



재무비율과 기술적 분석을 통한 AI 주식 트레이딩 알고리즘 모델링

저자 (Authors)	정해성, 김용현, 임한준, 정기백, 정진대, 최원화
출처 (Source)	대한산업공학회 추계학술대회 논문집 , 2019.11, 3821-3835(15 pages)
발행처 (Publisher)	대한산업공학회 Korean Institute Of Industrial Engineers
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09272563
APA Style	정해성, 김용현, 임한준, 정기백, 정진대, 최원화 (2019). 재무비율과 기술적 분석을 통한 AI 주식 트레이딩 알고리즘 모델링. 대한산업공학회 추계학술대회 논문집, 3821-3835
이용정보 (Accessed)	연세대학교 121.162.235.*** 2020/08/27 15:26 (KST)

저작권 안내
DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information
Copyright of all literary works provided by DBpia belongs to the copyright holder(s)and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s)for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

2019 대한산업공학회
대학생 프로젝트 경진대회

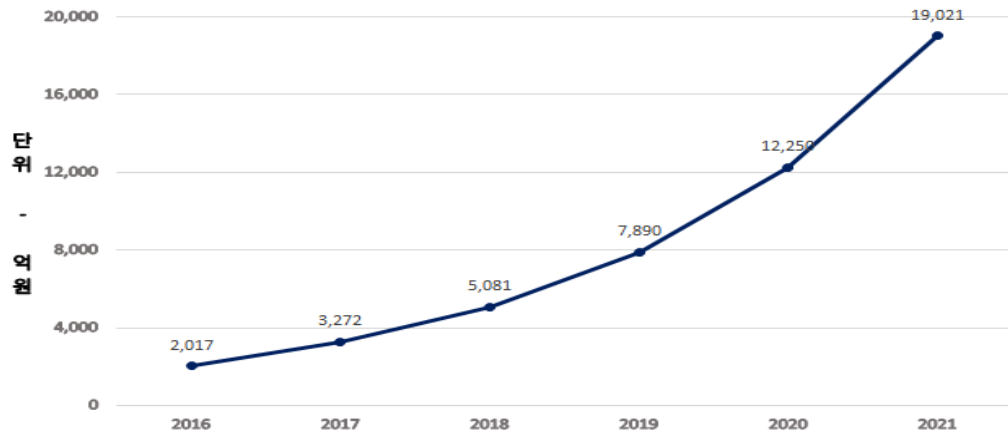
재무비율과 기술적 분석을 통한 AI 주식 트레이딩 알고리즘 모델링

< 한양대학교 ERICA 산업경영공학과 >

팀장 : 정해성
팀원 : 김용현
임한준
정기백
정진태
최원화

자산관리시장 트렌드 변화

국내 로보 어드바이저 시장 규모 전망

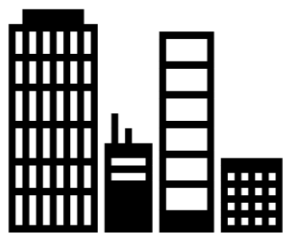


AI 자산관리 서비스 도입 활발

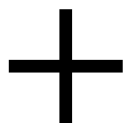
고액 자산가의
고액투자로보 어드바이저
서비스 도입대중 부유층의
소액투자

“중위험 중수익”

연구 목적



가치투자



모멘텀 투자

“안정적인 수익률 구현”

본 연구에서는 재무비율 데이터를 기반으로 한 기업선정모델과 주가데이터와 거래지표 데이터를 기반으로 한 모멘텀 투자모델의 앙상블을 통한 'AI 주식 트레이딩 봇'을 구현하여 '안정적인 수익률' 달성을 목적으로함

데이터 수집 & 전처리

모델링

검증

OUTPUT

기업 선정 모델

재무비율 데이터

수정주가 데이터

데이터
전처리

분류 알고리즘

투자 백테스팅

검증

최적
포트폴리오

모델 최적화

모멘텀 투자 모델

주가 데이터

거래 지표 데이터

데이터
전처리LSTM
알고리즘매수, 매도
예측MC Policy
Gradient투자
백테스팅최적 매수,
매도 예측

반복학습

최종수익률

데이터 수집



DATAGUIDE 5.0

재무비율
(42780행, 109열)수정주가
(7220행, 2139열)주가
(7220행, 8556열)거래지표
(7220행, 8556열)

