

머신러닝 기반 시계열 예측 시스템 비교 및 최적 예측 시스템 구현

한용희*, 고방원**

Comparison and Implementation of Optimal Time Series Prediction Systems Using Machine Learning

Yong Hee Han*, Bangwon Ko**

요약 본 연구는 시계열 데이터를 효과적으로 예측하기 위해 데이터를 Seasonal-Trend Decomposition on Loess 을 통해 추세, 계절성, 잔차 성분으로 분해한 후 추세 성분에는 ARIMA, 계절성 성분에는 Fourier Series Regression, 잔차 성분에는 XGBoost를 적용하는 하이브리드 예측 모델을 제안하였다. 또한, ARIMA, XGBoost, LSTM, EMD-ARIMA, CEEMDAN-LSTM 모델을 포함한 성능 비교 실험을 수행하여 각 모델의 예측 성능을 평가하였다. 실험 결과, 제안된 하이브리드 모델은 MAPE, MAAPE, RMSE 지표에서 각각 3.8%, 3.5%, 0.35로 가장 좋은 평가 지표 값을 보이며 기존의 단일 모델보다 우수한 성능을 보였다.

Abstract In order to effectively predict time series data, this study proposed a hybrid prediction model that decomposes the data into trend, seasonality, and residual components using Seasonal-Trend Decomposition on Loess, and then applies ARIMA to the trend component, Fourier Series Regression to the seasonality component, and XGBoost to the remaining components. In addition, performance comparison experiments including ARIMA, XGBoost, LSTM, EMD-ARIMA, and CEEMDAN-LSTM models were conducted to evaluate the prediction performance of each model. The experimental results show that the proposed hybrid model outperforms the existing single models with the best performance indicator values in MAPE(3.8%), MAAPE(3.5%), and RMSE(0.35) metrics.

Key Words : Hybrid forecasting, Seasonal-Trend Decomposition on Loess, Machine Learning, Time series, Performance evaluation

1. 서론

시계열 예측은 데이터 분석 및 의사 결정 과정에서 중요한 역할을 수행한다. 시계열 데이터는 시간 순서에 따라 수집된 데이터로, 경제, 금융, 에너지 등 다양한 분야에서 널리 사용된다. 시계열 예측을 통해 미래의 데이터를 예측함으로써 기업은 재고 관리, 자원 할당, 전략 계획 수립 등에 있어 중요한 의사 결정 수행이 가능하며, 최근 머신러닝과 딥러닝 기술의 발전으로 시계열 예측의 정확도와 효율성이 크게 향상되었다[1].

본 연구의 목적은 머신러닝 기반 시계열 예측 시스템을 비교하고 STL(Seasonal-Trend Decomposition based on Loess)[2]과 머신러닝 알고리즘을 결합하여 시계열 데이터를 효과적으로 예측할 수 있는 하이브리드 모델을 제안하는 것이다. 기존의 시계열 예측 방법들은 각각의 데이터 특성에 따라 한계가 있으며, 이를 보완하기 위해 다양한 접근 방식을 시도할 필요가 존재한다. 본 연구에서는 STL을 사용하여 시계열 데이터를 추세(trend), 계절성(seasonality), 잔차(residual) 성분으로 분해하고, 각 성분을 별도의 알고리즘으로

This work was supported by the Soongsil University Research Fund(Convergence Research) of 2020.

*Department of Entrepreneurship and Small Business, Soongsil University (amade@ssu.ac.kr)

**Corresponding Author: Department of Statistics and Actuarial Science, Soongsil University

Received June 17, 2024

Revised June 26, 2024

Accepted August 17, 2024

분석하여 예측 성능을 향상시키고자 한다.

