Arun Sharma

CONTACT INFORMATION	Computer Science Department 200 Union St SE Minneapolis, MN 55455	Phone: 6129465123 Linkedin: https://www.linkedin.com/in/arun08sharma/ Email: sharm485@umn.edu

RESEARCH INTEREST	Responsible AI, Machine Learning, Data Mining, Algorithms, Optimization, Natural Language Processing, Computer Vision, Database Systems, Distributed Systems

EDUCATION	**University of Minnesota**, Twin Cities *Ph.D. Candidate in Computer Science* *Advised by Prof. Shashi Shekhar*	September 2018 - May 2024
	University at Buffalo, New York *MS in Computer Science*	September 2016 - May 2018

APPOINTMENTS	**University of Minnesota** Twin Cities, MN *Graduate Research Assistant*	Jan 2020 - Present
	University of Minnesota Twin Cities, MN *Graduate Teaching Assistant*	September 2018 - December 2019
	NASA Ames Research Center Mountain View, CA *Graduate Project Assistant*	May 2017 - August 2017

TECHNICAL SKILLS
- **Languages:** Java, Python, R, SQL, Scala, Android.
- **Machine Learning:** PyTorch, Keras, TensorFlow, OpenCV.
- **Big Data:** Apache Spark, Hadoop, Hive, Pig, HDFS, Kafka, PySpark.
- **DevOps:** Jira, Git, Agile Methodologies, Scrum and Sprint planning, Docker, Tableau.

WORK EXPERIENCE

Graduate Research Assistant, UMN, Twin Cities — Jan 2020 - Present
Project: Identifying Aberration Patterns in Multi-attribute Trajectory Gaps
Affiliation: *National Geospatial-Intelligence Agency (US DOD)*
- Designed efficient algorithms using spatial data science methods within trajectory gaps used in mobility pattern mining (e.g., rendezvous).
- Explored physics-aware parameters to capture abnormal behavioral mobility patterns and applied in various application domain (e.g., maritime space, road networks etc.).
- Currently exploring machine learning and probabilistic methods to capture space-time uncertainty within trajectory data gaps.

Project: Understanding COVID-19 Effects on Mobility Patterns
Affiliation: *Minnesota Department Of Transportation, Minnesota Management and Budget*
- Designed a COVID-19 decision support system, based on close collaboration with end-users, who provided the system requirements such based on queries of interest (e.g., business categories with many long-duration visits, sampling bias, data privacy etc.)
- Designed Human-in-the-loop feedback system to improve understanding of the aggregated mobile device data and facilitates a richer set of queries.
- Report the user feedback on summary reports and visualizations generated by our system leveraging explainable interpretation posed by end-users.

Graduate Project Assistant, SUNY Buffalo, New York — May 2017 - August 2017
Project: WebGlobe - A cloud-based geospatial analysis framework for climate data
Affiliation: *NASA Ames Research Center*
- Designed end-to-end framework to understand climate change's impact using scientific data (NetCDF), Apache Spark, and Hadoop Distributed File System.

- Implemented Gaussian process (GP) based change detection algorithm events to detect extreme hot events (excessive heat) and extreme cold events (winter storm).
- Proposed efficient algorithm and reduce GP regression models from $O(N^3)$ to $O(N^2)$ via matrix factorization techniques and scaled on Apache Spark framework.
- Provided geographic visualization and user-friendly UX/UI to aid domain interpretation and provided reproducible Jupyter Notebooks for the data scientists.

RELEVANT PROJECTS

Project: Implementation over Machine Learning and Computer Vision Fundamentals
Courses: *Machine Learning and Computer Vision*
- Scratch implementations of machine learning fundamentals such as regression model and classification models (e.g., MLP BPNN, DT, Ensemble Methods), Probabilistic Models (e.g., Naive Bayes) and clustering based methods (e.g., Kmeans, DBSCAN, HAC)
- Scratch implementations of Computer Vision techniques (e.g., edge detection methods, Fourier Analysis, Laplacian Pyramid, Region Merging etc.)
- Implemented disparity estimation from rectified images using basic block matching and dynamic programming (Longest Common Subsequence) to perceive depth in Stereo Vision.

Project: Implementation of Deep Learning and Optimization Concepts
Courses: *Deep Learning and Non-Linear Optimization*
- Formulate a sparse non-negative matrix factorization (NMF) problem, called co-sparse non-negative matrix factorization (cosNMF) for exploring the sparseness of scRNA-seq data. We computed the factor matrices while improving scalability.
- Implemented RNN with LSTMs on on Imdb movie reviews dataset to perform sentiment analysis and determine if a review has a positive or negative polarity. The model is further fine-tuned using Adam Optimizer.
- Trained unsupervised models, Autoencoder and Restricted Boltzmann Machine, to classify fraudulent v/s non-fraudulent transactions using Autoencoder and RBM.