수집기간

2000년 1월 3일 ~ 2019년 9월 30일

사용사이트

DATAGUIDE 5.0

데이터 구성

재무비율			Features(Annual)	재무비율			Features(Annual)	수정주가	Features (Daily)	주가	Features (Daily)	거래지표	Features (Daily)
KOSPI KOSDAQ 상장기업 2138개	안정성 변수 21개	타인자본비율(%) ... 유보액/총자산(%)		KOSPI KOSDAQ 상장기업 2138개	주당지표 변수 6개	EPS(원) ... 수정EBITDAPS(원)		KOSPI KOSDAQ 상장기업 2138개	수정주가	KOSPI KOSDAQ 상장기업 2138개	시가 종가 고가 저가	KOSPI KOSDAQ 상장기업 2138개	거래량 외국인순매수량 기관순매수량 개인순매수량
	성장성 변수 27개	영업수익증가율 (전년동기)(%) ... DPS증가율(보통주, 전년동기)(%)			주가배수 16개	PER(배) ... PEGR(-)							
	수익성 변수 20개	매출총이익률(%) ... ROE(%)			배당관련 변수 9개	DPS(보통주, 현금 +주식)(원) ... 배당성향							
	활동성 변수 3개	총자산회전율(회) ... 총자본회전율(회)			기업관련 변수 4개	매출비중1(결산) ... 코스닥 대형주							

데이터 전처리

재무비율

수정주가

데이터 병합 & Target 생성

결측치 제거

결측치 대체
(선형보간법)

Scaling

최종 데이터 셋

데이터 병합 & Target 생성

❖ 재무비율

Name	회계년	타인자본비율	자기자본비율	...
삼성전자	2000	63.84	36.16	...
삼성전자	2001	60.86	39.14	...

- K-IFRS가 회계 기준 대상의 상장기업들은 사업연도의 **차기연도 3월 31일**까지 사업보고서를 제출하여 재무정보데이터를 얻을 수 있음

❖ 수정주가

날짜	삼성전자	SK하이닉스
2001년 4월 2일	3990	60656
...
2002년 3월 29일	7360	28267

- 매년 **공시마감일의 익일**을 투자 시작일로 가정하고 **차기연도 공시마감일**을 투자 종료일로 가정하여 각 투자 시작, 종료일의 수정주가를 재무비율 데이터와 병합

❖ Target 생성

시작주가	종료주가	수익률	수익
3990	7360	0.845	1
7480	5680	-0.241	0

- **1년 적정 수익률 5%를 기준으로** 5%이상인 경우 1, 이하인 경우 0으로 분류

사업보고서 제출일을 기준으로
1년간 주가의 수익을 타겟으로
구성하여, 주가가 상승할 기업을 분류

3825

결측치 처리

❖ 결측치 제거

- 상장일 이전, 상장폐지 기업의 데이터
총 7696개 데이터 삭제

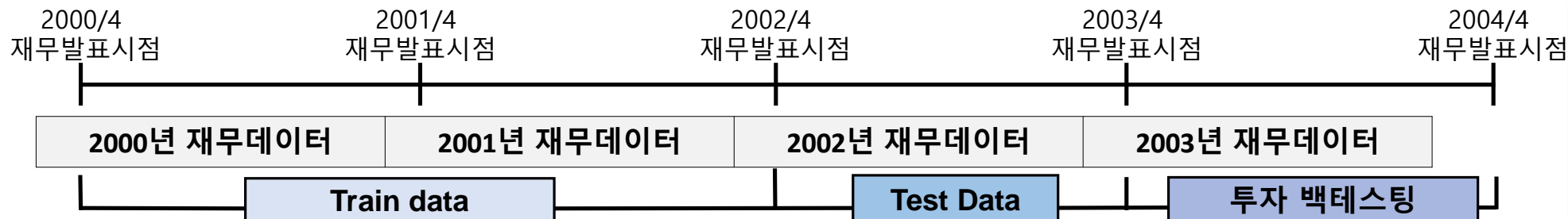
❖ 결측치 대체(선형보간법)

$$y = y_0 + (y_1 - y_0) \frac{x - x_0}{x_1 - x_0}$$

- GAAP에서 K-IFRS로 회계기준이 바뀐 연도에
대하여 총 7982개 결측치를 선형보간법을
이용하여 대체

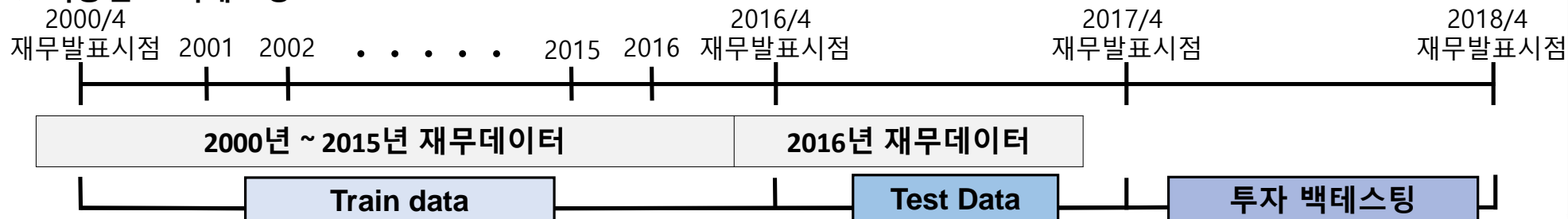
투자 백테스팅

❖ 초기년도 백테스팅



2000년과 2001년 데이터를 실제 값과 학습시켜 모델을 생성한 후, 2002년 데이터를 통해 분류된 결과를 통해 2003년 ~ 2004년까지 투자 백테스팅을 진행

