

온라인 뉴스 및 거시경제 변수를 활용한 주가예측 Combining Macro-economical Effects with Sentiment Analysis for Stock Index Prediction

성노윤 · 남기환
Seong, Nohyoon · Nam, Kihwan

Abstract

시장의 상태는 새로운 정보를 받으면서 변한다. 이는 투자자들의 결정에 영향을 주어 주식 시장의 변동을 만들어 내며, 주식 시장의 동향을 예측하는 데에 중요한 자료로써 사용된다. 외부의 정보는 크게 두 가지로 생각할 수 있는데, 뉴스의 정보와 거시 경제적인 지표이다. 하지만, 기존의 연구들은 뉴스만을 가지고 혹은 거시 경제 지표만을 가지고 예측을 하였으며, 뉴스의 정보와 거시 경제 지표를 함께 결합하여 외부의 영향을 모두 고려한 연구는 없었다. 본 논문에서는 주식 시장에 영향을 미치는 외부의 정보를 모두 고려하여 서로 다른 특성을 가지는 정보들을 딥러닝 알고리즘을 통해 통합하여 예측하는 방법을 제시한다. 뉴스 정보는 뉴욕타임스 2년 치 데이터를 Word2Vec 모델을 사용하여 정보 추출을 하였다. 거시 경제 지표는 다우 존스 지수에 영향을 줄 수 있는 지표들을 경제학적 이론에 기반을 두어 선정하였다. 이는 금값과 환율 시장이다. 뉴스에서 추출한 정보와 거시 경제적인 영향을 함께 통합하는 알고리즘은 시계열 데이터와 문자열 데이터를 다루기에 적절한 LSTM 네트워크를 응용하여 사용한다. 예측 기간을 바꾸며 실험을 진행한 결과, 단순 정보만을 가지고 예측한 결과나 거시 경제만을 가지고 예측한 결과보다 본 논문에서 제시한 두 가지를 통합하여 예측하는 것이 월등한 결과를 내었다. 또한, 기존의 연구에서 사용한 방법론과 제시한 딥러닝 알고리즘을 비교한 결과 본 논문에서 제시한 알고리즘이 기존 연구에서 사용한 방법론보다 월등한 결과를 나타냈다.

주제어: 주가예측, 딥러닝, 감성 분석, 경제 지표, Word2Vec

Market conditions change with new information. This affects investors' decisions and creates stock market volatility. This information is used as an important resource for forecasting stock market trends. At this time, there are two kinds of external information: news information and macroeconomic indicators. However, if we look at existing studies, they predicted only with news or macroeconomic indicators. There was no study that considered all the external influences by combining news information and macroeconomic indicators together. Therefore, in this paper, we propose a method of integrating and forecasting information with different characteristics by considering the external information affecting the stock market through the deep learning algorithm. News information was extracted from the New York Times's 2-year data using the Word2vec model. Macroeconomic indicators are based on economic theories that can influence the Dow Jones Index. This is the gold price and exchange rate market. Algorithms that integrate information extracted from news and macroeconomic effects together are applied to LSTM networks suitable for handling time series data and string data. As a result of experimenting with changing the forecasting period, it is better to predict the results with only simple information and the macroeconomic results than with the results from this study. In addition, the algorithm proposed in this paper is superior to the methodology used in previous research.

Keywords : Stock price forecasting, Deep learning, Emotional analysis, Economic index, Word2Vec

성노윤: KAIST, 경영대학원 경영공학부, 석사과정(제1저자)
남기환: 한양대학교, 경영대학, 겸임교수(교신저자)

1. 서론

주가 예측은 주식 가격 변동에 영향을 미치는 많은 요인 간의 상호작용으로 인해 매우 어렵다고 알려진 분야임에도 불구하고 오랜 기간 꾸준히 시도됐다. 이처럼 주가 예측이 어려운 이유는 단순한 데이터의 패턴을 분석하는 것뿐 아니라 투자자들의 심리적인 영향에 의해 주가가 움직일 수 있는 것과 같이 다양한 면에서 받는 영향도 고려해야 하기 때문이다. 이론적으로, 주가는 미래의 성과에 대한 기댓값이기 때문에, 투자자들은 투자할 때 주식의 가치를 평가하고 투자하려고 노력한다. 가치를 평가하기 위해서 다양한 정보를 활용할 수 있으며, 이로 인해 투자자들의 투자 심리 또한 변하게 된다. 즉, 미래 성과에 대한 기댓값을 볼 때는 투자자들의 심리와 정보가 매우 높은 상관관계를 가지며 정보는 매우 중요한 역할을 한다 [19][56]. 즉, 정보는 미래 주가를 예측할 때에 필수적이다. 특히 온라인 시대에서 모바일 시대까지 도래하면서 수많은 정보를 손쉽게 얻을 수 있게 되면서 이러한 현상은 더욱 두드러지게 나타나게 되었다. 이와 같은 수많은 정보가 다양한 매체를 통해 투자자들에게 전달되며, 시장의 새로운 정보로 활용되면서 주식 시장에 영향을 미치게 된다 [7][33]. 즉, 정형화된 주식의 시계열 데이터뿐 아니라 각종 언론과 같은 정보를 분석하는 것은 주가를 예측하는 데에 중요하다.

빅데이터 시대에 주식 시장에 영향을 미치는 뉴스 정보의 양은 급증하였다. 하지만, 수많은 양의 뉴스에서 필요한 정보만을 선별적으로 찾기는 쉽지 않다. 따라서, 주가를 예측하는 데 필수적인 질적인 정보를 얻기 위해 많은 연구가 진행되어 왔으며, 이를 자동으로 행하는 방법을 찾기 위해 많은 노력이 있었다 [19][36]. 위 연구들에서는 자동으로 뉴스를 분석하여 주가를 예측하는 것이 효율적임을 제시한다. 하지만, 이 연구들은 대부분 Bag-of-Words 모델을 채택하여 연구를 진행하는데, Bag-of-Words 모델은 단어의 순서나 내재적인 의미를 파악하지 못한다는 단점이 있다. 따라서, 본 논문에서는 이를 해결하기 위해 고안된 Word2Vec [47] 모델을 사용하여, 자동으로 주가의 흐름을 예측할 수 있는 시스템을 고안하였다.

또한, 주가에 영향을 미칠 수 있는 것은 정보뿐만 아니라, 거시 경제적인 효과 또한 있다. 예를 들어, 금값과 주가에 관련된 연구는 과거부터 꾸준히 되어 있으며, 금값은 주가에 영향을 준다고 알려져 있다 [4][48]. 또한, 외환 시장의 변동은 국제적 경쟁력에 영향을 주어서 무역적 균형에 영향을 주며, 기업의 가치에 영향을 준다. 따라서, 외환시장은 주식 시장과 밀접

한 관계가 있다 [14]. 즉, 환율은 주식 시장에 영향을 미친다 [9][32][42][45][49]. 특히, Phylaktis (2005) [42]에서 일본 시장과 같은 안정성 있는 시장의 환율은 미국의 주가에 영향을 미치지 않지만 태평양해변의 국가의 환율은 미국의 주가에 영향을 준다고 하였다. 그중에서도, Andreou et al. (2013) [3]에서는 신흥 시장인 한국과의 환율을 사용하는 것이 적합하다고 한다. 따라서, 본 논문에서는 금값과 원-달러 환율을 거시 경제적인 지표로 고려하여 사용하였다.

기존 논문들에서는 금융 뉴스를 분석하여 주가의 오르고 내림을 예측하는 것이 대부분이다. 기술적 분석을 통합하여 고려하는 분석은 종종 시행되고 있으나 [11], 거시 경제적인 영향을 고려하는 연구는 매우 드물다. 금융 연구에서는 거시 경제적 효과를 고려하나, 정보의 영향을 함께 분석하지는 않는다. 하지만, 두 가지 모두 주가에 영향을 주며, 예측하는 데 도움을 준다. 따라서, 본 논문에서는 정보와 거시 경제적인 효과를 함께 고려하는 새로운 방법을 제시한다.

