KIPS Trans. Softw. and Data Eng.
Vol.7, No.10 pp.387~396 pISSN: 2287-5905

https://doi.org/10.3745/KTSDE.2018.7.10.387

Prediction of Cryptocurrency Price Trend Using Gradient Boosting

Joo-Seong Heo[†] · Do-Hyung Kwon[††] · Ju-Bong Kim[†††] · Youn-Hee Han[††††] · Chae-Hun An[†††††]

ABSTRACT

Stock price prediction has been a difficult problem to solve. There have been many studies to predict stock price scientifically, but it is still impossible to predict the exact price. Recently, a variety of types of cryptocurrency has been developed, beginning with Bitcoin, which is technically implemented as the concept of distributed ledger. Various approaches have been attempted to predict the price of cryptocurrency. Especially, it is various from attempts to stock prediction techniques in traditional stock market, to attempts to apply deep learning and reinforcement learning. Since the market for cryptocurrency has many new features that are not present in the existing traditional stock market, there is a growing demand for new analytical techniques suitable for the cryptocurrency market. In this study, we first collect and process seven cryptocurrency price data through Bithumb's API. Then, we use the gradient boosting model, which is a data-driven learning based machine learning model, and let the model learn the price data change of cryptocurrency. We also find the most optimal model parameters in the verification step, and finally evaluate the prediction performance of the cryptocurrency price trends.

Keywords : Price Prediction, Cryptocurrency, Machine Learning, Supervised Learning, Gradient Boosting

그래디언트 부스팅을 활용한 암호화폐 가격동향 예측

허 주 성[†] · 권 도 형[††] · 김 주 봉[†††] · 한 연 희[††††] · 안 채 헌[†††††]

요 약

과거부터 주식시장의 주가 예측은 풀리지 않는 난제이다. 이를 과학적으로 예측하기 위해 다양한 시도 및 연구들이 있어왔지만 정확한 가격을 예측하는 것은 불가능하다. 최근 분산 원장이라는 개념을 기술적으로 구현한 최초의 암호화폐인 비트코인을 시작으로 다양한 종류의 암호화폐가 개발되면서 암호화폐 시장이 형성되었고, 그 가격을 예측하기 위해 다양한 접근들이 시도되고 있다. 특히, 기존의 전통적인 주식시장에서의 주가 예측 기법들을 적용하려는 시도부터 딥러닝과 강화학습을 적용하려는 시도까지 다양하다. 하지만 암호화폐 시장은 기존 주식 시장에는 없던 여러 가지 새로운 특징을 가지는 시장으로서 전통적인 주식 시장 분석 기술뿐만 아니라 암호화폐 시장에 적합한 새로운 분석 기술에 관한 수요가 증가하고 있는 상황이다. 본 연구에서는 우선 빗썸의 API를 통하여 7개의 암호화폐 가격 데이터를 수집 및 가공하였다. 이후, Data-Driven 방식의 지도학습 기반 기계학습 모델인 그래디언트 부스팅 모델을 채택하여 암호화폐 가격 데이터 변화를 학습하고, 검증단계에서 가장 최적의 모델 파라미터를 산출하고, 최종적으로 테스트 데이터를 활용하여 암호화폐 가격동향 예측 성능을 평가한다.

키워드 : 가격 예측, 암호화폐, 기계학습, 지도학습, 그래디언트 부스팅

1. 서 론

과거부터 주식시장의 주가 예측은 풀리지 않는 난제이다. 이를 과학적으로 예측하기 위하여 다양한 시도 및 연구들이 있어 왔지만, 아직까지 정확한 가격예측은 불가능하다[1]. 현재 주식시장에서는 이미 사람이 직접 투자하는 방식 대신, 알고리즘 트레이딩 프로그램을 활용하여 수익을 창출하는 것이 일반화 되어 있다.

최근 분산 원장이라는 개념을 기술적으로 구현한 최초의 암호화폐인 비트코인[2]을 시작으로 다양한 종류의 암호화폐가 개발되면서 암호화폐 시장이 형성되었고, 그 가격을 예측하기 위해 다양한 접근들이 시도되고 있다[3-5]. 특히 기존의 전통적인 주식시장에서의 예측 기법들을 적용하려는 시도부터 딥러닝과 강화학습을 적용하려는 시도까지 다양하다[8-10].

암호화폐 시장은 비트코인이 처음 등장한 이후의 상대적으로 짧은 역사를 갖고 있다. 그렇기 때문에 기존 주식시장에

※ 이 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2018R1A6A1A03025526).
† 준 회 원 : 한국기술교육대학교 컴퓨터공학부 석사수료
†† 준 회 원 : 한국기술교육대학교 창의융합공학협동과정 ICT융합 석사과정
††† 준 회 원 : 한국기술교육대학교 컴퓨터공학부 석사과정
†††† 종신회원 : 한국기술교육대학교 컴퓨터공학부 정교수
††††† 비 회 원 : 한국기술교육대학교 메카트로닉스공학부 조교수
Manuscript Received : July 13, 2018
First Revision : August 17, 2018
Accepted : August 20, 2018
* Corresponding Author : Youn-Hee Han(yhhan@koreatech.ac.kr)

서 활용하는 알고리즘 트레이딩 방식으로 암호화폐의 가격을 예측하기 위해서는 먼저 암호화폐 가격 데이터의 특성에 대한 고찰이 필요하다. 기존의 주가 분석과 비교할 때 암호화폐 시장에서 거래되는 암호화폐들에 대한 데이터 분석은 전무하다. 또한 본질적으로 주가 데이터와 암호화폐 가격 데이터는 시계열 데이터로서 무작위적인 특성을 갖고 있으나, 암호화폐 가격 데이터가 더욱 변동성이 크며 특정 암호화폐의 가격에 의하여 다른 암호화폐의 가격이 영향을 받는다는 특징이 있다. 따라서 암호화폐 시장은 기존 주식 시장과는 다르게 여러 가지 새로운 특징을 가지는 시장이라고 할 수 있으며, 전통적인 주식 시장 분석 기술뿐만 아니라 암호화폐 시장에 적합한 새로운 분석 기술에 관한 수요가 증가하고 있는 상황이라고 할 수 있다.

본 연구에서는 빗썸의 API를 통해 데이터를 수집 및 가공하여 활용한다[11]. 또한 기존의 전통적인 알고리즘 트레이딩 모델에 기반을 두는 Theory-Driven 방식인 평균회귀 테스트 등을 통해 암호화폐 데이터의 무작위성을 알아보고 나아가 Data-Driven 방식을 활용하여 지도학습 기반 기계학습 모델에 대해 알아보고 이를 활용해 암호화폐 가격동향을 예측한다.

본 논문의 구성은 다음과 같다. 제2장에서는 본 논문과 관련하여 암호화폐에 대해 알아보고, Theory-Driven과 Data-Driven 두 가지 접근 방식을 참고하여, 알고리즘 트레이딩 모델을 암호화폐 시장에 적용할 수 있는지를 살펴보기 위해 평균회귀 테스트를 수행한다. 그리고 대표적인 지도학습 기반 기계학습 모델들을 살펴본다. 제3장에서는 학습을 위한 암호화폐 데이터의 수집 및 전처리 과정에 대해 설명한다. 제4장에서는 모델 학습을 위한 최적 파라미터를 찾기 위해 그리드 탐색 기반 k-Fold 교차 검증 방법을 제안한다. 제5장에서는 모델의 성능평가 지표를 활용해 입력 데이터의 조건을 달리하여 학습한 예측 모델에 대해 상세히 실험 및 비교 분석을 하고, 제6장에서 본 논문의 결론을 제시한다.

