Omid Setayeshfar | Resume

omid.s@uga.edu • omid-s.github.io • www.linkedin.com/in/omidsetayeshfar

I am a learner and a re-learner. I have more than 15 years of experience with different programming languages in many different fields, both in academic and industry settings. I have a track record of analyzing the problems, extracting data from *large scale* enterprise and industrial systems, and building *scalable* and *distributed* systems that will automate tasks or help make better decisions using artificial intelligence models; many of my works are deployed as web sites, web services or mobile apps. I have extensive experience with both classic artificial intelligence methods and modern machine learning and deep learning techniques for *optimization*, *simulation*, *natural language processing*, *image processing and computer vision*, and *industrial control*, on both CPU and GPU platforms. I am most happy in interdisciplinary teams, and I have served in many leadership positions with teams of different sizes.

Education

Department of Computer Science, University of Georgia — **Athens, GA**
Ph.D., GPA: 3.87 — *2015–2020(expected)*
Advisor: Dr. Kyu H. Lee, Dissertation: Enhanced Cyber Attack Investigation Via AI Aided Provenance Tracking

Shahid Chamran University of Ahvaz — **Ahvaz, Iran**
B.S., Top of My Class — *2007-2012*
Thesis: Online Robot Path Planning Using Fuzzy Logic

Skills

Database: Neo4j, Postgres, MS SQL Server, MySql

Programming Languages: Python, Java, C#, Objective-c

ML Packages: Keras, Tensorflow, Networkx, Scikit-Learn, nltk, gensim, pandas

Distributed & Parallel Computing: Apache Spark, CUDA, AWS, Google Cloud Platform

Techniques: (Un)Supervised Learning, RNN, CNN, LSTM, RF, LR, Clustering, Topic Modeling, Neural Networks, Bayesian Networks, Markovian Decision Models, K-Means, XGBoost, Hierarchical Clustering, DBScan, Language Translations (Seq2Seq), Database Design, Internet of Things, Edge Computing, Time Series, Swarm Intelligence, Regression, Classification, Prediction

Experience

The University of Georgia — **Athens, GA**
Research Assistant — *Aug 2015 - Present*
- Designed and developed artificial intelligence (deep learning, machine learning, NLP) models to solve issues in different fields, mainly internet of things and cyber-security.
- Done research on more than 5 government and industry funded research projects, resulting in 10 published and under publication papers.

NEC Labs America — **Princeton, NJ**
Research Assistant — *May 2017-Aug 2018*
- Research in the field of Systems Security, causality analysis, automated security, which have resulted in 2 patents and 1 research paper.
- Developed data pipelines for cyber-threat analysis on an internet-scale using scrapping, image, and natural language processing tools.
- Contributed to Automated Security and Intelligence (ASI) system developed by NEC Lab America

Samix Co. — **Tehran, Iran**
Department Lead — *Jun 2014 - Aug 2015*
- Got Fast-Tracked into a lead position for exceptional performance and leadership initiatives.
- Managed and assessed more than 10 national and international projects.
- As 'Mobile Solutions Committee Member' participated in Mobile Software business line development.

Samix Co. — **Tehran, Iran**
Software Developer/ Analyzer — *Nov 2013 - Aug 2015*
- Utilized data-mining and artificial intelligence models to turn enterprise data into helpful insights for multiple enterprises.
- Designed distributed and edge-based systems for road traffic monitoring, banking transaction summarizing, help desks, enterprise network health monitoring, and resource optimizations.
- Finished first place in company-wide periodical technical tests twice.

Aghajari Data Processing Company (ADPC) — **Ahvaz, Iran**
Software Engineer and architect (part-time) — *Nov 2008 - Jan 2012*
- Developed 2 of the total 7 modules on a drilling engineering software(DRESS) using C#, Java and Microsoft SQL Server.
- Won National Energy Ministry's distinguished project award and the governor's distinguished research-based company award for our works.

