

CNN을 활용한 KOSPI 주가지수 시계열 분석

이재권 진영봉



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

목차

1. 배경 및 목표

- 배경
- 분석 목표
- 선행연구

2. 활용 데이터

- KOSPI 지수 데이터
- KOSPI 지수 증가
- Target : Trend

3. 모델링 기법

- CNN
- ResNet
- GAF
- MTF

4. 모델링

5. 결과

- 결과
- 활용성
- 향후 계획



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

배경

“ 2020년 주식결제대금 417조, 투자 열풍에 46.6% 급등 ” (디지털타임스)

“ 새해에도 이어가는 ‘동학개미’ 열풍, 1월 한 달 개인투자자 26조 순매수 ” (경향신문)

2030 토스 사용자 “주식투자 지속 혹은 확대할 것”

2030 토스 사용자 1093명 대상 설문조사 결과

현재 주식투자를
하고 계신가요?

1,093명이 이 질문에 답변했습니다



toss

주가는 경제적 소득에 직접
적인 영향을 미치기 때문에
주가 예측은 투자자들로
하여금 많은 관심을 받음



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

분석 목표

“ KOSPI 주가지수의 **시계열성**을 바탕으로 **CNN** 모델을 활용해
다음날의 **주가지수**를 3개의 클래스로 **분류**하여 예측 ”

- 이미지 인식 분야에서 좋은 성능을 보이는 딥러닝 모델인 CNN을
시계열 데이터에 적용
- 일반적으로 시계열 분석에는 RNN이 적합하다고 알려졌지만,
CNN을 활용한 시계열 분석 연구 또한 진행되고 있음
- CNN을 활용하면 비교적 적은 매개변수를 가지고 결과에 대한 해석이
시각적으로 가능하다는 장점 존재



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

선행 연구

KOSPI 시장에서의 주가 예측(분류)

“기술적 지표와 기계학습을 이용한 KOSPI 주가지수 예측”, 박재연, 유재필, 신현준, 2016, 정보기술아키텍처연구

- KOSPI200 지수에서 파생된 기술적 지표들을 사용하여 SVM으로 주가지수 예측
- Target : (buy, hold, sell) test accuracy : 0.55

“딥러닝을 이용한 주가예측”, 이지훈, 2016, 숭실대학교 석사학위 논문

- KOSPI 상위 20 종목에 대해 기술적 지표들을 사용하여 RNN으로 주가 예측
- Target : (Up, Down) test accuracy : 0.53

“강화학습을 이용한 주가 예측”, 이보미, 2018, 한양대학교 석사학위 논문

- KOSPI200 지수의 종가, 시가에 대한 파생변수들을 사용하여 강화학습으로 주가지수 예측
- Target : (Up, Down) test accuracy : 0.53