2. 관련 연구

2.1 암호화폐

암호화폐는 블록체인을 통한 분산원장기술을 이용한 디지털화폐를 말한다. 여기서 블록체인이란 거래정보가 기록된 관리 대상 데이터를 '블록'이라고 하는 소규모 데이터들이 P2P 방식을 기반으로 생성된 체인 형태의 연결고리 기반 분산 데이터 저장환경에 저장되어 누구도 임의로 수정할 수 없고 누구나 변경의 결과를 열람할 수 있는 분산 컴퓨팅 기술 기반의 데이터 위 변조 방지 기술이다. 대표적인 암호화폐인 비트코인은 인터넷상에서 개인 대 개인(P2P) 간에 이용될 목적으로 암호체계에 기초해 설계되어, 금전적 가치가 전자적 형태로 저장되어 지급수단으로 사용되지만 정부나 중앙은행에 의해 지급이 보장되지 않는다는 점에서 기존의 법정화폐와 차이가 있다 [12].

2.2 주가 예측 연구

암호화폐 가격 예측과 유사한 분야인 주가 예측 분야에서는 관련 연구들이 상당부분 진행되어 있다. 특히 최근 딥러닝 기술을 주가 가격 예측에 적용하려는 시도들이 증가하고 있는 추세이다. [8]에서는 주가 데이터에 서로 다른 15가지의 입력 피쳐를 이용하여 삼성전자 주가 예측을 위한 학습모델을 구축하여 50% 이상의 예측 성능이 나옴을 보이고 있다. [13]에서는 딥러닝 기반 학습 모델을 개발하여 56% 가량의 성능이 나옴을 보이고 있다. 특히, 시가, 종가, 고가, 저가 등의 단순 주식 데이터만으로 예측 성능을 높였다는 점이 주목할 만하다. Stanford University에서 수행된 다양한 기계학습 모델을 이용한 주식 가격 예측 연구 결과[14]에서는 2008년 9월 1일부터 2013년 11월 8일까지의 주식 가격 훈련 데이터를 통해 훈련한 각 모델들의 성능이 44.52%~58.2% 정도였다고 보고한 바 있다. [15]에서는 신경망을 기반으로 하는 새로운 주식 거래 시스템을 만들어 최적의 거래 정책을 선정해내고자 하였다.

주가에 대한 예측 연구결과는 많이 보고되고 있지만, 이제 대중에게 알려져서 관심을 받고 있는 암호화폐 가격 예측에 대한 연구 결과는 아직 미비하다. 현재까지 보고되고 있는 암호화폐 가격 데이터와 관련된 연구는 [5]와 [16]에서 보고되고 있다. 하지만, [5]와 [16]은 암호화폐 시장에 대한 투자에 있어서 어느 암호화폐에 많은 비중을 두고 투자를 해야 할지를 결정하는 포트폴리오 관리를 목적으로 강화학습으로 암호화폐 투자 포트폴리오를 구성하는 방안에 대해 제안하고 있다. 최근에는 산업체 위주로 단순한 기법으로 암호화폐 가격을 예측하여 가입자들에게 암호화폐 가격 등락 신호를 제공하는 사업이 많이 등장하고 있지만, 전통적인 기계학습 분석을 적용하여 암호화폐 가격 예측을 학술적으로 다루는 연구는 매우 최근에 시작되었다고 볼 수 있으며, 본 연구 논문은 그러한 연구 결과를 제시한다.

주가 예측 모델은 주식의 가치를 평가하고, 이에 따라 매수, 매도, 보유여부를 결정하며, 시장 평균수익보다 월등한 수익을 내는 알파 모델과 시장 평균이나 이를 약간 웃도는 정도의 수익을 내는 베타 모델로 나눌 수 있다. 주로 사용하는 모델은 알파 모델이며, Theory-Driven 방식과 Data-Driven 방식으로 나눌 수 있다. Theory-Driven은 어떤 가설을 세우고, 이를 기반으로 알파 모델을 만들기 때문에 모델 설계자의 측면에서 보면 자신의 모델이 무엇인지, 여러 변수가 어떻게 영향을 미치는지 파악하기가 비교적 용이하다. 대표적인 Theory-Driven 모델에는 평균회귀 모델이 있다. 평균회귀 모델은 시계열 데이터가 과거의 평균값으로 회귀하려는 경향을 가지며, 관심 데이터 값들이 정규분포를 따르면서 무작위적인 특성이 없어야 함을 가정한다. 하지만, 암호화폐 가격 데이터에 대하여 Augmented Dickey-Fuller (ADF) 테스트, 허스트 지수, Half Life 등으로 평균회귀 테스트[17, 18]를 수행한 결과, ADF 테스트의 경우, 실험을 진행한 암호화폐들 전부에서 검정 통계량값이 기각값을 넘지 못하기 때문에 평균회귀 모델을 적용하기 적합하지 않으며, Hurst Coefficient 값

의 경우, 대부분의 암호화폐가 0.5보다 작은 값을 나타내었으나 유의미한 수준이 아님을 확인하였다. 또한 Regression half life값은 EOS, BTG와 같은 암호화폐들의 경우에는 상대적으로 낮은 값을 보였으나, 해당 값이 회귀 모델을 적용하기에 적합한 값인지에 대해서는 확신할 수 없음이 기존 연구에서 밝혀진바 있다[19].

따라서, 본 논문에서는 기존의 전통적인 평균회귀 모델이 변동성이 매우 큰 암호화폐의 가격 예측에는 적합하지 않음에 착안하여, 그 대안으로서 기계학습 모델을 기반으로 하는 Data-Driven 방식을 활용하여 분석을 수행한다. Data-Driven 방식은 임의의 가설 기반의 모델을 세우지 않은 채 데이터 분석을 시작하며, 최근 이슈가 되고 있는 기계학습 방법론을 적용하여 암호화폐 가격 동향을 예측할 수 있는 모델을 직접 구성할 수 있다는 장점이 있다. 본 논문에서는 Data-Driven 방식의 기계학습 모델을 적용하여 암호화폐의 가격 변동을 예측한다.

2.3 기계학습 모델

기계학습에서 쓰이는 대표적인 모델로는 1)k-NN, 2)Logistic Regression 및 SVM 등의 선형 모델, 3)결정 트리, 그리고 4)랜덤 포레스트와 그래디언트 부스팅 등의 앙상블 기법 등이 있다. 암호화폐에 대한 가격 예측 분석 이전에 전통적으로 주식 가격 데이터에 대한 예측 분석은 금융공학에서 비교적 활발히 연구가 이루어진 편이다. k-NN의 경우 모델 자체가 이해하기 쉬운 모델이며, 매개변수를 딱히 조정하지 않아도 좋은 성능을 발휘하는 편이다[20]. 하지만 학습 데이터셋이 클 경우에는 예측 속도가 느려서 현업에서는 잘 쓰이지 않는 모델이다. 한편, 선형 모델은 학습 속도가 빠르고 추론 또한 빠른 속도로 이루어진다는 장점이 있으며, 학습 데이터셋이 큰 경우와 희소한 데이터셋인 경우에도 비교적 정확도가 높으며, 특성 데이터가 고차원일 경우에도 비교적 그 성능이 높다고 알려져 있다. 현재 본 논문의 암호화폐 가격 데이터의 단위는 거래량을 제외한 시가, 종가, 고가, 저가가 모두 비슷한 단위이므로, 암호화폐 가격 데이터 예측에 SVM을 적용하고자 하는 시도 또한 유의미하다고 할 수 있다[21]. 다만 데이터셋이 많을 때는 SVM 모델의 예측 속도가 떨어져 잘 작동하지 않는다는 단점이 있다. 또한 SVM은 데이터 전처리와 매개변수 설정에 많은 신경을 써야한다. 이러한 단점들로 인해, 최근에는 SVM보다는 랜덤 포레스트나 그래디언트 부스팅과 같은 전처리가 거의 필요 없는 트리 기반 모델을 통해 예측 분류기를 만들고자 하는 시도들이 많이 보고되고 있다[22, 23]. 랜덤 포레스트의 경우, 매개변수의 조정 없이도 기본적으로 좋은 성능을 낸다. 그러나 트리의 깊이가 깊어질수록 더 많은 메모리와 더 긴 훈련시간을 요구한다.

