



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0081699  
(43) 공개일자 2021년07월02일

(51) 국제특허분류(Int. Cl.)  
G06Q 40/04 (2012.01) G06F 40/30 (2020.01)  
(52) CPC특허분류  
G06Q 40/04 (2013.01)  
G06F 40/30 (2020.01)  
(21) 출원번호 10-2019-0173917  
(22) 출원일자 2019년12월24일  
심사청구일자 없음

(71) 출원인  
서강대학교산학협력단  
서울특별시 마포구 백범로 35 (신수동, 서강대학교)  
(72) 발명자  
박수진  
서울특별시 양천구 목동동로 430 목동신시가지아파트6단지 604동 402호  
김재훈  
서울특별시 마포구 토정로18길 11 래미안웰스트림 107동 1501호  
(뒷면에 계속)  
(74) 대리인  
이지연

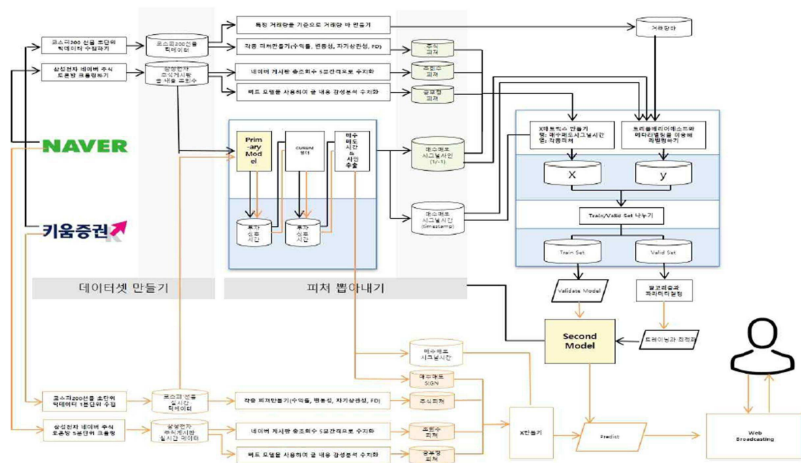
전체 청구항 수 : 총 1 항

(54) 발명의 명칭 자연어 피처를 결합한 주식 매도 및 매수 시그널 판별기

(57) 요약

본 발명은 주식 매도 및 매수 시그널 판별기에 관한 것이다. 상기 주식 매도 및 매수 시그널 판별기는, 자연어 데이터를 반영한 모델을 구축하고, 메타 라벨링을 비롯한 다양한 방법론을 사용하여 시장상황에 강건한 주식 매도 및 매수 시그널을 판별할 수 있도록 하는 것을 특징으로 한다.

대표도 - 도1



(72) 발명자

**박재윤**

울산광역시 남구 삼산로 87 삼환아르누보 1005호

**황수현**

서울특별시 영등포구 당산로 180 신우빌딩

**안성모**

대구광역시 달성군 유가읍 테크노대로5길 80 호반

베르디움THECLASS 2차 208동 401호

이 발명을 지원한 국가연구개발사업

과제고유번호	1711093716
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기술진흥센터
연구사업명	SW전문인력역량강화(R&D)
연구과제명	SW중심대학(서강대학교)
기 여 율	1/1
과제수행기관명	서강대학교 산학협력단
연구기간	2019.01.01 ~ 2019.12.31

---

**명세서**

**청구범위**

**청구항 1**

자연어 데이터를 반영한 모델을 구축하고,

메타 라벨링을 비롯한 다양한 방법론을 사용하여 시장상황에 강건한 주식 매도 및 매수 시그널을 판별할 수 있도록 하는 것을 특징으로 하는 자연어 피처를 결합한 주식 매도 및 매수 시그널 판별기.

**발명의 설명**

**기술분야**

[0001] 본 발명은 주가 예측 방법에 관한 것으로서, 더욱 구체적으로는 머신 러닝 방법을 이용하여 자연어 피처를 결합한 주식 매도 및 매수 시그널 판별기에 관한 것이다.

