

장단기 앙상블 모델과 이미지를 활용한 주가예측 향상 알고리즘 : 석유화학기업을 중심으로

방은지[†], 변희용^{**}, 조재민^{***}

Stock Price Prediction Improvement Algorithm Using Long-Short Term Ensemble and Chart Images: Focusing on the Petrochemical Industry

Eun Ji Bang[†], Huiyong Byun^{**}, Jaemin Cho^{***}

ABSTRACT

As the stock market is affected by various circumstances including economic and political variables, predicting the stock market is considered a still open problem. When combined with corporate financial statement data analysis, which is used as fundamental analysis, and technical analysis with a short data generation cycle, there is a problem that the time domain does not match. Our proposed method, LSTE the operating profit and market outlook of a petrochemical company and estimates the sales and operating profit of the company, it was possible to solve the above-mentioned problems and improve the accuracy of stock price prediction. Extensive experiments on real-world stock data show that our method outperforms the 8.58% relative improvements on average w.r.t. accuracy.

Key words: Stock Price Prediction, Deep Learning, Ensemble, Financial Investment, LSTE

1. 서 론

주식시장은 경제 및 정치적 변수를 포함한 다양한 상황에 영향을 받으므로 주식시장을 예측하는 것은 매우 어려운 문제로 여겨지고 있다. 주식 가격은 기업의 내·외부 요인들의 상호관계가 주가 형성에 많은 영향을 주는 가격 결정 메커니즘으로 인해 주가 변동을 설명할 수 없는 경우가 자주 발생하고 있기 때문이다. 일반적으로 주식 가격의 분석 방법은 기본적 분석(Fundamental analysis)과 기술적 분석(Technical analysis)으로 구분할 수 있다[1]. 기본적 분석은 기업의 재무 정보에서 나타나는 기업 내·외부의 가치를 분석하여 현재의 주가와 비교하는 방법이고, 기술

적 분석은 주식시장에서 실제로 거래되는 주식 가격을 기초로 기술적 지표를 도출하여 주식 가치를 분석하는 방법이다[2].

최근 여러 연구에서 심층학습(Deep Learning) 기술의 발달로 온라인 뉴스, 소셜 미디어 등의 비정형 데이터와 정형 데이터를 활용하여 다양한 방법으로 주가 예측 연구를 진행 중이나, 여전히 전통적인 금융기관에서는 기본적 분석과 기술적 분석 방법을 분리하여 주식시장을 분석하고 있다[3,4].

본 연구는 주가 분석을 위하여 장기적으로는 기업의 실적을 기반을 둔 기본적 분석과 단기적으로 영향을 미치는 시장 및 수급 상황, 투자 심리 등을 반영한 기술적 분석을 동시에 분석하고자 하였다.

※ Corresponding Author : Jaemin Cho, Address: (44610) Daehak-ro 93, Nam-gu, Ulsan, Korea, TEL : +82-52-259-2913, E-mail : jaemincho@ulsan.ac.kr
Receipt date : Feb. 10, 2022, Approval date : Feb. 16, 2022
[†] AleconoLab, Inc. (E-mail : eunji1293@gmail.com)

^{**} AleconoLab, Inc. (E-mail : expectbhyy@gmail.com)

^{***} Dept. of Business Administration, University of Ulsan

※ This work was supported by the 2019 Research Fund of University of Ulsan (20190270).

공시 기간이 최소 3개월 이상인 기업의 재무제표의 과거 정보와 데이터 발생주기가 짧은 기술적 분석의 시장 정보를 결합하려면 시간 영역이 일치하지 않는 문제가 있다[5]. 따라서, 본 연구에서는 이러한 정보의 분석 기간을 일치하기 위하여 미래 시장전망과 기업의 영업이익을 간접적으로 측정하는 지표를 개발하였다. 특히, 석유화학업종 중 NCC(Naphtha Cracking Center) 산업의 경우 기업의 미래 시장전망과 영업이익을 간접적으로 측정하는 지표를 추정할 수 있는데, 본 연구에서는 석유화학기업의 영업이익과 시장전망을 예측하여 해당 기업의 매출액과 영업이익을 우회 추정함으로써 앞서 기술한 문제를 해소함과 동시에 주가 예측 정확도를 향상하고자 하였다.