❖ 최종년도 백테스팅



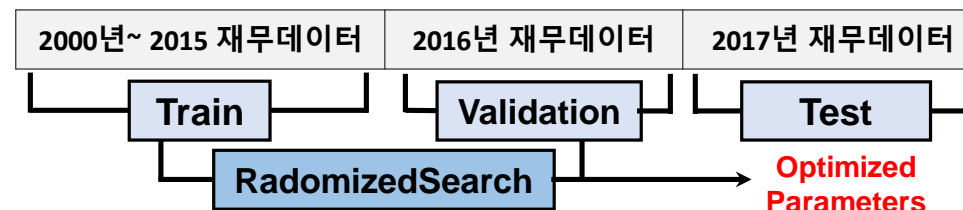
이러한 구조로 연도별 학습을 통한 모델 생성과 투자 백테스팅을 진행

분류 모델 선정

알고리즘	RandomForest	XGboost	Logistic Regression	SVM
정밀도 평균	0.65	0.49	0.44	0.38
정밀도 표준편차	0.18	0.27	0.21	0.31
수익률 평균	20%	11%	19%	8%
수익률 표준편차	22%	23%	22%	17%

- 알고리즘의 정밀도와 수익률을 검정기준으로 각 평균과 표준편차를 비교하여 알고리즘의 안정성을 평가하여 RandomForest를 최종모델로 선정

모델 최적화



- Train의 최종년도 데이터를 Validation 데이터로 활용하여 Sklearn에서 제공하는 RandomizedSearch를 통해 Model Fitting을 진행함

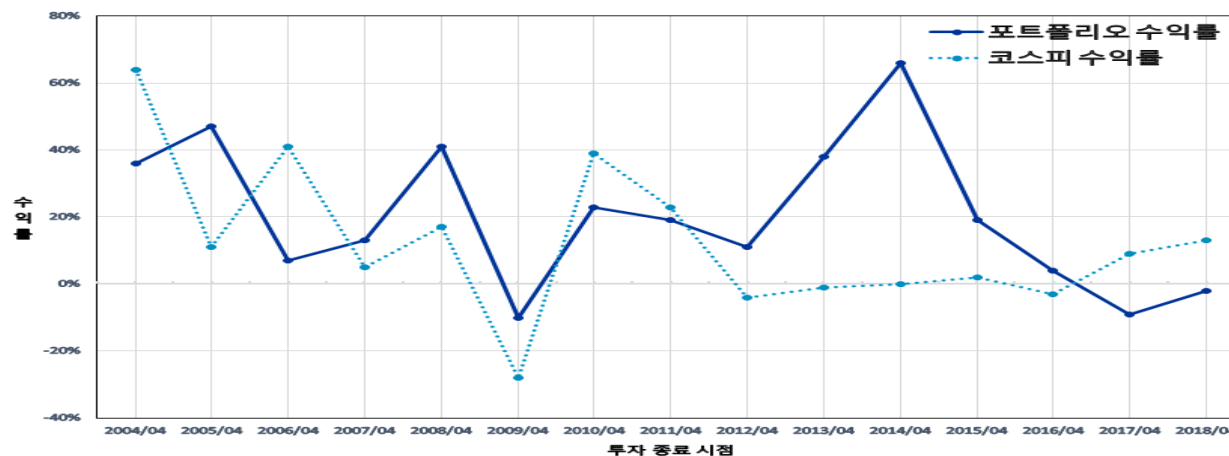
투자 포트폴리오

Name	시작주가	종료주가	수익률	수익	확률
지엠비코리아	5340.0	9820.0	83%	1	0.86
기아차	36850.0	31000.0	-15%	0	0.81
삼호개발	4395.0	5360.0	21%	1	0.80
대원산업	7390.0	7200.0	-2%	0	0.79
서연전자	3060.0	2300.0	-24%	0	0.78
유진기업	5490.0	6100.0	11%	1	0.77
유니크	3320.0	5980.0	80%	1	0.75
서연	10250.0	6370.0	-37%	0	0.74
...

