Mohammad Arvan

Machine Learning, Open Science, and Reproducibility

Education

2018-2023 (expected)

PhD, Computer Science; University of Illinois at Chicago (UIC), Chicago, IL

GPA: 4.0/4.0

Thesis title: Machine Leanring and Open Science: On Risks and Challenges

Summary: The principles of open science advocated by UNESCO have important implications for machine learning, and in particular, deep learning. However, the complexity of the field and the unique challenges posed by these methods have made reproducibility a significant concern. While current efforts have attempted to improve reproducibility through community-wide actions and checklists, we hypothesize that these measures are not sufficient to fully address the problem at hand. This thesis aims to investigate the impact of these actions and explore the extent to which the machine learning community has underestimated the challenges of achieving reproducibility in deep learning.

2012-2016

BSc, Software Engineering; Qazvin Islamic Azad University, Qazvin, Iran

Ranked 2nd among 180 students

Experience

Research Assistant at UIC Natrual Language Processing Laboratory:

- Explore ways of improving computational efficiency in training and inference of sequence processing neural networks in the language modeling task.
- Examine techniques for reducing the Time-to-Answer (TtA) of the Transformer-based neural networks on the `Mathematics Dataset,' a mathematical question answering dataset formulated as a character-level sequence to sequence problem.
- Implemented a neural reading comprehension model utilizing context-to-query and query-to-context attention on the `SQuAD 1.1' dataset.
- Implemented an optical character recognition with Convolutional filters and Conditional Random Fields (CNN-CRF) on `Tasker's OCR' dataset.
- Evaluated the performance of several machine learning algorithms such as Support Vector Machines (SVM), Random Forest, Logistic Regression, Naive Bayes, and Neural Networks on the MNIST and Fashion-MNIST datasets.

Data Engineer at Kara Intelligent System:

- Extracted, transformed, and loaded +100 GB data files into a data warehouse
- Created dashboard and visualization tools to present the analysis using Tableau, and Microsoft ASP.NET MVC.

- Developed an Android application for visualizing plots and charts
- Optimized the performance of the data warehouse by creating indexes and optimizing queries
- Designed a web application for visualizing the data using Angular

Co-Founder and Developer at Indooria Startup:

Indooria is a startup that provides indoor navigation and localization services for large indoor spaces such as shopping malls, airports, and hospitals. I was responsible for the development of the Android application.

- Developed an Android application for indoor navigation and localization using Bluetooth Low Energy (BLE) beacons.
- Implemented Djikstra's algorithm for finding the shortest path between two points on the map.
- Utilized Kalman filter to improve the accuracy of the localization.
- Explored outlier detection techniques to filter out invalid wireless signal strength values.
- Optimized map rendering engine libGDX which is based on OpenGL utilizing caching and batching techniques.

Research Assistant at Mechatronics Research Laboratory (MRL):

- Developed and evaluated a classifier using AdaBoost for detecting humans from a group of predefined geometry features computed from 2-dimensional range scans.
- Optimized code base to run on a edge device with limited computational resources.
- Worked on search algorithms, path planning (A*), motion planning (probabilistic roadmap), behavior planning, and obstacle avoidance of a 4-wheel steered mobile robot in indoor environments.
- Implemented an algorithm for detecting circle-shaped objects (barrels) from 2D range data.
- Designed and implemented an exploration strategy for search and rescue scenarios with the objective of maximum coverage while having minimum traveled distance by finding the shortest route for traveling all the nodes in the topological map of the environment. This problem is formulated as a Traveling Salesman Problem (TSP) and solved using a Particle Swarm Optimization (PSO) algorithm.
- Parallelized the code base of the PSO algorithm using OpenMP to improve the performance.

