

비트코인 가격 예측

컴퓨터공학과 2012104102 오영택

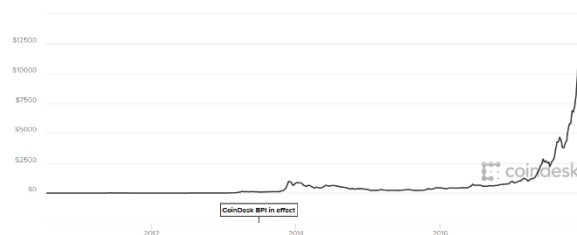
ABSTRACT

현재 기존의 통화 화폐와는 성격이 다른 온라인 가상 화폐인 비트코인은 가격 변동이 매우 심하여 예측하기가 어렵습니다. 이 보고서에서는 비트코인의 가격을 예측하기 위한 RNN모형을 소개합니다.

1. INTRODUCTION

비트코인(bitcoin)은 블록체인 기술을 기반으로 만들어진 온라인 가상화폐입니다. 비트코인의 화폐 단위는 BTC로 표시합니다. 2008년 10월 사토시 나카모토^[1]라는 가명을 쓰는 프로그래머가 개발하여, 2009년 1월 프로그램 소스를 배포했습니다. 중앙은행이 없이 전 세계적 범위에서 P2P 방식으로 개인들 간에 자유롭게 송금 등의 금융거래를 할 수 있게 설계되어 있습니다.

2009년 10월 5일 1 비트코인(BTC)에 0.000007563\$에 거래되었고, 꾸준히 가치가 상승하여 2017년 12월 6일 현재, 1BTC에 약 13000\$(16,000,000₩)에 거래되고 있습니다(<그림-1>). 또 <그림-2>에서 볼 수 있듯이 비트코인의 가격은 변동의 폭이 매우 심합니다. 이성적인 투자의 장이 아닌 투기장의 성격이 강하여 사람이 기존 데이터를 분석하여 대략적인 가격마저 예측할 수 없는 상황입니다. 현재까지의 시장의 종가 데이터를 바탕으로 기계학습을 시킨다면 대략적으로 가격을 예측할 수 있지않을까라는 생각으로 프로젝트를 시작하였습니다. 또한 시간 연속적인 성격을 가지고 있어 Recurrent Neural Network를 사용하였습니다.



<그림-1 : 2010년 4월~2017년 12월 6일 비트코인 가격 그래프(단위: \$)>



<그림-2: 2014년 12월 4일 최고가 경신 후 폭락, 2017년 4월부터 현재까지 급격히 상승 중>

2. MODEL

비트코인 가격 예측 모델을 만들기 위해 사용한 데이터셋은 2013년 4월부터 2017년 8월까지의 <날짜, 시작가, 당일 최고가, 당일 최저가, 종가(\$), 거래량(BTC)>으로 이루어진 데이터입니다.

이 데이터는 시간 연속적인 성격을 가지고 있습니다. 예를 들어 2017년 12월 6일의 비트코인 가격은 5일,4일,3일의 데이터와 밀접한 연관이 있습니다. 연속적인 데이터(sequence data)를 학습시키는데 RNN이 적합하다고 판단하여 RNN을 사용하였습니다.

2.1 data preprocessings

1) 시간순으로 데이터를 재정렬

시간 순으로 학습을 시키기 위해 정렬을 했습니다.

2) Missing value

거래량(\$)이 매우 적은 경우에 속해 없는 것이므로 Value가 없으면 해당 row는 제거했습니다.

3) MinMax Scaling을 통하여 값을 0~1범위로 변환

$$X = (X - \text{Min}) / (\text{Max} - \text{Min} + 0.000000001)$$

Max와 Min값이 같은 경우 0으로 나누게 되므로 이 경우를 피하고자 매우 작은 값을 더해 나누었습니다.

4) 학습 데이터와 예측할 데이터를 분할

Sequence length만큼 과거의 데이터를 기반으로 학습 시켜야하기때문에, i일 비트코인 종가를 결정하는 인자는 다음의 표와 같습니다.

