

투자 알고리즘 설명서

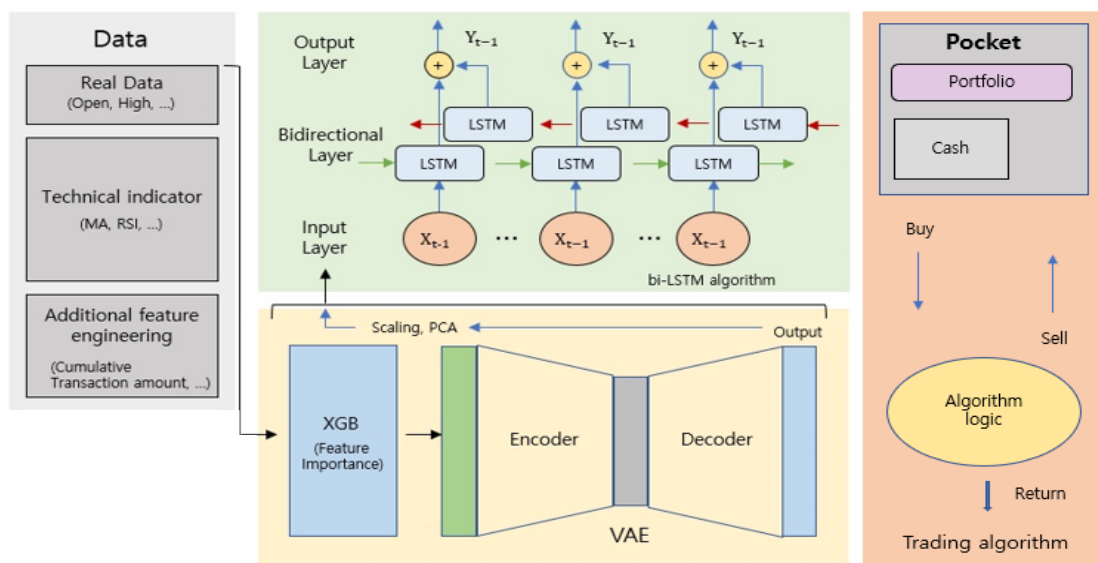
"딥러닝을 활용한 투자 알고리즘"

PAUL77MS

1. 소개

과거의 주식 시장은 기업 분석부터 투자 지표 분석을 통해 노동 집약적으로 주식투자가 이뤄졌다면 최근에는 인공지능(AI) 열풍이 주식투자에도 옮겨붙어, 다양한 인공지능 모델을 통한 주식 매매가 활발히 이뤄지고 있습니다. 하지만 인공지능 주식 매매에는 한계들이 존재합니다. 실제로 주식시장은 항상 바뀌기 때문에 수익을 내는 전문 투자자들, 금융 애널리스트들도 매년 현재 주식시장에 맞는 투자 지표를 찾는다고 합니다. 그리고 과거의 데이터만으로는 미래 주식을 예측할 수 없다는 것이고, 주식 시장에 영향을 미치는 팩터(factor)들이 다양하게 존재합니다. 이 한계들을 하나 하나 해결하면 과연 인공지능이 큰 성과를 낼 수 있을까? 고민해보았습니다. 인공지능에게 과거의 데이터는 물론이고, 바뀌는 주식시장에 맞는 투자 지표들을 선택하고 필요한 정보들만 모아서 압축하고 인공지능이 복잡한 패턴을 보다 정확하게 학습할 수 있도록 알고리즘을 구성하여 주가 예측에 존재하는 한계들을 하나 하나 해결하였습니다. 그리고 인공지능 추천 상위 종목과 우량주 상위 종목을 선택하여 포트폴리오로 만든 후 종목 특성에 맞게 투자전략을 적용하여 자동매매를 하도록 알고리즘을 구성하였습니다.

2. 작업 전체 아키텍처



3. 데이터

주가의 방향성을 예측할때 사용할 수 있는 필요한 모든 정보들을 인공지능이 가지고 있다면 인공지능은 다양한 측면과 각도에서 주식의 패턴을 할 수 있습니다. 기술적 지표부터 추가 피쳐엔지니어링등 가능한 많은 정보들을 통합했습니다. 실제 주가 데이터와 다음과 같은 정보들을 함께 통합했습니다.

우선 MultiProcessing 기술을 적용하여 Check-API를 통해 데이터를 빠르게 불러올 수 있도록 진행하였습니다. 그리고 거래량이 높은 순으로 상위 3개 종목을 제외한 196개 종목의 데이터를 추출하였습니다.