따라서, 본 논문에서는 대표적인 기계학습 모델인 그래디언트 부스팅(Gradient Boosting) 모델을 적용해 실험을 진행한다[24]. 그래디언트 부스팅의 경우, 랜덤 포레스트보다는 매개변수 조정에 더 신경써야 한다는 단점이 있으나, 본 논문에서는 이러한 단점을 그리드 탐색(grid search) 방식을 활용하여 가장 적합한 매개변수를 찾아냄으로써 해결하고자 하였으며, 이를 통해 그래디언트 부스팅의 주요 장점인 메모리를 적게 사용하면서도 빠른 예측이 가능한 이점을 살리고자 하였다. 또한 그래디언트 부스팅 모델은 여러 머신러닝 경연 대회에서 상위권에 입상을 한 팀들이 사용한 모델이며, 최근 주식 가격 데이터 예측 모델로도 흔히 사용되고 있고 있는 보편적인 모델이라고 알려져 있다[25, 26].

3. 암호화폐 데이터

본 장에서는 빗썸 거래소의 API를 활용하여 데이터를 수집한 과정을 설명한다. 또한 암호화폐 가격 예측을 위해 사용된 데이터의 구조에 대해 설명하고, 수집한 데이터의 여러 문제점을 해결하기 위한 비정상 데이터 처리 과정에 대해 설명한다. 더불어, 지도학습을 위한 데이터 준비를 위해 데이터를 새롭게 정의하고 가공한다.

3.1 데이터 수집 및 처리

빗썸에서 제공하는 API는 암호화폐별로 UNIX Timestamp, 시가, 종가, 고가, 종가, 거래량 데이터를 제공한다. 본 논문에서는 데이터 수집을 위해 Python 라이브러리를 활용했으며, 수집한 데이터는 전처리 과정을 거쳐 csv 파일의 형태로 로컬 드라이브에 저장 및 활용한다.

빗썸에서는 2018년 6월 1일 기준 최근에 추가된 암호화폐를 포함한 총 30개의 암호화폐에 대한 거래 가격을 포함한 각종 데이터 및 차트를 제공하고 있으며, API를 활용해 암호화폐 데이터들의 수집 시간 주기는 10분으로 하였다.

암호화폐 가격 예측을 위해 딥러닝 모델에 사용되는 데이터는 크게 학습 데이터, 테스트 데이터, 검증 데이터로 분류되고, API에서 제공하는 각 암호화폐별 데이터의 양이 다르기 때문에 본 논문에서는 데이터가 많은 상위 7개의 암호화폐(BTC, BCH, ETC, DASH, ETH, XRP, LTC)으로 실험을

Table 1. Collected Cryptocurrency Data

	Cryptocurrency Name	Collection Period	Unit (min.)	Number
1	BTC	2017-06-09 08:50:00~ 2018-05-18 12:00:00	10	37008
2	ETH	2017-06-09 09:00:00~ 2018-05-18 12:00:00	10	37007
3	XRP	2017-06-09 09:00:00~ 2018-05-18 12:00:00	10	37008
4	BCH	2017-08-04 21:40:00~ 2018-05-18 12:00:00	10	28867
5	LTC	2017-06-09 09:00:00~ 2018-05-18 12:00:00	10	37008
6	DASH	2017-06-09 09:00:00~ 2018-05-18 12:00:00	10	37008
7	ETC	2017-06-09 09:00:00~ 2018-05-18 12:00:00	10	37008

Table 2. Example of Collected BCH Price Data

	UNIX Timestamp	Date&Time (String)	Opening Price	Closing Price	High Price	Low Price	Volume
1	149632680000	2017-06-01 23:20	3094000	3118000	3127000	3094000	191.580
2	149632740000	2017-06-01 23:30	3116000	3119000	3124000	3116000	63.8428
3	149632794000	2017-06-01 23:39	3119000	3116000	3119000	3116000	52.9553
4	149632860000	2017-06-01 23:50	0	0	0	0	0
5	**Missing**	**Missing**	**Missing**	**Missing**	**Missing**	**Missing**	**Missing**
6	**Missing**	**Missing**	**Missing**	**Missing**	**Missing**	**Missing**	**Missing**
7	149632985400	2017-06-02 00:19	0	0	0	0	0
8	149633100000	2017-06-02 00:30	3116000	3119000	3124000	3116000	101.68

진행한다. 학습을 위해 암호화폐 7개의 공통 기간인 2017년 8월 4일 21시 40분부터의 데이터를 사용했으며, 암호화폐별 가격 차이가 크기 때문에 정규화한 가격 정보를 사용한다. Table 1은 각 암호화폐별 데이터 수집 현황으로서 7개의 암호화폐별 수집 데이터의 기간 및 10분 단위의 수집 개수를 보여준다. Table 2는 BCH에 대해 10분 간격으로 수집된 데이터 예시를 보여준다.

3.2 비정상 데이터 처리

실험에 사용된 데이터는 빗썸에서 제공하는 API로부터 수집되었다. 수집된 데이터에는 정상 데이터 이외에, 손실 데이터, 제로 데이터, 부적합 데이터 등 정상적으로 입력되지 않은 데이터들이 존재한다. 정상 데이터란 아무 문제가 없는 데이터이며, 손실 데이터는 10분 간격으로 수집되지 않은 데이터를 말한다(Table 2의 5, 6번째 행). 제로 데이터는 10분 간격으로 수집되었으나 실제 내용이 비어 있는 경우이다(Table 2의 4, 7번째 행). 마지막으로 부적합 데이터란 데이터가 10분 간격에 딱 맞게 이어지지 않은 채 수집된 데이터를 의미한다(Table 2의 3, 7번째 행). 이와 같은 비정상 데이터들은 직전의 정상 데이터의 정보를 복사하여 보정하는데, 손실 데이터의 경우 비어있는 시간만큼 10분 단위로 직전 데이터를 복사하였고, 제로 데이터의 경우에도 마찬가지로 처리하였다. 부적합 데이터의 경우 이미 10분 단위의 다른 정상 데이터가 있을 경우 삭제하고, 정상 데이터가 없으나 시간의 조정이 필요하다면 조정된 시간으로 업데이트 하였다.

4. 기계학습 모델

4.1 지도학습 데이터 준비

1) 특성 데이터

학습 및 테스트 데이터는 앞선 절에서 수집 및 전처리 과정을 거친 데이터를 활용한다. [16] 연구와 유사하게 특성 데이터는 수집된 데이터에서 종가, 고가, 저가, 거래량만을 활용한다. i번째 암호화폐에 대한 기간 t의 종가 데이터를 $v_{i,t}$로 표기하며, $v_{i,t}^{(hi)}$는 고가 데이터, $v_{i,t}^{(lo)}$는 저가 데이터, $v_{i,t}^{(vo)}$는 거래량 데이터를 의미한다. 기간 t는 10분 간격으로 증가한다.

