

K-평균 군집화 데이터 증강을 통한 주가 심층 예측[☆]

Deep Prediction of Stock Prices with K-Means Clustered Data Augmentation

한 경 훈¹ 양 희 규² 추 현 승^{1,2,3*}
Younghoon Han Huigyu Yang Hyunseung Choo

요 약

금융 분야에서 주가예측연구는 거래 안정성 및 이익 실현 등을 목적으로 한다. 기존의 통계적 예측기법은 무작위로 예측한 결과와 정확도 측면에서 비슷하거나 낮은 예측 신뢰도 때문에 실제 거래 결정에 참고 되기 어렵다. 인공지능 모델은 데이터특성과 변동 패턴을 학습해 예측하기 때문에 향상된 정확도를 달성한다. 그러나 장기간의 시계열 데이터를 사용해 주가를 예측하는 것은 여전히 어려운 문제이다. 본 논문에서는 K-means 클러스터링 기반의 데이터 증강 및 입력 시퀀스의 Window-size 별 정규화 기법과 시계열 학습에 특화된 LSTM 모델을 활용하여 안정적이고 신뢰성 있는 주가예측 방법을 제안한다. 이를 통해 더욱 정확하고 신뢰성 있는 예측 결과를 얻고, 나아가 시장 안정성에 기여할 뿐 아니라 높은 수익도 추구할 수 있다.

☞ 주제어 : 주가예측, 클러스터링, 데이터 증강, 딥러닝, 인공지능

ABSTRACT

Stock price prediction research in the financial sector aims to ensure trading stability and achieve profit realization. Conventional statistical prediction techniques are not reliable for actual trading decisions due to low prediction accuracy compared to randomly predicted results. Artificial intelligence models improve accuracy by learning data characteristics and fluctuation patterns to make predictions. However, predicting stock prices using long-term time series data remains a challenging problem. This paper proposes a stable and reliable stock price prediction method using K-means clustering-based data augmentation and normalization techniques and LSTM models specialized in time series learning. This enables obtaining more accurate and reliable prediction results and pursuing high profits, as well as contributing to market stability.

☞ keyword : Stock price prediction, Clustering, Data augmentation, Deep learning, Artificial intelligence

1. 서 론

인공지능은 산업 전반에 걸쳐 활용되고 있으며, 특히 금융 분야에서는 부정거래탐지, 자산관리, 자동화 거래, 신용평가, 환율 예측 등에 적용되고 있다. 특히, 주가 정

보는 실시간으로 시장에 공개되며 다른 분야에 비해 데이터의 확보가 쉽고 결과 비교가 용이해 많은 금융 관계자들이 인공지능을 활용해 주가예측에 도전하고 있다.

기존의 인공지능을 활용한 연구들은 주가, 뉴스, 경제 지표[1], 기업의 재무 상태 정보를 등을 가공하여 개별 주식의 주가를 예측하고자 했다. 예를 들어, 특정 주식 종목에 대해 참여자들의 글을 추출 후 감정분석을 통해 주가를 예측하는 연구[2, 3, 4]가 있다. 또한, 금융 기술적 지표와 머신러닝 알고리즘을 결합하여 주가를 예측하는 연구[5, 6]도 있다. 주가의 시계열성 뿐만 아니라 비선형적 요인도 추출하여 예측할 수 있도록 다양한 모델을 만들어 적용해본 연구[7, 8]에서는 50% 초반대의 예측 정확도를 기록하였다. 그 외에, 주가지수의 시계열적 특성을 고려하여 장단기 기억 신경망(Long Short-Term Memory) 모델을 활용한 연구[9, 10]들도 활발히 이루어지고 있다. 그러나 장시간의 시계열 데이터를 사용해 주가를 예측하는

¹ College of Computing and Informatics, Sungkyunkwan University, Suwon, 16419, Korea

² Dept. of Superintelligence Engineering, Sungkyunkwan University, Suwon, 16419, Korea

³ Dept. of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, 16419, Korea

* Corresponding author (choo@skku.edu)

[Received 16 April 2023, Reviewed 19 April 2023, Accepted 25 April 2023]

☆ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT명품인재양성(IITP-2023-2020-0-01821), 4단계 BK21 사업과, 인공지능 혁신 허브 (No.2021-0-02068), 지원을 받아 수행된 연구 결과임.