정확한 시계열 예측은 기업과 기관의 효율적인 운영 및 전략적 의사 결정에 필수적이다. 특히, 전력 수요 예측, 금융 시장 분석, 기후 변화 예측 등 다양한 분야에서 시계열 예측의 중요성이 부각되고 있다[3]. 본 연구에서 제안하는 하이브리드 모델은 기존의 단일 알고리즘 기반 예측 방법보다 높은 예측 정확도를 제공할 수 있으며, 다양한 시계열 데이터에 적용 가능하다. 이를 통해 기업과 기관은 더욱 정교한 예측을 바탕으로 자원을 효율적으로 관리하고, 리스크를 최소화할 수 있다. 본 연구를 통해 제안된 하이브리드 예측 모델이 시계열 예측 분야에서 높은 정확도를 제공할 수 있음을 입증하고, 다양한 응용 가능성을 제시하고자 한다.

본 논문의 구조는 다음과 같다. 2장에서는 시계열 예측의 기존 방법론과 최신 연구 동향을 요약하고, 3장에서는 제안된 하이브리드 예측 모델의 구성과 실험 결과를 상세히 설명하며, 4장에서는 연구의 주요 성과와 한계를 논의하고 향후 연구 방향을 제시한다.

2. 선행 연구

시계열 예측 시스템은 통계적 방법, 전통적 머신러닝 방법, 딥러닝 방법, 하이브리드 방법, AutoML 방법으로 구분 가능하다. 각 방법은 고유의 특징과 강점이 존재하므로 다양한 시계열 데이터 특성에 따라 적절한 방법을 선택하는 것이 중요하다.

통계적 방법은 시계열 데이터의 자기 상관성을 이용하여 미래의 값을 예측한다. ARIMA (AutoRegressive Integrated Moving Average) 모델은 자기회귀, 차분, 이동 평균의 세 가지 요소를 결합하여 시계열 데이터를 분석하고 예측한다. 이는 주로 선형 시계열 데이터에 적합하며, 데이터의 정상성을 확보하기 위해 차분 과정을 포함한다. SARIMA(Seasonal ARIMA)는 ARIMA 모델의 확장판으로 계절성을 고려한 시계열 예측에 적합하다[3]. STL은 Loess 평활화 기법을 순차적으로 적용하여 시계열 데이터를 추세, 계절성, 잔차 요소로 분해한 후 이를 개별적으로 모델링하는 기법으로, 데이터의 복잡한 패턴을 효과적으로 분해하고 각 구성 요소를 별도로 학습하여 종합적인 예측 성능을 향상시킨다[2].

Fourier Series Regression에는 주기적인 패턴을 가진 시계열 데이터를 분석하는 데 사용된다. 이 방법은 시계열 데이터를 푸리에 급수로 표현하여 주기적인 성분을 추출하고 예측하는 방식이다. 이 방법은 주기성이 뚜렷한 시계열 데이터에서 뛰어난 성능을 보이지만, 비주기적인 요소가 많은 데이터에는 적합하지 않다. 푸리에 급수 회귀는 특히 기상 데이터와 같이 계절성이 중요한 데이터에 유용하다.

전통적 머신러닝 방법은 데이터의 패턴을 학습하여 예측 성능을 높이는 데 중점을 두며, 주요 알고리즘은 KNN(k-Nearest Neighbor), SVR(Support Vector Regression), Random Forest, GBM(Gradient Boosting Machine) 등이다. Random Forest와 GBM은 앙상블 학습 방법을 사용하여 다수의 결정 트리를 결합하여 예측 성능을 높인다. 이러한 방법들은 비선형성과 상호작용 효과를 잘 포착하며 일반화 능력이 뛰어나다. 특히, XGBoost(eXtreme Gradient Boosting)와 같은 GBM은 각 단계에서 이전 모델의 오류를 보정하여 점진적으로 예측 성능을 개선한다. XGBoost는 Gradient Boosting Framework을 기반으로 하여, 여러 약한 학습기(weak learners)를 결합하여 강한 학습기(strong learner)를 만드는 방식이며, 예측 정확도가 중요한 시계열 예측 문제에 널리 사용된다. XGBoost의 장점은 높은 예측 정확도와 빠른 학습 속도이며, 과적합을 방지하기 위한 다양한 정규화 기법을 제공한다[4].

