

머신 러닝 기반 소셜 빅데이터 분석을 이용한 금융자산 트레이딩 모델의 성능 향상에 관한 연구

A Study on Improving Financial Asset Trading Model Using Social Big Data Analysis Based on Machine Learning

송성환 · 한경석 · 최수정 · 박성민

Song, Sunghwan · Han, Kyeongseok · Choi, Sujung · Park, Sungmin

Abstract

머신 러닝 기반 소셜 빅데이터 분석을 이용한 금융자산 트레이딩 모델의 성능향상에 관한 연구는 기존 연구들에 근간하여 사회를 구성하는 구성원들의 집단감성이 주가에 영향을 미친다는 가정하에서 출발했다. 최근 소셜 데이터가 양적인 측면에서 기하급수적으로 증가함에 따라 실시간으로 변화하는 사회구성원의 집단감성의 대표성을 갖게 되었다. 최근의 연구에 따르면, 인간의 의사결정이 이성과 깊은 성찰로부터 기인하기보다는 감성이 더 깊이 관여한다고 진단하고, 이러한 의사결정 모형을 경제사회로 확장시켜 Socio-economics라 명명한 바 있다. 본 연구는 시가총액이 한국 전체 주식시장의 70%를 차지하는 KOSPI200을 대상으로 한다. 본 모델은 자연언어 처리를 이용하여 소셜 빅데이터로부터 주가의 전망과 관련성이 깊은 감성데이터를 추출한다. 머신 러닝을 활용하여 각 기업의 Fundamental Ratio와 Technical Indicators, 소셜 감성데이터 등을 러닝하여 주식 시장을 전망한다. 현재 주요 펀드들의 운용에는 여러 시장 지표 데이터들이 활용되지만, 일반적으로는 펀드매니저 개개인의 지식과 경험을 활용한 통찰이 펀드의 운용에 결정적인 영향을 미친다. 본 모델에서는 사람의 개입 없이 머신 러닝을 통해 발굴된 종목들을 기반으로 포트폴리오를 형성한다. 이번 연구를 통해 소셜 빅데이터 분석을 이용한 머신 러닝 기반의 로봇 트레이딩이 특정 기간 동안 한국의 주식시장에서 실제 시장지표 대비 높은 수익률을 보임을 확인했다.

주제어 : 머신 러닝, 빅데이터 분석, 금융자산, 트레이딩

This research aims to improve the performance of machine learning based on financial trading model applying social big data analysis. Based on the existing studies, it started from the assumption that collective emotions of society members have an impact on stock prices. As social data has been rapidly increasing to the extent that it can be called as 'big data', it could represent the real time flow of emotion from society members. According to recent studies, it is evident that people's decisions result from neither rationality nor considerate introspection but emotion. This kind of decision making model has been extended to economics which was named as 'Socio-economics'. This study is conducted with KOSPI200 which occupies 70% of total Korea stock market capitalization. This model uses Natural Language Processing to extract emotion data closely related to stock prediction from social big data. It utilizes machine learning to learn fundamental ratio and social emotion data together with technical indicators of each company. On the current market, various indices are referred in managing major funds. However, it usually depends more on insights and knowledge of individual fund managers. In this study, portfolios are formulated based on confirmed stock items from machine learning without human intervention. Through this research, we proved machine learning based robot trading applying social big data analysis outperform higher than actual Korean stock market indices in a certain era.

Keywords : Machine Learning, Big Data, Financial Trading, Trading

송성환: 숭실대학교, 일반대학원 IT정책경영학과, 박사과정(주저자)

한경석: 숭실대학교, 경영대학, 교수(교신저자)

최수정: 숭실대학교, 경영대학, 조교수(공저자)

박성민: 숭실대학교, 일반대학원 경영학과, 석사과정(공저자)

1. 서론

전통적으로 주식시장의 움직임은 예측하기 힘들기로 악명이 높다. 최근 20여 년 동안 급격한 컴퓨터분야의 기술에 힘입어 금융기관과 학술연구자들은 위험을 최소화하고 이윤을 극대화하기 위한 수학적 모델을 개발하기 위하여 힘써 왔다[8]. 최근의 연구에 따르면 인간의 의사결정이 이성과 깊은 성찰로부터 기인하기보다는 감성이 더 깊이 관여한다고 진단하고, 이러한 의사결정 모형을 경제사회로 확장시켜 Socio-economics와 연관지어 설명하는 연구가 활발하다[5]. 이는 트위터로부터 그 사회집단의 감성을 추출하고 그 감성이 그 사회의 전체의 감성에 통계적으로 유의미한 대표성을 갖고 시장을 예측할 수 있다는 가설이다[4]. 본 연구에서는 이러한 이유로 트위터를 우리사회의 감성을 실시간으로 읽어 낼 수 있는 척도로 보았다. 또한 주식시장의 시장 참가자인 개개인은 이러한 사회적 집단 감성에 크게 좌우될 것이라고 가정하였다. 매일 발생하는 트윗들을 저장하고 처리하기 위하여 하둡을 이용한 빅데이터 처리 시스템을 구축하였다. 또한, 이렇게 저장된 빅데이터를 분석하기 위하여 정교한 자연언어처리시스템을 구축하였다. 이렇게 분석된 감성데이터와 주식데이터의 기본정보 및 기술적 지표를 바탕으로 기존에 시장에서 감지할 수 있었던 데이터와 기존에 시장에서 감지할 수 없었던 감성데이터를 융합하여 주가를 예측하고 실제로 투자하며 지속적으로 포트폴리오를 개선하는 인공지능 트레이딩 시스템을 구축하였다.

주식 트레이딩 분야에서 머신 러닝은 주로 과거 데이터를 통한 패턴 인식과 미래 예측을 위해 활용되어 왔다. 또한, 최근에 설립되는 핀테크 회사뿐 아니라 보수적인 기존의 다국적 금융회사인 골드만 삭스도 사회적 감성과 주식시장의 연관성에 주목하고 사회적 감성을 추출하는 기술과 인공지능 기술과 결합하여 새로운 상품과 서비스를 연구 개발하고 있다[9].

본 연구에서는 한국사회의 사회적 감성을 추출하고 추출된 사회적 감성과 한국의 대표적인 주식시장지표인 KOSPI200과의 상관관계의 유효성을 검증하고 주식시장의 고전적인 지표인 개별종목의 기본 재무적 비율과 기술적 지표를 바탕으로 머신 러닝을 통하여 주식시장을 예측하고자 한다. 또한 트레이딩 모델은 위험을 회피하고 최적의 포트폴리오를 구성하는 것을 목표로 한다. 포트폴리오 알고리즘은 대체로 시장에 대한 예측치를 요구하는데, 본 연구에서는 주식시장 관련 주요 지표와 소셜 빅데이터 바탕으로 머신 러닝을 적용하고 그 결과로 시장에서의 개별종목에 대한

예측치를 얻고자 한다. 이렇게 구성된 머신 러닝 기반 소셜 빅데이터 분석을 이용한 트레이딩 모델에 대한 검증이 이 연구의 목표이다.

