



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

암호화폐 분석을 위한 인공지능 기반의
통합모델 제안

Proposal of Integrated Model based on Artificial
Intelligence for Cryptographic Currency Analysis

조 성 근

한양대학교 공학대학원

2018 년 8 월

석사학위논문

암호화폐 분석을 위한 인공지능 기반의
통합모델 제안

Proposal of Integrated Model based on Artificial
Intelligence for Cryptographic Currency Analysis

지도교수 조 인 휘

이 논문을 공학 석사학위논문으로 제출합니다.

2018 년 8 월

한양대학교 공학대학원

컴퓨터공학 전공

조 성 근

이 논문을 조 성 근 의 석사학위논문으로 인준함

2018 년 8 월

심사위원장 이 병 호 (인)

심사위원 조 인 휘 (인)

심사위원 손 규 식 (인)

한양대학교 공학대학원

국문요지

최근 기술인 블록체인은 금융기관과 콘텐츠 유통 등 다양한 분야에 혁신을 주도할 것으로 예상된다. 또한 기업과 기업(B2B) 영역에서부터 기업과 개인(B2C) 영역까지 영향을 끼칠 것이다.

블록체인 기술이 관심을 받게 되면서 암호화폐들이 주목받고 있는 시점이다. 블록체인들의 원장 정보를 처리하면서 주어지는 보상이 암호화폐이다. 다양하고 빠르게 가격 변동이 이루어지는 암호화폐들의 가치에 대한 평가는 점점 더 복잡해지고 있다.

본 논문은 이와 같은 문제에 대해 가치를 정확하게 평가하기 위해, 알고리즘 모델과 인공지능 모델을 통합한 모델을 제안한다.

첫 번째로는 알고리즘 모델이다. 흔히 주식 분석에서 표현되는 여러 가지 기법들을 통칭한다. 기본적인 골든 크로스 기법을 활용하였다.

두 번째로는 인공지능 모델이다. 소셜 분석 모델과 가치 예측 모델로 구성되었다. 오피니언마이닝(Opinion mining)의 한 종류인 감성 분석은 텍스트 데이터를 분석하는 방법이다. 작성자의 글 내용의 감성을 긍정적인지 부정적인지에 대하여 구분하는 분석 기법이다.

기계학습과 인지과학에서 사용되는 통계학적 학습 알고리즘인 인공신경망 중 LSTM을 활용하였다. 기존의 데이터를 활용해서 트레이닝을 하고 이를 바탕으로 미래의 가치를 예측한다. 여기서는 예측된 가치의 기울기를 중심으로 확인한다.

암호화폐 중 이더리움을 기준으로 2018년 01월 29일부터 2018년 05월

08일까지의 데이터를 활용하여 위에서 언급한 기존의 알고리즘 모델과 통합 모델을 비교하여 평가하였을 때, 알고리즘 모델의 경우 -15%의 수익률을 나타내었고, 통합모델의 경우 76%의 수익률을 나타내었다.



차 례

국문요지.....	i
차 례.....	iii
그림 차례.....	v
표 차례.....	vi
제 1 장 서 론.....	1
1.1. 연구의 배경 및 필요성.....	1
1.2. 연구 내용.....	1
제 2 장 관련 연구.....	3
2.1. 기존연구.....	3
2.1.1. 감성분석에 관한 연구.....	3
2.1.2. 다변량 분석에 관한 연구.....	3
2.1.3. 머신러닝에 관한 연구.....	4
2.2. 알고리즘 모델에 대한 고찰.....	5
2.2.1. 이동 평균 개념.....	5
2.2.2. 골든 크로스 개념.....	6
2.3. 인공지능 모델에 대한 고찰.....	6
2.3.1. 블록체인과 암호화폐 개념 및 특징.....	6
2.3.2. 웹크롤링 개념 및 특징.....	9
2.3.3. 감성 분석 개념.....	10
2.3.4. RNN 개념 및 특징.....	11
2.3.5. LSTM 개념 및 특징.....	12

2.3.6. 로보 어드바이저 개념 및 특징.....	14
제 3 장 가치 기반 통합 모델.....	17
3.1. 통합 모델 제안 개요.....	17
3.2. 개발 도구.....	19
3.3. 개발 도구 설명.....	20
3.4. Application Architecture.....	23
제 4 장 통합 모델 구현 및 성능평가.....	26
4.1. 구현내용.....	26
4.2. 정성평가 및 정량평가.....	34
제 5 장 결론 및 향후 연구.....	37
5.1. 결론.....	37
5.2. 향후 연구.....	38
참고문헌.....	39
ABSTRACT.....	41
Résumé.....	43
감사의 글.....	45

그림 차례

그림 1 블록체인의 기본 구조 <출처:파이널셀타임즈>	7
그림 2 RNN 구조.....	11
그림 3 순환신경망의 구조 <출처: colah.github.io>.....	11
그림 4 LSTM 순환신경망의 셀 구조 <출처: colah.github.io>.....	13
그림 5 국내 로보 어드바이저 시장 규모전망 <출처: 한국과학기술정보연구원>.....	15
그림 6 Diagram.....	18
그림 7 Application Architecture.....	23
그림 8 시세 정보 구현 코드 일부.....	27
그림 9 시세 정보 실행 화면 일부.....	27
그림 10 SNS 글 수집 구현 코드 일부.....	28
그림 11 SNS 글 수집 실행 화면 일부.....	29
그림 12 가치 예측 모델 구현 코드 일부.....	30
그림 13 가치 예측 모델 실행 화면 일부.....	31
그림 14 가치치 예측 모델 실행 화면 일부.....	31
그림 15 소셜 분석 모델 구현 코드 일부.....	32
그림 16 소셜 분석 모델 실행 화면 일부.....	33
그림 17 골든 크로스 정의.....	33
그림 18 수익률 비교.....	34

표 차례

Table 1 개발 환경 및 도구.....	19
Table 2 모델의 형태.....	24
Table 3 정량 평가.....	34
Table 4 정성 평가.....	36



제 1 장 서 론

1.1. 연구의 배경 및 필요성

2018 년, 가트너에서 발표한 10 대 전략 기술 트렌드에는 인공지능 강화 시스템과 블록체인이 포함되어있다. 블록체인은 분산 장부를 처리하면서 암호화폐라는 것을 발급받는다. 이것은 암호화폐 거래소를 통하여 현물로 교환이 가능하다.

하지만 암호화폐는 중앙 은행이 없다는 점에서 거래의 책임이나 통제가 되지 않는다. 암호화폐 거래소는 주식시장의 거래소와 비슷하면서도 다르다. 기존의 알고리즘을 활용한 매매 기법 방식을 사용하고 있다. 인공지능을 활용하여 24 시간 거래가 되는 암호화폐 시장의 가치를 적절하게 평가하여 거래한다.

이 시스템의 목적은 인공지능을 활용하여 블록체인에서 사용되는 암호화폐의 가격을 적절하게 평가하는 데에 있다.

1.2. 연구 내용

제안하는 가치 기반 통합모델의 다음과 같다

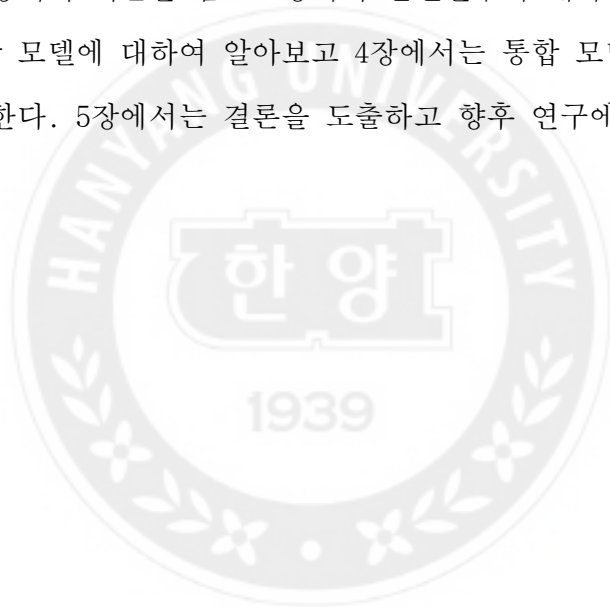
첫 번째로는 알고리즘 모델이다. 흔히 주식 분석에서 표현되는 여러 가지 기법들을 통칭한다. 기본적인 골든 크로스 기법을 활용하였다.

두 번째로는 인공지능 모델이다. 소셜 분석 모델과 가치 예측 모델로 구성되었다. 텍스트로 구성된 비정형 데이터를 분석하는 방법들 중의

하나인 감성 분석은 텍스트가 글의 주제에 대해 가지고 있는 감성이
긍정적인지 부정적인지에 대하여 연구하는 분석 기법이다.

기계학습과 인지과학에서 사용되는 통계학적 학습 알고리즘인 인공신경망
중 LSTM을 활용하였다. 기존의 데이터를 활용해서 트레이닝을 하고 이를
바탕으로 미래의 가치를 예측한다. 여기서는 예측된 가치의 기울기를
중심으로 확인한다.

본 논문은 1장에서 서론을 담고 2장에서 관련연구에 대하여 알아본다.
3장에서는 통합 모델에 대하여 알아보고 4장에서는 통합 모델을 통한
시스템을 구현한다. 5장에서는 결론을 도출하고 향후 연구에 대하여
알아본다.



제 2 장 관련 연구

2.1. 기존연구

2.1.1. 감성분석에 관한 연구

차은정은 뉴스 감성분석과 SVM을 이용한 주가지수 예측을 구현하였다.¹

감성분석과 기술적 분석을 결합한 SS-SVM 모델을 제시하였다. “주식시장은 기업의 내,외부 요인의 상호 관계가 주가를 형성하는데 영향을 주며 주식가격 변동에 뉴스의 출현이 중요한 원인으로 작용한다고 하였고 기존의 감성분석은 뉴스의 감성만을 활용하여 분석을 진행하였기 때문에 뉴스가 주식시장을 예측할 수 있는지에 관한 연구가 주를 이루었고 주식시장의 장기적 변화에 대한 예측에는 한계가 있다고 하였다.” [1]

2.1.2. 다변량 분석에 관한 연구

조아라는 다변량 분석을 통하여, 암호화폐에 미치는 영향을 연구하였다.²

“가격은 수요와 공급의 상황에 따라 변동하는 것이 원칙이지만, 그 밖

¹ 차은정, 뉴스 감성분석과 SVM을 이용한 주가지수 예측, 2016

² 조아라, 다변량 분석을 적용한 가상화폐 가격 등락에 미치는 영향 연구, 2017

에 정치, 경제, 시장 심리도 가격 등락에 영향을 미칠 수 있고 암호화폐의 경우 공급 측면은 가상화폐의 거래량, 가상화폐의 발행량, 알트 코인의 수, 거래소 수가 존재한다.” [2]

수요의 측면으로는 안전자산에 대한 수요, 랜섬웨어(Ransomware) 관련 수요가 있다.

