

Deep Learning-Based Stock Fluctuation Prediction According to Overseas Indices and Trading Trend by Investors

Tae Seung Kim[†] · Soowon Lee^{††}

ABSTRACT

Stock price prediction is a subject of research in various fields such as economy, statistics, computer engineering, etc. In recent years, researches on predicting the movement of stock prices by learning artificial intelligence models from various indicators such as basic indicators and technical indicators have become active. This study proposes a deep learning model that predicts the ups and downs of KOSPI from overseas indices such as S&P500, past KOSPI indices, and trading trends by KOSPI investors. The proposed model extracts a latent variable using a stacked auto-encoder to predict stock price fluctuations, and predicts the fluctuation of the closing price compared to the market price of the day by learning an LSTM suitable for learning time series data from the extracted latent variable to decide to buy or sell based on the value. As a result of comparing the returns and prediction accuracy of the proposed model and the comparative models, the proposed model showed better performance than the comparative models.

Keywords : Stock Price Fluctuation Prediction, Deep Learning, Overseas Indices, Trading Trends by Investor

해외지수와 투자자별 매매 동향에 따른 딥러닝 기반 주가 등락 예측

김 태 승[†] · 이 수 원^{††}

요 약

주가 예측은 경제, 통계, 컴퓨터 공학 등 여러 분야에서 연구되는 주제이며, 특히 최근에는 기본적 지표나 기술적 지표 등 다양한 지표로부터 인공지능 모델을 학습하여 주가의 변동을 예측하는 연구들이 활발해 지고 있다. 본 연구에서는 S&P500 등의 해외지수, 과거 KOSPI 지수, 그리고 KOSPI 투자자별 매매 동향으로부터 KOSPI의 등락을 예측하는 딥러닝 모델을 제안한다. 제안 모델은 주가 등락 예측을 위하여 비지도 학습 방법인 적층 오토인코더를 이용하여 잠재변수를 추출하고, 추출된 잠재변수로부터 시계열 데이터 학습에 적합한 LSTM 모델로 학습하여 당일 시가 대비 증가의 등락을 예측하며, 예측된 값을 기반으로 매수 또는 매도를 결정한다. 본 연구에서 제안하는 모델과 비교 모델들의 수익률 및 예측 정확도를 비교한 결과 제안 모델이 비교 모델들 보다 우수한 성능을 보였다.

키워드 : 주가 등락 예측, 딥러닝, 해외 지수, 투자자별 매매 동향

1. 서 론

주식은 기업이 주주의 출자에 대하여 교부하는 유가 증권이자 자본 국가의 핵심적인 역할을 한다. 주가 예측에 대한 연구는 통계, 경제, 컴퓨터 공학 등 여러 분야에서 수행되고 있으며, 특히 최근에는 다양한 지표로부터 인공지능 모델을 학

습하여 주가의 변동을 예측하는 연구들이 활발해 지고 있다.

기계학습에 사용되는 지표를 기준으로 이들 연구들을 구분하면 캔들 차트 및 보조 지표 등의 기술적 지표(Technical Indicator)로부터 인공지능 모델을 학습하여 주가를 예측하는 연구[1-11]들과 재무제표 등의 기본적 지표(Fundamental Indicator)로부터 인공지능 모델을 학습하여 주가를 예측하는 연구들이 수행되었다[12,13]. 또한, 연구의 목적에 따라 이들 연구들을 구분하면 국가별 주가 지수나 개별 종목에 대한 주가 등락을 예측하는 연구[2,8-11,14-16]와 주가 등락 예측결과를 기반으로 수익률을 평가하는 연구[11] 등이 있다.

본 연구에서는 S&P500 등의 해외지수, 과거 KOSPI (Korea Composite Stock Price Index) 지수, 그리고 KOSPI 투자자

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터 지원사업의 연구결과로 수행되었음(IITP-2021-2018-0-01419).

† 준 회 원 : 릿지트레이딩그룹 연구원

†† 정 회 원 : 송실대학교 소프트웨어학부 교수

Manuscript Received : February 15, 2021

First Revision : April 20, 2021

Accepted : May 24, 2021

* Corresponding Author : Soowon Lee(swlee@ssu.ac.kr)

별 매매 동향으로부터 KOSPI의 등락을 예측하는 딥러닝 모델을 제안한다. 제안 모델은 당일 KOSPI 시가 대비 증가 등락을 당일 장전 동시호가 전에 예측하여 주식 매매를 결정한다. 증가 등락 예측을 위하여 S&P500의 캔들 차트 데이터와 KOSPI 투자자별 순매매 동향 데이터로부터 적층 오토인코더(Stacked Auto-Encoder; SAE)를 이용하여 증가 등락에 영향을 끼치는 잠재변수를 추출한다. 이는 데이터를 재구성하는 학습과정을 통해 특징을 더욱 효율적으로 표현하는 데이터를 추출하기 위함이다. 추출된 잠재변수로부터 시계열 데이터 학습에 적합한 장단기 메모리 네트워크(Long Short-Term Memory; LSTM)를 학습하여 당일 시가 대비 증가의 등락을 예측하며, 예측된 값을 기반으로 매수 또는 매도를 결정한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구들을 소개하며, 3장에서는 본 연구에서 제안하는 증가 등락 예측 모델을 설명한다. 4장에서는 제안 모델의 성능을 다른 모델들과 실험적으로 비교, 분석하고, 마지막 5장에서는 결론과 향후 연구 방향을 기술한다.

