



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2021년12월22일

(11) 등록번호 10-2342321

(24) 등록일자 2021년12월17일

(51) 국제특허분류(Int. Cl.)

G06Q 40/04 (2012.01) G06N 3/08 (2006.01)

G06Q 10/04 (2012.01) G06Q 40/06 (2012.01)

(52) CPC특허분류

G06Q 40/04 (2013.01)

G06N 3/08 (2013.01)

(21) 출원번호 10-2019-0149341

(22) 출원일자 2019년11월20일

심사청구일자 2019년11월20일

(65) 공개번호 10-2021-0061598

(43) 공개일자 2021년05월28일

(56) 선행기술조사문헌

KR1020140005606 A*

KR1020190081373 A*

KR1020190094319 A*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

세종대학교산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

유성준

서울특별시 강동구 고덕로 210, 506동 1503호(명일동, 삼익그린맨션)

이상일

경기도 수원시 영통구 권광로260번길 36, 128동 502호(매탄동, 현대힐스테이트아파트)

(74) 대리인

양성보

전체 청구항 수 : 총 2 항

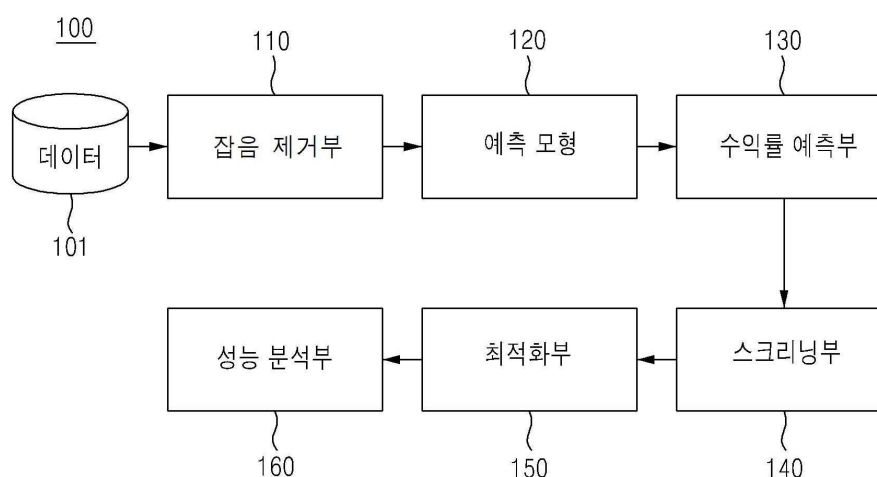
심사관 : 이지찬

(54) 발명의 명칭 딥러닝 기반의 주식 스크리닝과 이를 이용한 포트폴리오 자동화 및 고도화 방법 및 장치

(57) 요약

딥러닝 기반의 주식 스크리닝과 이를 이용한 포트폴리오 자동화 및 고도화 방법 및 장치가 제시된다. 일 실시예에 따른 딥러닝 기반 주식 스크리닝 방법은, 금융 데이터를 딥러닝(deep learning)에 적합한 데이터로 변환하기 위해 전처리하고, 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거하는 단계; 잡음이 제거된 데이터를 이용하여 딥러닝 기반 예측 모형에 입력하는 단계; 훈련된 상기 예측 모형을 통해 주식 종목별로 주가 예측 수익률을 예측하는 단계; 및 스크리닝 레벨을 설정하여, 설정된 상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택하는 단계를 포함하여 이루어질 수 있다.

대표도



(52) CPC특허분류

G06N 3/082 (2013.01)

G06Q 10/04 (2013.01)

G06Q 40/06 (2013.01)

이 발명을 지원한 국가연구개발사업

| | |
|-------------|-------------------------|
| 과제고유번호 | 1711080962 |
| 부처명 | 과학기술정보통신부 |
| 과제관리(전문)기관명 | 정보통신기획평가원 |
| 연구사업명 | ICT융합산업원천기술개발(R&D) |
| 연구과제명 | 자가진화형 인공지능 투자 기술 개발 |
| 기 여 율 | 1/1 |
| 과제수행기관명 | 세종대학교 산학협력단 |
| 연구기간 | 2007.04.01 ~ 2020.12.31 |

공지예외적용 : 있음

명세서

청구범위

청구항 1

컴퓨터 시스템에서 실행되는 딥러닝 기반 주식 스크리닝 방법에 있어서,

상기 컴퓨터 시스템은 메모리에 포함된 컴퓨터 판독가능한 명령들을 실행하도록 구성된 적어도 하나의 프로세서를 포함하고,

상기 딥러닝 기반 주식 스크리닝 방법은,

상기 적어도 하나의 프로세서에 의해, 금융 데이터를 딥러닝(deep learning)에 적합한 데이터로 변환하기 위해 전처리하고, 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거하는 단계;

상기 적어도 하나의 프로세서에 의해, 잡음이 제거된 데이터를 이용하여 딥러닝 기반 예측 모형에 입력하는 단계;

상기 적어도 하나의 프로세서에 의해, 훈련된 상기 예측 모형을 통해 주식 종목별로 주가 예측 수익률을 예측하는 단계; 및

상기 적어도 하나의 프로세서에 의해, 스크리닝 레벨을 설정하여, 설정된 상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택하는 단계

를 포함하고,

상기 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거하는 단계는,

입력층, 은닉층 및 출력층으로 구성되는 오토인코더를 사용하여 은닉층을 통해 특징을 추출하고, 은닉층을 다음 오토인코더의 입력층으로 사용하여 은닉층을 통해 특징을 추출하는 방식을 이용하여, 복수 개의 상기 오토인코더를 통해 은닉층을 쌓아 적층 잡음 제거 오토인코더(Stacked Denoising Autoencoder, SDAE)를 구축하고, 상기 적층 잡음 제거 오토인코더를 통해 고차원 레벨(high-level)의 특징을 추출하고,

