



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0098423
(43) 공개일자 2022년07월12일

(51) 국제특허분류(Int. Cl.)
G06Q 40/04 (2012.01) G06N 3/08 (2006.01)
G06T 5/00 (2019.01) G06T 5/40 (2006.01)
(52) CPC특허분류
G06Q 40/04 (2013.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2021-0000130
(22) 출원일자 2021년01월04일
심사청구일자 2021년01월04일

(71) 출원인
연세대학교 원주산학협력단
강원도 원주시 흥업면 연세대길 1
(72) 발명자
안재준
서울특별시 동작구 보라매로5가길 7 캐릭터그린빌
1903호
임진영
강원도 원주시 흥업면 복원로 1412-11 휴먼스빌
305호
(뒷면에 계속)
(74) 대리인
오영진

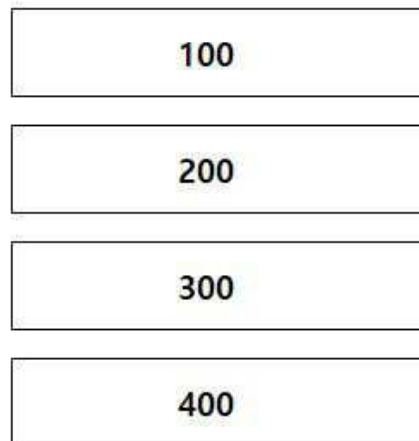
전체 청구항 수 : 총 5 항

(54) 발명의 명칭 객체탐지모형을 적용한 주가동락 예측장치

(57) 요약

본 발명의 일 실시예에 따른 객체탐지모형을 적용한 주가동락 예측장치는, 미리 설정된 기간 동안의 주가정보를 수집하는 수집모듈; 상기 주가정보를 히트맵 형식의 2D 이미지로 변환하고, 상기 2D 이미지에 라벨정보를 추가하여 적어도 하나의 입력이미지를 생성하는 입력이미지 생성모듈; 객체탐지모형 내에 상기 입력이미지를 입력하여 상기 객체탐지모형을 학습시키는 학습모듈; 및 학습된 상기 객체탐지모형을 이용하여 현재 시점에서 미리 설정된 기간 이후의 주가의 상승여부를 판단하는 주가예측모듈;을 포함한다.

대표도 - 도2



(52) CPC특허분류

G06T 5/007 (2013.01)

G06T 5/40 (2013.01)

(72) 발명자

장인호

강원도 원주시 예술관길 15 원주 삼성아파트 3동 303호

최세환

서울특별시 동대문구 한천로8가길 17-1

박성중

경기도 고양시 일산동구 산두로 54 정발마을3단지 건영빌라 302동 104호

이 발명을 지원한 국가연구개발사업

과제고유번호	1711118846
과제번호	2020-51-0194
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	정형 시계열데이터 분석을 위한 합성곱 기반의 새로운 딥러닝 모형 개발
기 여 율	1/1
과제수행기관명	연세대학교 원주산학협력단
연구기간	2020.06.01 ~ 2021.02.28

명세서

청구범위

청구항 1

미리 설정된 기간 동안의 주가정보를 수집하는 수집모듈;

상기 주가정보를 히트맵 형식의 2D 이미지로 변환하고, 상기 2D 이미지에 라벨정보를 추가하여 적어도 하나의 입력이미지를 생성하는 입력이미지 생성모듈;

객체탐지모형 내에 상기 입력이미지를 입력하여 상기 객체탐지모형을 학습시키는 학습모듈; 및

학습된 상기 객체탐지모형을 이용하여 현재 시점에서 미리 설정된 기간 이후의 주가의 상승여부를 판단하는 주가예측모듈;

을 포함하는 객체탐지모형을 적용한 주가등락 예측장치.

청구항 2

청구항 1에 있어서,

상기 주가정보는, 거래일 날짜, 상기 거래일의 시가, 고가, 저가 및 종가인 것을 특징으로 하는 객체탐지모형을 적용한 주가등락 예측장치.

청구항 3

청구항 2에 있어서,

상기 입력이미지는 9×4 그리드(Grid)를 갖는 히트맵 이미지이며, 상기 입력이미지의 각 행은 거래일 날짜로 정의되고, 상기 입력이미지의 4개의 열은 각각 해당 거래일의 시가, 고가, 저가 및 종가로 정의되는 것을 특징으로 하는 객체탐지모형을 적용한 주가등락 예측장치.

청구항 4

청구항 3에 있어서,

상기 라벨정보는 해당 거래일로부터 미리 설정된 기간 이후의 종가의 등락여부이고,

상기 라벨정보는 상기 히트맵 이미지의 각 영역에서의 그레이 스케일로 구분되어 표현되는 것을 특징으로 하는 객체탐지모형을 적용한 주가등락 예측장치.

청구항 5

청구항 3에 있어서,

상기 객체탐지모형은 합성곱 신경망을 기반으로 하는 YOLO v2 모형인 것을 특징으로 하는 객체탐지모형을 적용한 주가등락 예측장치.

발명의 설명

기술 분야

[0001] 본 발명은 주가등락 예측장치에 관한 것으로, 구체적으로는 객체탐지모형을 적용한 주가등락 예측장치에 관한

것이다.