주가 예측을 위하여 기본적 분석과 기술적 분석을 함께 분석하기 위하여 적절한 기초유분 스프레드 및 WTI 가격을 통해 해당 기업의 영업이익을 추정한다. 그리고 조건검색 및 딥러닝 기반 앙상블 방법을 통해 주가의 장단기 예측 및 변곡점 이미지 감지를 활용하여 학습시간 및 정확도를 효과적으로 향상했다. 장단기 예측과 변곡점 감지를 활용하여 그리고 제안한 알고리즘의 성능평가를 위해 기존 과거 데이터로만 활용하여 동일 모델을 실험을 수행하였다. 또한, 앙상블 방법의 유효성을 확인하기 위해 단일모델과 앙상블 방법의 정확도와 누적수익률을 비교하였다. 실험 결과 제안한 방법이 기존의 방법에 비하여 성능이 우수함을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서 본 논문의 기본 이론이 되는 기술적 및 기본적 분석과 딥러닝 모델에 대해 정의한 다음, 3장에서 제안하는 LSTE를 설명한다. 여기에서는 제안한 알고리즘의 전체적인 개요와 전처리법, 단기예측인 조건기반 검색 알고리즘(Rule-based Algorithm, RBA), 장기예측인 앙상블 알고리즘, 그리고 변곡점 탐지 알고리즘 순서로 서술한다. 그리고 4장에서는 제안한 알고리즘으로 실험한 결과를 기존의 알고리즘들과 객관적인 지표로 비교하여 성능을 평가하고 5장에서 결론을 맺는다.

2. 이 론

2.1 석유화학산업 주가의 분석

EMH(Efficient Market Hypothesis)에 따르면 주가는 모든 과거 정보를 반영하기 때문에, 주가 자체

와 거래량만으로도 미래의 가격 움직임을 예측할 수 없다. 즉, 주가의 움직임은 이미 시장의 모든 정보를 반영하고 있으며 이를 이용한 초과수익률은 불가능하다[6,7]. 그러나, EMH가 도입된 이후 연구자들은 EMH가 실제로 적용되는 주식시장은 존재하지 않으며, 시장이 비효율적이며 실제로 주가가 랜덤 워크를 정확히 따르지 않고 오히려 대략적인 랜덤 워크를 따르는 약형에서 준강형의 EMH를 따르고 있다고 주장했다. 따라서 본 연구에서는 국내 주식시장을 약형에서 준강형 효율적시장으로 전제하여 기본적 분석과 기술적 분석을 결합하여 주가를 예측하고자 한다. 일반적인 자본시장의 대상인 약형에서 준강형 효율시장에서는 주가가 단기적으로 시장 상황에 의해 내재가치를 반영하지 못한 잘못된 가격으로 움직이더라도 결국에는 기업의 내재가치를 반영한 가격으로 수렴하기 때문에 기업의 기본적 분석을 통해 초과수익률을 달성할 수 있다[7].

NCC는 수입된 나프타를 분해하여 석유화학의 기초 원료인 에틸렌, 프로필렌 등의 기초유분을 생산하는 시설이다. 장치산업의 특성상 365일 24시간 운영되므로 에틸렌(Ethylene)-납사(Naphtha) 스프레드를 이용하여 영업이익을 우회 추정할 수 있다. 또한, WTI(West Texas Intermediate) 가격은 단기적으로 납사 가격과 밀접하게 연관되어 있으므로 외부 충격에 의한 WTI 가격 하락은 납사 가격 하락으로 이어져 에틸렌-납사 스프레드 가격 상승으로 이어진다.

따라서 본 연구에선 석유화학기업과 NCC 기업들의 재무공시를 통해 주요 사업구조를 파악하여 영업이익을 추정하고 이를 모델에 활용하는 기술적 분석과 기본적 분석 방법을 결합하여 만든 새로운 알고리즘을 제안한다.

2.2 기계학습 및 심층학습의 적용

약형 효율적 시장에서의 주가가 자기 상관관계를 가지고 있다고 가정하면 일반적으로 ARIMA (Auto-regressive Integrated Moving Average) 모델과 같은 전통적인 시계열 모델링 기술을 사용하여 주가를 예측한다[8,9]. ARIMA 모델은 단기예측 측면에서 견고성과 효율성이 강점이지만 대부분의 ARIMA 모델은 일변량이며 이러한 ARIMA 모델과 관련된 데이터는 일반적으로 과거 시계열 주가 자체이다. 예를 들어, 일일 예측을 수행하려는 경우 일일 고가, 저가,

시가, 종가 및 거래량 정보를 잃게 된다. 이는 시장참여자들이 수요-공급 관점에서 추세를 예측하기 위해 사용하는 이 주가 정보에서 직접 파생된 기술적 지표를 활용할 수 없음을 의미한다. 또한, 기계학습이 대중화됨에 따라 일부 연구자들은 기계학습 알고리즘의 유연성을 활용하여 기술 지표를 모델에 활용한다. 예를 들어 방향 예측을 위해 RF(Random Forest) 또는 LSTM(Long Short-Term Memory) 모델에 제공할 관련 기술 지표 목록을 생성할 수 있다. 그러나, 일반적으로 시계열 데이터를 기계학습에 사용할 때 데이터의 발생주기가 달라 사용 가능한 데이터가 제한돼 사용 가능한 데이터가 충분하지 않는다는 문제가 있다[10]. 제안한 LSTE 모델은 다음과 같은 방법으로 서술한 문제를 해결하였다. 동일 모델에 대한 초매개변수(Hyper-Parameter)에 따른 예측값의 차이를 최소화하기 위해, 여러 매개변수를 그리드 검색(Grid Search)을 통해 해당 매개변수를 저장 후 앙상블을 적용했다. 또한, 다양한 모델들을 합쳐 성능을 더 끌어 올릴 수 있었고, 과적합(Over-fitting)도 방지할 수 있었다.

