<div align="center">

|  |  |  |
| --- | --- | --- |
| **2018** | **Report – 2018/ 11월 4주** | **Seong-gu** |

</div>

**I. Plan for this week**

***Personal Research Subject***

* 예심 논문 작성
* 정확도를 높이기 위한 모델링 최적화
* 시간, 공간,시공간 보간법 적용 후, 결과 비교

**II. Plan for next week**

* 논문 초안 작성 완료

**III. Response on Prof. Shin's feedback last week**

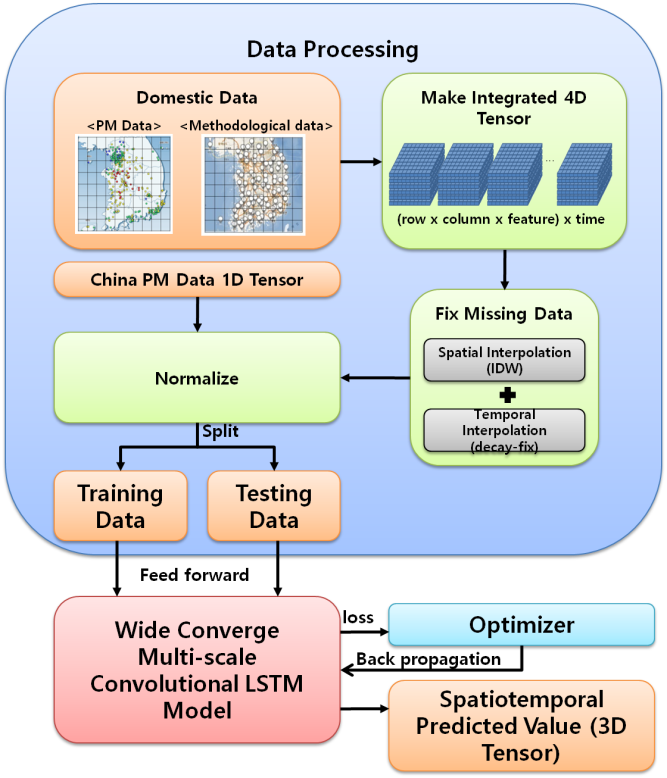
* 위치적으로 조밀하게 정확한 미세먼지 농도 데이터를 추정하기 위한 방법론 검토  
  (꼭 로라 게이트 웨이를 사용하지 않더라도)
* 예측시간 8시간은 너무 길고 예측 성능이 나쁘게 나오니, 예측시간을 줄일 것.

**IV. Milestone**

1. 실험실 환경에서 미세먼지 농도 예측 시스템 구축   
   1. 머신러닝 기법을 활용한 미세먼지 농도 예측  
        
      **- 한국의 지리학적 특성 및 계절풍의 영향을 고려하여, 활용할 측정소 범위를 국내 전역으로 하고, 중국의 미세먼지 농도를 모델링 데이터로 사용함.**(중국데이터를 확보하는 방법 조사할 것)

**- 결측치는 공간적 보간법인 IDW 기법과 시간적 보간법인 decay-fix 기법을 결합한 시공간 보간법을 사용하여 추정함.**

**- 모델링은 미세먼지 유입의 공간적인 특성을 적극 반영하기 위해서 최근에 연구되고 있는 Convolutional LSTM 모델을 사용.**(모델링 최적화 필요)



**V. Reports**

* 개인연구 진행사항
* **모델링 관련**

현재 진행되고 있는 모델링 단계입니다.

1. **측정소 선별, Convolutional LSTM input matrix에 data mapping – 진행완료**
2. **Fixing missing value – 진행완료**

**3) Convoltional LSTM 모델링 및 파라미터 최적화 - 진행중  
4) 다른 모델링(LSTM, CNN, CNN-LSTM 등)과 비교- 진행중**

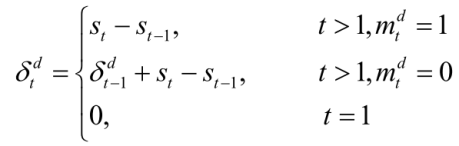
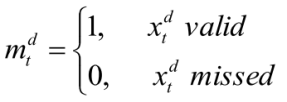
**2). Fixing Missing Values**

1. forward fix  
2. IDW fix

3. Forward & IDW(Inversed Distance Weight) -fix

들을 사용하여 보간한 데이터를 통해 결과값을 비교할 예정입니다. 해당 방법들은 구현이 되었습니다.

가장 나은 성능은 3으로 보이나, 2와 3은 큰 차이가 없을 수도 있습니다.

   
( : dimension d와 t에서의 측정값, : t에서의 timestep값)

라고 하면,

1. forward fix  
: 누락데이터의 있다면, 해당 누락 데이터가 측정되기 이전값을 그대로 가져옵니다.

