<div align="center">

|  |  |  |
| --- | --- | --- |
| **2019** | **Report – 2019/ 5월4주** | **Seong-gu** |

</div>

**I. Plan for this week**

***Personal Research Subject***

* 본심 발표 준비
* Milestone 테스트

**II. Plan for next week**

* SCI 포맷에 맞추어 한글버전 완성

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

* JEET에 내는 것을 목표로 함.

**IV. Milestone**

1. 3일 중국 데이터를 넣어볼 것. feature가﻿ 너무 커지는 게 문제가 된다면 중국 데이터를 시간 단위가 아니라 3시간 단위로 넣는 방법도 고려할 수 있음. -> 완료

2. 3달 단위로 명확한 특성을 가지고 있으면 제안한 계절 단위 인코딩이 효율적으로 판단되나 실제 특성들이 어느 경우에느 2달 어느 경우에는 5달 이런 식으로 분포될 것으로 생각됨. 따라서, scalar encoding을 제안 방법과 결합해서 4차원 벡터로 사용하면서 계절내의 달을 (1,2,3)과 같은 형태로 구분해 주는 것도 고려. -> 완료

3. Training error와 Test error를 비교하여 두차이가 상당히 있을 때에 Overfitting을 줄이는 다양한 방법을 실시.

(정규화, dropout 방법) -> 아직 의미있는 성능 개선 보이지 않음

4. 중국 데이터 부분 RNN기반으로 개선해보기

5. 모델링 파라미터 초기화의 최적화

6. GRU+Convgru

**V. Reports**

* 개인연구 진행사항

모델 성능을 개선하기 위해서 아래와 같이 데이터 전처리 과정 1,2,3을 추가로 거친 결과를 정리하였습니다.

1. 기존 C파트에 24시간 중국 데이터 -> data dimension은 똑같이 24로 주되, 3시간간격으로 평균하여 총 72시간 중국 데이터를 반영할 수 있도록 하였습니다.
2. month를 season(봄, 여름, 가을, 겨울)에 따라 4의 data dimension을 가지는 one-hot encoding vector로 변환 -> 각 계절별 기준(봄=4월, 여름=7월, 가을=11월, 겨울=1월)에 대해 떨어진 만큼 역가중치를 준 inverted-weight encoding vector로 변환

예시 : 1월 = (0,0,0,3), 2월 = (1,0,0,2), 3월=(2,0,0,1), 4월=(3,0,0,0)

1. hours를 4시간별 묶음으로 6의 data dimension을 가지는 one-hot encoding vector로 변환 -> 2와 비슷하게 각 시간대별 기준(0, 4, 8, 12, 16, 20,)에 대해 떨어진 만큼 역가중치를 준 inverted-weight encoding vector로 변환  
   예시 : 0시 = (4,0,0,0,0,0), 1시 = (3,1,0,0,0,0), 2시 = (2,2,0,0,0,0)…

나머지 설정에 대해서는 동일합니다.

결과를 요약하자면 1시간 뒤는 제안한 T, T+W에서 약간 나은 결과를 보였으나 성능이 개선되었다고 말하기에는 개선폭이 좁으며, 나머지 예측 결과에 대해서는 비슷하거나 더 나쁜 결과를 보였습니다. 모델간 성능 차이는 설정하기 이전 결과와 비슷한 차이를 보였지만 특별히 T(simple)에 대해서는 1시간 뒤를 제외하고 개선된 결과가 나왔음을 알 수 있습니다.

다음과 같이 결론을 내릴 수 있습니다.

- C파트에 24시간 중국 데이터를 72시간 데이터로 확장한 것이 유의미한 성능 개선을 가져오진 않았다.

- month와 hour의 one-hot encoding vector를 inverted-weight encoding vector로 바꾼 것은 유의미한 성능 개선을 가져오지 않고 오히려 성능을 악화시켰음. 따라서 기존 방법을 유지하거나, month와 hour을 묶지않고 one-hot encoding vector해서 성능을 평가해보는 것도 좋아 보인다.

- 단, 이전 설정을 유지해서 나온 결과이며 다음 실험에서 설정을 바꾼 뒤 더 좋은 결과를 가져온다면 해당 DATASET으로 다시 실험 해보는 것은 의미 있어 보인다.

다음에는 모델링 및 하이퍼파라미터를 직접 바꿔서 실험을 진행할 예정입니다.