본 연구에서는 두 가지 방법을 섞은 Long Short Term Ensemble(LSTE) method를 제안한다. 제안된 방법은 다음과 같이 구성된다. 첫째, 석유화학기업의 주요 Value Chain(밸류 체인)을 파악하여 매출액, 영업이익 등 실적을 추정하고, 향후 주가에 대한 방향성을 예측한다. 둘째, 딥러닝 기반 주가 예측을 통해 추세에 따른 변동 패턴을 찾아낸다. 마지막으로 실시간 주식투자 시 시장참여자들의 심리를 파악하여 최적의 주문집행 알고리즘을 시행한다. 이로써 투자자는 주식 거래 차트를 어느 시점에 사용하여 얼마나 거래를 할 것인지 결정을 할 수 있다.

2.2.1 LSTM(Long Short Term Memory)

주로 시계열 처리나, 자연어 처리에 사용되는 LSTM(Long Short Term Memory)는 기존의 RNN이 출력과 먼 위치에 있는 정보를 기억할 수 없다는 단점을 보완하여 장/단기 기억을 가능하게 설계한 신경망의 구조를 말한다. Fig. 1은 LSTM의 전체적인 내부 모습을 보여준다. LSTM은 은닉층의 메모리 셀에 입력 게이트, 망각 게이트, 출력 게이트를 추가하여 불필요한 기억을 지우고, 기억해야 할 것들을 정한다. LSTM은 은닉 상태(hidden state)를 계산하는 식이 전통적인 RNN보다 조금 더 복잡해졌으며 셀 상태(cell state)라는 비교하여 긴 시퀀스의 입력을 처리하는데 탁월한 성능을 보인다[11]. 셀 상태는 왼쪽에서 오른쪽으로 진행하며, 셀 상태 또한 이전에 배운 은닉 상태처럼 이전 시점의 셀 상태가 다음 시점의 셀 상태를 구하기 위한 입력으로 사용된다. 은닉 상태의 값과 셀 상태의 값을 구하기 위해서 새로 추가된 3개의 게이트를 사용하고, 각 게이트는 삭제 게이트, 입력 게이트, 출력 게이트라고 부르며 이 3개의 게이트에는 공통으로 시그모이드 함수가 존재한다. 시그모이드 함수를 지나면 0과 1 사이의 값이 나오게 되는데 이 값들을 가지고 게이트를 조절하게 된다.

2.2.2 앙상블 학습(Ensemble Learning) 방법

앙상블 학습(Ensemble Learning)은 여러 개의 분류기를 생성하고, 그 예측을 결합함으로써 더욱 정확한 예측을 도출하는 기법을 말한다. 강력한 하나의 모델을 사용하는 대신 보다 약한 모델 여러 개를 조합하여 더 정확한 예측에 도움을 주는 방식이다. 우리는 랜덤포레스트(Random Forest)와 XGBoost를 사용하였다. Bagging의 분산은 각 트리들의 분산과

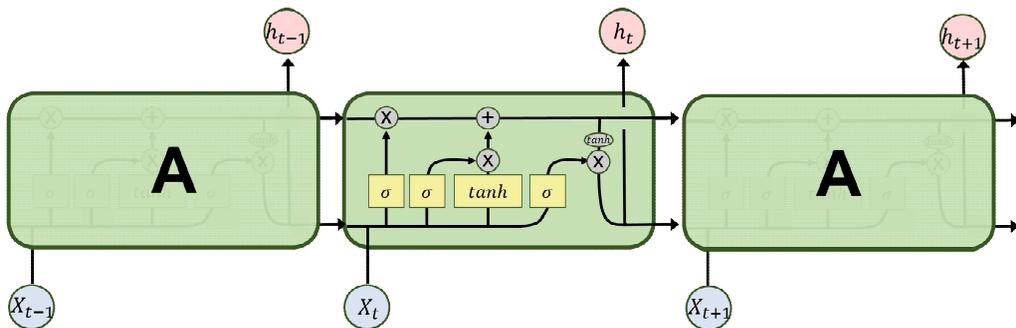


Fig. 1. Architecture of LSTM.

그들의 공분산으로 구성 되는데, 복원추출을 진행했으나 각 트리들은 중복데이터를 가지고 있기 때문에 독립이라는 보장이 없다. 공분산이 0이 아님에 따라 트리가 증가하면 모델 전체의 분산이 증가할 수가 있는데, 각 트리 간 공분산을 줄이는 방법인 랜덤포레스트 알고리즘을 활용하였다. 즉, 데이터뿐만 아니라 변수까지 무작위로 뽑아 기본모델(Base Learner) 간 공분산을 줄이고자 했다. 또한, 부스팅은 약분류기(Weak Classifier)를 이용해서 강분류기(Strong Classifier)를 만드는 방법으로 약분류기를 직렬로 연결해서 앞선 분류기에서 제대로 분류하지 못한 부분에 대하여 부스팅을 하는 알고리즘이다. 부스팅의 경우 차례로 모델을 학습시킴으로써 오답에 대하여 집중하여 후순위 모델을 학습하므로 정확도가 더 높게 나타난다.

