



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년11월25일
(11) 등록번호 10-2734026
(24) 등록일자 2024년11월20일

- (51) 국제특허분류(Int. Cl.)
G06Q 40/04 (2023.01) G06N 20/00 (2019.01)
G06Q 10/04 (2023.01) G06Q 30/06 (2023.01)
- (52) CPC특허분류
G06Q 40/04 (2023.01)
G06N 20/00 (2021.08)
- (21) 출원번호 10-2021-0192621
- (22) 출원일자 2021년12월30일
심사청구일자 2021년12월30일
- (65) 공개번호 10-2023-0102461
- (43) 공개일자 2023년07월07일
- (56) 선행기술조사문헌
KR101983752 B1
KR1020200039037 A
KR1020210063765 A
KR1020210015582 A

- (73) 특허권자
리벨리온 주식회사
경기도 성남시 분당구 정자일로 239, 102동 801호
(정자동, 아이파크분당1)
- (72) 발명자
이병재
경기도 성남시 분당구 정자일로 239, 102동 908호
- (74) 대리인
김한솔, 김세환, 김준식, 안제성

전체 청구항 수 : 총 17 항

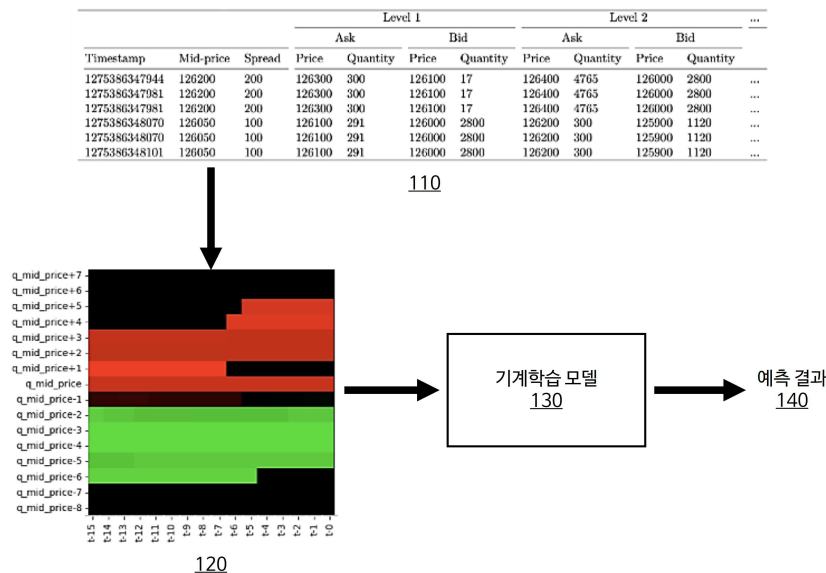
심사관 : 석상문

(54) 발명의 명칭 **오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법 및 시스템**

(57) 요약

본 개시는 적어도 하나의 프로세서에 의해 수행되는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법에 관한 것이다. 방법은, (a) 미리 결정된 기간 이전의 과거 시점으로부터 현재 시점까지의 시간에 따른, 제1 증권 거래소에서 거래되는 종목에 대한 복수의 매수 요청 가격에 대한 데이터 아이템 및 복수의 매도 (뒷면에 계속)

대표도 - 도1



요청 가격에 대한 데이터 아이템을 포함한 오더북 데이터를 획득하는 단계, (b) 복수의 매수 요청 가격 중에서 가장 높은 매수 요청 가격과 복수의 매도 요청 가격 중에서 가장 낮은 매도 요청 가격을 기초로 중간 가격을 산출하는 단계 및 (c) 일정한 시간 간격의 시간을 제1 축으로 하고, 틱 단위의 가격을 제2 축으로 하고, 복수의 매도 요청 가격의 각각의 수량에 대한 데이터 또는 복수의 매수 요청 가격의 각각의 수량에 대한 데이터를 제1 축의 시간과 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 대한 값으로 가진 하나 이상의 2D 데이터 아이템을 포함한 텐서 형태의 데이터를 생성하는 단계- 틱 단위의 가격은 산출된 중간 가격을 포함함 -를 포함한다.

(52) CPC특허분류

G06Q 10/04 (2023.01)

G06Q 30/0633 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711134476
과제번호	2021-0-00310-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	SW컴퓨팅산업원천기술개발(R&D, 정보화)
연구과제명	인공지능 학습/추론 효율성 향상을 위한 서버용 SW 프레임워크 개발
기 여 율	1/1
과제수행기관명	리벨리온 주식회사
연구기간	2021.04.01 ~ 2024.12.31

명세서

청구범위

청구항 1

적어도 하나의 프로세서에 의해 수행되는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법에 있어서,

(a) 미리 결정된 기간 이전의 과거 시점으로부터 현재 시점까지의 시간에 따른, 제1 증권 거래소에서 거래되는 종목에 대한 복수의 매수 요청 가격에 대한 데이터 아이템 및 복수의 매도 요청 가격에 대한 데이터 아이템을 포함한 오더북 데이터를 획득하는 단계;

(b) 상기 복수의 매수 요청 가격 중에서 가장 높은 매수 요청 가격과 상기 복수의 매도 요청 가격 중에서 가장 낮은 매도 요청 가격을 기초로 중간 가격을 산출하는 단계; 및

(c) 일정한 시간 간격의 시간을 제1 축으로 하고, 틱 단위의 가격을 제2 축으로 하고, 상기 복수의 매도 요청 가격의 각각의 수량에 대한 데이터 또는 상기 복수의 매수 요청 가격의 각각의 수량에 대한 데이터를 상기 제1 축의 시간과 상기 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 대한 값으로 가진 하나 이상의 2D 데이터 아이템을 포함한 텐서 형태의 데이터를 생성하는 단계- 상기 틱 단위의 가격은 상기 산출된 중간 가격을 포함함 -

를 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법.

청구항 2

제1항에 있어서,

상기 텐서 형태의 데이터는, 복수의 채널의 각각에 대응하는 2D 데이터를 포함하고,

상기 복수의 채널은 제1 채널 및 제2 채널을 포함하고,

상기 텐서 형태의 데이터는,

상기 제1 채널에 대응하는, 상기 시간에 따른 상기 제1 증권 거래소에서 거래되는 종목의 매수 요청 가격에 대한 2D 데이터; 및

상기 제2 채널에 대응하는, 상기 시간에 따른 상기 제1 증권 거래소에서 거래되는 종목의 매도 요청 가격에 대한 2D 데이터

를 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법.

청구항 3

제2항에 있어서,

상기 텐서 형태의 데이터는 제3 채널 및 제4 채널을 더 포함하고,

상기 텐서 형태의 데이터는,

상기 제3 채널에 대응하는, 상기 시간에 따른 제2 증권 거래소에서 거래되는 상기 종목의 매수 요청 가격에 대한 2D 데이터; 및

상기 제4 채널에 대응하는, 상기 시간에 따른 상기 제2 증권 거래소에서 거래되는 상기 종목의 매도 요청 가격에 대한 2D 데이터

를 더 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법.

청구항 4

제2항에 있어서,

상기 텐서 형태의 데이터는 제5 채널 및 제6 채널을 더 포함하고,

상기 텐서 형태의 데이터는,

상기 하나 이상의 2D 데이터의 각각에서의 상기 제1 축에 대한 좌표를 나타내는 값을 포함한 2D 데이터; 및

상기 하나 이상의 2D 데이터의 각각에서의 상기 제2 축에 대한 좌표를 나타내는 값을 포함한 2D 데이터

를 더 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법.

청구항 5

제1항에 있어서,

상기 (c) 단계는,

상기 복수의 매도 요청 가격의 각각의 수량 및 상기 복수의 매수 요청 가격의 각각의 수량을 정규화하는 단계; 및

상기 정규화된 수량을 상기 제1 축의 시간과 상기 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 연관시키는 단계

를 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법.

청구항 6

제1항에 있어서,

상기 현재 시점이 변함에 따라, 상기 변화된 현재 시점으로부터 상기 미리 결정된 기간 이전의 과거 시점까지의 오더북 데이터에 대해 상기 (a) 단계 내지 상기 (c) 단계를 반복적으로 수행하는 단계

를 더 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법.

청구항 7

제6항에 있어서,

상기 생성된 텐서 형태의 데이터를 복수의 레이어를 포함한 기계학습 모델에 입력하여 예측 결과를 출력하는 단계

를 더 포함하고,

상기 예측 결과는, 상기 현재 시점으로부터 미리 결정된 기간 이후의 미래 시점에서의 상기 종목의 가격에 대한 예측 정보를 포함하는,

오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법.

청구항 8

제7항에 있어서,

상기 예측 정보는,

상기 미래 시점에서의 상기 종목의 가격이 현재 시점의 가격보다 높을 것을 나타내는 제1 클래스;
 상기 미래 시점에서의 상기 종목의 가격이 현재 시점의 가격과 동일할 것을 나타내는 제2 클래스; 및
 상기 미래 시점에서의 상기 종목의 가격이 현재 시점의 가격보다 낮을 것을 나타내는 제3 클래스
 를 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법.

청구항 9

제1항 내지 제8항 중 어느 한 항에 따른 방법을 컴퓨터에서 실행하기 위해 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램.

청구항 10

오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 시스템에 있어서,
 하나 이상의 인스트럭션을 저장하는 메모리; 및

상기 메모리의 하나 이상의 인스트럭션을 실행함으로써, (a) 미리 결정된 기간 이전의 과거 시점으로부터 현재 시점까지의 시간에 따른, 제1 증권 거래소에서 거래되는 종목에 대한 복수의 매수 요청 가격에 대한 데이터 아이템 및 복수의 매도 요청 가격에 대한 데이터 아이템을 포함한 오더북 데이터를 획득하고, (b) 상기 복수의 매수 요청 가격 중에서 가장 높은 매수 요청 가격과 상기 복수의 매도 요청 가격 중에서 가장 낮은 매도 요청 가격을 기초로 중간 가격을 산출하고, (c) 일정한 시간 간격의 시간을 제1 축으로 하고, 틱 단위의 가격을 제2 축으로 하고, 상기 복수의 매도 요청 가격의 각각의 수량에 대한 데이터 또는 상기 복수의 매수 요청 가격의 각각의 수량에 대한 데이터를 상기 제1 축의 시간과 상기 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 대한 값으로 가진 하나 이상의 2D 데이터 아이템을 포함한 텐서 형태의 데이터를 생성하도록 구성된 적어도 하나의 프로세서

를 포함하고,

상기 틱 단위의 가격은 상기 산출된 중간 가격을 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 시스템.

청구항 11

제10항에 있어서,

상기 텐서 형태의 데이터는, 복수의 채널의 각각에 대응하는 2D 데이터를 포함하고,

상기 복수의 채널은 제1 채널 및 제2 채널을 포함하고,

상기 텐서 형태의 데이터는,

상기 제1 채널에 대응하는, 상기 시간에 따른 상기 제1 증권 거래소에서 거래되는 종목의 매수 요청 가격에 대한 2D 데이터; 및

상기 제2 채널에 대응하는, 상기 시간에 따른 상기 제1 증권 거래소에서 거래되는 종목의 매도 요청에 대한 2D 데이터

를 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 시스템.

청구항 12

제11항에 있어서,

상기 텐서 형태의 데이터는 제3 채널 및 제4 채널을 더 포함하고,

상기 텐서 형태의 데이터는,

상기 제3 채널에 대응하는, 상기 시간에 따른 제2 증권 거래소에서 거래되는 상기 종목의 매수 요청 가격에 대한 2D 데이터; 및

상기 제4 채널에 대응하는, 상기 시간에 따른 상기 제2 증권 거래소에서 거래되는 상기 종목의 매도 요청 가격에 대한 2D 데이터

를 더 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 시스템.

청구항 13

제11항에 있어서,

상기 텐서 형태의 데이터는 제5 채널 및 제6 채널을 더 포함하고,

상기 텐서 형태의 데이터는,

상기 하나 이상의 2D 데이터의 각각에서의 상기 제1 축에 대한 좌표를 나타내는 값을 포함한 2D 데이터; 및

상기 하나 이상의 2D 데이터의 각각에서의 상기 제2 축에 대한 좌표를 나타내는 값을 포함한 2D 데이터

를 더 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 시스템.

청구항 14

제10항에 있어서,

상기 텐서 형태의 데이터를 생성하는 것은,

상기 복수의 매도 요청 가격의 각각의 수량 및 상기 복수의 매수 요청 가격의 각각의 수량을 정규화하고, 상기 정규화된 수량을 상기 제1 축의 시간과 상기 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 연관시키는 것을 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 시스템.

청구항 15

제10항에 있어서,

상기 적어도 하나의 프로세서는,

상기 현재 시점이 변함에 따라, 상기 변화된 현재 시점으로부터 상기 미리 결정된 기간 이전의 과거 시점까지의 오더북 데이터에 대해 상기 (a) 내지 상기 (c)를 반복적으로 수행하도록 더 구성된, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 시스템.

청구항 16

제15항에 있어서,

상기 적어도 하나의 프로세서는,

상기 생성된 텐서 형태의 데이터를 복수의 레이어를 포함한 기계학습 모델에 입력하여 예측 결과를 출력하도록 더 구성되고,

상기 예측 결과는, 상기 현재 시점으로부터 미리 결정된 기간 이후의 미래 시점에서의 상기 종목의 가격에 대한

예측 정보를 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 시스템.

청구항 17

제16항에 있어서,
 상기 예측 정보는,
 상기 미래 시점에서의 상기 종목의 가격이 현재 시점의 가격보다 높을 것을 나타내는 제1 클래스;
 상기 미래 시점에서의 상기 종목의 가격이 현재 시점의 가격과 동일할 것을 나타내는 제2 클래스; 및
 상기 미래 시점에서의 상기 종목의 가격이 현재 시점의 가격보다 낮을 것을 나타내는 제3 클래스
 를 포함하는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 시스템.

발명의 설명

기술 분야

[0001] 본 개시는 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법 및 시스템에 관한 것으로, 구체적으로, 시간을 제1 축으로 하고, 틱 단위의 가격을 제2 축으로 하고, 복수의 매도 요청 가격의 각각의 수량에 대한 데이터 또는 복수의 매수 요청 가격의 각각의 수량에 대한 데이터를 제1 축의 시간과 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 대한 값으로 가진 하나 이상의 2D 데이터 아이템을 포함한 텐서 형태의 데이터를 생성하는 방법 및 시스템에 관한 것이다.

배경 기술

[0003] 최근, 기계학습 모델을 이용한 증권 가격 예측이 수행되고 있다. 특히, 증권 시장에 존재하는 매수 희망자의 매수 요청 가격에 대한 정보, 매도 희망자의 매도 요청 가격에 대한 정보를 포함하는 오더북 데이터를 이용한 증권 가격 예측이 활발하게 수행되고 있다. 기계학습 모델을 이용한 증권 가격 예측은 시장의 복잡한 정보로부터 기존의 고전적인 알고리즘이 찾지 못하는 특징을 찾을 수 있다는 장점이 있다.

[0004] 한편, 패턴 인식에 특화된 기계학습 모델(예를 들어, CNN)은 대부분 2D 데이터를 입력으로 받는데, 오더북 데이터는 일반적으로 테이블 형태로 저장되어 있어, 이러한 기계학습 모델의 입력으로 적절하지 않다는 문제점이 있다. 또한, 테이블 형태의 오더북 데이터를 2D 데이터로 변환하는 경우, 증권 가격 예측을 위한 중요한 정보가 유실될 수 있으며, 이에 따라 기계학습 모델의 예측 성능이 떨어질 수 있다는 문제점이 있다.

발명의 내용

해결하려는 과제

[0006] 본 개시는 상기와 같은 문제점을 해결하기 위한 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법, 시스템 및 기록매체에 저장된 컴퓨터 프로그램을 제공한다.

과제의 해결 수단

[0008] 본 개시는 방법, 장치(시스템) 또는 판독 가능 저장 매체에 저장된 컴퓨터 프로그램을 포함한 다양한 방식으로 구현될 수 있다.

