

딥러닝을 활용한 실시간 주식거래에서의 매매 빈도 패턴과 예측 시점에 관한 연구: KOSDAQ 시장을 중심으로

송현정* · 이석준**

〈목 차〉

I. 서론	IV. 실증분석 및 결과
II. 선행연구 고찰	4.1 데이터 수집 및 기간설정
2.1 통계학적 접근방법	4.2 패턴 생성 및 라벨링 결과
2.2 머신러닝과 딥러닝 접근방법	4.3 패턴 이미지 학습 및 비교 결과
III. 연구방법	V. 결론
3.1 패턴 생성 및 라벨링	참고문헌
3.2 패턴 이미지 학습	<Abstract>
3.3 매매 빈도 패턴과 예측 시점 탐색	

I. 서론

주식시장은 경제 및 정치적 변수를 포함한 다양한 상황에 영향을 받으므로 주식시장을 예측하는 것은 매우 어려운 문제로 여겨지고 있다. 이러한 주식시장을 예측하는 방법으로는 기본적 분석(fundamental analysis)방법과 기술적 분석(technical analysis)방법이 있다. 기본적 분석방법은 기업의 과거와 현재 상태를 보고 현재 주식의 가치와 미래의 주가를 분석(조희연, 김영민, 2003)하는 방법이다. 기술적 분석방법은 기업의 주가와 거래량의 흐름을 보고 미래

의 주가 흐름을 예측하는 방법을 말하며, 컴퓨터와 인터넷 기술의 발달로 인하여 손쉽게 차트를 접할 수 있게 됨에 따라 기술적 분석을 이용하는 투자자가 증가하는 추세이다(이운선, 2002). 기술적 분석은 주식의 가격 및 거래량에 의해 계산되는 다양한 기술적 지표들을 분석하는 것(오성민, 김성진, 2000; 김상환, 조태근, 2003; 김범준 등, 2005; 이운선, 2006;)과 차트 데이터를 이용한 분석 방법이 있다(이강희, 1997; 이운선, 2006; 김성동, 2012; 이재원, 2013). 차트 데이터는 분 단위(5분, 10분, 30분, 60분 등), 일(daily) 단위 등의 다양한 빈도로 시가(open), 고가(high), 저가(low), 종가(close),

* 광운대학교 경영학부, hyunjung12@kw.ac.kr(주저자)

** 광운대학교 경영학부, sjlee@kw.ac.kr(교신저자)

거래량(volume) 데이터를 이용하여 봉 모양을 작성하며 주식 가격의 흐름과 변동을 요약해준다(이재원, 2013).

기술적 분석 연구들은 최근 인공지능 기법에 대한 관심이 커지면서 이를 적용한 연구로 발전하고 있다(황희수, 2008; 김유일 등, 2004; 박강희, 신현정, 2011; 오세창, 2014; 이우식, 2017). 인공지능 기법 중 딥러닝(deep learning)은 인간의 뇌 구조를 모방한 기계학습이며, 기존 인공신경망(artificial neural network)의 문제점을 개선하여 영상인식, 음성인식, 자연어 처리, 미래 예측 등 다양한 영역에서 활용되고 있다. 딥러닝은 다른 형태의 기계학습과는 달리 입력 데이터의 특징 패턴을 있는 그대로 학습하기 때문에 입력 데이터의 특징 패턴에 따라 예측결과가 큰 차이가 있다는 특징이 있다(안성만, 2016). 기존 기술적 분석에 집중한 인공지능 연구는 MACD(moving average convergence divergence)를 포함한 이동 평균(moving average) 기반의 기술적 지표(technical indicator)들에 집중하여 연구를 진행하였다(성노윤, 남기환, 2017; 신동하 등, 2017; 송유정, 이종우, 2017; Nelson et al., 2017). 하지만 기술적 지표는 상당 부분 주가 추세가 진행된 이후에 매수 및 매도 신호가 발생하는 '추세 후행성'이라는 한계를 가지고 있다(이재원, 2017). 반면 차트 데이터를 이용한 분석은 다양한 주가 변동 유발요인이 존재한다 해도 최대한 인간의 개입을 최소화하여 차트의 모양, 주가의 추세에 따른 변동 패턴(pattern)을 찾아낼 수 있다면 정치·경제·사회적 사건 등과 관계 없이 주가의 변동을 예측할 수 있다(송유정, 이종우, 2017; 송유정 등, 2017)는 주장이 제기되

었다. 따라서 성공적인 주식시장 분석을 위해 차트 데이터를 이용한 많은 패턴 인식 방법론이 개발되었으며, 많은 투자자들은 주가 패턴을 관찰하고 이를 결정적인 요소로 믿는 것에 큰 관심을 가지고 있다(Shih et al., 2008; Chiu and Chian, 2010; Chen and Lin, 2012; Lee and Jeong, 2012). 그 결과 주식 패턴을 예측하기 위한 기술적 분석 연구가 활발하게 수행되고 있다(이재원, 2013; 유재필 등, 2017; 박상민 등, 2018).

따라서 본 연구는 차트 이미지 데이터를 대상으로 딥러닝에 적용하여 주식시장의 대표적인 주가패턴을 학습하고 미래 주가의 등락을 예측하고자 한다. 차트 데이터가 5분, 10분, 30분, 60분 등의 분 단위와 일(daily)단위 등 다양한 단위의 빈도로 작성됨에도 불구하고 기존 연구들은 주식시장을 예측할 때 일 단위의 차트 데이터에 집중하여 연구를 진행하였다(이강희 등, 1997; 이종우 등, 2003; 이재원, 2013). 따라서 본 연구는 5분, 10분, 30분, 60분과 같이 분 단위 데이터를 이용하여 주가 등락 예측에 유용한 매매 빈도를 탐색하고자 한다. 또한 패턴 분석을 통해 주가 등락 예측을 수행할 적절한 시점(timing)을 확인하고자 한다. 예측 시점은 몇 단위 이전에 생성된 패턴이 주가 등락 예측 확률을 높이는지 탐색하는 것을 의미하며 기술적 분석을 수행하는 투자자는 주식 거래 차트를 어느 시점에 사용하여 거래를 할 것인지를 결정하는 것이 중요하다(Deboeck, 1994). 따라서 본 연구에서는 한 단위 빈도 전(이하 -1 tick) 두 단위 빈도 전(이하 -2 tick), 세 단위 빈도 전(이하 -3 tick) 별 분석을 제안한다. 이를 위해 본 연구는 다양한 패턴을 보여주는

KOSDAQ 상위 150개 종목(김종선 등, 2014)을 선택하였으며 딥러닝 알고리즘 중 패턴인식에 적합하다고 평가되는 합성곱 신경망(convolutional neural network, 이하 CNN)으로 학습하여 매매 빈도 패턴과 예측 시점을 탐색한다. 실증분석 결과 임계치가 0.5인 10분 단위 빈도 데이터가 -1 tick에서 하락패턴을 보이면 주가 하락으로 장이 종료된다고 예측하는 것이 76%의 예측정확도를 보인 것으로 확인되었으며, 매매 빈도는 10분 단위, 예측 시점은 -1 tick에서 예측하는 것이 -2 tick과 -3 tick에서 예측하는 것보다 높은 예측정확도를 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구의 이론적 배경이 되는 여러 주가 예측 방법들을 소개하고, 3장에서는 실시간 주식거래를 위한 매매 빈도 패턴 및 주가 예측 타이밍에 대해 설명한다. 4장에서는 제안 모델의 실증분석 결과를 제시하였으며, 마지막 장 5장에서는 본 연구의 과정 및 결과를 간략하게 제시한 후 본 연구의 의의 그리고 한계 및 향후 연구 방향을 제시한다.