딥러닝 방법은 복잡한 시계열 데이터의 특성을 학습하는 데 매우 효과적으로, 주요 모델로는 Long Short-Term Memory(LSTM), Recurrent Neural Network(RNN), Convolutional Neural Network(CNN) 등이 존재한다. LSTM은 RNN의 일종으로, 데이터의 시간적 의존성을 효과적으로 처리할 수 있는 능력을 갖추고 있다. 이는 긴 시간 동안의 패턴을 포착할 수 있으며, 기율기 소실 문제를 해결하기 위해 게이트 메커니즘을 사용한다 [5]. CNN은 주로 이미지 처리에 사용되지만, 시계열 데이터에서도 특징 추출에 유용하다. CNN은 필터를 사용하여 데이터의 국소 패턴을 포착하며, 이를 통해 중요한 특징을 추출하고 예측에 활용할 수 있다. 또한, CNN과 RNN을 결합한 모델도 존재하며, 이는 시계열

데이터의 공간적 및 시간적 패턴을 동시에 학습할 수 있는 장점을 가진다.

하이브리드 방법은 다수의 알고리즘을 결합하여 각 알고리즘의 장점을 이용하여 예측 성능을 극대화하는 접근법으로, 예를 들어 하이브리드 딥러닝 모델은 딥러닝 모델과 기타 모델을 결합하여 예측 성능을 극대화한다. EMD-ARIMA 모델은 시계열 데이터를 Empirical Mode Decomposition(EMD) 기법으로 분해한 후 각 성분을 ARIMA 모델로 예측하여 종합하며, EMD는 비선형적이고 비정상적인 데이터를 다루는데 효과적이고 ARIMA는 추세 성분을 포착한다[6]. CEEMDAN-LSTM 모델은 CEEMDAN(Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) 기법으로 시계열 데이터를 분해한 후 각 성분을 LSTM 모델로 예측하는 방식으로, EMD의 단점을 보완하여 시계열 데이터를 더 정교하게 분해하며 LSTM은 시계열 데이터의 시간적 의존성을 학습한다[7]. 이러한 하이브리드 접근법은 데이터의 비선형성과 복잡한 패턴을 효과적으로 처리 가능하다. 하이브리드 방법의 일종인 앙상블 방법은 여러 머신러닝 모델을 결합하여 예측의 정확도를 높이는 방법으로, 개별 모델의 약점을 보완하고, 전반적인 예측 성능을 향상시킬 수 있다. 앙상블 방법에는 배깅, 부스팅, 스택킹 등이 포함되며, 이들 방법은 다양한 모델을 결합하여 안정적이고 일관된 예측 성능을 제공한다.

AutoML 방법은 머신러닝 모델의 자동화된 설계 및 최적화를 통해 시계열 예측의 효율성을 높이는 방법으로, 데이터 전처리, 특성 공학, 하이퍼파라미터 최적화, 모델 선택 및 앙상블을 자동화하는 시스템이다[1].

회전 기계 고장 진단 관련 기존 연구로는 SE-ResNet50 기반 대조 학습 네트워크와 Focal Loss를 사용하여 라벨링이 제한된 상황에서 높은 진단 정확도를 달성한 사례[8]가 있으나, 라벨링이 제한되지 않은 일반적인 상황에서는 진단 정확도가 낮은 문제가 존재한다. 다른 연구[9][10]에서 제시한 모델은 일반적인 상황에서도 진단 정확도를 제공하는 장점이 있으나, 고장 여부 진단에 특화되어 있어 일반적인 시계열 예측에는 적용할 수 없는 한계가 존재한다.