**배경기술**

[0002] 기존 연구에서는 예측 성능에 강점을 보이는 머신러닝 방법론들을 활용해 모델을 구축함으로써 주가를 예측하고자 하였다. XG부스트 모형을 활용한 코스피 200 주가지수 등락 예측에 관한 연구에서는 분류성능과 속도에 강점을 보이는 XG부스트 모형을 적용해 주가를 예측했다. LSTM 신경망 모형에서는 80%의 정확도를 보인 반면 XG부스트 모형은 86.67%라는 높은 정확도를 기록하여 그 성능을 증명하였다. 한편 주식시세 예측을 위한 딥러닝 최적화 방법 연구와 인공지능을 이용한 주가 방향성 예측 및 투자 포트폴리오 최적화에서는 대표적인 딥러닝 방식인 RNN-LSTM, CNN 방식을 이용해 주가 예측 모델을 수립하였다. 주식시세 예측을 위한 딥러닝 최적화 방법 연구에서는 종가, 거래시간 feature를 중심으로 하이퍼 파라미터 튜닝을 실시하고 이를 통해 주식시세 예측 모형을 수립하였다. 그 결과 기존 학습 정확도인 89.84%보다 약 9.6% 향상된 99.52%의 정확도를 기록했다. 인공지능을 이용한 주가 방향성 예측 및 투자 포트폴리오 최적화는 종목 별 주가 정보를 보조지표로 가공한 데이터, 경제 기사의 제목 데이터를 대상으로 선정된 종목의 투자 비율을 최적화하는 모형을 만들었다. 유전 알고리즘을 활용해 정밀도, 재현율, 정확도, 당일 예측 확률을 비교하고 이를 통해 선정한 종목의 수익성을 벤치마크 주가지표인 KOSPI, KOSPI 200와 비교하였다. 그 결과 상위 15개 종목의 정확도가 61.52%~68.20%를 기록했다.

[0003] 이러한 기존 연구들은 과적합 문제가 발생한다는 결정적인 한계를 지닌다. 금융 데이터는 기존의 머신러닝에서 사용하는 그림, 영상 데이터와는 다르게 과거의 사건이 일정기간 동안 지속적으로 영향을 주는 시계열 상관성이 존재한다는 특징이 있다. 따라서 금융 데이터를 대상으로 다른 분야에서 좋은 성능을 내는 머신러닝 알고리즘을 그대로 적용할 경우 모델의 과적합 현상이 발생하게 되는 것이다. 즉 특정 데이터에 대한 예측력은 매우 높을 수 있으나 시장의 상황이 조금이라도 변화하면, 매우 낮은 정확도를 보이게 된다.

**선행기술문헌**

**특허문헌**

[0004] (특허문헌 0001) 한국공개특허번호 제 10-2001-0008679호  
 (특허문헌 0002) 한국공개특허공보 제 10-2001-0089791호

**발명의 내용**

**해결하려는 과제**

[0005] 전술한 문제점을 해결하기 위한 본 발명의 목적은, 다양한 방법론을 활용해 과적합 문제를 극복하고, 시퀀셜 부트스트래핑을 사용한 배깅을 이용하여, 실제 시장의 주가 변동에도 대응할 수 있는 주식 매도 및 매수 시그널

판별기를 제공하는 것이다.

**과제의 해결 수단**

[0006] 전술한 기술적 과제를 달성하기 위한 본 발명의 특징에 따른 주식 매도 및 매수 시그널 판별기는, 시퀀셜 부트스트래핑을 사용한 배깅을 이용하여, 실제 시장의 주가 변동에도 대응할 수 있는 모델을 제공한다.

**발명의 효과**

[0007] 본 발명의 의의는 금융 시장 상황 예측에 자연어 데이터를 반영한 모델 구축을 시도했다는 것과 메타 라벨링을 비롯한 다양한 방법론을 사용하여 시장상황에 강건한 모델을 만든 데에 있다. 전체적으로 예상에 부합하는 결과가 나왔으며 74%의 투자 성공 분류 정확도는 고무적이라고 할 수 있다. 또한 실제 데이터의 결과값을 통해 Primary Model은 현재 주식 시장이 어떤 상황인가에 따라 변동되어야 함을 알 수 있었고, 따라서 투자 전략은 여러 개를 갖고 시장에 따라 다른 전략과 피처를 사용해야 하는 것 또한 알 수 있었다

**도면의 간단한 설명**

[0008] 도 1은 본 발명에 따른 자연어 피처를 결합한 주식 매도 및 매수 시그널 판별기를 도시한 시스템 구성도이며, 도 2는 각 구성 요소들을 개념적으로 도시한 모식도이다.