0번째 암호화폐의 가격 데이터는 기간 t에 관계없이 항상 1이고, 0번째 암호화폐는 나머지 암호화폐들에 대한 기준 통화의 역할을 하며 본 논문에서는 원화로 정했다. Equation (1)은 이를 수식으로 표현한 모습이다.

$$v_{0,t} = v_{0,t}^{(hi)} = v_{0,t}^{(lo)} = 1 \tag{1}$$

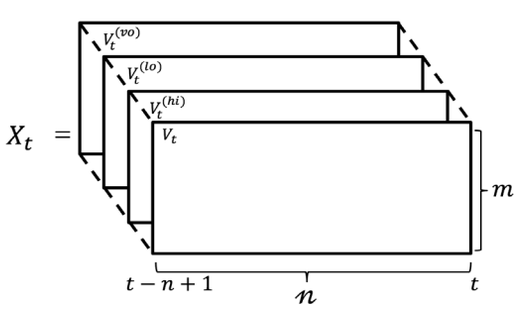


Fig. 1. Structure of Feature Data for n Window

기준 통화를 포함하여 8개(즉, $m=8$)의 암호화폐 가격 데이터를 포함하는 임의의 기간 t의 특성 데이터 X_t의 구조는 Fig. 1과 같다. 임의의 X_t는 8개 암호화폐에 대하여 기간 t기준에서 과거 n개의 종가, 고가, 저가, 거래량 데이터를 포함한다. 이 때 n은 윈도우 사이즈라고 일컫는다. 한편, n 윈도우에 대하여 마지막 종가 대비 벡터 요소별 나누기 연산($\oslash$)을 통해 [16] 연구에서 제시하는 것처럼 X_t 내의 모든 데이터를 정규화 하였다.

$$V_t = [v_{t-n+1} \oslash v_t | v_{t-n+2} \oslash v_t | \cdots | v_{t-1} \oslash v_t | 1]$$

$$V_t^{(hi)} = [v_{t-n+1}^{(hi)} \oslash v_t | v_{t-n+2}^{(hi)} \oslash v_t | \cdots | v_{t-1}^{(hi)} \oslash v_t]$$

$$V_t^{(lo)} = [v_{t-n+1}^{(lo)} \oslash v_t | v_{t-n+2}{}^{(lo)} \oslash v_t | \cdots | v_{t-1}^{(lo)} \oslash v_t]$$

$$V_t^{(vo)} = [v_{t-n+1}^{(vo)} \oslash v_t^{(vo)} | v_{t-n+2}{}^{(vo)} \oslash v_t^{(vo)} | \cdots | v_{t-1}^{(vo)} \oslash v_t^{(vo)} | 1] \tag{2}$$

Equation (2)에서 V_t 는 기간 t의 기준에서 과거 n 윈도우 내의 m개 암호화폐 가격의 종가 행렬을 의미하며, $V_t^{(hi)}$, $V_t^{(lo)}$, $V_t^{(vo)}$는 각각 고가, 저가, 거래량 행렬을 의미한다.

2) 타겟 데이터

암호화폐 i 의 시간 t 에 대한 타겟 데이터를 $v_{i,t}^{target}$ 이라 했을 때, $v_{i,t}^{target}$은 Equation (3)과 같다.

$$v_{i,t}^{target}=\begin{cases}1 & if\ v_{i,t}^{\theta}(1+\frac{\epsilon}{100})>v_{i,t}\\ 0 & otherwise\end{cases} \tag{3}$$

위 식에서, $v_{i,t}^{\theta}$는 시간 t 기준에서 암호화폐 i 의 θ 기간 이후의 종가를 의미한다. 이 때 Equation (3)에서 설명하는 $v_{i,t}^{target}$은 i번째 암호화폐에 대한 기간 t의 특성 데이터인 X_t를 기준으로, X_t가 지니는 전체 윈도우에서 마지막 종가 대비 다음 θ 기간이 지난 뒤의 종가가 ϵ%를 초과해 상승한 경우 1이 되며, 그렇지 않은 경우 0이 된다. 즉, ϵ은 상승률로서 타겟 데이터의 값을 정하는 기준이 된다. 시간 t에 대해 모든 암호화폐에 대하여 타겟 데이터 벡터를 구한 예시는 다음 Equation (4)와 같다.

$$v_t^{target}=[0,1,0,1,1,0,0,1]^T \tag{4}$$

3) 데이터 정규화

특성 데이터 X_t는 Equation (2)에 의하여 정규화가 되었지만 윈도우 내의 마지막 종가로만 요소별 나누기 연산을 통하여 정규화하기 때문에 그 값들의 분포가 다양해진다. 따라서, 기계학습에 X_t를 사용하기 전에 학습 특성 데이터 내 모든 값을 0과 1사이로 조정할 필요가 있다. 이를 위해 Equation (5)와 같은 Min-Max Scaler를 사용한다.

$$\frac{v}{Max(v)-Min(v)} \tag{5}$$

위와 같은 정규화 작업은 기계 학습의 훈련 데이터에만 활용하고, 테스트 데이터는 훈련 데이터를 정규화 할 때 사용한 최대, 최소값을 기준으로 정규화한다.

4.2 그래디언트 부스팅 모델

기계학습에서 부스팅(Boosting)이란 비교적 부정확한 약한 학습기(Weak Learner)를 묶어서 보다 정확하고 강한 학습기(Strong Learner)를 만드는 방식을 뜻한다. 일단 정확도가 낮더라도 첫 번째 트리 모델을 만들고, 드러난 약점(예측 오류)은 두 번째 트리 모델이 보완한다. 이와 같은 방법으로 다음 트리 모델에서 약점을 계속하여 보완하여 결국에는 강한 학습기를 구축한다.

손실함수(Loss Function, J)는 예측 모델의 오류를 정량화하며, 이러한 손실함수 값을 최소화하는 모델 내 파라미터를 찾기 위하여 일반적인 기계학습 모델들은 경사 하강(Gradient Descent) 방식을 사용한다. 그래디언트 부스팅은 이러한 파라미터 손실함수 최소화 과정을 모델 함수(f_i) 공간에서 수행하며, 손실함수를 모델 파라미터가 아니라 다음과 같은 (6)번 수식에 의해 현재까지 학습된 트리 모델 함수로 미분한다(아래 수식에서 ρ는 학습률).

$$f_{i+1}=f_i-\rho\frac{\delta J}{\delta f_i} \tag{6}$$

즉, 그래디언트 부스팅 모델에서 트리 모델 함수 미분값은 현재까지 학습된 모델의 약점을 나타내는 역할을 하며, 다음 트리 모델의 피팅을 수행할 때 그 미분값을 사용하여 약점을 보완하여 성능을 Boosting한다.

4.3 그리드 탐색 기반 k-Fold 교차 검증

지도학습 모델을 사용하여 기계학습을 수행할 때, 보다 좋은 성능을 내는 모델로 학습시키기 위해서는 모델이 지닌 하이퍼 파라미터에 대한 최적 설정이 필요하다. 따라서, 본 논문에서는 후보로 선정한 모든 파라미터 집합에 대해 최적의 하이퍼 파라미터를 찾는 그리드 탐색(Grid Search)[27] 기반 k-Fold 교차검증[28] 기법을 활용한다. 이와 같은 기법을 포함한 전체 모델 학습 절차는 Fig. 2와 같다.