주제



목차



배경 및 목표



활용 데이터



모델링 기법



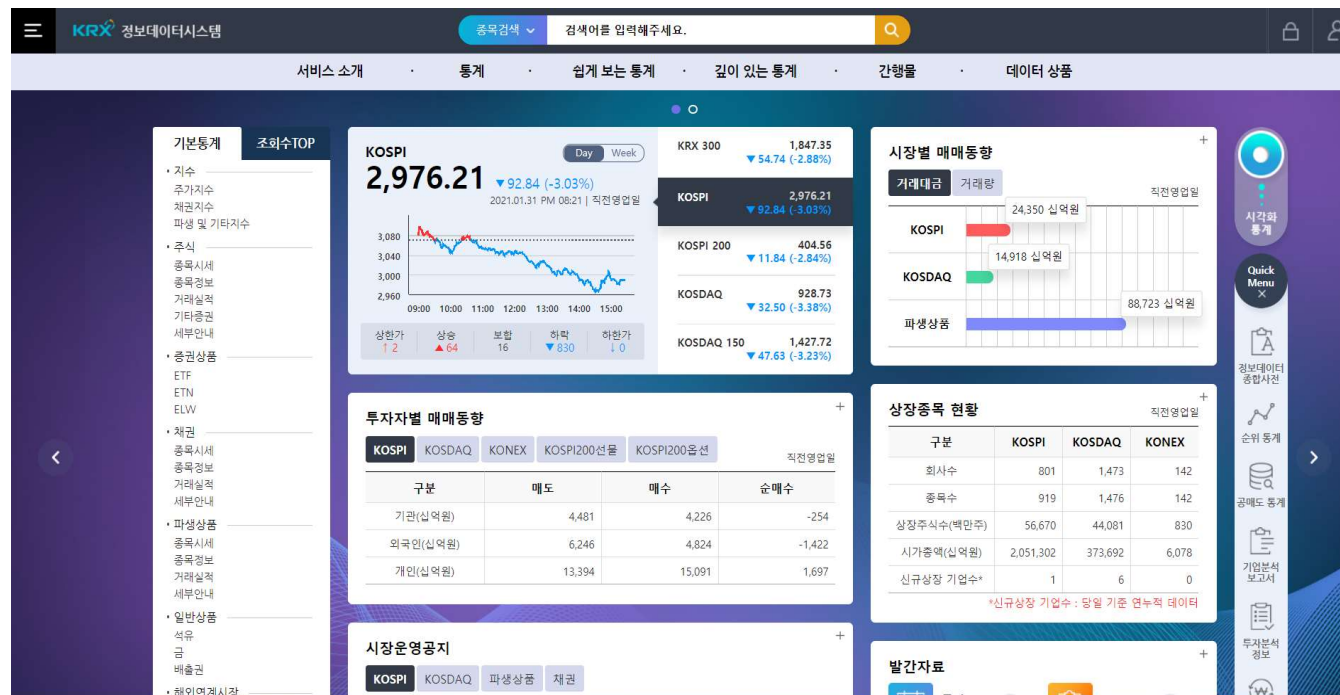
모델링



결과

KOSPI 지수 데이터

2016-2020 5년간의 일별 KOSPI 주가지수



한국거래소 정보데이터시스템 (www.krx.co.kr)



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

KOSPI 지수 증가



주가 지수는 **외부적인 영향**을 많이 받고, 매우 **불규칙적**임

→ 코로나의 영향으로 주가지수 급락, 이후 주식투자 열풍으로 급상승

Target : Trend



A hybrid stock trading framework integrating technical analysis with machine learning techniques

Rajashree Dash^{a,*}, Pradipta Kishore Dash^{b,2}

^a Computer Science & Engineering Department, ITER, Siksha 'O' Anusandhan University, Bhubaneswar, Odisha, 751030, India
^b Siksha 'O' Anusandhan University, Bhubaneswar, Odisha, 751030, India

Received 30 December 2015; revised 3 March 2016; accepted 8 March 2016
Available online 22 March 2016

Abstract

In this paper, a novel decision support system using a computational efficient functional link artificial neural network (CEFLANN) and a set of rules is proposed to generate the trading decisions more effectively. Here the problem of stock trading decision prediction is articulated as a classification problem with three class values representing the buy, hold and sell signals. The CEFLANN network used in the decision support system produces a set of continuous trading signals within the range 0–1 by analyzing the nonlinear relationship exists between few popular technical indicators. Further the output trading signals are used to track the trend and to produce the trading decision based on that trend using some trading rules. The novelty of the approach is to engender the profitable stock trading decision points through integration of the learning ability of CEFLANN neural network with the technical analysis rules. For assessing the potential use of the proposed method, the model performance is also compared with some other machine learning techniques such as Support Vector Machine (SVM), Naive Bayesian model, K nearest neighbor model (KNN) and Decision Tree (DT) model.

© 2016, China Science Publishing & Media Ltd. Production and hosting by Elsevier on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Stock trading; Stock trend analysis; Technical indicators; CEFLANN

“A hybrid stock trading framework integrating technical analysis with machine learning techniques”

Rajashree Dash, Pradipta Kishore Dash,
2016, The Journal of Finance and Data Science

(기계학습 기법과 기술 분석을 통합한
하이브리드 주식 거래 프레임워크)