것은 여전히 어려운 문제이고, 통계적 예측기법을 활용한 주가예측 모델은 예측 정확도가 50% ~ 60% 수준에 그쳐, 사실상 무작위로 예측하는 경우의 확률과 크게 다르지 않아 투자자들의 신뢰를 얻지 못하고 참고지표로 주로 사용되고 있는 것이 현실이다.

이러한 문제를 해결하기 위해 제안 예측모델은 시계열 주가 데이터에 LSTM 모델을 적용하고, 클러스터링 기반의 증강된 데이터셋을 학습한다. 모델 학습에 충분한 데이터를 확보하기 위해 유사한 특성을 가진 종목 데이터를 연결하여 데이터를 증강한다. 데이터셋 구성을 위해 제안 모델은 K-Means 알고리즘을 활용해 주가 변동 패턴을 기준으로 종목을 군집화한다. 실험은 KOSPI200 지수에 포함된 종목들의 시작가, 최고가, 최저가, 증가, 거래량을 학습데이터로 사용한다.

K-Means 알고리즘을 이용하여 유사한 주가 변동 패턴을 가진 종목들을 클러스터링하고 해당 그룹의 시간별 주가 정보를 합쳐 하나의 데이터셋으로 학습한 결과 최대 68%의 정확도를 보여주었다. 이러한 결과는 다양한 종목의 데이터를 활용하여 학습할 경우 주가예측의 성능이 향상될 수 있음을 보여준다.

본 논문은 총 5장으로 구성되어 있으며, 1장에서는 연구의 배경과 목적, 기존 기법과 제안된 기법을 비교하고 제안된 모델의 성능 결과를 제시한다. 2장에서는 주가예측 모델을 구축하기 위해 사용된 딥러닝 및 K-Means 클러스터링 기법을 설명하며, 금융 예측 관련된 시계열 모델들을 소개한다. 3장에서는 데이터 수집 및 전처리 방법, 모델 구축 방법 등에 대한 제안 모델과 연구 내용을 설명한다. 4장에서는 실험을 통한 성능 측정 결과를 평가하고, 5장에서는 결론 및 향후 연구계획에 관해 설명한다.

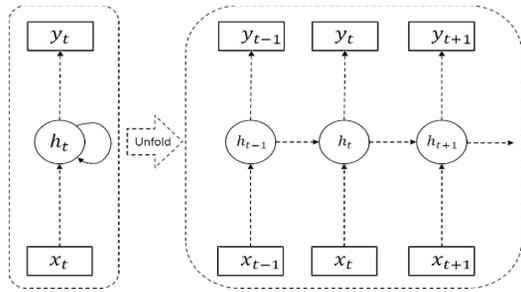
2. 관련 연구

주가, 환율, 금리, 가상화폐, 부동산 시장, 신용 리스크 [11, 12, 13] 등 다양한 금융 분야에서 시계열 모델을 활용한 연구가 이루어지고 있다. 예를 들어, 주가 시계열 모델은 시장의 변화를 예측하여 투자 전략을 수립하는 데에 사용되며, 환율 시계열 모델은 외환 시장의 변동성을 파악하여 환전 시점을 결정하는 데에 활용된다. 또한, 가상화폐와 같은 신생 분야에서도 시계열 모델을 이용한 예측 및 분석이 활발히 이루어지고 있으며, 이러한 연구는 금융 시장의 안정성을 높이는 데 중요한 역할을 한다.

K-Means 클러스터링은 비지도 학습 알고리즘으로, 사용자가 사전에 클러스터 개수 k와 초기값을 설정한 후 각

데이터를 그룹에 할당해 나가는 알고리즘이다. 패턴 공간상의 특징 벡터로 표현된 패턴 데이터를 몇 개의 부분집합으로 나누고, 클러스터들 중에서 안정성과 수익률이 높은 클러스터를 선정 후 포트폴리오를 구성하여 다양성을 증가시키고 위험을 최소화하면서 수익을 극대화 할 수 있는 방법[14]을 제안한 연구도 있다.

