Spatial Machine Learning in Atmospheric Sciences

Orhun Aydin, PhD
Sudhir Shrestha, PhD
Artificial Intelligence

Machine Learning

Deep Learning
Artificial Intelligence: Pre-Machine Learning

- Programs with common sense¹ (McCarthy, 1960) → Set of **predefined** logical operators
  
  *in* (Orhun, Claremont), *in* (Claremont, California) → *in* (Orhun, California)

- Complex and extensive representations of human knowledge - Knowledgebases
  - CYC²
  - SenticNet (1, 2, 3, 4, 5)³

---

Artificial Intelligence: Machine Learning

• Learn rules and patterns from data

• Data is represented explicitly, knowledge is NOT
  • Data-driven

• Explicit rules do NOT exist, instead inferred from data

“A computer program is said to learn from experience $E$ with respect to some class of tasks $T$ and performance measure $P$, if its performance ... improves with experience $E.$”

Spatial Machine Learning
Spatio-Temporal Machine Learning
time-window
sequences
duration
temporal trends
change
Machine Learning in ArcGIS

Classification
- Maximum Likelihood Classification
- Random Trees
- Support Vector Machine

Clustering
- Spatially Constrained Multivariate Clustering
- Multivariate Clustering
- Density-based Clustering
- Image Segmentation
- Hot Spot Analysis
- Cluster and Outlier Analysis
- Space Time Pattern Mining

Prediction
- Empirical Bayesian Kriging
- Areal Interpolation
- EBK Regression Prediction
- Ordinary Least Squares Regression and Exploratory Regression
- Geographically Weighted Regression
- Random Forest Classification/Regression
Machine Learning in ArcGIS

- Classification
- Clustering
- Prediction
Integration with External Frameworks

ArcGIS

TensorFlow

IBM Watson

scikit learn
Machine Learning in Atmospheric Sciences
Descriptive

- Data aggregation
- Trend detection
- Signal decomposition
- Pattern-mining

Predictive

- Predict Extreme Events
- Climate Prediction
- Delineate Climate Zones

Prescriptive

- Legislation
- Resilience plans
- Urban Design
- Routing & Evacuation
Artificial Intelligence

Machine Learning

Deep Learning

TensorFlow

CNTK

scikit-learn

IBM Watson

Theano

Keras
Overarching Problems

• Complex physical system interactions
• Limitation on weather forecast models to complement extreme events
• Labor intensive weather and climate monitoring workflow- NOAA drought monitor, drought outlook, Temperature-Precipitation outlook
Bottlenecks: An Example

• We will have the GEFS reforecast data that NOAA NCEP/EMC is set to deliver next year but without data-driven approaches we fall short because of it’s complexity and wider applications.
Automated Statistical Downscaling of Climate Models
Problem of Scale Representation

- Stations measure temperature data sensitive to regional characteristics
  - Elevation
  - Nearby water-bodies, cities, ...

- GCMs simulations reflect
  - Analytical relationships between large scale climate parameters
  - They can be simulated well into the future
Statistical Downscaling

• Regional climate is conditioned by the local physiographic characteristics + large scale atmospheric state

• Assumption $f_{today} = f_{future}$
Space or Not to Space?

1. Not worried about it
   • My data/simulation has spatial auto-correlation
   • Non-spatial model will detect the impact of aspect

2. Space is an important input
   • I will feed spatial representation into my model

3. Space is inherent/implicit
   • I need a method to represent space implicitly
The Regression Problem

1. Non-spatial regression
2. Non-spatial regression w. spatial inputs
3. Spatial Regression
Predictors

Canadian Centre for Climate Modelling and Analysis Simulation:

1. Mean sea level pressure
2. Surface airflow strength
3. 850 hPa airflow strength
4. Surface zonal velocity
5. 850 hPa zonal velocity
6. Surface meridional velocity
7. 850 hPa meridional velocity
8. Surface vorticity
9. 850 hPa vorticity
10. 850 hPa geopotential height
11. 500 hPa airflow strength
12. 500 hPa zonal velocity
13. Near surface relative humidity
14. 500 hPa meridional velocity
15. Specific humidity at 500 hPa
16. 500 hPa vorticity
17. Specific humidity at 850 hPa
18. 500 hPa geopotential height
19. Near surface specific humidity
20. 500 hPa wind direction
21. Mean temperature at 2 m
References


Representation of Climate Data

1. Representing Space in climate problems

2. Representing time

3. Representing space-time
Spatial Representation

1. Geoenrichment
   1. Distance features
   2. Vector field summary

2. Representing time
   1. Convolutions
   2. Decomposition

3. Representing space-time
   1. Spatio-temporal aggregation
   2. Mining patterns in space-time
Road Ahead
Some Challenges ML Pose

• Model interoperability
  • Data-driven model in Site A → good enough for site B

• Quality metrics
  • Assessing suitability and quality of results
  • Incorporating scientific information

• Model visualization and explanation
  • ML models can be harder to communicate
  • Model components can be abstracted fully from underlying mechanism
By 2020...

- **Smart Devices**: 20B IoT devices
- **People**: 1.5 GB per day
- **Smart Home**: 50 GB per day
- **Smart City**: 250 PB per day
- **Connected Factory**: 1 PB per day
- **Autonomous Vehicle**: 5 TB per day
- **Stadium**: 200 TB per game
- **Smart Office**: 150 GB per day