정보는 Word2Vec 모형으로 벡터 형태로 표현하며, 거시 경제적인 효과와 주가의 정보에서 추출한 특징 벡터는 딥러닝 알고리즘(순환 신경망)으로 통합하여 계산하였다. 순환 신경망은 금값과 주가지수, 환율 등의 시계열 데이터를 처리하는 데 효율적인 구조를 가지며, 특징 벡터를 포함할 수 있다. 순환신경망을 통해, 거시 경제 지표와 뉴스를 통합하여 주가를 예측하는 것이 거시 경제 지표만으로 분석하거나 뉴스만을 가지고 주가를 예측하는 것보다 더 나은 결과를 보여 주었다. 또한, 기존 연구들에서 제시된 방법론인 서포트 벡터 회귀(Support Vector Regression)와 인공 신경망과 비교를 한 결과 본 논문에서 제시한 딥러닝 알고리즘이 더 우월한 결과를 내었다.

본 논문이 기여한 부분은 크게 두 가지로 나누어진다. 첫 번째, 우리는 약형 효율적 시장 가설(Weak Efficient Market Hypothesis)에 기반을 두어 외부의 정보인 거시경제학적인 요소들과 뉴스의 문자열 데이터를 경제학 이론에 기반을 두어 성공적으로 통합하였다. 우리는 뉴스, 트위터 등의 텍스트 데이터 [6][19][36]만을 이용한 기존의 논문들과 시계열 데이터만을 이용한 기존의 논문들과는 다르게 [38][41][50] 두 가지를 모두 성공적으로 통합하였다. Deng et al. (2011) [11]에서는 기술적 분석과 뉴스 정보 분석을 같이하였다. 하지만, 이는 두 가지 점에서 한계를 가진다. 첫 번째, 약형 효율적 시장에서 기술적 분석은 주가를 예측하는 데 도움을 주지 않고, 외부로부터의 영향만을 반영해야 하는데, 내생적인 지표인 기술적 분석을 지표로 사용하면

서 이 결과가 특수한 시장 상태에서만 적용이 된다는 의문점을 가진다는 점에서 한계점을 갖는다. 따라서, 본 논문에서는 기술적 분석이 아닌 거시경제적 지표와 뉴스 정보를 함께 통합하여 예측하였다. 두 번째, 저자는 뉴스의 빈도와 감성 분석 결과만을 사용하여 분석을 시행하였는데, 이는 뉴스가 가지는 내재적인 의미를 파악하기 부족하다는 점에서 한계를 가진다. 따라서, 본 논문에서는 단순한 감성 분석이 아닌, 뉴스의 내재적 의미를 파악할 수 있는 방법을 사용하였다.

두 번째, 우리는 텍스트 벡터, 거시 경제 지표들의 특징 벡터를 통합하는 방법으로, 새로운 딥러닝 알고리즘을 제시하였다. Phua et al.(2003)[41]과 Hagenau et al.(2013)[19]의 논문에서 볼 수 있듯이, 우리는 인공 신경망과 서포트 벡터 회귀가 주가를 예측하는 좋은 방법이라는 것을 알 수 있다. 하지만, 입력값인 텍스트 벡터와 거시 경제 지표는 다른 특성이 있다. 따라서 우리는 서포트 벡터 회귀를 대체하는 다른 방법을 고안하였다. 또한, 인공신경망은 시계열 데이터를 잘 반영하지 못하였다. Deng et al. (2011)[11]에서는 이를 극복하기 위해, 다중 커널 분석(Multiple Kernel Learning)을 사용하였는데, 이는 서포트 벡터 회귀와 인공신경망과 마찬가지로 시간에 따른 흐름을 반영하지 못하였기 때문에, 본 논문에서는 이를 반영할 수 있는 딥러닝 알고리즘을 고안하였다.

본 논문의 뒷부분은 다음과 같이 구성된다. 2장에서는 주가에 대한 영향을 예측하는 기존의 관련 논문들에 대한 전반적인 흐름을 볼 것이며, 거시 경제적인 효과가 어떻게 영향을 주는지 자세하게 설명할 것이다. 3장에서는 데이터를 얻은 방법과 요약 정보를 설명할 것이며, 4장에서는 연구 모형에 관해 설명할 것이다. 5장에서는 연구 결과에 대한 소개와 해석을 한다. 6장에서는 본 연구에 대해 결론을 내며, 한계점과 미래 연구를 위한 지침에 관해 설명할 것이다.

2. 이론적 배경

2.1 주식 가격 이론

2.1.1 효율적 시장 가설

주식의 가격은 시장에서 판매자와 구매자의 실제 거래에서 오르고 내리며, 적절한 점을 찾아간다. 따라서, 주가는 수요의 공급 법칙에 따라 정해진다(Nassirtoussi et al., 2014). 이때, 주가는 수요는 투자자들이 생각하기에 기업의 성과가 뛰어나다면 올라갈 것이며, 투자자

들이 생각하기에 기업의 성과가 저조하다면 공급이 늘어날 것이다. 즉, 이론적으로 주가는 회사의 미래의 성과에 대한 기댓값이다. 하지만, 회사의 미래 성과에 대한 기댓값에 영향을 미칠 수 있는 요소는 매우 많다. 학계에서는 주가에 관련된 여러 이론을 제시하는데, 가장 대표적인 것이 Fama(1965)[17]의 효율적 시장 가설이다.

효율적 시장 가설은 시장의 가격은 가치를 정확하게 반영하며, 오직 새로운 정보에만 반응한다는 내용이다. 효율적 시장 가설은 정보가 어느 정도로 반영이 되는 지에 따라 3가지로 나뉜다. 이는 다음과 같다.

약형 효율적 시장

약형 효율적 시장에서 현재의 시장에서 거래가 가능한 주식, 채권, 유형 자산 등 모든 금융 자산의 가격은 이용 가능한 모든 과거의 주가에 대한 정보를 이미 현재의 가격에 반영하여, 기술적 분석을 통해서 분석하는 것은 초과 수익을 낼 수 없다. 즉, 내부의 정보를 이용하여 주가를 예측하는 것은 무의미하다.

중강형 효율적 시장

중강형 효율적 시장에서는 과거 주가의 변화, 공시, 뉴스 등의 모든 공개적인 외부 정보들이 이미 금융 자산에 반영되어 있어, 이를 이용한 예측은 시장에서 초과 수익을 낼 수 없다.

강형 효율적 시장

강형 효율적 시장에서는 과거의 가격과 공개된 정보에 더해서 비공개 정보까지도 시장의 가격에 반영되어 있으므로, 이를 이용한 예측은 시장에서 초과수익을 낼 수 없다.

2.1.2 허스트 지수(Hurst Exponent)

약형 효율적 시장은 과거의 주가가 이미 현재의 주가에 반영되어 있다는 개념이다. 이러한 시장의 상태를 정량적으로 측정하기 위하여 여러 가지 측정 방법이 시도됐는데, 그중 경제 물리학에서 주로 사용하는 개념 중 하나가 허스트 지수이다[8][10][16]. 허스트 지수는 주식의 장기 기억을 정량적으로 측정하는 개념이다[22]. x_i 를 시계열의 한 지점이라고 생각하자. DFA 방법[40]을 사용할 때, 허스트 지수는 다음과 같이 정의된다[12][54].

$$y(i) = \sum_{i=1}^N (x(i) - \bar{x})$$

$$F(n) = \sqrt{\frac{1}{N} \sum_{i=1}^N [y(i) - y_n(i)]^2}$$

$$F(n) = c * n^H$$

이때, 허스트 지수가 0.5보다 작다면 회귀형, 0.5이면 무작위 운동, 0.5보다 크다면 경향성을 가진다고 해석할 수 있다. 즉, 허스트 지수가 커질수록, 과거의 가격이 현재의 가격에 미치는 영향이 더 높아진다. 즉, 과거의 주가의 패턴이 현재의 패턴을 예측하는 데에 도움을 줄 수 있다는 것이다. 즉, 허스트 지수가 높다면, 주식 시장은 기술적 분석을 통해 예측할 수 있으므로, 약형 효율적 시장이 아니게 된다.

Eomet al.(2008)[16]에서 볼 수 있듯이, 다우 존스 지수의 허스트 지수는 0.5보다 약간 작은 값을 가진다. 즉, 다우 존스 지수는 장기 기억을 가지지 않고, 무작위 운동 형태를 가지기 때문에 다우 존스 지수는 약형 효율적 시장에 속함을 알 수가 있다. 따라서 다우 존스 지수를 분석할 때, 기술적인 지표를 사용하는 것이 아닌 외부의 정보만을 가지고 주식을 예측하는 것이 유의미하다. 따라서, 본 논문에서는 기술적 지표가 아닌, 외부의 정보인 뉴스와 다른 거시적인 지표들을 이용하고 예측하였다.

