산업융합연구 제21권 제1호 pp. 167-173, 2023
ISSN 2635-8875 / e-ISSN 2672-0124
DOI : https://doi.org/10.22678/JIC.2023.21.1.167

LDA와 LSTM를 응용한 뉴스 기사 기반 선물가격 예측

주진현[1], 박근덕[2*]
[1]호서대학교 AI융합학부 조교수, [2]호서대학교 컴퓨터공학부 교수

Futures Price Prediction based on News Articles using LDA and LSTM

Jin-Hyeon Joo[1], Keun-Deok Park[2*]
[1]Professor, College of AI Convergence, Hoseo University
[2]Professor, Dept. of Computer Engineering, Hoseo University

요 약 경제지표를 분석하는 방법으로 회귀 분석이나, 인공지능을 활용하여 미래의 데이터를 예측하는 연구가 발표되었다. 본 연구에서는 토픽모델링을 사용하여 과거 뉴스 기사로부터 얻은 주제 확률 데이터를 이용한 인공지능으로 미래 선물 가격을 예측하는 시스템을 구상하였다. 과거 뉴스 기사로부터 비지도학습을 통한 문서의 주제를 추출할 수 있는 LDA 방법으로 각 뉴스 기사 주제 확률 분포 데이터를 얻을 수 있고, 해당 데이터를 인공지능의 RNN의 파생 구조인 LSTM의 입력 데이터로 활용함으로써 미래 선물 가격을 예측하였다. 본 연구에서 제안한 방법에서는 선물 가격의 추세를 예측할 수 있었고, 이를 활용하여 추후 옵션 상품 등의 파생 상품에 대한 가격 추세도 예측할 수 있을 것으로 보인다. 다만, 일부 데이터에 대해 오차가 발생하는 것이 확인되어 정확도 향상을 위한 추가적인 연구가 필요하다.

키워드 : 토픽모델링, LDA, LSTM, 인공지능, 데이터분석

Abstract As research has been published to predict future data using regression analysis or artificial intelligence as a method of analyzing economic indicators. In this study, we designed a system that predicts prospective futures prices using artificial intelligence that utilizes topic probability data obtained from past news articles using topic modeling. Topic probability distribution data for each news article were obtained using the Latent Dirichlet Allocation (LDA) method that can extract the topic of a document from past news articles via unsupervised learning. Further, the topic probability distribution data were used as the input for a Long Short-Term Memory (LSTM) network, a derivative of Recurrent Neural Networks (RNN) in artificial intelligence, in order to predict prospective futures prices. The method proposed in this study was able to predict the trend of futures prices. Later, this method will also be able to predict the trend of prices for derivative products like options. However, because statistical errors occurred for certain data; further research is required to improve accuracy.

Key Words : Topic modeling, LDA, LSTM, Artificial intelligence, Data analysis

1. 서론

선물(Futures) 가격은 특정 시점에 수량과 규격이 표준화된 상품이나, 외환 등의 다양한 금융 자산을 거래하기 위하여 현재 시점에서의 합의한 가격을 의미한다. 이러한 선물 거래는 현물 거래와 서로 밀접한 관계를 맺고 있으므로[1] 선물 가격을 예측하는 것은 미래 현물의 가격에 대한 예측으로 이어질 수 있기에 중요하다. 이러한 선물 가격은 경제 상황에 따라 민감하게 가격 변동이 이루어지게 되는데, 경제 상황을 알 방법 중 가장 정확한 방법은 GDP를 활용하는 것이다. 하지만 GDP는 해당 분기가 끝난 후 책정되기 때문에 GDP를 활용하여 그해의 경제 상황을 파악하는 것은 불가능하다.

따라서 많은 연구에서 GDP를 활용하여 선물 가격을 파악하지 않으며, 연구에 자주 사용되는 방법이 선물 가격이나 변동성 지수 등 파생상품에 대한 가격 변동을 이

*Corresponding Author : Keun-Deok Park(gdpark@hoseo.edu)
Received December 10, 2022 Revised December 19, 2022
Accepted January 20, 2023 Published January 28, 2023

용한다. 연구 중에서는 인공지능을 활용하여 과거 선물 가격 데이터나 과거 옵션 가격 데이터를 학습하여 미래 선물 가격이나 옵션 가격을 예측하거나, 변동성 지수를 학습해서 미래 옵션 가격을 예측하는 방법이 있다. 이러한 방법은 과거에 없는 데이터가 출현하더라도 비교적 정확한 데이터 예측이 가능하지만, 과거 데이터와 미래 데이터 간의 관계를 정의할 수 없어 예측 데이터로 사용하기에는 부적절하다.