도 3은 Primary Model의 매수매도구간과 투자 시그널을 예시적으로 도시한 그래프이다.

도 4 및 도 5는 모델1과 모델 2를 통해 테스트 데이터 셋인 2019년 7월부터 9월 사이의 데이터에 적용해 본 결과에 대한 sklearn 모듈의 Classification Report와 ROC곡선 그리고 각각의 모델의 피 처 중요도를 나타내는 세 가지 지표인 MDI(Mean Decrease Impurity), MDA(Mean Decrease Accuracy), SFI(Single Feature Importance)를 나타낸 그림이다. 즉, 도 4는 배깅, 메타라벨링, Purged 4fold 적용 모델의 결과를 도시한 도표 및 그래프이며, 도 5는 시퀀셜 부트스트래핑 배깅, 메타라벨링, Purged 4fold 적용 모델의 결과를 도시한 도표 및 그래프이다.

도 6은 본 발명에 따른 판별기를 테스트한 결과를 도시한 그래프이다.

**발명을 실시하기 위한 구체적인 내용**

[0009] 우리는 과적합 문제를 비롯하여, 기존 금융시장에 머신러닝을 적용한 연구에서 제시된 문제점을 해결하고, 자연어에서 새로운 피처를 찾아내는 금융 머신러닝 모델을 만들고자 하였다. 본 ‘자연어 피처를 결합한 매수매도 시그널 판별기’는 1) 한국 선물(코스피200)시장의 모든 거래데이터(Tick Data)를 증권사 api를 통해 수집하고 매일 저장하여 데이터 베이스화 시켜 학습 데이터 로 사용하였고 2) 이렇게 생성된 고빈도 데이터(High Frequency data)를 이용하여 일/분/초 의 시 간단위가 아닌 거래량 단위의 거래량봉(Volume Bar)를 재구성하여 머신러닝 학습 데이터에 적합한 통계적 특성을 갖출 수 있게 하였다. 3)이 데이터를 바탕으로 이동평균선 투자 전략(Trend Following Strategy)을 제1투자전략(Primary Model)으로 CUSUM필터를 사용하여 도출된 시그널에 대해 메타라벨링(Meta-labeling)을 적용하여 라벨링하였다. 4)머신러닝에 사용되는 피처로 기술적 지표나 재무제표에 포함된 항목뿐만 아니라, 자연어 피처(네이버 주식계시판 글의 긍부정 정도, 실 시간 조회수)를 버트(BERT)1를 이용한 감성분석과, 실시간 게시판 크롤링을 통해 구현하여, 예측 성능을 향상시키고자 하였다. 5) 머신러닝 알고리즘으로는 금융데이터를 적용하고 과적합을 방지하는데 적합하다고 알려진 Sequential bootstrapping 을 사용한 배깅 모델을 사용하여 시장 환경 변화에 대한 강건성을 유지하는 동시에 성능을 향상시키고자 하였다. 5) 위 과정을 통해 학습한 모델(Second Model)을 바탕으로 6) 삼성전자의 주식거래데이터를 1분 간격, 삼성전자 네이버 주식 게시판 글을 5분 간격으로 실시간 크롤링하여, 매수 매도 시그널의 발생여부를 살핀 후 시그널이 발생한다면 각종 피처를 학습한 모델에 투입하여 그 결과를 웹을 통해 Broadcasting할 것이다.

[0010] 도 1은 본 발명에 따른 자연어 피처를 결합한 주식 매도 및 매수 시그널 판별기를 도시한 시스템 구성도이며, 도 2는 각 구성 요소들을 개념적으로 도시한 모식도이다.

[0011] 본 발명은 네이버 주식 게시판에서 추출한 여러 지표와 과적합을 줄이는 데 장점이 있는 배깅모델을 사용하여, 변화하는 시장에 대응하도록 하였다. 머신 러닝에 필요한 피처를 확보하기 위해 네이버 주식토론방의 게시글의 날짜, 제목, 내용, 작성자, 조회수, 공감수, 비공감수를 5분 간격으로 크롤링을 자동화시켰으며, 이 자료들 또한 일별로 csv파일을 저장하여 각종 자료와 수치들을 데이터베이스화 하였다. 또한 버트 모델을 파인 튜닝하는 학습 데이터 셋을 만들기 위해, 코스피 시가총액 상위20개 종목, 코스닥 시가총액 상위20개 종목에서 게시글과

내용을 크롤링하였다. 4명의 인원이 각각의 게시판 글을 보고, 긍정(1) 부정(0)으로 라벨링 하는 작업을 진행하여 ‘주식 게시판의 긍정과 부정에 관한 자연어 학습 데이터 셋’ (2019)를 완성하였다.