2. 관련 연구

인공지능 기반의 증가 등락 예측 연구는 크게 지도학습 기반의 연구와 비지도학습 기반의 연구로 구분된다.

지도학습 기반의 연구 중 기본적인 지표를 이용한 증가 예측 연구는 다음과 같다. 이윤환[12]은 기업 재무제표를 기반으로 순환 신경망 모델과 다중 퍼셉트론 모델을 학습하여 주가가 상승할 종목을 예측한 후 장기적으로 '매수 후 보유' 전략을 사용했을 때의 수익률을 평가하였다. 허준영[13]은 기업의 분기별 재무정보를 입력으로 하여 KOSPI200에 속한 기업의 분기 마지막일로부터 1~2개월 후의 증가를 SVM를 이용하여 예측하였다.

지도학습 기반의 연구 중 기술적 지표를 이용한 증가 예측 연구로, Siou Jhjh Guo[5]는 캔들 차트 데이터로부터 CNN을 학습하여 다음날 증가의 변동을 예측하는 Deep Candlestick

Predictor 프레임워크를 제안하였다. Rosdyana Mangir Irawan Kusuma[4]는 캔들 차트 데이터로부터 CNN을 학습하여 대만과 인도네시아 60개 상장사의 증가 등락을 예측하였다. 이우식[8]은 KOSPI의 상대강도지수, 모멘텀, 변화율지표 등 기술적 분석 지표들로부터 Decision Tree, SVM, Deep Learning 모델을 학습하여 KOSPI 일별 증가의 방향성 예측 성능을 비교 평가하였다. Chongda Liu[9]는 S&P500의 기술적 분석 지표들과 유가, 다른 나라의 종합 증가 지수 등으로부터 Logistic Regression, Naive Bayes 등의 모델을 학습하여 S&P500의 증가 등락 예측 정확도를 비교하였다.

기본적 분석 지표와 기술적 분석 지표를 함께 이용한 연구로 김성근[17]은 46개 상장 기업을 표본으로 선정한 후 자기 자본비용, 시가총액 증가율, 주식거래량 증가율 등 16개의 변수로부터 인공신경망을 학습하여 수익률을 예측하였다.

비지도 학습을 이용한 증가 예측 연구로 Guosheng Hu[1]는 Convolutional AutoEncoder를 통하여 512차원의 특징으로 표현된 캔들 차트를 군집화한 후 각 군집의 샵프 비율을 최적화하는 방법으로 포트폴리오를 구성하는 방법을 제시하였다. 샵프 비율은 투자성과를 평가함에 있어서 해당 투자의 위험을 조정하여 반영하는 지표이다. Seksan Sangsawad[18]는 증가 변수 OHLC(Open-High-Low-Close) 데이터에 K-Means 알고리즘을 적용하여 12비트의 캔들 차트 특징을 추출한 후 추출된 정보를 기반으로 차트를 재구성하여 실제 캔들 차트와의 유사도를 비교하였다.

지도 학습과 비지도 학습을 함께 이용한 증가 예측 연구로 Wei Bao[13]는 WSAEs(Wavelet Transforms + Stacked Auto-Encoder)를 사용하여 입력 데이터로부터 잠재 변수를 추출한 후 지도 학습을 이용하여 경제수준이 각기 다른 나라들의 증가 지수 동향을 예측하는 모델을 제시하였다.

3. 제안 모델

제안 모델의 구조도는 Fig. 1과 같다. 제안 모델은 파생변

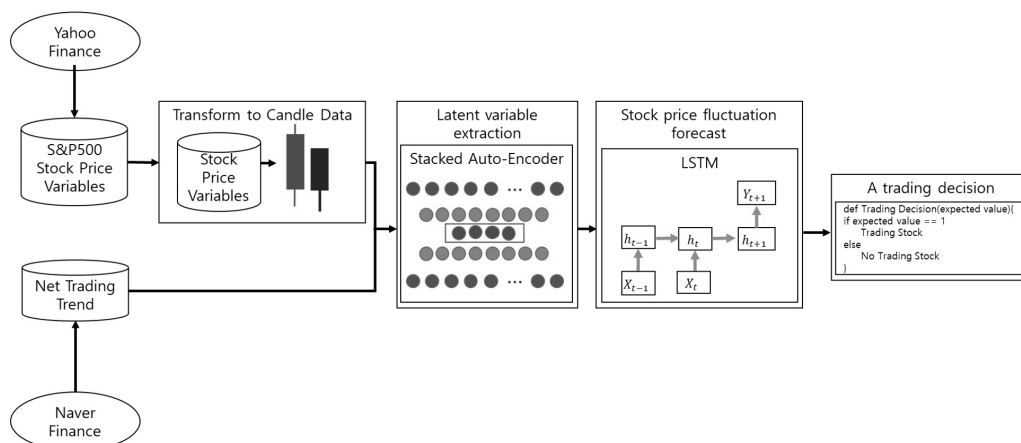


Fig. 1. Structure of the Proposed Model

수 생성 단계, 잠재변수 추출 단계, 주가 등락 예측 단계 및 매매 결정의 네 단계로 구성되어 있다. 파생변수 생성(Derived Variable Generation) 단계에서는 과거 주가 데이터로부터 캔들(Candle) 관련 파생변수를 생성하고 이를 투자자별 순매매 동향 데이터(Net Trading Trend)와 통합하여 입력 데이터를 생성한다. 잠재변수 추출(Latent Variable Extraction) 단계에서는 입력 데이터로부터 수익률 예측에 영향을 미치는 잠재변수를 추출한다. 주가 등락 예측(Stock Price Fluctuation Prediction) 단계에서는 추출된 잠재변수를 기반으로 주가의 등락을 예측하고 마지막으로 매매 결정(Trading Decision) 단계에서는 전단계의 예측값을 기반으로 매수 또는 매도를 결정한다.

