



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0099321
(43) 공개일자 2021년08월12일

(51) 국제특허분류(Int. Cl.)

G06Q 40/04 (2012.01)

(52) CPC특허분류

G06Q 40/04 (2013.01)

(21) 출원번호 10-2020-0013024

(22) 출원일자 2020년02월04일

심사청구일자 2020년02월04일

(71) 출원인

울산대학교 산학협력단

울산광역시 남구 대학로 93(무거동)

(72) 발명자

윤석훈

울산광역시 남구 문수로409번길 20, 106동 902호
(신정동, 문수로아이파크)

웬 티 투

울산광역시 남구 대학로 93, 7호관 201-2호 (무거동)

(74) 대리인

김종선, 이형석

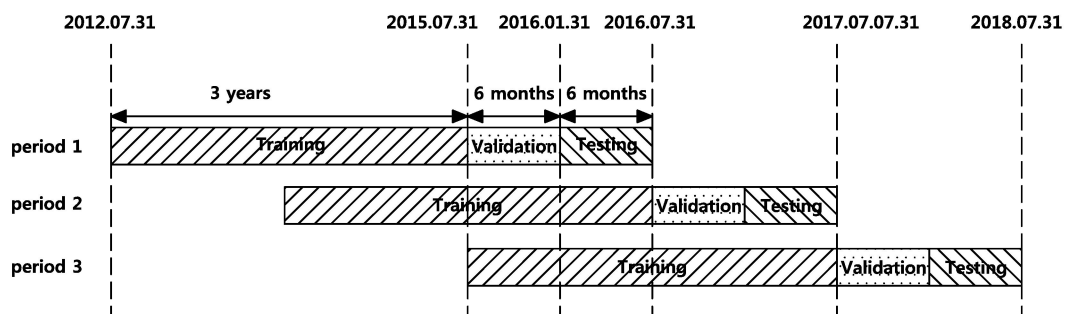
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 주가 변동 예측 장치 및 방법

(57) 요약

본 발명은 주가 변동 예측 장치 및 방법에 관한 것으로, 다수의 다른 종목의 주식 데이터 및 관심종목의 주식 데이터를 저장하는 저장부, 및 전이학습에 이용할 베이스 모델을 생성한 후 상기 다수의 다른 종목의 주식 데이터를 이용하여 훈련하고, 훈련된 베이스 모델의 모델 파라미터를 기반으로 생성되는 예측 모델을 상기 관심종목의 주식 데이터를 이용하여 훈련하고 훈련된 예측 모델을 통해 상기 관심종목의 주가 변동을 예측하는 처리부를 포함한다.

대표도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호 1345303144

부처명 교육부

과제관리(전문)기관명 한국연구재단

연구사업명 지역대학우수과학자지원사업(1년~3년)

연구과제명 사용자 이동성 예측 모델 및 기회적이동소셜네트워크에서의 메시지 라우팅 기술 개

발

기 여 율 1/2

과제수행기관명 울산대학교

연구기간 2016.11.01 ~ 2019.10.31

이 발명을 지원한 국가연구개발사업

과제고유번호 1711095211

부처명 과학기술정보통신부

과제관리(전문)기관명 한국연구재단

연구사업명 기본연구

연구과제명 악취 추적 및 대기질 모니터링을 위한 딥러닝 기반의 모바일 클라우드 센싱 네트워

크 기술 개발

기 여 율 1/2

과제수행기관명 울산대학교

연구기간 2019.06.01 ~ 2022.02.28

명세서

청구범위

청구항 1

다수의 다른 종목의 주식 데이터 및 관심종목의 주식 데이터를 저장하는 저장부, 및

전이학습에 이용할 베이스 모델을 생성한 후 상기 다수의 다른 종목의 주식 데이터를 이용하여 훈련하고, 훈련된 베이스 모델의 모델 파라미터를 기반으로 생성되는 예측 모델을 상기 관심종목의 주식 데이터를 이용하여 훈련하고 훈련된 예측 모델을 통해 상기 관심종목의 주가 변동을 예측하는 처리부를 포함하는 주가 변동 예측 장치.

청구항 2

제1항에 있어서,

상기 처리부는, 상기 다수의 다른 종목의 주식 데이터로부터 소스 데이터를 획득하고 획득된 소스 데이터를 이용하여 상기 베이스 모델을 훈련하는 것을 특징으로 하는 주가 변동 예측 장치.

청구항 3

제1항에 있어서,

상기 처리부는, 상기 관심종목의 주식 데이터로부터 타겟 데이터를 획득하고 획득된 타겟 데이터를 이용하여 상기 예측 모델을 훈련하는 것을 특징으로 하는 주가 변동 예측 장치.

청구항 4

제3항에 있어서,

상기 타겟 데이터는, 관심종목, 인덱스 및 관심종목과 관련된 다른 종목의 주식 수익 정보를 이용하여 생성되는 것을 특징으로 하는 주가 변동 예측 장치.

청구항 5

제1항에 있어서,

상기 처리부는, 타겟 일자를 기준으로 정해진 기간의 과거 관심종목 주식 데이터를 상기 훈련된 예측 모델에 적용하여 타겟 일자의 관심종목 주가 상승 여부를 결정하는 것을 특징으로 하는 주가 변동 장치.

청구항 6

전이학습에 이용할 베이스 모델을 생성하는 제1단계,

상기 베이스 모델을 다수의 다른 종목의 주식 데이터를 이용하여 훈련하는 제2단계,

상기 훈련된 베이스 모델의 모델 파라미터를 기반으로 예측 모델을 생성하는 제3단계,

상기 예측 모델을 관심종목의 주식 데이터를 이용하여 훈련하는 제4단계, 및

상기 훈련된 예측 모델을 이용하여 상기 관심종목의 주가 변동을 예측하는 제5단계를 포함하는 주가 변동 예측

방법.

청구항 7

제6항에 있어서,

상기 제2단계는,

상기 다수의 다른 종목의 주식 데이터로부터 소스 데이터를 획득하는 단계, 및

상기 소스 데이터를 이용하여 상기 베이스 모델을 훈련하는 단계를 포함하는 것을 특징으로 하는 주가 변동 예측 방법.

청구항 8

제6항에 있어서,

상기 제4단계는,

상기 관심종목의 주식 데이터로부터 타겟 데이터를 획득하는 단계, 및

상기 타겟 데이터를 이용하여 상기 예측 모델을 훈련하는 단계를 포함하는 것을 특징으로 하는 주가 변동 예측 방법.

청구항 9

제8항에 있어서,

상기 타겟 데이터는, 관심종목, 인덱스 및 관심종목과 관련된 다른 종목들의 주식 수익 정보를 이용하여 생성되는 것을 특징으로 하는 주가 변동 예측 방법.

청구항 10

제6항에 있어서,

상기 제5단계는,

타겟 일자를 기준으로 정해진 기간의 과거 관심종목 주식 데이터를 상기 훈련된 예측 모델에 적용하여 타겟 일자의 관심종목 주가 상승 여부를 결정하는 것을 특징으로 하는 주가 변동 예측 방법.

발명의 설명

기술 분야

[0001] 본 발명은 주가 변동 예측 장치 및 방법에 관한 것으로, 특히 전이학습 기법을 기반으로 소량의 관심종목 주식 데이터만 사용하여 주가 변동을 예측하는 장치 및 방법에 관한 것이다.

배경 기술

[0002] 인공지능경망 기반의 딥러닝(deep learning) 기술은 과거의 기계학습 모델에 비하여 높은 예측 성능을 보이므로, 주식 가격 및 주가 변동 예측에서도 딥러닝 기술을 이용하려는 시도가 있었다.

[0003] 딥러닝 기반 모델은 일반적으로 많은 양의 데이터를 필요로 한다. 일 단위로 관심종목(Company of Interest, COI)의 주가 변동을 예측하는 예측 모델을 만들기 위해서는 과거의 주식 가격 데이터(주가 데이터)가 필요하다.

