<Transformer 기반 주가 예측>

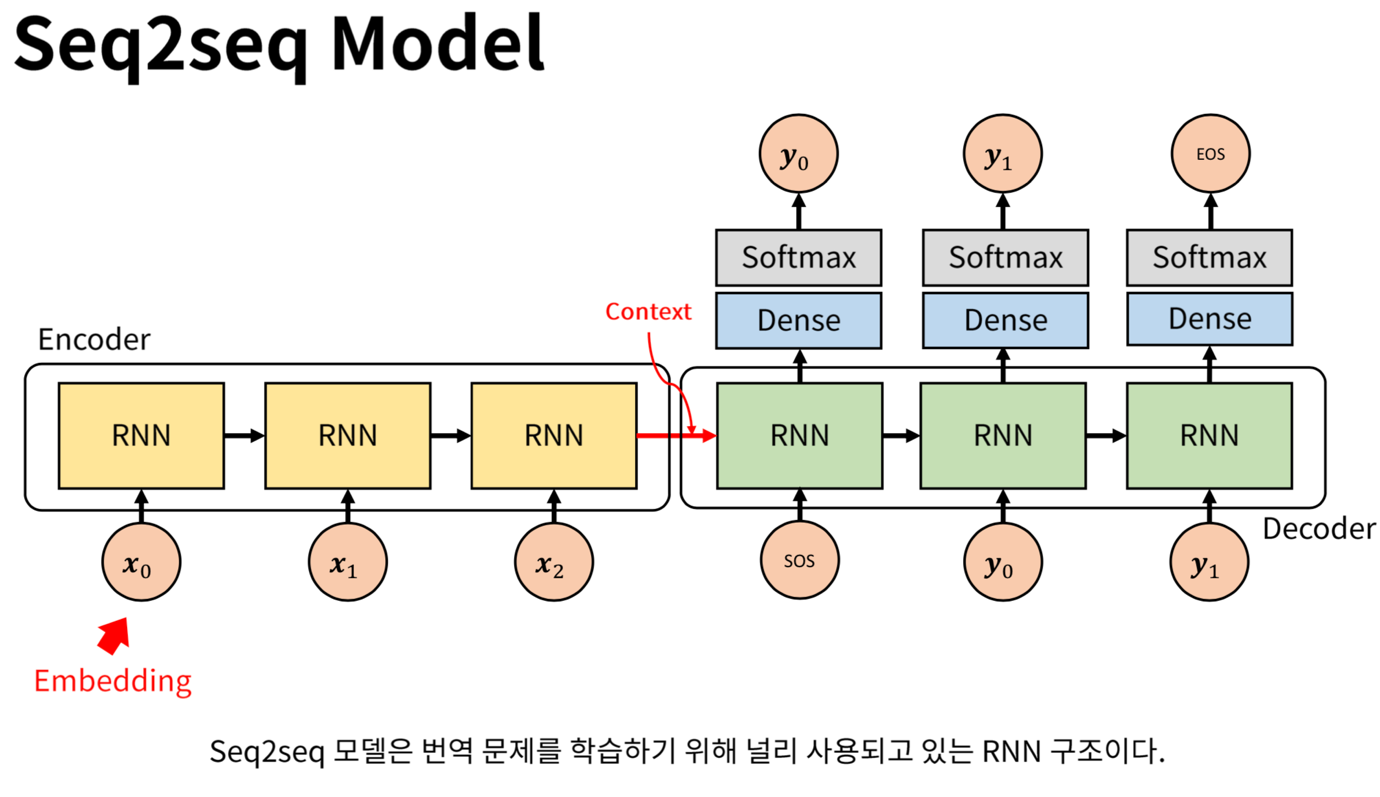
1. Transformer

- Transformer : Attention is all you need)