3. 제안

3.1 제안한 알고리즘의 개요

본 논문에서는 주식의 장단기 움직임 예측을 통해 정확도를 향상하고 수익률을 높이기 위한 장단기 예측 알고리즘을 제안한다. 제안한 방법론은 다음과 같다. 첫째, 석유화학기업의 주요 Value Chain을 파악하여 매출, 영업이익 등의 성과를 추정하고 향후 주가의 움직임을 예측한다. 둘째, 딥러닝 기반 주가 예측을 통해 추세에 따른 변동 패턴을 찾는다. 마지막으로 실시간 주식투자 시 시장참가자의 심리를 파악하여 최적의 주문실행 알고리즘을 구현하고, 이를 통해 투자자는 주식 거래 차트를 사용하여 거래 시기와

거래 금액을 결정할 수 있다.

또한, 제안한 LSTE 방법은 전처리 부분, 예측 알고리즘 부분 및 실시간 자동 거래 부분의 부분으로 구성돼 있다. 첫째, 주식의 증가를 활용하여 GT (Ground Truth)를 생성한다. 이때 우리는 6가지 주요 주식 움직임이 있다고 가정했다. 비지도 휴리스틱 알고리즘을 활용하여 각 주식의 거래 데이터를 미리 정의된 고정 길이의 여러 클립으로 자르고 6개의 주요 클래스로 분류하였다. 우리는 해당 길이를 5일로 세팅했다. 또한, 상향/하향 변곡점을 하기 감지 위해 반전 지점을 정의하였다. 캔들차트와 기술적 지표의 조합을 바탕으로 27개의 특징 후보군을 구, 측면 상향, 연속 하향, 측면 하향, 플랫폼, 그 외). 주가의 움직임과 그 상승(혹은 하락)을 예상하였고, 각 종목에 적합한 상향/하향 예측의 특징(Feature) 집합을 추출하였다. 우리는 상향, 하향, 그리고 변화 없음 세가지로 GT를 재설정했다.

두 번째, 예측 알고리즘은 다음날 주가를 예측하는 조건기반알고리즘, day+n일 예측하는 앙상블 알고리즘, 변곡점(Reversal point, RP) 예측 알고리즘이 있다. Fig. 3에 서술된 조건기반알고리즘은 주가를 예측하기 위해 특성의 조건을 결합하여 최상의 조건을 찾아낸다. 우리가 제안한 방법은 석유화학 품목과 가격에 대한 정보를 기반으로 다음 날 주식의 움직임을 예측하는 것이다. 앙상블과 변곡점 탐지 알고리즘은 모두 장기 절차를 예측한다. 앙상블 학습 방법은 여러 하위 모델의 결정을 새 모델로 결합한 다음 최종 출력을 만들어 예측 정확도 또는 전체 성능을 향상한다. 각 모델에는 고유한 장점이 있고 서

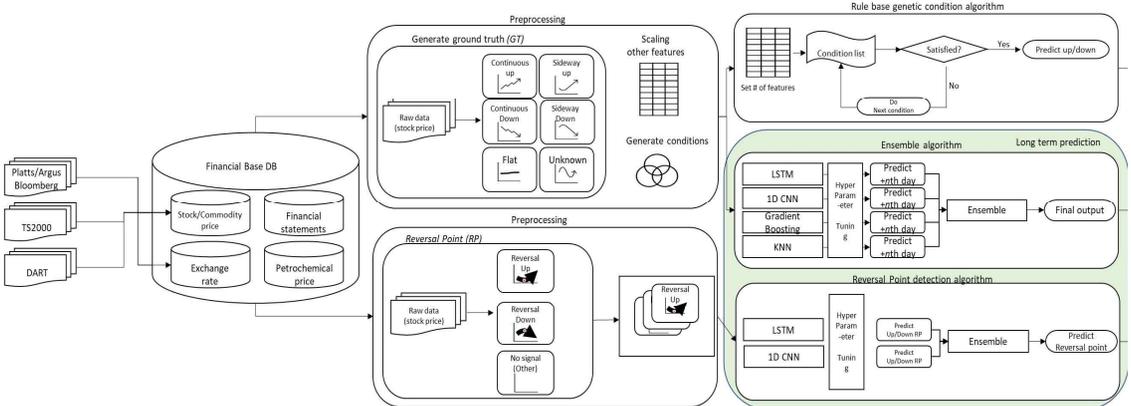


Fig. 2. Overview of proposed algorithms, Long Short Term Ensemble (LSTE).