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

Target : Trend

MA : 15일간 종가의 이동평균

$$MA_t = \frac{1}{t} \sum_{i=1}^t cp(i)$$



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

Target : Trend

UpTrend : 종가가 MA보다 큼 & 5일간 MA가 계속 증가한 상태

DownTrend : 종가가 MA보다 작음 & 5일간 MA가 계속 감소한 상태

Boxed : 그 외



🏠 주제

🔖 목차

📄 배경 및 목표

📄 활용 데이터

📄 모델링 기법

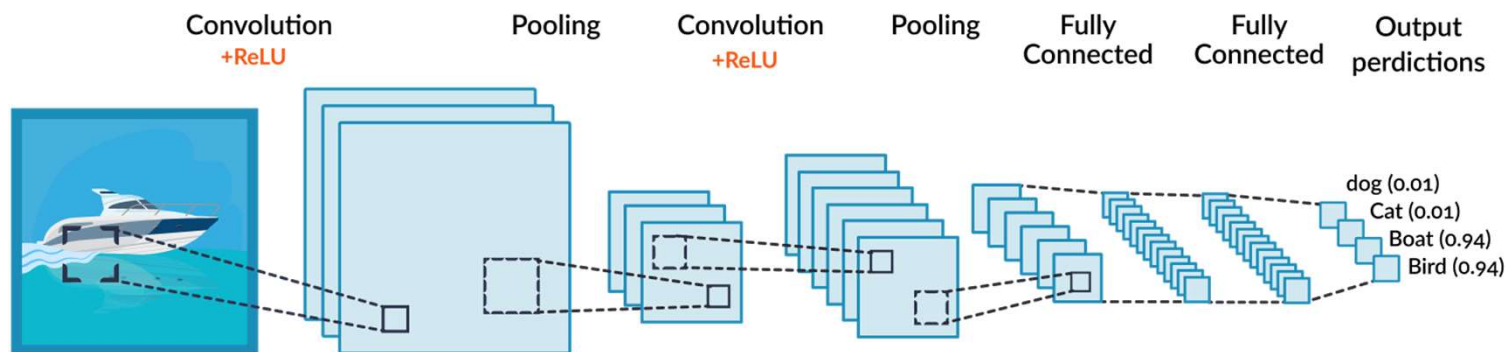
📄 모델링

📱 결과

C N N

Convolutional Neural Networks (합성곱 신경망)

- 공간적 특성을 추출할 수 있는 Convolution Filter 연산을 통해 **이미지의 특징**을 추출
- 입력데이터를 필터가 순회하며 합성곱을 계산하고, 그 계산 결과를 이용하여 Feature map을 만듦
- 필터를 사용하여 일반적인 DNN보다 상대적으로 **파라미터의 개수가 적음**
- 주로 이미지나 영상 데이터를 처리할 때 쓰임



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링

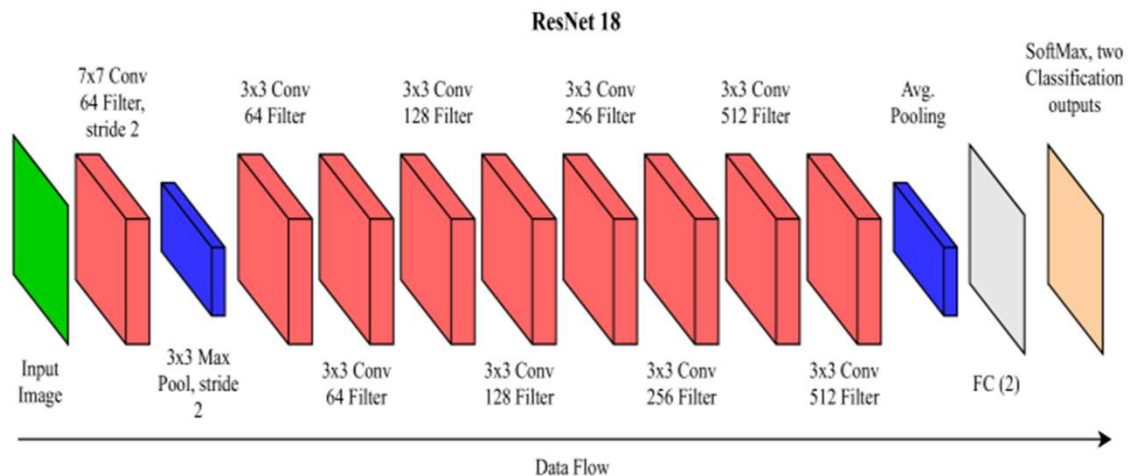


결과

ResNet

Residual Network

- ResNet은 이미지 인식 경진대회인 **ILSVRC**에서 우승한 모델
- **잔차(Residual)**를 **최소**가 되도록 학습시킴
- 층의 개수에 따라 ResNet-18/34/50/101/152
- Pytorch에서 제공하는 라이브러리 이용



