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, Geospatial Data Analytics, and Big Data Systems

Education

- **Arizona State University** — Tempe, Arizona
  PhD in Computer Science (CGPA: 4.00) — *Aug. 2018 – Jul. 2023 (Expected)*
  Advisor: Prof. Mohamed Sarwat

- **Chittagong University of Engineering and Technology** — Chittagong, Bangladesh
  Bachelor of Science in Computer Science and Engineering (CGPA: 3.76) — *Mar. 2010 – Nov. 2014*
  Advisor: Prof. Mohammed Moshiul Hoque

Experience

- **Data Systems Lab, Arizona State University** — Tempe, Arizona
  Research Assistant — *Aug. 2018 - Present*
  - **Spatio-Temporal ML Framework**: Developing a framework consisting of both data preprocessing and machine learning modules to support easy and efficient integration of spatio-temporal data into machine learning models.
  - **Spatial Data Repartitioning**: Reducing model training time by ML-aware repartitioning of spatial data.
  - **NLIDB-Bench**: Developing a benchmark for evaluating approaches of SQL synthesis from natural language query.

- **Arizona State University** — Tempe, Arizona
  Teaching Assistant — *Aug. 2018 - Present*
  - **Courses Taught**: Distributed Database Systems (Fall 2021, Spring 2021, Fall 2020, Spring 2020), Data Processing at Scale (Fall 2019), Object Oriented Programming & Data Structure (Spring 2019), Principles of Programming with C++ (Spring 2019), Principles of Programming with Java & Python (Fall 2018)

- **Gagagugu PTE LTD** — Dhaka, Bangladesh
  Software Engineer — *Jan. 2017 - Jun. 2018*
  - **Responsibilities**: Developing Android Applications featuring functionalities related to social networking such as calling, messaging, and post sharing.

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

Publications

- 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*, 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

Technical Skills

- **Languages & Databases**: Python, Java, C, Scala, SQL, HTML, MySQL, and SparkSQL
- **Others**: Apache Spark, Apache Sedona, PyTorch, Scikit-learn, Jupyter Notebook, Spatial and Statistical Data Analysis, Machine Learning and Deep Learning, Reinforcement Learning, Data Distributions, Sampling, Sequence Models, Pandas, Numpy, Matplotlib, Git, Rest API, and Android SDK

Projects

- **Spatio-Temporal ML Framework**: A framework consisting of two modules: 1) data preprocessing module on top of Apache Sedona to support all preprocessing of spatial and spatio-temporal data from data loading to tensor generation and 2) machine learning module on top of PyTorch to support easy implementation of custom spatio-temporal models.
- **Spatial Data Repartitioning**: A framework built on Apache Sedona for efficient re-partitioning of spatial data with an end objective to reducing training time and memory usage of spatial machine learning models. It can reduce the training time significantly without major impact on model accuracy.
- **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.
- **Hotspot Analysis on Apache Sedona**: This work writes functions in SparkSQL to perform spatial queries, performs range join between two spatial datasets, and calculates Getis-Ord statistic of NYC Taxi Trip dataset to perform hot-cell analysis.
- **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.
- **Hyper-parameter Optimization**: Comparing state-of-the-art approaches for searching the optimum hyper-parameters of a deep learning model, such as grid search and Bayesian optimization.

Projects with Collaboration

- **AI for Human-Database Interaction**: Comparing state-of-the-art machine learning techniques for SQL query prediction during an interaction session between a user and a DBMS system. Compared techniques include reinforcement learning, Q-learning, and sequence models.
- **Tabula**: A sampling middleware based on materialized sampling cube that sits between the data system and the geospatial visualization dashboard to accelerate the interactive visual analysis.

Participation and Awards

- 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 undergraduate University for maintaining good CGPA.

CourseWorks

Statistical Machine Learning, Fundamentals of Statistical Learning, Distributed Database Systems, Data mining, Semantic Web Mining, Social Media Mining, Programming in C/C++, Object Oriented programming in Java, Data Structure, Algorithms, Discrete Mathematics, Artificial Intelligence, Software Engineering, Big Data Analysis with Scala and Spark, Natural Language Processing, Deep Neural Networks with PyTorch, and Python for Data Science, AI & Development.