Algorithm 1: Rule Based Algorithm

```

1 function RuleBased(a, n, m):
  Input : 1. Array of a stock ground truth. Shape (a, 1)
          2. Array of petrochemical price (percentage change between
             today and yesterday), shape (a, n)
          3. comb_list = combinations of petrochemical list nCm,
             where m is the number of petrochemical items what we utilize
             to train
  Output: conditions, predicted value (up/down)
2 for comb_list(item1, item2, ..., itemn) do
3   for row_index in each item rows do
4     if row_index == True: then
5       | return pred_value = up;
6     else
7       | return pred_value = down;
8     end
9   end
10 end
    
```

Fig. 3. Explanation of Rule Based algorithm (RBA).

로 보완하여 개별적인 단점을 극복할 수 있으므로 앙상블 모델은 단일모델보다 더 나은 예측성능을 보여준다[12]. LSTM, 1D CNN, Gradient Boosting, KNN 4가지 모델을 사용하였고, 그리드 검색을 통해 각각의 하이퍼 파라미터들값을 저장하였다. 그 이후 각각의 모델에 대해 같은 모델로 앙상블을 하여 하이퍼 파라미터들의 값에 따른 성능 차이를 최소화하고자 했다. 그다음 최종적으로 각각의 모델을 앙상블 하였다. 해당 방법을 통해 성능 향상을 이룰 수 있었다(Table 1 참조).

3.2 변곡점 이미지의 활용

변곡점 탐지 알고리즘의 주요 목적은 주가의 반전점을 예측하는 것이다. 본 연구의 가장 중요한 문제는 주가의 상승 또는 하락 추세가 방향을 바꾸는 시점이 실제 반전 포인트인지, 일시적인 변동인지를 판단하는 것이다[13]. 추세, 변곡점 후보군, 그리고 반전 포인트를 정의하는 방법은 매우 중요하다. 시계열 데이터, 특히 주가 데이터의 경우 변곡점 포인트 예측에 항상 적합한 조합들이 존재하지 않으며, 또한 모든 시간 축에 대해서 균일하게 적합한 특징지도가 존재하지 않기 때문이다. 예측 정확도를 높이기 위해 기존에 도입된 기술적 지표와 캔들 차트를 서로 다른 조합으로 27개의 특징 후보 집합을 구성하고 실제 실험을 통해 각 종목에 가장 적합한 특징 집합을 선정했다. 27개의 특징 후보 집합 중 학습 과정을 통해 상하향 예측에 적합한 특징 집합을 각각 선정하고, 이에 따라 상향/하향 예측 모델의 구조적 파라미터를 결정하였다.

또한, 변곡점 변화 탐지를 위해 Reversal Signal Line(RSL)을 추가로 적용하여 GT의 Decision boundary를 그렸다. RSL은 알고리즘 2(Fig. 5)와 같이 정의된다. 5일 이동평균선이 지속해서 하락 중이나 t-1 시점을 기준으로 이점 시점보다 주가가 반전한 경우, t 시점을 기준으로 forward window size까지 주가가 상승/하락지지 신호선을 이탈하지 않으며, 기준시점

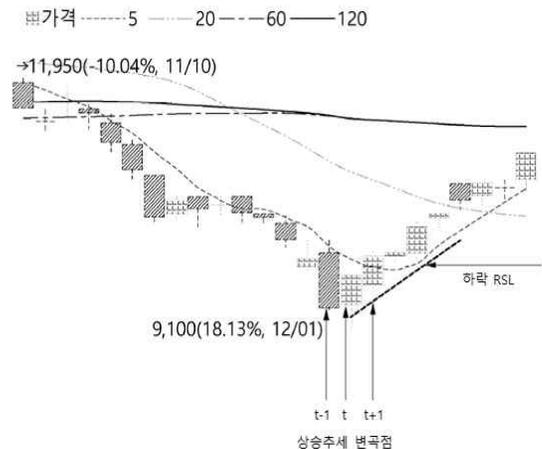
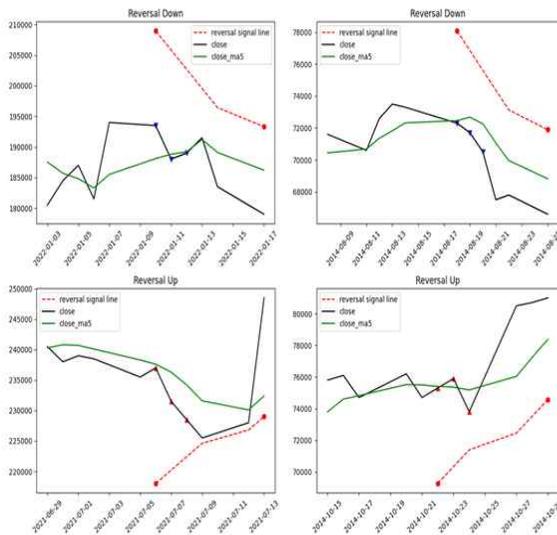


Fig. 4. (Left) Examples of Reversal Signal Line in train dataset, (Right) An illustration of candlestick chart with RSL.

```

Algorithm 2: Reversal Signal Line (RSL)
1 function RSL ( $\alpha_1, \alpha_2, \beta, \delta, G$ ):
  Input : 1. G= Growth rate in a window
          2. IS.RSL=False
  Output: RS Line
2 for Signal point between 50% and 80% in a window do
3   if The close price of Signal point in the range [ $\alpha_1 * G, \alpha_2 * G$ ]: then
4     Draw a RSL;
5     if a rate of  $\delta$  points between at the first and of RSL* $\beta$ : then
6       | IS.RSL=True
7     end
8   end
9 end
10
    
```

Fig. 5. How to generate Reversal Signal Line (RSL).

