Jason Liu

Currently investigating automatic topic labeling and applied deep learning.

Jason is the President of the Data Science Club this term at Waterloo and the creator and organizer of Data Hackers; a data science interest group with more than 3000 members worldwide. He is very interested applying data science to understand social networks, peer effects, and mental health. He is also interested in making realistic modeling assumptions to create models that are informative and interpretable.

If you're interested in contacting Jason for work or anything else, you can reach him at jason@jxnl.co.

Writing

Publications

Liu J., Weitzman E., Chunara R.
Assessing Behavior Stage Progression from Social Media Data *Conference on Knowledge Discovery and Data Mining* **Submitted** 2016, San Francisco, CA, USA

Rehman NA, **Liu J.**, Chunara R.
Using Propensity Score Matching to Understand the Relationship Between Online Health Information Sources and Vaccination Sentiment. *AAAI Spring Symposium.* 2016, Stanford University Palo Alto, CA, USA

Work Experience

StitchFix
Data Scientist — Algorithms (Data Labs)

- Designed various neural networks in Chainer to study problems such as image hashing, image generation, and image labeling.
- Successfully engineered interpretable features for content based recommendation at Stitch Fix using multi-modal item embeddings (text, image, context).
- Designed and analyzed human computation experiments to study ranking algorithms effects on human performance using mixed models.
- Developing dataset creation tools that minimized leakage for time series classification.

NYU Global Institute of Public Health
Research Assistant — ChunaraLab

Supervisor: Dr.Rumi Chunara
- Analyzed 50+ million tweets to study the geographic and temporal characteristics of alcohol consumption amoung twitter users.
- Leveraged Amazon Mechanical Turk to build an active learning pipeline for heirarchical classification using SVMs to minimize costs for dataset creation.
- Submitted research paper to SIGKDD Applied Data Science Track as Lead Author. (currently under review)

Marketwired
Data Scientist — Sysomos Labs

- Prototyped a proof-of-concept advertisement recommendation platform for offline targeted ads using multiarmed bandit algorithms.
- Improved clustering performance for twitter community detection on Sysomos MAP and Heartbeat.
- Created MapReduce and Spark applications for ad targeting generation and various ETL.

Areas of Interest
Classification + Social Media + Mental Health + Topic Modeling + Applied Machine Learning.

Summary of Qualifications

TOOLKIT

python + R + pandas + dplyr + SQL + Presto + Spark + Map Reduce + scikit-learn + Chainer + Gensim + BeautifulSoup + Requests.

Education

University of Waterloo: Computational Mathematics & Statistics, 2017

Classification + Comp. Stats + Comp. Phys + Data Structures and Algorithms + Computational Linear Algebra + Generalized Linear Models + Group Theory + Real Analysis.

Internet

Quora + Github + Dribbble + LinkedIn + Twitter.

Online articles

Liu J.,What is an intuitive explanation of singular value decomposition? *Quora* 2015, Waterloo, ON, Canada