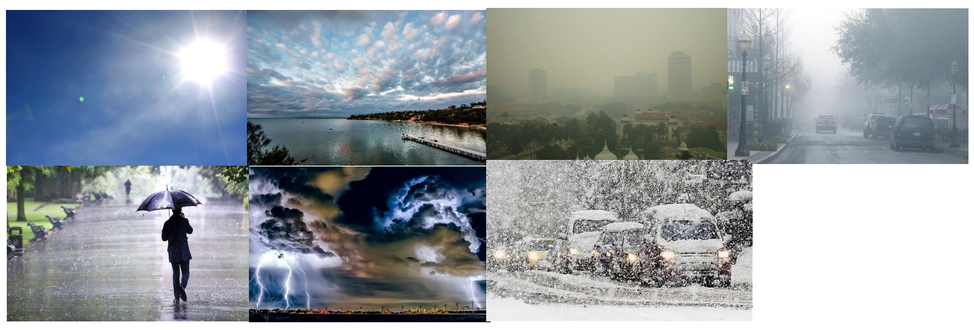
도시의 온도, 습도, 압력, 풍속 데이터를 이용하여 HMM으로 날씨 예측하기

Graduate Sch… onal University
Graduate Program in I… oul National University

2018 Project 2 포스터 모범사례

Backgrounds

- 도시의 날씨는 그 날의 온도, 습도, 압력, 풍속에 따라 다르고 그에 따라 여러 날씨가 존재 (Sky is clear, Clouds, Fog, Haze, Rain, Storm, Snow).
- HMM은 Markov Model 중 하나이고 날씨, 주식가격 등과 같은 어떠한 현상의 Hidden State와 그에 따른 Observation을 이용하여 Hidden State를 추론하는 모델임.
- 훈련 데이터를 이용하여 각 Hidden States의 Sequential 데이터에 대한 Transition & Emission Matrix 모델 형성
- 훈련 데이터 셋을 이용하여, 각 Hidden States에 대한 확률 계산 (decode process)

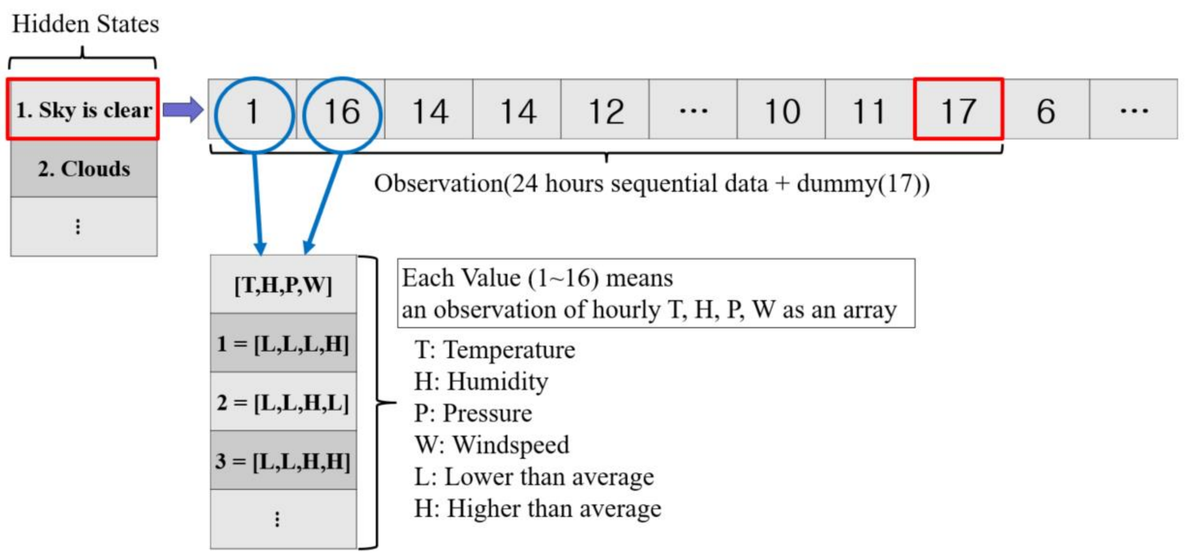


Research Goals

- 그 날에 대한 온도, 습도, 압력, 풍속 관측 값을 HMM을 이용하여, 그 날의 날씨를 예측.
- 계절별로 세분화하여 예측한 값과 사계절로 예측한 값이 달라질 것임.
  (데이터의 분산된 정도에 따른 예측도)
- 날씨 데이터 양에 따라 예측 정확도가 달라질 것임.
  (단순한 1년치 데이터, 5년치 데이터 비교)