[0004] 그런데, 일 단위의 주식 가격은 하루에 1개의 데이터 샘플(예: 종가 데이터 등)만 존재하며 1년에 240여개의 샘플 데이터만 수집이 가능하다. 즉, 지난 수년간의 데이터를 이용한다고 해도 획득할 수 있는 데이터 양에 한계가 있으며, 딥러닝 기술 적용 시 부족한 데이터로 인해 과적합 문제(overfitting problem)에 봉착하기 쉽다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 전이학습 기법을 기반으로 소량의 관심종목 주식 데이터만 사용하여 주가 변동을 예측하는 예측 모델을 훈련하는 주가 변동 예측 장치 및 방법을 제공하고자 한다.

[0006] 또한, 본 발명은 예측 모델 성능을 향상시키기 위해 관심종목과 관계성이 있는 다른 종목들의 주식 데이터들을 이용하여 관심종목의 주가 변동을 예측하는 주가 변동 예측 장치 및 방법을 제공하고자 한다.

과제의 해결 수단

[0007] 상기한 과제를 해결하기 위하여, 본 발명의 일 실시 예에 따른 주가 변동 예측 장치는 다수의 다른 종목의 주식 데이터 및 관심종목의 주식 데이터를 저장하는 저장부, 및 전이학습에 이용할 베이스 모델을 생성한 후 상기 다수의 다른 종목의 주식 데이터를 이용하여 훈련하고, 훈련된 베이스 모델의 모델 파라미터를 기반으로 생성되는 예측 모델을 상기 관심종목의 주식 데이터를 이용하여 훈련하고 훈련된 예측 모델을 통해 상기 관심종목의 주가 변동을 예측하는 처리부를 포함한다.

[0008] 상기 처리부는, 상기 다수의 다른 종목의 주식 데이터로부터 소스 데이터를 획득하고 획득된 소스 데이터를 이용하여 상기 베이스 모델을 훈련하는 것을 특징으로 한다.

[0009] 상기 처리부는, 상기 관심종목의 주식 데이터로부터 타겟 데이터를 획득하고 획득된 타겟 데이터를 이용하여 상기 예측 모델을 훈련하는 것을 특징으로 한다.

[0010] 상기 타겟 데이터는, 관심종목, 인덱스 및 관심종목과 관련된 다른 종목의 주식 수익 정보를 이용하여 생성되는 것을 특징으로 한다.

[0011] 상기 처리부는, 타겟 일자를 기준으로 정해진 기간의 과거 관심종목 주식 데이터를 상기 훈련된 예측 모델에 적용하여 타겟 일자의 관심종목 주가 상승 여부를 결정하는 것을 특징으로 한다.

[0012] 한편, 본 발명의 일 실시 예에 따른 주가 변동 예측 방법은 전이학습에 이용할 베이스 모델을 생성하는 제1단계, 상기 베이스 모델을 다수의 다른 종목의 주식 데이터를 이용하여 훈련하는 제2단계, 상기 훈련된 베이스 모델의 모델 파라미터를 기반으로 예측 모델을 생성하는 제3단계, 상기 예측 모델을 관심종목의 주식 데이터를 이용하여 훈련하는 제4단계, 및 상기 훈련된 예측 모델을 이용하여 상기 관심종목의 주가 변동을 예측하는 제5단계를 포함한다.

[0013] 상기 제2단계는, 상기 다수의 다른 종목의 주식 데이터로부터 소스 데이터를 획득하는 단계, 및 상기 소스 데이터를 이용하여 상기 베이스 모델을 훈련하는 단계를 포함하는 것을 특징으로 한다.

[0014] 상기 제4단계는, 상기 관심종목의 주식 데이터로부터 타겟 데이터를 획득하는 단계, 및 상기 타겟 데이터를 이용하여 상기 예측 모델을 훈련하는 단계를 포함하는 것을 특징으로 한다.

[0015] 상기 타겟 데이터는, 관심종목, 인덱스 및 관심종목과 관련된 다른 종목들의 주식 수익 정보를 이용하여 생성되는 것을 특징으로 한다.

[0016] 상기 제5단계는, 타겟 일자를 기준으로 정해진 기간의 과거 관심종목 주식 데이터를 상기 훈련된 예측 모델에 적용하여 타겟 일자의 관심종목 주가 상승 여부를 결정하는 것을 특징으로 한다.

발명의 효과

[0017] 본 발명에 따르면, 전이학습 기법을 기반으로 소량의 관심종목 주식 데이터를 사용하여 주가 변동을 예측하는 예측 모델을 훈련하므로, 데이터 부족으로 인해 야기되는 과적합 문제를 해결할 수 있다.

[0018] 또한, 본 발명에 따르면, 관심종목과 관계성이 있는 다른 종목들의 주식 데이터들을 이용하여 관심종목의 주가 변동을 예측하므로 주가 변동 예측 정확도를 향상시킬 수 있다.

도면의 간단한 설명

- [0019] 도 1은 본 발명의 일 실시 예에 따른 주가 변동 예측 장치의 블록구성도.
 도 2는 본 발명의 일 실시 예에 따른 데이터세트 전처리를 설명하기 위한 도면.
 도 3은 본 발명의 일 실시 예에 따른 전이학습 과정을 설명하기 위한 도면.
 도 4는 본 발명의 일 실시 예에 따른 베이스 모델의 구조.
 도 5a 내지 도 5c는 본 발명의 일 실시 예에 따른 베이스 모델의 입력특성을 설명하기 위한 도면.
 도 6은 본 발명의 일 실시 예에 따른 주가 이동 예측 방법을 도시한 흐름도.
 도 7 내지 도 9는 본 발명의 일 실시 예에 따른 예측 모델의 성능 실험 결과를 도시한 도면.

발명을 실시하기 위한 구체적인 내용

- [0020] 본 명세서에 기재된 "포함하다", "구성하다", "가지다" 등의 용어는 특별히 반대되는 기재가 없는 한 해당 구성 요소가 내재될 수 있음을 의미하는 것이므로 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [0021] 또한, 본 명세서에 기재된 "...부", "...기", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다. 또한, "일", "하나" 및 "그" 등의 관사는 본 발명을 기술하는 문맥에 있어서 본 명세서에 달리 지시되거나 문맥에 의해 분명하게 반박되지 않는 한, 단수 및 복수 모두를 포함하는 의미로 사용될 수 있다.
- [0022] 이하에서는 첨부된 도면들을 참조하여 본 발명의 실시 예를 상세하게 설명한다.
- [0023] 본 명세서에서는 주가 변동 예측 기술에 딥 러닝 기술 적용 시 부족한 데이터로 인해 야기되는 과적합 문제를 해결하기 위하여, 전이학습(Transfer learning)을 기반으로 소량의 관심종목(COI: Company of Interest) 주식 데이터만을 사용하여 예측 모델(prediction model)을 훈련하는 심층 전이 기반 프레임워크(deep transfer-based framework)를 제안한다.
- [0024] 또한, 본 명세서에서는 예측 모델의 성능을 향상시키기 위해 관심종목(타겟 종목)과 관계성이 있는 다른 종목의 주식 데이터를 이용하여 관심종목의 주가 변동을 예측하는 모델을 제안한다.
- [0025] 도 1은 본 발명의 일 실시 예에 따른 주가 변동 예측 장치의 블록구성도이고, 도 2는 본 발명의 일 실시 예에 따른 데이터세트 전처리를 설명하기 위한 도면이며, 도 3은 본 발명의 일 실시 예에 따른 전이학습 과정을 설명하기 위한 도면이고, 도 4는 본 발명의 일 실시 예에 따른 베이스 모델의 구조이며, 도 5a 내지 도 5c는 본 발명의 일 실시 예에 따른 베이스 모델의 입력특성을 설명하기 위한 도면이다.
- [0026] 도 1에 도시된 바와 같이, 주가 변동 예측 장치(100)는 통신부(110), 입력부(120), 저장부(130), 출력부(140) 및 처리부(150)를 포함한다.
- [0027] 통신부(110)는 외부의 단말 또는 서버 등과 데이터 통신을 수행할 수 있게 한다. 통신부(110)는 랜(LAN: Local Area Network), 이더넷(Ethernet), 와이파이(Wi-Fi: Wireless LAN), 와이브로(Wibro: Wireless broadband), Wimax(World Interoperability for Microwave Access) 및 HSDPA(High Speed Downlink Packet Access) 등의 네트워크를 통해 외부의 단말 또는 서버 등과 데이터를 주고 받을 수 있다.
- [0028] 통신부(110)는 처리부(150)의 지시에 따라 외부로부터 샘플 데이터(dataset)를 획득(수집)할 수 있다. 여기서, 샘플 데이터는 관심종목의 주식 데이터 및/또는 다른 종목들의 주식 데이터들을 포함한다. 통신부(110)는 외부의 주식 정보 제공 사이트에 접속하여 특정 주식 시장 인덱스(예: KOSPI 200 및 S&P 500 등)의 과거 종가 데이터(historical closing price data)를 수집한다. 이때, 통신부(110)는 지정해진 기간 내 인덱스별 모든 종목의 종가 데이터(주식 데이터 또는 주식 정보)를 수집한다. 예를 들어, 통신부(110)는 야후 금융 웹사이트(<https://finance.yahoo.com/>) 등에 접속하여 2012년 7월 31일부터 2018년 7월 31일까지의 인덱스별 모든 종목의 종가 데이터(주식 정보)를 수집할 수 있다.
- [0029] 또한, 통신부(110)는 기설정된 기준에 따라 지정해진 기간 내 관심종목의 주식 데이터 및/또는 다수의 다른 종목의 주식 데이터(다른 회사들의 주식 데이터)들을 수집할 수 있다. 기설정된 기준은 사전에 사용자에게 의해 정