2. 이론적 배경

2.1 소셜 빅데이터 기반 주식시장 예측이론

2.1.1 Bollen의 이론[5]

소셜 미디어 상에 표현된 인간의 감성이 주식시장에 미치는 영향에 관한 연구이다. Bollen의 논문은 개인이 아닌 사회의 감성 상태 또한 그 사회 구성원의 집단적 의사결정에 영향을 주는지에 관한 연구이다[5].

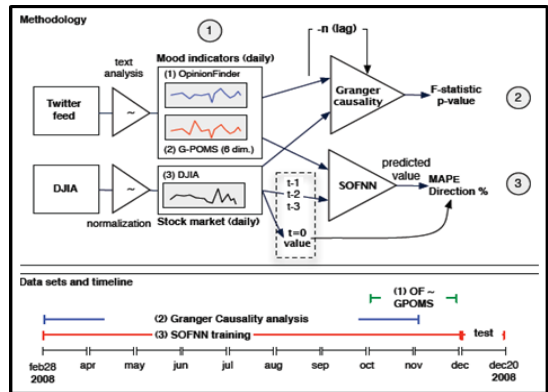


그림 1. Bollen의 소셜 감성과 주가와의 상관관계[6]

이러한 확장을 통하여 군중의 감성이 그 사회를 대변할 뿐 아니라 경제지표에 선행하여 이를 예측할 수 있을가에 대한 연구이다. 이를 논증하기 위해서 군중의 감성은 트위터를 사용하였으며 이를 다우존스산업평균과의 상관관계를 장기간에 걸쳐 비교 분석하였다. 매일 트위터를 수집하였고 Opinion Finder라는 툴을 사용하여 긍정어와 부정어를 판별하였고, 구글의 감성상태지표(GPOMS: Google-Profile of Mood States)를 활용하여 감성을 6개로 나누고 이 모델을 미국 대선 데이터를 통하여 검증하였다. 결과적으로 Bollen은 실험을 통하여 매일 DJIA 등락예측을 함에 있어 오차범위 6% 내외에서 87.6%의 예측 정확도를 보였다[5].

영국 UCL의 Th'arsis T. P. Souza는 '비선형 영향: 시장가격에 영향을 미치는 소셜 미디어의 효과(A Non-linear Impact: Evidences of Causal Effects of Social Media on Market Prices)'를 통하여 소셜 데이터가 기존에 해석하기 힘든 비선형적 주가 흐름에 대한 예측에 매우 유의

함을 검증하였다.

2.2 지식 기반 예측 접근법

지식 기반 예측 접근법은 투자에 있어서 인간이 가진 기반지식을 이용하여 시장을 전망하는 것을 통칭한다. 일반적으로 투자자는 투자를 함에 있어서 전문가의 평가에 의존하는 경향이 크다. 투자자들이 의존하는 전문가는 일반적으로 시중 증권사의 애널리스트이다. 증권사의 애널리스트는 개별기업이나 시장에 대한 재무분석, 시장분석 그리고 주식시장의 기술적 지표에 의거한 분석을 통하여 기업을 평가하고 의사결정을 한다. 재무분석은 기업의 재무제표를 보고 기업의 기대수익을 예측하고 현재 기업의 공시자료를 이용하여 성장전망을 통해 기업을 평가하는 것이다. 시장분석은 크게 시장의 수요와 공급의 균형관계를 분석하여 해당기업이나 섹터의 성장성을 전망한다. 하지만, 금융기관은 위험을 회피하고 수익을 극대화하기 위한 상품을 설계해야 한다[10]. 상품을 설계하는 데 있어서 필수적으로 사용하는 것이 포트폴리오전략이다. 일반적으로 포트폴리오는 주식에만 투자하는 것이 아니라, 채권과 같은 위험회피자산이나 부동산이나 인덱스, 또는 산업 군으로 묶는 섹터분산, 해외분산 등 다양한 방법으로 이루어지게 된다. 개인의 지식과 경험에 의한 투자전망을 자산 배분 전략으로 이용하여 최종적인 최적 자산 배분치를 얻게 된다.

2.3 통계 기반 예측 접근법

주식시장을 통계적으로 분석해보면 주가의 흐름이 유사한 종목이 많이 있다. 그래서 우리는 주가를 예측함에 있어서 통계적 방법을 통한 접근법을 사용하고자 한다. 본 연구에서는 크게 통계적 방법과 인공지능적 방법 두 가지를 이용하여 주가를 예측하고자 하는 것이다. 통계적으로 볼 때 유사한 업종에 속하는 종목들은 특별한 악재 또는 호재가 없고, 주가 자체가 주로 기업의 실적에 영향을 받는다고 가정할 때, 상호간에 유사한 흐름을 보이는 것이 일반적이다. 종합적으로 판단해 볼 때 항상 시장이 변화할 수 있음을 염두한 상태에서 적절한 시점에서 시장의 트렌드를 반영하여 예측할 수 있는 모형 수립이 중요함을 알 수 있다[3]. 본 연구에서는 별도의 알고리즘을 통해 가장 적절한 평균 회귀 전략을 선정하여 균형 가격을 계산하고 이를 활용하여 주가 수익률을 예측하는 모형을 만들었다. 평균 회귀 전략은 머신 러닝을 기반으로 시장을 전망할 때 발생할 수 있는 주식시장 고유의 특성을 보완하고자 하는 목적으로 이용할 것이다.

2.4 머신 러닝 기반 예측 접근법

머신 러닝 기반 예측 접근법은 본 연구에서 이용하고자 하는 데이터의 특성과 깊은 관련이 있다. 앞서 살펴본 지식 기반 예측 접근법의 경우 비록 보조적 지표로 통계적 방법을 이용한다 하더라도 주식시장의 전망에 대한 예측치의 정확도가 애널리스트나 상품 설계자의 식견에 의하여 크게 좌우됨을 관련 연구를 통해 확인한 바 있다[11]. 이용하고자 하는 데이터는 소셜 빅데이터와 주식시장에서 제공하는 많은 지표데이터를 기반으로 예측을 하고자 하기 때문에 기본적으로 개별주식에 대한 수많은 데이터를 양적 측면에서 인간이 지속적으로 보기 힘들다고 생각한다. 본래 인공지능의 대표적인 방법이었던 전문가 시스템은 접근 기반 방식으로서 사람이 알고 있는 경험치를 방식의 형태로 구성하는 것을 말한다. 이와 같은 접근은 과학적 방법론에 기반하여 인간이 지금까지 발견한 일반화된 과학적 규칙들을 일종의 데이터베이스 형태로 등록하고 유지하고 관리하는 것을 말한다.

앞서 언급한 여러 알고리즘 중에서도 SVM(Support Vector Machine)은 [13]에 의하여 개발된 학습 알고리즘으로서 현재 다양한 분야에 활용되고 있는데, 주식, 선물과 같은 노이즈가 많고 비정상적 특징이 있는 금융데이터에 분석에도 신경망 등 기존 모델보다 더 나은 성능을 보이는 것으로 기존 연구에서 밝혀진 바 있다[6][7]. 이에 금융 분야에서 가장 적합하다고 확인되어 널리 쓰이고 있다.