2.2 기존 주가 예측 논문의 흐름

주가를 예측하는 것은 학문적으로도 산업적으로도 중요한 문제이다. 따라서, 많은 주가 예측 연구가 시도되어 오고 있다. 기존의 연구에서 주가를 예측하는 연구는 시계열 데이터를 이용하여 예측하는 것이 일반적이었다. 시계열 데이터를 가지고 한 연구는 크게 2가지로 나뉘며, 기술적 분석(Technical Analysis)과 거시경제적 분석(Macro-economic analysis)이 있다[52].

기술적 분석은 흔히 말하는 차트 분석으로 주식 투자에 전통적으로 사용되어 왔다[24][31]. 이것은 과거 주가의 움직임과 거래량에 기반을 두어 미래를 예측하는 것이다. 가장 대표적인 예시로는 MA(Moving Average)[23][31], RSI(Relative Strength Index)[19] 등이 있다. 기술적 분석은 주식의 가치를 평가하여 주식 투자를 하는 데에 도움을 줄 수 있다[19]. 이에 대한 자세한 설명은 Nazário et al.(2017)[37]에서 확인할 수 있다.

거시 경제 분석은 경제 지표가 주가에 미치는 영향을 계산하여 미래를 예측하는 것이다[52]. 경제 지표는

주식에 직접 영향을 미치는데, 환율, 금값, 물가상승률, 고용률, 금리 등이 가장 대표적인 예시이다. 이들을 고려하여 투자자들은 주식을 구매할지, 하지 않을지에 대해 결정을 할 수 있다[23].

이런 기술적 분석과 거시 경제 지표로 주가를 통합하여 예측하는 연구는 꾸준히 진행돼 왔다[23][26][52]. Lawrence(1997)[26]에서는 금값, 외환 환율, 이자율과 같은 거시 경제적 지표와 기술적 지표를 통합하여 인공 신경망으로 예측하였다. 이때, 거시 경제적 지표와 기술적 지표를 통합하는 것이 자기 자신만을 가지고 예측하는 것보다 더 좋은 성능을 보인다고 하였다. Lam(2004)[23]에서는, 16개의 금융지표와 11개의 거시 경제적 지표를 인공 신경망을 통해 예측하였다. 하지만, 이 논문에서는 불필요한 지표를 첨부한다면 오히려 예측률을 떨어뜨릴 수 있다는 것을 명시하였다. 따라서, 모든 거시 경제 지표나 기술적 지표를 사용하는 것이 아니라, 적절한 지표를 선정하는 것이 중요한 문제임을 알 수 있다. Tsai and Hsiao(2010)[52]에서는 이를 해결하기 위하여 금융 지표들과 거시경제 지표, 과거 주식 데이터를 주성분 분석(Principal Component Analysis), 의사 결정 트리(Decision Tree), 유전자 알고리즘(Genetic Algorithms)을 이용해 특징 추출을 하고 앙상블 기법을 이용하여 예측하였다.

이렇듯 과거부터 거시 경제적 지표와 기술적 지표를 통합하여 예측하려는 연구는 많이 됐으며, 그중에서 필요한 변수만을 선별하는 것이 중요하다고 언급돼 왔다. 하지만, 2.1장에서 볼 수 있듯이, 다우 존스 지수를 예측할 때는 다우 존스 지수가 약형 효율적 시장에 속하기 때문에 기술적 분석을 사용하는 것은 무의미하다. 따라서, 본 논문에서는 주가를 예측할 때 필수적인 지표를 거시 경제적 지표에서 찾으며, 이는 2.3장에서 언급된다.

2.3 언론을 기반으로 한 주가 예측 연구

문자 형태의 정보를 이용하여 주가를 예측하는 연구에는 많이 시도 되고 있다[2][6]. 예를 들어, Lavrenko et al.(2000)[25]에서는 주가의 경향성과 금융 뉴스를 통합하고, 뉴스를 이용하여 경향성이 나타나기 전에 예측하였다. 또한, Schumaker and Chen(2009)[47]에서는 Bag-of-words, 고유명사, 명사 등을 이용한 모델들을 여러 가지를 비교하였다. 또한, Hagenau et al.(2013)[19]에서는 Bag-of-words, 2-gram, 명사, 고유명사, 2-Word combinations 등의 방법을 비교하였다. 이처럼, 기존의 많은 연구가 Bag-of-Words 모형과 그 응용을 채택하고 있다. 이들이 가지던 공통적인 연구 모형은 <그림 1>

에서 나타나는 것과 같다. 하지만, Bag-of-Words 모형은 단어의 순서나 내재한 의미를 파악하지 못한다는 점에서 한계를 가진다.

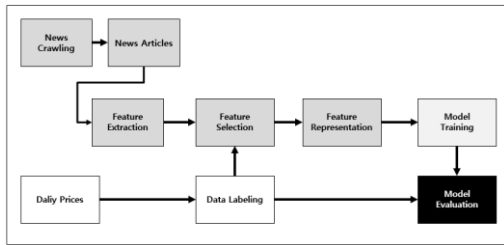


그림 1. Previous Studies of News

이러한 문제를 해결하기 위해, Word2Vec 모형[35]이 나오게 된다. Word2Vec은 딥러닝에 기반을 둔 단어를 벡터의 형태로 바꾸어 주는 알고리즘이다. Word2Vec은 크게 두 가지 형태를 가지는데, CBOW(Continuous Bag-of-words)모형과 Skip-gram 모형이다. CBOW는 문맥에 기반을 두어 현재 단어를 예측하는 것이고, Skip-gram은 현재 단어를 기반을 두어 주위 단어를 예측하는 것이다. 자세한 알고리즘은 Mikolov et al.(2013)[35]에 묘사되어있다. 최근에 Word2Vec 모형을 이용한 감성 분석, 주가 예측 등의 연구가 많이 시행되고 있는데, 기존 연구에서 많이 사용되던 Bag-of-Word 모형보다 뛰어나다는 평가를 받고 있다[55].

따라서, 본 논문에서는 내재적 의미와 단어의 순서를 파악할 수 있도록, Word2Vec 모형을 사용한다.

2.4 거시 경제 지표가 주가에 미치는 영향

Deng et al. (2011)[11]에서 적절한 지표를 사용하여 뉴스 정보와 함께 예측하는 것은 예측률을 더 향상할 수 있다. 기존의 연구들에서 볼 때, 이론적으로 주가에 영향을 미칠 수 있는 경제적 요인이 여러 가지가 있는데, 그중 금값과 환율이 대표적이다.

금값은 오랫동안 안전 자산으로 고려됐다. 금값은 가격변동이 크지 않고, 금융 시장의 안전성을 높여 준다[4]. 따라서, 2016년 브렉시트와 같은 사건이 일어나며 시장에 대한 불안감이 늘어날 때, 금값은 오른다. 따라서 이러한 측면에서 바라볼 때, 금값과 주식 시장은 밀접한 관계가 있다. Smith(2001)[48]에서는 1991년 1월부터 2001년 10월까지 금값과 미국의 주식 관계를 조사하였다. 이때, 월 단위에서 금값과 주가는 통계적으로 유의한 관계가 없으며, 일 단위에서는 오후 시간대에서 음의 상관관계를 가지며, 주 단위에서 또한 금

값과 주가는 영향을 가진다. 본 논문에서 제시하는 딥러닝 알고리즘은 주 단위의 기억을 기반으로 주가를 예측하는 것이기 때문에, 단기적으로 금값의 증가를 가지고 확인을 하는 것이 의미를 가진다. 따라서, 본 논문에서는 금값을 단기적 지표로 미국의 주가를 예측하는 하나의 변수로 사용하여 예측한다.

