

A Temporal Convolutional Network for Hotel Demand Prediction Based on NSGA3 Feature Selection

Keehyun Park*, Gyeongho Jung**, Hyunchul Ahn***

*PhD Candidate, Graduate School of Business IT, Kookmin University, Seoul, Korea

**Master's Candidate, Graduate School of Business IT, Kookmin University, Seoul, Korea

***Professor, Graduate School of Business IT, Kookmin University, Seoul, Korea

[Abstract]

Demand forecasting is a critical element of revenue management in the tourism industry. Since the 2010s, with the globalization of the tourism industry and the increase of different forms of marketing and information sharing, such as SNS, forecasting has become difficult due to non-linear activities and unstructured information. Various forecasting models for resolving the problems have been studied, and ML models have been used effectively. In this study, we applied the feature selection technique (NSGA3) to time series models and compared their performance. In hotel demand forecasting, it was found that the TCN model has a high forecasting performance of MAPE 9.73% with a performance improvement of 7.05% compared to no feature selection. The results of this study are expected to be useful for decision support through improved forecasting performance.

▶ **Key words:** Hotel Demand Forecasting, Time Series, Feature Selection, NSGA3, TCN

[요약]

수요 예측은 관광 산업에서 수익 관리의 중요한 요소이다. 2010년대 이후 관광 산업의 세계화와 SNS와 같은 다양한 형태의 마케팅 및 정보 공유가 증가함에 따라 비선형 활동과 비정형 정보로 인해 예측이 어려워졌다. 이러한 문제를 해결하기 위한 다양한 예측 모델이 연구되었으며, 기계 학습(ML) 모델이 효과적으로 사용되었다. 본 연구에서는 특징 선택 기법(NSGA3)을 시계열 모델에 적용하고 성능을 비교하였다. 호텔 수요 예측에서 TCN 모델은 MAPE 9.73%로, 특징 선택을 적용하지 않았을 때보다 7.05% 성능이 향상된 높은 예측 성능을 보였다. 본 연구 결과는 향상된 예측 성능을 통해 의사결정 지원에 유용할 것으로 기대된다.

▶ **주제어:** 호텔 수요 예측, 시계열, 특징 선택, 비지배정렬 유전 알고리즘3, 시간 컨볼루션 네트워크

- First Author: Keehyun Park, Corresponding Author: Hyunchul Ahn
- *Keehyun Park (keehpark@kookmin.ac.kr), Graduate School of Business IT, Kookmin University
- **Gyeongho Jung (gyeongho9878@kookmin.ac.kr), Graduate School of Business IT, Kookmin University
- ***Hyunchul Ahn (hcahn@kookmin.ac.kr), Graduate School of Business IT, Kookmin University
- Received: 2024. 09. 27, Revised: 2024. 10. 10, Accepted: 2024. 10. 18.

I. Introduction

호텔 및 여행 산업은 국가 경제에서 중요한 역할을 담당하며, 코로나19 안정화 이후 새로운 여행 트렌드가 등장하고 있다. 이와 함께 여행 및 호텔 수요 예측의 중요성이 대두되고 있으며, 이는 운영 수익을 극대화하는 핵심 의사결정 요소로 작용한다. 호텔 객실은 시간과 장소에 제약을 받는 특수한 상품이기에 때문에, 정확한 수요 예측이 수익 관리에 필수적이다. 기존 수요 예측 모델은 안정적인 시계열 데이터와 경제적 변수를 활용해 왔다. 그러나 2010년대 이후 관광 수요는 비선형적 특성을 보이며, 전통적인 방법으로는 정확한 예측이 어려워졌다. 이러한 한계를 극복하기 위해 다양한 전처리 기법과 기계 학습 모델이 연구되고 있지만, 호텔 수요 예측에 대한 적용 사례는 여전히 부족하다[1].

여러 분야에서 시계열 예측에 활용되는 머신러닝 기법은 개선된 모델을 통해 높은 성능을 보이고 있다. 특히, 순환 신경망(RNN) 기반의 LSTM, GRU 모델이 좋은 결과를 보여주었으며, 딥러닝에서 이미지 처리에 사용되던 합성곱 신경망(CNN)은 개선 과정을 거쳐 순차적 데이터 처리를 위한 TCN(Temporal Convolutional Network) 모델로 발전했다[2]. TCN은 순환 구조가 아닌 1D-Convolution을 사용하여 순차적 데이터를 학습하는 구조로, 과거 데이터를 바탕으로 미래를 예측하는 데 매우 효과적이다. 또한, 예측 성능 향상과 자원 최적화를 위해 특징 선택(feature selection)과 하이퍼파라미터 조정이 중요하며, 이 중 특징 선택은 많은 연구에서 다루어져 왔다.

본 연구에서는 호텔 운영 데이터에서 비지배 정렬 유전자 알고리즘(NSGA3, Non-dominated Sorting Genetic Algorithm III)을 적용하여 특징 변수를 추출하고, 이 변수를 기반으로 TCN 모델을 적용하여 성능을 향상시키고자 한다. 이를 통해 호텔 수요 예측의 정확성을 높이고, 의사결정 지원을 위한 최적의 정보를 제공할 수 있을 것으로 기대된다.