배경 기술

- [0003] 최근 학계에서는 주식시장에 대하여 주가의 흐름을 예측하고자 하는 연구가 활발히 이루어지고 있다.
- [0004] 주식시장은 매일 엄청난 양의 거래가 발생하는 세계 경제의 중추이기 때문에 주식시장의 흐름을 정확히 예측하는 것은 다양한 분야에 높은 가치 효과를 불러일으킬 수 있으며, 이는 주식시장의 행태를 분석하고 예측하는 것이 경제적인 목표를 달성하는 데에 큰 역할을 한다는 것을 반증한다.
- [0005] 조직화된 증권시장이 개설된 이후로 많은 투자자들 또한 수익 창출, 자산의 증대 등을 목적으로 하여 주가의 변동을 예측하고자 하지만 주식시장은 비선형적이고 급격한 변화의 정도를 지니는 복잡한 시스템일 뿐만 아니라 정치적 상황이나 경제 정책 등에 영향을 많이 받기 때문에 예측이 까다로우며, 주식시장의 흐름을 잘못 예측하여 투자를 진행하였을 경우 큰 손실을 볼 가능성이 존재한다.
- [0006] 따라서 주가의 흐름을 정확하게 예측하기 위한 다양한 방법이 연구되어 왔는데, 주식시장을 예측하는 방법은 크게 두 가지로 구분될 수 있다.
- [0007] 먼저 과거의 주식 가격들을 활용하여 미래의 주식 가격을 예측하는 기술적 분석(technical analysis)을 들 수 있으며, 또 다른 방법으로는 거시경제 지표들을 활용하여 예측을 진행하는 기본적 분석(fundamental analysis)을 들 수 있다.
- [0008] 아울러 최근에는 머신러닝(Machine Learning), 특히 딥러닝(Deep Learning)과 같은 AI 관련 기술들이 급격히 발전하며 과거의 주식 가격들을 활용한 기술적 분석 방법이 활발히 사용되고 있다.
- [0009] 특히 최근에는 이미지 데이터를 사용하여 주가 등락을 예측하는 방식이 제안된 바 있으며, 다른 방법 대비 주식시장의 흐름 예측에 있어서 높은 정확도를 보이는 것이 확인된 바 있다.
- [0010] 이는 주가 데이터를 기존의 정형화된 데이터를 이용하는 방법보다 비정형데이터인 이미지를 통해 예측을 시도하는 것이 주식시장의 변동성을 파악하는데 더 나은 예측력을 보일 수 있다는 가능성을 열어주었다.
- [0011] 그러나 지금까지 이미지를 기반으로 연구가 진행된 주가 예측 정확도를 기반으로 투자자들이 주식 매매를 시도하기에는 현저히 낮은 정확도를 보일 뿐만 아니라 연구에서 보여주고 있는 정확도는 전체적인 모델의 정확도일 뿐 확정적인 주가의 등락 시점에 대한 정확도가 아니라는 문제점이 있다.
- [0012] 따라서 상술한 종래 연구들의 문제점들을 극복함으로써 주가의 등락의 정확한 예측이 가능할 뿐만 아니라 투자 수익을 극대화할 수 있도록 정확한 매매시점까지 판단할 수 있는 주가등락 예측장치의 연구가 필요한 상황이다.
- [0013] 한편, 하기 선행기술문헌은 사용자 참여형 주가 예측 서비스 제공 방법에 관한 내용을 개시하고 있으며, 본 발명의 기술적 요지는 개시하고 있지 않다.

선행기술문헌

특허문헌

- [0015] (특허문헌 0001) 대한민국 공개특허공보 제10-2015-0094923호

발명의 내용

해결하려는 과제

- [0016] 본 발명의 일 실시예에 따른 객체탐지모형을 적용한 주가등락 예측장치는 전술한 문제점을 해결하기 위하여 다음과 같은 해결과제를 목적으로 한다.
- [0017] 컴퓨터 비전 영역에서 활용되고 있는 객체탐지 모형(object detection)을 활용하여 주가의 등락을 예측할 수 있을 뿐만 아니라 수익을 최대화하기 위한 정확한 매매 시점까지 제안할 수 있는 주가등락 예측장치를 제공하는

것이다.

[0018] 본 발명의 해결과제는 이상에서 언급된 것들에 한정되지 않으며, 언급되지 아니한 다른 해결과제들은 아래의 기재로부터 당해 기술분야에 있어서의 통상의 지식을 가진 자에게 명확하게 이해되어 질 수 있을 것이다.

과제의 해결 수단

[0020] 본 발명의 일 실시예에 따른 객체탐지모형을 적용한 주가등락 예측장치는, 미리 설정된 기간 동안의 주가정보를 수집하는 수집모듈; 상기 주가정보를 히트맵 형식의 2D 이미지로 변환하고, 상기 2D 이미지에 라벨정보를 추가하여 적어도 하나의 입력이미지를 생성하는 입력이미지 생성모듈; 객체탐지모형 내에 상기 입력이미지를 입력하여 상기 객체탐지모형을 학습시키는 학습모듈; 및 학습된 상기 객체탐지모형을 이용하여 현재 시점에서 미리 설정된 기간 이후의 주가의 상승여부를 판단하는 주가예측모듈;을 포함한다.

[0021] 상기 주가정보는, 거래일 날짜, 상기 거래일의 시가, 고가, 저가 및 종가인 것이 바람직하다.

[0022] 상기 입력이미지는 9×4 그리드(Grid)를 갖는 히트맵 이미지이며, 상기 입력이미지의 각 행은 거래일 날짜로 정의되고, 상기 입력이미지의 4개의 열은 각각 해당 거래일의 시가, 고가, 저가 및 종가로 정의되는 것이 바람직하다.

[0023] 상기 라벨정보는 해당 거래일로부터 미리 설정된 기간 이후의 종가의 등락여부이고, 상기 라벨정보는 상기 히트맵 이미지의 각 영역에서의 그레이 스케일로 구분되어 표현되는 것이 바람직하다.

[0024] 상기 객체탐지모형은 합성곱 신경망을 기반으로 하는 YOLO v2 모형인 것이 바람직하다.

발명의 효과

[0026] 본 발명의 일 실시예에 따른 객체탐지모형을 적용한 주가등락 예측장치는 이미지상의 객체를 검출하는데 주로 연구되어 온 객체탐지모형을 적용함으로써 변동이 크고 예측이 불확실한 주가의 흐름을 높은 정확도로 예측할 수 있는 효과를 기대할 수 있다.

[0027] 본 발명의 효과는 이상에서 언급된 것들에 한정되지 않으며, 언급되지 아니한 다른 효과들은 아래의 기재로부터 당해 기술분야에 있어서의 통상의 지식을 가진 자에게 명확하게 이해되어질 수 있을 것이다.

도면의 간단한 설명

[0029] 도 1은 합성곱 신경망 구조를 도시한 개념도이다.

도 2는 본 발명의 일 실시예에 따른 객체탐지모형을 적용한 주가등락 예측장치를 간략히 도시한 블록도이다.

도 3은 도 2에서의 수집모듈이 수집하는 주가정보의 일 예를 도시한 이미지이다.

도 4는 도 2에서의 입력이미지 생성모듈, 학습모듈 및 주가예측모듈의 전반적인 동작 프로세스를 설명하기 위한 이미지이다

도 5는 도 2에서의 입력이미지 생성모듈에 의한 데이터의 히트맵 형식으로 변환시키는 방법을 설명하기 위한 이미지이다.

도 6은 동일 시점에 예측되는 서로 다른 bounding box를 confidence score를 이용하여 결정하는 것을 설명하기 위한 이미지이다.

도 7(a)은 예측된 객체의 상승에 대한 혼동행렬이고, 도 7(b)는 하강에 대한 혼동행렬이다.

도 8 및 도 9는 본 발명의 일 실시예에 따른 객체탐지모형을 적용한 주가등락 예측장치에서 5거래일 이후에 주가가 상승할 것이라고 모델이 예측할 때에 최적의 confidence threshold를 정하기 위한 모델학습의 결과를 나타내는 표 및 그래프이다.

도 10 및 도 11 본 발명의 일 실시예에 따른 객체탐지모형을 적용한 주가등락 예측장치에서 5거래일 이후에 주

가가 하락할 것이라고 모델이 예측할 때에 최적의 confidence threshold를 정하기 위한 모델학습의 결과를 나타내는 표 및 그래프이다.

