<div align="center">

|  |  |  |
| --- | --- | --- |
| **2019** | **Report – 2019/ 3월3주** | **Seong-gu** |

</div>

**I. Plan for this week**

***Personal Research Subject***

* GCN 오픈 소스 분석 완료
* GCN 소스, LSTM에 결합 모델링
* GCN에 맞는 데이터 전처리
* GCN-LSTM 과 ConvLSTM 결과 비교 (Error 및 Trade off 관점)
* feature extraction 방법 검토

**II. Plan for next week**

* GCN-LSTM 과 ConvLSTM 결과 비교 (Error 및 Trade off 관점)
* 모델링 개선 방법 여러가지 도출   
  (1. wind effect와 local dependency를 잘 반영할 방법. : 방향성을 넣을 방법.  
   2. Location buffer? )
* Wind vector 분해 및 feature extraction 방법 검토
* 피크값 확인

**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**

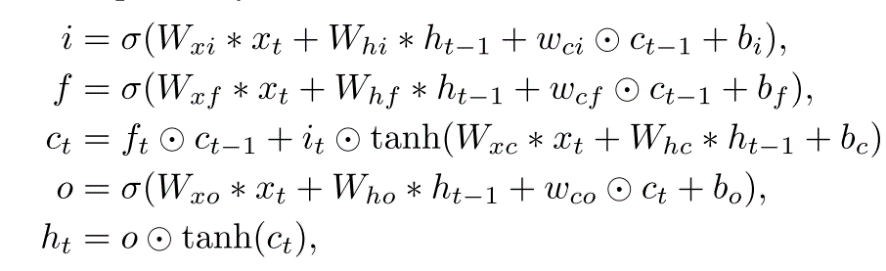
* 개인연구 진행사항 – GCN 모델링
* GCN 오픈 소스 분석 – 완료

- GCN에 맞는 데이터 전처리 - 완료

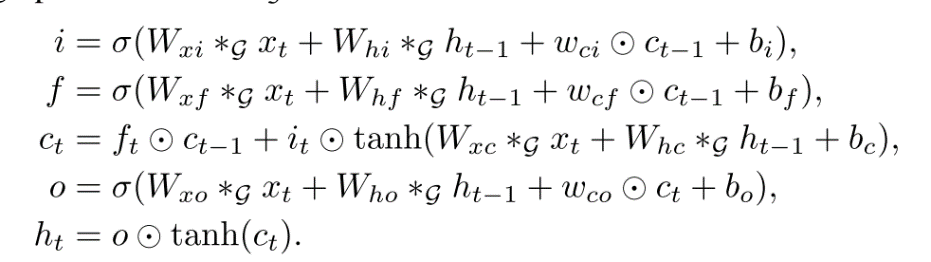
* GCN 소스, LSTM에 결합 모델링 – 진행중
* GCN-LSTM 과 ConvLSTM 결과 비교 (Error 및 Trade off 관점)

현재 LSTM모델에 GCN을 implement하는 작업을 거의 완료하였으나, LSTM 내부 처리 로직에 기존GCN 오픈 소스의 처리 로직과 충돌되는 부분이 있어 해당 부분을 수정 중입니다.

GCN-LSTM은 아래 convolution 연산을 graph convolution으로 대체합니다.



ConvLSTM 처리 식

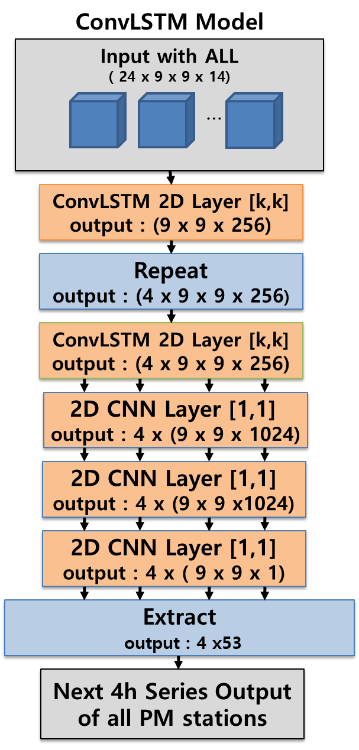
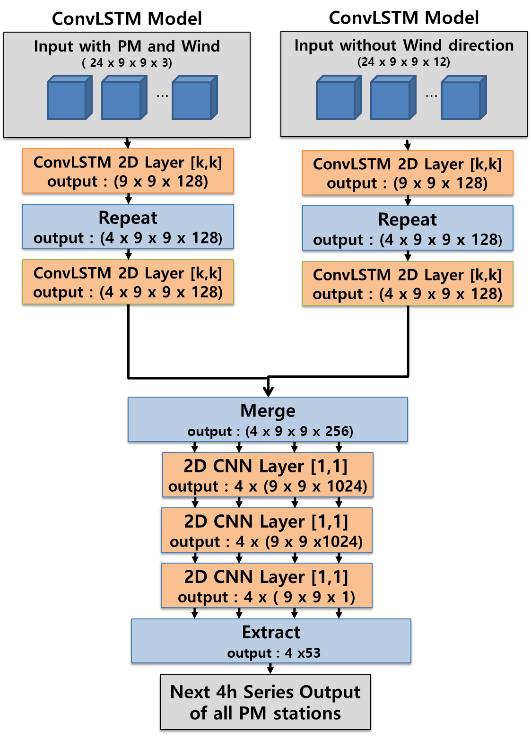


GCN-LSTM 처리 식

* 개인연구 진행사항 – feature 추가 결과

Time predictor(month, day of week, hour)와 2차 생성물질 (SO2, O3, CO 등)을 feature에 추가하였고, ConvLSTM을 1. Input 데이터를 wind effect만 고려한 것과 wind effect를 배제한 2가지 set으로 분리하고 ConvLSTM layer에 병렬로 feed한 구조와 2. Input 데이터를 모두 하나의 ConvLSTM 모델에만 넣은 구조를 비교해보았습니다.

| Model | Prediction hour | | | | Average |
| --- | --- | --- | --- | --- | --- |
| 1h | 2h | 3h | 4h |
| Hybrid 3x3 | 9.76 | 13.36 | 16.38 | 18.86 | 14.59 |
| ConvLSTM 3x3 | 9.85 | 13.61 | 16.69 | 19.19 | 14.84 |
| LSTM | 13.90 | 16.00 | 17.99 | 19.67 | 16.89 |
| Double ConvLSTM 3x3 | **9.48** | **12.67** | 15.25 | 17.41 | 13.70 |
| **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 |



Double ConvLSTM(왼쪽)과 ConvLSTM(오른쪽)

결과적으로 모델을 병렬로 이은 구조가 prediction time term이 길어질수록 좋은 결과를 보였습니다. 원인으로 두 가지를 뽑을 수 있는 데, 1. 모델을 병렬로 이은 구조 자체가 데이터 분석을 다각도로 가능하게 하기 때문 2. 바람에 의한 외부 미세먼지 유입 요인과, 기상 및 기타 내부 요인에 각각 영향을 미치는 요소를 분리하여 분석하였기 때문으로 보입니다. 또한, feature를 추가함으로써 3-4시간 예측에 대해서는 성능 개선의 효과도 얻을 수 있었습니다. 다만 1-2시간 예측에서는 feature를 추가하지 않는 모델이 더 좋은 성능을 보였습니다. 이는 추가된 feature들이 단기간 예측에서는 오히려 예상 미세먼지농도와 상관성이 떨어져서 모델 성능을 낮추게 한 것으로 보입니다.