의될 수 있다. 예를 들어, 통신부(110)는 처리부(150)의 지시에 따라 2012년부터 2018년까지의 시가총액 기준 상위 50개 종목의 주식 데이터를 수집할 수 있다.

[0030] 통신부(110)는 외부로부터 샘플 데이터를 수신하여 처리부(150)로 직접 전송하거나 또는 저장부(130)에 저장할 수 있다. 또한, 통신부(110)는 처리부(150)의 제어에 따라 주가 변동 예측 결과를 외부로 전송할 수 있다.

[0031] 입력부(120)는 통신부(110)를 통해 수집된 샘플 데이터 또는 저장부(130)에 저장된 샘플 데이터를 가공하여 처리부(150)로 전송할 수 있다. 즉, 입력부(120)는 샘플 데이터를 처리부(150)가 처리할 수 있는 데이터 형태로 전처리(pre-processing)하여 처리부(150)로 전송한다.

[0032] 입력부(120)는 수집된 샘플 데이터를 3개의 데이터세트 즉, 훈련 데이터세트(training set), 검증 데이터세트(validation set), 및 테스트 데이터세트(test set)으로 구분한다. 예를 들어, 도 2에 도시된 바와 같이 입력부(120)는 4년간의 주식 데이터 중 처음 3년간의 데이터를 훈련 데이터세트로, 다음 6개월 주식 데이터를 검증 데이터세트, 마지막 6개월 주식 데이터를 테스트 데이터세트로 분류한다.

[0033] 입력부(120)는 모든 종목의 주식 데이터에 대해 각 종목 s 의 t 일에 대한 1일($p=1$)의 주식 수익률 $r_{t,1}^s$ 을 계산한다. 그리고, 입력부(120)는 $(t+1)$ 일의 주가 변동을 예측하기 위하여 기정해진 기간 u 예컨대, $(t-19)$ 에서 t 까지의 일별 주식 수익률을 산출한다. 이때, 입력부(120)는 다음 [수학식 1]을 이용하여 $(t-p)$ 일로부터 t 일까지의 주식 수익률 즉, 지난 p 일간의 일별 주식 수익률을 산출할 수 있다.

수학식 1

$$r_{t,p}^s = \frac{P_t^s - P_{t-p}^s}{P_{t-p}^s}$$

[0034]

여기서, P_t^s 는 t 일(day t)에서 s 종목에 대한 주식 종가이고, $s \in \{s_1, s_2, \dots, s_n\}$ 는 시가총액 내림차순으로 종목의 위치이고, n 은 수집된 종목의 개수이다.

[0036] 입력부(120)는 산출된 t 일에 대한 1일($p=1$)의 수익률 및 지난 p 일간의 일별 주식 수익률을 하나의 특징 벡터 $R_t^s = \{r_{t-19,1}^s, r_{t-18,1}^s, \dots, r_{t,1}^s\}$ 로 생성한다. 다시 말해서, 입력부(120)는 가장 최근 20일간(즉, $u=20$)의 주식 수익 정보를 생성한다. 본 실시 예에서는 가장 최근 20일간의 주식 수익 정보를 생성하는 것을 예로 들어 설명하고 있으나, 이에 한정되지 않고 임의의 기간 데이터를 이용할 수 있다. 또한, 본 실시 예에서는 입력부(120)가 데이터세트의 전처리를 수행하는 것을 일 예로 설명하고 있으나, 이에 한정되지 않고 처리부(150)에 의해 데이터세트가 전처리될 수도 있다.

[0037] 입력부(120)는 사용자의 조작에 따라 입력 데이터(예: 샘플 데이터)를 발생시킬 수도 있다. 입력부(120)는 키보드(keyboard), 터치 패드(touch pad), 터치스크린(touch screen) 및/또는 QR(Quick Response) 코드 스캐너(code scanner) 등으로 구성될 수 있다.

[0038] 저장부(130)는 처리부(150)의 동작을 위한 프로그램을 저장할 수 있고, 처리부(150)의 입력 데이터 및/또는 출력 데이터들을 임시 저장할 수도 있다. 또한, 저장부(130)는 샘플 데이터, 훈련 데이터(training dataset), 다수의 다른 종목의 주식 데이터(source data) 및/또는 관심종목의 주식 데이터(target data) 등을 데이터베이스(database) 형태로 저장할 수 있다.

[0039] 저장부(130)는 전이학습 과정에서 발생하는 데이터, 베이스 모델(base model)(151)로부터 추출되는 파라미터 및/또는 예측 모델(prediction model)(152)에 의해 예측된 결과 등을 저장할 수 있다. 저장부(130)는 플래시 메모리(flash memory), 하드디스크(hard disk), SD 카드(Secure Digital Card), 램(Random Access Memory, RAM), SRAM(Static Random Access Memory), 롬(Read Only Memory, ROM), PROM(Programmable ReadOnly Memory), EEPROM(Electrically Erasable and Programmable ROM), EPROM(Erasable and Programmable ROM), 레지스터, 착탈형 디스크 및 웹 스토리지(web storage) 등의 저장매체 중 적어도 하나 이상의 저장매체(기록매체)를 포함할 수 있다.

[0040] 출력부(140)는 처리부(150)의 지시에 따라 주가 변동 예측 결과를 출력하기 위한 것으로, 디스플레이 등을 포함

할 수 있다. 디스플레이는 액정 디스플레이(liquid crystal display, LCD), 박막 트랜지스터 액정 디스플레이(thin film transistor-liquid crystal display, TFT LCD), 유기 발광 다이오드(organic light-emitting diode, OLED) 디스플레이, 플렉시블 디스플레이(flexible display), 3차원 디스플레이(3D display), 투명디스플레이, 헤드업 디스플레이(head-up display, HUD), 터치스크린 및 클러스터(cluster) 중에서 적어도 하나 이상을 포함할 수 있다.

