

인공신경망 모형을 이용한 투자신호 추정 및 분석: 중국 SSE50 지수를 중심으로*

임 결** · 박성용***

〈요 약〉

본 연구에서는 인공신경망 모형을 사용한 투자신호 분석방법을 제안하고 이를 기반으로 하는 투자 전략을 제안하고자 한다. 주가 변화에 따른 투자신호 분석을 위하여 2006년 6월 1일부터 2019년 1월 4일까지의 일별 중국 상하이주가지수와 바이두지수 자료를 사용하였다. 또한 이동평균수렴확산(Moving Average Convergence and Divergence)지표를 사용하여 주식시장의 강세 시장과 약세 시장에 따라 투자전략을 차별화하였으며, 투자 기회를 판단할 수 있는 신호를 가장 낮은 0번부터 5번까지 여섯 개 신호로 구분하여 보다 세분화된 투자전략 모형을 제안하였다. 또한 인터넷 검색지수 빅 데이터를 인공신경망 모형의 독립변수로 사용하여 추정한 결과 실제 신호와 유사한 신호를 예측할 수 있다는 것을 확인하였다. 2019년 1월 4일부터 3월 24일까지의 기간을 네 구간으로 나누어 투자신호를 예측한 결과 세분화된 신호에 따른 투자전략은 기존의 투자전략보다 미래의 투자 기회를 더 잘 예측할 수 있으며, 무의미한 거래를 감소시키는 효과가 있다는 것을 확인하였다.

주제어 : 투자신호 예측, 인공신경망, 바이두 지수, 이동평균수렴확산지수, 빅 데이터

논문접수일 : 2019년 11월 27일 논문수정일 : 2020년 01월 17일 논문게재확정일 : 2020년 02월 13일

* 이 논문은 2019년도 중앙대학교 연구장학기금 지원을 받아 수행되었습니다.

** 제1저자, 중앙대학교 경제학부 대학원생, E-mail: lingjie993@cau.ac.kr

*** 교신저자, 중앙대학교 경제학부 부교수, E-mail: sungpark@cau.ac.kr

I. 서 론

주식투자에 있어서 자산을 매수해야 할 시점을 정하는 일은 중요한 투자전략 중 하나이다. 주식 가격에는 투자를 위한 많은 정보가 포함되어 있으며, 따라서 주가 변화를 정확하게 분석하는 것은 투자전략의 기본요건이 된다. 최근 대용량 자료를 금융 시계열 분석에 사용할 수 있는 다양한 모형이 제안되고 있는데, 본 연구는 이러한 빅 데이터를 고려한 인공지능망 모형을 사용하여 기간별 주가 변화에 따른 투자신호를 예측하고 이에 따라 자산매입 시점을 정하기 위한 투자전략 모형을 제안하고자 한다. 본 연구에서는 중국 상하이 주식(SSE: Shanghai Stock Exchange) 지수를 사용하여 분석하였다.

본 연구에서는 기술분석에 기초하여 도출되는 기존의 투자신호의 과도매매 문제를 개선하고 보완할 수 있는 새로운 유형의 투자신호 시스템을 개발하고자 한다. 이 투자신호 시스템에는 투자신호에 대한 정의와 투자신호에 대한 예측이 포함된다. 따라서 고안된 모형은 투자자가 투자 의사 결정을 할 때 중요한 참고로 사용할 수 있으며, 이 시스템을 사용함으로써 특정 기간의 시장 투자 기회를 예측할 수도 있다.

본 연구는 다음과 같은 네 가지 점에서 선행연구들과 차별점이 있다. 첫째, 기존의 두 개로 주어지는 투자신호를 여섯 개의 투자신호로 세분화함으로써 기존의 모형에서 나타난 과도매매의 문제점을 해결하고자 하였다. 기존 연구들에서는 주로 기술분석을 기반으로 투자신호를 제작하는 것을 연구하고 있지만, 이들 연구에서는 일정 시간 동안 많은 투자신호를 방출하고 높은 거래 비용을 발생시킨다는 문제점이 공통으로 나타난다. 또한 기존의 모형에서 투자신호는 매수 여부가 0과 1에 해당하는 이항변수만을 방출할 수 있기 때문에 다양한 위험 감수 능력을 가진 투자자가 이 신호를 적절히 사용하는 데는 어려움이 있을 수 있다. 본 연구에서는 Appel and Hirschler(1990)이 제시한 이동평균 수렴 확산(MACD: Moving Average Convergence and Divergence)지표에 시장의 과열 및 침체 문제를 모두 고려하여 여섯 개 신호(Sig 0, Sig 1, Sig 2, Sig 3, Sig 4, Sig 5) 등급의 신호(관망, 신중추천, 추천, 강력추천, 매수, 매수)를 만들었다.

둘째, 본 연구에서 정의된 투자신호에 기초하여 특정 기간의 투자신호를 다양한 설명 변수들을 이용하여 예측한다. 투자전략에 대한 기존의 연구들은 주로 미래 주가의 추세 또는 미래 주가의 상승 및 하락 가능성을 다루고 있으나, 미래 투자 기회에 대한 매수 신호를 예측하는 연구는 많이 연구되지 않았다. 따라서 투자자들에게는 주가의 추세 예측 및 등락 가능성에 대한 예측보다는 투자 기회에 대한 매수 신호 예측이 투자의사 결정에 있어 더 직관적으로 도움이 될 수 있을 것으로 기대할 수 있다.

셋째, 본 연구에서는 투자신호를 분석함에 있어서 인공신경망 모형을 이용하였다. 많은 선행연구에서는 종속변수와 독립변수의 함수적 관계에 대한 연구가 이뤄져 왔지만 본 연구에서는 사용하고자 하는 인공신경망 모형은 종속변수와 독립변수의 관계를 선형적으로 설정할 필요가 없으며 다양한 비선형적 관계를 분석에 포함시킬 수 있다는 장점을 가진다. 또한, 인공신경망 모형을 이용한 분석은 예측성고가 우수하고 변수 간의 복잡한 관계를 잘 근사시키는 것으로 알려져 있다.

넷째, 인공신경망 모형을 이용하여 미래 매수 신호를 예측함에 있어 인터넷 검색지수를 독립변수로 사용하였다. 선행연구에서 주로 사용하였던 거시경제변수를 사용하는 대신 행동재무학에 기초한 투자자의 정서를 잘 반영하는 인터넷 검색지수(빅 데이터)를 인공신경망 모형의 독립변수로 사용하였다. Song et al.(2011)은 구글 트렌드(Google Trend)와 같은 인터넷 검색지수는 투자자 정서를 잘 반영할 수 있다고 주장하였으며, Ahmed et al.(2017)은 금융시장을 예측할 때 인터넷 검색 자료를 이용하여 예측한 결과가 좀 더 정확한 예측치를 가진다는 점을 실증적으로 보인 바 있다. 중국 주식시장은 시가총액 기준 1995년 3,474억에서 2019년 7조 6천 억으로 빠르게 증가하여 세계 2위의 주식시장이 되었다. 또한, 다른 주식시장과는 다르게 개인투자자 비율이 상대적으로 더 높다는 특징을 가진다. 개인투자자의 수가 전체투자자 수의 99.7%를 차지하는 중국 주식시장에서 효율적이고 투자 의사 결정에 도움이 되는 투자신호 시스템은 매우 중요한 의미를 지닌다고 할 수 있겠다. 본 연구에서는 중국의 주식시장에 직접적인 영향을 미칠 것으로 생각되는 바이두 지수(Baidu Index)를 사용한 인공신경망 모형을 토대로 미래의 투자신호를 예측하였다.

본 연구의 구성은 다음과 같다. 제Ⅱ장에서는 기존의 선행연구를 살펴보고, 제Ⅲ장에서는 MACD 지표를 토대로 만든 투자신호를 정의한다. 제Ⅳ장에서는 본 연구에서 사용된 자료와 실증분석 결과를 살펴본다. 제Ⅴ장에서는 연구의 주요 분석 결과들을 정리하였다.

Ⅱ. 주요 선행연구

최근에 기계학습(Machine Learning)을 이용하여 복잡한 금융시장 문제를 분석하는 연구가 많이 수행되어져왔다. 여러 기계학습 모형 중 인공신경망 모형은 이전부터 많은 관심을 받아오던 모형 중의 하나이다. Mendelsohn and Stein(1991)는 독일의 3년 거래 자료를 이용하여 투자신호를 예측한 결과 인공신경망 모형의 예측결과는 24개의 투자신호를 가장 적절히 설명하고 또한 50% 이상의 성공률을 가진다는 것은 보였다. 또한 인공신경망 모형을 주식시장분석에 활용한 기존 연구에서는 기본적으로 시장에 대한 투자자의 관심을 잘

반영하고 있는 자료를 독립변수로 사용하고 있다. Gervais et al.(2001)과 Barber and Odean (2008)은 투자자가 접하는 정보와 거래 방식에 관련된 연구를 통해, 관련된 정보를 바탕으로 투자전략을 구성하는 경우 높은 초과 수익률을 얻을 수 있다는 것을 보였다. 기술이 발전함에 따라 인터넷 검색을 통하여 훨씬 더 많은 정보를 쉽고 빠르게 구할 수 있게 되었지만, 수집한 정보를 바탕으로 정확한 분석을 하기 위한 모형은 아직 더 많은 연구가 필요한 실정이다.

