UNIVERSITÉ LIBRE DE BRUXELLES

Learning Operators for Geophysic Phenomena Spatio-Temporal Simulation and Forecasting

2023-2024 Support Committee Meeting

Pascal Tribel

Faculté des Sciences
Université Libre de Bruxelles

December 16, 2024

Who am I

- Pascal Tribel
- Bachelor's and Master's degrees in Computer Science from ULB
- Master Thesis conducted at IRIDIA under Prof. Bersini
- Currently a first-year PhD student at MLG and teaching assistant in computer science
- Areas of expertise and interest:
  - Machine learning and time series forecasting
  - Analysis and processing of time-dependent signals and data
  - Modeling of physical systems, data simulation, and machine learning
  - Applications in geophysics

Support Committee

- Thesis Support Committee:
  - Gianluca Bontempi (Supervisor)
  - Corentin Caudron (Co-supervisor)
  - Tom Lenaerts (Chair)
  - Matthieu Defrance

Learning Operators (1/3)

- Operators map function spaces to other function spaces
- Common examples are partial differential equations:
  - $F(x_1, x_2, ..., x_n, \frac{dw}{dx_1}, \frac{dw}{dx_2}, ..., \frac{dw}{dx_n}) = 0$ where $w(x_1, x_2, ..., x_n)$ is the solution function and F is a given function
- Especially, if there is a parameter of w which is the time, then the equation describes the spatio-temporal evolution of a system
- Most ML models approximate *functions* by mapping vector spaces but struggle to generalize to operator

---

- Further work - Hypothesis: some dynamics are measurable but too complex to be expressed by *a standard formulation of the underlying PDE*, but they can be approximated, making ML a useful tool for their estimation

Learning Operators (2/3)

- Recent methods proposed in the literature are neural network based: those are called *neural operators*
- Neural networks designed to approximate operators
- Various approaches have been tested on known PDE systems (e.g., Navier-Stokes, Darcy flow)

- Further work: It's still unclear which transformations are essential for mapping function spaces

Learning Operators (2/3)

- Advantages of ML-approaches of Operator Learning
- If the hypothesis that a given system can be described by a PDE is accepted
  - The simulation can be performed faster than with traditional solvers
  - **Systems for which we don't have a good describing PDE could be simulated and their behaviour could be forecasted**

Seismicity

- A good case of system which can be described by an operator is the propagation of a seismic wave
- It has mutliple applications: the study of earthquakes, of volcanic tremors, of tsunami tremors
- It is a good starting point: we have formulations for the underlying PDE, and real-life temporal data is available, measured by seismometers
- Current real-life spatial coverage is relatively poor: in most cases, the seismometers are separated of $> 10km$

Meetings with Prof. Caudron

- To study operator learning in seismicity, large, reliable and high-resolution spatio-temporal data is needed
- Prof. Caudron's team experiments with Distributed Acoustic Sensing (DAS) in volcanic fields
- DAS uses optic fibers to record ground displacement
- Provides high-frequency spatio-temporal data
- Data is unannotated (no known events are associated with the experiments)

- Further work - Hypothesis: the volcanic seismicity of the ground can be described by partial differential equations (PDEs)

Distributed Acoustic Sensing data

- DAS experiments are in early stages and not yet suitable for ML

- Further work - DAS data must first be transformed into a format suitable for machine learning

Development of PyAWD (1/2)

- Since high-resolution spatio-temporal data is not yet fully available, simulated data could allow proof-of-concepts of the ML procedures
- Seismic wave propagation is described by PDEs (acoustic or elastic)
- Solvers for these PDEs exist but are complex and not user-friendly for ML. They solve one scenario at a time and don't generate large datasets for varied initial conditions or parameters

---

- I developed PyAWD (PyTorch Acoustic Waves Dataset), a Python library that generates datasets of acoustic wave propagation simulations than can be queried at all spatio-temporal point

Development of PyAWD (2/2)

- Comprehensive documentation and multiple tutorials are available

---

- An article is in preparation but requires a concrete experiment to demonstrate its usefulness
- We chose *epicenter retrieval* as a use case to present the tool to the research community

Use Case: Epicenter Retrieval (1/3)

- The challenge: using only two sensors in a plane that record wave motion over time, how can we determine the wave's epicenter?
- Typically solved using triangulation with three or more sensors
- PyAWD generated datasets for training and testing, with varying epicenters, force amplitudes, and delays

Use Case: Epicenter Retrieval (2/3)

- Several ML architectures were trained on the data or extracted statistics
- Simple methods (e.g., linear regression, decision trees) failed, but deep learning (e.g., temporal convolutional networks) produced significantly better results:
  - The problem can be addressed with ML
  - PyAWD provides a valuable dataset for complex problems

PyAWD: Article

- Final analysis of the experiment results is needed
- Discussion of the tool's limitations and advantages remains

---

- The article is nearly finished
- PyAWD was presented at a PyTorch Meetup (hosted by Devoteam), generating interest at its early stages

Seismic Monitoring and Analysis Challenge

- 25 participants
- Task: using satellite imagery before and after an event:
  - Classify the event: earthquake/non-earthquake
  - If earthquake, predict the magnitude
- My solution, although simple, demonstrated that sometimes ML/DL isn't the best choice

---

- Secured 2nd Prize at ECML-KPDD 2024 in Vilnius
- Resulted in an article submitted to ECML-KPDD 2024

National Electricity Demand Very Short Forecasting

- Organized by Gian Marco Paldino during the 2024 TRAIL workshop
- Achieved excellent results compared to the baseline
- This is a relatively unexplored area in the literature

- We are in the process of writing a paper to present our findings and methodology

Conclusion (1/2)

- What has been done:
  - Development of the PyAWD library
  - Redaction of the documentation and the tutorials
  - Presentation at a PyTorch Meetup
  - 2nd Prize for the SMAC Challenge
  - Challenge article submitted for the ECML-KPDD 2024
- On the short term:
  - Finish the redaction of the PyAWD article
  - Finish the redaction of the National Electricity Demand Forecasting article
  - Experiment the notion of Operator Learning on PyAWD data

Conclusion (2/2)

- On longer term:
  - Determine if seismic data can be accurately described by learned operators
  - Make DAS data usable in ML pipelines
  - Determine if DAS data can be described by PDE
  - Determine which transformation of the neural operators are indeed leading to a good operator approximation $\rightarrow$ determine if simpler architectures than neural operators can be suitable
  - Determine to what extent other geophysical systems (volcanoes, sea currents, wind currents, ...) can be described by learned operators