II. 선행연구 고찰

주식시장 예측을 위해 경제, 수학, 물리, 그리고 전산학 등의 분야에서 다양한 방법으로 접근하고 있으며, 본 연구에서는 통계학적 접근방법과 머신러닝과 딥러닝 접근방법과 관련된 선행연구를 아래와 같이 정리하였다.

2.1 통계학적 접근방법

과거에는 주식시장 데이터에 근거하여 미래를 예측하는 통계학적 접근 방법이 널리 사용되었다(김진백, 김유일, 2000). 통계학적 기법은 숫자와 같이 객관화된 데이터를 정량적으로 분석하는 것(양유영 등, 2009)으로 주요 방법으로는 이동평균 분석과 몬테카를로(Monte Carlo) 시뮬레이션이 대표적이다.

이동평균 분석은 일정 기간 내의 주가를 산술 평균한 값으로 나누어 평균 주가로 표현하는 이동평균을 이용한 분석을 말한다(안성원, 조성배, 2010). 이운선(2003)은 단기 주가 이동평균선을 이용하여 매매하는 교차매매전략 뿐만 아니라 단일의 주가이동평균의 등락을 이용한 추세매매전략 주가와 이동평균선의 교차매매전략에 대한 검증은 시도하였으며 주가이동평균 전략이 유용함을 확인할 수 있었다. 남주하와 윤기향(2001)은 한국과 같은 신흥주식시장이 미국 주식시장으로부터 변동성 이전 효과를 갖고 있는지를 분석하였으며 모형이 유발하는 시계열 상관관계를 제거하기 위해 이동평균을 이용하였다. 그 결과 1999년부터 2000년의 기간은 변동성 이전이 보이지 않았지만 1999년부터 2000년의 기간에서는 변동성 이전이 발생하였음을 확인하였다. 김현지와 장우진(2010)은 5일, 20일, 60일, 120일 이동평균의 효과에 대해 한국, 미국, 일본의 주식시장을 대상으로 실증적인 통계 검증을 수행하였다. 그 결과 5일 이동평균선과 20일 이동평균선으로 조합된 투자 전략이 모든 국가별 테스트에서 가장 높은 초과 수익률을 얻는 것을 확인하였다.

두 번째 예측방법인 몬테카를로 시뮬레이션은 문제가 주어졌을 때 난수(random number)를 대량으로 발생시켜 충분히 많은 수의 조각

또는 무작위 실험의 결과를 종합해서 구하고자 하는 법칙을 근사적으로 얻고자 하는 방법을 말한다(신동백, 2007). 이재원(2000)은 주가 추세의 현재 상태를 평가하기 위해 강화 학습 모형을 도입하였고, 가치도 함수의 계산을 위해 강화 학습 모형 알고리즘의 일종인 몬테카를로 알고리즘을 적용하여 그 효용성을 검증하였다. 유진(2002)은 주식시장에 가격제한이 존재할 때 실제 시장에서의 주가에 대한 VaR(Value at Risk)을 산출하는 방법론을 제안하였다. 또한 미래 주가의 확률 분포를 구하기 위해 몬테카를로 시뮬레이션 기법을 사용하였고 그 결과 수익률제한을 감안한 실제 시장에서의 VaR과 감안하지 않은 VaR의 편차는 매우 클 수 있다는 사실을 제안했다. 하지만 주가의 등락은 비선형적이면서도 유동적인 방식에 결정(McNelis, 2005)이 되므로 선형적 문제에 보다 적합한 통계적 기법은 한계가 있음을 알 수 있다.

2.2 머신러닝과 딥러닝 접근방법

통계적 기법으로 시작한 주가 예측은 인공지능을 이용한 방법으로 발전되었으며 주가예측 기술 개발에 큰 공헌을 하였다(이석준과 오경주, 2011). 인공지능 기법과 통계적 기법을 비교해보면, 높은 비선형성과 동적인 성질을 가진 문제에서 인공지능이 예측과 분류를 더 정확하게 수행한다는 장점이 있다. 또한 신경망은 변화하는 환경에 쉽게 적응이 가능(손동우, 이웅규, 2000).

이러한 인공지능은 분류와 회귀분석에 널리 사용되는 기계학습 방법인 SVM(support vector machine), 인공신경망, 유전자 알고리즘

(genetic algorithm) 등을 사용한다. 특히 신경망이나 유전자 알고리즘을 이용한 방법은 예측 모델의 최적 패턴이나 가중치 변수를 찾기 위해 널리 사용되었으며 주가예측을 위한 연구에도 적용되었다. 특히 기술적 분석을 중심으로 인공지능 기법이 이용되었는데, 이우식(2017)은 기술적 분석 지표와 딥러닝 모형을 결합하여 한국 KOSPI 200 지수를 예측하는 모형을 개발하고 그 결과 기술적 분석 지표에 딥러닝 알고리즘을 결합한 모형이 주가지수 방향성 예측 문제에 응용될 수 있음을 확인하였다. 신동하 등(2017)은 최적의 보조지표 조합을 연구하고 추출된 보조지표 조합을 RNN(recurrent neural network)과 LSTM(long short-term memory)의 입력 값으로 제시한 주가예측 딥러닝 모델을 제안하였다. 대부분의 보조지표는 예측율을 저하시켰으나 추가 환경 예측요소를 통해 예측율이 향상되는 결과를 얻었으며 DNN(deep neural network)에 비해 RNN과 LSTM이 대략 15% 정도의 예측율 향상을 보였다. 김유일 등(2004)은 한국과 미국의 주식시장을 대표하는 KOSPI 200 지수와 S&P 500 지수를 대상으로 SVM과 신경망을 이용하여 상승과 하락을 예측하였다. 그 결과 SVM이 신경망보다 높은 예측율을 보였으며 모형 구축에 있어 신경망에 비해 수정해야 하는 파라미터의 수가 적어 용이하다는 이점을 보였다. 송유정 등(2017)은 삼성전자와 삼성전자 우선주의 주식 가격 데이터에 이동평균 데이터 등의 10가지 특성(feature)을 추가하여 딥러닝으로 주가예측을 진행하였다. 그 결과 특성을 추가한 모델이 아닌 모델보다 높은 성능을 보였으며 입력특성에 따라 예측성능이 달라짐을 주장하였다. 이모