2006년 구글 트렌드가 공개되면서 이를 활용한 주식시장 분석모형 또한 활발히 이루어지고 있다. Da, Engelberg and Gao(2011)은 구글 트렌드를 활용하여 투자자들의 관심에 기초한 인터넷 검색지수(SVI: Search Volume Index)를 구성하고 이를 이용한 투자분석 모형을 제시하였다. 이 연구에 따르면 SVI 지수는 주식시장을 설명할 수 있는 통계적으로 유의한 정보를 포함하고 있으며, 기존의 변수보다 높은 설명력을 보이는 것으로 나타났다. 중국 주식시장의 경우, 구글 트렌드를 사용하는 대신 중국 시장을 좀 더 잘 반영할 수 있는 바이두 지수를 이용하고 있으며, 검색 자료를 사용하여 중국 주식시장을 분석하고 투자전략을 구성할 때 상대적으로 높은 정확성을 보이는 것으로 나타났다(俞慶進, 張兵, 2012). 陳影, 王立夏, 周智敏(2017)은 바이두 지수를 사용하여 주가를 분석하고 변동성을 예측한 결과 바이두 지수를 사용하여 주가지수를 예측한 결과 높은 정확성을 보였으며, 자료가 주가지수를 분석하는 데 적합하다는 것을 확인하였다. 또한, 楊超, 姜昊, 雷崢嶸(2019)에서는 바이두 지수를 시계열 모형의 독립변수로 하여 중국의 환율을 예측할 때 더욱 정확한 결과를 얻을 수 있다는 것을 보였다.

인터넷 검색지수는 투자자의 행동, 경험, 인식 등을 빠르게 반영할 수 있다. 또한, 인공지능망 모형은 방대한 자료가 주는 신호를 보다 정확하게 분석할 수 있다는 장점이 있다. 따라서 본 연구에서는 인터넷 검색 자료와 인공지능망 모형을 통하여 투자신호를 분석하고 보다 정교한 투자전략 모형을 제안하고자 한다.

Ⅲ. 투자신호의 정의: SSE50 지수

1. MACD 지표

Granville(1963)은 이동평균(MA: Moving Average)분석법을 이용한 투자전략을 제안 하였으며, MA 분석법은 다음과 같이 나타낼 수 있다.

$$MA_{t,n} = \left(\frac{1}{n} \right) \sum_{i=t-n}^{t-1} P_t, \quad (1)$$

여기서 $MA_{t,n}$ 은 t 시점까지 n 일 동안의 이동평균, P_t 는 t 시점의 시장 증가를 의미한다. 이와 같은 MA 분석법은 모든 시기의 이동 평균값에 같은 가중치를 부여하기 때문에 정확한 결과를 얻을 수 없다는 단점이 있다. Appel and Hirschler(1990)는 과거의 주가가 현시점과 가까울수록 그 영향은 더 커지고 멀어질수록 영향이 약해지는 현상을 발견하고, 이를 토대로 MA 분석법을 보완한 MACD 지표를 활용한 투자신호를 제안하였다. MACD 지표란 12일 동안의 시장 증가 평균(단기 이동 평균값)과 26일 동안의 시장 증가 평균(장기 이동 평균값)을 이용하여 주가가 분리되고 취합되는 상태를 판단하고 이에 따라 매수, 매도의 신호를 판단하는 지표이다.

MACD 지표를 구하기 위해서는 다음과 같은 지수 이동평균선(EMA)를 구해야 한다. EMA 는 12일 동안의 단기 가중지수 이동평균선 $EMA_{t,1}$ 와 26일 동안의 장기 가중지수 이동평균선 $EMA_{t,2}$ 로 나눌 수 있다. 먼저 12일 동안의 단기이동평균선 $EMA_{t,1}$ 은 다음과 같이 구할 수 있다.

$$EMA_{t,1} = \frac{11}{13}EMA_{t-1,1} + \frac{2}{13}P, \quad (2)$$

여기서 $EMA_{t,1}$ 은 t 시점에서 12일 동안의 단기 가중지수 이동평균지수, P 는 해당일의 주식 증가를 의미한다. 이러한 EMA 지수를 구하기 위해서는 최근 가격에 2/13의 가중치를 부여하고 이전의 평균 가격에 11/13의 가중치를 준다. 다음은 26일 동안의 단기 이동평균선 $EMA_{t,2}$ 을 나타낸다.

$$EMA_{t,2} = \frac{25}{27}EMA_{t-1,1} + \frac{2}{27}P, \quad (3)$$

여기서 $EMA_{t,2}$ 는 t 시점의 26일 동안의 장기 가중지수 이동평균지수를 의미한다. $EMA_{t,2}$ 에서 또한 식 (2)와 동일한 방식으로 최근 가격에 2/27의 가중치를 부여하고 이전의 평균 가격에 25/27의 가중치를 부여한다. 앞서 계산된 $EMA_{t,1}$ 과 $EMA_{t,2}$ 를 사용하여 단기와 장기의 이동평균지표의 차분 값인 DIF_t 를 구할 수 있다.

$$DIF_t = EMA_{t,1} - EMA_{t,2}. \quad (4)$$

이러한 DIF_t 를 토대로 계산된 9일 동안의 이동 평균값을 DEA_t 라고 하며, 다음과 같이 나타낼 수 있다.

$$DEA_t = \frac{8}{10}DEA_{t-1} + \frac{2}{10}DIF_t. \quad (5)$$

MACD 지표는 주식시장의 단기적인 추세보다는 장기적인 추세에 주목하고 있다. 이러한 MACD 지표는 장단기 지수 이동 평균값을 차분하여 얻은 DIF 값을 통하여 주가의 단기 추세에 따른 무의미한 거래 신호를 효과적으로 제거할 수 있다.

2. 투자신호

1) 기존 연구의 MACD 투자신호

일반적으로 MACD를 이용한 투자전략은 Appel and Hirschler(1990)이 제시하는 MACD 거래법에 기초하여 만드는 것이다. 지금까지 가장 많이 활용되는 MACD를 이용한 투자전략은 DIF_t 와 DEA_t 값과 이들의 상대적인 위치를 기준으로 결정된다.

(1) DIF_t 와 DEA_t 를 활용한 투자전략

기본적으로 DIF_t 와 DEA_t 의 값이 모두 0보다 크게 나타나면 단기 지수 평균이 장기 지수 평균보다 클수록 DIF_t 가 커지게 된다. 따라서 지수는 상승할 것으로 판단할 수 있다. 이와는 반대로 DIF_t 와 DEA_t 의 값이 모두 0보다 작고 낮게 나타나면 단기 지수 평균이 장기 지수 평균보다 작을수록 DIF_t 가 작아지므로 지수는 하락할 것이라고 판단할 수 있다.

(2) 기존의 DIF_t 와 DEA_t 의 지수매수 투자전략¹⁾

DIF_t 와 DEA_t 의 값이 모두 0보다 크고, DIF_t 의 값이 DEA_t 의 값보다 크게 나타나면서 DIF_t 의 추세선이 DEA_t 의 추세선을 상향 돌파하는 경우에는 단기 지수 평균이 장기 지수 평균보다 커진다. 또한, DIF_t 가 점점 커지면 미래 단기 지수 평균의 값을 더 좋아지기 때문에 해당 주식을 매수해야 하는 신호로 판단할 수 있다. 이러한 경우는 투자 기회가 있다고 판단하여 주식을 매수해야 한다.

DIF_t 와 DEA_t 의 값이 모두 0보다 작지만, DIF_t 의 추세선이 DEA_t 를 상향 돌파하는 경우 단기 지수 평균이 장기 지수 평균보다 작지만, DIF_t 의 절대값이 점점 작아지면 미래 단기 지수 평균의 값이 하락에서 상승으로 바뀌기 때문에 해당 주식의 매수 신호로 해석할 수

1) 일반적인 개인투자자는 공매도가 제한되므로 본 연구에서는 공매도 거래를 고려하지 않는다.

있다. 이러한 경우 약세 시장에 해당한다. 하지만 주가가 상승하는 상황으로 전환될 수 있기 때문에 투자 기회가 있다고 판단하여 주식을 매수해야 한다.

Sig 0: DIF_t 의 추세선이 DEA_t 의 추세선을 상향 돌파하지 않은 경우 (관망)

Sig 1: $DIF_{t-1} < DEA_t$ and $DIF_t > DEA_t$ (매수)

기존의 MACD를 사용한 투자 교차전략은 신호가 주식을 매수하지 않아야 할 때는 0, 매수해야 할 때는 1로 표현되어진다. 이와 같은 방법은 분석 기간이 짧아짐에 따라 많은 투자신호를 보여주게 되고, 이에 따라 높은 거래 비용을 수반할 수 있다는 문제가 있다. 또한, 투자자의 성향에 따라 투자신호를 정확하게 구분할 수 없다는 단점이 있으며, 賈宇杰(2018)가 앞서 설명한 MACD 전략을 사용하여 실증분석한 결과에서 수익을 발생 시키기 어렵다는 점이 확인되었다. 따라서 본 연구는 이러한 단점을 보완하고 약세 시장과 강세 시장에 따라 세분화된 투자신호를 구할 수 있는 방법을 제안하고자 한다. 본 연구에서는 투자신호를 가장 부정적인 관망 신호인 Sig 0부터 긍정적인 매수 신호인 Sig 5까지 총 여섯 개 신호로 세분화하여 투자전략을 더욱 구체화하고자 하였다.

