

기술 분석과 환경요소를 이용한 주가 예측률 향상을 위한 딥러닝 병렬 모델

황주훈*, 김창복**

Deep Learning Parallel Model to Improve Stock Price Prediction Rate using Technical Analysis and Environmental Factors

Ju-Hoon Hwang*, Chang-Bok Kim**

요 약

본 연구는 주가 데이터, 기술 분석 데이터, 환경요소 데이터를 이용하여, 주가예측을 위한 딥러닝 병렬 모델을 제안하였다. 예측을 위한 데이터 셋은 3개로 나누었으며, 데이터 셋 1은 시가, 고가, 저가, 종가, 거래량이며, 데이터 셋 2는 기술 분석 데이터를 추가하였으며, 데이터 셋 3은 주가에 영향을 줄 수 있는 환율, 전산업 생산지수를 추가하였다. 딥러닝 모델은 기본 모델로서 DNN, LSTM, 1D-CNN 모델과 병렬 모델로서 DNN 모델을 기본으로 1D-CNN을 병합한 DCNN 모델과 LSTM을 병합한 DLSTM 모델을 제안하였다. 실험 결과, DNN과 CNN 보다는 LSTM과 BiLSTM 모델의 성능이 높았으며, 특히 병렬모델인 DLSTM 모델이 가장 성능이 좋았다. 병렬 모델인 DLSTM 모델에 대한 데이터 셋 1의 RMSE는 0.0091, 데이터 셋 2의 RMSE는 0.0080, 데이터 셋 3의 RMSE는 0.0071로서 모든 데이터가 합쳐진 데이터 셋 3의 성능이 가장 좋았다.

Abstract

This study proposed a deep learning parallel model for stock price prediction using stock price data, technical analysis data, and environmental factor data. The data set for prediction was divided into three, data set 1 is the opening price, high price, low price, closing price, and trading volume, data set 2 added technical analysis data, and data set 3 is the exchange rate that can affect the stock price. the overall industrial production index was added. The deep learning model proposed DNN, LSTM, and 1D-CNN models as basic models, and a DCNN model that merged 1D-CNN based on the DNN model as a parallel model, and a DLSTM model that merged LSTM as a parallel model. As a result of the experiment, the performance of LSTM and BiLSTM models was higher than that of DNN and CNN, and in particular, the DLSTM model, a parallel model, performed the best. For the DLSTM model, which is a parallel model, the RMSE of data set 1 was 0.0091, the RMSE of data set 2 was 0.0080, and the RMSE of data set 3 was 0.0071. Data set 3, which combined all data, had the best performance.

Keywords

stock forecast, deep neural network, parallel model, 1D-CNN, ResNet

* 가천대학교 에너지 IT학과 학사과정
- ORCID: <https://orcid.org/0009-0002-5584-6161>
** 가천대학교 컴퓨터공학과 교수
- ORCID: <https://orcid.org/0000-0002-1755-5033>

• Received: Oct. 15, 2023, Revised: Nov. 10, 2023, Accepted: Nov. 13, 2023
• Corresponding Author: Chang-Bok Kim
Dept. of Computer Engineering, Gachon University, 1342, Seongnam-daero, Sujeong-gu, Seongnam-si, Gyeonggi-do, Korea
Tel.: +82-31-750-4773, Email: cbkim@gachon.ac.kr

1. 서론

주식은 잡음(Noise), 비정상성(Non-stationarity), 비선형성(Non linearity) 등으로 인해 가격변동이나 기대수익을 예측하는 것은 매우 어렵다[1][2]. 주식은 2020년 가장 큰 이슈인 코로나-19와 같은 사회적 이슈에도 영향을 받지만, 기본적으로 경제 상황에 더 민감하게 반응한다. 유수 기관에서 미래의 경제 지표를 내고 주식을 전망하지만 맞아떨어지는 경우는 극히 드물다. 주식은 분석 지표를 고도화하는 측면에서 기술 분석(Technical analysis) 외에 기본 분석(Fundamental analysis)을 추가하거나 거시경제 지표 혹은 뉴스 정보 등 주식 시장 외의 정보들을 활용해 예측의 정확도를 높이려는 연구가 지속적으로 진행되고 있다[3]-[5].

예측 분야에서 딥러닝(Deep learning)은 인간의 능력을 능가하는 수준으로 진화하고 있으며, 실세계에 사용할 수 있도록 다양한 모델을 끊임없이 연구 개발하고 있다. 이러한 딥러닝을 이용한 다양한 주식 예측에 우수한 성능을 보이는 연구가 있었다[6]-[8].

본 연구는 기본 주식 데이터와 기술 분석 데이터 그리고 다양한 예측요소 데이터를 이용하여, 주가에 예측을 위한 딥러닝 병렬 모델을 연구하였다. 기본 주식 데이터는 시가, 고가, 저가, 종가, 거래량 등이며, 기술 분석 데이터는 11개의 보조 지표를 사용하였다. 또한, 주가에 영향을 미치는 예측요소 데이터는 환율, 전 산업생산 지수 데이터를 사용하였다.