- [0012] 한국의 대표적인 주식인 삼성전자 네이버 주식게시판에서, 작성 시간을 기준으로 바 1단위의 시간마다 생성된 게시글의 개수를 합하여 피쳐로 사용하였고, 조회수는 5분마다 시행된 크롤링 데이터를 전처리하여, 매 5분 사이 증가한 조회수를 피쳐로 사용하였다. 게시물의 긍부정에 관한 피쳐는, ‘주식 게시판의 긍정과 부정에 관한 자연어 학습 데이터 셋’ (2019)을 이용하여 버트 모델을 파인튜닝 시켜, ‘주가에 관한 긍/부정 정도 판별기’ (2019)를 만든 후, 이 판별기를 통해 삼성전자 네이버 주식 게시판에 존재하는 글의 긍부정 정도를 산출해내어, 바 1단위 시간마다 생성된 게시물의 긍부정도를 평균해 이를 피쳐로 사용하였다.
- [0013] 30 거래량 바의 이동평균선이 60 거래량 바의 이동평균선 위에 있는 경우에는 매수, 반대의 경우엔 매도하는 전략을 통해 각각의 거래량 봉마다 매수 혹은 매도를 판별하였고, 이후 CUSUM 필터를 사용하여 특정한 변동성을 포함한 투자 시점(Timestamp)만을 산출해 이 시점과, 그 시점의 투자 방향(매수1 / 매도-1)을 산출하고 이 시그널에 대해 트리플 베리어 메소드를 통한 메타 라벨링 방법론을 적용하였다.
- [0014] 이하, 본 발명에 따른 시스템에 사용될 학습 데이터 셋 구축에 대하여 설명한다.
- [0015] 먼저, 금융 데이터 셋 구축시, 본 방법론을 적용하기 위해서는, 기본적으로 매 순간 이뤄지는 거래 데이터가 필요하다. 해당 거래 데이터(틱데이터)를 확보하기 위해 증권사에 요청해 보았으나 1달 이상의 자료를 보유하고 있는 증권사가 없었고, 그 자료조차 높은 가격에 구매해야 했다. 따라서 우리는 한 국 주식시장을 대표할 수 있는 선물시장(코스피200선물)의 모든 거래데이터(Tick Data)를 증권 사 api를 이용해 직접 수집하고 이를 데이터 베이스화 시켜 학습 데이터로 사용하기로 하였다. 키움증권의 api를 이용하였으며, 이를 자동화하여 수집된 데이터를 일별로 csv파일로 저장하여 틱데이터를 데이터베이스화 하였다. 본 연구에는 2018년 10월 1일부터 2019년 9월 30일 까지 1년 간 수집한 코스피200 선물거래 틱데이터가 사용되었다.
- [0016] 다음, 네이버 주식게시판 데이터셋 구축시, 한국의 네이버 주식 토론방은, 한국 주식 시장에만 특징적으로 존재하는 것으로 많은 시 장 참여자들이 이곳에서 주식과 관련된 정보를 얻고 의견을 교환한다. Does Big Data Matter?: Predicting Stock Returns using Online Stock Message Boards(김재훈, 2019)에서는 네이버 주식 토론방의 게시물 수를 예측 변수로 활용할 경우 주가 예측의 성능을 향상시킬 수 있음을 보였다. 즉 게시판에 존재하는 어떤 정보나 사람들의 반응이 주가에 대해 설명력이 있음을 의미한다.