이 밖에 암호화폐와 관련된 정치적 이슈, 각국의 경제 상황을 비교해볼 수 있는 환율, 그리고 법적 요인인 법정통화 지위 여부, 정부의 규제, 또 심리적 요인인 투기, 기대, 불확실성도 암호화폐의 가격에 영향을 미칠 수 있다.

그 결과 모든 암호화폐 가격 등락에는 각국의 통화가치 평가 상태가 미치는 영향이 가장 컸으며, 안전자산에 대한 수요, 가상화폐 공급량이 그 뒤를 따랐다고 설명하였다.

2.1.3. 머신러닝에 관련 연구

정욱은 주식시세 예측을 딥러닝을 이용하여 최적화 방법을 도출하는 것을 보여주었다.³

“예측 모델의 생성 및 하이퍼 파라미터 최적화를 통한 Model 튜닝 및 관련 거래 시간 피처의 제공을 추가하여 주가 예측치의 보다 높은 예측 정확

³ 정욱, 주식시세 예측을 위한 딥러닝 최적화 방법 연구, 2018

도를 제공 하고 주가 예측에 사용되는 알고리즘은 순차적인 입력 데이터로 부터 이후의 데이터를 예측하는 순환 신경망(RNN : Recurrent Neural Network)을 사용하고 기본적으로 일련의 순차적인 데이터를 학습하는데 정확도를 떨어뜨리고 과적합 문제를 발생시키는 장기 의존성 문제를 가지고 있는데, 이를 개선하기 위해 고안된 LSTM(Long Short Term Memory) 을 사용하여 주가 예측을 수행하였고 추가로 각 종목의 시간별 주식 거래 패턴을 피쳐로 선정하여, 시간 피쳐의 적용 시 주가 예측의 정확도가 증가함을 확인하였다.” [3]

2.2. 알고리즘 모델에 대한 고찰

2.2.1. 이동 평균 개념

이동 평균은 현재 주세의 방향을 측정하는 데 사용된다. 모든 유형의 이동 평균(Moving Average)은 과거 데이터 요소의 수를 평균하여 계산된 수학적 결과이다. 이동 평균(Moving Average)의 종류는 Simple Moving Average(SMA)와 Exponential Moving Average(EMA)가 존재한다.

가장 단순한 이동 평균 (SMA)이라고 알려진 이동 평균의 형식은 주어진 값 집합의 산술 평균을 취하여 계산한다. 예를 들어, 기본 5일 이동 평균을 계산하려면 지난 5일의 증가를 더한 다음 결과를 5로 나눈다.(2.1)

$$SMA(n) = \frac{C_n + C_{n-1} + \dots + C_2 + C_1}{n} \quad (2.1)$$

2.2.2. 골든 크로스 개념

골든 크로스(Golden cross)는 단기 이동 평균 (예 : 5 일 이동 평균)이 장기 이동 평균 (예 : 20 일 이동 평균) 또는 저항 수준을 초과하는 크로스 오버에서 형성된 완고한 브레이크 아웃 패턴이다. 장기 지표가 더 많은 비중을 차지함에 따라 골든 크로스(Golden cross)는 강세장을 의미하며 높은 거래량으로 강화된다.

2.3. 인공지능 모델에 대한 고찰

2.3.1. 블록체인과 암호화폐 개념 및 특징

블록체인은 데이터 분산 처리 기술이다. 즉, “네트워크에 참여하는 모든 사용자가 모든 거래 내역 등의 데이터를 분산, 저장하는 기술을 지칭하는 말이다.” [4] “블록들을 체인 형태로 묶은 형태이기 때문에 블록체인이라는 이름이 붙여졌고, 블록체인에서 블록은 개인과 개인의 거래(P2P)의 데이터가 기록되는 장부가 된다.” [5] “이런 블록들은 형성된 후 시간의 흐름에 따라 순차적으로 연결된 체인의 구조를 가지게 된다.” [6]

“모든 사용자가 거래내역을 보유하고 있어 거래 내역을 확인할 때는 모든 사용자가 보유한 장부를 대조하고 확인해야 한다.” [7] 이 때문에 블록체인은 공공/분산 거래 장부라고 불린다.

블록체인은 분산저장을 한다는 점이 특징이다. 기존 방식은 데이터를 위·변조하기 위해서 중앙 서버 역할을 하는 은행을 공격하면 가능했다. 최근

발생한 은행 전산망 해킹 사건을 보면 현실적인 위협 발생됨을 알 수 있다. 그러나 블록체인은 여러 명이 데이터 트랜잭션을 처리하기 때문에 위·변조가 어렵다. “블록체인 네트워크를 위·변조하기 위해서는 참여자의 거래 데이터를 모두 공격해야 하기 때문에 사실상 해킹은 불가능하다고 여겨진다.” [8]

“블록체인은 중앙 관리자가 필요 없다. 은행이나 정부 등 중앙기관이나 중앙 관리자가 필요했던 것은 공식적인 증명, 등기, 인증 등이 필요했기 때문이었지만, 블록체인은 다수가 데이터를 저장, 증명하기 때문에 중앙관리자가 존재하지 않게 된다.” [9]

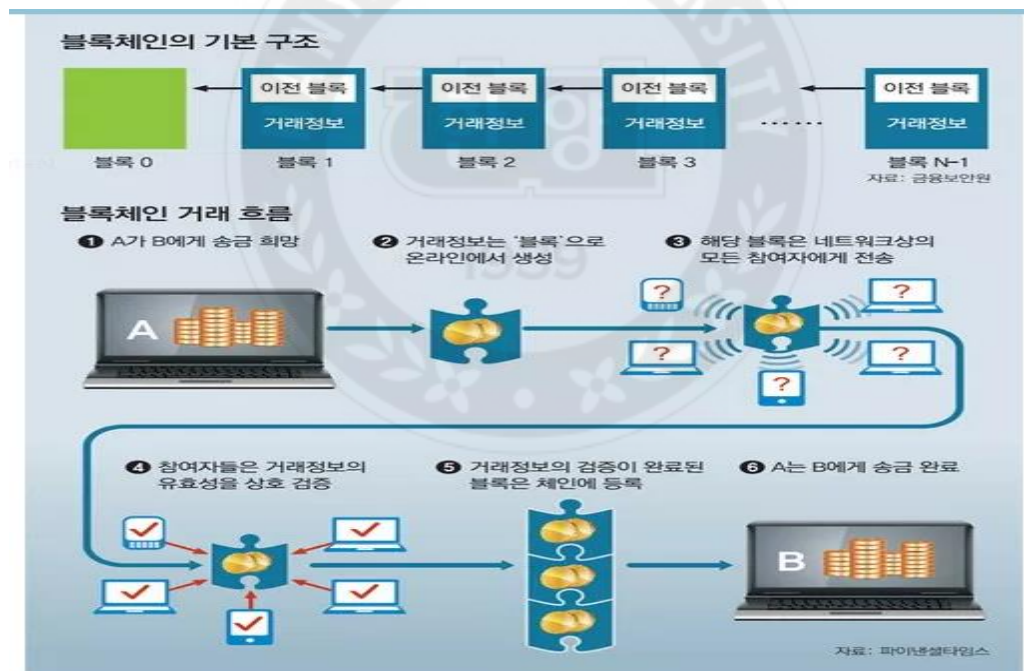


그림 1 블록체인의 기본 구조 <출처:파이널셀타임즈>

비트코인과 같은 암호화폐가 등장하게 된 것도 블록체인 덕분이다. 중앙 기관의 역할이 필요 없는 블록체인을 사용하게 되면서 ‘중앙은행’이 없더라도 화폐 발행이 가능해졌다. 암호화폐의 경우 기관에서 발행하지 않아 통제하거나 책임지는 일은 없다. 나카모토 사토시라는 개발자가 최초로 비트코인이라는 암호화폐 네트워크를 만들었다. 암호화폐의 경우 원하는 사람들이 직접 ‘채굴’을 통해 ‘발행’할 수 있다.

이더리움의 개념 및 특징

이더리움은 2015년 발표된 차세대 스마트 계약 분산 응용 프로그램 기술이며 Ethereum Foundation에서 개발하고 있는 오픈 소스 프로젝트이다.

“이더리움은 비트코인의 한계점을 극복하기 위하여 Solidity등의 튜링 완전성을 갖춘 확장용 언어를 갖춰 스마트 계약을 쉽고 간단하게 프로그램으로 만들 수 있도록 하였다.” [10]

현재 이더리움은 PoW(Proof of Work)로 합의 알고리즘을 사용하고 있지만 앞으로는 PoS(Proof of Stack)와 혼용으로 사용하다가 최종적으로 PoS(Proof of Stack)만으로 합의 알고리즘을 사용할 계획이다. 이더리움의 가상화폐 단위는 Ether를 사용한다.

이더리움도 비트코인과 마찬가지로 거래 이력을 블록체인에 기록할 수 있고 스마트 계약 그 자체나 실행 이력도 기록이 가능하다. “또한 비트코인과 마찬가지로 블록이 생성되면 블록에 저장된 스마트 계약이나 송금이 실행되며 스마트 컨트랙트나 송금 이력은 블록에 저장되고 정당성을 가지게

된다.” [11]

“스마트 계약의 실행 환경은 파이썬, C++, Go 언어와 같은 많은 언어로 구현되어 있지만 그중에 Go 언어가 많이 사용하는 Ethereum Client Soft에 포함되어 있고 Solidity로 작성된 Smart Contract는 Virtual Machine에서 동작하기 때문에 자바 가상 머신과 마찬가지로 운영체제에 종속되지 않는다.” [12]

2.3.2. 웹크롤링 개념 및 특징

스파이더라고도 불리는 웹 크롤러는 일반적으로 웹인덱스(웹스파이딩)의 목적을 위해 시스템적으로 월드와이드웹 브라우저 칭하는 인터넷 봇이다.

웹 검색엔진이고 다른 몇 개의 사이트를 웹 크롤링 혹은 스파이딩 소프트웨어로 그들의 웹 컨텐츠나 다른 사이트의 컨텐츠를 가리키기 위해 업데이트하는 소프트웨어이다. 웹 크롤링은 사용자들의 검색을 효율적으로 돕기 위해 페이지를 다운로드하여 인덱싱하는 검색엔진 프로세스에 의해서 페이지를 복사한다. 크롤러는 방문하는 시스템의 자원을 소비하고 종종 허용되지 않은 사이트에 접근하기도 한다. 대규모 페이지를 수집하기 위해 접근할 때 스케줄과 로딩의 이슈가 발생한다. 공개 사이트는 크롤링 에이전트에 수집되는 것을 원치 않을 때는 robots.txt를 포함한다. 웹사이트 일부나 전체가 수집 대상에서 제외된다. 크롤러는 하이퍼링크와 HTML 코드의 유효성을 확인한다. 그것들은 웹 스크래핑 하는데 사용된다.

