KMA Solaiman

Contact Information

Ph.D. Candidate
Department of Computer Science
Purdue University
West Lafayette, IN 47907

B132 #5 Lawson Computer Science Building
+1 765-775-8230
https://ksolaiman.github.io/
ksolaima@purdue.edu

Research Interests

Multi-modal Information Retrieval, Heterogeneous Data Mining, Novelties in Machine Learning, and Data Management Systems.

Education

Purdue University
Ph.D. in Computer Science — Fall 2016 - Spring 2023
M.Sc. in Computer Science — Fall 2022
Advisor: Bharat K. Bhargava

Bangladesh University of Engineering and Technology (BUET) — July 2014
B.Sc. in Computer Science and Engineering (CSE)
GPA: **3.79/4.00** (Ranked 16th in a class of 153 students)

Technical Skills

⋆ **Industry Knowledge:** Machine Learning, Data Science, Data Analysis, NLP, Computer Vision, Reinforcement Learning, Deep Learning, Graph Representation Learning

⋆ **Tools and Technologies:**

- **Software Stacks**: Docker, Kafka, Conda, Git (Version Control), Amazon Web Services (AWS)
- **Programming Language**: Python, Java, C, C++, 80x86 Assembly, Prolog, PHP, HTML, XML, CSS, JavaScript, Linux Shell Programming
- **Deep Learning Frameworks**: PyTorch (*High*), Caffe, Matlab (*Moderate*)
- **Database Systems**: Oracle, Microsoft SQL Server, MySQL, PL/SQL
- **Tools and Software**: CSim, Matlab, UML, ERD, Weka, Ajax, JQuery, Lex, Yacc, OpenGL, LaTeX, Google API, Jupyter Notebook
- **Operating Systems**: Linux, Unix, Windows

Collaborations

Massachusetts Institute of Technology (MIT), University of Southern California (USC-ISI), Northrop Grumman Corporation (NGC), Institute for Defense Analyses (IDA)

Research Experience

Research in Applications for Learning Machines (REALM).
Bharat K. Bhargava, Michael Stonebraker
We focused on heterogeneous data discovery for real-time contextual awareness. Collaborating with local authorities, we worked on finding a real-time scalable solution for *missing person search* where we worked with real-world noisy and high dimensional data collected by the agencies.

- Designed a *scalable cross-modal querying method* based on relational schema
- Proposed two *weakly supervised methods* for *cross-modal information retrieval*
- Proposed a novel *human attribute recognition model from unstructured text* and benchmarked attribute recognition models for video and image.

Science of Artificial Intelligence and Learning for Open-world Novelty (SAIL-ON).

- Worked on generating a graph state representation for a reinforcement learning agent for Monopoly
- Proposed empirical methods to estimate novelty difficulty in Monopoly
- Studied Prisoners' Dilemma to propose a theory of novelty
- Proposed intrinsic complexity measures for perception domains such as MNIST.

Understanding Political Bias in News Articles using Social Media. *(Dan Goldwasser)*

- Designed representations for newspaper articles using Contextualized Language Models
- Designed joint representations for text and social information
- Text classification for political bias with vanilla and neural network based classifiers
- Building weakly supervised models for collective classification using Probabilistic Soft Logic

Adversarial Attacks on Neural Networks. *(Miguel Villarreal Vasquez)* Spring 2019
Ran experiments to tackle the VGG-FACE deep neural network trojaning attack by introducing a healing dataset from Wild dataset and retraining it. Original attack pollutes the model by trojan trigger generation and retraining the model with reversed engineered training data.

Unsupervised Learning *(Advisor:* Md. Monirul Islam*)* Feb 2013 - Jul 2014
• Proposed a novel clustering algorithm for irregular and complex shaped data with only a single variable parameter, *filter-width* and described an empirical method to dynamically find optimal value of that parameter.
• Developed extension to Weka platform and showed performance comparison of our algorithm with K-means, EM, Cobweb & Classit, FarthestFirst, DBSCAN & Hierarchical Clustering algorithm across data sets like Aggregation, Compound, Jain, Spiral, and others.

Analysis & Visualization of Road Accident Data Oct 2011 - May 2013
• Implemented a novel web interface for collection of road accident data in Bangladesh and for performing dynamic data analysis of road accidents with contributing factors using Google APIs.

Data Mining, Complex Network Analysis
• Developed methods for finding hand-gesture patterns in smart watch sensor data using Recurrent Neural Networks, specifically LSTM. (*Supervisor: He Wang*) Sep 2017 - Oct 2017
• Investigated *TribeFlow* for mining and predicting user preferences using website hyperlink structure in Wikipedia. Jan 2017 - Jul 2017

Publications

• *(Conference)* Kma Solaiman and B. Bhargava, Multi-modal Information Retrieval for Systems with Explicit Information Needs and Object Properties (FemmIR), *Submitted in* ***SIGMOD 2023***.

