Kanchan Chowdhury

Email : kchowdh1@asu.edu
Webpage: kanchanchy.github.io
Mobile : +1-480-410-8677
Linkedin: linkedin.com/in/kanchan-chowdhury-5729699a

Research Interests

Machine Learning, Big Geospatial Data Analytics, and Database Systems

Technical Skills

- **Languages & Databases**: Python, Java, C, Scala, SQL, MySQL, PostgreSQL, and SparkSQL
- **Others**: Apache Spark, Apache Sedona, GeoPandas, PyTorch, Scikit-learn, AWS and Databricks cloud, Spatial and Statistical Data Analysis, Machine Learning and Deep Learning, Pandas, Numpy, Git, Rest API, Android SDK, IntelliJ IDEA, Eclipse, and Data Visualization with Matplotlib, Seaborn, and Plotly Express

Education

- **Arizona State University** — Tempe, Arizona
  MS (completed) and PhD (ongoing) in Computer Science — *Aug. 2018 – Present*
  CGPA: 4.00, Advisor: Prof. Mohamed Sarwat, Co-advisor: Prof. Jia Zou
- **Chittagong University of Engineering and Technology** — Chittagong, Bangladesh
  Bachelor of Science in Computer Science and Engineering — *Mar. 2010 – Nov. 2014*
  Advisor: Prof. Mohammed Moshiul Hoque, CGPA: 3.76

Experience

- **Wherobots Inc.** — Scottsdale, Arizona, USA
  Research and Development Intern — *Jan. 2023 - Aug. 2023*
  - **Responsibilities**: Developing spatial machine learning and deep learning tools, scalable map-matching and geospatial data analytical algorithms. Integrating the developed tools into Wherobots cloud platform.
- **Data Systems Lab, Arizona State University** — Tempe, Arizona
  Research Assistant — *Aug. 2018 - December 2022*
  - **Current Research Project**:
    When features in a machine learning inference application come from multiple datasets, joining the datasets to create feature vectors becomes a bottleneck before performing the actual inference. In my current research project, I am working on decomposing models into multiple parts and pushing the join operation down the decomposed models in a way that reduces the runtime and memory consumption of the overall flow.
  - **Past Research Projects**:
    1) Designing and implementing a deep learning framework on top of PyTorch for raster imagery and spatiotemporal non-imagery vector datasets.
    2) Re-partitioning training data of a spatial model to reduce training time and memory usage.
    3) Evaluating state-of-the-art approaches for the synthesis of SQL queries from natural language questions.
    4) Contributing to a geospatial entity matching framework and a satellite image labeling framework.
- **Arizona State University** — Tempe, Arizona
  Teaching Assistant — *Aug. 2018 - December 2022*
  - **Instructor**: Served as an instructor for the course Spatial Data Science and Engineering (CSE 594) in Fall 2022. Responsibilities included preparing lecture slides and conducting lectures, preparing assignments, projects, exam questions and grading rubrics, and helping students understand lectures and projects.
  - **Teaching Assistant**: Served as a Teaching Assistant for the following courses from Fall 2018 to Spring 2022: Distributed Database Systems, Data Processing at Scale, Object-Oriented Programming & Data Structure, Principles of Programming with C++, Principles of Programming with Java & Python
- **Gagagugu PTE LTD** — Dhaka, Bangladesh
  Software Engineer — *Jan. 2017 - Jun. 2018*

- **Responsibilities**: Developing Android Apps with social networking features such as calling, messaging, and posting.

- **Le Chef Plc** Dhaka, Bangladesh
  Android Application Developer *Jan. 2015 - Dec. 2016*
  - **Responsibilities**: Developing Android Apps featuring online order and reservation services for restaurants in UK.

Publications

- **Kanchan Chowdhury**, Mohamed Sarwat; A Demonstration of GeoTorchAI: A Spatiotemporal Deep Learning Framework. *ACM SIGMOD International Conference on Management of Data*, 2023
- **Kanchan Chowdhury**, Mohamed Sarwat; GeoTorch: A Spatiotemporal Deep Learning Framework. *30th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '22)*, 2022
- **Kanchan Chowdhury**, Vamsi Meduri, Mohamed Sarwat; A Machine Learning-Aware Data Re-partitioning Framework for Spatial Datasets. *38th International Conference on Data Engineering (ICDE)*, 2022
- Vamsi Meduri, **Kanchan Chowdhury**, Mohamed Sarwat; Evaluation of Machine Learning Algorithms in Predicting the Next SQL Query From the Future. *ACM Transactions on Database Systems (TODS)*, 2021
- Jia Yu, **Kanchan Chowdhury**, Mohamed Sarwat; Tabula in Action: A Sampling Middleware for Interactive Geospatial Visualization dashboards. *46th International Conference on Very Large Databases (VLDB)*, 2020.
- Vamsi Meduri, **Kanchan Chowdhury**, Mohamed Sarwat; Recurrent Neural Networks for Dynamic User Intent Prediction in Human-Database Interaction. *22nd International Conference on EDBT, 2019*
- **Kanchan Chowdhury**, Lamia Alam, Shyla Sarmin, Safayet Arefin, Mohammed Moshiul Hoque; A Fuzzy Features Based Online Handwritten Bangla Word Recognition Framework. *18th ICCIT*, 2015