2.3.3. 감성 분석 개념

자연어 프로세싱, 텍스트 분석, 전산 언어학 및 생체 인식을 통해 정서적인 상태와 주관적인 정보를 체계적으로 식별, 추출, 정량화 및 연구를 사용한다.

감정 분석은 리뷰, 설문 조사 응답, 온라인 및 소셜 미디어, 마케팅에서 고객 서비스, 임상 의학에 이르기까지 다양한 애플리케이션의 건강 관리 자료와 같은 고객 자료의 소리에 널리 적용된다.

일반적으로 정서 분석은 문서, 상호 작용 또는 사건에 대한 일부 주제 또는 전반적인 문맥 극성 또는 정서적 반응과 관련하여 연사, 작가 또는 기타 주제에 대한 태도를 결정하는 것을 목표로 한다. 태도는 판단이나 평가 (평가 이론 참조), 정서적인 상태 (즉, 저자 또는 연사의 감정 상태) 또는 의도된 감정적 의사소통이다.

“작성된 글의 주제를 찾는 것이 아니라 글을 작성한 사람이 주제에 대해 가지는 감성을 분석을 통해 찾아내는 방법인데 글을 작성한 사람이 글의 주제에 대해 긍정적인지 부정적인지를 글에 포함된 용어들의 분석을 통해 알아낸다.” [13] 글에 포함된 내용을 긍정, 부정, 중립으로 구분하고 모든 감정들을 종합하여 문서 안에 감정이 긍정인지 부정인지 혹은 중립인지를 구분한다. 감성분석은 텍스트마이닝 분야에서 연구되고 있다. 데이터마이닝, 웹마이닝, 텍스트마이닝의 분야에서 많이 활용되고 있다. 또한 연구 성과가 나타나고 있다.

2.3.4. RNN 개념 및 특징

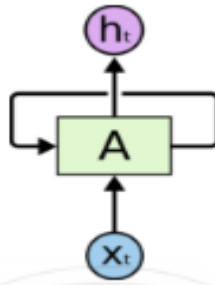


그림 2 RNN 구조

인공신경망의 한 종류인 RNN은 입력 데이터로 순서가 있는 정보를 사용한다. 이전의 데이터를 사용하여 학습된 뉴런의 상태 정보를 다음 데이터에서도 이용한다. “이를 통해 시간 순서에 대한 동적 시간 동작을 나타낼 수 있고 피드 포워드 신경 네트워크와 달리 RNN은 내부 상태 (메모리)를 사용하여 입력 시퀀스를 처리할 수 있다.” [14] 이는 세그먼트 화되지 않은 연결 필기 인식이나 음성 인식과 같은 작업에 적용할 수 있다.

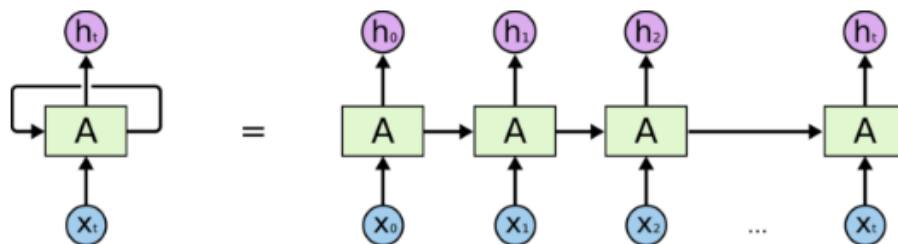


그림 3 순환신경망의 구조 <출처: colah.github.io>

시간에 따른 RNN을 보면 t 시점의 은닉 노드 H_t 는, t 시점의 노드 X_t 와 $t-1$ 시점의 은닉 노드 h_{t-1} 이 t 시점에서 어떤 연산을 통해 결합되어 산출되는 구조를 가지고 있다. h_{t+1} 도 마찬가지로 X_{t+1} 과 H_t 가 $t+1$ 시점에서 결합되어 나타나지는 구조로 순환신경망은 이전 시점에 생성된 H 가 다음 시점에 반영되어 새로운 h 가 만들어지는 순환 구조를 가지고 있다. H 는 이전 신경망의 뉴런에서 나온 정보가 숨겨져 있는 은닉 상태를 의미한다. 순환 신경망의 은닉 상태 H_t 는 이전 은닉 상태 h_{t-1} 과 현재의 입력값 x_t 에 가중치 W 를 곱하고 편차 b 를 더하여 활성화 함수 f 를 취함으로 계산할 수 있다.(2.2)

$$H_t = f(W(h_{t-1}, X_t) + b) \quad (2.2)$$

2.3.5. LSTM 개념 및 특징

1997 년 Hochreiter와 Schmidhuber는 LSTM (Long Short-Term Memory) 네트워크를 발명했으며 여러 응용 분야에서 정확성 기록을 세웠다. 2007 년 전후로 LSTM은 음성 인식에 혁명을 일으키기 시작하여 특정 음성 응용 프로그램에서 기존 모델을 능가했다. 2009 년에 CTC (Connectionist Temporal Classification) 방식의 LSTM 네트워크는 연결된 필기 인식에서 여러 번 경쟁했을 때 패턴 인식 경연 대회에서 우승 한 최초의 RNN이다. 2014 년 중국의 검색 자 Baidu는 기존의 음성 처리 방법을 사용하지 않고 스위치보드 Hub5'00 음성 인식 벤치마크를 위반하기 위해 CTC에서 훈련된 RNN을 사용했다. LSTM은 또한 큰 어휘 음성 인식과 텍스트 음성 변환 합성을 향상시켰고 Google 안드로이드에서 사용되었다. 2015 년 Google의 음성 인식 기술은

Google 음성 검색에서 사용된 CTC에서 교육받은 LSTM을 통해 49 %의 놀라운 성능 향상을 경험했다.

LSTM은 장기 의존성 문제를 피하기 위해 명시적으로 설계되었다. 오랜 기간 동안 정보를 기억하는 것은 실제 배우는 것이 아닌 자신의 기본 행동이다. 모든 반복적인 신경망은 신경망의 반복적인 모듈 체인의 형태를 갖는다. 표준 RNN에서 이 반복 모듈은 단일 탄층과 같은 매우 간단한 구조를 갖는다.

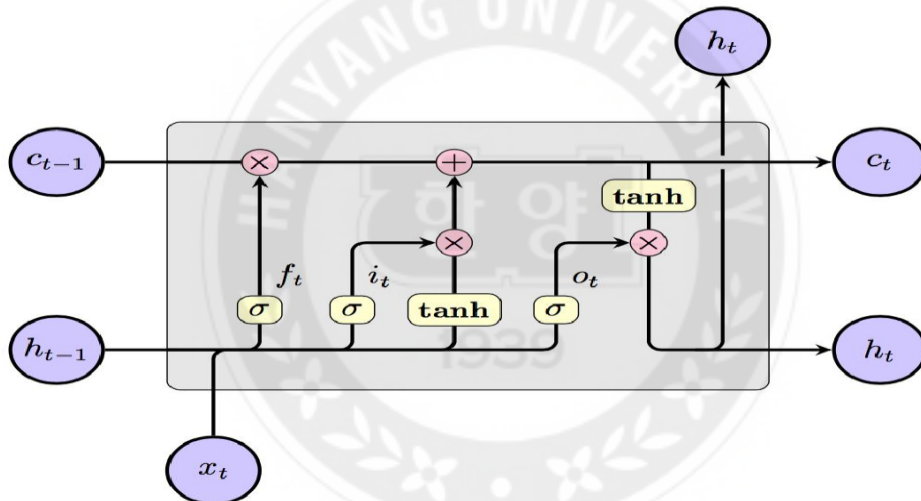


그림 4 LSTM 순환신경망의 셀 구조 <출처: colah.github.io>

긴 시퀀스의 데이터라도 그레이디언트 소실 없이 LSTM 셀을 사용하면 순환 신경망은 아무리 문제없다. LSTM 순환신경망의 셀 구조는 LSTM 셀을 가진 순환신경망을 보여주고 있다. 입력 게이트는 입력값의 크기를 결정하고(식 2.3), 망각 게이트는 이전의 셀 상태 정보를 얼마나 잊어버릴지의 크기를 결정하며(식 2.4), 출력 게이트는 출력의 크기를 결정한다(식 2.5).

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (2.3)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (2.4)$$

$$c_t = f_{tc_{t-1} + i_t \tan h(W_{xc}x_t + W_{hc}h_{t-1} + b_c)} \quad (2.5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \quad (2.6)$$

$$h_t = o_t \text{Tanh}(c_t) \quad (2.7)$$

여기서 c_t 셀 상태이며, h_t 는 셀의 출력이다. i_t 는 입력 게이트를 나타내며 f_t 는 망각 게이트를 나타내고 o_t 는 출력 게이트를 나타낸다. σ 와 Tanh 는 활성화 함수로 각각 시그모이드와 하이퍼 탄젠트를 나타낸다. “기존의 셀 상태에 망각 게이트의 출력값을 곱하여 셀 상태 값의 일정량을 잊어버리고, 입력값과 이전 단계의 출력값을 처리한 결과에 입력 게이트의 출력값을 곱하여 입력의 일정량을 받아들임으로써 새로운 셀 상태를 만든다(식 2.6).” [15] “이 셀 상태에 출력 게이트의 출력값을 곱하여 출력을 낸다(식 2.7). 이와 같이 셀 상태 값을 얼마나 잊어버리고, 새로운 입력값을 얼마나 받아들일지 결정하여 더하는 구조이므로 이 과정이 반복되어도 학습이 불가능해지는 일이 발생하지 않는다.” [16]

2.3.6. 로보 어드바이저 개념 및 특징

“미국에서 2000년 중후반 경 Wealthfront, Betterment 등의 로보 어드바이저 업체들이 생겨나기 시작했고 선두권 로보 어드바이저 업체들은 미국 스타트업 투자의 대호황에 따른 엄청난 투자금의 유입을 바탕으로 거대한 마케팅 예산 집행을 할 수 있게 되었다.” [17] “활발한 마케팅에 힘입어 많은 운용자산이 유입되기 시작하였고 선두업체인 웰스프론트와 베테르먼트 등

은 업체별로 운용자산이 3조 원 이상으로 성장 중이다.” [18] 로보 어드바이저 업체들의 총 운용자산은 약 20조 원으로 추정된다.