• *(Journal)* Kma Solaiman, Tao Sun, Alina Nesen, Bharat Bhargava, and Michael Stonebraker. Applying Machine Learning and Data Fusion to the *Missing Person* Problem. ***IEEE Computer****, (Volume: 55, Issue: 6, June 2022).*

• *(Symposium)* K. Solaiman & B. Bhargava, Open-Learning Framework for Multi-modal Information Retrieval with Weakly Supervised Joint Embedding. ***AAAI Spring Symposium****, 2022.*

• *(Symposium)* K. Solaiman & B. Bhargava, Measurement of Novelty Difficulty in Monopoly. ***AAAI Spring Symposium****, March 2022.* [Link]

• *(Conference)* A. Nesen, KMA Solaiman and B. Bhargava, Dataset Augmentation with Generated Novelties, *International Conference on Transdisciplinary AI (****TransAI****), IEEE, 2021.*

• *(Workshop)* M. Stonebraker et al, Surveillance Video Querying With A Human-in-the-Loop, *Workshop on Human-In-the-Loop Data Analytics* ***(HILDA)*** *with SIGMOD, 2020.*

• *(Workshop)* S. Palacios and K. Solaiman**, P. Angin, A. Nesen, B. Bhargava, Z. Collins, A. Sipser, M. Stonebraker, SKOD: A Framework for Situational Knowledge on Demand, In ***POLY****, co-located with VLDB, Springer 2019.* ****Co-first authors.**

• *(Conference)* S Roy, K Solaiman, C Li, D Goldwasser, Identifying Bias in News Narratives Using Distant Supervision, **(IWCS)**, 2019. [Submitted Work]

• *(Conference)* Kma Solaiman, MM Rahman, and N Shahriar, AVRA BANGLADESH: Collection, Analysis & Visualization of Road Accident Data in Bangladesh, **ICIEV**, *IEEE 2013.*

• *(Poster)* Kma Solaiman, AA Muzaddid, Minimal Parameter Clustering of Complex Shape Dataset with High Dimensional Dataset Compatibility, *Presented in CSE, BUET, 2014.*

• *(Undergraduate Dissertation)* Minimal Parameter Clustering of Complex Shaped & Different Sized Dataset with Noise & Outlier Handling (MPCACS).

Professional Experience

Research Assistant Spring 2020 - Present
Purdue University, Supervisor: Bharat K. Bhargava

Graduate Teaching Assistant
Computer Science, Purdue University
∗ *CS18000 : Problem Solving and Object Oriented Programming* Fall 2016 - Fall 2017
∗ *CS25100 : Data Structures* Spring 2017 - Fall 2018
∗ *CS54300 : Introduction To Simulation & Modeling Of Computer Systems* Spring 2019
∗ *CS44800 : Introduction To Relational Database Systems* Fall 2019
∗ *CS53600 : Data Communication and Computer Networks* Fall 2022

Lecturer, *Computer Science & Engineering,*
Ahsanullah University of Science & Technology Oct 2014 - Jul 2016

Academic Service

Program Committee Member / Reviewer #papers reviewed
∗ *ECML/PKDD 2022, 2021* 4 papers, 1 paper
∗ *IEEE PIMRC 2019* 2 papers

Course Projects

• Wikicapraim: Replica of Wikipedia to collect "user navigation traces". `MediaWiki`
• HOBORODH: A run & dodge game with a racing car modeled after Lamborghini Aventador and designed to avoid molotov cocktails and bombs. `OpenGL`
• Converting *Modular Accident Analysis Program* data to database query formats `C++`
• Predicted user ratings for films using Collaborative Filtering `Python`
• Personalized Daily Activity Recognition by analyzing data collected from accelerometer and gyroscope sensors using SVM & Logistic Regression `Matlab`
• Inference of Attributes from Crowd-sourced Annotations for Fake News Detection `Python`
• Diagnosis of cancer using Decision Tree Learning and Ada-boost ensemble classifier; Text classification using K-Nearest Neighbor and Naive Bayes. `Python`
• Mega Structure Modeling - *Himeji Castle*, Lighting and Texture, Ray Tracing `Graphics`
• Collision Avoiding Robot using E-puck
• Implementation of Transport, Network and DLL of OSI Layers `Java/C`
• 4 bit pipe-lined microprocessor, capable of executing 28 specific instructions `Quartus`
• A content management system (CMS) for college websites `PHP` `MySQL` `CodeIgniter`
• Thread synchronization, scheduling, multiprogramming, caching, process and virtual memory management of NACHOS `Java`
• Implementation of a Compiler (Symbol-Table, Lexical Analyzer, a Parser and an Intermediate Code Generator) for a subset of Pascal `Lex` `YACC` `C++`
• Four in a Row: A LAN multi-player game using Socket Programming `Java`
• Hangman: Word guessing game with enhanced audio & visual interface `C`

Relevant Courses

Crowd-sourcing and Social Computing, AI meets Sustainability, **Machine Learning, Data Mining, Pattern Recognition, Artificial Intelligence, Database**, **Distributed Database Systems, Compiling And Programming Systems**, Operating Systems, **Data Communication and Computer Networks**, Data Structure, Algorithms, Object Oriented and Structured Programming Language

---

Honors and Awards

2018	Graduate School Summer Research Grant
2009-2012	University Merit Scholarship
2008-2011	University Stipend
2009-2011	Dean's List Award
2006	Dhaka Education Board Scholarship

Reference

Bharat K Bhargava
Professor, Purdue University Fall 1984 - Present

Last updated on November 14, 2022.