

# LSTM 모델의 하이퍼 파라미터가 암호화폐 가격 예측에 미치는 영향 분석

박재현\*, 이동건\*, 서영석\*

\*영남대학교 컴퓨터공학과

kat3160@yu.ac.kr, dklee77@ynu.ac.kr, ysseo@yu.ac.kr

## Understanding the effect of LSTM hyperparameters tuning on Cryptocurrency Price Prediction

Jaehyun Park\*, Dong-Gun Lee\*, Yeong-Seok Seo\*

\*Dept. of Computer Engineering, Yeungnam University

### 요약

최근 암호화폐가 발전함에 따라 다양한 연구들이 진행되고 있지만 그 중에서도 암호화폐의 가격 예측 연구들이 활발히 진행되고 있다. 특히 이러한 예측 분야에서도 인공지능 기술을 접목시켜 암호화폐 가격의 예측 정확도를 높이려는 노력들이 지속되고 있다. 인공지능 기반의 기법들 중 시간적 정보를 가진 데이터를 기반으로 하고 있는 LSTM(Long Short-Term Memory) 모델이 다각도로 활용되고 있으나 급등락하는 암호화폐 가격 데이터가 많을 경우에는 그 성능이 상대적으로 낮아질 수밖에 없다. 따라서 본 논문에서는 가격이 급등락하고 있는 Bitcoin, Ethereum, Dash 암호화폐 데이터 환경에서 LSTM 모델의 예측 성능이 향상될 수 있는 세부 하이퍼 파라미터 값을 실험 및 분석하고, 그 결과의 의미에 대해 고찰한다. 이를 위해 LSTM 모델에서 향상된 예측률을 보일 수 있는 epoch, hidden layer 수, optimizer 에 대해 분석하였고, 최적의 예측 결과를 도출해 줄 수 있는 최소 training data 개수도 함께 살펴보았다.

### 1. 서론

Satoshi Nakamoto 가 Bitcoin 을 소개한 이후로 수많은 암호화폐가 등장했고, 관련 연구도 활발히 진행 중이다. 가격 예측 연구도 학습 데이터를 다각도에서 분석하거나, 시계열 데이터에서 높은 예측 성능을 보이는 RNN, LSTM 등의 모델을 사용해 예측 [1][2]하는 등 다양한 방법이 제안되었다.

이렇듯 예측 분야에서도 최근 기계학습이 도입되었으며, 특히 다양한 모델 중 시계열 데이터 예측에는 LSTM 이 자주 쓰인다 [1][2][3]. LSTM 은 RNN 의 느린 학습 속도와 state 간의 거리에 따른 Vanishing gradient problem 을 개선한 모델 [4][5]로서 시계열 데이터 예측에 자주 사용되고 있다.

그러나 LSTM 의 경우 RNN 에 비해 sigmoid 와 tanh 함수 연산이 많아 특정 데이터의 경우 예측 성능이 오히려 낮게 나타날 수 있다 [1][4]. 암호화폐 예측의 경우 가격의 급등락을 예측하는 것이 중요 요소인데 위에서 설명한 특정 데이터가 자주 발생해 LSTM 의 성능 저하가 나타날 수 있다.

따라서 본 논문에서는 이러한 경향이 나오지 않도록

최적화된 하이퍼 파라미터를 실험 및 분석한다. 하이퍼 파라미터는 간단하고 다양하게 그 값을 변경 시킬 수 있기 때문에 Bitcoin, Ethereum, Dash 암호화폐에서 가격이 급등락하는 데이터에 대해 LSTM 의 예측 성능을 향상시키기 위해 이를 변경하며,  $R^2$  와 Adjusted  $R^2$  값 [6]이 얼마나 차이 나는지 실험 및 분석한다.

초기 설정한 LSTM 모델에 대해 epoch, hidden layer 층 수, optimizer 와 같은 하이퍼 파라미터를 변경하면서 변경점을 각각 반영한 모델과 모든 변경점을 반영한 모델에 대해 실제 값이 얼마나 적합한지를 나타내는  $R^2$  와 Adjusted  $R^2$  값에서 얼마나 차이가 나는지 검증 및 원인을 실험 및 분석하였고, 최적의 예측 결과를 도출해 줄 수 있는 최소 training data 개수도 함께 살펴보았다. 이를 기반으로 가이드라인을 제시한다.