또한, 기존의 경제학 이론에서 주가와 환율과의 관계는 명백하게 언급되고 있다. 예를 들어, Dornbusch and Fisher(1980)[14]의 논문에 따르면 통화의 움직임이 국제적인 경쟁력에 영향을 미치고, 무역적 위치의 균형에 영향을 미친다. 따라서 환율의 변화는 국가의 실질적인 결과물에 영향을 미치며, 기업들의 미래 가치와 유동 자산 흐름에 영향을 미치며 주가에 영향을 미친다. 기존의 연구들을 살펴보면, 많은 연구가 미국의 주식 시장과 외환 시장의 관계를 연구하려고 노력하였다[9][30][32][45][49]. 예를 들어, Soenen and Hennigar(1988)[49]의 연구에서는 1980년부터 1986년까지의 주식 시장과 외환 시장의 관계를 연구하였는데, 통계적으로 유의한 음의 상관 관계를 가짐을 보였다. 특히, Abdalla and Murinde(1997)[1]와 Blose and Shieh (1995)[5]에서 신흥 시장의 환율과 주가지수의 관계가 명확함을 보여 주면서 신흥 시장인 한국의 환율을 하나의 지표로 사용하는 것이 정당하다고 입증한다. 또한, Phylaktis and Ravazzolo(2005)[42]에서는 태평양 연안에 인접한 국가들과 미국 달러의 환율이 미국의 주식 시장에 미치는 영향을 확인하며, 한국, 필리핀, 태국, 싱가포르의 환율이 미국 주가에 영향을 미치는 것을 확인하였다. 또한, Andreou et al.(2013)[3]에서는 신흥 국가들과의 환율과 미국의 주식 시장의 쌍방향의 연결 고리를 확인하였다. 이때, 베네수엘라, 한국, 필리핀, 태국 네 나라에서 환율이 미국의 주식 시장에 영향을 미침을 Granger Causality를 통해서 확인하였다. 따라서, 한국의 환율은 미국의 다우 존스 지수에 영향을 미친다고 확인할 수 있으며, 본 논문에서는 많은 연구에서 공통적으로 영향이 있다고 언급한 한국, 필리핀, 태국의 환율로 미국의 다우 존스 지수를 예측한다.

2.5 머신러닝 방법론

모든 준비 과정이 끝나면, 머신러닝 방법론을 이용하여 시장이 어떻게 될지를 예측을 한다. 많은 알고리즘이 주식 시장을 예측하는 데 사용되어 왔다. 그중에서도 특히 서포트 벡터 머신[23], 인공 신경망[18][41] 등이 많이 사용되었다. 또한, 서포트 벡터 회귀는 중국 주식 시장에서 금융 뉴스의 영향을 알아보기 위하여 사용되기도 하였으며[56], 주가의 오르고 내림이 아닌 주가 자체의 크기를 반영해서 알아낼 때 사용되기도

하였다[19][47]. 그러나, 서포트 벡터 회귀는 하나의 종류의 자료에만 적용된다[11]. 따라서, 본 논문의 거시 경제 지표와 뉴스 정보를 통합하기에 서포트 벡터 회귀는 사용하기 적절하지 않다. 또한, 기존의 신경망 구조에서는 모든 입력이 독립적이라고 가정하지만, 현실에서는 그렇지 않은 경우가 많다. 예를 들면, 사진 자료 같은 경우에는 현재 있는 픽셀이 옆의 픽셀과 많은 연관성을 가지고 있는데, 기존의 신경망으로는 이것을 반영할 수가 없다. 또한, 주식 데이터의 경우에도 이전의 기록들에 영향을 받지만, 이러한 점이 반영되지 않는다. 따라서 시계열 데이터를 주로 사용하는 본 논문에서는 인공신경망을 사용하기는 적합하지 않다.

기존의 신경망이 여러 한계점을 가지기에, 이것을 해결하기 위하여, 메모리를 가지고 있는 네트워크를 고안하였다. 이러한 메모리를 가진 딥러닝 알고리즘을 순환 신경망(Recurrent Neural Network)이라고 한다[15].

딥러닝 알고리즘인 순환 신경망 중에도 Long Short-Term Memory(이하 LSTM)[20] 은 순환 신경망 중에서 가장 유명한 구조이다. LSTM은 네트워크의 은닉층에서 기존의 전통적인 인공 신경망의 역할을 하는 기억 세포의 개념을 도입하면서, 시간에 따른 동적인 데이터를 효율적으로 처리할 수 있다.

이러한 LSTM은 순서를 가지는 데이터를 처리하는 데에 매우 적합하다. 예를 들어, 소리, 시계열 데이터, 자연어 등을 처리하는 데에 매우 효율적이다[30]. 이러한 특징을 사용해서, 순환 신경망은 시계열 데이터인 주가를 예측하는 데에도 사용되는 경우가 많다[44][46]. 특히, Manuel et al.(2017)에서는 일일 내에서 S&P 500 지수를 금융 뉴스와 기술적 분석을 딥러닝 기술을 예측하였다. 이때, 저자들은 CNN과 LSTM 같은 알고리즘들을 사용하였는데, LSTM이 정보의 문맥을 잡는 것에 유리하고, 복잡하고 일시적인 특성의 데이터들을 다루는데 유리하다는 것을 밝혀냈다. 따라서, 본 논문에서는 LSTM 네트워크를 사용하여 거시 경제 지표와 금융 뉴스를 통합하였다. 또한, Nelson et al.(2017)[39]에서는 브라질 주식 시장의 많은 회사의 주가들을 가지고, 어떤 회사의 주식이 15분 뒤에 어떤 방향성을 가지고 움직일지에 대해 LSTM 네트워크를 이용하여 예측하였다. 이때, 무작위 결정 트리, 무작위 예측, 인공 신경망 등과 비교를 하였는데, 몇 개의 예외를 제외하고는 LSTM으로 예측한 주식의 움직임이 다른 방법들 보다 우월한 예측력을 보이며, 예측의 위험도를 줄일 수 있다고 하였다. 하지만, 저자들은 단순히 다른 회사들의 주가의 영향만을 고려할 뿐, 주가에 미칠 수 있는 뉴스와 같은 외부 요인을 함께 고려하지 못했다는 한계점을 가진다. 따라서, 본 논문에서는 시계열 분석과 외

부 요인들을 함께 고려할 수 있는 모델을 제시한다.

본 논문에서는 자연어와 시계열 데이터를 모두 다룰 수 있는 딥러닝 알고리즘(LSTM)을 사용하여 두 데이터를 통합하고, 시간에 따른 흐름을 적절하게 고려하면서 서포트 벡터 회귀와 인공 신경망이 가지는 한계점을 극복하였다.

3. 데이터

제3장에서는 데이터들을 어떻게 수집하였는지와 이를 어떻게 처리하였는지에 대해 설명한다. 3.1.에서는 뉴스 데이터를 어떻게 얻어왔는지와 어떤 전처리 과정을 거쳤는지에 대해 설명하고, 3.2.에서는 거시 경제 데이터의 특성과 수집 과정에 관해 설명한다.

3.1 뉴스 데이터

뉴스 데이터는 2014년 11월부터 2016년 10월까지 2년 동안의 New York Times 기사를 사용하였다. 데이터는 New York Times Archive API를 사용하여 뉴스 제목, 날짜, 기사 종류, URL로 이루어진 리스트를 가져왔다. 각 URL을 파이썬의 BeautifulSoup 모듈을 이용하여 뉴스를 크롤링하였다. 데이터의 구조는 뉴스의 고유 식별 번호, 제목, 날짜, 본문, 본문의 유형, 자료의 유형, 기사의 종류가 있다. 전체 2년 동안의 기사 개수는 114,230개가 있다. 기존의 논문들에서 모든 뉴스를 사용하는 것보다 기사의 종류가 금융인 뉴스만을 사용하는 것이 효율성이 높다고 판단하여 사용하였기에, 본 논문에서 또한 금융 관련 기사만 채택하였다[13][19]. 이때, 금융 관련 뉴스는 18176개가 있었다.

3.2 거시 경제 데이터

거시 경제 변수로 사용할 금리와 환율, 다우 존스 지수는 Wisefn에서 제공하는 Quantiwise를 사용하여 주가, 지수, 금값, 환율의 일별 데이터를 가져온다.

주가, 금값, 환율 등의 데이터는 그 날의 종가를 기준으로 한다. 이 데이터들의 날짜는 특성에 따라 각각 다르다. 예를 들어, 한국의 증시 같은 경우에는 매년 8월 15일 광복절이 없으며, 미국 같은 경우에는 미국의 공휴일에 따라 시장이 열리지 않는다. 이에 패널 데이터의 형태가 시간에 따라서 모두 같지 않기에, 이를 해결하기 위하여 같은 날짜의 데이터만 사용하도록 한다. 이때, 변수들의 데이터의 특성은 아래 <표 1>과 같다.

