

A Study on Stock Trend Determination in Stock Trend Prediction

Chungsoo Lim*

*Professor, Dept. of Electronic Eng., Korea National University of Transportation, Chungju, Korea

[Abstract]

In this study, we analyze how stock trend determination affects trend prediction accuracy. In stock markets, successful investment requires accurate stock price trend prediction. Therefore, a volume of research has been conducted to improve the trend prediction accuracy. For example, information extracted from SNS (social networking service) and news articles by text mining algorithms is used to enhance the prediction accuracy. Moreover, various machine learning algorithms have been utilized. However, stock trend determination has not been properly analyzed, and conventionally used methods have been employed repeatedly. For this reason, we formulate the trend determination as a moving average-based procedure and analyze its impact on stock trend prediction accuracy. The analysis reveals that trend determination makes prediction accuracy vary as much as 47% and that prediction accuracy is proportional to and inversely proportional to reference window size and target window size, respectively.

▶ **Key words:** Stock investment, Stock trend prediction, Stock trend determination, Machine learning

[요 약]

본 연구에서는 주가 결정 방법이 주가 경향 예측에 미치는 영향을 확인하기 위한 분석을 수행한다. 주식시장에서 성공적인 투자를 위해서는 주가의 상승과 하락을 정확하게 예측하는 것이 큰 도움이 되므로 주가 경향 예측에 관해 많은 연구가 진행되고 있다. 예를 들어 근래에는 SNS나 뉴스의 내용을 텍스트 마이닝을 이용하여 분석하고, 이를 이용한 주가 등락의 예측 방법이 제안되었으며 다양한 기계학습 기법들이 활용되고 있다. 그러나 주가의 경향을 ‘상승’ 또는 ‘하락’으로 결정하는 방법은 제대로 분석된 적 없으며 일반적으로 쓰던 방법을 답습하고 있다. 이에 본 논문에서는 주가 경향 결정 방법을 이동평균을 이용해 일반화하고 주가 경향 결정 방법이 예측 정확도에 미치는 영향을 분석한다. 분석 결과, 다음 날의 주가 경향을 예측하는 경우, 주가 경향 결정 방법에 따라 예측 정확도가 47%까지 차이가 남을 발견하였다. 또한 경향 결정에 사용되는 기준값 윈도우의 크기와 예측의 정확도는 비례 관계이며, 대상값 윈도우의 크기와 정확도는 반비례 관계를 알 수 있었다.

▶ **주제어:** 주식 투자, 주가 경향 예측, 주가 경향 결정, 기계학습

-
- First Author: Chungsoo Lim, Corresponding Author: Chungsoo Lim
 - *Chungsoo Lim (clim@ut.ac.kr), Dept. of Electronic Eng., Korea National University of Transportation
 - Received: 2020. 08. 31, Revised: 2020. 11. 18, Accepted: 2020. 11. 19.

I. Introduction

높아진 생활 수준과 정보통신기술 (information and communication technology)의 발전으로 사람들은 쉽게 주식 거래를 할 수 있게 되었으며, 간편한 투자로 고수익을 얻을 수 있기에 많은 사람이 관심을 가지고 투자에 참여하고 있다 [1]. 그러나 문제는 투자가 항상 성공적이지 않다는 것이다. 즉 주가 변동에 대한 이해가 부족하거나 정보가 부족한 경우, 선부른 투자로 축적한 부를 잃어버릴 수 있다 [2]. 따라서 주식 투자자들이 합리적 투자를 돕는 거래시스템 (trading system)의 필요성이 높아지게 되었다. 그리고 높은 수익을 보장하기 위해서 거래시스템이 갖춰야 할 핵심 기능의 하나는 주가의 경향을 정확하게 예측하는 것이다.

보다 정확한 주가 경향의 예측을 위해 산업계뿐 아니라 학계에서도 활발하게 연구가 진행 중인데, 예측의 가능성이 어느 정도인지는 아직 밝혀진 바 없다 [3]. 연구 초기에는 효율적 시장에서의 주가는 모든 가용한 정보를 반영한다는 효율적 시장 가설 (efficient market hypothesis; EMH)과 주가의 변화는 과거의 어떤 패턴에 독립적으로 움직인다는 랜덤워크 이론 (random walk theory; RWT)에 의하여 주가 예측은 매우 어려우며 예측을 한다고 해도 그 효용성을 높지 않다고 생각되었다 [4]. 그러나 근래에 이와 다른 견해를 주장하는 연구가 발표되고 있다. Kumar et al.은 주가는 완전히 랜덤한 것이 아니라 비선형 (nonlinear)이고 동적 (dynamic)이라고 주장하였고 [5], Lei et al.은 EMH가 적용되지 않는 시장은 존재하며 이 시장에서의 주가는 완전히 랜덤이 아니라 예측이 가능하다고 주장하였다 [6]. 또한 Oztekin et al.은 신흥시장 (emerging market)에서는 상대적으로 적은 거래량, 상대적으로 느린 정보 흐름 (information flow)으로 주가 예측이 가능하다고 하였다 [4]. 또한 Zhang et al.도 신흥시장은 완전히 효율적인 시장이 아니기 때문에 미래의 주가 예측이 가능하다고 주장하였다 [2].

주가예측의 정확도를 높이고 주식거래의 질을 높여 보다 높은 수익을 얻기 위해 주로 두 가지 종류의 분석방법이 사용되고 있다. 첫 번째 방법은 기본 분석 (fundamental analysis)으로 주식의 내재적 (intrinsic) 가치를 분석하여 주가를 예측하는 방법으로, 해당 회사의 재무제표 (financial statements), 산업계과 경제의 상황, 거시경제 (macro-economy) 환경 등을 고려한다. 두 번째 방법은 기술 분석 (technical analysis)으로 주가에 영향을 주는 요인들이 주가, 거래량 등의 값에 집약되어 있다고 보고 이들을 분석하여 패턴을 찾아내고 이를 이용하

여 주가를 예측하는 방법이다. 분석을 위해 주가와 거래량으로부터 여러 가지 기술지표 (technical indices)를 구하여 사용한다.

그러나 기본적인 분석 방법은 세 가지 약점이 있다. 첫 번째는 분석에 사용되는 정보에 오류가 있을 수 있다는 것이고, 두 번째는 사용되는 정보와 주가 사이의 상관관계가 있다는 보장이 없다는 것이며, 세 번째는 시장이 semi-strong form efficient 시장이거나 strong form efficient 시장이라면 EMH에 의해서 기본적인 분석 방법은 수익을 보장할 수 없다는 점이다. 또한 기술적 분석의 약점은 weak form efficient 시장에서도 오직 과거 주가와 거래량만을 고려하며 회사에 대한 정보는 전혀 고려하지 않는 데 있다 [7].

이러한 기본적인 분석과 기술적 분석의 단점에도 불구하고 이 분석법들은 계속 주가 예측에 사용되고 있다. 그 이유는 앞서 기술한대로 EMH가 완벽하게 적용되지 않는 시장이 존재하며, 랜덤 또는 비선형 데이터에서 패턴을 찾아낼 수 있는 기계학습 (machine learning) 기법들이 컴퓨터 기술의 발전으로 그 활용 범위가 넓어졌기 때문이다 [7][8]. 실제로 기본적인 분석, 기술적 분석은 기계학습 기법과 함께 전문 트레이더들에 의해 사용되어지고 있으며, 기술분석과 기계학습 조합이 자주 사용되어지고 있다 [9].

