Smart Media Journal / Vol.14, No.6 / ISSN:2287-1322
https://dx.doi.org/10.30693/SMJ.2025.14.6.14

# 범주형 변수와 연속형 변수가 혼합된 시계열 데이터 예측 기법

## (Time Series Forecasting Method for Mixed Categorical and Continuous Variables)

**이재진*, 이주홍****

(Jaejin Lee, Ju Hong Lee)

**요 약**

기존 시계열 예측 연구는 범주형 변수의 사용을 고려하지 않아 중요한 패턴이나 정보를 활용하지 못하였다. 결과적으로 이러한 연구는 편향성과 과적합 문제로 고전하였다. 본 논문에서는 다음과 같이 두 가지 기법을 통합적으로 활용하여 시계열 예측 성능을 개선하는 방법을 제안한다. 첫째, Feature Tokenizer와 Dozer Attention를 활용해 범주형 변수와 연속형 변수를 통합적으로 처리한다. 둘째, Variational Auto-Encoder(VAE) 기반 데이터 증강 기법으로 시계열 패턴의 편향성과 과적합 문제를 완화한다. 실험 결과, 제안한 방법이 기존 모델보다 더 우수한 예측 성능을 보인다. 제안한 방법은 시계열 예측에서 범주형 변수의 중요성을 반영하고, VAE 기반으로 증강된 데이터를 사용함으로써 예측 성능을 개선하였다.

■ 중심어 : 시계열 예측 ; 딥러닝 ; 트랜스포머 ; 시계열 증강

**Abstract**

Existing time series forecasting studies have overlooked the use of categorical variables, failing to leverage important patterns and information. As a result, they suffer from pattern bias and overfitting issues. This paper proposes a novel approach to improve time series forecasting performance by integrating the following two techniques. First, the Feature Tokenizer and Dozer Attention are utilized to comprehensively process both categorical and continuous variables. Second, a Variational Auto-Encoder (VAE)-based data augmentation technique is employed to mitigate pattern bias and overfitting problems. Experimental results demonstrate that the proposed method outperforms existing models in predictive performance. By incorporating the importance of categorical variables in time series forecasting and leveraging VAE-augmented data, the proposed method achieves significant improvements in predictive accuracy.

■ keywords : Time Series Forecasting ; Deep Learning ; Transformer ; Time Series Augmentation

## Ⅰ. 서 론

시계열 데이터는 일정한 시간 주기로 획득한 데이터를 의미한다. 이러한 시계열 데이터는 금융, 의료, 날씨 등 다양한 분야에서 생성되고 있으며 이를 정확히 예측하는 것은 각 분야에서 적절한 자원 관리를 위해 필요한 작업이다.

시계열 데이터를 예측하는 기법은 과거 데이터의 기록이 유용한 정보나 패턴을 포함하고 있어 미래를 예측함에 도움이 된다는 생각에 근거하고 있다[1]. 금융 분야의 경우 주가나 환율 등의 시계열 데이터를 예측하기 위해 알고리즘 트레이딩에 대한 관심이 높아지고 있다. 알고리즘 트레이딩에 사용하는 기법들은 데이터의 과거 기록을 바탕으로 각자의 분석을 통한 매매 전략을 추구하게 되며 딥러닝을 통해 가격 흐름의 패턴

* 준회원, 인하대학교 전기컴퓨터공학과
** 정회원, 인하대학교 컴퓨터공학과

접수일자 : 2024년 12월 09일
수정일자 : 1차 2025년 02월 17일, 2차 2025년 03월 19일
게재확정일 : 2025년 03월 31일
교신저자 : 이주홍 e-mail : juhong@inha.ac.kr

을 찾아내려는 연구가 계속 진행되고 있다. 의료 분야에서는 환자의 병력과 같은 과거 기록을 분석하여 질병의 발생 가능성이나 건강 상태의 변화를 예측하고자 하며, 환경을 담당하는 기관에서 온도, 강수량, 바람의 방향 등을 이용하여 기후의 변화를 예측하는 것 역시 시계열 예측을 수행하는 유명한 사례이다.

딥러닝 기술의 발전으로 합성곱 신경망(Convolutional Neural Networks, CNN)과 순환 신경망(Recurrent Neural Networks, RNN), 트랜스포머(Transformer) 등 여러 구조의 딥러닝 모델이 등장했으며, 시계열 예측에도 다양한 구조의 모델들을 적용한 방법이 제안되었다[1-5]. 그중에서도 트랜스포머[6]는 자연어 처리(Natural Language Processing, NLP) 작업을 위해 등장한 모델로서 현재는 영상처리나 시계열 예측 등의 분야에서도 큰 성과를 거두고 있다.

그러나 시계열 예측 연구에서는 범주형 변수를 다루는 방식에 한계가 존재한다. 범주형 변수를 단순히 원-핫 인코딩이나 라벨 인코딩으로 변환하면, 값 간에 순서나 크기의 개념이 없는 변수는 시간 흐름에 따른 변화를 모델링하기 어렵기 때문이다[7]. 범주형 변수란 해당 변수의 값이 일정한 범주 혹은 그룹으로 나뉘는 변수를 의미한다. 범주형 변수의 값은 특정 그룹 또는 클래스에 속하게 되며, 텍스트 혹은 수학적 연산이 불가능한 숫자로 표현된다. 그러므로 시계열 예측에는 일반적으로 연속형 변수를 사용한다. 연속형 변수는 값의 범위가 연속적이며 수학적인 연산이 가능한 숫자로 표현되는 변수를 말한다. 시계열 예측에 범주형 변수를 고려하지 않는다면 해당 변수가 지닌 중요한 패턴이나 정보가 손실되어 예측 성능이 저하될 수 있다. 따라서 범주형 변수와 연속형 변수가 혼합된 데이터를 활용할 때, 범주형 변수를 효과적으로 모델링 하여 예측 성능을 향상시키는 방법이 필요하다. 본 연구에서는 이를 위해 FT-Transformer[8]에서 사용된 Feature Tokenizer 기법과 트랜스포머 기반의 최신 시계열 예측 모델인 Dozerformer[5]의 Dozer Attention 구조를 사용하였다.