순차 데이터는 일련의 데이터가 특정한 시간적 순서를 가지는 데이터로, 이러한 순서가 변경되면 의미가 손상되며, 대표적인 예로는 DNA 염기서열, 기온 변화, 지진파 등이 있다. 이러한 순차 데이터를 학습하기 위해 개발된 RNN(Recurrent Neural Networks)은 딥러닝의 한 종류이다. RNN은 내부의 순환 구조를 통해 입력을 받을 때마다 이전의 값을 기억할 수 있어, 순차적인 입력값들의 학습에 적합한 모델이다.



(그림 1) RNN 구조도
(Figure 1) Architecture of RNN

그림 1은 RNN의 순환 구조를 펼친 형태를 보인다. RNN은 일반적으로 시간에 따라 각각의 입력값과 출력값이 있다. RNN은 이전 시간 단계에서 계산된 출력값을 현재 시간 단계의 입력값과 함께 연산하여 현재 시간 단계의 출력값을 계산한다. 이 과정에서 RNN은 시퀀스 데이터의 이전 정보를 기억하고 현재 시점에서의 패턴을 학습할 수 있다. RNN을 수식으로 표현하면 다음과 같다.

$$y_t = \alpha(W_{hy}h_t + b_y) \quad (1)$$

$$h_t = \alpha(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (2)$$

그러나, RNN은 입력값 사이의 시간적인 차이가 매우 큰 경우, 시간 방향으로 펼쳐진 신경망의 역전파(BPTT, Back Propagation Through Time) 알고리즘을 통해 기울기 값을 보정하는 학습 능력이 저하되며, 이는 시계열 데이터에서 장기 의존 관계를 학습하기 어렵게 만든다[15].

이러한 RNN의 장기기억 문제를 해결하기 위해 고안

된 알고리즘인 LSTM(Long Short-Term Memory)은 RNN 계층의 신경망 구성에 입력, 삭제 및 출력 게이트를 추가하는 방식으로, 이전의 불필요한 기억은 지우고, 필요한 기억을 유지하며, 이를 새로운 입력과 결합하여 다음 출력을 생성하도록 도움을 주어 장기 의존 관계를 쉽게 학습할 수 있도록 한다. 따라서 LSTM은 시계열 데이터에 좀 더 적합하고, 주가예측에도 널리 활용되고 있다.

3. 제안 모델 및 연구내용

3.1 데이터 수집 및 전처리

주가 분석을 위해 Finance Data Reader API[16]를 사용하여 KOSPI 200 내 16개 종목의 1998년 12월부터 2023년 3월까지의 일별 데이터를 수집하였다. 또한, Yahoo Finance API[17]를 통하여 KOSPI 200내의 16개 종목의 2020년 3월부터 2023년 3월까지의 시간별 데이터를 수집하였다. 수집한 데이터는 날짜, 시초가, 최대가, 최저가, 종가, 거래량, 등락 여부로 구성되어 있으며, 시간별 데이터셋은 일별로 6개의 데이터로 구성되어 있다.

(표 1) 일자별 데이터셋
(Table 1) Daily dataset

Date	Open	High	Low	Close	Volume	Change
1998-12-22	976	996	967	975	832580	0
1998-12-23	975	1009	957	981	879790	1
1998-12-24	976	990	961	967	346480	0
1998-12-28	957	965	932	942	428840	0
...
2023-03-20	62100	62300	61100	61800	964632	1

(표 2) 시간별 데이터셋
(Table 2) Hourly dataset

Date Time	Open	High	Low	Close	Volume	Change
2020-03-15 09:00:00	9160	9200	9120	9190	0	0
2020-03-15 10:00:00	9190	9190	9150	9150	134576	0
2020-03-15 11:00:00	9150	9170	9150	9160	64637	0
2020-03-15 12:00:00	9160	9170	9150	9170	80907	0
2020-03-15 13:00:00	9160	9170	9140	9160	102564	0
2020-03-15 14:00:00	9160	9170	9140	9160	113220	0
...
2023-03-15 14:00:00	14210	14270	14160	14170	109822	1

이 연구에서는 주가 등락을 예측하기 위해 사용되는 주가 정보의 양이 부족하거나, 단일 종목만으로 분석하여 주가 변동의 특징을 제대로 파악하지 못했던 기존 연구들의 한계를 극복할 수 있도록 새로운 방법을 도입하였다. 먼저 K-Means 알고리즘을 사용하여 주가 변동 패턴이 유사한 종목들을 클러스터링 하는 방법을 사용하였다. 이를 통해 종목들 간의 상관관계를 파악하고, 더 우수한 예측 정확도를 보여줄 수 있게 한다. 그 후 군집화 한 종목들의 데이터셋을 연결하여 데이터의 양을 늘려 보다 정확한 분석 결과를 얻을 수 있도록 하였다. 이 방법은 심플하지만 데이터 부족으로 인해 정확도가 저하되는 문제를 극복할 수 있으며, 데이터의 다양성도 보장할 수 있다.