2 장에서는 초기 모델 설정과 이와 관련된 세부 정보를 설명한다. 다음 3 장에서는 제시된 방법으로 얼마나 가격 예측 성능이 향상되는지 보인다. 마지막으로 4 장에서는 결론과 향후연구를 기술한다.

## 2. 초기 모델 설정 및 세부 정보

Bitcoin과 같이 암호화폐 가격은 2020년 하반기를 기준으로 가격이 급등하고 있다. 따라서 본 논문에서는 이와 같은 경우를 가진 데이터에 초점을 맞췄다.

LSTM 모델에서는 1장의 <표 1>과 같이 가격이 큰 변동이 없다가 특정 시점에서 크게 올라가는 경우와 같이 급등하는 데이터에 대해서는 RNN에 비해 낮은 성능을 보인다 [1]. 현재 가장 널리 알려진 Bitcoin과 Ethereum, 익명성에서 장점이 있는 Dash 총 세 가지의 암호화폐에서 이러한 데이터를 찾아 모델을 개선한 뒤  $R^2$ 와 Adjusted  $R^2$  값을 측정했다. 본 실험에 대한 딥러닝 라이브러리는 pytorch를 사용했다.

예측 모델은 기존의 hybrid 예측 모델 연구 [1]과 동일한 하이퍼 파라미터를 사용했다. 아래 <표 1>은 LSTM 예측 모델의 초기 하이퍼 파라미터 설정 값이다. 입력 feature로는 Open(시가), High(고가), Low(저가), Price(가격) 네 가지를 사용한다. hidden layer의 노드 개수는 30개이며, hidden layer는 2층으로 구성하였다. optimizer는 Adam optimizer를, learning rate는 0.01로 하여 epoch 300번 만큼 학습한다. 최종적으로 다음 날의 가격을 예측한다. 이 모델을 init\_model로 정의한다.

<표 1> 예측 모델의 초기 하이퍼 파라미터 값

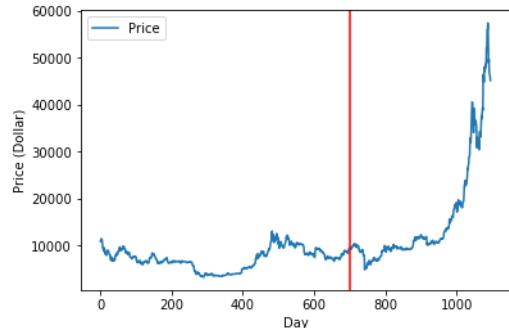
	Value
input_size	4
hidden_size	30
num_layers	2
optimizer	Adam
learning rate	0.01
epoch	300
train set 크기	700

Bitcoin은 '18.03.01 ~ '21.02.28 까지 3년 (1,096일)간의 데이터를 날짜 오름차순으로 정렬한 데이터를 사용한다. Ethereum은 '18.01.01 ~ '20.09.26 까지 1,000일간의 데이터를 날짜 내림차순으로 정렬한 데이터를 사용한다. Dash는 '18.01.01 ~ '20.12.31 까지 3년 (1,096일)간의 데이터를 날짜 내림차순으로 정렬한 데이터를 사용한다. 또한 급등하는 가격 데이터를 구성하기 위해 Ethereum과 Dash는 날짜 내림차순으로 정렬했다.

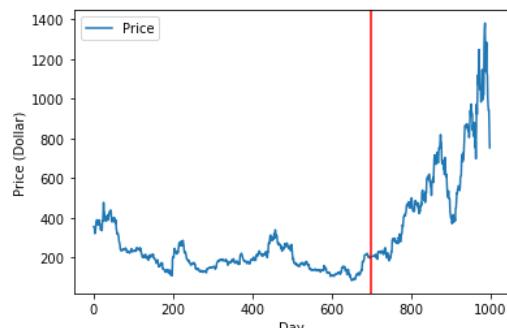
실험에 사용한 데이터 중 700개를 train data로 정했으며, Bitcoin과 Dash는 나머지 396개를, Ethereum은 나머지 300개를 test data로 정했다.

아래 (그림 1), (그림 2), (그림 3)은 각 암호화폐 종류 별 Price 그래프이다. x 좌표 700의 수직선을 기준으로 왼쪽이 train data, 오른쪽이 test data이다.

