john chen

johnc@cs.toronto.edu johntiger1.github.io 4168398823 linkedin.com/in/johnn-chen/

+ employment

Amazon
Toronto, Canada
Applied Scientist Intern
Fall 2020

Alexa AI - Natural Understanding

Improving Entity Resolution using multiple hypothesis ASR and vector search information retrieval techniques

Mentors: Majid Laali, PhD and Haoyu Wang

Lyft
San Francisco (remote)
Software Engineer in Machine Learning Intern
Summer 2020

Conversion Learning for Dynamic Pricing (Lyft Prime Time)

- Developed **non-parametric, real-time machine learning models** to replace existing econometric, statistical models for conversion elasticity. Production models produce real-time price estimates for **700+ million events daily.**
- **Decreased relative error by 20%** in binary conversion prediction model by leveraging offline personalized features. Model has been put into production impacting **over 1 million customers** in rollout regions.

Mentors: Nile Beesinger and Dean Grosbard, PhD

Air Miles (a LoyaltyOne Company)
Toronto, Canada
Data Scientist Intern
May 2019 to Aug. 2019

- Developed several machine learning models end-to-end, including flight redemption propensity model. FRP **xgboost** model achieved **0.97 AUC ROC**, and has **9x lift** compared to baseline.
- Improved model via **seasonality adjustments**, **model-free feature selection** algorithm and **model-based genetic feature selection** algorithm, as well as hyperparameter tuning via **Bayesian Optimization** using scikit-optimize.
- Model has been used in business and has delivered **40% e-mail cost savings** in an acquisition campaign targeting over **500 000 customers**.
- Performed research and development in recommendation engine prototype for leading Canadian grocery chain. Achieved **111% basket coverage improvement** over baseline by using market basket analysis and frequency-based model.

Mentor: Somayyeh Aghababei, PhD

Vector Institute
Toronto, Ontario
Deep Learning Research Intern
June 2018 to Current

Developed and implemented several end-to-end deep learning solutions including:

- **character-level recurrent neural network language model** as part of industry partnership with DataX
- medical image **captioning using convolutional neural network encoder and recurrent neural network decoder**. **Doubled F1 performance** by reformulating problem as multilabel classification problem.
- paragraph-level style transfer in text between doctor and patient corpora (research ongoing)
- massively multilabel, highly imbalanced classification. **Improved F1 performance by 156% over baseline pretrained** network via **image augmentation** and **finetuning**.

University of Toronto
Teaching Assistant
Sept. 2016 to Current

- TA for 8 **different courses**: CSC209 (Software Tools and Systems Programming), CSC263 (Data Structures and Analysis), CSC165 (Mathematical Expression and Reasoning for Computer Science), CSC324 (Functional Programming and Principles of Programming Languages), CSC384 (Introduction to Artificial Intelligence), CSC420 (Computer Vision and Image Understanding), CSC413 (Neural Networks and Deep Learning) and JSC370 (Data Science II)

Centre for Computational Medicine at SickKids Hospital
Machine Learning Research Student
May 2017 to Aug. 2017

- **Awarded $6000 NSERC research grant** to conduct summer research into Bayesian networks to improve automatic disease diagnosis algorithm used in Phenotips web platform. **Sped up naive algorithm by 500x** via dynamic programming and multithreading.

+ awards

University of Toronto and Province of Ontario · Ontario Graduate Scholarship — July 2020

Awarded for excellence in graduate studies.

Vector Institute · Vector Research Grant — June 2020

Two-time recipient (2018 and 2020). Awarded to conduct deep learning research under supervision of Dr. Frank Rudzicz

Rotary Club of Toronto and Toronto Argonauts · Rotary Club of Toronto Community Champion Scholarship — June 2014

Full tuition scholarship to University of Toronto

+ education

Doctor of Philosophy in Computer Science
Sept. 2019 to Current

University of Toronto 2023

Research areas: Deep Learning and Natural Language Processing. Advised by Professor Frank Rudzicz. Transitioning from Masters. Cumulative GPA: 4.00/4.00

Bachelor of Science in Computer Science
Sept. 2014 to May 2019

University of Toronto 2019

Focus in Artificial Intelligence and Deep Learning, Graduated with High Distinction. Cumulative GPA: 3.98/4.00

+ publications

Towards an Expert-Based Question Answering System for COVID-19 Literature

Keywords: Transformer, Question Answering, T5, Active Learning. Accepted (**oral**) at *Scholarly Document Processing @ EMNLP 2020*

Exploring Text Specific and Blackbox Fairness Algorithms in Multimodal Clinical NLP

Keywords: Fairness, Multimodality, Debiased Word Embeddings. **First author.** Accepted (**oral**) at *Clinical NLP @ EMNLP 2020*

Reinforcement Learning for Pool-Based Active Learning

Keywords: DQN, Policy Gradient, Uncertainty Sampling. *In submission at ICML 2020*

+ skills

DATA SCIENCE AND DEEP LEARNING

PyTorch
scikit-learn
PySpark
Tensorflow
xgboost
SparkSQL
NumPy
Pandas
nltk
SpaCy
matplotlib
Spark MLLib
mlflow
Keras
transformers
t5
lightGBM

FULL STACK PROGRAMMING

Python
R
Node.js
React
JavaScript
C
Java
Bootstrap
Postgres
MongoDB
C#
Racket
HTML5
CSS3
AngularJS