II. Preliminaries

1. Hotel Demand Time Series Forecasting

호텔 경영에서는 세심한 계획과 체계적인 운영이 필수적이다. 비즈니스 전략 수립, 자금 및 조직 운영, 고객 및 파트너 관계, 회계 등의 다양한 업무가 긴밀하게 협력해야 한다. 특히, 미래 수요를 예측하는 것은 호텔 운영의 효율

성 향상과 수익 극대화를 위해 핵심적인 역할을 한다. 수요 예측은 자원의 효율적 활용을 가능하게 하며, 손실 최소화 및 서비스 품질 유지를 위한 중요한 요소이다. 이러한 요구를 충족시키기 위해 수요 예측과 관련된 다양한 연구가 진행되어 왔다.

기존의 통계적 기법은 비교적 안정적인 데이터에 기반한 장기적 예측에 유용하게 사용되었으며, ARIMA, SARIMA 등의 시계열 분석 모델이 대표적으로 사용되었다. 최근에는 기계 학습과 딥러닝 기술이 수요 예측에 적극적으로 활용되고 있다. 수요 예측의 정확성을 개선하는 것은 호텔 운영 최적화와 안정적인 수익 관리를 위한 핵심적인 요소로 인식되고 있다.

순환신경망(RNN)은 순차적 데이터 처리에 적합하여 자연어 처리, 음성 인식, 시계열 데이터 분석에 많이 사용되었으나, 과거 정보 기억의 한계, 병렬 처리 불가로 인한 학습 속도 지연, 기울기 소실 등의 문제를 가지고 있다. 이를 해결하기 위해 LSTM, GRU 등의 개선된 모델이 개발되어 여러 분야에서 활발히 연구되고 있다[3]. 합성곱 신경망(CNN)은 데이터 특징 추출과 패턴 파악에 효율적이어서 공간적 데이터, 특히 이미지 데이터 처리에 적합하다. 또한, 시계열 데이터의 시간 의존성을 모델링하기 위해 RNN과 결합하여 좋은 성과를 보인 사례가 있다[4].

CNN은 2D 또는 3D 공간적 정보 처리를 위해 설계되었으며, 병렬 처리로 인한 빠른 처리 능력과 높은 특징 추출 능력, 노이즈에 강한 특성 덕분에 여러 분야에서 연구되고 있는 모델이다. CNN의 개선된 모델인 TCN은 시계열 데이터 연구 분야에서 RNN, LSTM 등에 대한 대안으로 설계되었으며[2], 1D-CNN을 사용하여 1차원 배열 형태의 시계열 또는 순차적 데이터 처리 모델에 적용되어 좋은 성과를 보여주고 있다[5].

2. Feature Selection Models

특징 선택(FS, Feature Selection)은 기계 학습 모델의 성능을 높이기 위한 주요한 기능 요소이다. 여러 특징 선택 방법 중에서 유전자 알고리즘(GA, Genetic Algorithm)은 찰스 다윈의 적자생존(survival of the fittest)의 원리를 기반으로 한 최적화 기법이다. 병렬 처리, 전역 최적화, 확률적 접근 등의 특징을 가지고 있으며, 기계 학습 모델의 성능 향상과 복잡성을 줄이기 위해 사용된다. GA는 많은 변수 중에서 최적의 변수를 추출할 가능성이 높고, 변수 간 비선형적인 상호작용이 반영되므로 전역 최적해를 찾을 가능성이 높은 알고리즘이다.

Ahn(2014)은 기업 신용등급 평가 모델에서 GA를 이용한

하이퍼파라미터 설정 및 입력 변수 집합 탐색 모델을 다분류 SVM과 융합하였으며, 다항 로지스틱 회귀 분석(MLOGIT), 사례 기반 추론(CBR), 인공지능망(ANN) 등과 비교하여 일관성 있는 우수한 성능을 확인하였다[6]. Pethe et al.(2024)은 소프트웨어 결함 예측 연구에서 GA를 적용한 특징 선택(FS)을 통해 나이브 베이즈 분류기(NBC), K-Nearest Neighbor, 의사결정트리 분류기(DTC), 랜덤포레스트(RF) 모델 모두에서 성능이 향상된 것을 확인하였다[7]. Alghamdi(2023)은 호텔 상품 구매 의사결정에 영향을 미치는 소셜 데이터 분석 연구에서 필터 모델인 상관관계 기반 특징 선택(CFS, Correlation-based Feature Selection)을 적용하여 성능 향상을 확인하였다[8].

NSGA3은 다목적 최적화 문제 해결을 위한 GA의 개선된 모델이다. 여러 개의 목적 함수를 동시에 최적화하는 해를 찾는 과정에서 변수 간 상충 관계(trade-off)를 최소화하는 최적 해결책을 찾기 위한 방법이다. 이를 위해 가중치 적용, 제약 조건 적용 등을 포함하는 지속적인 개선 모델이 연구되었으며, NSGA (1994), NSGA2 (2002)를 거쳐 NSGA3 (2014) 개선 모델이 완성되었다. 이 개선 과정에서 계산의 복잡도는 줄어들었으며, 적합도 공유, 혼잡 거리, 참조점 방식을 통해 메커니즘이 발전하였다. 또한 목적 함수의 수가 2~3개에서 더 많은 수로 확장되면서 효율성과 성능이 향상되었다.