2.5 자산 배분 전략

초기에 자산 배분은 보다 나은 투자를 통하여 더 많은 수익을 얻고자 하는 아이디어에서 출발하였으나 인류가 수세기에 거쳐 세계적 경제위기를 경험하면서 자산배분 모델의 역사는 자산배분초점을 맞춘 초기이론에서 점차 자산의 위험을 배분함으로써 위험을 최소화하는 전략으로 진화중이다[1][2].

가장 잘 알려진 마코위츠의 평균분산 모형은 특정 주식에 자산 배분이 많이 되는 쏠림 현상이 나타날 수 있고 재 표본 추출 모형은 수학적 근거가 약하다는 단점이 있다. 부트스트랩 방법을 통한 분산 투자 모형은 복잡도와 시간 비용에서 약점을 가진다. 따라서 본 연구에서는 블랙 리터만 모형을 사용하여 SVM과 평균 회귀에 따른 예측치를 분산 투자 모형에 투입하여 최적의 자산 배분안을 도출한다.

투자자산배분모형을 보다 자세히 살펴보자면, 먼저,

평균-분산 모형은 세 단계로 이루어 진다. 첫째로, 자산군을 선정하는데 이것은 전략적인 자산배분의 기반이 된다. 둘째, 변수를 예측한다. 여기서 요구되는 변수는 기대수익률과 기대위험과 자산 간 상관관계수 등이다. 마지막으로, 선정된 자산과 예측된 변수를 활용하여 최적화 알고리즘을 적용하여 포트폴리오를 얻게 된다. 이는 투자자의 예측 능력에 의존하여 기대수익률, 표준편차, 그리고 포트폴리오 내의 다른 자산과의 상관관계가 결정된다는 것이다. 이러한 이유로 포트폴리오가 어느 한 자산에 집중적으로 배치되어 비현실적인 평균 포지션이 구성될 수 있다는 점이다. 두 번째로, 변수에 미쳐한 변동에도 크게 반응하는 포트폴리오의 안정성이 문제가 된다. 이는 입력된 변수에 오류가 있는 경우 포트폴리오에 극단적인 영향을 줄 수 있는 위험이 있다. 셋째로, 통계를 기반으로 하기 때문에 직관적이지 못하다. 이는 포트폴리오가 위험적 기여도의 특성을 통계적 보수성으로 인하여 직관적으로 반영하지 못하는 결과를 낳는다. 앞서 제기한 기대수익률에 관한 예측오류 문제를 해결하기 위한 방법론으로 블랙 리터만(Black-Littermann) 모형을 제시한다.

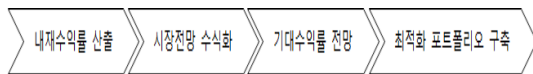


그림 2. 블랙 리터만 모형 구축순서

2.6 자본 자산 가격 결정 모형

2.6.1 CAPM

CAPM(Capital Asset Pricing Model)은 자본비용을 계산하는 모형이다. 또한 CAPM은 가격을 결정하는 모형이다. Capital Asset은 투자자본인데 산업사회에서 투자자본은 투기자본과 달리 장기간에 걸쳐 유의미한 생산활동에 쓰이므로 자본을 장기간 사용하는 자원으로 볼 수 있다[12].

CAPM은 이론적으로 리스크 프리미엄 계산방법을 제시하였으며 여기에서 파생된 켄넬의 알파를 사용해서 포트폴리오 성과를 검증할 수 있다. 켄넬은 그의 논문에서 가능한 모든 매매 전략을 사용하더라도 시장 대비 초과 수익률인 알파가 존재하지 않음을 증명하려 했다.

2.6.2 Jensen's Alpha

Jensen's Alpha는 펀드의 수익률이 시장의 평균적인 수익률, 즉 균형상태에서의 수익률에 비하여 높은 정

도를 알 수 있는 지표이다. 이는 펀드의 수익률에서 시장평균수익률을 뺀 값을 나타낸다. 따라서, Jensen's Alpha가 높을수록 시장 대비 펀드가 우수하다는 것을 알 수 있다.

3. 연구모형

본 연구의 연구모형은 <그림 3>과 같다. 본 연구의 연구모형은 2장의 관련 연구를 통하여 살펴본 바와 같이 소셜 네트워크 서비스에 표현된 감성이 주식시장을 예측한다는 이론을 추가예측에 반영하기 위하여 주식 시장 관련 여러 지표들과 소셜 네트워크 서비스로부터 추출된 감성을 학습시키는 머신 러닝을 활용한다. 머신 러닝의 과적 합을 보조하기 위하여 통계 기반 모델인 평균회귀전략(MRS: Mean Reversion Strategy)을 사용한다[4].

SVM에 대해서 깊이 살펴보면, 기존 신경망 계열의 모형에서는 입력데이터와 출력데이터 간의 오류율을 최소화하는 데 중점을 둔 반면 SVM에서는 이와 더불어 두 데이터 간의 여백을 최대화 하여 일반화 능력의 극대화를 목표로 둔다.

SVM은 비교적 높은 분류 정확도를 가지고 있다. 추가 예측에서 은닉층이 1개 이상인 다층 퍼셉트론(Multiple Layer Perceptron: MLP) 방식의 예측보다 SVM을 사용한 추정 결과가 더 높은 정확도를 나타낸다. 또한 인공신경망과 지지벡터회귀분석 모형과 비교하여 볼 때에 SVM 모형이 더 좋은 예측력을 나타내고 있다. 나아가서 SVM을 단독으로 사용할 때보다 유전자 알고리즘과 SVM을 동시에 사용한 하이브리드 추가예측 모형을 추정한 결과가 더 우수한 성과를 보였다.

머신 러닝에 필요한 최적의 속성들만을 SVM의 입력치로 필터링하기 위하여 추가적으로 유전자알고리즘을 활용하고 학습구간을 N개의 구간으로 나누어 특정기간의 특이점이 전체 학습치를 오염시키는 것을 방지하기 위하여 N-Fold Validation Theory를 이용한다. 최적의 자산배분을 얻기 위하여 블랙 리터만 모형을 이용한다. 최종적으로 본 모형과 타 모형을 비교하고 검증하기 위하여 CAPM을 활용한다. CAPM을 통하여 검증된 값을 바탕으로 시장성과 대비 머신 러닝 기반 소셜 빅데이터 분석을 이용한 금융자산 트레이딩 모델에 피드백하여 지속적으로 모델을 개선하는 것이 본 연구의 연구모형이다[12].