[0009] 본 개시의 일 실시예에 따르면, 적어도 하나의 프로세서에 의해 수행되는, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법은, (a) 미리 결정된 기간 이전의 과거 시점으로부터 현재 시점까지의 시간에 따른, 제1 증권 거래소에서 거래되는 종목에 대한 복수의 매수 요청 가격에 대한 데이터 아이템 및 복수의 매도 요청 가격에 대한 데이터 아이템을 포함한 오더북 데이터를 획득하는 단계, (b) 복수의 매수 요청 가격 중에서 가장 높은 매수 요청 가격과 복수의 매도 요청 가격 중에서 가장 낮은 매도 요청 가격을 기초로 중간 가격을 산출하는 단계 및 (c) 일정한 시간 간격의 시간을 제1 축으로 하고, 틱 단위의 가격을 제2 축으로 하고, 복수의 매도 요청 가격의 각각의 수량에 대한 데이터 또는 복수의 매수 요청 가격의 각각의 수량에 대한 데이터를 제1 축의 시간과 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 대한 값으로 가진 하나 이상의 2D 데이터 아

이템을 포함한 텐서 형태의 데이터를 생성하는 단계- 틱 단위의 가격은 산출된 중간 가격을 포함함 -를 포함한다.

- [0010] 본 개시의 일 실시예에 따르면, 텐서 형태의 데이터는, 복수의 채널의 각각에 대응하는 2D 데이터를 포함하고, 복수의 채널은 제1 채널 및 제2 채널을 포함하고, 텐서 형태의 데이터는, 제1 채널에 대응하는, 시간에 따른 제1 증권 거래소에서 거래되는 종목의 매수 요청 가격에 대한 2D 데이터 및 제2 채널에 대응하는, 시간에 따른 제1 증권 거래소에서 거래되는 종목의 매도 요청 가격에 대한 2D 데이터를 포함한다.
- [0011] 본 개시의 일 실시예에 따르면, 텐서 형태의 데이터는 제3 채널 및 제4 채널을 더 포함하고, 텐서 형태의 데이터는, 제3 채널에 대응하는, 시간에 따른 제2 증권 거래소에서 거래되는 종목의 매수 요청 가격에 대한 2D 데이터 및 제4 채널에 대응하는, 시간에 따른 제2 증권 거래소에서 거래되는 종목의 매도 요청 가격에 대한 2D 데이터를 더 포함한다.
- [0012] 본 개시의 일 실시예에 따르면, 텐서 형태의 데이터는 제5 채널 및 제6 채널을 더 포함하고, 텐서 형태의 데이터는, 하나 이상의 2D 데이터의 각각에서의 제1 축에 대한 좌표를 나타내는 값을 포함한 2D 데이터 및 하나 이상의 2D 데이터의 각각에서의 제2 축에 대한 좌표를 나타내는 값을 포함한 2D 데이터를 더 포함한다.
- [0013] 본 개시의 일 실시예에 따르면, (c) 단계는, 복수의 매도 요청 가격의 각각의 수량 및 복수의 매수 요청 가격의 각각의 수량을 정규화하는 단계 및 정규화된 수량을 제1 축의 시간과 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 연관시키는 단계를 포함한다.
- [0014] 본 개시의 일 실시예에 따르면, 현재 시점이 변함에 따라, 변화된 현재 시점으로부터 미리 결정된 기간 이전의 과거 시점까지의 오더북 데이터에 대해 (a) 단계 내지 (c) 단계를 반복적으로 수행하는 단계를 더 포함한다.
- [0015] 본 개시의 일 실시예에 따르면, 생성된 텐서 형태의 데이터를 복수의 레이어를 포함한 기계학습 모델에 입력하여 예측 결과를 출력하는 단계를 더 포함하고, 예측 결과는, 현재 시점으로부터 미리 결정된 기간 이후의 미래 시점에서의 종목의 가격에 대한 예측 정보를 포함한다.
- [0016] 본 개시의 일 실시예에 따르면, 예측 정보는, 미래 시점에서의 종목의 가격이 현재 시점의 가격보다 높을 것을 나타내는 제1 클래스, 미래 시점에서의 종목의 가격이 현재 시점의 가격과 동일할 것을 나타내는 제2 클래스 및 미래 시점에서의 종목의 가격이 현재 시점의 가격보다 낮을 것을 나타내는 제3 클래스를 포함한다.
- [0017] 본 개시의 일 실시예에 따른 방법을 컴퓨터에서 실행하기 위해 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램이 제공된다.
- [0018] 본 개시의 일 실시예에 따르면, 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 시스템은 하나 이상의 인스트럭션을 저장하는 메모리 및 메모리의 하나 이상의 인스트럭션을 실행함으로써, (a) 미리 결정된 기간 이전의 과거 시점으로부터 현재 시점까지의 시간에 따른, 제1 증권 거래소에서 거래되는 종목에 대한 복수의 매수 요청 가격에 대한 데이터 아이템 및 복수의 매도 요청 가격에 대한 데이터 아이템을 포함한 오더북 데이터를 획득하고, (b) 복수의 매수 요청 가격 중에서 가장 높은 매수 요청 가격과 복수의 매도 요청 가격 중에서 가장 낮은 매도 요청 가격을 기초로 중간 가격을 산출하고, (c) 일정한 시간 간격의 시간을 제1 축으로 하고, 틱 단위의 가격을 제2 축으로 하고, 복수의 매도 요청 가격의 각각의 수량에 대한 데이터 또는 복수의 매수 요청 가격의 각각의 수량에 대한 데이터를 제1 축의 시간과 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 대한 값으로 가진 하나 이상의 2D 데이터 아이템을 포함한 텐서 형태의 데이터를 생성하도록 구성된 적어도 하나의 프로세서를 포함하고, 틱 단위의 가격은 산출된 중간 가격을 포함한다.
- [0019] 본 개시의 일 실시예에 따르면, 텐서 형태의 데이터는, 복수의 채널의 각각에 대응하는 2D 데이터를 포함하고, 복수의 채널은 제1 채널 및 제2 채널을 포함하고, 텐서 형태의 데이터는, 제1 채널에 대응하는, 시간에 따른 제1 증권 거래소에서 거래되는 종목의 매수 요청 가격에 대한 2D 데이터 및 제2 채널에 대응하는, 시간에 따른 제1 증권 거래소에서 거래되는 종목의 매도 요청에 대한 2D 데이터를 포함한다.
- [0020] 본 개시의 일 실시예에 따르면, 텐서 형태의 데이터는 제3 채널 및 제4 채널을 더 포함하고, 텐서 형태의 데이터는, 제3 채널에 대응하는, 시간에 따른 제2 증권 거래소에서 거래되는 종목의 매수 요청 가격에 대한 2D 데이터 및 제4 채널에 대응하는, 시간에 따른 제2 증권 거래소에서 거래되는 종목의 매도 요청 가격에 대한 2D 데이터를 더 포함한다.
- [0021] 본 개시의 일 실시예에 따르면, 텐서 형태의 데이터는 제5 채널 및 제6 채널을 더 포함하고, 텐서 형태의 데이터는, 하나 이상의 2D 데이터의 각각에서의 제1 축에 대한 좌표를 나타내는 값을 포함한 2D 데이터 및 하나 이

상의 2D 데이터의 각각에서의 제2 축에 대한 좌표를 나타내는 값을 포함한 2D 데이터를 더 포함한다.

- [0022] 본 개시의 일 실시예에 따르면, 텐서 형태의 데이터를 생성하는 것은, 복수의 매도 요청 가격의 각각의 수량 및 복수의 매수 요청 가격의 각각의 수량을 정규화하고, 정규화된 수량을 제1 축의 시간과 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 연관시키는 것을 포함한다.
- [0023] 본 개시의 일 실시예에 따르면, 적어도 하나의 프로세서는, 현재 시점이 변함에 따라, 변화된 현재 시점으로부터 미리 결정된 기간 이전의 과거 시점까지의 오더북 데이터에 대해 (a) 내지 (c)를 반복적으로 수행하도록 더 구성된다.
- [0024] 본 개시의 일 실시예에 따르면, 적어도 하나의 프로세서는, 생성된 텐서 형태의 데이터를 복수의 레이어를 포함한 기계학습 모델에 입력하여 예측 결과를 출력하도록 더 구성되고, 예측 결과는, 현재 시점으로부터 미리 결정된 기간 이후의 미래 시점에서의 종목의 가격에 대한 예측 정보를 포함한다.
- [0025] 본 개시의 일 실시예에 따르면, 예측 정보는, 미래 시점에서의 종목의 가격이 현재 시점의 가격보다 높을 것을 나타내는 제1 클래스, 미래 시점에서의 종목의 가격이 현재 시점의 가격과 동일할 것을 나타내는 제2 클래스 및 미래 시점에서의 종목의 가격이 현재 시점의 가격보다 낮을 것을 나타내는 제3 클래스를 포함한다.

발명의 효과

- [0026] 본 개시의 일부 실시예에 따르면, 미래 시장 상황 예측을 위한 중요한 정보의 유실 없이, 테이블 형태의 오더북 데이터가 2D 데이터로 변환될 수 있다.
- [0027] 본 개시의 일 실시예에 따르면, 매수 요청 가격에 대한 정보와 매도 요청 가격에 대한 정보가 별개의 채널에 분리되어 포함됨으로써, 기계학습 모델이 매수 요청 가격에 대한 정보와 매도 요청 가격에 대한 정보를 명확하게 구별할 수 있으며, 그 결과 기계학습 모델의 예측 성능이 향상될 수 있다.
- [0028] 본 개시의 일 실시예에 따르면, 입력 데이터에 포함된 2D 데이터와 해석 맵이 동일한 크기를 가지도록 구성된 기계학습 모델이 사용됨으로써, 예측 결과 출력에 영향을 미친 부분에 대한 정확한 정보가 제공될 수 있다.
- [0029] 본 개시의 일 실시예에 따르면, 설명가능한 데이터에 해석 맵이 적용됨으로써, 직관적으로 이해 가능한 해석 정보가 제공될 수 있다.
- [0030] 본 개시의 효과는 이상에서 언급한 효과로 제한되지 않으며, 언급되지 않은 다른 효과들은 청구범위의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자(“통상의 기술자”라 함)에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [0032] 본 개시의 실시예들은, 이하 설명하는 첨부 도면들을 참조하여 설명될 것이며, 여기서 유사한 참조 번호는 유사한 요소들을 나타내지만, 이에 한정되지는 않는다.
- 도 1은 본 개시의 일 실시예에 따른 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법의 예시를 나타내는 개요도이다.
- 도 2는 본 개시의 일 실시예에 따른 정보 처리 시스템이 복수의 사용자 단말과 통신 가능하도록 연결된 구성을 나타내는 개요도이다.
- 도 3은 본 개시의 일 실시예에 따른 사용자 단말 및 정보 처리 시스템의 내부 구성을 나타내는 블록도이다.
- 도 4는 본 개시의 일 실시예에 따른 오더북 데이터를 2D 데이터로 변환하는 방법의 예시를 나타내는 도면이다.
- 도 5는 본 개시의 일 실시예에 따른 2D 데이터를 생성하고, 생성된 2D 데이터를 정규화하는 방법의 예시를 나타내는 도면이다.
- 도 6은 본 개시의 일 실시예에 따른 텐서 형태의 데이터를 생성하는 방법의 예시를 나타내는 도면이다.
- 도 7은 본 개시의 일 실시예에 따른 텐서 형태의 데이터의 예시를 나타내는 도면이다.
- 도 8은 본 개시의 일 실시예에 따른 3개 이상의 채널을 포함하는 텐서 형태의 데이터의 예시를 나타내는 도면이다.

도 9는 본 개시의 일 실시예에 따른 기계학습 모델의 해석을 제공하는 방법의 예시를 나타내는 도면이다.

도 10은 본 개시의 일 실시예에 따른 입력 데이터에 포함된 2D 데이터와 동일한 크기를 갖는 해석 맵을 제공하기 위한 기계학습 모델의 예시를 나타내는 도면이다.

도 11은 본 개시의 일 실시예에 따른 입력 데이터에 포함된 2D 데이터와 동일한 크기를 갖는 해석 맵을 제공하는 방법의 예시를 나타내는 도면이다.

도 12는 본 개시의 일 실시예에 따른 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법의 예시를 나타내는 흐름도이다.

도 13은 본 개시의 일 실시예에 따른 기계학습 모델의 해석을 제공하는 방법의 예시를 나타내는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0033] 이하, 본 개시의 실시를 위한 구체적인 내용을 첨부된 도면을 참조하여 상세히 설명한다. 다만, 이하의 설명에서는 본 개시의 요지를 불필요하게 흐릴 우려가 있는 경우, 널리 알려진 기능이나 구성에 관한 구체적 설명은 생략하기로 한다.
- [0034] 첨부된 도면에서, 동일하거나 대응하는 구성요소에는 동일한 참조부호가 부여되어 있다. 또한, 이하의 실시예들의 설명에 있어서, 동일하거나 대응되는 구성요소를 중복하여 기술하는 것이 생략될 수 있다. 그러나, 구성요소에 관한 기술이 생략되어도, 그러한 구성요소가 어떤 실시예에 포함되지 않는 것으로 의도되지는 않는다.
- [0035] 개시된 실시예의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 개시는 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 개시가 완전하도록 하고, 본 개시가 통상의 기술자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것일 뿐이다.
- [0036] 본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 개시된 실시예에 대해 구체적으로 설명하기로 한다. 본 명세서에서 사용되는 용어는 본 개시에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 관련 분야에 종사하는 기술자의 의도 또는 관례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서, 본 개시에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 개시의 전반에 걸친 내용을 토대로 정의되어야 한다.
- [0037] 본 명세서에서의 단수의 표현은 문맥상 명백하게 단수인 것으로 특정하지 않는 한, 복수의 표현을 포함한다. 또한, 복수의 표현은 문맥상 명백하게 복수인 것으로 특정하지 않는 한, 단수의 표현을 포함한다. 명세서 전체에서 어떤 부분이 어떤 구성요소를 포함한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다.
- [0038] 또한, 명세서에서 사용되는 '모듈' 또는 '부'라는 용어는 소프트웨어 또는 하드웨어 구성요소를 의미하며, '모듈' 또는 '부'는 어떤 역할들을 수행한다. 그렇지만, '모듈' 또는 '부'는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. '모듈' 또는 '부'는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서, '모듈' 또는 '부'는 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로 코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 또는 변수들 중 적어도 하나를 포함할 수 있다. 구성요소들과 '모듈' 또는 '부'들은 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 '모듈' 또는 '부'들로 결합되거나 추가적인 구성요소들과 '모듈' 또는 '부'들로 더 분리될 수 있다.
- [0039] 본 개시의 일 실시예에 따르면, '모듈' 또는 '부'는 프로세서 및 메모리로 구현될 수 있다. '프로세서'는 범용 프로세서, 중앙 처리 장치(CPU), 마이크로프로세서, 디지털 신호 프로세서(DSP), 제어기, 마이크로제어기, 상태 머신 등을 포함하도록 넓게 해석되어야 한다. 몇몇 환경에서, '프로세서'는 주문형 반도체(ASIC), 프로그램가능 로직 디바이스(PLD), 필드 프로그램가능 게이트 어레이(FPGA) 등을 지칭할 수도 있다. '프로세서'는, 예를 들어, DSP와 마이크로프로세서의 조합, 복수의 마이크로프로세서들의 조합, DSP 코어와 결합한 하나 이상의 마이크로프로세서들의 조합, 또는 임의의 다른 그러한 구성들의 조합과 같은 처리 디바이스들의 조합을 지칭할 수도 있다. 또한, '메모리'는 전자 정보를 저장 가능한 임의의 전자 컴포넌트를 포함하도록 넓게 해석되어야 한

다. '메모리'는 임의 액세스 메모리(RAM), 판독-전용 메모리(ROM), 비-휘발성 임의 액세스 메모리(NVRAM), 프로그램가능 판독-전용 메모리(PROM), 소거-프로그램가능 판독 전용 메모리(EPROM), 전기적으로 소거가능 PROM(EEPROM), 플래쉬 메모리, 자기 또는 광학 데이터 저장장치, 레지스터들 등과 같은 프로세서-판독가능 매체의 다양한 유형들을 지칭할 수도 있다. 프로세서가 메모리로부터 정보를 판독하고/하거나 메모리에 정보를 기록할 수 있다면 메모리는 프로세서와 전자 통신 상태에 있다고 불린다. 프로세서에 집적된 메모리는 프로세서와 전자 통신 상태에 있다.

[0040] 본 개시에서, '시스템'은 서버 장치와 클라우드 장치 중 적어도 하나의 장치를 포함할 수 있으나, 이에 한정되는 것은 아니다. 예를 들어, 시스템은 하나 이상의 서버 장치로 구성될 수 있다. 다른 예로서, 시스템은 하나 이상의 클라우드 장치로 구성될 수 있다. 또 다른 예로서, 시스템은 서버 장치와 클라우드 장치가 함께 구성되어 동작될 수 있다.