도 12 및 도 13은 도 8 내지 도 11의 결과에 기초하여 설정된 Parameter를 적용할 경우의 주가의 상승 및 하락시의 수익도출 결과를 도시한 도표이다.

발명을 실시하기 위한 구체적인 내용

- [0030] 첨부된 도면을 참조하여 본 발명에 따른 바람직한 실시예를 상세히 설명하되, 도면 부호에 관계없이 동일하거나 유사한 구성 요소는 동일한 참조 번호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다.
- [0031] 또한, 본 발명을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 또한, 첨부된 도면은 본 발명의 사상을 쉽게 이해할 수 있도록 하기 위한 것일 뿐, 첨부된 도면에 의해 본 발명의 사상이 제한되는 것으로 해석되어서는 아니 됨을 유의해야 한다.
- [0033] 본 발명의 일 실시예에 따른 객체탐지모형을 적용한 주가등락 예측장치를 설명하기에 앞서, 본 발명에 적용되는 합성곱 신경망(Convolutional Neural Networks, CNN), 히트맵 이미지(Heatmap Image) 및 YOLO v2를 포함하는 객체탐지모형(Object Detection Model)에 대하여 먼저 설명하도록 한다.
- [0034] 합성곱 신경망은 공간적 구조(spatial structure)를 가지는 2D 이미지의 패턴을 분석하는 하나의 방법으로, 합성곱 신경망은 합성곱 연산을 통해 2-D 이미지를 일반화하며 이미지가 내포하고 있는 특징을 추출하여 이를 효과적으로 분석해준다.
- [0035] 도 1은 일반적인 합성곱 신경망의 구조를 도식화 한 것으로, 합성곱 신경망은 2D 이미지를 입력 이미지로 받아 합성곱 층(convolutional layer)에서의 연산을 거친다.
- [0036] 합성곱 층에서는 합성곱 연산을 거친 결과를 받아 활성 함수(activation function), pooling 계층을 통과하는 과정들이 여러 번 반복되며 이미지의 특징값들이 출력되며, 합성곱 층에서 도출된 특징값들은 완전연결계층(fully connected layer)에서 기존 신경망 구조를 통하여 연산을 진행하여 결과값을 산출한다.
- [0037] 특히 본 발명의 주가등락 예측장치에 적용되는 객체탐지모형은 합성곱 신경망을 기반으로 한 딥러닝 모형을 사용하여 주식의 매매 타이밍을 도출하기 때문에 모델에 사용되는 입력값을 비정형데이터인 2D 이미지 형태로 만들어주는 작업을 진행한다.
- [0038] 본 발명에서는 정형데이터인 주가 정보를 히트맵(heatmap) 형식인 비정형 이미지 데이터로 변환하는데, 일반적으로 히트맵 기법은 분석하고자 하는 대상의 전반적인 특징에 대해서 쉽게 파악할 수 있게끔 해주는 효과적인 시각화 기법으로 알려져 있다.
- [0039] 한편, 객체탐지모형은 이미지 상에서 관심 있는 객체를 이미지의 배경과 구분하고 식별해주는 알고리즘으로, 객체탐지 모형은 2D 영상을 입력값으로 받아 합성곱 신경망을 통해 유의미한 특징 값(feature map)을 생성하고 이를 통해 객체를 판별한다.
- [0040] 즉 이러한 객체탐지모형은 이미지를 분류(classification)하는 방법과 우리가 관심 있는 객체의 위치에 대해서 예측(regression)하는 문제가 합쳐진 hybrid 모델이라고 할 수 있다.
- [0041] 객체탐지 모형의 방법으로 종래에 제시되었던 two-stage 기반 모형들은 낮은 연산 속도라는 한계점을 보여 빠른 속도로 변화하는 주가의 흐름을 예측하기에는 부적합하다는 문제점이 있다.
- [0042] 이러한 문제점을 극복할 수 있는 one-stage 기반 객체탐지 알고리즘이 제안되고 있으며, 이러한 one-stage 기반 객체탐지 알고리즘은 연산 속도가 빠르며 그 효율성 또한 뛰어나며, One-stage 기반 객체탐지 모형에는 대표적으로 YOLO 알고리즘이 있다.
- [0043] YOLO는 region proposal 후에 classification을 순차적으로 진행하는 two-stage method와는 다르게 입력값으로 받는 이미지를 grid로 나누어 bounding box regression과 classification을 한 번에 처리하기 때문에 연산 속도가 빠르다는 장점이 있다.
- [0044] 그러나 YOLO 알고리즘은 작은 물체를 탐지하는 것에 있어서 탐지 정확도가 떨어진다는 단점을 갖고 있기 때문에, 본 발명에서는 주가 데이터를 히트맵 형식인 2-D 이미지로 변환할 때에 학습 이미지의 size를 조정함으

로써 YOLOv2모형을 이용해 등락을 예측하도록 하여 기존 YOLO가 가지는 문제점을 극복할 수 있도록 구성된다.