다른 연구에서는 토픽 모델링을 사용하였다. 해당 연구에서는 경제 상황이 좋아지면 긍정적인 뉴스 기사가 주로 작성되지만, 경제 상황이 안 좋은 경우 부정적인 뉴스 기사 보도가 빈번하게 보도되는 특징에서 뉴스 기사가 경제에 민감하게 반응한다는 점을 활용하여 과거 뉴스 데이터의 주제를 학습하여 KOSPI 지수를 예측하기도 하였다[2]. 하지만 해당 연구에서는 회귀 분석을 통한 예측을 진행하였기 때문에 데이터가 없는 상황에서는 데이터 예측이 힘들다.

본 연구에서는 기존 연구에서 언급되었던 경제 상황에 따라 뉴스 기사에서 보도되는 주제가 민감하게 반응한다는 점을 활용하여, 과거 뉴스 기사 주제를 추출하여 신경망의 입력 데이터로 활용해서 미래 선물 가격을 예측해보았다.

2. 선행연구

2.1 토픽 모델링

모든 문서는 표현하고자 하는 주제를 가지고 있고, 이러한 주제와 관련된 단어들은 주제와 연관이 없는 단어들보다 많이 문서 내에 등장하게 된다. 따라서 주제와 관련된 단어들은 문서에 같이 등장하게 되고, 이렇게 같이 등장하는 단어들은 일반적으로 유사한 의미를 지니게 되므로 이를 주제라 할 수 있다. 이렇듯 토픽 모델링이란 문서를 구성하는 주제, 토픽을 파악하는 비지도학습 분류 방법론이다[3].

정보화 시대가 되면서 인터넷 문화가 크게 발전하였고, 이로 인하여 인터넷 상에는 무수하게 많은 데이터들이 존재하게 되었다. 이러한 데이터 중에서 목적을 가지고 원하는 데이터를 찾으려면 사람이 일일이 찾는 것은 불가능하다. 때문에 봇(Bot) 프로그램을 작성하여 인터넷 상의 데이터를 수집하는 웹 크롤러(Web Crawler)를 만들어 데이터를 수집한다.[4] 하지만 프로그램인 크롤러는 사람들이 사용하는 자연어에 대해 이해를 할 수 없다. 이를 해결하고자 크롤러가 수집한 문서에 대하여 토픽 모델링으로 웹 문서의 주제를 파악하여 사용자의 목적에 맞는 문서를 웹 상에서 수집할 수 있다.

Papadimitriou, Raghavan, Tamaki, Vempala(1998)는 문서와 단어 간 행렬(Matrix)을 이용하여 문서 내 잠재의미를 도출하는 알고리즘을 LSI(Latent Semantic Analysis)를 개발하였다. Blei, Ng, Jordan(2003)는 현재 많이 사용되고 있는 LDA(Latent Dirichlet Allocation)을 제시하였다.[6] 하지만 LSI와 LDA 모두 비지도학습(unsupervised learning) 기반으로 컴퓨터가 자동으로 분석을 수행하기 때문에 사용자가 의도한 바와 다른 결과가 발생하기도 한다. 이러한 문제를 해결하기 위해 Mcauliffe, Blei(2008)은 sLDA(supervised LDA)를 제안하였다.[7] Roberts, Stewart, Tingley, Lucas, Leder-Luis, Gadarian, Albertson, Rand(2014)은 LDA에서 확장되어 단어의 빈도수 뿐만아니라 문서 내 메타데이터(meta-data)를 활용하여 메타 데이터와 토픽들 간의 상관관계 추정 및 토픽 간 관계를 구분하여 해석할 수 있는 STM(Structural Toopic Model)을 제안하였다.[8] LSI와 LDA와 달리 시간의 흐름에 따라 토픽의 내용 변화 파악에 활용되는 DTM(Dynamic Topic Model)이 있다.[9]

2.1.1 LDA

LDA는 잠재 디리클레(Latent Dirichlet) 확률 기반의 비지도 학습을 통해 문서의 주제를 파악하여 문서를 분류할 수 있는 비지도 학습 방법론이다.