로보 어드바이저의 장점은 첫 번째는 인간이기 때문에 실수나 부정을 저지르는 리스크가 존재하지만 로보 어드바이저의 경우는 데이터와 알고리즘을 바탕으로 투자를 운용한다. 두 번째는 시간과 장소에 구애받지 않는 편리함이 존재한다. 투자 자문을 받거나 일임할 경우 직접 상담을 받고 이용하였다. 하지만 로보 어드바이저는 컴퓨터 프로그램이 이를 모두 담당하기 때문에 시간과 비용을 절약할 수 있다. 세 번째는 오프라인 자산관리 서비스에 비해 수수료가 30% 수준으로 저렴하다. 서비스 가입에 필요한 최소 거래 금액도 낮다. 기존 개인 자산관리(PB) 서비스는 주로 고액 자산가만을 대상으로 접근성이 낮다.



그림 5 국내 로보 어드바이저 시장 규모 전망 <출처: 한국과학기술정보연구원>

한국에서도 로보 어드바이저를 통한 자산 관리는 빠르게 성장하고 있다. 금융사들은 자체 검증을 거친 로보 어드바이저 업체와 협력하여 관련 서비스를 출시했다.

자문형 로보 어드바이저의 경우는 투자자에게 몇 가지 질문을 제시한 뒤 이를 기반으로 자체 알고리즘을 통하여 투자자에게 맞는 포트폴리오를 추천해준다. 고객은 자문에 따라서 직접 상품을 선택하여 운용할 수 있다. 일임형 로보 어드바이저의 경우는 자문형 로보 어드바이저 처럼 질문에 답변하면 알고리즘에 의해서 로보 어드바이저가 직접 운용을 한다.



제 3 장 가치 기반 통합 모델

3.1. 통합 모델 제안 개요

이번 논문이 제안하는 개요는 다음과 같다.

첫 번째로 인공지능 모델에서 활용되는 웹 크롤러를 구현하였다.
LSTM의 기본 인풋 데이터로 사용될 암호화폐 시세 정보와 감성분석에
사용하기 위한 SNS의 글 정보를 수집한다.

두 번째로 인공지능 모델에는 크게 두 가지 모델로 다시 구분된다.
가치 예측 모델과 소셜분석 모델로 구분된다. 첫 번째로 가치 예측 모델은
이전 시세 정보들을 활용하여 다음날 가격을 LSTM을 활용하여 예측하는 것
이다.

두 번째로는 소셜분석 모델은 수집된 SNS 글 정보를 활용하여 대상이 되는
암호화폐에 대한 긍정/부정인지를 평가한다.

세 번째로 알고리즘 모델이다. 이동평균선을 기반으로 하는 골든 크로스
기법을 활용하여 구현하였다.

네 번째로는 인공지능 모델과 알고리즘 모델의 결과값을 조합하여 판단
하는 가치 기반 통합 모델이다.

제안하는 해당 모델에 대한 Diagram은 다음과 같다.

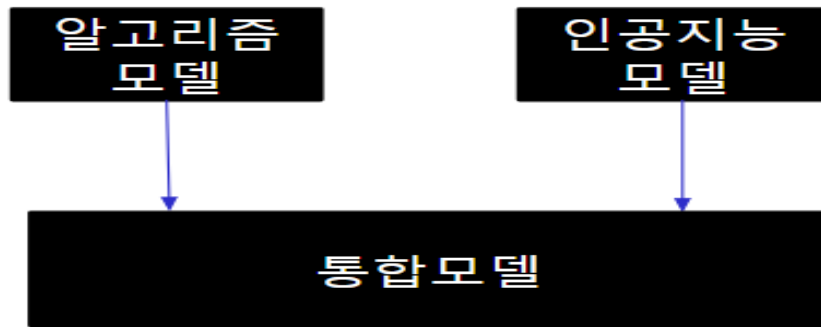


그림 6 Diagram

특정 종목에 대하여 100일 동안의 표준 편차 값이 3%라고 가정했을 경우 금일 시가가 알고리즘 모델에 의하여 매수 포지션이라고 한다면 시가에 사서 3%의 수익이 나면 매도한다. 인공지능 모델 중 가치 예측 모델에서 다음날 예상 금액이 시가 금액 보다 기울기가 45도 이상일 경우 급등을 예상하고, 금일 매도 %를 2배로 변경한다. 반대로 기울기가 음수가 나올 경우에는 금일 매도 %를 절반으로 변경한다.

알고리즘 모델과는 별개로 소셜 분석 모델에서 긍정의 점수가 높을 경우에는 알고리즘 모델에서 매수 포지션이 아닌 경우라도 매수 포지션을 가져간다.

위에서 설명한 내용을 수식화하면 100일간의 종가 기준으로 표준편차를 구한 값을 구하고(3.1) 이동평균선 5일과 20일 값을 구하고(3.2) 인공지능 모델 중 가치 예측 모델의 다음날 종가를 예측하고(3.3) 기울기 값을 구하고(3.4) 알고리즘 모델을 통해 매수가 가능한지를 확인하고(3.5) 소셜 분석 모델을 통해 매수가 가능한지를 확인한다(3.6)

$$\sigma = 100 \quad (3.1)$$

$$SMA(n) = \frac{C_n + C_{n-1} + \dots + C_2 + C_1}{n} \quad (3.2)$$

$$h_t = o_t \tanh(c_t) \quad (3.3)$$

$$\frac{\text{atan2}(5ma, h) * 180}{\pi} = f(h) \begin{cases} 2, & h > 45 \\ 1, & 1 < h < 45 \\ \frac{1}{2}, & h < 1 \end{cases} \quad (3.4)$$

$$a = \begin{cases} 1, & SMA(5) > SMA(20) \\ 0, & SMA(5) < SMA(20) \end{cases} \quad (3.5)$$

$$\beta = \begin{cases} 1, & r > 0.85 \\ 0, & r < 0.85 \end{cases} \quad (3.6)$$

$$S = \frac{\sigma f(h)}{\left| \frac{\alpha + \beta}{2} \right|} \quad (3.7)$$

σ 는 특정 기간 동안의 표준 편차 값이다. SMA(n)을 통해서 5일과 20일의 이동 평균 값을 구하여, a 를 구할 수 있다. 시가에 골든 크로스라면 알고리즘 모델에 의해 매수를 진행한다. 이와는 별개로 β 는 소셜 감성 분석 모델에 의해서 매수를 진행한다. 두 조건 중 한 개 이상이라면 매수는 진행된다. h_t 은 가치 예측 모델 결과값으로 다음 날 종가를 예측하는데 $f(h)$ 함수를 활용하여 σ 값을 가중치를 조정하면서 매도를 진행한다.

3.2. 개발 도구

해당 시스템에 대한 개발 도구 및 사양은 다음의 표에 정리하여 두었다.

Table 1 개발 환경 및 도구

환경설정명	버전 및 이미지명
프로세서	Intel® Pentium® XEON E3-1231 V3 @ 3.40GHz

메모리	8.00GB
시스템종류	64 비트 운영체제, x64 기반 프로세서
윈도우버전	Window Server 2008 R2
Anaconda	4.4
Python	3.6
beautifulsoup4	4.6.0
Keras	2.1.5
Pandas	0.22
Selenium	3.11.0
Tensorflow	1.8.0
Nltk	3.2.5
Numpy	1.14.2
Konlpy	0.4.4
Request	2.18.4
urllib3	1.22
Jupyter	1.0.0

표

3.3. 개발 도구 설명

파이썬을 중심으로 제안하는 통합 모델의 구현을 진행하였다. 파이썬의 특

정으로는 크로스 플랫폼 호환이 되는 라이브러리가 대부분 존재하여 구조가 간단하고, 명확하게 정의된 구문을 사용하여 가독성이 뛰어나서 유지 보수가 쉽다.

텐서플로우(TensorFlow)는 구글에서 만든 기계 학습과 딥러닝을 위한 오픈 소스 라이브러리이다. “데이터 플로우 그래프(Data Flow Graph) 방식을 사용하였고 이는 수학 계산과 데이터의 흐름을 노드(Node)와 엣지(Edge)를 사용한 방향 그래프(Directed Graph)로 표현한다.” [19] “노드는 수학적 계산, 데이터 입/출력, 그리고 데이터의 읽기/저장 등의 작업을 수행한다. 엣지는 노드들 간 데이터의 입출력 관계를 나타내고 엣지는 동적 사이즈의 다차원 데이터 배열(=텐서)을 실어 나르는데, 여기에서 텐서플로우라는 이름이 지었다.” [20]

텐서 플로우의 랩핑 라이브러리로 케라스를 활용하였다. 케라스의 특징은 첫 번째는 독립적인 모듈 설정이 가능하여 최소한의 제약사항으로 서로 연결한다. 모델은 시퀀스 또는 그래프로 모듈들을 구성한다.

두 번째는 반복 속도와 혁신성은 떨어질 수 있지만 각 모듈간의 간결하여 모든 코드를 쉽게 이해할 수 있다.

세 번째는 확장성으로 새로운 클래스나 함수로 모듈을 아주 쉽게 추가할 수 있다.

케라스의 가장 핵심적인 데이터 구조는 바로 모델인데 제공하는 시퀀스 모델로 원하는 레이어를 쉽게 순차적으로 쌓을 수 있다.

첫 번째는 데이터 세트 생성한다. 원본 데이터를 불러오거나 시뮬레이션을 통해 데이터를 생성하거나 데이터로부터 훈련 세트, 검증 세트, 시험 세트를 생성한다.

두 번째는 모델 구성한다. 시퀀스 모델을 생성한 뒤 필요한 레이어를 추

가하여 구성한다.

세 번째는 모델 학습과정 설정한다. 학습하기 전에 학습에 대한 설정을 수행한다.

네 번째는 모델 학습시킨다. 훈련 세트를 이용하여 구성된 모델로 학습시킨다.

다섯 번째는 학습과정 확인하기다. 모델 학습 시 훈련 세트, 검증 세트의 손실 및 정확도를 측정한다.

여섯번째는 모델 평가이다. 준비된 시험 세트로 학습한 모델을 평가한다.

마지막으로 모델 사용한다. 임의의 입력으로 모델의 출력을 얻는다.



3.4. Application Architecture

이번 논문에서 구현한 시스템의 구조를 정리를 하면 다음과 같다.