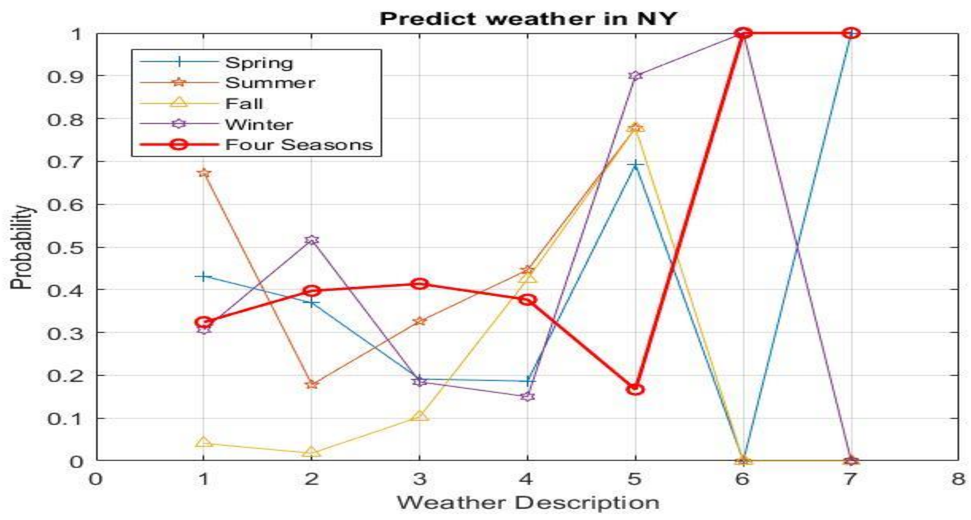
Methodology

- **날씨 예측을 위한 데이터 가공 및 구성**
  - X = {Sky is clear, Clouds, Fog, Haze, Rain, Storm, Snow}; Hidden States (#: 4)
  - O = {Temperature(High, Low), Humidity(High, Low), Pressure(High, Low), Wind Speed(High, Low)}; Observations (#: 2^4=16);
  - Observation 값의 평균을 기준으로 High, Low로 분류
  - 24 시간의 날씨 데이터에서 아침 7시에서 저녁 10시 중 많이 관측된 날씨를 그 날의 대표 날씨로 설정.
  - 시간별로 나타나는 관측 값(Observation)을 1~16으로 변환
  - 각 대표 날씨에 맞춰서 24 Sequential 데이터와 dummy(17)로 훈련 데이터 셋 구성.



Experimental Results

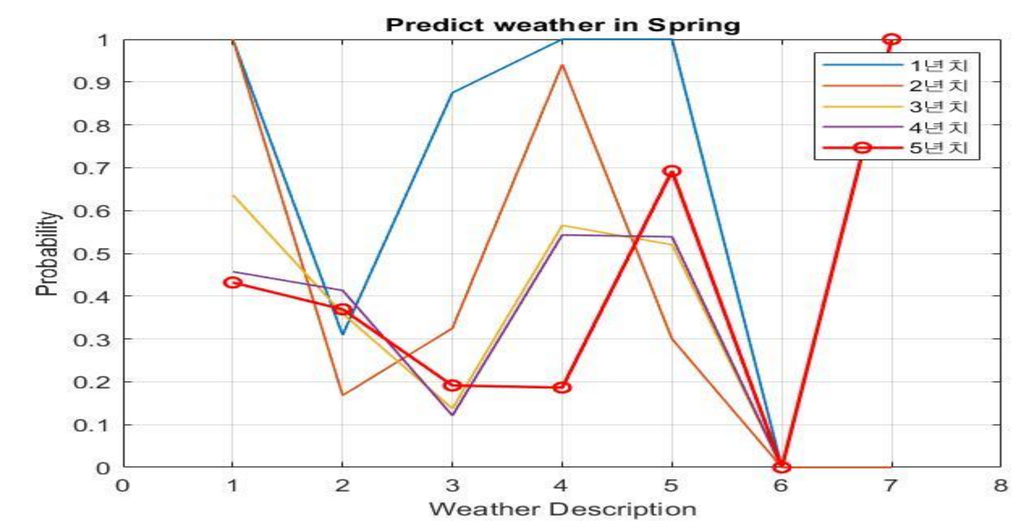
- 계절별로 세분화하여 예측한 값과 1년 전체의 데이터의 예측 정확도 비교