- [0017] 머신 러닝에 필요한 피쳐를 확보하기 위해 네이버 주식토론방의 게시글의 날짜, 제목, 내 용, 작성자, 조회수, 공감수, 비공감수를 5분 간격으로 크롤링을 자동화 시켰으며, 이 자료들 또한 일별로 csv파일을 저장하여 각종 자료와 수치들을 데이터베이스화 하였다. 또한 버트 모 델을 파인 튜닝하는 학습 데이터 셋을 만들기 위해, 코스피 시가총액 상위20개 종목, 코스닥 시가총액 상위20개 종목에서 게시글과 내용을 크롤링하였다. 5명의 인원이 각각의 게시판 글을 보고, 긍정(1) 부정(0)으로 라벨링 하는 작업을 진행하여 ‘주식 게시판의 긍정과 부정에 관한 자연어 학습 데이터 셋’ (2019)를 완성하였다.
- [0018] 이하, 네이버 주식 게시판의 자연어와 관련된 피쳐 생성에 대하여 설명한다. 시장에 변화의 영향력과 이에 따른 사람들의 반응을 변수로 사용하기 위해, 네이버 주식 게시판에서 수치화가 가능한 세 가지 피쳐를 생성하였다. 게시물 개수와 조회수, 그리고 게시물의 내용의 긍부정 정도가 그것이다. 게시물의 개수와 조회수는 사람들의 실시간 관심 정도를 나타내는 지표로 볼 수 있고, 게시물 긍부정 정도는 사람들의 주식에 대한 심리적 반응에 관한 대리 지표로 삼았다.
- [0019] 한국의 대표적인 주식인 삼성전자 네이버 주식게시판에서, 작성 시간을 기준으로 바 1단 위의 시간마다 생성된 게시글의 개수를 합하여 피쳐로 사용하였고, 조회수는 5분마다 시행된 크롤링 데이터를 전처리하여, 매 5분 사이 증가한 조회수를 피쳐로 사용하였다. 게시물의 긍부정 에 관한 피쳐는, ‘주식 게시판의 긍정과 부정에 관한 자연어 학습 데이터 셋’ (2019)을 이용하여 버트 모델을 파인튜닝 시켜, ‘주가에 관한 긍부정 정도 판별기’ (2019)를 만든 후, 이 판별기를 통해 삼성전자 네이버 주식 게시판에 존재하는 글의 긍부정 정도를 산출해내어, 바 1단위 시간마다 생성된 게시물의 긍부정도를 평균해 이를 ‘sentiment’ 피쳐로 사용하였다.
- [0020] 다음, 자연어 피쳐를 결합한 금융 머신러닝 모델 구축하는 과정을 설명한다. 우리는 수집된 거래량 바를 사용하여, Primary 모델로, 이동평균선 투자전략 (Trend Following Strategy)을 사용하여 500거래대금봉의 이동평균선이 4000거래대금봉의 이동평균선 위 에 있는 경우에는 매수, 반대의 경우엔 매도하는 전략을 통해 각각의 거래대금봉마다 매수(검정표 시구간) 혹은 매도(비표시구간)를 판별하였고, 이후 CUSUM 필터를 사용하여 특정한 변동성을 포 함한 투자 시점(Timestamp)만을 산출해 이 시점과, 그 시점의 투자 방향(매수1 / 매도-1)을 산출하

