

1. 서론

해수면 온도 예측이 주제인만큼 해수면 온도 상승의 파급력에 대해 설명한다. 해수면 온도는 해양-대기 순환 현상을 결정하는 중요한 변수이며, 다양한 요인들로 한반도 해역의 고수온 현상이 자주 발생하고 있다. 이에 따라 다양한 방법으로 해수면 온도 예측이 이루어지고 있는데, 대표적으로 수치모델과 자료중심(Data-driven) 모델을 사용한다. 수치모델에는 ROMS, NEMO와 같이 일사량, 해류 등의 입력자료로 해수면 온도를 예측하는 모델이 있고, 자료 중심 방법인 통계 및 기계학습 기법에는 Markov model, SVM을 활용한 모델 등 다양한데, Fully connected LSTM (FC-LSTM) 모델이 상대적으로 우수한 예측성능을 가진다. 추가적으로 시/공간 학습을 동시에 하기 위해 합성곱을 활용한 Convolutional LSTM (ConvLSTM) 하기도 한다. 이때, 기존 LSTM, SVM을 활용한 모델에 비해 ConvLSTM 모델이 RMSE 측면에서 더 우수한 성능을 보였다.

한편, 한반도 남해 연안은 최근 특정 기간동안 고수온 현상이 오랫동안 지속되는 등 양식업적으로 큰 피해가 발생하고 있다. 이에, 해당 부근의 양식장 등에 해황에 대한 지속적 모니터링을 바탕으로 실시간 속보를 제공하는데, 몇가지 기준에 따라 고수온 주의보, 고수온 경보 등이 발령된다. 이때, 현장 관측 센서로부터의 관측값이 갖는 국지적 한계를 극복하기 위해 위성 자료를 활용하기도 한다. 실제로 많은 연구에서 위성 자료로부터 얻은 합성 해수면 온도자료로 이상치 산출 및 탐지를 진행했으며, 현장 관측자료와 해당 결과가 유사함을 보이기도 했다. 이 연구에서는 LSTM, ConvLSTM 딥러닝 기법을 활용해 한반도 남해 해역의 해수면 온도를 중단기로 예측하고 고수온 사례를 분석한다.

2. 연구 지역 및 연구 자료

1) 연구 지역

위도 34.475°E~33.025°N, 경도 126.972°E~128.925°E로, 해당 지역은 대마난류(Tsushima Warm Current)이 흐르는 지역이다.

2) 해수면 온도 합성 자료

인공위성으로 측정한 해수면 온도는 열적외, 마이크로파 기반의 위성 자료를 기반으로 산출하는데, 이는 여러가지 이유로 자료의 소실이 발생한다. 이를 해결하기 위해 자료 동화 방식을 사용한다. 이 연구에서는 유럽 기상청에서 제공하는 OSTIA 자료를 입력 데이터로 사용하며, 이 데이터의 공간 해상도는 0.05° , 시간 해상도는 1일이다. (2007년 1월 1일 ~ 2019년 12월 31일까지 13년의 합성자료 사용)

3. 연구방법

1) 연구방법

이 연구에서는 LSTM, ConvLSTM 모델을 10년 자료로 학습하고 이후 3년간의 자료로 검증하도록 하여 일일 기준 해수면 온도를 1-7일 예측하는 모델과 고수온 사례 예측 모델을 구축했다. 특히, 거문도 지역의 픽셀값 누락에 따라 Cubic interpolation을 사용해 보간하였으며, 시공간 차이로 인한 성능 저하를 방지하기 위해 범위를 0~1로 조정하는 정규화를 진행했다. 이외의 파라미터 등은 선행 연구의 값을 그대로 따랐다. 이때, 고수온 주의보 및 경보 기준선에 따라 픽셀 값을 1, 2로 범주화해 공간적 분포를 확인할 수 있게 했다.