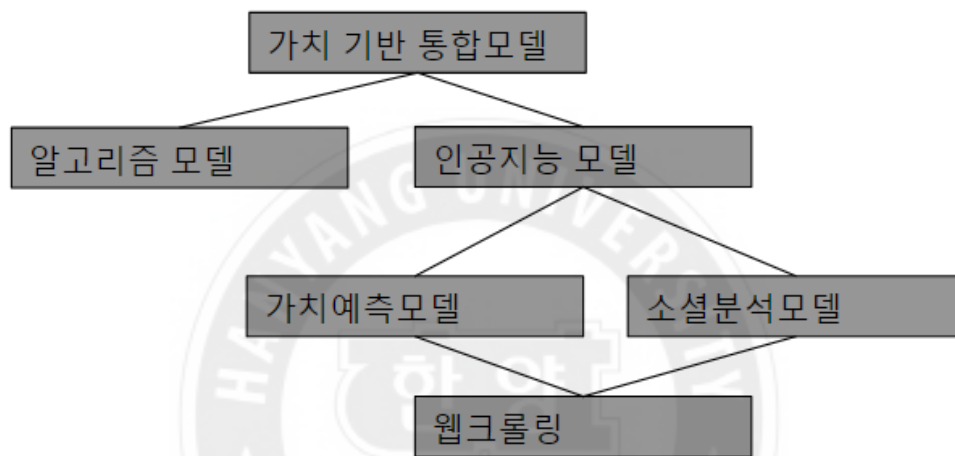


그림 7 Application Architecture

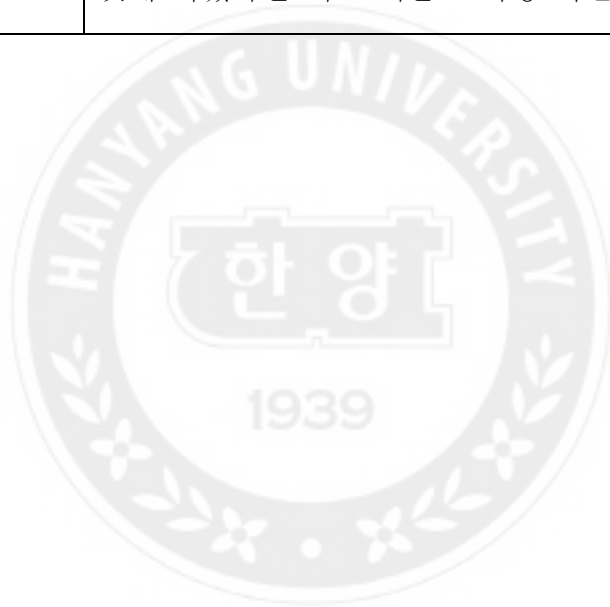
제안하는 통합모델은 골든 크로스 기반의 알고리즘 모델과 가치 예측 모델과 소셜 분석 모델을 포함한다.

위의 그림에 대하여 아래 표에 정리해보았다.

Table 2 모델의 형태

구분	설명
웹크롤링	가치 예측 모델과 소셜분석 모델의 기본 정보가 되는 거래소 시세정보 데이터와 SNS글 데이터를 수집하는 필요하다.
가치예측모델	웹크롤링으로 수집된 거래소의 시세 정보를 기반으로 RNN-LSTM을 활용하여, 미래에 평가되는 가격을 예측한다
소셜분석모델	웹크롤링으로 수집된 SNS의 글을 기반으로 형태소 분석을 진행하고, 긍정,부정 표현하는 감성 분석을 진행하여 점수를 도출한다.
인공지능모델	가치 예측 모델과 소셜분석 모델을 기반으로 가치 기반 통합 모델과 인터페이스 한다.
알고리즘모델	이동평균선을 이용하여 5일 선 - 20일 선 골든 크로스 기법을 활용하고 가치 기반 통합 모델과 인터페이스 한다.
가치 기반 통합모델	인공지능 모델과 알고리즘 모델을 활용하여, 매수와 매도의 포지션을 결정한다. 시세 정보를 활용하여 과거 등락률에 대한 표준편차를 매도의 기준 %로 결정한다. 매도가 미체결 됐다면 종가에는 매도한다. 매수 전략은 크게 두 가지 경우로 구분된다.

	<p>첫 번째로는 알고리즘 모델에 의해 골든 크로스가 진행되었을 경우에 시가에 매수를 한다. 두 번째로는 소셜분석 모델에서 분석된 값이 매우 긍정적인 경우에 시가에 매수를 한다. 매도 전략은 당일 표준편차에 의한 매도 기준 %가 기본이 된다. 만약 가치 예측 모델을 통한 다음날 가치의 기울기가 높게 나왔다면 매도 기준 % 상향 시키고, 기울기가 낮게 나왔다면 매도 기준 % 하향 시킨다.</p>
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------



제 4 장 통합 모델 구현 및 성능평가

4.1. 구현내용

구현된 내용 중 첫 번째로 암호화폐의 시세 정보 수집 관련 부분이다.
암호화폐 거래소의 API 주소에 접근하다. 필요한 데이터를 요청한다.

예를 들어 <https://crix-api-endpoint.upbit.com/v1/crix/candles/minutes/5?code=CRIX.UPBIT.KRW-ETH&count=1>의 경우 아래와 같은 출력된다.

```
[{"code": "CRIX.UPBIT.KRW-ETH", "candleDateTime": "2018-06-10T14:40:00+00:00", "candleDateTimeKst": "2018-06-10T23:40:00+09:00", "openingPrice": 628800.00000000, "highPrice": 628900.00000000, "lowPrice": 628400.00000000, "tradePrice": 628500.0, "candleAccTradeVolume": 99.49904336, "candleAccTradePrice": 62538139.923129000, "timestamp": 1528641865081, "unit": 5}]
```

위와 같은 JSON 형태의 return 값을 보다 안정적으로 대용량 데이터들을 쉽게 처리할 수 있다고 알려진 파이썬 패키지 중 Pandas를 통해서 필요한 부분만 추출하여 데이터 분석, 처리를 쉽게 처리한다.

```

import json
import time
import requests
import pandas as pd

headers = {'User-Agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/39.0.2171.95 Safari/537.36'}

coinName = "BTH";
url = "https://crix-api-endpoint.upbit.com/v1/crix/candles/minutes/5?code=CRIX.UPBIT.KRW-BTH&count=1";

while 1:
    try:
        res = requests.get(url, headers=headers)
    except requests.exceptions.HTTPError as err:
        print (err)
        exit(1)

    data = res.json() # json 구조도 변환
    code = data[0]['code']
    print(code)
    print("=====")
    print(" date      open high low final    vol")
    print("=====")

    for i in range(len(data)):
        date = data[i]['candleDateTimeRet']
        onlyDate = date.split('T') # 날짜정보와 시간정보 분리
        print(onlyDate[0], "%d"%data[i]['openingPrice'], "%d"%data[i]['highPrice'], "%d"%data[i]['lowPrice'], "%d"%data[i]['tradePrice'], "%d"%data[i]['candleAccTradeVolume']]);

    print("=====")

```

그림 8 시세 정보 구현 코드 일부

```

관리자: Anaconda Prompt - python coin.py
CRIX.UPBIT.KRW-ETH
=====
date      open high low final    vol
=====
2018-06-07 663700 663700 662000 662000 81
=====
CRIX.UPBIT.KRW-ETH
=====
date      open high low final    vol
=====
2018-06-07 662100 662100 661100 662100 40
=====
CRIX.UPBIT.KRW-ETH
=====
date      open high low final    vol
=====
2018-06-07 662200 663400 662200 663300 24
=====
CRIX.UPBIT.KRW-ETH
=====
date      open high low final    vol
=====
2018-06-07 663000 663000 660200 660800 103
=====

```

그림 9 시세 정보 실행 화면 일부

두 번째로 SNS 정보 수집 관련 부분이다. 카페의 글을 수집하기 위해서 로그인 및 세션 정보가 유지되어야 한다. 이 부분은 Python의 패키지인 Selenium과 BeautifulSoup4를 활용하여 HTML을 파싱 한다. “Selenium은 웹 어플리케이션을 위한 테스트 프레임워크로 자동화 테스트를 위한 여러 가지

강력한 기능을 지원하는데 다양한 브라우저들도 지원하며, 다양한 테스트 작성 언어(Java, Ruby, Groovy, Python, PHP, and Perl.)를 지원한다.” [21]

“현재 WebDriver라는 웹 자동화 툴과 통합하는 작업이 진행되고 있다.” [22] Selenium으로 세션 정보를 유지하는 것을 활용하여 해결하고, 필요한 정보만 추출하는 것은 BeautifulSoup4를 활용한다. HTML document 안에 있는 수많은 HTML 태그들을 사용하기 편한 Python 객체 형태로 만들어 HTML 태그들을 객체화 시켜서 파싱 하는데 활용한다.

카페에 로그인 하여, 해당 게시물로 이동한다. 필요한 정보를 파싱하여 가공한다. 예로 글의 제목, 내용, 조회수, 댓글 내용의 글을 추출하여 사용한다.

```
# 네이버 로그인 완료 => 추후에 @토 만들어보기
submit_element = driver.find_element_by_css_selector(".btn_global")
submit_element.click()

#driver.get("https://cafe.naver.com/nexontv//ArticleList.nhn?search.clubid=24978815&search.menuid=345&search.boardtype=L")
driver.get("https://cafe.naver.com/nexontv/ArticleList.nhn?search.clubid=24978815&search.menuid=345&userDisplay=5&search.boardtype=L&search.specialmenutype=6&search.questionid=1")

# iframe으로 프레임 전환
driver.switch_to_frame("cafe_main")

# href 속성을 찾아 url을 리스트로 저장한다.
article_list = driver.find_elements_by_css_selector('span.aaa > a.m-tool-c')
article_urls = [ i.get_attribute('href') for i in article_list ]

res_list = []
# BeautifulSoup 활용
for article in article_urls:
    driver.get(article)
    # article도 switch_to_frame이 필수
    driver.switch_to_frame("cafe_main")
    soup = bs(driver.page_source, 'html.parser')
    # 제목 검색
    title = soup.select('div.tit-box span.b')[0].get_text()
    # 게시글을 띄어쓰기 단위로 합친다.
    content_tags = soup.select('#tbody')[0].find_all(lambda tag:tag.name == "p" and len(tag.attrs) == 0 )
    #content = ' '.join([ tags.get_text().strip() for tags in content_tags ])
    content = ' '.join([ tags.get_text().strip() for tags in content_tags ])
    # 게시물 조회수
    readcount = soup.select('span.b.m-tool-c.reply')[1].get_text()
    # 댓글
    reply = ' '.join([ x.get_text().strip() for x in soup.select('span.comm_body') ])
    # dict형태로 만들어 결과 list에 저장
    res_list.append({'title': title, 'readcount': readcount, 'reply': reply, 'content': content+'\n'})
    # time.sleep 작업도 필요하다.
# 결과 데이터프레임화
cafe_df = pd.DataFrame(res_list)
```

