로 출력된다. 하이퍼 파라미터의 경우 시간적 비용이 적게 소모되면서 최적의 성능을 보이는 값으로 설정하였으며, Table 6에서 상세한 값을 확인할 수 있다.

Fig. 4-7은 하이퍼 파라미터 중 Window size에 따른 실



Fig. 4. Experimental results by changing the window size (R2 Score)

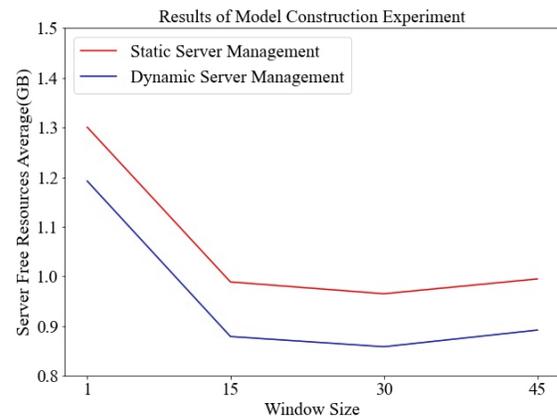


Fig. 5. Experimental results by changing the window size (Average)

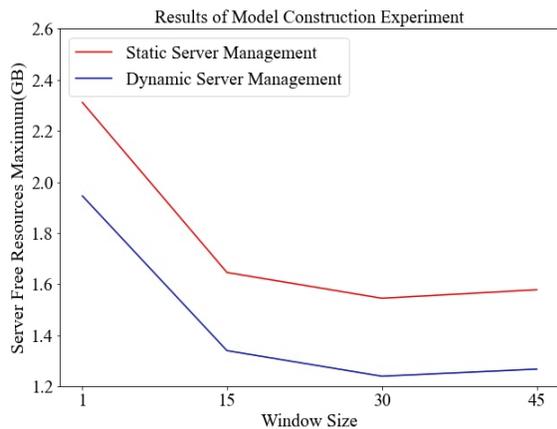


Fig. 6. Experimental Results by Changing the Window Size (Maximum)

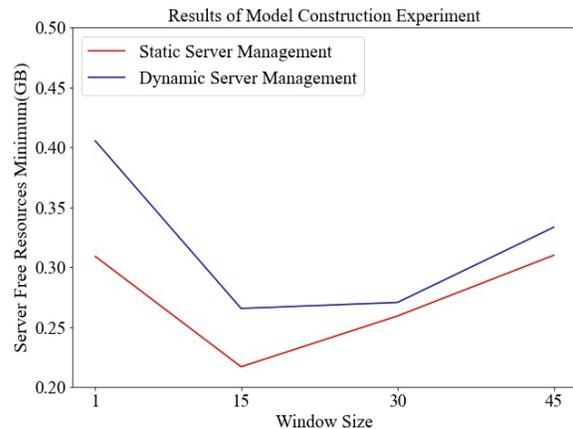


Fig. 7. Experimental Results by Changing the Window Size (Minimum)

Table 7. Summary of Experimental Results

	Static Correction Ranges	Dynamic Correction Ranges	Rate of Change
R2 score	0.991	0.991	
Server Free Resources Average	0.8679GB	0.7768GB	-10.5%
Server Free Resources Maximum	2.3785GB	2.0407GB	-14.2%
Server Free Resources Minimum	0.2017GB	0.2662GB	+31%

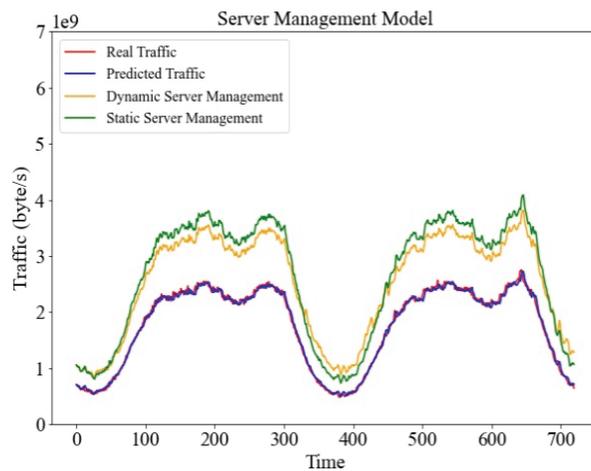


Fig. 8. Traffic Graphs of Experimental Results

험 결과이다. 트래픽 데이터는 4분 간격으로 수집되기 때문에 1시간동안 수집된 트래픽 데이터는 15개이다. 그러므로 Window size를 15개씩 늘려서 실험하였다.