딥러닝 모델은 기본모델로서 DNN(Deep Neural Network), LSTM(Long Short Term Memory), BiLSTM(Bidirectional LSTM), 1D-CNN(Convolution Neural Network) 모델과 기본모델을 병합한 병렬 모델을 제안하였다. DNN은 중간층을 6층으로 깊게 하였으며, 기울기 소실(Gradient vanishing)을 줄이기 위해 4층과 6층에 2층과 4층의 결과를 합하여 출력하도록 구성하였다. 출력층은 예측할 증가로서 1개의 노드로 구성하였다. CNN 모델은 수치 데이터를 처리하기 위하여 1D-CNN을 사용하였다. LSTM 모델은 주가 데이터가 시계열 데이터 셋이므로 사용하였다. 병렬 모델은 DNN 모델과 1D-CNN을 병합한 DCNN 모델과 DNN 모델과 LSTM을 병합한 DLSTM 모델을 구축하였다.

본 논문은 2장에서 관련연구로서 주가 예측에 관련된 보조 지표와 예측을 위한 환경요소 그리고 딥러닝에 대해서 서술하였다. 또한, 3장에서 주가 예측 모델을 제안하였으며, 4장에서 연구 결과를 나타내고, 마지막으로 결론에 대해서 서술하였다.

II. 주가 예측과 딥러닝

2.1 주가 예측

주가는 상당 기간 같은 방향으로 움직임을 지속하려는 경향을 갖고 있다. 주가모형은 투자의 심리가 반영된 결과이며, 심리적 속성이 바뀌지 않는 한 패턴은 반복될 가능성이 높다. 일반적으로 주가 분석을 위해 기술 분석과 기본분석을 이용한다.

기술 분석은 회사의 내부가치를 무시하고 가격의 움직임에 대한 패턴만을 분석하여 예측하는 방법이다. 기본 분석은 기업의 재무제표와 경제 지표에 관련된 분석으로 회사의 가치를 이해하고 그것이 현재의 주가에 올바르게 반영되었는지를 판단하게 된다. 기술 분석은 다양한 보조 지표를 이용한다 [9]-[11]. 표 1에 본 연구에서 사용한 11개의 보조 지표 중 주가 예측에 효과적인 결과를 보이는 보조 지표와 공식에 대해서 나타냈다.

표 1. 보조 지표와 공식
Table 1. secondary indicators and formula

Secondary indicators	Formula
Moving average	$MA_t^n = \frac{1}{n} \left(\sum_{i=t-n+1}^t close_i \right)$
Fast stochastic	$K_f = \frac{close_t - MIN(t:t-m)}{MAX(t:t-m) - MIN(t:t-m)}$
slow stochastic	$K = \frac{K_{f1} + K_{f2} + K_{f3}}{3}$
CCI (Commodity Channel Index)	$CCI = \frac{M_t - MA_t^n}{D \times 0.015}$ $M_t = \frac{high_t + low_t + close_t}{3}$ $D = \frac{1}{n} \left(\sum_{i=t-n+1}^t M_t - MA_t^n \right)$

이동평균(Moving average)은 주가 및 거래량이 특정기간의 방향성을 수치화한 것이다.

지수 이동평균(Exponential moving average)은 최근 값에 좀 더 큰 가중치를 부여하는 방법이다. MACD(Moving Average Convergence and Divergence)는 지수 이동평균 값 간의 거리를 나타낸다.

골든 크로스(GC, Golden Cross)는 단기 이동평균선이 장기 이동평균선을 상향으로 돌파하는 시점으로 매수 신호로 간주한다. 반대인 경우는 데스크로스(DC, Death Cross)로 매도 신호로 간주한다[12]. 이격도(Disparity)는 주가와 이동 평균선과의 거리이며, 이동 평균선의 단점을 보완한 지표이다. RSI(Relative Strength Index)는 현재의 주가 추세 강도를 백분율로 나타내어, 추세 전환을 예측하는 지표이다. CCI(Commodity Channel Index)는 이동평균으로부터 주가의 변동성을 측정하는 지표로 추세의 강도와 방향을 표시한다. 스토캐스틱(Stochastic Oscillator)은 최근 N 일간의 최고가와 최저가의 범위 내에서 현재 가격의 위치를 백분율로 표시한 지표이다.

주가에 영향을 미치는 요소들은 금리, 환율, 원자재, 경기변동 등이 있다. 이 요소들은 주식뿐만 아니라 경제의 상태를 알 수 있으며, 금융의 흐름을 파악할 수 있다. 기본적으로 금리가 내리면 주가는 오른다. 금리는 은행 예금 이자율을 의미하며 금리가 높을 때 적금만으로도 큰 수익을 볼 수 있다. 그러나 금리가 낮아지면 적금보다는 다른 투자처를 찾게 된다. 가장 대표적인 예가 부동산이나 주식 시장이다. 부동산의 경우에는 상대적으로 큰 자본이 필요하므로 접근성이 비교적 떨어진다. 따라서 주식 시장으로 몰릴 가능성이 크다. 또한 기업은 낮은 금리로 인해 쉽게 대출할 수 있으므로 사업하기 좋은 환경이 조성된다. 그러므로 주가 흐름은 상승 방향으로 나타날 가능성이 크다.

