

데이터 기반 한강 수질 예측

2018. 10. 04

빅데이터과제 progress seminar

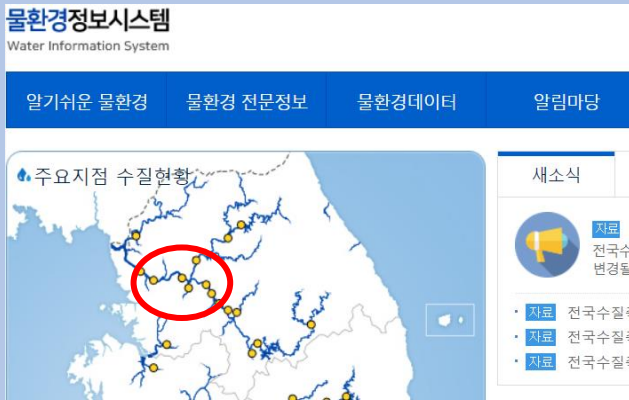
홍한움

연구의 필요성 및 목적

- 효율적인 수질 관리를 위한 수질 예측의 필요성
 - 딥러닝 기술의 발달로 가지고 있는 자료를 최대한으로 이용한 수질 예측 가능
 - 물환경 정보 시스템의 정보를 최대한으로 활용 cf) 물리모형 (QUAL2K, WASP)
 - 기존 통계모형의 선형성, 안정성 등의 가정으로부터 비교적 자유로움
- 다층 퍼셉트론 신경망, 순환신경망(RNN)을 이용한 수질 예측
 - 한국 정보화진흥원의 녹조예측 선행연구: 초기 형태의 RNN
 - 최근 10년간 크게 발전한 순환 신경망 알고리즘 이용
 - 분석 대상 변수: 클로로필-a(Chl-a)
 - 예측 대상 지역: 가양-노량진-팔당
 - 자료 범위: 2008년 1월 – 2017년 12월 (시간 해상도: 주)
 - Training : 2008.01.10-2017. 08. 05 400개 (Validation : 2008.01.10-2013. 5. 18 (120개))
 - Test set: : 2017.08.05-2018. 12. 30 120개



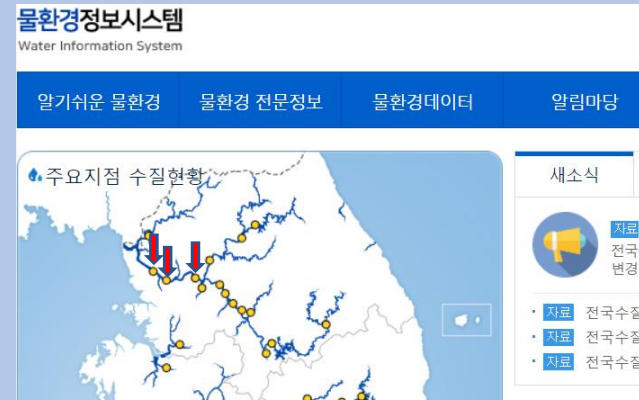
Input 변수



현재 및 과거 Chl-a

예측&상류 지점
과거 Chl-a

(가양, 노량진, 팔당 & 삼봉리, 경안천5, 강상, 강천)



예측지점 현재 및 과거 수질

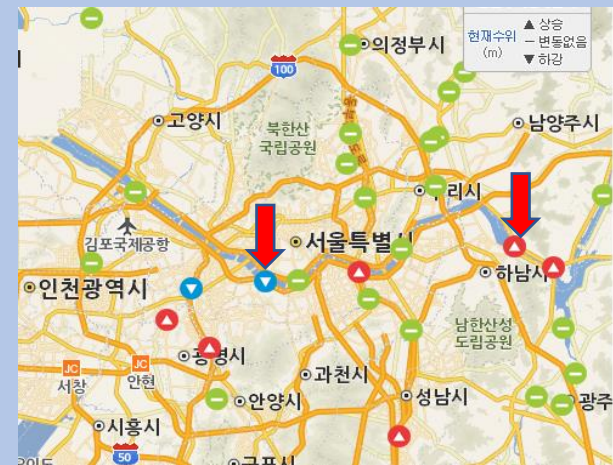
pH, DO, BOD, TN, TP, 수온, ...
전기전도도
총대장균군수
분원성대장균군수



기상 자료

기온, 강수량, 풍속, 습도, 기압, 일사량

(서울, 양평 관측소)