Software Engineer at Abara Fountain:

- Developed cross-platform desktop and embedded applications for the design and visualization of multimedia fountains using C++ and Qt framework
- Designed 2d fountain simulator using QGraphicView and QGraphicScene
- Developed a timeline editor for multimedia fountain shows using Qt framework
- Impleneted export/import, undo/redo, and copy/paste, batch editing, and other functionalities for the timeline editor and the fountain simulator
- Designed multi-threaded architecture for the timeline editor and the fountain simulator to improve the performance

Publications

- **Mohammad Arvan**, A. Seza Doğruöz, Natalie Parde. "Investigating Reproducibility at Interspeech Conferences:A Longitudinal and Comparative Perspective". The 24th INTERSPEECH Conference (INTERSPEECH 2023)
- Maja Popović, **Mohammad Arvan**, Natalie Parde, Anya Belz. "Exploring Variation of Results from

Different Experimental Conditions". The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023)

- **Mohammad Arvan**, Mina Valizadeh, Parian Haghighat, Toan Nguyen, Heejin Jeong, Natalie Parde. "Linguistic Cognitive Load Analysis on Dialogues with an Intelligent Virtual Assistant". The 45th Annual Meeting of the Cognitive Science Society (CogSci 2023)
- Anya Belza, Craig Thomson, Ehud Reiter, Gavin Abercrombie, Jose M. Alonso-Moral, **Mohammad Arvan**, Jackie Cheung, Mark Cieliebak, Elizabeth Clark, Kees van Deemter, Tanvi Dinkar, Ondrej Dušek, Steffen Eger, Qixiang Fang, Albert Gatt, Dimitra Gkatzia, Javier González-Corbelle, Dirk Hovy, Manuela Hürlimann, Takumi Ito, Emiel van Miltenburg, Chris van der Lee, John D. Kelleher, Filip Klubicka, Saad Mahamood, Margot Mieskes, Malvina Nissim, Natalie Parde, Ondrej Plátek, Verena Rieser, Pablo Mosteiro Romero, Joel Tetreault, Xiaojun Wan, Leo Wanner, Lewis Watson, Diyi Yang. "Missing Information, Unresponsive Authors, Experimental Flaws: The Impossibility of Assessing the Reproducibility of Previous Human Evaluations in NLP". The Forth Workshop on Insights from Negative Results in NLP (Insights 2023)
- **Mohammad Arvan**, Luís Pina, Natalie Parde. "Reproducibility in Computational Linguistics: Is Source Code Enough?". The 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)
- Parian Haghighat, Toan Nguyen, Mina Valizadeh, **Mohammad Arvan**, Natalie Parde, Myunghee Kim, Heejin Jeong. "Effects of an Intelligent Virtual Assistant on Office Task Performance and Workload in a Noisy Environment". Applied Ergonomics, 109, 103969.
- Parian Haghighat, Toan Nguyen, Mina Valizadeh, **Mohammad Arvan**, Natalie Parde, Myunghee Kim, and Heejin Jeong. "Human Interaction with Intelligent Virtual Assistant in a Noisy Environment". The 66th Proceedings of the Human Factors and Ergonomics Society Annual Meeting
- **Mohammad Arvan**, Luís Pina, Natalie Parde. "Reproducibility of *Exploring Neural Text Simplification Models*: A Review". The 15th International Natural Language Generation Conference (INLG 2022)
- Farshid Najafi, Mehdi Dadvar, Soheil Habibian, Alireza Hosseini, Hossein Haeri, **Mohammad Arvan**, Behzad Peykari, Hamed Bagheri."RoboCup Rescue 2016 Team Description Paper MRL". Robocup Rescue 2016 TDP Collection

Honors and Awards

- Ranked top 8% on Stackoverflow's 2020 year
- Recipient of 2020 Provost's Graduate Research Award ($5000)
- Graduate Assistantship, University of Illinois at Chicago (UIC), 2018-2023
- Undergraduate Assistantship, Qazvin Islamic Azad University, 2013-2016
- Ranked 3rd, RoboCup World Championship, Rescue Robot League in Nagoya, Japan, 2017
- Ranked 2nd, RoboCup World Championship, Rescue Robot League in Leipzig, Germany, 2016
- Innovative User Interface Award, RoboCup World Championship, Rescue Robot League in Hefei, China, 2015
- Ranked 1st, RoboCup World Championship, Rescue Robot League in Hefei, China, 2015
- Ranked 2nd, RoboCup World Championship, Rescue Robot League in João Pessoa, Brazil, 2014

Skills

Languages Python, C++, java, C#, SQL, Octave/MATLAB

Frameworks and Libraries	PyTorch, NumPy, Numba, TensorFlow, Keras, OpenCV, sklearn, SciPy, Pandas, Matplotlib, Spacy, Docker

Email: **mo.arvan@gmail.com**
Personal Website: **https://mo-arvan.github.io**
LinkedIn: `https://linkedin.com/in/mo-arvan`