LDA 방법을 사용하기 위해서 우선 문서 내 단어들에 대한 빈도수를 설정해야 한다. 해당 빈도수에는 일반적으로 단어의 순서와는 상관없이 빈도수만을 표현하는 BoW(Bag of Words) 혹은 TF-IDF를 사용한다.

BoW로 전달된 문자 데이터들은 LDA에 전달되어 BoW 내 N개의 문자 중 주제를 선정하게 된다. 이때 LDA에는 선정할 주제 개수인 K, K개의 주제에 대한 α, N개의 단어에 대한 β를 정한다. 주어진 α에 대해서 디리클레 분포 $Dir(\alpha)$로부터 d번째 문서에 대한 주제 가중치 θ_d을 무작위로 추출하고 주어진 β에 대해서 $Dir(\beta)$ 분포로부터 k번째 주제에 대한 가중치 ϕ_k를 추출한다. 앞서 추출된 θ_d에 대하여 다항분포 $Mlti(\theta_d)$ 로부터 문서 d에서 n번째 단어 ω_{dn}과 주제 z_{dn}이 확률적으로 할당된다.

Fig. 1과 같은 과정을 여러 문서를 거쳐 진행하면서 확률이 안정적으로 변하면 이를 통해 특정 문서가 어떤 주제를 담고 있는지 확률적으로 추정할 수 있게 된다.

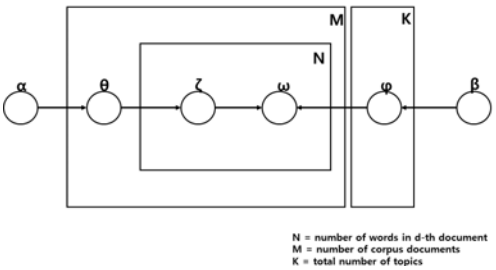

Fig. 1. LDA topic selection process

2.1.2 DTM

DTM(Dynamic Topic Model)은 주로 시계열에 의한 문서 집합에서 주제 트랜드(Trend) 변화를 분석하는데 사용한다.[10] LDA 방법에서는 BoW 혹은 TF-IDF를 사용하여 문서 내 단어들을 제공하기 때문에 단어가 문서에 나타나는 순서와 문서가 코퍼스(corpus)에 나타나는 순서를 무시하기 때문에 다양한 주제가 서로에게 영향을 주어 변화하는 과정을 나타내지 못한다.[11] 하지만 DTM의 경우 시간별로 그룹화하여 주제가 고정된 시간 간격으로 진화할 수 있도록 한다. 따라서 DTM 방법에서는 주제가 시간 흐름에 따라 변화하고, 시계열 데이터를 분석하는데 적합한 방법이다.

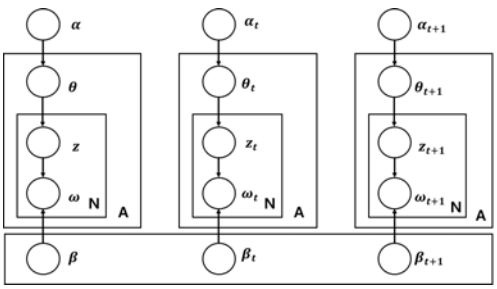

Fig. 2. DTM Topic selection process

2.2 신경망(Neural Network)

인공지능(AI, Artificial Intelligence)는 기계에 인간이 지니고 있는 지적 능력의 일부 또는 전체를 인공적으로 구현한 것을 의미한다.[12] 인공지능이라는 단어는 17세기부터 언급되어 왔지만 학습 데이터셋의 부족과 연산처리장치 성능의 한계로 구글 딥마인드(Google DeepMind) 사에서 개발한 알파고(Alphago) 이전에는 연구에 큰 진보가 없었다.[13] 하지만 이후 GPU를 비롯한 연산장치 성능의 향상으로 인공지능 기술의 발전은 빠른 속도로 진행되며, 급격한 성장을 보여주고 있다.

이러한 인공지능의 기초가 되는 방법이 인공신경망(ANN, Artificial Neural Network)이다. 인공신경망은 가중치(weight)를 갖는 망(network) 형태로써, 입력된 데이터를 각각의 퍼셉트론(Perceptron)에 입력하고, 퍼셉트론들은 임계값(threshold)를 넘은 경우 활성화되어 데이터를 출력하게 된다. 이러한 구조가 여러 개의 층(layer)로 이루어져 있고, 입력층(input layer)에서 출력층(output layer)까지 순차적으로 각각의 층으로 데이터가 전파된다.