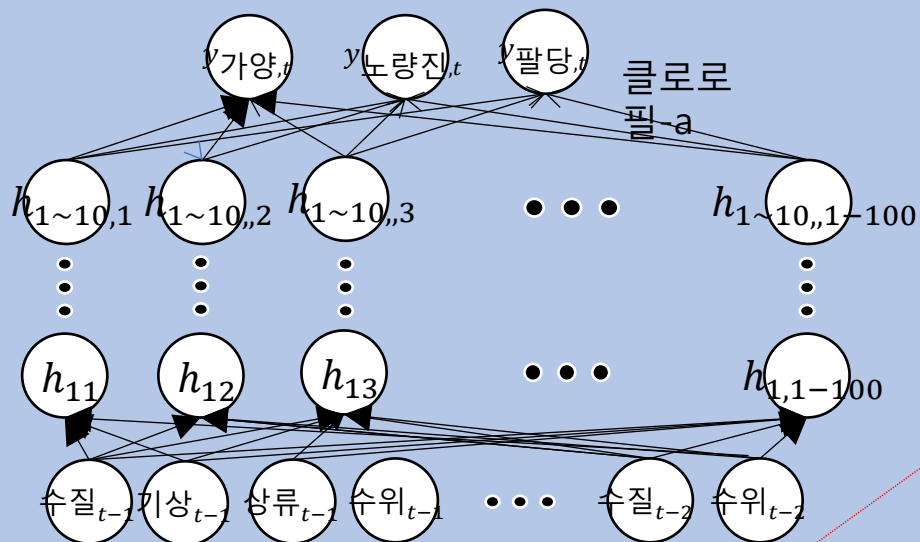


수위, 유량

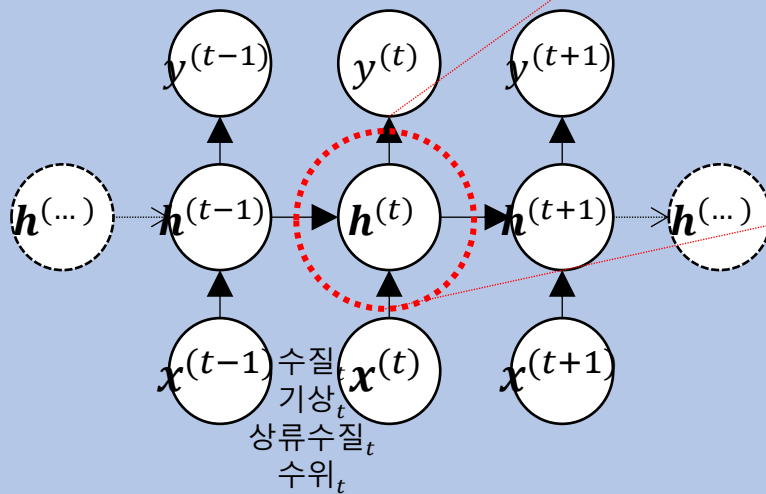
팔당대교: 수위, 유량
한강대교: 수위, (유량은 2008년 자료가 없어 제외)

모형

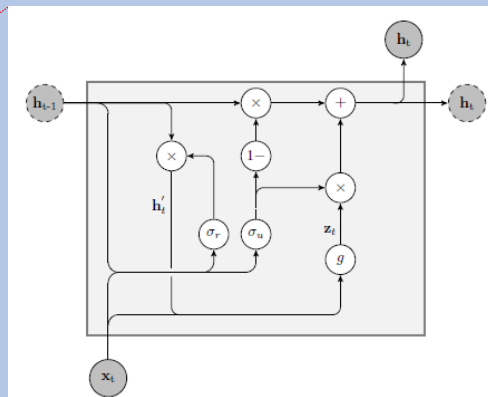
- 선형회귀분석모형
- MLP
- simpleRNN
- GRU
- LSTM
- stateful RNN