2) 세분화된 MACD 투자신호

세분화된 투자신호를 설정하는데 있어 중요한 것은 신호등급의 개수라고 할 수 있겠다. 가장 간단한 경우 2개의 신호등급을 고려할 수 있는데 이 경우 신호는 명확하지만, 그 신호의 세기가 명확하지 않게 된다. 그와 반대로 신호의 등급이 많아지면 신호의 세기를 더 명확하게 정의할 수 있지만, 어느 신호에 주식을 구매해야 하는지가 분명하지 않을 수 있다. 이항변수 투자신호를 정의하게 되면 짧은 기간 내에 많은 투자신호를 방출하게 되지만 효율적인 투자신호가 많이 없는 것이 일반적이다. 신호등급의 개수를 통계적인 검정을 통하여 선택하는 연구가 필요하지만, 일반화되고 많이 사용되는 검정 통계량은 아직 많지 않다. 따라서 본 연구에서는 이항변수를 사용하여 투자신호를 정의하기보다는 李愛泉(2019)에 따라 투자신호를 가장 낮은 등급의 Sig 0부터 가장 높은 등급의 Sig 5까지 여섯 개 신호등급으로 구분하였다. 본 연구에서는 관망, 신중추천, 추천, 강력추천, 매수, 매수로 구분한다. DIF 와 DEA 의 값이 모두 0보다 큰 경우 강세 시장으로 해석하며, DIF 의 추세가 DEA 보다 크게 나타나 상향 돌파하는 경우 주가는 상승할 것으로 판단할 수 있으며 투자 기회가 있는 것으로 해석할 수 있다. 두 값이 모두 0보다 큰 위치에서 DIF 의 추세선이 DEA 의 추세선을 돌파하는 경우 주식시장이 강세이지만 DIF 의 절대값이 매우 크게 나타나면

과열된 것으로 판단하여 투자를 보류해야 할 것으로 판단할 수 있다. 이와는 반대로 DIF 와 DEA 의 값이 모두 0과 가까운 지점에서 DIF 의 추세가 DEA 를 돌파하는 경우 주식시장이 강세임과 동시에 상승할 수 있는 신호이므로 적극적으로 투자를 해야 한다는 신호로 해석할 수 있다. 만약 DIF 의 추세선이 DEA 의 추세선을 고려한 기간 내에 상향 돌파하지 않는 경우 투자 기회가 없는 경우로 해석할 수 있다. 위와 같은 상황을 기반으로 투자 기회가 있는 경우를 Sig 1부터 Sig 5까지 총 다섯 개로 나누고, DIF_t 의 추세선이 DEA_t 의 추세선을 상향 돌파하지 않은 경우를 Sig 0으로 하여 관망 신호로 간주하였다. 투자신호를 세분화하기 위하여 사용한 $Index$ 는 기존의 MACD 투자신호에 따라 매수신호가 주어질 때 DIF 와 DEA 의 추세선이 교차하는 값으로 정의하였다. 모두 신호를 기반으로 다음과 같은 투자모형을 구성하였다.

Sig 0: DIF_t 의 추세선이 DEA_t 의 추세선을 상향 돌파하지 않은 경우 (관망)

Sig 1: $\frac{4\gamma_{1,t,k}}{5} < Index_{1,t} \leq \gamma_{1,t,k}$ (신중추천)

Sig 2: $\frac{3\gamma_{1,t,k}}{5} < Index_{1,t} \leq \frac{4\gamma_{1,t,k}}{5}$ (추천)

Sig 3: $\frac{2\gamma_{1,t,k}}{5} < Index_{1,t} \leq \frac{3\gamma_{1,t,k}}{5}$ (강력추천)

Sig 4: $\frac{1\gamma_{1,t,k}}{5} < Index_{1,t} \leq \frac{2\gamma_{1,t,k}}{5}$ (매수)

Sig 5: $0 < Index_{1,t} \leq \frac{\gamma_{1,t,k}}{5}$ (매수)

$Index_{1,t}$ 는 t 시점에서 DIF 가 DEA 를 상향 돌파하는 순간($DIF=DEA$)과 강세 시장에만 값을 갖게 되며, 따라서 강세 시장에서만 신호를 보이게 된다. $\gamma_{1,t,k}$ 는 t 시점의 k 일 동안의 DIF_t 와 DEA_t 의 이동 최댓값으로 정의하며, 다음과 같이 나타낼 수 있다.

$$\gamma_{1,t,k} = \text{Max}_k(DIF_t, DEA_t).$$

만약 DIF_t 와 DEA_t 가 모두 0보다 작을 때 약세 시장으로 판단하며, 0으로부터 멀어질수록 주가는 하락하는 것으로 해석할 수 있다. 이러한 상황에서 DIF_t 의 추세선이 DEA_t 를 상향 돌파하면 주가가 상승하는 것으로 판단할 수 있지만, 두 추세선의 교차점이 0보다 큰 지점에서 DIF_t 의 값이 DEA_t 를 상향 돌파하게 된다면 시장이 과열된 이후 침체될 수 있기 때문에

투자신호는 가장 낮은 값을 보이게 된다. 두 추세선의 교차점이 모두 0보다 작은 지점에서 DIF_t 의 값이 DEA_t 을 상향 돌파하게 된다면 시장은 약세 시장에서 강세 시장으로 쉽게 바꿀 수 있기 때문에 투자신호는 가장 높은 값을 보이게 된다. 만약 DIF 의 추세선이 DEA 의 추세선을 고려한 기간 내에 상향 돌파하지 않는 경우 투자 기회가 없는 경우로 해석할 수 있다. 위와 같은 상황을 기반으로 투자 기회가 있는 경우를 Sig 1부터 Sig 5까지 총 다섯 개 신호로 나누고, DIF_t 의 추세선이 DEA_t 의 추세선을 상향 돌파하지 않은 경우를 Sig 0으로 하여 관망 신호로 간주하였다.

Sig 0: DIF_t 의 추세선이 DEA_t 의 추세선을 상향 돌파하지 않은 경우 (관망)

Sig 1: $\frac{4\gamma_{2,t,k}}{5} \geq Index_{2,t} > \gamma_{2,t,k}$ (신중추천)

Sig 2: $\frac{3\gamma_{2,t,k}}{5} \geq Index_{2,t} > \frac{4\gamma_{2,t,k}}{5}$ (추천)

Sig 3: $\frac{2\gamma_{2,t,k}}{5} \geq Index_{2,t} > \frac{3\gamma_{2,t,k}}{5}$ (강력추천)

Sig 4: $\frac{1\gamma_{2,t,k}}{5} \geq Index_{2,t} > \frac{2\gamma_{2,t,k}}{5}$ (매수)

Sig 5: $0 \geq Index_{2,t} > \frac{\gamma_{2,t,k}}{5}$, (매수)

$Index_{2,t}$ 는 약세 시장에서의 t 시점에서 DIF 가 DEA 을 상향 돌파하는 순간의 신호를 나타내고 있으며, 따라서 $Index_{2,t} \leq 0$ 로 주어진다. $\gamma_{2,t,k}$ 는 t 시점의 k 일 동안의 DIF_t 와 DEA_t 의 이동 최솟값으로 정의하며, 다음 식으로 고려하였다.

$$\gamma_{2,t,k} = \text{Min}_k(DIF_t, DEA_t),$$

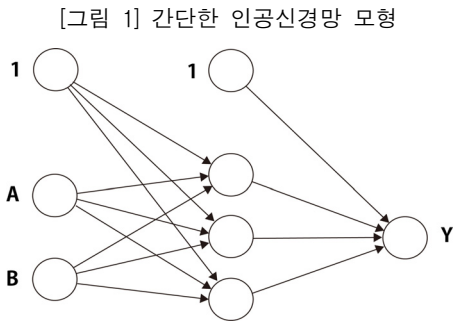
여기서 k 는 γ_t 의 기간을 나타내며, k 가 작을수록 짧은 기간 내에 많은 신호를 얻게 되지만 시장의 상황을 반영할 수 없으며, k 가 클수록 시장의 상황은 반영할 수 있지만, 기간 내에 얻어지는 극단 값에 따라 부정확한 신호를 보일 수 있다.

3. 인공신경망 모형(Artificial Neural Network)

많은 실증분석에서 신호의 경제변수 간의 함수적 관계에 대한 연구를 행하고 있다. 이러한 변수 간 가장 기본적으로 그리고 많이 설정하는 관계는 선형관계이며 필요한 경우 이들

변수의 관계를 비선형으로 설정하여 분석하고 있다. 인공신경망 모형은 종속변수와 독립변수의 관계를 설정할 필요가 없으며 비선형 관계를 분석할 때 적합한 모형이다. 또한, 변수 간의 복잡한 관계를 잘 설명할 수 있는 모형으로 예측결과와 높은 정확성을 보인다(Hornik et al., 1989). Rumelhart et al.(1986)이 제안한 인공신경망 모형은 정보를 입력하는 입력층, 처리하는 은닉층, 결과를 보여주는 출력층으로 나누어지며, 각 층이 연결되는 구조는 사람의 신경망과 유사한 형태를 가지고 있다. 이러한 모형은 사람의 신경망이 정보를 처리하는 과정과 같이 정보를 처리함으로써 복잡한 변수간 관계를 분석하는 데 적합한 모형이다.

본 연구에서 사용된 인공신경망 모형은 오류역전파 인공신경망(Back Propagation Artificial Neural Network)으로, 추정 과정에서 자료의 구성 및 각 층을 연결해주는 정보를 입력해줄 필요 없이 훈련을 통하여 자료를 분석하고 오차를 최소화하여 결과를 보여주는 모형이다.