그림 10 SNS 글 수집 구현 코드 일부

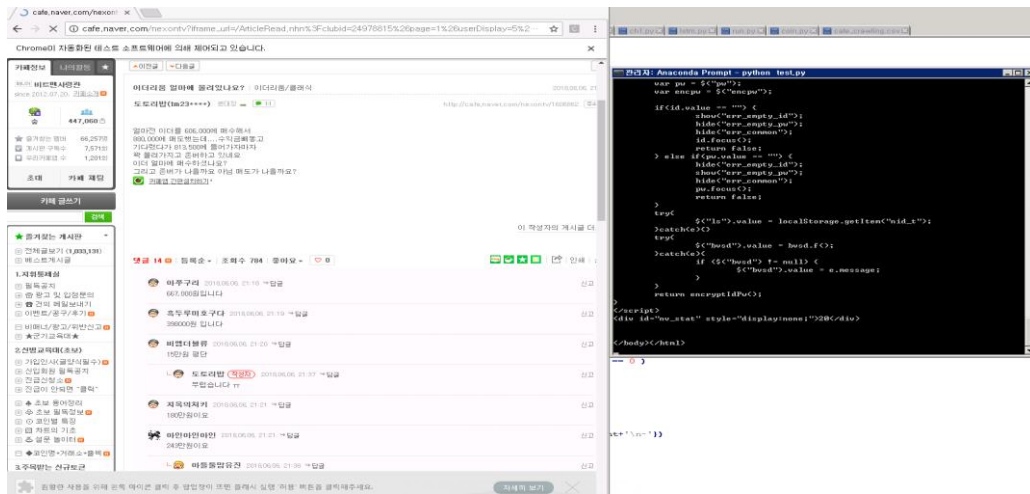


그림 11 SNS 글 수집 실행 화면 일부

세 번째로는 가치 예측 모델의 구현이다. Tensorflow를 활용한 keras를 이용하여 LSTM을 실행한다. 웹크롤링으로 수집된 시세 정보를 로드한다.

연산을 하다 보면 데이터의 크기가 들쭉날쭉하여 기준을 잡기가 어렵다. 데이터의 스케일을 일정하게 맞출 수 있다. 이러한 과정을 데이터 전처리라고 한다. Python 패키지 중 Numpy는 과학적 계산을 위한 핵심 라이브러리이다. 고성능 다차원 배열 객체와 이들 배열과 함께 작동하는 도구들을 제공한다. 이를 통해서 데이터 세트를 만든다.

Sequential 선형 스택의 레이어 모델로 구성한다. 계층 인스턴스 목록을 생성자에 전달하여 Sequential 모델을 만들 수 있다. 모델은 예상되는 입력 모양을 알아야 한다. 이러한 이유 때문에 Sequential 모델의 첫 번째 레이어는 입력 모양에 대한 정보를 받아야 한다. 또한 Dense 레이어는 입력과 출력을 모두 연결한다. “예를 들어 입력 뉴런이 4개, 출력 뉴런이 8개 있다면 총 연결선은 32개($4 \times 8 = 32$) 이고 각 연결선에는 가중치(weight)를 포함하고 있는데, 이 가중치가 나타내는 연결강도이다.” [23] 연결선이 32개이므로 가중치도 32개이다.

모델의 옵티마이저의 종류는 여러 가지가 존재한다.

예로 `keras.optimizers.Adam(lr=0.001, beta_1=0.9, beta_2=0.999, epsilon=None, decay=0.0, amsgrad=False)` 은
`model.compile(loss='mean_squared_error', optimizer='adam')`
으로 나타낸다.

`predict (self, x, batch_size = None, verbose = 0, steps = None)`으로
일괄적으로 계산되는 입력 샘플에 대한 출력 예측을 생성한다.
Python 패키지 중 `matplotlib.pyplot`을 활용하여 차트를 그릴 수 있다.

```
# normalize the dataset
scaler = MinMaxScaler(feature_range=(0, 1))
dataset = scaler.fit_transform(dataset)
# split into train and test sets
train_size = int(len(dataset) * 0.67)
test_size = len(dataset) - train_size
train, test = dataset[0:train_size,:], dataset[train_size:len(dataset),:]
# reshape into X=t and Y=t+1
look_back = 1
trainX, trainY = create_dataset(train, look_back)
testX, testY = create_dataset(test, look_back)
# reshape input to be [samples, time steps, features]
trainX = numpy.reshape(trainX, (trainX.shape[0], 1, trainX.shape[1]))
testX = numpy.reshape(testX, (testX.shape[0], 1, testX.shape[1]))
# create and fit the LSTM network
model = Sequential()
model.add(LSTM(7, input_shape=(1, look_back)))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
model.fit(trainX, trainY, epochs=100, batch_size=1, verbose=1)
# make predictions
trainPredict = model.predict(trainX)
testPredict = model.predict(testX)
# invert predictions
trainPredict = scaler.inverse_transform(trainPredict)
trainY = scaler.inverse_transform([trainY])
testPredict = scaler.inverse_transform(testPredict)
testY = scaler.inverse_transform([testY])
# calculate root mean squared error
trainScore = math.sqrt(mean_squared_error(trainY[0], trainPredict[:,0]))
print('Train Score: %.2f RMSE' % (trainScore))
testScore = math.sqrt(mean_squared_error(testY[0], testPredict[:,0]))
print('Test Score: %.2f RMSE' % (testScore))
# shift train predictions for plotting
trainPredictPlot = numpy.empty_like(dataset)
trainPredictPlot[:, :] = numpy.nan
trainPredictPlot[look_back:len(trainPredict)+look_back, :] = trainPredict
# shift test predictions for plotting
testPredictPlot = numpy.empty_like(dataset)
testPredictPlot[:, :] = numpy.nan
testPredictPlot[len(trainPredict)+(look_back*2)+1:len(dataset)-1, :] = testPredict
# plot baseline and predictions
plt.plot(scaler.inverse_transform(dataset))
plt.plot(trainPredictPlot)
plt.plot(testPredictPlot)
plt.show()
```

그림 12 가치 예측 모델 구현 코드 일부


```

관리자: Anaconda Prompt
52/62 [=====] - 0s 1ms/step - loss: 0.0029
Epoch 37/100
52/62 [=====] - 0s 1ms/step - loss: 0.0029
Epoch 38/100
52/62 [=====] - 0s 1ms/step - loss: 0.0027
Epoch 39/100
52/62 [=====] - 0s 1ms/step - loss: 0.0027
Epoch 40/100
52/62 [=====] - 0s 1ms/step - loss: 0.0027
Epoch 41/100
52/62 [=====] - 0s 1ms/step - loss: 0.0027
Epoch 42/100
52/62 [=====] - 0s 1ms/step - loss: 0.0026
Epoch 43/100
52/62 [=====] - 0s 1ms/step - loss: 0.0027
Epoch 44/100
52/62 [=====] - 0s 2ms/step - loss: 0.0026
Epoch 45/100
52/62 [=====] - 0s 2ms/step - loss: 0.0027
Epoch 46/100
52/62 [=====] - 0s 1ms/step - loss: 0.0027
Epoch 47/100
52/62 [=====] - 0s 1ms/step - loss: 0.0027
Epoch 48/100
52/62 [=====] - 0s 1ms/step - loss: 0.0027
Epoch 49/100
52/62 [=====] - 0s 1ms/step - loss: 0.0027
Epoch 50/100
52/62 [=====] - 0s 1ms/step - loss: 0.0028
Epoch 51/100
52/62 [=====] - 0s 1ms/step - loss: 0.0026
Epoch 52/100
52/62 [=====] - 0s 1ms/step - loss: 0.0028
Epoch 53/100

```

그림 13 가치 예측 모델 실행 화면 일부

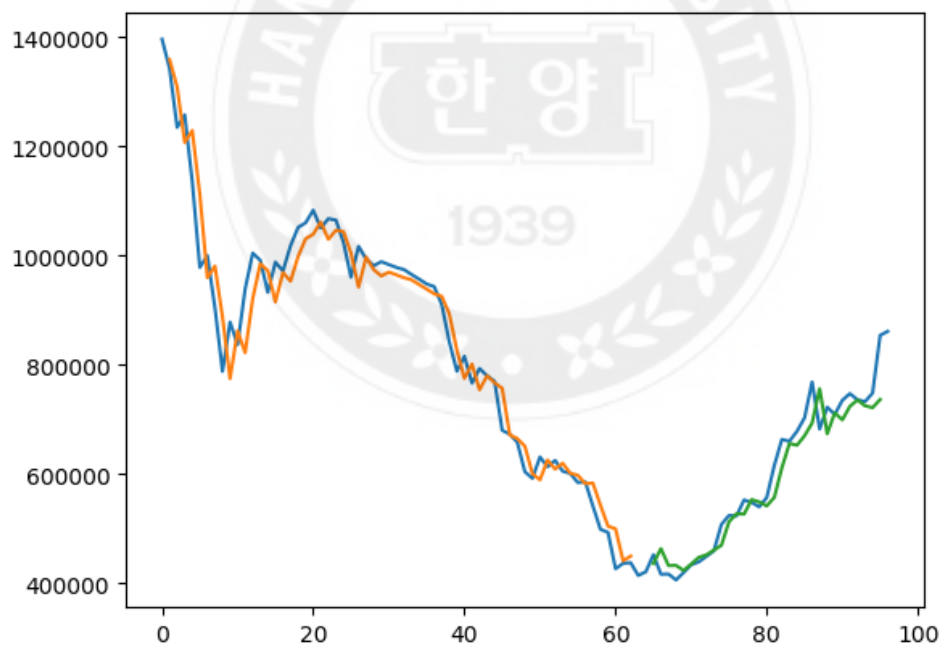


그림 14 가치치 예측 모델 실행 화면 일부

네 번째로는 SNS 정보를 활용한 소셜분석 모델이다. 형태소 분석을 한 후 감성분석을 진행한다.

KoNLPy로 형태소 분석을 진행하였다. KoNLPy는 한국어 정보처리를 위한 간결하고 우아한 문법구조를 가지고 있을 뿐 아니라 크롤링, 웹 프로그래밍, 그리고 데이터 분석을 수행할 수 있는 다양한 패키지와 함께 사용할 수 있는 파이썬으로 구현되어 있다.