- 상승할 것으로 분류된 기업 중 **확률값 상위 5개 기업**을
포트폴리오로 구성하여 투자를 진행

투자 백테스팅 결과

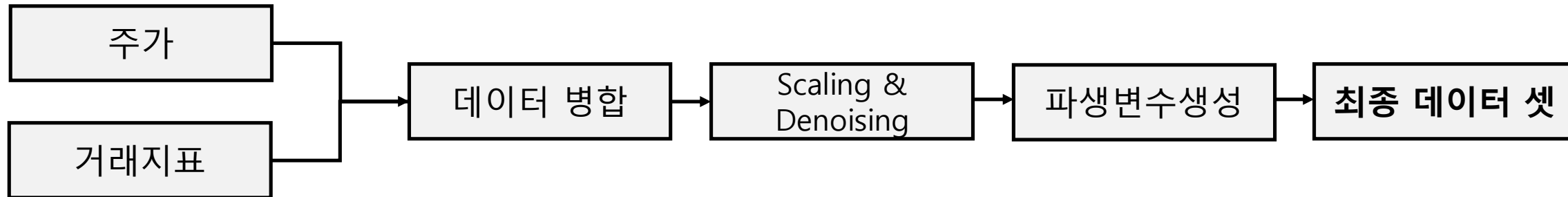
투자 연도별 수익률



투자종료	2004/04	2005/04	2006/04	2007/04	2008/04	2009/04	2010/04	2011/04
포트폴리오 수익률	36%	47%	7%	13%	41%	-10%	23%	19%
KOSPI 수익률	64%	11%	41%	5%	17%	-28%	39%	23%
투자종료	2012/04	2013/04	2014/04	2015/04	2016/04	2017/04	2018/04	-
포트폴리오 수익률	11%	38%	66%	19%	4%	-9%	-2%	-
KOSPI 수익률	-4%	-1%	0%	2%	-3%	9%	13%	-

종목선정 모델의 연도별 백테스팅 결과를 KOSPI 수익률과 비교한 결과 매년 KOSPI 수익률을 상회하는 수익률을 올림, 하지만 1년간 경제상황을 고려하지 않는 알고리즘의 특징 상 경제위기 상황에서 수익률이 저하되는 현상을 보임, 이를 LSTM과 강화학습을 통한 모멘텀 투자 모델로 수익의 안정성을 더함

전처리 Process



데이터 전처리

❖ 원 데이터

시가	고가	저가	종가	거래량	순매수량(외국인)	순매수량(기관)	순매수량(개인)
11900	13650	11900	13650	26150	-15200	16580	-1480
17950	17950	16600	17950	31250	54470	-56570	900



❖ 전처리 데이터

시가	고가	저가	종가	거래량(5일)	거래량(10일)
0.007042	0.009802	0.006902	0.009782	0.0693	0.07345
0.009201	0.012932	0.007102	0.009923	0.06969	0.08203

3828

❖ Standard Scaler

$$Z = \frac{x - \mu}{\sigma}$$

– 주가 데이터에 대해 Standard Scaler를 적용해 데이터의 범위를 조정

❖ 5일 / 10일 / 20일 / 60일 이동평균 비율 생성

$$MA = \frac{\sum_{i=0}^{n-1} P_{d-i}}{n}$$

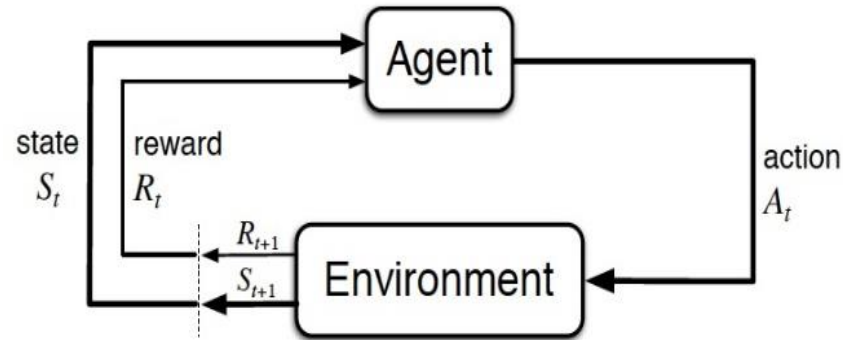
– 이동평균비율을 통해 거래량과 주가 데이터의 단위 통일

❖ 액면분할, 액면병합, 기업분할, 기업병합 등 데이터 처리



– 액면분할 및 병합비율로 전처리 진행

강화학습 설정

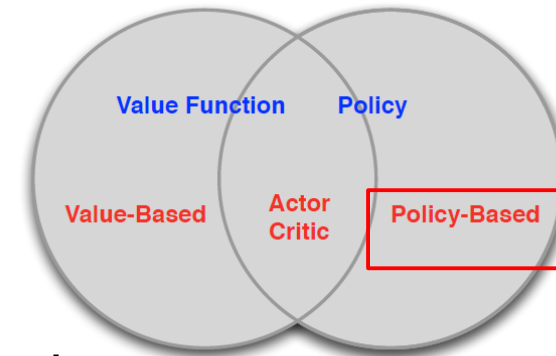


Parameter setting

Parameter setting	
Action	매수, 매도, 관망
Environment	주가데이터, 거래지표
state	주식 보유개수, 계좌 잔액, 현재 주시가격...
Agent	행동을 수행하는 투자자
Reward	긍정적 학습 : +1, 부정적 학습 : -1
Reward 규칙	손익률 3% 기준