인공지능이 학습하는 방법으로 대표적인 방법이 오차역전파(Back propagation) 알고리즘이다. 해당 알고리즘은 입력부터 출력까지 순방향으로 이루어지지만, 가중치 학습할 때는 역방향으로 이루어지기 때문이다. 이때 출력에 발생하는 오차를 가중치로 미분하여 오차를 감소하는 방식으로 가중치를 수정한다.

퍼셉트론의 활성화를 위한 함수로는 계단 함수, 항등함수, 시그모이드(sigmoid) 함수 등이 있다. 이러한 활성화 함수는 입력된 데이터의 합에 따라 활성화 여부를 결정하게 된다. 활성화 함수 중 ReLU (Rectified Linear Unit)은 0보다 작은 입력 값에는 0을 출력하고, 0보다 크면 입력된 값을 그대로 출력하는 함수이다.

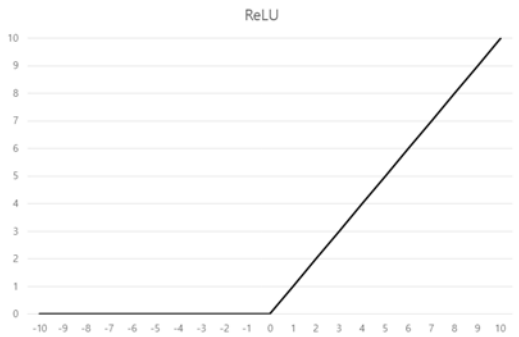

Fig. 3. ReLU activation function

인공지능은 영상인식(Recognition), 영상분류(Classfication), 계층적 분류(Hierarchical Classifier), 객체 인식(Object Detection), 세메틱 분할(Semantic segmentation)과 같은 영상 처리 분야에서 큰 활약을 하고 있다. 또한 자율주행 자동차, 스마트 팩토리 등과 같이 다양한 분야에서 인공지능을 사용하고 있다.

2.2.1 LSTM

LSTM(Long Short-Term Memory)은 RNN (Recurrent

Neural Network)의 한 구조로, RNN의 기본 셀(cell)에 망각, 입력, 출력 3개의 게이트가 하나의 셀을 이루는 구조를 갖는다.[14] 이러한 구조는 가중치뿐만 아니라 메모리에 대한 추가 정보를 셀 상태에 저장하여 시계열 데이터의 장기 기억 문제를 해결하기 위한 구조이다.[15]

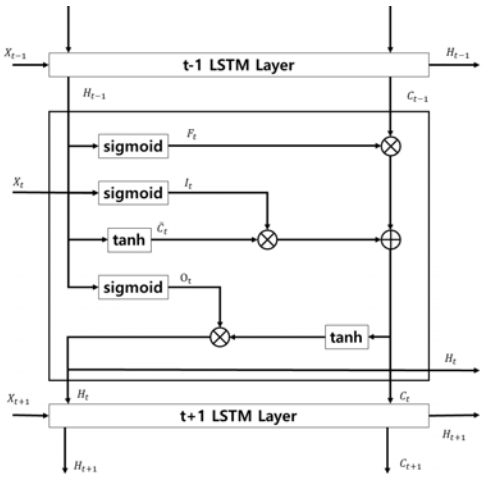

Fig. 4. LSTM learning process

LSTM 내 셀에 있는 망각 게이트는 시그모이드 함수를 통해 0에서 1 사이의 값으로 이전 셀 상태(C_{t-1})의 정보 망각율을 결정한다. 입력 게이트(i_t)는 시그모이드 함수로 활성화하여 업데이트 여부를 결정하고 후보 값($\widetilde{C}_t$)을 만들어 이전 셀 상태를 업데이트한다. 출력 게이트(o_t)는 생성된 시그모이드 레이어를 통해 현재 셀 상태에서의 출력값을 산출한다.