표 1. Summary Statistics

	count	mean	std	min	25%	50%	75%	max
Dow Jones Index	484	17458.99	703.4035	15660.18	16938.72	17623.46	17977.62	19083.18
Gold Prices	484	44058.19	2480.847	39900	42422.5	43050	46175	50910
KRW-USD FX	484	1116.408	56.99234	1008.5	1081.6	1116.4	1163.1	1238.8
THB-USD FX	484	34.4610	1.27987	32.0519	32.9350	34.9655	35.54112	36.4310
PHP-USD FX	484	45.8355	1.1319	43.882	44.6985	46.1625	46.805	47.8890

이 데이터를 시간 축에 따라서 그려보면 <그림 2>와 같다.

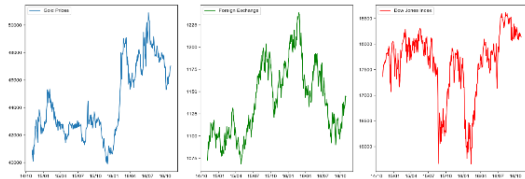


그림 2. Graphs of Variables

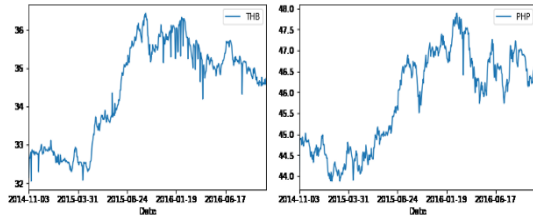


그림 3. Graph of Variables

4. 연구 모형

이번 장에서는 연구가 진행되는 자세한 방향을 제시한다. 먼저, 문자열 데이터를 다루는 방법에 관해서 설명하고, 그 기초 작업이 어떤 것이 있는지에 대해 설명한다. 또한, 문자열 데이터가 어떻게 시계열 데이터와 함께 통합되는지 설명하며, 어떻게 머신 러닝으로 이것을 예측하는지에 대해 설명한다. 연구의 전체적인 개요는 <그림 4>에서 확인할 수 있다.

4.1 연구개요

2장에서 언급하였듯이, 다우존스 지수는 금값과 한국 미국 환율과 같은 거시 경제 지표에 영향을 받으며, 뉴스의 정보를 나타낸다. 이를 수식으로 표현하면 다음과 같다.

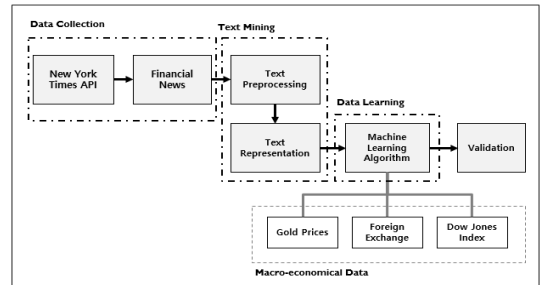


그림 4. 연구 모형

$$\ln(p_t) < -\ln(G_{t-1}) + Ex_{t-1} + text_{t-1} + \dots$$

p_t 는 날짜 t 에서의 다우 존스 지수를 의미하며, G_t 는 날짜 t 에서의 금값을 의미한다. Ex_t 는 날짜 t 에서의 한국-미국 환율, 태국-미국 환율, 필리핀-미국 환율의 벡터를 의미하며, $text_t$ 는 날짜 t 에서의 뉴스 정보를 의미한다. 이때 뉴스 정보는 날짜 t 에 나온 모든 금융 뉴스를 의미한다. 하지만 장이 마감된 후에 나온 뉴스는 날짜 t 의 다우 존스 지수에 영향을 준다고 보기 힘들기 때문에 $t+1$ 에 영향을 준다고 본다[27]. 뉴스 정보는 원형으로 만들어진 후, Word2Vec 모형을 통해 300 차원의 벡터로 표현이 되며, 이는 4.2에서 자세하게 다룰 것이다. 이후 이들은 위의 공식과 같이 LSTM을 통해 통합되어 예측될 것이며, 이는 4.3장에서 자세히 다룰 것이다. 오차는 Deng et al.(2011)[11]과 마찬가지로 RMSE(Root Mean Square Error)로 측정을 한다. 이때, 본 연구에서 주장하고자 하는 바인, 거시 경제 지표와 뉴스 정보를 통합하여 예측하는 것이 예측력을 높여 준다는 것을 증명하기 위해, 각각의 요소로 예측하고, 통합하고 예측하여 비교를 한다. 또한, 기존 논문들에서 제시한 예측 방법과 LSTM을 비교하여 타당성을 입증한다. 또한, 이 결과가 예측 기간에 한정적이지 않다는 것을 보이기 위하여 4.4에서와 같이 상호 검증을 하였다.

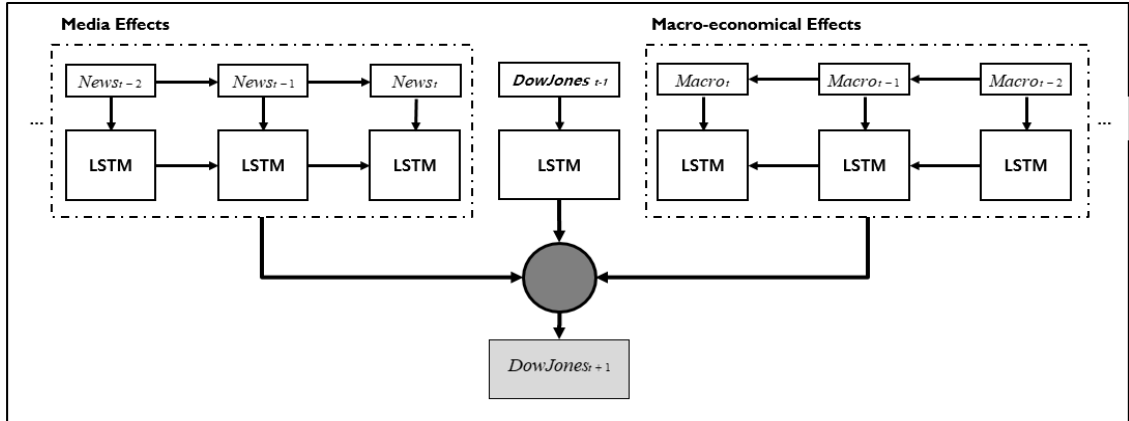


그림 5. 딥러닝 예측 방법

4.2 뉴스 기사 분석

뉴스를 받은 후 Toutanova et al.(2003)[51]의 방법을 사용하여 형태소 분석을 진행한다. 후에 Porter(1980)[43]를 통해 어간만을 추출한다. 또한, 불용어(stopword)를 모두 제거한 후, 영문 소문자로 바꾸는 전처리 과정을 거친다. 이때, 모든 단어는 “단어 원형/형태소”의 형태를 가진다. 이후 Mikolov et al.(2013)[35]의 Word2Vec을 skip-gram에 기반하여 파이썬의 gensim 패키지를 이용하여 훈련 기간의 뉴스 기사를 300차원으로 축소를 시킨다. 이때, 하루의 $text_t$ 는 단어들의 벡터에 TF-IDF 가중치를 주어 만들어진다. 또한 훈련 기간에 훈련된 단어들을 이용한 Word2Vec 모델을 통해 예측 기간의 뉴스 기사를 변형하여 주가를 예측하는 데 사용한다.

4.3 딥러닝 알고리즘

본 논문에서 거시 경제 지표와 뉴스 정보를 통합한 모형은 <그림 5>에서 제시된 바와 같다. 먼저, 뉴스를 4.2에서 제시된 방법으로 뉴스를 300차원의 데이터로 합친 후 일별로 통합한 것을 <그림 5>에서 ‘News’라고 하였으며, 거시 경제 지표를 나타낸 것들을 ‘Macro’라고 하였다. 뉴스와 거시 경제 지표 모두 Deng et al.(2011)[11]에서와 마찬가지로 시간 $t-7 \sim t-1$ 까지의 데이터를 사용하여 시간 t 의 데이터를 예측하였다.