[0041] 본 개시에서, '기계학습 모델'은 주어진 입력에 대한 해답(answer)을 추론하는데 사용하는 임의의 모델을 포함할 수 있다. 일 실시예에 따르면, 기계학습 모델은 입력 레이어(층), 복수 개의 은닉 레이어 및 출력 레이어를 포함한 인공신경망 모델을 포함할 수 있다. 여기서, 각 레이어는 복수의 노드를 포함할 수 있다. 또한, 본 개시에서, 기계학습 모델은 인공신경망 모델을 지칭할 수 있으며, 인공신경망 모델은 기계학습 모델을 지칭할 수 있다.

[0042] 본 개시에서, '디스플레이'는 컴퓨팅 장치와 연관된 임의의 디스플레이 장치를 지칭할 수 있는데, 예를 들어, 컴퓨팅 장치에 의해 제어되거나 컴퓨팅 장치로부터 제공된 임의의 정보/데이터를 표시할 수 있는 임의의 디스플레이 장치를 지칭할 수 있다.

[0043] 본 개시에서, '복수의 A의 각각' 또는 '복수의 A 각각'은 복수의 A에 포함된 모든 구성 요소의 각각을 지칭하거나, 복수의 A에 포함된 일부 구성 요소의 각각을 지칭할 수 있다. 예를 들어, 하나 이상의 채널의 각각은 하나 이상의 채널에 포함된 모든 채널의 각각을 지칭하거나, 하나 이상의 채널에 포함된 일부 채널의 각각을 지칭할 수 있다. 이와 유사하게, 2D 이미지에 포함된 픽셀의 각각은, 2D 이미지에 포함된 모든 픽셀의 각각을 지칭하거나 2D 이미지에 포함된 일부 픽셀의 각각을 지칭할 수 있다.

[0044] 본 개시에서, '데이터'는 데이터 아이템을 지칭하고, '데이터 아이템'은 데이터를 지칭할 수 있다.

[0045] 본 개시에서, '종목'이란, 증권 시장에서 매매 거래의 대상이 되는 주식, 채권, 파생상품(옵션, 선물 등)과 같은 유가 증권을 내용과 형식에 따라 분류한 것을 지칭할 수 있다. 종목은 개별 종목 이외에 지수 관련 종목, 산업 섹터와 관련된 종목, 특정 상품(예를 들어, 원유, 농산물, 금 등)에 대한 종목, 환율 관련 종목 등도 포함할 수 있다.

[0046] 본 개시에서, '증권 거래소'란, 적어도 하나의 국가에서 유통되는 유가 증권을 유통하는 곳으로서, 각 기업이나 정보 등이 발행한 유가 증권을 상장하여 거래할 수 있도록 중개하는 곳을 지칭할 수 있다. 일 실시예에서, 증권 거래소는 증권 거래소의 시스템을 포함할 수 있다.

[0047] 본 개시에서, '오더북(Order Book; OB 또는 Limit Order Book; LoB)' 또는 '오더북 데이터'는 증권 시장에 존재하는 매수 희망자의 매수 요청 가격에 대한 정보(매수 요청 가격, 수량 등), 매도 희망자의 매도 요청 가격에 대한 정보(매도 요청 가격, 수량 등)를 포함하는 데이터를 지칭할 수 있다. 일 실시예에 따르면, 오더북 또는 오더북 데이터는 테이블 형태의 데이터를 포함할 수 있다.

[0048] 본 개시에서, '오더북의 상단(Top of the Book; ToB)'은 가장 높은 매수 요청 가격에 대한 정보(가격, 수량 등) 및 가장 낮은 매도 요청 가격에 대한 정보(가격, 수량 등)를 포함할 수 있다.

[0049] 본 개시에서, '스프레드(spread)'는 가장 높은 매수 요청 가격과 가장 낮은 매도 요청 가격의 차이 즉, 두 오더북의 상단의 차이를 지칭할 수 있다.

[0050] 본 개시에서, '이미지'는 '텐서 형태의 데이터'와 혼용되어 사용될 수 있다.

[0051] 도 1은 본 개시의 일 실시예에 따른 오더북 데이터(110)를 기계학습 모델(130)을 위한 2D 데이터로 변환하는 방법의 예시를 나타내는 개요도이다. 일 실시예에 따르면, 정보 처리 시스템은 기계학습 모델(130)을 이용하여, 특정 증권 거래소에서 거래되는 특정 종목에 대한 과거부터 현재까지의 마켓 데이터를 기초로 해당 종목에 대한 예측 결과(140)를 생성할 수 있다. 예를 들어, 정보 처리 시스템은 기계학습 모델(130)을 이용하여, 특정 종목에 대한 과거부터 현재까지의 오더북 데이터(110)를 기초로, 해당 종목의 미래 시점에서의 가격에 대한 예측 정보를 생성할 수 있다. 여기서, 기계학습 모델(130)이 2D 데이터의 학습에 적합한 구조를 가진 기계학습 모델

(130)(예를 들어, CNN(Convolutional Neural Network))인 경우, 입력 데이터는 하나 이상의 2D 데이터 아이템을 포함한 텐서 형태의 데이터(120)로 입력되어야 할 수 있다.

[0052] 본 개시에서, 하나 이상의 2D 데이터 아이템을 포함하는 텐서 형태의 데이터(120)는 2D 이미지를 포함하는 개념일 수 있다. 예를 들어, 그레이 스케일의 2D 이미지는 하나의 채널에 대응되는 2D 데이터를 포함하는 텐서 형태의 데이터일 수 있으며, RGB 2D 이미지는 Red 채널, Green 채널 및 Blue 채널의 3개의 채널 각각에 대응되는 2D 데이터를 포함하는 텐서 형태의 데이터일 수 있다.

[0053] 한편, 오더북 데이터(110)는 특정 증권 거래소에서 거래되는 특정 종목에 대한 복수의 매수 요청 가격에 대한 데이터 아이템 및 복수의 매도 요청 가격에 대한 데이터 아이템을 포함하는 데이터일 수 있으며, 복수의 매수 요청 가격에 대한 데이터 아이템 및 복수의 매도 요청 가격에 대한 데이터 아이템 각각은 요청 가격 및 수량에 대한 데이터를 포함할 수 있다. 일반적으로, 이러한 오더북 데이터(110)는 도 1에 도시된 바와 같이 테이블 형태로 표현된다. 따라서, 본 개시의 일 실시예에 따르면, 정보 처리 시스템은 기계학습 모델(130)(예를 들어, CNN)을 이용하여, 특정 종목에 대한 오더북 데이터(110)를 기초로, 해당 종목의 미래 가격에 대한 예측 정보를 생성하기 위해, 오더북 데이터(110)를 하나 이상의 2D 데이터 아이템을 포함한 텐서 형태의 데이터(120)로 변환할 수 있다. 여기서, 테이블 형태의 오더북 데이터(110)를 텐서 형태의 데이터(120)로 변환하는 방법에 따라, 오더북 데이터(110)의 정보가 손실될 수 있으며, 변환된 텐서 형태의 데이터(120)가 기계학습 모델이 학습하거나 추론하기에 용이한 형태인지에 따라, 기계학습 모델의 학습 또는 추론 성능에 영향을 미칠 수 있다. 이에 따라, 본 개시는 오더북 데이터(110)로부터, 데이터의 손실 없이, 기계학습 모델이 학습하거나 추론하기에 용이한 형태의 텐서 형태의 데이터(120)로 변환하는 방법을 제시한다.

[0054] 일 실시예에 따르면, 변환된 텐서 형태의 데이터(120)는 매수 요청 가격과 관련된 2D 데이터 아이템 및 매도 요청 가격과 관련된 2D 데이터 아이템을 포함할 수 있다. 매수 요청 가격과 관련된 2D 데이터 아이템 및 매도 요청 가격과 관련된 2D 데이터 아이템은 텐서 형태의 데이터(120)의 별개의 채널에 포함될 수 있다. 이에 따라, 매수 요청과 관련된 데이터와 매도 요청과 관련된 데이터가 뚜렷이 구분되며, 기계학습 모델은 매수 요청과 관련된 데이터와 매도 요청과 관련된 데이터를 정확히 구분하여 학습 또는 추론할 수 있다.

[0055] 일 실시예에 따르면, 텐서 형태의 데이터(120)에 포함되는 2D 데이터 아이템 각각은, 시간 축 및 가격 축을 가질 수 있으며, 시간 축에는 과거부터 현재까지의 시점 중 일정한 시간 간격의 시점들이 포함될 수 있고, 가격 축에는 틱 단위의 가격들이 포함될 수 있다. 또한, 2D 데이터 아이템 각각은 각 시간에서의 각 매도 요청 가격의 매도 요청 수량에 대한 데이터 또는 각 시간에서의 각 매수 요청 가격의 매수 요청 수량에 대한 데이터를 시간 축 및 가격 축에 따라 정의되는 좌표에 대한 값으로 가질 수 있다.

[0056] 일 실시예에 따르면, 매수 요청 수량 및 매도 요청 수량이 전혀 없는 가격도 가격 축에 포함될 수 있다. 또한, 일 실시예에 따르면, 정보 처리 시스템은 오더북 데이터(110)에 포함된 복수의 매수 요청 가격 중에서 가장 높은 매수 요청 가격과 오더북 데이터(110)에 포함된 복수의 매도 요청 가격 중에서 가장 낮은 매도 요청 가격을 기초로 중간 가격(mid-price)을 산출하여, 2D 데이터 아이템의 가격 축에 포함시킬 수 있다. 이와 같은 구성에 의해, 매수, 매도의 각 주문 레벨 별 수량에 대한 정보, 스프레드(spread), 중간 가격 등 미래 시점의 가격 예측을 위한 중요한 정보의 유실 없이 오더북 데이터(110)를 텐서 형태의 데이터(120)로 변환시킬 수 있다. 오더북 데이터(110)를 하나 이상의 2D 데이터 아이템을 포함한 텐서 형태의 데이터(120)로 변환하는 구체적인 방법은 도 4 내지 도 6을 참조하여 후술된다.

[0057] 일 실시예에 따르면, 정보 처리 시스템은 기계학습 모델(130)에 변환된 텐서 형태의 데이터(120)를 입력하여, 예측 결과(140)를 출력할 수 있다. 예를 들어, 정보 처리 시스템은 기계학습 모델(130)에 변환된 텐서 형태의 데이터(120)를 입력하여, 현재 시점으로부터 미리 결정된 기간 이후의 미래 시점에서의 특정 종목의 가격(예를 들어, 중간 가격, 시장 가격 등)에 대한 예측 정보를 출력할 수 있다. 일 실시예에 따르면, 특정 종목의 가격에 대한 예측 정보는 미래 시점에서의 특정 종목의 가격이 현재 시점의 가격보다 높을 것을 나타내는 제1 클래스, 미래 시점에서의 특정 종목의 가격이 현재 시점의 가격과 동일할 것을 나타내는 제2 클래스 및 미래 시점에서의 특정 종목의 가격이 현재 시점의 가격보다 낮을 것을 나타내는 제3 클래스를 포함할 수 있다.

[0058] 일 실시예에 따르면, 정보 처리 시스템은 오더북 데이터(110)를 텐서 형태의 데이터(120)로 변환하는 과정 및/또는 변환된 텐서 형태의 데이터(120)를 기계학습 모델(130)에 입력하여 예측 결과(140)를 출력하는 과정을 시간이 흐름에 따라 복수 번 반복적으로 수행할 수 있다. 이와 같은 구성에 의해, 최근 데이터를 이용한 최신 예측 결과를 반복적으로 생성할 수 있다.

- [0059] 도 2는 본 개시의 일 실시예에 따른 정보 처리 시스템(230)이 복수의 사용자 단말(210_1, 210_2, 210_3)과 통신 가능하도록 연결된 구성을 나타내는 개요도이다. 도시된 바와 같이, 복수의 사용자 단말(210_1, 210_2, 210_3)은 네트워크(220)를 통해 증권 가격 예측 서비스를 제공할 수 있는 정보 처리 시스템(230)과 연결될 수 있다. 여기서, 복수의 사용자 단말(210_1, 210_2, 210_3)은 증권 가격 예측 서비스를 제공받을 사용자 및/또는 증권 가격 예측 서비스의 운영자의 단말을 포함할 수 있다. 일 실시예에서, 정보 처리 시스템(230)은 증권 가격 예측 서비스 등과 관련된 컴퓨터 실행 가능한 프로그램(예를 들어, 다운로드 가능한 애플리케이션) 및 데이터를 저장, 제공 및 실행할 수 있는 하나 이상의 서버 장치 및/또는 데이터베이스, 또는 클라우드 컴퓨팅 서비스 기반의 하나 이상의 분산 컴퓨팅 장치 및/또는 분산 데이터베이스를 포함할 수 있다.
- [0060] 정보 처리 시스템(230)에 의해 제공되는 증권 가격 예측 서비스는, 복수의 사용자 단말(210_1, 210_2, 210_3)의 각각에 설치된 증권 가격 예측 애플리케이션, 주식 거래 애플리케이션, 파생상품 거래 애플리케이션, 모바일 브라우저 애플리케이션 또는 웹 브라우저 등을 통해 사용자에게 제공될 수 있다. 예를 들어, 정보 처리 시스템(230)은 증권 가격 예측 애플리케이션 등을 통해 사용자 단말(210_1, 210_2, 210_3)로부터 수신되는 가격 예측 요청 또는 가격 예측에 대한 해석 요청에 대응하는 정보를 제공하거나 대응하는 처리를 수행할 수 있다.
- [0061] 복수의 사용자 단말(210_1, 210_2, 210_3)은 네트워크(220)를 통해 정보 처리 시스템(230)과 통신할 수 있다. 네트워크(220)는 복수의 사용자 단말(210_1, 210_2, 210_3)과 정보 처리 시스템(230) 사이의 통신이 가능하도록 구성될 수 있다. 네트워크(220)는 설치 환경에 따라, 예를 들어, 이더넷(Ethernet), 유선 홈 네트워크(Power Line Communication), 전화선 통신 장치 및 RS-serial 통신 등의 유선 네트워크, 이동통신망, WLAN(Wireless LAN), Wi-Fi, Bluetooth 및 ZigBee 등과 같은 무선 네트워크 또는 그 조합으로 구성될 수 있다. 통신 방식은 제한되지 않으며, 네트워크(220)가 포함할 수 있는 통신망(일례로, 이동통신망, 유선 인터넷, 무선 인터넷, 방송망, 위성망 등)을 활용하는 통신 방식뿐만 아니라 사용자 단말(210_1, 210_2, 210_3) 사이의 근거리 무선 통신 역시 포함될 수 있다.
- [0062] 도 2에서 휴대폰 단말(210_1), 태블릿 단말(210_2) 및 PC 단말(210_3)이 사용자 단말의 예로서 도시되었으나, 이에 한정되지 않으며, 사용자 단말(210_1, 210_2, 210_3)은 유선 및/또는 무선 통신이 가능하고 증권 가격 예측 애플리케이션, 주식 거래 애플리케이션, 파생상품 거래 애플리케이션, 모바일 브라우저 애플리케이션 또는 웹 브라우저 등이 설치되어 실행될 수 있는 임의의 컴퓨팅 장치일 수 있다. 예를 들어, 사용자 단말은, AI 스피커, 스마트폰, 휴대폰, 내비게이션, 컴퓨터, 노트북, 디지털방송용 단말, PDA(Personal Digital Assistants), PMP(Portable Multimedia Player), 태블릿 PC, 게임 콘솔(game console), 웨어러블 디바이스(wearable device), IoT(internet of things) 디바이스, VR(virtual reality) 디바이스, AR(augmented reality) 디바이스, 셋톱 박스 등을 포함할 수 있다. 또한, 도 2에는 3개의 사용자 단말(210_1, 210_2, 210_3)이 네트워크(220)를 통해 정보 처리 시스템(230)과 통신하는 것으로 도시되어 있으나, 이에 한정되지 않으며, 상이한 수의 사용자 단말이 네트워크(220)를 통해 정보 처리 시스템(230)과 통신하도록 구성될 수도 있다.
- [0063] 일 실시예에 따르면, 정보 처리 시스템(230)은 복수의 사용자 단말(210_1, 210_2, 210_3)로부터 특정 종목의 가격 예측 요청을 수신할 수 있다. 이 경우, 정보 처리 시스템(230)은 미리 결정된 기간 이전의 과거 시점으로부터 현재 시점까지의 시간에 따른, 해당 종목에 대한 복수의 매수 요청 가격에 대한 데이터 아이템 및 복수의 매도 요청 가격에 대한 데이터 아이템을 포함한 오더북 데이터를 획득하고, 복수의 매수 요청 가격 중에서 가장 높은 매수 요청 가격과 복수의 매도 요청 가격 중에서 가장 낮은 매도 요청 가격을 기초로 중간 가격을 산출할 수 있다. 그런 다음, 정보 처리 시스템(230)은 일정한 시간 간격의 시간을 제1 축으로 하고, 틱 단위의 가격을 제2 축으로 하고, 복수의 매도 요청 가격의 각각의 수량에 대한 데이터 또는 복수의 매수 요청 가격의 각각의 수량에 대한 데이터를 제1 축의 시간과 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 대한 값으로 가진 하나 이상의 2D 데이터 아이템을 포함한 텐서 형태의 데이터를 생성할 수 있다. 그 후, 정보 처리 시스템(230)은 생성된 텐서 형태의 데이터를 복수의 레이어를 포함한 기계학습 모델에 입력하여 예측 결과를 출력하고, 예측 결과를 네트워크(220)를 통해 사용자 단말(210_1, 210_2, 210_3)로 제공할 수 있다.
- [0064] 다른 실시예에 따르면, 정보 처리 시스템(230)은 복수의 사용자 단말(210_1, 210_2, 210_3)로부터 가격 예측에 대한 해석 정보 요청을 수신할 수 있다. 이 경우, 정보 처리 시스템(230)은 기계학습에 포함된 복수의 레이어 중 적어도 하나의 레이어로부터의 출력인 특징 맵을 이용하여 예측 결과에 대한 해석 맵을 생성할 수 있으며, 테이블 형태의 데이터 및 텐서 형태의 데이터를 기초로 설명가능한 데이터를 생성할 수 있다. 그런 다음, 정보 처리 시스템(230)은 설명가능한 데이터에 해석 맵을 적용하여, 해석 맵이 적용된 설명가능한 데이터를 네트워크

(220)를 통해 사용자 단말(210_1, 210_2, 210_3)로 제공할 수 있다.