[0041] 처리부(150)는 주가 변동 예측 장치(100)의 전반적인 동작을 제어한다. 처리부(150)는 ASIC(Application Specific Integrated Circuit), DSP(Digital Signal Processor), PLD(Programmable Logic Devices), FPGAs(Field Programmable Gate Arrays), CPU(Central Processing unit), 마이크로 컨트롤러(microcontrollers) 및 마이크로 프로세서(microprocessors) 중 적어도 하나를 포함할 수 있다.

[0042] 처리부(150)는 전이학습에 이용할 베이스 모델(151)과 관심종목의 주가 변동을 예측하기 위한 예측 모델(152)을 포함한다. 처리부(150)는 전이학습 기법을 기반으로 베이스 모델(151)로부터 예측 모델(152)을 구축한다. 여기서, 전이학습 기법으로는 라벨된 소스 데이터(labeled source data)와 라벨된 타겟 데이터(labeled target data) 간의 관계에 따라 학습하는 귀납 전이 학습(inductive transfer learning)이 사용된다.

[0043] 처리부(150)는 DTRSI(Deep Transfer with Related Stock Information) 프레임워크를 기반으로 관심종목 즉, 타겟종목의 일 단위 주가 변동을 예측할 수 있다. DTRSI 프레임워크는 도 3에 도시된 바와 같이 베이스 모델(151)을 구축하는 과정(210)과 예측 모델(152)을 훈련하는 과정(220)을 포함한다. 베이스 모델(151)을 구축하는 과정(210)은 소스 데이터 준비 단계(211)와 베이스 모델 훈련 단계(212)를 포함한다. 예측 모델(152)을 훈련하는 과정(220)은 타겟 데이터 준비 단계(221)와 예측 모델 훈련 단계(222)를 포함한다.

[0044] 이하에서는 베이스 모델(151)을 구축하는 과정(210)과 예측 모델(152)을 훈련하는 과정(220)을 보다 구체적으로 설명한다.

[0045] 먼저, 베이스 모델(151)을 구축하는 과정(210)을 설명한다. 처리부(150)는 다수의 가용 종목의 주식 데이터(available stocks)로부터 소스 데이터를 획득한다. 다시 말해서, 처리부(150)는 다수의 다른 종목(예: 50개 다른 종목)의 주식 데이터를 사용하여 소스 데이터를 준비한다. 소스 데이터는 베이스 모델(151)의 입력으로 사용된다.

[0046] 베이스 모델(151)의 입력으로 3가지 타입의 입력 특성을 가지는 3가지 소스 데이터를 사용한다. 제1타입 입력 특성은 한 종목의 주식 수익 정보(주식수익률)에 의해 결정된다. 주식 수익 정보는 각 종목에 대한 기정해진 기간 u 의 일 단위 주식수익률을 포함한다. 여기서, 기정해진 기간 u 은 사전에 사용자에게 의해 설정될 수 있다. 예를 들어, 처리부(150)는 시가총액 기준 상위 50개 종목 각각에 대하여 20일간의 일 단위 주식 수익률(일별 주식 수익률)을 이용하여 제1타입 입력 특성의 소스 데이터(훈련 데이터)를 구성한다. 제2타입 입력 특성은 한 종목과 주가 인덱스(이하, 인덱스)의 주식 수익 정보의 조합으로 결정(구성)된다. 처리부(150)는 한 종목 및 주가 인덱스(예: KOSPI200 또는 S&P500)에 대한 기정해진 기간 u 의 일 단위 주식 수익률의 조합으로 제2타입 입력 특성의 소스 데이터를 생성한다. 제3타입 입력 특성은 관심종목, 인덱스 및 관심종목과 관련된 다수의 다른 종목의 주식 수익 정보의 조합으로 구성된다. 처리부(150)는 관심종목, 인덱스 및 관심종목과 관련된 다수의 다른 종목의 기정해진 기간 u 의 일 단위 주식 수익률의 조합으로 제3타입 입력 특성의 소스 데이터를 생성한다. 처리부(150)는 전체 종목 중 관심종목과 코사인 유사도(cosine similarity, CS)가 높은 상위 복수 개(예: 50개)의 종목, 관심종목과 업종 유사도(similar field, SF)가 높은 상위 복수 개의 종목, 또는 전체 종목 중 주가총액 기준 상위 복수 개의 종목(highest market capitalization, HMC)을 관심종목과 관련된 다른 종목으로 선택할 수 있다.

[0047] x_t^s 는 $(t+1)$ 일에 s 종목의 증가 변동(주가 변동)을 예측하는데 사용되는 입력이다. R_t^s , R_t^{index} , 및 R_t^{rs} 는 t 일에 s 종목, 인덱스 및 s 종목과 관련된 다른 종목들 $rs \in \{rs_1, rs_2, \dots, rs_m\}$ 각각의 특징 벡터 즉, $R_t^s = \{r_{t-19,1}^s, r_{t-18,1}^s, \dots, r_{t,1}^s\}$, $R_t^{index} = \{r_{t-19,1}^{index}, r_{t-18,1}^{index}, \dots, r_{t,1}^{index}\}$, 및 $R_t^{rs} = \{r_{t-19,1}^{rs}, r_{t-18,1}^{rs}, \dots, r_{t,1}^{rs}\}$ 이다. 여기서, m 은 s 종목과 관련된 종목수이다. 베이스 모델에서 제1타입 입력 특성, 제2타입 입력 특성 및 제3타입 입력 특성은 각각 $x_t^s = \{R_t^s\}$, $x_t^s = \{R_t^s, R_t^{index}\}$ 및 $x_t^s = \{R_t^s, R_t^{index}, R_t^{rs_1}, R_t^{rs_2}, \dots, R_t^{rs_m}\}$ 이다.

[0048] 처리부(150)는 3가지 타입의 입력 특성의 소스 데이터에 해당하는 $n' \times 20 \times k$ 크기(dimensions)를 가지는 3차원

텐서(tensors) (X_1 , X_2 , 및 X_3)를 생성한다. 여기서, n' 및 n_{samp} 은 소스 데이터의 훈련 샘플 개수 및 각 종목 s 의 훈련 샘플 개수이고, k 는 입력 특성의 수이다. 3차원 텐서 X_1 , X_2 , 및 X_3 는 도 5a 내지 도 5c에 도시된 바와 같이 하나의 훈련 샘플에 대응하는 $20 \times k$ 크기의 각 매트릭스(matrix)로 나타낼 수 있다.

[0049] 특히, 제1타입 입력 특성과 제2타입 입력 특성에서, 시가총액 기준 상위 복수 개 종목(예: 50개 종목)의 주식 데이터 $s \in \{s_1, s_2, \dots, s_{50}\}$ 를 선택하여 하나의 데이터셋으로 스택한다. 이에, 소스 데이터의 훈련 샘플 개수 ($n' = n_{\text{samp}} \times 50$)가 증가하게 된다. 한 종목의 훈련 샘플을 사용하는 대신 복수 개 종목의 훈련 샘플을 사용하여 모델 파라미터(model parameters)를 최적화할 수 있다. 제3타입 입력 특성은 각 관심종목 주식과 가장 유사한 상위 복수 개의 종목 주식을 선택하고, 선택된 종목 주식의 특징은 베이스 모델 성능을 개선할 수 있다. 복수 개 관련 종목 주식은 랜덤하게 복수의 그룹으로 분류하고, 각 그룹은 m 개의 종목을 포함하게 한다. 예를 들어, 전체 주식종목 중 관심종목과 코사인 유사도가 높은 50개 종목을 선택한 상태에서 m 이 5인 경우, 10개의 그룹이 생성될 수 있다. 각 그룹은 도 5c에 도시된 바와 같이 제3타입 입력 특성을 설계하는데 사용된다. 도 5c에 도시된 바와 같이 각 관심종목에 대해 10개 그룹으로부터 설계된 입력은 $x_t^{COI} = \{R_t^{COI}, R_t^{index}, R_t^{rs_1}, R_t^{rs_2}, \dots, R_t^{rs_m}\}$, $x_t^{COI} = \{R_t^{COI}, R_t^{index}, R_t^{rs_{m+1}}, R_t^{rs_{m+2}}, \dots, R_t^{rs_{2m}}\}$, ..., $x_t^{COI} = \{R_t^{COI}, R_t^{index}, R_t^{rs_{46}}, R_t^{rs_{47}}, \dots, R_t^{rs_{50}}\}$ 이다. 제3타입 입력 특성에서, x_t^{COI} 와 함께 n_{samp} 훈련 샘플은 관심종목 주가 변동을 예측하기 위해 사용된다. 예를 들어, $m = \{3, 5,$

$7\}$ 인 경우, 훈련 샘플 개수 n' 는 $n' = n_{\text{samp}} \times \frac{50}{m}$ 이고, 입력 특성의 수 k 는 $2+m$ 이다. 이와 같이, 다른 그룹들을 고려하므로 베이스 모델을 훈련시키는데 있어 2가지 이점이 있다. 첫번째는 훈련 샘플의 개수가 $50/m$ 으로 증가한다. 두번째는 보다 적합하고 효과적인 특성의 모델 입력을 구성하는데 사용될 수 있다.