3. Temporal Convolutional Network(TCN)

시간 컨볼루션 네트워크(TCN, Temporal Convolutional Network)은 시계열 데이터 처리를 위해 설계된 CNN 모델이다. 현재 시점의 출력이 과거 시점의 입력에만 의존하는 인과 컨볼루션(causal convolution)과 확장 컨볼루션(dilated convolution)을 통해 효과적인 장기 의존성을 모델링할 수 있다. TCN은 병렬 처리로 인한 빠른 학습 속도와 고정된 메모리 길이를 가지는 장점이 있다. 여러 연구에서 RNN 계열 모델과 비교하여 상대적으로 우수한 성능을 보인다는 결과가 도출되었다. 특히, 여러 연구를 통해 특징 선택과 TCN 결합을 통한 성능 향상이 확인되었다.

Li et al.(2021)은 화력 발전기의 과열기 온도 예측 연구에서 그래이너 인과관계 검정(Granger causal test)을 통한 특징 선택과 2D-TCN 모델을 적용하여 RNN, LSTM, GRU 대비 높은 예측 성능을 보여주었으며, 속도 측면에서 RNN 대비 6.4배, GRU 대비 2.2배의 빠른 속도가 확인되었다[9]. Zha et al.(2024)은 항공 엔진의 잔여 수명 예측 연구에서 XGBoost 모델을 이용한 특징 선택과 개선된

TCN 모델을 적용하였으며, 하이퍼파라미터 조정을 통해 생성한 두 가지 모델 연구에서 DCNN, BiLSTM 등 다른 비교 모델 대비 우수한 성능을 보여주었다[10].

III. Research Model

이전 연구에서 적용된 다양한 예측 모델들은 예측 성능이 제한적이었다. 전통적인 선형 회귀 모델은 계절적 변화와 같은 장기적인 관점에서 안정적인 예측 성능을 보였으나, 급변하는 경제 상황이나 재해 등의 국제적 환경 변화 및 비선형적인 패턴에 적응하는 데 한계가 있었다. 또한, 기계 학습 기반의 예측 모델들은 데이터의 다양성이 확보되지 못하는 경우가 많았다. 특히 호텔 수요 예측에 대한 대부분의 연구는 특정 지역이나 특정 호텔 데이터를 대상으로 한 제한된 모델이었으며, 비록 높은 예측 정확도를 보였으나 일반화하기 어려운 상황이었다.

이에 대한 대안으로 Park et al.(2024)는 다양한 호텔 데이터를 이용하여 전역 모델과 그룹화된 지역 모델을 구축하였다. 실시간 데이터 분석 후 전역 또는 지역 모델을 선택적으로 적용하여 예측 성능을 개선하려고 하였으나, 사용된 기계 학습 모델의 예측 성능에 한계가 있었다[1].

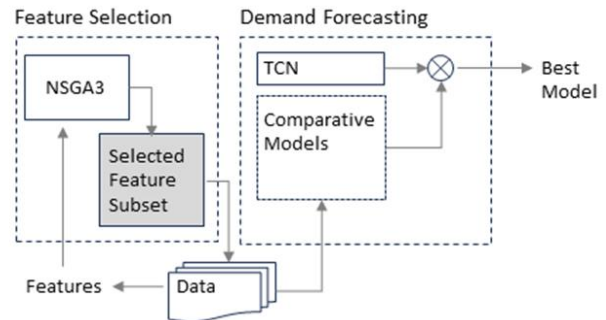


Fig. 1. Proposed NSGA3-TCN architecture

본 연구에서는 Fig. 1에 제시된 바와 같이 호텔 수요 예측 모델 구축을 위해 특징 선택과 시계열 CNN 융합 모델을 제안한다. 특징 선택(FS) 기법으로는 개선된 유전자 알고리즘인 NSGA3를 적용하여 모델에 반영할 변수 목록을 추출하고, 예측 모델에는 최근 시계열 분석에서 좋은 성과를 보이는 TCN 모델을 적용한다. 그리고 이상 제안모델의 우수성은 다른 다양한 머신러닝 모델들과의 비교를 통해 확인하게 된다.

본 제안모델의 핵심적인 학술적 기여는 이전 호텔 수요 예측에서 사용되지 않은 개선된 유전자 특징 선택 모델과

시간 컨볼루션 네트워크(TCN)의 결합을 통해 호텔 수요 예측의 일반화 성능을 높이고, 하이퍼파라미터를 최적화하여 예측 성능을 향상시키는 것이다. LSTM, GRU, CNN 융합 모델들은 다양한 분야에서 높은 성능을 보였지만 호텔 수요 예측에서는 그렇지 못했다. 이에 성능 개선을 위한 시도로 NSGA3-TCN 융합 모델을 제안한다.