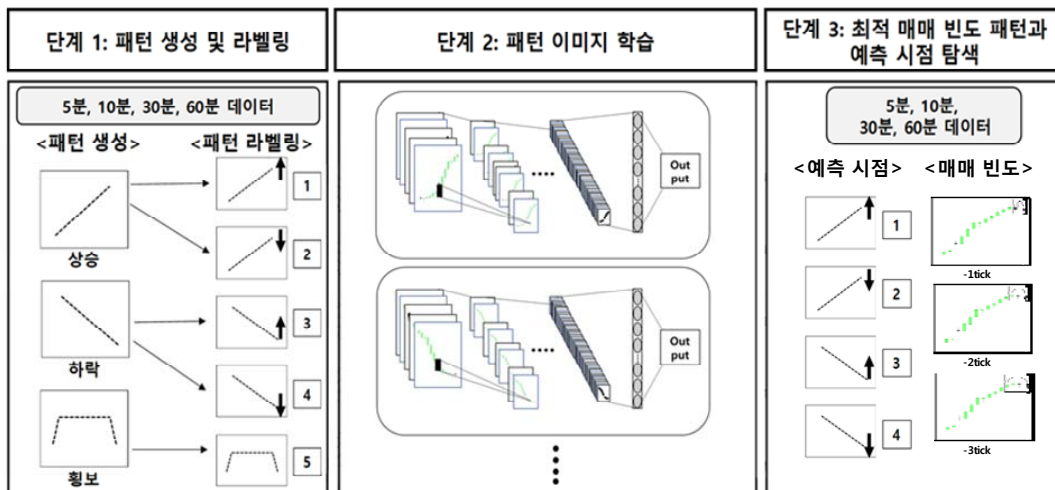
세와 안현철(2018)은 기술적 지표의 숫자 데이터를 이용한 기존 딥러닝 연구와는 다르게 기술적 지표를 꺾은선 그래프로 작성하여 이미지 형태로 변환한 후 딥러닝 방법 중 CNN에 적용하여 주가를 예측하는 시도를 하였다. 그 결과 로지스틱 회귀분석, 인공신경망, SVM의 다른 여러 방법보다 우수한 예측 정확도를 보여주었고 평균 예측 정확도는 57.74%로 나타났다.

기존 선행연구들은 기술적 분석 방법 중 기술적 지표를 이용하여 인공지능 기법을 적용하는 것에 집중하였으며 차트 이미지를 활용하여 딥러닝을 적용한 연구는 거의 없다. 이재원(2013)은 12개의 기술적 지표를 분석할 때 생기는 추세 후행성 문제를 보완하기 위해 차트 데이터인 일본식 봉(Japanese candlestick)지표를 패턴의 정의와 신경망 학습에 활용하였고 이를 통해 가장 최근의 주가 변동을 더욱 세밀히 요약할 수 있음을 주장하였으며 이강희 등(1997), 이운선(2006), 황희수(2008) 등의 선행연구들에 따르면 차트 분석을 통해 주가를 예측하는

것이 유의하다는 것을 보여준다. 최근 박상민 등(2018)은 차트 이미지를 이용하여 주가 예측을 진행하였다. 차트 이미지 중 상승 추세의 패턴 중 하나인 “Ascending Triangle”을 보이는 이미지 100개를 활용하여 Faster R-Cnn모델에 학습을 진행하였다. 그 결과 60.71%의 정확도를 보였다. 하지만 여러 상승 패턴 중 하나를 이용하여 학습을 진행하였고 작은 양의 데이터로 학습을 진행하여 정확도가 낮다는 한계를 가진다. 따라서 본 연구는 상승, 하락 패턴을 보이는 다량의 데이터를 이용하여 다양한 매매 빈도 패턴과 예측 시점 탐색 및 주가의 등락을 예측하고자 한다.

Ⅲ. 연구방법

본 장에서는 실시간 주식거래를 위한의 매매 빈도 패턴 및 주가 예측 타이밍을 탐색하기 위해 <그림 1>과 같은 제안 연구 절차에 대해 설



<그림 1> 제안 연구 절차

명한다. 첫 번째 단계에서는 학습을 수행하기 위한 패턴 이미지를 생성하며, 두 번째 단계에서는 생성한 패턴 이미지를 CNN의 학습데이터로 적용하여 학습을 진행한다. 마지막 단계에서는 학습 결과를 바탕으로 매매 패턴과 예측 시점을 예측하는 연구 절차를 수행하며, 각 절차에 대한 자세한 내용은 다음과 같다.

3.1 패턴 생성 및 라벨링

패턴 이미지 학습을 위해 패턴 생성과 라벨링은 다음과 같은 순서로 진행된다.

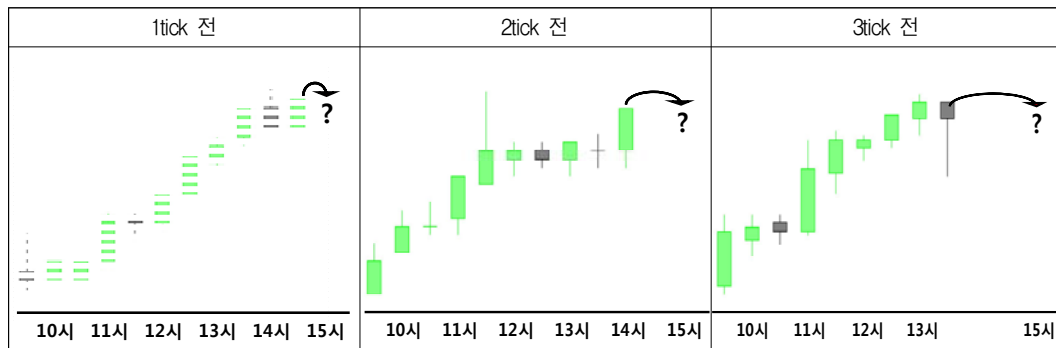
첫째, 수집한 주가 데이터를 약 70:30의 비율로 학습용 데이터 셋(training data set)과 테스트용 데이터 셋(test data set)으로 구분한다. 둘째, 패턴 생성을 위해 5분, 10분, 30분, 60분 빈도 데이터를 가지고 -1 tick, -2 tick, -3 tick과 같은 예측 시점별 캔들차트 이미지를 작성한다(<표 1 참고>). 캔들차트는 Nison(1991)에 의해 널리 알려진 방법으로 가격의 움직임을 선과 막대로 표기한 초 모양의 차트이며, 주식의 종가로 표현된 선 모양의 라인차트와 주식의 고가와 저가, 그리고 종가로 표현된 바차트가 결합된 형태로 볼 수 있다. 하나의 캔들은 주어진

시간 간격 동안 거래가 체결되는 가격의 시가, 고가, 저가, 종가의 정보를 담는다. 또한 시가보다 종가의 가격이 올라가면 초록색의 캔들, 시가보다 종가의 가격이 내려가면 회색의 캔들로 표현하여 가격 움직임의 방향까지 나타낸다. 캔들차트는 가격의 움직임을 시각적으로 표현하여 직관적이므로 다양한 분야에서 사용되고 있으며, 기술적 분석을 위한 전제 조건이다(김현호와 오경주, 2012).