[0050] 처리부(150)는 준비된 소스 데이터를 이용하여 베이스 모델(151)을 훈련한다. 베이스 모델(151)을 훈련하기 위해, 다수의 다른 레이어(different number of layers), 은닉 유닛(hidden units) 및 활성화 기능을 가진 다양한 네트워크 구조 중 LSTM(Long Short-Term Memory) 모델을 사용한다.

[0051] LSTM 모델의 토폴로지는 다음과 같다.

[0052] - 입력 레이어는 입력 특성 수와 20개의 타임스텝(timesteps)과 동일한 유닛들을 가진다.

[0053] - 각 입력 특성에 대해 2개의 LSTM 레이어가 있으며, 각 LSTM 레이어의 출력 벡터 차원이 16이다.

[0054] - 2개의 LSTM 레이어 위에 2개의 FC(Fully Connected) 레이어가 연결되고, 각 레이어는 16개의 유닛(뉴런)을 가지며 각 은닉 레이어의 출력에 드롭아웃 정규화를 적용한다. 유지 확률 값은 0.5로 설정된다.

[0055] - 출력 레이어는 하나의 시그모이드(sigmoid) 활성 유닛으로 구성되어 있다.

[0056] 따라서, 가중치 및 바이어스(bias)는 정규 분포로 초기화된다. RMSprop 최적화에서 학습 속도, 배치 크기(batch size) 및 최대 학습 에포크(epoch)는 각각 0.001, 512 및 5000으로 선택된다.

[0057] 다음, 처리부(150)는 예측 모델(152)을 훈련(학습)한다. 예측 모델(152)은 베이스 모델(151)과 같은 구조를 가지며 모델 파라미터의 초기값은 베이스 모델(151)로부터 추출된 모델 파라미터 값으로 설정된다. 예측 모델(152)은 관심종목의 주식 데이터만을 이용하여 훈련을 한다. 처리부(150)는 관심종목의 주식 데이터로부터 타겟 데이터를 획득(준비)한다. 타겟 데이터는 예측 모델(152)의 입력으로 사용된다.

[0058] 예측 모델(152)의 입력특성은 3가지 타입의 입력 특성 제1타입 입력 특성, 제2타입 입력 특성 및 제3타입 입력 특성으로 분류한다. 제1타입 입력 특성은 관심종목의 주식 수익 정보만 이용하여 결정된다. 주식 수익 정보는 각 종목에 대한 지정해진 기간 u 의 일 단위 주식수익률을 포함한다. 여기서, 지정해진 기간 u 은 사전에 사용자에 의해 설정될 수 있다. 예를 들어, 처리부(150)는 관심종목에 대하여 정해진 기간 u (예컨대, $u=20$)의 일 단위 주식수익률을 이용하여 제1타입 입력 특성의 타겟 데이터(훈련 데이터)를 생성한다. 제2타입 입력 특성은 관심종목 및 인덱스의 주식 수익 정보를 이용하여 구성한다. 처리부(150)는 관심종목 및 인덱스(예: KOSPI200 또는 S&P500)에 대한 지정해진 기간 u 의 일 단위 주식 수익률의 조합으로 제2타입 입력 특성의 타겟 데이터를 생성한다. 제3타입 입력 특성은 관심종목, 인덱스 및 관심종목과 관련된 다수의 다른 종목의 주식 수익 정보를 이용하여 결정한다. 처리부(150)는 관심종목, 인덱스 및 전체 종목 중 관심종목과 관련된 m 개 다른 종목의 지정해진

기간 u 의 일 단위 주식 수익률의 조합으로 제3타입 입력 특성의 타겟 데이터를 생성한다. 이때, 처리부(150)는 관심종목과 코사인 유사도(cosine similarity)가 높은 상위 m 개 종목, 관심종목과 업종 유사도가 높은 상위 m 개 종목, 및/또는 전체 종목 중 주가총액 기준 상위 m 개 종목을 관심종목과 관련된 종목(관련성이 높은 종목)으로 선별한다.

[0059]

6가지 다른 입력 특성 (SI_t^{COI} , SI_t^{CI} , SI_t^{CH+CS} , SI_t^{CH+SF} , SI_t^{CH+HMC} , 및 SI_t^{CH+RND})은 $(t+1)$ 일의 관심종목 주가 변동을 예측하기 위해 예측 모델(152)의 입력으로 공급된다. 여기서, SI_t^{COI} 는 t 일에 관심종목의 일 단위 주식수익률을 이용한 입력, SI_t^{CI} 는 t 일에 관심종목 및 인덱스의 일 단위 주식수익률을 사용한 입력, SI_t^{CH+CS} 는 t 일에 관심종목, 인덱스 및 관심종목과 최고 코사인 유사성을 가지는 종목의 일 단위 주식수익률을 이용한 입력, SI_t^{CH+SF} 는 t 일에 관심종목, 인덱스 및 관심종목과 유사한 업종에 속하는 종목의 일 단위 주식수익률을 이용한 입력, SI_t^{CH+HMC} 는 t 일에 관심종목, 인덱스 및 최고 시가총액을 가지는 종목의 일 단위 주식수익률을 이용한 입력, SI_t^{CH+RND} 는 t 일에 관심종목, 인덱스 및 임의 선택된 종목의 일 단위 주식수익률을 이용한 입력이다. 각 입력 벡터는 $SI_t^{COI} = \{R_t^{COI}\}$, $SI_t^{CI} = \{R_t^{COI}, R_t^{index}\}$, $SI_t^{CH+CS} = \{R_t^{COI}, R_t^{index}, R_t^{CS_1}, \dots, R_t^{CS_m}\}$, $SI_t^{CH+SF} = \{R_t^{COI}, R_t^{index}, R_t^{SF_1}, \dots, R_t^{SF_m}\}$, $SI_t^{CH+HMC} = \{R_t^{COI}, R_t^{index}, R_t^{HMC_1}, \dots, R_t^{HMC_m}\}$ 및 $SI_t^{CH+RND} = \{R_t^{COI}, R_t^{index}, R_t^{RND_1}, \dots, R_t^{RND_m}\}$ 이다. 여기서, R_t^{COI} 는 t 일에 관심종목 주식의 특징 벡터, R_t^{index} 는 t 일에 인덱스의 특징 벡터, $R_t^{CS_1}, \dots, R_t^{CS_m}$ 는 t 일에 관심종목과 코사인 유사도가 높은 상위 m 개 주식종목의 특징 벡터, $R_t^{SF_1}, \dots, R_t^{SF_m}$ 는 t 일에 관심종목과의 업종 유사도 기준 상위 m 개 종목의 특징 벡터, $R_t^{HMC_1}, \dots, R_t^{HMC_m}$ 는 t 일에 전체 종목 중 주가총액 기준 상위 m 개 종목의 특징 벡터, $R_t^{RND_1}, \dots, R_t^{RND_m}$ 는 t 일에 임의 선택된 종목의 특징 벡터이다.