날짜	시작가	최고값	최저값	거래량	종가
i-1	value	value	value	value	Value
i-2	value	value	value	value	Value
....
i-length	value	value	value	value	Value

또한 전체의 데이터중 80%만을 학습시키는데 사용하고, 나머지 20%의 데이터로 비트코인 가격을 예측하였습니다.

2.2 modeling

1) RNN

위에서 말했듯이, 비트코인 가격을 예측하기위해서는 연속적인 데이터를 학습 시켜야 한다고 판단을 했기 때문에 RNN을 사용하였습니다.

2) drop out wrapper

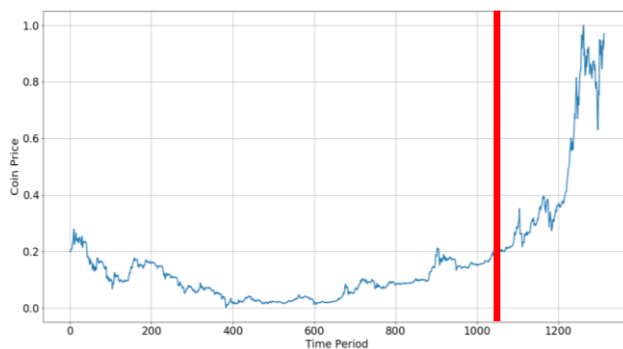
Over fitting을 방지하고 training time을 줄이기 위해 0.7 비율의 dropout을 사용했습니다.

3) Multi-RNN

보다 deep하게 NN을 구성하기 위해 MultiRNN을 사용하였습니다.

3. RESULTS AND ANALYSIS

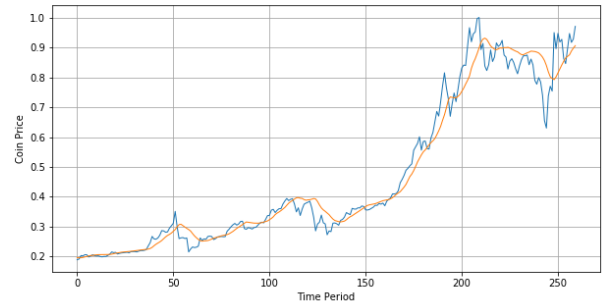
다음의 <그림-3>은 정규화된 전체 비트코인의 종가를 나타낸 그림입니다.



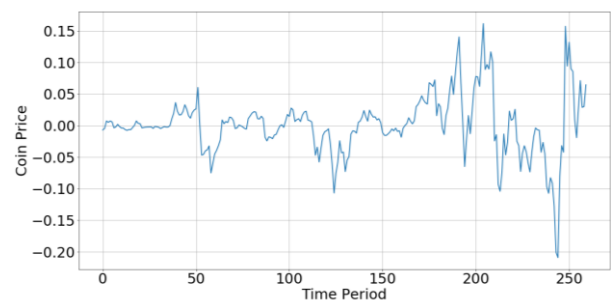
<그림-3> 2013년 4월부터 2017년 8월까지의 비트코인 종가, 빨간색 선을 기점으로 이전 데이터

터를 학습시키고, 이후의 데이터로 예측을 하였음.>

첫번째 실험에서 사용한 모델의 상세사항은 다음과 같습니다. 단일 RNN cell을 이용하고, hidden dimension은 100, sequence length는 15, learning rate는 0.0001, iteration은 500입니다. 이에 대한 결과는 <그림-4>와 <그림-5>에 있습니다.