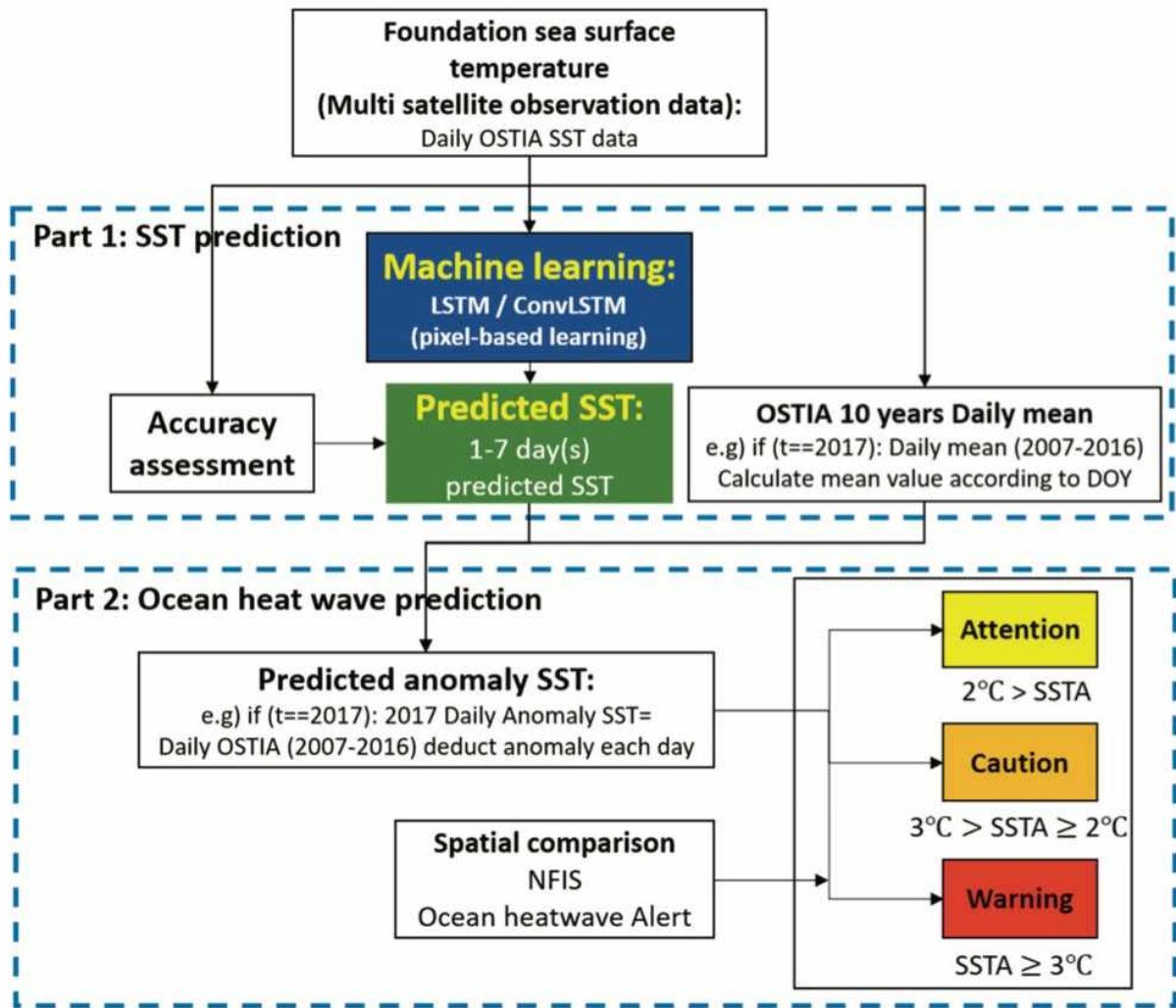
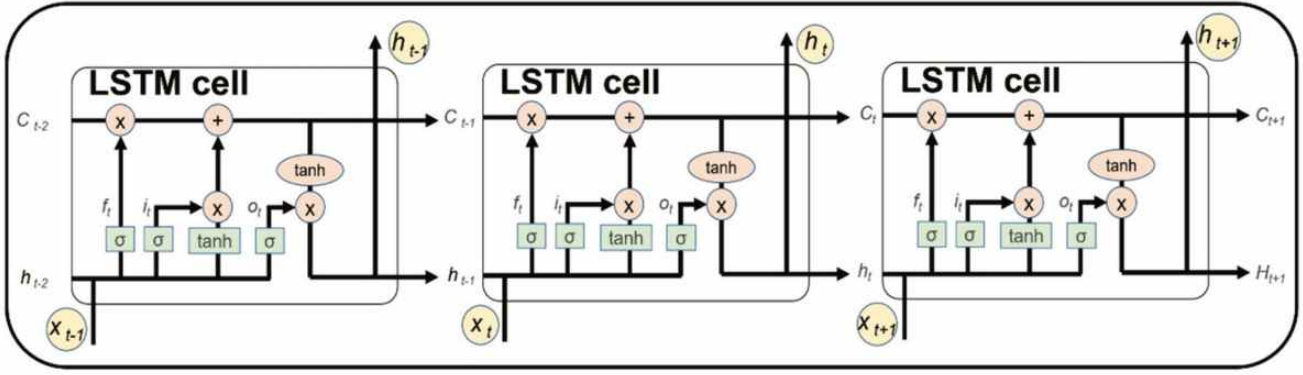


