Tianyi Sun

Tel: 612-309-0898 | E-mail: tianyisun@uchicago.edu | Web: https://tianyisuntt.github.io/

EDUCATION

University of Chicago **Sep. 2021 – Now**

M.S. in Computational and Applied Mathematics

University of Minnesota, Twin - Cities **Aug. 2018 – May. 2021**

B.A. in Mathematics (Computer Specialization), Minor in Statistics and Computer Science GPA: 3.73/4.00

- Relevant coursework: Linear Algebra & Differential Equations, Multivariable Calculus, Intro to College Physicals, Applied Linear Algebra, Regression and Correlated Data, Discrete Structures, Algorithms and Data Structures, Formal Languages and Automata Theory, Basic Theory of Probability and Statistics, Numerical Method I & II, Cryptology and Number Theory, Mathematical Logic I, Theory of Statistics II, Intro to Artificial Intelligence, Intro to Machine Learning.

Central University of Finance and Economics, Beijing **Sep. 2016 – Jun. 2018**

Mathematics and Economics GPA: 3.84/4.00 (top5%)

- Relevant coursework: Linear Algebra, Calculus I II, Fundamentals of Statistics, Business Statistics, Macroeconomics, Microeconomics, Fundamentals of Accounting, Public Finance, Marketing Management, Psychology, and Sociology.

EXPERIENCES

- UChicago Environmental Data Science Bootcamps (NSF supported), Selected Student Aug. 2021 – Oct. 2021
- University of Notre Dame HAI Lab, Research Intern Jun. 2021 – Now
- University of Minnesota, Student Researcher May. 2020 – May. 2021
- Ecolab – UMN Collaboration, Student Researcher May. 2020 – May. 2021
- CenterPoint Energy, Research Intern Apr. 2019 – May. 2020

HONORS & SCHOLARSHIPS

- University of Minnesota's Undergraduate Research Opportunities Program (USD $1,500) Spring 2021
- Maroon Global Excellence Scholarship (USD $15,000) Fall 2018 – Fall 2021
- Vice president candidate of Tau Sigma National Honor Society Uni. of Minnesota Twin Cities Chapter Fall 2020
- Membership of Tau Sigma National Honor Society University of Minnesota Chapter Spring 2019 – Fall 2021
- Dean's List of College of Library Arts at the University of Minnesota, Twin - Cities Spring 2019 – Fall 2021

RESEARCH INTERESTS

My research interests lie in the general area of machine learning, particularly in unsupervised learning, and Bayesian deep learning, as well as their applications in meta-learning, sequential decision making, human computer interaction, and natural language understanding. I am also interested in applying AI techniques to address societal challenges.

RESEARCH EXPERIENCES

Statistics for Research and Computing for Research **Aug. 2021- Sep. 2021**

Team Member, Advisor: Dr. Emily Padston.

- 24 hours meeting per week with team members and mentors working in team/independently on non-trivial projects to prepare/train for advanced research.

An Empirical Study on Model Errors & Users' Error Repair Strategies in Natural Language Interfaces for Dataset Queries (NL2SQL) **June. 2021 – Now**

Visiting Scholar at the University of Notre Dame, Advisor: Prof. Toby Li.

- Implemented three state-of-the-art machine learning models to translate natural language instructions into SQL queries for help users access natural language queries from databases.
- Statistical analysis on the cooccurrences of error types and categories from the result of the above models on Spider and NL2SQL datasets as well as the reliability of annotations.
- Sum up 46 error types in 14 error categories and built the first comprehensive taxonomy of NL2SQL errors.
- Designed natural language query interfaces through implementing three state-of-the-art strategies for error discovery and designed two strategies for error repair.
- A paper is in preparation.

Senior Capstone Project: On the Truth Assignment Theorem of the Language of Sentential Logic **Jan. 2021 – May. 2021**

Senior Undergraduate Researcher, Advisor: Prof. Karel Prikry.

- Used the Language of Sentential Logic to prove the Truth Assignment Theorem.

- Used the Recursion Theorem to prove the Truth Assignment Theorem.
- A senior capstone project submitted to the Department of Mathematics at the University of Minnesota, Twin-Cities.

Funded & Self-designed: Improve Natural Language Understanding **Oct. 2020 – Apr. 2021**

Independent Researcher, Advisor: Prof. Maria Gini.

- Used meta-learning to address the task-agnostic problem in natural language understanding.
- Designed a prototype neural-symbolic model.
- **Supported by the Office of Undergraduate Research at the University of Minnesota.**

Self-designed: How personal perceptions of COVID-19 have changed over time **Jun. 2020 – Sep. 2020**

Independent Researcher, Advisor: Prof. Maria Gini.