기계학습에는 많은 알고리즘이 있는데 근래에는 특히 딥러닝 (deep learning) 알고리즘을 이용한 연구가 많이 발표되고 있다 [1,10-11]. 이는 딥러닝이 영상인식 [12], 자연어 처리 [13], 그리고 생물정보학 분야 [14]에서 뛰어난 성능을 보였기에 주식 경향 예측에서도 좋은 성능을 보일 것이라고 기대되기 때문이다. 또한 기본 분석과 기술 분석에서 사용되는 지표 및 주가 자체를 특징 (feature)로 사용하여 기계학습 모델을 만드는 경우가 대부분이었는데, 근래에는 뉴스나 SNS에 올라온 글들을 텍스트 마이닝을 사용해 감성분석을 하는 방법 [15], 그리고 Google Trends나 Wikipedia에서 검색된 단어를 이용하는 크라우드소싱 (crowd-sourcing) 방법 [16]이 새롭게 제안되고 있다. 주가의 경향을 예측하는 데는 기존의 이동평균을 이용하는 방법 등이 있지만 기계학습을 이용하여 주가의 경향을 예측하는 것은 사람이 이동평균선을 보고 예측하는 것이 아니라 사람이 관찰할 수 없는 데이터에 내재된 정보를 이용하여 좀 더 예측의 정확도를 높이고자 함이다.

기계학습을 주가의 경향 예측에 이용하기 위해서는 학습과 테스트 데이터를 구성하는 각 특징벡터 (feature vector)의 레이블 (label)을 지정해야 하며, 일반적으로 '상승'과 '하락'이 레이블로 사용된다. 그리고 대부분의 연

구에서 이 레이블은 연속된 두 개의 종가를 비교하여 정해진다. 예를 들어 내일의 경향을 결정하기 위해 오늘의 종가와 내일의 종가를 비교하는데 내일의 종가가 높으면 ‘상승’으로 낮으면 ‘하락’으로 정의한다. 몇몇 연구에서는 연속된 두 종가를 사용하지 않고 이동평균을 사용한 사례도 있지만 하나의 결정 방법만을 선택하여 사용했을 뿐 결정 방법을 일반화하여 분석하지는 못했다. 따라서 본 논문에서는 두 개의 값을 비교하여 경향을 결정하는 방법을 이동평균을 이용하는 방법으로 일반화하고 이동평균을 구하는 기간을 변화시키며 정확도에 미치는 영향을 분석하고자 한다.

II. Motivation

이번 장에서는 본 논문에서 수행하고자 하는 분석에 대한 배경을 자세히 설명한다. Table 1은 주식 경향의 예측을 주제로 하는 논문에서 사용된 경향 결정 방법 (trend determination)을 정리해 놓은 것으로 주식의 경향을 결정할 때 비교되는 두 개의 값을 보여준다. 예를 들어 내일의 경향을 결정할 때 오늘의 주가와 내일의 주가를 비교한다면, 이 경우를 “내일의 주가 (tomorrow price)/오늘의 주가 (today price)”로 표현하였다. 본 논문에서는 앞에 있는 “내일의 주가”는 예측의 대상인 대상값 (target value), 그리고 뒤에 있는 “오늘의 주가”는 예측의 기준이 되는 기준값 (reference value)으로 호칭하도록 한다.

표에서 볼 수 있듯이 많은 논문에서 예측값으로 내일의 주가를, 기준값으로 오늘의 주가를 이용하여 경향을 결정하고 있다. 반면에 [6]에서는 대상값으로 향후 n 일의 주가 평균값 (Avg. price of next n days)을, 그리고 기준값으로 오늘의 주가를 이용하였다. [19]에서 대상값은 향후 n 일의 주가 평균값이고 기준값은 오늘포함 과거 n 일의 주가 평균값 (Avg. price of last n days)이다. [32]에서 대상값은 내일의 주가이고 기준값은 오늘포함 과거 n 일의 주가 평균값이다. 이렇게 대상값과 기준값에 따라서 여러 가지 다른 경향값이 결정될 수 있다.

표에 있는 논문들을 놓고 보면, 기준값으로는 오늘의 주가가 가장 많이 사용되었으며 (77%), 예측값으로는 내일의 주가가 가장 많이 사용되었다 (68%). 이렇게 경향의 결정에 주로 사용되는 대상값과 기준값이 있다는 것은 다양한 경향 결정방법이 시도되지 못했다는 것을 의미한다. 그리고 다른 경향 결정방법이 시도되어도 한 가지 방법만이 선택되어 사용되었을 뿐, 경향이 달라짐으로 인해 생기는 영향이 제대로 분석되지 못했다. 따라서 다른 경향 결정방법

이 사용된 경우, 그 연구 결과를 어떻게 해석해야 하는지 알 수 없다. 더욱이 기준값과 예측값의 선택은 학습과 예측에 직접적인 영향을 주며, 이 선택에 따라 예측의 정확도가 크게 달라질 수도 있다. 따라서 본 논문에서는 주가 경향 결정에 사용되는 기준값과 예측값을 변화시키면서 주가예측 정확도가 받는 영향을 분석하고자 한다.

Table 1. Stock trend determination in recent papers

Ref.	Year	Trend determination (target/reference)
[1]	2019	tomorrow price/today price
[2]	2014	polarity of excess return
[3]	2018	tomorrow price/today price
[4]	2016	tomorrow price/today price
[5]	2016	tomorrow price/today price
[6]	2015	Avg. price of next n days/today price
[7]	2018	pattern-based
[8]	2016	tomorrow price/today price
[9]	2018	tomorrow price/today price
[10]	2019	tomorrow price/today price
[11]	2019	price in 10 min./current price
[17]	2010	Avg. price of next n days/today price
[18]	2010	tomorrow price/today price
[19]	2011	Avg. price of next n days/ Avg. price of last n days
[20]	2011	tomorrow price/today price
[21]	2015	price in n days./today price
[22]	2013	Avg. price of next year/ Avg. price of current year
[23]	2013	tomorrow price/today price
[24]	2014	tomorrow price/today price
[25]	2013	tomorrow price/today price
[26]	2003	tomorrow price/today price
[27]	2013	tomorrow price/today price
[28]	2014	tomorrow price/today price
[29]	2012	price in n days./today price
[30]	2014	tomorrow price/today price
[31]	2000	tomorrow price/today price
[32]	2015	tomorrow price/ Avg. price of last n days
[33]	2010	Avg. price of next quarter/ Avg. price of current quarter
[34]	2009	tomorrow price/today price
[35]	2011	tomorrow price/today price
[36]	2018	tomorrow price/today price

III. Analysis

1. Analysis setting

분석에 사용되는 주식에 대한 정보, 특징벡터 (feature vector), 전처리 (pre-processing), 학습/분류 데이터 분할, 그리고 주가 경향 예측에 사용되는 분류기를 Table 2에 정리하였다.