Fig. 2. The overall flowchart of this study. Part 1 is to predict sea surface temperature and Part 2 is to predict ocean heatwave.

2) 시계열 딥러닝 기법

(1) LSTM

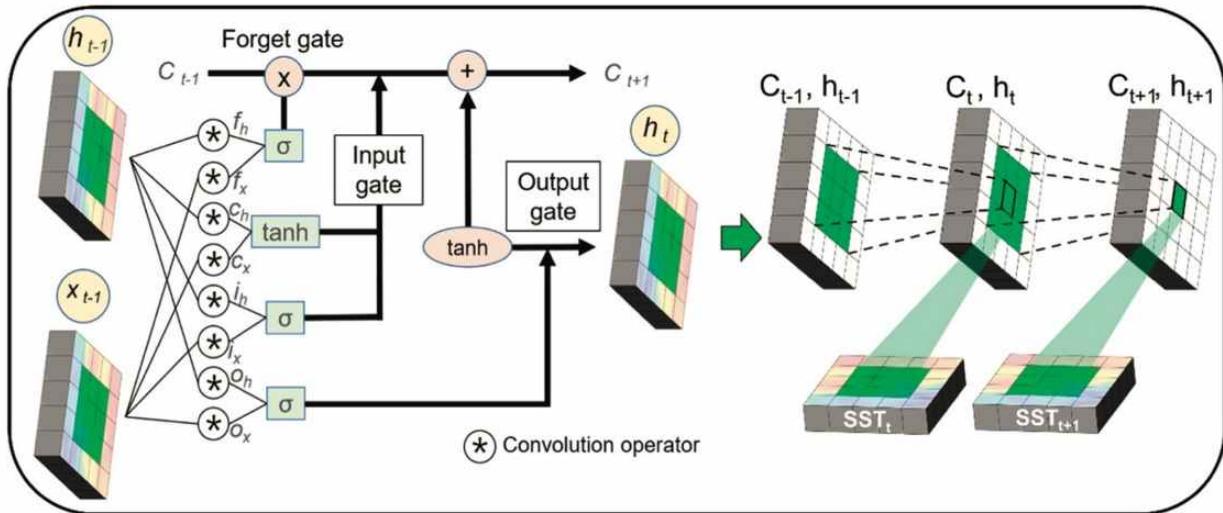
LSTM은 RNN의 일종으로 기존 RNN의 장기간 데이터에 대한 학습능력 저하와 기울기 소실 현상을 보완한 모델이다. 아래 그림과 같이 LSTM은 이전 cell ($t-1$)의 정보를 받아 현재의 state(t)를 update하고 다음 셀($t+1$)로 전달하는 구조이다. 각각의 cell에서는 sigmoid 함수로 다음 셀에 전달할 정보와 누락할 정보를 구분한다. 이 연구에서는 과거 14일의 해수면 온도 정보로 다음 날의 온도를 예측하는 모델을 각 픽셀별로 구축했다. 이때, 과거 정보는 14일이 최적화된 결과를 도출한다. 추가로, 픽셀별 시계열 특성값 차이로 인한 노이즈는 low-pass Gaussian filter를 적용해 smoothing했다. 더불어, 옵티마이저로는 adam을, 학습률은 0.005로 설정했다.



