



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위 청구논문

2018학년도

순환신경망을 활용한 호가 데이터  
기반의 주식시장 예측에 대한 연구

A Study on the stock market prediction based on  
bid and ask information using recurrent neural  
network

광운대학교 대학원

경영학과

조 윤 재

# 순환신경망을 활용한 호가 데이터 기반의 주식시장 예측에 대한 연구

A Study on the stock market prediction based on  
bid and ask information using recurrent neural  
network

지도교수 이 석 준

이 논문을 경영학 석사학위논문으로 제출함.

2019년 6월 일

광운대학교 대학원

경영학과

조 윤 재

조운재의 경영학 석사학위논문을 인준함

심사위원장 정 석 재                      ①인

---

심 사 위 원 한 재 현                      ①인

---

심 사 위 원 이 석 준                      ①인

---

광운대학교 대학원

2019년    6월    일

## 감사의 글

BIRC에 들어오고 벌써 3년하고도 반이 지나갔습니다. 많은 분들의 조언과 격려가 있었기에 배우고 또 즐길 수 있었습니다. 이 자리를 빌어, 그동안 응원해주신 모든 분들께 감사의 말씀을 전하고자 합니다.

먼저, 늘 유쾌하고 섬세하게 지도해주신 이석준 교수님께 감사의 말씀을 드립니다. 교수님께 배운 학문적 지식과 삶의 지혜를 통해 한층 더 발전할 수 있었습니다. 그리고 본 논문을 여러 차례에 걸쳐 다듬어 주시고, 연구 방향에 대해 깊게 고민하고 지도해 주신 정석재 교수님과 한재현 교수님께 감사의 말씀을 전합니다. 두 분 교수님의 조언을 통해 제 논문이 논문답게 완성될 수 있었습니다. 또한 경영학과의 다른 교수님들께도 진심으로 감사드립니다. BIRC 동료들에게도 감사의 말씀 전해봅니다. 우선 대한민국은 학연, 지연, 흡연이라고 이쯤이면 친구라고 할 수 있는 재웅이형, 2년째 집들이를 못하고 있는 재현이형, 아끼고 의지하는 만큼 못되게 굴어서 미안한 현정이, 반말해도 되는 형 같은 든든한 종수, 내가 제일 좋아하는 원거리 딜러 성식이, 마지막으로 늘 밝은 에너지 전해주는 미영이, 잠시나마 형들에겐 너무 버릇없었던 건 아닌지, 친구 그리고 동생들에겐 너무 함부로 대했던 것은 아닌지 반성해봅니다. 지치는 순간마다 여러분이 곁에 있어서 무너지지 않을 수 있었습니다. 친구들과 가족들에게도 감사의 말씀 전합니다. 항상 너무 감사한 마음이지만, 주시는 격려와 응원에 보답하지 못하고 바쁘다는 핑계로 소홀했던 점에 대해 반성하며 앞으로는 그러지 않겠노라 다짐해봅니다. 마지막으로 가장 큰 버팀목이 되어주던 친구에게도 정말 고마웠다고 감사의 말씀 전합니다.

2019년 6월 조윤재 올림

## 국문 요약

### 호가 데이터를 활용한 주식시장 예측에 대한 연구

주식시장은 다양한 변수들의 영향을 받으며 예측하기 어려운 난제로 알려져 있지만, 과거부터 이를 예측하기 위해 다양한 연구들이 수행되어 왔으며, 선행연구들을 통해 주식시장 예측에 적합한 다양한 방법론들과 데이터들이 제시되었다. 하지만 주식시장에서 발생하는 변수에는 선행연구들에서 제시된 데이터 이외에도 다양한 변수들이 존재하며, 호가 정보 또한 그 중 하나이다. 호가 정보는 시장참여자들의 매매행태 및 주문이 실시간으로 반영되는 정보이며 매도 및 매수에 대한 상황을 제공함으로써 주가 변동의 방향성에 대한 예측을 가능하게 하는 중요한 투자지표 중 하나이다. 하지만 호가 정보는 주가에 대한 영향이 지연되어 나타날 수 있고 유동성 등에 의해 그 영향력이 종목 별로 상이하게 나타나며, 시간대 및 단계 별로 가격 발견에 대한 공헌도가 모두 다르다. 따라서 호가 정보를 활용하는 일부 투자자들이 객관적인 의사결정을 하기 어려운 실정이다. 특히 2002년 이후 국내 증권시장의 호가 공개 단계가 확장됨으로써 호가 정보에 대한 중요성이 부각되었고 재무 분야에서는 이와 관련된 다양한 연구들이 수행되었으며 주식시장에 미치는 영향력 등을 증명하였지만, 아직 주식시장 예측 분야에서는 이를 활용한 연구가 전무하다. 따라서 본 연구에서는 딥러닝 기법 중 하나인 순환신경망과 호가 정보를 활용하여 주식시장을 예측하는 연구를 수행하였다. 데이터 수집 및 정제, 라벨링 및 기계학습용 데이터셋 생성, Recurrent Neural Network(RNN), Long Short-term memory(LSTM), Gated Recurrent Unit(GRU)를 활용한 주식시장 예측모형 학습 및 검증이 수행되었다. 실

험 결과에서 종목 별 유동성에 따라 정확도를 비교했을 때, 유동성 하위 구간에서는 RNN이 57.75%로 정확도가 가장 높게 나타났으며, 상위 구간에서는 GRU가 60.74%로 정확도가 가장 높게 나타났다. 해당 결과를 통해, 유동성이 작은 종목에서는 호가 정보의 변동이 적기 때문에 시계열 전체에 걸쳐 맥락을 학습하는 RNN이 가장 효율적이고, 유동성이 큰 종목에서는 호가 정보의 변동이 많기 때문에 시계열 전체에 대해서 선택적으로 장기 기억과 단기 기억을 학습하는 방식이 효율적인 것으로 사료된다. 본 연구에서 제시된 주식시장 예측모형은 호가 정보를 활용하는 투자자들의 객관적 의사결정에 기여할 수 있을 것으로 기대된다.

## **ABSTRACT**

### **A Study on the stock market prediction using bid and ask information**

The stock market is hard to predict because it is affected by many variables. Nevertheless, many studies have been conducted to predict the market from the past and have presented many methodologies and variables. However, in stock market, there are many variables besides the presented in previous researches and bid and ask information is also one of them. Bid and ask information reflects the orders of market participants in real time. Thus, it is one of the important investment indicators that enables forecasting of the direction of stock price fluctuation. However, there is a possibility that the influence of the bid and ask information on the stock price is delayed, and contribution to price discovery is different by time, price step, liquidity, and etc. Therefore, it is difficult for some investors who use of bid and ask information to make objective decisions. Especially, the importance of bid and ask information has been emphasized due to the expansion of public offerings in the korean stock market since 2002. Many studies, which is related to bid and ask information, have been carried out in the financial research, and have proved their impact on the stock market. But, there is no research utilizing bid and ask information in the stock market prediction field.



Therefore, stock market prediction was conducted using bid and ask information through recurrent neural network, which is one of the deep learning techniques. For this research, data gathering and refining, labeling and dataset generation, and training and testing stock market prediction model using Recurrent Neural Network(RNN), Long Short-term memory(LSTM), Gated Recurrent Unit(GRU) were preformed. In the experiment results, when comparing accuracy according to liquidity of each stock, accuracy of RNN was the highest at 57.75% in the liquidity lower section, and accuracy of GRU was the highest at 60.74% in the liquidity upper section.

The result shows that RNN, which is learning contexts over full time series, is the most efficient method for lower liquidity stocks because of less fluctuation of bid and ask volume. In the case of stocks with high liquidity GRU, which selectively learns long-term memory and short-term memory for time series, is considered to be efficient because there is a large fluctuation in stock price information. The stock market forecasting model proposed in this study is expected to contribute to the objective decision making of investors using bid and ask information.

## 차 례

국문요약 .....	i
영문요약 .....	iii
차 례 .....	v
표 목 차 .....	vii
그림목차 .....	viii
<b>I. 서론 .....</b>	<b>1</b>
1. 연구의 배경 및 목적 .....	1
2. 연구의 구성 .....	3
<b>II. 문헌 고찰 .....</b>	<b>4</b>
1. 호가 정보에 대한 고찰 .....	4
1.1. 호가 정보의 개요 .....	4
1.2. 호가 정보에 대한 선행연구 .....	5
2. 주식시장 예측에 대한 고찰 .....	6
2.1. 주식시장 예측 방법 .....	6
2.2. 주식시장 예측에 대한 선행연구 .....	8
3. 순환신경망에 대한 고찰 .....	10
3.1. RNN의 개요 .....	10
3.2. LSTM의 개요 .....	15
3.3. GRU의 개요 .....	18
3.4. 순환신경망에 대한 선행연구 .....	22
4. 선행연구의 한계점 .....	23

<b>III. 연구 모형</b> .....	<b>25</b>
1. 데이터 라벨링 및 기계학습용 데이터셋 생성 .....	25
2. 모형 학습 및 검증 .....	26
<b>IV. 실증분석 및 결과</b> .....	<b>28</b>
1. 데이터 기간 및 수집 .....	28
2. 호가 데이터 정제 .....	29
3. 호가 데이터 라벨링 및 기계학습용 데이터셋 생성 .....	30
4. 주식시장 예측모형 학습 및 검증 .....	33
<b>V. 결론</b> .....	<b>41</b>
<b>참고문헌</b> .....	<b>44</b>

## 표 목차

<표 1> 코스닥 시가총액 상위 50종목 .....	28
<표 2> 수집된 주식시장 데이터 예시 .....	29
<표 3> 데이터 분할 정보 .....	33
<표 4> RNN을 활용한 주식시장 예측 검증 정확도(%) .....	34
<표 5> LSTM을 활용한 주식시장 예측 검증 정확도(%) .....	35
<표 6> GRU를 활용한 주식시장 예측 검증 정확도(%) .....	36
<표 7> 종목 별 거래량 정보 .....	38

## 그림 목차

<그림 1> RNN의 구조 .....	11
<그림 2> 펼쳐진(un-rolled) RNN 구조 .....	12
<그림 3> RNN 은닉층의 내부 구조 .....	13
<그림 4> BPTT의 흐름도 .....	14
<그림 5> LSTM의 구조 .....	15
<그림 6> LSTM의 내부 구조 .....	16
<그림 7> GRU의 구조 .....	19
<그림 8> GRU의 내부 구조 .....	20
<그림 9> 연구 모형 .....	25
<그림 10> RNN, LSTM, GRU를 활용한 예측모형 학습 예시 .....	27
<그림 11> 시간 중복 데이터 처리 예시 .....	30
<그림 12> 데이터 정제 예시 .....	30
<그림 13> 무빙 윈도우 기법을 활용한 데이터 라벨링 예시 .....	31
<그림 14> 최소-최대 정규화 과정을 거친 기계학습용 데이터셋 .....	32
<그림 15> 유동성 별 모형 정확도 비교 결과 .....	39

# I. 서론

## 1. 연구의 배경 및 목적

과거부터 주식시장을 예측하기 위해 통계학적 분석 방법, 인공지능 분석 방법 등 다양한 분석 방법들이 활용되어 왔다. 또한, 최근 컴퓨터 기술이 발달함에 따라 딥러닝(Deep Learning) 기법들이 각종 분야에서 뛰어난 성능을 보이며 주식시장에서도 주목을 받고 있다. 특히 딥러닝 기법 중 하나인 순환신경망은 데이터의 시계열 맥락 정보를 학습할 수 있다는 장점 때문에 분석 방법으로서의 활용이 증가되고 있다. 주식시장의 경우 순환신경망을 이용한 다양한 주가예측 연구가 수행되었으며(Lin et al., 2009; Nelson et al., 2017; Roondiwala et al., 2017; Zhuge et al., 2017), 시계열 속성을 띄는 주가의 예측에 있어서 순환신경망이 적합함이 증명되었다.

주식시장의 예측은 경제 및 정치적 변수를 포함한 다양한 상황에 영향을 받고 잡음이 포함되어 있는 경우가 많기 때문에 매우 어려운 문제로 여겨지고 있다(이형용, 2008; 송현정 & 이석준, 2018). 하지만 선행연구들을 통해 주가의 움직임은 무작위적이지 않고, 비선형성을 띄지만 유동적인 방식에 의해 결정된다는 결과들이 제시되고 있으며(McNelis, 2005), 다양한 주식시장 내·외부의 변수들이 주식시장 예측에 유용함이 증명되었다. 주식시장 예측에 활용된 대표적인 변수로는 기업의 내재가치를 반영하는 기본적 지표, 과거 주식의 가격 및 거래량 등을 이용해 산출된 기술적 지표, 주식 관련 SNS 텍스트 데이터 등이 있다.

한편 주식시장에서 발생하는 정보 중에는 ‘호가 정보’가 있다. 한국거래소 유가증권시장 업무규정에 의하면 호가 정보란 시장참여자가 시장에서 매매거래를 위해 제출하는 매도 또는 매수의 의사표시를 의미한다.

호가 정보는 시장참여자들의 매도 및 매수, 그리고 매매행태가 모두 반영된다는 점을 고려할 때, 시장 방향성에 대한 예상을 가능하게 하는 중요한 투자판단의 지표 중 하나이며(엄세용, 2017) 실시간으로 주식의 가격에 영향을 미친다는 특징을 갖고 있다. 이러한 호가의 특성은 주식시장 분석 방법 중 하나인 기술적 분석의 배경이 ‘주가는 수요와 공급의 원리에 따라 결정된다’라는 점을 고려하면 주식시장 예측에 적합한 변수임을 알 수 있다.

최근 재무 분야에서는 호가 정보를 이용하여 시장의 미시구조(micro structure)를 분석하는 등 다양한 연구가 수행되고 있으며, 그 중요성이 더욱 부각되고 있다. 특히 재무 관련 선행 연구(이우백 & 최혁, 2006a; 이우백 & 최혁, 2006b; 이우백 & 최혁, 2007; 이우백 & 최혁, 2012; 우민철 & 김지현, 2014; 정재만 외, 2014)들은 호가 정보가 가격 발견에 공헌하고 단기 수익률 변화 등을 분석할 때 유용함을 주장했지만, 호가 정보를 활용한 주식시장 예측 연구가 전무하다.

따라서 본 연구에서는 딥러닝 기법 중 하나인 순환신경망과 호가 정보를 활용하여 주식시장을 예측하는 연구를 수행하고자 한다. 이 때, 주식시장의 종목들은 가격 발견에 대한 호가 정보의 공헌도에 영향을 미치는 유동성 등의 특징이 모두 상이하기 때문에 호가 정보 분석 시 고려되어야 하므로(우민철 & 김지현, 2014), 장·단기적 시계열 정보의 학습에 있어서 방법이 상이한 Recurrent Neural Network(RNN), Long Short-term memory(LSTM), Gated Recurrent Unit(GRU)를 활용하고자 한다. 본 연구를 통해 제안하는 주식시장 예측모형은 호가 정보를 투자 지표로 사용하고 있는 투자자들의 객관적인 의사결정에 기여할 것으로 기대된다.

