

ConvLSTM

- 2차원 이미지 데이터를 시계열로 모델링하여 예측하는 모델 (강우량 예측 모델)
- 바로 LSTM 내부 연산을 컨볼루션으로 대체한다.
- 이는 Fully-Connected Layer가 Convolutional Layer으로 대체되었을 때의 효과와 같으며, 전체적으로 모델의 웨이트 수를 크게 줄일 수 있습니다.

RNN

- 히든 레이어가 순환구조를 가지고 있어서, 직전 시점의 히든 state가 현재 state에 영향을 주는 구조
- parameter는 $W_{\{xh\}}$, $W_{\{hh\}}$, $W_{\{hy\}}$ 가 존재

Vanishing gradient

- **relu** 사용시 1보다 큰 값이 반복적으로 곱해진다.
- **sigmoid**나 **tanh** 사용시 역전파시 gradient가 점차 줄어드는 문제 발생

LSTM의 기본 구조

- cell state

가장 눈에 띄는 부분은 더하기(+)기호입니다. 일반적인 RNN에서는 곱하기 연산으로만 이루어져 있었는데 LSTM에서는 피드백을 더하기(+)로 이음으로써 Vanishing Gradient와 같은 문제를 해결할 수 있습니다.

- forget gate

LSTM의 첫번째 단계는 정보를 얼마나 잊을지에 관한 단계입니다. 이전상태의 hidden state(h)와 현재 상태의 input(x)이 시그모이드 함수를 거치면 0~1의 값이 나오게 됩니다.

- input gate

그 다음 Gate에서는 현재 상태의 input(x)를 얼마나 기억할 것인지에 대한것을 계산합니다.

- output gate

cell state의 계산은 끝났으니 이제 다음 상태로 보낼 (output)hidden state를 구해야합니다.