최근 머신러닝 기술의 발전은 시계열 예측에서 중

요한 역할을 하고 있다. 딥러닝 기술은 복잡한 시계열 데이터를 처리하고 예측 정확도를 높이는 데 매우 효과적이다. 주요 딥러닝 모델로는 LSTM, GRU, CNN 등이 있으며, 이들 모델은 시계열 데이터의 특성을 효과적으로 학습할 수 있다. 머신러닝 기반 시계열 예측 시스템은 각기 다른 접근 방식을 취하고 있으며, 각 모델의 성능은 데이터의 특성과 문제의 성격에 따라 다르게 나타난다. ARIMA 모델은 시계열 데이터의 자기상관 구조를 잘 반영하지만, 비정상성 데이터를 처리하는 데 한계가 있다. STD는 데이터의 구조적 이해에 유용하지만, 직접적인 예측보다는 분석에 초점을 맞춘다. 푸리에 급수 회귀는 주기적인 데이터에 뛰어난 성능을 보이지만, 비주기적인 요소가 많은 데이터에는 적합하지 않다. XGBoost는 높은 예측 정확도를 자랑하며 다양한 시계열 예측 문제에 적용 가능하지만, 복잡한 하이퍼파라미터 튜닝이 필요하다. 각 모델의 특성을 고려할 때, 최적의 시계열 예측 시스템은 데이터의 특성과 문제의 요구사항에 맞춰 모델을 선택하고, 필요에 따라 여러 모델을 결합하여 사용해야 한다는 결론을 도출할 수 있다. 예를 들어, 계절성이 뚜렷한 데이터는 푸리에 급수 회귀를, 비정상성 데이터는 ARIMA와 XGBoost를 결합하여 사용할 수 있다. 이러한 선행 연구를 바탕으로, 3장에서는 최적 예측 시스템을 제안하고 그 구현 방법을 논의한다.

3. 최적 예측 시스템 제안 및 실험 결과

본 연구에서는 통계적 방법인 STL과 다양한 머신러닝 알고리즘을 결합하여 시계열 데이터를 효과적으로 예측하는 하이브리드 시스템을 제안한다. 이 장에서는 제안된 모델의 워크플로우와 각 성분에 적용된 알고리즘의 선택 근거를 상세히 설명한다.

3.1 워크플로우 요약

본 논문에서 제안하는 최적 예측 시스템의 워크플로우는 그림 1과 같다.

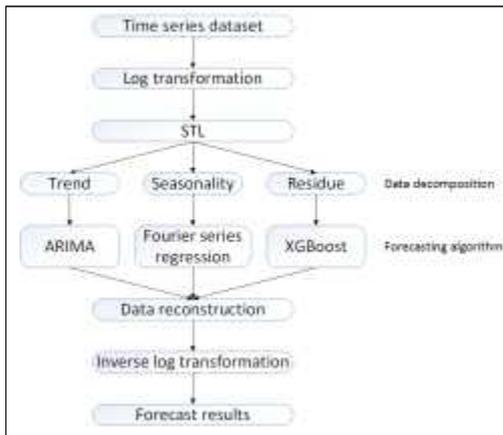


그림 1. 제안하는 시스템 워크플로우
 Fig. 1. Proposed system workflow

- 1) 시계열 데이터 전처리: 로그 변환을 통해 데이터의 분산을 안정화하고 이상치를 처리한다.
- 2) 데이터 분해: STL 알고리즘을 사용하여 데이터를 추세(trend), 계절성(seasonality), 잔차(residue) 성분으로 분해한다.
- 3) 개별 성분 예측
 - 3-1) 추세 데이터: ARIMA 모델을 사용하여 추세 성분을 추정한다.
 - 3-2) 계절성 데이터: Fourier Series Regression을 사용하여 계절성 성분을 추정한다.
 - 3-3) 잔차 데이터: XGBoost 알고리즘을 사용하여 잔차 성분을 추정한다.
- 4) 예측치 합산: 각 성분의 예측치를 합산한다.
- 5) 로그 역변환: 종합된 예측값을 원 데이터 스케일로 변환하여 최종 예측값을 도출한다.

이러한 워크플로우를 통해 시계열 데이터의 다양한 패턴을 효과적으로 포착하고 예측 정확도를 높일 수 있다.