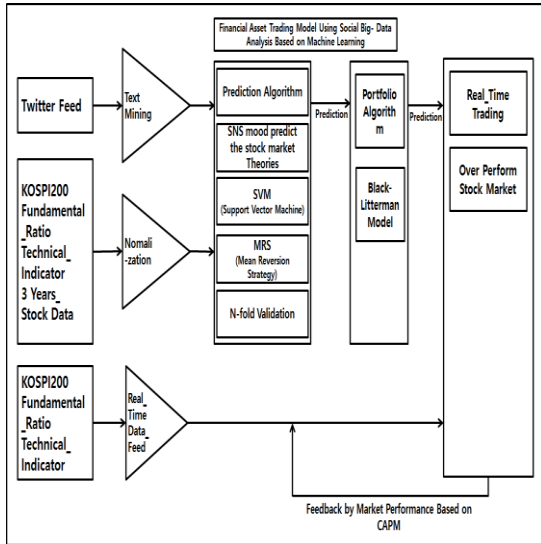


그림 3. 연구모형

세계적으로 가장 즉각적이고 파급력이 큰 소셜 네트워크 서비스인 트위터가 한국에서도 의미 있는 숫자로 증가한 시기가 2011년이다. 앞서 서론에서 전술한 바와 같이 소셜 네트워크 서비스상에서 표현되는 집단감성과 주가 사이의 상관관계를 분석하기 위하여 2011년 01월부터 2014년 10월까지의 국내에서 생성된 모든 트위터 데이터와 동 기간의 주가 데이터 비교를 통해 실험을 한 것도 트위터 데이터로부터 추출된 집단감성이 우리사회의 보편적 감성이라는 가정에서 출발한 것이다[5]. 트위터 데이터를 수집하기 위하여 트위터사에서 제공하는 REST API와 Streaming API를 사용하였다. 수집된 트위터 데이터는 그 양이 2014년 기준으로 일 700만 건에서 1000만 건까지 이를 정도로 방대하기 때문에 가장 보편적으로 활용되는 빅데이터 저장구조인 하둡 시스템을 이용하여 저장된다. KOSPI200, 선물 옵션 등 주식시장과 관련 데이터에서 75개의 키워드를 선정하였다. 시장관련 데이터들을 긍정어와 부정어로 나누어 소셜 데이터를 사용하여 차달의 시장을 예측하였다. 하둡 시스템에 저장된 트위터 데이터는 온톨로지 시스템에 기반한 자연언어처리 시스템을 통하여 분석된다. 자연언어처리 시스템을 통하여 분석된 데이터는 각각의 데이터의 속성에 따라서 다양한 감성체제로 분류된다. 또한 주식사이트를 통해 KOSPI200에 대한 기업별 기초비율(Fundamental Ratio)과 각 기업의 일별 주가데이터 또한 수집된다. 트위터 데이터는 국내에서 생성된 데이터를 준거로 전수 데이터를 사용하였다. 주식데이터는 증시비중의 70% 이상을 차지하는 KOSPI200 기업 중 일거래량 5만

주 이하의 기업과 부실기업을 제외한 종목을 활용하였다.

3.1 예측 알고리즘

3.1.1 예측 알고리즘 개요

SVM을 통한 주가 예측에서 주요 변수인 사회적 집단감성을 추출하는 것은 매우 중요하다. 본 연구에서는 소셜 네트워크 서비스를 통해 시장전망을 얻고 집단감성을 파악하기 위하여 필수적인 자연언어처리를 이용하였다. 자연언어처리에 사용될 감성들은 추출 기간, 포함어, 제외어에 따라 긍정, 부정이 변화하기 때문에 주가 수익률을 잘 설명할 것으로 예상되는 단어들을 포함어에 추가하고 노이즈가 지속적으로 발생하는 단어들을 제외어로 선정하였다. 거시경제지표에 대한 긍정, 부정 값도 동시에 얻어서 최종적인 소셜 빅데이터가 전망하는 1개월 후의 주식시장에 대한 긍정, 부정 전망을 얻는다. 더불어 거시경제지표에 대한 100,000건당 빈도수를 빅데이터 분석을 통해 얻게 된다. 통상적으로 2016년 02월 현재 국내에서 발생하는 트위터 수는 하루 1천만 건 안팎이다. 본 연구에서는 소셜 빅데이터로부터 추가로 월별 아파트 매매 가격 지수, KOSPI, 원 달러, 콜금리, CD 및 국민주택채권금리 금리와 단기금리, 환율 등 거시경제지표와 관련된 단어를 추출하여 트위터 데이터 100,000건당 빈도로 정규화하여 시장전망 데이터로 활용한다. 소셜 빅데이터의 분석에는 자연언어처리기술을 활용하는데, 여기에는 불필요한 데이터를 제거하는 스팸 필터링(Spam Filtering) 기술과 유의미한 데이터를 추출하는 텍스트 마이닝(Text Mining) 기술 그리고 추출된 데이터로부터 긍정, 부정 값을 얻어 내는 감성분석 기술과 상관관계 분석 기술 등이 활용된다.

소셜 감성분석은 기존의 연구[3]들은 키워드 자체를 긍정, 부정, 중립으로 구분하는 데 그쳤으나 본 연구에서는 키워드분석에서 그치지 않고 자연언어처리를 통한 주술관계 또는 목적어와 술어의 구분분석을 통하여 예측의 정교함을 더하였다. 예를 들면 코스피라고 하는 단어 자체에는 긍정 혹은, 부정의 감성이 존재하지 않으므로 ‘코스피가 오를 것으로 예상된다.’라는 문장에서 주어를 ‘코스피’로 ‘오를 것으로 예상된다.’를 긍정감성으로 뽑아내기 위하여

- 1) 트위터사에서 제공하는 Rest API와 Streaming API를 이용하여 문서를 수집한다.
- 2) 수집된 문장들을 개별 문장으로 분리한다.
- 3) 분리된 개별 문장들의 형태소를 분석한다.

4) 분석된 형태소에 품사를 태깅한다.

5) 그리고, 표현을 정규화한다. 예를 들면, ‘오를 것으로 예상된다.’ 와 ‘오를 것이다.’ 또는 ‘상승할 것으로 전망한다’를 모두 ‘오를 것이다.’라는 표현으로 인식 하는 것이다.

6) 다음으로는, 구문단위화를 거친다. 예를 들자면 “저평가된 종목을 발굴한다.”를 “저평가된 종목”을 발굴 한다.”로 단위화한다.

7) 마지막으로 구문분석을 거치는데, 여기서 분석대상 문장의 긍정과 부정을 결정한다.

이러한 언어분석(Natural Language Processing)을 위해서는 다양한 사전이 필요한데, 본 연구에서는 형태소 사전, 언어통계데이터, 언어규칙, 구문사전, 온톨로지 등이 사용되었다.

특정한 달의 소셜 감성 E_t 는 해당 월의 긍정 감성 P_t 를 긍정 감성 P_t 에서 부정 감성 N_t 를 뺀 값으로 나눈 값이다.

$$E_t = \frac{P_t}{P_t - N_t}$$

본 논문에서 사용한 독립변수인, 특정 달의 소셜 감성 예측치는 ΔE_t 가 양의 값을 가질 때 긍정의 전망을 음의 값을 가질 때는 부정의 전망을 하게 된다.

$$\Delta E_t = E_t - E_{t-1}$$

본 소셜 빅데이터의 분석규모는 일 평균 1천만 건을 기준으로 불필요한 스팸 트윗을 제거한 5년치 데이터 약 200억 건을 기준으로 분석하였다. 분석을 위한 서버는 약 300대의 하둡 플랫폼을 기반으로 하였으며, 분석데이터의 총량은 약 1페타바이트(Peta Byte)이다.

시물레이션 결과 이상의 방법으로 예측에 사용된 주식은 총 119개 종목이며, 총 10개월간의 결과를 바탕으로 매수 후 보유보다 16% 초과 수익을 보였다.