3.1. 기술 지표 - Cycle Indicators, Math Operator, Momentum Indicator, Overlap Studies, Pattern Recognition, Price Trasform, Statistic Functions, Volatility Indicators, Volume Indicators (총 154개 열)

3-2. 추가 피쳐 엔지니어링 - 누적 거래량, 누적 거래대금, Rolling, Lagging, 등 (총 19개 열)

4. 기능 공학

4-1. 기능 중요도 - 각 종목을 볼때 도움이 되는 기술 지표들이 있을 것이고, 각각 접근하는 방식이 모두 다를 것입니다. 따라서 각 종목에 맞는 정보들을 추출하기 위해 먼저 학습 환경을 만들어주기 위해 Train셋과 Test셋을 80 : 20 나눈 후 그리고 주가 데이터의 특성을 보면 주가 폭락, 폭등과 같이 값이 너무 커지거나 작아지는 경우가 존재하고, 우리는 종목 마다의 가격 수준이 아닌 주가 흐름과 관련된 필요한 정보들을 추출할 것이므로 정규화를 진행하였습니다. 성능이 가장 좋은 MinMaxScaler를 선택하였습니다. 외 나머지 독립변수들에 대해서는 StandardScaler를 사용하여 정규화를 진행하였습니다. 그리고 복잡한 데이터에서도 성능이 뛰어나고 설명력이 높은 트리기반 앙상블 알고리즘인 XGBoost를 통해 영향력이 있는 변수들을 추출하는 작업을 진행하였습니다. 과적합 방지를 위해 파라미터들은 초기화 상태로 두고 학습을 진행하였습니다. 변수 중요도를 수치화하고 주가 데이터에 영향을 조금이라도 주는 변수들을 추출하였습니다.

4-2. 상위 수준 기능 추출 - Variational Auto-Encoder(VAE) 알고리즘을 수행하여 주어진 데이터를 잠재 공간으로 압축(인코딩)하고 잠재 공간에서 임의의 노이즈를 더해서 새로운 특성들을 생성(디코딩)하는 작업을 반복 학습하여 상위 수준의 데이터로 재구성하였습니다.

4-3. Scaling 및 PCA를 통한 최적의 정보 생성 - 각 주식 종목마다 주가 가격은 천지차이입니다. 특성간의 규모 차이를 해소하기 위해 MinMaxScaling

작업을 해주었습니다. 그리고 신경망 알고리즘이 빠르고 정확한 학습을 하도록 주성분 기법을 통해 데이터의 분산의 80%를 설명할 수 있는 주성분을 선택하여 중요한 특성들을 보존하면서 더 낮은 차원으로 압축을 해주었습니다.

5. bi-LSTM (양방향 장단기 기억 장치)

우선 우리가 예측해야하는것은 종목별 미래 한달치 주가의 수익률을 구해야 하므로 슬라이딩 윈도우기법 을 적용해 다단계 예측을 하도록 구성하였습니다. 주가 데이터의 경우 과거의 가격과 정보들이 미래의 가격에 영향을 미칠 수 있습니다. 따라서 시간적 의존성을 모델링하기에 적합한 순환 신경망의 일종인 bi-LSTM 알고리즘을 선택하였습니다. bi-LSTM은 이전 타임 스텝과 이후 타임 스텝의 정보를 고려하는 양방향 순환 신경망으로 주가 방향성의 숨겨진 패턴을 더 잘 이해하고 모델링 할 수 있습니다.

(예선) 알고리즘 실험 결과를 바탕으로 최종적으로 양방향 LSTM을 선택하였습니다. * 성능 순서 : RNN-> DNN-> Transformer-> 단방향 LSTM, GRU-> 양방향 LSTM, GRU 양방향 LSTM *

각 종목별 미래 30일치의 주가 방향성을 예측 후 예상 수익률을 구하고 예상 수익률을 바탕으로 AI 추천 종목 포트폴리오를 만들었습니다.

6. 자동 매매 알고리즘

자동 매매 알고리즘을 설명하기에 앞서 코스피의 경우 코스닥보다 외국인, 기관 매도세 비율이 크고 투자에 리스크 요소가 많기 때문에 코스닥 종목만으로 투자를 진행하였고 개별 종목으로 투자함으로써 위험성을 분산시킬 수 있었고, AI 추천 상위 3개 종목과 거래량 단위 상위 3개의 종목을 두고 분할 종목 투자를 진행하였습니다. AI 추천 종목의 경우는 롱 전략 투자법으로 선택하였고 손실이 발생할 시 단가를 낮추기 위해 적은 수량을 추가 매수하고, 수익률이 3% 이상일 시 또는 수익률 -3% 많은 손실이 발생할 시 더 많은 손실을 멈추기 위해 전략 매도를 하도록 구성하였습니다. 그리고 거래량이 높은 우량주의 경우 보유 종목의 비중 20%를 꾸준히 추가 매수 하고 수익률이 1% 이상일시 전략 매도 -1% 이하일시 손절을 하여 꾸준히 누적 수익을 볼 수 있는 단기 매매법을 구성하였고, 매매를 통해 보유하지 않은 종목이 생기면 다시 대량 매수를 하고 다시 재매매를 하도록 하였고 매매를 할 때 발생하는 수수료, 액면분할 및 액면병합 이벤트 처리, 보유 잔고에 맞게 매매할 수 있도록 자동 매매 알고리즘을 구성하였습니다.