<그림-4> 첫번째 실험, 파랑은 실제 비트코인의 가격이고 주황은 예측한 비트코인의 가격

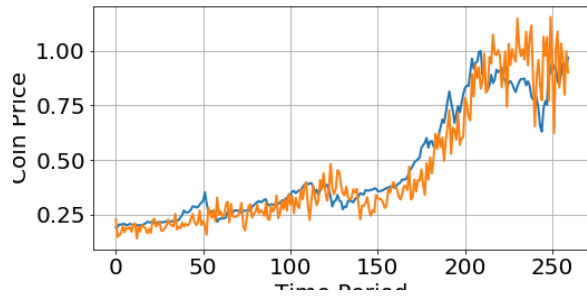


<그림-5> 첫번째 실험, 실제 가격에서 예측 가격을 뺀 오차

첫번째 실험의 결과는 놀라웠습니다. <그림-3>에서 볼 수 있듯이, 학습을 시킨 데이터와는 확연하게 다르게 비트코인의 가격이 가파르게 상승하고, 하락하는 등 불규칙한 변동을 보였음에도 첫번째 모델은 어느 정도의 트렌드를 예측할 수 있는 것처럼 보였습니다. 더 정확하게 분석해 보기 위해 실제 가격과 예측 가격의 차이, 오차를 구했습니다(그림-5). 가격 트렌드는 어느정도 맞추는 것처럼 보였지만, 급격히 상승하거나 급격히 하락하는 구간에서 오차가 최대 0.24까지 난 것을 볼 수 있습니다.

두번째 실험에서는 NN을 더 deep하게 만들기 위해 5개의 layer로 이루어진 Multi RNN cell을 이용했고 dropout을 0.7 주었다는 것을 제외하고는 동일한 조건을 주었습니다. 실험의 결과는 <그림-6>, <그림-7>에 있습니다. 이번에도 어느정도의 트렌드를 예측할 수 있는 것처럼 보였으나(그림-6), 실제 가격과 예측 가격의 차이(그림-7)을 보면 가파르게 상승하는 구간(time period > 150)에서의 차이는 첫번째 실험과 비슷합니다. 또 그 전 구간(time period

<150>에서는 오히려 첫번째 실험에 비해 오차가 커진 것을 볼 수 있습니다.

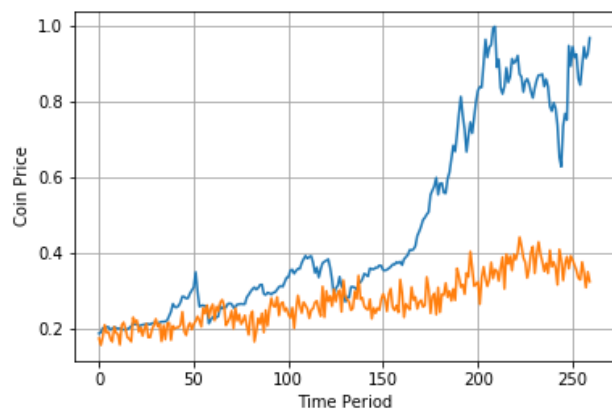


<그림-6: 두번째 실험, 파랑은 실제 비트코인의 가격, 주황은 예측한 비트코인의 가격>

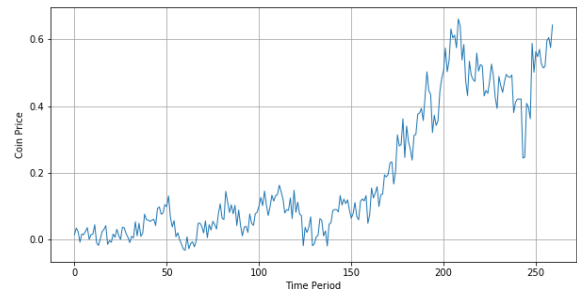


<그림-7: 두번째 실험, 실제 가격에서 예측 가격을 뺀 오차>

세번째 실험에서는 두번째 실험보다 더 deep하게 만들기 위해 10개의 layer로 이루어진 Multi RNN cell을 이용한 것을 제외하고는 동일한 조건에서 진행하였습니다. 실험의 결과는 <그림-8>,<그림-9>에 있습니다.