- [0045] 따라서 진행하고자 하는 주가 등락 예측에는 구조가 단순하며 검출 속도가 빠른 YOLOv2모형을 적용한다.
- [0047] 이하에서는 상술한 내용 및 도 2 내지 도 13을 참조하여 본 발명의 일 실시예에 따른 객체탐지모형을 적용한 주가등락 예측장치에 대하여 설명하도록 한다.
- [0048] 본 발명의 일 실시예에 따른 객체탐지모형을 적용한 주가등락 예측장치는 도 2에 도시된 바와 같이 수집모듈(100), 입력이미지 생성모듈(200), 학습모듈(300) 및 주가예측모듈(400)을 포함하도록 구성된다.
- [0049] 수집모듈(100)은 미리 설정된 기간 동안의 주가정보를 수집하는 기능을 수행하며, 이러한 주가정보로는 코스피 지수 등의 데이터에서 추출된 거래 당일의 시가, 고가, 저가, 종가 그리고 해당 거래일 날짜로 구성될 수 있다.
- [0050] 목적 변수로는 종가를 사용하고, 해당 거래일의 종가보다 5거래일 이후의 종가가 더 높은 값을 가진다면 상승, 낮은 값을 가진다면 하강으로 판단하여 목적 변수를 상승과 하강을 갖는 이진 변수로 처리한다.
- [0051] 본 발명의 연구에서 사용된 주가정보는 도 3에 도시된 바와 같이 코스피 지수에서의 종가를 이용하였으며, 전체 데이터는 2001년 1월 2일부터 2018년 12월 27일까지, 총 4,449 거래일로 이루어져 있다. 이 중 모델을 학습시키기 위한 학습데이터는 2001년부터 2016년까지의 3,962거래일을, 모델을 평가하기 위한 평가데이터는 2017년부터 2018년까지 약 2년간의 487거래일을 사용하였다. 모델에 학습에 사용된 데이터는 2008년부터 2009년까지 하락하는 구간을 제외하고 전반적으로 상승하는 시장의 흐름을 보인다.
- [0052] 입력이미지 생성모듈(200)은 수집모듈(100)에 의하여 수집된 주가정보를 히트맵 형식의 2D 이미지로 변환하고, 2D 이미지에 라벨정보를 추가하여 입력이미지를 생성하는 기능을 수행한다.
- [0053] 학습모듈(300)은 객체탐지모형 내에 입력이미지를 입력하여 객체탐지모형을 학습시키는 기능을 수행하며, 주가 예측모듈(400)은 학습된 객체탐지모형을 이용하여 현재 시점에서 미리 설정된 기간 이후의 주가의 상승 및 하강 여부를 판단하는 기능을 수행한다.
- [0054] 상기 입력이미지 생성모듈(200), 학습모듈(300) 및 주가예측모듈(400)의 전체적인 프로세싱을 도 4를 참조하여 설명하도록 한다.
- [0055] 먼저 수집모듈(100)이 수집한 데이터인 KS11 지수의 데이터를 입력이미지 생성모듈(200)이 히트맵 형식의 2D 이미지를 생성하며 입력 이미지에 객체에 대한 정보인 라벨 정보를 추가해준다.
- [0056] 이때 이미지의 크기는 9거래일을 하나의 입력 이미지로 생성하게 되고 1거래일씩 미루어 가며 전체 데이터에 대하여 입력 이미지를 만들어준다.
- [0057] 학습데이터를 기준으로 보면 원 데이터에서는 3,962거래일이 존재하지만 9×4 크기로 히트맵 이미지를 만들어주어 총 3,954장의 입력 이미지가 생성된 것을 볼 수 있다.
- [0058] 다음으로 생성된 이미지들 각각에 대해 bounding box를 기준으로 5거래일 이후 종가 등락 여부에 따라 라벨을 생성하는데, 이는 객체탐지모형에서 예측, 분류할 객체를 본 연구에서는 하나의 크기의 bounding box로 보아 최종적으로 5거래일 이후 종가의 등락 여부를 예측하고자 함이다.
- [0059] Bounding box 역시 입력 이미지를 만들 때와 같이 하나의 입력 이미지 내에서 1거래일씩 미루어 가며 생성해주며, 학습데이터를 기준으로 모든 입력 이미지 내에서 5거래일 이후 종가의 상승 정보를 담고 있는 bounding box는 총 2,224개, 하락 정보를 가지고 있는 것은 총 1,734개가 생성된다.
- [0060] 이후 주가예측모듈(400)은 라벨정보를 포함하고 있는 입력 이미지를 모델에 적용하고 5거래일 이후 종가의 등락을 예측하며, 이때 사용된 YOLOv2 모델은 customize하지 않은 vanilla model을 이용해서 2-D 이미지로 변형된 KS11 데이터를 학습모듈(300)에 의하여 학습시켜 모델을 최적화했다.
- [0061] 학습모듈(300)에서 학습에 사용되는 입력 이미지는 6거래일부터 9거래일까지 하루씩 늘려서 모델의 예측성능을 비교하였으며 Adam optimizer를 사용하였고 epoch은 200으로 고정하였으며, 학습된 모델에 test 이미지를 사용하여 결과를 도출하되 모델학습에 사용된 이미지의 크기와 test 이미지의 크기는 동일하게 사용되었다.
- [0062] 상술한 본 발명의 전반적인 프로세스에 기초하여 본 발명의 일 실시예에 따른 객체탐지모형을 적용한 주가등락 예측장치의 각 구성에 대하여 좀 더 구체적으로 설명하도록 한다.

- [0063] 입력이미지 생성모듈(200)은 수집모듈(100)에 의하여 수집된 KS11 지수 데이터를 합성곱 신경망을 기반으로 하는 객체탐지모형에 적용하기 위해 2D 이미지로 변환하되, 데이터에 존재하는 변수들인 시가, 고가, 저가, 종가의 값을 기반으로 히트맵 형식으로 변환한다.
- [0064] 도 5는 본 발명의 일 실시예에 따른 객체탐지모형을 적용한 주가등락 예측장치의 입력이미지 생성모듈(200)이 수집모듈(100)에서 수집한 데이터를 히트맵 형식으로 변환시키는 방법의 일 예를 도시한 그림이다.
- [0065] 먼저 학습모듈(300)에서 학습에 사용될 이미지의 입력 크기를 설정해 주어야 하는데, 이미지 하나의 row 값은 각 거래일(date)을 나타내고 column 값은 당일의 시가, 고가, 저가, 종가를 나타낸다.
- [0066] 학습에 사용될 이미지의 크기는 도 5에 도시된 바와 같이 9×4 로 설정하는데, 이는 9 거래일 동안 데이터에 속한 값들이 하나의 히트맵 이미지로 변환된 것을 나타내며, 9 거래일을 하나의 입력 크기로 설정한 이유는 객체탐지모델에서 예측에 필요한 bounding box의 설정과 관련이 있다.
- [0067] 하나의 입력 이미지 내에서 예측될 수 있는 bounding box의 개수는 최대 5개인데, 이때 첫 번째 bounding box는 입력 이미지 내의 1~5 거래일까지이며, 해당 bounding box로는 bounding box에서 가장 마지막 row의 거래일보다 5거래일 후 종가의 등락 여부를 예측하게 된다.
- [0068] 하지만 하나의 입력 이미지 내에서 첫 번째 bounding box에서 예측될 값이 동일 입력 이미지에 존재하게 된다면 이미 존재하는 정보를 다시 한번 예측하게 되는 오류를 범하게 된다.
- [0069] 변환된 이미지의 오른쪽에 있는 컬러바는 해당 이미지가 변환될 때에 색상을 띄는 정도를 표시해주며 각 학습 이미지에 따라서 다른 scale을 지니게 되며, 히트맵은 해당 기간 내에 상대적으로 높은 값을 가질수록 하양에 가까운 색상을, 낮은 값을 가질수록 검정에 가까운 색상을 띈다.
- [0070] 학습모듈(300) 및 주가예측모듈(400)에 적용되는 객체탐지모형인 YOLOv2는 이미지 내에서 찾고자 하는 객체가 있을만한 곳인 bounding box와 해당 bounding box 내에 객체가 존재하는지 여부인 class probability를 single regression problem으로 바라보며, 이때 예측되는 각각의 bounding box는 하기 수식 (1)을 통해 각각의 confidence score을 계산한다