3.1 파생변수 생성

캔들 차트는 주가의 변동을 시각적으로 분석하는데 사용되는 지표이다. 파생변수 생성 단계에서는 수집된 과거 주가 변수로부터 캔들 차트의 정보를 표현하는 파생변수의 값을 추출한다. 캔들은 몸통(Body), 윗꼬리(Upper Tail), 아랫꼬리(Lower Tail)로 구성되어 있으며, 각 부분은 OHLC (Open-High-Low-Close) 변수(시가($Open_t$), 고가($High_t$), 저가(Low_t), 종가($Close_t$))로부터 계산된다. Table 1은 OHLC 변수로부터 파생된 캔들 관련 변수의 정의이다.

본 연구에서는 Table 1의 캔들 정보와 별개로 ($Open_t - Close_{t-1}$)로 계산되는 '전일종가 대비 당일시가' 변수를 추가로 사용한다.

3.2 잠재변수 추출

잠재변수 추출 단계는 Table 1에 정의된 변수들로부터 비지도 학습 방법 중 하나인 적층 오토인코더(SAE)를 사용하여 수익률 예측에 영향을 미치는 잠재변수를 추출한다. 잠재변수 추

Table 1. Derived Variables for Candles

Derived Variables	Definition
$Body_t$	$Close_t - Open_t$
$Fluctuation_t$	$High_t - Low_t$
$Upper Tail_t$	$High_t - Close_t$ (Bullish Candle)
	$High_t - Open_t$ (Bearish Candle)
$Lower Tail_t$	$Open_t - Low_t$ (Bullish Candle)
	$Close_t - Low_t$ (Bearish Candle)

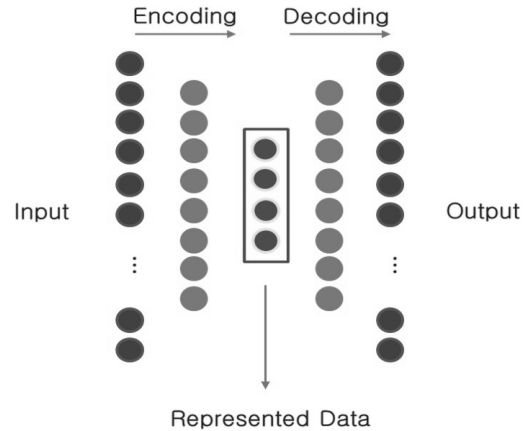


Fig. 2. Structure of SAE for Extracting Latent Variables

출 단계에서 사용된 적층 오토인코더의 구조는 Fig. 2와 같다.

잠재변수 추출단계에서 사용된 적층 오토인코더는 인코더와 디코더 층으로 구성되며, 입력값과 출력값 오차의 최소화를 목표로 하여 모델을 학습한다. 제안 모델에서는 수익률 예측에 필요한 잠재변수를 표현하기 위하여 적층 오토인코더의 파라미터를 최적화하는 과정과 압축된 값으로 학습한 주가 등락 예측 모듈의 수익률을 평가하는 과정을 반복하며 파라미터를 최적화한다. 파라미터 최적화의 상세 내용은 4장에서 기술한다.

3.3 주가 등락 예측

주가 등락 예측 단계에서는 잠재변수 추출 단계를 통하여 재구성된 데이터가 입력으로 주어졌을 때 LSTM을 이용하여 당일 시가 대비 당일 종가 등락을 예측한다. 이 단계에서 사용되는 LSTM은 당일 시가 대비 당일 종가 등락 여부를 Target으로 하여 Loss Function을 최소화하는 방식으로 학습된다. 입력 데이터는 시퀀스 정보를 학습하기 위하여 Time Step의 형태로 구성된다. Time Step은 과거 n 일의 데이터를 사용하여 Target을 예측할지 결정한다. 주가 등락 예측 단계의 LSTM은 단층의 은닉층으로 구성되어 있으며 도식화된 LSTM 구조는 Fig. 3과 같다.