오류역전파 인공신경망에서는 $X(x_1, x_2, \dots, x_n)$, $Y(y_1, y_2, \dots, y_n)$ 는 인공신경망의 입력변수와 출력변수를 각각 나타낸다. [그림 1]에서 볼 수 있듯이 두 개 투입 신경세포, 세 개 은닉 신경세포, 한 개 출력 신경세포로 구성된 인공신경망 모형이 나타난다. 좌측에 위치하는 층은 입력층이며 우측에 위치하는 층이 출력층, 가운데 위치하는 층이 은닉층이다. 가장 간단한 인공신경망 모형은 다음과 같이 나타낼 수 있다.

$$f(x) = F\left(\beta_0 + \sum_{j=1}^q \beta_j x_j\right). \tag{6}$$

위의 인공신경망 모형은 은닉층이 존재하지 않은 모형이고 은닉층을 추가하면 다음의 식 (7)과 같이 표현할 수 있다.

$$f(x, \theta) = F(\beta_0 + \sum_{j=1}^q G(x' \gamma_j) \beta_j), \quad (7)$$

여기서 $F: R \rightarrow R$ 와 $G: R \rightarrow R$ 은 활성화 함수를 의미한다. β_0 는 출력 신경세포의 상수항을, γ_j 는 독립변수의 가중치를, β_j 는 j 의 은닉 신경세포에서부터 출력 신경세포로의 연결 가중치를 각각 의미한다. 일반적으로 위의 식 (7)과 같이 주어진 하나의 은닉층이 있는 모형이 경제학과 재무학에 일반적으로 쓰이는 간단한 모형이다. 은닉층 활성화 함수로는 일반적으로 로지스틱함수(Logistic)²⁾가 많이 사용되고 있다. 본 연구에서는 출력층은 선형 활성화 함수 $F(x) = x$ 를 가정하였고 은닉층은 다음의 식 (8)과 같은 로지스틱함수로 사용하였다.³⁾

$$G(a) = \frac{1}{1 + \exp(-a)}. \quad (8)$$

많은 은닉 신경세포를 사용하여 모형의 표본내 정확도를 높일 수도 있지만, 이 높아진 표본내 정확도가 표본외 예측의 정확도를 항상 시키진 않는다. 또한, 현재 문헌에서 최적의 은닉층 개수와 세포갯수를 검정하는 일반적이고 통상적인 검정기법을 찾아보기 힘들다. 따라서 본 연구에서는 투자신호를 추정하기 위해 식(7)과 (8)로 주어진 인공신경망 모형에서 (i) 한 개의 은닉층과 한 개 은닉 신경세포 $j=1(L1)$, (ii) 한 개의 은닉층 세 개의 은닉 신경세포 $j=3(L2)$, 그리고 입력층은 앞서 고려한 총 20개의 인터넷 검색 자료를 토대로 추정한 주성분의 1분기 전(lag 1), 2분기 전(lag 2), 3분기 전(lag 3), 4분기 전(lag 4), 5분기 전(lag 5)까지의 시차 변수 총 40개 변수를 사용하였다. 실증분석을 하기 위해 교차 검증법(Cross Validation)을 사용하였으며, 이를 위하여 전체 자료를 훈련 집합(training set), 검증 집합(validation set), 테스트 집합(test set)으로 구분하였다. 인공신경망 모형을 사용하기 위하여 2006년 6월 1일부터 2017년 8월 18일까지의 자료를 훈련 집합으로, 2017년 8월 18일부터 2019년 1월 4일까지를 검증 집합으로, 모형을 사용하여 다음 4개 구간의 신호를 예측하고 테스트 집합으로 사용하였다. 또한, 인공신경망 모형을 검증하기 위하여 2017년 8월 18일부터 2019년 1월 4일까지의 일별 바이두 지수 시계열 자료를 사용하여 모형의 검증 집합으로 간주하였다. 추정 모형을 가지고 다음 네 구간의 일별 투자신호를 예측하여 테스트 집합으로 간주하였다:

2) 일반적으로 은닉층에는 쌍곡탄젠트, 로지스틱, ReLU함수 등이 주로 사용된다. 로지스틱함수의 경우 다른 활성화 함수보다 쉽게 미분이 가능하다는 장점이 있으며, 다른 활성화 함수에 비하여 비선형적인 문제를 푸는 데 적합하다는 장점이 있다.

3) 인공신경망 모형의 추정법과 이론적인 부분은 Bishop(1995)을 참고하라.

- ① 2019년 1월 4일부터 2019년 1월 24일까지, ② 2019년 1월 25일부터 2019년 2월 14일까지, ③ 2019년 2월 15일부터 2019년 3월 5일까지, ④ 2019년 3월 6일부터 2019년 3월 24일까지.

Shanker et al.(1996)은 독립변수와 종속변수를 $[-1, 1]$ 로 정규화하여 사용하는 것이 모형설정과 모형추정에 더 효율적이라는 것을 보였다. 따라서 본 연구에서도 인공신경망 모형의 독립변수와 종속변수들을 $[-1, 1]$ 로 정규화하여 사용하였다. 모든 변수에 대한 정규화 방법은 다음과 같이 나타낼 수 있다.

$$X^* = \frac{x - \min(x)}{\max(x) - \min(x)}.$$

최종적으로 본 연구에서 사용된 신호의 범위에 따라 종속변수의 최솟값을 0부터 최댓값 5까지 구분하여 마지막 출력층에서의 추정결과에 5를 곱한 값을 투자신호로 간주하였다.

IV. 실증분석

본 장에서는 제Ⅲ장에서 정의된 투자신호를 사용하여 투자신호 모형을 추정하고 이를 이용하여 2019년 1월 4일부터 2019년 1월 24일까지, 2019년 1월 25일부터 2019년 2월 14일까지, 2019년 2월 15일부터 2019년 3월 5일까지와 2019년 3월 6일부터 2019년 3월 24일까지 네 구간의 중국 SSE50 지수에 대한 투자신호를 예측하였다. 투자자 정서가 주가에 큰 영향을 미치고 인터넷 검색 자료가 투자자의 정서에 영향을 미칠 수 있기 때문에 인터넷 검색 자료는 주가에 큰 영향을 미치는 요인으로 고려될 수 있다(Ahmed et al., 2017). 본 연구에서는 인터넷 검색지수를 고려하여 투자신호 예측에 사용되는 모형으로써 종속변수와 독립변수의 복잡한 관계를 잘 근사시키는 것으로 알려진 인공신경망(ANN: Artificial Neural Network) 모형을 사용하였다.

1. 분석자료

실증분석을 위해 2006년 6월 1일부터 2019년 1월 4일까지의 일별 바이두 지수 시계열 자료를 사용하였다. 이는 중국에서 구글 트렌드보다 바이두 지수가 더 큰 시장을 차지하고 있음을 반영한 것이다. 본 연구에서는 주가 변동과 관련이 있는 대표적인 20개 검색어를 사용하였는데, 경제, 채권, 외국인 투자자, 회사 가치, 선물, 국내총생산(GDP: Gross Domestic Product), 금 가격, 성장주, 이자율, 투자자, 주식공개상장(IPO: Initial Public Offerings), 원유가, 주요시장, 시장 증가, 부동산 가격, 신주, 환매조건부채권(REPO: Repurchase

Agreements), 예금 지급 준비금, 대주거래, 워런 버핏이다. 이상의 20개 바이두 지수 검색 자료는 독립변수로 모두 사용할 수 있지만, 인공신경망의 적합 효율을 늘리기 위해 주성분 분석을 이용하여 검색 자료의 8개 주성분을 인공신경망 모형에 사용하였다. <표 1>은 2006년 6월 1일부터 2019년 1월 4일까지의 SSE50 지수와 대표적인 20개 인터넷 검색지수의 기초 통계량을 보여주고 있다.

<표 1> SSE50 지수와 SVI 기초 통계량

이 표는 2006년 6월 1일부터 2019년 1월 4일까지의 SSE50 지수와 대표적인 20개 인터넷 검색지수 ‘경제’, ‘채권’, ‘외국인 투자자’, ‘회사 가치’, ‘선물’, ‘GDP’, ‘금 가격’, ‘성장주’, ‘이자율’, ‘투자자’, ‘IPO’, ‘원유가’, ‘주요시장’, ‘시장 증가’, ‘부동산 가격’, ‘신주’, ‘REPO’, ‘예금 지급 준비금’, ‘대주거래’, ‘워런 버핏’의 기초 통계량을 보여주고 있다.

변수명	표본수	표준편차	평균값	최댓값	최솟값
SSE50	3075	636.96	2182.76	4726.08	977.84
경제	3075	383.17	1385.68	2548.00	284.00
채권	3075	15002.74	10983.26	128699.00	921.00
외국인 투자자	3075	766.41	1960.81	9616.00	391.00
회사 가치	3075	121.08	209.47	972.00	0.00
선물	3075	1460.94	2994.67	20099.00	654.00
GDP	3075	1527.92	2405.37	30268.00	135.00
금가격	3075	6652.95	7585.02	132671.00	734.00
성장주	3075	5503.12	2519.39	108506.00	0.00
이자율	3075	970.69	1488.73	17650.00	110.00
투자자	3075	4131.62	3071.14	53022.00	319.00
IPO	3075	3859.00	2748.82	99191.00	112.00
주요시장	3075	52.12	106.17	409.00	0.00
시장 증가	3075	20214.57	6377.65	400062.00	0.00
신주	3075	2369.45	1686.00	39725.00	0.00
유가	3075	1835.33	1462.71	12570.00	0.00
부동산 가격	3075	1838.34	2798.83	17510.00	198.00
REPO	3075	875.99	678.09	11455.00	0.00
지급준비율	3075	492.72	383.26	9864.00	0.00
대주거래	3075	1290.39	1102.78	16645.00	0.00
워런 버핏	3075	1408.27	2053.72	20817.00	187.00

<표 2>에서 인터넷 검색지수의 첫 번째 주성분은 전체 데이터에 대해 42.46%에 해당하는 분산을 설명할 수 있고, 여덟 번째 주성분까지 전체 데이터에 대해 80.80%에 해당하는 분산을 설명할 수 있음을 확인할 수 있다. 여덟 번째 이후의 주성분은 총 데이터의 분산 중에서 상대적으로 작은 부분만을 설명할 수 있기 때문에 본 연구에서는 첫 번째부터 여덟 번째까지, 총 8개 주성분을 인공신경망 모형의 독립변수에 포함하여 사용하였다.

<표 2> PCA의 분산 총비율

이 표는 2006년 6월 1일부터 2019년 1월 4일까지의 주가 변동과 관련이 있는 대표적인 20개 검색어를 주성분 분석을 통하여 얻은 주성분 분산의 결과를 나타낸다.