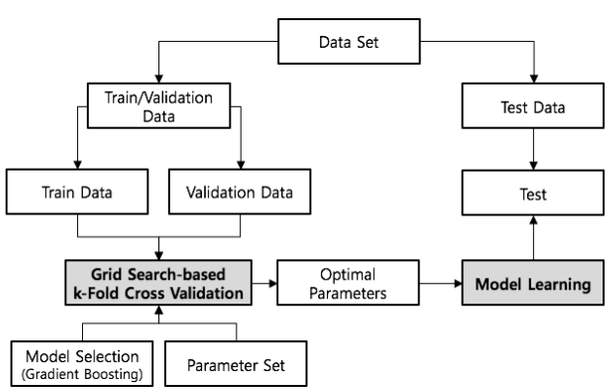


Fig. 2. Flow Chart of Model's Learning & Testing Procedure

전체 모델 학습 절차 중에서 그리드 탐색 기반 k-Fold 교차 검증은 다음과 같은 과정으로 수행한다. 먼저, 전체 데이터셋을 훈련 및 검증 데이터와 테스트 데이터로 분할한다. 이 때 미리 구성한 각 파라미터 집합에 대하여 훈련 및 검증 데이터를 5개의 훈련 데이터와 검증 데이터 폴드로 구성한다(즉, $k=5$). 각 파라미터 집합마다 5개 폴드를 활용하면서 모델을 구성하고 검증까지 수행하기 때문에 탐색 시간은 오래 걸리지만 더 정확한 최적 파라미터를 찾을 수 있다.

각 파라미터 집합마다 5개 폴드에 대한 훈련 및 검증을 하게 되고, 이러한 검증을 통하여 산출한 평균 정확도가 가장 높은

파라미터 집합을 하나 선정한다. 이렇게 선정한 최적 파라미터들을 활용하여 지도학습 기반 모델을 학습을 시킬 수 있다.

5. 실험 평가

5.1 성능 평가 지표

학습한 모델의 성능을 평가하기 위해서 본 논문에서는 분류 모델의 성능 평가 지표 중 정확도, 정밀도, 재현율, F1 Score 네 가지의 평가 지표를 활용하여 지도학습 모델의 결과를 분석 및 평가하였다.

Confusion Matrix는 실제 값과 예측한 값이 일치하는 것을 양성으로, 실제 값과 예측 값이 일치하지 않는 것을 음성으로 나누어 그 개수를 나타낸 표이다. Fig. 3은 Confusion Matrix를 나타낸 것으로서 행은 음성 클래스와, 양성 클래스를 의미하며 열은 음성으로 예측한 경우와, 양성으로 예측한 경우를 의미한다. 예를 들어, TN (True Negative)은 모델의 예측 결과는 음성 클래스로 나왔는데, 실제로도 음성 클래스인 경우를 말한다. FP (False Positive)는 모델은 양성 클래스로 예측했지만 실제로는 음성 클래스인 경우를 말한다. FN (False Negative)은 음성 클래스로 예측했지만 실제 결과는 양성 클래스인 경우를, TP (True Positive)는 예측한 결과와 실제 결과 모두 양성 클래스일 때를 의미한다.

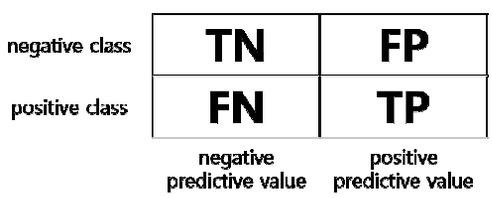


Fig. 3. Confusion Matrix

Confusion Matrix를 기반으로 산출한 첫 번째 성능 지표인 정확도는 예측 클래스와 실제 클래스가 일치한 경우의 비율을 뜻하며 정확도의 수식은 Equation (7)과 같다.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{7}$$

정밀도는 양성 클래스로 예측한 예측 결과들 중에서 실제 양성 클래스인 경우의 비율을 뜻하며 정밀도의 수식은 Equation (8)과 같다. 예를 들어, 암호화폐 시장에서 암호화폐 가격이 오르는 경우를 양성 클래스로 가정할 때, 정밀도가 낮다는 것은 암호화폐 가격이 떨어지는 것을 정확하게 예측하지 못한다는 것을 의미한다. 따라서 낮은 정밀도는 재화의 손실을 의미한다.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

재현율은 양성 검출율이라고도 불리우며, 전체 양성 클래스 중 실제로 양성 클래스로 예측한 경우의 비율을 뜻한다. 재현율의 수식은 Equation (9)와 같다. 예를 들어, 암호화폐 시장에서 암호화폐 가격이 오르는 경우를 양성 클래스로 가정할 때, 재현율이 낮다는 것은 암호화폐 가격이 오르는 것을 정확하게 예측하지 못하는 것을 의미한다. 결국 재현율이 낮다는 것은 높은 가치를 지닌 암호화폐를 얻을 기회를 놓쳤다고 해석 할 수 있다.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

마지막으로, F1 Score는 정밀도와 재현율의 조화 평균을 의미하며 Equation (10)과 같다. 재현율과 정밀도는 서로 상충관계에 있다. F1 Score는 이 두 지표를 잘 통합하여 정확성을 한 번에 나타내는 지표다. 일반적으로 높은 F1 Score를 얻기 위해서는 재현율과 정밀도가 모두 높아야 한다. 예를 들어, 암호화폐 가격 동향 예측에서, 양성을 암호화폐 가격이 오르는 것으로 가정할 때, 모델의 F1 Score가 높다는 것은 암호화폐 가격이 실제로 하락할 때 하락한다고 예측하고, 암호화폐 가격이 실제로 상승할 때 상승한다고 예측하는 것을 의미한다.

$$F1\ Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{10}$$

5.2 실험 환경

실험 평가를 위하여 구축한 실험 환경으로는 Python 3.6 및 scikit-learn 라이브러리를 사용했고, 좀 더 원활하게 최적 파라미터를 구하기 위하여 컴퓨터 6대에 Python 기반의 대용량 데이터 처리 및 분석에 적합한 DASK [29] 병렬 및 분산화 모듈을 설치하여 실험하였다.

BTC, ETH, XRP, BCH, LTC, DASH, ETC 암호화폐 가격을 포함하는 특성 및 타겟 데이터를 3.1절에서 언급한 방법대로 수집 및 가공하여 총 41,098개를 준비하였고, 이 중 36,988개는 훈련 데이터로 사용하고, 4,110개는 테스트 데이터로 사용하여 모델 학습 및 테스트를 수행하였다.

실험에 필요한 기본 실험 파라미터로는 $\theta = 10$, $n = 25$, $\tau = 1$, $\epsilon = 0.1$을 사용하였다. 이와 같은 4개의 실험 파라미터 중 τ는 예측 범위를 나타내며, 타겟 데이터를 산출할 때 특성 데이터의 마지막 종가로부터 $\tau * \theta$ 시간 이후 종가를 기준으로 산출함을 의미한다. 예를 들어 τ가 2이고 θ가 10분이면, 타겟 데이터는 20분 후의 종가를 기준으로 Equation (3)을 이용하여 산출된다. 다음 절부터는 언급한 4개의 실험 파라미터 각각에 대하여 해당 파라미터에 다른 값을 할당하면서 모델 성능을 비교 평가한다.