(표 3) 주가 변동 패턴에 기반한 클러스터링 결과
(Table 3) The clustering results based on pattern of stock price fluctuations

업종 일치	은행업	신한지주, KB금융, 하나은행, IBK은행
	증권업	삼성증권, 한국금융지주, 미래에셋증권, NH증권
업종 비일치	전자, 전기, SW개발	삼성전자, SK하이닉스, 삼성전기, 삼성SDS
	사행시설, 유통, 지주, 운송	강원랜드, 코웨이, CJ, 대한통운

일반적으로 데이터를 정규화 할 때는 모든 Feature 들을 0~1 범위 내의 값으로 변환하여 값의 크기에 따른 영향을 줄이게 된다. 그러나 주가 데이터의 경우, 10년 전과 현재의 주가 사이에는 일반적으로 큰 가격 차이가 있다. 따라서, 전체 기간을 대상으로 정규화 된 데이터는 변동 민감도를 감소시키고, 학습 과정에서 모델은 주가 변동의 특징을 파악하지 못한다. 이를 극복하기 위해 제안 모델은 입력 시퀀스의 Window-size 별로 정규화를 수행하여, 학습에 활용되는 기간의 주가 변동 특성을 극대화한다. 이 방식은 LSTM 모델의 시계열 데이터 특성 추출 성능을 향상시키며, 비교 테스트 결과 전체기간에 대한 정규화보다 Window-size 별 정규화 시 0.1% ~ 0.8% 정도 더 높은 예측 정확도를 보였다.

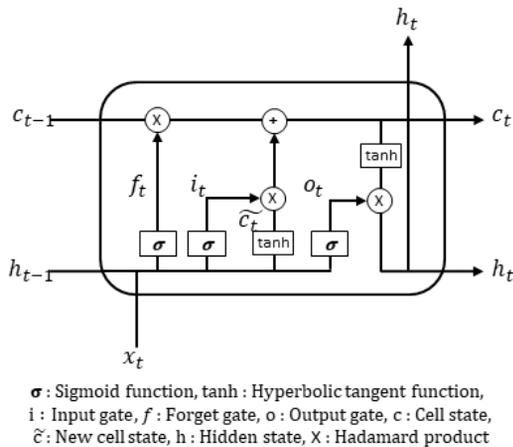
3.2 딥러닝 모델

LSTM(Long Short-Term Memory)은 Input gate, Cell state, Hidden state, Forget gate, Output gate의 다섯 가지 요소로 구성되어 있으며, 각 요소는 다음과 같은 역할을 수

Date	Close Price	Normalized
1999. 1. 15	1692	0.652542
1999. 1. 18	1745	0.877119
1999. 1. 19	1719	0.766949
1999. 1. 20		0.94491
1999. 1. 21		0.78389
1999. 1. 22		0.27966
1999. 1. 23		0.29661
1999. 1. 24		0.61864
1999. 1. 25		0.68644
1999. 1. 26		1.00000
1999. 1. 27		0.91101
1999. 1. 28		0.95339
1999. 1. 29		0.85169
1999. 1. 30		0.61864
1999. 2. 1		0.52966
1999. 2. 2		0.48728
1999. 2. 3		0.547170
1999. 2. 4		0.509434
1999. 2. 5		0.27966
1999. 2. 6		0.44491
1999. 2. 7		0.44491
1999. 2. 8		0.54661
1999. 2. 9		0.61016
1999. 2. 10		0.27118
1999. 2. 11		0.19067
1999. 2. 12		0.00000
1999. 2. 13		0.17372
1999. 2. 14		0.16525
1999. 2. 15		0.48728
1999. 2. 16		0.50423
1999. 2. 17		0.41101
1999. 2. 18		0.849057
1999. 2. 19		1.000000
1999. 2. 20		0.886792
1999. 2. 21		0.622642
1999. 2. 22		0.603774
1999. 2. 23		0.603774
1999. 2. 24		0.924528