3. 연구방법

본 연구에서는 뉴스 기사를 웹 크롤링을 통해 수집한 후 LDA 방법을 통해 해당 뉴스 기사가 어떠한 주제를 표현하는지 분류과정을 거쳐 해당 날짜에 보도된 뉴스 기사의 주제들이 선물 가격에 어떠한 영향을 미치는지 알아보기 위하여 LSTM으로 주제 빈도수 입력 데이터로, 선물 종가(Last price)를 출력 데이터로 설정하여 학습을 진행하여 예측 결과를 출력하였다.

3.1 한국어 전처리 과정

뉴스 기사를 LDA 방법으로 주제를 추정하기 전에 먼저 뉴스 기사에서 불용어 제거 및 형태소 분석을 진행해야 한다. 해당 과정이 처리되지 않으면 필요 없는 데이터가 많이 생겨 LDA 모델 학습 시 학습이 정상적으로 이루어지지 않는다.

본 연구에서는 불용어에 대한 처리는 트위터에서 만든 Okt(Open Korean Text) 형태소 처리기를 사용하여 조사나 접속사 같은 명사가 아닌 것들만 제거하였다.

3.2 뉴스 기사 데이터 처리

본 연구에서 사용된 뉴스 기사의 경우 한국언론진흥재단의 빅카인즈(BigKinds)에서 2000년 1월 4일부터 2019년 12월 30일까지 총 46,961개의 뉴스 기사를 사용하였다. 뉴스 기사는 경기변동 관점에서 경기중립적인 '경기전망', '경제전망', '경제동향', '경기변동', '경기분석' 및 '경기추이'를 19개의 중앙지 및 경제지에서 추출하였다.

해당 뉴스 기사들은 LDA 방법을 사용하여 어느 주제에 해당하는지 확률 분포를 구한다. 해당 확률 분포는 추후 LSTM에 입력 데이터로서 사용된다.

Table 1. Topic distribution table with the highest probability by news article

	counts	proportion
1	1,333	2.8%
2	3,966	8.4%
3	1,090	2.3%
4	6,886	14.7%
5	4,361	9.3%
6	6,184	13.2%
7	184	0.4%
8	1,730	3.7%
9	2,757	5.9%
10	8,075	17.2%
11	619	1.3%
12	2,687	5.7%
13	3,156	6.7%
14	1,164	2.5%
15	2,770	5.9%
total	46,962	100%

Table 1은 뉴스 기사별 가장 높은 주제에 대한 분포 확률을 나타내고 있다. 가장 빈번하게 나타나는 10번 주제는 '한국은행', '총재', '소비자', '물가', '이주열', '금융통화위원회', '심리', '공개시장', '금통위', '동향' 단어를 포함하며 국내 물가에 따른 소비자 심리에 대한 주제를 나타낸다. 가장 적게 나타나는 7번 주제의 경우 '기획재정부', '정부', '장관', '경제', '회의', '대통령', '국회', '부총리', '정책', '경제정책' 단어를 포함하며, 국내 정부 정책에 대한 주제임을 추측할 수 있다. 7번 주제가 가장 적게

Table 2. Correlation coefficient by topic

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.0	−0.1	−0.1	−0.1	0.1	−0.1	0.0	0.0	0.0	−0.1	0.0	−0.1	−0.1	0.2	−0.1
2	−0.1	0.0	−0.1	−0.3	−0.2	0.0	−0.1	0.0	−0.1	−0.1	−0.1	−0.1	−0.1	0.0	−0.2
3	−0.1	−0.1	0.0	0.0	−0.1	−0.1	0.0	0.1	0.0	−0.1	0.0	0.1	0.2	−0.1	−0.1
4	−0.1	−0.3	0.0	0.0	−0.1	−0.1	0.0	0.1	−0.1	−0.2	0.0	−0.1	0.1	−0.1	0.2
5	0.1	−0.2	−0.1	−0.1	0.0	−0.1	0.0	−0.1	0.1	−0.1	0.0	−0.1	−0.2	−0.1	−0.1
6	−0.1	0.0	−0.1	−0.1	−0.1	0.0	0.0	−0.1	−0.1	0.0	0.0	−0.2	−0.2	0.0	−0.2
7	0.0	−0.1	0.0	0.0	0.0	0.0	0.0	−0.1	−0.1	0.2	0.0	−0.1	−0.1	0.0	0.0
8	0.0	0.0	0.1	0.1	−0.1	−0.1	−0.1	0.0	−0.1	−0.2	−0.1	0.0	0.1	−0.1	0.0
9	0.0	−0.1	0.0	−0.1	0.1	−0.1	−0.1	−0.1	0.0	−0.2	0.1	−0.1	−0.1	−0.1	−0.1
10	−0.1	−0.1	−0.1	−0.2	−0.1	0.0	0.2	−0.2	−0.2	0.0	0.0	−0.1	−0.2	−0.1	−0.1
11	0.0	−0.1	0.0	0.0	0.0	0.0	0.0	−0.1	0.1	0.0	0.0	−0.1	−0.1	0.0	−0.1
12	−0.1	−0.1	0.1	−0.1	−0.1	−0.2	−0.1	0.0	−0.1	−0.1	−0.1	0.0	0.2	−0.1	0.1
13	−0.1	−0.1	0.2	0.1	−0.2	−0.2	−0.1	0.1	−0.1	−0.2	−0.1	0.2	0.0	−0.2	0.0
14	0.2	0.0	−0.1	−0.1	−0.1	0.0	0.0	−0.1	−0.1	−0.1	0.0	−0.1	−0.2	0.0	0.0
15	−0.1	−0.2	−0.1	0.2	−0.1	−0.2	0.0	0.0	−0.1	−0.1	−0.1	0.1	0.0	0.0	0.0