"Attention is All You Need"는 2017년 Google Brain의 연구팀에서 발표한 논문으로, 기존의 RNN이나 CNN과 같은 네트워크 아키텍처에서 사용되는 장기 의존성 문제를 해결하기 위해 제안된 Transformer 모델을 소개한다. 논문에서는 기존의 시퀀스 모델링에 사용되는 RNN이나 CNN은 장기 의존성을 처리하는데 한계가 있으며, 병렬 처리가 어렵다는 문제점이 있음을 지적하고 있습니다. 이에 대한 대안으로, Transformer는 self-attention 메커니즘을 사용하여 장기 의존성 문제를 해결하고, 입력 시퀀스의 각 위치가 서로 독립적으로 처리될 수 있도록 구성된다. 논문에서는 Transformer 모델의 구조와 self-attention 메커니즘에 대해 상세하게 설명하고, 이를 기반으로 하여 기계 번역 및 언어 모델링 작업에서 다른 모델들과 비교한 결과를 제시하고 있다. 실험 결과, Transformer 모델은 RNN, CNN과 같은 모델과 비교하여 좋은 성능을 보이며, 병렬 처리가 가능하므로 훈련 속도가 빠르다는 장점이 있다. 또한, 논문에서는 Transformer 모델을 적용할 수 있는 다양한 분야들에 대해 언급하고 있다. 자연어 처리 분야에서의 번역, 생성 모델링, 문서 분류 및 질문 응답 등 다양한 분야에서 Transformer 모델의 활용 가능성을 제시하고 있다. 최근에는 Transformer 모델이 시계열 데이터 분석 분야에도 적용되어 많은 연구가 진행되고 있다. 따라서, "Attention is All You Need"는 딥러닝 분야에서 매우 중요한 논문으로 평가되며, 자연어 처리부터 다른 분야의 문제에 대한 해결책으로 제안된 Transformer 모델은 현재까지도 많은 연구자들에게 큰 영향력을 끼치고 있다.

- 기존 Seq2seq의 한계점과 어텐션 메커니즘

기존의 seq2seq 모델은 context vector에 문장의 정보를 압축하는 과정에서 병목(bottle neck)현상이 발생해 성능 하락의 원인이 된다. 기계 번역을 예로 들면, 각각의 단어들로 구성된 하나의 문장(sequence)가 입력으로 들어왔을 때, 이를 고정된 크기의 context vector로 바꾼 뒤에 만들어진 context vector로부터 출력 문장을 만들어 낸다. 즉, 단어들이 입력될 때마다 이전까지 입력되었던 단어들의 정보를 포함하고 있는 hidden state 값을 받아서 hidden state 값을 새롭게 갱신하게 된다. 따라서, 마지막 단어가 입력되었을 때, 마지막 hidden state 값은 문장 전체를 대표하는 context vector로 활용할 수 있다. 그러므로 context vector는 앞의 문장에 대한 문맥적인 정보를 담고 있다. 하지만, 입력 문장에 대해서 모든 문장의 정보를 하나의 고정된 벡터에 포함시켜야 하기 때문에 병목현상이 발생하게 되고, 결과적으로 성능 하락의 원인이 된다. 이를 개선하기 위해 매번 인코더(encoder)의 모든 출력(hidden state)을 디코더(decoder)에서 참조하는 방법을 제시하는데, 이것을 어텐션 매커니즘(attention mechanism)이라고 한다. 따라서, 디코더는 매번 이전 출력과 비교해 인코더의 모든 출력 중에서 어떤 정보가 중요한지를 나타내는 energy 값을 계산한다. 이후 energy 값에 softmax를 적용해 가중치 값을 인코더의 hidden state 값에 적용해 다음에 올 단어를 예측하게 된다. 이렇게 어텐션 메커니즘을 활용하면 성능 개선에 효과적일 뿐만 아니라, 각 출력이 입력의 어떤 정보를 참고했는지 또한 알 수 있다.

<Seq2seq 모델 구조>

- Transformer 모델의 구조

Transformer 구조는 self-attention 구조를 활용해 입력 단어들 간의 상관 관계를 보다 잘 표현함으로써 BERT, GPT 등 자연어 처리 분야에서 SOTA(State Of The Art) 모델을 만드는 데에 초석이 되었다. BERT 계열의 모델은 Transformer의 인코더 부분을 활용하고, GPT 계열의 모델의 경우 Transformer의 디코더 부분을 활용해 주로 텍스트를 생성하는 모델로 활용하고 있다. 기존에 RNN, LSTM 등을 활용해 과거의 패턴을 분석해 주가를 예측하고자 하는 시도는 많았다. 따라서, Transformer를 활용한 시계열 예측이나 주가 예측이 가능할 것이라고 생각했고, 이번 프로젝트에서는 Transformer 구조를 주가를 예측하는 모형으로 활용해 Transformer가 주가를 예측하는 데에 어떻게 활용될 수 있는지 알아보고자 한다. 특히, Transformer의 self-attention 구조와 인코더 구조를 활용하고자 한다.