(milestone 3,4,5,6)

표 10: 1시간 뒤 예측결과 비교 (설정 이전 결과 / 설정 이후 결과)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **이전 결과** | T | T+W | T+W+C | T (simple) | Conv LSTM | CNN+ ConvLSTM  [1x1] |
| RMSE | 6.33 | 6.25 | 6.32 | 6.35 | 6.30 | 7.41 |
| MAE | 4.08 | 4.04 | 4.11 | 4.11 | 4.08 | 4.82 |
| IA | 0.9857 | 0.9859 | 0.9857 | 0.9856 | 0.9858 | 0.9804 |
| Accuracy | 90.93 | 91.06 | 90.87 | 90.93 | 90.95 | 89.43 |
| **이후 결과** | T | T+W | T+W+C | T (simple) | Conv LSTM | CNN+ ConvLSTM  [1x1] |
| RMSE | 6.30 | **6.22** | 6.41 | 6.43 | 6.42 | 7.48 |
| MAE | 4.12 | **4.01** | 4.18 | 4.18 | 4.19 | 4.89 |
| IA | 0.9858 | **0.9864** | 0.9852 | 0.9851 | 0.9852 | 0.9802 |
| Accuracy | 90.94 | **91.11** | 90.79 | 90.69 | 90.70 | 89.13 |

표 11: 4시간 뒤 예측결과 비교 (설정 이전 결과 / 설정 이후 결과)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **이전 결과** | T | T+W | T+W+C | T (simple) | Conv LSTM | CNN+ ConvLSTM  [1x1] |
| RMSE | 12.84 | **12.29** | 12.60 | 13.53 | 13.48 | 15.07 |
| MAE | 8.18 | **7.74** | 8.13 | 8.74 | 8.51 | 9.45 |
| IA | 0.9318 | **0.9406** | 0.9360 | 0.9242 | 0.9252 | 0.9049 |
| Accuracy | 81.79 | **83.02** | 81.98 | 80.73 | 81.18 | 79.04 |
| **이후 결과** | T | T+W | T+W+C | T (simple) | Conv LSTM | CNN+ ConvLSTM  [1x1] |
| RMSE | 12.82 | 12.40 | 12.71 | 12.78 | 13.46 | 15.23 |
| MAE | 8.19 | 7.76 | 8.32 | 8.11 | 8.47 | 9.71 |
| IA | 0.9324 | 0.9369 | 0.9342 | 0.9330 | 0.9231 | 0.9024 |
| Accuracy | 81.93 | 83.02 | 81.54 | 82.04 | 81.35 | 78.23 |

표 12: 12시간 뒤 예측결과 비교 (설정 이전 결과 / 설정 이후 결과)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **이전 결과** | T | T+W | T+W+C | T (simple) | Conv LSTM | CNN+ ConvLSTM  [1x1] |
| RMSE | 20.03 | 19.39 | **19.38** | 21.62 | 20.29 | 22.02 |
| MAE | 12.86 | 12.79 | **12.29** | 13.40 | 13.46 | 14.25 |
| IA | 0.7768 | 0.8006 | **0.8071** | 0.7364 | 0.7600 | 0.7399 |
| Accuracy | 70.89 | 70.96 | **73.14** | 69.70 | 68.90 | 68.43 |
| **이후 결과** | T | T+W | T+W+C | T (simple) | Conv LSTM | CNN+ ConvLSTM  [1x1] |
| RMSE | 21.05 | 20.39 | 19.51 | 21.04 | 20.63 | 22.29 |
| MAE | 13.92 | 13.63 | 12.83 | 13.52 | 13.93 | 14.70 |
| IA | 0.7583 | 0.7680 | 0.8013 | 0.7576 | 0.7793 | 0.7252 |
| Accuracy | 68.33 | 69.08 | **71.52** | 69.31 | 67.91 | 67.46 |

표 13: 24시간 뒤 예측결과 비교 (설정 이전 결과 / 설정 이후 결과)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **이전 결과** | T | T+W | T+W+C | T (simple) | Conv LSTM | CNN+ ConvLSTM  [1x1] |
| RMSE | 23.51 | 22.91 | 22.41 | 23.91 | 23.66 | 23.66 |
| MAE | 16.47 | 15.57 | 15.13 | 16.82 | 16.34 | 16.67 |
| IA | 0.6236 | 0.6269 | **0.7093** | 0.6087 | 0.6338 | 0.6452 |
| Accuracy | 62.06 | 63.58 | **66.15** | 60.15 | 63.63 | 64.08 |
| **이후 결과** | T | T+W | T+W+C | T (simple) | Conv LSTM | CNN+ ConvLSTM  [1x1] |
| RMSE | 23.77 | 22.80 | **22.25** | 23.15 | 24.06 | 24.56 |
| MAE | 16.37 | 15.24 | **14.94** | 16.07 | 16.74 | 17.05 |
| IA | 0.5847 | 0.6671 | 0.6848 | 0.6373 | 0.6160 | 0.6115 |
| Accuracy | 61.75 | 65.23 | 65.99 | 62.16 | 62.05 | 60.85 |