[0072] <수식 (1)>

$$confidence\ score = Pr(Object) \times IOU\ value$$

$$IOU\ value = \frac{Area\ of\ Overlap(a\ \&\ b)}{Area\ of\ Union(a\ \&\ b)}$$

- [0073]
- [0075] Confidence score는 실제 객체가 bounding box 내에 존재할 확률인 $Pr(object)$ 에 ground truth bounding box(a)와 predicted bounding box(b)간의 관계로 도출되는 IOU value의 곱으로 이루어져 있다.
- [0076] Confidence score는 기준이 되는 threshold를 통하여 종가의 등락 여부 예측에 있어서 사용되며 적절한 threshold의 결정을 통해 모델의 성능을 향상할 수 있다.
- [0077] 한편, 도 6은 동일 시점에 예측되는 서로 다른 bounding box를 confidence score를 이용하여 결정하는 것을 설명하기 위한 그림이다.
- [0078] YOLOv2 모델을 학습시키게 되면 하나의 입력 이미지에서 상승 또는 하강이라고 판단되는 bounding box가 regression 되는데, 이때 예측되는 bounding box는 도 6에서 확인되는 바와 같이 서로 다른 시점에서 생성되는 입력 이미지이더라도 동일 거래일을 갖는 bounding box가 여러 개 생성될 수 있다.
- [0079] 본 발명에서는 이 중 가장 정확한 bounding box를 결정하기 위해서 동일 거래일을 가지는 bounding box 중 예측된 confidence score가 가장 높은 box를 선택하여 해당 거래일에서 생성될 수 있는 bounding box로 판단하며, 최종적으로 거래일 하루마다 5거래일 이후 종가의 등락을 판단하는 하나의 bounding box가 regression된다.

[0080] 한편, 본 발명에서 제안한 모델을 통해 예측하는 객체는 현재 시점에서 5거래일 이후의 종가가 상승할 것인지, 하강할 것인지를 나타내는 bounding box 두 가지이며, 각 객체의 상승과 하강에 대해서 도 7에 도시된 바와 같이 두 개의 혼동행렬을 도출할 수 있다.

[0081] 모델을 통한 결과로 0~1 사이의 confidence score마다 recall, precision이 나오게 되며, 객체가 정확하게 검출되었는지를 나타내는 최적의 confidence score threshold를 구해주기 위해서 하기 수식 (2)를 통해 수익 ($\text{Income}_{\text{PR/PF}}$)을 나타내는 지표를 계산하였다.

[0083] <수식 (2)>

$\text{Revenues}_{\text{PR/PF}} (\%) = \text{Cumulative sum of PR/PF in stock price rate from } t \text{ to } t+5 \text{ time that predicted exactly in 'True Positive' during the training period.}$

$\text{Loss}_{\text{PR/PF}} (\%) = \text{Cumulative sum of PR/PF in stock price rate from } t \text{ to } t+5 \text{ time that predicted exactly in 'False Positive' during the training period.}$

$$\text{Income}_{\text{PR/PF}} = \text{Revenues}_{\text{PR/PF}} + \text{Loss}_{\text{PR/PF}}$$

[0084]

[0086] 모델이 예측한 주가의 상승/하강이 실제 주가의 상승/하강과 일치하는 ‘True Positive(TP)’ 영역에서 ‘ $\text{Revenues}_{\text{PR/PF}}$ ’가 산출된다. $\text{Revenues}_{\text{PR/PF}}$ 은 ‘특정 시점 t로부터 5거래일 이후인 t+5 시점에 주가의 등락이 변화할 것이다’라는 의미를 내포한다.

[0087] Revenues 의 값은 ‘True Positive’ 영역에서 도출한 모든 변화율의 값을 누적해서 구해주며 결과적으로 모델이 예측한 주가의 상승/하강률로 얻어낸 이득을 의미한다.

[0088] 모델이 5거래일 이후에 상승/하강할 것이라 예상하였지만 실제값은 그렇지 않은, ‘False Positive(FP)’ 영역에서 산출되는 $\text{Loss}_{\text{PR/PF}}$ 역시 $\text{Revenues}_{\text{PR/PF}}$ 과 동일한 방식으로 등락률의 누적 합을 통해 구해진다.

[0089] $\text{Revenues}_{\text{PR/PF}}$ 은 제안한 모델을 통해 이득을 볼 수 있는 총 등락률이고 $\text{Loss}_{\text{PR/PF}}$ 은 모델이 잘못 예측하여 손해를 볼 수 있는 총 등락률이며, $\text{Income}_{\text{PR/PF}}$ 은 $\text{Revenues}_{\text{PR/PF}}$ 과 $\text{Loss}_{\text{PR/PF}}$ 의 값을 합해주어 모델이 예측한 손해를 배제하고 순수하게 이득을 볼 수 있는 부분만을 나타내 준다.

[0090] 구해진 $\text{Income}_{\text{PR/PF}}$ 을 기반으로 0~1사이의 confidence level마다의 값을 구해주고 가장 $\text{Income}_{\text{PR/PF}}$ 이 높은 지점을 최적의 confidence score threshold로 지정하여 결과를 도출한다.

[0091] 다음으로 모델을 통해 알아낸 매매 시점에서 얻은 $\text{Income}_{\text{PR/PF}}$ 이 얼마나 효과적인지 검증하기 위해서 buy-and-hold(BH)전략을 이용, 앞서 산출한 $\text{Income}_{\text{PR/PF}}$ 과 비교한다.

[0092] BH전략은 단순히 특정 거래기간 동안 해당 기간의 초에 주식을 매수하고 거래 기간의 마지막에 매도한 가격의 차이를 의미하지만 일반적으로 상승하는 시장에 있어서는 BH전략을 통해 산출한 buy-and-hold는 높은 값을 갖는다는 것을 알 수 있다.

[0093] 본 발명에서는 시장이 가지는 특성을 배제하고 일반적인 buy-and-hold 값과 연구에서 제시한 모델로부터 산출된 $\text{Income}_{\text{PR/PF}}$ 을 비교하기 위해 $\text{Income}_{\text{PR/PF}}$ 이 산출되는 매매 시점을 따로 설정해주며, $\text{Income}_{\text{PR/PF}}$ 과 buy-and-hold 값을 도출하는 기간은 한 달, 분기, 반기, 연 단위로 동일하게 구성하였다.

[0094] 앞서 구한 상승과 하강의 $\text{Income}_{PR/PF}$ 을 buy-and-hold 값과 비교해준 comparison rate는 하기 수식 (3)과 같이 전체 기간을 본 발명에서 구성한 기간 각각에서 나오는 값들의 평균으로 산출할 수 있다.

[0096] <수식 (3)>

$$\text{Comparison rate} = \frac{1}{M} \sum \frac{\text{Income}_{PR/PF}}{\text{buy and hold}}$$

$$M = \frac{N}{t}$$

$$N = \text{Test period}$$

$$t = \text{a month, quarter year, half year, a year}$$

[0097]

[0099] 상기 수식 (3)에서와 같이 구성한 기간 내에 시장의 형성이 상승 시장인지 하강 시장인지를 반영하여 비교해주기 위해서 buy-and-hold 값이 음수로 하강하는 시장의 경향을 보이면 비교 값으로 Income_{PF} 을 사용해주고 양수 값으로 상승 시장의 경향에는 Income_{PR} 을 사용하여 comparison rate을 구한다.