두 번째로 평균 회귀 전략을 사용하였다. 총 76개 종목을 대상으로 분석하였으며 SVM 결과와 마찬가지로 10개월 동안 실험하였다. 그 결과 매수 후 보유보다 약 27%의 초과 수익률을 보고하였다. 하지만 시물레이션에서는 매수 포지션만을 수행하였으므로 만약 주가가 예상된 것과 반대로 저평가 지점에서 더욱 하락할 경우 막대한 손실을 불러일으킬 수 있다. 따라서

손실과 포트폴리오의 분산을 최소화하기 위해 자산 간의 가중치를 최적화 하였다.

3.1.2 머신 러닝

본 연구에서 가장 핵심적인 요소가 머신 러닝 기반 예측 모델이다. 머신 러닝기반의 예측모델의 입력 값은 독립변수로서 재무, 기술적 지표, 소셜 빅데이터로부터 추출된 특징들로 구성된다. 예측모델로부터 출력되는 종속변수는 해당 종목에 대한 다음달 종가의 상승 또는 하락에 대한 예측치이다.

일정기간을 통하여 학습된 모델은 교차검증을 통하여 검증을 한다. 모델의 성능을 측정하고 검증하기 위하여 5 Fold Cross-Validation을 이용한다. 우선 학습기간을 5구간으로 나눈 후 한 구간씩 제외하고 학습한 것과 5구간 전체를 학습한 것을 비교, 분석함으로써 어느 한 구간 값이 전체의 예측 성능에 편향되지 않는 안정된 성능의 예측 모델을 얻고자 함이다.

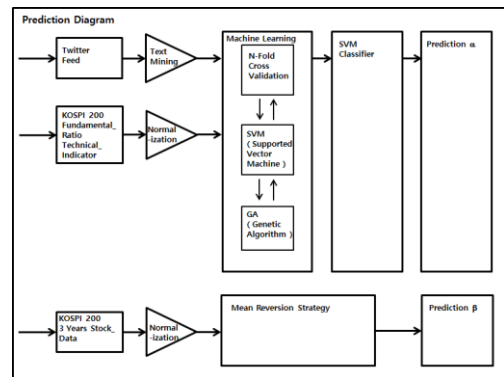


그림 4. 예측모형

머신 러닝과 유전자 알고리즘, 교차검증을 통하여 우리는 SVM Classifier를 얻게 된다. 이렇게 얻어진 SVM Classifier에 소셜 분석데이터와 개별종목의 기본 재무적 비율과 기술적 지표를 입력하면 우리는 가상의 종목에 대한 Prediction 값을 얻을 수 있다.

3.1.3 유전자 알고리즘

본 연구에서 주가의 예측을 위한 목적으로 사용하는 머신 러닝 알고리즘 SVM 모델은 속성의 숫자가 많을수록 연산량이 기하급수적으로 증가하기 때문에 여기서 발생하는 시간독립(Time-independent)적인 문제를 해결하기 위하여 유전자 알고리즘을 사용한다.

첫 단계에서는 통계적 방법을 활용하여 초기유전자를 세팅한다. 여기서 통계적 방법이란 계속 진화하겠지만 처음에는 거의 무작위로 특징들을 초기유전자로 세팅하는 것을 말한다. 두 번째 단계에서는 주가와 특징 간의 시계열적 패턴을 찾아내기 위하여 적합도 함수를 사용하여 속성을 추출한다. 세 번째 단계에서는 재생산, 교배, 돌연변이 과정을 거쳐서 다음 세대로 우수한 유전자를 물려 준다. 여기서 불필요한 특징은 삭제되고 어떤 특징들은 교배를 통하여 변경되며 초기에 세팅된 유전자 이외에 새로운 유전자를 삽입한다. 네 번째 단계에서 다시 적합도 평가가 이루어지고, 다섯 번째 단계에서 모든 특성이 평가되어 종료조건이 만족되면 해당 종목에 대한 최종적인 유전자집합을 선택하게 되고, 종료조건이 만족되지 않아 모든 속성에 대한 연산이 이루어지지 않으면 다시 두 번째 단계인 적합도 함수 단계로 돌아가게 된다. 이러한 기전을 통하여 각 종목별로 해당 종목의 예측에 가장 관련이 깊다고 여겨지는 속성집합을 찾아내어 SVM에 입력치로 넘겨 주게 된다.

3.2 자산 분배 최적화

자산 분배 최적화를 위해서 본 연구에서는 블랙 리터만 모형을 사용한다. 이 모형은 베이지안 접근법(Bayesian Approach)을 이용하여 개별 주식의 기대 수익률에 주관적 전망을 결합 후 내재 균형 수익률을 도출한다. 예를 들어, 머신 러닝과 평균 회귀 전략에 의해 특정 주식에 대한 전망이 좋을 것으로 예측된 해당 주식의 내재 균형 수익률이 상승되고 다음 달에 투자되어야 될 비율이 증가하게 된다. 다음 <그림 5>의 최적 전망 행렬은 SVM과 평균 회귀 전략을 이용하여 구성한다.

블랙 리터만 모형에서는 SVM 예측치와 평균 회귀 전략의 예측과, 두 가지를 사용하여 종목별 투자가중치를 결정한다. 이 모형의 경우 전망행렬을 구성하는 것이 매우 중요하다. 또한, 종목별 공분산이 투자 가중치를 계산하는 데 있어 큰 영향을 끼친다. 이는 SVM의 결과보다 우수하고 평균 회귀 전략의 결과보다 저조하지만 주가 상승 구간에서 수익이 극대화되는 현상을 관찰할 수 있었으며, 동시에 주가가 계속해서 하락할 경우 변동성이 다소 커지는 경향이 있음을 알 수 있었다.

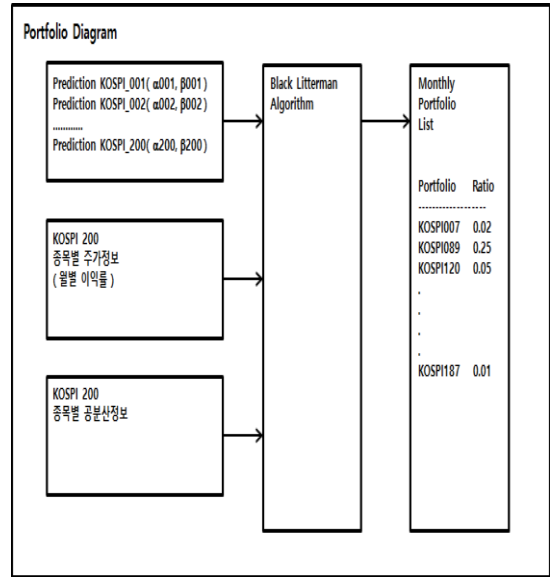


그림 5. Portfolio Diagram

3.3 자산거래 시스템

트레이딩 시스템은 실시간 트레이딩을 위하여 국내 증권사인 E-Best가 제공하는 시스템 트레이딩 등을 위한 API를 사용하였다. 전체적인 시스템은 예측치와 마찬가지로 하둡 기반 멀티 프로세싱을 사용하였다. 통계분석을 위해서는 대표적인 통계 오픈소스인 R을 사용하였다. R Programming Language는 주로 통계 계산 그 결과를 적절하게 보여 주는 그래픽을 위한 통계전용 통합소프트웨어라고 할 수 있다. 뉴질랜드 오클랜드 대학교수인 로버트 젠틀맨과 로스 이하카가 개발을 시작하였고 현재는 R Core Team에 의해 개발이 주도되고 있다. R은 통계 및 자료분석에 폭 넓게 사용되며 최근 빅데이터 분석을 위한 개발 언어로서 큰 위상을 차지하게 되었다.