3.4 매매 결정

매매 결정 단계에서는 주가 등락 예측 결과를 기반으로 매수 또는 매도를 결정한다. 매매 결정을 위해서는 다양한 전략

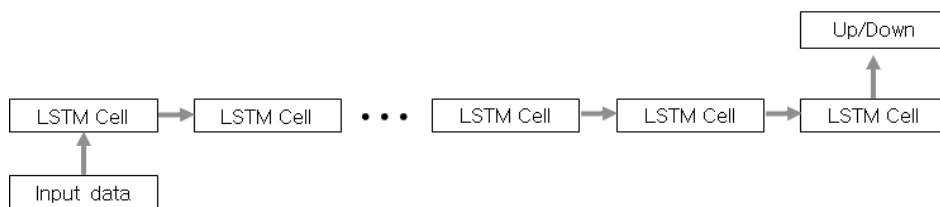


Fig. 3. Structure of LSTM for Stock Price Fluctuation Prediction

Table 2. Buy/Sell Strategy by Prediction Result

Prediction result	Buy/Sell Strategy
Up	Buy Opening Price, Sell Closing Price
Down	No Trading

이 사용될 수 있으나, 본 연구에서는 앞 단계까지의 주가 등락 예측 정확도 향상에 초점을 맞추기 위하여 Table 2와 같은 단순화된 매매 전략을 사용한다. Table 2의 전략은 장전 예측 결과가 '상승'이면 당일 시가에 매수하고 증가에 매도하며 '하락'이면 매수/매도를 하지 않는 전략이다.

4. 실험 및 결과

4.1 실험 데이터

제안 모델의 성능 평가를 위하여 본 연구에서는 2014년 4월 21일부터 2019년 4월 20일까지 5년간의 일별 S&P500과 KODEX200의 주가 데이터와 및 KOSPI 투자자별 매매 동향 데이터를 수집하였다. 수집된 주가 데이터는 S&P500과 KODEX200의 OHLC(Open-High-Low-Close) 데이터이며 Table 3은 수집된 주가 데이터에 대한 설명이다.

투자자별 매매 동향 데이터는 KOSPI에 포함된 종목들에 대한 투자자별 일단위 순매매 금액(단위: 억)으로 투자자는 개인, 외국인, 은행, 연기금 등 10개의 투자주체로 구분되어 있다. 본 데이터는 네이버 금융의 투자자별 순매매 동향 탭으로부터 수집하였다.

본 연구에서는 해외 지수 S&P500 데이터와 국내지수 KODEX200 데이터 및 KOSPI 투자자별 매매 동향 데이터를 통합하는 과정에서 발생하는 양국 증시 개장일의 차이로 인한 일관성(Consistency) 문제를 해결하기 위하여 미국과 한국의 증권시장이 모두 개장한 날의 정보만을 사용하였다.

제안 모델의 평가를 위하여 수집된 데이터들을 Training Set, Validation Set, 그리고 Test Set으로 분할하였다. Training Set은 모델의 학습에 사용되고, Validation Set은 모델의 검증에 사용되며, Test Set은 최종적으로 모델의 성능 평가에 사용되는 데이터이다. Table 4는 각 데이터들에 대한 설명이다.

Table 3. Description of Collected Stock Price Data

Data	Description	Source
S&P500	Index including stocks of 500 large corporations owned and managed by US Standard & Poor's, an international credit rating agency	Yahoo Finance
KODEX 200	ETF to track Korea's stock index KOSPI 200. The largest market cap and trading volume among ETFs, operated by Samsung Asset Management	Investing.com

Table 4. Description of the Collected Data

Items	Terms	# of Days
Training Set	2014.04.21 ~ 2017.04.20	716
Validation Set	2017.04.21 ~ 2018.04.20	236
Test Set	2018.04.21 ~ 2019.04.20	233

Table 5. Model Prediction Based Marketing Strategy

Predicted	Strategy	Actual	Change in Assets
Up	Buy Opening Price, Sell Closing Price	Up	Investment amount * fluctuation rate (Revenue)
		Down	Investment amount * fluctuation rate (Loss)
Down	No Trading Stock	Up	No change
		Down	No change

4.2 평가 방법

본 연구에서는 제안하는 모델의 성능을 평가를 위하여 Table 2에서 기술한 매매 전략에 따라 매매를 하였을 때의 수익률을 평가하였다. Table 5는 예측결과에 따른 수익률 계산 과정이다. KODEX200의 경우 매매차익에 대해 비과세가 적용되고 국내 주식에 대한 수수료가 무료인 계좌가 존재하므로 각 매매에 대한 수수료와 세금은 고려하지 않았다.

본 연구에서는 비교 모델로 사용할 통계 모델 생성을 위하여 Scikit-Learn 라이브러리를, 딥러닝 모델 생성을 위하여 Tensorflow 라이브러리를 사용하였다.

4.3 파라미터 최적화 실험 결과

잠재변수 추출 단계에서 SAE 파라미터 최적화를 위하여 첫 번째와 세 번째 은닉층의 노드 수, 그리고 잠재변수로 사용될 두 번째 은닉층의 노드 수를 변경하며 실험을 진행하였다. SAE의 특성상 입력과 출력의 크기가 동일하므로 첫 번째와 세 번째의 은닉층의 노드 수는 동일하다. 데이터 압축이 목적이기 때문에 두 번째 은닉층의 노드 수는 첫 번째, 세 번째 은닉층의 노드 수보다 같거나 작도록 설정하였다. Fig. 4는 SAE의 노드 수 변경에 따른 제안 모델의 수익률(Yield) 변화이다.

실험 결과에 따르면 SAE의 첫 번째, 세 번째 은닉층의 노드 수가 8, 두 번째 은닉층의 노드 수가 4일 때의 수익률이 7.53으로 가장 높은 수치를 기록하였다. Table 6은 본 실험에서 SAE의 최적 파라미터이다.