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링

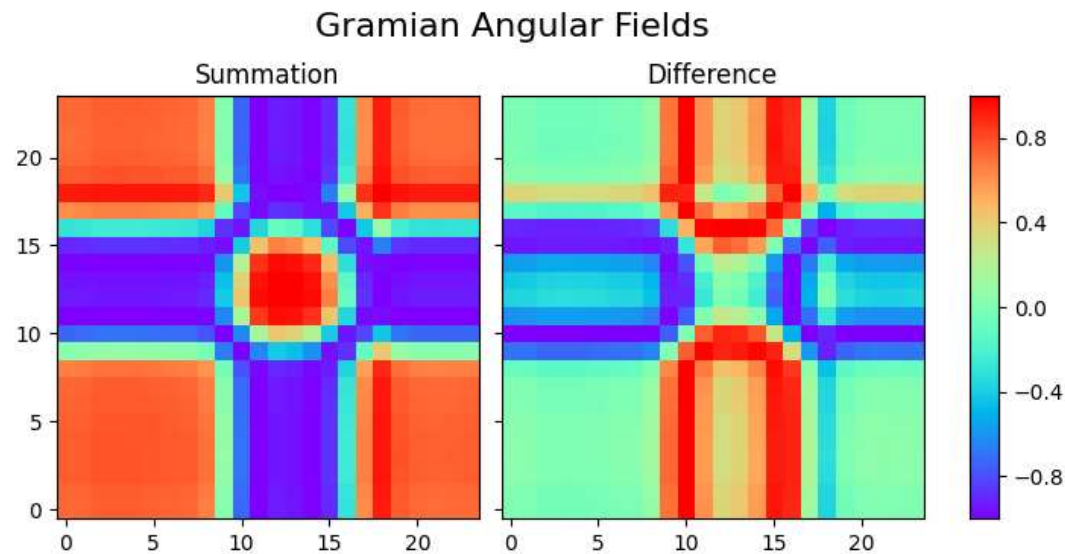


결과

G A F

Gramian Angular Field

- Temporal Correlation Matrix를 생성하고, \arccos 을 사용하여 시계열의 극좌표 계산
 - **GASF**(Gramian Angular Summation Field) : 각도 합계의 코사인
 - **GADF**(Gramian Angular Difference Field) : 각도 차이의 사인