Select Projects

GrAALF
Finding what has caused a security breach is like finding a needle (malicious actions) in a haystack (billions of benign activities each day). In GrAALF, I used anomaly detection, graph embedding, and expert system methods as supervised learning models to find cyber-threads and their causality trace from multi-terabyte logs from multiple sources in realtime. The resulting graphs are stored in Graph and tabular databases.
Java, Python, Networkx, Neo4J, PostgreSQL, Distributed Computing

Active Deception using Factored Interactive POMDPs to Recognize Cyber Attacker's Intent (Funded by US Army)
Understanding an attacker's intention by observing their actions is challenging, just like understanding what people really mean by what they say. I built simulation platforms that enable the interaction of players for large scale experiments. I also contributed to developing Markovian Decision Models that deceive the attacker into an engagement to understand their intent.
Python, Partial Observable Markov Decision Processes, Java, Simulation, Virtualization

Data Provenance Infrastructure towards Robust and Reliable Data Sharing and Analytics (Funded By NSF)
Finding the source of a fault or malicious action in a world of connected internet of things is challenging; information travels through many layers and comes from heterogeneous device platforms. I use supervised and semi-supervised machine learning models to find inputs with foul data and trace them back to devices responsible with high accuracy. Also, I developed simulation platforms to support our large scale experiments with message passing protocols.
Python, Linear Regression, Neural Networks, Graph Embedding, Graph Search, Random Forest, SGD, Causality Analysis, Simulation, MQTT, Internet of Things

Banking Surplus System
Extracting high-level statistics from high-frequency data such as banking transactions is a lengthy task. Using cascading data filtering and data mining models pipeline, I calculated the profit/loss/surplus for each branch of an international bank with more than 3000 branches. Also, I created dashboards and visualizations to present these insights. Deployed since mid 2014, this system processes multiple Terabytes of data each week to extract the weekly reports.
SQL Server, Data Mining, C#, ASP.NET, MS SharePoint Server, Visualization

141.ir
An advanced nationwide data gathering and reporting system with crowdsourcing navigation system, route recommendation, road state reporting, and online hazard alert capabilities by Iranian Road Maintenance and Transportation Organization. I participated in algorithm designs for routing and data gathering and used data mining models to extract and filter information; I was also one of the system architects on this project. Deployed since 2013, this system has an average of 1000 concurrent users and peaks at several more. I also developed the iOS app (ranked in top 50 of navigation on iTunes with more than 100K active users)
Recommender Systems, Path Finding, Full-Stack, Web API, REST APIs, Objective-c, C#

Microsoft Malware Classification Challenge
Identifying malware families helps us understand how to defend against them, so much as developing vaccines needs identification of the disease family. However, the malware code might be a tiny portion of a program. Furthermore, it might be perturbed to avoid detection. I designed scalable classification models using, and ensemble of DeepLearning and Natural Language Processing techniques to find the family the malware in a program belongs to with more than 99% accuracy.
PySpark, Google Cloud Platform, MLlib, Convolutional Neural Networks (CNN), Bag of Word, nltk, tf-idf, LSTM, RNN

Landmark Classification
Tourist attractions and scenes seen in the real world have a lot of similarities. Identifying landmarks in a photo is a very challenging task for both human and the machine. I designed deep learning pipelines, including classical image processing and modern deep learning, to segment information from the images and classify them into one of 15K classes in Google Landmark Dataset of more than 1.4 million images.
OpenCV, Keras, Tensorflow, CNN, Segmentation, GPU Computing, Computer Vision, Image Processing

Select Publications

See my Scholar page (`https://scholar.google.com/citations?user=gDwPmR0AAAAJ`) ask me for other submitted and in progress papers.

Select Awards

2016-19: Ranked 3rd and 2nd, 2nd, 2nd in Poster Session, UGA CS-AI Day

2018: Runner up, Home Depot Code Athlon, Atlanta, GA