IV. Empirical Test

1. Experimental Procedure

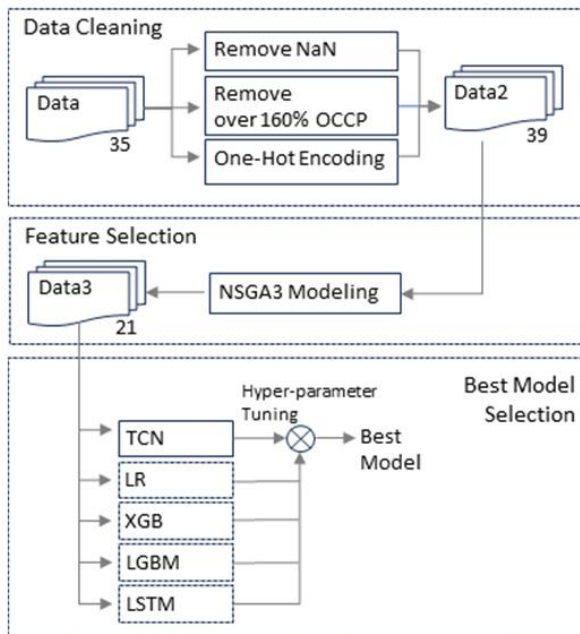


Fig. 2. The flow of the experimental process

본 연구의 제안모델을 검증하기 위한 전체적인 실증분석은 다음과 같은 흐름으로 진행된다. 먼저 호텔 수요 예측을 위한 데이터 수집을 위해, 실제 호텔에서 운영되는 시스템에 저장된 데이터 중 예측 성능에 영향을 줄 수 있는 변수 35개를 수집하였다. 이어, 수집된 데이터에 대해 결측값 제거, 일투속률 160% 이상의 이상치(outlier) 데이터 제거를 진행하였다. 그런 다음, 호텔등급, 지역과 같은 범주형 변수를 '원핫인코딩(One-Hot Encoding)'을 통해 숫자형 변수로 전환하여 총 39개의 항목으로 재정리하였다. 이렇게 정리된 39개 변수 데이터를 기준으로 NSGA3 특징 선택 모델을 적용하여, 예측 성능에 영향을 줄 수 있는 것으로 분류되는 21개 입력 변수 그룹을 추출한다. 끝으로 추출된 변수 데이터를 통해 TCN, LGBM, XGB, LR, LSTM 등 다양한 기계학습 모델을 적용하여 예측 성능을

측정함으로써, 제안 모델의 우수성을 확인하였다. Fig. 2는 이와 같은 본 연구에서 수행된 실증분석의 흐름을 도식으로 나타내고 있다.

2. Data Collection and Preprocessing

본 연구에서 활용된 데이터는 국내 A사의 시스템에 기록된 데이터를 사용하였다. A사는 국내 호텔들을 대상으로 호텔 관리 시스템을 제공하고 있으며, A사가 보유한 데이터 중에서 90개 이상의 객실을 보유한 12개 호텔을 선정하여 2017년부터 2019년까지의 일별 판매 관련 데이터를 추출하였다. 추출된 데이터에서 범주형 변수는 '원핫인코딩'을 통해 2차 분류하였고, 이후 NSGA3 특징 선택 알고리즘을 통해 추출된 변수들로 데이터를 재생성하였다. 재생성된 데이터를 바탕으로 제안 모델과 비교 모델의 기계학습을 진행하였으며, 제안 모델과 비교 모델들의 측정지표를 비교하여 예측 성능을 검증하였다.

3. Experimental Environment

본 연구의 모형 구축 과정에서 수집된 데이터 중, 앞서 설명한 방법을 통해 실험에 적용할 13,057건의 데이터를 추출하였다. 전체 데이터 세트의 70%는 훈련용으로, 15%는 검증용으로 설정하였으며, 나머지 15%는 최종 시험용 데이터로 사용하였다. 실험에 사용된 데이터와 하드웨어, 소프트웨어 사양은 표 1과 표 2에 각각 정리되어 있다.

Table 1. Data Configuration

Data Type	Portion	No. of Samples
Training	70%	9,134
Validation	15%	1,956
Test	15%	1,967

Table 2. Experimental Environment

Type	Environment
H/W1	CPU : Ryzen 9 7950x3D GPU : RTX 4090 M/M 64GB
H/W2	CPU : Intel i7-11370H 3.3GHz GPU : NVIDIA MX450 M/M 32GB
S/W	OS1 : Ubuntu 22.04.4 LTS OS2 : Windows 11 Python 3.10.12 GPU driver : nvidia-driver-535 Cuda 12.2 Pytorch 2.4.1 Tslearn 0.6.3 scikit-learn 1.2.2 Numpy 1.23.5 Pandas 1.5.3

아울러, 본 연구에서 제안모델 학습에 사용된 옵티마이저, NSGA3, TCN 모델에 적용된 하이퍼파라미터(hyperparameter)는 표 3과 같다.