주제



목차



배경 및 목표



활용 데이터



모델링 기법



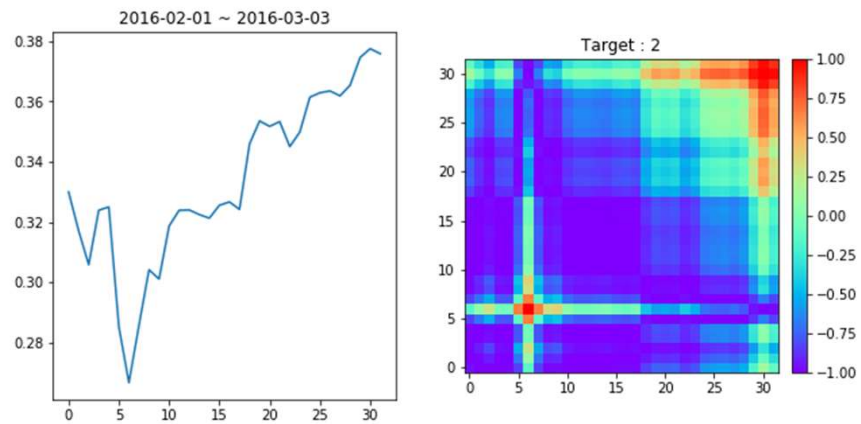
모델링



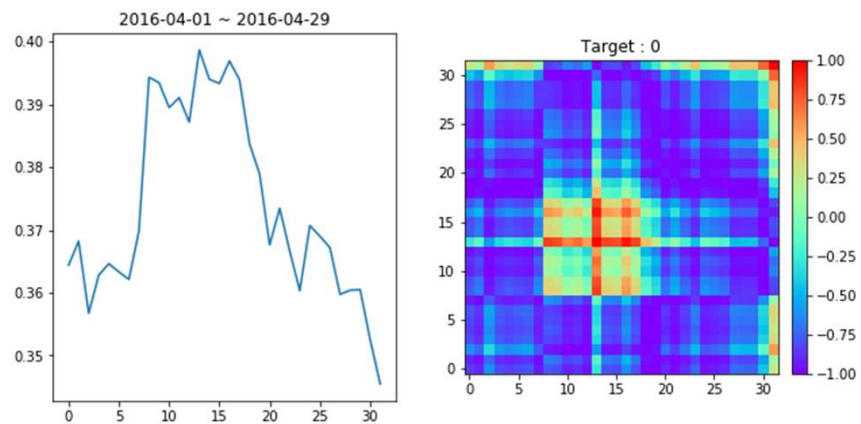
결과

G A F

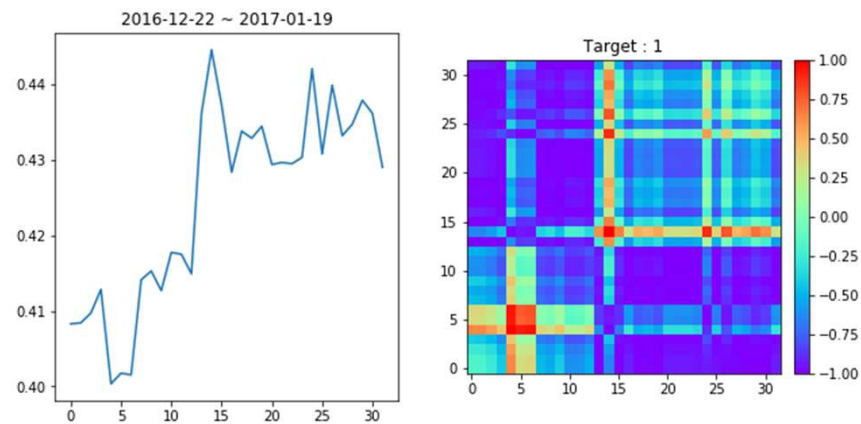
UpTrend



DownTrend



Boxed



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링

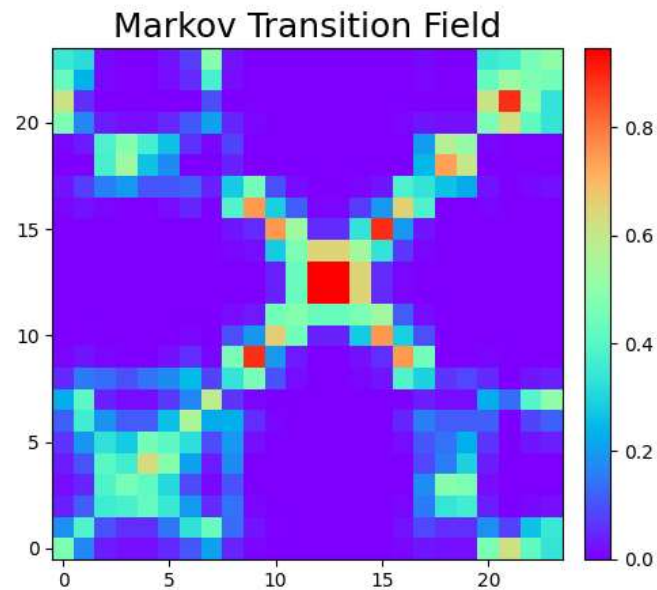


결과

M T F

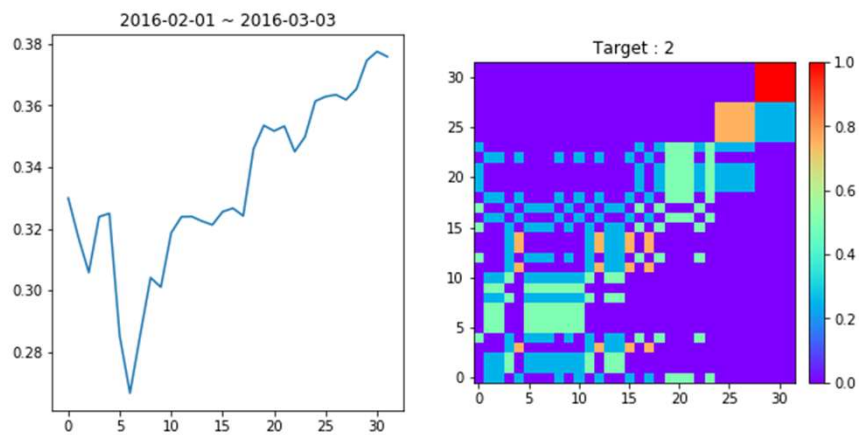
Markov Transition Field

- 시계열을 bin으로 나누고, 나뉘진 시계열의 **Markov Transition Matrix**를 계산
- 시간 정보 손실을 줄이기 위해 전이 행렬을 필드로 분산