분석에 사용되는 데이터는 2009년부터 2018년까지 10

년 간 KOSPI200을 비롯하여 삼성전자, 기업은행, 한국전력, 모비스의 주식 데이터를 한국 거래소 웹사이트에서 내려받아 사용하였다. 이 주식들을 선택한 이유는 이 주식들이 투자자들이 주목하는 대기업과 공기업의 주식이기 때문에 이들을 분석의 대상으로 하는 것이 더욱 의미가 있기 때문이다. 데이터 구성의 특징은 ‘상승’과 ‘하락’ 두 경향의 비율을 동일하게 조정한 것이다. 주가 경향 예측은 50%대의 예측 정확도가 나오는 경우가 많은데, 만약 ‘상승’의 비율이 55%이고 ‘하락’의 비율이 45%라면 예측이 ‘상승’으로만 실행되어도 55%의 정확도를 가지는 경우가 생긴다. 이를 방지하기 위하여 두 경향의 비율을 동일하게 조정하였다. 각 주식 데이터의 70%는 학습에 사용하고 나머지 30%는 검증용으로 사용하였다. 특징벡터 (feature vector)는 시가, 종가, 상한가, 하한가, 거래량과 같은 기본 주식 정보에 기술지표를 추가하여 총 27개의 특징으로 구성하였다. 그리고 각 특징별로 값을 -1에서 1 사이로 조정하기 위하여 정규화를 수행하였고 예측을 위한 분류기로 주식 경향 예측에서 많이 사용되는 서포트벡터머신 (support vector machine)을 사용하였다. 서포트벡터머신의 커널함수로는 RBF (radial basis function)을 사용하였다.

Table 2. Analysis setting

Stock data (2009 ~ 2018)	KOSPI200, Samsung electronics, KEPCO, Mobis, Industry bank
Data partitioning	Training: 70%, Validation: 30%
Feature vector (27 features)	Open, close, high, and low prices, volume, stochastic %K, stochastic %D, slow stochastic %D, momentum, ROC, williams' %R, A/D oscillator, disparity5, disparity10, OSCP, CCI, RSI, ROV, EMA, MACD, MFI, MAO, VROC, MAs of {open, close, high, and low prices}
Preprocessing	Min-max normalization (-1 ~ 1)
Classifier	SVM (kernel = RBF)

2. Analysis of stock trend determination

2.1 Formulating stock trend determination

주가의 경향을 예측하기 위해서는 먼저 모든 데이터 벡터들의 경향이 정해져야 하는데 일반적으로 주가의 ‘상승’과 ‘하락’으로 결정된다. 각 벡터의 상승과 하락을 결정하기 위해서는 주로 현재 및 과거의 주가를 기준값 (reference)로, 그리고 미래의 주가를 대상값 (target)으로 놓고 대상값이 기준값보다 더 크면 상승으로, 작으면 하락으로 판정한다. Table 1에 포함된 논문들에서 사용된 주가의 경향 결정 방법을 Table 3에 정리하였다.

Table 3. Reference and target pairs used in trend determination (summary of the papers in Table 1)

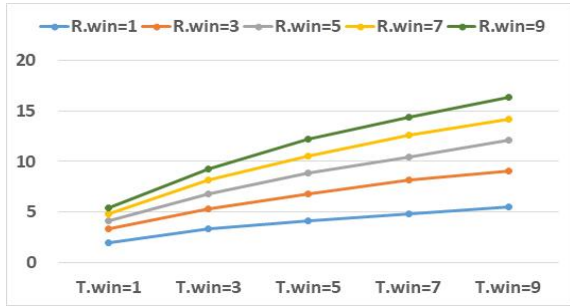
Reference	Target	Ratio
Today's closing price	Tomorrow's closing price	65%
Average closing price of past n days	Average closing price of future m days	10%
Today's closing price	Average closing price of future m days	6%
Average closing price of past n days	Tomorrow's closing price	3%
Etc.		16%

Table 3에서 알 수 있듯이 기준값으로 두 가지 종류의 값들 (오늘의 종가, 과거 n일 평균종가), 그리고 대상값으로도 두 종류의 값들 (내일의 종가, 향후 n일 평균종가)이 주로 사용되었다. 즉 총 네 가지의 조합들이 가능하며, 이 조합들로 조사된 연구의 84%에서 사용한 방법을 나타낼 수 있다. 이를 이용하여 주가 예측 연구에서 사용된 주가의 일별 경향을 결정하는 방법을 일반화하여 표현하면 다음과 같다.

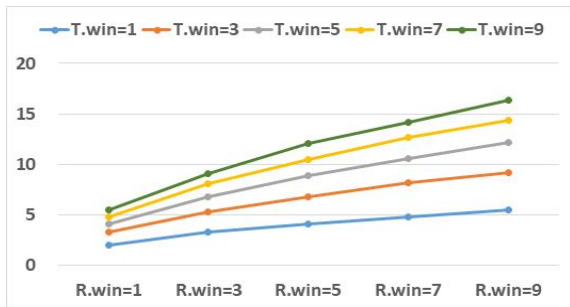
$$\begin{aligned} & \text{if } R.MAn < T.MAm, \text{ trend} = \text{'up'} \\ & \text{if } R.MAn \geq T.MAm, \text{ trend} = \text{'down'} \end{aligned} \quad (1)$$

여기서 $R.MAn$ 은 당일 포함 과거 n일 동안의 주식 종가 이동평균 (moving average)으로 계산되는 기준값 (reference)을 의미하고 $T.MAm$ 은 다음날 포함 m일 동안의 주식 종가 이동평균으로 계산되는 대상값 (target)을 뜻한다. 예를 들어 $R.MA1$ 은 기준값으로 당일의 종가를 사용한다는 의미이며, $T.MA3$ 은 다음날 포함 3일 동안의 주식 종가 이동평균을 대상값으로 사용한다는 의미이다.

이동평균을 구할 때 사용되는 데이터의 개수를 나타내는 n 과 m 이 1보다 큰 값이라면 n 과 m 이 1인 경우와 비교했을 때 기준값과 대상값이 스무딩 (smoothing)된 효과가 있으므로 주식 데이터에 내재되어 있는 노이즈가 감소하는 효과가 있다. 따라서 노이즈에 의해 주가의 경향이 일별로 오르내리는 현상이 감소하며, 동일한 경향이 연속되는 횟수가 증가한다. Fig. 1에 기준값의 이동평균을 구하는데 사용되는 윈도우의 크기 ($R.win$)와 대상값의 이동평균을 구하는데 사용되는 윈도우의 크기 ($T.win$)를 변화시킬 때 동일한 경향이 연속되는 평균 횟수를 나타내었다. 이 값들은 다섯 개 주식의 평균값이며 Fig. 1(a)는 $R.win$ 을 고정시키고 $T.win$ 을 변화시킨 그래프이고 Fig. 1(b)는 $T.win$ 을 고정시키고 $R.win$ 을 변화시킨 그래프이다. 그래프에서 $R.win$ 과 $T.win$ 이 커질수록 연속되는 횟수가 증가하는 것을 볼 수 있다. 또한 고정된 윈도우의 크기가 클수