(a) LSTM

2) Convolutional LSTM

ConvLSTM은 기존 Fully Connected LSTM (CNN + LSTM)이 공간적 특성을 반영하지 못하는 단점을 보완한 모델이다. 이 모델은 기존 모델과 학습 방법은 동일하지만 입/출력, 상태 레이어가 3차원 벡터로 연산되며 일반 행렬곱 대신 합성곱으로 이루어져 시간적, 공간적 특성을 동시에 학습할 수 있다는 장점이 있다. 이 모델은 기존 LSTM, FC-LSTM 모델보다 우수한 성능을 갖는다. 이 연구에서는 20개의 필터와 3 * 3 커널 크기, 128의 배치크기, 4개의 레이어를 선택했다. 추가로 마지막에는 1개의 필터, 1 * 1 * 1 size의 커널을 가진 Conv3D 레이어를 추가했다. 최종적으로 28일치의 과거 자료(input)를 통해 1일 예측치(output)를 산출하도록 했다. 이때, 학습에 사용되는 기간은 예측 날짜의 4배로 설정했다. 마지막으로, 학습 시 예측된 자료를 다시 입력 자료로 활용하도록 했다.



(b) ConvLSTM

Fig. 3. The structure of (a) Long Short Term Memory (LSTM) and (b) Convolution Long Short Term Memory (ConvLSTM) models used in this research. The x, h, c, i, f, and o are input sequence, hidden state, memory cell, input gate, forget gate and output gate, respectively.

3) 정확도 평가

정확도 측정을 위해 평균오차 (Mean Difference: MD) (1)와 RMSE(Root Mean Square Error) (2) 를 사용했다. MD는 과추정, 저추정과 같은 공간적 오차의 패턴 분석에 용이하며, RMSE는 공간적 오차의 절대적 분포를 비교하는데 용이하다.

$$MD = \frac{1}{n} \sum_{i=1}^n (y_i - x_i) \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \quad (2)$$

4. 연구결과 및 토의

1) 예측 오차 변화 분석

예측오차 분석에 있어 대상을 전체 테스트 기간과 여름 계절로 나눠 비교했다. 전체 기간에 대한 RMSE는 LSTM의 경우 0.69°C, ConvLSTM은 0.56°C로 ConvLSTM이 상대적으로 우수했다. 더불어, 선행 연구와 비교했을 때, ConvLSTM의 경우, 7일 예측 결과에서 0.71°C로 선행 연구의 0.85°C보다 RMSE 오차 측면에서 개선된 성능을 보였는데 이는 연산에 사용된 픽셀량, 레이어 수에 따른 차이로 보인다. 여름 기간에 대해서는 LSTM이 0.90°C, ConvLSTM이 0.71°C로 마찬가지로 ConvLSTM이 더 높은 성능을 가진다. 전반적으로, 예측 기간이 길어짐에 따라 LSTM 모델의 RMSE가 ConvLSTM에 비해 빠르게 증가하며, 여름철의 예측 오차가 더 빠르게 증가함을 확인했다. 오차 증가폭에 대해서는 LSTM이 ConvLSTM보다 크며, 이는 합성곱을 활용한 연산의 성능 향상에 기인한다.