한편, 일부 시계열 데이터는 도메인에 따라 각기 다른 패턴과 분포를 보이거나, 동일한 데이터라도 시간이 지남에 따라 추세와 분산 등 통계적 특성이 변하는 비정상성(Non-stationary)이 나타날 수 있다. 이로 인해 동일한 도메인 내에서 유사한 패턴과 분포를 지닌 학습 데이터를 충분히 확보하기가 어려운 실정이다. 이러한 문제를 해결하기 위해 본 연구에서는 VAE(Variational Auto-Encoder)[9]를 활용하여 데이터의 과거 불확실성을 복원하고, 현실에서 발생 가능한 다양한 시나리오의 시계열 데이터를 생성해 학습에 활용하였다.

본 논문의 구성은 2장에서 관련 연구를 소개하고, 3장에서는 제안 방법에 대하여 설명하며, 4장에서는 실험 및 결과를 나타낸다. 그리고 5장에서는 결론과 시사점을 살펴본다.

## Ⅱ. 관련 연구

### 1. 범주형 변수와 시계열 예측의 관계

범주형 변수는 시계열 데이터의 수집 주기에 따라 데이터가 확보되었을 때, 해당 범주형 변수의 값 중 어디에 속해있는지 기록한 것이다. [10]에서는 성별과 나이에 따라 그룹을 나누어 도로 교통 사망자 수의 추이를 분석하였고 [11]에서는 전력 부하 예측에 평일, 공휴일, 명절 등 범주형 변수를 원-핫 벡터로 구성된 더미 변수로 활용하여 요일별 부하 패턴의 중요성을 입증하였다. 이러한 연구들은 범주형 변수를 효과적으로 반영하지 않으면 시계열 예측에 중요한 정보를 놓칠 수 있음을 보여준다.

### 2. 시계열 예측 연구

시계열 예측 기법은 ARIMA, SARIMA와 같은

통계적 기법, Linear Regression과 같은 기계 학습 기법 그리고 RNN이나 Transformer 등을 포함한 딥러닝 기법 세 가지로 분류할 수 있다[2].

최근 시계열 예측에는 딥러닝, 그중에서도 트랜스포머 구조가 자주 사용되며 [3-5]와 같은 모델들이 등장하였다. 이러한 모델들은 각 키가 일부의 쿼리만 상호작용하도록 제안하는 Sparse Attention 메커니즘을 도입하였다.

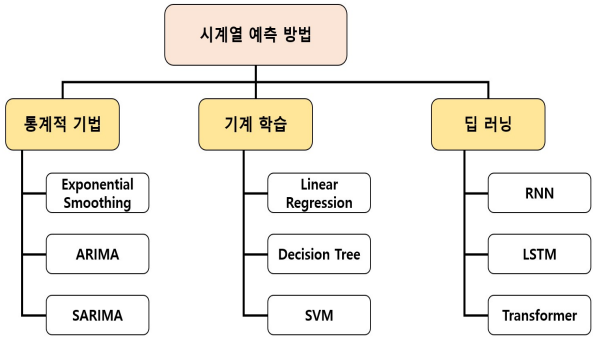


그림 1. 시계열 예측 방법 분류

그중에서 Dozerformer[5]는 트랜스포머 기반의 시계열 예측 모델 중 가장 최신 모델로서 시계열 예측에서 가장 우수한 성능을 입증한 모델이다. 기존 트랜스포머의 Full Attention과 달리 Dozerformer의 Dozer Attention은 Local, Stride, Vary로 구성된다. Local Attention은 예측 시점 주변의 짧은 시간 범위에서 중요한 정보에만 집중하며, Stride Attention은 일정 간격으로 떨어진 시점들을 참조하여 주기적 또는 간헐적 패턴을 학습한다. 디코더에서만 사용되는 Vary Attention은 예측 시점이 멀어질수록 더 긴 과거 데이터를 참조하여 장기적 패턴을 학습한다.

그림 2는 일반적인 트랜스포머 아키텍처의 Full Attention과 Dozerformer에서 사용되는 Dozer Attention (Local, Stride, Vary)이 예측을 위해 어떤 과거 데이터를 선택하는지 나타낸 그림이다. I는 학습에 사용되는 입력 시퀀스 길이를 나타내며, O는 예측하고자 하는 미래시점 t부터의 예측 길이를 나타낸다. Dozer Attention은 선택적으로 과거 데이터를 참조하여 계산량을 줄이고 중요한 패턴을 효과적으로 학습할 수 있도록 설계되었다. 하지만 Dozerformer와 같은

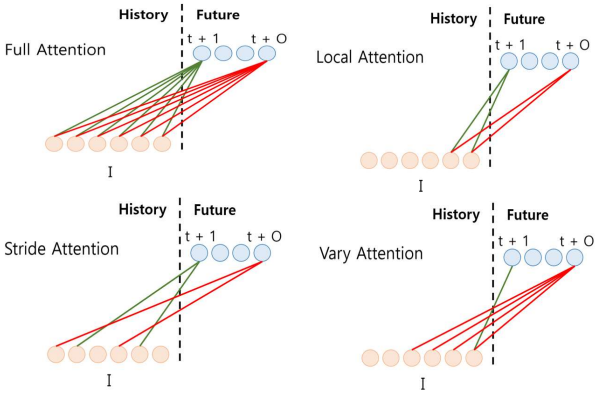


그림 2. 트랜스포머 아키텍처의 Full Attention과 Dozer Attention (Local, Stride, Vary)

최신 트랜스포머 기반 시계열 예측 모델은 Attention 구조 설계에만 집중하고 있어 범주형 변수 처리와 데이터 편향성 극복 측면에서는 한계를 가진다.