부터 G만큼 상승/하락한 지점을 정의한 후, 변곡점을 탐지할 Signal point 범위를 50%~80%로 정했다. 그 다음, $\alpha_1 * G \sim \alpha_2 * G$ 범위 내에 있는 Signal point를 계산하였고, 이때 α_1 와 α_2 값은 각각 0.05, 0.1로 설정했다. 해당 값이 임계치($\beta=0.95$) 내에 있는 경우 RSL로 지정하였다.

새로운 모멘텀 변화지점을 관찰함으로써 추세가 급격하게 변화한 지점에 영향을 주는 변수를 탐지할 수 있었다. (Fig. 6 참고) 해당 특징 세트, 상하향 예측 모델은 모두 입력 및 출력 레이어와 히든 레이어로 구성된 LSTM 네트워크를 활용하며 7개의 값(10, 20, 30, 두 모델에서 각각 60, 100, 120, 200). 변곡점 예측 모델은 과거 10일 데이터를 입력 데이터로 사용하여 미래 10일 추세를 예측하도록 세팅했다.

4. 실험 결과 및 고찰

4.1 실험 세팅

4.1.1 데이터셋

본 연구의 실험을 위해 국내 NCC 기업인 대한유화, 롯데케미칼, 그리고 OCI 주식에 대해 주가 예측 실험을 진행했다. 실험데이터 셋은 2011년 01월 02일부터 2021년 10월 05일까지 포함하고 있으며, 학습, 검증, 테스트의 비율은 7: 2: 1로 진행했다. 또한, 데이터에서 주말과 공휴일 등의 거래일을 정리하였다[14, 15]. 본래 데이터는 움직임이 아닌 주가 예측을 위해 수집되므로 레이블을 만들어내는 다른 접근 방식을 설정하였다. 원 OHLCV(Open, High, Low, Close, Volume) 데이터를 사용한 것이 아니라 Zhang[16]에 따르면 주가의 움직임은 6가지로 정의가 되는데, 패턴 듀레이션 안에서의 패턴 인식을 통해 주가의 움직임을 분류하였다. 패턴 듀레이션은 증가 트렌드를 예측하는 특정 윈도우 크기를 말한다. 일반적인 인공지능망의 경우 전달의 값을 대부분 다음날의 예측 결과로 사용하는 단점이 있다. 패턴 듀레이션과 그에 따른 주가의 움직임을 정의함으로써 단점을 해소할 수 있었다.

4.1.2 평가방법

정확도(Accuracy)를 측정항목으로 예측성능을 평가하며, 범위는 [0, 100]이다. 측정항목 값이 클수록 더 나은 예측 분류 성능을 가진다. 그리고 동일 실험을 10번 반복함으로써, 초기매개변수 값의 의존도를 낮췄다.

4.2 정확도

Table 1은 단기예측을 하는 RBA의 정확도를 보여준다. 롯데케미칼이 제일 높은 정확도를 보여주었

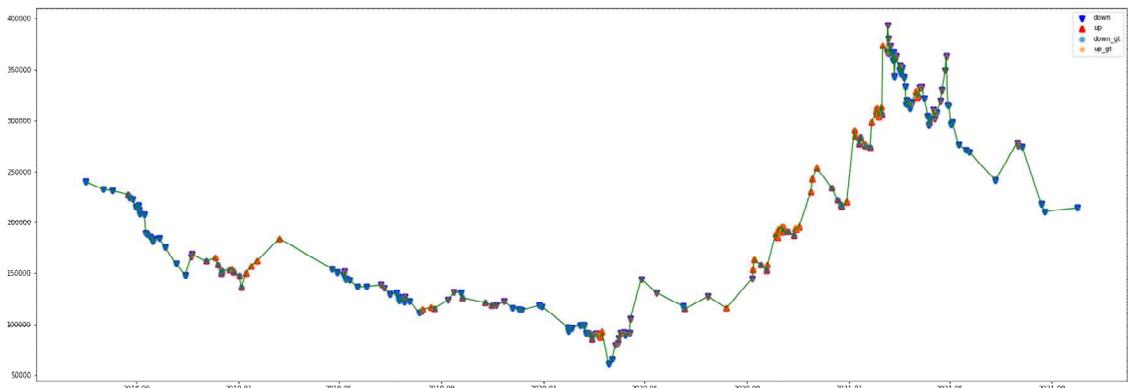


Fig. 6. Reversal Signal Line (RSL).

Table 1. Accuracy for Rule-Based Algorithm (RBA).