(a) When the reference window size is fixed and the target window size is varied



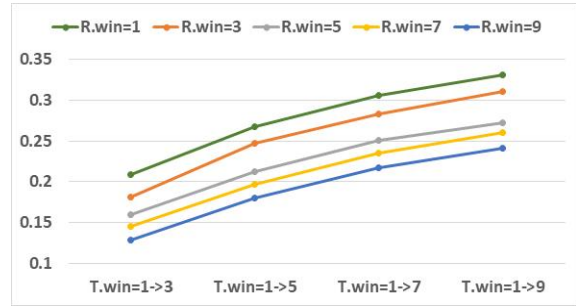
(b) When the target window size is fixed and the reference window size is varied

Fig. 1. Impact of the averaging window sizes on the length of the same trend sequence

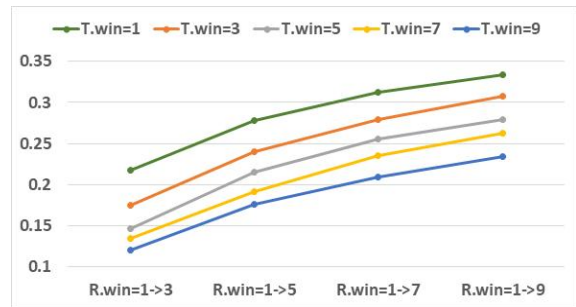
록 변화되는 윈도우 크기에 따른 변화의 기울기가 더 큰 것이 관찰되며, Fig. 1(a)와 Fig. 1(b)의 모양이 비슷하므로 $T.win$ 과 $R.win$ 이 동일한 경향이 연속되는 횟수에 미치는 영향은 비슷한 것으로 보인다.

또한 $R.win$ 과 $T.win$ 은 주식의 경향이 상승에서 하락으로 또는 하락에서 상승으로 전환되는 비율에도 영향을 미친다. Fig. 2는 $R.win$ 과 $T.win$ 이 변할 때 전체 일별 데이터 개수 대비 경향이 전환된 데이터의 비율을 표시한다.

예를들어 $R.win=1$ 인 경우, $T.win=1 \rightarrow 3$ 의 값은 $R.win$ 이 1이고 $T.win$ 이 1일 때와 비교하여 $R.win$ 은 변하지 않고 $T.win$ 이 3으로 증가할 때 경향이 전환되는 데이터 벡터의 개수를 전체 데이터의 개수와의 비율로 나타낸 것이다. 그래프를 보면 윈도우 크기가 증가되는 정도가 커질수록 경향이 전환되는 비율이 증가하는 것을 알 수 있다. 또한, 고정된 윈도우의 크기가 크면 고정된 윈도우의 크기가 상대적으로 작은 경우에 비해 경향 전환의 비율이 작은 것도 관찰된다. 그 이유는 경향 전환의 원인은 스무딩 효과로 인한 노이즈 감소나 급격한 경향 변환의 억제인데, 고정된 윈도우의 크기가 큰 경우는 이미 스무딩 효과를 많이 적용받고 있으므로 변화되는 윈도우의 크기가 증가해도 추가되는 스무딩 효과가 크지 않는 것으로 생각된다.



(a) When the reference window size is fixed and the target window size is varied



(b) When the target window size is fixed and the reference window size is varied

Fig. 2. Impact of the averaging window sizes on the trend conversion ratio

경향의 전환은 학습 데이터와 테스트 데이터 모두에 영향을 주게 된다. 특히 주식의 경향은 상승과 하락이 +1과 -1과 같은 숫자로 표현되어 각 데이터 벡터의 레이블이 되는데, 이렇게 전환된 레이블이 우수한 예측력을 가지는 모델로 이어지는지 다음 절에서 살펴보도록 한다.

2.2 Impact of trend determination on prediction accuracy

III.2.1절에서 본 것과 같이 일별 경향을 결정하는 방법에 있어서 기준값과 대상값을 구하는데 사용되는 윈도우의 크기가 커질수록 스무딩 효과가 강해져서 주식의 동일 경향 연속 횟수가 증가하고 경향 전환 비율이 높아지게 된다. 이러한 주가의 일별 경향의 변화는 학습과 분류 데이터의 레이블에 반영이 되어 모델 구축과 분류 결과에 영향을 주게 된다. 그러나 지금까지의 연구에서는 하나의 경향 결정방법을 선택하여 사용하였으며, 대부분의 경우 오늘의 종가를 기준값으로 내일의 종가를 대상값으로 사용하는 방법이 채택되었다. 따라서 본 절에서는 기준값을 구하기 위한 윈도우와 대상값을 구하기 위한 윈도우를 변화시키며 주가 예측 정확도가 받는 영향을 측정해 본다.