출현하는 이유는 뉴스 기사를 선정할 때 경제에 관련된 뉴스 기사를 선정하였기 때문이다.

주제별 상관계수는 Table 2와 같다. 1번 주제의 경우 '기업', '경기', '지수' 등 기업과 관련된 주제이고, 14번 주제는 '연구기관', '국책'과 같이 국책 사업과 관련된 주제로 1번 주제와 14번 주제는 0.24의 상관계수를 가지며 관련성이 높은 것으로 나타났다. 2번 주제는 국내 분기별 성장 전망에 관련된 주제이고, 4번은 해외 투자 동향에 대한 주제로 −0.24라는 낮은 상관계수를 나타낸다.

3.3. 선물 가격

선물 가격은 한국거래소(KRX) 사이트 내 정보데이터 시스템에서 추출하였다. 해당 사이트에서 데이터를 다운로드하기 위해서는 일일이 날짜를 입력해야 하므로 2000년 1월 4일부터 2019년 12월 30일까지의 날짜별 선물 가격을 가져오기 위하여 웹 크롤러 제작을 통해 데이터를 가져왔다.

해당 데이터는 LDA를 통해 추출된 뉴스 기사의 주제 확률 분포와 함께 LSTM의 출력 데이터로 사용되었다.

4. 연구결과

본 연구에서는 LSTM의 입력에는 15개의 주제에 대한 확률 분포가 설정되고, 출력에는 선물 종가가 설정되었다. 총 4,875개의 데이터 중 4,674개의 데이터가 훈련 데이터로 사용되었고, 남은 200개의 데이터는 검증 데이터로 사용되었다.