상기 잡음이 제거된 데이터를 이용하여 딥러닝 기반 예측 모형에 입력하는 단계는,

딥러닝 기반 예측 모형인 전방 심층 인공경망(Feedforward Deep Neural Network, FNN)의 입력층에 상기 적층 잡음 제거 오토인코더를 연결하여 하이브리드 모형(SDAE-FNN)을 구축하고, 역전파(Backpropagation) 알고리즘을 활용하여 상기 하이브리드 모형(SDAE-FNN)을 훈련하고,

상기 예측 모형을 통해 주식 종목별로 주가 예측 수익률을 예측하는 단계는,

상기 적층 잡음 제거 오토인코더의 출력을 상기 전방 심층 인공경망의 입력층으로 사용하여 상기 역전파 알고리즘으로 하이퍼 파라미터(hyper-parameter)를 튜닝함으로써 주가 예측 수익률을 예측하는 것으로,

상기 하이퍼 파라미터의 튜닝으로 적층 개수가 결정된 상기 하이브리드 모형을 통해 코스피 지수와 국외 지수의 트레이딩 데이터와 기술적 지표로 구성된 상기 금융 데이터를 이용하여 상기 주가 예측 수익률을 예측하고,

상기 하이퍼 파라미터는 베이시안 최적화 알고리즘(Bayesian optimization algorithm)을 이용하여 평균 제곱근 오차(mean square root)를 최소화하는 파라미터로 선택되고,

상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택하는 단계는,

상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택하여 포트폴리오를 구축하고,

상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식에 대해 평균-분산 최적화 과정으로 주식 종목별 포트폴리오 비중을 도출하여 상기 포트폴리오를 최적화하고,

테스트 데이터를 이용하여 상기 포트폴리오의 누적 수익률과 변동성을 포함하는 성능을 분석하고,

상기 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거하는 단계는,

상기 금융 데이터를 0~1 사이의 값으로 스케일링하여 전처리하는 것을 특징으로 하는, 딥러닝 기반 주식 스크리닝 방법.

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

삭제

청구항 6

삭제

청구항 7

삭제

청구항 8

삭제

청구항 9

삭제

청구항 10

삭제

청구항 11

딥러닝 기반 주식 스크리닝 장치에 있어서,

금융 데이터를 딥러닝(deep learning)에 적합한 데이터로 변환하기 위해 전처리하고, 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거하는 잡음 제거부;

잡음이 제거된 데이터를 이용하여 딥러닝 기반 예측 모형에 입력하는 예측 모형;

훈련된 상기 예측 모형을 통해 주식 종목별로 주가 예측 수익률을 예측하는 수익률 예측부; 및

스크리닝 레벨을 설정하여, 설정된 상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택하는 스크리닝부

를 포함하고,

상기 잡음 제거부는,

입력층, 은닉층 및 출력층으로 구성되는 오토인코더를 사용하여 은닉층을 통해 특징을 추출하고, 은닉층을 다음 오토인코더의 입력층으로 사용하여 은닉층을 통해 특징을 추출하는 방식을 이용하여, 복수 개의 상기 오토인코더를 통해 은닉층을 쌓아 적층 잡음 제거 오토인코더(Stacked Denoising Autoencoder, SDAE)를 구축하고, 상기 적층 잡음 제거 오토인코더를 통해 고차원 레벨(high-level)의 특징을 추출하고,

상기 예측 모형은,

딥러닝 기반 예측 모형인 전방 심층 인공경망(Feedforward Deep Neural Network, FNN)의 입력층에 상기 적층 잡음 제거 오토인코더를 연결하여 하이브리드 모형(SDAE-FNN)을 구축하고, 역전파(Backpropagation) 알고리즘을 활용하여 상기 하이브리드 모형(SDAE-FNN)을 훈련하고,

상기 수익률 예측부는,

상기 적층 잡음 제거 오토인코더의 출력을 상기 전방 심층 인공경망의 입력층으로 사용하여 상기 역전파 알고리즘으로 하이퍼 파라미터(hyper-parameter)를 튜닝함으로써 주가 예측 수익률을 예측하는 것으로,

상기 하이퍼 파라미터의 튜닝으로 적층 개수가 결정된 상기 하이브리드 모형을 통해 코스피 지수와 국외 지수의 트레이딩 데이터와 기술적 지표로 구성된 상기 금융 데이터를 이용하여 상기 주가 예측 수익률을 예측하고,

상기 하이퍼 파라미터는 베이지안 최적화 알고리즘(Bayesian optimization algorithm)을 이용하여 평균 제곱근 오차(mean square root)를 최소화하는 파라미터로 선택되고,

상기 스크리닝부는,

상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택하여 포트폴리오를 구축하고,

상기 딥러닝 기반 주식 스크리닝 장치는,

상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식에 대해 평균-분산 최적화 과정으로 주식 종목별 포트폴리오 비중을 도출하여 상기 포트폴리오를 최적화하는 최적화부; 및

테스트 데이터를 이용하여 상기 포트폴리오의 누적 수익률과 변동성을 포함하는 성능을 분석하는 성능 분석부를 더 포함하고,

상기 잡음 제거부는,

상기 금융 데이터를 0~1 사이의 값으로 스케일링하여 전처리하는 것

을 특징으로 하는, 딥러닝 기반 주식 스크리닝 장치.

청구항 12

삭제

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

발명의 설명

기술 분야

[0001]

아래의 실시예들은 딥러닝 기반의 주식 스크리닝(screening)과 이를 이용한 포트폴리오 자동화 및 고도화 방법 및 장치에 관한 것으로, 더욱 상세하게는 딥러닝을 이용하여 주가 예측 수익률의 정확도를 높이고 예측된 수익률을 기반으로 포트폴리오를 구축하는 방법 및 장치에 관한 것이다.

배경 기술

[0002]

최근 인공지능 기술에 관한 관심이 증폭되면서, 기계학습(machine learning), 딥러닝(deep learning) 등에 대한 연구가 활발히 진행되고 있다. 특히 핀테크(Fin-tech) 산업에서는 결제 시스템, 금융 서비스, 은행용자 서비스, 투자 자산운용, 보험 및 시장 인프라 등을 중심으로 기존의 금융 서비스 양사에 큰 변화를 줄 수 있는

혁신이 이루어지고 있다.