시스템 트레이딩의 프로세스는 먼저 증권시장 개장 전에 API를 호출하고 로그인을 하게 된다. 다음으로, 현재 계좌정보를 조회하고 블랙 리터만에 의하여 최적화된 포트폴리오를 호출하게 된다. 현재 보유주식과 최적화된 포트폴리오를 비교하고 최적화된 포트폴리오를 실현하기 위하여 트레이딩을 실시한다. 트레이딩 시스템은 이를 위해, 실시간으로 주가를 전송 받고, 투자시점과 투자금액을 분석한다. 분석이 완료되면 최적화된 포트폴리오를 구현하기 위하여 리밸런싱을 하게 되는데 이때는 주식을 한 번에 매수하지 않고 나누어서 분할매수를 하게 된다. 매도시에도 수익을 극대화하기 위하여 균형가 매도 전략을 취한다.

4. 실험 결과

4.1 실험 세팅

실험을 위하여 갖추어야 할 시스템은 크게 데이터를 수집하는 시스템과 수집된 데이터를 저장하는 구조가 필요하다. 두 번째로는 수집된 데이터를 필터링하고 정규화하는 과정이 요구된다. 주요 입력 변수로 사용되는 소셜 데이터의 경우에는 자연언어처리과정을 통하여 본 연구에서 요구되는 입력변수들을 추출하고 저장한다. 주요 입력변수 중 추가정보와 종목별 기초비율(Fundamental Ratio)과 기술지표(Technical Indicator) 등은 모두 주식사이트에서 실시간으로 수집되어 정규화 과정을 거쳐 저장된다. 소셜 빅데이터는 스팸 필터링 과정을 통과시켜 불필요한 데이터가 제거되고, 텍스트 마이닝 과정을 통하여 우리는 본 실험에서 요구하는 주식전망 관련 유의미한 데이터들만을 추출한다. 추출된 데이터들은 다시 감성분석과 상관관계분석 과정을 거쳐 머신 러닝을 위한 최종적인 특징들로 선별된다.

투자전략시스템은 크게 포트폴리오최적화(Portfolio Optimization) 시스템과 실시간 거래시스템으로 나누어진다. 예측은 한 달에 한 번 이루어지며 매달 그 예측에 맞게 포트폴리오도 재구성되고 재구성된 포트폴리오를 달성하기 위하여 트레이딩 시스템은 지속적으로 매도/매수 작업을 수행한다.

4.2 실험결과

실험기간은 2014년 1월부터 2014년 12월까지로 제한하였다. 시스템의 성능을 측정하기 위한 벤치마크는 대한민국 주가를 대표하는 KOSPI200으로 하였으며, 종목 수는 50종목 이하로 제한하였고, 포트폴리오 내에서 개별주식은 전체의 8% 이하로 한정하였다. 이것은 포트폴리오의 비중을 늘리고, 더 많은 종목을 투자하기 위함이다. 더불어, 하락장에 대비하여 역 ETF를 추가하였는데 총 포트폴리오의 14.5% 이하로 구성하는 원칙을 고수하였다. 이것은 투자의 안정적인 측면을 늘리고, 시장 변동성을 최소한으로 줄이기 위해 설정한 것이다. 본 연구에서 설정한 모델은 편의상 DS Fund라 칭한다. <그림 6>은 동 기간에 운용된 실적을 보여 준다.

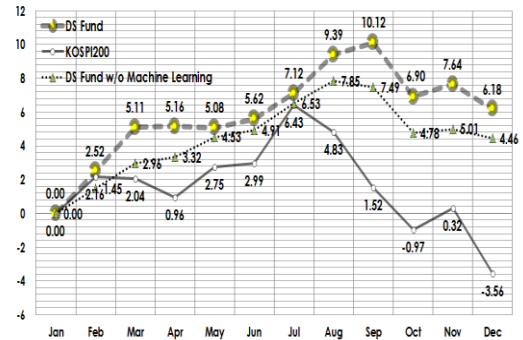


그림 6. 2014년 연구모델 시뮬레이션 결과

4.2.1 실험결과 통계분석

주식을 위주로 하는 고위험 자산의 성과평가모형으로 CAPM을 통하여 산출한 알파(alpha)를 벤치마크인 시장 평균 수익률을 초과하는 기준으로 채택한 Jensen's alpha(1968) 이후에도 여러 방법론이 발표되었다. 비교적 최근의 연구는 시장에 미치는 성과평가모형을 제시하였는데 이들은 알파와 베타를 상정한 아래의 수식을 통하여 논리적으로 일관된 성과 측정을 한다고 주장하고 있고 이는 펀드의 성과의 지속성을 측정하는 주요한 도구로 이용되고 있다. 알파는 시장 대비 초과 수익률을, 그리고 베타는 시장 대비 변동성에 대한 민감도를 측정하는 주요 지표이다. 베타가 1보다 작으면 시장의 변동성에 비해 보다 덜 민감하게 반응한다는 것을 의미한다. 알파가 양의 값을 가지고 베타가 1보다 작다면, 우리가 예측한 결과에 따른 포트폴리오가 시장에 비하여 우수한 성과를 내는 동시에 시장의 변동성에도 적게 흔들리는 수익성과 안정성을 동시에 얻을 수 있는 포트폴리오라고 할 수 있다. 우리는 최적의 연구모델을 얻기 위하여 총 2개의 모델을 설정하고 시뮬레이션하였다.

표 1. 시뮬레이션 결과에 대한 유의성 검증

패널 1 : Model_1(Mean Reversion & Machine Learning)			
	계수	t 통계량	P-값
α	0.788389695	2.180009678	0.043603554
β	0.460545928	3.358310143	0.00372939
패널 2 : Model_2(Mean Reversion)			
	계수	t 통계량	P-값
α	0.484183872	1.943628145	0.068684262
β	0.39654702	4.197859736	0.000604409

4.3 실제운용 결과

4.3.1 트레이딩 시스템 운용환경

트레이딩 시스템의 운용기간은 2015년 1월부터 2016년 2월로 하였다. 머신 러닝 기반 소셜 빅데이터를 이용한 트레이딩 시스템의 유효성을 검증하기 위한 벤치마크로 가장 대표적인 지수인 KOSPI200을 선정하였다. 운용하는 포트폴리오의 구성 종목 수는 최대 50개 종목을 넘지 않게 제한하였다. 2015년 실제운용에서는 현금을 거래하는 만큼 2014년 실험 대비 투자환경을 보다 보수적으로 설정하였다. 안정성을 더하기 위하여 1개 종목당 최대 투자비중은 5%를 넘지 않게 배분하였고, 하락장에 보다 적극적으로 대응하기 위하여 역 ETF를 기존 14.5% 이하에서 20% 이하로 상향하여 구성하였다. 투자금액은 4,000만 원으로 시작하였다. <표 2>는 운용된 실적을 보여 준다.