[0100] 앞서 설명한 본 발명의 제안모델을 통해 주가의 상승률과 하락률에 대한 결과를 각각 산출하였는데, 우선적으로 학습모델(300)에서의 모델의 training을 이용하여 최적의 confidence threshold를 찾아주는 것이 요구된다.

[0101] 도 8 및 도 9는 기준일로부터 5거래일 이후에 주가가 상승할 것이라 모델이 예측할 때에 최적의 confidence threshold를 정하기 위한 모델학습의 결과이다.

[0102] 결과를 보면 모델의 confidence threshold를 0.5로 설정하였을 때 16년간의 학습 기간 동안 약 2001%의 누적 상승률을 예측하며 가장 좋은 성능을 보인 것을 확인할 수 있다.

[0103] 도 10 및 도 11은 기준일로부터 5거래일 이후에 주가가 하강할 것이라 모델이 예측할 때에 최적의 confidence threshold를 정하기 위한 모델학습의 결과로서, 모델의 confidence threshold를 0.7로 설정하였을 때에 모델이 약 1570%의 누적 하강률을 예측하며 가장 좋은 성능을 보이는 것을 확인하였다.

[0104] 결과적으로 모델이 주가의 상승이라는 객체를 예측할 때에는 bounding box의 confidence score를 0.5로 설정하게 되며 주가가 하강할 것이라는 객체를 예측할 때에는 confidence score를 0.7로 설정하여 test 기간인 2017~2018년의 결과를 도출할 수 있게 된다.

[0105] 앞서 설정한 parameter값들을 적용하여 test 기간인 2016~2017년의 결과를 산출할 수 있으며, 도 12는 기준일로부터 5거래일 이후에 주가의 상승과 하강에 대한 결과이다.

[0106] 주가의 상승에 대해서 confidence threshold를 0.5로 설정한 결과, test period 2년간의 누적 상승 Income_{PR} 의 결과는 약 165%를 얻었으며 precision 0.77, recall 0.52의 값을 보인다.

[0107] 다음으로 누적 하강 수익 도출에 있어서 confidence threshold를 0.7로 설정한 결과로는 약 174%의 Income_{PF} 을 얻었고 precision 0.77, recall 0.59의 결과를 보였다.

[0108] 모델이 구한 결과를 buy-and-hold 전략과 비교한 결과는 도 13에서 확인할 수 있는데, 구성 기간을 한 달로 설정할 경우에는 본 연구의 모델이 기본적인 buy-and-hold 전략을 이용한 결과보다 평균적으로 5.78배 더 많은 이익을 얻을 수 있으며 분기, 반기, 일 년의 기간으로 구성한 결과 순차적으로 2519.73, 9.01, 6.59배 더 많은 이익을 낸 것을 볼 수 있다.

[0109] 구성 기간을 분기로 설정한 경우에 제안모델의 결과가 buy-and-hold의 결과보다 약 2519 배 정도로 상당히 높은

이익을 산출하는 결과를 보인다.

[0110] 이는 주식시장이 설정기간(분기)동안 큰 변화가 없었지만 제안 모델의 같은 경우에는 시장의 흐름과 상관없이 주가의 상승/하락을 예측해주어 이익을 산출해 준다는 것을 반증하는 결과이다.

[0111] 추가적으로 일반적인 buy-and-hold 전략을 사용한 값과 제안한 모델의 결과를 비교하기 위해서 기간을 test 데이터셋 의 전체 기간(2년)으로 설정하고 결과를 도출하였을 경우에도 약 507 배만큼 제안 모델이 좋은 결과를 보이고 있다.

[0112] 모든 설정 기간에서 본 연구에서 제시한 모델 결과가 bench mark model보다 더 우수한 성능을 보이고 있으며 모델이 예측한 적절한 매매 시점에서의 거래가 더 많은 이익을 취할 수 있다는 점을 확인할 수 있다.

[0114] 본 명세서에서 설명되는 실시예와 첨부된 도면은 본 발명에 포함되는 기술적 사상의 일부를 예시적으로 설명하는 것에 불과하다. 따라서 본 명세서에 개시된 실시예들은 본 발명의 기술적 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이므로, 이러한 실시예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것이 아님은 자명하다. 본 발명의 명세서 및 도면에 포함된 기술적 사상의 범위 내에서 당해 기술분야에 있어서의 통상의 지식을 가진 자가 용이하게 유추할 수 있는 변형 예와 구체적인 실시예는 모두 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0116] 100: 수집모델