환율은 타국의 화폐 가치에 따른 자국의 화폐 가치이다. 환율이 상승하면 수출 가격은 상승하기 때문에, 우리나라와 같은 수출국에서는 중요하다고 할 수 있다. 또한, 환율이 낮으면 외국인 자본이 국내 시장으로 많이 유입되어 주식 시장이 좋아질 수 있다. 대한민국은 제조업을 기반으로 하는 수출 국가이기 때문에 원자재 가격도 중요하다. 원자재 가격이 상승하게 되면 기업으로서는 수익성이 낮아질 수밖에 없다. 기업의 실적이 낮아지고 주가는 하락할 가능성이 커진다. 원자재 가격은 다양한 변수들

에 의해서 결정되기 때문에 쉽게 예측하기 힘들다. 하지만 기본적으로 원자재 가격이 상승하면 주가는 하락할 여지가 크다.

주식 시장은 경기회복, 경기활황, 경기침체, 경기후퇴를 반복한다. 이러한 사이클은 GDP 경제성장률과 연관성이 크기 때문에 현재 우리나라의 경제성장률 및 GDP를 살펴볼 필요가 있다. 표 2에 주가에 영향을 미치는 요소들에 대해서 나타냈다.

표 2. 주가에 영향을 미치는 요소들
Table 2. Factors affecting stock prices

Variable	Change	Stock price impact
Interest rate	increase	degradation
	reduction	discount
Exchange rate	rise	degradation
	discount	Increase
Raw materials	price rise	Increase
	price discount	degradation
Economic change	boom	Increase
	downturn	degradation

2.2 딥러닝 모델

본 연구는 기본모델과 병렬 모델을 이용하였다. 기본모델은 DNN, 1D-CNN, Multi kernel 1D-CNN, LSTM, BiLSTM 모델을 이용하였다.

DNN은 중간층을 6층으로 깊게 하였으며, 기울기 소실을 줄이기 위해 ResNet 모델을 적용하였다. LSTM은 DNN을 병렬로 연결한 모델로서 순차적인 시계열 데이터 학습에 적합한 모델이다. 1-D CNN은 패턴을 학습하는 모델이다.

기존 연구에서는 주식 예측에 RNN과 LSTM 모델을 이용하여 예측 결과를 산출하였다[13]. 그러나 본 연구는 다양한 모델을 병렬로 연결하여 주식 예측을 하고자 한다. 주식 예측 병렬 모델은 동일한 모델 또는 서로 다른 모델을 병렬로 구성한 모델이다. 병렬 모델은 각 기본모델로 중간층을 구성하고, 모델의 결과를 연결하여 추가로 밀집 모델을 연결한 모델이다. 본 연구에서 병렬 모델은 DNN과 Multi kernel 1D-CNN을 병렬 연결한 DMCNN 모델 그리고 DNN과 LSTM을 병렬연결한 DLSTM 모델을 구축하였다. 그림 1에 병렬 모델에 대해서 나타냈다.

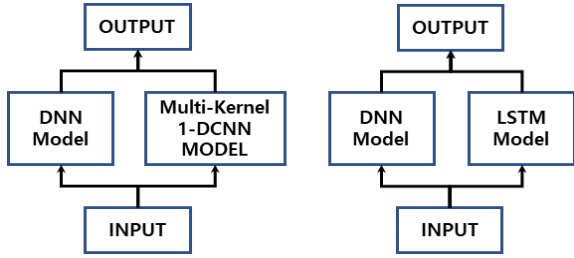


그림 1. 병렬 모델
Fig. 1. Parallel models

III. 주식 예측 모델

3.1 주식 예측 모델

3.1.1 기본모델

DNN 모델은 입력층은 주식 데이터 입력 노드이며, 중간층 6층, 중간층 노드 256 노드로 구성하였다. 또한, 기울기 소실(Gradient vanishing)을 줄이기 위해 4층과 6층에 2층과 4층의 결과를 합하여 출력하도록 구성하였다. 출력층은 예측할 증가로서 1개의 노드로 구성하였다.

LSTM과 BiLSTM은 DNN과 입력은 동일하며, 타임스탬프(Time stamp)는 일주일 장이 5일이므로 5개로 하였으며, 은닉층 노드 128개, 은닉층의 결과를 256 노드의 밀집(Dense) 층과 32개의 밀집 중간층으로 구성하였다. 출력층은 예측할 증가로서 1개의 노드로 구성하였다. 표 3에 DNN과 LSTM 기본모델에 대해서 나타냈다.

표 3. DNN, LSTM, BiLSTM 기본모델
Table 3. DNN, LSTM, BiLSTM basic model

Model	Input layer		Hidden		Dense	Output layer
	Time stamp	Input	Layer	Node	Dense 1 Dense 2	
DNN		5, 10, 14	6	256		1
LSTM	5	5, 10, 14	1	128	256 32	1
BiLSTM	5	5, 10, 14	1	128	256 32	1

입력층은 기본데이터가 5개 입력, 기술 분석 데이터가 5개, 환경 요소데이터가 4개이므로 5, 10, 14

입력이 전달한다, 그림 2와 그림 3에 DNN과 LSTM 기본모델에 대해서 나타냈다.