Fig. 5는 주가 등락 예측 단계에서 LSTM의 Layer 수에 따른 제안 모델의 수익률 변화를 보여준다. Fig. 5에 따르면 Layer가 단층일 때 수익률이 7.53으로 가장 높은 것으로 평가되었다. 이는 이미 압축된 데이터의 특성을 학습하는데 복잡한 구조의 신경망이 필요하지 않기 때문인 것으로 사료된다.

Fig. 6은 LSTM의 Time Step에 따른 제안 모델의 수익률 변화를 보여준다. LSTM의 Time Step이 2일 때 가장 높은 수익률(7.53)을 보였다. Layer 수에 따른 수익률 변화를 측

Table 6. Optimal Parameters for SAE

Parameter	Value
Learning Rate	0.05
Epoch	100
Batch Size	32
# of 1st, 3rd Layer's Hidden Units	8
# of 2nd Layer's Hidden Units	4

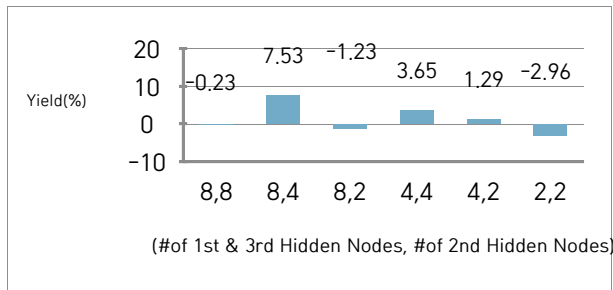


Fig. 4. Yield Change by # of Nodes in SAE

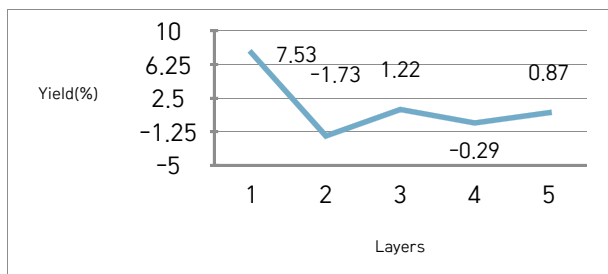


Fig. 5. Yield Change by # of Layer in LSTM

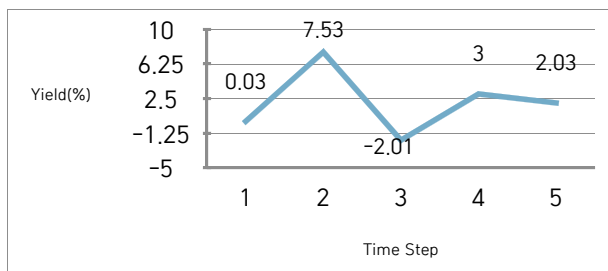


Fig. 6. Yield Change by # of Time Steps in LSTM

Table 7. Optimal Parameters for LSTM

Parameter	Numbers
Learning Rate	0.001
Epoch	10
Batch Size	32
Hidden Layer	1
Time Step	2
Hidden Unit	16

정할 때는 Time Step을 2로, Time Step에 따른 수익률 변화를 측정할 때는 Layer 수를 1로 고정하였다. Table 7은 LSTM의 최적 파라미터이다.

4.4 모델별 수익률 비교 실험 결과

Fig. 7은 본 연구에서 제안하는 모델과 비교 모델들과의 수익률을 비교한 것이다. 수익률 비교는 모의실험을 통해서 수행하였으며, 장중에 매매하는 것이 아니라 동시호가에만 전량 매수 및 매도하는 것으로 가정하였으므로 단기 매매에 중요한 거래량, 슬리피지 등의 요소들은 고려하지 않았다. Fig. 7의 B1, B2, B3, B4는 모두 BaseLine을 의미한다. B1은 KODEX200 지수에 대해 Test Set의 첫날인 2018년 4월 21일 시가에 매수하고 마지막 날인 2019년 4월 20일 종가에 매도하였을 때의 수익률이며, B2는 매일 시가에 매수하고 종가에 매도하였을 때의 수익률이다. B3는 매일 증가에 매수하고 익일 시가에 매도하였을 때의 수익률이며, B4는 전날 S&P500의 시가 대비 증가가 상승인 경우에만 KODEX200을 시가에 매수하고 증가에 매도하였을 때의 수익률이다. Test Set의 기간 1년 동안에는 주가가 -9.96% 하락하는 구간이므로 B1 값이 -9.96이며 B3(+2.98)을 제외한 B2, B4에서 모두 음의 수익률이 평가되었다.

비교 모델인 Decision Tree(DT)와 Logistic Regression (LR)의 수익률은 B1, B2, B4 보다는 높지만 B3보다는 낮은 -2.07과 -3.61로 평가되었다. 제안 모델은 SAE와 LSTM을 결합한 모델로 다른 신경망 기반 비교 모델 (MLP, LSTM, PCA*LSTM, SAE*MLP)에 비해서 가장 높은 수익률(8.17)을 보였다. 그 이유는 잠재변수 추출 단계에서 입력데이터가 가지는 조합의 영향력을 잘 추출했기 때문인 것으로 사료된다.