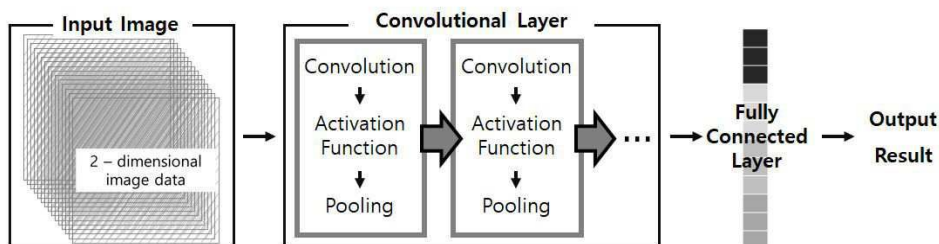
200: 입력이미지 생성모델

300: 학습모델

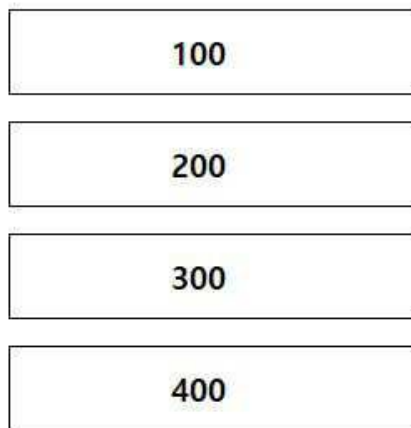
400: 주가예측모델

도면

도면1



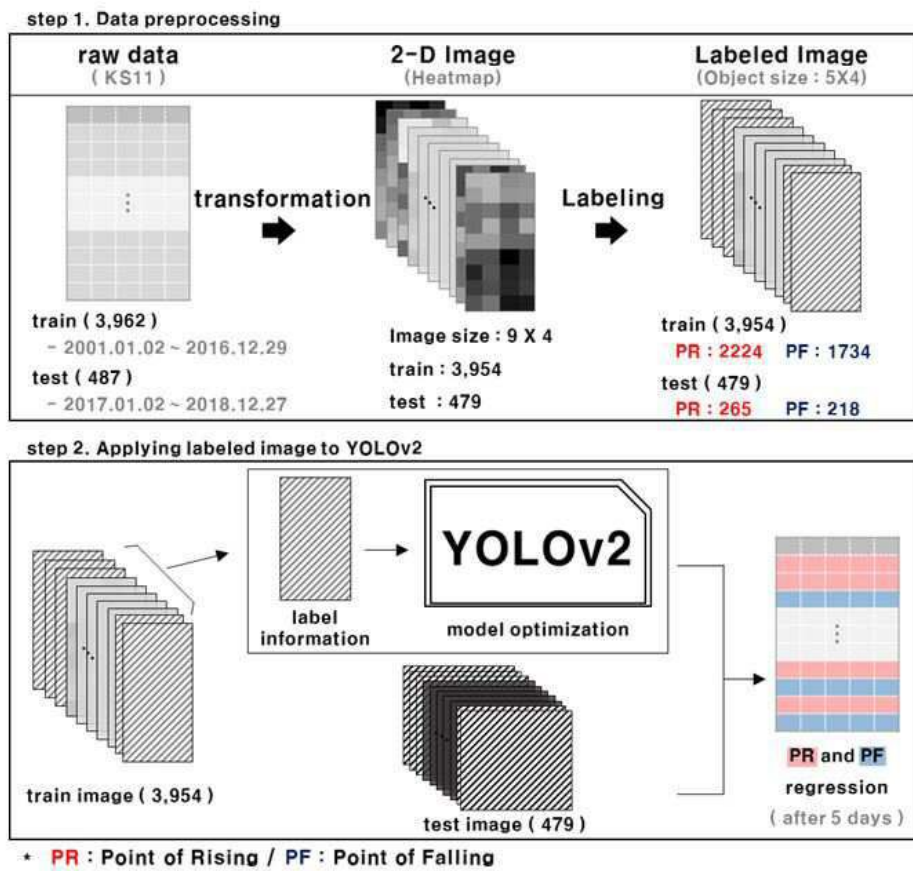
도면2



도면3

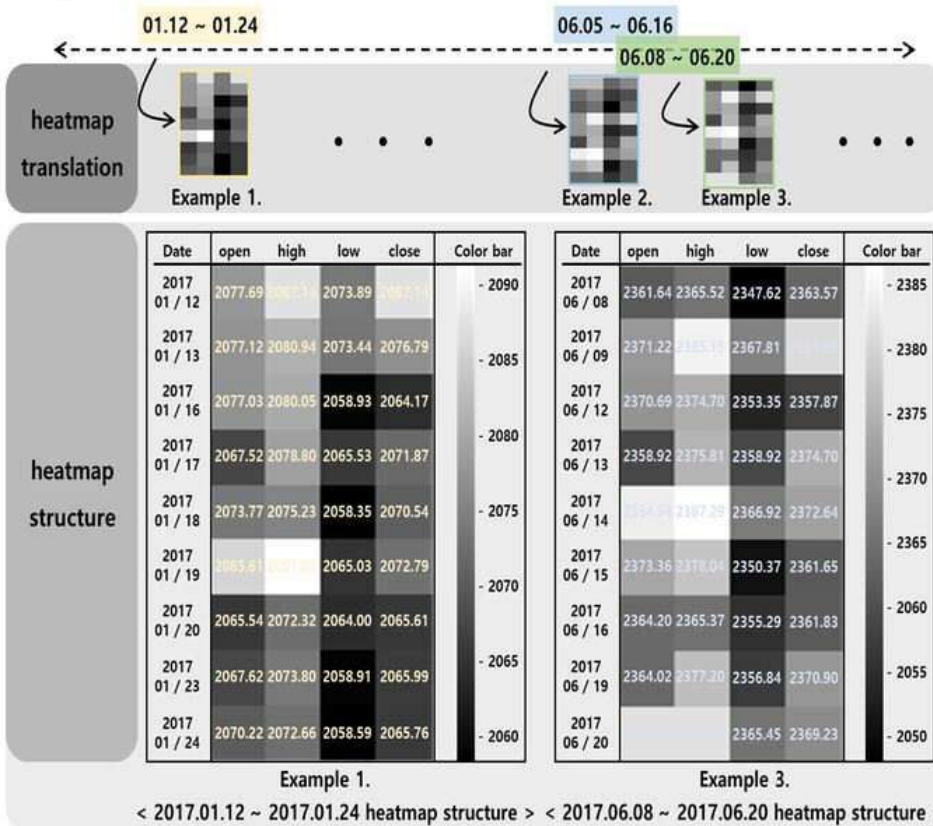


도면4

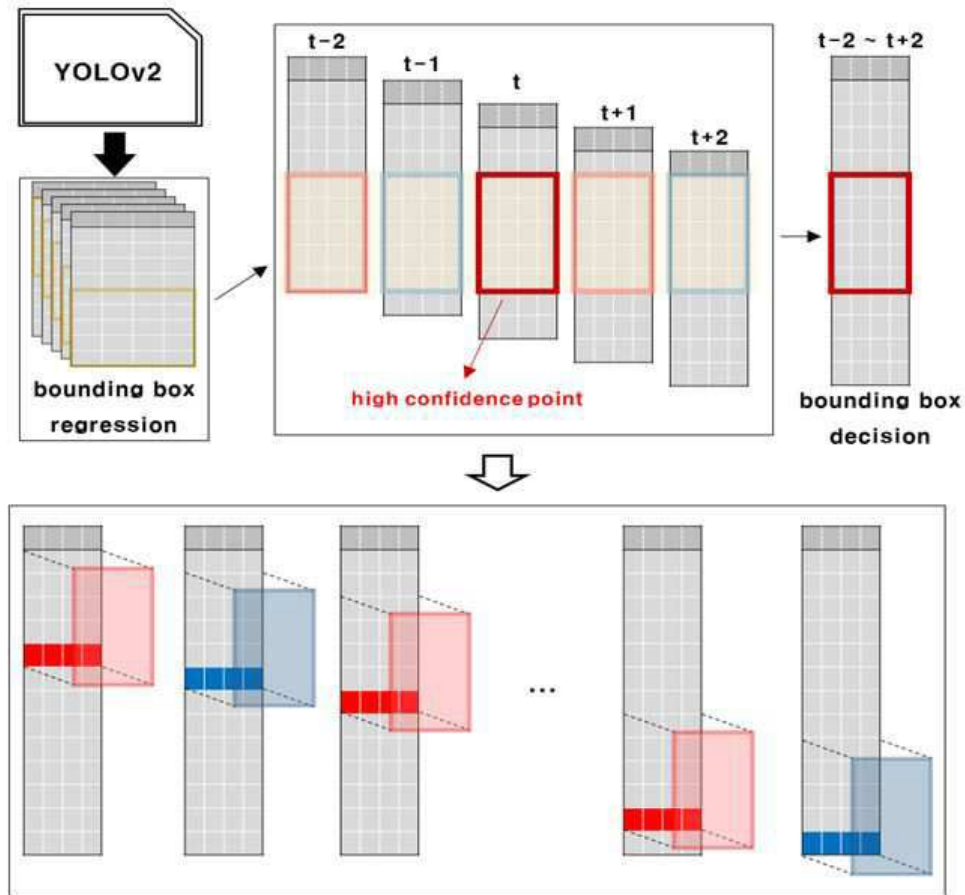


도면5

Example Date : 2017.01.01 ~ 2017. 07.31



도면6



도면7

	Actual _{PR}	Actual _{Background}
Predicted _{PR}	TP _{PR}	FP _{PR}
Predicted _{Background}	FN _{PR}	TN _{PR}