	Stand Deviation	Proportion of Variance	Cumulative Proportion
PC1	2.9141	0.4246	0.4246
PC2	1.5460	0.1195	0.0554
PC3	1.1643	0.0678	0.6119
PC4	0.9637	0.0464	0.6583
PC5	0.9378	0.4397	0.7023
PC6	0.9100	0.0414	0.7437
PC7	0.8275	0.0342	0.7779
PC8	0.7758	0.0301	0.8080
PC9	0.7149	0.0256	0.8336
PC10	0.6959	0.0242	0.8578
PC11	0.6853	0.0235	0.8813
PC12	0.6542	0.0214	0.9026
PC13	0.6368	0.0214	0.9299
PC14	0.5835	0.0170	0.9400
PC15	0.5623	0.0158	0.9558
PC16	0.5437	0.0148	0.9705
PC17	0.4495	0.0101	0.9807
PC18	0.4264	0.0091	0.9897
PC19	0.3636	0.0066	0.9963
PC20	0.2704	0.0037	1.0000

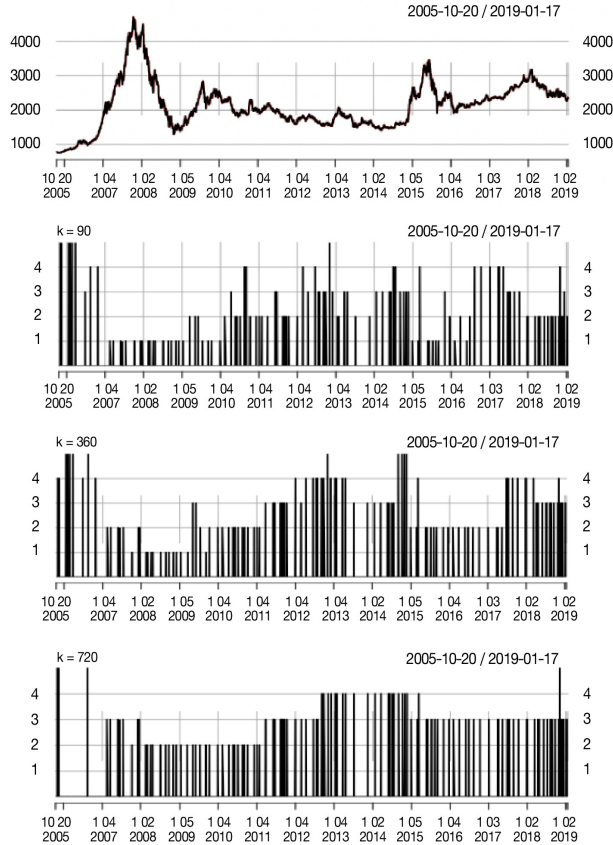
2. SSE50 지수 투자신호

[그림 2]는 기간별로 추정된 SSE50 지수의 투자신호를 보여주고 있다. 신호를 구하기 위한 기간 k 를 90, 360, 720으로 구분하였다. k 의 값이 90으로 가장 낮을 때 신호가 상대적으로 급격하게 변화하는 모습을 확인할 수 있으며, 이는 k 의 값이 높아질수록 더 완만하게 나타나게 된다.

[그림 2]에서 볼 수 있듯이 k 의 값이 가장 낮을 때($k=90$) 신호가 급격하게 변화하며, k 의 값이 높아질수록 상대적으로 신호가 완만하게 나타나는 것을 확인할 수 있다. 본 연구에서는 k 의 값을 360으로 설정하여 분석을 진행하였다. 분석 결과에서 2006년 1월에는 강세 시장과 함께 지수가 상승하기에 앞서 적극적으로 투자를 해야 한다는 Sig 5를 보이고 있고, 2014년 12월, 2015년 1월 또한 시장 상황이 잘 반영되어 투자에 긍정적인 Sig 5를 보이고 있음을 확인할 수 있다. <표 3>의 첫 번째 표는 Sig 5의 주가 수익률이다. Sig 5의 주가 수익률은 2006년 6월 1일부터 2019년 1월 6일까지의 MACD 신호를 기반으로

[그림 2] SSE50 지수와 k 기간 동안의 세분화된 투자신호

순서대로 첫 번째 그림은 SSE50 지수 시계열이며, 두 번째, 세 번째, 네 번째 그림은 각각 k 의 값이 90, 360, 720일 때의 SSE50 지수 투자신호에 대한 그래프이다.



하는 투자전략의 수익률을 보여준다. 분석 결과 해당 기간에 Sig 5 매수 신호에 따라 투자를 하는 경우 수익률이 높게 나타나고 있다. 따라서 본 연구에서 제안한 신호가 좋은 성과를 낼 수 있음을 확인할 수 있다. <표 3>의 두 번째 표는 누적수익률⁴⁾을 보여주는데, Sig 4와 5의 매수 신호에 따라 30~360일의 보유 기간 중 보유 기간이 180일 일 때 Sig 4와 5의 누적수익률 최대화에 이를 수 있음을 보여준다. 또한, Sig 5의 30일, 60일, 180일, 360일 보유주기의 누적수익률이 해당 주기 Sig 4의 누적수익률보다 높게 나타남을 확인할 수 있다. Sig 5 매수 신호의 180일 누적수익과 360일 누적수익이 모두 30일 누적수익과 60일

4) 누적수익률 계산법: 매수 신호가 주어지는 경우 이에 해당하는 일수 동안 매수하고 보유하며, 보유 일수 내에 신호가 다시 주어지는 경우 보유 기간을 연장한다. 보유 기간 중 매수 신호가 주어지지 않는 경우 기간이 끝나는 시점에서 매도하고, 다시 매수 신호가 주어지는 경우 매수하는 전략으로 얻어진 총수익률을 계산하였다.

누적수익의 두 배가 넘으므로, 우리는 Sig 5 매수 신호가 단기 추세에 비해 180일 이상의 장기 추세에 투자하는 데 더 적합하다는 결론을 내릴 수 있다. Sig 4 매수 신호의 30일, 60일, 180일의 누적수익은 모두 70%를 초과하지만 360일 누적수익은 결손이 나므로 Sig 4 매수 신호는 360일의 장기 추세보다 180일 이내의 단기 추세에 투자하기에 적합함을 보여준다.

<표 3>의 세 번째에 해당하는 투자 성공률⁵⁾ 표를 보면, 30일 보유에서 360일 보유까지의 Sig 5 매수 신호 투자 성공률은 높은 수준을 유지하고 있음을 알 수 있다. 그러나 Sig 4 매수 신호 투자 성공률은 30일, 60일 보유가 50%보다 크게 나타나지만 반면에 180일, 360일 보유에 대해서는 Sig 4 매수 신호의 표현이 크게 떨어졌다.

이 기간에 Sig 5 신호수는 Sig 4 신호수보다 적었지만, Sig 5는 투자 성공률과 투자수익 모두 Sig 4보다 현저히 높다. 여기서 Sig 5는 180일 이상의 장기투자에 더 적합하고 Sig 4는 180일 이내의 투자에 더 적합하다는 결과를 확인할 수 있다.

<표 3> 매수 신호의 주기 투자 결과

첫 번째 표는 Sig 5의 주기 수익률이다. 이 표는 2006년 6월 1일부터 2019년 1월 4일까지의 Sig 5 매수 신호가 나올 때 SSE50 지수를 구입함으로써 얻을 수 있는 수익률을 나타낸다. 두 번째 표는 Sig 4와 5의 주기 누적수익률이다. 이 표는 2006년 6월 1일부터 2019년 1월 4일까지의 Sig 4와 5 매수 신호를 나올 때 상응한 기간 동안 보유하는 총 누적수익률을 나타낸다. 세 번째 표는 Sig 4와 5의 투자 성공률을 보여준다. 여기서는 2006년 6월 1일부터 2019년 1월 4일까지의 Sig 4와 5 매수 신호가 나올 때 상응한 일수 동안 보유하면 투자 성공률 및 투자 평균 성공률이 나타난다.

신호	날짜	30일	60일	180일	360일	
5	2006/08/17	6.86	26.93	181.16	259.23	
5	2012/11/06	7.03	22.68	-4.53	-9.14	
5	2014/09/05	-3.29	34.38	101.49	24.8	
5	2014/10/10	7.59	57.99	68.28	32.04	
5	2014/11/03	39.26	47.91	66.02	30.69	
5	2014/11/24	53.68	42.60	50.29	18.46	

신호	신호 수(개)	30일 누적수익률 (%)	60일 누적수익률 (%)	180일 누적수익률 (%)	360일 누적수익률 (%)	투자 평균 누적수익률 (%)
5	6	82.22	84.17	319.23	307.1	199.68
4	23	70.88	82.11	83.2	-8.02	62.79

신호	신호 수 (개)	30일 투자 성공률(%)	60일 투자 성공률(%)	180일 투자 성공률(%)	360일 투자 성공률(%)	투자 평균 성공률(%)
5	6	83.33	100	83.3	83.3	87.49
4	23	56	60.8	47.8	26	47.65

5) 투자 성공률: 신호를 방출할 때 상응한 요구일수 동안 매수하고 보유할 때 수익이 발생할 수 있는 확률이다.

<표 4>는 기존 MACD 투자신호 모형의 투자 결과와 본 연구에서 제시한 여섯 개 신호 투자모형의 투자 결과를 비교한 것이다. <표 4>의 첫 번째 표를 보면, 0과 1의 매수 여부 이항변수가 기존의 MACD 투자신호 모형과 비하여 Sig 0-5로 세분화되어 MACD 투자모형의 30일, 60일, 180일, 360일 보유주기의 투자 성공률이 해당 주기의 기존의 MACD 투자신호 모형보다 높게 나타났다. 또한 세분화된 MACD 투자신호 모형의 매수 신호는 기존의 MACD 투자신호 모형의 매수 신호보다 더 적게 나타남을 확인 할 수 있다.

<표 3>과 <표 4>로 부터 기존의 투자신호 모형에 비해, 세분화된 투자신호 모형의 총 신호수는 적어지고 투자 성공률은 전반적으로 크게 높아져 더 높은 투자 효율성이 나타난다는 점을 확인할 수 있다.