Project: Distributed Key-Value Storage using Amazon Dynamo
Course: *Distributed Systems*
- Implemented Group Messenger with Total and FIFO Ordering Guarantees. Passed different test cases based on a total-FIFO ordering of messaging with or without failures of message passing nodes.
- Design a simple DHT based on Chord via implementing 1) ID space partitioning/re-partitioning, 2) Ring-based routing, and 3) Node joins. Passed different point and range queries in testing phases with 5 message passing nodes.
- Implemented a simplified version of Dynamo-style key-value storage via three main pieces 1) Partitioning, 2) Replication, and 3) Failure handling. Passed all test phases based on concurrency control and failure handling of message passing nodes.

Project: Design and implement voting algorithms for US elections (Team of 5)
Course: *Software Engineering*
- Created a voting system that is capable of performing both the STV with droop quota and Plurality (popularity) algorithms using Waterfall and Agile Scrum model.
- Conducted state-of-the-art software engineering practices (e.g., Requirements, Design, Implementation, Testing etc.) with documentation and unit testing.
- Conducted detailed unit and system testing under state-of-the-art AGILE practices.

SELECTED PUBLICATIONS

[1] **Analyzing Trajectory Gaps for Possible Rendezvous Regions**
ACM Transactions in Intelligent Systems and Technology, 2022
Arun Sharma and Shashi Shekhar

[2] **Towards a Tighter Bound on Possible-Rendezvous Areas: Preliminary Results**
30th International Conference on Advances in Geographic Information Systems, 2022
Arun Sharma, Jayant Gupta, Subhankar Ghosh, and Shashi Shekhar

[4] **Abnormal Trajectory-Gap Detection: A Summary (Short Paper)**
15th International Conference on Spatial Information Theory (COSIT 2022)
Arun Sharma, Jayant Gupta, and Shashi Shekhar

[3] **Mining Taxonomy-aware Colocations: A Summary of Results**
30th International Conference on Advances in Geographic Information Systems, 2022
Jayant Gupta, **Arun Sharma**, and Shashi Shekhar

[5] **Understanding COVID-19 Effects on Mobility: A Community-Engaged Approach**
AGILE: GIScience Series (2022)
Arun Sharma, M Farhadloo, Yan Li, Aditya Kulkarni, Jayant Gupta, and Shashi Shekhar

[6] **Analyzing Trajectory Gaps for Possible Rendezvous: A Summary of Results**
30th International Conference on Advances in Geographic Information Systems, 2022
Arun Sharma, Xun Tang, Jayant Gupta, Majid Farhadloo and Shashi Shekhar

[6] **Towards geographically robust statistically significant regional colocations**
11th *International Conference in Geographic Information Science, 2021 (Oral)*
Subhankar Ghosh, Jayant Gupta, **Arun Sharma**, Shuai An, Shashi Shekhar

[7] **Spatiotemporal Data Mining: A Survey**
Handbook of Spatial Analysis for the Social Sciences, Edward Elgar, 2022
Arun Sharma, Zhe Jiang, and Shashi Shekhar

[8] **WebGlobe: A cloud-based framework for interacting with climate data**
International Workshop on Analytics for Big Geospatial Data (SIGSPATIAL) 2018
Arun Sharma, SM Arshad Zaidi, Varun Chandola, Melissa R Dumas, Budhendra L Bhaduri

TEACHING EXPERIENCE	**CSCI 5715 Spatial Data Science** **Fall 2019**

TEACHING EXPERIENCE

CSCI 5715 Spatial Data Science   **Fall 2019**
Graduate Student Instructor
- Guest Lecturer on topics: Spatial Networks and Spatial Data Mining
- Involved in designing homework assignment, labs, exams, lecture slides etc.

CSCI 5708 Advanced Database Systems   **Spring 2019**
Graduate Student Instructor
- Guest Lecturer on topics: Concurrency Control, Database Security and Data Mining.
- Involved in designing homework assignment, labs, exams, lecture slides etc.

CSCI 4041 Data Structures and Algorithms   **Fall 2018**
Graduate Teaching Assistant
- Instructed and proctored weekly recitation sessions with over 40+ students.
- Involved in grading of 400+ students and testing weekly assignments and labs.

HONORS AND ACHIEVEMENTS

Doctoral Dissertation Fellowship   **2022-2023**
University of Minnesota, Twin Cities

Travel Grant   **2019-2022**
SIGKDD 2019, NeurIPS 2020-2021, ICML 2021, AGILE 2022-2023, SIGSPATIAL 2022

SERVICES AND LEADERSHIP

Monitoring COVID-19 for Minnesota Management and Budget   **2020-2021**
Reporting State-Level Mobility Traffic to Research Scientists and Policymakers
- Published periodic mobility reports for informed decision making by domain experts.
- Presented periodic updates on mobility patterns during COVID-19 pandemic.
- Involved and advised high school students who are considering a career in research.

Reviewer
SIGSPATIAL, SIGKDD, ICDM, SSTD, AAAI, AGILE, Geoinformatica