[0065] 도 3은 본 개시의 일 실시예에 따른 사용자 단말(210) 및 정보 처리 시스템(230)의 내부 구성을 나타내는 블록도이다. 사용자 단말(210)은 증권 가격 예측 애플리케이션, 주식 거래 애플리케이션, 파생상품 거래 애플리케이션, 모바일 브라우저 애플리케이션 또는 웹 브라우저 등을 실행 가능하고 유/무선 통신이 가능한 임의의 컴퓨팅 장치를 지칭할 수 있으며, 예를 들어, 도 2의 휴대폰 단말(210_1), 태블릿 단말(210_2), PC 단말(210_3) 등을 포함할 수 있다. 도시된 바와 같이, 사용자 단말(210)은 메모리(312), 프로세서(314), 통신 모듈(316) 및 입출력 인터페이스(318)를 포함할 수 있다. 이와 유사하게, 정보 처리 시스템(230)은 메모리(332), 프로세서(334), 통신 모듈(336) 및 입출력 인터페이스(338)를 포함할 수 있다. 도 3에 도시된 바와 같이, 사용자 단말(210) 및 정보 처리 시스템(230)은 각각의 통신 모듈(316, 336)을 이용하여 네트워크(220)를 통해 정보 및/또는 데이터를 통신할 수 있도록 구성될 수 있다. 또한, 입출력 장치(320)는 입출력 인터페이스(318)를 통해 사용자 단말(210)에 정보 및/또는 데이터를 입력하거나 사용자 단말(210)로부터 생성된 정보 및/또는 데이터를 출력하도록 구성될 수 있다.

[0066] 메모리(312, 332)는 비-일시적인 임의의 컴퓨터 판독 가능한 기록매체를 포함할 수 있다. 일 실시예에 따르면, 메모리(312, 332)는 RAM(random access memory), ROM(read only memory), 디스크 드라이브, SSD(solid state drive), 플래시 메모리(flash memory) 등과 같은 비소멸성 대용량 저장 장치(permanent mass storage device)를 포함할 수 있다. 다른 예로서, ROM, SSD, 플래시 메모리, 디스크 드라이브 등과 같은 비소멸성 대용량 저장 장치는 메모리와는 구분되는 별도의 영구 저장 장치로서 사용자 단말(210) 또는 정보 처리 시스템(230)에 포함될 수 있다. 또한, 메모리(312, 332)에는 운영체제와 적어도 하나의 프로그램 코드(예를 들어, 사용자 단말(210)에 설치되어 구동되는 증권 가격 예측 애플리케이션 등을 위한 코드)가 저장될 수 있다.

[0067] 이러한 소프트웨어 구성요소들은 메모리(312, 332)와는 별도의 컴퓨터에서 판독가능한 기록매체로부터 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독가능한 기록매체는 이러한 사용자 단말(210) 및 정보 처리 시스템(230)에 직접 연결가능한 기록 매체를 포함할 수 있는데, 예를 들어, 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록매체를 포함할 수 있다. 다른 예로서, 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록매체가 아닌 통신 모듈을 통해 메모리(312, 332)에 로딩될 수도 있다. 예를 들어, 적어도 하나의 프로그램은 개발자들 또는 애플리케이션의 설치 파일을 배포하는 파일 배포 시스템이 네트워크(220)를 통해 제공하는 파일들에 의해 설치되는 컴퓨터 프로그램에 기반하여 메모리(312, 332)에 로딩될 수 있다.

[0068] 프로세서(314, 334)는 기본적인 산술, 로직 및 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(312, 332) 또는 통신 모듈(316, 336)에 의해 프로세서(314, 334)로 제공될 수 있다. 예를 들어, 프로세서(314, 334)는 메모리(312, 332)와 같은 기록 장치에 저장된 프로그램 코드에 따라 수신되는 명령을 실행하도록 구성될 수 있다.

[0069] 통신 모듈(316, 336)은 네트워크(220)를 통해 사용자 단말(210)과 정보 처리 시스템(230)이 서로 통신하기 위한 구성 또는 기능을 제공할 수 있으며, 사용자 단말(210) 및/또는 정보 처리 시스템(230)이 다른 사용자 단말 또는 다른 시스템(일례로 별도의 클라우드 시스템 등)과 통신하기 위한 구성 또는 기능을 제공할 수 있다. 일례로, 사용자 단말(210)의 프로세서(314)가 메모리(312) 등과 같은 기록 장치에 저장된 프로그램 코드에 따라 생성한 요청 또는 데이터(예를 들어, 증권 가격 예측 요청, 예측 결과에 대한 해석 정보 요청 등)는 통신 모듈(316)의 제어에 따라 네트워크(220)를 통해 정보 처리 시스템(230)으로 전달될 수 있다. 역으로, 정보 처리 시스템(230)의 프로세서(334)의 제어에 따라 제공되는 제어 신호나 명령이 통신 모듈(336)과 네트워크(220)를 거쳐 사용자 단말(210)의 통신 모듈(316)을 통해 사용자 단말(210)에 수신될 수 있다. 예를 들어, 사용자 단말(210)은 정보 처리 시스템(230)으로부터 통신 모듈(316)을 통해 예측된 증권 가격, 예측 결과에 대한 해석 정보 등을 수신할 수 있다.

[0070] 입출력 인터페이스(318)는 입출력 장치(320)와의 인터페이스를 위한 수단일 수 있다. 일 예로서, 입력 장치는 오디오 센서 및/또는 이미지 센서를 포함한 카메라, 키보드, 마이크로폰, 마우스 등의 장치를, 그리고 출력 장치는 디스플레이, 스피커, 햅틱 피드백 디바이스(haptic feedback device) 등과 같은 장치를 포함할 수 있다. 다른 예로, 입출력 인터페이스(318)는 터치스크린 등과 같이 입력과 출력을 수행하기 위한 구성 또는 기능이 하나로 통합된 장치와의 인터페이스를 위한 수단일 수 있다. 예를 들어, 사용자 단말(210)의 프로세서(314)가 메모리(312)에 로딩된 컴퓨터 프로그램의 명령을 처리함에 있어서 정보 처리 시스템(230)이나 다른 사용자 단말이 제공하는 정보 및/또는 데이터를 이용하여 구성되는 서비스 화면 등이 입출력 인터페이스(318)를 통해 디스플레이

이에 표시될 수 있다. 도 3에서는 입출력 장치(320)가 사용자 단말(210)에 포함되지 않도록 도시되어 있으나, 이에 한정되지 않으며, 사용자 단말(210)과 하나의 장치로 구성될 수 있다. 또한, 정보 처리 시스템(230)의 입출력 인터페이스(338)는 정보 처리 시스템(230)과 연결되거나 정보 처리 시스템(230)이 포함할 수 있는 입력 또는 출력을 위한 장치(미도시)와의 인터페이스를 위한 수단일 수 있다. 도 3에서는 입출력 인터페이스(318, 338)가 프로세서(314, 334)와 별도로 구성된 요소로서 도시되었으나, 이에 한정되지 않으며, 입출력 인터페이스(318, 338)가 프로세서(314, 334)에 포함되도록 구성될 수 있다.

[0071] 사용자 단말(210) 및 정보 처리 시스템(230)은 도 3의 구성요소들보다 더 많은 구성요소들을 포함할 수 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다. 일 실시예에 따르면, 사용자 단말(210)은 상술된 입출력 장치(320) 중 적어도 일부를 포함하도록 구현될 수 있다. 또한, 사용자 단말(210)은 트랜시버(transceiver), GPS(Global Positioning system) 모듈, 카메라, 각종 센서, 데이터베이스 등과 같은 다른 구성요소들을 더 포함할 수 있다. 예를 들어, 사용자 단말(210)이 스마트폰인 경우, 일반적으로 스마트폰이 포함하고 있는 구성요소를 포함할 수 있으며, 예를 들어, 가속도 센서, 자이로 센서, 카메라 모듈, 각종 물리적인 버튼, 터치패널을 이용한 버튼, 입출력 포트, 진동을 위한 진동기 등의 다양한 구성요소들이 사용자 단말(210)에 더 포함되도록 구현될 수 있다. 일 실시예에 따르면, 사용자 단말(210)의 프로세서(314)는 증권 가격 예측 서비스를 제공하는 애플리케이션 등이 동작하도록 구성될 수 있다. 이 때, 해당 애플리케이션 및/또는 프로그램과 연관된 코드가 사용자 단말(210)의 메모리(312)에 로딩될 수 있다.

[0072] 증권 가격 예측 서비스를 제공하는 애플리케이션 등을 위한 프로그램이 동작되는 동안에, 프로세서(314)는 입출력 인터페이스(318)와 연결된 터치 스크린, 키보드, 오디오 센서 및/또는 이미지 센서를 포함한 카메라, 마이크로폰 등의 입력 장치를 통해 입력되거나 선택된 텍스트, 이미지, 영상, 음성 및/또는 동작 등을 수신할 수 있으며, 수신된 텍스트, 이미지, 영상, 음성 및/또는 동작 등을 메모리(312)에 저장하거나 통신 모듈(316) 및 네트워크(220)를 통해 정보 처리 시스템(230)에 제공할 수 있다. 예를 들어, 프로세서(314)는 특정 종목에 대한 가격 예측을 요청하는 사용자 입력을 수신하여, 통신 모듈(316) 및 네트워크(220)를 통해 정보 처리 시스템(230)에 제공할 수 있다. 다른 예로서, 프로세서(314)는 입력 장치(320)를 통해 예측 결과에 대한 해석 정보를 요청하는 사용자 입력을 수신하여, 네트워크(220) 및 통신 모듈(316)을 통해 정보 처리 시스템(230)에 제공할 수 있다.

[0073] 사용자 단말(210)의 프로세서(314)는 입력 장치(320), 다른 사용자 단말, 정보 처리 시스템(230) 및/또는 복수의 외부 시스템으로부터 수신된 정보 및/또는 데이터를 관리, 처리 및/또는 저장하도록 구성될 수 있다. 프로세서(314)에 의해 처리된 정보 및/또는 데이터는 통신 모듈(316) 및 네트워크(220)를 통해 정보 처리 시스템(230)에 제공될 수 있다. 사용자 단말(210)의 프로세서(314)는 입출력 인터페이스(318)를 통해 입출력 장치(320)로 정보 및/또는 데이터를 전송하여, 출력할 수 있다. 예를 들면, 프로세서(314)는 수신한 정보 및/또는 데이터를 사용자 단말의 화면에 디스플레이할 수 있다.

[0074] 정보 처리 시스템(230)의 프로세서(334)는 복수의 사용자 단말(210) 및/또는 복수의 외부 시스템으로부터 수신된 정보 및/또는 데이터를 관리, 처리 및/또는 저장하도록 구성될 수 있다. 프로세서(334)에 의해 처리된 정보 및/또는 데이터는 통신 모듈(336) 및 네트워크(220)를 통해 사용자 단말(210)에 제공할 수 있다. 일 실시예에서, 정보 처리 시스템(230)의 프로세서(334)는 복수의 사용자 단말(210)로부터의 증권 가격 예측 요청에 기초하여, 오더북 데이터를 텐서 형태의 데이터로 변환하고, 해당 데이터를 기계학습 모델에 입력하여 출력된 예측 결과를 통신 모듈(336) 및 네트워크(220)를 통해 사용자 단말(210)에 제공할 수 있다.

[0075] 정보 처리 시스템(230)의 프로세서(334)는 사용자 단말(210)의 디스플레이 출력 가능 장치(예: 터치 스크린, 디스플레이 등), 음성 출력 가능 장치(예: 스피커) 등의 출력 장치(320)를 통해 처리된 정보 및/또는 데이터를 출력하도록 구성될 수 있다. 예를 들어, 정보 처리 시스템(230)의 프로세서(334)는, 특정 종목에 대한 가격 예측 결과를 통신 모듈(336) 및 네트워크(220)를 통해 사용자 단말(210)로 제공하고, 상품 기획진을 사용자 단말(210)의 디스플레이 출력 가능 장치 등을 통해 출력하도록 구성될 수 있다. 다른 예로서, 정보 처리 시스템(230)의 프로세서(334)는 통신 모듈(336) 및 네트워크(220)를 통해 사용자 단말(210)로 가격 예측 결과에 대한 해석 정보 등을 제공하고, 사용자 단말(210)의 디스플레이 출력 가능 장치 등을 통해 출력하도록 구성될 수 있다.

[0076] 도 4는 본 개시의 일 실시예에 따른 오더북 데이터(410, 420, 430)를 2D 데이터(412, 422, 432)로 변환하는 방법의 예시를 나타내는 도면이다. 오더북 데이터(410, 420, 430)는, 하나 이상의 시점에서의 특정 종목에 대한 유효한 매도 요청(ask) 가격 및 수량, 매수 요청(bid) 가격 및 수량에 대한 정보 중 적어도 일부를 포함할 수

있다. 예를 들어, 제1 오더북 데이터(410)는 시점 t에서의 제1 종목에 대한 유효한 매도 요청 가격 및 수량, 매수 요청 가격 및 수량에 대한 정보 중 적어도 일부를 포함하며, 제1 오더북 데이터(420)는 시점 t에서의 제2 종목에 대한 유효한 매도 요청 가격 및 수량, 매수 요청 가격 및 수량에 대한 정보 중 적어도 일부를 포함할 수 있다. 또한, 제3 오더북 데이터(430)는 시점 t 및 시점 t-1(시점 t로부터 1 단위 시간 전의 시점)에서의 제2 종목에 대한 유효한 매도 요청 가격 및 수량, 매수 요청 가격 및 수량에 대한 정보 중 적어도 일부를 포함할 수 있다.

[0077] 한편, 틱은 매수 또는 매도 요청을 할 수 있는 최소 가격 단위를 지칭할 수 있으며, 증권 거래소에서 거래되는 종목 별로 미리 지정된 값이 틱으로 사용될 수 있다. 예를 들어, 제1 오더북 데이터(410)의 제1 종목, 제1 오더북 데이터(420) 및 제3 오더북 데이터(430)의 제2 종목은 틱이 100원일 수 있다. 즉, 100원 단위로 매수 요청 또는 매도 요청이 수행될 수 있으며, 이와 유사하게, 100원 단위로 시장 가격이 형성될 수 있다.

[0078] 일 실시예에 따르면, 정보 처리 시스템은 2D 데이터(412, 422, 432)의 가격 축을 결정하기 위해, 시점 t(예를 들어, 2D 데이터(412, 422, 432)의 시간 축에 포함될 시간 중 가장 최근 시점)에서의 중간 가격(mid-price)을 산출할 수 있다. 예를 들어, 중간 가격은 다음과 같은 수학적 식 1에 의해 산출될 수 있다.