(그림 2) LSTM Window-size 별 정규화

(Figure 2) Normalization of LSTM based on window-size

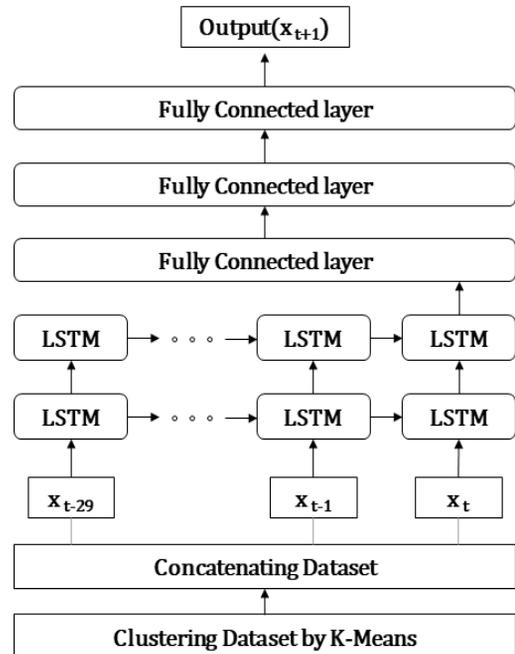


(그림 3) LSTM 구조도

(Figure 3) Architecture of LSTM

행한다. 그림 3에서 Forget gate는 기억을 삭제하고자 삭제하고자 하는 정도를 Sigmoid function을 사용해 0~1 사이의 값으로 전달한다. LSTM은 다차원 특징으로 구성되어 있어, 특징별로 어떤 정보를 저장할지 결정한다. Input gate는 Sigmoid function을 사용하여 새로운 입력을 받고자 하는 정도를 0~1 사이의 값으로 전달하고, LSTM의 다차원 특징의 구성을 통해, 특징별로 어떤 정보를 수용할지 결정한다. Cell state는 기억을 총괄하며, Hadamard 연산을 통해 Entity와 Relation 간 상호작용을 높이고, 다차원 정보를 특징별로 저장할 수 있다. Output gate는 Cell state에서 어떤 특징을 출력할지 결정하며, 마찬가지로 Sigmoid activation function을 사용하여 0~1 사이의 값을 전달한다. Hidden state는 Cell state에 tanh activation을 적용하고 Output gate와 내적하여 출력한다. 이를 통해 출력값의 범위를 -1 ~ 1로 수렴시킨다. 이러한 LSTM의 특징적 구조를 통해, 기존의 RNN에서 발생했던 기울기 소멸 (Gradient vanishing) 문제[18]와 장기의존성 (Long-term dependency) 문제를 해결할 수 있다.

그림 4는 데이터셋 전처리 과정과 예측모델 구조를 설명한다. 데이터셋은 K-means알고리즘을 통해 주가 변동 특성을 기준으로 종목별로 군집화된다. 군집화된 종목들



(그림 4) 제안하는 알고리즘의 구조

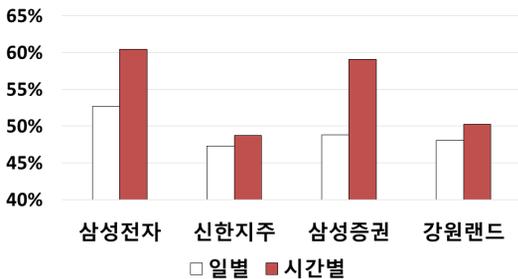
(Figure 4) Structure of the proposed algorithm

은 시간 순서를 유지하는 하나의 데이터셋으로 통합된다. 데이터셋에 포함된 입력 시퀀스는 30일의 시간 단계 길이를 Window-size 단위로 정규화되어 LSTM 기반 예측모델에 전달된다. 시퀀스의 단일 시간 단계는 주식의 시가, 고가, 저가, 종가, 거래량을 데이터로 포함한다. 예측모델은 30개의 LSTM셀로 구성된 2개 계층과 3개의 Dense Layer를 통해 다음 시간 단계의 주가 등락을 예측한다.