[0060]

예측 모델 훈련에서, 6차원 텐서 SI_1 , SI_2 , SI_{3a} , SI_{3b} , SI_{3c} 및 SI_{3d} 는 6가지 입력 벡터 SI_t^{COI} , SI_t^{CI} , SI_t^{CH+CS} , SI_t^{CH+SF} , SI_t^{CH+HMC} , 및 SI_t^{CH+RND} 에 대한 훈련 데이터로 구성된다. 6가지 훈련 데이터를 사용하여 6가지 예측 모델을 훈련한다. 예측 모델(152)의 모든 훈련 데이터는 다음 [표 1]과 같이 베이스 모델(151)에 공급될 수 있다.

표 1

[0061]

베이스 모델	예측 모델
X_1	SI_1
X_2	SI_2
X_3	SI_{3a} , SI_{3b} , SI_{3c} , SI_{3d}

[0062]

학습 속도 및 최대 에포크는 0.0005 및 500이다. 학습 속도는 기훈련된 파라미터가 너무 빨리 왜곡되는 것을 방지하기 위해 작은 값이 선택된다. 처리부(150)는 준비된 타겟 데이터 즉, 소량의 관심종목 주식 데이터를 사용하여 미세 조정(fine-tuned)한다. 즉, 처리부(150)는 타겟 데이터를 사용하여 예측 모델(152)을 훈련(재학습)한다.

[0063]

처리부(150)는 훈련된 예측 모델(152)에 관심종목의 과거 주식 데이터를 입력하여 관심종목의 주가 변동을 예측한다. 예를 들어, 처리부(150)는 t 로부터 과거 20일의 주식 데이터를 훈련된 예측 모델(152)에 입력하여 $t+1$ 일

에 관심종목의 주가 상승 여부를 결정한다. 처리부(150)는 $t+1$ 일의 관심종목 주가가 t 일의 관심종목 주가를 초과하는 경우 주가상승으로 판정하여 '1'을 출력한다. 한편, 처리부(150)는 $t+1$ 일의 관심종목 주가가 t 일의 관심종목 주가 이하인 경우 주가미상승으로 판정하여 '0'을 출력한다.

[0064] 도 6은 본 발명의 일 실시 예에 따른 주가 이동 예측 방법을 도시한 흐름도이다.

[0065] 처리부(150)는 전이학습에 이용할 베이스 모델(151)을 생성(구축)한다(S110). 처리부(150)는 임의로 선택된 종목의 주식 데이터를 이용하여 베이스 모델(151)의 모델 파라미터를 초기화할 수 있다.

[0066] 처리부(150)는 베이스 모델(151)을 다수의 다른 종목의 주식 데이터를 이용하여 훈련한다(S120). 처리부(150)는 다수의 다른 종목의 주식 데이터로부터 소스 데이터를 획득하고, 획득된 소스 데이터를 사용하여 베이스 모델(151)을 훈련한다. 소스 데이터는 관심종목의 주식 수익 정보뿐만 아니라 인덱스 및 관심종목과 관련된 다른 종목의 주식 수익 정보를 이용하여 구성된다.

[0067] 처리부(150)는 훈련된 베이스 모델(151)의 모델 파라미터를 기반으로 예측 모델(152)을 생성(구축)한다(S130). 처리부(150)는 훈련된 베이스 모델(151)로부터 모델 파라미터를 추출하고, 추출된 모델 파라미터를 예측 모델(152)의 초기 모델 파라미터로 설정한다. 다시 말해서, 처리부(150)는 훈련된 베이스 모델(151)의 모델 파라미터로 예측 모델(152)을 초기화한다.

[0068] 처리부(150)는 예측 모델(152)을 관심종목의 주식 데이터를 이용하여 훈련한다(S140). 처리부(150)는 관심종목의 주식 데이터로부터 타겟 데이터를 획득하고 획득된 타겟 데이터를 사용하여 예측 모델(152)을 훈련한다. 타겟 데이터는 관심종목의 주식 수익 정보뿐만 아니라 인덱스 및 관심종목과 관련된 다른 종목의 주식 수익 정보를 이용하여 구성된다.

[0069] 처리부(150)는 훈련된 예측 모델(152)을 이용하여 관심종목의 주가 변동을 예측한다(S150). 처리부(150)는 타겟 일자 $t+1$ 를 기준으로 정해진 기간($t-p$ 일부터 t 일 까지)의 과거 관심종목 주식 데이터를 훈련된 예측 모델(152)에 적용하여 타겟 일자 $t+1$ 의 관심종목 주가 상승 여부를 결정한다. 처리부(150)는 t 일에 비하여 $t+1$ 일의 관심종목 주가가 큰 경우 '1'을 출력하고, t 일에 비하여 $t+1$ 일의 관심종목 주가가 작거나 같은 경우 '0'을 출력한다.

[0070] 도 7 내지 도 9는 본 발명의 일 실시 예에 따른 예측 모델의 성능 실험 결과를 도시한 도면이다.

[0071] 도 7을 참조하면, 관심종목, 인덱스, 및 관심종목과 유사한 업종의 다른 종목의 일 단위 주식 수익 정보를 사용한 제3타입 입력 특성을 적용한 경우, 예측 모델의 예측 정확도가 가장 높다.

[0072] 도 8을 참조하면, 본 명세서에서 제안된 예측 모델의 성능을 기존의 다른 예측 모델들의 성능과 비교한 바, 제안된 예측 모델의 성능이 기존의 다른 예측 모델들보다 보다 우수함을 확인할 수 있다.

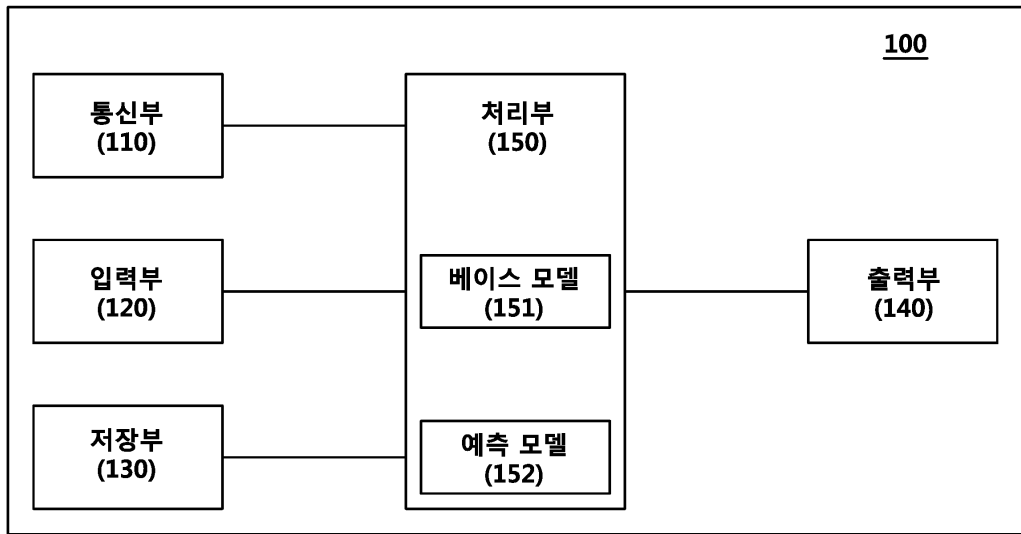
[0073] 도 9를 참조하면, 관심종목과 관련된 종목 수 m 은 예측 모델의 하이퍼파라미터(hyperparameter)로 고려되며 종목별로 신중하게 선택되어야 함을 알 수 있다.