[0003] 최적 포트폴리오(주식의 집합)를 구축할 때, 각 종목의 미래 예측 수익률이 필요하다. 대표적인 포트폴리오 방법론 중 하나인 평균-분산 포트폴리오의 경우, 미래 예측 수익률이 입력 변수로 사용되고 평균-분산 최적화 공식에 따라 각 종목의 비중이 도출되고, 포트폴리오의 비중은 입력변수에 큰 영향을 받는다. 따라서 정확한 미래 예측 수익률이 포트폴리오 성과의 핵심적인 역할을 한다.

[0004] 하지만, 종래의 포트폴리오 기술은 전문가의 주관적인 예측에 따라 추정되지만, 주가 수익률 예측은 지극히 어렵기 때문에 포트폴리오의 성과가 나쁜 경우가 많다.

[0005] 한국공개특허 10-2019-0064749호는 이러한 지능형 증권 투자 의사결정 지원 방법 및 그 장치에 관한 기술을 기재하고 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국공개특허 10-2019-0064749호

발명의 내용

해결하려는 과제

[0007] 실시예들은 딥러닝 기반의 주식 스크리닝과 이를 이용한 포트폴리오 자동화 및 고도화 방법 및 장치에 관하여 기술하며, 보다 구체적으로 딥러닝을 이용하여 주가 예측 수익률의 정확도를 높이고, 예측된 수익률을 기반으로 포트폴리오를 구축하는 기술을 제공한다.

[0008] 실시예들은 딥러닝을 이용하여 계량적인 방법으로 예측 정확도를 높이고 예측 수익률이 높은 종목들(스크리닝)로 포트폴리오를 구축함으로써, 포트폴리오 자동화 성과를 향상시켜 투자에 대한 리스크를 낮추고 수익률 높일 수 있는, 딥러닝 기반의 주식 스크리닝과 이를 이용한 포트폴리오 자동화 및 고도화 방법 및 장치를 제공하는데 있다.

과제의 해결 수단

[0009] 일 실시예에 따른 딥러닝 기반 주식 스크리닝 방법은, 금융 데이터를 딥러닝(deep learning)에 적합한 데이터로 변환하기 위해 전처리하고, 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거하는 단계; 잡음이 제거된 데이터를 이용하여 딥러닝 기반 예측 모형에 입력하는 단계; 훈련된 상기 예측 모형을 통해 주식 종목별로 주가 예측 수익률을 예측하는 단계; 및 스크리닝 레벨을 설정하여, 설정된 상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택하는 단계를 포함하여 이루어질 수 있다.

[0010] 상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택하는 단계는, 상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택하여 포트폴리오를 구축할 수 있다.

[0011] 상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택한 후, 평균-분산 최적화 과정으로 주식 종목별 포트폴리오 비중을 도출하여, 구축된 상기 포트폴리오를 최적화하는 단계를 더 포함할 수 있다.

[0012] 테스트 데이터를 이용하여 구축된 상기 포트폴리오의 성능을 분석하는 단계를 더 포함할 수 있다.

[0013] 상기 금융 데이터는, 코스피 지수와 국외 지수의 트레이딩 데이터와 기술적 지표를 사용하여 구성될 수 있다.

[0014] 상기 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거하는 단계는, 입력된 상기 금융 데이터를 0~1 사이의 값으로 스케일링하여 전처리할 수 있다.

[0015] 상기 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거하는 단계는, 입력층, 은닉층 및 출력층으로 구성되는 복수 개의 오토인코더의 은닉층을 쌓아 적층 잡음 제거 오토인코더 구성하고, 상기 적층 잡음 제거 오토인코더를 통해 고차원 레벨(high-level)의 특징을 추출할 수 있다.

[0016] 상기 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거하는 단계는, 입력층, 은닉층 및 출력

층으로 구성되는 오토인코더를 사용하여 상기 은닉층을 통해 특징을 추출하고, 상기 은닉층을 다음 오토인코더의 입력층으로 사용하여 은닉층을 통해 특징을 추출하는 방식을 이용하여, 복수 개의 상기 오토인코더를 통해 상기 적층 잡음 제거 오토인코더를 구축하고 고차원 레벨(high-level)의 특징을 추출할 수 있다.

[0017] 상기 딥러닝 기반 예측 모형은, 전방 심층 인공경망(Feedforward deep neural network)을 이용하여 상기 딥러닝 기반 예측 모형이 구성될 수 있다.

[0018] 상기 잡음이 제거된 데이터를 이용하여 딥러닝 기반 예측 모형에 입력하는 단계는, 딥러닝 기반 예측 모형인 전방 심층 인공경망(Feedforward Deep Neural Network, FNN)의 입력층에 상기 적층 잡음 제거 오토인코더(Stacked Denoising Autoencoder, SDAE)를 연결하여 하이브리드 모형(SDAE-FNN)을 구축하고, 역전파(Backpropagation) 알고리즘을 활용하여 상기 하이브리드 모형(SDAE-FNN)을 훈련할 수 있다.

[0019] 다른 실시예에 따른 딥러닝 기반 주식 스크리닝 장치는, 금융 데이터를 딥러닝(deep learning)에 적합한 데이터로 변환하기 위해 전처리하고, 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거하는 잡음 제거부; 잡음이 제거된 데이터를 이용하여 딥러닝 기반 예측 모형에 입력하는 예측 모형; 훈련된 상기 예측 모형을 통해 주식 종목별로 주가 예측 수익률을 예측하는 수익률 예측부; 및 스크리닝 레벨을 설정하여, 설정된 상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택하는 스크리닝부를 포함하여 이루어질 수 있다.

[0020] 상기 스크리닝부에서 상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택한 후, 평균-분산 최적화 과정으로 주식 종목별 포트폴리오 비중을 도출하여, 구축된 포트폴리오를 최적화하는 최적화부를 더 포함하고, 상기 스크리닝부는, 상기 스크리닝 레벨보다 큰 상기 주가 예측 수익률을 갖는 주식을 선택하여 포트폴리오를 구축할 수 있다.

[0021] 테스트 데이터를 이용하여 구축된 상기 포트폴리오의 성능을 분석하는 성능 분석부를 더 포함할 수 있다.