## 3. VAE(Variational Auto-Encoder)

VAE는 인코더와 디코더로 구성된 딥러닝 기반의 생성 모델이다. 인코더에서는 입력 데이터를 잠재 공간(latent space)으로 매핑하고 잠재 공간을 확률 분포로 모델링하여 데이터의 특성이 함축된 잠재 변수(latent variables)를 생성한다. 인코더의 출력 값은 잠재 변수의 평균과 분산이며 이를 통해 확률 분포를 모델링 할 수 있게 된다. 그리고 디코더는 잠재 변수들을 복원하여 새로운 데이터를 생성하는 역할을 한다. [12]에서는 시계열 데이터를 분해했을 때 구해지는 세 개의 구성 성분(추세, 계절성, 잔차)의 확률 분포를 각각 모델링하고 샘플링하여 잠재 변수들을 획득하는 방식을 사용하였다. 그리고 추세와 계절성을 각각 복원하는 디코더와 추세, 계절성, 잔차 잠재 변수의 곱으로 구해지는 예측 잠재 변수의 값을 복원하는 디코더로 구성된다.

해당 방법을 이용하여 시계열 데이터를 증강시키는 것의 장점은 첫째로 안정적인 학습이 가능하다는 점이다. GAN 기반의 시계열 증강 모델은 훈련 과정에서 모드 붕괴(mode collapse)나 수렴 문제가 발생하여 비정상적인 데이터를 생

성할 수 있다. 두 번째로는 잠재 공간 기반의 데이터 증강을 사용한다는 점이다. 추세와 분산, 잔차의 잠재 공간을 명시적으로 샘플링하여 생성 데이터에 대한 설명 가능성을 높이는 것으로 데이터의 구조적 이해를 도울 수 있으며, 세 개의 인코더와 디코더를 사용함으로써 불확실성이 보다 현실적으로 반영된 시계열 데이터의 시나리오를 생성할 수 있게 된다.

### 4. FT-Transformer

FT-Transformer는 범주형 변수와 연속형 변수가 함께 존재하는 표 형태의 데이터를 다루기 위해 제안된 모델이다. 범주형 변수와 연속형 변수가 혼합된 데이터셋을 Feature tokenizer에 입력하고 모든 변수를 개인이 설정한 d-크기의 차원으로 임베딩하여 변수들이 동일한 공간에서 표현되도록 한다.

Feature tokenizer가 변수를 임베딩하는 과정은 아래의 수식과 같다.

$$T_j^{(num)} = b_j^{(num)} + x_j^{(num)} \cdot \mathrm{W}_{\mathrm{j}}^{(\mathrm{num})} \in \mathrm{R}^{\mathrm{d}}, \quad (1)$$

$$T_j^{(cat)} = b_j^{(cat)} + e_j^T W_j^{(cat)} \in \mathrm{R}^{\mathrm{d}}, \quad (2)$$

$$T = stack[T_1^{num}, \cdots, T_{k^{(num)}}^{(num)}, T_1^{(cat)}, \cdots, T_{k^{(cat)}}^{(cat)}] \in \mathrm{R}^{\mathrm{k}\times\mathrm{d}}. \quad (3)$$

여기서 k는 전체 변수의 개수를 나타내고 j는 변수의 인덱스를 의미한다. $b_j$는 j번째 변수의 bias이며 $e_j^T$는 j번째 변수에 대한 원-핫 벡터를 나타낸다. 최종적으로 사전에 설정한 d-차원으로 매핑된 변수들을 이어붙여 임베딩을 완료한다.

## Ⅲ. 제안 방법

본 논문에서 제안하는 방법은 크게 세 단계로 구성된다. 첫째, 원본 데이터를 기반으로 VAE를 활용해 시계열 데이터를 증강한다. 증강 데이터는 시계열의 추세, 계절성, 잔차를 모델링하여 현실의 다양한 시나리오가 반영되며, 예측 모델이 일반화된 패턴을 학습할 수 있도록 도와주는 역할을 한다. 둘째, 증강 데이터와 원본 데이터를 이용하여 모델을 학습시킨다. 셋째, 학습된 모델을 기반으로 원본 데이터의 시계열 예측을 수행한다. 앞서 말한 세 가지 단계의 전체 프레임워크는 그림 3과 같다.