Policy-based

❖ 강화학습 모델의 범주



❖ Value based

- 미래의 보상을 최대화하기 위한 알고리즘

특정 행동을 취함으로써 얻어지는 보상에 대해 계산

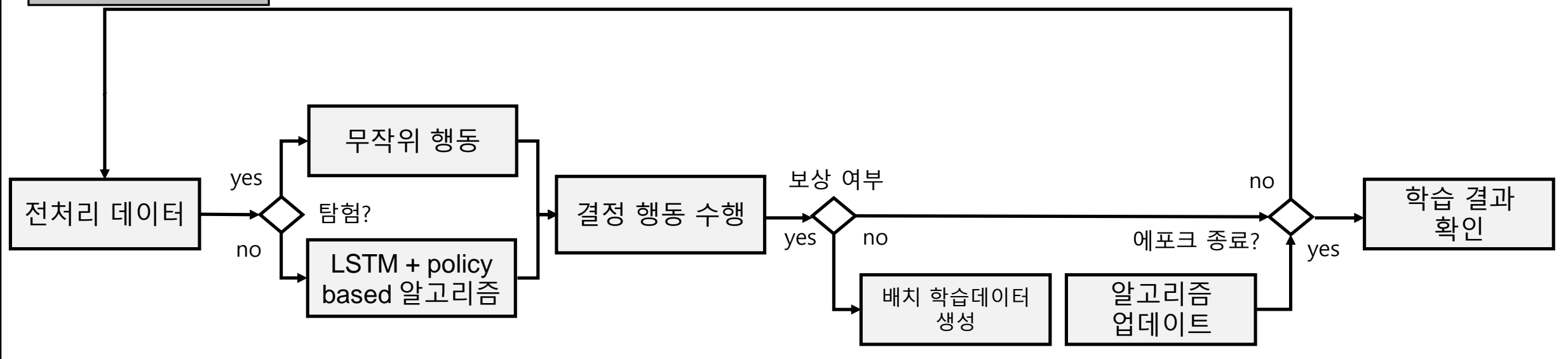
$$v_{\pi}(s) = \mathbb{E}_{\pi}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots | S_t = s]$$

❖ Policy based

- 특정 상태에서 Agent가 어떤 행동을 해야 하는가에 대한 알고리즘
- 현재 상태를 바탕으로 최적화된 행동을 Stochastic 한 방법으로 판단