<표 4> 기존의 MACD 투자신호와 세분화된 MACD 투자신호 비교결과

첫 번째 표는 2006년 6월 1일부터 2019년 1월 4일까지 기존의 MACD 투자신호와 세분화된 MACD 투자신호 모형의 각각 신호수이다. 이 표는 2006년 6월 1일부터 2019년 1월 4일까지 기존의 MACD 투자신호와 세분화된 MACD 투자신호 모형 각각의 신호수가 나타난다. 두 번째 표는 기존의 MACD 투자신호와 세분화된 MACD 투자신호 모형의 성공률이다. 2006년 6월 1일부터 2019년 1월 4일까지의 기존의 MACD 투자신호 모형과 세분화된 MACD 투자신호 모형 각각 매수 신호가 나올 때 상응한 일수 동안 보유하면 투자 성공률 및 투자 평균성공률이 나타난다.

전략	총 신호수 (개)	총 신호 수(개)	1	2	3	4	5
기존의 MACD	3075	106	106	-	-	-	-
세분화된 MACD	3075	106	13	35	29	23	6

전략	총 신호수 (개)	매수 신호 수(개)	30일 투자 성공률(%)	60일 투자 성공률(%)	180일 투자 성공률(%)	360일 투자 성공률(%)	투자 평균 성공률(%)
기존의 MACD	3075	106	40.56	43.39	47.16	24.52	38.90
세분화된 MACD	3075	29	61.65	68.91	55.14	37.85	55.88

3. 인공신경망 모형을 사용한 투자신호의 추정 결과

오류역전파 방법은 인공신경망의 가중치 갱신을 위해서 오차합수를 정의하고 그 오차를 최소화시키도록 신경망을 학습시키는 방법이다. 오류역전파 방법은 오류의 역전파와 신호의 순전파로 나뉜다. 신호를 입력층에서 출력층까지 순전파 시 오차를 계산하여 오차를 역으로 전파하는 동시에 각 세포의 가중치를 조정한다. 각 신경세포에 상응되는 오차는 계산을

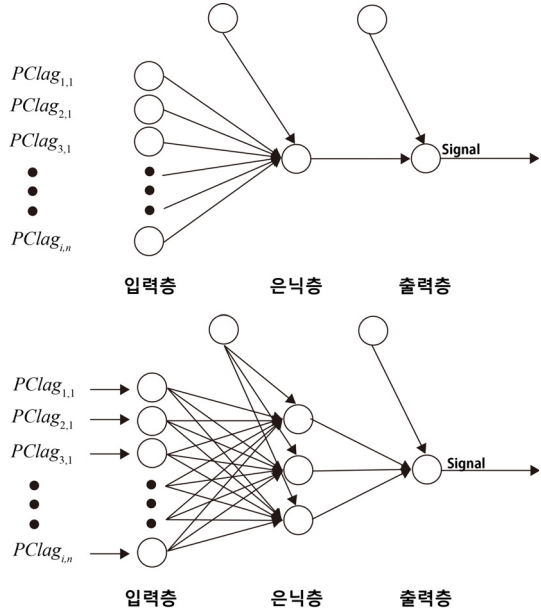
통해 손실 함수의 경사(Gradient)를 계산할 수 있으며 이 경사를 이용함으로써 경사 하강법(Gradient Descent Method)을 사용할 수 있게 된다.⁶⁾

인공신경망 모형에서 최적 은닉 세포의 수를 정하는 것을 중요한 문제이다. 하지만 이러한 최적 은닉 세포 수를 정하는 문제는 아직까지 많은 연구가 필요한 부분이다. 따라서 본 연구에서는 일반적으로 많이 사용되는 은닉 세포의 수를 사용하여 분석을 실시하였다.

오류역전과 범을 사용하여 세분화된 MACD 투자신호를 설명한 인공신경망 모형 L1($j=1$), L2($j=3$)의 추정 과정은 [그림 3]에 나타나 있다.

[그림 3] 인공신경망 모형 추정 과정

순서대로 첫 번째 그림은 은닉층 한 개, 은닉 세포 수 한 개로 설정한 인공신경망 추정 과정(L1)이다. $PCLag_{i,n}$ 는 i 번째 주성분의 n 일 과거 변수를 의미한다($i=1,\dots,8, n=1,\dots,5$). 두 번째 그림은 은닉층 한 개, 은닉 세포 수 세 개로 설정한 인공신경망 추정 과정(L2)이다.

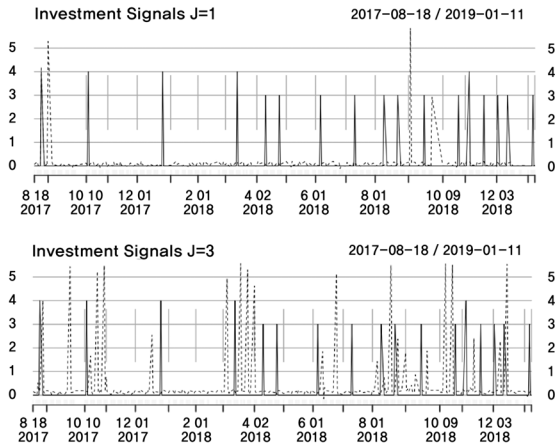


[그림 4]는 2017년 8월 18일부터 2019년 1월 4일까지의 검증 집합의 인공신경망 모형을 이용하여 추정하는 신호지수가 실제 신호지수에서의 적합도를 비교한 것이다. 점선은 오류역전과인공신경망 모형을 토대로 추정하는 신호의 시계열이고 실선은 실제 투자신호의 시계열을 각각 나타낸다.

6) 경사 하강법은 cost function의 1차 근사값을 계산하기 위한 최적화 알고리즘이다(Cauchy, 1847). 기본 개념은 함수의 기울기를 구하여 기울기가 낮은 쪽으로 계속 이동시켜서 극값에 이를 때까지 반복시키는 것이다.

[그림 4] 인공신경망 모형 결과

순서대로 첫 번째 그림은 은닉층 한 개, 은닉 세포 수 한 개로 설정하여 SSE50 지수의 세분화된 투자신호 실제값 및 예측값을 나타내는 그래프이다. 여기서 실선은 실제 신호를 의미하고 점선은 예측 신호를 나타낸다. 두 번째는 은닉층 한 개, 은닉 세포 수 세 개로 설정하여 SSE50 지수의 세분화된 투자신호 실제값 및 예측값을 나타내는 그래프이다.



앞서 세분화된 MACD 투자모형은 Sig 1-Sig 3에 기초하여 추천 신호를 주목하기보다는 Sig 4와 5 매수 신호에 기초하여 60일에서 360일 사이의 장기 투자 매매 점을 보유하는 것이 더 효과적이었다. 따라서 본 연구에서는 Sig 4와 5가 주어졌을 때만 매수하는 전략으로 모형의 실증결과를 비교하였다. 실제 투자신호가 추정 신호와 하루 차이가 나면 전통적인 평균 제곱 오차(MSE: Mean Squared Error)에 큰 오차가 생길 수 있지만, 실제 누적수익률에는 큰 영향을 주지 않는다. 그러므로 전통적인 MSE 테스트 모형은 인공신경망 모형으로 추정하는 신호 결과를 판단하는 데 사용될 수 없다. 이에 본 연구에서는 모형의 결과를 측정하기 위한 새로운 검출 방법을 설계하였다.

<표 5> L1과 L2 검증 집합 유효 신호수와 누적수익률 결과

첫 번째 표는 2017년 8월 18일부터 2019년 1월 4일까지의 검증 집합의 L1과 L2 추정 신호의 유효 매수 신호수를 나타낸다. 두 번째 표는 2017년 8월 18일부터 2019년 1월 4일까지의 검증 집합의 L1과 L2 추정 신호의 누적수익률을 나타낸다.

실제 총 매수 신호수		L1	L2
추정신호 유효 ‘매수’ 신호수(개)	5	1	4
L1		L2	
실제신호 누적수익률(%)	20.14	20.14	
추정신호 누적수익률(%)	12.23	28.12	

본 연구에서는 유효 매수 신호⁷⁾의 개수 및 누적수익률을 바탕으로 모형 결과를 측정하였다. <표 5>의 첫 번째 표와 같이 L1 모형은 적은 은닉 세포로 인해 모형이 너무 간단하므로 기간 내의 유효의 매수 투자신호 한 개만을 방출했다. 그러나 L2 모형은 세 개의 은닉 세포로 모형이 더 복잡한 비선형 문제를 처리할 수 있어서 기간 내에 4개의 유효한 투자신호를 보였다.

누적수익률로 결과를 측정하는 경우, Sig 0 - 3의 신호에는 매수하지 않고, Sig 4와 Sig 5의 신호가 주어지는 경우 매수하여 60일 동안 보유하고, 보유하고 있는 기간 내에 다시 매수 신호가 주어지는 경우 그 시점부터 보유 기간을 60일 연장한다. 보유 주식의 경우에는 60일에 매도되고 다음에 Sig 4와 5가 나올 때 재매수하여 누적수익률을 계산한다. <표 5>의 두 번째 표의 추정된 결과에서 보면 2017년 8월 18일에서 2019년 1월 4일 사이의 실제 투자신호의 누적수익률이 20.14%인 것을 확인할 수 있다. 이 테스트 방법에 따르면 L1 모형 추정된 결과의 누적수익률은 12.23%이다. L1 모형은 단 한 개의 은닉 세포로 인해 모형이 너무 간단하므로 기간 내의 유효 투자신호를 한 개만 방출했다. L2 모형 추정된 결과의 수익률은 28.12%이며, L2 모형의 추정 결과를 보면 실제 신호가 나오면 추정 결과가 실제 신호가 방출되는 날과 유사한 신호를 추정할 수는 없지만, 실제 신호 방출 전후 15일 동안 연속적으로 투자신호를 방출한다는 것을 알 수 있다. L2 모형의 추정 결과를 보면 일정한 지연의 문제가 있지만, 누적수익률에 영향이 없음을 확인할 수 있다.