고 이 시그널에 대해 트리플 베리어 메소드를 통한 메타 라벨링 방법론을 적용하였다.

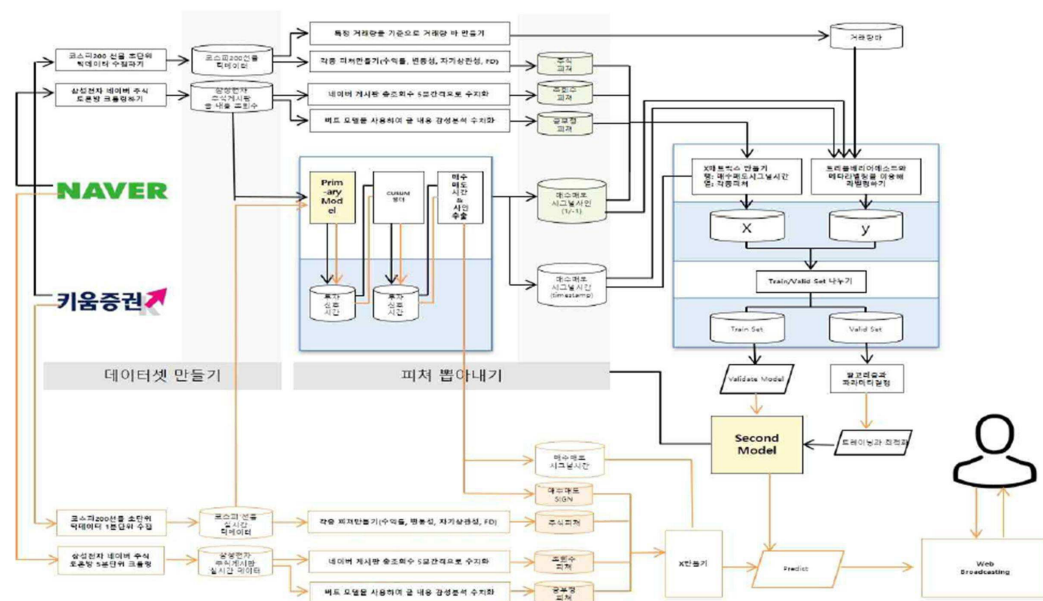
- [0021] 도 3은 Primary Model 의 매수매도구간과 투자 시그널을 예시적으로 도시한 그래프이다. 도 3을 참조하면, 각각의 투자 시그널에 대한 피쳐 매트릭스를 만들기 위해, 여러가지 기술적 지표들을 추가하였다. 주식과 관련한 기술적 지표인 수익률 자기상관도 변동성을 1바 5바 10바 30바 50바 단위로 산출하였고, 시장에 대한 추가 정보로 RSI와 FracDiff도 변수로 사용하였다. 자연어 데이터와 관련한 피쳐는 앞서 설명한 것과 같이, 해당 거래량바 기간 동안의 공부정도를 평균하여 'sentiment' 변수로 추가하였다.
- [0022] 이렇게 생성된 피쳐 매트릭스와 라벨링 데이터를 2018년 10월부터 2019년 6월까지의 학습데이터로 사용하고, 2019년 7월부터 2019년 9월까지를 테스트 데이터 셋으로 나눈 후, 학습데이터를 이용하여 첫 번째는 Purged K fold CV방식을 이용한 배경모델, 두 번째는 시퀀셜 부트스트래핑을 이용한 배경모델을 사용하여 학습하였다. 적절한 파라미터를 찾기 위해 max depth를 2,3,5,8 estimator의 개수를 10,20,50,100개로 설정하여 그리드 서치를 진행한 후, 여기서 산출된 최적학습 파라미터를 통해 모델1(일반배경, max\_depth:5 estimator 100개) 모델2(시퀀셜부트스트래핑배경, max\_depth:3, estimator:10개)모델을 결정하고 학습하였다.
- [0023] 이하, 본 발명에 따른 판별기의 결과를 설명한다. 도 4 및 도 5는 모델1과 모델 2를 통해 테스트 데이터 셋인 2019년 7월부터 9월 사이의 데이터에 적용해 본 결과에 대한 sklearn 모듈의 Classification Report와 ROC곡선 그리고 각각의 모델의 피쳐 중요도를 나타내는 세 가지 지표인 MDI(Mean Decrease Impurity), MDA(Mean Decrease Accuracy), SFI(Single Feature Importance)를 나타낸 그림이다. 즉, 도 4는 배경, 메타라벨링, Purged 4fold 적용 모델의 결과를 도시한 도표 및 그래프이며, 도 5는 시퀀셜 부트스트래핑 배경, 메타라벨링, Purged 4fold 적용 모델의 결과를 도시한 도표 및 그래프이다.
- [0024] 도 4 및 도 5를 통해, 첫째 Recall Rate를 통해 우리가 Primary Model로 삼았던 Trend Following Strategy가 테스트 기간에 적합한 전략인지를 가늠해볼 것이다. 둘째 두 분류기의 투자 성공에 대한 예측과 투자 실패에 대한 Accuracy를 비교하여, 해당 분류기가 리스크 관리에 적합한 모델인지, 수익 추구에 적합한 모형인지를 살펴볼 것이다. 마지막으로 피쳐 중요도에 대해 다 각도로 검토하여, 어떤 피쳐가 분류에 있어서 중요한 역할을 하였는지, 자연어 피쳐의 영향력은 어떠한지도 확인해볼 것이다.