이때, $News_t$ 와 $Macro_t$ 와 $DowJones_t$ 를 직접 통합하기에는 서로 다른 차원의 벡터를 가지고 있으므로 쉽지 않다. $News_t$ 이 300차원으로 $Macro_t$ 와 $DowJones_t$ 보다 매우 큰 벡터를 가지고 있기에, 단순히 통합한다면 $News_t$ 에 많은 영향을 받게 될 것이다. 이것은 예측을 방해하는 요소가 된다. 따라서, 이를 해

결하기 위하여, 본 논문에서는 LSTM 네트워크를 거쳐서 차원을 조절해준다. LSTM 네트워크에서 각각 1, 2, 1차원으로 축소를 한 뒤에 그 값을 입력값으로 하여 예측하였다.

이를 수식으로 표현하면 다음과 같다.

$$N_t = f(News_t, News_{t-1}, News_{t-2} \dots)$$

$$M_t = f(Macro_t, Macro_{t-1}, Macro_{t-2} \dots)$$

$$x_t = [N_t, M_t]$$

$$DowJones_{t+1} = D(x_t) = Wx_t + b$$

$$f(x) = LSTM(x)$$

$$D(x_t) = DNN(x)$$

N_t, M_t 는 뉴스 정보, 거시 경제 정보 두 가지 변수가 LSTM을 거친 것을 의미하며, LSTM함수는 Hochreiter and Schmidhuber(1997)[20]에서 언급한 LSTM 네트워크 함수이다. 또한, $DNN(x)$ 는 LSTM 함수로 차원 축소를 시행한 값을 입력값으로 하는 인공신경망으로 다우 존스 지수를 최종적으로 예측한다.

자료형은 모두 입력될 때의 0~1까지의 값을 가지도록 정규화하여 훈련에 걸리는 시간을 단축했다. 활성화 함수는 선형 함수(Linear)가 사용되었으며, ‘Adam’을 통하여 최적화를 하였다. 또한, Early Stopping 기술을 사용하여, 가장 많은 딥러닝 반복 횟수 중에 가장 좋은 반복 횟수로 예측을 한다.

4.4 상호 검증(Rolling Window with 10)

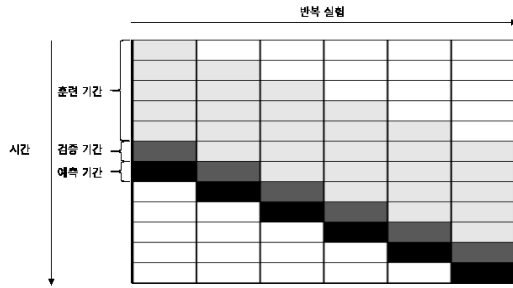


그림 6. 롤링 윈도우 방법

3장에서 언급하였던 것과 같이, 우리는 2014년 11월부터 2016년 10월까지 총 2년간의 뉴스와 거시 경제 지표들에 대한 정보를 수집하였다. 본 논문에서는 예측 기간에 따라 실험 결과가 다르지 않다는 것을 보이기 위하여, 예측 기간을 바꾸어 가면서 실험을 하였다. 특히, 시장의 상태는 시간에 따라서 바뀌기 때문에, 모든 기간을 대상으로 예측을 하는 것보다, 특정한 기간을 대상으로 시스템을 훈련하고 예측을 하는 것이 효율적이다[28][29]. 따라서, 본 논문에서는 롤링 윈도우 기법을 사용한다. 롤링 윈도우 기법은 예측 기간을 바꾸어 가며 예측을 하는 것이다. 이에 대한 대략적인 그림은 <그림 6>에 나타나 있다. 연한 회색은 훈련 기간을 의미하고, 짙은 회색은 검증 기간을 의미하며, 검은색은 예측 기간을 의미한다. 본 논문에서는 롤링 윈도우 기법을 사용하여 1년을 예측 기간으로 하고, 1개월을 검증 기간으로 하여 파라미터들을 최적화하고, 나머지 1개월을 예측 기간으로 하여 실제 참값과 비교를 하였다.

총 24개월의 기간 동안 롤링 윈도우 기법을 사용하면 총 10번의 실험이 있다. 이를 실험 1부터 10으로 명명하였다. 이 10번의 반복되는 실험에서, 본 연구에서 제시한 모델이 타당하다는 것을 보이기 위하여, 두 가지로 실험을 진행한다. 첫 번째는 거시 경제 지표와 뉴스 정보를 통합하는 것이 각각을 사용하는 것보다 효과적이라는 것을 확인하기 위하여, 각 요소만 가지고 LSTM 네트워크로 예측을 한 결과를 보여 주며, 이를 통합하여 예측한 결과와 비교한다. 또한, 두 번째로는, 기존의 연구들과는 다르게, LSTM 네트워크를 사용하는 것의 타당성을 확인하기 위하여, 기존의 연구들에서 사용한 서포트 벡터 회귀와 인공 신경망과 비교를 한다. 이때, Manjula et al.(2011)[34]과 같이 인공신경망은 은닉층을 1개를 가지며, 은닉층의 노드는 입력 값과 같은 개수를 가지고, dropout은 0.1, 0.2...0.9 중에

그리드 서치로 검증 기간에서 가장 좋은 값을 가지는 것으로 예측을 한다. 또한, 서포트 벡터 회귀는 Hsu et al.(2010)[21]에서 제시한 것과 같이, C 와 γ 를 $\{2^{-15}, \dots, 2^7\}$ 의 범위에서 그리드 서치를 하고 검증 기간에서 가장 좋은 값으로 예측을 한다.

5. 결과

이번 장에서는 4장에서 제시한 연구 모델로 나온 결과에 대해 논의한다. 5.1에서는 거시 경제 지표와 뉴스 정보를 비교한 결과를 의미하며, 5.2에서는 타 논문에서 사용한 통합 알고리즘과 본 논문에서 제시한 LSTM 네트워크를 비교한다. 5.1에서와 5.2에서 모두 본 논문에서 제시하는 거시 경제 지표와 뉴스 정보를 LSTM 네트워크로 예측하는 것이 가장 정확하다는 결과를 보여 주었다.

5.1 요소들과 통합 비교

표 2. 통합과 거시 경제 지표들과 뉴스의 비교

실험	통합	거시	뉴스
1	0.008821	0.0155623	0.0104888
2	0.008808	0.0109528	0.0182282
3	0.011848	0.0193510	0.0163601
4	0.014722	0.0186257	0.0197336
5	0.014501	0.0180821	0.0168508
6	0.006952	0.0079353	0.0088280
7	0.007665	0.0150662	0.0139053
8	0.014452	0.0146271	0.0156028
9	0.014237	0.0166006	0.0132151
10	0.006912	0.0087240	0.0092470
평균	0.010892	0.0145527	0.0142460

<표 2>는 실험을 1~10을 진행할 때 예측 기간에서 본 논문에서 제시한 LSTM 네트워크로 예측한 RMSE(Root Mean Square Error)를 나타낸 것이다. ‘통합’은 거시경제 지표와 뉴스 정보를 함께 통합하여 예측한 결과를 나타내며, ‘거시’는 거시경제 지표만을 가지고 예측한 결과를 나타내고, ‘뉴스’는 뉴스 정보만을 가지고 예측한 정보를 의미한다. ‘실험’은 4.4장에서 언급한 롤링 윈도우 방법으로 훈련 기간과 예측 기간을 바꾸는 실험 0~10을 의미한다. <표 2>에서는 대다수의 실험에서 본 연구에서 제시한 연구 모형이 가장 낮은 것을 보여 주며, 기존 연구에서 사용되던 모형들보다 좋은 결과를 나타낸다는 것을 보여 준다. 이것은 거시 경제 지표와 뉴스 정보를 함께 통합하여 예측한 것이 각각 하나의 정보만을 가지고 예측을 하는 것보

다 낮다는 것을 의미하며, 거시 경제 지표와 뉴스 데이터는 다른 특성을 가져 함께 예측하는 것이 더 우월하다는 것을 의미한다.