2) 일일 해수면 온도 예측 결과 정확도 분석

LSTM과 ConvLSTM의 해수면온도의 일일 예측값의 오차와 계절별 예측 정확도를 비교했을 때, 전 계절 결합적으로도, 개별적으로도 ConvLSTM 모델의 예측 정확도가 LSTM 모델에 비해 높음을 확인했다. 이상 해수면 온도 현상이 발생하는 여름의 경우 전반적인 모델의 예측 정확도가 떨어졌지만, ConvLSTM의 정확도가 LSTM 모델보다 높은 것은 동일했다. 또한, 수심이 얇은 연안에 가까워질수록 예측 정확도가 떨어졌으며, 이는 대마 난류로 인한 공간적 분포 오차 혹은 OSTIA 자료의 특성으로 인한 오류일 수 있다.

또한, 계절별 공간 분포 예측 결과를 비교분석하였는데, 여름을 제외한 계절에는 수심이 깊어질수록 온도가 증가하는 패턴을 보였으며, 겨울을 제외하고는 RMSE 측면에서 ConvLSTM의 성능이 LSTM보다 높게 측정됐다. 한편, LSTM은 공간분포에 대해 높은 노이즈 현상을 보였는데, 이를 해결하기 위해 픽셀의 공간적 특성을 반영하는 CNN을 적용한 FC-LSTM, ConvLSTM)을 활용할 수 있다. 또한, MD를 확인한 결과 7/1, 12/1에 과추정된 부분이 있다.

3) 예측 기간별 공간적 오차 분석

LSTM, ConvLSTM 두 모델 모두 예측일수 증가에 따라 예측오차가 증가했으며, MD 측면에서 LSTM은 불연속적인 오차패턴을, ConvLSTM은 연속적인 양의 오차를 보였다. ME 측면에서는 LSTM이 상대적으로 높은 노이즈와 과추정 패턴을 보였다. RMSE 측면에서도 LSTM 기반 모델이 상대적으로 낮은 예측 정확도를 보였다. 두 모델 모두 4일을 기준으로 예측 오차가 급격히 증가했으며, 연안지역의 예측정확도가 상대적으로 더 낮았다. 더불어, ConvLSTM 모델에서는 예측을 거듭할수록 추가되는 연산의 수가 늘어나 Smoothing되는 경향을 보였다.

4) 고수온 영역 탐지 활용 결과

국립수산과학원에서 제공하는 고수온 속보와 모델의 예측 결과를 비교했는데, 이때 몇가지 원인 (OSTIA 데이터의 주간 효과 제거, 공간적 분포의 비동일성 등)으로 인해 직접적인 비교에 한계가 있을 수 있다. 한편, 고수온이 발생한 2017-8-17일 레퍼런스 자료로부터 ConvLSTM의 예측결과가 LSTM에 비해 좀 더 유사한 공간적 분포를 가졌다. 이상치 측면에서도 LSTM 모델은 예측 기간이 2일 이상일 때 급격한 성능 저하가 나타나는 반면, ConvLSTM 모델은 예측기간이 5일을 넘어서는 순간부터 예측 오차가 서서히 증가하는 경향이 있다.

5. 결론

해수면 온도 예측, 고수온 해역 탐지에 있어 LSTM 보다 ConvLSTM이 여러 방면에서 낫다!! 예측 기간이 길어짐에 따라 과추정이 심화되는 현상은 앙상블 기법(ex. Adaboost, 편차 보정 랜덤 포레스트 등)을 활용해 성능을 개선할 수 있다. 이로써, 장기간 예측 시의 오차도 줄이고 고수온 탐지 영역의 축소 현상도 보완할 수 있다.