- [0025] 재현율(Recall)의 경우 주식의 투자 방향과 시점을 알려주는 Primary Model의 지표로 활용할 수 있다. 본 발명에서 Primary Model로 Trend Following Strategy를 적용하였다. 모델1의 Recall Rate는 투자 성공에 대해서 0.54를 보이고, 모델2의 경우 투자 성공에 대해서 0.50을 보였다. 모델의 낮은 Recall Rate는 Primary Model이 투자의 방향과 시점 자체를 잘못 잡아냈다는 의미로 해석할 수 있다. Trend Following Strategy는 일반적으로 주식의 대세 상승기 혹은 대세 하락기에 적합하다고 알려진 투자 전략이다. 테스트 기간 동안인 2019년 7월에서 9월 사이는 한국 주식이 박스권에서 횡보하던 시기로, 이 시기에 모델이 낮은 Recall Rate를 보이는 것은 예상에 부합하는 결과이다. 즉 분류기의 성능을 높이기 위해서는 좋은 머신 러닝 알고리즘 외에도 Primary Model 즉, 투자 전략 자체가 매우 중요한 요소가 될 수 있음을 확인할 수 있었다.
- [0026] 정확도(precision)를 살펴보면, 모델 1은 투자 성공에 대해서 77%의 분류 정확도, 투자 실패에 대해서는 0.63에 분류 정확도를 보이고, 모델2는 투자 성공에 대해서 74% 분류 정확도 투자 실패에 대해서는 0.63에 분류 정확도를 보인다. 모델2, 즉 시퀀셜 부트스트래핑 방법론을 사용한 분류기가 통해 일반 배경 방법론을 사용한 분류기보다 과적합을 낮추도록 설계되었으므로 분류 정확도가 떨어지는 것은 연구 의도에 부합하는 결과라 할 수 있다. 또한 전반적으로 투자 실패의 경우보다 투자 성공의 경우에 accuracy가 높았고, 이는 분류기가 리스크 관리를 하는 분야보다, 수익을 추구하는 모델에서 좀 더 잘 활용될 수 있다.
- [0027] 도 6은 본 발명에 따른 판별기를 테스트한 결과를 도시한 그래프이다.
- [0028] 본 발명의 의의는 금융 시장 상황 예측에 자연어 데이터를 반영한 모델 구축을 시도했다는 것과 메타 라벨링을 비롯한 다양한 방법론을 사용하여 시장상황에 강건한 모델을 만든 데에 있다. 전체적으로 예상에 부합하는 결과가 나왔으며 74%의 투자 성공 분류 정확도는 고무적이라고 할 수 있다. 또한 실제 데이터의 결과값을 통해 Primary Model은 현재 주식 시장이 어떤 상황인가에 따라 변동되어야 함을 알 수 있었고, 따라서 투자 전략은 여러 개를 갖고 시장에 따라 다른 전략과 피쳐를 사용해야 하는 것 또한 알 수 있었다는 점은 새롭게 얻은 시사점이었다.
- [0029] 하지만 데이터의 양이 부족함에서 오는 아쉬움이 있었다. 데이터를 수집한 기간이 1년 정도로 짧으며, 코스피 200선물 데이터 한종목으로 그 양도 제한적이었다. 만약 코스콤과의 제휴를 통해 모든 종목의 과거 고빈도 데이터에 위 방법론을 적용할 수 있다면, 산업별, 시기별 다양한 시뮬레이션을 통해 다양한 투자 전략과 각 상황에