Table 3. Hyperparameters

Type	Parameters
Optimizer	Torch optimizer : Adam Batch size : 256 Epochs : 5 Optimizer learn rate : 0.001
NSGA3	Reference direction : das-dennis, 2 dimension, 12 partitions Generation : 250 Pop_size : 200 Best_solution : >= 0.2
TCN	Selected_features : 180 Learn_rate : 0.001 Epochs : 80 Batch_size : 256 Num_inputs : 180 Num_channels : [64,64,64] Kernel_size : 3 Dropout : 0.2

제안 모델과 비교 모델의 성능 비교를 위해 사용한 정확도(Accuracy) 평가지표로는 MAPE, MSE, RMSE를 사용하였다. 이 중, MAPE(Mean Absolute Percentage Error)는 예측값과 실제값 사이의 백분율 오차의 평균을 나타내며, 값이 작을수록 예측 성능이 우수함을 의미한다. MSE(Mean Squared Error)는 실제값과 예측값 간의 차이를 제공한 후 그 평균을 계산하여 오차의 크기를 측정하며, 이 값 또한 작을수록 예측 성능이 뛰어나다. RMSE(Root Mean Squared Error)는 MSE에 제곱근을 적용한 값으로, 예측값과 실제값 간의 차이를 제공한 뒤 평균하고, 그 결과에 제곱근을 취한 값이다. 이 역시 값이 작을수록 성능이 우수하다. 다음 식 (1)~(3)은 이 지표들을 산출하는 산식들을 나타내고 있다.

$$MAPE(\%) = \frac{1}{n} \sum \left(\left| \frac{a-f}{a} \right| \right) \times 100 \quad (1)$$

$$MSE = \frac{1}{n} \sum (a-f)^2 \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (a-f)^2} \quad (3)$$

(a: 실제값, f: 예측값)

4. Experimental Results

본 연구에서는 호텔 수요 예측 모델에 유전자 알고리즘(NSGA3)을 적용한 특징 선택(FS)과 개선된 컨볼루션 네

트워크인 TCN을 적용하여 제안 모델의 성능을 평가하고자 하였다. 이를 위해 제안 모델을 포함한 5가지 머신러닝 알고리즘의 성능을 비교하고 분석하였다.

초기 35개의 변수 그룹 중 정제 과정과 NSGA3를 적용한 특징 선택을 통해 총 21개의 변수 그룹이 추출되었다. 추출된 변수들은 구체적으로 객실수(rms), 직원이용객실수(hu), 총가용객실수(tot_use), 판매수(sold), 성인수(adlt), 국내아동(chld_local), 외국인성인(adlt_foreign), 객실가격(rrate), 부가서비스가격(srater), 개인예약_객실가격(rrate_f), 개인예약_부가서비스가격(srater_f), 개인예약_총금액(trate_f), 숙박수(stay), 투숙일자(arr), 퇴실일자(dep), 예약 없는 고객수(wki), 취소수(cxl), 2등급호텔여부(class_2), 경기지역여부(area_gyeonggi), 서울지역여부(area_seoul), 제주지역여부(area_jeju)이다. 이들은 객실 수, 인원 수, 판매가, 투숙-퇴실일 등 객실 수요에 직접적인 영향을 미치는 변수들로 확인되었다.

한편, 제안 모델의 병렬 처리 능력을 검증하기 위해 측정된 처리 시간은 표 4에 정리되어 있다. 일반적으로 머신러닝 계열 모델이 딥러닝 계열 모델보다 빠른 처리 시간을 보이지만, 본 연구에서는 LSTM 모델이 우수한 처리 성능을 보였다. 특히, 선형 회귀(LR) 모델의 경우, 특징 선택 적용 이후 82.92%의 처리 속도 개선이 나타났으며, 모든 비교 모델에서 처리 속도 향상이 확인되었다. 이는 특징 선택 기법이 처리 성능을 향상시키는 전형적인 사례라 할 수 있다.

Table 4. Models' Processing Time

Model	Processing Time		
	Without FS	With FS	Improvement(%)
TCN	7.16s	6.97s	2.65
LSTM	11.57s	10.81s	6.57
LR	15.40s	2.63s	82.92
XGB	32.97s	22.15s	32.82
LGBM	41.42s	28.81s	30.44

각 모델의 성능, 즉 예측 정확도는 표 5에 제시되어 있다. 이 표를 통해 알 수 있듯이, TCN이 다른 기계학습 모델에 비해 괄목할 만한 성과를 보였으며, 특히 FS가 추가된 제안 모델인 NSGA3-TCN의 경우 가장 뛰어난 예측 정확도를 보였음을 확인할 수 있었다.

비교 모델의 경우, 이전 연구들에서 확인된 예측 성능 지표와 유사한 값이 나타났다. 특정 지역과 특정 호텔에 대한 수요 예측은 여러 연구에서 높은 예측 성능을 보였으나, 다수의 호텔을 대상으로 한 통합 및 일반화 모델에서는 높은 예측 성능을 찾기 어려웠다. 본 연구에서도 비교 모델들은 기존 연구와 유사한 성능을 보인 것이다.