셋째, 패턴 이미지 학습을 진행하기 위해 작성된 차트 이미지 중 학습용 데이터 셋으로 패턴을 생성한다. Lee et al.(2010)은 단기상승, 단기하락, 대세상승, 대세하락, 천정횡보, 바닥횡보와 같이 대표적인 주가패턴 6개를 제시하였으며 각각의 패턴 기간을 제안하였다. 본 연구에서는 패턴의 기간을 고려하지 않으므로 대표 패턴 모양인 상승패턴, 하락패턴, 횡보패턴만을 활용한다.

마지막으로, 생성된 패턴에 라벨링을 진행한다. 패턴 분석으로 주가의 등락을 예측하기 위해 생성한 패턴에 장 종료 시 상승 마감하는지 하락 마감하는지를 반영하여 총 5개로 라벨링을 진행한다. 횡보 패턴은 주가 등락 예측 대상이 아니므로 상승과 하락마감을 구분하지 않았

<표 1> 30분 단위 데이터 빈도의 틱(tick) 별 예시



—딥러닝을 활용한 실시간 주식거래에서의 매매 빈도 패턴과 예측 시점에 관한 연구: KOSDAQ 시장을 중심으로

으며 본 연구에서 사용하는 대표 패턴과 라벨링의 세부 정보는 <표 2>와 같다.

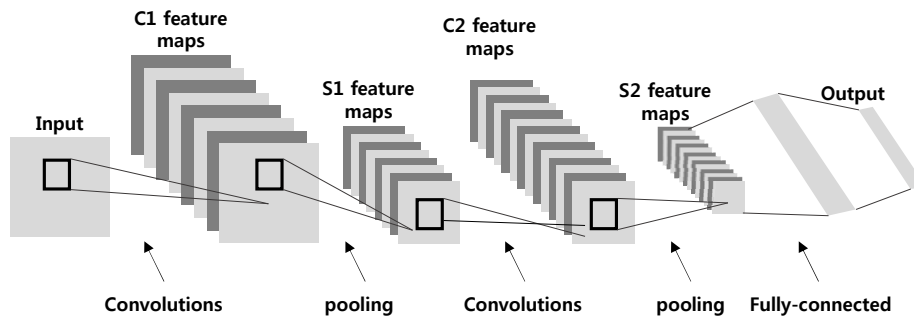
3.2 패턴 이미지 학습

본 단계에서는 학습용 데이터 셋을 대표적인 딥러닝 알고리즘인 CNN에 학습시킨다. CNN은 1990년대 후반 필기 인식(Lecun et al., 1998)에 활용되며 알려진 모델이며, 생물의 시신경이 동작하는 원리에서 영감을 얻어 고안(김지원 등, 2015)되었다. 이미지를 추상화하여 표현할 수 있는 특성(feature) 추출 기능으로 패턴 인식과 이미지 처리에 좋은 성능을 보이고

있어 이미지의 분류, 인식 분야에 주로 사용되고 있다(Lecun et al., 1998; 유제훈 등, 2016; 이우영 등, 2017). <그림 2>는 CNN 구조를 보여주고 있으며 합성곱 계층(convolutional layer), 다운 샘플링 계층(pooling layer), 그리고 완전 연결 계층(fully-connected layer)으로 구성되어 있다. 합성곱 계층은 필터를 이용하여 이미지로부터 특성들을 추출하는 역할을 하며, 다운 샘플링 계층은 이미지의 복잡도를 줄이기 위해 이미지의 크기를 점차적으로 줄이는 역할을 한다. 이러한 두 계층의 반복을 통해 이미지 데이터에 내재된 고차원의 정보가 효과적으로 추출되어 완전연결 계층의 입력 자료로 활용되

<표 2> 본 연구에서 사용된 대표 패턴과 라벨링 예시

대표패턴	대표패턴 예시	라벨링	설명
단기상승		1	상승패턴이며 상승 마감
		2	상승패턴이며 하락 마감
단기하락		3	하락패턴이며 상승 마감
		4	하락패턴이며 하락 마감
횡보		5	횡보패턴



<그림 2> CNN 구조(Lecun et al., 1988)

며, 완전 연결 계층에서는 두 계층의 반복을 통해 추출된 정보를 활용하여 이미지를 최종적으로 클래스 별로 분류하는 작업을 수행한다. CNN의 대표적인 모델로는 AlexNet (Krizhevsky et al., 2012), GoogLeNet(Szegedy et al., 2014), VGG-Net(Simonyan and Zisserman, 2014) 등이 있다. 본 연구에서는 ImageNet(Deng, 2009)이 개최하는 ILSVRC (ImageNet Large Scale Visual Recognition Challenge)에서 높은 성능을 보였으며 간단한 구조를 가져 여러 응용 분야에서 많이 사용되고 있는(고광은, 심귀보, 2017) VGG-Net으로 분석을 수행하고자 한다.

VGG-Net은 기본적인 CNN 구조와 마찬가지로 합성곱 계층과 다운 샘플링 계층에 완전 연결 계층이 연결된 구조이다(Simonyan & Zisserman, 2014). 합성곱 계층의 수에 따라 VGG-Net16과 VGG-Net19로 구분되며 <그림 3>은 합성곱 계층의 수가 16개인 VGG-Net16의 구조이다.

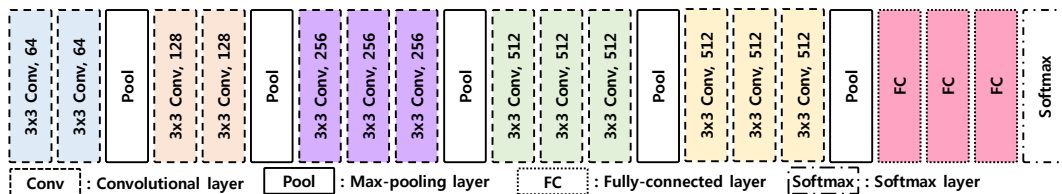
VGG-Net이 발표되기 전까지는 가장 깊은 계층이 8개였으나 VGG-Net에서 처음으로 16개, 19개의 깊은 계층이 발표되었으며 Katoaka et al.(2015)은 과적합 문제가 발생하지 않는다면 계층의 깊이가 깊을수록 이미지 인식 정확도가 높아진다는 것을 밝혀냈다. 본 연구에서는 ImageNet 분류 작업을 위해 훈련되었던 VGG-

Net의 선행 학습 모형(pre-trained model)의 파라미터 값을 초기값으로 이용하며 VGG-Net16과 VGG-Net19 중 선행 학습 모형의 정확도가 71.5%로 VGG-Net19보다 높은 VGG-Net16을 활용하여 연구를 수행한다.

3.3 매매 빈도 패턴과 예측 시점 탐색

본 단계에서는 주가 데이터 빈도와 예측 시점 별로 결과를 확인하여 정확성을 파악하고 이를 토대로 본 연구의 목적인 주가 등락 예측을 위해 매매 빈도 패턴과 예측 시점을 탐색한다. 이를 위해 학습된 CNN 모델에 테스트용 데이터 셋을 대상으로 다음과 같은 과정을 수행하여 검증이 이루어진다.