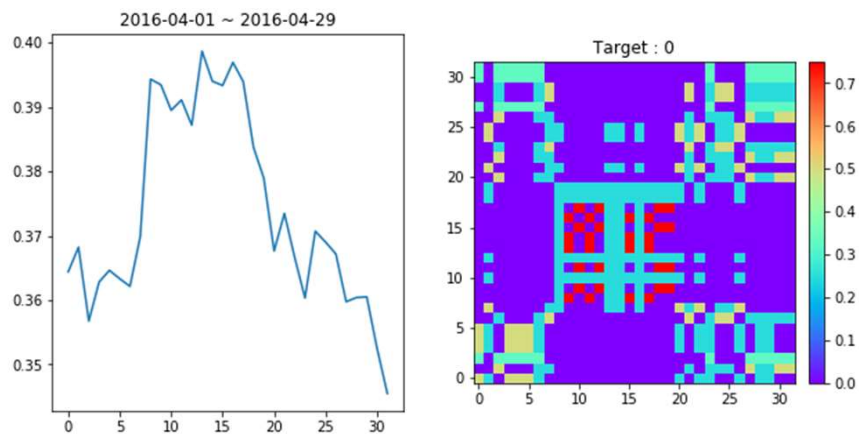


M T F

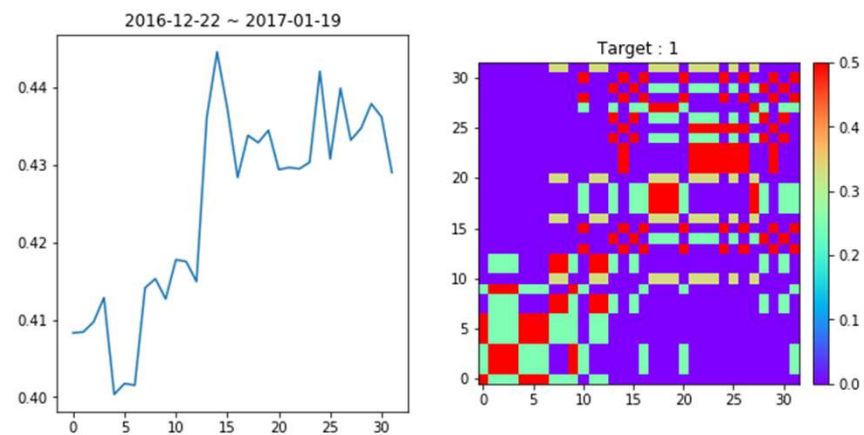
UpTrend



DownTrend



Boxed



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

모델링

- 5년간의 데이터를 7 : 3 으로 나누어 Train과 Test로 분리 (기준 : 2019-07-05)
- **MinMaxScaler**를 이용하여 최소값을 0으로, 최대값을 1로 정규화



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

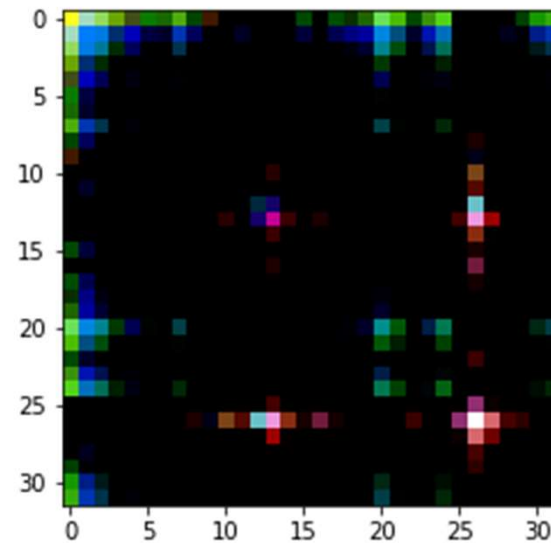
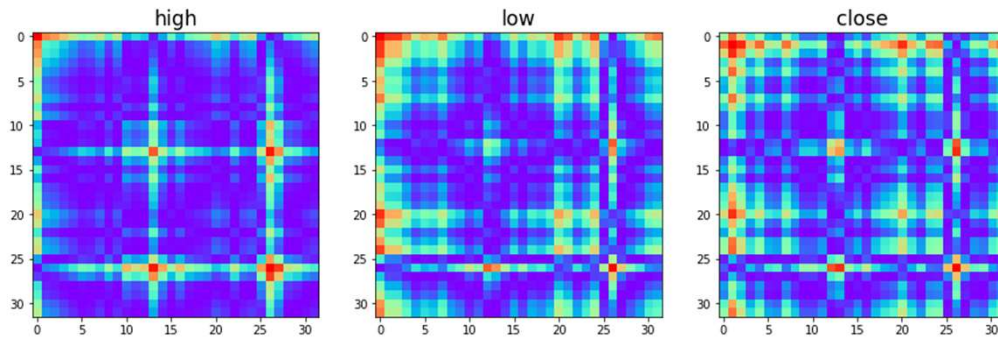
모델링