LSTM 학습 시 window size 파라미터를 설정할 수 있다. 해당 파라미터는 LSTM이 출력 값을 결정하기 위한

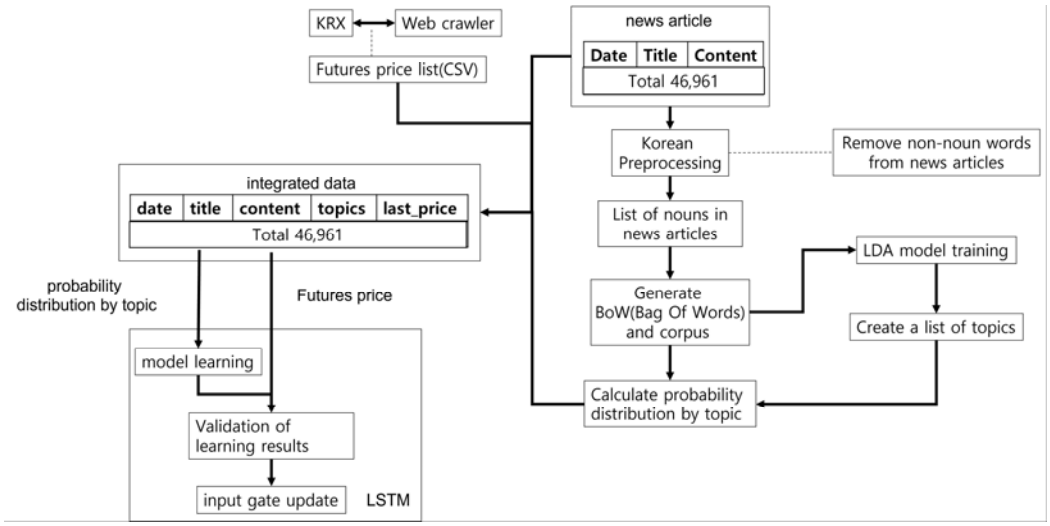

Fig. 5. Experimental process structure

입력 데이터 수를 의미한다. 따라서 해당 파라미터는 뉴스 기사가 보도된 후 선물 가격에 영향을 주기 전까지의 시차로써 사용된다. 뉴스 기사에 따라 선물 가격이 영향을 즉시 받는 것이 아닌 약간의 시차가 존재할 것이기 때문에 시차로 사용할 수 있는 window size 파라미터의 값을 10부터 40까지 10단위로 증가하며 실험을 진행하였다.

window size에 따른 예측 데이터와 검증 데이터 간의 차이값 중에서 평균 차이가 가장 적은 window size는 20으로 약 9 정도의 차이였고, 표준편차 값 역시 두 번째로 적었다.

Table 3. Difference between prediction data and verification data according to window size

wsize	max	min	avg	stddev
10	37.88	-160.85	-20.3	26.73
20	23.38	-83.91	-9.76	18.06
30	29.06	-75.61	-16.1	18.26
40	17	-72.96	-11.96	17.55

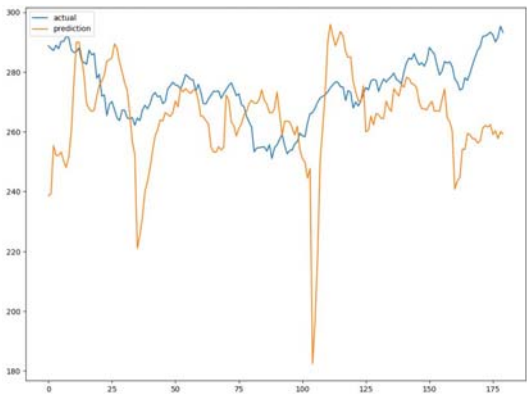

Fig. 6. Graph of prediction data and validation data in window size 20

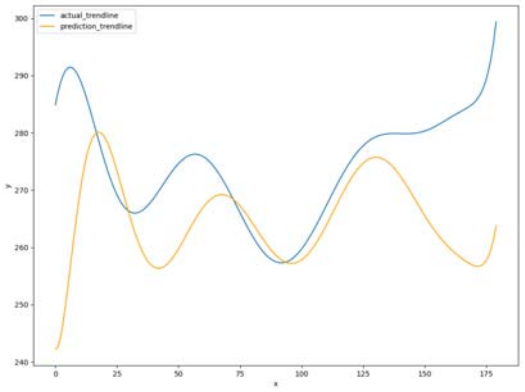

Fig. 7. Trendline graph with a window size of 20

Fig. 7 추세선 그래프에 나타나는 것처럼 예측 데이터 값과 검증 데이터 값의 추세가 크기에 차이가 있지만, 증가 및 감소에 대한 추세는 따라가는 것을 확인할 수 있다.

5. 결론

본 연구에서는 토픽 모델링 방법 중 LDA와 인공 신경망 중 LSTM을 활용하여 뉴스 기사로부터 얻은 주제 확률 분포를 사용하여 선물 가격을 예측하는 시스템은 제안하였다. 한국어 전처리, 주제 상관관계 분석, 그리고 LSTM을 통한 과거 뉴스 기사 주제 확률로부터 미래 선물 가격을 예측하는 순으로 연구를 진행하였으며, 제안된 방법으로 양의 차이가 있지만 뉴스 기사 주제 확률을 통해 선물 가격의 추세를 예측할 수 있었다.

선물 가격은 현물 가격과 밀접한 관계를 갖기 때문에 선물 가격을 예측하여 미래의 경기 지표를 추정할 수 있다. 이를 통해 실업률, 소비자 물가지수 등의 다양한 부분에 대한 대책을 미리 준비할 수 있다.