## 2. 연구의 구성

본 연구는 서론, 문헌고찰, 연구모형, 실증분석 및 결과, 결론 총 5장으로 구성되어 있으며, 각 장에 대한 내용은 다음과 같다.

제1장 서론에서는 본 연구를 수행하게 된 배경 및 목적과 전체적인 연구의 구성을 요약하여 서술했다.

제2장 문헌 고찰에서는 우선 호가 정보에 대한 개념을 정리했으며, 주식시장 예측과 관련된 연구들을 고찰하였고, 순환신경망과 이를 구성하는 요소들에 대한 개념 및 관련된 연구들을 살펴보았다.

제3장 연구 모형에서는 순환신경망을 활용한 주식시장 예측모형을 개발하기 위해 데이터 라벨링 및 분석용 데이터셋 생성, 모형 학습 및 검증에 대한 방법을 제시했다.

제4장 실증 분석 및 결과에서는 제3장에서 제시한 연구 모형을 활용하여 주식시장 예측모형에 대한 학습 및 검증이 수행되었다.

제5장 결론에서는 연구 결과를 요약하고 이를 바탕으로 연구의 의의를 부여했으며, 본 연구의 한계와 향후 연구 방향을 제시했다.



## II. 문헌 고찰

본 장에서는 주식시장 및 호가 정보, 주식시장 예측, 그리고 순환신경망에 대한 문헌을 고찰했으며 선행연구의 한계점을 서술하였다. 주식시장 및 호가 정보에 대한 고찰에서는 호가 정보의 개요 및 선행연구를 살펴보고자 한다. 또한 주식시장 예측에 대한 고찰에서는 주식시장 예측의 개요, 주식시장 예측에 대한 선행연구를 살펴보고자 한다. 순환신경망에 대한 고찰에서는 RNN, LSTM, GRU에 대한 개요 및 선행연구에 대해 살펴보고자 한다. 마지막으로 선행연구의 한계점에서는 선행연구들에 대한 고찰을 바탕으로 한계점에 대해 서술하고자 한다.

### 1. 호가 정보에 대한 고찰

#### 1.1. 호가 정보의 개요

호가란 시장참여자가 시장에서 매매거래를 위해 제출하는 매도 또는 매수의 의사표시를 의미하고, 제출된 호가는 매매거래시간 내에서 거래소가 접수한 때로부터 매매거래가 성립될 때까지 유효하며, 제출된 호가 중 매매거래가 성립되지 않은 수량의 전부 또는 일부의 취소가 가능하다. 한국주식시장은 2002년 1월을 기준으로 10단계 공개호가를 시행하였으며 시장참여자들은 매도 및 매수 10단계 호가에 대한 정보를 제공받게 되었다.

호가 정보와 관련된 선행 연구들을 통해 시장구조의 제도적인 측면으로 인해 단기적으로 호가 정보는 주가에 지체되어 반영될 수 있다는 점과, 거래자들도 주문 흐름 정보를 추론해서 이익을 실현하기까지는 일정한 시간이 소요됨이 밝혀졌다. 이 같은 이유로 호가 정보로부터 단기적인 주가 방향을 추론할 수 있는 가능성이 제기되었다(이우백 & 최혁, 2007). 이와 같이 호가정보의 공개는 시장참여자들에게 매도 및 매수 호

가 상황을 공개함으로써 시장의 방향성에 대한 예상을 가능하게 하여 투자판단의 지표로 활용될 수 있다는 장점이 있으며(엄세용, 2017), 이는 호가 정보의 공개가 가격 발견에 대해 공헌한다는 선행연구들의 주장을 뒷받침한다. 호가 정보의 특징으로는 시장참여자들의 지정가 주문, 매매 등에 의해 변동되고 각 단계 별로 가격 발견에 공헌하는 정도가 다르며, 종목의 거래량 규모 및 시간대 등 다양한 요인에 의해 가격 발견 기능의 효율성이 변화된다는 점이 있다(이우백 & 최혁, 2006a).

## 1.2. 호가 정보에 대한 선행연구

이우백 & 최혁 (2006a)에 따르면 호가 정보의 공개는 가격 발견 기능에 공헌하며 각 단계 별로 공헌하는 정도가 다르며, 특히 최우선 호가 이하 단계에서는 가격발견의 주도권이 거래량 규모에 따라 체계적으로 증가함을 주장했다. 또한 이우백 & 최혁 (2006b)은 호가 정보로부터 측정되는 잔량과 스프레드의 패턴이 지정가의 가격발견과정을 설명하는 요인이라는 것을 실증적으로 분석하였으며, 최우선평가의 경우 장 시작 직후가 가장 효율적이며 오후장에 접어들수록 효율성이 지체되며, 최우선평가 이하 단계의 경우 장 시작 직후 효율성이 떨어지지만, 시간이 경과함에 따라 효율성이 증가하여 오후장에서는 안정적으로 가격 발견에 공헌한다고 주장했다. 이우백 & 최혁 (2012)은 2002년 1월에 실시된 호가 범위 확대와 관련하여 호가 공개범위의 확대에 따른 시장의 질적 수준의 변화를 관찰하였다. 해당 연구에서는 벡터오차수정모형을 활용하여 Hasbrouck (1995)의 정보량을 추정하였으며, 그 결과 차우선평가 이하 단계에서 가격발견기능이 유의함을 밝혔고 특히 유동성 소비규모가 높은 종목에서 2~5단계 호가의 가격발견에 대한 공헌도는 최우선평가와 유사한 수준이며, 비거래 효과를 완화시킨 결과에서는 6~10단계도 유의적인

공헌을 한다는 점을 밝혔다. Boehmer et al. (2005)은 호가주도형 시장인 NYSE에서 2002년 1월부터 스페셜리스트가 보유한 주문원장이 외부에 전면적으로 공개됨에 따라 투자자의 거래 패턴과 시장효율성에 미친 영향을 분석하였으며, 정보 공개 이후 투자자들은 브로커에 주문을 위탁하지 않고 스스로 주문을 관리하는 경향이 있음을 밝혔다. 또한 과거 스페셜리스트들이 주문원장을 독점함으로써 가졌던 정보 우월성은 상실되어 전체적으로 시장의 효율성과 질을 향상시킨다고 주장했다. Harris & Panchapagesan (2005)은 NYSE에서 스페셜리스트들이 주문원장을 독점적으로 활용하며, 주문원장에 내포된 호가 정보가 단기적 수익률 변화를 예측하는데 유용함을 주장했다. Cao et al. (2009)는 호주주식시장(ASX)을 대상으로 호가의 정보내용을 측정하였으며, 최우선�호가 다음 10단계까지의 호가가 가격발견기능과 단기적 수익률 예측에 공헌함을 밝혔다. Broggard (2010)는 고빈도 호가 및 매매가 저빈도 호가 및 매매에 비해 가격 발견기능 및 시장효율성 제고에 기여한다고 주장했다.

## 2. 주식시장 예측에 대한 고찰

### 2.1. 주식시장 예측 방법

주식시장 예측 방법은 분석하는 방법론을 기준으로 기본적 분석(fundamental analysis)과 기술적 분석(technical analysis)으로 구분할 수 있으며(김선웅 & 안현철, 2010), 기술적 분석은 다시 분석방법론을 기준으로 통계학적 분석 방법과 인공지능 분석 방법으로 구분할 수 있다.

기본적 분석은 기업의 내재가치에 영향을 미치는 모든 요소들을 분석하는 방법으로 기업의 재무 정보와 같은 기업의 과거 성과를 평가하여 주가를 예측하는 방법이다(박재연 외, 2016). 기본적 분석에 사용되는 자료는 재무정보 중 기업의 유동자산(current assets), 고정자산

(non-current assets), 자산총액(total assets), 주주자본(shareholders equity), 주식자본(share capital), 유동부채(current liabilities), 고정부채(non-current liabilities) 등의 재무정보 등이 있다(홍태호 외, 2017). 기본적 분석은 시장에서 형성되는 주식의 가격이 그 주식을 발행한 기업의 가치에 의하여 결정된다는 이론을 배경으로 주시가격을 예측하게 된다(구승환 & 장성용, 2012).

기술적 분석은 ‘주가는 수요와 공급의 원리에 따라 결정되며 수급의 결과는 그래프로 나타난다’는 이론을 배경으로, 그래프 분석을 통해 미래 주가를 예측하는 방법이다(이우식, 2017). 기술적 분석에 사용되는 자료는 대부분 과거 주식의 시가, 고가, 저가, 종가, 거래량 등의 주시가격을 이용해 산출된 기술적 지표(technical indicator)이다(홍태호 외, 2017). 기술적 분석은 주식의 수요와 공급에 대한 변화는 주식시장의 움직임과 관련되는 변수에 관찰되며, 이러한 변수를 분석함으로써 궁극적으로 미래 주식 가격의 변화를 파악할 수 있다는 믿음에 기초하고 있다(김진화 외, 2011).

과거에는 기술적 지표를 활용하여 주식시장을 예측하는 방법으로 통계학적 분석 방법이 활용되었으며(김진백 & 김유일, 2000), 주로 이동 평균 분석법(moving average analysis)이나 몬테카를로 시뮬레이션(monte carlo method) 등의 방법으로 주식시장을 예측하였다.

하지만 시간이 흐름에 따라 통계학적 분석 방법으로 시작한 주식시장 예측은 인공지능을 이용한 방법으로 발전하였으며, 인공지능 분석 방법은 주가예측기술 개발에 큰 공헌을 했다(이석준 & 오경주, 2011). 인공지능 분석 방법은 높은 비선형성과 동적인 성질을 가진 문제에서 통계적 방법에 비해 예측과 분류를 더 정확하게 수행한다는 장점이 있다. 인공 신경망(artificial neural network), 지지벡터머신(support vector machine),

의사결정나무(decision tree) 등의 머신러닝(machine learning) 기법에서 시작된 인공지능 분석 방법은 최근 컴퓨팅 기술이 발달함에 따라 심층신경망(deep neural network), 순환신경망(recurrent neural network) 등의 딥러닝(deep learning) 기법으로 발전하였다. 기존의 머신러닝 기법들은 데이터를 분석함에 있어서 특징 추출(feature extraction) 등의 과정을 거쳐야 한다는 점에 있어서 한계가 있었지만, 딥러닝 기법은 단순하지만 강력한 비선형 모듈로 네트워크를 구성하여 입력되는 데이터를 추상적인 정보로 변환하는 과정을 반복함으로써 매우 복잡한 문제도 해결할 수 있게 되었다(Lecun et al., 2015).

## 2.2. 주식시장 예측에 대한 선행연구

본 연구에서는 주식시장 예측에 대한 선행연구 중 통계학적 분석 방법, 인공지능 분석 방법 위주로 살펴보았다. 우선 통계학적 분석 방법에 대한 선행연구는 다음과 같다. 이재원 (2000)은 통계학적 분석 방법 중 하나인 몬테카를로 시뮬레이션 기법이 주가 추세의 현재 상태를 평가하고 가치도 함수의 계산을 위해 사용하여 주식시장 예측에 대한 효용성을 검증하였다. 남주하 & 윤기향 (2001)은 통계학적 분석 방법 중 하나인 이동 평균 분석법을 활용하여 국내 주식시장에 대한 미국 주식시장의 변동성 이전 효과를 분석하였으며, 1999년부터 2000년의 기간에서 변동성 이전이 발생하였음을 확인하였다. 유진 (2002)은 몬테카를로 시뮬레이션 기법을 활용하여 주식시장에 가격제한이 존재할 때, 실제 시장에서의 주가에 대한 VaR (Value at Risk)을 산출하는 방법을 제안하였다. 이윤선 (2003)은 단기 이동 평균선을 활용하여 매매하는 교차매매전략, 단일 이동평균의 등락을 이용한 추세매매전략, 주가와 이동평균선의 교차매매전략에 대한 검증을 시도하였으며, 주식시장 예측에 이동 평균 분석

이 유용함을 확인하였다. 신동백 (2007)은 몬테카를로 시뮬레이션 기법과 환율 시계열 자료의 통계수치를 활용하여 환율의 변화를 예측하였으며, 전문예측기관의 예측값과 비교하여 몬테카를로 시뮬레이션 기법의 유용성을 확인하였다. 김현지 & 장우진 (2010)은 한국, 미국, 일본의 주식시장에 대한 이동 평균 분석법의 효과성에 대한 검증을 수행하였으며, 5일 이동 평균과 20일 이동 평균으로 조합된 투자 전략이 모든 국가에서 효과적이라는 점을 확인하였다.

인공지능 분석 방법에 대한 선행연구는 다음과 같다. 이형용 (2008)은 기술적 지표와 로지스틱 회귀 분석, 인공신경망, 유전 알고리즘을 활용하여 코스피 주가 지수의 일별 등락을 예측하였으며, 인공지능 분석 방법이 주식시장 예측에 유용함을 확인하였다. 안성원 & 조성배 (2010)의 경우 텍스트 마이닝을 활용하여 코스피에 상장된 종목과 관련된 뉴스 기사를 분석했고, 해당 종목들의 가격등락을 예측하였으며 텍스트 마이닝 기법이 주식시장 예측에 유용함을 확인하였다. 이우식 (2017)은 인공지능 분석 방법 중 하나인 딥러닝 분석을 활용하여 기술적 분석 지표를 분석하여 코스피 주가 지수의 일별 등락을 예측하였으며, 그 결과 딥러닝 분석이 주식시장 예측에 유용함을 확인하였다. Kim & Han (2000)은 유전 알고리즘 기반의 인공신경망을 활용하여 기술적 지표를 분석했고, 코스피에 상장된 종목들의 일별 가격 등락을 예측하였다. 그 결과 인공신경망이 주식시장 예측에 유용함을 확인하였다. Ince & Trafalis (2004)는 인공지능 분석 방법 중 하나인 서포트벡터머신 알고리즘과 주성분분석을 활용하여 기술적 지표를 분석했고, 나스닥(NASDAQ)에 상장된 종목들의 가격변화를 예측하였다. 그 결과 서포트벡터머신이 주식시장 예측에 유용함을 확인하였다. Huang et al. (2008)은 인공신경망과 군집 분석, 서포트벡터머신 등의 알고리즘으로 대만 거래소와 코스피에 상장된 종목

들의 가격 등락을 예측하였으며, 그 결과 인공지능 분석 방법이 주식시장 예측에 유용함을 확인하였다. Lai et al. (2009)은 인공지능 분석 방법 중 하나인 의사결정나무와 군집 분석을 활용하여 기술적 지표를 분석했고, 대만 거래소에 상장된 종목들의 가격 등락을 예측하였다. 그 결과 의사결정나무가 주식시장 예측에 적합함을 확인하였다.