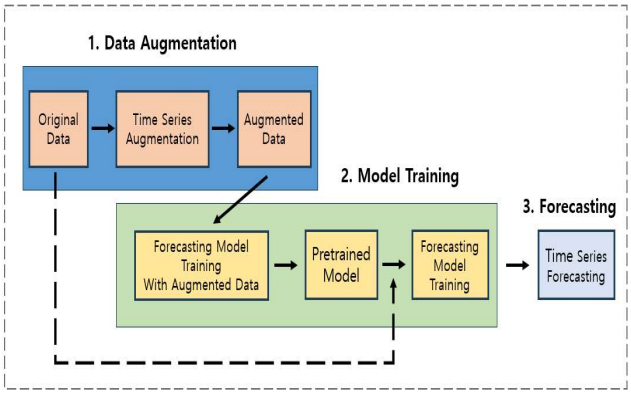


그림 3. 제안된 전체 프레임워크

### 1. VAE를 이용한 시계열 데이터 증강

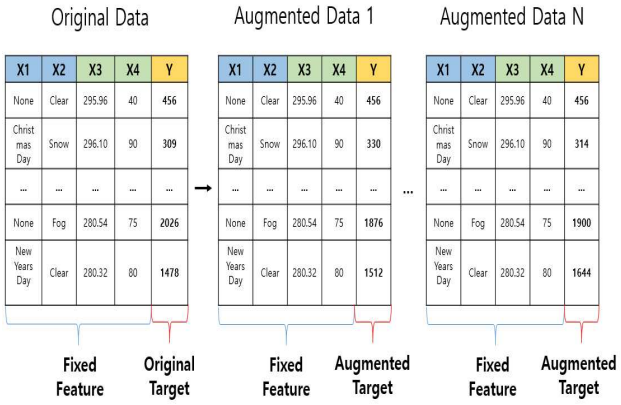


그림 4. VAE로 증강된 데이터셋. 종속 변수를 제외한 나머지 변수들은 고정된 값을 사용한다.

원본 데이터만 학습에 사용한다는 것은 시계열의 패턴을 파악할 때, 원본 데이터의 시나리오를 넘어서는 패턴의 정보를 반영할 수가 없음을 의미한다. 따라서 학습에 사용된 패턴의 편향성, 그리고 과적합으로 인하여 일반화 성능이 감소할 수 있다. 이를 방지하고자 VAE를 활용하여 시계열 데이터를 증강시켜 예측하고자 하는 시계열의 다양한 변동성을 학습하고자 한다.

실험에 사용할 증강 데이터는 그림 4와 같이 원본 데이터에서 종속 변수(Y)를 제외한 독립 변수(X)를 고정하여 구성한다(Fixed Feature). 종속 변수(Target)에만 변화를 주어 새로운 데이터셋(Augmented Data)을 생성함으로써, 동일한 독립 변수 조건에서 다양한 시나리오를 반영한 시계열 데이터를 확보한다. 원본 데이터만으로는 동일한 독립 변수 조건에서 나타나는 종속 변수의 변화와 상호작용을 충분히 학습하기 어렵다. 그러므로 증강 데이터를 활용함으로써 변수 간의 잠재적 상호작용을 보다 효과적으로 학습할 수 있는 환경을 제공하는 것이 목적이다.

### 2. 시계열 예측 과정

범주형 변수는 연속형 변수와 달리 값 자체가 특정한 의미를 지니기보다는 구분을 위해 할당된 경우가 많다. 따라서 변수의 특성을 적절히 반영하여 모델이 효과적으로 학습할 수 있도록 변환하는 것이 중요하다. 본 연구에서는 FT-Transformer의 Feature Tokenizer를 사용하여 범주형 변수를 동일한 d-차원 공간에 임베딩하는 방식을 적용하였다. 그리고 Dozer Attention 구조를 가진 시계열 예측 모델을 함께 사용하여 범주형 변수와 연속형 변수를 통합적으로 처리하여 시계열 예측을 수행하는 방법을 제안한다. 해당 방법은 기존 Dozerformer가 범주형 변수를 처리하지 못한다는 한계를 극복할 수 있다.

그림 5는 시계열 데이터가 예측에 사용되는 과정을 나타낸다. 시계열 예측과정은 먼저 시계열 데이터에서 범주형 변수를 제거(split)하는 것에서 시작한다. 그리고 연속형 변수만 남은 데이터는 시계열 분해 블록을 통과해 추세(x_trend)와 계절성(x_seasonal) 성분으로 분해된다. 여기서 추세 성분은 시계열 데이터의 이동 평균으로 구해지며 장기적인 변화 패턴을 나타낸다. 계절성은 주기적인 변동 패턴을 나타내며 추세의 완만한 곡선으로부터 각 데이터가 추세로부터 얼마나 벗어나는지를 나타낸다. 시계열 데이터의 분해를 진행하기 전, 데이터의 가장 앞과 뒤에 제로 패딩(Zero Padding)을 수행하여 데이터의 시계열 길이를 그대로 유지한다. 그로 인해 떼어낸 범주형 변수를 그대로 다시 붙일 수 있으며(Concatenate), 범주형 변수와 계절성으로 구성된 데이터를 Feature Tokenizer에 입력하여 임의의 차원으로 매핑된 임베딩을 얻는다. 이것을 시계열 예측 모델인 Dozerformer에 넣어 특정 예측 시점 t에 대한 계절성 예측값을 획득한다. 시계열 분해로 구한 추세 성분은 단순 Linear 모델에 입력하여 예측하며, 앞에서 획득한 계절성과 더하는 것으로 최종 예측값(Y_pred)을 구한다.

