<div align="center">

|  |  |  |
| --- | --- | --- |
| **2019** | **Report – 2019/ 3월4주** | **Seong-gu** |

</div>

**I. Plan for this week**

***Personal Research Subject***

* GCN-LSTM 과 ConvLSTM 결과 비교 (Error 및 Trade off 관점)

**II. Plan for next week**

* 모델링 개선 방법 여러가지 도출   
  (1. wind effect와 local dependency를 잘 반영할 방법. : 방향성을 넣을 방법.  
   2. Location buffer를 넣을지?  
   3. GCN-LSTM 내부 연산 방법 개선 –(주변 station 결과를 잘 반영하도록)
* Wind vector 분해 및 feature extraction 방법 검토
* 결과에서 피크값 확인-대조
* Attention 구조 적용 반드시 고려 및 시간대별 모델링 다르게 하는 것도 고려..

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

* 랜드마크 등 지형, 지역적 특성 수기 추가
* Feature를 잘 선택 (데이터 프로세싱으로 정확도 개선)
* 지역은 한국을 전범위로 하지 말고 적당한 수준으로 결정함.
* 시간대 설정
* 효율성 및 정확도 면에서 트레이드 오프 분석 설계 방안을 제기
* 온도, 습도, 대기압, 바람세기, 풍향 데이터 외에 날씨(눈/비/안개 등)도 포함하면 좋을듯함
* 앙상블 기법 (기초 모듈을 조합해서 좋은 모델 만들기)

**IV. Milestone**

* 시공간 데이터 보간 기법 활용 (forward + IDW)
* 비선형 데이터 관계를 고려한 다층 구조 설계
* Attention 구조를 결합한 모델링도 고려. <https://github.com/zxj32/GCN-GRU/blob/master/git_figure.PNG>’
* feature extraction 방법 검토
* 추가 feature에 대한 성능 개선 효과 확인 (location predictor, weather, china PM ..)
* feature extraction 방법
* : forward-fixed 기법과 IDW 기법에 대해 가중치를

**V. Reports**

* 개인연구 진행사항 – GCN 모델링
* GCN 오픈 소스 분석 – 완료

- GCN에 맞는 데이터 전처리 - 완료

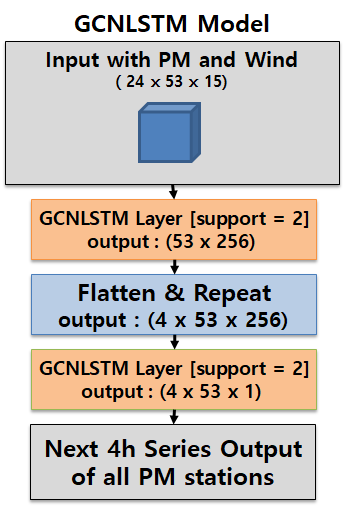
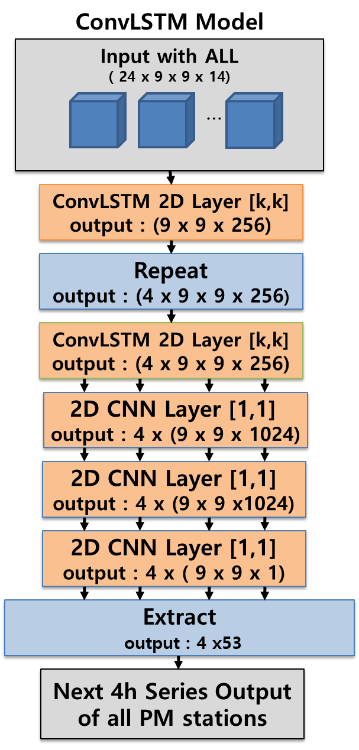
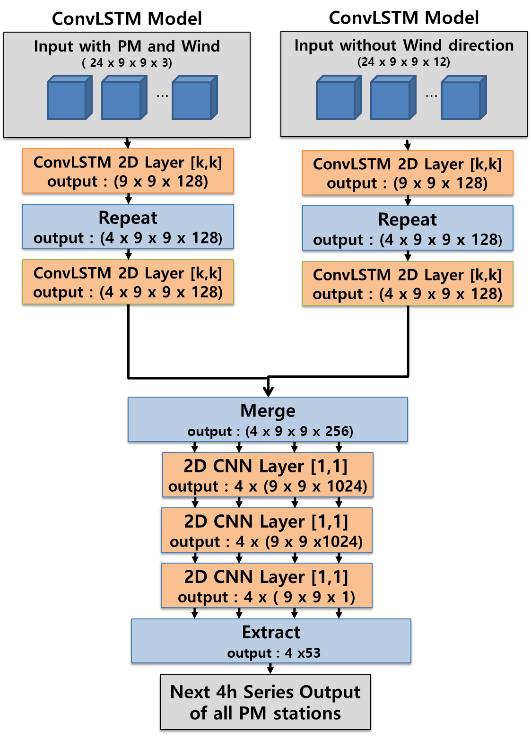
* GCN 소스, LSTM에 결합 모델링 – 완료
* GCN-LSTM 과 ConvLSTM 결과 비교 - 완료

GCN-LSTM input data는 (input time-step, stations number, features) 형태로 기존 ConvLSTM (input time-step, row, column, features) 보다 더 적은 데이터만을 사용합니다. Graph convolution 연산을 위해서는 이 input data 이외에 그래프 연결 상태(인접된 station 구조)를 나타내기 위해 가중치 인접행렬을 만듭니다.

인접 행렬의 가중치는 해당 station에서 가장 가까운 8개의 station에 대해서만 유효하도록 합니다. 곧, 인접한 8개의 staions만 직접적으로 연결되어 각 time-step마다 해당 station의 미세먼지 예측 연산에 사용되게 됩니다. 가중치는 (1/거리)로 계산되어 거리가 먼 station은 영향을 덜 주게 됩니다.

* 개인연구 진행사항 – 결과 비교

| Model | Prediction hour | | | | Average |
| --- | --- | --- | --- | --- | --- |
| 1h | 2h | 3h | 4h |
| Hybrid 3x3 | **9.76** | 13.36 | 16.38 | 18.86 | 14.59 |
| **Double ConvLSTM 3x3**  **+ time predictor, 2차 생성물 (SO2,CO,O3)** | 10.18 | **12.77** | **14.99** | **16.84** | **13.69** |
| ConvLSTM 3x3  + time predictor, 2차 생성물 (SO2,CO,O3) | 10.41 | 13.15 | 15.71 | 17.81 | 14.27 |
| GCN-LSTM (simple) | 13.34 | 15.32 | 17.17 | 18.67 | 16.12 |



Double ConvLSTM(왼쪽)과 ConvLSTM(가운데) 및 GCN-LSTM (오른쪽)

GCN-LSTM의 결과는 기존 Hybrid 3x3보다 4h 예측부분에서는 약간 좋은 결과를 보였으나, 나머지 부분에서는 나쁜 결과를 보였으며 현재까지 가장 잘나온 Double ConvLSTM 3x3보다는 전체적인 시간대에서 나쁜 결과를 보였습니다. 다만, 아직 기초 모델이기 때문에 개선의 여지가 있습니다.