$$\pi(a|s) = \mathbb{P}[A_t = a | S_t = s]$$

모델 전개도



탐험 Algorithm

❖ Decaying ϵ - greedy algorithm

– ϵ 의 확률로 임의의 행동 결정

– LSTM+Policy based 알고리즘만으로 모델을 구성하면 최적 값이 아닌 값에 수렴할 수 있으므로 일정 비율로 랜덤 행동을 수행

❖ 목적식 : $\text{MIN}(\text{regret function}) = \text{MAX}(\text{누적 reward})$

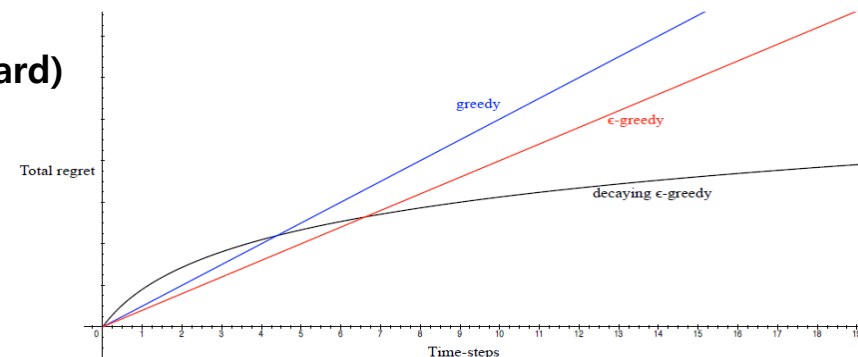
Regret function

$$L_t = \mathbb{E} \left(\sum_{\tau=1}^t V^* - Q(a_\tau) \right)$$

V^* = 최적 value

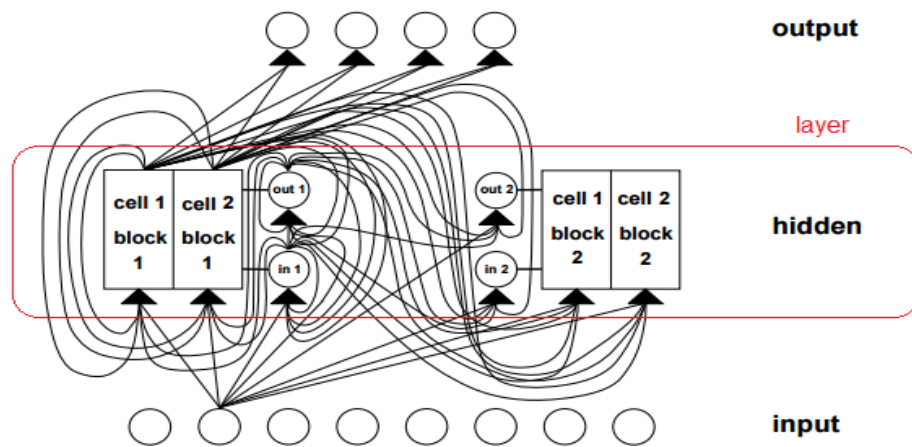
$Q(a_\tau)$ = 행동에 대한 평균 보상값

3830



– ϵ -greedy, greedy 함수에 비해 Decaying ϵ -greedy 함수가 total regret을 최소화

LSTM model



Parameter setting

Parameter setting	
Input layer	전처리된 주가 데이터, 거래지표
Hidden layer	256차원 : 2개 128차원 : 1개
Output layer	매수, 매도, 관망
Optimizer	Adam Optimizer
Loss function	Mean Squared Error
Overfitting 방지	Dropout 설정 및 배치 정규화

Policy based algorithm

❖ Monte Carlo policy gradient

$$\Delta \theta_t = \alpha \nabla_{\theta} \log \pi_{\theta}(s_t, a_t) v_t$$



LSTM 신경망으로
parameter 업데이트

❖ Softmax policy

$$\pi_{\theta}(s, a) = \frac{e^{\varphi(s, a)^T \theta}}{\sum_{k=1}^N e^{\varphi(s, a_k)^T \theta}}$$

매수/매도/관망

주가 데이터는 현재 데이터를 기준으로 다음 주가에 영향을 미치기 때문에 Markov process라고 가정

model-free 알고리즘인 Monte Carlo Policy Gradient 사용

policy gradient에서는 stochastic한 policy 학습을 통해 확률이 가장 높은 행동 출력

3831

Output

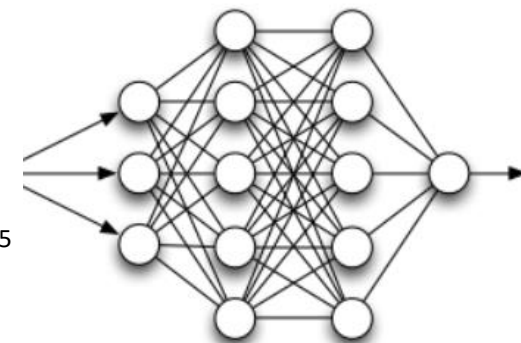
Input

2019-10-04
시가: 0.007042
고가: 0.09802
저가: 0.006902
종가: 0.009782
거래량: 0.07345

⋮
⋮
⋮

“탐험률 : 0.000 매수 : 1 매도 : 0 관망 : 0 PV : ₩10,060,000”

Final Model



Output

2019-10-05

“경동 나비엔
1주 매수”