첫째, 테스트용 데이터를 학습이 완료된 CNN 모델에 입력한 후 CNN 모델의 Softmax 계층으로부터 테스트용 데이터에 대한 패턴(\hat{Y})과 해당 패턴에 대한 확률($P(\hat{Y})$)을 산출한다. 둘째, Softmax 계층으로부터 산출된 패턴에 대한 확률 값 중 가장 높은 수치를 나타낸 확률 값($P(\hat{Y}_1)$)과 본 연구에서 설정한 임계치(θ)를 확인한다. 셋째, 확률 값이 임계치보다 높다면 CNN 모델에 입력된 테스트용 데이터가 해당 패턴에 해당된다고 판단하며, 임계치가 확률 값보다 낮다면 해당 데이터는 어느 패턴에



<그림 3> VGG-Net16 계층 구조

—딥러닝을 활용한 실시간 주식거래에서의 매매 빈도 패턴과 예측 시점에 관한 연구: KOSDAQ 시장을 중심으로

도 속하지 않는다고 판단한다. 또한 홍보 패턴은 모델 학습에 혼란을 줄이기 위한 목적으로 이용된 것이므로 매매 빈도 패턴의 정확도 측정에서 제외된다. 마지막으로, CNN 모델에 의해 분류된 패턴별 테스트용 데이터 중 올바르게 분류된 데이터의 비율을 임계치를 설정했을 때와 설정하지 않았을 때로 나누어 확인하며, 가장 비율이 높은 패턴을 매매 패턴으로 정의한 후 데이터 빈도 별, 예측 시점별로 결과를 비교한다.

빈도 데이터에서 생성되는 캔들의 개수가 변경되었다. 따라서 그 이전의 데이터인 2013년 4월 1일부터 2016년 7월 31일까지 코스닥 150 종목의 시가, 고가, 저가, 종가, 거래량 등의 기본 데이터가 수집되었다. <그림 4>는 수집된 주가 데이터의 예시를 보여주고 있다.

수집된 코스닥 150 종목의 주가 데이터 중 2013년 4월 1일 이후에 등록되어 데이터의 시작일이 맞지 않은 종목을 제외하고 총 104개의 종목에 대해 분석을 진행하였으며 데이터 셋에 대한 세부 정보는 아래 <표 3>과 같다.

IV. 실증분석 및 결과

4.1 데이터 수집 및 기간설정

본 연구에서는 CNN 학습에 사용될 차트 패턴 이미지를 생성하기 위해 코스닥 상위 150종목을 선정하였으며, 대신증권에서 제공하는 OpenAPI를 이용하여 주가 데이터를 수집하였다. 2016년 8월 1일부터 주식시장 종료시간이 오후 3시에서 3시 30분으로 변경되어 동일한

<표 3> 데이터 셋 현황

	5분	10분	30분	60분
실험 기간	2013.04.01.~ 2016.07.31			
속성 종류	Open, High, Low, Close, Volume			
데이터 수	5,760,139	2,996,117	1,016,230	509,588

4.2 패턴 생성 및 라벨링 결과

수집된 데이터는 학습용 데이터 셋과 테스트용 데이터 셋으로 구분하는 작업을 수행하며, 2013년

일자	시간	시가	고가	저가	종가	거래량
20160412	1000	9040	9050	9010	9040	8957
20160412	955	9060	9060	9010	9040	10574
20160412	950	9030	9080	9020	9060	26357
20160412	945	9020	9030	8990	9030	17691
20160412	940	9010	9060	9000	9030	9879
20160412	935	8950	9110	8940	9020	65994
20160412	930	8910	8980	8910	8940	14985
20160412	925	8890	8910	8880	8910	8983
20160412	920	8870	8910	8870	8890	9745
20160412	915	8840	8870	8840	8870	2582
20160412	910	8880	8890	8830	8840	9998
20160412	905	8840	8880	8820	8880	35936

<그림 4> 수집된 주가 데이터 예시

4월 1일부터 2015년 7월 31일까지의 데이터는 학습용 데이터 셋, 2015년 7월 31일부터 2016년 7월 31일까지는 테스트용 데이터 셋으로 활용된다.

분류된 데이터 셋은 데이터 빈도와 예측 시점별로 R programming의 'plotly' 패키지를 이용하여 캔들차트 이미지를 생성하며 주식시장 운영 중 특수한 상황으로 인한 거래정지, 단일

가 거래 등이 발생한 날을 제외하기 위해 오전 9시부터 오후 3시까지 하루 동안 수집된 시가, 고가, 저가, 종가의 데이터 수가 5분 단위 빈도 데이터는 71개, 10분은 36개, 30분은 12개, 60분은 6개인 경우에만 차트 이미지를 작성하는 과정을 진행한다. 작성된 차트 이미지의 세부정보는 <표 4>와 같다.

<표 4> 작성된 차트 이미지 데이터 수

5분		10분		30분		60분	
학습용 데이터	테스트용 데이터	학습용 데이터	테스트용 데이터	학습용 데이터	테스트용 데이터	학습용 데이터	테스트용 데이터
38,406	21,244	49,110	23,913	57,849	24,999	59,169	25,181




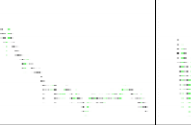
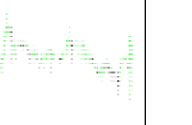

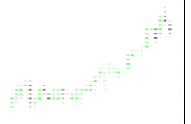



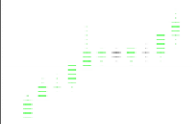



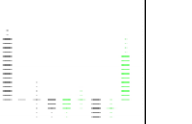





CNN 학습에 적용되기 위한 학습용 패턴 이미지 데이터의 수는 <표 5>와 같다.

<표 5> 생성된 패턴 이미지 데이터의 수

생성된 학습용 패턴			
5분	10분	30분	60분
6,632	8,246	10,910	15,802

생성한 패턴은 종가에 따라 상승마감과 하락마감으로 구분되었으며, 아래 <표 6>은 데이터 빈도 별로 생성된 패턴에 라벨링한 결과이며 CNN 학습의 입력 데이터(input data)로 이용한다.

<표 6> 생성된 패턴의 분봉 별 예시

빈도	패턴				
	1(상승패턴이며 상승마감)	2(상승패턴이며 하락마감)	3(하락패턴이며 상승마감)	4(하락패턴이며 하락마감)	5(횡보패턴)
5분					
10분					
30분					
60분					

4.3 패턴 이미지 학습 및 비교 결과

주가 등락 예측을 위한 매매 빈도 패턴과 예측 시점을 탐색하기 위해 VGG-Net16의 선형 학습 모형을 초기값으로 이용하였으며, 본 연구에 맞게 배치크기(batch size), 학습율(learning rate), 모멘텀(momentum), 가중치 감소율(weight decay), 반복 학습 횟수(learning step)를 미세 조정하는 방법(fine-tuning)을 적용하였다. 모델별 형평성을 위해 배치크기는 32, 학습율은 0.01, 모멘텀은 0.9, 가중치 감소율은 0.00004, 마지막으로 반복 학습 횟수는 20,000으로 동일하게 설정하였다. 또한 본 연구에서는 모델의 임계치를 0.5로 설정하였으며 적어도 절반 이상의 확률을 가져야 해당 패턴과 유사하다고 말할 수 있다고 판단하였기 때문이다. 모형 학습이 완료된 후 임계치가 0일 때와 0.5일 때로 나누어 테스트용 데이터 셋을 패턴 별로 분류하고 결과를 비교하였다.