## Ⅳ. 실험 방법 및 결과

### 1. 실험 데이터

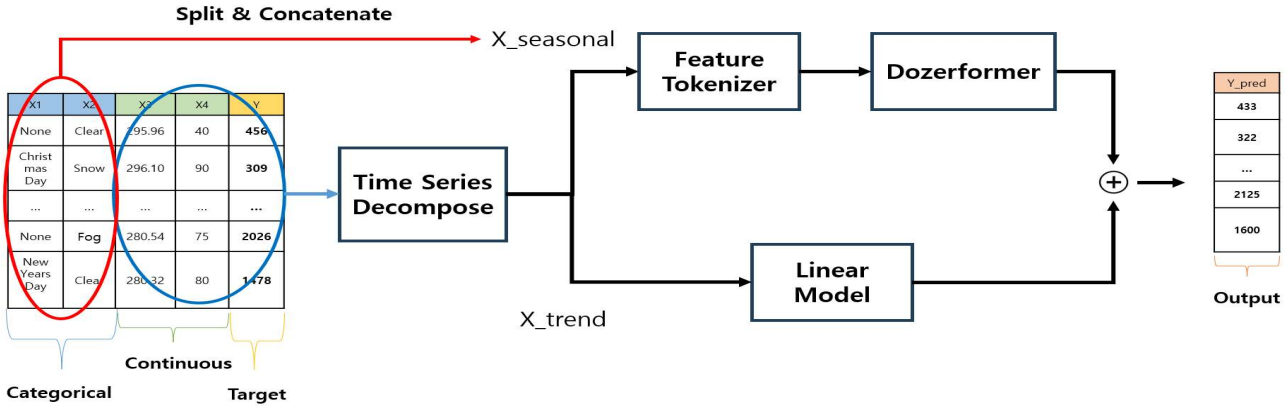


그림 5. 시계열 예측 프레임워크

실험에 사용하는 데이터는 UCI Machine Learning Repository에서 구한 Traffic volume 및 Air pollution 데이터셋을 사용한다. Traffic volume 데이터는 미국 미니애폴리스 지역 관측소에서 측정된 시간별 교통량 데이터셋이다. 2016년 7월 24일부터 2018년 8월 17일까지 1시간 간격으로 수집된 18000개의 타임 스텝과 공휴일과 날씨를 나타내는 2개의 범주형 변수 그리고 4개의 연속형 변수로 구성되어있다. Air pollution 데이터는 중국 베이징의 대기질 관측소 중 한 곳에서 측정된 시간 별 대기 오염 물질 데이터셋이다. 해당 데이터는 2015년 1월 7일부터 2017년 2월 3일까지 1시간 간격으로 수집된 17100개의 타임 스텝과 풍향을 나타내는 1개의 범주형 변수 그리고 11개의 연속형 변수로 구성되어있다.

실험 데이터는 7:1:2의 비율로 train, validation, test set을 나누었으며 예측 길이는 48, 96, 192로 설정하였다.

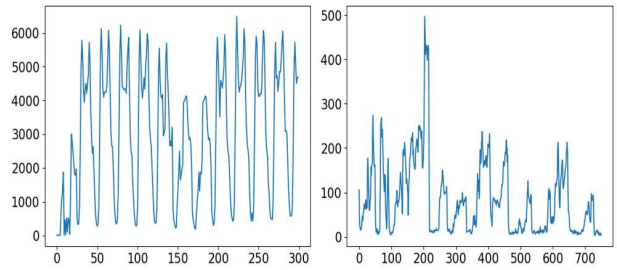


그림 6. 실험 데이터 (좌) Traffic volume (우) Air pollution

## 2. 실험 구성

본 논문에서는 범주형 변수가 시계열 예측 성능에 미치는 영향을 분석하기 위해 실험을 설계하였다. 최신 시계열 예측 모델인 Dozerformer를 비롯한 기존 모델은 범주형 변수를 직접 입력하지 못하는 한계를 가지고 있다. 이를 해결하고 성능 차이를 비교하기 위해, 먼저 범주형 변수를 제거한 상태에서 Informer, Autoformer, Dozerformer를 사용하여 예측 성능을 평가하였다. 이후, 범주형 변수가 포함된 데이터셋을 FT-Transformer에 입력하여 시계열 예측을 수행하였다. 마지막으로, 범주형 변수가 포함된 데이터셋에 Feature Tokenizer와 Dozer Attention 알고리즘을 함께 적용한 기법(FT-Dozerformer)을 사용하여 예측을 진행하고 각 방법의 성능을 비교 및 분석하였다. 해당 실험을 통해 범주형 변수를 활용함으로써 시계열 예측 성능에 미치는 영향을 검증하고자 한다.

이어서 종속 변수의 다양성이 시계열 예측에 미치는 영향을 확인하기 위해 VAE를 활용하여 시계열 데이터를 생성하고, 이를 바탕으로 사전 학습 모델을 구성하였다. 사전 학습 모델은 train set에서 학습하고 validation set에서의 성과를 기준으로 최적의 결과를 낸 모델을 선택한다. 이후, 원본 데이터를 학습하고 예측결과를 비교하는 방식으로 성능을 평가하였다(Our Method). 평가지표는 MAE(Mean Absolute Error)와 MSE(Mean Squared Error)를 사용하였으며, random seed를 1부터 5까지 고정하여 실험을 진행한 후, 획득한 결과의 평균값을 산출하였다. 각 모델에 사용된 하이퍼파라미터는 아래의 표 1에 제시된 범위 내에서 가장 좋은 성능을 나타낸 값을 선택하였다.