3 Channels (종가, 고가, 저가)

- 주식시장에서 중요하게 사용되는 캔들 차트에서 아이디어를 얻어, 종가, 고가, 저가를 각각 한 채널로 하여 분석

→ 예측 결과가 편향되는 문제 발생

이미지 간 특징의 차이가 분명하지 않아서 발생하는 문제로 추정



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링

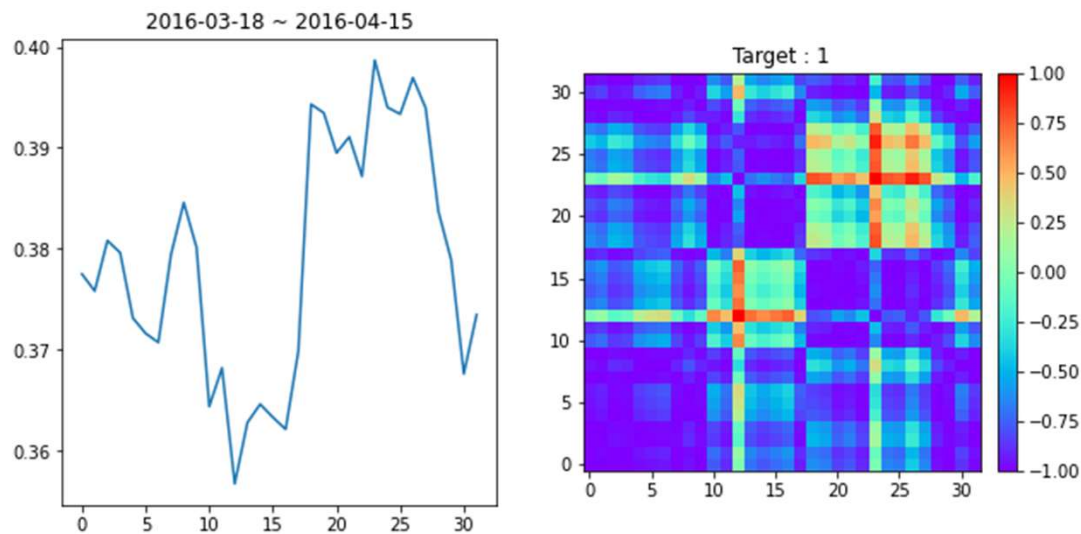


결과

모델링

1 Channel (종가)

- 32일간 KOSPI 주가지수의 종가를 **MTF** 기법을 이용하여 **32*32** size의 이미지 생성
- Pytorch에서 제공하는 **ResNet-18** 모델로 학습
- ResNet-18의 첫번째 Convolutional layer를 Channel 1개로 변형