다만, 본 연구에서 제안한 방법은 예측 데이터 추세와 검증 데이터 추세 간에 차이값이 존재하는 것을 확인하였다. 추후 연구에서는 이러한 차이값을 줄이기 위해 시간 흐름에 따른 주제 추세를 나타낼 수 있는 DTM 방법을 사용하도록 한다.

REFERENCES

[1] Yang Cheol Won, Lu Bing. (2019). Do Futures Prices Lead Spot Prices? : Evidence from the Chinese Market. The Research Institute of Future Industry Dankook University, 43(1), 55-71.

[2] Ko, K., Oh, S., & Baek, J. (2020). Development of economic fluctuation topic indices and topic indices regression model for KOSPI200 index. Journal of the Korean Data And Information Science Society, 31(4), 579-594. DOI : 10.7465/jkdi.2020.31.4.579

[3] Kim, I., Lee, A., Kim, J., & Choi, J. (2022). Analysis of Owner's Detached House Housing Needs Sentences Using LDA Topic Modeling. Korean Journal of Computational Design and Engineering, 27(4), 435-445. DOI : 10.7315/cde.2022.435

[4] Han, D.-H., & Lee, Y.-K. (2021). Design of Action-Based Web Crawler Structural Configuration for Multi-Website Management. KIISE Transactions on Computing Practices, 27(2), 98-103. DOI : 10.5626/ktcp.2021.27.2.98

[5] Choi, S. C., & Park, H. W. (2020). A Study on the Trend of Topic Modeling in South Korea using KCI Journal Publications. The Korean Data Analysis Society, 22(2), 815-826.
DOI : 10.37727/jkdas.2020.22.2.815

[6] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, null (3/1/2003), 993-1022.

[7] Blei, D.M., & McAuliffe, J.D. (2007). Supervised Topic Models. NIPS.
DOI : 10.48550/arXiv.1003.0783

[8] Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B. and Rand, D.G. (2014), Structural Topic Models for Open-Ended Survey Responses. American Journal of Political Science, 58: 1064-1082.
DOI : 10.1111/ajps.12103

[9] David M. Blei and John D. Lafferty. 2006. Dynamic topic models. In Proceedings of the 23rd international conference on Machine learning (ICML '06). Association for Computing Machinery, New York, NY, USA, 113-120.
DOI : 10.1145/1143844.1143859

[10] C. H. Kang (2021). A study on the trend of COVID-19 perception through dynamic topic modeling and semantic network analysis using tweeter text data, master dissertation, Sungkyunkwan University, Seoul

[11] Wang, X., & McCallum, A. (2006). Topics over time. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '06.
DOI : 10.1145/1150402.1150450

[12] S. W. Lee, C. K. Ahn. (2017). Introduction and Development trend of Artificial Neural Networks. The Korean Institute of Electrical Engineers, 66(8), 36-41.

[13] Kihwan Choi. (2018). Recent Applications in Convolutional Neural Networks. Communications of the Korean Institute of Information Scientists and Engineers, 36(2), 25-31.

[14] Kim, S.-J., & Choi, B.-J. (2022). LSTM Model based Prediction of Daily confirmed cases of COVID-19 in Korea using Google Mobility Data. Journal of Korean Institute of Intelligent Systems, 32(4), 292-298. DOI : 10.5391/jkiis.2022.32.4.292

[15] Kim, S. W. (2022). Long Short-Term Memory-based Prediction Performance of COVID-19 Fear Index on Asset Prices: Stocks vs Cryptocurrencies. Asia-Pacific Journal of Convergent Research Interchange, 8(8), 45?58.
DOI : 10.47116/apjcri.2022.08.05

주 진 현(Jin-Hyeon Joo) [정회원]

- 2011년 2월 : 호서대학교 컴퓨터공학(학사)
- 2013년 2월 : 호서대학교 컴퓨터공학전공(석사)
- 2020년 3월~현재 : 호서대학교 AI융합학부 조교수

- 관심분야 : 인공지능, 데이터마이닝, 소프트웨어공학
- E-Mail : joojin4381@hoseo.edu

박 근 덕(Keun-Deok Park) [정회원]

- 2005년 8월 : 서울대학교 컴퓨터공학부(박사)
- 2006년 3월~현재 : 호서대학교 컴퓨터공학과 교수

- 관심분야 : 소프트웨어공학, 데이터마이닝
- E-Mail : gdpark@hoseo.edu