표 2. 트레이딩 시스템 투자 환경(Universe)

기준		KOSPI 200
유형		자기자본 파생상품
평균 일일 거래량에 의한 주식 선택 기준		일일 기준 5만주 이상 거래되는 주식을 만을 선택했음.
자기자본 비율	주식	총 보유자본의 80%이하
	역 ETF	총 보유자본의 20%이하
연간 회전율		총 보유자본의 300%이하

4.3.2 트레이딩 시스템 운용결과

2016년 2월 현재 트레이딩 시스템의 수익률은 0.22%이다. 반면, KOSPI200의 수익률은 -6.30%이다. 따라서 벤치마크 대비 초과 수익률은 6.52%가 된다. 운용결과는 <그림 7>과 같다.

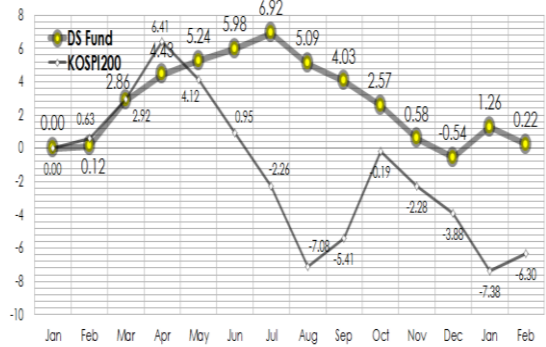


그림 7. 트레이딩 시스템 운용결과1

4.3.3 Mann-Whitney U test

독립표집에 의한 두 평균치의 차에 관한 t검증의 대안으로 사용될 수 있는 비모수적 통계방법으로 이 방법은 비모수적 방법 중 가장 유용한 방법의 하나로서 일반적 조건하에서 t검증만큼의 통계적 검증력이 있다. 두 독립표집의 사례수가 각각 n_1 , n_2 하고 하고 두 집단의 사례수의 합, 즉 n_1+n_2 를 N 이라고 한다[11].

본 연구에서는 트레이딩 시스템을 통한 이익률을 Mann-Whitney U test를 활용하여 검증하고자 하였으며, 트레이딩 시스템에서 기본 모형에 SNS와 인공지능을 기반으로 한 빅데이터를 적용했을 때 이익률과 머신 러닝만을 적용 했을 때 이익률 간의 차이를 측정하고자 하였다.

그 결과는 다음과 같이 그래프와 표를 통하여 확인할 수 있다.

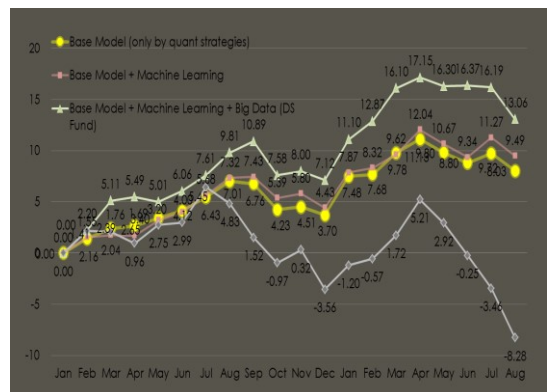


그림 8. 트레이딩 시스템 운용결과2

<그림 8>과 같이 SNS와 인공지능을 기반으로 한 빅데이터와 머신 러닝 간의 이익률 차이를 보였으며,

Mann-Whitney U test 분석을 통한 결과는 <표 4>와 같이 유의 수준 0.05 이하 기준으로 유의 확률 또한 이상이 없으므로 나타났다.

표 3. Mann-Whitney U test 순위

그룹구분	N	평균순위	순위 합
Big Data	20	24.33	486.50
Machine Learning	20	16.68	333.50

표 4. Mann-Whitney U test 검정 통계량

구분	이익률
Mann-Whitney U	123.500
Wilcoxon W	333.500
Z	-2.069
근사 유의확률	0.039
정확한 유의확률	0.038

5. 결론

최근 주식시장은 유래 없는 불확실성을 보이고 있다. 실제 주식시장은 그 속성상 가치투자가 기본이지만, 지난 8월 24일 미연준의 금리인상발표와 한반도에 발생한 북한 관련 불확실성이 보여 주듯이 기업의 활동과는 무관한 투자심리가 주가에 막대한 영향을 준다는 사실을 확인하는 계기가 되었다. 본 연구에서는 사전실험을 통하여 본 연구에서 제안하는 머신 러닝 기반의 소셜 빅데이터를 이용한 금융자산 트레이딩 모델을 검증하였다. 검증은 2014년 1월부터 2014년 12월까지 1년으로 제한하였고 학습기간은 3년으로 제한하였다. 실험은 통계 기반 예측 접근방법과 소셜 빅데이터 분석을 이용한 머신 러닝 기반 예측 접근법으로 나누어 이루어졌다. 실험결과 CAPM에 기반한 통계분석결과 시장초과수익률인 α 값과 시장변동성에 대한 민감도 지표인 β 값 모두 우수하거나 유의한 결과를 보인 소셜 빅데이터 분석을 이용한 머신 러닝 기반 접근방법이 우수함을 확인하였다. 실험결과를 바탕으로 2015년 1월부터 2016년 2월까지 실전투자를 진행하였다. 실전투자를 진행한 목적은 시뮬레이션을 통한 결과가 다양한 동인에 의해 변화하는 시장상황에서도 여전히 유의한지 평가하기 위해서 진행하였다. 실전투

자 결과 현재 벤치마크인 KOSPI200 대비 6.52%를 초과하는 성과를 보였다. 현재 사람의 경험에 의존하는 동일유형의 펀드들의 수익률 평균이 -5%에서 -6%에 이르는 상황이다. 우리는 본 연구를 통하여 빠르게 변화하는 집단의 심리와 주가를 결정하는 기업의 기본 지표와 기술적 지표를 바탕으로 머신 러닝 기반의 소셜 빅데이터 분석을 이용한 금융자산 트레이딩 모델이 기존의 통계 기반 예측 접근법을 통한 트레이딩 시스템이나 사람이 직접 운용하는 상업적 펀드들에 비하여 여러 약재에도 불구하고 유의한 실적을 보인 것으로 평가할 수 있음을 실증분석으로 검증하였다. 따라서 머신 러닝 기반의 소셜 빅데이터 분석을 이용한 금융자산의 트레이딩 모델이 기존의 금융자산 트레이딩 모형의 성능향상에 도움을 줄 수 있다고 할 수 있다.

결론적으로, 본 연구를 통하여 인공지능 기반의 빅데이터 분석을 이용한 로보 트레이딩이 실제 시장지표 대비 높은 수익률을 보일 수 있는 가능성을 확인할 수 있었다.