NLTK(Natural Language Toolkit) 패키지는 교육용으로 개발된 자연어 처리 및 문서 분석용 파이썬 패키지다. 다양한 기능 및 예제를 가지고 있으며 실무 및 연구에서도 많이 사용된다.

```
# 형태소 분류
train_docs = [(tokenize(row[1]), row[2]) for row in train_data[1:]]
test_docs = [(tokenize(row[3]), row[3]) for row in test_data[1:]]

# Training data의 token 모으기
tokens = [t for d in train_docs for t in d[0]]
print(len(tokens))

# Load tokens with nltk.Text()
text = nltk.Text(tokens, name='NM48C')
print(text.vocab().most_common(10))

# 텍스트간의 연어 빈번하게 등장하는 단어 구하기
text.collocations()

# term이 존재하는지에 따라서 문서들 분류
selected_words = [f[0] for f in text.vocab().most_common(2000)] # 여기서는 최빈도 단어 2000개를 피쳐로 사용
train_docs = train_docs[:5000] # 시간 단축을 위한 함수로 training corpus의 일부만 사용할 수 있음
train_xy = [(term_exists(d), c) for d, c in train_docs]
test_xy = [(term_exists(d), c) for d, c in test_docs]

# nltk의 NaiveBayesClassifier으로 데이터를 트레이닝 시키고, test 데이터로 확인
classifier = nltk.NaiveBayesClassifier.train(train_xy) # Naive Bayes classifier 적용
print(nltk.classify.accuracy(classifier, test_xy))
# => 0.80418
```

그림 15 소셜 분석 모델 구현 코드 일부

'131', '이더리움 킬러 이오스타세요', '이더리움에서 다른거로 갈아타야할지
 요?' '1
 row131
 '이제 여기서 조금만 더 떨어지면 채굴장 차단기 떨어집니다. 채굴장에 수많은
 그래픽카드가 매물로 쏟아지면 이제 할값에 개입을 할수 있는 기회가 오겠네요.'
 '140' '채굴장도 이제 스위치를 만지작 만지작심정지 상태인 이더리움부터 돈하
 일 바랍니다~비탈락 살 좀 찌르면 물물이 현상태 이더리움 같군요' '788' '작년에
 와서 코인시장=거품 이라고 소개했죠.한달전에 했던말 :이더리움의 개발자 비탈락
 루테런이 암호화폐는 언제든 제로로 떨어질 수 있다'며 투자 위험성을 다시 한번
 경고했다.자시기가 열귀놓고 주워먹는 남입;오면 죽빵날리러 가야겠다;말못하게 해야
 합니다;할 ~개발자가 세력임? 내 이더 살려내라 이놈들아' '이더리움 개발자 서울오
 네요']
 row140
 '힘 좀 내보장..한때 70층도 갔잖니....' '409' '언제쯤 오를까요?ㅜㅜ;4월은 와
 죠... 선을 만기기간엔 강 빼두시는게 정신건강에 이롭죠;뽀 흥기도없네요ㅡㅡ;물
 시거나..조금더 오르면 빠져 단타 열심히 하셔야 겨우 복구되죠.;그래야하는데 원금
 손실이 넘겨 이러지도저러지도 못하고ㅜㅜ아이고아이고;물탱이로 다 빠지마시고 분할
 정리하면서 그만타요.;네네 ㅜㅜㅜㅜ;고인..;코인 아니라 고인 ㅋㅋㅋ' '이름아.
 ... 이름아....']
 row409
 'Ethereum 2015. 7. 30 ~ 2018. 3. 27 <1> '1' '121' '30'만년 회귀..;웃프
 요...진짜 웃프네 ㅋㅋㅋ 아오;;;either 이도저도 아닌...;오 재치쟁이.. 슬프데 또
 웃기네요;웃긴데 진짜 웃을수가없어요 ㅠ 전 이더 즐더는 아님데 ㅠ' '이더리움
 ... ㅜㅜㅜㅜ']
 row1
 194539
 ['<./Punctuation', 68629], '<영화/Noun', 51365], '<하다/Verb', 50281], '<이/Jos.
 ', 39123], '<보다/Verb', 34764], '<의/Josa', 30480], '<../Punctuation', 29055],
 '<에/Josa', 27108], '<가/Josa', 26696], '<을/Josa', 23481]]
 이/Determiner 것/Noun; 적/Suffix 인/Josa; 이/Determiner 거/Noun; 안/Noun
 리다/Verb; 것/Noun 은/Josa; 10/Number 적/Noun; 배우/Noun 들/Suffix; 수/Noun

그림 16 소셜 분석 모델 실행 화면 일부

다섯 번째로 알고리즘 모델이다. 기본적으로 이동평균선의 5 일선과 20
 일선 위에서 시가가 시작되는지 판단하고, 위에 있다면 시가로 매수한다.
 표준 편차를 활용하여 기준 매도가를 설정한다. 기준 매도가에 도달하지
 않을 경우에는 종가에 손해를 보더라도 매도한다.

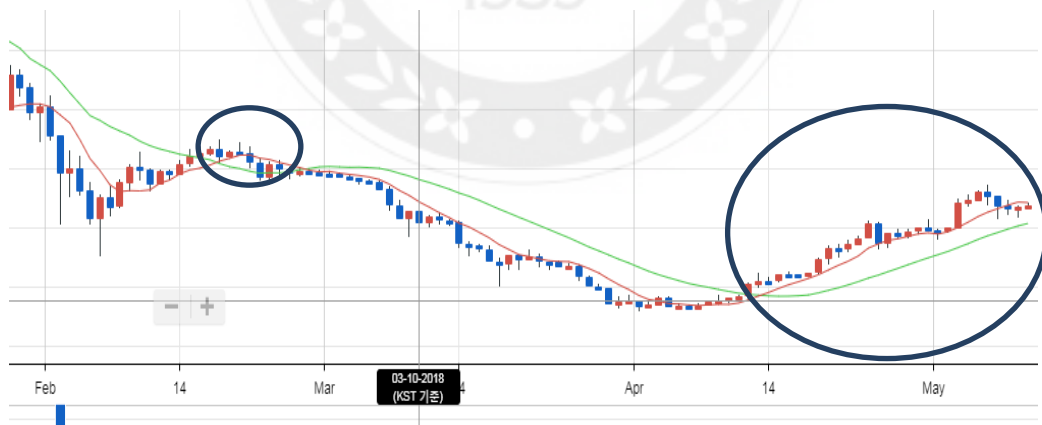


그림 17 골든 크로스 정의

4.2. 정성평가 및 정량평가

정량평가를 위하여 백 테스트를 진행하였다. 알고리즘 모델만으로 가치를 예측하여 매수와 매도를 진행했을 경우와 가치 기반 통합 모델로 가치를 예측하여 매수와 매도를 진행한 수익률을 비교하였다.

암호화폐의 종류인 이더리움으로 진행했다. 2018.01.29 ~ 2018.05.08의 데이터 세트로 진행하였다.

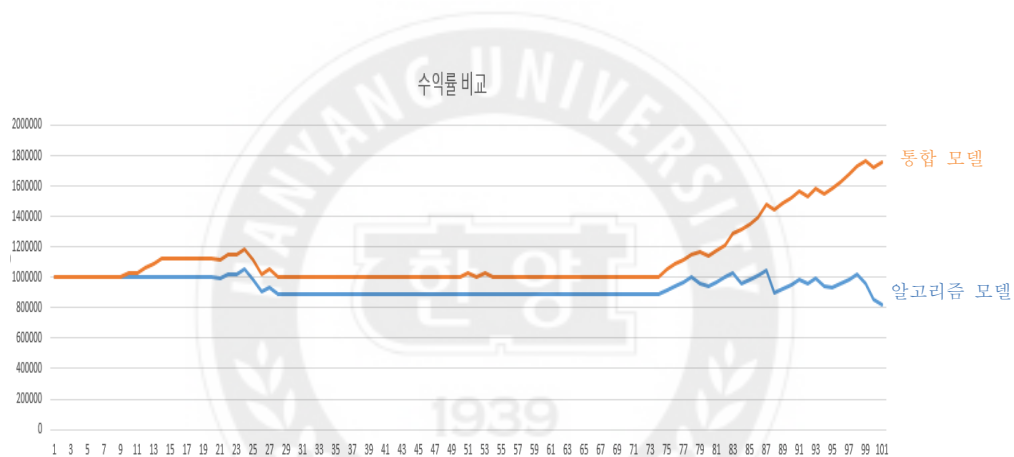


그림 18 수익률 비교

위의 구현 내용에 대하여 정성평가 및 정량평가를 한 내용을 정리하면 다음 표와 같다.

Table 3 정량 평가

구분	알고리즘 모델	통합 모델
시작 금액	1,000,000 원	1,000,000 원
최종 금액	855,882 원	1,760,511 원
수익률	-15.5%	76.0%

알고리즘 모델의 경우 5일 선과 20일 선의 이동평균선을 활용한 골든 크로스 기법만을 사용하였다. 골든 크로스가 발생하는 시점에만 시가에 매매하고 기준 % 도달시 매도하는 전략이나 기준 %까지 도달하지 않을 경우 종가에 매도하였다. 21~27일 사이 골든 크로스 부분이어서 매수와 매도가 진행되었으나 기준 % 도달하지 못하는 경우가 많아 손해가 발생하였다. 75~100일 사이에도 골든 크로스가 발생하였으나 수익과 손실이 반복되면서 결국 손해가 더 많이 발생하였다.

통합모델의 경우에는 9~11일 사이에 소셜 분석 모델에 의하여 매수가 진행되었다. 21~27일에는 알고리즘 모델과 같은 패턴으로 손실이 발생하였다. 50~55일에는 다시 소셜 분석 모델에 의하여 거래가 진행되었으나 큰 효과를 얻지 못했다. 75~100일 사이에는 가치 예측 모델의 기울기 변화에 의해 기준 %가 변경되면서 안정적인 수익률을 나타내었다.

또한 위의 모델을 로보 어드바이저로 구축하였을 경우에 대한 정성평가는 표로 정리하면 아래와 같다.

Table 4 정성 평가

구분	사용자	로봇어드바이저	비고
접근성	1일 이상	1일 이내	기초지식이 필요
수익성	높다	낮다	
거래 시간	높다	낮다	
데이터 검증 (신뢰성)	낮다	높다	
편리성	낮다	높다	
시스템 구축 비용	낮다	높다	

제 5 장 결론 및 향후 연구

5.1. 결론

위의 시스템에 대한 결론은 다음과 같이 장점 및 단점을 정리할 수 있다.

첫 번째 장점으로 가치 기반 통합모델을 활용하여 매매를 진행할 시 기존의 알고리즘 모델만으로 진행했던 방법 보다 수익률이 높아짐을 확인할 수 있었다.

두 번째 장점으로 편리성을 들 수 있다. 사용자가 별도의 행위를 통하여 거래 진행을 필요로 하지 않으며 상세한 금융지식이 없어도 사용할 수 있다.

세 번째로 활용성 면에서 로보 어드바이저로 신규 사용자에게 편리성을 어필함으로써 수익률을 높일 수 있고 그에 따라 신규 사용자에 대한 접근성을 높일 수 있다.

반면 단점으로는 다양한 항목을 빠르게 연산 처리하기 위해서는 고성능의 병렬처리를 할 수 있는 그래픽카드가 다량으로 필요하기 때문에 비용이 발생한다.