Table 5. Models' Accuracy

Metric	Model	Base Model	NSGA3+ Base Model
MAPE	TCN	10.47	9.73
	LGBM	21.88	21.58
	XGB	22.69	22.03
	LR	27.95	31.10
	LSTM	27.89	27.90
MSE	TCN	9.83	9.51
	LGBM	18.20	17.97
	XGB	18.97	18.30
	LR	26.12	27.90
	LSTM	21.11	21.16
RMSE	TCN	96.71	90.43
	LGBM	331.14	322.94
	XGB	359.87	335.01
	LR	682.18	778.56
	LSTM	445.74	447.73

그러나 TCN 모델은 호텔 수요 예측에서 MAPE 10.47%의 높은 예측 성능을 보였으며, 특히 특징 선택 (FS) 모델과 융합한 제안 모델은 미적용 모델 대비 7.05%의 추가적인 성능 향상을 보이면서 MAPE 9.73%의 최고 성능을 기록하였다. 또한, TCN의 예측 성능은 모델 학습 횟수(epoch)와 단위 처리 용량(batch size)에 많은 영향을 받는 것으로 확인되었다. 표 6에서는 하이퍼파라미터 값에 따라 예측 성능이 8%에서 최대 2배까지 변화함을 확인할 수 있었으며, 이는 최적의 하이퍼파라미터 설정이 예측 성능에 중요한 영향을 미친다는 것을 보여준다.

Table 6. Performance according to hyperparameter

TCN	epoch	batch size	MAPE(%)
1	40	128	10.51
2	40	256	10.03
3	40	512	20.58
4	80	128	10.42
5	80	256	9.73
6	80	512	10.13
7	120	128	9.95
8	120	256	9.88
9	120	512	10.21

또한, NSGA3를 통해 추출된 변수 그룹을 대상으로 변수 변경 테스트를 진행하였다. 특징 선택 기법을 통해 추출된 변수 그룹 중 일부 유사한 변수(개인예약_객실가격 (rrate_f), 개인예약_부가서비스가격(srata_f), 개인예약_총금액(trate_f))을 금액관련 변수들이 중복되는 것으로 보인다는 전문가 의견에 따라 임의로 제거하고 예측 성능을 측정된 결과, 특징 선택 기법으로 추출된 변수 그룹을 적용한 모델에 비해 낮은 예측 성능을 보임을 확인하였다. 이는 알고리즘에 의해 추출된 변수가 전문가 지식보다 예

측 성능에 더 긍정적인 영향을 미친다는 것을 의미한다.

종합적으로, 특징 선택(FS) 기법을 통한 변수 추출과 TCN 모델 적용을 통해 호텔 수요 예측에서 높은 예측 성능이 확인되었다. 본 연구에서 제안한 유전자 알고리즘을 이용한 특징 선택 기법과 TCN 융합 모델은 다수 호텔에 대한 데이터 통합 및 일반화 모델 구축을 위한 좋은 사례가 될 수 있음을 보여준다.

V. Conclusions

본 연구에서는 호텔 데이터 분석을 통한 수요 예측을 위해 NSGA3-TCN 모델을 제안하였으며, 이를 통해 호텔 수요 예측 성능을 획기적으로 향상시킬 수 있음을 확인하였다. NSGA3-TCN 모델은 1D-Convolution을 적용한 병렬 처리 기법을 통해 빠른 처리 속도를 제공하였고, TCN의 인과성 특징을 활용하여 과거 데이터만으로도 예측 성능을 높이는 동시에 그래디언트 소실 문제를 해결하는 데 기여하였다. 본 연구의 결과는 국내 호텔 운영 솔루션에서 수집된 실제 데이터를 기반으로 한 실증 분석을 통해 검증되었으며, 모델의 예측 정확도와 실용성이 매우 높음을 확인하였다. 변수 그룹에 대한 특징 선택 기법을 적용한 결과, 호텔 수요 예측에 사용된 TCN 모델은 MAPE 9.73%로 상당히 우수한 예측 성능을 보였다.

또한, 본 연구에서 사용된 데이터가 다수 호텔에서 확보된 실제 운영 데이터임을 감안할 때, 호텔 수요 예측의 일반화 가능성을 높였으며, 이는 다수 호텔 및 관련 기업들에게 실질적인 혜택을 제공할 수 있음을 의미한다. 이러한 본 연구의 성과는 호텔 및 여행 산업에서 적합한 모델 적용을 통해 정확한 수요 예측 활동에 기여할 수 있다는 점에서 의미가 있다. 이는 호텔 및 여행 관련 사업자의 효율성 향상과 수익 증대에 중요한 의사결정 도구로 작용할 수 있다.