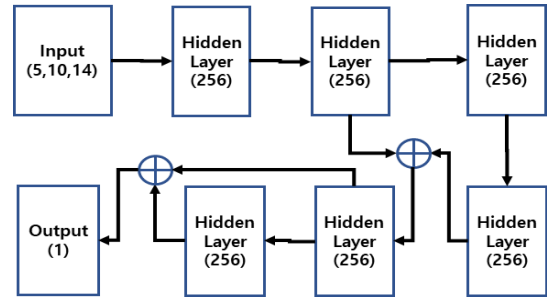


그림 2. DNN 기본모델
Fig. 2. DNN basic models

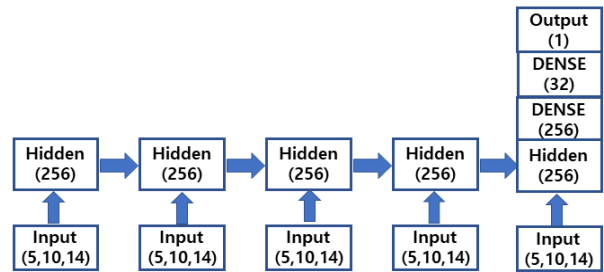


그림 3. LSTM 기본모델
Fig. 3. LSTM basic models

1D-CNN 모델의 입력층은 주식 데이터 입력 노드이며, 합성곱(Convolution) 층으로 필터 256개, 커널 3, 스트라이드 1로 구성하였으며, 풀링(Pooling)층으로 맥스(Max) 풀링을 이용하였다. 또한, 풀링 결과는 커널 수와 동일한 256개의 스칼라 값이며, 이를 256 노드의 밀집층과 32개의 밀집층으로 구성하였다. 출력층은 예측할 증가로서 1개의 노드로 구성하였다.

Multi kernel 1D-CNN 모델의 입력층은 주식 데이터 입력 노드이며, 합성곱층으로 필터 256개, 커널 3, 4, 5로 구성하였다. 따라서 3개의 커널을 거친 특성값은 1, 2, 3개 노드를 가진 256개의 벡터가 3개로 구성되며, 맥스 풀링의 값은 768개의 스칼라 값이다. 이를 256 노드의 밀집 중간층과 32개의 밀집 중간층으로 구성하였다. 출력층은 예측할 증가로서 1개의 노드로 구성하였다.

표 4에 1-CNN과 Multi kernel 1D-CNN 기본모델에 대해서 나타냈다. 또한, 그림 4에 1-D CNN에 대해서 나타냈으며, 그림 5에 Multi kernel 1D-CNN 구조에 대해서 나타냈다.

표 4. CNN 기본모델

Table 4. CNN basic model

model	input layer	Convolution		Pooling result	output layer	
		kernel	filter		Dense1	Dense2
1D-CNN	5, 10, 14	3	256	256	256	1
					32	
1D-MCNN	5, 10, 14	3,4,5	256	768	256	1
					32	

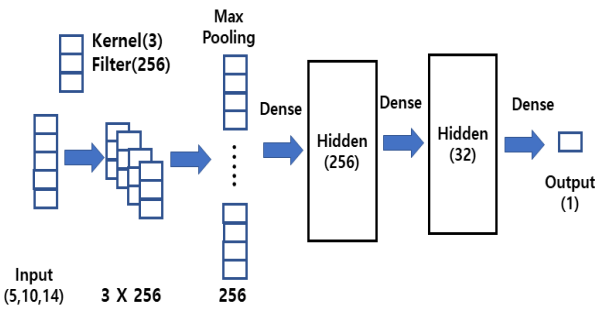


그림 4. 1D-CNN 모델
Fig. 4. 1D-CNN model

3.1.2 병렬 모델

병렬 모델은 DMCNN 모델과 DLSTM 등 2개의 모델을 제안하였다. DMCNN 모델은 DNN 모델과 Multi kernel 1D-CNN 모델을 병렬한 모델로서 DNN 모델과 Multi kernel 1D-CNN 모델의 출력을 8노드로 하여, 출력 결과를 연결한 16노드에 대해서 256노드의 밀집층과 32개의 밀집층으로 구성하였다. DLSTM 모델은 DNN 모델과 LSTM 모델을 병렬한

모델로서 DNN 모델과 LSTM 모델의 출력을 8노드로 하여, 출력 결과를 연결한 16노드에 대해서 256노드의 밀집층과 32개의 밀집층으로 구성하였다. 표 5에 병렬 모델 구조에 대해서 나타냈다. 그림 6에 병렬 모델에 대해서 나타냈다.