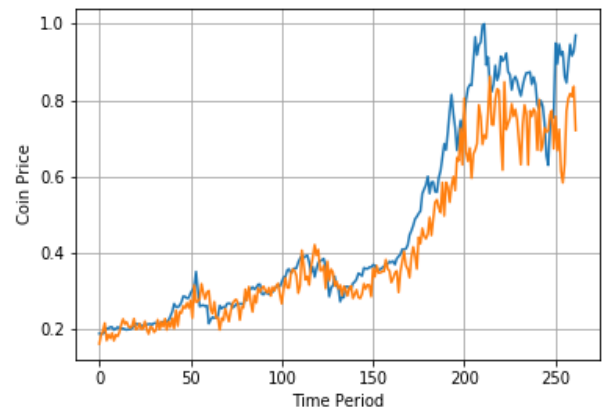
<그림-8: 세번째 실험, 파랑은 실제 비트코인의 가격, 주황은 예측한 비트코인의 가격>



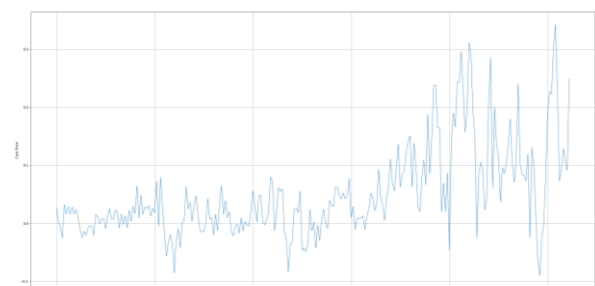
<그림-9: 세번째 실험, 실제 가격에서 예측 가격을 뺀 오차>

이 경우, 트렌드 예측에 아예 실패했습니다. 급격히 상승하는 구간에서 최대 0.65의 오차가 발생했습니다. 지금까지의 실험에서는 sequence length를 15로 설정하였는데, 이는 급격히 변하는 비트코인의 가격을 예측하는 것에 적합하지 않다고 판단을 했습니다.

그래서 sequence length를 7로 설정하고 **네번째 실험**을 하였습니다. 실험의 결과는 <그림-10>,<그림-11>에 있습니다.



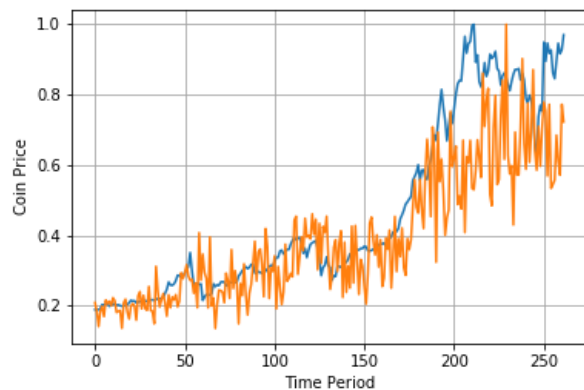
<그림-10: 네번째 실험, 파랑은 실제 비트코인의 가격, 주황은 예측한 비트코인의 가격>



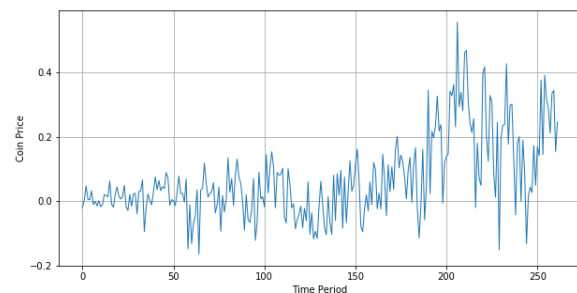
<그림-11: 네번째 실험, 실제 가격에서 예측 가격을 뺀 오차>

세번째 실험에 비해, 확연하게 나은 결과가 나온 것을 확인할 수 있습니다. 세번째 실험에서 최대 0.65의 오차가 발생한 것에 비해(그림-9) 최대 오차 0.33으로 오차가 줄어든 것을 볼 수 있습니다.