본 연구는 다음과 같은 몇 가지 한계점을 가지고 있으므로, 향후 연구에서는 이를 보완하여 진행해야 할 것이다. 본 연구의 제약사항은 본 연구의 주요학습기반인 소셜 빅데이터가 국내에 본격적으로 도입된 시기가 2011년으로 비교적 짧은 기간의 학습데이터만을 가지고 있다는 점이다. 예를 들어 2008년 미국 발 서브프라임모기지 사태나 연이은 2009년 이후의 유로존 금융위기 등에 대한 학습기반이 없다는 것이다. 그러나 소셜 빅데이터는 양태를 달리할 뿐 계속 확산되고 있는 추세이기 때문에 지속적인 연구를 통하여 사람들의 감성이 실물경제에 주는 영향에 대한 연구를 지속적으로 해 나갈 수 있다고 생각한다.

또한 본 연구가 대한민국 주식시장에 한정되어 이루어졌다는 점이다. 2015년 우리 주식시장의 경우 삼성전자의 환차익에 대한 Earnings Surprise가 주식시장을 견인하여 겉으로 보면 인덱스가 크게 변동성이 없어 보였지만 대부분의 주식은 폭락하였다. 이러한 시장환경에 대한 비정상적 대장주의 왜곡현상은 우리나라 주식시장의 구조적 편향에서 기인한다. 그래서 보다 합리적인 머신 러닝 기반의 소셜 빅데이터 분석으로 이용한 금융자산 트레이딩 모델에 관한 연구는 시장편향이 작은 미국 S&P500 종목을 대상으로 한 보다 심도 있는 연구가 요구된다고 생각한다.

마지막으로, 본 연구에서 이용된 정형데이터는 주로 주식시장에서 얻을 수 있는 기업의 Fundamental Ratio와 Technical Indicators만을 활용하였으나, 앞으로

는 세계경제에 영향을 주는 금융 인덱스에서 실물에 대한 인덱스까지 다양한 인덱스 데이터를 학습데이터로 활용하는 연구를 지속적으로 전개한다면 종목 개개의 머신 러닝을 통한 전망을 위한, 보다 의미 있는 속성들을 얻을 수 있을 것으로 생각한다. 이로써 머신 러닝을 통한 더 나은 전망치를 얻을 수 있을 것으로 기대한다.

참고문헌

- [1] 김상환, 2014. “공분산행렬 추정방법과 공매도 제한 제약의 자산배분성과에 대한 영향분석”, *한국증권학회지*, 43집, 4호, pp.705-729.
- [2] 김선웅, 안현철, 2010. “Support Vector Machines와 유전자 알고리즘을 이용한 지능형 트레이딩 시스템의 개발”, *지능정보연구*, 16권, 1호, pp.71-92.
- [3] 김유신, 김남규, 정승렬, 2012. “뉴스와 주가: 빅데이터 감성분석을 통한 지능형 투자의사 결정모형”, *지능정보연구*, 18권, 2호, pp.143-156.
- [4] Balvers, R., Wu, Y. and Gillilan, E., 2000. “Mean Reversion across National Stock Markets and Parametric Contrarian Investment Strategies”, *Journal of Finance*, Vol.55, No.2, pp.745-772.
- [5] Bollen, J., Mao, H. and Zeng, X., 2011. “Twitter Mood Predicts the Stock Market”, *Journal of Computer Science*, Vol.2, No.1, pp.1-8.
- [6] Bollen, J., Mao, H. and Pepe, A., 2010. “Determining the public mood state by analysis of microblogging posts”, *Proceedings of the ALife XII*, MIT Press.
- [7] Cao, L. J., 2003. “Support vector machine experts for time series forecasting”, *Neurocomputing*, Vol.51, pp.321-339.
- [8] Cao, L.J., Chua, K.S. and Guan, L.K., 2003. “Combining KPCA with Support Vector machine for Time Series Forecasting”, *Proceedings of IEEE international conference on computational intelligence for financial engineering*, pp.325-329.
- [9] Chen, R. and Lazra, M., 2011. “Sentiment Analysis of Twitter Feeds for the Prediction of Stock Market Movement”, Stanford University, Working Paper.
- [10] Evry, 2016. *The New Wave of Artificial Intelligence*.
- [11] Fabre, J. and Snape, M., 2007. “Liquidity Surrounding Sell-Side Equity Analyst Recommendation Revisions on the Australian Securities Exchange”, University of Sydney, Working Paper.
- [12] Gibbons, J.D. and Chakraborti, S., 1991. “Comparisons of the Mann-Whitney, Student’s, and Alternate t Tests for Means of Normal Distributions”, *Journal of Experimental Education*, Vol.59, No.3, pp.258-267.
- [13] Li, H., Zhang, X. and Zhao, R., 2011. “Investing in Talents: Manager Characteristics and Hedge Fund Performances”, *Journal of Financial and Quantitative Analysis*, Vol.46, No.1, pp.59-82.
- [14] Sigman, K., 2005. *Capital Asset Pricing Model*, Columbia University.
- [15] Vapnik, V., 1995. *The Nature of Statistical Learning Theory*, Springer Verlag.

저자소개



송성환
(Song, Sunghwan)

한양대학교에서 전자계산학 석사, 숭실대학교 IT 정책경영학과 박사학위를 취득하였다. 한국무역정보통신(KTNET)에서 한국무역자동화 및 미국통관자동화, 한국조달EDI를 담당하였고, 다음소프트에서 사업개발, 재무, 마케팅 임원을 역임하였으며, 현재 대표이사로 재직중이다. 주요 관심분야는 IT 정책 및 전략, 자연언어처리, 빅데이터, 핀테크, 인공지능, 창업지원 등이다.

E-mail: shsong@daumsoft.com

Tel: +82-2-565-0531



한경석
(Han, Kyeongseok)

서울대학교에서 경영학 석사학위를 취득하였고, 미국 Purdue대학교에서 경영학 박사학위를 취득하였으며, 현재 숭실대학교 경영학부 교수로 재직중이다. 주요 관심분야는 디지털정책, 경영정보시스템(Technical MIS), Digital Economy, Agent-Eased Simulation, Web Programming, ERP, 회계정보시스템, e-Business, 전자상거래, 중소기업정보화, 기업컨설팅, 기업자금지원 정책연구 등이다.

E-mail: kshan@ssu.ac.kr

Tel: +82-2-820-0585



최수정
(Choi, Sujung)

미국 스탠포드대학교에서 통계학 석사, 미국 캘리포니아대학교 재무분야 경영학과 박사학위를 취득하였다. 현재 숭실대학교 경영학부 조교수로 재직중이며, 주요 관심분야는 행태재무, 투자론 등이다.

E-mail: sjchoi@ssu.ac.kr

Tel: +82-2-820-0579



박성민
(Park, Sungmin)

숭실대학교 경영학 석사과정을 밟고 있다. 주요 관심분야는 디지털정책, 정보프라이버시, 빅데이터 등이다.

E-mail: sungmin30@hanmail.net

Tel: +82-2-820-0585

◇ 이 논문은 2016년 11월 21일에 접수하여 2017년 1월 24일에 1차 수정, 2017년 2월 25일에 2차 수정을 거쳐 2017년 4월 13일에 게재확정되었습니다.