하지만 본 연구는 다음과 같은 점에서 한계를 갖는다. 첫째, 본 연구는 NSGA3의 단일 기법으로만 특징 선택이 이루어져 특징 선택 기법들의 차이가 성과 개선에 어떤 영향을 미치는지 확인하지 못했다는 한계가 있다. 둘째, 사용된 데이터가 제한적이라는 한계도 있다. 현재 사용된 데이터는 코로나 팬데믹의 영향을 최소화하기 위해 2019년 이전 데이터를 사용하고 있어서, 최신의 경향을 반영하지 못했다는 한계가 있다. 또한, 운영 시스템에서 추출한 정형 데이터 외에 SNS 콘텐츠, 고객 리뷰와 같은 비정형 데이터의 활용은 이루어지지 않았다는 점에서도 한계를 가진다. 셋째, 다양한 하이퍼파라미터 설정에 따른 성능 비

교가 이루어지지 못했으며, 실험 시 탐색된 하이퍼파라미터가 다른 데이터셋에서도 유효한지에 대한 오버피팅 등의 문제도 충분히 검증되지 못했다는 한계가 있다.

따라서 향후 수행할 후속 연구에서는 우선 지역, 규모 등이 고려된 다양한 최신의 데이터셋을 통한 연구가 필요하다. 또한, 정치, 경제, 날씨 등의 외부 요인의 반영과 비정형 데이터가 예측 성능에 미치는 영향을 검증하는 것이 필요하며, 지속적으로 연구되고 있는 다양한 개선 및 융합 모델 적용을 통해 모델의 일반화와 성능 향상을 도모할 필요가 있다. 추가적으로 운영시스템에서 수집된 실운영 데이터 외에도 호텔 경영상황을 반영하기 위한, 직원수, 매출, 수익, 경영방침 등의 정형, 비정형 정보를 적절히 벡터화하여 수요 예측을 위한 변수그룹에 반영하는 시도도 필요할 것이다. 이러한 연구의 고도화를 통해 호텔 및 여행 산업에서 수요 예측의 정확성과 실용성을 확장하고, 해당 산업과 다른 산업 분야에서 생산성을 높이기 위한 의사 결정을 효과적으로 지원할 수 있을 것으로 기대한다.

REFERENCES

- [1] K. Park, G. Jung, and H. Ahn, "A Time Series Forecasting Model with the Option to Choose between Global and Clustered Local Models for Hotel Demand Forecasting," *The Korean Journal of BigData*, vol. 9, no. 1, pp. 31-47, Jun. 2024. DOI: 10.36498/kbigdt.2024.9.1.31
- [2] S. Bai, J. Z. Kolter and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," *arXiv*, Mar. 2018. DOI: 10.48550/arXiv.1803.01271
- [3] J. A. Rusman, K. Chunady, S. T. Makmud, K. E. Setiawan, and M. F. Hasani, "Crude Oil Price Forecasting: A Comparative Analysis of ARIMA, GRU, and LSTM Models", *IEEE 9th International Conference on Computing, Engineering, and Design (ICCED)*, 2023. DOI: 10.1109/ICCED60214.2023.10425576
- [4] N. Dowlut, B. Gobin-Rahimbux, "Forecasting resort hotel tourism demand using deep learning techniques - A systematic literature review", *Heliyon*, vol. 9, no. 7, e18385, 2023. DOI: 10.1016/j.heliyon.2023.e18385
- [5] J. Kim, Y. Yang, M. Oh, S. Lee, S. Kwon, and W. Cho, "Demand Prediction of Furniture Component Order Using Deep Learning Techniques", *The Korean Journal of BigData*, vol. 5, no. 2, pp. 111-120, Dec. 2020. DOI: 10.36498/kbigdt.2020.5.2.111
- [6] H. Ahn, "Optimization of Multiclass Support Vector Machine Using Genetic Algorithm: Application to the Prediction of Corporate Credit Rating," *Journal of MIS Research*, vol. 16, no. 3, pp. 161-177, Dec. 2014. DOI: 10.14329/isr.2014.16.3.161
- [7] Y. Pethe and H. Das, "Feature Selection Using Genetic Algorithm for Software Fault Prediction," *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAC)*, pp. 1132-1137, Jun. 2024. DOI: 10.1109/ICAAC60222.2024.10575523
- [8] A. Alghamdi, "A Hybrid Method for Big Data Analysis Using Fuzzy Clustering, Feature Selection and Adaptive Neuro-Fuzzy Inference System Techniques: Case of Mecca and Medina Hotels in Saudi Arabia," *Arabian Journal for Science & Engineering*, vol. 48, no. 2, pp. 1693-1714, Feb. 2023. DOI: 10.1007/s13369-022-06978-0
- [9] X. Li, C. Liu and Y. He, "Efficient Time Series Predicting with Feature Selection and Temporal Convolutional Network," *2021 IEEE 4th International Conference on Computer and Communication Engineering Technology (CCET)*, pp. 141-145, Aug. 2021. DOI: 10.1109/CCET52649.2021.9544317
- [10] W. Zha and Y. Ye, "An Aero-Engine Remaining Useful Life Prediction Model Based on Feature Selection and the Improved TCN," *Franklin Open*, vol. 6, March 2024. DOI: 10.56094/fo.2024.1004
- [11] P. Lara-Benítez, M. Carranza-García, J. M. Luna-Romera and J. C. Riquelme, "Temporal Convolutional Networks Applied to Energy-Related Time Series Forecasting," *Applied Sciences*, vol. 10, no. 7, 2322, 2020. DOI: 10.3390/app10072322
- [12] R. Wan, S. Mei, J. Wang, M. Liu and F. Yang, "Multivariate Temporal Convolutional Network: A Deep Neural Networks Approach for Multivariate Time Series Forecasting," *Electronics*, vol. 8, no. 8, 876, Aug. 2019. DOI: 10.3390/electronics8080876
- [13] G. Selva Jeba and P. Chitra, "River Flood Prediction through Flow Level Modeling Using Multi-Attention Encoder-Decoder-Based TCN with Filter-Wrapper Feature Selection," *Earth Science Informatics*, pp. 1-17, 2023. DOI: 10.1007/s12145-024-01446-9
- [14] M. Liu, X. Sun and Q. Wang, "Short-Term Load Forecasting Using EMD with Feature Selection and TCN-Based Deep Learning Model," *Energies*, vol. 15, no. 19, 7170, Oct. 2022. DOI: 10.3390/en15197170
- [15] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Computation*, vol. 1, no. 4, pp. 541-551, Dec. 1989. DOI: 10.1162/neco.1989.1.4.541
- [16] N. Kalchbrenner, L. Espeholt, K. Simonyan, A. van den Oord, A. Graves and K. Kavukcuoglu, "Neural Machine Translation in Linear Time," *arXiv*, Oct. 2016. *arXiv preprint arXiv:1610.10099*
- [17] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735
- [18] J. Chung, C. Gulcehre, K. Cho and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence

- Modeling," arXiv, Dec. 2014. DOI: 10.48550/arXiv.1412.3555
- [19] J. Long, E. Shelhamer and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431-3440, June 2015. DOI: 10.1109/CVPR.2015.7298965
- [20] R. Urraca, A. Sanz-Garcia, J. Fernandez-Ceniceros and F. J. Martinez-De-Pison, "Improving Hotel Room Demand Forecasting with a Hybrid GA-SVR Methodology Based on Skewed Data Transformation, Feature Selection and Parsimony Tuning," Lecture Notes in Computer Science, vol. 9121, Dec. 2015. DOI: 10.1007/978-3-319-19644-2_52
- [21] A. M. Usman, U. K. Yusof and S. Naim, "Filter-Based Multi-Objective Feature Selection Using NSGA III and Cuckoo Optimization Algorithm," IEEE Access, vol. 8, pp. 76333-76356, 2020. DOI: 10.1109/ACCESS.2020.2987057
- [22] Y. Xue, Y. Tang, X. Xu, J. Liang and F. Neri, "Multi-Objective Feature Selection with Missing Data in Classification," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 6, no. 2, pp. 355-364, Apr. 2022. DOI: 10.1109/TETCI.2021.3074147
- [23] A. Papisani, R. Durgam and N. Devarakonda, "Adaptive Neighborhood Adjustment Strategy Based on MOHHO and NSGA-III Algorithms for Feature Selection," IAENG International Journal of Applied Mathematics, vol. 54, no. 5, pp. 917-935, May 2024.
- [24] M. S. Almutairi, K. Almutairi and H. Chiroma, "Selecting Features That Influence Vehicle Collisions in the Internet of Vehicles Based on a Multi-Objective Hybrid Bi-Directional NSGA-III," Applied Sciences, vol. 13, no. 4, 2064, Feb. 2023. DOI: 10.3390/app13042064
- [25] Md Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. S. Awwal and V. K. Asari, "A State-of-the-Art Survey on Deep Learning Theory and Architectures," Electronics, vol. 8, no. 3, 292, Mar. 2019. DOI: 10.3390/electronics8030292
- [26] F. Karim, S. Majumdar, H. Darabi and S. Harford, "Multivariate LSTM-FCNs for Time Series Classification," Neural Networks, vol. 116, Aug. 2019, pp. 237-245. DOI: 10.1016/j.neunet.2019.04.014
- [27] A. Ayodeji, Z. Wang, W. Wang, W. Qin, C. Yang, S. Xu and X. Liu, "Causal Augmented ConvNet: A Temporal Memory Dilated Convolution Model for Long-Sequence Time Series Prediction," ISA Transactions, vol. 123, pp. 200-217, April 2022. DOI: 10.1016/j.isatra.2021.05.026

Authors



Keehyun Park received his Master in Business Administration from Yonsei University, Korea in 2018. He is currently enrolled in the Doctoral's program at the Graduate School of Business IT at Kookmin

University. His main research areas are Process Intelligence, AI-powered CX and Revenue Management by ML forecasting.



Gyeongho Jung received his B.S. in Smart Automotive Engineering from Halla University, Korea in 2023. He is currently enrolled in the master's program at the Graduate School of Business IT at Kookmin

University. His main research areas of interest are business analytics, deep learning and machine learning.



Hyunchul Ahn received a BS in Industrial Management from KAIST, and a ME and PhD from KAIST Graduate School of Management. He is currently working as a professor of the Graduate School of Business

IT at Kookmin University. His main research area is AI applications in finance and marketing.