지도학습에서 과적합 된 모델은 학습에 활용된 데이터셋에서만 좋은 예측 성능을 보이고 실제 추론 과정에서 입력된 새로운 데이터에서는 좋은 성능을 보이지 못한다. Keras 라이브러리의 Early Stopping 함수를 활용하여 학습 중 개선이 없을 경우 이를 강제 종료하여 과도한 학습으로 인한 과적합(Overfitting) 문제를 방지한다. 일반적으로 훈련과정에서 모델의 Loss가 연속적으로 10번 증가하는 경우 학습을 종료하나, 모델을 충분히 학습시키기 위해 실험에서는 연속된 30번의 추론에서 모델의 loss를 추적하고 개선되지 않는 경우에 학습을 종료한다. 전체 데이터셋은 8대2의 비율로 훈련데이터와 테스트 데이터로 나누어 활용한다. 이를 통해 모델의 일반화 능력을 평가하고, 훈련데이터에 과적합 되는 것을 방지한다.

4. 실험 결과

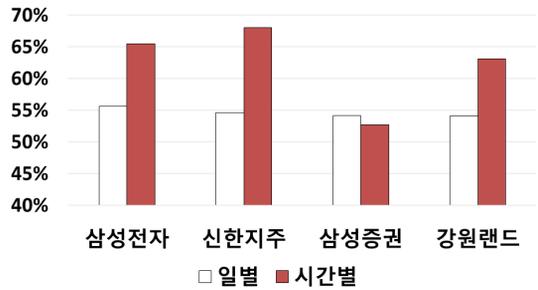
데이터 증강 없이 개별 종목으로만 학습한 경우, 시간별 데이터를 활용한 경우에 상대적으로 더 높은 예측 정확도를 보여주는 것을 확인할 수 있었다.



(그림 5) 개별종목으로 학습 후 일별과 시간별 주가예측 정확도 비교

(Figure 5) Comparison of stock price prediction accuracy between daily and hourly datasets after training on individual stocks

클러스터링 후 증강된 데이터셋을 이용하여 학습하고 테스트한 결과, 시간별 데이터를 활용한 경우에는 최대 68%까지 예측 정확도가 향상되었다.



(그림 6) 군집화 된 데이터셋으로 학습 후 일별과 시간별 주가예측 정확도 비교

(Figure 6) Comparison of stock price prediction accuracy between daily and hourly datasets after training on clustered datasets

표 4에서는 16개 종목에 대한 일자별, 시간별 데이터를 사용하여 개별 종목으로 학습한 결과와 클러스터링으로 데이터를 증강한 결과를 종합적으로 비교하였다. 그 결과, 대부분의 종목에서 클러스터링으로 증강한 데이터를 사용한 경우 예측 정확도가 향상된 것을 확인할 수 있다.

(표 4) 정확도 결과표

(Table 4) Accuracy results table

구분	개별종목	클러스터링			
		일별	시간별		
은행업	신한지주	47.3%	48.8%	54.6%	68.0%
	KB금융	46.7%	46.8%	56.4%	65.2%
	하나은행	51.5%	48.8%	54.7%	60.6%
	IBK은행	56.6%	47.3%	52.8%	51.7%
증권업	삼성증권	48.8%	59.1%	54.1%	52.7%
	한국금융지주	49.8%	47.5%	55.1%	63.6%
	미래에셋증권	51.3%	43.8%	54.4%	65.8%
	NH투자증권	53.5%	48.3%	55.6%	59.4%
전자, 전기, SW개발	삼성전자	52.7%	60.5%	55.6%	65.4%
	SK하이닉스	48.4%	52.7%	54.2%	57.1%
	삼성전기	50.2%	47.8%	50.5%	60.1%
	삼성SDS	50.1%	40.4%	46.2%	51.0%
사행시설, 유통, 지주, 운송	강원랜드	48.6%	50.3%	54.8%	63.5%
	코웨이	51.9%	55.2%	59.9%	58.9%
	CJ	50.6%	50.0%	54.4%	57.9%
	CJ대한통운	50.9%	56.4%	50.3%	44.8%

었다. 특히, 시간별 데이터를 활용한 경우 예측 정확도가 전반적으로 높았으며, 이는 주가 예측에서 시간적인 패턴이 중요한 역할을 한다는 것을 보여준다. 더불어, 유사한 주가 변동 특징을 가진 종목들을 모아 데이터를 증강하고 학습하는 것이 주가 예측을 위한 패턴을 찾는 데 유용하게 활용될 수 있음을 보여준다.

