Skyler Seto

EDUCATION

Cornell University August 2014 - Present
Ph.D. in Statistics with Special Masters in Computer Science (tentative)
M.Sc. in Statistics awarded January 2018
Advising Committee: Martin T. Wells, Andrew G. Wilson, Thorsten Joachims

Massachusetts Institute of Technology September 2010 - June 2014
B.Sc. in Mathematics with Computer Science

GRADUATE COURSEWORK

Bayesian Machine Learning, Big Data Analysis, Generalized Linear Models, Machine Learning, NLP, Statistical Learning Theory, Statistical Computing

RESEARCH EXPERIENCE

Word Embeddings as Generalized Linear Models August 2017 - Present

- Developing word embedding models based on glms and GloVe embeddings

Statistical Tests for Topic Models December 2016 - Present

- Developing bootstrap tests to check the fit of NMF-based topic models
- Accepted at NIPS 2017 Symposium on Interpretable Machine Learning

Automatic Bayesian ARMA Models August 2016 - Present

- Developing probabilistic approximate inference algorithm for automatic ARMA order determination

WORK EXPERIENCE

Amazon Alexa Speech and ML Group May 2017 - August 2017

- Experimented with neural word-character hybrid language models for entity recognition and generation tasks

Riot Games Data Science Team May 2016 - August 2016

- Used time series forecasting models to detect network connection anomalies
- Automated models for real-time anomaly tracking and recording

MIT Anyscale Learning For All Group May 2015 - August 2015

- Used statistical models for car destination prediction
- Built tools for parsing and cleaning car trip signal data for ML prediction problems

ML AND AI PROJECTS

Predicting Student Final Grades January 2016 - May 2016

- Used approximate inference algorithm for predicting the distribution of final grades
- Presentation to the Cornell Learning Analytics Group

Human Activity Time Series Classification January 2015 - September 2015

- Designed algorithm based on dynamic time warping for automatic feature selection
- Published in 2015 IEEE Symposium Series on Computational Intelligence

Linguistic Differences in Reddit Users August 2015 - December 2015

- Used language models and lexicon statistics to classify user involvement on reddit

TECHNOLOGY

Big Data Software: Hadoop, Spark
Programming Languages: Python, R, SQL, Java, MATLAB
Deep Learning Frameworks: PyTorch, Tensorflow