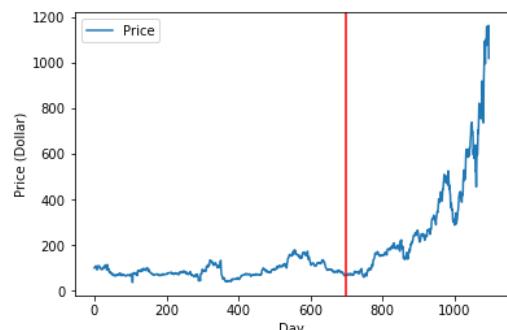
세 가지 암호화폐 모두 약 800일 전후로 가격이 급등하는 것을 볼 수 있다.



(그림 1) Bitcoin Price 그래프



(그림 2) Ethereum Price 그래프



(그림 3) Dash Price 그래프

평가 기준으로는  $R^2$ 와 Adjusted  $R^2$  값을 사용했다.  $R^2$  값은 제안한 모델이 주어진 자료에 얼마나 적합한지를 0~100% 사이로 측정하는 척도이다. 그러나  $R^2$  값은 input과 같은 독립변수의 종류가 많아질수록 함께 증가한다는 문제점이 있다. 따라서 이러한 문제점을 해결한 Adjusted  $R^2$  값도 같이 사용했다.

위에서 제안된 init\_model과 동일한 하이퍼 파라미터를 가진 RNN으로 예측한 값에 대한  $R^2$ 와 Adjusted  $R^2$  값은 아래 <표 2>와 같다.

<표 2> init\_model의 결과 (단위: %)

	Bitcoin	Ethereum	Dash
RNN $R^2$	94.71	93.13	91.71
RNN Adjusted $R^2$	94.66	93.06	91.63
LSTM $R^2$	68.90	68.89	32.59
LSTM Adjusted $R^2$	68.58	68.57	31.90

<표 2>의 결과처럼 RNN보다 예측 정확도가 떨어지는 상황을 개선하기 위해 다음 4 가지 LSTM의 하이퍼 파라미터를 변경시키며 실험 및 분석을 진행했다.

epoch의 경우 기준 300이 아닌, 400에서 100까지 10 단위로 줄여가면서 실험을 진행했다. hidden layer의 경우 이를 의미하는 하이퍼 파라미터인 num\_layer를 기준 2층이 아닌, 3층에서 1층까지 1층씩 줄여가면서 실험을 진행했다. 마지막으로 주어진 데이터에 대해 가격이 급등락하는 구간은 test data로 정의될 수 있다. Adam optimizer는  $L_2$  Norm을 사용하기 때문에 급등락하는 데이터에 민감하고 이에 적합하지 않아 예측을 못할 수 있다. 따라서 Adam optimizer 대신 pytorch 라이브러리에서 기본으로 지원하는 모든 optimizer를 하나씩 변경해가며 실험을 진행했다.

마지막으로 위의 세가지 변경을 모두 적용한 fin\_model에 대해 test data는 그대로 두고, train data를 10개씩 차례로 줄여가면서 며칠까지의 데이터를 학습시켰을 때 기준 연구 [3]를 기반으로 3장에서 정한 정확성의 기준을 넘어가는지를 확인해 새로운 코인이 나오더라도 특정 일자의 데이터가 모이는 순간 바로 예측을 할 수 있도록 최소한의 가이드라인을 제공한다.

### 3. 실험 결과

기준의  $R^2$  값 연구 [3]에서 최대 99.2%의  $R^2$  값을 달성했다. 그러나 모델 세부 설정에 따라 차이가 있기 때문에 최소 99.0%를 넘어야 예측 성능이 높은 것으로 정했다.  $R^2$  값이 99.0%라는 의미는 제시한 모델이 실제 값에 대해 99.0% 적합하다는 의미이다. 또한 같은 연구에서 0.1~0.2% 차이로 성능의 순위를 매겼다. 따라서 이 정도의 차이도 유의미한 차이로 판단한다.

#### 3.1. epoch 분석

아래 <표 3>는 init\_model에서 epoch를 300에서 200으로 줄였을 때 결과이다.