4.5 모델별 예측 정확도 비교 실험 결과

Fig. 8은 모델별 예측 정확도(Accuracy)를 비교한 것이다

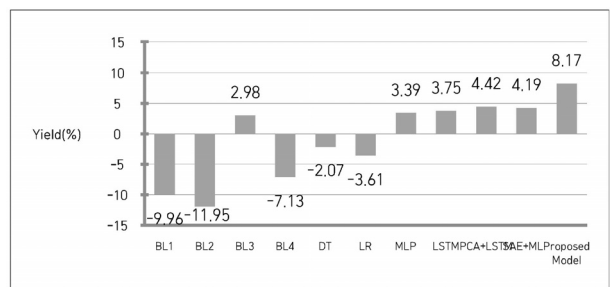


Fig. 7. Comparison of Yields by Prediction Models

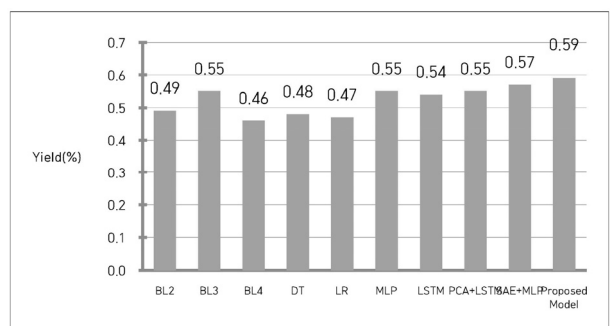


Fig. 8. Comparison of Accuracy by Precision Models

다. B2의 예측 정확도는 0.49로, 이는 당일 시가 대비 증가의 상승일과 하락일 비율이 거의 비슷함을 의미한다. B3의 정확도는 0.55로, 이는 당일 증가 대비 익일 시가의 상승일이 하락일 보다 더 많다는 것을 의미한다. B4은 다른 비교모델들에 비해 낮은 정확도(0.46)를 보였다. 이는 전일 S&P500의 시가 대비 증가 등락 여부만으로는 KOSPI200의 등락 예측이 어렵다는 것을 의미한다. Fig. 8에 따르면 제안 모델이 비교 모델 대비 가장 높은 정확도(0.59)를 보였다.

각 모델들의 정확도와 수익률 순위를 비교하면, MLP와 LSTM의 경우 수익률은 각각 3.39, 및 3.75로 LSTM이 약간 높지만 정확도는 MLP 0.55, LSTM 0.54로 MLP가 높은 것으로 평가되었다. 또한, PCA+LSTM과 MLP의 정확도는 0.55로 거의 동일하지만 수익률은 MLP 3.39, PCA+LSTM 4.42로 PCA+LSTM이 높은 것으로 평가되었다. 이는 더 높은 정확도가 반드시 더 높은 수익률을 의미하지 않는다는 것을 의미하며, 상승 하락의 폭에 따라 수익률이 영향을 받기 때문인 것으로 해석된다.

Table 8은 제안 모델의 예측 결과에 대한 혼동행렬 (Confusion Matrix)이다. 제안 모델은 실제 상승과 하락의 횟수가 115회 및 118회로 거의 비슷한 것에 비해 상승 예측 횟수 136회, 하락 예측 횟수 97회로 하락 예측보다 상승 예측을 더 많이 한 것으로 평가되었으며, Precision=0.57, Recall=0.68, F1-score=0.62로 산출되었다.

Table 9는 제안 모델의 수익률을 월별로 분석한 것이다. Test 기간 동안의 일 등락률의 월별 누적값(Cumulative Daily Fluctuation)과 월 등락률(Monthly Fluctuation), 그리고 같은 기간 제안 모델의 월 수익률(Monthly Yield)이 비교되어 있다. Table 9의 예측 성공 평균(Average^{Up}), 예

Table 8. Confusion Matrix for the Proposed Model

Actual \ Prediction	Rise	Decline	Total
	Rise	78	37
Decline	58	60	118
Total	136	97	233

측 실패 평균(Average^{Down})은 각각 예측을 성공하였을 때, 실패하였을 때의 등락률의 평균이다.

일 등락률의 월별 누적값과 월 등락률을 비교하였을 때 대부분의 달에서 거의 비슷한 값을 보이는 것으로 평가되었다. 그러나 12월에는 일 등락률 월별 누적값은-1.39, 월 등락률은 5.16으로 큰 차이를 보이는데 이는 12월에는 당일 증가 대비 익일 시가의 상승(갭 상승)이 많았던 것이 그 이유인 것으로 추측된다.

전체 기간 중 상승 횟수는 115회, 하락 횟수는 118회로 월 등락률 누적값은 -6.29이지만 제안 모델은 총 136회의 상승 예측에서 78회의 상승 예측에 성공하여(약 57%) +8.26의 누적 수익률(Monthly Yield의 Total)을 기록하였다. 특히 9월의 경우 상승 7회, 하락 12회로 월 등락률이-6.89이지만 제안 모델에서는 10회의 상승 예측에서 6회의 성공, 9회의 하락예측에서 8회 성공하여 +1.08의 누적 수익률을 기록하였다. Table 9와 Fig. 7의 수익률이 서로 다른 이유는 매 거래 시 발생하는 수익이나 손실을 투자금 변화에 적용하였는지의 여부이다. Fig. 7에서 측정된 수익률의 경우 이전 거래의 수익이나 손실에 따른 투자금의 변화를 다음 거래의 투자금에 적용한 경우이고, Table 9에서 측정된 수익률의 경우 이전 거래의 수익이나 손실에 따른 투자금의 변화를 적용하