표 5. DMCNN과 DLSTM 병렬 모델

Table 5. DMCNN and DLSTM parallel model

Model	Network	Output1	Hidden1	Hidden2	Output
DMCNN	DNN	8	256	32	1
	MCNN	8			
DLSTM	DNN	8	2564	32	1
	LSTM	8			

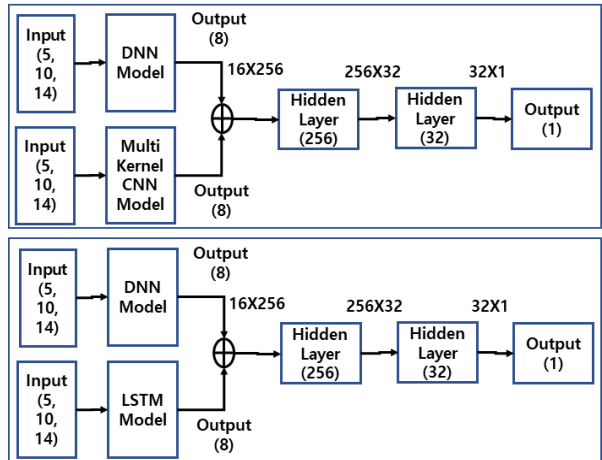


그림 6. DMCNN과 DLSTM 병렬 모델
Fig. 6. DMCNN and DLSTM parallel model

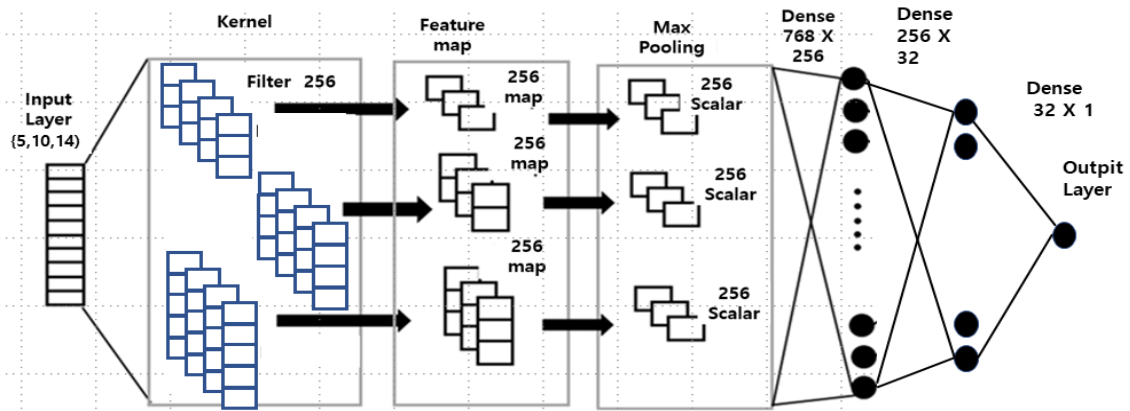


그림 5. Multi kernel 1D-CNN 구조
Fig. 5. Multi kernel 1D-CNN structure

3.2 실험 데이터

주가 데이터는 2000년 2월에서 2022년 12월 기간의 가격 안정성이 높은 KOSPI 200 종목에서 4 종목을 선택하였다. 전체 샘플은 4767개이며, 학습데이터는 4400개 테스트 데이터는 367개이다.

기본 데이터는 일별 시가, 고가, 저가, 거래량, 종가 데이터이며, 기본분석 데이터로 표 1을 이용하여 보조 지표 데이터를 사용하였으며, 또한, 주가 예측에 필요한 환경요소인 환율과 전산업생산지수를 사용하였다. 본 실험은 3개의 데이터 셋으로 구성하였다. 데이터 셋 1은 시가, 고가, 저가, 종가, 거래량 등의 기본 주가 데이터이며, 데이터 셋 2는 데이터 셋 1을 포함하여, 기술 분석을 이용한 11개의 보조 지표 중 우수한 예측 결과를 가진 지표인 5일 이동평균, 20일 이동평균, CCI, Fast Stochastic 등의 보조 지표 데이터이다[13] 데이터 셋 3은 데이터 셋 2을 포함하여, 환율, 환율 5일 이동 평균, 환율 20일 이동 평균, 전산업생산지수 등 환경 요소 데이터로 구성되어 있다.

데이터 셋 1은 파이선의 finance 모듈을 통해 크롤링하였으며, 데이터 셋 2는 엑셀로 보조 지표(이동평균, 스토캐스틱, CCI) 구하여 작성 하였다. 또한, 데이터 셋 3은 FinanceDataReader 모듈을 통해 환율 데이터 크롤링하고, 통계청에서 전산업생산지수 데이터 다운로드하여 전부 통합 후 작성하였다.

표 6. 데이터 셋
Table 6. Dataset

DataSet	Independent and dependent variables
DataSet 1	Open, high, low, volume, close (Price)
DataSet 2	DataSet1 and 5-day and 20-day moving average, CCI, fast stochastic, slow stochastic
DataSet 3	DataSet2 and exchange rate, exchange rate moving average, index of all industry production

병렬 모델은 모델마다 입력 데이터의 형태가 다르기 때문에, 입력 데이터의 매칭이 필요하다. LSTM 데이터 구축 시 5번째 샘플까지 타임스탬프에 입력되는 데이터이며, 6번째 샘플의 종가가 라벨

이다. 따라서 DNN과 CNN의 5개 샘플을 제거해야만 LSTM의 실제 데이터와 일치한다.