Window size를 설정하지 않았을 경우 다른 결과들에 비해 트래픽 데이터양을 정확하게 예측을 하지 못하며, 서버관리를 비효율적으로 하고 있다. 이는 R² Score와 서버 여유 자원의 최솟값이 다른 결과들에 비해 낮은 것에서 확인할 수 있으며, 서버 여유 자원의 평균값과 최댓값이 다른 결과들에 비해 높은 것에서 확인할 수 있다.

Window size를 설정했을 경우 트래픽 데이터양을 정확하게 예측하며, 서버관리를 효율적으로 하는 것을 평가지표를 통해 확인할 수 있다. Window size를 15, 30, 45로 설정했을 경우 모두 실험 결과가 준수하므로 Window size를 시간적 비용이 가장 적게 소모되는 경우인 15로 설정하였다.

4.3 실험 결과

Test set을 활용하여 3.2장에서 서술된 평가지표 측면과 R² score 측면에서 정리한 결과는 Table 7과 같다. R² Score 측면에서는 Table 7의 두 결과는 동일한 모델에서 나온 결과에 대해 진행한 내용이다. R² Score가 0.991라는 의미는 제안된 모델이 실제 값에 대해 99.1% 적합하다는 의미이다. 따라서 제안된 모델이 높은 수준으로 실제 트래픽

데이터를 예측함으로써 서버 관리에 적합함을 알 수 있다.

3.2장에서 서술된 평가지표 측면에서는 서버 여유 자원의 평균값과 최댓값에 대해 보정치를 고정적으로 사용했을 때 보다 보정치를 가변적으로 바꾸었을 때가 약 12% 감소하는 결과를 보였다. 또한, 서버 여유 자원의 최솟값에 대해 보정치를 가변적으로 바꾸었을 때가 보정치를 고정적으로 사용했을 때 0.2017에서 0.2662으로 약 31% 증가하는 결과를 보였다.

Fig. 8과 같이 트래픽 데이터양이 적을 때와 트래픽 데이터양이 급증할 때 보정치를 가변적으로 두었을 경우가 고정적으로 두었을 경우보다 서버 여유 자원이 더 많은 것을 볼 수 있다. 또한 트래픽 데이터양이 많을 때와 트래픽 데이터양이 급감할 때 보정치를 가변적으로 두었을 경우에 서버 여유 자원이 더 적은 것을 볼 수 있다. 이는 서버를 관리할 때 보정치를 가변적으로 바꾸며 관리했을 때가 더 안정적이고 효율적으로 서버를 관리됨을 의미한다.

본 실험을 통해 내린 결론은 다음과 같다. 서버를 관리할 때 고정적인 값을 보정치로 설정한 뒤 LSTM 기반 트래픽 데이터양 예측 모델을 사용할 때가 LSTM 기반 트래픽 예측 모델을 사용하지 않았을 때 보다 서버를 효율적으로 관리한다. 또한, 고정적인 값을 보정치로 사용할 때 보다 보정치를 서버 상황에 따라 가변적으로 바꾸며 서버를 관리했을 때가 항상

된 서버 관리 성능을 보임을 3.2장의 평가지표를 통해 알 수 있다. 또한, 높은 수준의 트래픽 예측 성능을 가지고 있어 서버관리에 사용되기 적합한 모델임을 R² Score 평가지표를 통해 알 수 있다.

5. 결 론

본 논문에서는 효율적인 서버 관리를 위해 LSTM 기반 트래픽 예측 모델을 제안했다. 모델 검증을 위해 Wikitech에서 Wikipedia의 데이터 센터 중 6개의 데이터 센터들의 전송 및 수신 트래픽 데이터들을 수집했으며, R² Score와 3.2장에서 서술된 평가지표를 객관적인 실험과 검증을 위해 사용했다. LSTM 기반 트래픽 예측 모델은 R² Score가 0.991라는 결과를 보였으며, 보정치를 고정적으로 두었을 때 서버를 원활하게 관리하는 모습을 볼 수 있었다.

그러나 고정적으로 두었을 때보다 보정치를 서버 상황에 따라 가변적으로 바꾸었을 때 서버 여유 자원들의 최솟값은 약 31% 증가했으며, 서버 여유자원들의 평균값과 최댓값이 약 12% 낮아진 모습을 볼 수 있었다. 제안된 모델은 두 평가지표를 통해 유의미한 성능임을 보여주었으며, 서버 관리를 안정적이고 효율적으로 수행함을 보여주었다. 본 논문을 통해 현업환경에서 이전보다 서버를 더 효율적으로 관리 할 수 있으며, 궁극적으로 에너지 소모량을 감소해 탄소중립에 기여할 것으로 기대된다.