<표 5>의 유효 매수 신호수와 누적수익률의 결과를 결합하면, 은닉 세포 수 세 개의 인공신경망 모형이 은닉 세포 수 한 개의 인공신경망 모형보다 더 강한 설명력을 가진다는 점을 확인할 수 있다.

4. 투자신호 예측

기존의 MACD 투자전략과 세분화된 MACD 전략의 신호 예측결과를 비교하기 위하여 2019년 1월 4일부터 3월 24일까지의 기간을 다음과 같이 네 구간으로 나누어 결과를 확인하였다.⁸⁾

기간 1: 2019/01/04~2019/01/24,

기간 2: 2019/01/25~2019/02/14,

7) 유효 매수 신호란 목표 실제 신호 전후 15일 내에 추정 신호가 나오는 신호이다.

8) 목표 실제 신호 전후 15일내에 추정 신호가 나오는 신호는 유효신호로 보기 때문에 본연구에서는 15일을 한 구간으로 선택하였다.

기간 3: 2019/02/15~2019/03/05,

기간 4: 2019/03/06~2019/03/24.

본 연구에서는 네 구간의 투자신호를 예측하기 위하여 앞서 고려한 20개 자료의 8개 주성분을 독립변수로 사용하였다. 즉, 투자신호를 예측하기 위하여 모든 구간은 벡터 자기회귀(VAR: Vector Autoregressive) 모형으로 얻어진 예측값을 독립변수로 사용하였으며, 각각 주성분의 1분기 전(lag 1)부터 5분기 전(lag 5)까지의 시차 변수 40개 변수를 사용하여 투자신호를 예측하였다. VAR 모형의 시차는 AIC(Akaike Information Criterion)를 기준으로 선택하였다. 본 연구에서는 VAR(3) 모형을 사용한다.⁹⁾

<표 6>은 VAR(3) 모형에서 추정된 검색어 지수를 사용한 인공신경망 모형으로 2019년 1월 4일부터 3월 24일까지의 기간을 네 구간으로 나누어 기존의 MACD 투자신호와 세분화된 MACD 투자신호를 종속변수로 한 예측결과를 나타낸다. 예측결과에 따라 2019년 1월 4일부터 1월 24일까지의 기간 1에 세분화된 MACD 투자신호 종속변수로 L1 인공신경망을 이용할 때 예측하는 투자신호의 결과에 따르면 매수를 의미하는 Sig 4 이상의 신호는 나타나지 않아 해당 기간에는 투자 기회가 없는 것으로 나타난다.

L1 인공신경망 모형은 간단한 비선형 문제를 처리할 수 있기 때문에 L2 모형에 비하여 매수 신호를 정확하게 예측하지 못하는 경향을 보인다. 하지만 해당 기간에 세분화된 MACD 투자신호를 종속변수로 L2 인공신경망 모형을 이용할 때 투자 기회를 예측하는 경우 매수를 의미하는 Sig 4 매수 신호를 보이고 있다. 따라서 미래 투자신호의 날짜를 정확하게 예측할 수는 없지만, 해당 기간의 매수 기회가 있는지를 예측할 수 있다. 또한, 예측 신호가 실제 신호와 같은 등급의 예측결과가 나타났다. 한편 기존의 MACD 투자신호에 대하여 L1과 L2 인공신경망 모형을 이용할 때 예측하는 투자신호의 결과에 따라 매수를 의미하는 Sig 1 매수 신호는 나타나지 않아 해당 기간에는 투자 기회가 없는 것으로 나타났다. 결과적으로 해당 기간의 예측 신호와 실제 신호 결과가 다르다는 것을 확인하였다. 기간 1의 예측 신호 결과를 보면 세분화된 MACD 투자신호를 종속변수로 사용하여 L2 인공신경망 모형으로 투자 기회를 예측하는 경우에 가장 좋은 예측결과를 보였지만, 이에 반하여 기존의 MACD 투자신호를 종속변수로 예측한 결과 매수 신호를 잘 못 예측하는 결과가 발생하였다.

9) VAR(3)을 추정할 때는 변수들의 로그증감률을 사용해 추정하였다.

<표 6> 투자신호 예측결과(2019/01/04~2019/03/24)

이 표는 2019년 1월 4일부터 3월 24일까지의 기간을 네 구간으로 나누어 투자신호를 예측한 결과이다. VAR(3) 모형을 통하여 추정된 검색어 지수를 사용하여 PCA의 1분기 전부터 5분기 전까지의 일별 데이터 독립변수로 기존의 MACD 투자신호와 세분화된 MACD 투자신호를 종속변수로 추정하는 은닉층 세로 한 개(L1), 세 개(L2)의 인공신경망 모형을 이용하여 예측한 투자신호이다.

날짜	실제신호 (0, 1, 2, 3, 4, 5)	기존의 MACD 투자신호 (0,1)		세분화된 MACD 투자신호 (0, 1, 2, 3, 4, 5)	
		L1	L2	L1	L2
기간 1	2019/01/04	0	0	0	0
	2019/01/07	4	0	0	0
	2019/01/08	0	0	0	0
	2019/01/09	0	0	0	4
	2019/01/10	0	0	0	0
	2019/01/11	0	0	0	0
	2019/01/14	0	0	0	0
	2019/01/15	0	0	0	0
	2019/01/16	0	0	0	0
	2019/01/17	0	0	1	0
	2019/01/18	0	0	0	0
	2019/01/21	0	0	0	0
	2019/01/22	0	0	0	0
	2019/01/23	0	0	0	0
	2019/01/24	0	0	0	0
기간 2	2019/01/25	0	1	0	0
	2019/01/28	0	0	0	3
	2019/01/29	0	0	0	0
	2019/01/30	0	0	1	0
	2019/01/31	0	0	0	0
	2019/02/01	0	0	0	0
	2019/02/04	0	0	1	0
	2019/02/05	0	0	0	0
	2019/02/06	0	0	0	0
	2019/02/07	0	0	0	0
	2019/02/08	0	0	1	0
	2019/02/11	0	0	0	0
	2019/02/12	0	0	0	0
	2019/02/13	0	0	0	0
	2019/02/14	0	0	0	0

<표 6> 투자신호 예측결과(2019/01/04~2019/03/24) (계속)

날짜	실제신호 (0, 1, 2, 3, 4, 5)	기존의 MACD 투자신호 (0, 1)		세분화된 MACD 투자신호 (0, 1, 2, 3, 4, 5)	
		L1	L2	L1	L2
기간 3	2019/02/15	0	0	0	0
	2019/02/18	0	0	0	0
	2019/02/19	0	0	1	0
	2019/02/20	0	0	0	0
	2019/02/21	0	0	0	0
	2019/02/22	0	0	0	0
	2019/02/25	0	0	0	0
	2019/02/26	0	0	0	0
	2019/02/27	0	0	2	0
	2019/02/28	0	0	0	0
	2019/03/01	0	0	0	0
	2019/03/02	0	0	0	0
	2019/03/03	0	0	0	0
	2019/03/04	0	0	0	0
	2019/03/05	0	0	0	0
기간 4	2019/03/06	0	0	0	0
	2019/03/07	0	0	0	0
	2019/03/08	0	0	1	0
	2019/03/09	0	0	0	0
	2019/03/10	0	0	0	0
	2019/03/13	0	0	0	0
	2019/03/14	0	0	0	0
	2019/03/15	0	0	0	0
	2019/03/16	0	0	4	0
	2019/03/17	0	0	1	0
	2019/03/20	0	0	0	0
	2019/03/21	0	0	0	0
	2019/03/22	0	0	0	1
	2019/03/23	0	0	0	0
	2019/03/24	0	1	0	0

2019년 1월 25일부터 2월 14일까지의 기간 2에 세분화된 MACD 투자신호 종속변수로 L1과 L2 인공신경망 모형을 이용할 때 매수를 의미하는 Sig 4 이상의 신호는 나타나지 않아 해당 기간에는 투자 기회가 없는 것으로 나타났다. 하지만 해당 기간 투자신호 Sig 1과 2의 신호가 주어진다는 것을 확인하였다. 해당 기간에 실제 신호의 *DIF*와 *DEA*의

추세선을 교차하지 않아 투자 기회가 없는 것으로 보일 수 있지만, 기존의 MACD 투자신호에 대한 L1 모형의 예측결과에 따르면 매수를 의미하는 Sig 1가 주어져 실제와 다른 결과가 주어진다는 것을 확인할 수 있다. 또한, 기존의 MACD 투자신호의 L2 모형 결과에서는 매수를 의미하는 Sig 1이 빈번하게 나타났다. 이러한 결과에 따라 거래를 하는 경우 과도한 매매를 불러올 수 있다.

VAR(3) 모형을 통하여 추정된 검색어 지수를 독립변수로 예측하는 결과에 따르면 기존의 이항변수 MACD 투자신호는 극단적인 요인에 영향을 받기 쉬우며 이에 따라 높은 거래 비용을 수반할 수 있다는 것을 확인하였다. 하지만 세분화된 MACD 투자신호를 사용하는 경우 신뢰할 만한 투자신호 예측결과를 보여준다는 것을 확인하였다.

2019년 2월 15일부터 2019년 3월 5일까지의 기간 3에 세분화된 MACD에 대하여 L1와 L2 인공신경망 모형을 사용한 결과 매수 신호는 나타나지 않았다. L1 인공신경망 모형을 사용한 기존 MACD 신호의 결과에서 매수를 의미하는 Sig 1이 나타나지 않고 이러한 결과는 실제와는 일치하게 나온 결과임을 볼 수 있다. 하지만 L2 인공신경망 모형을 사용한 기존 MACD 신호의 결과에서 매수를 의미하는 Sig 1이 확인되었고 이러한 결과는 실제와는 다른 결과임을 볼 수 있다.