표 1. 하이퍼파라미터

| Hyper Parameter | Value |
|---|---|
| learning rate | {1e-5, 5e-4, 1e-4, 5e-3, 1e-3} |
| lradj | CosineAnnealing |
| batch size | {32, 64, 128} |
| drop out | {0.1, 0.2, 0.3} |
| epoch | 300 |
| loss | {L1, L2} |
| d (FT embedding size) | 8 |

## 3. 실험 결과

실험 결과는 표 2와 같다. 모델이 예측하고자 하는 time step 크기를 pred 48, 96, 192로 표현하였으며, 예측에 사용된 모델과 예측 길이 그리고 사용 데이터에 따른 예측결과를 나타낸다. Traffic Volume 데이터와 Air pollution 데이터 모두 제안된 모델이 더 좋은 성과를 보였다. Air

pollution 데이터셋의 MAE, MSE 값이 Traffic Volume 데이터셋보다 높은 이유는 해당 데이터에서 나타나는 데이터의 패턴이 상대적으로 더 불규칙적이기 때문이다. 실험 결과를 통해 시계열 예측에 있어서 범주형 변수의 중요성을 알아볼 수 있었으며 예측에 도움이 되는 적절한 범주형 변수가 있다면 제안한 예측 기법이 기존의 예측 기법보다 더 적합하다고 할 수 있다. 또한, 시계열 예측을 진행하기 전에 원본 데이터를 증강하여 미리 학습시키는 것이 가장 높은 예측결과를 보였다. 이는 비슷한 분포를 따라가는 다양한 패턴을 미리 학습 후 원본 데이터를 학습시키는 것이 과적합을 방지하고 종속 변수와 다른 변수 간 상호작용을 더 잘 학습시키는 것을 나타낸다.

표 2. 실험 결과

| | | Traffic Volume | | Air Pollution | |
|---|---|---|---|---|---|
| | | MAE | MSE | MAE | MSE |
| Pred 48 | Informer | 0.2539 | 0.1454 | 0.9478 | 1.8263 |
| | Autoformer | 0.2845 | 0.1667 | 1.0326 | 1.8696 |
| | FT-Transformer | 0.3312 | 0.1945 | 1.0889 | 1.9062 |
| | Dozerformer | 0.1773 | 0.0997 | 0.8626 | 1.4787 |
| | FT-Dozerformer | 0.1682 | 0.0961 | 0.8538 | 1.4329 |
| | Our Method | **0.1636** | **0.0930** | **0.8459** | **1.4226** |
| Pred 96 | Informer | 0.2927 | 0.1873 | 0.9709 | 1.8581 |
| | Autoformer | 0.3009 | 0.1794 | 1.0517 | 1.9130 |
| | FT-Transformer | 0.3575 | 0.2208 | 1.1151 | 1.9524 |
| | Dozerformer | 0.2135 | 0.1437 | 0.9626 | 1.7765 |
| | FT-Dozerformer | 0.1997 | 0.1411 | 0.9547 | 1.7281 |
| | Our Method | **0.1951** | **0.1376** | **0.9496** | **1.7183** |
| Pred 192 | Informer | 0.3668 | 0.2664 | 1.0091 | 1.8809 |
| | Autoformer | 0.3352 | 0.2117 | 1.1261 | 2.1462 |
| | FT-Transformer | 0.3916 | 0.2952 | 1.1536 | 1.9834 |
| | Dozerformer | 0.2633 | 0.1894 | 1.0088 | 1.8546 |
| | FT-Dozerformer | 0.2347 | 0.1836 | 1.0084 | 1.8535 |
| | Our Method | **0.2316** | **0.1820** | **1.0060** | **1.8526** |

추가적인 실험으로 Traffic 데이터를 7:1:2의 비율이 아닌 4:3:3의 비율로 분할 하여 시계열 예측을 진행하였다. 표 3은 두 가지 데이터 분할 방식에서 예측실험을 진행한 결과를 나타낸 것이다. 실험 결과, 기존과 새로운 분할 비율 모두에서 제시한 방법이 가장 우수한 성능을 보였으며, 새로운 분할 비율에서 제시한 예측 방법과 다른 예측 기법 간의 성능 차이가 더 크게 나타났다. 이는 학습 데이터셋의 비율이 작아 VAE의 효과가 더 크게 작용한 것이 원인으로 생각된다.

표 3. Traffic 데이터 분할 비율에 따른 결과

| | | Split 7:1:2 | | Split 4:3:3 | |
|---|---|---|---|---|---|
| | | MAE | MSE | MAE | MSE |
| Traffic volume (pred 48) | Informer | 0.2539 | 0.1454 | 0.3432 | 0.2278 |
| | Autoformer | 0.2845 | 0.1667 | 0.3338 | 0.2116 |
| | FT-Transformer | 0.3312 | 0.1945 | 0.3561 | 0.2348 |
| | Dozerformer | 0.1773 | 0.0997 | 0.1933 | 0.1180 |
| | FT-Dozerformer | 0.1682 | 0.0961 | 0.1870 | 0.1076 |
| | Our Method | **0.1636** | **0.0930** | **0.1829** | **0.1010** |
| Traffic volume (pred 96) | Informer | 0.2927 | 0.1873 | 0.3801 | 0.2675 |
| | Autoformer | 0.3009 | 0.1794 | 0.3576 | 0.2280 |
| | FT-Transformer | 0.3575 | 0.2208 | 0.3850 | 0.2669 |
| | Dozerformer | 0.2135 | 0.1437 | 0.2258 | 0.1453 |
| | FT-Dozerformer | 0.1997 | 0.1411 | 0.2155 | 0.1444 |
| | Our Method | **0.1951** | **0.1376** | **0.2132** | **0.1363** |
| Traffic volume (pred 192) | Informer | 0.3668 | 0.2664 | 0.4383 | 0.3408 |
| | Autoformer | 0.3352 | 0.2117 | 0.3689 | 0.2542 |
| | FT-Transformer | 0.3916 | 0.2952 | 0.4421 | 0.3511 |
| | Dozerformer | 0.2633 | 0.1894 | 0.2717 | 0.2033 |
| | FT-Dozerformer | 0.2347 | 0.1836 | 0.2631 | 0.1999 |
| | Our Method | **0.2316** | **0.1820** | **0.2586** | **0.1928** |

아래의 그림 7, 8은 각각 Traffic volume 및 Air pollution 데이터셋의 원본 데이터와 VAE를 이용해 생성된 종속 변수 시계열 데이터의 분포를 비교한 것이다. 이를 통해 원본 데이터의 변화 패턴을 증강 데이터가 얼마나 잘 반영하고 있는지 확인할 수 있다. 파란색으로 표시된 것이 원본 데이터이며 노란색으로 표시된 것이 생성된 데이터를 나타낸다. 원본 데이터의 분포와 조금씩 차이가 있으나 증강된 데이터가 원본 데이터의 분포를 잘 따라가며 변화를 주었음을 알 수 있다.