수학적 식 1

[0079]
$$mid\ price = \frac{ToB_{ask} + ToB_{bid}}{2}$$

[0081] 여기서, ToB_{ask} 는 가장 낮은 매도 요청 가격, ToB_{bid} 는 가장 높은 매수 요청 가격을 나타낸다. 즉, 수학적 식 1에 따른 중간 가격은 가장 낮은 매도 요청 가격과 가장 높은 매수 요청 가격의 평균으로 산출될 수 있다. 예를 들어, 제1 오더북 데이터(410)로부터 산출된 시점 t에서의 수학적 식 1에 따른 중간 가격은 69400원이다.

[0082] 한편, 제1 오더북 데이터(420) 및 제3 오더북 데이터(430)로부터 산출된 시점 t에서의 수학적 식 1에 따른 중간 가격은 69350원으로서, 해당 가격은 제2 종목 및 제2 종목을 틱 단위의 가격에 해당하지 않는 가격이다. 즉, 해당 가격은 매수 요청 및 매도 요청이 수행될 수 없는 가격이며, 형성될 수 없는 시장 가격이다. 따라서, 일 실시예에 따르면, 정보 처리 시스템은 중간 가격이 틱 단위의 가격이 되도록, 양자화된 중간 가격(quantized mid-price)을 산출할 수 있다. 예를 들어, 양자화된 중간 가격은 아래와 같은 수학적 식 2로 산출될 수 있다.

수학적 식 2

[0083]
$$quantized\ mid\ price = ceil\left(\frac{mid\ price}{tick}\right) \times tick$$

[0085] 여기서, $mid\ price$ 는 수학적 식 1에 따른 중간 가격, $tick$ 은 틱, $ceil$ 은 소수점 첫째 자리에서 수행하는 올림 연산(일부 실시예에서, 올림 연산을 대체하여 내림 연산 가능)을 나타낸다. 본 개시의 일부 실시예에서, 중간 가격은 틱 단위의 가격으로 맞춰진 양자화된 중간 가격을 지칭할 수 있다. 수학적 식 2에 따라, 제1 오더북 데이터(410), 제1 오더북 데이터(420), 제3 오더북 데이터(430) 각각으로부터 양자화된 중간 가격을 산출해보면, 모두 69400원으로서, 틱 단위의 가격이 산출된다.

[0086] 일 실시예에 따르면, 정보 처리 시스템은 산출된(양자화된)중간 가격을 포함하여, 틱 단위의 연속한 가격들을 2D 데이터(412, 422, 432)의 가격 축에 포함시킬 수 있다. 예를 들어, 도시된 바와 같이, 중간 가격을 기준으로 중간 가격 -3틱부터 중간 가격 +2틱까지의 가격을 가격 축에 포함시킬 수 있다. 일 실시예에서, 매도 요청 가격과 매수 요청 가격을 균형적으로 포함시키기 위해, 중간 가격이 가격 축의 중간 부근에 위치하도록 가격 축에 포함되는 가격들을 결정할 수 있다.

[0087] 일 실시예에 따르면, 오더북 데이터(410, 420, 430)는 매수 요청 수량 및 매도 요청 수량이 0인 가격에 대한 정보는 포함하지 않을 수 있다. 예를 들어, 제1 오더북 데이터(410)는 시점 t에서 매수 요청 수량 및 매도 요청 수량이 0인 69100원, 69400원에 대한 정보는 포함하지 않으며, 제1 오더북 데이터(420)는 시점 t에서 매수 요청 수량 및 매도 요청 수량이 0인 69000원, 69300원, 69400원에 대한 정보는 포함하지 않는다. 오더북 데이터

(410, 420, 430)에는 가격 자체에 대한 정보가 포함되어 있어, 수량이 0인 가격에 대한 정보가 없더라도 오더북 데이터(410, 420, 430)로부터 스프레드 등에 대한 정보가 추출될 수 있지만, 2D 데이터(412, 422, 432)에는 가격 측에 포함된 가격 자체에 대한 정보는 포함되지 않아, 수량이 0인 가격에 대한 정보를 포함시키지 않는 경우, 스프레드 등 미래 시점의 가격 예측에 중요한 정보의 일부가 유실될 수 있다. 따라서, 본 개시의 일 실시예에 따르면, 틱 단위의 연속한 가격들을 2D 데이터(412, 422, 432)의 가격 측에 포함시킴으로써, 매도 요청 수량 및 매수 요청 수량이 0인 가격에 대한 수량 정보도 포함되어, 스프레드 등의 중요한 정보가 유실되지 않을 수 있다.

[0088] 2D 데이터(412, 422, 432)의 값은 각 시간에서의 각 가격의 매도 요청 수량 또는 매수 요청 수량으로 결정될 수 있다. 특정 시간에서의 특정 가격의 매도 요청 수량 및 매수 요청 수량이 모두 0인 경우, 해당 시간 및 해당 가격의 좌표에 대응되는 값은 0으로 결정될 수 있다. 도 4에서는 명확한 이해를 위해, 오더북 데이터(410, 420, 430)를 2D 데이터(412, 422, 432)로 변환하는 방법을 크기가 작은 데이터를 예시로, 구체적으로 설명하였다. 도 5 내지 도 6에서는 2D 데이터(또는 하나 이상의 2D 데이터 아이템을 포함하는 텐서 형태의 데이터)를 생성하는 방법의 예시를 좀 더 일반적으로 설명하고자 한다.

[0089] 도 5는 본 개시의 일 실시예에 따른 2D 데이터(510)를 생성하고, 생성된 2D 데이터를 정규화하는 방법의 예시를 나타내는 도면이다. 일 실시예에 따르면, 정보 처리 시스템은 2D 데이터(510)를 생성하기 위해 먼저 2D 데이터(510)의 시간 축을 결정할 수 있다. 예를 들어, 정보 처리 시스템은 현재 시점(예를 들어, 가장 최근 시점)인 시점 t (512)를 선택하고, 시점 t (512)를 포함하여, 일정한 시간 간격으로 연속한 하나 이상의 시점을 선택할 수 있다. 구체적 예로, 도시된 바와 같이 시점 t (512)부터 시점 $t-15$ (시점 t 로부터 15 단위 시간 전의 시점)까지의 16개의 시점을 선택할 수 있다. 그런 다음, 시간 축으로 결정된 하나 이상의 시점에서의 데이터를 포함한 오더북 데이터를 획득할 수 있다.

[0090] 그런 다음, 정보 처리 시스템은 2D 데이터(510)의 가격 측을 결정할 수 있다. 예를 들어, 정보 처리 시스템은 현재 시점(예를 들어, 가장 최근 시점)인 시점 t 에서의 중간 가격(514)을 산출할 수 있다. 여기서, 중간 가격(514)은 도 4에서 상술한 양자화된 중간 가격일 수 있다. 그 후, 산출된 중간 가격(514)을 포함하여, 틱 단위의 연속한 하나 이상의 가격을 2D 데이터의 가격 측에 포함시킬 수 있다. 구체적 예로, 도시된 바와 같이, 중간 가격-8틱으로부터 중간 가격+7틱까지의 가격을 가격 측에 포함시킬 수 있다. 일 실시예에 따르면, 매도 요청 가격과 매수 요청 가격을 균형적으로 포함시키기 위해, 중간 가격(514)이 가격 측의 중간 부근에 위치하도록 가격 측을 결정할 수 있다.

[0091] 그런 다음, 정보 처리 시스템은 시간 축에 포함된 각 시점에서의 가격 측에 포함된 각 가격의 매도 요청 수량 또는 매수 요청 수량을 2D 데이터(510)의 값으로 결정할 수 있다. 특정 시간에서의 특정 가격의 매도 요청 수량 및 매수 요청 수량이 모두 0인 경우, 해당 시간 및 해당 가격의 좌표에 대응되는 값은 0으로 결정될 수 있다. 이와 같은 방식으로 2D 데이터(510)가 생성될 수 있다.

[0092] 일 실시예에 따르면, 정보 처리 시스템은 생성된 2D 데이터(510)를 정규화할 수 있다. 예를 들어, 정보 처리 시스템은 도시된 바와 같이, 생성된 2D 데이터(510)를 min-max normalization 방식을 통해 정규화하여, 정규화된 2D 데이터(520)를 생성할 수 있다. 도 5에서는 min-max normalization 방식을 통해 정규화된 2D 데이터(520)를 예시로 도시하였지만, 이에 한정되지 않으며, mean-zero variance-one normalization 등 다양한 정규화 방식을 통해 정규화된 2D 데이터(520)를 생성할 수 있다. 본 개시의 일부 실시예에서, 2D 데이터는 정규화된 2D 데이터(520)를 지칭할 수 있다.

[0093] 도 6은 본 개시의 일 실시예에 따른 텐서 형태의 데이터를 생성하는 방법의 예시를 나타내는 도면이다. 일 실시예에 따르면, 정보 처리 시스템은 (정규화된)2D 데이터(520)를 두 개의 2D 데이터 아이템(610, 620)으로 분리할 수 있다. 예를 들어, 정보 처리 시스템은 2D 데이터(520)를 매도 요청 가격에 대한 2D 데이터(610) 및 매수 요청 가격에 대한 2D 데이터(620)로 분리할 수 있다. 정보 처리 시스템은 분리된 두 개의 2D 데이터 아이템(610, 620)을 채널 축으로 쌓아, 텐서 형태의 데이터를 생성할 수 있다. 즉, 일 실시예에 따르면 텐서 형태의 데이터는 제1 채널에 대응하는 매도 요청 가격에 대한 2D 데이터(610) 및 제2 채널에 대응하는 매수 요청 가격에 대한 2D 데이터(620)를 포함할 수 있다. 이와 같이, 매도 요청 가격에 대한 2D 데이터(610)와 매수 요청 가격에 대한 2D 데이터(620)를 별개의 채널로 분리시킴에 따라, 이를 입력받는 기계학습 모델은 매도 요청 관련 정보와 매수 요청 관련 정보를 명확하게 구분하여 학습 및 추론할 수 있어, 기계학습 모델의 성능이 향상될 수 있다.

[0094] 일 실시예에 따르면, 정보 처리 시스템은 현재 시점이 변함에 따라 오더북 데이터를 텐서 형태의 데이터로 변환

하는 과정을 복수 번 반복하여 수행할 수 있다. 이 경우, 정보 처리 시스템은 시점 t 가 변함에 따라, 시점 t 에서의 새로운 오더북 데이터를 이용하여, 중간 가격을 새롭게 산출할 수 있으며, 이에 따라 가격 속도 새롭게 결정될 수 있다.

[0095] 도 7은 본 개시의 일 실시예에 따른 텐서 형태의 데이터(710)의 예시를 나타내는 도면이다. 일 실시예에 따르면, 텐서 형태의 데이터는 제1 채널에 대응하는 매도 요청 가격에 대한 2D 데이터 및 제2 채널에 대응하는 매수 요청 가격에 대한 2D 데이터를 포함할 수 있다. 도 7에서는 텐서 형태 데이터를 이미지로 도시하고 있는데, 텐서 형태의 데이터에 포함된 각 채널을 별개의 색 채널에 대응시켜 표현한 이미지로 도시하고 있다. 제1 채널에 대응하는 매도 요청 가격에 대한 2D 데이터는 Red 채널, 제2 채널에 대응하는 매수 요청 가격에 대한 2D 데이터는 Green 채널로 표현되어, 두 개의 채널을 포함하는 텐서 형태의 데이터가 하나의 이미지로서 표현되어 있다. 또한 픽셀의 색이 어두울수록, 해당 칸의 값이 작고, 픽셀의 색이 밝을수록 해당 칸의 값이 클 수 있다.

[0096] 도시된 텐서 형태의 데이터는 각 시점에서의 각 가격의 매수 요청 수량 또는 매도 요청 수량에 대한 정보는 물론, 매수 요청 수량 및 매도 요청 수량 정보가 모두 0인 가격에 대한 수량 정보, 각 수량에 대응하는 가격이 중간 가격에서 떨어진 정도에 대한 정보까지 포함하고 있으며, 매도 요청 가격 관련 채널과 매수 요청 가격 관련 채널이 분리되어 있다. 이에 따라, 도시된 텐서 형태의 데이터는 가격 예측에 중요한 정보들을 유실 없이 포함할 수 있다. 예를 들어, 매도 요청 가격 사이의 빈 부분(712)은, 매도 요청 레벨 사이에 빈 가격이 존재한다는 정보를 포함할 수 있으며, 가장 낮은 매도 요청 레벨(ask level 1) 및/또는 가장 높은 매수 요청 레벨(bid level 1)의 가격 변화(714)는 중간 가격이 변화하였다는 정보를 포함할 수 있으며, 가장 낮은 매도 요청 레벨(ask level 1) 및/또는 가장 높은 매수 요청 레벨(bid level 1) 사이의 픽셀(716)은 스프레드에 대한 정보를 포함할 수 있다. 또한, 각 매도/매수 요청 가격의 수량 데이터의 차이(718)는 각 매수/매도 레벨 사이의 수량 차이에 대한 정보를 포함할 수 있다. 이와 같이, 본 개시에 따른 텐서 형태의 데이터는 가격 예측에 중요한 정보들을 유실 없이 포함할 수 있으며, 이에 따라, 미래 시장 상황 예측의 성능이 향상될 수 있다.

[0097] 도 8은 본 개시의 일 실시예에 따른 3개 이상의 채널을 포함하는 텐서 형태의 데이터(800)의 예시를 나타내는 도면이다. 상술한 설명에서, 특정 증권 거래소에서 거래되는 특정 종목의 매도 요청 가격에 대한 2D 데이터 및 특정 증권 거래소에서 거래되는 특정 종목의 매수 요청 가격에 대한 2D 데이터 각각에 대응되는 두 개의 채널을 포함하는 텐서 형태의 데이터를 예시로 들어 설명하였으나, 일 실시예에 따르면, 텐서 형태의 데이터(800)는 3개 이상의 채널을 포함할 수 있다. 예를 들어, 텐서 형태의 데이터(800)는 제1 증권 거래소에서 거래되는 제1 종목의 매도 요청 가격에 대한 2D 데이터(810), 제1 증권 거래소에서 거래되는 제1 종목의 매수 요청 가격에 대한 2D 데이터(820), 제2 증권 거래소에서 거래되는 제1 종목의 매도 요청 가격에 대한 2D 데이터(830) 및 제2 증권 거래소에서 거래되는 제1 종목의 매수 요청 가격에 대한 2D 데이터(840) 각각에 대응되는 4개의 채널을 포함할 수 있다. 즉, 텐서 형태의 데이터(800)는 제1 증권 거래소에 대한 데이터뿐만 아니라 하나 이상의 증권 거래소에 대한 데이터를 포함할 수 있으며, 제1 종목에 대한 데이터뿐만 아니라 제1 종목과 관련된 하나 이상의 종목에 대한 데이터를 포함할 수 있다. 여러 증권 거래소에 대한 데이터 및/또는 여러 종목에 대한 데이터를 채널 축으로 쌓는 경우, 동일한 시점에서의 동일한 매수 가격 레벨의 수량 데이터들은 텐서 형태의 데이터의 동일한 위치에 포함될 수 있으며, 이에 따라, 기계학습 모델은 이러한 정보를 취합하여 예측 결과를 추론할 수 있다.

[0098] 추가적으로 또는 대안적으로, 텐서 형태의 데이터(800)는 좌표 정보를 나타내는 하나 이상의 채널을 더 포함할 수 있다. 예를 들어, 텐서 형태의 데이터(800)는 X축에 대한 좌표를 나타내는 값을 포함하는 2D 데이터 및 Y축에 대한 좌표를 나타내는 값을 포함하는 2D 데이터 각각에 대응되는 2개의 채널을 더 포함할 수 있다. 여기서, X축에 대한 좌표를 나타내는 값을 포함하는 2D 데이터는 X축 좌표에 따라서만 값이 달라지며, Y축에 대한 좌표를 나타내는 값을 포함하는 2D 데이터는 Y축 좌표에 따라서만 값이 달라질 수 있다. 이에 따라, 좌표 정보를 나타내는 하나 이상의 채널을 포함하는 텐서 형태의 데이터를 입력받은 기계학습 모델은 좌표에 대한 정보를 학습 및 추론에 추가적으로 이용함으로써, 각 수량 데이터 값의 위치에 대한 정확한 정보를 토대로, 정확한 예측 결과를 출력할 수 있다.

