

Jason (Zhihang) Dong

Ph.D. Student | Looking for Research/Data/Machine Learning Scientist Intern Opportunities

in [linkedin.com/in/zhihangdong](https://www.linkedin.com/in/zhihangdong) | github.com/zdong1 | +1 814 548 6383 | zdong@uw.edu

📍 211 Savery Hall, Box 353340, Seattle, Washington, USA 98115 | <https://zdong.org>

I am interested in the mathematical and statistical applications of **AI and machine Learning**, specifically **Online & Adaptive Learning**, **Graphical and Geometric Modeling** (e.g. manifold learning) and **Speech Recognition/ Natural Language Processing** (Question Answering and Embedding Methods). Additionally, I am interested in **Statistical Methodologies**, such as spatial-temporal models in **demography**. I have exposure to major AI-Platforms such as *Glun, aws SageMaker and YOLO*.

Skills

Note: Skills with **bold** mean 50+ hours experience (400+ for languages) & ranked with familiarity

Computer Languages: **Python, R, Java**, Scala, C(Statistics Libraries)

Machine Learning Toolkits: **Spark, Torch, xgBoost, caffè, Tensorflow, H2O**, Hadoop, CTNK, EMR, Keras

Databases: **Neo4j, Cassandra, MongoDB, postgre, mySQL, RDS, mySQL, Hive**, CouchDB, Lucene

Natural Language Processing: **Tika, NTLK, SpaCy, TextBlob**, Stanford Core NLP

Softwares: **ArcGIS, Stata, tableau**, GeoDa, SAS, Qt

Research Projects

Current
June, 2017 | **Swiss Cell Data Record Project, Seattle, WA**

- > Managed 100G+ flow of GPS record through cluster computing service; wrangled to overcome the noisiness and high dimensionality of space-time data; Implemented the topological data methods
- > Modeled activity space and social interactions with innovated rank-based convergence algorithm

R **AWS** **d3** **Tensorflow** **TDN** **Python** **SQL** **Demography** **Mapping**

Current
January 2018 | **Knowledge Transfer and Predictive Question Answering via Reinforcement Learning, Seattle, WA**

- > Predictive Graphical Models for Knowledge Graph with Reinforcement Learning, Bayesian Network and Multi-layer bidirectional long short-term memory network (LSTM)
- > Designed end-to-end training of multiple documents across the Document Retriever and Document Reader pipeline for Question & Answer (QA) NLP Tasks

GAN **Reinforcement Learning** **SpaCy** **Python** **NLP** **NTLK** **caffé** **LSTM**

May 2017
January 2017 | **Neighborhood Sensing : Investigating Crimes with Network Effects, Seattle, WA**

- > Scrapped Police data into City Gov APIs and created Integrative **R-Shiny** Interactive Visualization applications that mapping and modeling the network effects on neighborhood safety

R **Neighborhood Effects** **Time Series** **Social Network**

Mar 2018
May 2017 | **Anomaly Detections of NYC Taxi Ridership using Adapted Bayesian Neural Network, Seattle, WA**

- > Designed an adapted Bayesian neural network with benchmarked SARIMA time series model
- > Improved predicted accuracy by 4% and implemented the inference system on Python
- > Applied manifold learning methods to conduct dimensionality reduction

Bayes **Time Series** **SARIMA** **Anomalies** **Python** **LSTM**

Work Experience

September, 2017
June, 2017 | **Center for Studies in Demography and Ecology, Summer Research Assistant, Seattle, WA**

- > Provided 400+ hours of statistical/programming supports to 5 different mini-projects;
- > Need-Based Assistance such as implementing GIS ID Matrix; identification algorithms of fake patient ID in the system with < 0.1% error rate and Viz Project such as **leaflet.js** on interactive mapping

R **AWS** **d3** **leaflet** **Python** **SQL** **Demography** **Mapping**

Current
Summer 2015 | **Penn State and Univ. of Washington, Teaching and Research Assistant, —**