5.2 연구 알고리즘의 비교

표 3. 알고리즘별 비교

실험	LSTM	ANN	SVR
1	0.008821	0.0140249	0.0106256
2	0.0088086	0.0210480	0.0098922
3	0.0118485	0.012766	0.0599569
4	0.0147221	0.0170010	0.0591931
5	0.014501	0.0189676	0.0319179
6	0.0069525	0.0162275	0.0121563
7	0.0076650	0.0088857	0.0096192
8	0.0144528	0.0145418	0.0137195
9	0.0142370	0.0143156	0.0343070
10	0.0069121	0.0098132	0.0423755
평균	0.0108920	0.0147591	0.0283763

<표 3>은 실험을 1~10을 진행할 때 예측 기간에서 거시경제 지표와 뉴스 정보를 함께 통합하여 여러 가지 알고리즘으로 비교해 본 결과를 나타낸다. 즉, 본 논문에서 제시한 LSTM 네트워크로 예측한 RMSE(Root Mean Square Error)와 다른 논문에서 서포트 벡터 회귀(Support Vector Regression)와 인공신경망(Artificial Neural Network) 방법을 비교하여 나타낸 것이다. 'LSTM'은 본 논문에서 제시한 LSTM 네트워크를 의미하며, 'ANN'은 인공 신경망을 의미하며, 'SVR'은 서포트 벡터 회귀를 의미한다. '실험'은 4.4장에서 언급한 롤링 윈도우 방법으로 훈련 기간과 예측 기간을 바꾸는 실험 1~10을 의미한다. <표 3>에서는 대부분의 예측 기간에서 본 연구에서 제시한 연구 모형이 가장 낮다는 것을 보여 주며, 기존 연구에서 사용되던 모형들보다 좋은 결과를 나타낸다는 것을 보여 준다. 이것은 LSTM이 인공 신경망과 서포트 벡터 회귀보다 데이터를 잘 통합하여 예측한다는 것을 의미한다. 서포트 벡터 회귀는 다른 특성을 가지는 여러 개의 데이터에는 적합하지 않기 때문에[11], 기존의 연구들과는 다르게 우월한 결과를 보이지 못하였으며, 인공 신경망은 여러 데이터를 통합할 수 있으나, 시계열 데이터를 반영할 수 있는 구조가 아니므로, LSTM이 더 나은 결과를 보여 주었다고 판단할 수 있다.

6. 결론 및 향후 연구

주가에 영향을 줄 수 있는 요소는 무수히 많다. 다

른 나라의 경제 상황에 따라서 주가가 변하기도 하며, 새로운 정보가 유입될 때 주가가 변하기도 한다. 본 논문에서는 거시 경제 지표와 뉴스 정보를 통합하는 것이 어떻게 주가를 더 잘 예측할 수 있는지에 대해 연구하였다.

데이터는 뉴욕 타임스의 2014년 11월부터 2016년 10월까지의 금융 뉴스와 그 시기의 금값, 한-미 환율, 다우 존스 지수 증가를 사용하였다. 논문에서 제시한 방법은 다음과 같다. 먼저, 뉴스를 전처리 과정을 거친 후 Word2Vec 모형으로 벡터화를 시킨다. 이후 거시 경제 지표와 함께 LSTM으로 예측하는 것이다. 이 방법은 기존의 연구에서 사용되던, 서포트 벡터 회귀와 인공 신경망을 이용한 것보다 월등한 결과를 보였다. 또한, 뉴스 정보만을 가지고 예측하였을 때와 거시 경제 지표만을 가지고 예측하는 것보다 두 요소를 통합하여 예측하는 것이 더 월등한 성능을 보였다. 또한, 본 논문에서는 특정 기간에만 연구에서 제시한 모델이 우월한 결과를 보여 주는 것이 아님을 증명하기 위하여 예측하는 기간을 바꿔가면서 실험을 하였으나, 대부분 기간에서 본 연구에서 제시한 LSTM 네트워크로 거시 경제 지표와 뉴스 정보를 함께 통합한 것이 우월한 결과를 보였다.

후속 연구에서는 다른 추가적인 지표들을 이용하여 연구할 수 있을 것이다. 이자율과 같은 더 많은 경제적인 지표를 연구하여 포함할 수 있다면 더 좋은 결과를 나타낼 수 있을 것이다. 마지막으로, 뉴스의 정보뿐 아니라 대중의 심리 또한 중요한 역할을 미치므로, 트위터, 페이스북 등의 데이터를 함께 사용하는 것도 좋은 연구 결과를 보일 수 있을 것이다.

참고문헌

- [1] Abdalla, ISA., and V. Murinde, 1997. "Exchange rate and stock price interactions in emerging financial markets: evidence on India, Korea, Pakistan and the Philippines", *Applied financial economics*, Vol.7, No.1, pp.25-35.
- [2] Akita, R., A. Yoshihara, T. Matsubara, and K. Uehara, 2016. "Deep learning for stock prediction using numerical and textual information", *Computer and Information Science (ICIS)*, 2016 IEEE/ACIS 15th International Conference on. IEEE.
- [3] Andreou, E., Maria M., and Andreas S., 2013. "Stock and foreign exchange market linkages in emerging economies", *Journal of International Financial Markets, Institutions and Money*, Vol.27, No.1, pp.248-268.

- [4] Baur, D. G., and B. M. Lucey., 2010. "Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold", *Financial Review*, Vol.45, No.2, pp.217-229.
- [5] Blose, Laurence E., and Joseph CP Shieh, 1995. "The impact of gold price on the value of gold mining stock", *Review of Financial Economics*, Vol.4, No.2, pp.125-139.
- [6] Bollen, J., H. Mao, and X. Zeng, 2011. "Twitter mood predicts the stock market", *Journal of computational science*, Vol.2, No.1, pp.1-8.
- [7] Campbell, J. Y., and Shiller, R. J., 1987. "Cointegration and tests of present value models", *Journal of political economy*, Vol.95, No.1, pp.1062-1088.
- [8] Cajueiro, D. O., and Tabak, B. M., 2004. "The Hurst exponent over time: testing the assertion that emerging markets are becoming more efficient", *Physica A: Statistical Mechanics and its Applications*, Vol.336, No.3, pp.521-537.
- [9] Chow, E. H., W. Y. Lee, and M. E. Solt, 1997. "The exchange-rate risk exposure of asset returns", *Journal of Business*, pp.105-123.
- [10] Couillard, M., and Davison, M., 2005. "A comment on measuring the Hurst exponent of financial time series", *Physica A: Statistical Mechanics and its Applications*, Vol.348, pp.404-418.
- [11] Deng, S., T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai, 2011. "Combining technical analysis with sentiment analysis for stock price prediction", Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on. IEEE.
- [12] Di Matteo, T., 2007. "Multi-scaling in finance", *Quantitative finance*, Vol.7, No.1, pp.21-36.
- [13] Ding, X., Y. Zhang, T. Liu, and J. Duan, 2015. "Deep Learning for Event-Driven Stock Prediction", Ijcai.
- [14] Dornbusch, R. and S. Fischer, 1980. "Exchange rates and the current account", *The American Economic Review*, Vol.70, No.5, pp.960-971.
- [15] Elman, J. L., 1988. "Finding structure in time (CRL Technical Report 8801). La Jolla: University of California, San Diego", Center for Research in Language.
- [16] Eom, C., Choi, S., Oh, G., and Jung, W. S., 2008. Hurst exponent and prediction based on weak-form efficient market hypothesis of stock markets. *Physica A: Statistical Mechanics and its Applications*, Vol.387, No.18, pp.4630-4636.
- [17] Fama, E.F., Random walks in stock market prices, *Financial Analysis Journal* 21, 1965, pp.55-59.
- [18] Groth, S. S., and J. Muntermann, 2011. "An intraday market risk management approach based on textual analysis", *Decision Support Systems*, Vol.50, No.4, pp.680-691.
- [19] Hagenau, M., M. Liebmann, and D. Neumann, 2013. "Automated news reading: Stock price prediction based on financial news using context-capturing features", *Decision Support Systems*, Vol.55, No.3, pp.685-697.
- [20] Hochreiter, S., and J. Schmidhuber, 1997. "Long short-term memory", *Neural computation*, Vol.9, No.8, pp.1735-1780.
- [21] Hsu, C. W., Chang, C. C., and Lin, C. J., 2010. A practical guide to support vector classification. Taipei: Department of Computer Science National Taiwan University.
- [22] Hurst, H. E., 1951. Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116, 770-808.
- [23] Lam, M., 2004. "Neural network techniques for financial performance prediction: integrating fundamental and technical analysis", *Decision support systems*, Vol.37, No.4, pp.567-581.
- [24] Leigh, W., N. Modani, and R. Hightower, 2004. "A computational implementation of stock charting: abrupt volume increase as signal for movement in New York stock exchange composite index", *Decision Support Systems*, Vol.37, No.4, pp.515-530.
- [25] Lavrenko, V., Schmill, M., Lawrie, D., Ogilvie, P., Jensen, D., and Allan, J., 2000. "Mining of concurrent text and time series", In *KDD-2000 Workshop on Text Mining* Vol.2000, pp.37-44.
- [26] Lawrence, R., 1997. "Using neural networks to forecast stock market prices", University of Manitoba.
- [27] Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., and Chen, Y., 2014. "The effect of news and public mood on stock movements", *Information Sciences*, Vol.278, pp.826-840.
- [28] Lim, K. P., and Brooks, R. D., 2006. The evolving and relative efficiencies of stock markets: Empirical evidence from rolling bivariate correlation test statistics.