개인거래를 기준으로 주식 및 파생상품 거래에 따른 수수료와 세금을 0.21%로, 그리고 특정 종목에 대해서 Long Position(주식매수)과 Short Position(공매도)을 무조건 모두 취할 수 있다고 가정하고 수익률을 시뮬레이션 해보았다. 삼성전자지의 지난 일년간 일평균 주가 변동폭은 1.17%로, 가장 좋은 정확도인 65.43%를 보여준 모델을 활용하여, 오르는 것으로 예측했을 때는 Long Position 비중을 해당 확률만큼 높이고, 떨어지는 것으로 예측했을 경우에는 Short Position 비중을 높이는 방식을 적용하였다. 이렇게 할 경우 나올 수 있는 연간 수익률은 이론적으로는 37.6%가 된다. 엄격한 전제 조건이 있지만, 글로벌 헤지펀드의 경우 보통 연 7~15%의 절대수익[19]을 추구하는 것을 고려하면 비교적 높은 수익을 취할 수 있다.

5. 결 론

본 논문에서는 주가 등락 예측을 위한 패턴 학습에 대해 다루고 있다. 이를 위해 클러스터링을 통한 데이터 증강과 주가 측정 시간 단위를 조정하는 방식을 적용하여 성능을 높일 수 있다고 보고, 모델을 학습시켜 개별 주식의 주가 등락을 예측할 수 있도록 연구하였다. 또한, 주가의 시계열적 특성을 반영하기 위해 LSTM을 활용해 모델링을 수행하였고, 최대 68%의 예측 정확도를 보였다. 개별 종목의 주가 정보만으로 학습하는 것보다는 유사한 주가 패턴을 보이는 종목을 모아 군집화하여 학습하는 것이 더 좋은 성능을 보였다. 또한, 일별 주가 데이터를 사용하는 것보다는 시간별 주가 데이터를 사용하는 경우 더욱 높은 예측 정확도를 보여주었다.

이번 연구는 데이터를 클러스터링 할 때 특정 시점부터 현재까지의 기간을 모두 반영하여 주가 변동 패턴이 유사한 주식 종목들을 클러스터링 하는 방식을 사용한다. 그러나, 이를 여러 종목 간의 주가 수익률 그래프에서 종목 간 크로스가 발생하는 지점을 찾아 유사한 종목만을 기간별로 추려내는 방식을 적용하고, 최대 3년만 활용하고 있는 시간별 주가 데이터의 수집 기간을 늘리거나, 주가 수집 시점의 간격을 좁혀서 데이터를 추가 확보하면

개선된 결과를 보여줄 수 있을 것이다.

참고문헌(Reference)

- [1] Akhilesh Kumar, Ajay Kumar, "Macroeconomic Indicators Alone can Predict the Monthly Closing Price of Major U.S. Indices: Insights from Artificial Intelligence, Time-Series Analysis and Hybrid Models", *International Journal of Economics and Financial Issues*, Volume 7, pp. 603-613, 2017
<https://doi.org/10.1016/j.asoc.2018.07.024>
- [2] Shri Bharathi1, Angelina Geetha, "Sentiment Analysis for Effective Stock Market Prediction", 2017.
https://www.researchgate.net/publication/317214679_Sentiment_Analysis_for_Effective_Stock_Market_Prediction
- [3] Kim, Y., Jung, S., & Lee, S., "A study on predicting stock price fluctuations using sentiment analysis of social media", *Entrue Journal of Information Technology*, 2014.
<https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ARTO01936710>
- [4] Narayana Darapaneni, Anwesh Reddy Paduri, Himank Sharma, Milind Manjrekar, Nutan Hindlekar, Pranali Bhagat, Usha Aiyer, Yogesh Agarwal, "Stock Price Prediction using Sentiment Analysis and Deep Learning for Indian Markets", 2022.
<https://arxiv.org/abs/2204.05783>
- [5] Alice Zheng, Jack Jin, "Using AI to Make Predictions on Stock Market", 2017.
<http://cs229.stanford.edu/proj2017/final-reports/5212256.pdf>
- [6] Hong, S., "Automated stock trading research using macro indicators and deep learning algorithms", 2020. <https://doi.org/10.14400/JDC.2020.18.11.267>
- [7] Hwang, S., "Predicting stock market trends and optimizing investment portfolios using artificial intelligence", 2018.
http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=0f549fc955501ee7ffe0bdc3ef48d419&outLink=K

