

Olga Dorabiala

🏠 Seattle, WA ✉️ olgad400@uw.edu 🔗 OlgaD400 🌐 olga-dorabiala

Applied mathematician passionate about using machine learning to address challenging, real-world problems.

EDUCATION

University of Washington, Seattle Campus

2018 - 2023

PhD in Applied Mathematics, MS in Applied Mathematics

Pennsylvania State University, University Park

2014-2018

BS in Mathematics, Minors in Chemistry and German

PROFESSIONAL EXPERIENCE

The Alan Turing Institute

2023-present

Visiting Postdoctoral Researcher

- Designing and analyzing algorithms for unsupervised learning using an optimization framework
- Applying unsupervised learning methods to better characterize and understand patterns in real-world datasets

Disaster Simulation

- Exploring data-driven reduced order models of large scale agent-based systems
- Developing a simulation and visualization framework in Python for mathematically modeling the impact of natural disaster on the critical infrastructure and population of an urban area

Spatiotemporal Graph k-means (STGkM)

- Introduced a novel method for identifying multi-scale relationships in dynamic networks through vertex clustering
- Implemented STGkM in Rust and Python and demonstrated its utility on synthetic and real world dynamic networks

University of Washington

2018-2023

Graduate Research Assistant with Dr. Aleksandr Aravkin and Dr. Nathan Kutz

Robust Trimmed k-means (RTKM)

- Proposed a non-convex extension of the k-means objective function that allows outliers to be identified during the clustering process and provides flexibility for points to belong to multiple clusters
- Implemented RTKM in Python and showed that it achieves more consistent results across varying datasets compared with existing methods and outperforms other methods on multi-membership data containing outliers

Spatiotemporal k-means

- Proposed a two-phase spatio-temporal clustering method with a unified formulation over space and time that is able to identify both the temporary associations between moving objects as well as long-term, stable clusters
- Achieved an improvement in performance, particularly in the low-data domain, when compared against other st-clustering methods on a benchmark collective animal behavior dataset
- Applied STkM as part of a computer vision pipeline for Region of Interest (ROI) detection in videos

Ensemble Principal Component Analysis (EPCA)

- Developed a scalable, noise-resistant extension of PCA that lends itself naturally to uncertainty quantification
- Showed that EPCA performs competitively across noise domains, with a clear advantage on datasets containing outliers and orders of magnitude reduction in computational cost compared to Robust PCA

Stripe

Summer 2022

Machine Learning Engineer

- Developed an active and adaptive document identification pipeline for document verification
- Trained a computer vision model to generate low-dimensional embeddings of document images
- Clustered embeddings and validated model performance for open-set classification
- Utilized Apache AirFlow to execute document clustering as a daily task

Pacific Northwest National Laboratory

Summer 2021

Applied Mathematician

- Devised spatiotemporal clustering methods to track clusters of moving objects
- Summarized the movement of clusters and particle interactions using Kepler.gl and NetworkX

- Demonstrated the feasibility of non-speech audio geolocation by creating a labeled training dataset for supervised learning, securing funding for a permanent project
- Extracted non-speech segments from audio recordings stored in S3 buckets using librosa and pydub
- Trained an autoencoder using PyTorch on AWS to differentiate non-speech audio segments from one another and clustered the output to extract features relevant to geolocation

NLP

- Contributed to a text summarization pipeline running in Databricks
- Created a custom rule-based tokenizer in spaCy
- Aggregated the results of part of speech tagging, named entity recognition, and WordNet synsets to create an activity log

National Security Internship Program

Summer 2020

- Trained a Siamese Neural Network with PyTorch on biological signals to classify fake videos from the Deepfake Detection Challenge Dataset

PREPRINTS AND PUBLICATIONS

1. Olga Dorabiala, J. Nathan Kutz, and Aleksandr Y. Aravkin. “Robust trimmed k-means.” Pattern Recognition Letters (2022).
2. Olga Dorabiala, Jennifer Webster, J. Nathan Kutz, and Aleksandr Aravkin “Spatiotemporal k-means”, arxiv: 2211.05337 (2022)
3. Dabke Devarat and Olga Dorabiala “A Novel Method for Vertex Clustering in Dynamic Networks.” In R.M. Benito, C. Cherifi, H. Cherifi, E. Moro, L. M. Rocha, and M. Sales-Pardo (Eds.), *Complex Networks and Their Applications XII. Springer. In Preparation.*
4. Dorabiala, Olga, Aleksandr Aravkin, and J. Nathan Kutz. ”Ensemble Principal Component Analysis.” IEEE Access (2024).

EXTRACURRICULARS AND OUTREACH

Applied Mathematics Graduate Student Representative	2021-2022
Acted as a liason between students and faculty to represent the interests of the graduate student body	
Women in Applied Math Mentorship Program	2018-2021
Led three female undergraduate students in the exploration of a mathematical modeling problem	
Ran weekly research meetings and supervised preparation of final research presentations	
University of Washington SIAM Chapter Treasurer	2019-2021
Managed chapter budget and helped plan and run departmental and local outreach events	

AWARDS AND HONORS

Data Science Training Program	2020-2022
Fellowship offered jointly with Pacific Northwest National Laboratory and the University of Washington	
Boeing Fellowship	2018-2020
Two year fellowship awarded to select incoming doctoral students at the University of Washington Applied Mathematics department	
Achievement Rewards for College Scientists (ARCS) Fellowship	2018-2021
Three year fellowship sponsored by the ARCS Foundation given to outstanding STEM scholars at the University of Washington and Washington University	

SKILLS

Python (numpy, scipy, pandas, scikit-learn), clustering methods, PyTorch, Matlab, SQL, AWS, data preparation, machine learning, mathematical modeling and simulation