🏠 주제

🔖 목차

📄 배경 및 목표

📄 활용 데이터

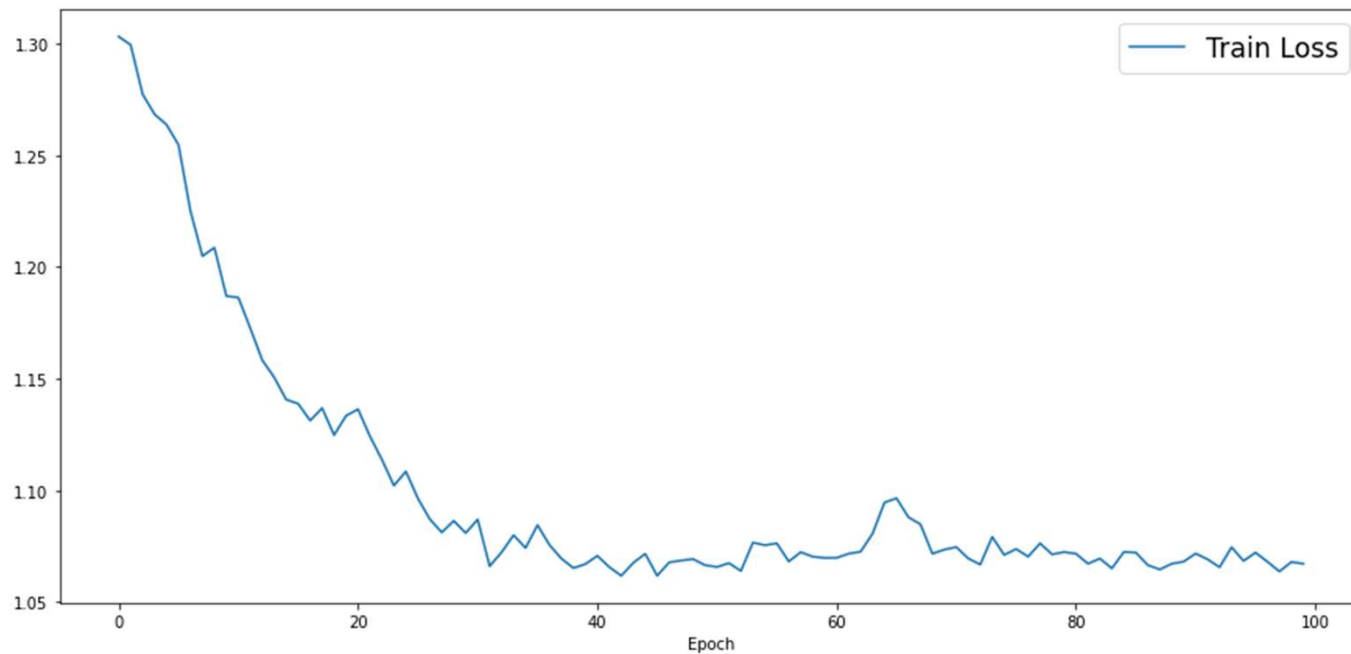
📄 모델링 기법

📄 **모델링**

📱 결과

모델링

- Loss Function : **Cross Entropy Loss**
- Optimizer : **Adam**
- Google Colaboratory GPU 환경에서 Pytorch 사용



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

결과



Train accuracy : 0.47

Test accuracy : 0.50



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

결과



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

결과

선행 연구와 비교

- SVM으로 3개의 클래스 분류 : 0.55
- RNN으로 2개의 클래스 분류 : 0.53
- 강화학습으로 2개의 클래스 분류 : 0.53

→ 모두 주가지수에서 파생된 **기술적 지표들을 활용**

주가지수의 **시계열적인 특징**만을 바탕으로 3개의 클래스를 분류하여 0.50의 성능
증가의 시계열 정보가 미래의 주가지수를 예측하기에는 부족함

“ 주식시장에서는 복리로 수익이 쌓이기 때문에 2개의 클래스를 분류하는데 51%를
초과하는 승률도 의미가 있다. ” (허준영, 양진용, “SVM 기반의 재무 정보를 이용한 주가 예측”. 2015)



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

활용성

주식시장 예측의 보조적인 수단

- 주식시장은 **외부적인 영향**이 크기에, 머신러닝 모델이 절대적인 방법은 될 수 없음
- 하지만 주식시장에서의 **의사결정을 도와주는 보조적인 역할** 가능

시각적 해석

CAM (Class Activation Map)

- 클래스 분류에 얼마나 영향을 주는지에 대해 히트맵으로 표현
- 다른 머신러닝 모델들과는 달리 CNN은 CAM을 이용하여 어느 시기의 주가지수를 보고 Trend를 판단했는지 알 수 있음 → **결과에 대한 해석이 시각적으로 가능**
- 결과에 대한 근거를 알 수 없는 머신러닝의 문제점인 **Black Box 해소** 가능



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

향후계획

경제적 지식 활용

- 주식시장에서 사용되는 경제적인 지식을 공부한 후 활용하여 **Target 변수와 이미지 size를 알맞게 조절**하여 성능을 개선해보고 싶음

모델 개선

- ResNet 모델은 224×224 이미지 분류에 최적화되어 복잡도가 높음
- 주가예측에 적합하도록 조금 더 가벼운 **모델을 직접 설계**하면 더 좋은 성능이 기대됨

CAM 모델 사용

- CAM 모델을 공부하여 이 분석에 적용해보고, **어떤 시기를 중심으로** 주가지수 예측이 이루어졌는지 알아보고 싶음



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과

감사합니다



주제



목차



배경 및 목표



활용 데이터



모델링 기법



모델링



결과