Table 9. Monthly Yield Comparison for the Proposed Model

Month	4	5	6	7	8	9	10	11	12	1	2	3	Total
# Up	8	8	9	11	11	7	9	10	10	11	5	16	115
Average ^{Up}	0.32	0.49	0.61	0.24	0.37	0.69	0.88	0.55	0.44	0.75	0.53	0.47	6.34
# Down	11	11	10	8	8	12	10	9	9	8	14	8	118
Average ^{Down}	-0.47	-0.83	-0.56	-0.62	-0.29	-0.68	-0.91	-0.67	-0.64	-0.55	-0.49	-0.31	-7.02
Cumulative Daily Fluctuation	-2.62	-5.22	-0.10	-2.32	1.70	-3.32	-1.23	-0.46	-1.39	3.85	-4.18	5.01	-10.28
Monthly Fluctuation	-0.45	-3.87	-1.97	-2.75	3.65	-6.89	-2.67	-0.41	5.16	4.45	-2.45	1.91	-6.29
Predicted Up (Proposed Model)													
# Success	6	5	4	7	6	6	8	8	8	6	3	11	78
Average ^{Success}	0.27	0.60	0.85	0.24	0.37	0.79	0.97	0.63	0.46	0.77	0.69	0.54	7.18
# Failure	5	6	5	3	3	4	6	4	6	4	8	4	58
Average ^{Failure}	-0.46	-0.77	-0.45	-0.80	-0.37	-0.91	-0.93	-0.76	-0.81	-0.57	-0.49	-0.37	-7.69
Predicted Down (Proposed Model)													
# Success	6	5	5	5	5	8	4	5	3	4	6	4	60
Average ^{Success}	-0.49	-0.90	-0.66	-0.51	-0.24	-0.57	-0.89	-0.59	-0.31	-0.54	-0.48	-0.24	-6.42
# Failure	2	3	5	4	5	1	1	2	2	5	2	5	37
Average ^{Failure}	0.48	0.31	0.42	0.24	0.37	0.12	0.17	0.24	0.35	0.73	0.29	0.29	4.01
Monthly Yield	-0.66	-1.65	1.12	-0.70	1.09	1.08	2.16	2.00	-1.14	2.34	-1.86	4.48	8.26

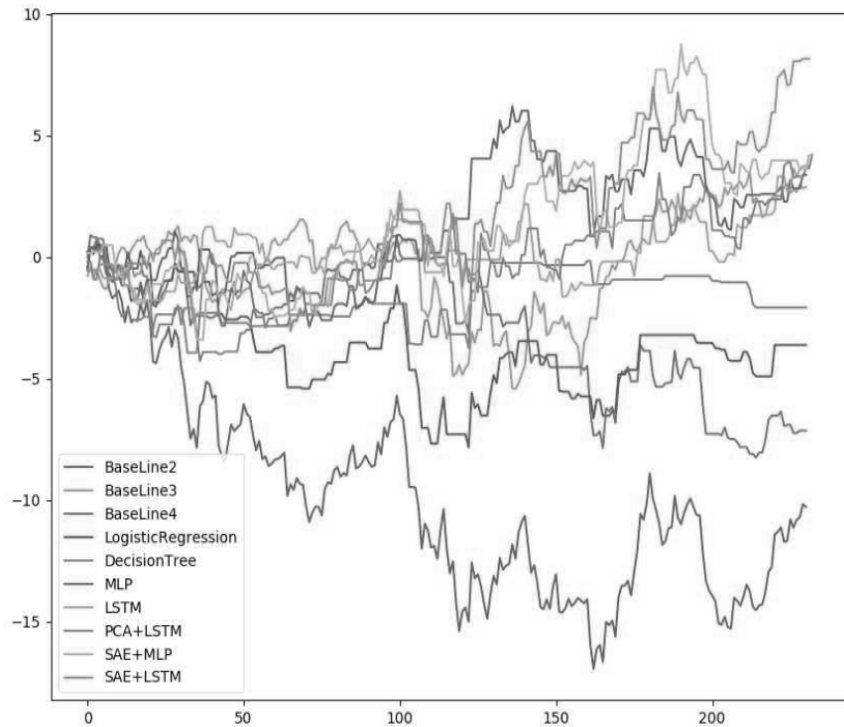


Fig. 9. Cumulative Yield Changes over Testing Period

지 않고 수익률의 누적값을 측정한 경우이므로 차이가 있다.

Fig. 9는 Test 기간에서의 누적 수익률 변화이다. Test 기간의 후반부로 갈수록 제안 모델인 SAE+LSTM과 비교 모델들과의 누적 수익률의 차이가 벌어지는 것으로 평가되었다.

5. 결론 및 향후 연구

본 연구에서는 KOSPI200의 당일 시가 대비 종가의 등락을 당일 장전 동시호가 전에 예측하여 매매를 결정하는 딥러닝 기반의 새로운 예측 모델을 제안하였다. 제안 모델은 해외지수, 과거 KOSPI 지수, 그리고 KOSPI 투자자별 매매 동향으로부터 SAE를 이용하여 잠재변수를 추출한 후 LSTM을 통하여 주가의 등락을 예측하고, 예측 결과를 기반으로 매매를 결정한다. 5년간의 일별 KOSPI 데이터로 제안 모델의 타당성을 실험적으로 평가한 결과 제안 모델이 비교 모델보다 수익률과 예측 정확도 측면에서 우수한 것으로 평가되었다.