Transformer의 경우, RNN이나 CNN을 전혀 사용하지 않고, 인코더와 디코더를 여러 개 사용하며, 문장 내 각 단어들의 순서에 대한 정보를 주기 위해서 positional encoding(위치(순서) 정보를 포함하고 있는 embedding)을 사용한다. 이후 attention을 수행하게 되는데, 이때, 인코더에서 사용하는 어텐션을 self-attention이라고 한다. 이는 각각의 단어가 어떤 연관성을 가지고 있는지 구하기 위해 사용된다. 즉, 어텐션은 전반적인 입력 문장에 대한 문맥에 대한 정보를 잘 학습하도록 만드는 것이다. Transformer는 인코더 층을 반복해 가장 마지막의 인코더 출력값을 각 디코더 층의 입력값으로 전달한다. 그 이유는 위에서 설명한 바와 같이 디코더 층에서는 입력된 문장 중에서 어떤 단어에게 가장 많은 초점을 두어야 하는지 알려주는 것이다. Transformer의 디코더 또한 여러 개의 디코더 층으로 구성된다. 따라서 마지막 층의 출력값이 실제로 번역을 수행한 결과가 나오게 되는 것이다.

도표이(가) 표시된 사진

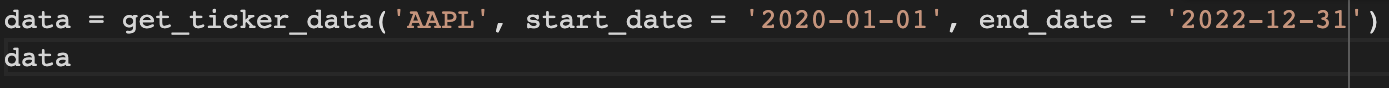
자동 생성된 설명

<Transformer의 구조>

2. Dataset 구성

- Yahoofinancial

Yahoofinancial은 Python 라이브러리로, Yahoo finance에서 제공하는 API를 활용해 시계열 주가 데이터를 쉽게 불러올 수 있다. Ticker라는 종목명과 원하는 날짜를 입력하면, 해당 기간의 주가 데이터를 가져올 수 있다. 다음의 예시는 AAPL이라는 ticker를 가진 종목(애플)의 주가를 2020년 1월 1일부터 2022년 12월 31일까지 모두 가져올 수 있는 코드이다.



<예시>

데이터의 각 열은 date, high, low, open, close, volume, adjclose로 구성되어 있다.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Date | High | Low | Open | Close | Volume | Adjclose |
| 날짜 | 고가 | 저가 | 시가 | 종가 | 거래량 | 수정 주가 |

<데이터 column 별 의미>

위의 데이터 중에서 종가를 기준으로 주가를 예측하고자 한다.

3. 기대효과

최근 Transformer를 활용한 시계열 예측 연구가 많이 진행되고 있다. 시계열 데이터를 처리할 때 Transformer의 encoder 구조와 self-attention 기법을 사용하면 시간 정보를 고려하면서 데이터를 처리할 수 있기 때문에 활용하여 주가 예측 모델을 구축할 수 있다. 결과적으로 Transformer 모델을 활용하면 이전 일정 기간 동안의 주가 데이터를 입력으로 사용하고, 해당 일자의 주가를 출력으로 예측하는 것이 가능하다. 이때, Transformer의 encoder 구조와 self-attention 기법을 사용하면 다음과 같은 기대 효과를 얻을 수 있다.

1) 장기간의 패턴 파악

Transformer의 encoder 구조와 self-attention 기법을 사용하면 장기간에 걸쳐서 발생하는 패턴을 파악할 수 있다. 이를 통해 주가 변동의 추세나 패턴을 더욱 정확하게 파악할 수 있다.

2) 다양한 입력 데이터 활용

Transformer의 encoder 구조와 self-attention 기법은 입력 데이터의 길이나 구조에 구애받지 않는다. 이를 활용하여 주가 예측에 필요한 여러 가지 데이터를 입력으로 사용할 수 있다. 예를 들어, 주가 예측 모델에 이전 주식 거래량, 외부적인 경제 지표 등 다양한 데이터를 입력으로 사용할 수 있다.

3) 예측 결과 해석의 용이성

위에서 설명한 바와 같이 Transformer의 self-attention 기법은 모델이 예측에 사용한 입력 데이터 중에서 어떤 부분을 더 중요하게 생각하는지를 분석할 수 있다. 이를 활용하여 모델의 예측 결과를 해석하는 것이 가능하다.

따라서, Transformer의 encoder 구조와 self-attention 기법을 활용하여 주가 예측 모델을 구축하면 보다 정확하고 유용한 예측 모델을 구축할 수 있다.