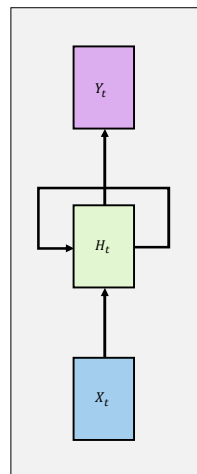
### 3. 순환신경망에 대한 고찰

최근, 딥러닝 기법 중 순환신경망은 데이터의 시계열 속성을 활용하여 학습할 수 있다는 장점 때문에 주식시장, 연상 메모리(associative memory), 시공간 패턴(spatiotemporal pattern) 분류, 제어, 최적화, 패턴 시퀀스의 예측 및 일반화를 비롯한 다양한 응용 분야에서 활용되고 있으며(Petrosian et al., 2000; Petrosian et al., 2001; Shieh et al., 2004), 대표적인 순환신경망 모형으로는 RNN, LSTM, GRU가 있다. 세 모형 모두 은닉층에서 순환하는 구조를 통해 시계열 맥락 정보를 포착하여 학습한다는 점에서 동일하지만, 은닉층에서 구조적 차이가 있으며 이로 인해 학습하는 방식이 다르다. 우선 RNN은 온전히 연결된(fully connected) 은닉층 구조를 가지고 있으며 시계열 전체에 걸쳐 전반적인 맥락을 학습하게 되고, LSTM은 메모리셀 상태와 은닉층을 활용하여 선택적으로 장기기억과 단기기억을 나누어 학습하게 된다. GRU는 은닉층을 활용하여 장기기억과 단기기억을 선택적으로 전달하여 학습하게 된다는 차이점이 있다. 각 모형에 대한 자세한 내용은 다음과 같다.

#### 3.1. RNN(Recurrent Neural Network)

최초의 RNN은 Elman (1990)에 의해 제안되었으며, 기존의 전방 전달 신경망(feed forward neural network)과 달리 이전 시점의 출력을 현재

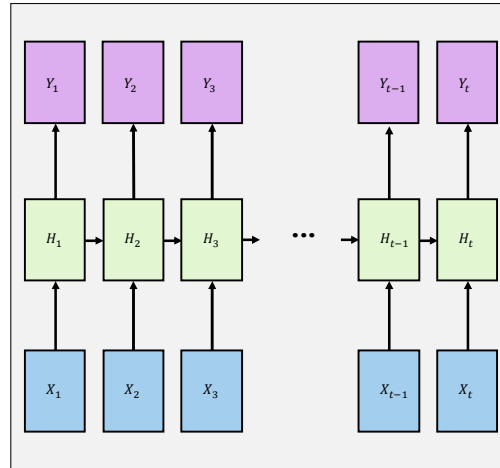
시점의 입력으로 공급하는 순환구조를 가지고 있으며(Zheng et al., 2019), 시계열 데이터에 존재하는 시간적 맥락 정보를 포착하도록 설계되었다(Liu & Sullivan, 2019). RNN의 기본적인 구조는 아래 <그림 1>과 같다.



<그림 1> RNN의 구조

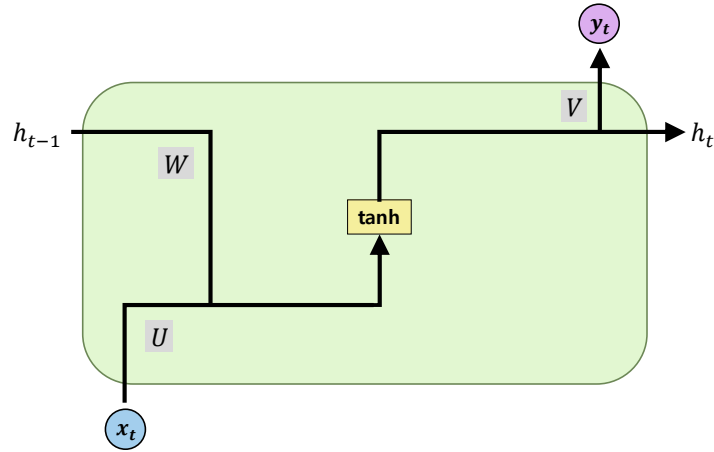
RNN은 이전 시점의 출력을 현재 시점의 입력으로 공급하기 위해 은닉층에서 순환하는 구조를 확인할 수 있으며, RNN은 다시 아래 <그림 2>와 같이 펼쳐서 표현할 수 있다.





<그림 2> 펼쳐진(un-rolled) RNN 구조

<그림 2>에서  $X_t$ 는  $t$ 시점의 입력값을 의미하고,  $H_t$ 는  $t$ 시점의 은닉층으로부터 계산된 은닉값을 의미하며,  $Y_t$ 는  $t$ 시점의 출력값을 의미한다. RNN은 각 시점의 입력값( $X_t$ )으로부터 산출된 은닉값( $H_t$ )을 다음 시점의 은닉값( $H_{t+1}$ )의 산출에 활용한다는 점에서 시계열 데이터에 존재하는 시간적 맥락 정보를 포착할 수 있다. 아래의 <그림 3>은 RNN의 핵심이라 할 수 있는 은닉층의 구조에 대한 내용이다.



<그림 3> RNN 은닉층의 내부 구조

순환신경망은  $t$  시점의 입력값( $x_t$ )과  $t-1$  시점의 은닉값( $h_{t-1}$ )을 활용하여 현재 시점의 은닉값( $h_t$ ) 및 출력값( $y_t$ )을 산출하게 된다. 현재 시점의 은닉값은 이전 시점의 은닉값과 현재 시점의 입력값에 하이퍼볼릭 탄젠트(hyperbolic tangent) 함수를 적용하여 산출되고, 현재 시점의 출력값은 산출된 현재 시점의 은닉값을 활용하여 산출되게 된다. 아래 식 (1)은  $t$  시점의 은닉값에 대한 수식이고, 식 (2)는 출력값에 대한 수식이다.

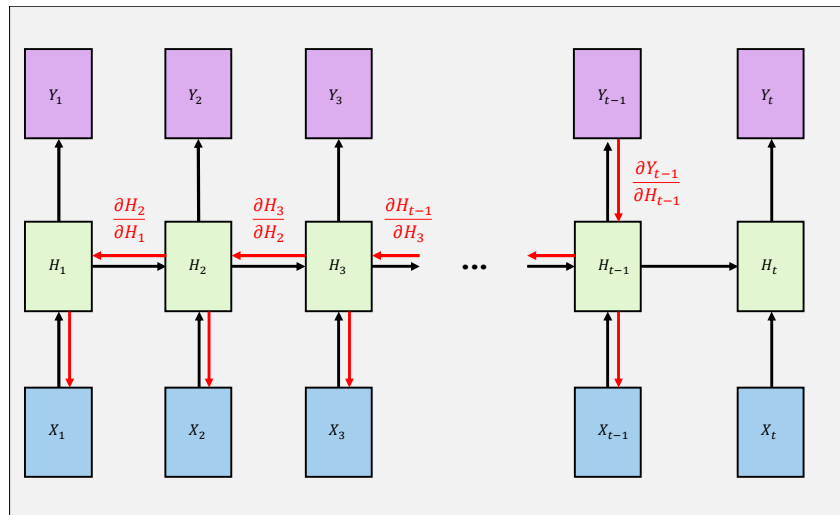
$$h_t = \tanh(W \times h_{t-1} + U \times x_t + b_h) \quad (1)$$

이 때,  $W$ 는  $H_{t-1}$ 과 선형결합하게 되는 가중치이며,  $U$ 는  $x_t$ 와 선형결합하게 되는 가중치이다.

$$y_t = V \times h_t + b_y \quad (2)$$

이 때,  $V$ 는  $h_t$ 와 선형결합하게 되는 가중치이다.

일반적으로 RNN을 포함한 순환신경망은 BPTT(Back-Propagation Through Time) 기법을 통해 학습하게 된다(Williams & Peng, 1990). BPTT는 기본적으로 오류역전파(backpropagation) 알고리즘과 같지만, 은닉층의 가중치가 서로 공유되어 학습되는 동안 누적되어 조정된다는 차이가 있으며(Werbos, 1990), 자세한 내용은 아래 <그림 4>와 같다.



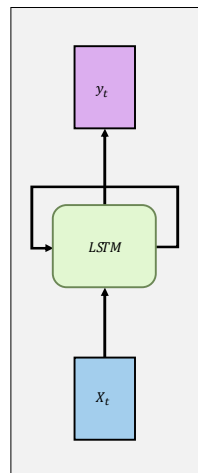
<그림 4> BPTT의 흐름도

위 <그림 4>와 같이 BPTT는 실제값과 예측값 사이의 차이를 손실함수로 정의한 후, 네트워크의 역방향으로 손실함수를 입력에 대해 편미분하여 값이 최소화시키는 방식으로 가중치들을 업데이트하며 학습을 진행하게 된다. 이론적으로 RNN은 장기 의존성 문제를 다룰 수 있다고 하지만 자연어 처리 등의 분야에서는 실제로는 그렇지 못하다(서지혜, 2017). RNN은 입력 데이터의 시계열 길이가 긴 복잡한 문제를 다룰 때에는 기울기가 소실(vanishing gradient)되어 과거 학습 결과가 사라지는 장기 의존성(long-term dependencies) 문제가 발생한다는 단점이 있다(Bengio

et al., 1994; Güler et al., 2005).

### 3.2. LSTM(Long Short-Term Memory)

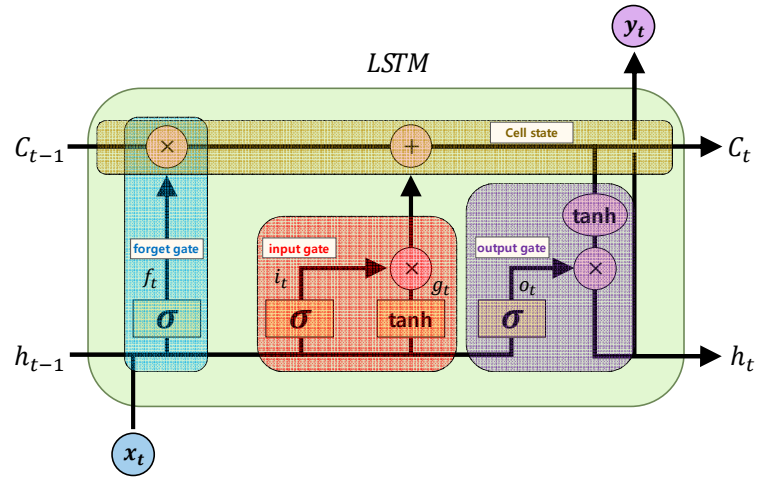
LSTM은 Hochreiter & Schmidhuber (1997)에 의해 제안된 모형이며 자세한 내용은 다음과 같다. RNN은 시계열 예측에 있어서 우수한 성능을 보인다는 것이 증명되었지만, 시계열의 길이가 늘어남에 따라 장기 의존성 문제가 발생할 수 있는 가능성이 있다(Bengio et al., 1994; Tian et al., 2018). LSTM은 이러한 장기 의존성 문제의 발생 가능성을 개선하기 위해 고안된 알고리즘이며(Hochreiter & Schmidhuber, 1997), 자세한 내용은 아래 <그림 5>와 같다.



<그림 5> LSTM 구조

LSTM의 구조는 RNN과 유사하지만 은닉층에서 차이가 있다. LSTM은 은닉층 내에 메모리셀(memory cell)이라는 기본 구조를 가지고 있으며(Liu & Sullivan, 2019), 메모리셀의 상태(state)를 이용하여 시간적 맥락 정보를 기억한다. LSTM의 메모리셀에 대한 자세한 설명은 아래 <그

림 6>과 같다.



<그림 6> LSTM의 내부 구조

LSTM의 핵심은 <그림 6>의 메모리셀 상태( $C_t$ )를 결정하는 각 게이트(forget 게이트, input 게이트, output 게이트)에 있다. LSTM은 각 게이트를 이용하여 어느 정보를 얼마만큼 메모리셀 상태에 포함시킬지 결정하게 되며(Graves & Schmidhuber, 2005), 게이트에 대한 자세한 내용은 다음과 같다.

첫 번째로 forget 게이트는 더 이상 과거 정보가 필요하지 않을 때 과거 정보를 잊기 위한 게이트(Gers et al., 1999)이며 자세한 내용은 다음과 같다. forget 게이트는 이전 시점의 메모리셀 상태( $C_{t-1}$ )를 현재 시점의 메모리셀 상태( $C_t$ )로 얼마나 전달할지를 결정하는 역할을 한다.

forget 게이트는  $f_t$  값에 의해 제어되며  $f_t$  값은 이전 시점의 은닉값( $h_{t-1}$ )과 현재 시점의 입력값( $x_t$ )에 대해 시그모이드(sigmoid) 함수를 취해 얻게 된다. 시그모이드 함수는 0에서 1 사이의 값을 출력하며, 만약  $f_t$ 의 값이 0이라면 이전 시점의 정보는 완전히 잊어버리게 되고, 값이 1

이라면 이전 시점의 정보를 온전히 기억하게 된다. 아래 식 (3)은  $f_t$ 에 대한 수식이다.

$$f_t = \sigma(W_f \times h_{t-1} + U_f \times x_t + b_f) \quad (3)$$

이 때,  $W_f$ 는  $h_{t-1}$ 과 선형결합하게 되는 가중치이며,  $U_f$ 는  $x_t$ 와 선형결합하게 되는 가중치이다.

두 번째로 input 게이트는 새로운 정보를 기억하기 위한 게이트이며 자세한 내용은 다음과 같다. input 게이트는  $i_t$ 값에 의해 제어되며,  $g_t$ 를 현재 시점의 셀 상태로 얼마나 전달할지 결정하게 된다.  $i_t$ 값은  $f_t$ 값과 마찬가지로 이전 시점의 은닉값과 현재 시점의 입력값에 대해 시그모이드 함수를 취해 얻게 되며,  $g_t$ 값은 이전 시점의 은닉값과 현재 시점의 입력값에 대해 하이퍼볼릭 탄젠트(hyperbolic tangent) 함수를 취해 얻게 된다. 아래의 식 (4)는  $i_t$ 에 대한 수식이며, 식 (5)는  $g_t$ 에 대한 수식이다.

$$i_t = \sigma(W_i \times h_{t-1} + U_i \times x_t + b_i) \quad (4)$$

이 때,  $W_i$ 는  $h_{t-1}$ 과 선형결합하게 되는 가중치이며,  $U_i$ 는  $x_t$ 와 선형결합하게 되는 가중치이다.

$$g_t = \tanh(W_g \times h_{t-1} + U_g \times x_t + b_g) \quad (5)$$

이 때,  $W_g$ 는  $h_{t-1}$ 과 선형결합하게 되는 가중치이며,  $U_g$ 는  $x_t$ 와 선형결합하게 되는 가중치이다.