서 중요한 피쳐들을 산출해 내는데 도움이 될 것이다. 또한 이번 연구에서는 확인할 수 없었던, 그러나 중요한 자연어 데이터의 영향력을 향상시키기 위해서, 긍/부정도 이외에 다른 조합을 시도해 볼 필요가 있음을 깨달았다. 향후 위의 문제점들을 보완하고, 다양한 산업군 별로 거래데이터를 수집하고, 지속적인 시뮬레이션을 통해 시장 상황별 다양한 Primary Model을 구축하는 것이 중요할 것이다. 이 후 각 시장상황별로 중요한 Feature들을 찾아내어 피쳐 매트릭스를 구성한 뒤, 실시간 데이터에서 시장상황에 적절한 Feature를 산출한다면, 시장 상황에 영향을 받지 않는 매수 매도 판별기의 성능을 향상시킬 수 있을 것이다.

[0030] 이상에서 본 발명에 대하여 그 바람직한 실시예를 중심으로 설명하였으나, 이는 단지 예시일 뿐 본 발명을 한정하는 것이 아니며, 본 발명이 속하는 분야의 통상의 지식을 가진 자라면 본 발명의 본질적인 특성을 벗어나지 않는 범위에서 이상에 예시되지 않은 여러 가지의 변형과 응용이 가능함을 알 수 있을 것이다. 그리고, 이러한 변형과 응용에 관계된 차이점들은 첨부된 청구 범위에서 규정하는 본 발명의 범위에 포함되는 것으로 해석되어야 할 것이다.

도면

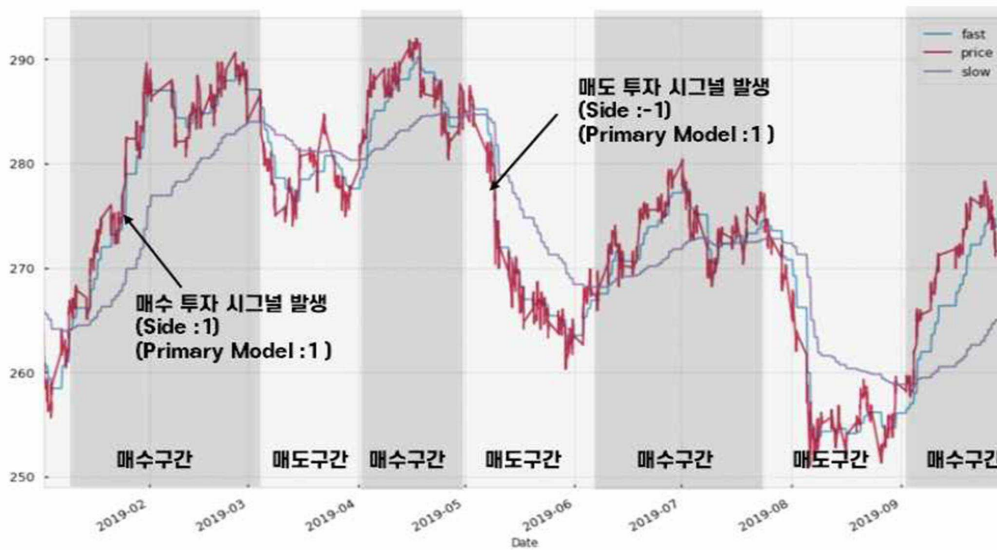
도면1



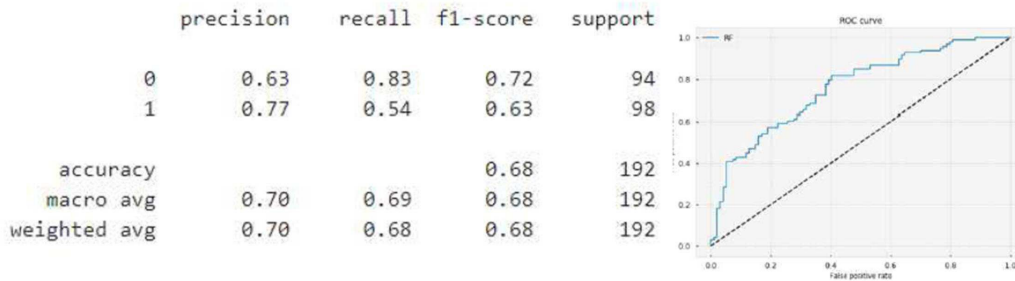
도면2



도면3

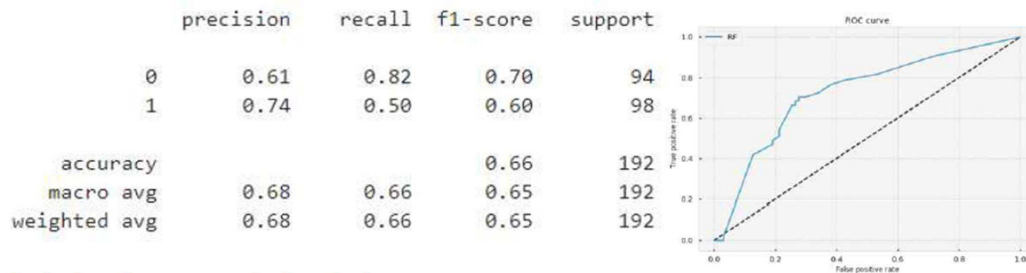


도면4





도면5



도면6