표 4와 5는 VAE를 통해 생성한 데이터를 학습에 활용한 개수에 따라 효과가 어떻게 변하는지를 나타낸다. Air pollution 데이터셋의 경우, 예측 길이에 관계없이 VAE로 증강된 데이터를 추가할수록 점진적으로 성능이 개선되는 모습을 보였다. Traffic volume 데이터의 경우, 예측 길이가 짧을 때는 사용한 데이터의 개수가 증가할수록 성능이 향상되었으나 예측 길이가 길어질수록 2개를 사용하는 것이 최적의 결과를 가져오고 이후로 데이터를 학습할수록 성능이 낮아졌다. 이러한 결과를 통해, 증강된 데이터를 많이 사용하는 것보다는 데이터에 따라 적절한 사용 개수를 찾는 것이 중요함을 알 수 있다.

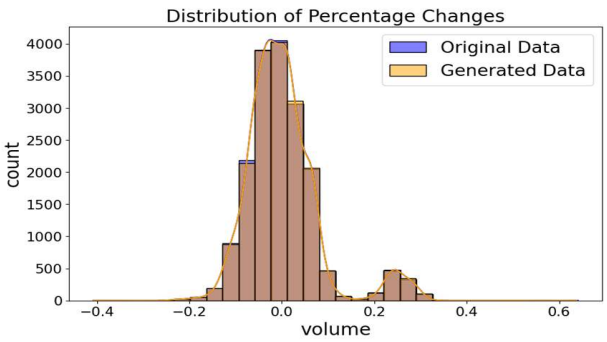


그림 7. Traffic volume 데이터와 증강 데이터의 분포 비교

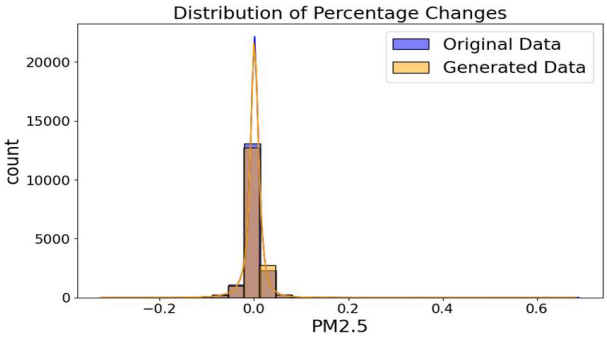


그림 8. Air pollution 데이터와 증강 데이터의 분포 비교

표 4. 증강 데이터 사용 개수에 따른 Traffic volume 데이터셋 예측 결과

| Traffic Volume | Pred 48 | | Pred 96 | | Pred 192 | |
|---|---|---|---|---|---|---|
| | MAE | MSE | MAE | MSE | MAE | MSE |
| VAE 1개 사용 | 0.1676 | 0.0956 | 0.1981 | **0.1376** | 0.2309 | 0.1848 |
| VAE 2개 사용 | 0.1649 | 0.0940 | 0.1997 | 0.1384 | **0.2316** | **0.1820** |
| VAE 3개 사용 | 0.1668 | 0.0955 | 0.1958 | 0.1385 | 0.2326 | 0.1859 |
| VAE 4개 사용 | 0.1656 | 0.0943 | 0.1957 | 0.1387 | 0.2318 | 0.1881 |
| VAE 5개 사용 | **0.1634** | **0.0930** | **0.1951** | 0.1396 | 0.2323 | 0.1901 |

표 5. 증강 데이터 사용 수에 따른 Air pollution 데이터셋 예측 결과

| Air pollution | Pred 48 | | Pred 96 | | Pred 192 | |
|---|---|---|---|---|---|---|
| | MAE | MSE | MAE | MSE | MAE | MSE |
| VAE 1개 사용 | 0.8487 | 1.4275 | 0.9518 | 1.7223 | 1.0072 | 1.8550 |
| VAE 2개 사용 | 0.8469 | 1.4257 | 0.9510 | 1.7205 | 1.0067 | 1.8539 |
| VAE 3개 사용 | **0.8459** | 1.4243 | 0.9503 | 1.7195 | 1.0063 | 1.8533 |
| VAE 4개 사용 | 0.8476 | 1.4234 | 0.9497 | 1.7188 | 1.0062 | 1.8529 |
| VAE 5개 사용 | 0.8536 | **1.4226** | **0.9496** | **1.7184** | **1.0061** | **1.8527** |

마지막으로, VAE와 또 다른 시계열 증강 기법인 Window Warping(WW)을 사용한 결과를 비교하고자 한다. WW 기법은 시계열의 특정 구간을 늘리거나 줄이는 방식으로 변형함으로써, 전체적인 흐름은 유지하면서 다양한 데이터를 생성하는 기법이다. 제안한 VAE와 WW 기법 모두 GAN 기반의 시계열 증강기법에서 발생하는 Mode collapse 문제를 피할 수 있어 정상적인 증강 데이터를 확보할 수 있으며, 시계열의 전체적인 패턴이 원본과 유사하다는 공통된 장점이 존재한다. 그러므로 해당 기법을 활용해 획득한 데이터를 학습에 사용하고 VAE 기법을 사용한 성능과 비교한다.