2019년 3월 6일부터 2019년 3월 24일까지의 기간 4에 대하여 L1 모형에 세분화된 MACD를 사용한 결과 Sig 4가 주어졌지만, 이는 실제 신호와는 다른 결과임을 확인하였다. 하지만 L2를 사용한 결과 Sig 4와 Sig 5와 같은 매수 신호는 주어지지 않음으로 실제 결과와 동일한 결과를 보인다는 것을 확인하였다. L1과 L2에 대하여 기존의 MACD를 사용한 결과 실제로는 매수 기회가 없음에도 불구하고 매수 신호를 보임으로써 실제와 일치하지 않는 결과를 보인다는 것을 확인할 수 있었다.

2019년 1월 4일부터 2019년 3월 24일까지의 모든 예측 구간에 대하여 L2 모형에 세분화된 MACD를 사용한 결과 과도매매 문제를 해결함과 동시에 매수 신호를 더욱 정확하게 예측한다는 것을 확인할 수 있었다. L1 모형에서는 과도매매 문제는 나타나지 않았지만 실제 신호를 정확하게 예측할 수는 없었다. 이러한 세분화된 MACD와는 달리 기존의 MACD의 경우 모두 과도매매 문제를 보였으며, 매수 신호 또한 정확하게 분석할 수 없었다.

네 개 구간의 예측결과에 따르면 기존의 MACD 신호를 인공신경망 모형으로 예측하는 경우, 부정확한 예측결과로 인하여 투자 기회의 올바른 판단이 어렵다는 것을 보았다. 또한, 기존 신호의 경우 매수 신호가 빈번하게 주어져 높은 거래 비용이 수반될 수 있다는 것을 확인할 수 있었다. 하지만 본 연구에서 사용된 세분화된 MACD 신호를 사용한 결과 해당 구간의 투자신호를 보다 명확하게 예측할 수 있었다. 이와 함께 매수 신호가 없음에도

불구하고 매수 신호를 보인 기존의 결과와는 다르게 Sig 1-3의 신호를 보임으로써 무의미한 거래를 감소시킬 수 있는 효과가 있다는 것을 결과를 통하여 확인하였다.

V. 결 론

본 연구에서는 인공신경망 모형을 사용하여 MACD 지표를 예측하고 이를 통하여 투자전략을 구성할 수 있는 방법을 제안하였다. MACD 지표에 따라 투자신호를 분석하기 위하여 SSE50 지수를 사용하였으며, 지수의 이동 평균값의 변화에 따라 투자신호를 Sig 0-5의 여섯 개 신호로 구분하였다. 신호에 따라 투자전략을 구성하고 이를 통한 결과를 비교하고자 Sig 4 이상의 신호에서 매수하여 60일 이상 보유하는 장기투자전략으로 결과를 살펴보았다. 이와 같은 투자신호를 추정하기 위하여 20개 바이두 지수의 검색어를 독립변수로 사용하여 신호를 추정하였다. 또한, 미래의 투자신호를 예측하기 위하여 VAR 모형을 사용하여 얻어진 검색어 자료의 예측값으로 투자신호를 예측하고 이를 통한 수익률을 함께 비교하였다.

인공신경망 모형의 추정 결과에 따라 투자를 하는 경우 실제값에 가까운 높은 수익률을 얻을 수 있다는 것을 확인할 수 있었다. 또한, 인공신경망 모형의 은닉 세포 수가 증가할수록 정확한 예측으로 상대적으로 높은 수익률을 얻을 수 있다는 것을 확인하였다. 검색어 자료의 예측값으로 투자신호를 예측한 결과에 따르면 해당 기간에는 세분화된 MACD 투자신호를 종속변수로 사용하여 인공신경망으로 미래 투자 기회를 예측하는 결과가 기존의 MACD 투자신호를 종속변수로 사용하는 경우보다 상대적으로 정확하게 신호를 분석하는 것으로 나타났다.

본 연구는 기존 연구와는 다르게 투자신호를 여섯 개 신호로 세분화함으로써 투자자의 성향에 따라 투자전략을 다르게 구성할 수 있는 투자 방법을 제시하였다. 또한 시장이 강세일 때와 약세일 때의 신호를 구분하여 투자전략의 정확성을 높이고자 하였다. 인공신경망 모형은 대용량의 복잡한 자료 분석에 적합한 모형으로써 이번 연구의 분석모형으로 사용되었다. 결과적으로 세분화된 투자신호를 인공신경망 모형에 사용함으로써 기존의 이항 투자신호에서 나타나는 과도매매 문제를 해결할 수 있었으며, 투자신호를 보다 정확하게 예측함으로써 높은 투자 성공률을 얻을 수 있었다.

본 연구를 바탕으로 몇 가지 흥미로운 연구 방향을 제시하면 다음과 같다. 첫 번째로, 최근에 많은 주목을 받고 있는 딥러닝(deep learning)의 주요한 모형인 잡음제거 오토인코더(stacked de-noising auto encoder, Vincent et al., 2010) 신경망모형을 투자신호예측에

사용함으로써 투자신호의 예측력을 높일 수 있을 것이라 생각된다. 두 번째, 본 연구의 한계점이라 할 수 있는 갑작스러운 시장 변화에 따른 위험을 감지할 수 있도록 학습을 통한 신호예측모형을 개발하는 것이다. 마지막으로 현재 기계학습에 대한 많은 모형이 대두되고 있는바 이러한 여러 가지 모형들의 결합(mixture)을 통하여 더 뛰어난 신호예측 모형을 설정하는 것도 의미 있는 연구 방향이 될 수 있을 것이다.

참 고 문 헌

- 陳影, 王立夏, 周智敏, “投資者情緒對股票价格指数影響研究－以百度指數與微博指數為指標,” 价格理論與實踐, 9, (2017), 56-59.
- 賈宇杰, “MACD指標在股市實戰中的應用及其存在的問題研究,” 中國商論, 12, (2018), 34-36.
- 李愛泉, “基于KDJ指標的期貨程序化交易策略研究,” 广西質量監督導報, 4, (2019), 79-81.
- 楊超, 姜昊, 雷崢嶸, “基于文本挖掘和百度指數的匯率預測,” 統計與決策, 13, (2019), 85-87.
- 俞慶進, 張兵, “投資者有限關注與股票收益－以百度指數作為關注度的一項實証研究,” 金融研究, 8, (2012), 120-138.
- Ahmed, F., R. Asif, S. Hina, and M. Muzammil, “Financial Market Prediction using Google Trends,” *International Journal of Advanced Computer Science and Applications*, 8(7), (2017), 388-391.
- Appel, G. and W. F. Hirschler, *Stock Market Trading Systems*, Traders Press: New York, 1990.
- Barber, B. M. and T. Odean, “The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investor,” *Review of Financial Studies*, 21(2), (2008), 785-815.
- Bishop, C. M., *Neural Networks for Pattern Recognition*, Clarendon Press: London, 1995.
- Cauchy, A., “Methode Generale pour la Resolution des Systemes D'equations Simultanees,” *Comp. Rend. Sci. Paris.*, 25, (1847), 536-538.
- Da, Z., J. Engelberg, and P. Gao, “In Search of Attention,” *Journal of Finance*, 66(5), (2011), 1461-1499.
- Gervais, S., R. Kaniel, and D. H. Mingelgrin, “The High-volume Return Premium,” *Journal of Finance*, 56(3), (2001), 255-279.
- Granville, J. E., *New Key to Stock Market Profits*, Martino Publishing: New York, 1963.
- Hornik, K., M. Stinchcombe, and H. White, “Multilayer Feed Forward Networks are Universal Approximators,” *Neural Networks*, 2(5), (1989), 359-366.
- Mendelsohn, L. and J. Stein, “Fundamental Analysis Meets the Neural Network,” *Future*, 20(10), (1991), 22-24.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams, “Learning Representations by Back-Propagating Errors,” *Nature*, 323(6088), (1986), 533-536.

- Shanker, M., M. Y. Hu, and M. S. Hung, "Effect of Data Standardization on Neural Network Training," *Omega*, 24(4), (1996), 385-397.
- Song, S. j., H. Cao, and K. Yang, "Investor Attention and IPO Anomalies – Evidence from Google Trend Volume," *Economic Research Journal*, 1, (2011), 145-155.
- Vincent, P., H. Larochelle, I. Lajoie, I. Bengio, and P. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," *Journal of Machine Learning Research*, 11, (2010), 3371-3408.

THE KOREAN JOURNAL OF FINANCIAL MANAGEMENT
Volume 37, Number 1, March 2020

Estimation and Analysis of Investment Signal for SSE50 Index Using Neural Network*

Ling Jie** · Sung Y. Park***

〈Abstract〉

This research demonstrates the development of an investment strategy model based on investment signal analysis using artificial neural networks. In order to analyze the investment signal analysis based on the stock price change, the China Shanghai Stock Exchange Index and the Baidu Index of the daily unit from June 1, 2006 to January 11, 2019 were used. Based on the MACD index, we will take both the short market and long market effects into consideration, and put forward a differentiated investment strategy model which can classify the investment opportunities from the lowest signal level 0 to the highest signal level 5. Meanwhile, it is for sure that the predicted results of the model are similar to the actual results as we use the big data of Internet search index as the independent variable of artificial neural network. In the process of predicting the model, the period from January 4 to March 24, 2019 is divided into four time intervals for prediction. And according to the result, the investment strategy prediction based on different levels of signals can predict the future investment opportunities better than the previous investment strategy model, and certainly reduce the meaningless transactions as well.

Keywords : Investment Signals, Baidu Index, Artificial Neural Network, MACD, Big Data

* This research was supported by the Chung-Ang University Research Scholarship Fund in 2019.

** First Author, MA, School of Economics, Chung-Ang University, E-mail: lingjie993@cau.ac.kr

*** Corresponding Author, Associate Professor, School of Economics, Chung-Ang University,
E-mail: sungpark@cau.ac.kr