Fig. 3는 다섯 가지 주가에 대해 기준값을 계산하기 위한 윈도우 ($R.win$)의 크기는 고정하고, 대상값을 계산하기

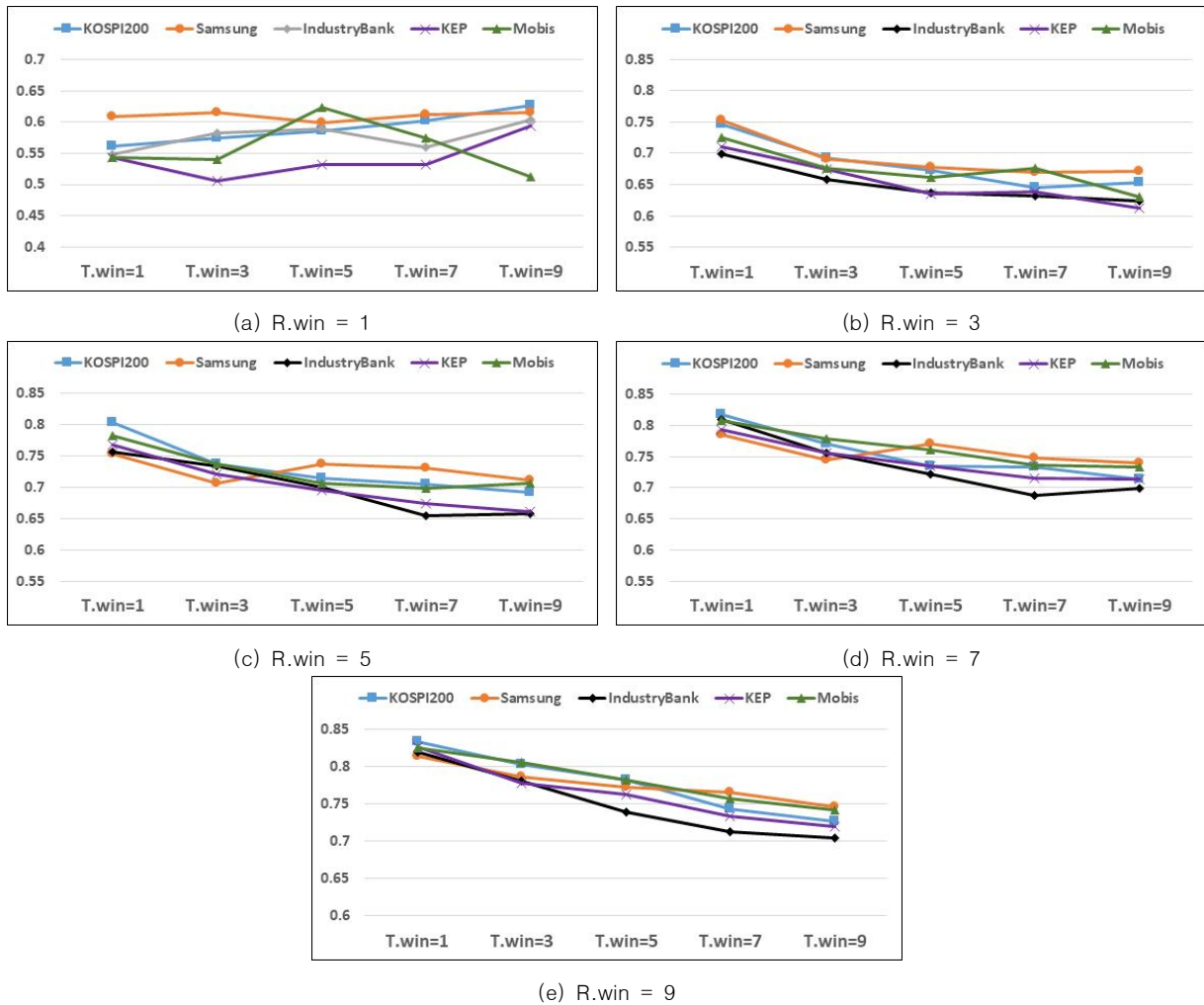


Fig. 3. Trend prediction accuracy when the target window is varied and the reference window is fixed

위한 윈도우 ($T.win$)의 크기를 1, 3, 5, 7, 9로 변화시킬 때 정확도의 변화를 보여주고 있다. 그리고 Fig. 4는 $T.win$ 은 고정하고, $R.win$ 을 1, 3, 5, 7, 9로 변화시킬 때 정확도의 변화를 보여주고 있다. $R.win$ 이 고정되고 $T.win$ 이 변화된 경우 $R.win$ 이 1로 고정되는 경우를 제외하면 $T.win$ 의 크기가 증가할수록 정확도가 완만하게 감소하는 경향이 관찰된다. 즉 $R.win$ 이 결정되었다면, 가장 높은 예측 정확도를 얻기 위해서는 가장 작은 $T.win$ 이 사용되어야 한다. 반대로 $T.win$ 이 고정되고 $R.win$ 이 변화된 경우 $R.win$ 이 증가할수록 정확도가 증가하는 것이 관찰된다. 즉 예측하고자 하는 $T.win$ 이 결정되었다면, 가장 높은 예측 정확도를 얻기 위해서는 가장 큰 $R.win$ 을 이용해야 한다. 종합하면 $R.win$ 이 가장 크고 $T.win$ 이 가장 작은 경우에 가장 높은 정확도를 얻을 수 있다. 따라서 $R.win$ 과 $T.win$ 조합 중 $R.win$ 이 9 (가장 큰 $R.win$)이고 $T.win$ 이 1 (가장 작은 $T.win$)일 때 가장 높은 정확도가 얻어진다. 이는 $T.win$ 이 커지면 더 먼 미래의 값까지 예측의 대상에

포함되는 것이므로 정확도가 떨어지고, $R.win$ 이 커지면 과거의 더 많은 정보를 바탕으로 경향을 정하는 것이므로 정확도가 올라가는 것으로 생각할 수 있다. 이렇게 생각하면 Fig.1(a)에서 $R.win$ 이 1로 고정되고 $T.win$ 이 변화될 때 정확도가 감소하는 패턴이 관찰되지 않는 것이 어느 정도 설명이 된다. 즉 적은 정보 (오늘의 종가)만으로 경향이 결정되기 때문에 정확도가 감소하는 패턴을 보여주지 못하는 것으로 생각된다. 그러나 $R.win$ 과 $T.win$ 이 커질수록 미래와 과거의 정보가 더 이용된다고 하지만 평균을 계산하는 것이며, 정보를 이용하여 단순히 이진값 (+1 또는 -1)을 결정하는 것이므로 정보의 이용율이 높다고 단정할 수는 없다. 또한 SVM 알고리즘이 학습을 통해 모델을 만들기 때문에 사용된 정보가 어떻게 SVM 모델에게 영향을 줬는지 알기 어렵다. 따라서 좀 더 면밀한 분석이 필요하며, 다음 절에서 기준값 윈도우와 대상값 윈도우의 크기가 주가 방향 예측 정확도에 미치는 영향을 경향의 전환 관점에서 분석하려고 한다.