본 실험에서는 트래픽 데이터만 사용했으나 서버 네트워크는 다양한 요인들로부터 영향을 받는다. 그리고 서버의 전력 감소량과 연산 성능은 trade-off 관계에 있다. 따라서, 향후에는 이 다양한 요인들을 분석한 후 딥러닝 모델에 적용하고 에너지 감소량 대비 성능 저하를 분석하는 방식으로 연구를 발전시켜 나갈 계획이다.

References

- [1] A. Shehabi et al., "United states data center energy usage report," Lawrence Berkeley National Laboratory, Berkeley, California. LBNL-1005775, 2016.
- [2] D. Bouley, "Estimating a data center's electrical carbon footprint," *Schneider Electric White Paper Library*, pp.14-22, 2011.
- [3] S. J. Davis et al., "Net-zero emissions energy systems," *Science*, Vol.360, No.6396, pp.eaas9793, 2018.
- [4] J. Judge, J. Pouchet, A. Ekbote, and S. Dixit, "Reducing data center energy consumption," *Ashrae Journal*, Vol.50, No.11, pp.14-26, 2008.
- [5] K. J. Lee and E. S. Kim "Data placement and power management for energy saving in multimedia servers," *Journal of Digital Contents Society*, Vol.19, No.1, pp.43-49, 2018.
- [6] H. Y. Kim, C. H. Ham, H. K. Kwak, H. U. Kwon, Y. J. Kim, and K. S. Chung, "A dynamic server power mode control for saving energy in a server cluster environment," *The KIPS Transactions: PartC*, Vol.19C, No.2, pp.135-144, 2012.
- [7] S. G. Ma, J. H. Park, and Y. S. Seo, "LSTM-based server management model for carbon-neutral data center operation," *Proceedings of the Annual Spring Conference of Korea Information Processing Society Conference (KIPS) 2022*, Vol.29, No.1, pp.487-490, 2022.
- [8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, Vol.9, No.8, pp.1735-1780, 1997.
- [9] Y. J. Jang, "Network prediction of traffic generation amount using time series prediction model," [Internet], <https://repository.hanyang.ac.kr/handle/20.500.11754/168385>.
- [10] T. Y. Kim and S. B. Cho, "Web traffic anomaly detection using C-LSTM neural networks," *Expert Systems with Applications*, Vol.106, pp.66-76, 2018.
- [11] B. R. Park, Y. J. Choi, J. Y. Hyun, Y. R. Tae, and J. W. Moon, "Development of optimal chilled water mass flow rate prediction and control algorithm for data center-cooling energy saving," *Journal of the Korea Institute of Ecological Architecture and Environment*, Vol.21, No.3, pp.47-53, 2021.
- [12] J. Y. Kim, H. J. Chang, and Y. H. Jung, "Study on the development of energy conservation system for data center by utilizing multi-staged outdoor air cooling," *Conference of Architectural Institute of Korea*, Vol.34, No.2, pp.267-268, 2014.
- [13] L. Jin and S. C. Myers, "R² around the world: New theory and new tests," *Journal of financial Economics*, Vol.79, No.2, pp.257-292, 2006.
- [14] O. Israeli, "A Shapley-based decomposition of the R-square of a linear regression," *The Journal of Economic Inequality*, Vol.5, No.2, pp.199-212, 2007.
- [15] B. Jacob, C. Jingdong, and H. Yiteng, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*, pp.1-4, 2009.
- [16] Wikitech [Internet], <https://wikitech.wikimedia.org/>.



마 상 균

<https://orcid.org/0000-0003-0243-7743>
e-mail : wodon326@yu.ac.kr
2020년 ~ 현 재 영남대학교 컴퓨터공학과
(학사)
관심분야 : Data Mining, AI, and Deep Learning



서 영 석

<https://orcid.org/0000-0002-5319-7674>
e-mail : ysseo@yu.ac.kr
2008년 KAIST 전산학과(석사)
2012년 KAIST 전산학과(박사)
2014년 ~ 2016년 한국산업기술시험원
선임연구원
2016년 ~ 현 재 영남대학교 컴퓨터공학과 교수
관심분야 : Software Engineering, Artificial Intelligence,
Internet of Things, and Big Data Analysis



박 재 현

<https://orcid.org/0000-0002-7807-4048>
e-mail : kat3160@yu.ac.kr
2021년 영남대학교 컴퓨터공학과(학사)
2021년 ~ 현 재 영남대학교 컴퓨터공학과
석사과정
관심분야 : Data Mining, Software
Engineering, AI, and Deep Learning