마지막으로 output 게이트는 현재 시점의 은닉값과 출력값( $y_t$ )을 출력하기 위해 현재 시점의 메모리셀 상태의 어느 부분을 채택할지 선택하기

위한 게이트이며 자세한 내용은 다음과 같다. output gate는  $o_t$  값에 의해 제어되며,  $o_t$  값은  $f_t$ ,  $i_t$ 와 마찬가지로 이전 시점의 은닉값과 현재 시점의 입력값에 대해 시그모이드 함수를 취해 얻게된다. 아래의 식 (6)은  $o_t$ 에 대한 수식이며, 식 (7)은 forget 게이트, input 게이트를 거친 현재 시점의 메모리셀 상태에 대한 수식이고, 식 (8)은 최종적으로 출력되는 현재 시점의 은닉값과 출력값에 대한 수식이다.

$$o_t = \sigma(W_o \times h_{t-1} + U_o \times x_t + b_o) \quad (6)$$

이 때,  $W_o$ 는  $h_{t-1}$ 과 선형결합하게 되는 가중치이며,  $U_o$ 는  $x_t$ 와 선형결합하게 되는 가중치이다.

$$C_t = f_t \times C_{t-1} + i_t \times g_t \quad (7)$$

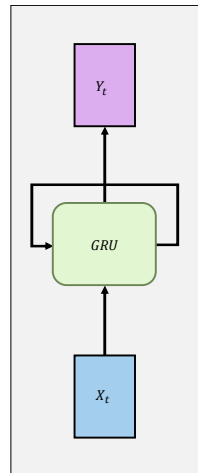
$$y_t, h_t = o_t \times \tanh(C_t) \quad (8)$$

모든 시점에 걸쳐 정보를 덮어쓰는 기존의 RNN과 달리, LSTM은 각 게이트들을 이용하여 선택적으로 정보를 채택하고 유지할 수 있다. 따라서 LSTM이 학습 초기 단계에서 검출한 중요한 정보는 시계열의 길이가 길어져도 장기 의존성 문제가 발생하지 않고 잘 유지될 수 있다 (Chung et al., 2014).

### 3.3. GRU(Gated Recurrent Unit)

GRU는 Cho et al. (2014)에 의해 제안된 모형이며 자세한 내용은 다음과 같다. LSTM은 메모리셀 상태 구조를 이용하여 장기 의존성 문제를 개선하였지만 여전히 모델의 복잡성으로 인해 연산량이 크다는 문제점이 있다(Shin & Sung, 2016). GRU는 이러한 연산량의 문제를 개선하

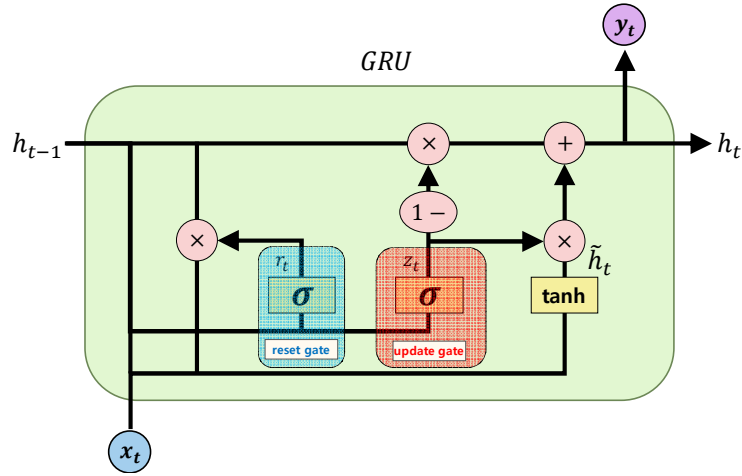
기 위해 LSTM의 구조를 보다 단순하게 처리한 LSTM 변형 모델 중 하나이며(Cho et al., 2014), 자세한 내용은 아래 <그림 7>과 같다.



<그림 7> GRU의 구조

GRU는 LSTM과 구조적으로 유사하지만, LSTM은 메모리셀 상태를 결정하기 위해 사용하는 3개의 게이트(forget 게이트, input 게이트, output 게이트)를 이용하게 되어 연산량이 증가하게 되는데, GRU는 총 2개의 게이트(update 게이트와 reset 게이트)로 구성되어 있으며, 해당 게이트들을 활용하여 LSTM과 비슷한 연산을 수행한다. 또한, LSTM 상에 존재하는 메모리셀 상태( $C_t$ )와 은닉값( $h_t$ )을 GRU에서는 하나의 은닉값으로 대체하여 사용한다는 차이점이 있다(송경환, 2018). GRU 내부 구조에 대한 자세한 내용은 아래 <그림 8>과 같다.





<그림 8> GRU의 내부 구조

GRU는 현재 시점의 은닉값을 산출하기 위해 중간은닉값( $\tilde{h}_t$ )을 사용하게 된다. 우선 update 게이트는 LSTM의 forget 게이트와 input 게이트의 역할을 수행하게 된다. update 게이트는  $z_t$  값에 의해 제어되며, 이전 시점의 은닉값( $h_{t-1}$ )과 현재 시점의 중간은닉값( $\tilde{h}_t$ )을 어떻게 조합하여 은닉값을 산출할지 결정하게 된다.  $z_t$  값은 이전 시점의 은닉값과 현재 시점의 입력값( $x_t$ )에 대해 시그모이드 함수를 취해 얻게 되며, 아래의 식 (9)은  $z_t$ 에 대한 수식이다.

$$z_t = \sigma(W_z \times h_{t-1} + U_z \times x_t + b_z) \quad (9)$$

이 때,  $W_z$ 는  $h_{t-1}$ 과 선형결합하게 되는 가중치이며,  $U_z$ 는  $x_t$ 와 선형결합하게 되는 가중치이다.

reset 게이트는  $r_t$  값에 의해 제어되며, 현재 시점의 중간은닉값을 산출

할 때 이전 시점의 은닉값을 얼마만큼 반영할지 결정하게 된다.  $r_t$ 는  $z_t$ 와 마찬가지로 이전 시점의 은닉값과 현재 시점의 입력값에 대해 시그모이드 함수를 취해 얻게 되며, 아래의 식 (10)은  $r_t$ 에 대한 수식이다.

$$r_t = \sigma(W_r \times h_{t-1} + U_r \times x_t + b_r) \quad (10)$$

이 때,  $W_r$ 는  $h_{t-1}$ 과 선형결합하게 되는 가중치이며,  $U_r$ 는  $x_t$ 와 선형결합하게 되는 가중치이다.

최종적으로 reset 게이트와 update 게이트를 거쳐 중간 은닉값과 은닉값이 산출되게 된다. 중간은닉값은 reset 게이트를 거친 이전 시점의 은닉값과 현재 시점의 입력값에 하이퍼볼릭 탄젠트 함수를 취하여 얻게 된다. 마지막으로 은닉값은 중간은닉값과 이전 시점의 은닉값이 update 게이트를 거쳐 산출되게 된다. 아래 식 (11)은 중간은닉값에 대한 수식이며, 식 (12)은 은닉값에 대한 수식이다.

$$\tilde{h}_t = \tanh(r_t \times W_h \times h_{t-1} + U_h \times x_t + b_h) \quad (11)$$

이 때,  $W_h$ 는  $h_{t-1}$ 과 선형결합하게 되는 가중치이며,  $U_h$ 는  $x_t$ 와 선형결합하게 되는 가중치이다.

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \quad (12)$$

LSTM은 은닉값과 메모리셀 값을 활용하여 학습을 진행하기 때문에 연산량으로 인해 학습속도가 느리다는 단점이 있다. 반면 GRU는 은닉값만을 활용하기 때문에 연산량이 적어 학습속도가 빠르지만, LSTM과 유사하게 각 시점마다 서로 다른 시간 규모의 정보를 적응적으로 포착할

수 있다는 장점이 있다.

### 3.4. 순환신경망에 대한 선행연구

순환신경망과 관련된 국내외 선행연구 조사 결과 대부분 주식시장 예측에 활용된 것으로 나타났으며, 먼저 국내에서 수행된 순환신경망 모형에 관한 선행연구를 살펴보면 다음과 같다. 신동화 외 (2017)는 순환신경망 모형 중 RNN을 활용하여 KOSPI 200 일부 종목을 대상으로 기술적 지표를 분석하였으며, 종목 별로 차이는 있으나 RNN이 심층 신경망에 비해 약 15% 정도 예측 성능이 우수함을 밝혔다. 주일택 & 최승호 (2018)는 순환신경망 모형 중 LSTM을 활용하여 구글의 일별 주식 가격 및 거래량 데이터를 분석하여 종가를 예측하였으며, 비선형적인 주식 가격의 예측에 LSTM이 적합함을 확인하였다. 김성수 & 홍광진 (2017)은 RNN을 이용하여 KOSPI 200 선물 지수를 분석하여 가격의 등락을 예측하였으며 입력 변수로 가격만을 사용하여도 예측이 가능함을 확인하였고, LSTM의 예측 결과를 토대로 투자 시뮬레이션을 수행한 결과, 수익이 발생함을 확인하였다. 성노윤 & 남기환 (2017)은 순환신경망 모형 중 하나인 LSTM과 텍스트 마이닝 기법을 활용하여 뉴스 기사 데이터와 거시경제 변수를 분석하여 금값을 예측하였으며, 경제 변수와 텍스트 데이터가 함께 분석될 때 예측 성능이 증가하는 것을 확인하였다. 해외에서 수행된 순환신경망 관련 선행연구를 살펴보면 다음과 같다. Lin et al. (2009)은 순환신경망 모형 중 하나인 ESN (Echo State Network)과 주성분분석을 활용하여 S&P 500 종목을 대상으로 일별 가격 및 거래량 데이터를 분석하여 종가를 예측하였으며, 주식시장 예측에 순환신경망이 유용함을 확인하였다. Roondiwala et al. (2017)은 신경망 모형 중 하나인 LSTM을 활용하여 NIFTY 50 지수를 대상으로 가격 및 거래량 데이

터를 분석하였으며, 주식시장 예측에 순환신경망이 유용함을 확인하였다. Zhuge et al. (2017)은 순환신경망 모형 중 하나인 LSTM과 텍스트 마이닝 기법을 활용하여 주식관련 온라인 게시물, 가격 및 거래량 등을 분석하여 상해종합주가지수(Shanghai Stock Composite Index)를 예측하였으며, 주식시장 예측에 있어서 LSTM이 유용함을 확인하였다.

#### 4. 선행연구의 한계점 및 본 연구의 차별점

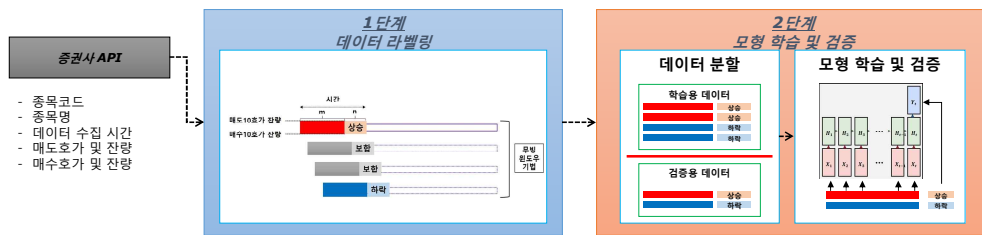
다양한 국내외 선행연구들을 고찰한 결과 선행연구들의 한계점은 다음과 같다. 주식시장 예측 관련 연구들은 S&P 500, KOSPI 200, NIFTY 50, 상해종합주가지수, 구글 등 다양한 지수 및 종목들에 대해 통계학적 방법 및 인공지능 분석 방법 등을 활용하여 예측하였으며 다양한 방법론들이 주식시장 예측에 적합함을 밝혔다. 하지만 해당 연구들을 살펴보면 대부분이 가격과 및 거래량과 관련된 기술적 지표를 활용하여 예측 연구를 수행하였고, 기술적 지표는 과거 시점의 가격과 거래량 정보를 이용하여 추세를 파악한다는 점에서 기술적 지표만으로 미래 주가지수의 움직임을 예측하는 것은 한계가 있다(이우식, 2017). 반면 본 연구에서 순환 신경망에 적용하고자 하는 호가정보는 주식에 대한 수요와 공급의사 및 매매행태가 반영된 정보이므로 시장의 방향성에 대한 예상을 가능하게 하여 실시간으로 주식의 가격에 영향을 미친다는 특징을 갖고 있다.

이런 호가 정보의 특성은 기술적 분석의 배경이 ‘주가는 수요와 공급의 원리에 따라 결정된다’라는 점을 고려하면 주식시장 예측에 적합하다는 점을 알 수 있다. 또한 호가 정보의 중요성으로 인해 재무 분야에서는 호가 정보에 대한 연구들이 활발하게 수행되고 있으며 다양한 연구결과들이 제시되고 있지만 이를 활용한 주식시장 예측연구는 전무하다. 따라서 본 연구에서는 순환신경망 기법을 활용하여 호가 정보를 분석함으

로써 주식시장을 예측하는 연구를 수행하고자 한다.

### III. 연구 모형

순환신경망 기법을 활용한 주식시장 예측모형을 개발하기 위해 본 연구에서는 <그림 9>와 같은 실험이 수행되며, 연구 모형의 세부적인 내용은 다음과 같다.



<그림 9> 연구 모형

#### 1. 데이터 라벨링 및 기계학습용 데이터셋 생성

입력 변수(input variables)와 출력 변수(output variables)를 구성하기 위해 무빙 윈도우 기법(moving window method)(Jang et al., 1993; Lee et al., 2017)을 이용한 ‘데이터 라벨링(data labeling)’ 과정이 수행되며 자세한 내용은 다음과 같다.