표 6은 WW 기법을 활용하여 증강한 Air pollution 데이터셋을 학습에 사용한 개수에 따른 영향을 나타낸다. 표 5와 비교하여 시계열 예측 성능을 확인한 결과, VAE로 증강한 데이터셋의 학습 성능이 더 좋은 것으로 나타났다. 또한, WW 기법으로 증강한 데이터는 대략 2개 정도 사용하였을 때 가장 성능이 좋았으며, 그보다 많은 수의 데이터를 사용할 경우 오히려 성능이 저하되는 경향을 보였다. 이는 WW 기법이 VAE에 비해 종속 변수의 다양성을 충분히 표현하지 못하기 때문으로 해석할 수 있다.

표 6. Window Warping(WW) 기법으로 생성된 데이터를 사용하여 Air pollution 데이터셋을 예측한 결과

| Air pollution | Pred 48 | | Pred 96 | | Pred 192 | |
|---|---|---|---|---|---|---|
| | MAE | MSE | MAE | MSE | MAE | MSE |
| WW 1개 사용 | 0.8520 | 1.4303 | 0.9534 | 1.7209 | 1.0073 | 1.8555 |
| WW 2개 사용 | **0.8481** | 1.4272 | **0.9531** | **1.7207** | **1.0072** | 1.8553 |
| WW 3개 사용 | 0.8498 | 1.4272 | 0.9539 | 1.7290 | 1.0074 | **1.8539** |
| WW 4개 사용 | 0.8488 | **1.4266** | 0.9537 | 1.7250 | 1.0077 | 1.8548 |
| WW 5개 사용 | 0.8489 | 1.4481 | 0.9537 | 1.7327 | 1.0088 | 1.8543 |

## V. 결 론

본 논문에서는 기존 연구에서 범주형 변수를 사용하지 않았던 한계를 극복하기 위해 범주형 변수의 정보 손실 없이 시계열 예측을 수행하는 방법을 제안하였다. 또한, 원본 데이터만으로 학습할 경우 발생할 수 있는 시계열 패턴의 편향성

과 과적합 위험을 해소하기 위해, VAE를 활용한 데이터 증강 기법을 적용하였다.

실험 결과는 Traffic Volume 데이터셋과 Air pollution 데이터셋 모두 범주형 변수를 사용하지 않은 것보다 함께 사용하는 것이 더 좋은 결과를 보였으며, VAE를 통해 생성한 데이터를 이용하는 것이 가장 좋은 결과를 보였다. 이를 통해 시계열 예측에서 범주형 변수를 사용하는 것에 대한 중요성을 제고 하였으며, 더 나아가 VAE 기반 데이터 증강을 활용해 시계열 예측 성능 개선 방안을 제시하였다.

## REFERENCES


[1] S.-Y. Shih, F.-K. Sun, and H.-Y. Lee, "Temporal Pattern Attention for Multivariate Time Series Forecasting," *Machine Learning*, vol. 108, pp. 1421 - 1441, 2019.

[2] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks," *in The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 95 - 104, 2018.

[3] H. Zhou et al., "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting," *in The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, pp. 11106 - 11115, AAAI Press, 2021.

[4] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting," *in Advances in Neural Information Processing Systems*, 2021.

[5] Y. Zhang et al., "Sparse Transformer with Local and Seasonal Adaptation for Multivariate Time Series Forecasting," *Scientific Reports*, vol. 14, 2024.

[6] A. Vaswani, "Attention is All You Need," *Advances in Neural Information Processing Systems*, 2017.

[7] E. Poslavskaya, A. Korolev, "Encoding Categorical Data: Is There Yet Anything 'Hotter' Than One-Hot Encoding?," *arXiv preprint arXiv*:2312.16930, 2023.

[8] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko, "Revisiting Deep Learning Models for Tabular Data," *Advances in Neural Information Processing Systems*, Vol. 34, pp. 18932 - 18943, 2021.

[9] D.P. Kingma and M. Welling, "An Introduction to Variational Autoencoders," *Foundations and Trends® in Machine Learning*, vol. 12, no. 4, pp. 307 - 392, 2019.

[10] Van den Bossche, F.A.M. et al., "Analysis of road risk by age and gender category," Transportation Research Record, vol. 2019, no. 1, 2007.

[11] 이경훈, 김진오, "더미변수를 포함하는 다변수 시계열 모델을 이용한 단기부하예측," *전기학회논문지 A권*, 제52권, 제8호, 450-456쪽, 대한전기학회, 2003년 8월

[12] B. Kalina, J.H. Lee, and K.T. Na, "Enhancing Portfolio Performance through Financial Time-Series Decomposition-Based Variational Encoder-Decoder Data Augmentation," *Symmetry*, vol. 16, no. 3, p. 283, 2024.


## 저 자 소 개


**이재진(준회원)**

2023년 계명대학교 경영학과 학사 졸업.

2024년 인하대학교 컴퓨터공학과 석사 졸업.

<주관심분야 : 시계열 데이터, 트랜스포머>

**이주홍(정회원)**

1983년 서울대학교 전자계산기공학과 학사 졸업.

1985년 서울대학교 컴퓨터공학과 석사 졸업.

2001년 한국과학기술원 컴퓨터공학 박사 졸업.

<주관심분야 : 머신러닝, 금융>