- Analyzed personal perceptions towards the COVID-19 pandemic with the main challenge emanating from the limited amount of data and paucity of previous works.
- Proposed a perception analysis method combining sentiment analysis with topic extraction and sequential prediction, discovering the first ground truth COVID-19 emotion responds dataset at ACL2020.
- Designed a model evaluation scheme to select the optimal one for sentiment analysis among Naïve Bayes, Random Forests, SVM, Logistic Regression, LSTM, BERT, RoBERTa, and DistilBERT.
- Extracted five topics from the first ground truth dataset using LDA.
- Made sequential prediction of trends in five topics and thirteen sentiments using ARIMA and Encoder-Decoder LSTMs.
- Estimated the health status of users in Reddit and discovered their consistent nervousness.
- **Presented in Prof. Maria's Chatbot Group.**

Clustering U.S. counties to find patterns for COVID-19 pandemic **Apr. 2020 – Sep. 2020**

Clustering team member in Ecolab-UMN collaboration, Leader: Dr. Sarah Milstein.

- Discovered patterns relevant to the COVID-19 pandemic for each U.S county and found the core transition factors.
- Constructed a dataset of data relevant to the spread of COVID-19 from WHO and Johns Hopkins University.
- Implemented and evaluated K-Means, Fuzzy c-Means, Mini Batch K-Means, Gaussian Mixture Models, and tuned their hyperparameters using matrices choosing from Silhouette Metrics, Calinski-Harabasz Index, Davies-Bouldin Index, Elbow Score, AIC, and BIC.
- Optimized clustering interpretation method using Jenks Natural Breaks Optimization and Decision Tree.
- **Paper submitted in SIAM.**

Forecasting daily COVID-19 spread in regions around the world **Mar. 2020 – Jul. 2020**

Forecasting best team member in Ecolab-UMN collaboration, Leader: Dr. Jimmy Broomfield.

- Predicted the COVID-19 confirmed cases and fatalities for each region and country around the world.
- Implemented Encoder-Decoder BiLSTM, Transformed ARIMA, Multiphase Logistic Model, and Fill Forward Model to select the optimal one for prediction.
- Optimized epidemiological SIR model into SEEAIRD model using incubation, infections not yet identified, asymptomatic carriers, and death compartments. **Improved rank by 31, from 66 in week3 to 35/250+ in week5.**
- Derived ordinary differential equation, transition probabilities, system of stochastic differential equations, numerical simulation and parameters estimation of SEEAIRD model.

FDA COVID-19 Risk Factor Modeling Challenge **May. 2020 – Jul. 2020**

Forecasting best team member in Ecolab-UMN collaboration, Leader: Dr. Jimmy Broomfield.

- Investigated how demographic features and history of comorbidities affect the infection of COVID-19 on veterans.
- Found the risk factor, which is the history of chronic comorbidities, and the protective factor, which is PCV vaccine.
- Made the final prediction of Alive or Deceased Status by integrating comorbidities records, COVID-19 Status, Days in ICU, and Controlled Ventilation Status.
- Discovered the inconsistency of categorical values between train and test sets and proposed a strategy of transforming the values in trainset to match the ones in test set, which improved the modeling accuracy.

MUDAC 2020: Investigating Disparities in Outcomes across Venues **Mar. 2020**

Member in UMN competition team, Advisor: Prof. Gilad Lerman.

- Predicted count venues' tendency to favor the plaintiffs/defendants.
- Predicted the probability that a case will be closed by a summary judgment.
- Evaluated the performance of Logistic Regression, Support Vector Machine, and Decision Tree for prediction.
- Used Random Forests Feature Selection to improve modeling accuracy by 30+ percent.

INTERNSHIP EXPERIENCE

CenterPoint Energy **Apr. 2019 – Mar. 2020**

Data Analyst, *Joblogic-X Corporation*, Supervisor: Tengran Liu.

- Optimized the customer entry methods by designing a model to automatically duplicate the entry context into another cell, largely saving customer's entry time and being used in other projects.
- Developed SQL Server Integration Service(SSIS) data flow to ingest data from various sources and leveraged the SSIS source reader to process flat files and XML documents.
- Designed standard data quality routine to clean the source data and keep track of data quality matrices.
- Time series predicted the products' weekly inventory from suppliers and created reports use SAP Business Objects.
- Reached out to suppliers if the information was unclear and sought opportunities to develop long-term cooperation.
- Analyzed prices, promotions, distances, delivery time, and qualities of suppliers of different types of products and customer requirements to design optimized purchasing solutions for customers.

AWARDS & LEADERSHIP & VOLUNTEER EXPERIENCES

- 2nd Place in National Collegiate DanceSport Championships in Chicago — Spring 2019
- 2nd Place in Dance Fest, Amateur Silver International Latin Dance in St. Paul, Minneapolis — Spring 2019
- Vice President of the Central Uni. of Finance and Econ. Students Union — Spring 2016 – Spring 2018
- Communication Coordinator of Beijing Daxing district No.1 middle school — Fall 2014 – Spring 2016
- Volunteer English teacher at a local primary school in Galle, Sri Lanka — Winter 2018
- Volunteer English teacher at Beijing No.2 primary school — Fall 2018 – Fall 2015
- Received the Gold Medal Certificate from China Ballroom Dance Federation — Fall 2014

SKILLS

- Programming Language: Python, Java, SQL, MATLAB, RStudio, and LaTeX.
- Tools: PyCharm, IntelliJ, NLTK, TensorFlow, Atom, and IDLE.
- BI Tools: SAP Business Objects, Tableau, and Power BI.