[0022] 상기 잡음 제거부는, 입력층, 은닉층 및 출력층으로 구성되는 복수 개의 오토인코더의 은닉층을 쌓아 적층 잡음 제거 오토인코더 구성하고, 상기 적층 잡음 제거 오토인코더를 통해 고차원 레벨(high-level)의 특징을 추출할 수 있다.

[0023] 상기 예측 모형은, 딥러닝 기반 예측 모형인 전방 심층 인공경망(Feedforward Deep Neural Network, FNN)의 입력층에 상기 적층 잡음 제거 오토인코더(Stacked Denoising Autoencoder, SDAE)를 연결하여 하이브리드 모형(SDAE-FNN)을 구축하고, 역전파(Backpropagation) 알고리즘을 활용하여 상기 하이브리드 모형(SDAE-FNN)을 훈련할 수 있다.

발명의 효과

[0024] 실시예들에 따르면 딥러닝을 이용하여 계량적인 방법으로 예측 정확도를 높이고 예측 수익률이 높은 종목들(스크리닝)로 포트폴리오를 구축함으로써, 포트폴리오 자동화 성과를 향상시켜 투자에 대한 리스크를 낮추고 수익률 높일 수 있는, 딥러닝 기반의 주식 스크리닝과 이를 이용한 포트폴리오 자동화 및 고도화 방법 및 장치를 제공할 수 있다.

도면의 간단한 설명

[0025] 도 1은 일 실시예에 따른 딥러닝 기반 주식 스크리닝 장치를 나타내는 블록도이다.

도 2는 일 실시예에 따른 딥러닝 기반 주식 스크리닝 방법을 나타내는 흐름도이다.

도 3은 일 실시예에 따른 적층 잡음 제거 오토인코더의 구축 방법을 설명하기 위한 도면이다.

도 4는 일 실시예에 따른 다른 적층 잡음 제거 오토인코더의 특징 추출을 설명하기 위한 도면이다.

도 5는 일 실시예에 따른 모든 종목이 동일한 비중으로 구축된 포트폴리오 예시를 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0026] 이하, 첨부된 도면을 참조하여 실시예들을 설명한다. 그러나, 기술되는 실시예들은 여러 가지 다른 형태로 변형될 수 있으며, 본 발명의 범위가 이하 설명되는 실시예들에 의하여 한정되는 것은 아니다. 또한, 여러 실시예들은 당해 기술분야에서 평균적인 지식을 가진 자에게 본 발명을 더욱 완전하게 설명하기 위해서 제공되는 것

이다. 도면에서 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있다.

- [0028] 아래의 실시예들은 딥러닝 기반의 주식 스크리닝과 이를 이용한 포트폴리오 자동화 및 고도화 방법 및 장치에 관한 것으로, 딥러닝을 이용하여 주가 예측 수익률의 정확도를 높이고, 예측된 수익률을 기반으로 포트폴리오를 구축하는 방법을 제공한다. 보다 구체적으로 실시예들은 딥러닝을 이용하여 계량적인 방법으로 예측 정확도를 높이고 예측 수익률이 높은 종목들(스크리닝)로 포트폴리오를 구축함으로써, 포트폴리오 자동화 성과를 향상시켜 투자에 대한 리스크를 낮추고 수익률 높일 수 있다.
- [0029]도 1은 일 실시예에 따른 딥러닝 기반 주식 스크리닝 장치를 나타내는 블록도이다.
- [0030]도 1을 참조하면, 일 실시예에 따른 딥러닝 기반 주식 스크리닝 장치(100)는 잡음 제거부(110), 예측 모형(120), 수익률 예측부(130) 및 스크리닝부(140)를 포함하여 이루어질 수 있다. 실시예에 따라 딥러닝 기반 주식 스크리닝 장치(100)는 최적화부(150) 및 성능 분석부(160)를 더 포함하여 이루어질 수 있다. 여기서, 일 실시예에 따른 딥러닝 기반 주식 스크리닝 장치(100)는 딥러닝 기반의 주식 스크리닝과 이를 이용한 포트폴리오 자동화 및 고도화 장치에 포함되거나 포함할 수 있다.
- [0031]잡음 제거부(110)는 금융 데이터(101)를 딥러닝(deep learning)에 적합한 데이터로 변환하기 위해 전처리하고, 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거할 수 있다.
- [0032]여기서, 주가 예측 모형(120)을 구축하기 위해 금융 데이터(101)(즉, 입력 데이터)는 코스피 지수와 국외 지수(S&P500 지수)의 트레이딩 데이터(시초가, 종가, 저가, 고가, 거래량)와 기술적 지표를 사용할 수 있다.
- [0033]잡음 제거부(110)는 입력층, 은닉층 및 출력층으로 구성되는 복수 개의 오토인코더의 은닉층을 쌓아 적층 잡음 제거 오토인코더 구성하고, 적층 잡음 제거 오토인코더를 통해 고차원 레벨(high-level)의 특징을 추출할 수 있다. 이는 아래에서 보다 상세히 설명하기로 한다.
- [0034]잡음이 제거된 데이터를 이용하여 딥러닝 기반 예측 모형(120)에 입력할 수 있다. 예측 모형(120)은 전방 심층 인공경망(Feedforward Deep Neural Network, FNN) 또는 장단기 메모리(long short-term memory)를 사용할 수 있다.
- [0035]특히, 딥러닝 기반 예측 모형(120)인 전방 심층 인공경망(FNN)의 입력층에 적층 잡음 제거 오토인코더(Stacked Denoising Autoencoder, SDAE)를 연결하여 하이브리드 모형(SDAE-FNN)을 구축하고, 역전파(Backpropagation) 알고리즘을 활용하여 하이브리드 모형(SDAE-FNN)을 훈련할 수 있다.
- [0036]수익률 예측부(130)는 훈련된 예측 모형(120)을 통해 주식 종목별로 주가 예측 수익률을 예측할 수 있다.
- [0037]스크리닝부(140)는 스크리닝 레벨을 설정하여, 설정된 스크리닝 레벨보다 큰 주가 예측 수익률을 갖는 주식을 선택할 수 있다. 여기서, 스크리닝부(140)는 스크리닝 레벨보다 큰 주가 예측 수익률을 갖는 주식을 선택하여 포트폴리오를 구축할 수 있다.
- [0038]최적화부(150)는 스크리닝부(140)에서 스크리닝 레벨보다 큰 주가 예측 수익률을 갖는 주식을 선택한 후, 평균-분산 최적화 과정으로 주식 종목별 포트폴리오 비중을 도출하여, 구축된 포트폴리오를 최적화할 수 있다.
- [0039]성능 분석부(160)는 테스트 데이터를 이용하여 구축된 포트폴리오의 성능을 분석할 수 있다.
- [0040]예를 들어, 본 실시예들에 따른 딥러닝 기반 주식 스크리닝 장치는 컴퓨터 시스템을 통해 구현될 수 있다. 컴퓨터 시스템은 딥러닝 기반 주식 스크리닝 방법을 실행하기 위한 구성요소로서 프로세서, 메모리, 영구 저장 장치, 버스, 입출력 인터페이스 및 네트워크 인터페이스를 포함할 수 있다. 예를 들어, 프로세서는 명령어들의 임의의 시퀀스를 처리할 수 있는 임의의 장치를 포함하거나 그의 일부일 수 있다. 프로세서는 예를 들어 컴퓨터 프로세서, 이동 장치 또는 다른 전자 장치 내의 프로세서 및/또는 디지털 프로세서를 포함할 수 있다. 프로세서는 예를 들어, 서버 컴퓨팅 디바이스, 서버 컴퓨터, 일련의 서버 컴퓨터들, 서버 팜, 클라우드 컴퓨터, 컨테이너 플랫폼, 이동 컴퓨팅 장치, 스마트폰, 태블릿, 셋톱 박스, 미디어 플레이어 등에 포함될 수 있다. 프로세서는 버스를 통해 메모리에 접속될 수 있다.
- [0042]도 2는 일 실시예에 따른 딥러닝 기반 주식 스크리닝 방법을 나타내는 흐름도이다.
- [0043]도 2를 참조하면, 일 실시예에 따른 딥러닝 기반 주식 스크리닝 방법은, 금융 데이터를 딥러닝에 적합한 데이터로 변환하기 위해 전처리하고, 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거하는 단계(S110), 잡음이 제거된 데이터를 이용하여 딥러닝 기반 예측 모형에 입력하는 단계(S120), 훈련된 예측 모형을