a. Confusion matrix of PR

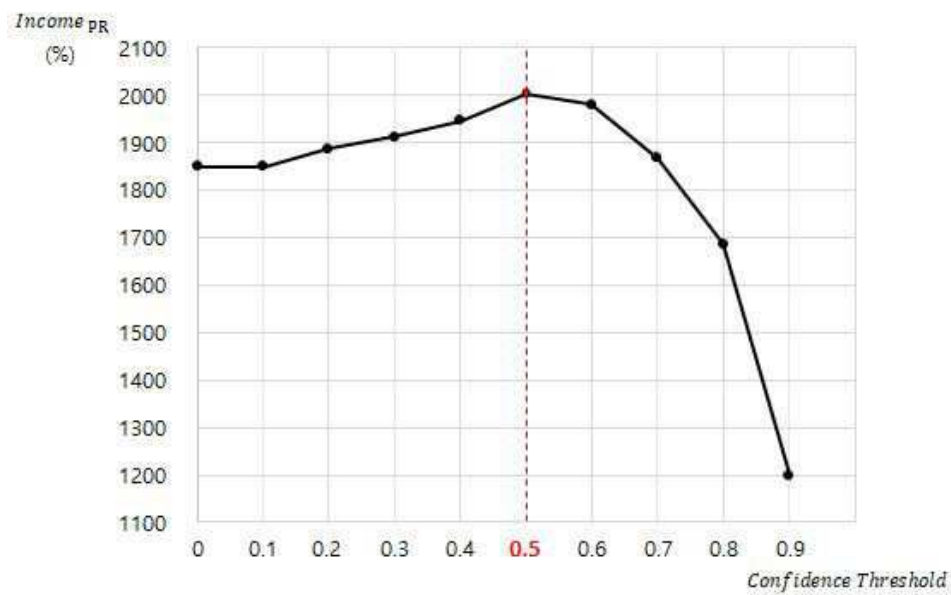
	Actual _{PF}	Actual _{Background}
Predicted _{PF}	TP _{PF}	FP _{PF}
Predicted _{Background}	FN _{PF}	TN _{PF}

b. Confusion matrix of PF

도면8

Confidence threshold	Precision	Recall	Revenues (%)	Loss (%)	Income _{PR} (%)
0	0.68	0.60	3057.11	-1206.96	1850.14
0.1	0.68	0.60	3057.11	-1206.96	1850.14
0.2	0.69	0.60	3057.11	-1170.61	1886.50
0.3	0.69	0.59	3051.52	-1139.50	1912.02
0.4	0.70	0.59	3024.80	-1079.26	1945.53
0.5	0.71	0.58	2973.27	-971.49	2001.77
0.6	0.73	0.55	2794.06	-813.75	1980.31
0.7	0.74	0.49	2550.50	-683.47	1867.03
0.8	0.75	0.42	2213.56	-527.54	1686.02
0.9	0.78	0.29	1527.31	-328.20	1199.10

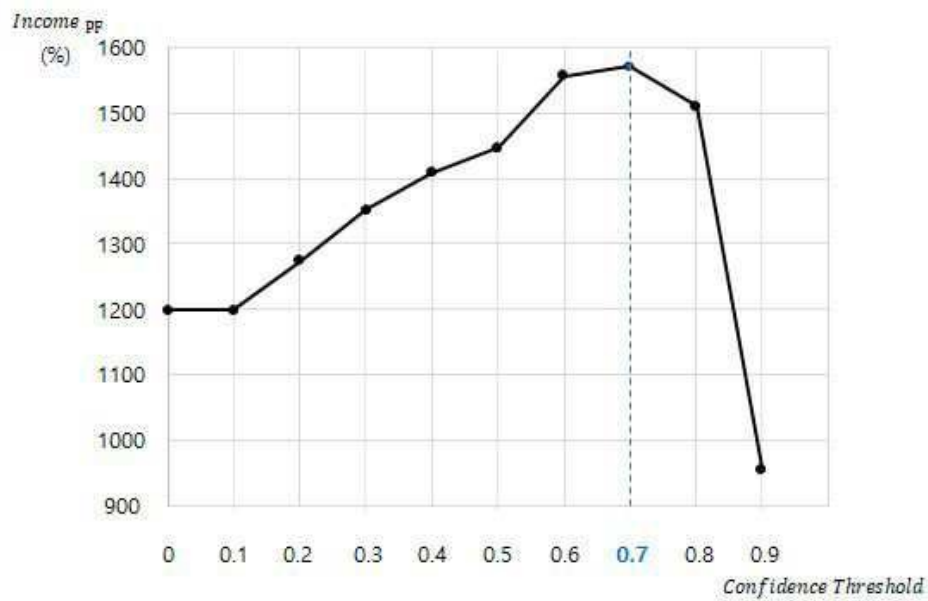
도면9



도면10

Confidence threshold	Precision	Recall	Revenues (%)	Loss (%)	Income _{pf} (%)
0	0.58	0.60	-2827.03	1627.76	1199.27
0.1	0.58	0.60	-2827.03	1627.76	1199.27
0.2	0.59	0.60	-2816.68	1542.88	1273.79
0.3	0.60	0.60	-2801.44	1449.66	1351.78
0.4	0.61	0.58	-2757.77	1349.67	1408.09
0.5	0.62	0.56	-2693.62	1246.58	1447.04
0.6	0.65	0.53	-2578.53	1021.43	1557.10
0.7	0.69	0.46	-2295.37	724.74	1570.63
0.8	0.76	0.35	-1838.84	328.14	1510.70
0.9	0.89	0.17	-1010.63	54.25	956.38

도면11



도면12

Stock Price	Precision	Recall	Revenues (%)	Loss (%)	Income _{pr} (%)
(PR)	0.77	0.52	195.32	-30.20	165.12
Stock Price	Precision	Recall	Revenues (%)	Loss (%)	Income _{pf} (%)
(PF)	0.66	0.59	-232.05	57.52	174.53

도면13

Period	Comparison Rate
a month	5.78
quarter year	2519.73
half year	9.01
a year	6.59
whole period (2Y)	507.72