3.2 워크플로우 상세

3.2.1 데이터 전처리 (로그 변환) 및 STL

시계열 데이터의 로그 변환을 통해 데이터의 분산을 안정화하고 이상치를 처리한다. STL은 어떤 계절성도 처리 가능하며, 이상치에 대해 강건한 등의 여러 장

점을 가지고 있으나, 덧셈 분해만 가능한 단점이 존재하므로 이를 극복하기 위해 로그 변환을 적용하였다. 로그 변환은 데이터의 분포를 정규화하여 모델의 학습 성능이 향상하는 부수적인 장점이 존재한다. 또한, 결측값이 있는 경우 이를 보간법이나 다른 적절한 방법으로 처리한다. 이후 STL 알고리즘을 사용하여 시계열 데이터를 추세, 계절성, 잔차 성분으로 분해한다. STL은 데이터의 비선형성과 다양한 계절성을 처리할 수 있는 강력한 기법으로, 계절적 변동을 정확하게 추출하고 추세 및 잔차 성분과의 분리를 용이하게 한다. 이러한 특성은 특히 경제 및 금융 시계열 데이터의 분석에서 중요한 역할을 한다.

3.2.2 개별 성분 예측

다음과 같은 단계로 구성된다.

추세 추정: ARIMA 모델을 사용하여 데이터를 평활화하고 추세 성분을 추정한다. 추세 성분은 시계열 데이터에서 장기적인 상승 또는 하강 패턴을 나타낸다. ARIMA 모델은 이러한 장기적인 패턴을 포착하고 예측하는 데 적합하다. 이는 ARIMA 모델이 데이터의 자기 상관성과 차분을 기반으로 장기적인 추세를 효과적으로 분석할 수 있기 때문이다. 특히, 추세 데이터는 주기적인 변동보다는 일정한 방향성을 가지는 경우가 많으므로 ARIMA 모델의 활용이 적합하다. ARIMA 모델은 과거 데이터의 패턴을 분석하여 미래의 추세를 예측할 수 있어 추세 성분을 정확하게 모델링 가능하다.

계절성 추정: 데이터에서 추세 성분을 제거하고 남은 데이터에서 주기적인 패턴을 추출하여 계절성 성분을 추정한다. 계절성 성분은 시계열 데이터에서 주기적으로 반복되는 패턴을 나타낸다. Fourier Series Regression은 이러한 주기적인 패턴의 수학적 모델링에 매우 효과적이다. 이는 Fourier Series Regression이 다양한 주기를 가지는 데이터를 분석할 수 있는 유연성을 제공하기 때문이다. 계절성 데이터는 특정 주기마다 반복되는 패턴을 보이므로, Fourier Series Regression을 통해 이러한 패턴을 정확하게 예측할 수 있다. Fourier Series Regression은 주기적인 변동을 수학적으로 표현하여 계절성 성분의 예측 정확도

를 높일 수 있다.

잔차 추정: 데이터에서 추세 성분과 계절성 성분을 제거하고 남은 잔차 성분을 추정한다. 잔차 성분은 추세와 계절성을 제거한 후 남은 데이터로, 주로 비선형적이고 불규칙한 패턴을 보인다. XGBoost 알고리즘은 이러한 비선형적이고 복잡한 패턴을 효과적으로 학습하고 예측하는 데 강력한 성능을 보인다. 이는 XGBoost가 여러 약한 학습기를 결합하여 강력한 예측 모델을 구성하는 부스팅 기법을 사용하기 때문이다. 잔차 데이터는 예측하기 어려운 비선형적 패턴을 포함할 수 있으므로, XGBoost의 비선형 학습 능력이 큰 장점이 된다. XGBoost는 잔차 성분의 복잡한 패턴을 효과적으로 포착하고 예측 가능하다.

3.2.3 예측치 합산 및 로그 역변환

각 성분의 예측치 합산을 통해 최종 예측값을 도출하고 최종 예측값을 원래 데이터 스케일로 변환하기 위해 로그 역변환을 적용한다. 이는 예측값을 원본 데이터와 동일한 스케일로 만들어 예측 결과를 해석하기 쉽게 한다.