하지만 높은 개발 비용 및 값비싼 하드웨어가 필요하지만 데이터의 높은 신뢰도, 편리성, 활용성이 더 크기 때문에 개발 및 구축이 필요하다.

5.2. 향후 연구

위의 논문을 기반으로 향후 필요한 연구에는 다음과 같은 것들이 존재한다.

첫 번째로 다양한 알고리즘 모델을 추가한다. 월간 단위 ~ 분 단위까지의 많은 시간 단위와 다양한 매개 기법을 활용하여 진행하면 더 좋은 연구 결과가 나올 것이다.

두 번째로 인공지능 모델 중 LSTM을 사용하고 있는데 다양한 알고리즘으로 진행한다.

세 번째로 가치 기반 통합 모델의 결과를 다시 한번 머신 러닝으로 학습시킨다.

참고문헌

- [1] 차은정, 뉴스 감성분석과 SVM을 이용한 주가지수 예측, 부산대학교 대학원 경영학과, 2016
- [2] 조아라, 다변량 분석을 적용한 가상화폐 가격 등락에 미치는 영향 연구, 숭실대학교 소프트웨어특성화 대학원, 2017
- [3] 정욱, 주식시세 예측을 위한 딥러닝 최적화 방법 연구, 경희대학교 소프트웨어융합과 대학원, 2018
- [4] 공인인증서를 대체하는 블록체인 기술,
<https://blog.naver.com/papabus/221202436028>
- [5] 블록체인 개념 완벽 정리,
<https://banksalad.com/contents/%EB%B8%94%EB%A1%9D%EC%B2%B4%EC%9D%B8-%EA%B0%9C%EB%85%90-%EC%99%84%EB%B2%BD-%EC%A0%95%EB%A6%AC-dh1do>
- [6] 안전한 금융 거래, 블록체인과 비트코인,
https://blog.naver.com/evergreen_sound/221150925304
- [7] 물류, 블록체인 기술을 입다,
https://blog.naver.com/dhl_korea/221249443137
- [8] 에너지시장의 새로운 바람, 블록체인, <http://blog.komipo.co.kr/598>
- [9] 블록체인 개념 완벽 정리,
<https://banksalad.com/contents/%EB%B8%94%EB%A1%9D%EC%B2%B4%EC%9D%B8-%EA%B0%9C%EB%85%90-%EC%99%84%EB%B2%BD-%EC%A0%95%EB%A6%AC-dh1do>
- [10] 블록체인 개념 완벽 정리,
<https://banksalad.com/contents/%EB%B8%94%EB%A1%9D%EC%B2%B4%EC%9D%B8-%EA%B0%9C%EB%85%90-%EC%99%84%EB%B2%BD-%EC%A0%95%EB%A6%AC-dh1do>
- [11] 이더리움 특징, <https://steemit.com/kr/@dongsamb/ethereum>

- [12] 양민희, 블록체인을 활용한 보험사 건강체 특약 할인 시스템 설계 및 구현, 한양대학교 전기및전자공학과 컴퓨터공학, 2018
- [13] 차은정, 뉴스 감성분석과 SVM 을 이용한 주가지수 예측, 부산대학교 대학원 경영학과, 2016
- [14] RNN(Recurrent Neural Networks),
<https://blog.naver.com/apr407/221238601658>
- [15] 김호현, LSTM/GRU 순환신경망을 이용한 시계열데이터 예측, 한국방송통신대학교 대학원 : 정보과학과 2017
- [16] 장옥주, LSTM 를 이용한 금융 시계열 예측 모델의 후행성 문제 해결에 관한 연구, 가톨릭대학교 일반대학원 : 수학과 2018
- [17] 로보어드바이저,
<https://namu.wiki/w/%EB%A1%9C%EB%B3%B4%EC%96%B4%EB%93%9C%EB%B0%94%EC%9D%B4%EC%A0%80>
- [18] 로보어드바이저, <http://blog.daum.net/babupapa/83>
- [19] 오픈 소스 딥러닝 소프트웨어, <http://se-arts.com/220656954920>
- [20] 빅데이터 분석을 위한 Python 으로 텐서플로우(Tensorflow) 시작하기,
<https://blog.naver.com/koys007/221053249605>
- [21] SW 테스트 자동화 동향, <http://www.sw-eng.kr/member/customer/Webzine/BoardView.do?boardId=000000000000000038710>
- [22] Selenium 을 이용한 UI 테스트, <https://testudo.blog.me/80162675973>
- [23] 다층 퍼셉트론 레이어 이야기,
https://tykimos.github.io/2017/01/27/MLP_Layer_Talk/

ABSTRACT

Proposal of integrated model based on artificial intelligence for cryptographic currency analysis

Cho, SeongGeun

Major in Computer Science and
Engineering

The Graduate School of Engineering
Hanyang University

Recent technical utility chains are leading innovation in areas such as financial institutions and logistics. Businesses and businesses (B2B) are influenced by businesses and individuals (B2C).

Block Chain technology can solve this problem. blockchains is to process the ledger information. More information is needed to get more information.

In this paper, we propose a model that integrates an algorithm model and an artificial intelligence model to accurately evaluate the value of the problem.

The first is the algorithm model. It often refers to various techniques that are often expressed in stock analysis. I used the basic golden cross technique.

The second is an artificial intelligence model. Social analysis model and value prediction model. Emotional analysis is a method of analyzing text based unstructured data. The subject of the article is an analysis technique that distinguishes whether the emotion is positive or negative.

LSTM was used in artificial neural network which is a statistical learning algorithm used in machine learning and cognitive science. It uses existing dating to train and predict future value based on it. Here, we focus on the slope of the predicted value.

When comparing the above algorithm and the integrated model using the data from January 29, 2018 to May 08, 2018 based on Ethereum in the cryptography, the algorithm model shows -15% And the combined model shows a yield of 76%.

Résumé

Proposition de modèle intégré basé sur l'intelligence artificielle pour l'analyse de monnaie cryptographique

Cho, SeongGeun

Faculté de la Science et

l'Ingénierie Informatique

École d'Ingénieur, l'Université

Hanyang

Les chaînes d'utilité techniques récentes mènent l'innovation dans des domaines tels que les institutions financières et la logistique. Les entreprises et les entreprises (B2B) sont influencées par les entreprises et les particuliers (B2C).

La technologie Block Chain peut résoudre ce problème. blockchains est de traiter les informations du grand livre. Plus d'informations sont nécessaires pour obtenir plus d'informations.

Dans cet article, nous proposons un modèle qui intègre un modèle d'algorithme et un modèle d'intelligence artificielle pour évaluer avec précision la valeur du problème.

Le premier est le modèle d'algorithme. Il fait souvent référence à diverses techniques qui sont souvent exprimées dans l'analyse des stocks. J'ai utilisé la technique de base de la croix d'or.

Le second est un modèle d'intelligence artificielle. Modèle d'analyse sociale et modèle de prédiction de valeur. L'analyse émotionnelle est une méthode d'analyse de données non structurées basées sur du texte. Le sujet de l'article est une technique d'analyse qui distingue si l'émotion est positive ou négative.

LSTM a été utilisé dans un réseau neuronal artificiel qui est un algorithme d'apprentissage statistique utilisé dans l'apprentissage automatique et la science cognitive. Il utilise la datation existante pour former et prédire la valeur future basée sur elle. Ici, nous nous concentrons sur la pente de la valeur prédite.

En comparant l'algorithme ci-dessus et le modèle intégré en utilisant les données du 29 janvier 2018 au 08 mai 2018 basé sur Ethereum dans la cryptographie, le modèle de l'algorithme montre -15% Et le modèle combiné montre un rendement de 76%.

감사의 글

지난 2년간의 시간을 돌이켜 보면 너무도 빠르게 지나갔습니다. 처음 다짐했던 후회 없이 열심히 노력하자는 마음가짐을 기억하면서 석사의 마지막인 이번 논문으로 끝을 맺습니다. 그동안 제 주변의 많은 분들이 도움을 주셨기 때문에 무사히 졸업을 할 수 있었다고 생각합니다. 모든 분들께 감사의 마음을 전합니다.

먼저 논문을 완성하는데 많은 조언을 해 주신 지도교수 조인휘 교수님께 진심을 다해 감사를 드립니다.

중간에 너무도 힘들어서 포기하고 싶었지만 뒤에서 밀어주고 앞에서 당겨준 나의 영원한 멘토들인 동기들에게 감사의 말을 전합니다.

사랑으로 저를 낳아주시고 헌신으로 길러주신 부모님과 친아들처럼 다정하게 항상 대해 주시는 장모님과 예쁜 조카 리아와 갓 태어난 신생아 쏘이의 두 엄마가 된 나의 사랑하는 동생 성희, 아프리카에서 열심히 일하는 매제 류종걸과 하나뿐인 나의 처제 헤미에게 감사의 말을 전합니다.

나에게 행복을 알려준 100일 된 나의 아들 조우진과 마지막으로 많은 격려로 항상 힘이 되어 논문을 쓸 수 있게 배려해준 사랑하는 나의 아내 김혜연에게 이 논문을 바칩니다.

연구 윤리 서약서

본인은 한양대학교 대학원생으로서 이 학위논문 작성 과정에서 다음과 같이 연구 윤리의 기본 원칙을 준수하였음을 서약합니다.

첫째, 지도교수의 지도를 받아 정직하고 엄정한 연구를 수행하여 학위논문을 작성한다.

둘째, 논문 작성시 위조, 변조, 표절 등 학문적 진실성을 훼손하는 어떤 연구 부정행위도 하지 않는다.

셋째, 논문 작성시 논문유사도 검증시스템 "카피킬러"등을 거쳐야 한다.

2018년06월14일

학위명 : 석사

학과 : 전기·전자·컴퓨터공학과

지도교수 : 조인휘

성명 : 조성근



한 양 대 학 교 공 학 대 학 원 장 귀 하

Declaration of Ethical Conduct in Research

I, as a graduate student of Hanyang University, hereby declare that I have abided by the following Code of Research Ethics while writing this dissertation thesis, during my degree program.

"First, I have strived to be honest in my conduct, to produce valid and reliable research conforming with the guidance of my thesis supervisor, and I affirm that my thesis contains honest, fair and reasonable conclusions based on my own careful research under the guidance of my thesis supervisor.

Second, I have not committed any acts that may discredit or damage the credibility of my research. These include, but are not limited to : falsification, distortion of research findings or plagiarism.

Third, I need to go through with Copykiller Program(Internet-based Plagiarism-prevention service) before submitting a thesis."

JUNE 14, 2018

Degree : Master
Department : DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
Thesis Supervisor : INWHEE JOE
Name : CHO SEONG GEUN


Signature