IV. 실험 및 성능평가

본 실험은 Window10 환경에서 Anaconda, Python 3.6, Keras 2.6.0을 사용하였다,

모델의 하이퍼 파라미터는 학습 알고리즘으로 Adam을 사용하였으며, 학습률은 0.001, 배치 크기는 1로 하여, 50번 반복 학습하였다. 모델의 중간층 활성화 함수는 ReLU를 사용하였으며, 출력층 활성화 함수는 linear를 사용하였다. 표 7에 실험에 사용한 하이퍼 파라미터에 대해서 나타냈다.

표 7. 하이퍼 파라미터
Table 7. Hyper parameter

Hyper parameter	Value
Learning algorithm	Adam
Learning rate	0.001
Batch size	1
Epoch	50
Hidden activation function	Relu
Output activation function	Linear

데이터는 각 독립변수의 크기가 다르기 때문에, 데이터의 범위를 일치시켜야 한다. 또한, 딥러닝은 입력 데이터의 크기가 작아야 학습이 효과적이기 때문에, 0과 1 사이로 정규화 하였다. 모델의 성능 평가는 RMSE(Root Mean Square Error)를 사용하였다. 실험은 3번 반복하여 평균을 나타냈으며, 각 결과에 대해 표준 편차를 나타냈다.

$$Normalize = \frac{d_i - MIN(d)}{MAX(d) - MIN(d)} \quad (1)$$

$$RMSE(x', x) = \frac{1}{N} \sqrt{\sum_{n=1}^N (x'_n - x_n)^2} \quad (2)$$

$$STDEV = \sqrt{\frac{\sum_{n=1}^N |x_n - \mu|^2}{N}} \quad (3)$$

식 (1)에서 d_i 는 실제 데이터이며, $\text{MIN}(d)$ 는 각 행에서 가장 작은 값이고 $\text{MAX}(d)$ 는 가장 큰 값이다. 식 (2)에서 N 은 샘플수이고 x'_n 은 오차값이며, x_n 은 실제값이다. 식 (3)에서 x_n 은 실제값이고 μ 는 평균 값을 나타낸다.

표 8. 기본 데이터 결과
Table 8. Basic data results

Model		Co. 1	Co. 2	Co. 3	Co. 4	AVG
DNN	RMSE	0.1824	0.1809	0.2461	0.2588	0.2170
	STDEV	0.0013	0.0017	0.0016	0.0022	0.0017
CNN	RMSE	0.1775	0.1790	0.2434	0.2476	0.2119
	STDEV	0.0024	0.0009	0.0033	0.0023	0.0022
MCNN	RMSE	0.1807	0.1806	0.2447	0.2521	0.2145
	STDEV	0.0019	0.0028	0.0029	0.0023	0.0025
LSTM	RMSE	0.0219	0.0291	0.1203	0.0373	0.0521
	STDEV	0.0035	0.0058	0.0993	0.0009	0.0274
BiLSTM	RMSE	0.0260	0.0208	0.0304	0.0288	0.0265
	STDEV	0.0092	0.0065	0.0088	0.0066	0.0078
DMCNN	RMSE	0.1781	0.1745	0.2472	0.2516	0.2129
	STDEV	0.0012	0.0004	0.0005	0.0024	0.0011
DLSTM	RMSE	0.0070	0.0063	0.0174	0.0057	0.0091
	STDEV	0.0020	0.0017	0.0099	0.0005	0.0035

기본 데이터인 데이터 셋 1의 실험 결과 기본 모델인 DNN, CNN, MCNN 그리고 병렬 모델인 DMCNN의 오차는 비슷하였으나 시계열 예측에 적합한 LSTM과 BiLSTM 모델의 성능이 월등히 높음을 알 수 있다. 또한, 병렬 모델 중에서 DLSTM 모델의 예측결과가 가장 우수한 것으로 나타났다.

표 9. 기본 데이터와 보조 지표 데이터 결과
Table 9. Primary data and secondary indicator data results

Model		Co. 1	Co. 2	Co. 3	Co. 4	AVG
DNN	RMSE	0.1835	0.1794	0.2442	0.2603	0.2168
	STDEV	0.0004	0.0019	0.0005	0.0011	0.0010
CNN	RMSE	0.1742	0.1693	0.2415	0.2663	0.2128
	STDEV	0.0028	0.0044	0.0036	0.0016	0.0031
MCNN	RMSE	0.1725	0.1760	0.2459	0.2424	0.2092
	STDEV	0.0009	0.0008	0.0040	0.0019	0.0019
LSTM	RMSE	0.0342	0.0572	0.0176	0.1286	0.0594
	STDEV	0.0044	0.0073	0.0000	0.0051	0.0042
BiLSTM	RMSE	0.0358	0.0353	0.0259	0.1192	0.0540
	STDEV	0.0031	0.0068	0.0019	0.0038	0.0039
DMCNN	RMSE	0.1822	0.1759	0.2407	0.2527	0.2129
	STDEV	0.0009	0.0014	0.0001	0.0021	0.0011
DLSTM	RMSE	0.0093	0.0063	0.0106	0.0060	0.0080
	STDEV	0.0021	0.0011	0.0017	0.0010	0.0015

기본 데이터와 보조 지표 데이터인 데이터 셋 2의 결과에서 대부분 기본 데이터와 비슷한 결과를 나타냈으나, 병렬 모델인 DLSTM 모델의 결과는 0.0080으로 기본 모델 보다 좀 더 오차가 적었음을 알 수 있다.