All Year	Observation						
Sky Clear	0.324	0.270	0.137	0.070	0.016	0.094	0.090
Clouds	0.171	0.397	0.265	0.032	0.021	0.056	0.057
Fog	0.081	0.144	0.414	0.185	0.053	0.083	0.041
Haze	0.051	0.051	0.228	0.377	0.102	0.112	0.080
Rain	0.000	0.037	0.241	0.260	0.167	0.056	0.241
Storm	0.000	0.000	0.000	0.000	0.000	1.000	0.000
Snow	0.000	0.000	0.000	0.000	0.000	0.000	1.000

Spring	Observation						
Sky Clear	0.432	0.273	0.045	0.000	0.159	0.000	0.091
Clouds	0.294	0.369	0.103	0.009	0.187	0.000	0.037
Fog	0.139	0.130	0.191	0.139	0.391	0.000	0.009
Haze	0.051	0.130	0.034	0.186	0.644	0.000	0.017
Rain	0.000	0.038	0.38	0.000	0.692	0.000	0.231
Storm	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Snow	0.000	0.000	0.000	0.000	0.000	0.000	1.000

- 데이터 양에 따라 예측 정확도와 신뢰도 비교



2013	Observation						
Sky Clear	1.000	0.000	0.000	0.000	0.000	0.000	0.000
Clouds	0.218	0.310	0.382	0.091	0.000	0.000	0.000
Fog	0.042	0.042	0.875	0.042	0.000	0.000	0.009
Haze	0.000	0.000	0.000	1.000	0.000	0.000	0.000
Rain	0.000	0.000	0.000	0.000	1.000	0.000	0.000
Storm	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Snow	0.000	0.000	0.000	0.000	0.000	0.000	0.000

2013~14	Observation						
Sky Clear	1.000	0.000	0.000	0.000	0.000	0.000	0.000
Clouds	0.477	0.168	0.103	0.131	0.121	0.000	0.000
Fog	0.075	0.000	0.325	0.400	0.200	0.000	0.009
Haze	0.000	0.000	0.000	0.941	0.059	0.000	0.000
Rain	0.000	0.000	0.000	0.000	0.300	0.000	0.000
Storm	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Snow	0.000	0.000	0.000	0.000	0.000	0.000	0.000

2013~15	Observation						
Sky Clear	0.636	0.061	0.000	0.061	0.121	0.000	0.121
Clouds	0.359	0.359	0.028	0.127	0.070	0.000	0.056
Fog	0.216	0.137	0.137	0.373	0.137	0.000	0.000
Haze	0.087	0.043	0.087	0.565	0.087	0.000	0.130
Rain	0.000	0.000	0.000	0.240	0.520	0.000	0.240
Storm	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Snow	0.000	0.000	0.000	0.000	0.000	0.000	1.000

2013~16	Observation						
Sky Clear	0.457	0.200	0.029	0.029	0.171	0.000	0.114
Clouds	0.201	0.413	0.056	0.145	0.140	0.000	0.045
Fog	0.077	0.121	0.121	0.308	0.374	0.000	0.000
Haze	0.029	0.029	0.029	0.543	0.343	0.000	0.029
Rain	0.000	0.000	0.000	0.240	0.538	0.000	0.231
Storm	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Snow	0.000	0.000	0.000	0.000	0.000	0.000	1.000

- 계절별로 세분화하여 예측한 값보다 사계절 데이터를 이용하여 예측한 값이 더 다양한 확률 분포를 보임.
- 관측 값의 평균에 대하여 분산된 정도가 1년 전체의 관측 값에서는 더 크게 되어, 예측 정확도가 떨어질 것이라는 예상과는 다르게 전체적으로 더 신뢰도가 높고 좋은 예측을 보임.
- 단순한 1년치 데이터 양을 보고 예측한 값보다 더 많은 연차 데이터를 보고 예측한 값이 대체적으로 더 좋은 정확도와 신뢰도를 보이지만, 일부 예측 정확도는 오히려 값이 감소.

Concluding Remarks

- 단순히 관측 값의 Low, High로 구성한 Sequential 데이터로도 패턴을 만들어 날씨의 예측이 가능함.
- 데이터의 양에 따라 예측 정확도가 바뀌지만, 꼭 많은 양의 데이터가 더 좋은 예측 정확도를 보이지 않음.
- 데이터의 양이 부족한 날씨(Storm, Snow…)에 대해서는 전혀 신뢰할 수 없는 결과가 나옴.
- 정확히 Hidden State를 예측하기 위해서는 일정 개수 이상의 데이터가 필요하지만, 오히려 너무 많은 데이터의 양은 정확도를 낮출 수 있음.