5.3 시간 단위에 따른 비교

시간 단위, 윈도우 사이즈, 예측 범위, 상승률 이렇게 네 개의 변인에 대하여, 시간 단위에 따른 모델 성능을 비교하기 위해 윈도우 사이즈, 예측 범위, 상승률을 각각 25, 1, 0.1로 고정한 후, 시간 단위의 값을 10분, 30분, 60분으로 변경해가면서 실험을 진행하였다. 암호화폐 가격이 상승할 때를 양성으로 정할 때 Fig. 4는 시간 단위에 따라 암호화폐별 예측 정확도와 F1 Score를 나타낸다.

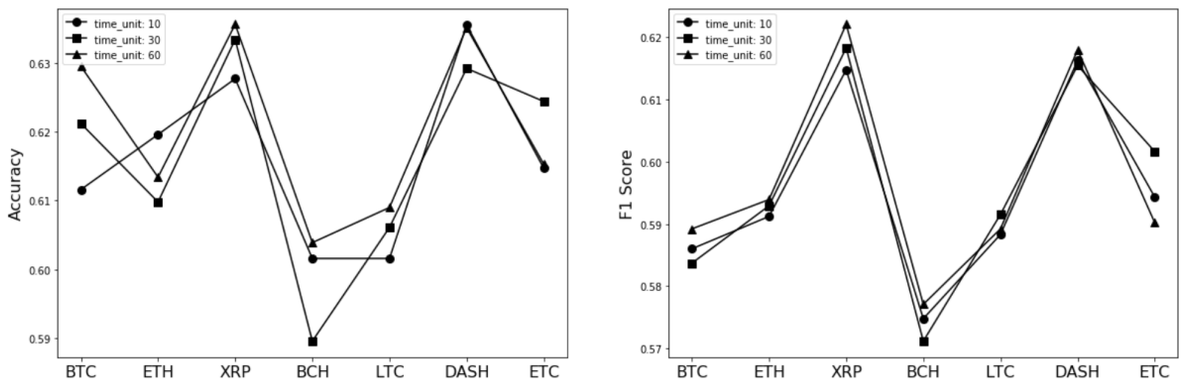


Fig. 4. Model Performance in Terms of Time Unit ($n=25$, $\tau=1$, $\epsilon=0.1$)

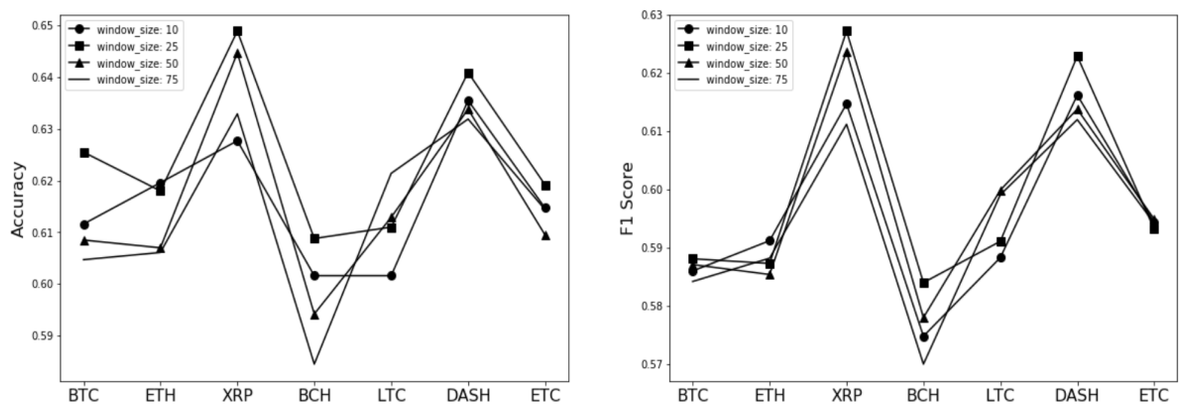


Fig. 5. Model Performance in Terms of Window Size ($\theta=10$, $\tau=1$, $\epsilon=0.1$)

그래프에서 볼 수 있듯이 다양한 시간 단위에 대해 대체적으로 XRP와 DASH의 정확도와 F1 Score는 높고 BCH의 경우는 상대적으로 낮았다(시간 단위가 60일 때, XRP 정확도는 0.635, DASH의 정확도는 0.635, BCH의 정확도는 0.593이며, XRP의 F1 Score는 0.622, DASH의 F1 Score는 0.618, BCH의 F1 Score는 0.578). 한편, 대체적으로 시간 단위가 60분 일 때 좋은 성능을 보이고 있지만 전반적으로 시간 단위 10분, 30분, 60분에 따른 정확도 차이는 크지 않음을 알 수 있었다.

5.4 윈도우 사이즈에 따른 비교

학습 데이터의 양에 많은 영향을 주는 윈도우 크기 n에 따른 비교를 위해 시간 단위, 윈도우 사이즈, 상승률을 각각 10, 1, 0.1로 고정한 후, n의 값을 10, 25, 50, 75로 변경해 가면서 실험을 진행하였다. 암호화폐 가격이 오르는 것을 양성으로 했을 때 Fig. 5는 윈도우 크기에 따른 암호화폐별 예측 정확도와 F1 Score를 나타낸다.

대체적으로 모든 암호화폐들에 대해서 윈도우 크기가 25일 때 정확도 및 F1 Score 성능 지표들이 높으며, 윈도우 크기가 75일 때는 두 성능 지표 값이 낮게 나왔다. 이는 오랜 기간 동안 수집된 데이터를 학습 데이터로 활용할 때, 암호화폐 가격 상승 및 하락을 예측하는 데에 오히려 부정적인 영향을 주는 것으로 분석된다. 이와 같은 경향은 시간 단위가 30분, 60분일 때도 비슷하게 나타난다.

본 실험에서도 대체적으로 XRP와 DASH의 성능 지표가 높게 나타났으며, BCH는 상대적으로 낮은 성능 지표를 산출한다(윈도우 크기가 25일 때, XRP 정확도는 0.648, DASH 정확도는 0.64, BCH 정확도는 0.608, XRP F1 Score는 0.63, DASH F1 Score는 0.628, BCH F1 Score는 0.584).

5.5 데이터 예측 범위(gap)에 따른 비교

데이터 예측 범위에 따른 비교를 위해 시간 단위, 윈도우 사이즈, 상승률의 값을 각각 10, 25, 0.1로 고정한 후, τ 값을 1, 2, 3으로 변경하면서 실험을 진행하였다. 암호화폐 가격이 오르는 것을 양성으로 할 때, Fig. 6은 데이터 예측 범위에 따른 암호화폐별 정확도와 F1 Score를 보여준다.