최종연도 백테스팅 결과

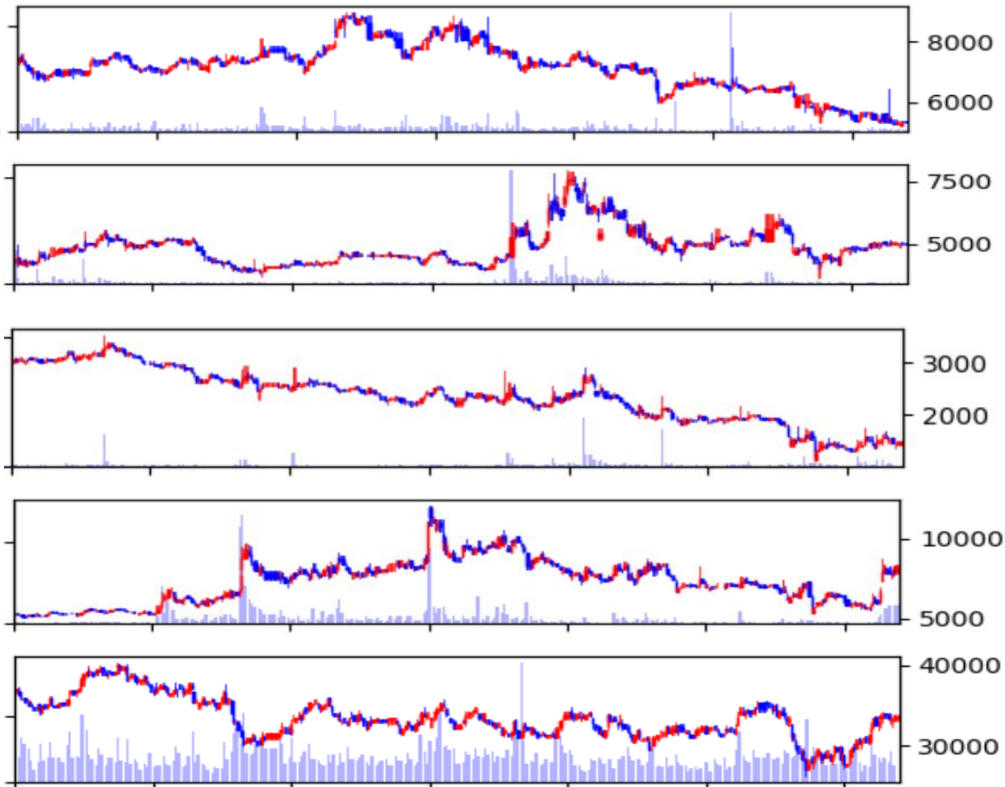
❖ 백테스팅 가정

- 투자 기간 : 2017년 3월 31일 ~ 2018년 4월 1일
- 자본금 : 기업 당 1000만원 총 5000만원
- 포트폴리오 기업명 : 대원산업, 삼호개발, 서연전자, 지엠비코리아, 기아차

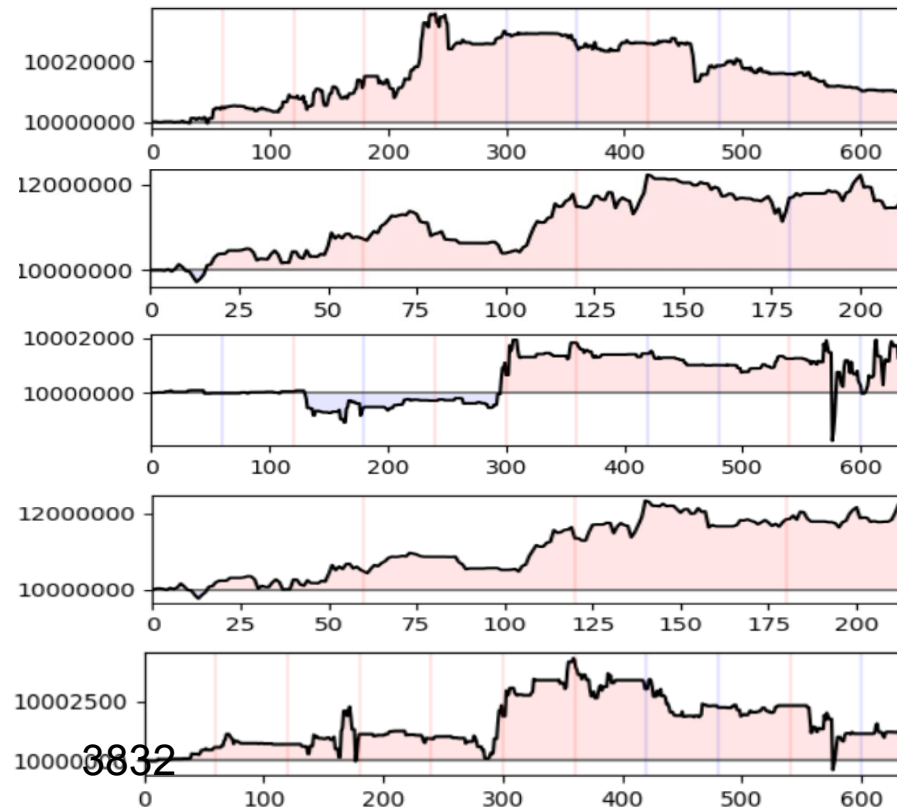
❖ 백테스팅 결과

포트폴리오 수익률	코스피 수익률	모멘텀 투자 수익률
-2%	13%	7.52%

❖ 차트그래프(X축 : 날짜, Y축 : 주가)



❖ 수익률 그래프(X축 : 행동 수, Y축 : 수익)