매매 빈도 패턴과 예측 시점 탐색의 결과는 아래 <표 7> ~ <표 10>과 같다. <표 7>은 -1 tick의 빈도 별 실험 결과이며 모든 빈도 데이터에서 4번 패턴의 예측 정확도가 가장 높게 나타났다. 그 중에서도 임계치가 0.5인 10분 단위 빈도 패턴이 76%로 가장 높은 정확도를 보였으며 그 뒤로 임계치가 0일 때와 0.5일 때의 5분(74%) 단위 빈도 패턴 순으로 정확도가 측정되었다.

<표 8>은 -2 tick의 빈도 별 실험결과를 나타내며 이를 살펴보면 대부분의 빈도 데이터에서 4번 패턴의 예측 정확도가 가장 높은 것으로 확인되었지만 임계치가 0인 30분 단위 빈도 데이터에서는 2번 패턴의 예측 정확도가 가장 높게

확인되었다. 그 중에서도 임계치가 0.5인 10분과 60분 빈도 패턴이 63%로 가장 높게 나타났으며 그 뒤로 임계치가 0인 5분(61%) 단위 빈도 패턴 순으로 높은 정확도가 측정되었다.

<표 9>는 -3 tick의 빈도 별 실험결과를 보여주며 임계치가 0인 5분, 10분 빈도 데이터와 임계치가 0.5인 5분, 30분, 60분 빈도 데이터에서는 4번 패턴의 예측 정확도가 가장 높지만 임계치가 0인 30분, 60분 빈도 데이터와 임계치가 0.5인 10분 단위 빈도 데이터에서는 2번 패턴이 정확도가 가장 높은 매매 빈도 패턴으로 확인되었다. 또한 임계치가 0.5인 10분 단위 빈도 데이터에서 2번 패턴이 71%로 가장 높은 정확도를 보였으며 그 뒤로 임계치가 0.5인 5분 단위 빈도 패턴(64%), 임계치가 0인 5분 단위 빈도 패턴(60%) 순으로 정확도가 측정되었다.

마지막으로 <표 10>과 같이 데이터 빈도 및 예측 시점별로 실험결과를 비교해보면 -1 tick의 임계치가 0.5인 10분 단위 빈도 데이터에서 4번 패턴이 76%로 가장 높은 예측 확률을 보인 매매 빈도 패턴으로 확인되었다. -1 tick의 임계치가 0.5인 5분 단위 빈도 데이터의 4번 패턴(74%), -1 tick의 임계치가 0.5인 60분 단위 빈도 데이터의 4번 패턴(73%) 순으로 정확도가 측정되었다. 예측 결과 대부분 하락패턴에 해당하는 4번 패턴과 2번 패턴의 정확도가 높았다. 이는 전반적으로 증시가 하락패턴인 경우가 많기 때문에 학습용 데이터의 양이 많았기 때문이라고 사료된다. 따라서 향후 상승패턴의 데이터가 더 수집된다면 상승패턴의 정확도가 더 높아질 것으로 판단된다. 또한 전체적으로 높은 예측률을 보인 매매 빈도는 10분 단위였으며 예측 시점은 -1 tick인 것으로 확인되었다.

<표 7> -1 tick 매매 빈도 패턴 예측의 정확도

패턴	(a) 임계치 0				(b) 임계치 0.5			
	5분	10분	30분	60분	5분	10분	30분	60분
1	34%	36%	40%	28%	36%	38%	50%	30%
2	64%	68%	54%	51%	63%	63%	54%	57%
3	38%	27%	45%	46%	28%	27%	45%	50%
4	74%	72%	59%	70%	74%	76%	60%	73%

<표 8> -2 tick 매매 빈도 패턴 예측의 정확도

패턴	(a) 임계치 0				(b) 임계치 0.5			
	5분	10분	30분	60분	5분	10분	30분	60분
1	43%	45%	49%	47%	49%	41%	49%	48%
2	56%	52%	53%	51%	57%	55%	51%	46%
3	35%	40%	48%	47%	35%	36%	48%	46%
4	61%	60%	52%	54%	60%	63%	54%	63%

<표 9> -3 tick 매매 빈도 패턴 예측의 정확도

패턴	(a) 임계치 0				(b) 임계치 0.5			
	5분	10분	30분	60분	5분	10분	30분	60분
1	44%	47%	48%	45%	49%	35%	51%	48%
2	56%	54%	51%	54%	56%	71%	49%	53%
3	40%	42%	48%	0%	40%	40%	50%	0%
4	60%	59%	50%	53%	64%	53%	53%	55%

<표 10> 매매 빈도 패턴 예측 결과 정리

틱	5분			10분			30분			60분		
	임계치	패턴	정확도	임계치	패턴	정확도	임계치	패턴	정확도	임계치	패턴	정확도
-1	0.5	4	74%	0.5	4	76%	0.5	4	60%	0.5	4	73%
-2	0	4	61%	0.5	4	63%	0.5	4	54%	0.5	4	63%
-3	0.5	4	64%	0.5	2	71%	0.5	4	53%	0.5	4	55%

V. 결론

본 연구는 주가 등락을 예측하기 위한 매매 빈도 패턴과 예측 시점을 탐색하기 위해 차트 이미지를 대상으로 딥러닝을 적용하였으며 다

음과 같은 과정으로 연구가 수행되었다. 첫째, 대신증권에서 제공하는 OpenAPI를 이용하여 주가 데이터를 수집하였으며 수집한 주가 데이터를 이용하여 데이터 빈도와 예측 시점별로 캔들차트 이미지를 작성하였다. 둘째, 차트 이

미지로 패턴을 생성하였으며 생성한 패턴을 대상으로 딥러닝 알고리즘 중 CNN을 이용하여 학습을 진행하였다. 마지막으로 검증용 데이터로 매매 빈도 패턴과 예측 시점을 탐색했다.

전술된 절차를 통해 연구를 수행한 결과 임계치가 0.5인 10분 단위 빈도 데이터가 -1 tick에 하락패턴으로 판단되면 주가 하락으로 장이 종료된다고 예측하는 것이 76%로 가장 정확도가 높은 매매 빈도 패턴으로 확인되었다. 또한 매매 빈도는 10분 단위 데이터, 예측 시점은 -1 tick에 예측하는 것이 높은 정확도를 보이는 것을 확인하였다.