3.3 실험

본 연구에서 제안된 시스템의 성능을 평가하기 위해 인공지능제조플랫폼(KAMP)에서 제공하는 회전 기계 고장 유형 AI 데이터셋[11]을 사용하였다. 데이터셋은 4개의 센서로부터 수집된 z 방향 진동 데이터로 구성되어 있으며, 약 150만 개의 샘플로 구성되어 있다. 시계열 데이터는 시간적 홀드아웃(holdout)을 적용하여 분할하였으며 테스트셋은 5%로 설정하였다. 평가 지표로는 회귀 시스템의 평가에 일반적으로 사용되는 MAPE(Mean Absolute Percentage Error) 및 RMSE(Root Mean Square Error)에 추가하여 MAPE보다 이상치에 덜 민감하고 시계열 모형의 예측 성능을 더 잘 반영하는 MAAPE(Mean Arctangent Absolute Percentage Error)를 사용하였다. 성능 비교 실험에는 ARIMA, XGBoost, LSTM, EMD-ARIMA, CEEMDAN-LSTM 모델이 포함되었으며, 그 이유는 다음과 같다.

ARIMA: 전통적인 통계적 시계열 예측 모델로, 선

형 시계열 데이터의 예측에 널리 사용된다[3].

XGBoost: 강력한 부스팅 알고리즘으로, 비선형적이고 복잡한 패턴을 효과적으로 학습 가능하다[4].

LSTM: 순환 신경망의 일종으로, 시계열 데이터의 시간적 의존성 학습에 뛰어난 성능을 보인다[5].

EMD-ARIMA: EMD를 통해 데이터를 분해한 후, 각 성분을 ARIMA 모델로 예측하여 비선형성과 비정상성을 처리할 수 있다[6].

1D CNN-LSTM: 1D CNN이 입력 데이터에서 국부적 패턴이나 특징을 자동으로 추출한 후, LSTM이 긴 시퀀스의 종속성을 모델링한다[12]. 잡음에 대한 내성이 강하고 긴 시퀀스의 모델링에 강하여 음성 인식, 자연어 처리 등에 많이 이용되나, 본 논문에서 제안된 모델 대비 과적합되기 쉽고 설계 과정이 복잡하여 모델 최적화가 어려운 단점이 존재한다.

CEEMDAN-LSTM: CEEMDAN을 통해 데이터를 분해한 후, 각 성분을 LSTM 모델로 예측하여 더 정교한 분해와 시간적 의존성을 동시에 처리 가능하다[7]. 각 시스템의 성능 비교 실험 결과는 표 1 및 그림 2와 같다.

표 1. 실험 결과
Table 1. Experimental results

System	MAPE	MAAPE	RMSE
ARIMA	5.2%	4.8%	0.45
XGBoost	4.8%	4.4%	0.42
LSTM	4.5%	4.1%	0.40
EMD-ARIMA	4.2%	3.9%	0.38
1D CNN-LSTM	4.1%	3.9%	0.37
CEEMDAN-LSTM	4.0%	3.7%	0.37
Proposed system	3.8%	3.5%	0.35