기업명 : 대원산업
수익률 : 0.015%

기업명 : 삼호개발
수익률 : 16.89%

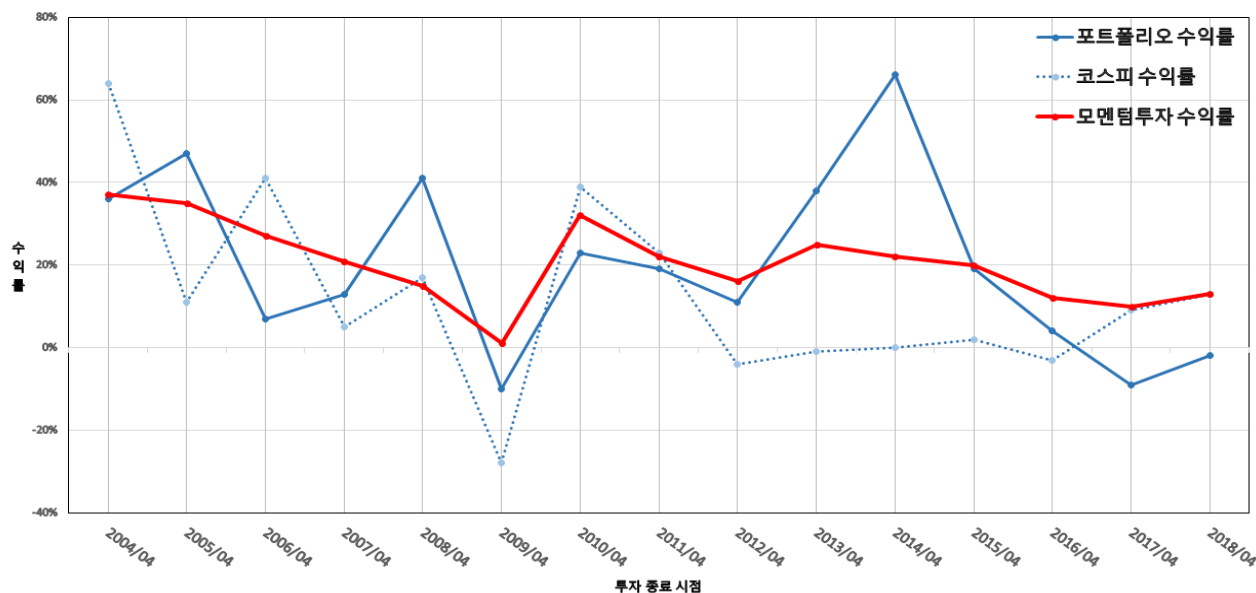
기업명 : 서연전자
수익률 : 0.0016%

기업명 : 지엠비코리아
수익률 : 19.215%

기업명 : 기아차
수익률 : 0.0014%

최종결과

투자 연도별 수익률



투자종료	2004/04	2005/04	2006/04	...	2017/04	2018/04
포트폴리오 수익률	36%	47%	7%	...	-9%	-2%
KOSPI 수익률	64%	11%	41%	...	9%	13%
모멘텀 모델 수익률	37%	35%	27%	...	10%	13%

	기업선정모델 수익률	KOSPI 수익률	모멘텀 투자 모델 수익률
평균	20%	13%	21%
표준편차	21.8	22.3	9.8

❖ 결론

기업선정모델과 모멘텀투자모델의 앙상블을 통해 15년이라는 장기간 동안 KOSPI수익률을 상회하고, 안정적인 수익률을 창출하는 안정적인 투자 모델을 구현

기대효과 및 발전방향

장기투자를 통해 '중위험 중수익'을 원하는 '대중부유층'을 위한 AI자산관리 서비스로 발전

실시간 주가데이터를 사용하여 모멘텀의 단위를 하루에서 분단위로 나누어 더 높은 수익률을 창출하는 모델 개발 가능

역할분담

김용현	PPT총괄, 데이터 전처리
임한준	강화학습모델 구현, PPT제작
정기백	데이터수집, 논문서치, PPT제작
정진태	기업모델선정 알고리즘 연구
정해성	모멘텀 투자모델 알고리즘 연구
최원화	자료조사, 결과 분석, PPT제작

프로젝트 개발환경

	항목	환경
1	Programming Language	Python
2	Programming Tool	Pycharm, Jupyter Notebook, Excel
3	Data Analysis Library	Scikit-learn, Keras, Tensorflow

NO.	참고 자료
1	Convolutional LSTM network: A machine learning approach for precipitation nowcasting -Xingjian SHI, 2015
2	LSTM-based deep learning models for non-factoid answer selection – M.Tan, 2015
3	LSTM fully convolutional networks for time series classification – F.Karim, 2017
4	Analysis of a classification-based policy iteration algorithm – A.Lazaric, 2010
5	A LSTM-based method for stock returns prediction: A case study of China stock market – K.Chen, 2015
6	국내 기업의 주요 재무비율과 해외 직접투자 간의 관련성 -신현대, 2011
7	The behavior of stock-market prices – E.F Fama, 1965
8	양방향 LSTM 순환신경망 기반 주가예측모델 - 주일택, 2018
9	Moving Average Prediction for Continuous Time Linear Systems - IY Song, 2019
10	Prediction of stock market price using hybrid of wavelet transform and artificial neural network - S.K Chandar, 2016
11	금융시장 로보어드바이저 산업에 대한 고찰: 현황과 개선방안 -임혜진, 2018
12	Stock market prediction using an improved training algorithm of neural network – M.Billah, 2016
13	Robo-Advisors: A Substitute for Human Financial Advice? – L.Brenner, 2019