수집된 데이터는 아래 식 (13)과 같이 일정한 시계열 길이( $l$ )를 기준으로 가격의 상승 또는 하락 여부를 판단하여 라벨(label)이 부여된다.

$$Label = \begin{cases} 0(\text{하락}) & \text{if } P_{t+l} < P_t \\ 1(\text{상승}) & \text{if } P_{t+l} > P_t \end{cases} \quad (13)$$

이 때  $P_t$ 는  $t$ 시점의 주가가격을 의미하고, 라벨이 부여되는 입력 변수에 대한 설명은 아래 식 (14)와 같다.

$$Input = [V_1, V_2, \dots, V_t] \quad (14)$$

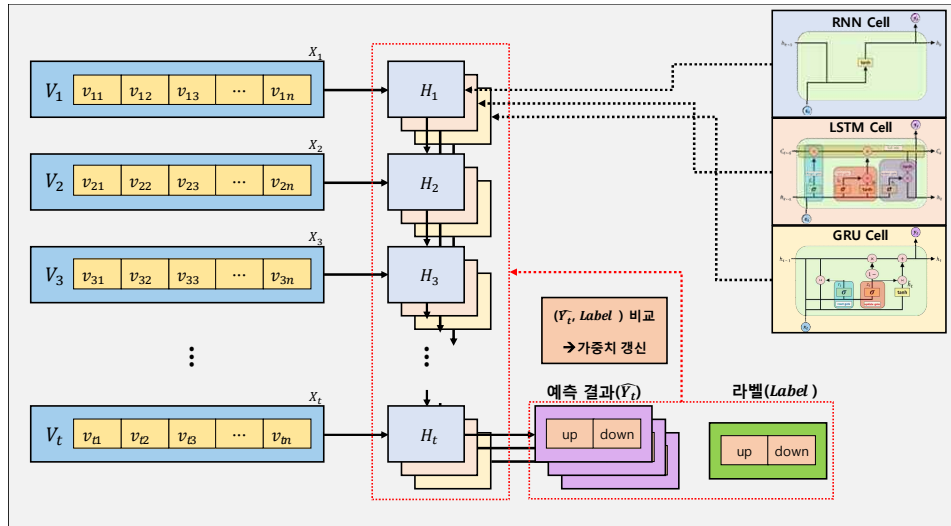
$Input$ 은  $[1, 2, \dots, t]$ 기간 동안의  $V_t$ 로 구성되고,  $V_t$ 는  $t$ 시점의 호가 잔량을 의미하며,  $V_t$ 에 대한 설명은 아래 식 (15)과 같다.

$$V_t = [v_{t1}, v_{t2}, \dots, v_{tn}] \quad (15)$$

$V_t$ 는  $t$ 시점의 각 매도 및 매수 호가 잔량에 해당되는  $n$ 개의 변수( $v_{tn}$ )로 구성된다. 라벨링 과정에서 생성되는  $Input$ 과  $Label$ 은 각각 주식시장 예측모형의 입력 변수와 출력 변수로 채택되게 된다. 또한, 라벨 불균형을 보이는 데이터를 대상으로 분류(classification)를 하는 경우 일반적으로 주요 라벨로만 분류가 되는 경향이 있으므로(서재현, 2018), 라벨 비율을 맞춰주는 다운 샘플링(down-sampling) 과정이 수행되며 기계학습용 데이터셋이 생성된다.

## 2. 모형 학습 및 검증

모형 학습 및 검증 단계에서는 ‘데이터 분할’, ‘모형 학습’, ‘모형 검증’ 과정이 수행되며 자세한 내용은 다음과 같다. ‘데이터 분할’ 과정에서는 모형의 학습 및 검증을 위해 앞선 단계에서 생성된 기계학습용 데이터셋이 학습용(training) 데이터셋과 검증용(validation) 데이터셋으로 분할되게 된다. ‘모형 학습’ 과정에서는 아래 <그림 10>과 같이 앞서 분할된 학습용 데이터셋을 활용하여 주식시장 예측모형의 학습이 수행되며, 순환신경망 모형 중 가장 대표적인 RNN, LSTM, GRU가 본 연구의 실험 모형으로 활용된다.



<그림 10> RNN, LSTM, GRU를 활용한 예측모형 학습 예시

마지막으로 ‘모형 검증’ 과정에서는 검증용 데이터셋을 활용하여 학습이 완료된 예측모형에 대한 성능 검증이 수행된다.



## IV. 실증분석 및 결과

### 1. 데이터 기간 및 수집

본 연구에서는 주식시장 예측모형 학습에 사용될 학습 데이터를 수집하기 위해 키움증권에서 제공하는 Open API 서비스를 이용하였다. 데이터 수집대상은 아래 <표 1>과 같이 코스닥(KOSDAQ) 시가총액 상위 50종목으로 선정되었다.

<표 1> 코스닥 시가총액 상위 50종목

시가총액순위	종목코드	종목명	시가총액순위	종목코드	시가총액순위
1	091990	셀트리온헬스케어	26	035900	코오롱티슈진
2	215600	CJ ENM	27	178920	차바이오텍
3	035760	신라젠	28	022100	삼천당제약
4	084990	바이로메드	29	036830	하림지주
5	003670	포스코케미칼	30	038540	메지온
6	086900	에이치엘비	31	048260	카페24
7	028300	메디톡스	32	240810	솔브레인
8	253450	스튜디오드래곤	33	102940	이오테크닉스
9	263750	팔어비스	34	000250	에스엠
10	025980	셀트리온제약	35	066970	SKC코오롱PI
11	068760	SK머티리얼즈	36	122870	포스코 ICT
12	034230	제넥신	37	183490	유텔렉스
13	036490	휴젤	38	214370	리노공업
14	078340	파라다이스	39	140410	위메이드
15	145020	에이비엘바이오	40	058470	상상인
16	095700	에스에프에이	41	086520	케어젠
17	056190	컴투스	42	036420	콜마비엔에이치
18	046890	아난티	43	092040	에이치엘비생명과학
19	041960	코미팜	44	083790	오스템임플란트
20	028150	원익IPS	45	069080	티씨케이
21	003380	서울반도체	46	200230	나노스
22	098460	고영	47	030190	제이콘텐츠트리
23	041510	에코프로비엠	48	073070	NICE평가정보
24	042000	GS홈쇼핑	49	007390	웹젠
25	192080	JYP Ent.	50	267980	천보

데이터 수집기간은 2019년 3월 20일부터 2019년 5월 7일까지이며, 2002년 1월 2일 한국거래소에 의해 지정된 공개범위인 매도 및 매수호가 10단계에 한하여 수집되었다. 또한 수집대상인 코스닥 상위 50종목 중 케어젠(214370) 종목은 매출채권 감사 증거 확보 등의 문제로 거래가 정

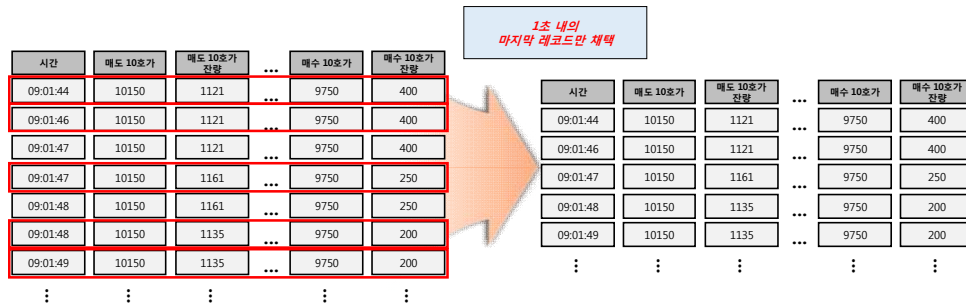
지되어 분석대상에서 제외되었다. 아래 <표 2>는 수집된 주식시장 데이터에 대한 예시이다.

<표 2> 수집된 주식시장 데이터 예시

시간	매도 10호가	매도 10호가 잔량	...	...	매수 10호가	매수 10호가 잔량
09:01:44	10,150	1121	...	...	9,750	400
09:01:46	10,150	1121	...	...	9,750	400
09:01:47	10,150	1161	...	...	9,750	250
09:01:47	10,150	1161	...	...	9,750	250
09:01:48	10,150	1161	...	...	9,750	250
09:01:49	10,150	1116	...	...	9,750	250

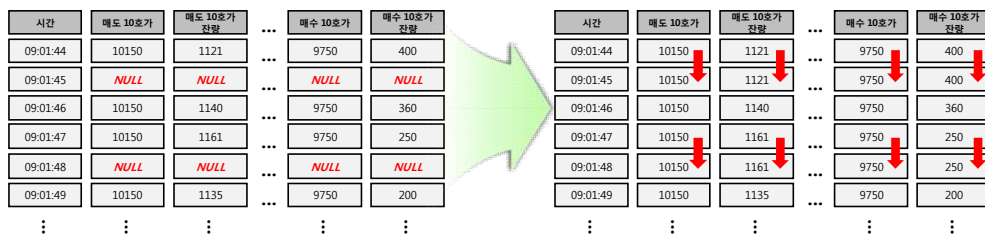
## 2. 호가 데이터 정제

데이터 정제 단계는 총 2단계로 구성되어 있으며 자세한 내용은 다음과 같다. 첫 번째 단계에서는 데이터 레코드(record) 간의 시간 간격이 일정한 간격을 갖도록 수정되었다. 위의 <표 2>를 살펴보면, 수집된 데이터는 매매가 발생하거나 시장참여자들이 가격지정주문(stop order)을 하여 호가 잔량이 변화될 때 기록되는 이벤트(event) 데이터이므로 레코드 간의 시간 간격이 일정하지 않다. 따라서 수집된 데이터는 시장참여자들의 행동에 따라 1초 내에도 다수의 레코드가 존재할 수 있고, 또는 긴 시간동안 레코드가 존재하지 않을 수도 있다. 본 연구에서는 일정한 시간 간격을 기준으로 가격의 상승 또는 하락을 예측하기 위해 전처리 과정을 통해 1초 내에서 중복된 레코드 중 마지막 레코드만을 채택하였으며, 자세한 내용은 아래 <그림 11>과 같다.



<그림 11> 시간 중복 데이터 처리 예시

두 번째 단계에서는 데이터 상에서 레코드가 존재하지 않는 시점에 대한 정제가 수행되었다. 수집된 데이터 상에서 레코드가 존재하지 않는다는 것은 시장참여자들의 행동이 발생하지 않았다는 것을 의미하며, 이는 호가 상태에 아무런 변화가 발생하지 않았다는 것을 의미한다. 따라서 레코드가 존재하지 않는 시점은 1초 전 레코드의 상태로 대체하였으며, 아래 <그림 12>와 같이 모든 레코드의 간격이 1초가 되도록 정제되었다.

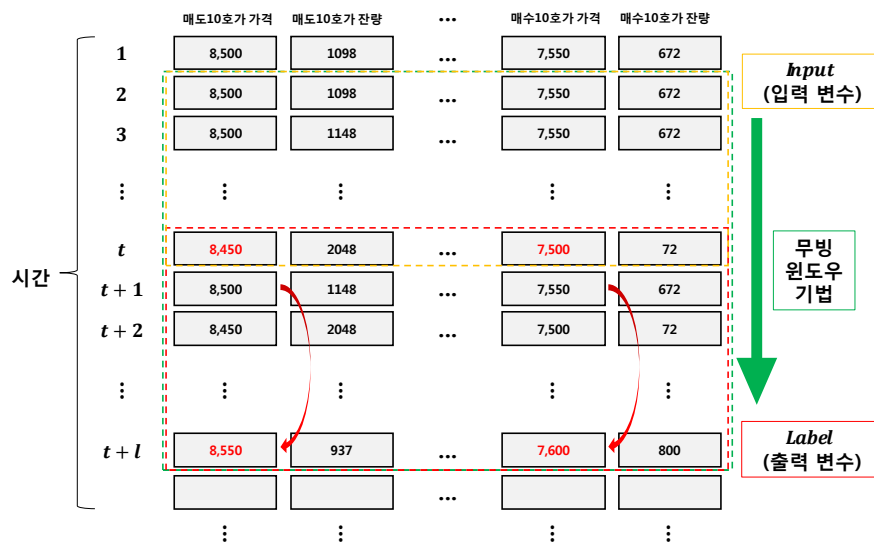


<그림 12> 데이터 정제 예시

### 3. 호가 데이터 라벨링 및 기계학습용 데이터셋 생성

주식시장 예측모형의 학습 및 검증을 위해 ‘데이터 라벨링’ 과정이 수행되었으며, 기계학습용 데이터셋이 생성되었다. ‘데이터 라벨링’ 과정에

서는 아래 <그림 13>과 같이 무빙 윈도우 기법을 이용하여 마지막  $l$ 초 동안 가격의 등락이 발생한 경우 처음  $t$ 기간 동안의 매도 및 매수 10 단계 호가 잔량이 입력변수로 채택되었고, 상승 혹은 하락으로 라벨링이 부여되었다.



<그림 13> 무빙 윈도우 기법을 활용한 데이터 라벨링 예시

이 때,  $t$ 는 우민철 & 최혁 (2013), 정재만 외 (2014)의 연구에서 호가 데이터 분석을 위해 제시된 10초를 채택하였으며, 이우백 & 최혁 (2007)의 연구에서 주장한 가격에 대한 호가 정보 반영의 지연을 고려하여 최적의 매매 타이밍을 탐색하기 위해  $l$ 은 정제된 데이터의 최소 시간 간격인 1초부터 입력 변수의 길이에 해당하는 10초까지를 채택하였다.  $l$ 값에 따라 각 종목 별로 10개의 기계학습용 데이터셋이 생성되었으며, 각 호가의 가격은 분석에 사용되지 않으므로 라벨링 과정에서 활용된 후 제거되었다. 라벨링 과정이 완료된 기계학습용 데이터셋은 각 종목 별로 상

승 혹은 하락 중 데이터 개수가 적은 쪽으로 다운샘플링이 수행되었다. 마지막으로, 최소-최대 정규화(min-max scaling)를 각 레코드 별로 수행한 후 데이터셋이 작성되었다. 최소-최대 정규화는 딥러닝을 위한 가장 보편적인 정규화 방법 중 하나로(Kim et al., 2017; Fang et al., 2017; Zhang et al., 2018), 하나의 값이 그 영역에서 최소 보다 얼마나 더 큰가를 보고 그 영역의 범위를 스케일링(scaling)하는 정규화 기법이다(정병욱 외, 2008). 최소-최대 정규화를 사용하면 데이터는 0과 1 사이의 값으로 만들어지게 된다. 아래 수식 (16)은 최소-최대 정규화에 대한 수식이다.

$$d' = \frac{d - \min(p)}{\max(p) - \min(p)} \quad (16)$$

이 때,  $d$ 는 대상 자료이고,  $\max(p)$ 는 대상 자료의 최댓값이며,  $\min(p)$ 는 대상 자료의 최솟값이다. 아래의 <그림 14>는 최소-최대 정규화 과정을 거쳐 생성된 분석용 데이터셋에 대한 예시이다.

매도 10호가	매도 9호가	매도 8호가	...	매수 8호가	매수 9호가	매수 10호가
498	4489	984	...	1597	776	1365
498	4489	984	...	1597	776	2416
498	4489	984	...	1597	776	2416
498	4489	984	...	1597	776	2416
498	4489	984	...	1597	776	2416
498	4489	984	...	1597	776	2416
498	4489	984	...	1597	776	2416
498	4489	984	...	1597	776	2416
1068	498	4489	...	1597	776	2416
498	4489	984	...	1597	776	2416

매도 10호가	매도 9호가	매도 8호가	...	매수 8호가	매수 9호가	매수 10호가
0.071429	1	0.184504	...	0.327129	0.13611	0.27315
0.071429	1	0.184504	...	0.327129	0.13611	0.517683
0.071429	1	0.184504	...	0.327129	0.13611	0.517683
0.071429	1	0.184504	...	0.327129	0.13611	0.517683
0.071429	1	0.184504	...	0.327129	0.13611	0.517683
0.071429	1	0.184504	...	0.327129	0.13611	0.517683
0.071429	1	0.184504	...	0.327129	0.13611	0.517683
0.071429	1	0.184504	...	0.327129	0.13611	0.517683
0.101733	1	0.211119	...	0.349088	0.164303	0.533423
0.204048	0.071429	1	...	0.327129	0.13611	0.517683
0.110343	1	0.21868	...	0.355328	0.172314	0.537896

기존 데이터

최소-최대 정규화 데이터

<그림 14> 최소-최대 정규화 과정을 거친 분석용 데이터 예시

#### 4. 주식시장 예측모형 학습 및 검증

주식시장 예측모형의 학습에 앞서, ‘데이터 라벨링 및 분석용 데이터셋 생성’ 과정이 완료된 기계학습용 데이터셋을 대상으로 ‘데이터 분할’ 과정이 수행되었다. 데이터 분할은 아래 <표 3>과 같이 수집된 데이터의 약 70%에 해당하는 3월 20일~4월 10일 데이터셋이 학습용 데이터셋으로 선정되었으며, 나머지 30%에 해당하는 4월 11일~4월 22일 데이터셋이 검증용 데이터셋으로 선정되었다.