통해 주식 종목별로 주가 예측 수익률을 예측하는 단계(S130), 및 스크리닝 레벨을 설정하여, 설정된 스크리닝 레벨보다 큰 주가 예측 수익률을 갖는 주식을 선택하는 단계(S140)를 포함하여 이루어질 수 있다.

[0044] 또한, 스크리닝 레벨보다 큰 주가 예측 수익률을 갖는 주식을 선택한 후, 평균-분산 최적화 과정으로 주식 종목별 포트폴리오 비중을 도출하여, 구축된 포트폴리오를 최적화하는 단계(S150)를 더 포함할 수 있다.

[0045] 또한, 테스트 데이터를 이용하여 구축된 포트폴리오의 성능을 분석하는 단계(S160)를 더 포함할 수 있다.

[0046] 이와 같이, 실시예들에 따르면 딥러닝을 이용하여 주가 수익률 예측의 정확도를 높일 수 있고, 예측된 주가를 이용하여 임계화 기반 포트폴리오를 구축함으로써 투자에 대한 리스크를 낮추고 수익률을 높일 수 있다.

[0047] 아래에서 일 실시예에 따른 딥러닝 기반 주식 스크리닝 방법의 각 단계를 보다 구체적으로 설명한다.

[0049] 일 실시예에 따른 딥러닝 기반 주식 스크리닝 방법은 일 실시예에 따른 딥러닝 기반 주식 스크리닝 장치(100)를 예를 들어 보다 상세히 설명할 수 있다. 일 실시예에 따른 딥러닝 기반 주식 스크리닝 장치(100)는 잡음 제거부(110), 예측 모형(120), 수익률 예측부(130) 및 스크리닝부(140)를 포함하여 이루어질 수 있고, 실시예에 따라 딥러닝 기반 주식 스크리닝 장치(100)는 최적화부(150) 및 성능 분석부(160)를 더 포함하여 이루어질 수 있다.

[0050] 단계(S110)에서, 잡음 제거부(110)는 금융 데이터를 딥러닝에 적합한 데이터로 변환하기 위해 전처리하고, 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음을 제거할 수 있다. 보다 구체적으로, 잡음 제거부(110)는 전처리 작업으로서 딥러닝에 적합한 데이터로 변환하기 위해 입력 데이터를 0~1 사이의 값으로 스케일링하고, 잡음 제거 오토인코더를 이용하여 금융 시계열에 내포된 잡음 제거할 수 있다. 여기서, 금융 데이터는 코스피 지수와 국외 지수의 트레이딩 데이터와 기술적 지표를 사용하여 구성될 수 있다.

[0051] 잡음 제거부(110)는 입력층, 은닉층 및 출력층으로 구성되는 복수 개의 오토인코더의 은닉층을 쌓아 적층 잡음 제거 오토인코더 구성하고, 적층 잡음제거 오토인코더를 통해 고차원 레벨의 특징을 추출할 수 있다. 보다 구체적으로, 잡음 제거부(110)는 입력층, 은닉층 및 출력층으로 구성되는 오토인코더를 사용하여 은닉층을 통해 특징을 추출하고, 은닉층을 다음 오토인코더의 입력층으로 사용하여 은닉층을 통해 특징을 추출하는 방식을 이용하여, 복수 개의 오토인코더를 통해 적층 잡음제거 오토인코더를 구축하고 고차원 레벨의 특징을 추출할 수 있다.