다섯번째 실험에서는, 네번째 실험에서 학습 반복 횟수가 부족하여 학습이 충분히 되지않은지를 확인하기 위해 학습 반복 횟수를 1000(네번째 실험의 2배)로 설정하고 실험하였습니다.



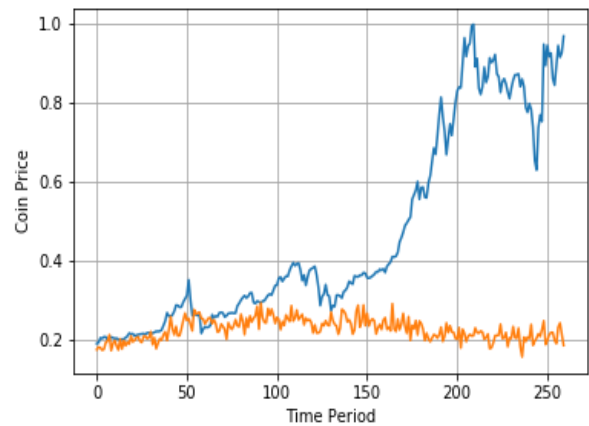
<그림-12: 다섯번째 실험, 파랑은 실제 비트코인의 가격, 주황은 예측한 비트코인의 가격>



<그림-13: 다섯번째 실험, 실제 가격에서 예측 가격을 뺀 오차>

이 경우, <그림-12>에서 볼 수 있듯이 over-fitting이 된 것으로 보입니다.

마지막 실험에서는 기존의 실험과는 달리, hidden dimension은 200으로 2배 늘리고, learning rate는 0.00001로 1/10로 줄였고, 학습 횟수는 10000으로 대폭 증가시켰고, 20개의 layer로 이루어진 Multi RNN cell을 이용했습니다. Learning rate를 줄이고, 보다 모델을 deep하고 wide하게 만든 후, 학습 횟수를 증가 시키면 더 좋은 결과가 나올 것이라고 예상했습니다. 결과는 <그림-14>에 있습니다.



<그림-14: 마지막 실험, 파랑은 실제 비트코인의 가격, 주황은 예측한 비트코인의 가격>

이 경우, 학습시간이 12시간이 넘게 걸렸고, loss, RMSE가 모든 실험 통틀어 가장 적게 나왔음에도 불구하고 트렌드 예측에 완벽하게 실패하였습니다. 학습 구간의 데이터에 over fitting이 되어 예측 구간의 급격한 상승, 하강을 예측 예측하지 못한 것으로 보입니다.

4. CONCLUSION

위의 실험들을 종합해볼 때, NN을 deep하고 wide하게 만들수록 예측 결과가 안좋게 나옵니다. 이는 NN을 deep하고 wide하게 만드는 것이 잘못되었다는 것이 아니라, 비트코인 가격을 현재 사용한 데이터 셋을 사용하여 예측한다는 것은 불가능하다는 것을 의미합니다. 학습 데이터의 트렌드를 정교하게 학습할수록 예측 구간의 급격한 변화를 예측할 수가 없습니다. <그림-3>을 보면 비트코인 가격이 학습 구간(빨간선 이전)에서의 변화보다 예측 구간(빨간선 이후)에서의 변화가 크고 더 자주 일어난다는 것을 볼 수 있습니다.

그러나 <그림-4>, <그림-6>에서 볼 수 있듯이 정확한 가격 예측은 힘들지만 어느정도 대략적인 가격 형성 방향은 참고할 수 있습니다.

현재 비트코인 거래는 투자의 성격보다는 투기의 성격을 가지고 있습니다. 이는 비이성적이고 비상식적인 가격의 등락이 있을 수 있다는 것입니다. 2009년 첫 거래를 시작으로 2017년 현재까지 비트코인 거래는 8년차에 접어들었지만, 이전과는 다르게 2017년 3월이후의 가격 변화폭이 기하급수적으로 커져서 현 시점에서 트렌드를 예측한다는 것은 불가능하고 다만 대략적인 가격 형성 방향은 참고할 수 있다고 결론을 내렸습니다.