표 10. 모든 데이터 결과
Table 10. All data results

Model		Co. 1	Co. 2	Co. 3	Co. 4	AVG
DNN	RMSE	0.1003	0.1012	0.0860	0.1732	0.1152
	STDEV	0.0001	0.0008	0.0015	0.0006	0.0290
CNN	RMSE	0.1046	0.1054	0.0844	0.1736	0.1170
	STDEV	0.0003	0.0041	0.0019	0.0003	0.0283
MCNN	RMSE	0.1069	0.1019	0.0842	0.1768	0.1175
	STDEV	0.0036	0.0025	0.0007	0.0015	0.0297
LSTM	RMSE	0.0190	0.0237	0.0248	0.0175	0.0213
	STDEV	0.0046	0.0077	0.0101	0.0007	0.0030
BiLSTM	RMSE	0.0230	0.0202	0.0170	0.0183	0.0196
	STDEV	0.0063	0.0031	0.0014	0.0030	0.0020
DMCNN	RMSE	0.1023	0.1005	0.0879	0.1719	0.1157
	STDEV	0.0007	0.0009	0.0013	0.0018	0.0281
DLSTM	RMSE	0.0060	0.0064	0.0091	0.0069	0.0071
	STDEV	0.0016	0.0020	0.0058	0.0023	0.0001

데이터 셋 3의 결과는 데이터 셋 1과 데이터 셋 2의 결과보다 전체적으로 우수한 결과를 보였음을 알 수 있다.

모든 실험 결과 병렬 모델인 DLSTM 모델의 예측결과가 가장 우수했으며, 데이터 셋 1의 RMSE는 0.0091, 데이터 셋 2의 RMSE는 0.0080, 데이터 셋 3의 RMSE는 0.0071이다. 그림 7, 8, 9, 10에 DNN, CNN, LSTM, DLSTM 모델에 대한 결과 그래프를 나타냈다.

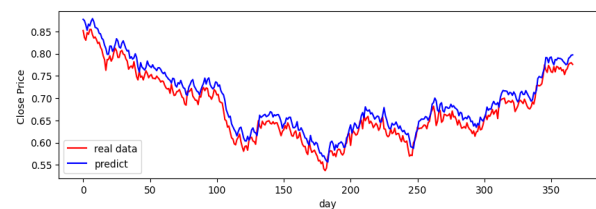


그림 7. DNN 모델 결과 그래프
Fig. 7. DNN model result graph

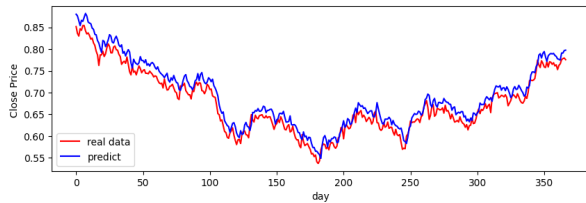


그림 8. CNN 모델 결과 그래프
Fig. 8. CNN model result graph

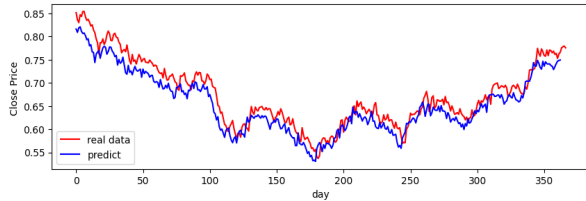


그림 9. LSTM 모델 결과 그래프
Fig. 9. LSTM model result graph

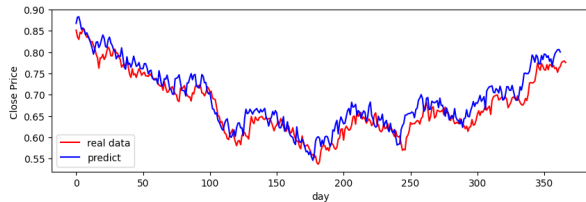


그림 10. DLSTM 모델 결과 그래프
Fig. 10. DLSTM model result graph

V. 결론 및 향후 과제

본 연구는 기본 주가 데이터와 기술 분석 데이터 그리고 다양한 예측 환경요소 데이터를 이용하여, 주가예측을 위한 딥러닝 병렬 모델을 제안하였다. 본 실험에 사용한 주식 데이터는 시계열 데이터이므로, DNN과 CNN 보다는 시계열 예측에 적합한 LSTM과 BiLSTM 모델의 성능이 월등히 높음을 알 수 있다. 또한, 병렬 모델 중에서 본 논문에서 제안한 DLSTM 모델의 예측결과가 가장 우수한 것으로 나타났다. 본 연구를 통해 기존의 주식 예측에 RNN과 LSTM 모델만을 사용한 것 보다 우수한 예측 결과를 산출한 것을 알 수 있다.