<표 3> 데이터 분할 정보

데이터 용도	일자		데이터 용도	일자
학습용 데이터셋	2019-03-20	2019-04-01	검증용 데이터셋	2019-04-11
	2019-03-21	2019-04-02		2019-04-12
	2019-03-22	2019-04-03		2019-04-15
	2019-03-25	2019-04-04		2019-04-16
	2019-03-26	2019-04-05		2019-04-17
	2019-03-27	2019-04-08		2019-04-18
	2019-03-28	2019-04-09		2019-04-19
	2019-03-29	2019-04-10		2019-04-22

주식시장 예측모형 학습을 위해 대표적인 순환신경망 모형인 RNN, LSTM, GRU가 활용되었으며, 학습용 데이터셋을 이용하여 각 종목 별로 모형 학습이 수행되었다. 본 연구에서 각 종목 별 모형 학습 시 학습 파라미터(parameter)로 학습 방법은 adamoptimizer, 학습 비율(learning rate)은 0.001, 최대 학습반복 횟수(epochs)는 1,000회, 배치 크기(batch size)는 256으로 설정하였다. 또한 모형의 은닉층을 2개 층으로 구성하여 각 노드 수는 16, 8로 설정하였고, 입력 길이(sequence length)를 10, 입력 차원(data-dimension)은 20으로 설정하였다. 모형 학습이 완료 된 후 검증용 데이터셋을 활용하여 모형의 정확도를 측정하였으며 각 실험 모형 별 검증 정확도는 아래의 <표 4>, <표 5>, <표 6>과 같다.

<표 4> RNN을 활용한 주식시장 예측 검증 정확도(%)

종목코드	$l$									
	1	2	3	4	5	6	7	8	9	10
000250	51.69	53.96	61.23	56.25	53.27	53.30	64.41	56.28	54.17	55.95
003380	62.60	59.94	66.11	65.42	60.45	55.46	57.42	63.15	60.59	61.83
003670	51.37	51.64	64.53	59.53	52.03	56.80	57.99	53.60	52.73	55.45
007390	66.41	51.58	67.34	64.10	50.52	53.87	50.44	59.46	65.08	77.17
022100	51.83	70.83	61.79	63.11	68.52	64.24	69.72	63.10	66.00	55.36
025980	53.66	50.13	69.98	73.52	69.10	66.21	68.96	65.04	60.70	57.12
028150	53.97	52.98	54.99	53.78	54.95	56.86	52.59	54.43	57.09	55.40
028300	53.75	56.68	62.29	66.74	67.45	63.77	63.04	65.99	66.92	67.84
030190	57.65	59.89	60.12	50.60	56.50	68.46	56.36	59.48	57.93	57.20
034230	51.56	52.93	60.63	50.00	53.93	50.98	54.46	51.63	55.82	50.50
035760	62.47	61.79	62.64	63.25	60.80	64.02	62.13	62.42	62.98	58.78
035900	55.58	64.71	56.51	65.91	65.89	50.58	63.60	56.89	73.62	55.37
036420	64.41	64.62	67.75	70.67	66.73	65.77	62.79	62.18	55.95	67.65
036490	54.21	55.49	54.72	54.17	57.27	55.62	59.48	54.19	53.87	57.06
036830	59.19	60.89	58.52	62.30	60.88	60.12	53.73	59.36	58.87	60.25
038540	59.44	53.24	54.28	59.59	55.16	53.37	53.68	54.73	54.89	52.91
041510	52.59	56.19	58.55	50.28	50.17	65.42	56.44	53.90	63.89	63.67
041960	57.66	64.76	63.08	54.08	66.16	60.18	61.94	57.15	61.48	58.19
042000	55.46	51.50	63.57	51.40	56.82	60.47	58.34	52.71	53.94	56.17
046890	58.10	61.74	69.09	75.00	67.53	72.34	68.92	72.21	56.36	75.41
048260	61.22	60.20	51.13	50.96	50.70	54.06	55.31	51.19	55.37	56.74
056190	52.71	64.66	58.15	59.43	57.99	60.73	53.34	57.65	56.18	55.43
058470	58.86	56.82	56.30	53.21	60.78	59.18	58.62	56.13	56.61	58.68
066970	61.50	57.00	57.40	58.88	59.78	59.96	60.45	64.89	60.63	63.16
068760	55.38	63.14	50.79	59.11	61.57	53.74	51.86	60.46	54.71	52.83
069080	56.91	52.70	59.24	53.69	62.50	59.01	59.05	55.49	54.86	57.93
073070	54.50	61.26	59.52	56.45	61.64	58.34	51.08	52.11	57.00	64.18
078340	53.74	56.31	63.22	58.82	51.62	55.32	54.84	58.87	56.06	52.43
083790	62.79	69.08	77.84	74.84	52.74	64.27	74.87	69.62	68.12	54.21
084990	57.34	56.15	59.75	58.09	60.08	59.96	60.17	61.69	58.74	58.82
086520	59.65	55.21	52.81	61.74	58.58	56.00	52.35	55.78	53.77	61.19
086900	55.40	51.10	52.28	54.69	55.54	54.31	54.34	54.24	54.84	54.08
091990	66.42	69.68	57.64	51.32	54.96	55.81	53.21	51.02	50.37	52.01
092040	60.84	54.28	63.04	62.99	53.80	59.08	61.25	61.26	58.20	63.21
095700	64.86	63.74	63.58	53.73	60.24	58.82	61.68	59.44	60.65	57.37
098460	62.58	59.59	57.71	60.82	59.97	51.78	53.71	63.75	53.38	57.84
102940	62.96	55.10	62.94	63.31	59.61	58.82	56.66	62.19	61.38	56.22
122870	53.16	62.69	56.13	56.38	55.12	52.70	54.10	55.33	59.42	55.63
140410	57.81	50.53	58.54	54.24	52.77	53.64	53.20	54.24	56.99	56.56
145020	53.28	51.17	56.55	56.88	52.35	53.80	55.30	51.39	57.14	53.06
178920	68.25	69.49	51.02	70.26	58.74	70.55	63.26	54.06	54.70	61.17
183490	61.61	63.08	61.30	58.57	61.30	60.51	55.27	59.84	59.02	60.66
192080	51.72	53.59	55.23	53.94	56.93	61.09	73.78	61.43	64.41	56.90
200230	68.11	63.23	59.79	69.25	67.13	67.27	66.45	63.66	72.06	62.71
215600	70.83	63.53	56.25	67.11	65.22	77.85	59.11	68.82	71.47	69.00
240810	52.94	51.42	69.22	65.83	53.25	63.12	67.70	51.44	63.88	58.89
253450	67.36	62.81	54.42	57.45	65.58	63.80	68.15	66.12	68.62	61.54
263750	50.81	67.01	62.36	55.77	62.09	57.18	56.16	58.30	57.89	52.10
267980	57.46	59.59	60.50	54.52	59.26	55.73	58.01	52.71	56.86	59.65
평균	<b>58.26</b>	<b>58.77</b>	<b>60.05</b>	<b>59.63</b>	<b>58.90</b>	<b>59.47</b>	<b>59.18</b>	<b>58.59</b>	<b>59.31</b>	<b>58.85</b>

<표 5> LSTM을 활용한 주식시장 예측 검증 정확도(%)

종목코드	$l$									
	1	2	3	4	5	6	7	8	9	10
000250	57.78	59.00	53.29	52.53	56.64	54.17	54.17	52.19	53.93	55.39
003380	61.63	58.07	64.42	58.91	57.71	58.83	54.56	64.84	60.82	61.47
003670	59.91	68.01	53.34	68.76	61.77	58.19	61.81	54.68	52.98	50.06
007390	69.63	66.75	69.95	69.20	69.76	72.39	74.89	66.39	66.12	61.44
022100	61.16	58.33	60.92	66.96	60.25	59.00	58.57	61.06	57.26	60.57
025980	54.77	59.78	59.52	70.00	75.78	77.24	68.15	72.82	73.40	58.58
028150	52.09	52.50	54.43	52.79	52.19	54.27	54.16	55.63	53.08	54.89
028300	60.89	64.08	61.74	58.02	64.60	60.61	58.32	59.92	54.99	58.89
030190	60.70	60.16	55.74	56.27	59.63	61.04	54.78	61.47	56.48	54.17
034230	63.80	64.11	57.38	64.87	55.96	53.59	65.68	60.40	61.37	63.35
035760	56.79	58.48	55.90	58.02	58.25	55.23	58.59	58.84	52.07	59.43
035900	56.74	59.04	55.14	62.79	60.43	54.84	65.96	66.99	53.80	60.17
036420	59.21	64.02	59.45	55.19	63.87	64.91	67.80	57.33	53.11	68.60
036490	55.33	53.45	52.87	52.26	52.00	53.87	53.27	55.96	51.76	50.54
036830	56.15	56.20	58.38	55.87	51.27	56.45	53.78	57.96	56.63	59.33
038540	56.04	57.78	53.28	55.33	52.89	55.03	57.00	55.52	56.68	53.92
041510	58.60	62.31	65.10	58.25	58.21	61.52	62.17	59.97	56.54	51.95
041960	52.01	52.61	58.79	54.37	55.53	55.67	58.59	65.58	58.81	58.63
042000	56.28	54.81	54.21	54.03	56.34	55.56	54.26	56.27	57.85	60.55
046890	59.63	70.18	59.66	49.13	56.25	62.55	60.41	51.96	54.80	58.54
048260	55.41	55.39	55.34	56.49	53.72	53.56	55.68	52.41	59.03	55.22
056190	53.28	56.48	51.32	59.28	56.32	61.07	56.70	60.04	54.91	56.89
058470	53.96	53.89	51.25	55.99	54.44	54.12	57.99	53.14	54.02	54.46
066970	57.61	62.03	56.02	55.97	56.76	55.11	57.75	59.14	57.91	59.64
068760	58.60	56.61	57.02	60.95	60.31	52.71	57.47	53.73	57.54	57.27
069080	57.88	61.15	61.03	57.41	60.68	57.39	60.41	59.32	56.66	57.48
073070	52.74	58.83	54.97	58.37	58.63	54.89	52.10	58.09	59.49	58.22
078340	56.76	57.37	57.99	52.80	56.51	52.92	54.34	56.76	56.52	54.54
083790	61.11	58.36	64.04	58.27	59.88	51.08	56.26	63.59	60.84	64.85
084990	56.06	53.88	53.30	53.68	53.62	55.63	55.94	54.03	53.70	55.56
086520	61.11	51.99	57.64	55.50	55.77	55.97	54.38	57.50	60.04	55.89
086900	53.80	52.04	55.25	53.46	55.84	51.38	54.27	54.52	54.06	53.78
091990	52.61	57.61	61.52	52.52	62.82	54.46	62.74	57.91	56.89	61.54
092040	56.05	55.84	56.92	56.77	59.75	53.02	50.77	55.19	55.59	52.65
095700	54.33	62.35	55.34	57.70	53.88	55.00	57.09	54.61	59.02	55.50
098460	55.70	58.18	58.80	57.28	55.44	55.68	55.54	56.15	56.43	58.02
102940	57.72	62.42	54.46	58.25	58.00	58.42	56.51	58.32	56.68	57.79
122870	55.75	55.95	53.02	55.09	61.77	57.35	61.75	56.33	62.90	54.71
140410	50.96	51.47	53.51	54.24	51.72	51.96	51.29	52.70	52.35	53.91
145020	52.75	52.03	52.40	53.40	54.30	53.93	53.64	53.33	54.06	54.35
178920	59.44	52.13	59.67	64.99	51.02	57.60	52.22	59.88	52.54	50.66
183490	52.65	52.34	52.09	52.46	51.57	49.98	51.77	51.26	50.64	52.72
192080	53.28	60.76	59.97	58.27	57.59	52.22	56.02	54.16	54.32	59.08
200230	58.55	55.43	61.71	61.71	59.48	64.78	61.06	57.65	61.41	61.60
215600	62.50	78.79	70.67	63.53	65.59	72.76	62.53	61.39	72.73	56.97
240810	63.12	60.59	61.32	59.81	54.35	66.15	62.16	57.25	57.80	64.33
253450	58.23	58.30	62.40	54.41	53.93	59.86	58.20	58.37	59.76	56.25
263750	57.25	55.82	54.03	52.24	56.84	58.14	53.11	56.44	54.78	55.60
267980	57.78	53.25	56.84	55.66	59.56	55.47	57.02	51.12	52.09	52.74
평균	<b>57.27</b>	<b>58.39</b>	<b>57.62</b>	<b>57.55</b>	<b>57.74</b>	<b>57.50</b>	<b>57.83</b>	<b>57.76</b>	<b>57.09</b>	<b>57.20</b>



<표 6> GRU를 활용한 주식시장 예측 검증 정확도(%)