[0074] 이상에서, 본 발명의 실시예를 구성하는 모든 구성 요소들이 하나로 결합되거나 결합되어 동작하는 것으로 설명되었다고 해서, 본 발명이 반드시 이러한 실시예에 한정되는 것은 아니다. 즉, 본 발명의 목적 범위 안에서라면, 그 모든 구성 요소들이 하나 이상으로 선택적으로 결합하여 동작할 수도 있다. 또한, 그 모든 구성 요소들이 각각 하나의 독립적인 하드웨어로 구현될 수 있지만, 각 구성 요소들의 그 일부 또는 전부가 선택적으로 조합되어 하나 또는 복수 개의 하드웨어에서 조합된 일부 또는 전부의 기능을 수행하는 프로그램 모듈을 갖는 컴퓨터 프로그램으로서 구현될 수도 있다. 그 컴퓨터 프로그램을 구성하는 코드들 및 코드 세그먼트들은 본 발명의 기술 분야의 당업자에 의해 용이하게 추론될 수 있을 것이다. 이러한 컴퓨터 프로그램은 컴퓨터가 읽을 수 있는 저장매체(Computer Readable Media)에 저장되어 컴퓨터에 의하여 읽혀지고 실행됨으로써, 본 발명의 실시예를 구현할 수 있다.

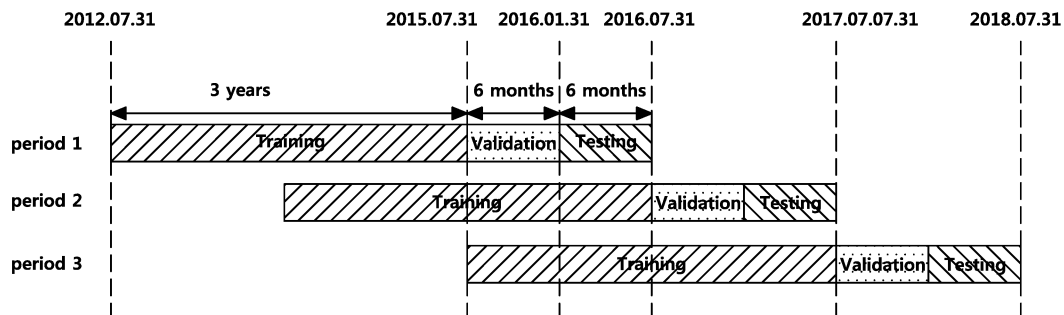
[0075]