Corporation	KPIC	Lotte Chemical	OCI
RANDOM ACC(%)	49.30	52.60	51.00
LSTM ACC(%)	53.21	51.09	53.74
RBA ACC(%)	58.67	59.42	59.78

고 OCI, 대한유화가 그 뒤를 따랐다. 기존의 LSTM 분석 방법[17]인 LSTM을 썼을때보다 RBA를 이용하였을 때 더 높은 정확도를 보였다.

또한, Table 2는 대한유화의 앙상블 알고리즘 실험에 관한 결과를 보여준다. 과거 데이터로만 학습할 경우 평균적으로 57.024%의 정확도를 기록했으나, 석유화학제품 데이터를 활용할 경우 평균적으로 61.21% (7.4% 향상)의 정확도를 보여주었다.

좀 더 자세히 보면, 학습 데이터 종류와 상관없이 앙상블 방법이 제일 정확도가 높았으나, 과거 데이터를 활용한 앙상블의 정확도(58.41%)보다 석유화학 제품 데이터를 사용한 실험에서 LSTM을 제외한 나머지 알고리즘들이 정확도가 훨씬 높았다. 추산한 매출액의 효과성을 보여줄 뿐만 아니라 과거 데이터의 의존성을 해소함으로써 일반화(generalization)를 향상할 수 있었다.

그리고 Table 3을 참조하면 앙상블 알고리즘에선

RBA와 마찬가지로 롯데케미칼이 제일 높은 정확도를 보였고, 대한유화와 OCI 정확도는 각각 65.61%, 63.15%이다. 또한, 변곡점(Reversal point, RP) 예측 알고리즘의 경우 대한유화, OCI, 롯데케미칼 순으로 정확도가 높았다.

마지막으로 제안한 알고리즘을 통합하여 정확도를 측정해 보았다. 최종 알고리즘인 LSTE는 단기예측 알고리즘인 RBA보다 상당한 성능 향상(14.47%)을 보였다. 해당 실험들을 통해 다음과 같은 관찰 결과가 있었다. 첫째, 기존 방법인 과거 데이터만을 사용했을 때, 과적합 되거나 손실함수가 수렴하지 않는 문제가 있었다. 이는 새로운 데이터셋에선 모델이 잘 작동을 못 하거나 학습을 하지 못 해 무작위에 가까운 예측을 한다는 것이다. 우리는 주가와 밀접한 데이터를 특징으로 정함으로써 서술한 두 가지 문제를 해결하였다. 둘째, 장단기 예측 알고리즘을 동시에 구축함으로써 더 높은 정확도를 가질 수 있었다.

Table 2. Experimental results on different train dataset.

	Train Dataset: Historical closed price					Train Dataset: petrochemical price				
	LSTM	1D-CNN	Gradient Boosting	KNN	Ensemble	LSTM	1D-CNN	Gradient Boosting	KNN	Ensemble
Accuracy (%)	53.54	55.80	57.61	59.76	58.41	55.24	62.45	59.33	63.42	65.61

Table 3. Performance on each stock with ensemble and reversal point detection algorithms.

Corporation	KPIC	Lotte Chemical	OCI
Ensemble ACC(%)	65.61	67.74	63.15
Reversal point ACC(%)	71.54	70.22	72.38

Table 4. Evaluation on the consolidated algorithm of proposed methods.

	RBA	RBA+Ensemble	Ensemble+RP	RBA+Ensemble+RP
KPIC ACC(%)	58.67	66.30	67.32	68.23
Lotte Chemical ACC(%)	59.42	61.71	68.10	65.49
OCI ACC(%)	59.78	64.42	66.57	69.89

5. 결 론

본 연구는 그동안 기존 연구에서 기본적 분석 또는 기술적 분석만을 이용하여 주가 예측을 시도한 것과는 달리, 두 가지 분석 방법을 결합하여 분석하고자 한 연구이다. 실험 분석을 위하여 기본적 분석 이론을 적용하기 쉬운 석유화학산업을 표적으로 선정하였다. 일일 단위로 업데이트되는 기초유분 스프레드 및 WTI 가격을 통해 해당 기업의 영업이익을 추정하였으며, 조건검색 및 딥러닝 기반 앙상블 방법을 통해 주가의 단기예측을 시도하였다. 또한, 변곡점 이미지 감지를 활용하여 학습시간 및 정확도를 효과적으로 향상했다. 실험 결과 본 연구에서 예측 알고리즘을 석유화학제품의 가격들로 특징 셋으로 구성함으로써 기존의 방법에 비하여 성능이 우수함을 확인하였다.