보다 정확한 주가 예측을 위해서는 주가 예측에 필요한 다양한 환경 데이터 추가가 필요하며, 예측 모델도 주식에 관련된 자연어를 처리하기 위한 BERT(Bidirectional Encoder Representations from Transformers)와 같은 모델도 추가로 병렬처리가 필요할 것이다.

References

- [1] E. J. Lee, C. H. Min, and T. S. Kim, "Development of the KOSPI (Korea Composite Stock Price Index) forecast model using neural network and statistical methods", The Institute of Electronics Engineers of Korea, Computer and Information, Vol 45, No. 5, pp. 95-101, Sep. 2008.
- [2] K. Y. Kim and K. R. Lee, "A Study on the Prediction of Stock Price Using Artificial Intelligence System", Korean Journal of Business Administration, Vol. 21, No. 6, pp. 2421-2449, Dec. 2008.
- [3] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions", Artificial Intelligence Review, Vol. 53, pp. 3007-3057, Aug. 2019. <https://doi.org/10.1007/s10462-019-09754-z>.
- [4] E. Jang, H. R. Choi, and H. C. Lee, "Stock prediction using combination of BERT sentiment Analysis and Macro economy index", Journal of the Korea Society of Computer and Information, Vol. 25, No. 5, pp. 47-56, May 2020. <https://doi.org/10.9708/jksci.2020.25.05.047>.
- [5] S. W. Ahn and S. B. Cho, "Stock Prediction Using News Text Mining and Time Series Analysis", Proc. of the Korean Information Science Society Conference, Vol. 37, No. 1C, pp. 364-369, Jun. 2010.
- [6] S. M. Nam, J. E. Kim, and Z. K. Lee, "Time-Invariant Stock Movement Prediction After Golden Cross Using LSTM", Journal of The Korea Society of Computer and Information, Vol. 28, No. 8, pp. 59-66, Aug. 2023. <https://doi.org/10.9708/jksci.2023.28.08.059>.
- [7] H. Choi, "Stock prediction analysis through artificial intelligence using big data," Journal of the Korea Institute of Information and Communication Engineering", Vol. 25, No. 10, pp.

1435-1440, Oct. 2021. <http://doi.org/10.6109/jkiice.2021.25.10.1435>.

- [8] T. D. Han, "Stock Price Prediction Using LSTM: Focusing on the Combination of Technical Indicators, Macroeconomic Indicators, and Market Sentiment", *The Society of Convergence Knowledge Transactions*, Vol. 9, No. 4, pp. 189-198, Dec. 2021. <https://doi.org/10.22716/sckt.2021.9.4.055>.
- [9] K. J. Jeong and J. S. Choi, "Deep Recurrent Neural Networks", *Communications of the Korean Institute of Information Scientists and Engineers*, Vol. 33, No. 8, pp. 39-43, Aug. 2015.
- [10] A. Y. Wiiava, C. Faticah, and A. Saikhu, "Stock Price Prediction with Golden Cross and Death Cross on Technical Analysis Indicators Using Long Short Term Memory", in *ICOIACT 2022 - 5th International Conference on Information and Communications Technology*, Yogyakarta, Indonesia, pp. 278-283, Aug. 2022. <https://doi.org/10.1109/ICOIACT55506.2022.9971844>.
- [11] T.-J. Hsieh, H.-F. Hsiao, and W.-C. Yeh, "Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm", *Applied Soft Computing*, Vol. 11, No. 2, pp. 2510-2525, Mar. 2011. <https://doi.org/10.1016/j.asoc.2010.09.007>.
- [12] Y.-S. Tsai, C.-P. Chang, and S.-W. Tzang, "The Impact of Golden Cross and Death Cross Frequency on Stock Returns in Pre- and Post-financial Crisis", *IMIS 2017: Innovative Mobile and Internet Services in Ubiquitous Computing*, Torino, Italy, Vol. 612, pp. 708-713, Jul. 2017. https://doi.org/10.1007/978-3-319-61542-4_70.
- [13] 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, Oct. 2017. <http://dx.doi.org/10.14801/jkiit.2017.15.10.9>.

저자소개

황 주 훈 (Ju-Hoon Hwang)



2018년 3월 ~ 현재 : 가천대학교
에너지 IT학과 학사과정
관심분야 : 머신러닝, 인공지능

김 창 복 (Chang-Bok Kim)



1986년 2월 : 단국대학교
전자공학과(공학사)
1989년 2월 : 단국대학교
전자공학과(공학석사)
2009년 2월 : 인천대학교
컴퓨터공학과(공학박사)
1994년 ~ 현재 : 가천대학교

IT대학 컴퓨터공학과 교수
관심분야 : 사물인터넷, 딥러닝, 클라우드 컴퓨팅