• MLP



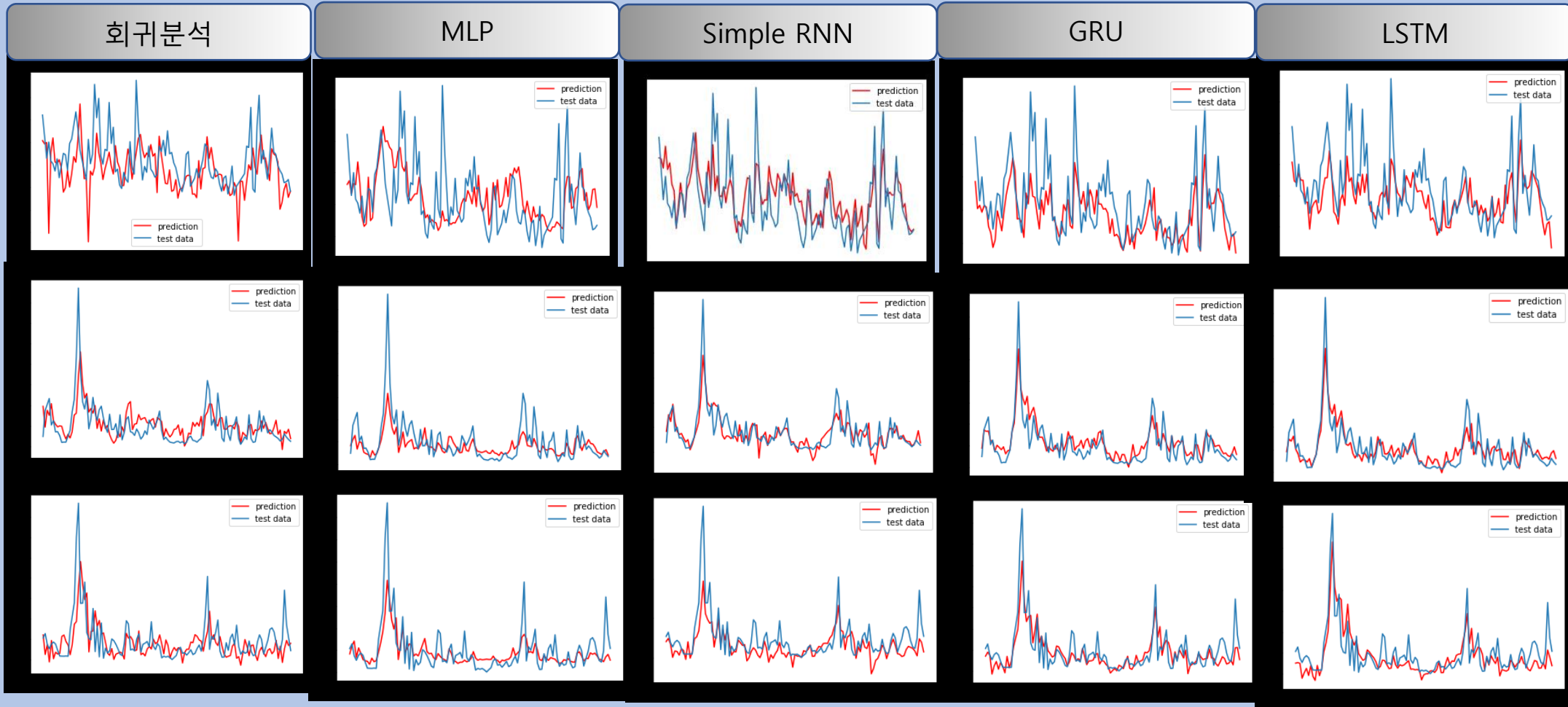
• Simple RNN



reset gate : $r[t] = \sigma(W_r h[t-1] + R_r x[t] + b_r)$,
 current state : $h'[t] = h[t-1] \odot r[t]$,
 candidate state : $z[t] = g(W_z h'[t-1] + R_z x[t] + b_z)$,
 update gate : $u[t] = \sigma(W_u h[t-1] + R_u x[t] + b_u)$,
 new state : $h[t] = (1 - u[t]) \odot h[t-1] + u[t] \odot z[t]$.

• GRU, LSTM

결과



팔당댐

노랑진

가양

RMSE: 16.95

RMSE: 13.55

RMSE: 12.62

RMSE: 10.93

RMSE: 11.51

요약 및 결론

- Chl-a 의 변화 추세와 큰 값을 잘 예측함
 - 회귀분석, ARIMA의 지연예측과 비교됨
 - 데이터가 더 모일 경우 RMSE를 더 줄일 수 있을 것으로 기대
 - 학습을 진행하며 Validation set의 비율을 줄이면서 trainging set 의 크기를 조금씩 늘릴 때마다 RMSE가 크게 줄어들었음
 - 큰 값의 예측 성능이 좋으므로 녹조 예측에 활용 가능

-
- Hyperparameter tuning
 - Staful RNN 적용시 예측력 향상 기대