- [8] Lee, J., Stock price prediction model using deep learning, 2016.
http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=80c3f1a991ea239bffe0bdc3ef48d419&outLink=K
- [9] Jung, J., Kim, J., “Performance analysis of stock price prediction models using LSTM according to their learning methods”, Journal of Digital Convergence Vol. 18. No. 11, pp. 259-266, 2020.
<https://koreascience.kr/article/JAKO202034352379471.pdf>
- [10] David M. Q. Nelson, Adriano C. M. Pereira, Renato A. de Oliveira, “Stock Market’s Price Movement Prediction with LSTM Neural Networks”, 2017.
https://www.researchgate.net/publication/318329563_Stock_market’s_price_movement_prediction_with_LSTM_neural_networks
- [11] Joseph BONELLO, Xavier BRÉDART, Vanessa VELLA, “Machine Learning Models for Predicting Financial Distress”, 2018.
https://www.researchgate.net/profile/Joseph-Bonello-2/publication/329604057_MACHINE_LEARNING_MODELS_FOR_PREDICTING_FINANCIAL_DISTRESS/links/5c162b544585157ac1c5ce1a/MACHINE-LEARNING-MODELS-FOR-PREDICTING-FINANCIAL-DISTRESS.pdf
- [12] Lee, J., Yoo, J., “Predicting housing price indices using artificial neural networks”, 2021.
<https://doi.org/10.5762/KAIS.2021.22.4.228>
- [13] Jonathan Felix Pfahler, “Exchange Rate Forecasting with Advanced Machine Learning Methods”, 2021.
<https://doi.org/10.3390/jrfm15010002>
- [14] Kim, Y., Kim, Y., Lee, I., Lee, H., “Improving the accuracy of deep learning-based stock pattern prediction through data augmentation”, 2019.
<https://scienceon.kisti.re.kr/srch/selectPORSrchArticle.do?cn=JAKO201919163740696&dbt=NART>
- [15] Hochreiter, S., & Schmidhuber, J., “Long Short-Term Memory”, Neural Computation, Volume 9, pp. 1735 - 1780, 1997.
https://www.researchgate.net/publication/13853244_Long_Short-term_Memory
- [16] FinanceDataReader API Tutorial,
<https://github.com/financedata-org/FinanceDataReader>
- [17] Yahoo! Finance’s API Tutorial,
<https://github.com/ranaroussi/yfinance>
- [18] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio, On the Difficulty of Training Recurrent Neural Networks, 2013.
https://www.researchgate.net/publication/233730646_On_the_difficulty_of_training_Recurrent_Neural_Networks
- [19] William Fung and David A. Hsieh, “Hedge Funds: Performance, Risk and Capital Formation”, 1997.
https://www.researchgate.net/publication/4913580_Hedge_Funds_Performance_Risk_and_Capital_Formation

● 저 자 소개 ●



한 경 훈(Kyoung-hoon Han)

2008년 동국대학교 컴퓨터공학과(공학사)
2020년~현재 성균관대학교 소프트웨어학과 석사과정
관심분야 : Fintech, AI, etc.
E-mail : khhan7813@gmail.com



양 희 규(Hui-gyu Yang)

2019년 충북대학교 컴퓨터공학과(공학사)
2021년 성균관대학교 대학원 컴퓨터공학과(공학석사)
2021년~현재 성균관대학교 Superintelligence학과 박사과정
관심분야 : 차세대 무선 네트워크, AI 기반 지능형 네트워크, etc.
E-mail : huigyuu@g.skku.edu



추 현 승(Hyunseung Choo)

1988년 성균관대학교 수학과(이학사)
1990년 MS, University of Texas at Dallas, USA
1996년 Ph.D., University of Texas at Arlington, USA
1998년~현재 성균관대학교 컴퓨터공학과 교수
관심분야 : 모바일 센서 신경망, 데이터 딥 러닝, 사물인터넷, 지능형 네트워크, 의료 영상 처리
E-mail : choo@skku.edu