Projects

- **GeoTorchAI**: A spatiotemporal deep learning and data processing framework on top of PyTorch and Apache Sedona, formerly known as **GeoTorch**. It enables spatiotemporal machine learning practitioners to easily and efficiently implement deep learning models targeting the applications of raster imagery datasets and spatiotemporal non-imagery datasets, besides supporting scalable data preprocessing.
- **ML Aware Spatial Data Re-partitioning**: This framework aims at reducing the training time and memory usage of a spatial machine learning model by reducing the number of partitions in a spatial grid dataset. Experiments on four datasets achieved significant reduction in training time and memory consumption while bounding the difference in prediction error within 5%.
- **Named Entity Recognition**: This work tunes various steps of state-of-the-art methods for named entity recognition in order to experiment the changes in performance. Evaluation is done with two popular datasets: CoNLL-2003 and OntoNotes-5.0.
- **Hotspot Analysis on Apache Sedona**: This work performs spatial queries and range joins between two spatial datasets and calculates Getis-Ord statistic of NYC Taxi Trip dataset to perform hot-cell analysis.
- **Climate Change Forecasting**: A data science project to perform data cleaning, feature engineering, and data preprocessing operations on raw temperature data and predict temperature trend with LSTM model.
- **NLIDB-Bench**: A benchmark for evaluating state-of-the-art approaches of SQL query generation from natural language text queries. Besides proposing a set of evaluation metrics, we evaluate all approaches with three state-of-the-art datasets along with our own proposed dataset.
- **Fake News Detection**: A data science project to perform data cleaning, feature engineering, and data preprocessing operations on news datasets and classifying fake and real news with Bidirectional LSTM model.

Participation and Awards

- Recipient of ACM SIGMOD 2023 student travel award to attend the conference and present a paper.
- Recipient of ACM SIGSPATIAL 2022 travel award to attend the conference and present a paper.
- Recipient of Arizona State University Graduate and Professional Student Association (GPSA) travel grant to attend ACM SIGSPATIAL 2022 and present a paper.
- Recipient of CIDSE Doctoral Fellowship at Arizona State University for the academic year 2018-2019.
- 2nd Runner-up at National Hackathon organized by ICT Division of Bangladesh in 2014. The challenge of the hackathon was to design a project based solution to solve a national problem of the country.
- 2nd Runner-up at National Mobile Application Code Hub organized by BUET, Bangladesh in 2014.
- 6th at Inter University Programming Contest organized by CUET, Bangladesh in 2012.
- Recipient of Honors award from my undergraduate university for maintaining academic excellence.

Additional Services

- **Paper Reviewer**: Served as a reviewer for the conference IEEE/ACM ICCAD 2023 and reviewed three papers.
- **Secondary Paper Reviewer**: Reviewing papers as a secondary reviewer of my Ph.D. supervisor for the conferences ACM SIGMOD, ACM VLDB, ACM SIGSPATIAL, and IEEE ICDE since 2018.
- **Grant Reviewer**: Served as a travel grant reviewer for the Graduate and Professional Student Association at Arizona State University from May 2022 to August 2023.
- **Conference Volunteer**: Volunteered to organize two conferences - ACM SIGSPATIAL 2022 and ACM SIGMOD 2023.
- **Open Source Contribution**: Contributed to Apache Sedona, an open-source geospatial cluster computing framework, by adding support for two new spatial data types. I am also serving as a PPMC member of Apache Sedona.

CourseWorks

Distributed Database Systems, Statistical Machine Learning, Fundamentals of Statistical Learning, Data mining, Semantic Web Mining, Programming in C/C++, Object Oriented programming, Data Structure, Algorithms, Discrete Mathematics, Artificial Intelligence, Software Engineering, Operating Systems, Database Management Systems, Social Media Mining, Big Data Analysis with Scala and Spark, Deep Neural Networks with PyTorch, and Effective Programming in Scala.