본 연구의 의의로는 첫째, 기존 연구와는 다르게 차트 이미지를 대상으로 CNN을 활용하였으며 데이터 빈도와 예측 시점별로 다양하게 차트 이미지 분석을 수행하는 방법을 제안했다는 점에서 의의를 갖는다. 둘째, 본 논문에서 제안한 방법은 딥러닝을 이용한 다른 주가 등락 예측 연구들보다 높은 정확도를 가지며 매매 전략에 관한 연구에 기여할 것으로 사료된다. 셋째, 본 논문에서 제안하는 방법은 주식시장 참여자들이 주가 등락 예측 시 객관적인 의사 결정에 기여할 것으로 기대된다.

본 연구의 한계점 및 향후 연구방향으로는 첫째, 2017년 8월 1일부터 주식시장의 종료시간이 변경되어 하루 동안 생성되는 빈도 데이터의 수가 달라지므로 본 연구는 2013년 4월 1일부터 2016년 7월 31일까지의 데이터를 이용하여 학습을 진행하였다. 따라서 향후 연구에서는 2016년 8월부터 새롭게 수집되는 데이터의 양이 충분해진다면 최신 데이터에 적용해 볼 수 있을 것으로 사료된다. 둘째, 차트 이미지 데이터에 금융시장 상황에 영향을 주는 변수나

최근 주가 패턴 연구에서 선행되고 있는 텍스트 마이닝을 이용한 뉴스 데이터 분석 등을 추가한다면 향후 연구과제가 될 수 있을 것이다. 셋째, 본 연구는 딥러닝을 활용하여 차트 이미지를 분석하는 것에 대한 가능성을 보기 위한 연구로 상승, 하락을 예측하는 단순한 모델이다. 따라서 본 연구에서 제외한 횡보 패턴뿐 아니라 다양한 주가 예측 패턴들을 추가하고 더 다양한 매매 빈도와 타이밍을 추가한다면 실제적 예측 모형 구축을 위한 향후 연구과제가 될 수 있을 것으로 사료된다.

참고문헌

- 고광은, 심귀보, “딥러닝을 이용한 객체 인식 및 검출 기술 동향,” 제어로봇시스템학회지, 제23권, 제3호, 2017, pp. 17-24.
- 김범준, 신민규, 박종훈, 이용준, “기술지표를 이용한 주가예측,” 경영경제 데이터 분석 사례집, 제8권, 2005, pp. 199-215.
- 김상환, 조태근, “기술적 거래전략의 예측력 검증,” 재무연구, 제16권, 제2호, 2003, pp. 67-93.
- 김유일, 신은경, 홍태호, “신경망과 SVM 을 이용한 주가지수예측의 비교,” 인터넷전자상거래연구, 제4권, 제3호, 2004, pp. 221-243.
- 김종선, 윤세현, 김철중, “고성장 코스닥시장 벤처기업의 개별특성이 IR공시 효과에 미치는 영향,” 벤처창업연구, 제9권, 제4호, 2014, pp. 97-109.
- 김지원, 표현아, 하정우, 이찬규, 김정희, “다양

- 한 딥러닝 알고리즘과 활용,” 정보과학회지, 제33권, 제8호, 2015, pp. 25-31.
- 김진백, 김유일, “인공 신경망의 학습에 있어 가중치 변화방법과 은닉층의 노드수가 예측정확성에 미치는 영향,” 정보시스템연구, 제9권, 제1호, 2000, pp. 27-44.
- 김현지, 장우진, “지수 이동 평균선을 이용한 거래 전략,” 대한산업공학회 춘계공동학술대회 논문집, 2010, pp. 1124-1130.
- 김현호, 오경주, “러프집합을 활용한 캔들스틱 트레이딩 최적화 전략,” 한국데이터정보과학회지, 제23호, 제5호, 2012, pp. 881-893.
- 남주하, 윤기향, “미국 주식시장에서 한국 주식 시장으로의 변동성 이전효과 분석,” 공동학술대회 발표논문, 2001, 한국금융학회.
- 박강희, 신현정, “시계열 네트워크에 기반한 주가예측,” 경영과학, 제28권, 제1호, 2011, pp. 53-60.
- 박상민, 김바다, 원태연, 허준영, “딥러닝을 이용한 이미지 기반 주가예측,” 한국정보과학회 학술발표논문집, 2018, pp. 1212-1213.
- 성노윤, 남기환, “온라인 뉴스 및 거시경제 변수를 활용한 주가예측,” Entrue Journal of Information Technology, 제16권 제2호, 2017, pp. 41-54.
- 손동우, 이용규, “약체연결뉴런 제거법에 의한 부도예측용 인공신경망 모형에 관한 연구,” 한국정보시스템학회 학술대회 논문, 2000, pp. 115-121.
- 송유정, 이재원, 이종우, “텐서플로우를 이용한 주가 예측에서 가격-기반 입력 피쳐의 예측 성능 평가,” 정보과학회 컴퓨팅의 실제 논문지, 제23권 제11호, pp. 625-631.
- 송유정, 이종우, “텐서플로우를 이용한 주가 변동 예측 딥러닝 모델 설계 및 개발,” 한국정보과학회 학술발표논문집, 2017, pp. 799-801.
- 신동백, “몬테카를로 시뮬레이션방법을 이용한 환율예측분석,” 산업경제연구, 제20권, 제5호, 2007, pp. 2075-2093.
- 신동하, 최광호, 김창복, “RNN 과 LSTM 을 이용한 주가 예측을 향상을 위한 딥러닝 모델,” 한국정보기술학회논문지, 제15권, 제10호, 2017, pp. 9-16.
- 안성만, “딥러닝의 모형과 응용사례,” 지능정보연구, 제22권, 제2호, 2016, pp. 127-142.
- 안성원, 조성배, “뉴스 텍스트 마이닝과 시계열 분석을 이용한 주가예측,” 한국정보과학회 학술발표논문집, 2010, pp. 364-369.
- 양유영, 박상성, 신영근, 장동식, “데이터 특성에 따른 예측모형 개발 동향에 관한 연구,” Proceedings of KIIT Summer Conference, 2009, pp. 591-597.
- 오성민, 김성집, “주식의 상한가시 잔량과 일일 거래량의 관계를 통한주가의 단기예측에 관하여-kospi 200 종목을 중심으로,” 한국경영과학회 학술대회논문집, 2000, 457-460.
- 오세창, “기술적 분석에 의한 주식투자에서 유전 알고리즘을 이용한 최적화 방법,” 한