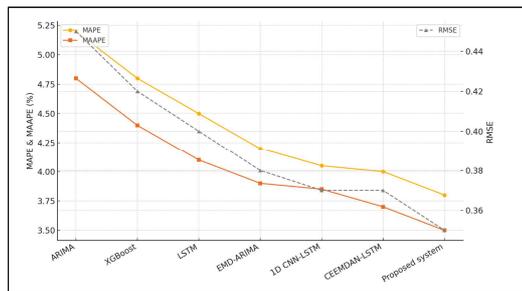


그림 2. 실험 결과
Fig. 2. Experimental results

ARIMA 모델은 MAPE 5.2%, MAAPE 4.8%, RMSE 0.45로, 비교 모델 중 가장 낮은 성능을 보였으며, 이는 단일 통계적 모델인 ARIMA가 비선형적이고 복잡한 패턴을 충분히 포착하지 못했음을 의미한다. XGBoost 모델은 MAPE 4.8%, MAAPE 4.4%, RMSE 0.42로, ARIMA에 비해 더 낮은 오차율을 보였으며, 이는 XGBoost가 비선형 패턴을 효과적으로 학습하여 예측 성능을 향상시켰음을 시사한다. LSTM 모델은 MAPE 4.5%, MAAPE 4.1%, RMSE 0.40으로, 시계열 데이터의 시퀀스 특성을 잘 학습하여 높은 예측 성능을 보였다. EMD-ARIMA 모델은 MAPE 4.2%, MAAPE 3.9%, RMSE 0.38로, 데이터 분해 기법을 사용하여 예측 성능을 개선하였다. CEEMDAN-LSTM 모델은 MAPE 4.0%, MAAPE 3.7%, RMSE 0.37로, EMD-ARIMA보다도 더 높은 성능을 보였다. 이는 CEEMDAN을 통해 데이터의 비선형성과 비정상성을 더욱 효과적으로 처리할 수 있음을 보여준다. 본 연구에서 제안된 하이브리드 모델은 MAPE 3.8%, MAAPE 3.5%, RMSE 0.35로, 모든 평가 지표에서 가장 좋은 평가 지표를 기록하였다. 이는 추세, 계절성, 잔차 성분을 개별적으로 분석하고 종합하는 접근 방식이 예측 정확도를 크게 향상시켰음을 입증한다. 제안된 모델이 가장 낮은 오차율을 기록한 것은 각 성분의 특성에 맞는 최적화된 알고리즘을 사용하여 데이터를 효과적으로 예측했기 때문으로 판단된다.

4. 결론

본 연구는 시계열 데이터를 효과적으로 예측하기 위해 STL과 다양한 머신러닝 알고리즘을 결합한 하이브리드 예측 모델을 제안하였다. 제안된 모델은 시계열 데이터를 추세, 계절성, 잔차 성분으로 분해한 후, 추세 성분에는 ARIMA, 계절성 성분에는 Fourier Series Regression, 잔차 성분에는 XGBoost를 적용하였다. 본 논문에서 제안된 예측 모델을 KAMP에서 제공하는 회전 기계 고장 유형 AI 데이터셋을 사용하여 실험한 결과, 제안된 모델이 타 모델(ARIMA, XGBoost, LSTM, EMD-ARIMA, CEEMDAN-LSTM)에 비해 MAPE, MAAPE, RMSE

지표에서 모두 가장 좋은 평가 지표 값을 기록하였으며(MAPE 3.8%, MAAPE 3.5%, RMSE 0.35), 모든 평가 지표에서 하이브리드 모델(EMD-ARIMA, 1D CNN-LSTM, CEEMDAN-LSTM, 제안된 모델)이 개별 모델(ARIMA, XGBoost, LSTM)보다 더 좋은 값을 보였다.

본 연구는 시계열 예측 모델의 성능을 향상시키기 위해 다양한 방법을 시도했지만, 몇 가지 한계가 존재한다. 여러 알고리즘을 결합한 하이브리드 모델은 단순한 모델보다 복잡성이 높아 계산 비용이 증가하며, 이는 대규모 데이터셋이나 실시간 예측에 적용 시 부담될 수 있다. 또한 제안된 모델은 특정 시계열 데이터셋에서 높은 성능을 보였지만, 다른 유형의 데이터셋에서의 성능 저하 가능성이 존재하므로 다양한 데이터셋에 대한 추가 검증이 필요하다. 마지막으로, 다양한 패턴을 포착하기 위해 각 성분에 최적화된 알고리즘을 사용하였지만, 이는 데이터의 특성에 따라 모델의 일반화 성능을 저하시킬 수 있다.