- > Assisted and co-taught undergraduate-level methodological and substantive courses; Shared responsibilities for lectures, exams, assignments; Best Rating : **4.4/5.0**, 15 Summer; Reference Available

Experimental Design **Teaching** **Public Speaking**

May 2015
August 2015

Survey Research Center of SSRI, Undergraduate Summer Analyst, University Park, PA

> Implement a validation algorithm to detect ineffective survey entries and non-random missingness

R Stata

Languages

English ● ● ● ● ●
Chinese ● ● ● ● ●
Japanese ● ● ○ ○ ○

+ Honors & Awards

- > **Clarence and Elissa M. Schrag Endowed Fellow**
University of Washington, 2016 (2 per Ph.D. Cohort)
- > **UW Cloud Computing Credits Research Awards**
University of Washington, 2016
- > **Undergraduate Research Travel Awards**
Penn State University, 2014 - 2016

Education

2020/12 (est.) Ph.D. — University of Washington, Seattle (CSSS Option, Sociology)

Research Areas: Machine Learning, Graphical Modeling and Geometric Data Analysis

Advisor: Prof. Adrian Dobra, Prof. Ross Matsueda (co-chair)

2019/3 (est.) M.S. Statistics — University of Washington, Seattle

2016/05 B.A. — Pennsylvania State University, Sociology with Concurrent Majors and Minors in Statistics, Geography, Media Studies, Psychology (5 majors and 3 minors)

“ Publications

— — Conferences — —

- > **Dong Zhihang**, Yen-Chi Chen and Adrian Dobra (2018). “Projecting the Short-term Population Mobility using Cell Data Records”. *Accepted to 2018 Joint Statistical Meetings (JSM)*.
- > **Dong Zhihang** and Tongshuang Wu (2018). “Benchmarking Open Source NoSQL Databases Performance on NLP Queries”. *Work in Progress*.
- > **Dong Zhihang** (2017). “Estimation and Extrapolation of Spatial Trends in Mortality Data using Bayesian APC Modeling”. In : *International Conference on Population Geography*.
- > — (2016). “Theorizing Urban Neighborhoods : Mapping the Interneighborhood and Intranighborhood Networks and Criminogenic Factors on Street Crime Victimization”. In : *American Society of Criminology Annual Meeting*.
- > — (2016). “Structural Covariates of Crime in Texas Counties using Spatial Effects.” In : *Annual Meetings of the Population Association of America*, Washington, D.C., March, 2016.

— — Journals — —

- > Dong, Zhihang. “Modeling Age Homogeneity : Age Homogamy And Marital Happiness Over The Life Course”. *Undergraduate Thesis, Under Review by JMF*.

Course Works

- > 0. STAT 535 : **Statistical Learning** [Link]
- > 1. CSE 599i : **Online and Adaptive Learning** [Link]
- > 2. STAT 564 : **Bayesian Statistics**[Link]
- > 3. MATH 515 : **Optimization**[Link]
- > 4. CSE 547 : **Machine Learning for Big Data**[Link]
- > 5. EE 576 : **Computer Vision**[Link]
- > 6. CSE 599d : **Advanced NLP Methods** [Link]
- > 7. CSE 544 : **Database Management** [Link]
- > 8. CSE 512 : **Data Visualization** [Link]
- > 9. CSE 599d1 : **Advanced Topics NLP** [Link]
- > 10. EE 595 : **Representation Learning** [Link]
- > 11. LING 575 : **Novel NLP Applications** [Link]

+ Labs & Working Groups

- > **Geometric Data Analysis Reading Group**
2018 – Current [Link]
- > **Working Group for Applied, Bayesian and Computational Statistics**
2016 – Current [Link]
- > **Context Working Group (Sociology)**
2016 – Current POC : Prof. Kyle Crowder
- > **Human Factors in GI Science Lab, PSU Geography**
2014 – 2015 [Link]