그림에서 볼 수 있듯이 예측하고자 하는 기간 범위가 길어질수록 모델의 성능이 낮아지는 것을 확인 할 수 있다. 하지만 XRP, BCH, LTC는 윈도우 내에 존재하는 마지막 시간 단위의 종가보다 두 번의 시간 단위가 지났을 때(즉, $\tau=2$일 때)의 종가 예측에 대한 성능 지표가 더 높음을 알 수 있다. 특히 앞서 살펴본 시간 단위에 따른 성능 분석과 윈도우 크기에 따른 성능 분석에서 가장 낮은 성능을 보이는 BCH에서 데이터의 예측 범위가 1일 때 보다 2일 때 그 성능이 높아지는 것에 주목할 만하다. 반면 DASH는 반대의 상황으로써,

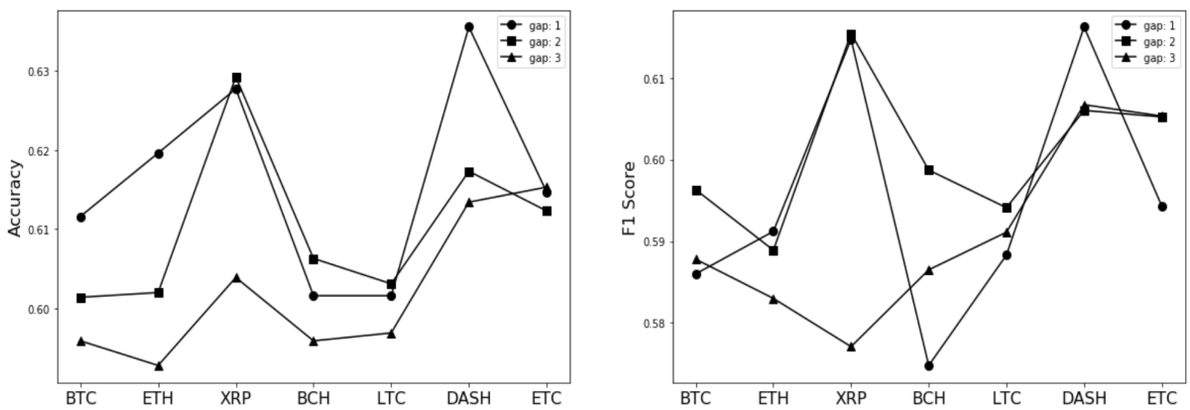


Fig. 6. Model Performance in Terms of Data Prediction Range (Gap) ($\theta = 10$, $n = 25$, $\epsilon = 0.1$)

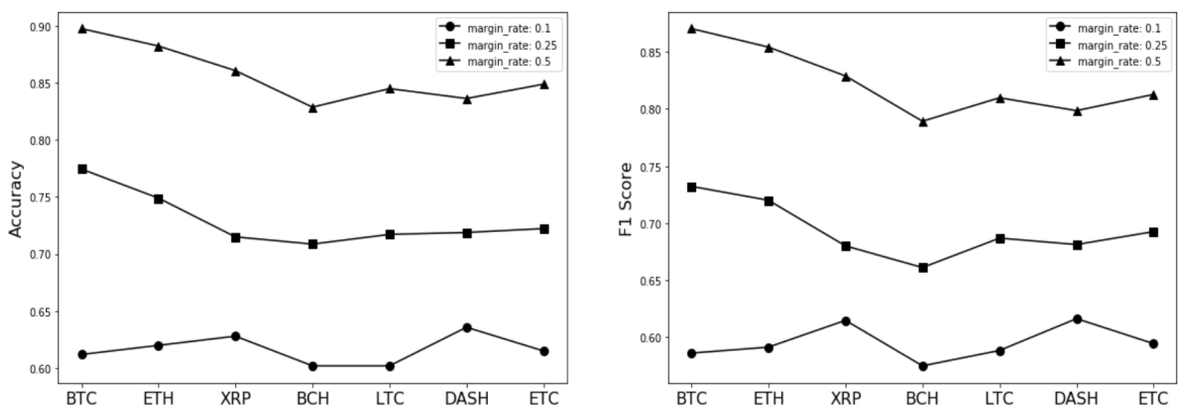


Fig. 7. Model Performance in Terms of Margin Rate ($\theta = 10$, $n = 25$, $\tau = 1$)

데이터 예측 범위가 2 또는 3일 때 보다 1일 때의 성능이 더 높음을 확인할 수 있다.

5.6 상승률에 따른 비교

마지막으로, 상승률 변경에 따른 비교를 위해 ϵ 값을 0.1, 0.25, 0.5로 변경하여 실험을 진행하였다. 암호화폐 가격이 오르는 경우를 양성으로 할 때 Fig. 7은 상승률 변경에 따른 암호화폐별 정확도와 F1 Score를 나타낸다.

그래프에서 볼 수 있듯이, 타겟 데이터의 값을 결정하는 상승률이 커질수록 모델의 성능이 크게 향상되는 것을 알 수 있다. 상승률의 값이 0.5일 때가 가장 높은 정확도와 F1 Score값을 가지며, 상승률의 값이 0.1일 때 가장 낮은 정확도와 F1 Score값을 갖는다. 하지만 상승률이 0.5일 때의 Confusion Matrix를 살펴보면 실제 양성 클래스인 경우가 극히 드물어서 전체 데이터의 5%이고, 대부분은 음성 클래스임을 알 수 있다. 따라서 상승률이 높아질수록 양성과 음성 클래스의 개수에 차이가 커지며, 이는 곧 데이터의 불균형이 심해진다고 이해할 수 있다. 즉, 모델은 대부분의 경우에서 예측 결과를 음성으로 내기 때문에 정확도 및 기타 성능 지표들이 좋게 나오는데 이는 학습 및 테스트 데이터의 불균형 정도가 높은 것이 그 이유라고 볼 수 있다.

6. 실험 결과 토의

5장에서 제시한 실험 결과를 종합한 결과 암호화폐별 예측 정확도는 XRP가 대체로 가장 높고 BCH가 가장 낮음을 알 수 있다. 그러므로, 향후 기계학습 모델을 통해 투자 수익을 높이는 데 있어서 XRP가 가장 유리한 암호화폐이며, BCH는 가급적 투자를 회피해야 할 암호화폐임을 알 수 있다.

한편, 각 암호화폐별로 시간 단위에 따른 모델의 성능에서는 뚜렷한 성능 차이가 없으며, 학습 데이터의 양이 많을수록 오히려 예측 성능에는 부정적인 영향을 주는 것을 관찰할 수 있다. 이는 암호화폐의 가격 변동 추이가 워낙 심하게 변하다 보니 오래 전의 데이터를 학습하는 것이 오히려 가까운 미래의 예측에 방해가 될 수 있음을 나타낸다. 또한 데이터 예측 범위의 실험 결과에 따르면, 더 먼 시점의 데이터를 예측하고자 할수록 성능이 낮아질 것이라는 직관과 일치한 실험 결과를 얻었다. 마지막으로, 상승률을 높게 설정할수록 예측 정확도가 높게 나오는 결과를 얻었으나, 이에 대해서는 데이터의 불균형한 정도가 영향을 준 것으로 해석된다.

마지막으로, 전반적인 실험 결과에서 볼 수 있듯, 그래디언트 부스팅 모델의 암호화폐 가격 동향 예측 모델의 정확도 및 성능 지표들은 60%정도로 비교적 높은 편이다.

7. 결 론

본 논문에서는 그레디언트 부스팅 모델을 활용해 암호화폐 가격 동향을 시간 단위, 윈도우 크기, 데이터 예측 범위, 상승률 변경에 따라 암호화폐 별로 예측하고, 분류 모델의 성능 평가 지표로서 정확도와 F1 Score를 활용하여 예측 성능을 비교 분석하였다. 대부분의 기존 주가 예측 관련 연구들에서는 50%대의 정확도를 보이나, 본 논문에서는 비교적 다른 기계학습 모델들 보다 우수하다고 입증된 그레디언트 부스팅 모델을 택하고, 그리드 탐색을 활용한 최적 하이퍼 파라미터 선정을 통해 약 60% 정도의 예측 정확도를 산출하였다. 향후 연구로는 Convolutional Neural Network나 Recurrent Neural Network와 같은 딥러닝 모델을 적용하여 더 높은 예측 성능을 산출해 볼 계획이다.