[0052] 단계(S120)에서, 예측 모형(120)은 잡음이 제거된 데이터를 이용하여 딥러닝 기반 예측 모형(120)에 입력할 수 있다. 여기서, 딥러닝 기반 예측 모형(120)은 전방 심층 인공경망을 이용하여 딥러닝 기반 예측 모형(120)이 구성될 수 있다. 딥러닝 기반 예측 모형(120)인 전방 심층 인공경망(FNN)의 입력층에 적층 잡음제거 오토인코더(SDAE)를 연결하여 하이브리드 모형(SDAE-FNN)을 구축하고, 역전파 알고리즘을 활용하여 하이브리드 모형(SDAE-FNN)을 훈련할 수 있다.

[0053] 단계(S130)에서, 수익률 예측부(130)는 훈련된 예측 모형(120)을 통해 주식 종목별로 주가 예측 수익률을 예측할 수 있다. 예를 들어, 수익률 예측부(130)는 훈련한 예측 모형(120)으로 주가 일별/주별/월별 수익률을 예측할 수 있다.

[0054] 단계(S140)에서, 스크리닝부(140)는 스크리닝 레벨을 설정하여, 설정된 스크리닝 레벨보다 큰 주가 예측 수익률을 갖는 주식을 선택할 수 있다. 여기서, 스크리닝부(140)는 스크리닝 레벨보다 큰 주가 예측 수익률을 갖는 주식을 선택하여 포트폴리오를 구축할 수 있다.

[0055] 단계(S150)에서, 최적화부(150)는 스크리닝 레벨보다 큰 주가 예측 수익률을 갖는 주식을 선택한 후, 평균-분산 최적화 과정으로 주식 종목별 포트폴리오 비중을 도출하여, 구축된 포트폴리오를 최적화할 수 있다.

[0056] 단계(S160)에서, 성능 분석부(160)는 테스트 데이터를 이용하여 구축된 포트폴리오의 성능을 분석할 수 있다. 여기서, 성능 분석부(160)는 테스트 데이터를 이용하여 구축된 포트폴리오의 누적 수익률, 변동성 등을 분석함으로써 포트폴리오의 성능을 분석할 수 있다.

[0058] 아래에서는 일 실시예에 따른 딥러닝 기반 주식 스크리닝 방법을 하나의 예를 통해 보다 상세히 설명한다.

[0059] 본 실시예들의 핵심적 목표는 다른 시간대에 있는 주식 시장 데이터를 효과적으로 결합하는 것이다. 예측 모형을 구축하기 위해 미국 주식 데이터 S&P500에 속한 10 개의 주식을 선택할 수 있다. 여기에서는 하나의 예시로서, Apple(AAPL), Amazon(AMZN), Bank of America Corporation(BAC), Berkshire Hathaway Inc. Class B(BRK-B), General Electric Company(GE), Johnson & Johnson(JNJ), JPMorgan Chase & Co.(JPM), Microsoft

Corporation(MSFT), AT & T Inc.(T), 및 Wells Fargo & Company(WFC)의 10 종목을 선택한다. 각 데이터는 예컨대 yahoo finance에서 내려 받을 수 있다. 또한, 각 데이터의 앞 70%는 학습용, 뒤 30%는 예측 모형 평가를 위해 사용할 수 있다. 예측 모형의 구축에 사용되는 데이터는 두 지수의 거래 데이터(시가, 고가, 저가, 종가, 거래량)와 기술적 지표로 구성될 수 있다.

[0060] 오토인코더는 기본적으로 3 개의 층, 즉 입력층, 은닉층, 출력층으로 구성될 수 있다. 또한, 적층 잡음 제거 오토인코더는 여러 오토인코더의 은닉층을 쌓음으로써 구성될 수 있다. 오토인코더의 작동 원리는 입력층의 입력 데이터와 출력층의 출력 데이터 간의 오차를 최소화하는 방식으로 최적화 되며, 이 과정에서 은닉층은 고차원 레벨 특징을 추출할 수 있다. 또한, 은닉층으로 추출된 특징은 다음 오토인코더의 입력 데이터로 사용되고, 위 과정과 동일한 방식으로 고차원 레벨을 추출할 수 있다. 이런 방식으로 추출된 은닉층을 쌓음(아래 그림의 왼쪽 그림, SDAE)으로써 적층 잡음 제거 오토인코더를 구축 할 수 있다.

[0061] 동일한 방식으로 S&P500 지수에 대한 적층 잡음 오토인코더를 구축할 수 있다.

[0062] 도 3은 일 실시예에 따른 적층 잡음 제거 오토인코더의 구축 방법을 설명하기 위한 도면이다.

[0063] 도 3을 참조하면, 적층 잡음 제거 오토인코더(SDAE)의 구축 방법을 나타내는 것으로, 예를 들어 한국 주식 데이터(Korea stock data)에 일정한 잡음으로 부여해 오류가 생긴 데이터(corrupted data)를 만들 수 있다. 그리고 입력층-은닉층-출력층으로 구성된 단위 오토인코더로 사용하여 특징을 추출할 수 있다. 이 때 오토인코더의 은닉층을 통해 특징을 추출할 수 있다. 은닉층은 다음 오토인코더의 입력층으로 사용할 수 있다. 이러한 과정을 반복하여 여러 은닉층을 통해 특징을 추출하고, 추출한 특징을 쌓음으로써 적층 잡음 제거 오토인코더(SDAE)를 구축할 수 있다.

[0064] 다시 말하면, 전방 심층 인공경망(FNN)의 입력층에 적층 잡음 제거 오토인코더(SDAE)를 연결해 하이브리드 모형(SDAE-FNN)을 구축하고, 역전파 알고리즘을 활용해 하이브리드 모형(SDAE-FNN)을 훈련할 수 있다.

[0065] 도 4는 일 실시예에 따른 다른 적층 잡음 제거 오토인코더의 특징 추출을 설명하기 위한 도면이다.

[0066] 도 4를 참조하면, 적층 잡음 제거 오토인코더(SDAE, 410)를 통해 고차원 레벨의 특징을 추출할 수 있다. 각각의 적층 잡음 인코더(즉, 오토인코더)의 출력은 전방 심층 인공경망(FNN)의 입력층에 연결될 수 있다. 적층 잡음 제거 오토인코더(SDAE, 410)와 전방 심층 인공경망(FNN, 420)으로 구성된 전체 신경망은 역전파 알고리즘으로 하이퍼 파라미터를 튜닝할 수 있다. 즉, 적층 잡음 제거 오토인코더(SDAE, 410)와 전방 심층 인공경망(FNN, 420)으로 구성된 전체 신경망을 통해 금융 데이터와 기술적 지표(401)를 이용하여 주가 예측 수익률(402)를 산출할 수 있다.