종목코드	$l$									
	1	2	3	4	5	6	7	8	9	10
000250	51.10	58.11	63.12	56.62	55.81	52.67	54.68	56.50	55.63	55.71
003380	64.92	56.40	57.55	60.73	71.02	59.74	64.91	62.00	52.76	58.08
003670	58.99	62.99	54.44	62.09	56.61	61.95	58.19	61.21	69.79	58.47
007390	58.90	69.17	68.97	74.44	64.86	53.62	68.11	60.38	64.16	52.01
022100	64.07	67.39	64.63	70.02	64.67	66.86	64.99	58.14	60.27	60.31
025980	53.60	61.47	78.08	67.33	75.64	73.49	73.95	72.35	76.86	73.00
028150	53.14	52.39	54.81	57.35	51.18	53.55	54.15	54.97	53.71	54.22
028300	62.51	64.85	65.61	62.17	59.02	62.58	67.65	67.51	65.40	58.91
030190	57.95	63.84	55.94	59.28	59.93	53.41	54.66	58.69	58.24	62.10
034230	63.02	65.22	67.32	56.39	50.00	52.66	56.28	57.35	70.90	65.55
035760	64.05	60.46	57.94	59.31	57.13	59.60	56.07	55.99	56.05	59.90
035900	56.81	51.00	58.41	53.45	51.22	55.14	65.40	51.87	58.07	58.78
036420	70.11	67.31	58.04	67.78	68.81	69.81	59.10	62.25	64.91	64.77
036490	53.35	59.04	56.83	51.65	55.25	59.29	50.88	52.24	50.86	59.68
036830	59.81	57.44	52.64	58.95	55.67	53.98	58.84	62.68	50.83	57.65
038540	59.44	57.94	55.94	57.29	58.16	57.06	58.12	61.37	58.13	60.14
041510	58.11	62.16	63.95	61.74	65.00	64.28	59.05	61.84	66.89	56.98
041960	59.87	54.72	67.30	59.00	60.23	58.30	55.82	51.55	61.93	55.97
042000	50.76	56.84	53.21	55.69	56.20	50.58	57.26	54.78	55.68	57.45
046890	53.21	66.96	58.55	73.39	62.72	57.01	57.63	51.81	70.92	58.40
048260	60.32	57.35	57.68	59.30	51.78	57.66	50.56	53.15	55.76	53.47
056190	57.34	55.53	55.07	52.26	58.60	50.60	54.74	55.05	62.73	59.73
058470	60.00	55.20	54.82	56.37	52.87	58.27	53.85	58.37	55.80	52.21
066970	56.19	61.34	65.92	61.64	56.39	54.02	63.04	62.05	63.62	60.31
068760	60.89	59.02	64.23	59.99	58.38	61.09	64.21	64.55	59.54	58.31
069080	55.06	52.90	52.44	58.68	50.20	62.99	52.25	60.27	65.83	63.99
073070	55.93	56.39	57.18	55.22	53.73	58.20	53.45	60.10	53.51	53.77
078340	57.39	61.07	53.24	53.23	57.51	54.76	57.36	60.73	55.29	58.78
083790	58.92	51.92	58.95	69.06	60.28	55.50	58.06	62.57	59.75	71.15
084990	51.05	52.76	53.21	56.57	51.14	56.82	57.54	51.70	50.70	51.24
086520	64.18	64.88	68.93	52.22	56.89	66.26	57.51	58.52	55.00	66.34
086900	52.94	50.88	51.57	51.26	50.71	54.37	52.94	53.76	51.98	51.65
091990	57.59	56.00	59.29	55.69	62.85	58.28	63.25	53.27	63.56	52.48
092040	56.45	55.89	58.92	54.04	57.55	51.59	52.99	59.28	58.81	64.16
095700	60.74	61.53	54.36	58.38	58.79	57.79	59.43	56.19	58.89	52.74
098460	53.67	61.82	52.60	52.39	59.21	52.24	54.22	58.96	58.03	52.31
102940	60.14	61.38	57.45	58.68	57.03	54.33	57.54	55.92	57.57	56.04
122870	56.83	63.62	55.57	52.49	65.03	65.64	52.72	53.54	58.19	53.73
140410	53.15	51.42	50.57	52.61	52.17	53.98	53.33	50.94	52.64	52.61
145020	51.38	53.68	50.95	53.17	54.67	49.26	52.62	54.53	54.42	52.21
178920	55.10	57.92	51.44	58.58	54.67	67.48	69.98	52.61	55.93	51.39
183490	57.03	58.19	51.95	54.75	55.89	56.87	53.83	55.17	59.14	56.83
192080	55.00	57.67	61.32	62.66	63.19	59.57	56.04	65.86	60.67	63.18
200230	57.73	61.79	67.28	60.80	57.52	65.94	64.92	64.62	71.29	58.66
215600	56.20	74.83	57.25	62.19	60.32	58.72	70.08	50.47	60.31	62.60
240810	75.26	56.97	70.34	62.07	59.63	66.53	67.55	65.37	67.59	63.18
253450	61.39	61.12	61.63	63.44	64.38	58.83	62.42	59.05	59.55	65.57
263750	51.02	55.18	59.04	54.56	54.67	55.78	55.27	57.56	54.38	52.70
267980	60.34	58.52	57.36	54.76	58.64	55.49	56.78	59.66	54.75	56.11
평균	<b>58.02</b>	<b>59.23</b>	<b>58.85</b>	<b>58.81</b>	<b>58.24</b>	<b>58.25</b>	<b>58.66</b>	<b>58.15</b>	<b>59.54</b>	<b>58.28</b>

결과를 살펴보면, 주식시장 예측모형의 정확도는 RNN의 경우 평균 59.10%, LSTM의 경우 57.59%, GRU의 경우 58.60%로 나타났으며, 3개의 모형(046890 종목 LSTM ( $l=4$ ), 183490 종목 LSTM ( $l=6$ ), 145020 종목 GRU( $l=6$ ))을 제외한 나머지 예측모형은 모두 50%이상의 정확도를 보였다. 각 모형 별 정확도를 비교해보면, 장기 의존성 문제의 발생 가능성이 대두되던 RNN의 평균 정확도가 가장 높게 나타났는데 RNN이 주로 활용되는 분야가 자연어 처리 등의 분야인 점을 고려하면 본 연구에서 사용한 입력 데이터의 시계열 길이( $t$ ) 10의 수준에서는 오히려 선택적으로 장기기억과 단기기억을 학습하는 LSTM이나 GRU에 비해 전체의 맥락을 학습하는 RNN이 효과적일 수 있음을 시사한다.

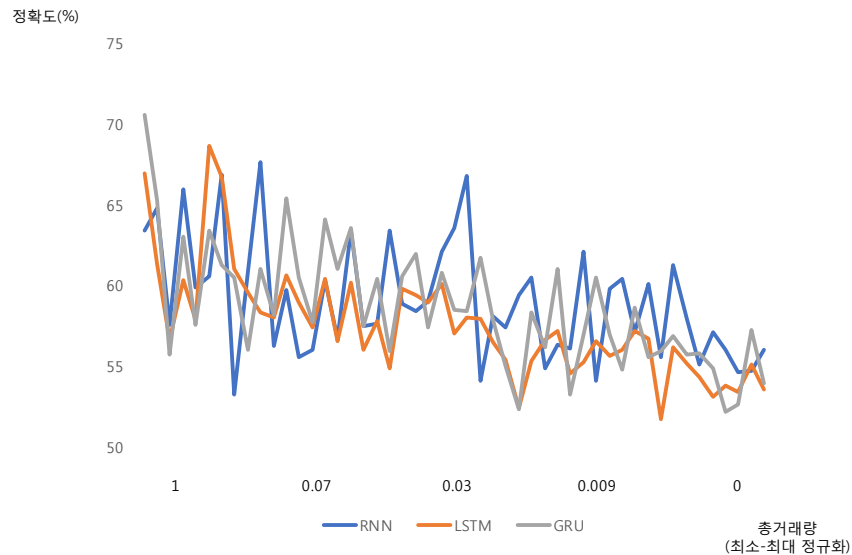
본 연구에서는 호가 정보가 주가에 지연되어 반영될 수 있다는 점(이우백 & 최혁, 2007)을 고려하여 최적의 매매 타이밍( $l=[1,2,3,...,10]$ )에 대한 탐색을 수행하였지만, 모든 종목에 있어서 일관되게 성능이 우수한 매매 타이밍은 나타나지 않았다. 이는 주식시장에 참여하는 투자자들의 매매행태와 각 종목들에 내재되어 있는 특징들이 모두 상이하기 때문에 나타난 결과로 사료된다.

또한 각 종목의 거래량 및 유동성 등 호가 정보의 가격 발견 공헌도에 영향을 미칠 수 있는 특징(이우백 & 최혁, 2006a; 우민철 & 김지현, 2014)을 고려하여 정확도를 비교하기 위해 종목 별 거래량 정보를 활용하였으며, 자세한 내용은 아래 <표 7>과 같다.

〈표 7〉 종목 별 거래량 정보

종목코드	학습구간 거래량	검증구간 거래량	총거래량	총거래량 (min-max)
000250	4,849,822	1,555,231	6,405,053	0.043267
003380	1,720,812	2,205,321	3,926,133	0.032866
003670	6,427,826	3,362,022	9,789,848	0.072792
007390	10,666,027	4,956,290	15,622,317	0.115028
022100	5,267,943	2,782,288	8,050,231	0.059601
025980	108,831,561	33,678,995	142,510,556	1
028150	161,520	95,431	256,951	0
028300	6,395,091	1,879,767	8,274,858	0.055746
030190	2,852,189	1,379,696	4,231,885	0.029909
034230	8,950,268	5,576,611	14,526,879	0.112326
035760	660,674	483,276	1,143,950	0.00725
035900	10,875,488	3,068,647	13,944,135	0.094749
036420	21,634,364	21,428,285	43,062,649	0.364761
036490	471,804	211,811	683,615	0.003088
036830	777,676	257,710	1,035,386	0.00535
038540	1,423,513	1,186,358	2,609,871	0.019585
041510	4,741,786	1,523,427	6,265,213	0.042291
041960	2,997,321	1,455,905	4,453,226	0.031602
042000	928,408	400,359	1,328,767	0.00783
046890	7,230,913	4,468,480	11,699,393	0.089944
048260	597,509	279,043	876,552	0.004568
056190	1,852,763	955,170	2,807,933	0.019397
058470	487,367	230,712	718,079	0.003392
066970	5,589,659	1,339,851	6,929,510	0.045024
068760	1,511,671	757,846	2,269,517	0.015213
069080	3,814,987	1,625,123	5,440,110	0.038177
073070	20,248,821	5,075,374	25,324,195	0.170881
078340	1,061,619	448,882	1,510,501	0.009139
083790	4,837,941	1,557,543	6,395,484	0.043225
084990	1,852,957	552,074	2,405,031	0.014813
086520	6,506,096	1,899,978	8,406,074	0.056607
086900	452,922	155,035	607,957	0.002335
091990	5,845,172	4,669,686	10,514,858	0.084352
092040	1,680,029	448,856	2,128,885	0.012656
095700	1,776,590	1,162,312	2,938,902	0.02132
098460	622,647	492,021	1,114,668	0.007133
102940	7,066,863	7,379,431	14,446,294	0.122121
122870	7,340,173	1,963,390	9,303,563	0.062072
140410	1,926,923	937,799	2,864,722	0.019621
145020	248,622	179,153	427,775	0.001448
178920	4,082,253	2,538,093	6,620,346	0.050081
183490	776,616	359,268	1,135,884	0.006499
192080	788,714	650,715	1,439,429	0.009883
200230	13,624,593	5,419,454	19,044,047	0.137122
215600	12,739,021	3,746,819	16,485,840	0.113061
240810	6,548,019	3,867,984	10,416,003	0.07923
253450	1,836,988	1,453,360	3,290,348	0.024974
263750	2,573,172	638,887	3,212,059	0.019897
267980	295,060	146,188	441,248	0.001337

<표 7>의 거래량 정보를 토대로 거래량 별로 각 순환신경망 모형의 정확도를 비교하였으며, 결과는 아래 <그림 15>과 같다.



<그림 15> 유동성 별 모형 정확도 비교 결과

<그림 15>를 살펴보면, RNN, LSTM, GRU 모두 거래량이 감소함에 따라 정확도 또한 하락하는 것으로 나타났다. 이는 이우백 & 최혁 (2006a)의 연구에서 주장된 종목의 거래량 규모 및 시간대 등 다양한 요인에 의해 가격 발견 기능의 효율성이 변화된다는 점과 일치하는 결과로 사료된다. 각 모형 별로 정확도를 비교해보면, 총거래량 정규화 값이 0.03 이하인 하위 구간(하위 50%)에서 대체적으로 RNN의 정확도가 가장 준수한 것으로 나타났으며 평균 57.75%의 정확도를 보였다. 이 결과는 거래량이 적어 호가 정보의 변동이 완만한 종목에 대해서는 시계열 전체에 걸쳐 정보를 학습하는 RNN이 가장 효율적이라는 것을 의미한다. 그리고 총거래량 정규화 값이 0.03 이상인 상위 구간(상위 50%)에서는

GRU의 정확도가 대체적으로 가장 우수한 것으로 나타났으며 평균 60.74%의 정확도를 보였다. 이 결과는 거래량이 많아 유동성이 충분히 공급된 종목에 대해서는 입력되는 호가 시계열 정보에 대해 선택적으로 학습하는 것이 전체 시계열에 걸쳐 학습하는 것보다 효율적이라는 점을 의미한다. 또한 거래량이 가장 많은 종목에 대해서 RNN은 63%의 정확도를 보인 반면 LSTM과 GRU가 각각 67%, 70.58%의 정확도를 나타냈다는 점을 미루어볼 때, 유동성이 높은 종목들에 대해서 LSTM과 GRU와 같이 장기 및 단기 기억을 선택적으로 학습하는 모형이 효과적일 것으로 사료된다.

## V. 결론

주식시장은 내·외부적으로 다양한 변수들의 영향을 받고 잡음이 포함되는 경우가 잦아 예측하기 어려운 난제로 여겨졌지만, 과거부터 주식시장을 예측하기 위해 많은 연구들이 수행되어 왔으며, 다양한 방법론들이 제시되었다. 특히 최근 컴퓨터 기술의 발달에 힘입어 딥러닝 분석이 주목을 받고 있으며 주식시장에서도 활용되고 있다. 특히 딥러닝 기법 중 하나인 순환신경망은 시계열 데이터의 맥락 정보를 학습할 수 있다는 점에서 다양한 주가지수, 종목 등의 예측에 활용되고 있으며 그 유용성이 증명되었다. 주식시장 예측과 관련된 이들 대부분의 연구는 예측에 있어서 가격과 관련된 기술적 지표, 혹은 재무 정보 등의 기본적 지표를 활용하였으며 우수한 예측 성능을 보이고 있다.

하지만 주식시장에서 발생하는 데이터 중에는 선행연구들에서 사용한 데이터 이외에도 다양한 변수들이 존재하며 ‘호가 정보’ 또한 그 중 하나이다. 호가 정보는 재무분야의 다양한 선행연구들에 의해 가격에 대해 미치는 영향력과 투자지표로써의 활용 가능성 등 그 중요성이 부각되고 있지만, 주식시장 예측분야에서는 아직 이를 활용한 주식시장 예측 연구가 전무하다. 또한 주식시장 예측과 관련된 대부분의 연구들은 일별 자료를 활용하여 일 단위 주가 예측을 수행하였지만, 최근 주식시장에서 고빈도 매매가 보편화된다는 점과 대부분의 개인투자자들이 단기투자를 선호한다는 점(김성동, 2012)을 고려할 때 주식시장 예측 또한 이런 변화에 대응할 필요가 있다.

따라서 본 연구에서는 호가 정보를 대상으로 순환신경망을 적용하여 주식시장 예측을 수행하였으며 다음과 같은 과정으로 연구가 수행되었다. 첫 번째로 키움증권에서 제공하는 Open API 서비스를 이용하여 주

식시장 호가 데이터를 수집하였다. 두 번째로 순환신경망의 학습 데이터를 생성하기 위해 수집된 데이터를 대상으로 전처리를 수행하였으며 세 번째로 무빙 윈도우 기법을 활용하여 라벨링을 수행하였다. 마지막으로 종목 별로 특징이 상이하다는 점을 고려하여, 입력 변수를 학습과정에서 다른 RNN, LSTM, GRU를 활용하여 최적의 매매 타이밍에 대한 탐색이 수행되었다.