본 연구의 결과를 바탕으로 다음과 같은 향후 연구 방향을 제안한다. 제안된 모델의 일반화 성능을 높이기 위해 다양한 도메인의 시계열 데이터를 사용하여 모델을 검증할 필요가 존재하며, 이를 통한 모델의 범용성 평가 및 개선이 가능하다. 또한 하이브리드 모델의 계산 비용을 감소시키는 방법을 연구하여 실시간 예측 시스템 적용 가능성에 대한 탐색이 필요하며, 이는 효율적인 데이터 처리 및 알고리즘 최적화가 포함된다. 마지막으로, 각 성분에 최적화된 알고리즘을 자동으로 선택 및 조합하는 시스템을 개발하여 다양한 데이터셋에 대한 최적 성능을 가지는 모델을 구축할 필요가 존재하며, 이를 위해 하이퍼파라미터 최적화 및 AutoML 기술을 활용 가능하다.

REFERENCES

- [1] Gijssbers, Peter, Marcos L. P. Bueno, Stefan Coors, Erin LeDell, Sebastien Poirier, Janek Thomas, Bernd Bischl, Joaquin Vanschoren. (2024). "Amlb: an automl benchmark." *Journal of Machine Learning Research*, 25-101, 1-65, 2024.

[2] Cleveland, Robert. B., William S. Cleveland, Jean E. McRae, & Irma Terpenning. "STL: A seasonal-trend decomposition." *Journal of Official Statistics*, 6-1, 3-73, 1990.

[3] Box, George. E. P., Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung. "Time series analysis: forecasting and control." John Wiley & Sons, 2015.

[4] Chen, Tiangi, Carlos Guestrin. "XGBoost: A scalable tree boosting system." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794, 2016.

[5] Hochreiter, Sepp, Jurgen Schmidhuber. "Long short-term memory". *Neural Computation*, 9-8, 1735-1780, 1997.

[6] Wang, Haizhong., Liu Liu, Shangjia Dong, Zhen Qian, Heng Wei. "A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD-ARIMA framework." *Transportmetrica B: Transport Dynamics*, 4-3, 159-186, 2016.

[7] Lin, Yu., Yan Yan, Jiali Xu, Ying Liao, Feng Ma. "Forecasting stock index price using the CEEMDAN-LSTM model." *The North American Journal of Economics and Finance*, 57, 101421, 2021.

[8] Liu, Zhuolin, Yan Zhang, Qingqing Huang. "A lifting contrastive learning method for rotating machinery fault diagnosis." *IEEE 3rd Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)*, 547-551, 2023.

[9] Xie, Xuyang, Zichun Yang, Lei Zhang Jintao Wang, Guoqing Zeng, Xuefeng Wang, Guobing Chen. "Fault Diagnosis Method for Rotating Machinery Based on Improved CBR." *Journal of Physics: Conference Series*, 2528-1, 012030, 2023.

[10] Bai, Yihao, Weidong Cheng, Weigang Wen, Yang Liu. (2023). "Application of Time-Frequency Analysis in Rotating Machinery Fault Diagnosis." *Shock and Vibration*, 2023-1, 9878228, 2023.

[11] Ministry of SMEs and Startups of Korea,

Rotating Machine AI Dataset, Korea AI Manufacturing Platform(KAMP), <https://www.kamp-ai.kr/>

[12] Daehee Lee, Se-Won Kim, Jung-Min Yoo, Ye-Won Jeong, Seo-Yoon Chae, Chae-Gyu Lee, Hyun-Seung Choo, Jong-Pil Jeong, "Anomaly detection based on 1D-CNN LSTM for rotating machine." *Proceedings of IIBC Conference*, pp. 4-6, 2023.

저자약력

한 용 희 (Yong Hee Han)

[정회원]



- 한양대학교 산업공학과 (학사)
- 조지아공대 산업공학과 (석사)
- 조지아공대 산업공학과 (박사)
- (현)충실대학교 벤처중소기업학과 교수

〈관심분야〉 머신러닝, 딥러닝, 스마트팩토리

고 방 원 (Bangwon Ko)

[정회원]



- 서울대학교 수학과 (학사)
- 서울대학교 통계학과 (석사)
- 아이오와 주립대 통계학과 (박사)
- (현)충실대학교 정보통계·보험수리학과 교수

〈관심분야〉 보험수리, 금융공학, 머신러닝