****

2. IDW fix

****

**: IDW 추정값 ,**

3. Forward & IDW(Inversed Distance Weight) -fix

****

****

** : 최신 측정값**

**: 최신 측정값과 추정할 값과의 time-step  
(= 주변 지점 측정치, = 대상 지점과 주변 지점 i까지의 거리, b= 거리가중계수)**

**3) Convoltional LSTM 모델링 및 파라미터 최적화 – 진행중**

**아래 파라미터를 적용하여 모델링하였으며, 해당 모델링은 basic한 모델링 이기 때문에, 최적화를 진행하면서 변경할 예정입니다.**Epochs = 200

batch\_size = 128

**ConvLSTM2D**number of filter = 32

kernel size = 1x1, 3x3 ,5x5

**LSTM**hidden state = 200  
**Dence**state = 100

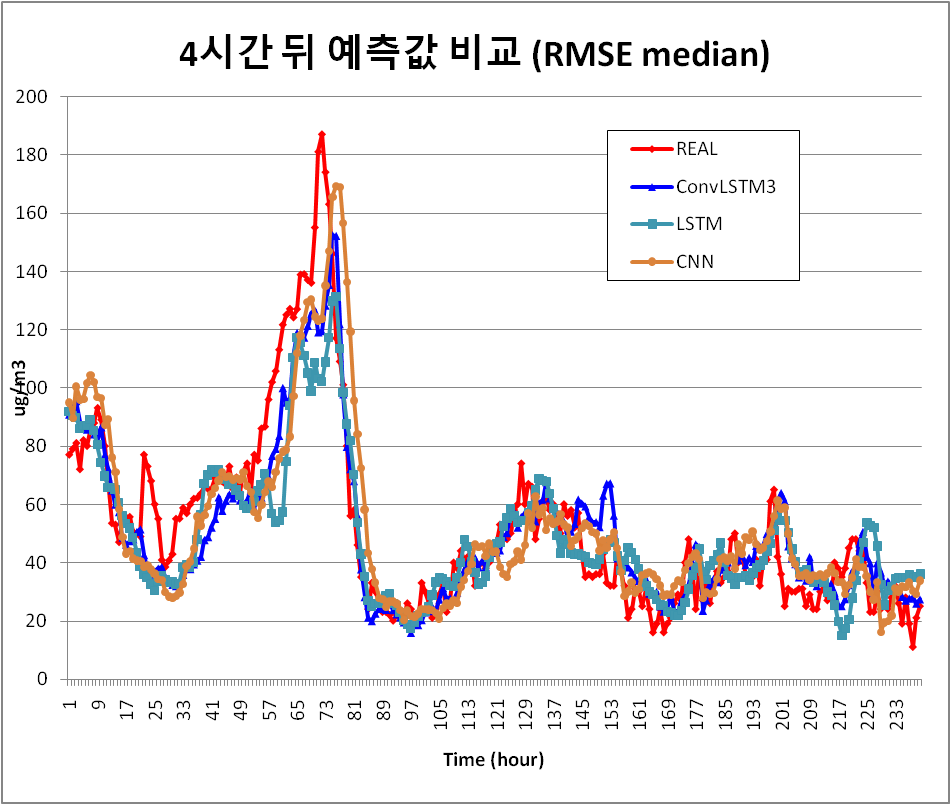
**4) 다른 모델링(LSTM, CNN, CNN-LSTM 등)과 비교- 진행중**

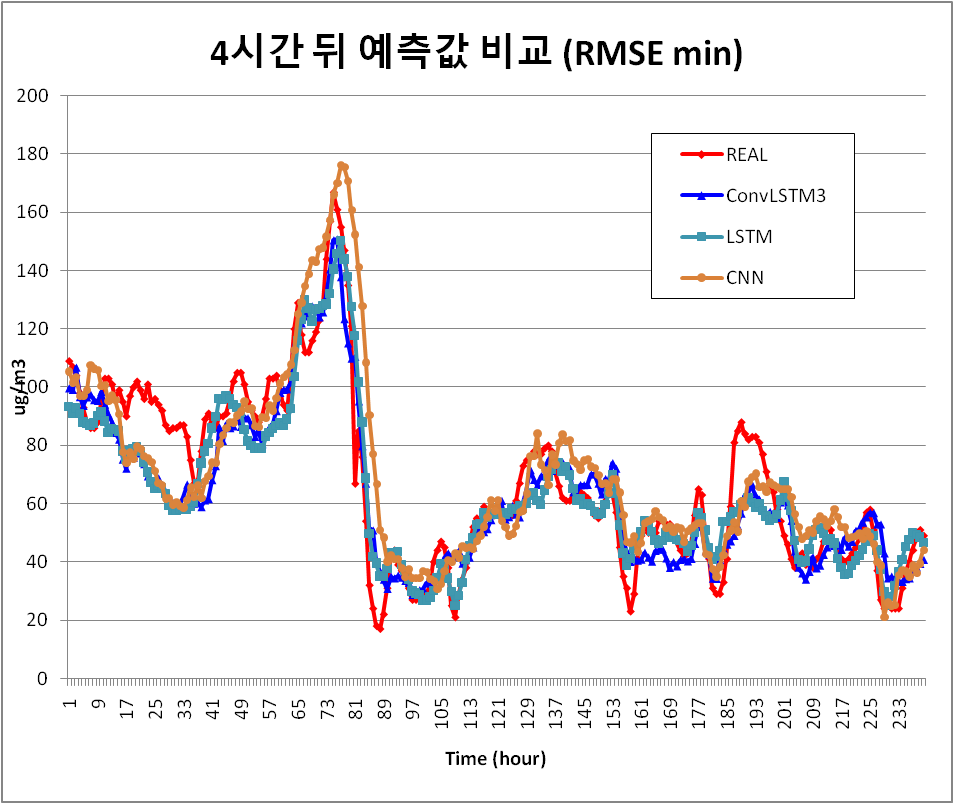
이전과 다르게측정소 91개 전부를 대상으로 RMSE를 계산하였습니다.