본 연구의 결과는 다음과 같은 학술적 실무적 의의가 있다. 첫째, 훈련 데이터에 대해 서로 다른 빈도와 예측 시점을 설정하여 장단기 예측을 통해 위험을 관리한다는 점에서 의의가 있다. 둘째, 딥러닝을 이용한 다른 주가 변동 예측 연구보다 정확도가 높아 실시간 매매 전략에 활용할 수 있다. 셋째, 이유를 설명하기 어려운 기존 딥러닝과 달리 조건기반 알고리즘을 추가로 사용하여 주가 변동을 예측할 때 주식 시장 참여자의 객관적인 의사 결정에 도움을 줄 수 있다. 다만 특정 산업 분야와 특정 기업을 대상으로 분석을 한 점은 연구의 한계점으로 볼 수 있으며, 본 알고리즘을 활용하여 다른 산업 분야로의 적용은 추가 연구과제로 남겨둔다.

REFERENCE

- [1] S. Boonpeng and P. Jeatrakul, "Decision Support System for Investing in Stock Market by using OAA-Neural Network," *8th International Conference on Advanced Computational Intelligence*, pp. 1-6, 2016.
- [2] B. Hurst, Y.H. Ooi, and L.H. Pedersen, "A Century of Evidence on Trend-Following Investing," *The Journal of Portfolio Management*, Vol. 44, No. 1, pp. 15-29, 2017.
- [3] P. Malo, A. Sinha, P. Korhonen, J. Wallenius, and P. Takala, "Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts," *Journal of the Association for Information Science and Technology*, Vol. 65, No. 4, pp. 782-796, 2014.
- [4] H. Kim, S. Hong, and H. Hong, "Machine Learning Based Stock Price Fluctuation Prediction Models of KOSDAQ-listed Companies Using Online News, Macroeconomic Indicators, Financial Market Indicators, Technical Indicators, and Social Interest Indicators," *Journal of Korea Multimedia Society*, Vol. 24, No. 3, pp. 448-459, 2021.
- [5] J.D. Piotroski, "Value Investing: Use of Historical Financial Statement Information to Separate Winners from Losers: An Indian Evidence," *Journal of Accounting Research* Vol. 38, 2000, pp. 1-41. 2000.
- [6] B.G. Malkiel. *Efficient Market Hypothesis. Finance*, Palgrave Macmillan, London, pp. 127-134, 1989.
- [7] E.F. Fama. "The Behavior of Stock-Market Prices," *The Journal of Business*, Vol. 38, No. 1, pp. 34-105, 1965.
- [8] T. Lux and T. Kaizoji, "Forecasting Volatility and Volume in the Tokyo Stock Market: Long Memory, Fractality and Regime Switching," *Journal of Economic Dynamics and Control*, Vol. 31, No. 6, pp. 1808-1843, 2007.
- [9] A.A. Ariyo, A.O. Adewumi, and C.K. Ayo, "Stock Price Prediction Using the ARIMA model," *UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pp. 106-112, 2014.
- [10] Y. Park and J. Lee, "An Empirical Study on the Relevance of Applying Relative Valuation Models to Investment Strategies in the Japanese Stock Market," *Japan and the World Economy*, Vol. 15, No. 3, pp. 331-339, 2003.
- [11] D.M. Nelson, A.C. Pereira, and R.A. Oliveira, "Stock Market'S Price Movement Prediction with LSTM Neural Networks," *International Joint Conference on Neural Networks (IJCNN)*,

pp. 1419-1426. 2017.

[12] T.G. Dietterich, *Ensemble Learning, The Handbook of Brain Theory and Neural Networks*, 2nd edition, The MIT Press, Cambridge, 2002.

[13] J. U, P. Lu, C. Kim, U. Ryu and K. Pak “A New LSTM Based Reversal Point Prediction Method Using Upward/ Downward Reversal Point Feature Sets,” *Chaos, Solitons & Fractals*, Vol. 132, 109559, 2020.

[14] Feng, Chen, He, Ding, Sun, and Chua “Enhancing Stock Movement Prediction with Adversarial Training,” *arXiv Preprint*, arXiv: 1810.09936, 2018.

[15] N. Cohen, T. Balch, and M. Veloso, “Trading via Image Classification,” *Proceedings of the First ACM International Conference on AI in Finance*, pp. 1-6, 2020.

[16] J. Zhang, S. Cui, Y. Xu, Q. Li and T. Li “A Novel Data-Driven Stock Price Trend Prediction System,” *Expert Systems with Applications*, Vol. 97, pp. 60-69, 2018.

[17] H. Sepp and S. Jürgen, “Long Short-Term Memory,” *Neural Computation*, Vol. 9, No. 8, pp. 1735-1780, 1997.



방 은 지

울산과학기술원 컴퓨터공학 석사
 울산과학기술원 에너지상품거래
 및 금융공학 석사
 울산대학교 회계학 학사
 주식회사 알레코노랩 연구원
 (’21.11.~현재)



변 희 용

방송통신대학교 통계·데이터과학
 과 재학
 주식회사 알레코노랩 연구원
 (’21.11.~현재)



조 재 민

포항공과대학교 기술경영대학원
 공학박사
 연세대학교 경영학사/기계공학사
 현, 울산대학교 경영학부 부교수
 현, 주식회사 알레코노랩 대표이
 사(’21.11.~현재)

전, 전국은행연합회 기술정보부 조사역
 전, 삼성중공업 기획팀 과장