&lt;표 3&gt; epoch 차이에 따른 결과 (단위: %)

	Bitcoin	Ethereum	Dash
epoch = 300 $R^2$	68.90	68.89	32.59
epoch = 300 Adjusted $R^2$	68.69	68.59	31.90
epoch = 200 $R^2$	72.41	77.96	36.67
epoch = 200 Adjusted $R^2$	72.10	77.74	36.26

<표 3>의 마지막 두 행에서 볼 수 있듯이 epoch가 줄어들었을 때  $R^2$ 와 Adjusted  $R^2$  값이 약 4~9% 증가했다. 이는 epoch에 의한 과적합 (overfitting)이 사라졌기 때문으로 분석되었다.

#### 3.2. hidden layer 층 수 분석

아래 <표 4>는 init\_model에서 hidden layer 층 수를 2층에서 1층으로 줄였을 때 결과이다.

&lt;표 4&gt; hidden layer 층 수 차이에 따른 결과 (단위: %)

	Bitcoin	Ethereum	Dash
num_layer = 2 $R^2$	68.90	68.89	32.59
num_layer = 2 Adjusted $R^2$	68.69	68.59	31.90
num_layer = 1 $R^2$	98.59	98.01	98.21
num_layer = 1 Adjusted $R^2$	98.58	97.99	98.20

<표 4>의 마지막 두 행에서 볼 수 있듯이 hidden layer 층 수가 줄어들었을 때  $R^2$ 와 Adjusted  $R^2$  값이 약 30~66% 증가했다. LSTM은 RNN에 비해 state 간의 거리가 멀어도 예측이 잘 되지만 [3], sigmoid 와 tanh 함수 연산이 많아 [1][7] hidden layer 층이 늘어날수록 Gradient vanishing 현상이 일어나 예측을 못할 수 있다. 따라서 Gradient vanishing의 영향이 줄어들었기 때문으로 분석되었다.

#### 3.3. optimizer 분석

아래 <표 5>는 init\_model에서 optimizer를 Adam optimizer에서 Adamax optimizer로 바꿨을 때 결과이다.

&lt;표 5&gt; optimizer 차이에 따른 결과 (단위: %)

	Bitcoin	Ethereum	Dash
Adam $R^2$	68.90	68.89	32.59
Adam Adjusted $R^2$	68.69	68.59	31.90
Adamax $R^2$	75.82	85.54	55.87
Adamax Adjusted $R^2$	75.58	85.39	55.41

<표 5>의 마지막 두 행에서 볼 수 있듯이 Adamax optimizer를 사용한 경우  $R^2$ 와 Adjusted  $R^2$  값이 약 7~27% 증가했다. Adamax optimizer는 Adam optimizer와 달리  $L_2$  Norm 대신  $L_\infty$  Norm을 사용한다 [8]. 따라서 급등락하는 데이터의 영향이 줄어들었기 때문으로 분석되었다. 따라서 Adamax optimizer가 Adam optimizer에 비해 급등락하는 데이터에 대해서 더 좋은 성능을 낸다는 것을 확인했다.

#### 3.4. 최종 결과 분석

최종적으로 세 변경점을 모두 적용한 fin\_model의 결과는 아래 <표 6>과 같다.

<표 6>의 init\_model과 fin\_model의  $R^2$ 와 Adjusted  $R^2$  값을 보았을 때 그 차이가 매우 크며, fin\_model이 실제 test data에 약 99.2~99.4% 정도로 적합한 것을 알 수 있다.

&lt;표 6&gt; 최종 결과 (단위: %)

	Bitcoin	Ethereum	Dash
init_model $R^2$	68.90	68.89	32.59
init_model Adjusted $R^2$	68.69	68.59	31.90
fin_model $R^2$	99.45	99.30	99.43
fin_model Adjusted $R^2$	99.44	99.29	99.43

추가적으로 test data 의 개수에 따른 예측 정확도를 분석해보기 위해 train data 의 개수는 고정시킨 채 test data 의 개수를 변경시켜가며 몇 개가 되었을 때 특정  $R^2$  값을 넘어가는지 추이를 실험 및 분석해 보았다. 그 결과는 아래 <표 7>과 같다.