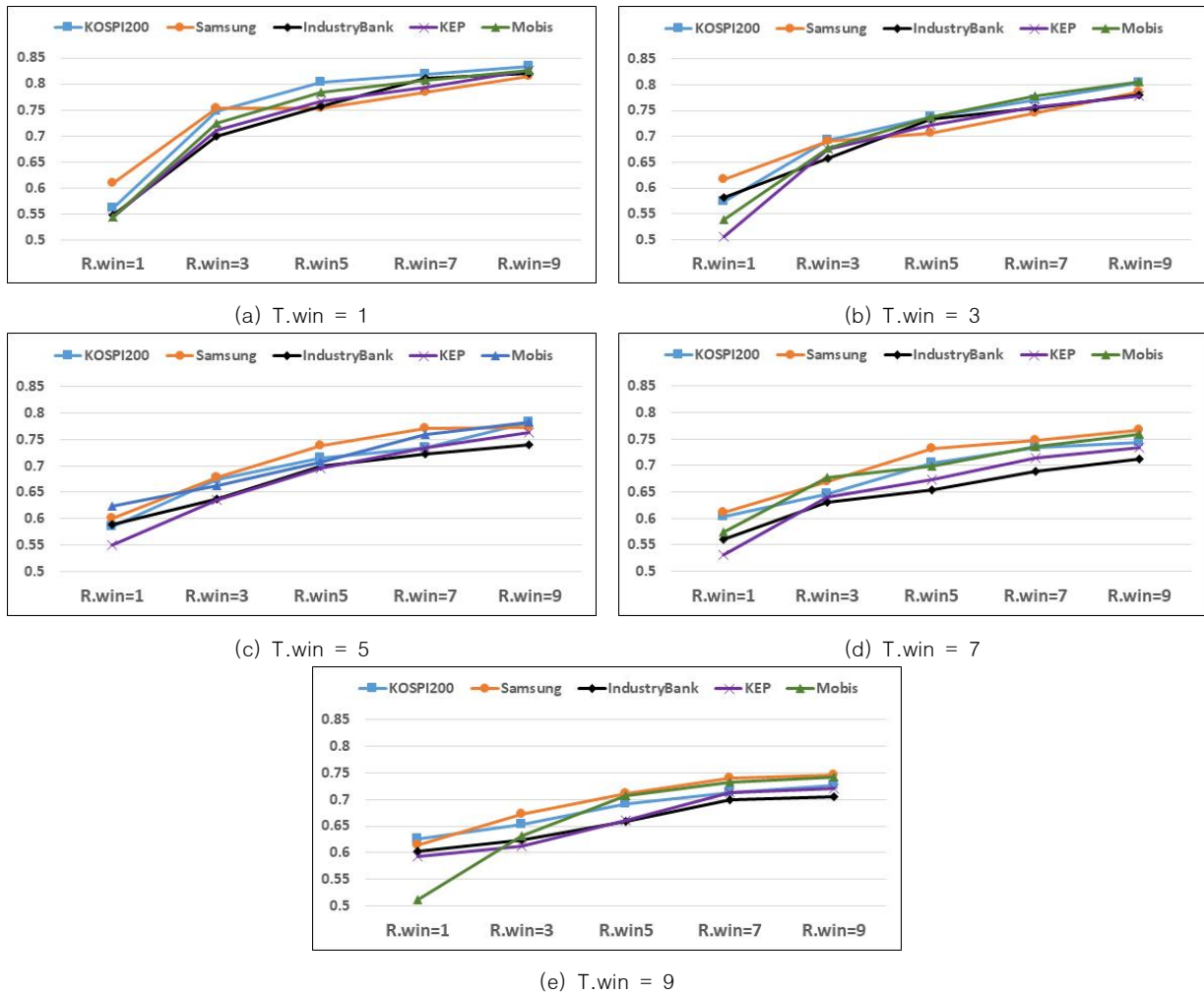


Fig. 4. Trend prediction accuracy when reference window is varied and target window is fixed

2.3 Analyzing prediction accuracy from the perspective of label change

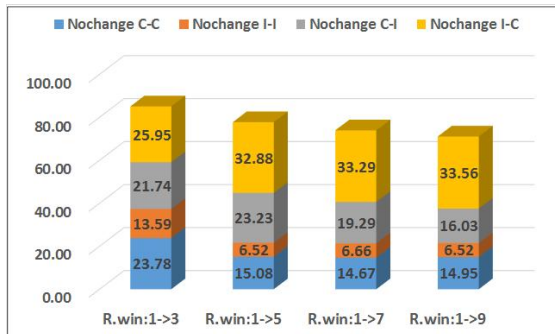
경향 결정방법이 변하면 각 데이터 벡터의 레이블 (경향)이 달라지고, 결과적으로 예측 모델을 만드는 학습과정과 그 모델을 사용하는 예측과정의 정확도가 영향을 받게 된다. 그러나 이 영향에 대해 보다 깊은 이해를 얻기 위해서는 다른 관점에서의 분석이 필요하므로, 레이블의 변화에 따라 데이터 벡터들을 나누어 예측 결과를 분석하고자 한다. 즉 테스트 데이터 벡터를 경향 결정방법이 바뀌어도 레이블이 변하지 않는 데이터 벡터 그룹, 그리고 레이블이 바뀌는 그룹으로 나누고, 기준값과 대상값의 이동평균을 구하는데 사용되는 윈도우의 크기가 변할 때 이 두 가지 그룹에 속한 데이터 벡터들의 예측 성공과 실패를 관찰하여, Fig. 3과 Fig. 4에서 관찰된 패턴의 원인을 찾고자 한다.

Fig. 5는 $R.win$ 과 $T.win$ 이 변할 때 레이블의 전환을 보여준다. Fig. 5(a)는 KEP 주식 데이터를 사용하였으며, $T.win$ 은 5로 고정하였고 $R.win$ 이 증가할 때 레이블의 전

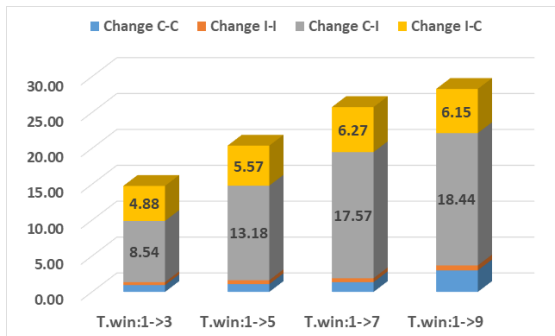
환이 없는 테스트 벡터들에 대한 결과이다. 레이블 전환이 생기는 테스트 벡터들은 정확도의 변화에 기여도가 적기 때문에 나타내지 않았다. Fig. 5(b)는 Industry bank 주식 데이터를 사용하였으며 $R.win$ 을 7로 고정하였고 $T.win$ 이 증가할 때 레이블이 전환된 테스트 벡터들에 대한 결과이다. 역시 레이블이 전환되지 않은 벡터들은 정확도의 변화에 영향을 주지 않기 때문에 그림에 포함시키지 않았다.

그래프의 각 막대마다 이름이 있는데 이는 $R.win$ 의 증가 (Fig. 5(a))나 $T.win$ 의 증가 (Fig. 5(b))를 나타낸다. 예를 들어 $R.win: 1 \rightarrow 3$ 은 $R.win$ 이 1에서 3으로 증가한 경우를 의미하며, $T.win: 1 \rightarrow 5$ 는 $T.win$ 이 1에서 5로 증가한 경우를 뜻한다. 그리고 각 막대는 4가지 종류의 색으로 구성되는데 각 색은 윈도우의 크기가 변하기 전과 후에 예측의 성공 (C로 표시)과 실패 (I로 표시)를 나타낸다. 예를 들어 $R.win: 1 \rightarrow 3$ 의 이름을 가지는 막대에서 $Nochange C-C$ 가 의미하는 것은 레이블이 변하지 않은 벡터 중 $R.win$ 이 1일 때에도 예측이 맞았고, $R.win$ 이 3일 때에도

예측이 맞은 벡터들을 의미한다. 그리고 각 색이 막대에서 차지하는 부피와 각 색에 쓰여있는 값은 전체 데이터 벡터에서 각 색에 속하는 벡터의 비율을 나타낸다. 예를 들어 Fig. 5(a)의 $R.win:1 \rightarrow 3$ 의 이름을 가지는 막대에서 $Nochange\ C-C$ 에 속하는 데이터 벡터들의 전체 데이터 벡터에 대한 비율은 25.95%이다.