- [29] Lim, K. P., Brooks, R. D., and Kim, J. H., 2008. "Financial crisis and stock market efficiency: Empirical evidence from Asian countries", *International Review of Financial Analysis*, Vol.17, No.3, pp.571-591.
- [30] Lipton, Z. C., Berkowitz, J., and Elkan, C., 2015. "A critical review of recurrent neural networks for sequence learning", arXiv preprint arXiv:1506.00019.
- [31] Lo, A. W., H. Mamaysky, and J. Wang, 2000. "Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation", *The Journal of Finance*, Vol.55, No.4, pp.1705-1765.
- [32] Ma, C. K., and G. W. Kao, 1990. "On exchange rate changes and stock price reactions", *Journal of Business Finance & Accounting*, Vol.17, No.3, pp.441-449.
- [33] MacKinlay, A. C., 1997. "Event studies in economics and finance", *Journal of economic literature*, Vol.35, No.1, pp.13-39.
- [34] Manjula, B., S. S. V. N. Sarma, R. L. Naik, and G. Shruthi, 2011. "Stock prediction using neural network", *International Journal of Advanced Engineering Sciences and Technologies*, Vol.10, No.1, pp.13-18.
- [35] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J., 2013. "Distributed representations of words and phrases and their compositionality", *In Advances in neural information processing systems*, pp.3111-3119.
- [36] Nassirtoussi, A. K., S. Aghabozorgi, T. Y. Wah, and D. C. L. Ngo, 2014. "Text mining for market prediction: A systematic review", *Expert Systems with Applications*, Vol.41, No.16, pp.7653-7670.
- [37] Nazário, R. T. F., e Silva, J. L., Sobreiro, V. A., and Kimura, H., 2017. "A Literature Review Of Technical Analysis On Stock Markets", *The Quarterly Review of Economics and Finance*.
- [38] Neely, C. J., 1998. "Technical analysis and the profitability of US foreign exchange intervention", *Federal Reserve Bank of St. Louis Review*, (Jul), pp.3-17.
- [39] Nelson, D. M., Pereira, A. C., and de Oliveira, R. A., 2017. "Stock market's price movement prediction with LSTM neural networks. In Neural Networks (IJCNN)", 2017 International Joint Conference on (pp.1419-1426). IEEE.
- [40] Peng, C.K., Buldyrev, S.V., S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, *Physical Review E* 49, 1994, 1685.
- [41] Phua, P. K. H., Zhu, X., and Koh, C. H., 2003. "Forecasting stock index increments using neural networks with trust region methods", *In Neural Networks, 2003. Proceedings of the International Joint Conference on. IEEE*, Vol.1, pp.260-265.
- [42] Phylaktis, K., and F. Ravazzolo, 2005. "Stock prices and exchange rate dynamics", *Journal of International Money and Finance*, Vol.24, No.7, pp.1031-1053.
- [43] Porter, M. F., 1980. "An algorithm for suffix stripping", *Program*, Vol.14, No.3, pp.130-137.
- [44] Rather, A. M., A. Agarwal, and V. N. Sastry, 2015. "Recurrent neural network and a hybrid model for prediction of stock returns", *Expert Systems with Applications*, Vol.42, No.6, pp.3234-3241.
- [45] Roll, R., 1992. "Industrial structure and the comparative behavior of international stock market indices", *The Journal of Finance*, Vol.47, No.1, pp.3-41.
- [46] Saad, E. W., D. V. Prokhorov, and D. C. Wunsch, 1998. "Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks", *IEEE Transactions on neural networks*, Vol.9, No.6, pp.1456-1470.
- [47] Schumaker, R. P., and H. Chen, 2009. "Textual analysis of stock market prediction using breaking financial news: The AZFin text system", *ACM Transactions on Information Systems, (TOIS)* Vol.27, No.2, pp.12-23.
- [48] Smith, G., 2001. "The price of gold and stock price indices for the United States", *The World Gold Council*, Vol.8, pp.1-16.
- [49] Soenen, L. A., and E. S. Hennigar, 1988. "An analysis of exchange-rates and stock-prices-the united-states experience between 1980 and 1986", *Akron Business and Economic Review*, Vol.19, No.4, pp.7-16.
- [50] Taylor, M. P., and H. Allen, 1992. "The use of technical analysis in the foreign exchange market", *Journal of International Money and Finance*, Vol.11, No.3, 304~314.
- [51] Toutanova, K., Klein, D., Manning, C. D., and Singer, Y., 2003. "Feature-rich part-of-speech tagging with a cyclic dependency network", *In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology* Vol.1, Association for Computational Linguistics, pp.173-180.

- [52] Tsai, C. F., and Y. C. Hsiao, 2010. "Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches", *Decision Support Systems*, Vol.50, No.1, pp.258-269.
- [53] Vargas, M. R., de Lima, B. S., and Evsukoff, A. G. 2017. "Deep learning for stock market prediction from financial news articles", In *Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 2017 IEEE International Conference on (pp. 60-65). IEEE.
- [54] Weron, R., 2002. "Estimating long-range dependence: finite sample properties and confidence intervals", *Physica A: Statistical Mechanics and its Applications*, Vol.312, No.1, pp.285-299.
- [55] Zhang, D., H. Xu, Z. Su, and Y. Xu, 2015. "Chinese comments sentiment classification based on word2vec and SVM perf", *Expert Systems with Applications*, Vol.42, No.4, pp.1857-863.
- [56] Zhao, X., Yang, J., Zhao, L., and Li, Q., 2011. "The impact of news on stock market: Quantifying the content of internet-based financial news", In *Proc. of the 11th Intl DSI & 16th APDSI Joint meeting*, pp. 12-16.

저자소개

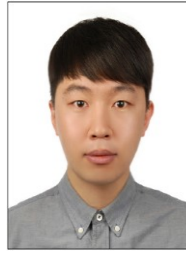


성노윤
(Seong, Nohyoon)

KAIST에서 물리학 학사 학위를 취득하였다. 현재 KAIST 경영대학원 경영공학부 MIS 석사 과정에 재학 중이다. 주요 관심분야는 자연어 처리, 머신러닝, 딥러닝, 빅데이터 분석, 계량 경제학, 경제물리학 등이다. 기존 경제학 이론에 기반하여 머신러닝과 딥러닝을 하여 경영학적 문제와 경제학적 문제를 해결하는 데에 관심을 가지고 있다.

E-mail: nyseong@business.kaist.ac.kr

Tel: +82-10-4542-4525



남기환
(Nam, Kihwan)

연세대학교에서 통계학 학사, KAIST 경영대학원 경영공학부 MIS 박사학위를 취득하였다. 현재 한양대학교 경영대학 겸임교수, UNIST 경영대학 대우교수로 재직 중이다. 주요 관심 분야는 비즈니스 애널리틱스, 빅데이터 분석, 데이터 마이닝, 통계분석, 추천 시스템, 계량경제모델, 기계학습, 딥러닝, 소셜 네트워크 분석 등이다. IT기업들의 경영 전반에 걸친 문제들을 데이터 기반으로 접근함으로써 보다 효과적인 마케팅 및 경영전략을 수립하는데 관심을 가지고 있다.

E-mail: namkh@kaist.ac.kr

Tel: +82-10-4930-8317

◇ 이 논문은 2017년 8월 21일에 접수하여 2017년 11월 2일에 1차 수정, 2017년 11월 17일에 2차 수정을 거쳐 2017년 11월 17일에 게재확정되었습니다.