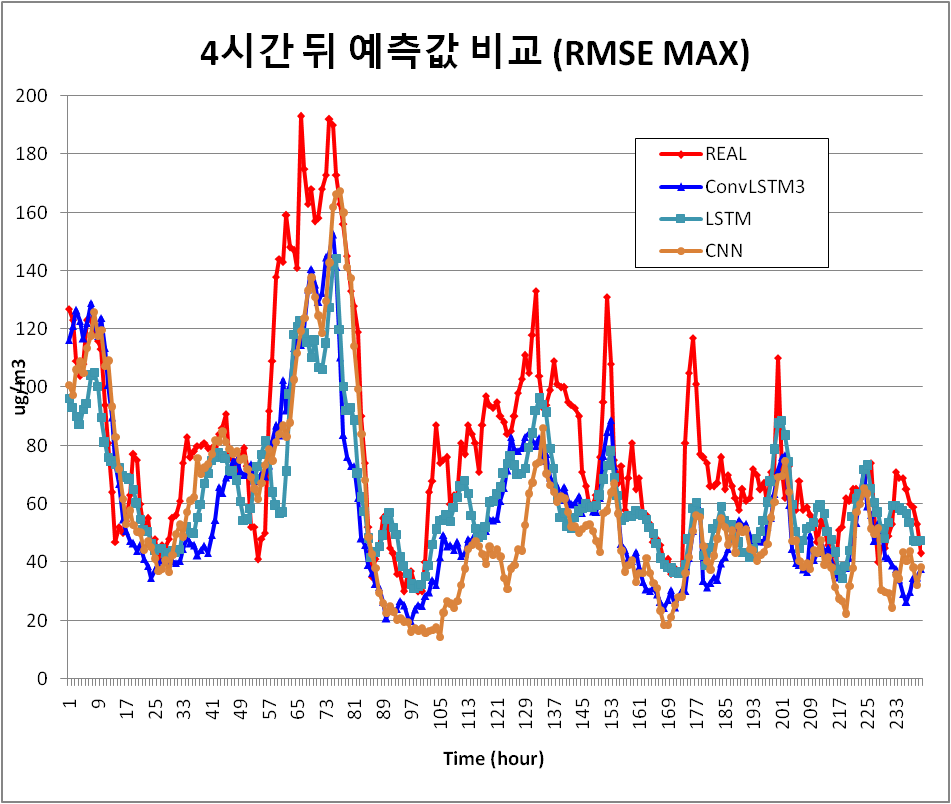
* **결과**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | RSME(average) | RMSE(min) | RMSE(max) | RMSE(median) |
| ConvLSTM 1x1 | 21.79948 | 12.9068 | 38.88621 | 20.87167 |
| ConvLSTM 3x3 | 21.00426 | 12.69053 | 35.50324 | 19.95437 |
| ConvLSTM 5x5 | 21.77158 | 13.56223 | 36.58099 | 21.25255 |
| LSTM | 22.6799 | 13.72855 | 36.60621 | 21.73275 |
| CNN | 22.44803 | 14.22062 | 39.43956 | 22.0319 |

아래는 4시간 뒤 예측값 비교 그래프입니다.







* **Future Work**

1. 테스트 시간이 오래걸리기 때문에, 테스트 시간을 줄이기 위한 최적화 진행
2. 정확도를 높이기 위한 모델링 최적화 (계속)
3. 결과값 비교
4. 비교 대상으로 기존 논문에서 사용되고 있는 아래 모델 추가 (A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities Chiou-Jye / Published: 10 July 2018)  
     
   