도면

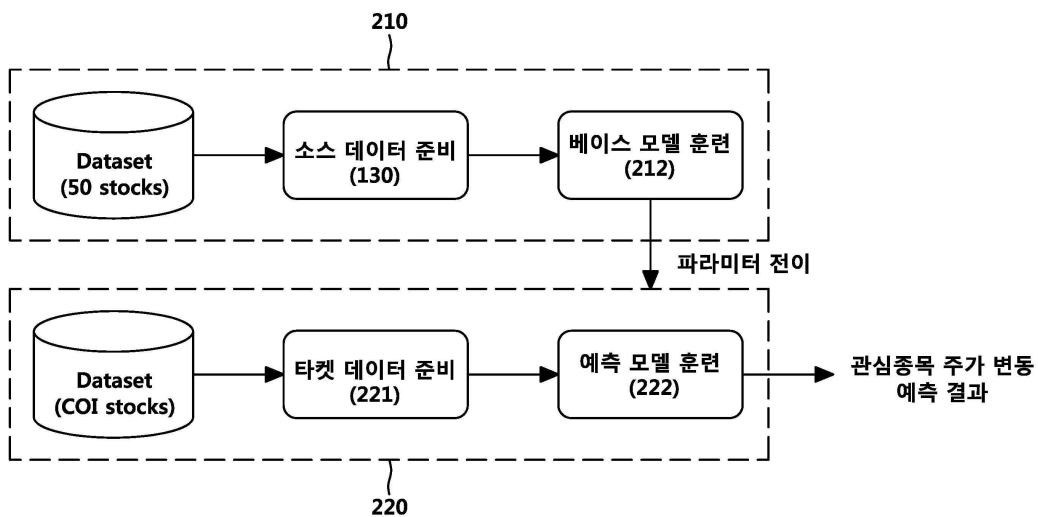
도면1



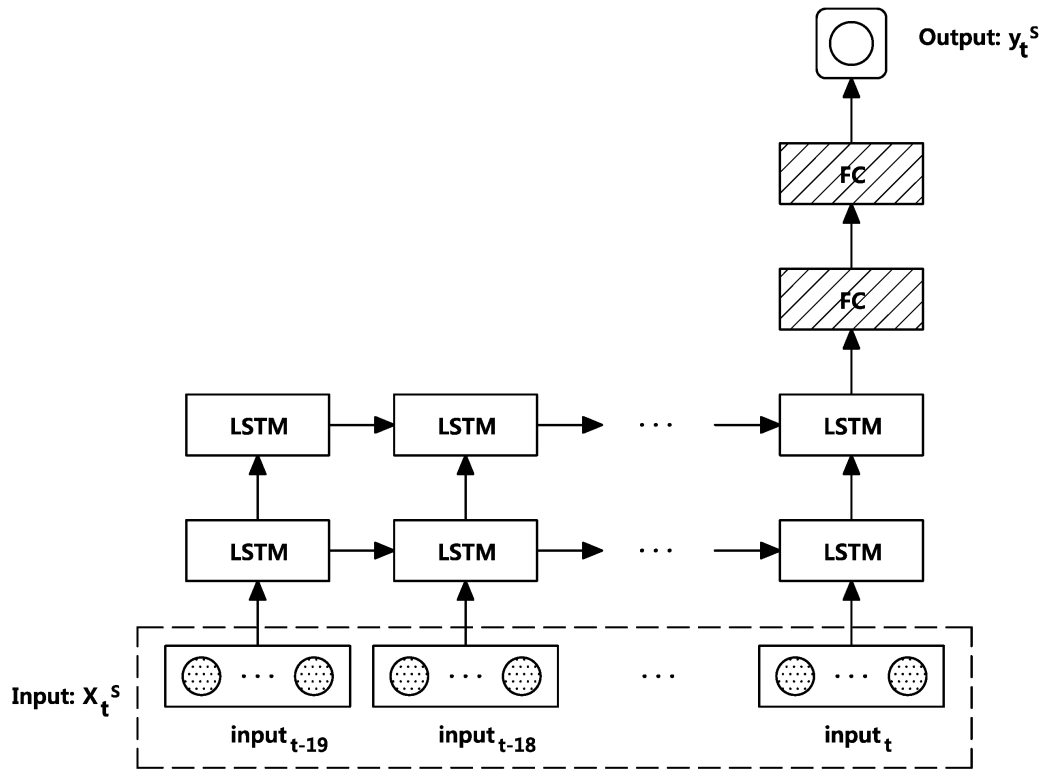
도면2



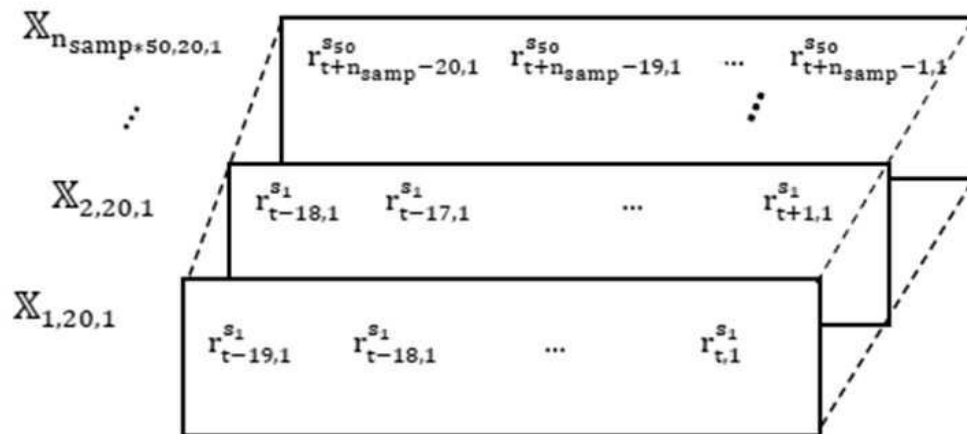
도면3



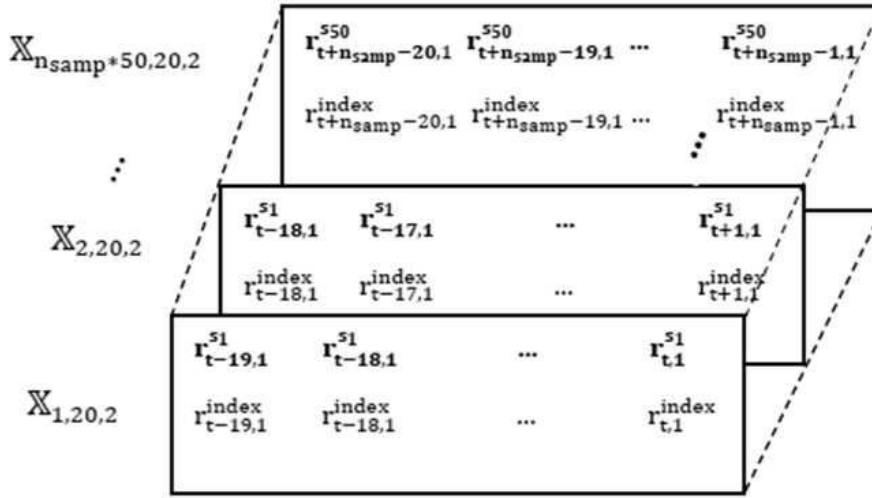
도면4



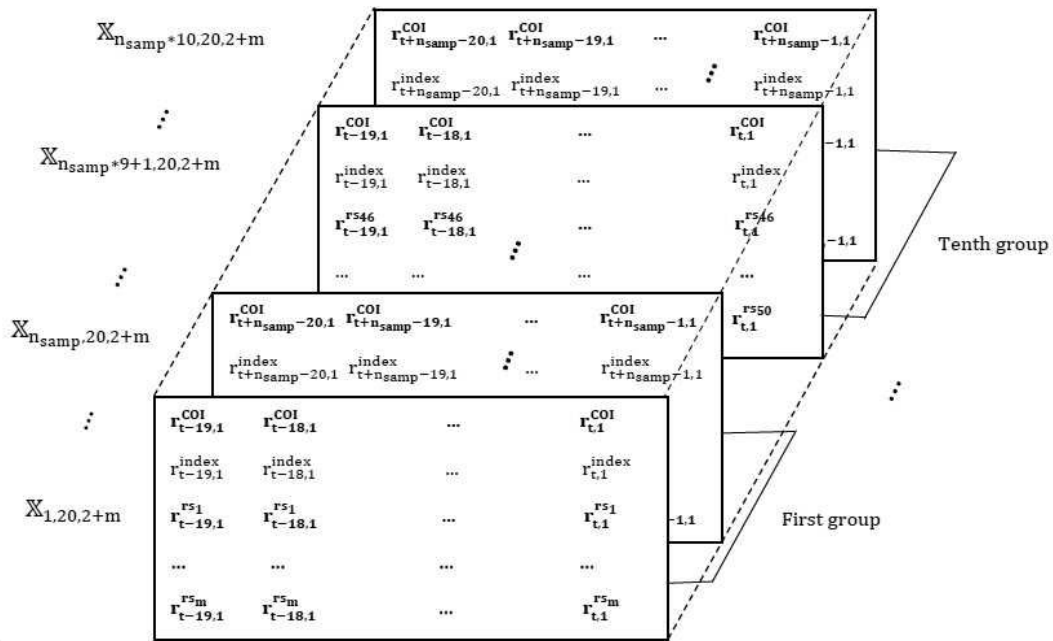
도면5a



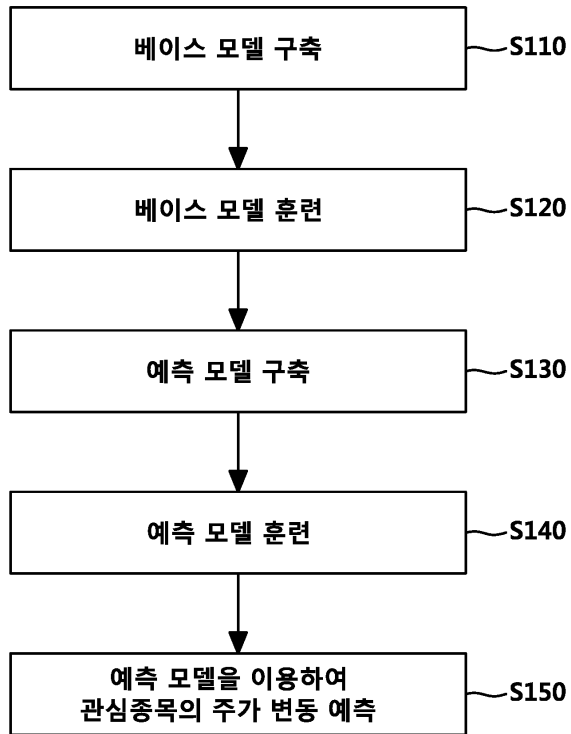
도면5b



도면5c



도면6



도면7

Stock Name	COI	CI	CI+CS	CI+SF	CI+HMC	CI+RND
Samsung Electronics Co	49.84	53.36	57.51	59.74	58.15	55.91
SK Hynix Inc	53.02	54.63	58.46	60.38	58.78	55.91
Celltrion	55.77	55.28	58.17	61.02	56.87	55.27
Hyundai Motor Co.	58.16	60.07	62.94	64.54	61.98	58.79
LG Chem	51.43	56.23	59.75	60.05	60.00	56.23
Boeing Company	50.93	55.56	59.57	60.19	57.71	55.86
Apple Inc.	55.87	60.49	57.41	59.88	59.88	54.63
Amazon.com Inc.	55.56	57.72	60.19	62.65	61.11	58.33
Johnson & Johnson	53.08	54.94	57.71	59.26	57.10	55.25
JPMorgan Chase & Co.	53.71	55.25	58.02	58.64	54.01	52.47
Average	53.74	56.35	58.91	60.64	58.59	55.87

도면8

Stock Name	Our Approach	Baselines			
		SVM	RF	KNN	Fisher and Krauss [20]
Samsung Electronics Co.	59.74	52.07	48.25	50.80	54.95
SK Hynix Inc	60.38	55.28	51.43	51.44	55.27
Celltrion	61.02	56.55	53.67	54.97	59.74
Hyundai Motor Co.	64.54	54.96	57.51	58.47	57.83
LG Chem	60.05	50.47	49.98	50.15	58.78
Boeing Company	60.19	48.78	52.78	50.00	58.64
Apple Inc.	60.49	43.83	46.91	50.62	57.72
Amazon.com Inc.	62.65	54.01	59.95	50.93	58.02
Johnson & Johnson	59.26	48.77	50.62	50.31	56.48
JPMorgan Chase & Co.	58.64	50.94	52.16	50.00	56.17
Average	60.70	51.57	51.49	51.77	57.36

도면9

Stock Name	$m = 3$	$m = 5$	$m = 7$
Samsung Electronics Co.	59.42	58.14	52.39
SK Hynix Inc	60.06	56.56	53.66
Celltrion	54.94	60.70	55.59
Hyundai Motor Co.	55.26	57.83	64.22
LG Chem	53.33	56.21	56.87
Boeing Company	60.19	57.41	56.48
Apple Inc.	58.96	53.09	52.78
Amazon.com Inc.	62.04	61.11	57.41
Johnson & Johnson	57.79	58.33	53.70
JPMorgan Chase & Co.	57.72	56.79	56.17
Average	57.97	57.62	55.93