- 국인터넷정보학회 학술발표대회 논문집, 2014, pp. 319-320.
- 유재필, 신현준, 김미희, 백종관, (2017). 기계학습과 데이터 시각화 기법을 이용한 주가 패턴 분석. 정보기술아키텍처연구, 14(2), 189-197.
- 유제훈, 고광은, 심귀보, “컨볼루션 신경망과 케스캐이드 안면 특징점 검출기를 이용한 얼굴의 특징점 분류,” 제어로봇시스템학회 논문지, 제22권, 제3호, 2016, pp. 241-246.
- 유진, “가격제한 존재 시 VaR 추정 방법론: 몬테카를로 시뮬레이션,” 선물연구, 제10권, 제1호, 2002, pp. 81-111.
- 이강희, 양인실, 조근식, “캔들스틱 분석을 이용한 주식 매매 타이밍 예측을 위한 전문가 시스템,” 한국전문가시스템학회지, 제3권, 제2호, 1997, pp. 57-70.
- 이모세, 안현철, “효과적인 입력변수 패턴 학습을 위한 시계열 그래프 기반 합성곱 신경망 모형,” 지능정보연구, 제24권, 제1호, 2018, pp. 167-181.
- 이석준, 오경주, “선물시장의 시스템트레이딩에서 동적시간외평 알고리즘을 이용한 최적매매빈도의 탐색 및 거래전략의 개발,” 한국데이터정보과학회지, 제22권, 제2호, pp. 2011, 255-267.
- 이우식, “딥러닝분석과 기술적 분석 지표를 이용한 한국 코스피주가지수 방향성 예측,” 한국데이터정보과학회지, 제28권, 제2호, 2017, pp. 287-295.
- 이우영, 박승민, 장인훈, 김태형, 심귀보, “CNN 기반 신발갑피 패턴인식 및 접촉제도 포를 위한 경로점 생성,” 제어로봇시스템학회 논문지, 제23권, 제9호, 2017, pp. 725-731.
- 이운선, “시간흐름을 반영하는 캔들스틱과 거래량차트,” 금융공학연구, 제5권, 제1호, 2006, pp. 113-127.
- 이운선, “주가이동평균선을 이용한 기술적 분석의 효과,” 한국금융공학회 학술발표논문집, 제2002권, 제2호, 2002, pp. 1-13.
- 이운선, “한국증권시장에서 기술적 분석의 경제적 효과-주가이동평균선을 중심으로,” 금융공학연구, 제2권, 제2호, 2003, pp. 1-21.
- 이재원, “Monte Carlo 알고리즘에 의한 주가 추세의 평가,” 한국정보과학회 학술발표논문집, 제27권, 제1호, 2000, pp. 235-237.
- 이재원, “고변동 주가 패턴의 감독 학습에 기반한 주식 거래 시스템,” 정보과학회논문지, 2013.
- 이종우, 김유섭, 김성동, 이재원, 채진석, “인공지능: 패턴 매칭과 자동 규칙 생성에 기반한 2 단계 주식 트레이딩 시스템,” 정보처리학회논문지 B, 제10권, 제3호, 2003, pp. 257-264.
- 조희연, 김영민, “유전자 알고리즘을 이용한 주식투자 수익률 향상에 관한 연구,” 정보시스템연구, 제12권, 제2호, 2003, pp. 1-20.
- 황희수, “퍼지 모델을 이용한 일별 주가 예측,” 정보처리학회논문지 B, 제15-B권, 제6호, 2008, pp. 603-608.

- Chen, M., and Lin, H., "Order Imbalance, Individual Stock Returns and Volatility: Evidence from China," *JCIT: Journal of Convergence Information Technology*, Vol. 7, No. 2, 2012, pp. 101-107.
- Chiu, D. Y., and Chian, S. Y., "Exploring stock market dynamism in multi-nations with genetic algorithm, support vector regression, and optimal technical analysis," In *Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference on* (pp. 694-699). IEEE.
- Deboeck, G. J., "Trading on the Edge," Wiley, Canada, 1994.
- Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L., "Imagenet: A large-scale hierarchical image database," *In Computer Vision and Pattern Recognition*, 2009, pp. 248-255.
- Kataoka, H., Iwata, K., and Satoh, Y., "Feature Evaluation of Deep Convolutional Neural Networks for Object Recognition and Detection," *arXiv preprint*, 2015, arXiv:1509.07627.
- Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Network," *in Proceeding of the Neural Information Processing Systems 25(NIPS)*, 2012, pp. 1097-1105.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, 1998, Vol. 86, pp. 2278-2324.
- Lee, S. J., Ahn, J. J., Oh, K. J., and Kim, T. Y., "Using rough set to support investment strategies of real-time trading in futures market," *Applied Intelligence*, Vol. 3, 2010, pp. 364-377.
- Lee, S. J., and Jeong, S. J., "Trading Strategies based on Pattern Recognition in Stock Futures Market using Dynamic Time Warping Algorithm," *Journal of Convergence Information Technology*, Vol. 7, No. 10, 2012, pp. 185-196.
- McNelis, P. D., "Neural networks in finance: gaining predictive edge in the market," Academic Press, 2005.
- Nelson, D. M., Pereira, A. C., and de Oliveira, R. A., "Stock market's price movement prediction with LSTM neural networks," *In Neural Networks (IJCNN), International Joint Conference*, 2017, pp. 1419-1426.
- Nison S., "Japanese Candlestick charting techniques: A contemporary guide to the ancient investment technique of the far east Institute of Finance," New York, 1991.
- Shih, M. L., Hsiao, S. H., and Chen, F. S., "The association of China stock index with Japan and US," *Journal of Convergence Information Technology*, Vol. 3, No. 2,

—딥러닝을 활용한 실시간 주식거래에서의 매매 빈도 패턴과 예측 시점에 관한 연구: KOSDAQ 시장을 중심으로

2008, pp. 13-22.

Simonyan, K., and Zisserman, A., “Very deep convolutional networks for large-scale image recognition,” arXiv preprint, 2014, arXiv:1409.1556.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., “Going deeper with convolutions,” in Proceeding of the Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9.

송 현 정 (Song, Hyun-Jung)



현재 광운대학교 경영학부 석사과정 중이다. 주요 관심분야는 머신러닝, 딥러닝, 네트워크 분석, 데이터 마이닝 등이다.

이 석 준 (Lee, Suk-Jun)



연세대학교 박사학위를 취득하였다. 현재 광운대학교 경영학부 교수로 재직하고 있으며, 주요 관심분야는 주가예측, 인공지능, 빅데이터 등이다.

<Abstract>

A Study on the Optimal Trading Frequency Pattern and Forecasting Timing in Real Time Stock Trading Using Deep Learning: Focused on KOSDAQ

Song, Hyun-Jung · Lee, Suk-Jun

Purpose

The purpose of this study is to explore the optimal trading frequency which is useful for stock price prediction by using deep learning for charting image data. We also want to identify the appropriate time for accurate forecasting of stock price when performing pattern analysis.

Design/methodology/approach

In order to find the optimal trading frequency patterns and forecast timings, this study is performed as follows. First, stock price data is collected using OpenAPI provided by Daishin Securities, and candle chart images are created by data frequency and forecasting time. Second, the patterns are generated by the charting images and the learning is performed using the CNN. Finally, we find the optimal trading frequency patterns and forecasting timings.

Findings

According to the experiment results, this study confirmed that when the 10 minute frequency data is judged to be a decline pattern at previous 1 tick, the accuracy of predicting the market frequency pattern at which the market decreasing is 76%, which is determined by the optimal frequency pattern. In addition, we confirmed that forecasting of the sales frequency pattern at previous 1 tick shows higher accuracy than previous 2 tick and 3 tick.

Keyword: Stock price forecasting, Deep learning, Convolutional Neural Network, Optimal trading frequency

* 이 논문은 2018년 7월 3일 접수, 2018년 8월 21일 1차 심사, 2018년 9월 26일 게재 확정되었습니다.