(a) Prediction change of test vectors without label conversion as $R.win$ increases and $T.win$ is fixed



(b) Prediction change of test vectors with label conversion as $T.win$ increases and $R.win$ is fixed

Fig. 5. Label conversion-based analysis of prediction change with varying $T.win$ and $R.win$

Fig. 4에서 $T.win$ 이 고정된 상태에서 $R.win$ 이 증가하면 정확도가 증가하는 패턴이 관찰되는데 이는 레이블 전환이 된 테스트 벡터들보다는 Fig. 5(a)에서처럼 레이블 전환이 되지 않은 테스트 벡터들에서 그 이유를 찾을 수 있다. 여기서 $Nochange\ C-C$ 와 $Nochange\ I-I$ 는 $R.win$ 이 증가하기 전과 후 모두 동일한 예측 결과를 나타내는 것이고, $Nochange\ C-I$ 와 $Nochange\ I-C$ 가 윈도우의 증가 전과 후에 예측 결과가 다른 경우이다. 그런데 $Nochange\ C-I$ 보다 $Nochange\ I-C$ 가 더 크며 그 차이가 $R.win$ 이 증가함에 따라 증가함을 볼 수 있고, 이것이 $R.win$ 의 증가에 따라 정확도가 향상되는 이유이다. 좀 더 자세히 살펴보면, $Nochange\ I-C$ 에 속하는 벡터들은 $R.win$ 이 증가될 때 레이블은 변하지 않았지만 학습된 SVM 모델에 의해 예측 값이 바뀌어 결과적으로 맞는 예측이 된 벡터들이다. 즉

$R.win$ 이 증가하여 생긴 일부 학습 데이터 벡터들의 레이블 전환으로 인해 학습 모델이 영향을 받아 $R.win$ 의 증가 전에는 맞게 예측하지 못했던 테스트 데이터 벡터들을 맞게 예측하게 된 것이다.

반면에 Fig. 3에서 $R.win$ 이 고정된 상태에서 $T.win$ 이 증가하면 정확도가 감소하는 패턴이 관찰된다. 분석결과, 이 패턴의 주된 원인은 Fig. 5(b)에서 보여주는 것처럼 $T.win$ 이 증가할 때 레이블이 전환되는 테스트 벡터들 중에 증가되기 전에는 맞게 예측되었으나 증가 후에 예측이 틀리게 되는 벡터들 ($Change\ C-I$)의 수가 증가하는 것이다. 이런 벡터들은 $T.win$ 의 증가 전과 후에 동일하게 예측이 되는 벡터들로 $T.win$ 의 증가로 테스트 벡터의 레이블이 바뀌어 예측이 틀리게 된 것이다. 이는 $T.win$ 의 증가로 일부 학습 데이터의 레이블이 전환되었지만, 이를 반영하는 학습 모델이 만들어지지 않아 기존의 테스트 벡터들이 윈도우 증가 후에도 동일하게 예측되었기 때문으로 생각된다.

IV. Conclusions

주가 경향 예측에서 예측의 목표가 되는 경향을 결정하는 방법은 예측의 정확도에 영향을 미치는 중요한 요소인데, 지금까지의 대부분의 연구에서는 주로 오늘의 주가와 내일의 주가를 가지고 경향을 결정해 왔다. 그러나 경향을 결정하는 방법은 고려하는 기간에 따라 다양해질 수 있으며, 이런 다양한 방법을 실험하고 비교하는 연구는 없었다. 따라서 본 논문에서는 경향 결정 방법을 이동평균을 이용한 기준값 윈도우 크기와 대상값 윈도우 크기의 식으로 일반화하였고, 실험을 통해 경향 결정 방법이 정확도에 미치는 영향을 분석하였다. 분석을 통해 기준값 윈도우가 증가하면 예측의 정확도가 증가하고 대상값 윈도우가 증가하면 정확도가 감소함을 알 수 있었다. 또한 이 예측 정확도에 미치는 영향을 레이블의 전환 관점에서 분석해보았다. 정확도가 향상된 경우는 기준값 윈도우 크기가 증가됨에 따라 학습된 모델이 향상되어 예전에는 제대로 예측하지 못하던 테스트 벡터들을 맞게 예측했기 때문이다. 정확도가 감소한 경우는 대상값 윈도우의 크기가 증가됨에 따라 레이블이 바뀐 테스트 벡터들이 생겨나고, 이 레이블의 전환 때문에 대상값 윈도우가 증가하기 전에는 맞게 예측이 되었지만 증가된 후에는 틀리게 예측되는 벡터들이 증가되기 때문이다.

이 분석을 통해 밝혀진 것은 다음과 같다. 첫째, 주가 경향 예측에 사용되는 분류기나 특징벡터 등은 변경하지

않고, 오직 주가 경향 결정 방법만을 조정하여 주가 경향 예측의 정확도를 향상시킬 수 있다. 즉 예측 대상 기간이 정해지면, 예측 경향 정확도 향상을 위해 경향 결정 방법에서의 기준값을 어떻게 설정해야 하는지 알게 되었다. 둘째, 서로 다른 주가 경향 결정 방법을 사용한 예측 모델들을 비교하면 확연한 성능 차이가 날 수 있다. 이 경우에 주가 경향 결정 방법에서 오는 영향을 감안하여 모델들을 비교 및 해석할 수 있다.

그러나 예측의 정확도는 높아지지만, 기준값이 달라지는 것이므로 이 향상된 예측 능력이 어떻게 활용될 수 있는지는 향후 연구로 진행할 예정이다. 예를 들어 보다 높은 수익률을 위해 활용한다면, 일단 변경된 경향 결정 방법이 주식 거래와 수익률에 미치는 영향을 분석하고, 이를 바탕으로 보다 높은 수익률을 얻는 활용법을 개발할 계획이다.