&lt;표 7&gt; train data 수에 따른 결과 (단위: 개)

	Bitcoin	Ethereum	Dash
$R^2 \geq 80\%$	270	190	240
$R^2 \geq 90\%$	290	210	280
$R^2 \geq 99\%$	340	500	360

Train data 는 <표 7>에 적인 개수만큼 사용했으며, test data 는 2 장에서 언급한 것과 하게 사용했다. Ethereum 에서 99% 이상의  $R^2$  값을 얻기 위해서는 train data 가 500 개 (500 일 치) 필요하며, 이는 99%를 넘어갈 때 세 종류의 암호화폐에 data 수에 대해서 최댓값이다. 따라서 99%의  $R^2$  값을 얻기 위해서는 적어도 500 개이상의 데이터가 필요하는 것을 알 수 있다.

암호화폐 예측은 다양한 하이퍼 파라미터에 영향을 받는다. LSTM 모델에서 epoch 는 너무 많으면 과적합(overfitting)을, 너무 적으면 학습이 제대로 되지 않기 때문에 가장 좋은 성능을 보이는 epoch 수를 찾아야 한다. 두번째로 LSTM 은 RNN 에 비해 sigmoid 와 tanh 함수 연산의 수가 많아 [1][7] hidden layer 층이 많으면 Gradient vanishing 현상의 영향을 크게 받게 된다 [1]. 마지막으로 급등락하는 데이터가 있는 경우 Adam optimizer 에 비해 Adamax otpimizer 가 더 좋은 성능을 낼 수 있다.

#### 4. 결론 및 향후연구

본 논문에서는 LSTM 으로 가격을 예측했을 때 예측 성능이 낮은 데이터에 초점을 맞추었다. 초기 LSTM 모델에 대해 epoch 와 hidden layer 층 수를 줄이고, optimizer 를 급등락하는 데이터에 덜 민감한 Adamax 로 변경하였으며, training data 의 개수를 줄여보면서 실험을 수행하였다. 또한 언급한 하이퍼 파라미터를 각각 반영한 모델에 대해 실제 값이 얼마나 적합한지를 나타내는  $R^2$  와 Adjusted  $R^2$  값에서 얼마나 차이가 나는지 검증 및 원인을 실험 및 분석했다.

본 실험에서는 epoch 는 200, hidden layer 는 1 층, optimizer 는 Adamax optimizer 를 썼을 때 <표 6>의 마지막 두 행에서 볼 수 있듯이 성능이 가장 좋은 결과를 보여주었다. 또한 test data 개수에 따른  $R^2$  값의 차이에서도 500 개 정도의 데이터를 수집할 수 있다면 새로운 코인이 등장해도 보다 정확한 예측을 할 수 있을 것으로 기대된다.

이번 실험에서는 LSTM 으로만 실험을 수행했지만,

LSTM 뿐만 아니라 다양한 시계열 모델들이 있고, 서로를 결합한 hybrid 모델 [1][2]도 있다. 향후에는 이와 같은 모델에 대해서도 실험을 해볼 예정이다.

#### Acknowledgement

이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2020R1I1A3073313).

#### 참고문헌

- [1] M. Patel and S. Tanwar, "A deep learning-based cryptocurrency price prediction scheme for financial institutions", Journal of Information Security and Applications, vol.55, pp. 1-13, 2020
- [2] MA. Hossain and R. Karim, "Hybrid deep learning model for stock price prediction", IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, pp. 1837-1844
- [3] T. Phaladisailoed and T. Numnonda, "Machine Learning Models Comparison for Bitcoin Price Prediction", 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), Bali, Indonesia, 2018, pp. 506-511
- [4] S. Hochreiter and J. Schmidhuber, "Long short-term memory", Neural computation, vol.9, no.8, pp. 1735-1780, 1997
- [5] D. Wei and B. Wang, "Research on unstructured text data mining and fault classification based on RNN-LSTM with malfunction inspection report", Energies, vol.10, no.3, pp. 1-406, 2017
- [6] O. Kazuhiro, "Bootstrapping R2 and adjusted R2 in regression analysis", Economic Modeling, vol.17, no.4, pp. 473-483, 2000
- [7] J. Feng and S. Lu, "Performance Analysis of Various Activation Functions in Artificial Neural Networks", Journal of Physics: Conference Series, vol.1237, no.2, pp. 1-6, 2019
- [8] DP. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", 2015 3rd International Conference on Learning Representations (ICLR), San Diego, USA, 2015, pp. 1-15