제안 모델은 입력데이터의 구성 및 잠재변수를 추출하고 주가 등락을 예측하는 방법에 있어서 독창성을 가지고 있다. 한국거래소 등으로부터 더욱 다양한 종류의 데이터가 확보된다면 심층적인 학습을 통하여 예측 정확도의 향상을 기대할 수 있다. 또한, 예측 결과에 대하여 본 연구에서는 상승 예측 시에만 매수하는 매우 단순한 매매 전략을 사용하였으나 인버스 ETF 활용 등 매매 전략의 변화가 상승장 또는 하락장에서의 수익률에 어떻게 영향을 미치는지에 대한 추가적인 연구도 필요하다.

References

- [1] G. Hu, et al., "Deep stock representation learning: From candlestick charts to investment decisions," *arXiv preprint arXiv:1709.03803*, 2018.
- [2] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," *Expert Systems with Applications*, Vol.42, No.4, pp.2162-2172, 2015.
- [3] R. Singh, and S. Srivastava, "Stock prediction using deep learning," *Multimedia Tools and Applications*, Vol.76, No.18, pp.18569-18584, 2016.
- [4] R. M. I. Kusuma, T. T. Ho, W. C. Kao, Y. Y. Ou, and K. L. Hua, "Using deep learning neural networks and candlestick chart representation to predict stock market," *arXiv:1903.12258v1*. 2019.
- [5] S. J. Guo, F. C. Hsu, and C. C. Hung, "Deep candlestick predictor: A framework toward forecasting the price movement from candlestick charts," In *Proceedings of 9th International Symposium on Parallel Architectures, Algorithms and Programming*, pp.219-226, 2018.
- [6] T. Kamo, and C. Dagli, "Hybrid approach to the Japanese candlestick method for financial forecasting," *Expert Systems with Applications*, Vol.36, No.3, pp.5023-5030. 2009.

- [7] D. H. Shin., K. H. Choi, and C. B. Kim, "Deep learning model for prediction rate improvement of stock price using RNN and LSTM," *Journal of Korean Institute of Information Technology*, Vol.15, No.10, pp.9-16, 2017.
- [8] W. S. Lee, "A deep learning analysis of the KOSPI's directions," *Journal of the Korean Data & Information Science Society*, Vol.28, No.2, pp.287-295, 2017.
- [9] C. Liu, J. Wang, D. Xiao, and Q. Liang, "Forecasting S&P 500 stock index using statistical learning models," *Open Journal of Statistics*, Vol.6, No.6, pp.1067-1075, 2016.
- [10] H. Y. Lee, "A combination model of multiple artificial intelligence techniques based on genetic algorithms for the prediction of Korean Stock Price Index(KOSPI)," *Entrue Journal of Information Technology*, Vol.7, No.2, pp.33-43, 2008.
- [11] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and longshort term memory," *PloS One*, Vol.12, No.7, e0180944, 2017.
- [12] Y. H. Lee, "Value investing using deep neural networks," Master's Thesis, Seoul National University, Korea, 2018.
- [13] J. Y. Heo, and J. Y. Yang, "SVM based stock price forecasting using financial statements," *Journal of KIISE Transactions on Computing Practices*, Vol.21, No.3, pp.167-172, 2015.
- [14] B. Liao, C. Ma, L. Xiao, R. Lu, and L. Ding, "An arctan-activated WASD neural network approach to the prediction of dow jones industrial average," In *Preceedings of the 14th International Symposium on Neural Networks*, pp.120-126, 2017.
- [15] S. A. Hamid, and Z. Iqbal, "Using neural networks for forecasting volatility of S&P 500 Index futures prices," *Journal of Business Research*, Vol.57, No.10, pp.1116-1125, 2004.
- [16] G. J. Kim, "Stock fluctuation prediction based on news sentiment analysis and time series prediction," Master's Thesis, Graduate School of Soongsil University, Korea, 2017.
- [17] S. K. Kim, "Stock price prediction neural network system based on financial ratio," In *Proceedings of the Korea Intelligent Information System Society*, Vol.1, pp.293-307, 1994.
- [18] S. Sangsawad, and C. C. Fung, "Extracting significant features based on candlestick patterns using unsupervised approach," In *Preceedings of 2th International Conference on Information Technology*, pp.1-5, 2017.



김 태 승

<https://orcid.org/0000-0001-8632-0889>
 e-mail : rlaxotmd9960@gmail.com
 2020년 송실대학교 융합소프트웨어학과 (석사)
 2020년~2021년 Cy2Code 연구원
 2021년~현 재 릿지트레이딩그룹 연구원

관심분야 : 인공지능, 추가등락예측



이 수 원

<https://orcid.org/0000-0001-5863-1188>
 e-mail : swlee@ssu.ac.kr
 1982년 서울대학교 계산통계학과(학사)
 1984년 한국과학기술원 전산학과(석사)
 1994년 University of Southern California 전산학과(박사)

2003년~2004년 한국정보과학회 인공지능연구회 분과위원장
 2008년~2009년 한국정보과학회 논문지(SA) 편집위원장
 1995년~현 재 송실대학교 소프트웨어학부 교수
 관심분야 : 데이터사이언스, 인공지능