[0067] 여기서, 튜닝을 통해 최적의 적층 개수를 결정할 수 있으며, 예컨대 10 개의 층으로 적층될 수 있다.

[0068] 예측 모형의 하이퍼 파라미터(hyper-parameter)는 베이지안 최적화 알고리즘(Bayesian optimization algorithm)을 이용하여 검증 데이터에서 평균 제곱근 오차(mean square root)를 최소화하는 파라미터를 선택할 수 있다.

[0069] 여기서, 베이지안 최적화 알고리즘은 모든 경우의 하이퍼 파라미터를 고려하는 대신 계산된 파라미터 값을 이용하여 우선적으로 고려해야 될 파라미터를 찾음으로써 계산 과정에 필요한 시간을 단축할 수 있다.

[0070] 그리고, 스크리닝 레벨을 선택할 수 있다. 예컨대, 본 실험에서는 두 가지 예시로써 0.0005, 0.001 값의 두 레벨을 선택할 수 있다. 즉, 각 종목의 딥러닝 예측 수익률이 0.0005(또는 0.001)보다 높은 종목만을 선택함으로써 향후 수익률이 좋은 것으로 예측되는 종목만을 선택해 포트폴리오를 구성할 수 있다.

[0071] 선택된 종목에 대해 평균-분산 최적화 과정을 거쳐 주식 종목별 비중을 산출할 수 있다. 여기서, 평균-분산 최적화 과정은 Markowitz의 Portfolio selection(1952)에서 제안된 방법을 이용할 수 있다. 위 방법은 리스크 대비 기대수익률을 최대화 하는 종목 비중을 구하는 방법으로 금융 투자에서 가장 널리 사용되는 방법이다. 기존 Markowitz 포트폴리오 모형은 입력 정보로서 사람이 추정한 미래 주가 수익률을 사용한다. 반면, 본 실시예에 따른 딥러닝 기반 주식 스크리닝 방법 및 장치는 이러한 과정을 딥러닝 방법으로 대체해 최적 포트폴리오 구축 과정을 자동화하고, 딥러닝의 높은 예측 정확도로 포트폴리오의 수익률을 높일 수 있다.

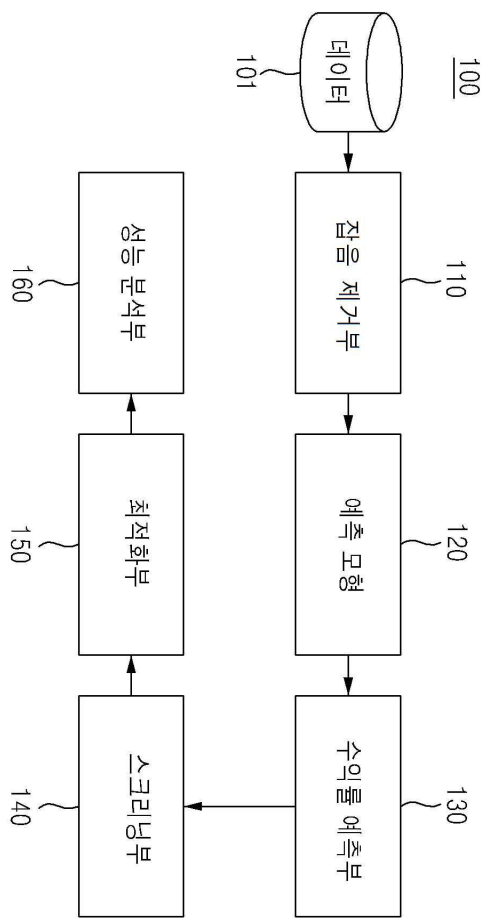
[0072] 스크리닝 레벨 0.001에 대해서도 동일한 계산을 반복할 수 있다.

[0073] 본 실시예에 따른 딥러닝 기반 주식 스크리닝 방법 및 장치의 평가는 테스트 기간(모형 구축에 사용된 기간 외 구간) 동안 포트폴리오의 누적 수익률 계산할 수 있다.