[0099] 도 9는 본 개시의 일 실시예에 따른 기계학습 모델의 해석을 제공하는 방법의 예시를 나타내는 도면이다. 일 실시예에 따르면, 정보 처리 시스템은 텐서 형태의 입력 데이터(910)를 복수의 레이어를 포함하는 기계학습 모델에 입력하여, 예측 결과를 출력할 수 있다. 여기서, 예측 결과는 미래 특정 시점에서의 특정 종목의 가격이 현재 시점의 가격보다 높을 것을 나타내는 up 클래스, 미래 특정 시점에서의 특정 종목의 가격이 현재 시점의 가격과 동일할 것을 나타내는 stationary 클래스 및 미래 특정 시점에서의 특정 종목의 가격이 현재 시점의 가

격보다 낮을 것을 나타내는 up 클래스를 포함할 수 있다.

- [0100] 정보 처리 시스템은 기계학습 모델에 의해 출력된 예측 결과에 대한 해석을 제공할 수 있다. 예를 들어, 정보 처리 시스템은 CAM(Class Activation Map) 기법 또는 Grad-CAM 기법 등을 이용하여, 기계학습 모델이 예측 결과를 출력하는데 영향을 미친 픽셀의 위치에 대한 정보를 포함하는 해석 맵(map)(920)을 제공할 수 있다. 일 실시예에 따르면, 해석 맵(920)은 2D 이미지로 표현될 수 있다. 예를 들어, 해석 맵(920)은 해석 맵(920)에 포함된 픽셀의 각각에, 예측 결과를 출력하는데 영향을 미친 정도에 따라 색상이 표시되는 히트맵으로 표현될 수 있다. 구체적 예로, 특정 픽셀이 예측 결과를 출력하는데 영향을 미친 정도가 클수록 해당 픽셀이 빨간색으로 표시되고, 특정 픽셀이 예측 결과를 출력하는데 영향을 미친 정도가 작을수록 해당 픽셀이 파란색으로 표시되고, 영향을 미친 정도에 따라 빨간색과 파란색 사이의 연속된 색 중 하나로 표시되는 연속적인 히트맵으로 해석 맵(920)이 표현될 수 있다. 다른 구체적 예로, 특정 픽셀이 예측 결과를 출력하는데 영향을 미친 정도 또는 순위가 속하는 범위에 따라 불연속적인 색상으로 표시되는 불연속적인 히트맵으로 해석 맵(920)이 표현될 수 있다.
- [0101] 또한, 일 실시예에 따르면, 정보 처리 시스템은 해석 맵(920)을 입력 데이터(910)에 적용하여, 예측 결과 출력에 영향을 미친 픽셀의 입력 데이터(910)에서의 위치를 파악하기 쉽도록, 해석 맵이 적용된 입력 데이터(940)를 제공할 수 있다. 예를 들어, 정보 처리 시스템은 입력 데이터(910)에 해석 맵(920)을 오버랩시킴으로써, 해석 맵이 적용된 입력 데이터(940)를 생성하여 제공할 수 있다. 다른 예로, 정보 처리 시스템은 입력 데이터 중 예측 결과 출력에 가장 많은 영향을 미친 n개의 지점에 대한 정보를 제공할 수 있다.
- [0102] 일반적으로, 복수의 레이어를 거친 해석 맵(920)의 크기(x축 및 y축의 크기)는 입력 데이터의 크기보다 작기 때문에, 이를 입력 데이터(910)에 적용하기 위해서는 해석 맵(920)의 크기를 입력 데이터(910)의 크기와 동일하도록 조정하여야 한다. 크기가 조정된 해석 맵(930)은 예측 결과를 출력하는데 영향을 미친 픽셀의 위치에 대한 정확한 정보를 포함하지 못할 수 있으며, 대략적인 위치에 대한 정보만을 포함할 수 있다는 문제점이 있다.
- [0103] 또한, 입력 데이터(910)는 기계학습 모델을 위한 하나 이상의 2D 데이터 아이템을 포함하는 텐서 형태의 데이터(예를 들어, 2D 이미지)일 수 있는데, 이러한 데이터에는 축에 대한 정보는 전혀 포함되어 있지 않아, 해석 맵이 적용된 입력 데이터(940)를 사람이 보는 경우, 어떠한 정보가 예측 결과의 출력에 영향을 미친 것인지를 파악하기 쉽지 않을 수 있다.
- [0104] 도 10은 본 개시의 일 실시예에 따른 입력 데이터에 포함된 2D 데이터와 동일한 크기를 갖는 해석 맵을 제공하기 위한 기계학습 모델의 예시를 나타내는 도면이다. 도 10에는 기계학습 모델의 예시로 CNN 모델이 도시되어 있다. 기계학습 모델에 포함된 복수의 레이어는 특징 맵의 크기를 줄이도록 구성된 레이어를 포함할 수 있다. 예를 들어, CNN 모델에 포함된 풀링 레이어(Pooling layer)(예: 도 10의 Maxpool), 컨볼루션 필터 적용시 2 픽셀 이상씩 이동하면서 연산을 수행하도록 설정된 스트라이드 레이어(Stride layer) 등이 특징 맵의 크기를 줄일 수 있다. 이에 따라, GAP(Global Average Pooling) 이전의 복수의 레이어 중 마지막 레이어로부터의 출력인 특징 맵(1010)의 크기가 입력 특징 맵(1000)의 크기보다 작을 수 있다. 또한, 해석 맵의 크기 역시 입력 특징 맵(1000)의 크기보다 작을 수 있다.
- [0105] 이와 같은 문제점을 해결하기 위해, 본 개시의 일 실시예에 따르면, 정보 처리 시스템은 특징 맵의 크기를 줄이도록 구성된 모든 레이어가 제거된 기계학습 모델을 이용할 수 있다. 특징 맵의 크기를 줄이도록 구성된 모든 레이어가 제거된 기계학습 모델을 이용하는 경우, GAP 이전 마지막 레이어로부터의 출력 특징 맵(1020)의 크기는 입력 특징 맵(1000)의 크기와 동일할 수 있으며, 이에 따라, 해석 맵의 크기 역시 입력 특징 맵(1000)의 크기와 동일할 수 있다.
- [0106] 도 11은 본 개시의 일 실시예에 따른 입력 데이터에 포함된 2D 데이터와 동일한 크기를 갖는 해석 맵(1130)을 제공하는 방법의 예시를 나타내는 도면이다. 일 실시예에 따르면, 정보 처리 시스템은 테이블 형태의 데이터(1110)를 하나 이상의 채널의 각각에 대응하는 2D 데이터를 포함하는 텐서 형태의 데이터(1120)로 변환하고, 변환된 텐서 형태의 데이터(1120)를 복수의 레이어를 포함한 기계학습 모델에 입력하여 예측 결과를 출력할 수 있다. 여기서, 기계학습 모델은, 특징 맵의 크기를 줄이도록 구성된 레이어가 모두 제거된 모델일 수 있다. 즉, 기계학습 모델에서 특징 맵의 크기를 줄이도록 구성된 레이어가 모두 제거되어, 각 레이어로부터의 출력인 특징 맵의 크기는 텐서 형태의 데이터(1120)에 포함된 2D 데이터의 크기와 동일할 수 있다.
- [0107] 일 실시예에 따르면, 정보 처리 시스템은 기계학습 모델이 예측 결과를 출력하는데 영향을 미친 픽셀의 위치에 대한 정보를 포함하는 해석 맵(1130)을 생성할 수 있다. 예를 들어, 정보 처리 시스템은 복수의 레이어 중 적

어도 하나의 레이어로부터의 출력인 특징 맵을 이용하여 해석 맵(1130)을 생성할 수 있다. 여기서, 각 레이어로부터 출력인 특징 맵의 크기는 입력 데이터(1120)에 포함된 2D 데이터의 크기와 동일하기 때문에, 이를 기초로 생성된 해석 맵(1130)의 크기 역시 입력 데이터(1120)에 포함된 2D 데이터의 크기와 동일할 수 있다. 따라서, 해석 맵(1130)의 크기를 입력 데이터(1120)에 포함된 2D 데이터의 크기와 동일하게 조정할 필요가 없으며, 예측 결과를 출력하는데 영향을 미친 픽셀의 대략적인 위치가 아닌 정확한 위치에 대한 정보를 포함하는 해석 맵(1130)이 생성될 수 있다.

[0108] 추가적으로, 정보 처리 시스템은 테이블 형태의 데이터(1110) 및 텐서 형태의 데이터(1120)를 기초로 설명가능한(explainable) 데이터(1140)를 생성할 수 있다. 설명가능한 데이터는, 사람이 해당 데이터를 봤을 때, 직관적인 이해 및/또는 해석이 가능한 데이터를 지칭할 수 있다. 예를 들어, 오더북 데이터를 기초로 변환된 텐서 형태의 데이터(1120)에 대해, 2D 데이터의 구조 및 크기를 그대로 유지하면서, 2D 데이터의 축에 대한 정보(예를 들어, 시간 축, 가격 축에 대한 정보)를 추가하고, 각 픽셀 값을 원래 데이터의 값(예를 들어, 정규화되지 않은 수량 정보)으로 변경함으로써, 설명가능한 데이터를 생성할 수 있다.

[0109] 그런 다음, 정보 처리 시스템은 설명가능한 데이터(1140)에 해석 맵(1130)을 적용할 수 있다. 예를 들어, 정보 처리 시스템은 설명 가능한 데이터(1140)에 해석 맵(1130)을 오버랩시킴으로써, 해석 맵이 적용된 설명가능한 데이터(1150)를 생성할 수 있다. 이와 같은 구성에 의해, 예측 결과 출력에 영향을 준 데이터를 정확하고 직관적으로 파악할 수 있다.

[0110] 도 12는 본 개시의 일 실시예에 따른 오더북 데이터를 기계학습 모델을 위한 2D 데이터로 변환하는 방법(1200)의 예시를 나타내는 흐름도이다. 일 실시예에 따르면, 방법(1200)은 프로세서가 미리 결정된 기간 이전의 과거 시점으로부터 현재 시점까지의 시간에 따른, 제1 증권 거래소에서 거래되는 종목에 대한 복수의 매수 요청 가격에 대한 데이터 아이템 및 복수의 매도 요청 가격에 대한 데이터 아이템을 포함한 오더북 데이터를 획득함으로써 개시될 수 있다(S1210).

[0111] 그런 다음, 프로세서는 복수의 매수 요청 가격 중에서 가장 높은 매수 요청 가격과 상기 복수의 매도 요청 가격 중에서 가장 낮은 매도 요청 가격을 기초로 중간 가격을 산출할 수 있다(S1220).

[0112] 그런 다음, 프로세서는 일정한 시간 간격의 시간을 제1 축으로 하고, 틱 단위의 가격을 제2 축으로 하고, 상기 복수의 매도 요청 가격의 각각의 수량에 대한 데이터 또는 상기 복수의 매수 요청 가격의 각각의 수량에 대한 데이터를 상기 제1 축의 시간과 상기 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 대한 값으로 가진 하나 이상의 2D 데이터 아이템을 포함한 텐서 형태의 데이터를 생성할 수 있다(S1230). 여기서, 제2 축에 포함되는 틱 단위의 가격은 산출된 중간 가격을 포함할 수 있다. 일 실시예에 따르면, 프로세서는 복수의 매도 요청 가격의 각각의 수량에 대한 데이터 또는 상기 복수의 매수 요청 가격의 각각의 수량에 대한 데이터를 정규화하여 제1 축의 시간과 제2 축의 가격에 따라 정의되는 복수의 좌표의 각각에 연관시킬 수 있다.

[0113] 일 실시예에 따르면, 텐서 형태의 데이터는 복수의 채널의 각각에 대응하는 2D 데이터를 포함할 수 있으며, 예를 들어, 제1 채널에 대응하는, 시간에 따른 제1 증권 거래소에서 거래되는 종목의 매수 요청 가격에 대한 2D 데이터 및 제2 채널에 대응하는, 시간에 따른 제1 증권 거래소에서 거래되는 종목의 매도 요청 가격에 대한 2D 데이터를 포함할 수 있다. 추가적으로, 텐서 형태의 데이터는, 제3 채널에 대응하는, 시간에 따른 제2 증권 거래소에서 거래되는 종목의 매수 요청 가격에 대한 2D 데이터 및 제4 채널에 대응하는, 시간에 따른 제2 증권 거래소에서 거래되는 종목의 매도 요청 가격에 대한 2D 데이터를 더 포함할 수 있다.

[0114] 추가적으로 또는 대안적으로, 텐서 형태의 데이터는 하나 이상의 2D 데이터의 각각에서의 제1 축에 대한 좌표를 나타내는 값을 포함한 2D 데이터 및 하나 이상의 2D 데이터의 각각에서의 제2 축에 대한 좌표를 나타내는 값을 포함한 2D 데이터를 더 포함할 수 있다.

[0115] 그 후, 프로세서는 생성된 텐서 형태의 데이터를 복수의 레이어를 포함한 기계학습 모델에 입력하여 예측 결과를 출력할 수 있다. 여기서, 예측 결과는, 현재 시점으로부터 미리 결정된 기간 이후의 미래 시점에서의 종목의 가격에 대한 예측 정보를 포함할 수 있다. 예를 들어, 미래 시점에서의 종목의 가격에 대한 예측 정보는 미래 시점에서의 특정 종목의 가격이 현재 시점의 가격보다 높을 것을 나타내는 제1 클래스, 미래 시점에서의 특정 종목의 가격이 현재 시점의 가격과 동일하다는 것을 나타내는 제2 클래스 및 미래 시점에서의 특정 종목의 가격이 현재 시점의 가격보다 낮을 것을 나타내는 제3 클래스를 포함할 수 있다.

[0116] 일 실시예에 따르면, 프로세서는 현재 시점이 변함에 따라, 변화된 현재 시점으로부터 미리 결정된 기간 이전의 과거 시점까지의 오더북 데이터에 대해 상술한 과정 중 적어도 일부의 과정을 반복적으로 수행할 수 있다.