References

[1] M. S. Helen, C. Chester, A. Adam, K. Y. Dror, S. Eugene, and P. Tobias, "Quantifying Wikipedia Usage Patterns Before Stock Market Moves," Scientific Reports, May 2013.

[2] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," [Internet], http://www.bitcoin.org, 2008.

[3] R. Phillips and D. Gorse, "Predicting Cryptocurrency Price Bubbles Using Social Media Data and Epidemic Modelling," *IEEE Symposium Series on Computational Intelligence*, 2017.

[4] A. Radityo, Q. Munajat, and I. Budi, "Prediction of Bitcoin exchange rate to American dollar using artificial neural network methods," *International Conference on Advanced Computer Science and Information Systems*, 2017.

[5] Z. Jiang and J. Liang, "Cryptocurrency portfolio management with deep reinforcement learning," *Intelligent Systems Conference*, 2017.

[6] L. Li, Y. Wu, Y. Ou, Q. Li, Y. Zhou, and D. Chen, "Research on machine learning algorithms and feature extraction for time series," *The 28th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct. 2017.

[7] J. W. Lee, "A Stock Trading System based on Supervised Learning of Highly Volatile Stock Price Patterns," *Journal of KIISE : Computing Practices and Letters*, Vol.19 No.1, pp.23-29, 2013.

[8] Y. Song, J. W. Lee, and J. Lee, "Performance Evaluation of Price-based Input Features in Stock Price Prediction using Tensorflow," *KIISE Transactions on Computing Practices*, Vol.23, No.11, pp.625-631, 2018.

[9] Y. Kim, E. Shin, and T. Hong, "Comparison of Stock Price Index Prediction Performance Using Neural Networks and Support Vector Machine," *The Journal of Internet Electronic Commerce Resarch*, Vol.4, No.3, pp.221-243, 2004.

[10] A. M. Ho and R. M. Hyun, "Algorithm trading system development using machine learning," Hanbit Media(Inc), Apr. 2016, ISBN: 9788968488030.

[11] Bithumb [Internet], https://www.bithumb.com/

[12] T. Yook, "Change of Financial Systems by Virtual Currency or Cryptocurrency and its Legal Implications," *Kangwon Law Review*, Vol.53, pp.225-270, 2018.

[13] Y. Song and J. Lee, "A Design and Implementation of Deep Learning Model for Stock Predictions using TensorFlow," *Processing of Korea Information Science Society Conference*, pp.799-801, June 2017.

[14] Y. Dai and Y. Zhang, "Machine Learning in Stock Price Trend Forecasting," Stanford University, 2013.

[15] J. W. Lee and J. M. O, "Artificial Intelligence: Integrated Multiple Simulation for Optimizing Performance of Stock Trading Systems based on Neural Networks," *KIPS Journal B (2001~2012)*, Vol.14B, No.2, pp.127-134, Feb. 2007.

[16] Z. Jiang, D. Xu, and J. Liang, "A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem," *Journal of Machine Learning Research*, arXiv:1706.10059v2, 2017.

[17] I.-S. Baek, "Local Hurst Exponent Indicator to Sell the Abruptly Rising Stock," *Korean Association of Financial Engineering*, Vol.9, No.3, pp.149-165, Sept. 2010.

[18] S. Cho and J.-S. Choi, "A Monte Carlo Experiment on the Power of Augmented Dickey-Fuller Unit Root Test," *Journal of The Korean Official Statistics*, Vol.10, No.1, 2005.

[19] D.-H. Kwon, J.-S. Heo, J.-B. Kim, H.-K. Lim, and Y.-H. Han, "Correlation Analysis and Regression Test on Cryptocurrency Price Data," *Proceedings of Spring Korea Information Processing Society Conference*, May 2018.

[20] P. Hall, B. U. Park, and R. J. Samworth, "Choice of Neighbor Order in Nearest-neighbor Classification," *Annals of Statistics*, Vol.36, No.5, pp.2135-2152, 2008.

[21] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, Vol.2, No.2, pp.121-167, Jun. 1998.

[22] Y. D. Kim, K. H. Kim, and S. H. Song, "Comparison of Boosting and SVM," *Journal of the Korean Data And Information Science Sociaty*, Vol.16, No.4, pp.999-1012, 2005.

[23] J. H. Jung and D. K. Min, "The study of foreign exchange trading revenue model using decision tree and gradient boosting," *Journal of the Korean Data And Information Science Society*, Vol.24, No.1, pp.161-170, 2013.

[24] Alexey Natekin and Alois Knoll "Gradient Boosting Machines, a Tutorial," *Front Neurorobot*, Vol.7, No.21, 2013.

[25] S. Kar, S. Saha, L. Khaidem, and S. R. Dey, "Predicting the

Direction of Stock Market Price Using Tree Based Classifiers," *Elsevier North American Journal of Economics and Finance*, Jul. 2018. (available online).

[26] B. Gorman, "A Kaggle Master Explains Gradient Boosting," [Internet], http://blog.kaggle.com/2017/01/23/a-kaggle-master-explains-gradient-boosting, Jan 2017.

[27] J. Bergstra and Y. Bengio. "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, Vol.13 pp.281-305, 2012.

[28] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Statistics and Computing*, Vol.21, No.2, pp.137-146, 2011.

[29] DASK: Scalable analytics in Python [Internet], https://dask.pydata.org/

허 주 성

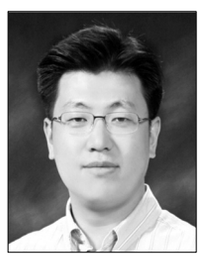
https://orcid.org/0000-0002-2486-9515
e-mail : chil1207@koreatech.ac.kr
2016년 한국기술교육대학교 컴퓨터공학부(학사)
2016년～현 재 한국기술교육대학교 컴퓨터공학부 석사수료
관심분야 : Machine Learning, Social Network Analysis

권 도 형

https://orcid.org/0000-0002-5951-2081
e-mail : dohk@koreatech.ac.kr
2017년 한국기술교육대학교 컴퓨터공학부(학사)
2017년～현 재 한국기술교육대학교 창의융합공학협동과정 ICT 융합 석사과정
관심분야 : Machine Learning, Stock Price Prediction

김 주 봉

https://orcid.org/0000-0002-8234-1030
e-mail : rlawnqhd@koreatech.ac.kr
2017년 한국기술교육대학교 컴퓨터공학부(학사)
2017년～현 재 한국기술교육대학교 컴퓨터공학부 석사과정
관심분야 : Reinforcement Learning, Stock Price Prediction



한 연 희

https://orcid.org/0000-0002-5835-7972
e-mail : yhhan@koreatech.ac.kr
1998년 고려대학교 컴퓨터학과(석사)
2002년 고려대학교 컴퓨터학과(박사)
2002년 삼성종합기술원 전문연구원
2013년～2014년 미국 SUNY at Albany, Department of Computer Science 방문교수
2006년～현 재 한국기술교육대학교 컴퓨터공학부 정교수
관심분야 : Mobility Management, Internet of Things, Machine Learning, Social Networks, Future Internet

안 채 헌

https://orcid.org/0000-0002-0989-5299
e-mail : ach@koreatech.ac.kr
2002년 한국기술교육대학교 기계공학과 (석사)
2011년 한국기술교육대학교 메카트로닉스공학부(박사)
2013년 한국생산기술연구원 연구원
2017년～현 재 한국기술교육대학교 메카트로닉스공학부 조교수
관심분야 : Mechanical Dynamics, Machine Learning Based Control System, Internet of Things, Smart Factory