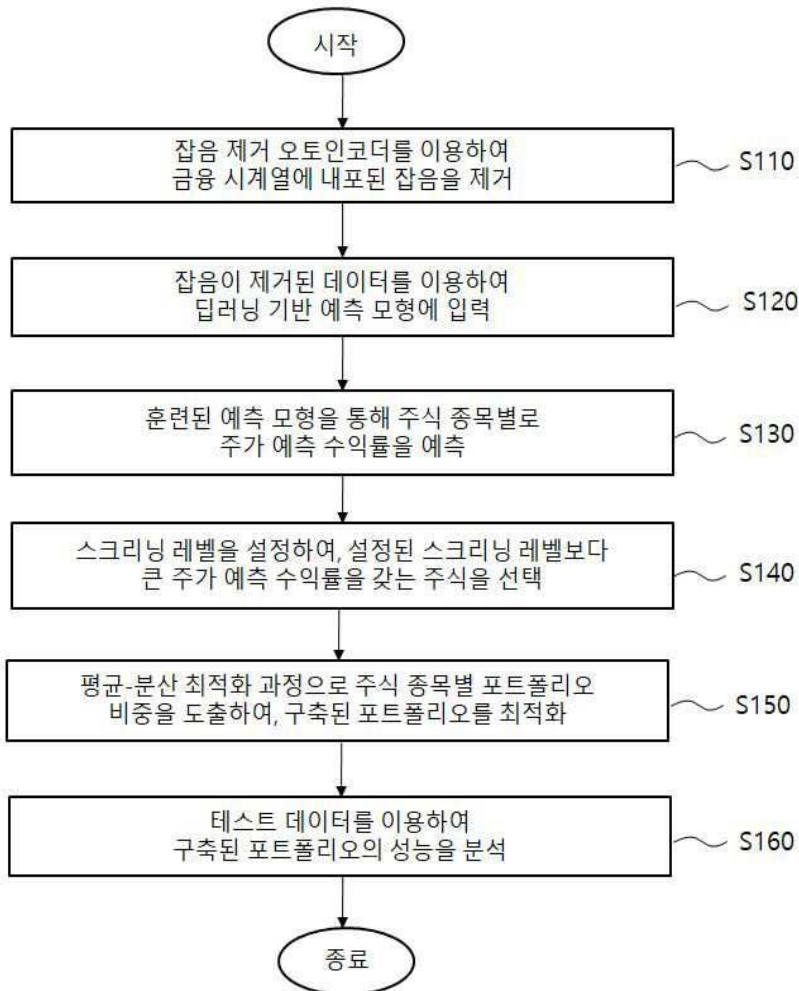
- [0074] 도 5는 일 실시예에 따른 모든 종목이 동일한 비중으로 구축된 포트폴리오 예시를 나타내는 도면이다.
- [0075] 도 5를 참조하면, 계산된 결과를 벤치마크(S&P500 지수, 동일 비중 포트폴리오)와 비교한 결과를 나타낸다. 여기서, EW는 모든 종목이 동일한 비중으로 구축된 포트폴리오이고, SP500은 S&P500 지수이며, Screening level: 0.0005 및 Screening level: 0.001은 본 실시예에 따라 구축된 포트폴리오이다. 본 실시예에 따라 구축된 포트폴리오는 동일 비중(EW) 및 시장 지수(SP500)와 비교했을 때 보다 높은 수익률을 달성하는 것을 확인할 수 있다.
- [0076] 즉, 벤치마크 지수 대비 우수한 수익률(0.001의 경우 S&P500지수 대비 약 2배, EW 대비 1.5배)은 본 실시예에 따라 구축된 포트폴리오의 우수한 성과를 보여준다.
- [0078] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 컨트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.
- [0079] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.
- [0080] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.
- [0081] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.
- [0082] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

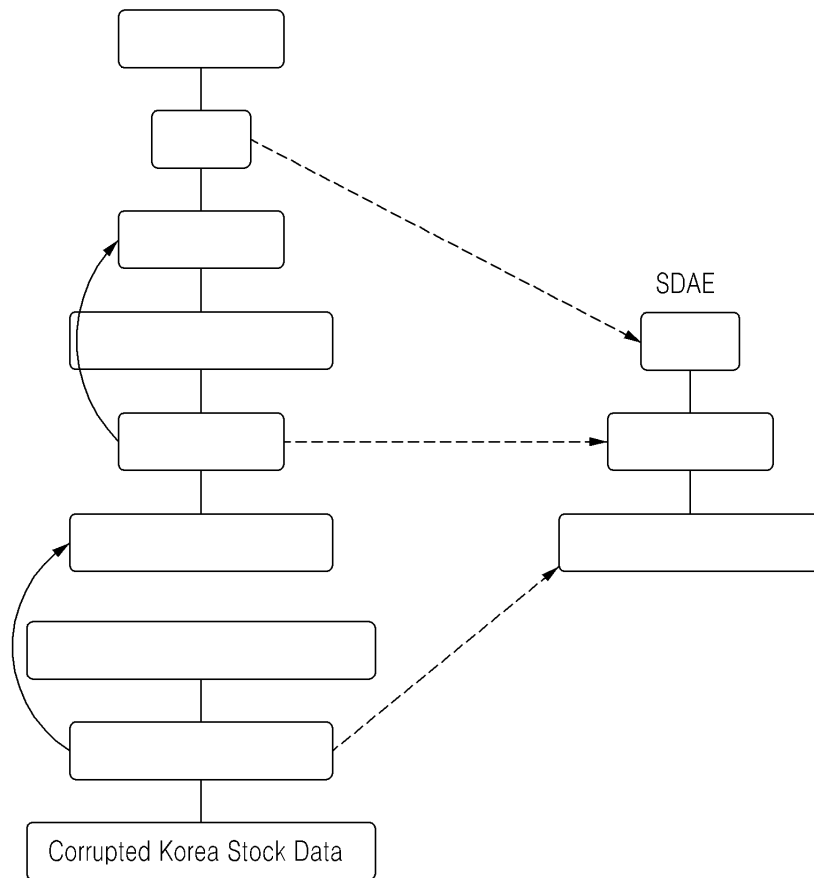
도면1



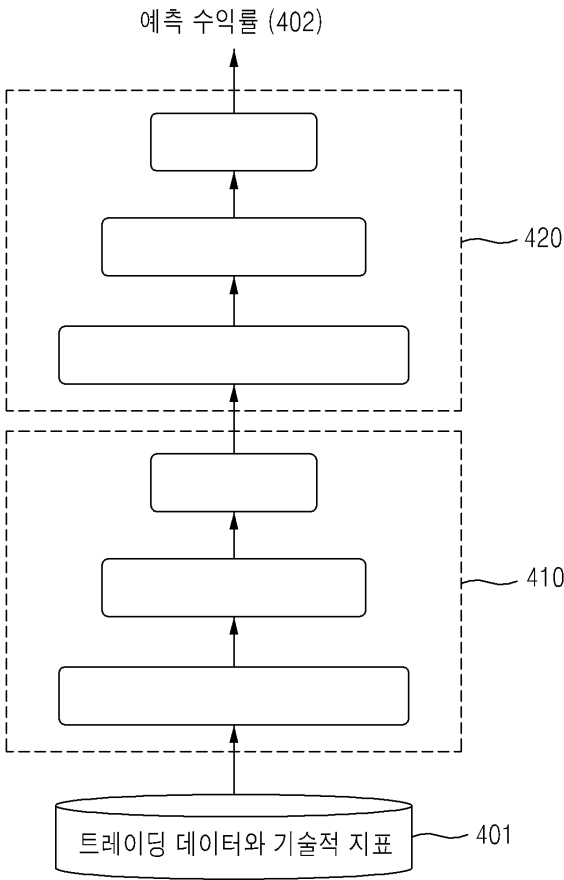
도면2



도면3



도면4



도면5