- [0117] 도 13은 본 개시의 일 실시예에 따른 기계학습 모델의 해석을 제공하는 방법(1300)의 예시를 나타내는 흐름도이다. 일 실시예에 따르면, 프로세서(예를 들어, 정보 처리 시스템의 적어도 하나의 프로세서)가 테이블 형태의 데이터를 하나 이상의 채널의 각각에 대응하는 2D(dimensional) 데이터를 포함하는 텐서 형태의 데이터로 변환함으로써 방법(1300)이 개시될 수 있다(S1310). 일 실시예에서 테이블 형태의 데이터는, 시간에 따른 증권 거래소의 종목의 매수 요청 가격 및 매도 요청 가격에 대한 정보를 포함할 수 있다.
- [0118] 일 실시예에 따르면, 텐서 형태의 데이터는, 제1 채널에 대응하는 시간에 따른 제1 증권 거래소에서 거래되는 종목의 매수 요청 가격에 대한 2D 데이터 및 제2 채널에 대응하는 시간에 따른 제1 증권 거래소에서 거래되는 종목의 매도 요청 가격에 대한 2D 데이터를 포함할 수 있다.
- [0119] 추가적으로, 텐서 형태의 데이터는, 제3 채널에 대응하는 시간에 따른 제2 증권 거래소에서 거래되는 종목의 매수 요청 가격에 대한 2D 데이터 및 제4 채널에 대응하는 시간에 따른 제2 증권 거래소에서 거래되는 종목의 매도 요청 가격에 대한 2D 데이터를 더 포함할 수 있다.
- [0120] 그런 다음, 프로세서는 변환된 텐서 형태의 데이터를 복수의 레이어를 포함한 기계학습 모델에 입력하여 예측 결과를 출력할 수 있다(S1320). 일 실시예에 따르면, 예측 결과는 현재 시점으로부터 미리 결정된 기간 이후의 미래 시점에서의 특정 종목의 가격에 대한 예측 정보를 포함할 수 있다. 구체적인 예로, 특정 종목의 가격에 대한 예측 정보는 미래 시점에서의 특정 종목의 가격이 현재 시점의 가격보다 높을 것을 나타내는 제1 클래스, 미래 시점에서의 특정 종목의 가격이 현재 시점의 가격과 동일하다는 것을 나타내는 제2 클래스 및 미래 시점에서의 특정 종목의 가격이 현재 시점의 가격보다 낮을 것을 나타내는 제3 클래스를 포함할 수 있다.
- [0121] 그런 다음, 프로세서는 복수의 레이어 중 적어도 하나의 레이어로부터의 출력인 특징 맵을 이용하여 상기 예측 결과에 대한 해석 맵을 생성할 수 있다(S1330). 예를 들어, 프로세서는 기계학습 모델의 마지막 레이어로부터의 출력인 특징 맵을 이용하여 예측 결과에 대한 해석 맵을 생성할 수 있다. 여기서, 기계학습 모델의 마지막 레이어로부터의 출력인 특징 맵 및/또는 해석 맵은 2D 데이터를 포함할 수 있으며, 기계학습 모델은 기계학습의 마지막 레이어로부터의 출력인 특징 맵의 크기 및/또는 해석 맵의 크기가 텐서 형태의 데이터에 포함된 2D 데이터의 크기와 동일한 크기를 갖도록 구성될 수 있다.
- [0122] 일 실시예에 따르면, 예측 결과에 대한 해석 맵은 2D 이미지로 표현될 수 있으며, 해석 맵의 2D 이미지는 예측 결과를 출력하는데 영향을 미친 해석 맵의 2D 이미지에 포함된 픽셀들에 대한 정보를 포함할 수 있다. 예를 들어, 해석 맵의 2D 이미지는 이미지에 포함된 픽셀의 각각에, 예측 결과를 출력하는데 영향을 미친 정도에 따라 표현된 색상이 표시되는 히트맵을 포함할 수 있다.
- [0123] 일 실시예에 따르면, 프로세서는 테이블 형태의 데이터 및 상기 변환된 텐서 형태의 데이터를 기초로 설명가능한 데이터를 생성할 수 있다(S1340). 여기서, 설명가능한 데이터는 사람이 해당 데이터를 봤을 때, 직관적으로 이해 및/또는 해석될 수 있는 데이터를 지칭할 수 있다.
- [0124] 프로세서는 설명가능한 데이터에 해석 맵을 적용할 수 있다(S1350). 예를 들어, 프로세서는 설명가능한 데이터에 히트맵을 오버랩시킴으로써, 설명가능한 데이터에 해석 맵을 적용할 수 있다.
- [0125] 도 12 및 도 13의 흐름도 및 상술한 설명은 일 예시일 뿐이며, 다른 실시예에서는 다양한 방식으로 구현될 수 있다. 예를 들어, 각 단계의 순서가 변경되거나, 하나 이상의 단계가 추가되거나, 하나 이상의 단계가 생략될 수 있다. 다른 예로, 하나 이상의 단계가 다른 구성에 의해 수행될 수 있다.
- [0126] 상술한 방법은 컴퓨터에서 실행하기 위해 컴퓨터 판독 가능한 기록매체에 저장된 컴퓨터 프로그램으로 제공될 수 있다. 매체는 컴퓨터로 실행 가능한 프로그램을 계속 저장하거나, 실행 또는 다운로드를 위해 임시 저장하는 것일 수도 있다. 또한, 매체는 단일 또는 수개 하드웨어가 결합된 형태의 다양한 기록 수단 또는 저장수단일 수 있는데, 어떤 컴퓨터 시스템에 직접 접속되는 매체에 한정되지 않고, 네트워크 상에 분산 존재하는 것일 수도 있다. 매체의 예시로는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM 및 DVD 와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto optical medium), 및 ROM, RAM, 플래시 메모리 등을 포함하여 프로그램 명령어가 저장되도록 구성된 것이 있을 수 있다. 또한, 다른 매체의 예시로, 애플리케이션을 유통하는 앱 스토어나 기타 다양한 소프트웨어를 공급 내지 유통하는 사이트, 서버 등에서 관리하는 기록매체 내지 저장매체도 들 수 있다.
- [0127] 본 개시의 방법, 동작 또는 기법들은 다양한 수단에 의해 구현될 수도 있다. 예를 들어, 이러한 기법들은 하드웨어, 펌웨어, 소프트웨어, 또는 이들의 조합으로 구현될 수도 있다. 본원의 개시와 연계하여 설명된 다양한

예시적인 논리적 블록들, 모듈들, 회로들, 및 알고리즘 단계들은 전자 하드웨어, 컴퓨터 소프트웨어, 또는 양자의 조합들로 구현될 수도 있음을 통상의 기술자들은 이해할 것이다. 하드웨어 및 소프트웨어의 이러한 상호 대체를 명확하게 설명하기 위해, 다양한 예시적인 구성요소들, 블록들, 모듈들, 회로들, 및 단계들이 그들의 기능적 관점에서 일반적으로 위에서 설명되었다. 그러한 기능이 하드웨어로서 구현되는지 또는 소프트웨어로서 구현되는지의 여부는, 특정 애플리케이션 및 전체 시스템에 부과되는 설계 요구사항들에 따라 달라진다. 통상의 기술자들은 각각의 특정 애플리케이션을 위해 다양한 방식으로 설명된 기능을 구현할 수도 있으나, 그러한 구현들은 본 개시의 범위로부터 벗어나게 하는 것으로 해석되어서는 안된다.

[0128] 하드웨어 구현에서, 기법들을 수행하는 데 이용되는 프로세싱 유닛들은, 하나 이상의 ASIC들, DSP들, 디지털 신호 프로세싱 디바이스들(digital signal processing devices; DSPD들), 프로그램가능 논리 디바이스들(programmable logic devices; PLD들), 필드 프로그램가능 게이트 어레이들(field programmable gate arrays; FPGA들), 프로세서들, 제어기들, 마이크로제어기들, 마이크로프로세서들, 전자 디바이스들, 본 개시에 설명된 기능들을 수행하도록 설계된 다른 전자 유닛들, 컴퓨터, 또는 이들의 조합 내에서 구현될 수도 있다.

[0129] 따라서, 본 개시와 연계하여 설명된 다양한 예시적인 논리 블록들, 모듈들, 및 회로들은 범용 프로세서, DSP, ASIC, FPGA나 다른 프로그램 가능 논리 디바이스, 이산 게이트나 트랜지스터 로직, 이산 하드웨어 컴포넌트들, 또는 본원에 설명된 기능들을 수행하도록 설계된 것들의 임의의 조합으로 구현되거나 수행될 수도 있다. 범용 프로세서는 마이크로프로세서일 수도 있지만, 대안으로, 프로세서는 임의의 종래의 프로세서, 제어기, 마이크로제어기, 또는 상태 머신일 수도 있다. 프로세서는 또한, 컴퓨팅 디바이스들의 조합, 예를 들면, DSP와 마이크로프로세서, 복수의 마이크로프로세서들, DSP 코어와 연계한 하나 이상의 마이크로프로세서들, 또는 임의의 다른 구성의 조합으로서 구현될 수도 있다.

[0130] 펌웨어 및/또는 소프트웨어 구현에 있어서, 기법들은 랜덤 액세스 메모리(random access memory; RAM), 판독 전용 메모리(read-only memory; ROM), 비휘발성 RAM(non-volatile random access memory; NVRAM), PROM(programmable read-only memory), EPROM(erasable programmable read-only memory), EEPROM(electrically erasable PROM), 플래시 메모리, 콤팩트 디스크(compact disc; CD), 자기 또는 광학 데이터 스토리지 디바이스 등과 같은 컴퓨터 판독가능 매체 상에 저장된 명령들로서 구현될 수도 있다. 명령들은 하나 이상의 프로세서들에 의해 실행 가능할 수도 있고, 프로세서(들)로 하여금 본 개시에 설명된 기능의 특정 양태들을 수행하게 할 수도 있다.

[0131] 소프트웨어로 구현되는 경우, 상기 기법들은 하나 이상의 명령들 또는 코드로서 컴퓨터 판독 가능한 매체 상에 저장되거나 또는 컴퓨터 판독 가능한 매체를 통해 전송될 수도 있다. 컴퓨터 판독가능 매체들은 한 장소에서 다른 장소로 컴퓨터 프로그램의 전송을 용이하게 하는 임의의 매체를 포함하여 컴퓨터 저장 매체들 및 통신 매체들 양자를 포함한다. 저장 매체들은 컴퓨터에 의해 액세스될 수 있는 임의의 이용 가능한 매체들일 수도 있다. 비제한적인 예로서, 이러한 컴퓨터 판독가능 매체는 RAM, ROM, EEPROM, CD-ROM 또는 다른 광학 디스크 스토리지, 자기 디스크 스토리지 또는 다른 자기 스토리지 디바이스들, 또는 소망의 프로그램 코드를 명령들 또는 데이터 구조들의 형태로 이송 또는 저장하기 위해 사용될 수 있으며 컴퓨터에 의해 액세스될 수 있는 임의의 다른 매체를 포함할 수 있다. 또한, 임의의 접속이 컴퓨터 판독가능 매체로 적절히 칭해진다.

[0132] 본원에서 사용된 디스크(disk) 와 디스크(disc)는, CD, 레이저 디스크, 광 디스크, DVD(digital versatile disc), 플로피디스크, 및 블루레이 디스크를 포함하며, 여기서 디스크들(disks)은 보통 자기적으로 데이터를 재생하고, 반면 디스크들(disks)은 레이저를 이용하여 광학적으로 데이터를 재생한다. 위의 조합들도 컴퓨터 판독가능 매체들의 범위 내에 포함되어야 한다.

[0133] 소프트웨어 모듈은, RAM 메모리, 플래시 메모리, ROM 메모리, EPROM 메모리, EEPROM 메모리, 레지스터들, 하드 디스크, 이동식 디스크, CD-ROM, 또는 공지된 임의의 다른 형태의 저장 매체 내에 상주할 수도 있다. 예시적인 저장 매체는, 프로세서가 저장 매체로부터 정보를 판독하거나 저장 매체에 정보를 기록할 수 있도록, 프로세서에 연결될 수 있다. 대안으로, 저장 매체는 프로세서에 통합될 수도 있다. 프로세서와 저장 매체는 ASIC 내에 존재할 수도 있다. ASIC은 유저 단말 내에 존재할 수도 있다. 대안으로, 프로세서와 저장 매체는 유저 단말에서 개별 구성요소들로서 존재할 수도 있다.

[0134] 이상 설명된 실시예들이 하나 이상의 독립형 컴퓨터 시스템에서 현재 개시된 주제의 양태들을 활용하는 것으로 기술되었으나, 본 개시는 이에 한정되지 않고, 네트워크나 분산 컴퓨팅 환경과 같은 임의의 컴퓨팅 환경과 연계하여 구현될 수도 있다. 또 나아가, 본 개시에서 주제의 양상들은 복수의 프로세싱 칩들이나 장치들에서 구현될 수도 있고, 스토리지는 복수의 장치들에 걸쳐 유사하게 영향을 받게 될 수도 있다. 이러한 장치들은 PC들,

네트워크 서버들, 및 휴대용 장치들을 포함할 수도 있다.

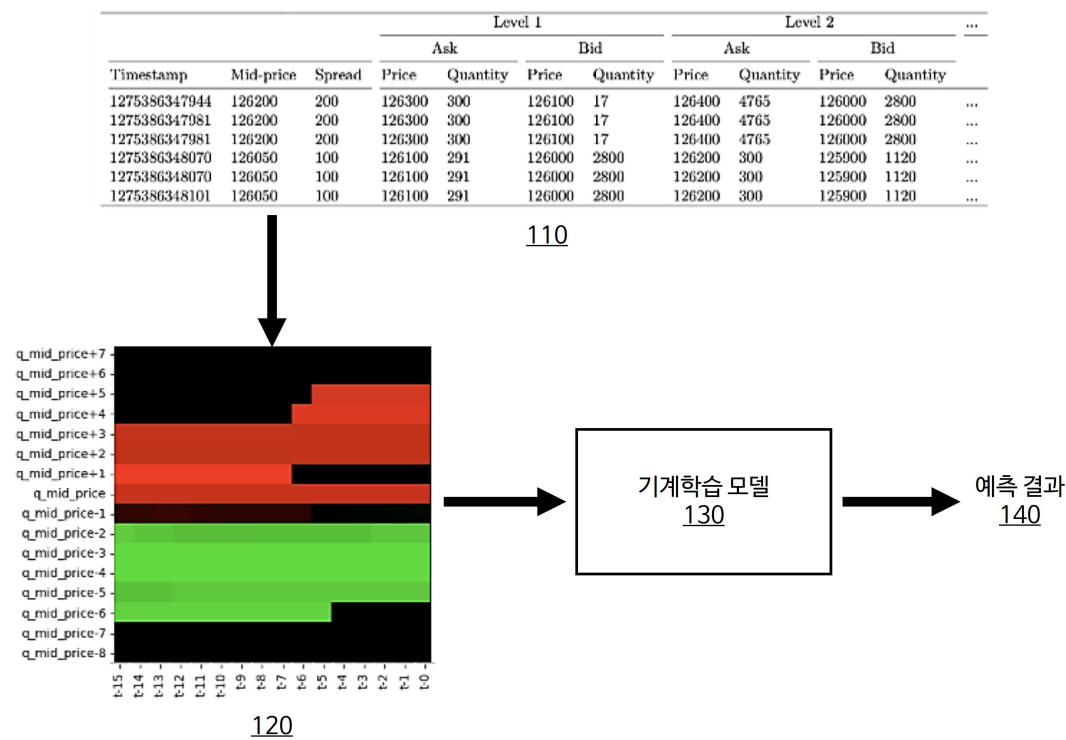
[0135] 본 명세서에서는 본 개시가 일부 실시예들과 관련하여 설명되었지만, 본 개시의 발명이 속하는 기술분야의 통상의 기술자가 이해할 수 있는 본 개시의 범위를 벗어나지 않는 범위에서 다양한 변형 및 변경이 이루어질 수 있다. 또한, 그러한 변형 및 변경은 본 명세서에 첨부된 특허청구의 범위 내에 속하는 것으로 생각되어야 한다.

부호의 설명

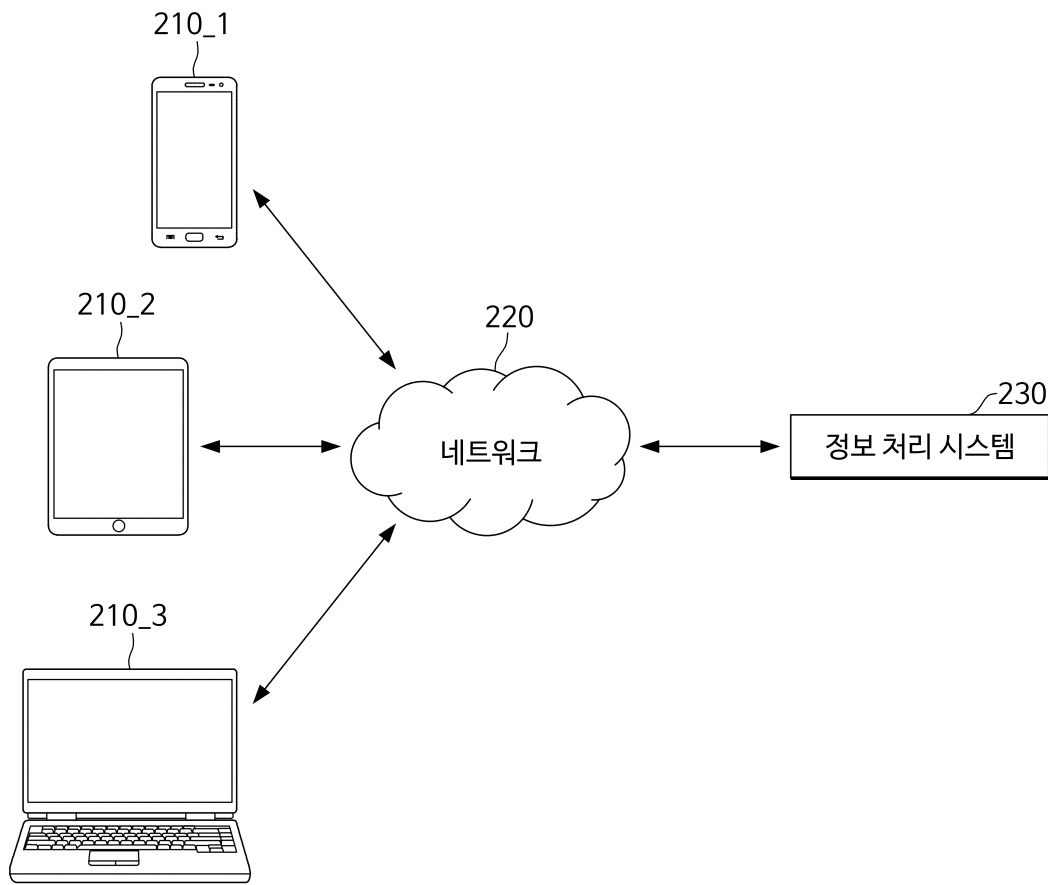
- [0136] 110: 오더북 데이터
- 120: 텐서 형태의 데이터
- 130: 기계학습 모델
- 140: 예측 결과

