Tianyi Sun

Tel: 612-309-0898 | E-mail: sun00234@umn.edu

EDUCATION

University of Minnesota Twin Cities Aug. 2018 – May 2021 (Expected)

B.A. in Mathematics (Computer Applications), Minor in Statistics and Computer Science GPA: 3.67/ 4.0

Relevant coursework **(upper division and graduate courses)**: Artificial Intelligence, Machine Learning, Algorithm and Data Structure, Formal Language and Automata, Probability and Statistics, Applied Linear Algebra, Numerical Method, Cryptology and Number Theory, Mathematical Logic, Theory of Statistics, and Regression and Corelated Data.

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

Major in Mathematics GPA: 3.96/ 4.0

Relevant coursework: Macroeconomics, Microeconomics, Fundamentals of Accounting, Public Finance, Business Statistics, Marketing Management, and Psychology.

HONORS & SCHOLOARSHIP

- 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 Uni. of Minnesota - Twin Cities Chapter Spring 2019 – Present
- Dean's List of College of Library Art at the University of Minnesota Spring 2019 – Present

RESEARCH INTEREST

My research interests lie in the general area of machine learning, specifically in deep learning, supervised learning, unsupervised learning, as well as their applications in natural language processing, sequential decision making, specifically, GPT-3, Natural Language Generation and Natural Language Understanding. I am also interested in applying AI techniques to address societal challenges, such as COVID-19 pandemic.

RESEARCH EXPERIENCE

Improve Natural Language Understanding Through Logical Reasoning Oct. 2020 – Present

Independent Researcher, Advisor: Prof. Maria Gini.

- Optimize leading language pre-training models including BERT, and RoBERTa.
- Design a model combining logical reasoning and the leading language pre-training model.
- Implement Few-Shot Learning and GPT-3 to generate texts using the first ground truth dataset of emotion responds to COVID-19.

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

Independent Researcher, Advisor: Prof. Maria Gini.

- Aimed to analyze 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 topics extraction and sequential prediction, discovering the first ground truth COVID-19 emotion dataset at ACL-2020.
- Designed a model evaluation scheme and selected optimal models for three tasks from approaches: Naïve Bayes, Random Forests, Linear SVM, Linear SVC, Logistic Regression, LSTM, XLNet, BERT, RoBERTa, and DistilBERT, ARIMA, and Encoder-Decoder LSTMs.
- Estimated the health status of authors in Reddit based on the predictive trend of extracted five topics and discovered authors' consistent nervousness about COVID-19 based on the predictive trend of classified thirteen sentiments.
- Paper submitted at AAAI-2021.

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

Group member of Ecolab-UMN Collaboration, Leader: Sarah Milstein

- Discovered patterns relating to the COVID-19 pandemic by clustering U.S counties.
- Constructed a dataset of data relevant to the spread of COVID-19 from WHO and JHU.
- Implemented and evaluated clustering methods, including K-Means, Gaussian mixture models, and tuned their hyperparameters using metrics silhouette, Cainski-Harabasz Index, Davies-Bouldin Index, elbow, AIC, and BIC.
- Innovated a clustering interpretation methods using Jenks Natural Breaks Optimization to determine which features were useful in distinguishing our clusters.
- Long paper submitted at SIAM.

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

Group member of Ecolab-UMN Collaboration, Leader: Jimmy Broomfield

- Investigated how race, ethnicity, age and history of comorbidities affect the progress of COVID-19 infected veterans.

- Combined teammates' predictions of COVID-19 Status, Days Hospitalized, Days in ICU, and Controlled Ventilation Status for a given individual to make the final prediction of Alive or Deceased Status.
- Discovered the inconsistency of categorical values between training and test set, and proposed a strategy of transforming the categorical values in train set to match the ones in test set, which significantly improved our model.
- Found many COVID-19 infections died due to history of chronic comorbidities instead of COVID-19 and infections who got PCV vaccines are less likely to die from COVID-19.

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

Group member of Ecolab-UMN Collaboration, Leader: Jimmy Broomfield

- Implemented and evaluated models, including Bidirectional LSTM model with encoder-decoder layers, ARIMA with Square Root Transform, ARIMA with Log Transform, Multiphase Logistic Model and Fill Forward Model, in order to predict the confirmed cases and fatalities for each country around the world.
- Submitted model to Kaggle competition, which ranked 66th out of 250+ participant groups.
- Developed compartmental SIR model to SEEAIRD model using exposed infectious, exposed not yet infectious, asymptomatic, and death features, based on the complete transition pattern of COVID-19 to improve the result.
- Paper is currently in the process of preparation.

MUDAC2020: Investigating Disparities in Outcomes across Venues Mar. 28th – 29th, 2020

Data Science Challenge participant, Advisor: Prof. Gilad Lerman

- Investigated the disparities in outcomes across count venues, including discovering count venues' tendency to favor the plaintiffs or defendants and predicting the probability that a case will be closed by a summary judgment.
- Implemented and evaluated supervised learning methods, including Logistic Regression, Support Vector Machine, Decision Tree, and Random Forests.
- Used feature importance to select features to use in the models, which improved modelling accuracy by more than 30 percent on average.

INTERNSHIP EXPERIENCE

CenterPoint Energy Apr. 2019 – Mar. 2020

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

- Optimize the customer entry methods by designing a model to automatically duplicate the entry context into another cell, which save customer's entry time by xx.
- Developed SSIS (SQL Server Integration Service) data flows to ingest data from various sources and leveraged the SSIS source reader to process flat files, XML documents and other related sources.
- Designed standard data quality routine to clean the source data and keep track of data quality matrixes.
- Implemented time series model to predict the weekly inventory of each product from each supplier and created new reports through SAP Business Objects.
- Reached out to suppliers if information was unclear and sought opportunities to develop long-term cooperation.
- Analyzed prices, promotions, distances, delivery time and qualities of suppliers to design optimized purchasing solutions for customers, based on different types of products and customer requirements.
- Awarded return offer.

AWARDS & LEADERSHIP EXPERIENCES & VOLUNTEER EXPERIENCES

- 2nd Place in National Collegiate DanceSport Championships, Amateur Collegiate Championship Latin Spring 2019
- 2nd Place in Dance Fest, Cody Arndston Amateur Silver International Latin S/C/R Spring 2019
- Vice President of the Central Uni. of Finance and Econ. Students Union Fall 2016 – Spring 2018
- Communications Coordinator of Beijing Daxing district No. 1 middle school Fall 2014 – Spring 2016
- Volunteer English teacher at Beijing's No.2 primary school and Galle, Sri Lanka. Winter 2018 – Fall 2015

SKILLS

- Programming Language: Python, Matlab, R-studio, SQL, and LaTeX.
- Tools: SciPy, Scikit-learn, PyTorch, Keras, fast.ai, TensorFlow, Transformers, and NLTK.
- Data analysis: dataset construction, text preprocessing, and machine learning and deep learning modeling and tuning.
- BI Tools: SAP Business Objects, Tableau, and Power BI.