본 연구의 의의는 다음과 같다. 첫 번째로 기존의 연구들과는 다르게 입력변수로 호가 정보를 사용하였으며 호가 정보를 이용하여 주식시장을 예측했다는 점에서 의의를 갖는다. 두 번째로, 본 연구에서 제안하는 예측모형은 호가정보를 지표로 사용하고 있는 투자자들의 객관적인 의사결정에 기여할 것으로 기대된다. 주식시장에서 일부 투자자들은 투자 의사결정에 있어서 호가 정보를 활용하고 있지만, 투자자들이 활용하고 있는 호가 정보 그 자체는 고빈도 매매나 알고리즘 매매에 의해 주문이 체결되어 빠르게 변화되며 주가에 영향을 미치기까지 지연될 가능성도 있으므로, 객관적이고 정확하게 판단하기 어렵다. 하지만 본 연구에서 제안한 예측모형은 상승과 하락패턴에 대한 학습을 기반으로 예측결과를 제공하므로 호가 정보를 활용하는 투자자들의 의사결정을 지원할 수 있을 것으로 사료된다.

본 연구의 한계점 및 향후 연구방향은 다음과 같다. 본 연구에서는 종목 별 특징을 고려하기 위해 유동성을 대표할 수 있는 거래량을 활용하였다. 그 결과 유동성이 높아질수록 호가 정보를 활용한 예측 정확도가 향상되며 특히 시계열 정보를 선택적으로 학습하는 LSTM, GRU가 적합함을 확인하였다. 하지만 호가 정보는 제출된 후 취소가 가능하다는 점에서 유동성(거래량)만으로는 호가 정보의 단계 별 변화량을 모두 대표할 수는 없다. 그러므로 호가 정보의 단계 별 변화량을 모두 대표할

수 있는 분산이 고려될 필요가 있다. 특히, 주식시장에서는 시세 견인·조종 등을 목적으로 허수성 호가가 제출되고 있다는 점을 고려할 때 비슷한 수준의 유동성을 가진 종목들 간에도 분산에서의 차이점이 나타날 것으로 사료된다. 따라서 향후 연구에서 유동성과 단계 별 분산이 모두 고려된다면 예측 성능이 크게 향상될 것으로 기대된다.

두 번째로 본 연구에서는 공개범위에 해당하는 10단계 호가를 모두 활용하여 주식시장을 예측하였다. 하지만 이우백 & 최혁 (2012) 등의 연구에서 제시된 바와 같이 각 단계 별 호가는 가격발견에 기여하는 정도가 다르다는 점을 고려하여 일부 단계의 호가만을 선택적으로 활용하거나, 이우백 & 최혁 (2006b)의 연구에서 시간대별로 호가 정보의 가격 발견 공헌도가 단계 별로 상이하단 점을 고려하여 가중치를 부여하게 된다면, 예측 성능이 향상될 수 있을 것으로 판단된다. 마지막으로 주식시장은 경제 및 정치적 변수를 포함한 다양한 상황에 영향을 받는다는 점을 고려할 때, 호가 정보 이외의 금융시장에 영향을 주는 변수가 추가되어 예측한다면 성능이 더욱 향상될 것으로 사료된다.



## 참고문헌

- [1] 구승환, 장성용. (2012). 기본적 분석방법을 통한 주식 투자 전략에 관한 시뮬레이션 연구. 경영과학, 29(2), 53-64.
- [2] 김선웅, 안현철. (2010). Support Vector Machines 와 유전자 알고리즘을 이용한 지능형 트레이딩 시스템 개발. 지능정보연구, 16(1), 71-92.
- [3] 김성수, 홍광진. (2017). 순환 신경망 기술을 이용한 코스피 200 지수에 대한 예측 모델 개발 및 성능 분석 연구. 한국산업정보학회논문지, 22(6), 23-29.
- [4] 김진백, 김유일. (2000). 인공 신경망의 학습에 있어 가중치 변화방법과 은닉층의 노드수가 예측정확성에 미치는 영향. 정보시스템연구, 9(1), 27-44.
- [5] 김진화, 홍광현, 민진영. (2011). 지식 누적을 이용한 실시간주식시장 예측. 지능정보연구, 17(4), 109-130.
- [6] 김현지, 장우진. (2010). 지수 이동 평균선을 이용한 거래 전략. 한국경영과학회 학술대회논문집, 1124-1130.
- [7] 남주하, 윤기향. (2001). 미국 주식시장에서 한국 주식시장으로의 변동성 이전효과 분석. 국제경제연구, 7(3), 23-46.
- [8] 박재연, 유재필, 신현준. (2016). 기술적 지표와 기계학습을 이용한 KOSPI 주가지수 예측. 정보화연구 (구 정보기술아키텍처연구), 13(2), 331-340.

- [9] 서지혜. (2017). 텐서플로우를 이용한 순환신경망들의 알고리즘 성능 평가. 석사학위논문, 이화여자대학교
- [10] 서재현. (2018). 딥러닝 기반 불균형 침입탐지 데이터 분류에 관한 비교 연구. 한국지능시스템학회 논문지, 28(2), 152-159.
- [11] 성노윤, 남기환. (2017). 온라인 뉴스 및 거시경제 변수를 활용한 주가예측. Entrue Journal of Information Technology, 16(2), 41-54.
- [12] 송경환. (2018). GRU 기계학습 알고리즘을 이용한 델타 상관관계 프리페치 기법. 석사학위논문, 인하대학교
- [13] 송현정, 이석준. (2018). 딥러닝을 활용한 실시간 주식거래에서의 매매 빈도 패턴과 예측 시점에 관한 연구: KOSDAQ 시장을 중심으로. 정보시스템연구, 27(3), 123-140.
- [14] 신동백. (2007). 몬테카를로 시뮬레이션방법을 이용한 환율예측분석. 산업경제연구, 20(5), 2075-2093.
- [15] 안성원, 조성배. (2010). 뉴스 텍스트 마이닝과 시계열 분석을 이용한 주가예측. 한국정보과학회 학술발표논문집, 37(1C), 364-369.
- [16] 엄세용. (2017). 허수성 호가에 대한 시세조종과 교란행위 규제의 평가. 증권법연구, 18(2), 183-221.
- [17] 우민철, 김지현. (2014). 분할호가 제출의 목적에 관한 연구: 위탁매매를 중심으로. 재무관리연구, 31(4), 41-68.
- [18] 우민철, 최혁. (2013). 고빈도거래자의 매매양태 분석. 한국증권학회지, 42(4), 699-732.
- [19] 유진. (2002). 가격제한 존재 시 VaR 추정 방법론: 몬테카를로 시뮬

- 레이션. 선물연구, 10(1), 81-111.
- [20] 이석준, 오경주. (2011). 선물시장의 시스템트레이딩에서 동적시간과  
평 알고리즘을 이용한 최적매매빈도의 탐색 및 거래전략의 개발. 한  
국데이터정보과학회지, 22(2), 255-267.
- [21] 이우백, 최혁. (2006a). 거래 전 정보공개와 가격발견. 한국증권학회  
지, 35(4), 143-190.
- [22] 이우백, 최혁. (2006b). 하루중 유동성 형성 과정과 가격발견. 재무  
연구, 19(2), 1-38.
- [23] 이우백, 최혁. (2007). 공개주문원장 정보의 단기수익률 예측력 분석.  
한국증권학회지, 36(6), 963-1007.
- [24] 이우백, 최혁. (2012). 공개주문집계장 가격발견의 횡단면 분석. 재무  
관리연구, 29(3), 131-176.
- [25] 이우식. (2017). 딥러닝분석과 기술적 분석 지표를 이용한 한국 코스  
피주가지수 방향성 예측. 한국데이터정보과학회지, 28(2), 287-295.
- [26] 이윤선. (2003). 한국증권시장에서 기술적 분석의 경제적 효과-주가  
이동평균선을 중심으로. 금융공학연구, 2(2), 1-21.
- [27] 이재원. (2000). Monte Carlo 알고리즘에 의한 주가 추세의 평가. 한  
국정보과학회 학술발표논문집, 27(1B), 235-237.
- [28] 이형용. (2008). 한국 주가지수 등락 예측을 위한 유전자 알고리즘  
기반 인공지능 예측기법 결합모형. Entrue Journal of Information  
Technology, 7(2), 33-43.
- [29] 정재만, 전용호, 최혁. (2014). 전략적 반복주문을 통한 고빈도 거래

- 가 한국 주식시장에 미치는 영향. 재무연구, 27(2), 177-211.
- [30] 주일택, 최승호. (2018). 양방향 LSTM 순환신경망 기반 주가예측모델. 한국정보전자통신기술학회 논문지, 11(2), 204-208.
- [31] 홍태호, 김은미, 차은정. (2017). 뉴스 감성분석과 SVM 을 이용한 다우존스 지수와 S&P500 지수 예측. 인터넷전자상거래연구, 17(1), 23-36.
- [32] Bengio, Y., Simard, P., Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks, 5(2), 157-166.
- [33] Boehmer, E., Saar, G., Yu, L. (2005). Lifting the veil: An analysis of pre trade transparency at the NYSE. The Journal of Finance, 60(2), 783-815.
- [34] Brogaard, J. (2010). High frequency trading and its impact on market quality. Northwestern University Kellogg School of Management Working Paper, 66.
- [35] Cao, C., Hansch, O., Wang, X. (2009). The information content of an open limit order book. Journal of Futures Markets: Futures, Options, and Other Derivative Products, 29(1), 16-41.
- [36] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
- [37] Chung, J., Gulcehre, C., Cho, K., Bengio, Y. (2014). Empirical

evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

- [38] Elman, J. L. (1990). Finding structure in time. *Cognitive science*, 14(2), 179–211.
- [39] Fang, Y., Yu, B., Tang, Y., Liu, L., Lu, Z., Wang, Y., Yang, Q. (2017, July). A new malware classification approach based on malware dynamic analysis. In *Australasian Conference on Information Security and Privacy* (pp. 173–189). Springer, Cham.
- [40] Graves, A., Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5–6), 602–610.
- [41] Gers, F. A., Schmidhuber, J., Cummins, F. (1999). Learning to forget: Continual prediction with LSTM.
- [42] Güler, N. F., Übeyli, E. D., Güler, I. (2005). Recurrent neural networks employing Lyapunov exponents for EEG signals classification. *Expert systems with applications*, 29(3), 506–514.
- [43] Harris, L. E., Panchapagesan, V. (2005). The information content of the limit order book: evidence from NYSE specialist trading decisions. *Journal of Financial Markets*, 8(1), 25–67.
- [44] Han, K. H., Kim, J. H. (2000). Genetic quantum algorithm and its application to combinatorial optimization problem. In *Proceedings of the 2000 Congress on Evolutionary Computation. CEC00* (Cat. No. 00TH8512) (Vol. 2, pp. 1354–1360). IEEE.

- [45] Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *The journal of Finance*, 50(4), 1175–1199.
- [46] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.
- [47] Huang, C. J., Yang, D. X., Chuang, Y. T. (2008). Application of wrapper approach and composite classifier to the stock trend prediction. *Expert Systems with Applications*, 34(4), 2870–2878.
- [48] Ince, H., Trafalis, T. B. (2008). Short term forecasting with support vector machines and application to stock price prediction. *International Journal of General Systems*, 37(6), 677–687.
- [49] Jang, G. S., Lai, F., Jiang, B. W., Parng, T. M., Chien, L. H. (1993). Intelligent stock trading system with price trend prediction and reversal recognition using dual-module neural networks. *Applied Intelligence*, 3(3), 225–248.
- [50] Kim, J., Le, T. T. H., Kim, H. (2017). Nonintrusive load monitoring based on advanced deep learning and novel signature. *Computational intelligence and neuroscience*, 2017.
- [51] Lai, R. K., Fan, C. Y., Huang, W. H., Chang, P. C. (2009). Evolving and clustering fuzzy decision tree for financial time series data forecasting. *Expert Systems with Applications*, 36(2), 3761–3773.
- [52] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *nature*,

521(7553), 436.

- [53] Lee, S., Enke, D., Kim, Y. (2017). A relative value trading system based on a correlation and rough set analysis for the foreign exchange futures market. *Engineering Applications of Artificial Intelligence*, 61, 47–56.
- [54] Lin, X., Yang, Z., Song, Y. (2009). Short-term stock price prediction based on echo state networks. *Expert systems with applications*, 36(3), 7313–7317.
- [55] Liu, Z., Sullivan, C. J. (2019). Prediction of weather induced background radiation fluctuation with recurrent neural networks. *Radiation Physics and Chemistry*, 155, 275–280.
- [56] McNelis, P. D. (2005). *Neural networks in finance: gaining predictive edge in the market*. Academic Press.
- [57] Nelson, D. M., Pereira, A. C., de Oliveira, R. A. (2017, May). Stock market's price movement prediction with LSTM neural networks. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 1419–1426). IEEE.
- [58] Petrosian, A., Prokhorov, D., Homan, R., Dasheiff, R., Wunsch II, D. (2000). Recurrent neural network based prediction of epileptic seizures in intra-and extracranial EEG. *Neurocomputing*, 30(1-4), 201–218.
- [59] Petrosian, A. A., Prokhorov, D. V., Lajara-Nanson, W., Schiffer, R. B. (2001). Recurrent neural network-based approach for early

- recognition of Alzheimer's disease in EEG. *Clinical Neurophysiology*, 112(8), 1378–1387.
- [60] Roondiwala, M., Patel, H., Varma, S. (2017). Predicting stock prices using LSTM. *International Journal of Science and Research (IJSR)*, 6(4), 1754–1756.
- [61] Shieh, J. S., Chou, C. F., Huang, S. J., Kao, M. C. (2004). Intracranial pressure model in intensive care unit using a simple recurrent neural network through time. *Neurocomputing*, 57, 239–256.
- [62] Shin, S., Sung, W. (2016, May). Dynamic hand gesture recognition for wearable devices with low complexity recurrent neural networks. In *2016 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 2274–2277). IEEE.
- [63] Tian, Y., Zhang, K., Li, J., Lin, X., Yang, B. (2018). LSTM-based traffic flow prediction with missing data. *Neurocomputing*, 318, 297–305.
- [64] Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10), 1550–1560.
- [65] Williams, R. J., Peng, J. (1990). An efficient gradient-based algorithm for on-line training of recurrent network trajectories. *Neural computation*, 2(4), 490–501.
- [66] Zhang, J., Zhang, D., Ma, W., Jiao, L. (2018). Deep Self-Paced Residual Network for Multispectral Images Classification Based



on Feature-Level Fusion. IEEE Geoscience and Remote Sensing Letters, (99), 1-5.

[67] Zheng, C., Wang, S., Liu, Y., Liu, C. (2019). A novel RNN based load modelling method with measurement data in active distribution system. Electric Power Systems Research, 166, 112-124.

[68] Zhuge, Q., Xu, L., Zhang, G. (2017). LSTM Neural Network with Emotional Analysis for Prediction of Stock Price. Engineering Letters, 25(2).