도면

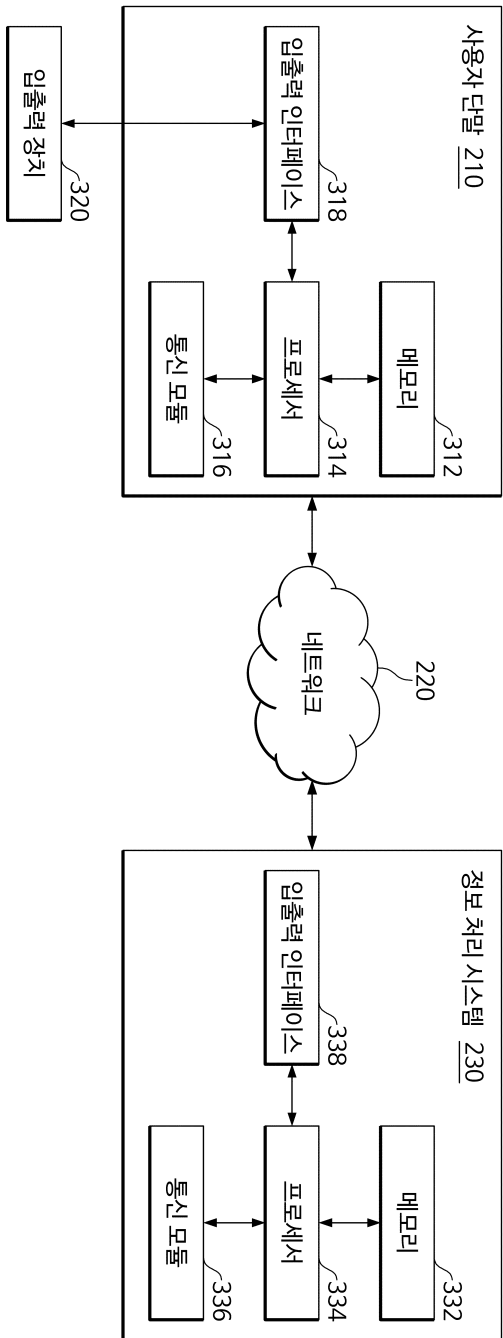
도면1



도면2



도면3



도면4

	Quantity(가격)	Price(수량)
Ask level 3	31	69700
Ask level 2	24	69600
Ask level 1	6	69500
Bid level 1	41	69300
Bid level 2	47	69200
Bid level 3	43	69000

410

		t
Q_mid_price+2tick	(69600)	24
Q_mid_price+1tick	(69500)	6
Q_mid_price	(69400)	0
Q_mid_price-1tick	(69300)	41
Q_mid_price-2tick	(69200)	47
Q_mid_price-3tick	(69100)	0

412

	Quantity(가격)	Price(수량)
Ask level 3	31	69700
Ask level 2	24	69600
Ask level 1	6	69500
Bid level 1	41	69200
Bid level 2	47	69100
Bid level 3	43	68900

420

		t
Q_mid_price+2tick	(69600)	24
Q_mid_price+1tick	(69500)	6
Q_mid_price	(69400)	0
Q_mid_price-1tick	(69300)	0
Q_mid_price-2tick	(69200)	41
Q_mid_price-3tick	(69100)	47

422

	Quantity	Price
Ask level 3	25	69700
Ask level 2	21	69600
Ask level 1	3	69400
Bid level 1	31	69300
Bid level 2	37	69200
Bid level 3	39	68900

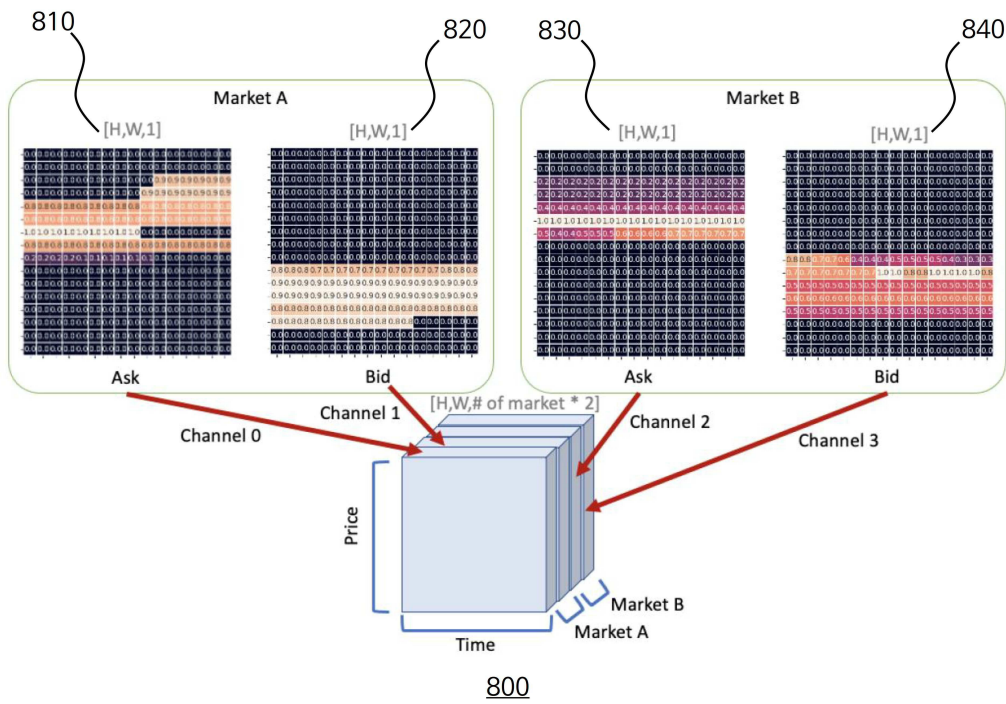
430

	Quantity	Price
Ask level 3	31	69700
Ask level 2	24	69600
Ask level 1	6	69500
Bid level 1	41	69200
Bid level 2	47	69100
Bid level 3	43	68900

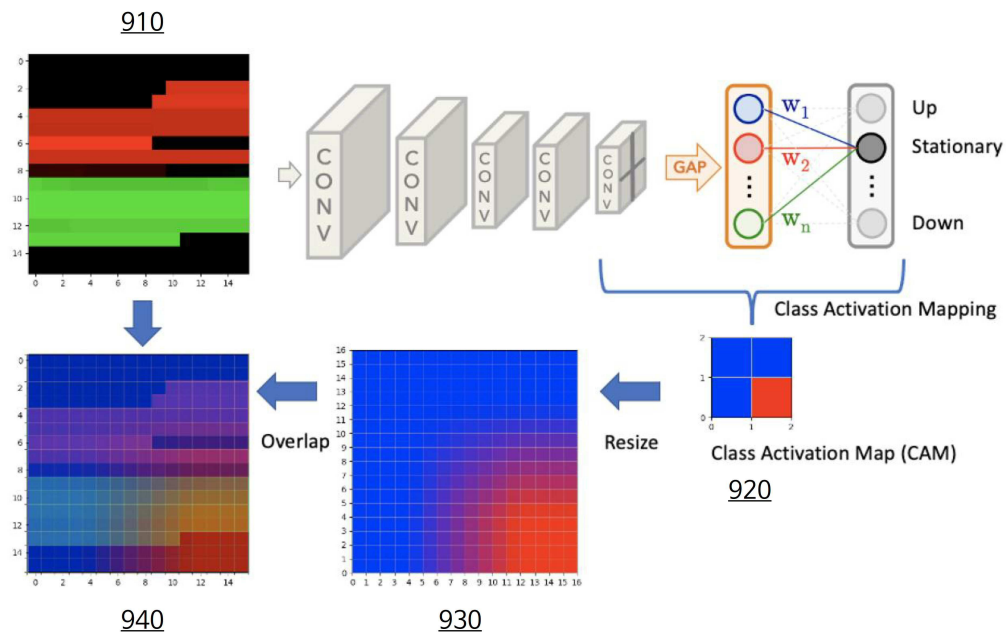
		t-1δ	t
Q_mid_price+2tick	(69600)	21	24
Q_mid_price+1tick	(69500)	0	6
Q_mid_price	(69400)	3	0
Q_mid_price-1tick	(69300)	31	0
Q_mid_price-2tick	(69200)	37	41
Q_mid_price-3tick	(69100)	0	47

432

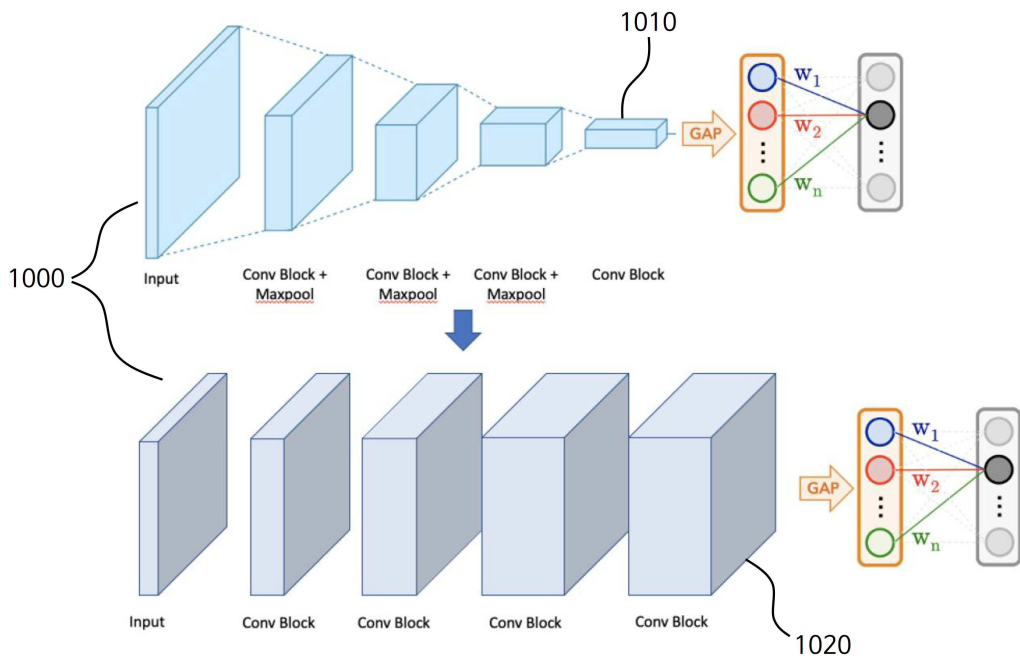
도면8



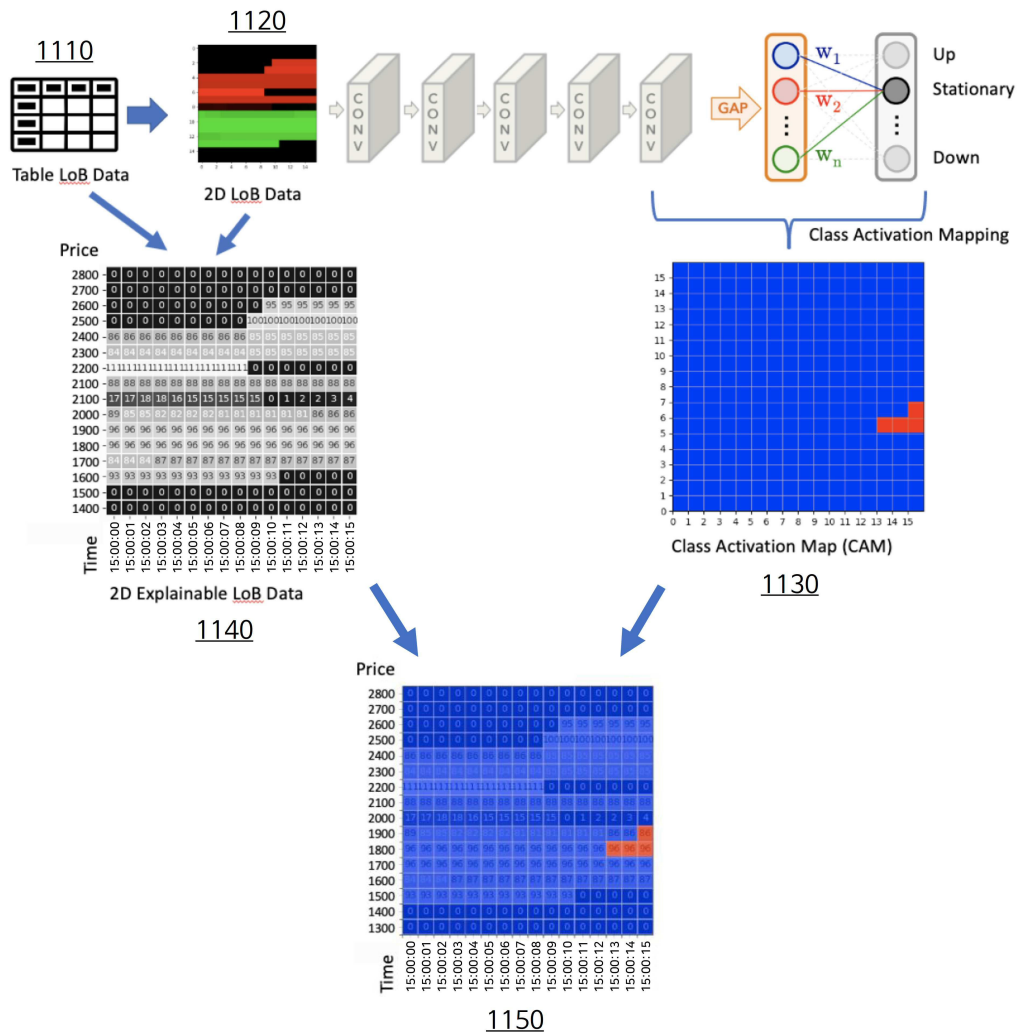
도면9



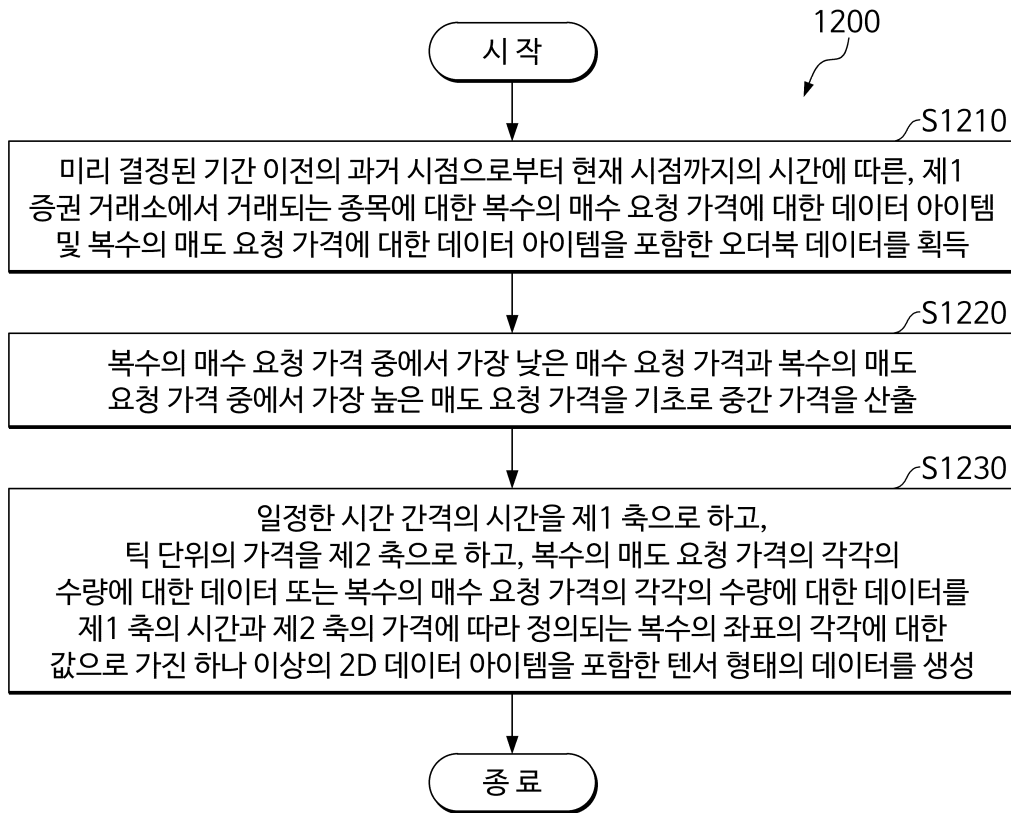
도면10



도면11



도면12



도면13