REFERENCES

- [1] H. S. Sim, H. I. Kim, and J. J. Ahn, "Is deep learning for image recognition applicable to stock market prediction?," *Complexity*, Vol.2019, pp. 1-10, Feb. 2019. DOI: 10.1155/2019/4324878
- [2] X. Zhang, Y. Hu, K. Xie, S. Wang, E. Ngai, and M. Liu, "A causal feature selection algorithm for stock prediction modeling," *Neurocomputing*, Vol. 142, pp. 48-59, Oct. 2014. DOI: 10.1016/j.neucom.2014.01.057
- [3] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, and P. Yu, "Improving stock market prediction via heterogeneous information fusion," *Knowledge-Based Systems*, Vol. 143, pp. 236-247, March 2018. DOI: 10.1016/j.knsys.2017.12.025
- [4] A. Oztekin, R. Kizilaslan, S. Freund, and A. Iseri, "A data analytic approach to forecasting daily stock returns in an emerging market," *European Journal of Operational Research*, Vol. 253, pp. 697-710, Sept. 2016. DOI: 10.1016/j.ejor.2016.02.056
- [5] D. Kumar, S. S. Meghwani, and M. Thakur, "Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets," *Journal of Computational Science*, Vol. 17, pp. 1-13, Nov. 2016. DOI: 10.1016/j.jocs.2016.07.006
- [6] Z. Lei and W. Lin, "Price trend prediction of stock market using outlier data mining algorithm," *Proceedings of IEEE International Conference on Big Data and Cloud Computing*, pp. 93-98, 2015. DOI: 10.1109/bdcloud.2015.19
- [7] J. Zhang, S. Cui, Y. Xu, Q. Li, and T. Li, "A novel data-driven stock price trend prediction system," *Expert Systems with Applications*, Vol. 97, pp. 60-69, May 2018. DOI: 10.1016/j.eswa.2017.12.026
- [8] W. Chiang, D. ENke, T. Wu, and R. Wang, "An adaptive stock index trading decision support system," *Expert Systems with Applications*, Vol. 59, pp. 195-207, Oct. 2016. DOI: 10.1016/j.eswa.2016.04.025
- [9] Y. Chen and Y. Hao, "Integrating principle component analysis and weighted support vector machine for stock trading signals prediction," *Neurocomputing*, Vol. 321, pp. 381-402, Dec. 2018. DOI: 10.1016/j.neucom.2018.08.077
- [10] E. Hoseinzade and S. Haratizadeh, "CNNpred: CNN-based stock market prediction using a diverse set of variables," *Expert Systems with Applications*, Vol. 129, pp. 273-285, Sept. 2019. DOI: 10.1016/j.eswa.2019.03.029
- [11] W. Long, Z. Lu, and L. Cui, "Deep learning-based feature engineering for stock price movement prediction," *Knowledge-Based Systems*, Vol. 164, pp. 163-173, Jan. 2019. DOI: 10.1016/j.knsys.2018.10.034
- [12] R. Socher, B. Huval, B. Bhat, C. Manning, and A. Ng, "Convolutional-recursive deep learning for 3D object classification," *Proceedings of International Conference on Neural Information Processing Systems*, pp. 656-664, 2012.
- [13] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6645-6649, 2013. DOI: 10.1109/icassp.2013.6638947
- [14] D. Chicco, P. Sadowski, and P. Baldi, "Deep autoencoder neural networks for gene ontology annotation predictions," *Proceedings of ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 533-540, 2014. DOI: 10.1145/2649387.2649442
- [15] A. Sun, M. Lachanski, and F. Fabozzi, "Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction," *International Review of Financial Analysis*, Vol. 48, pp. 272-281, Dec. 2016. DOI: 10.1016/j.irfa.2016.10.009
- [16] H. Hu, L. Tang, S. Zhang, and H. Wang, "Predicting the direction of stock markets using optimized neural networks with Google Trends," *Neurocomputing*, Vol. 285, pp. 188-195, April 2018. DOI: 10.1016/j.neucom.2018.01.038
- [17] G. Park and H. Shin, "Stock price forecasting using semi-supervised learning," *Proceedings of KORMS Conference*, pp. 110-116, Oct. 2010.
- [18] S. Ahn and S. Cho, "Stock prediction using news text mining and time series analysis," *Proceedings of KIISE Conference*, pp. 364-369, June 2010.
- [19] D. Shin and K. Jung, "Forecasting short-term KOSPI using wavelet transforms and fuzzy neural network," *Journal of the Korea Contents Association*, Vol. 11, No. 6, pp. 1-7, June 2011. DOI: 10.5392/JKCA.2011.11.6.001
- [20] J. Jin and J. Min, "A real-time stock market prediction using

- knowledge accumulation,” *Journal of Intelligent Information Systems*, Vol. 17, No. 4, pp. 109-130, Dec. 2011. DOI: 10.13088/jiis.2011.17.4.109
- [21] J. Huh and J. Yang, “SVM based stock price forecasting using financial statements,” *KIISE Transactions on Computing Practices*, Vol. 21, No. 3, pp. 167-172, March 2015.
- [22] Z. Hu, J. Zhu, and K. Tse, “Stocks market prediction using support vector machine,” *Proceedings of International Conference on Information Management, Innovation Management and Industrial Engineering*, pp. 115-118, Nov. 2013. DOI: 10.1109/iciiii.2013.6703096
- [23] G. Dong, K. Fataliyev, and L. Wang, “One-step and multi-step ahead stock prediction using backpropagation neural networks,” *Proceedings of International Conference on Information, Communicatio, and Signal Processing*, pp. 1-5, 2013. DOI: 10.1109/iciics.2013.6782784
- [24] F. Wang, Z. Zhao, X. Li, and H. Zhang, “Stock volatility prediction using multi-kernel learning based extreme learning machine,” *Proceedings of International Joint Conference on Neural Networks*, pp. 3078-3085, 2014. DOI: 10.1109/ijcnn.2014.6889651
- [25] Y. Xu, Z. Li, and L. Luo, “A study on feature selection for the trend prediction of stock trading price,” *Proceedings of International Conference on Computational and Information Sciences*, pp. 579-582, 2013. DOI: 10.1109/iccis.2013.160
- [26] K. Kim, “Financial time series forecasting using support vector machine,” *Neurocomputing*, Vol. 55, pp. 307-319, Sept. 2003. DOI: 10.1016/s0925-2312(03)00372-2
- [27] Y. Lin, H. Guo, and J. Hu, “An SVM-based approach for stock market trend prediction,” *Proceedings of International Joint Conference on Neural Networks*, pp. 1-7, 2013. DOI: 10.1109/ijcnn.2013.6706743
- [28] A. K. Sirohi, P. K. Mahato, and V. Attar, “Multiple kernel learning for stock price direction prediction,” *Proceedings of IEEE International Conference on Advances in Engineering & Technology Research*, pp. 1-4, 2014. DOI: 10.1109/icaetr.2014.7012901
- [29] D. Kato and T. Nagao, “Stock prediction using multiple time series of stock prices and news articles,” *Proceedings of IEEE Symposium on Computers & Informatics*, pp. 11-16, 2012. DOI: 10.1109/isci.2012.6222659
- [30] Y. Luo, J. Hu, X. Wei, D. Fang, and H. Shao, “Stock trends prediction based on hypergraph modeling clustering algorithm,” *Proceedings of IEEE International Conference on Progress in Informatics and Computing*, pp. 27-31, 2014. DOI: 10.1109/pic.2014.6972289
- [31] K. Kim and I. Han, “Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index,” *Expert Systems with Applications*, Vol. 19, pp. 125-132, Aug. 2000. DOI: 10.1016/s0957-4174(00)00027-0
- [32] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, “Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques,” *Expert Systems with Applications*, Vol. 42, pp. 259-268, Jan. 2015. DOI: 10.1016/j.eswa.2014.07.040
- [33] C. Tsai and Y. Hsiao, “Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches,” *Decision Support Systems*, Vol. 50, pp. 258-269, Dec. 2010. DOI: 10.1016/j.dss.2010.08.028
- [34] M. Lee, “Using support vector machine with hybrid feature selection method to the stock trend prediction,” *Expert Systems with Applications*, Vol. 36, pp. 10896-10904, Oct. 2009. DOI: 10.1016/j.eswa.2009.02.038
- [35] L. Ni, Z. Ni, and Y. Gao, “Stock trend prediction based on fractal feature selection and support vector machine,” *Expert System with Applications*, Vol. 38, pp. 5569-5576, May 2011. DOI: 10.1016/j.eswa.2010.10.079
- [36] A. N. Kia, S. Haratizadeh, and S. B. Shouraki, “A hybrid supervised semi-supervised graph-based model to predict one-day ahead movement of global stock markets and commodity prices,” *Expert Systems with Applications*, Vol. 105, pp. 159-173, Sept. 2018. DOI: 10.1016/j.eswa.2018.03.037

Authors



Chungsoo Lim received the M.S. degree in electrical engineering from University of Maryland at College Park in 2004 and the Ph.D. degree in computer engineering from North Carolina State University in 2009, respectively.

Since October 2013, he has been with the Department of Electronic Engineering, Korea National University of Transportation, Chungju, Korea, where he is currently an associate professor. His research interests include stock trend prediction, biometric verification, and machine learning accelerator.