

Jason (Zhihang) Dong

Ph.D. Student | Looking for Research/Applied/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 broadly interested in the statistical applications of **AI, ML and data mining**, specifically **geometric methods, online methods** and **architectures for time series forecasting**. Application wise, I am interested in the engineering related to **Speech/Natural Language Processing**. Additionally, I am interested in **methodologies** for **population sciences (demography)**.

Skills

Notes : 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 System : Git, AWS

Research Projects

Notes : Titles are followed by relevant publication reference tags (sample links found there) and starting year.

- | | |
|--------------------------|--|
| Current
June, 2017 | Swiss Cell Data Record Project, [5], 2017 <ul style="list-style-type: none">➢ Managed 400G+ sensor data through Spark and cluster computing; Designed spatio-temporal models for complex human mobility; Implemented the topological model resulting a 35% improvement on activity space coverage; Implemented tensor deconvolutional network to reveal social interaction <div style="display: flex; gap: 5px;">R AWS d3 Tensorflow TDN Python SQL Demography Mapping</div> |
| Current
January 2018 | Online Methods for Knowledge Graph Completion Tasks, [-], 2018 <ul style="list-style-type: none">➢ Designed a new online learning mechanism for knowledge graph completion. Built a pipeline architecture for document reading and information retrieval dedicated to Question & Answer (QA) tasks; Achieved a 0.04 improvement on F-1 score for the SQuAd 2.0 dataset; <div style="display: flex; gap: 5px;">NLP Deep Learning Online Learning</div> |
| May 2017
January 2016 | Neighborhood Sensing : Investigating Crimes with Network Effects, [2,3], 2016 <ul style="list-style-type: none">➢ 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 <div style="display: flex; gap: 5px;">R Neighborhood Effects Time Series Social Network</div> |
| May 2018
May 2017 | st-ANN : Spatial-Temporal Attention Network for Air Quality Prediction, [6], 2017 <ul style="list-style-type: none">➢ Designed a temporal-spatial attention nets architecture for time series forecasting with dual attention on spatially local contingency; Applied the methods to Beijing Air Quality Prediction Forecasting data (KDD Cup 2018) with a 1.2% improvement on accuracy than the benchmark seq-to-seq model; <div style="display: flex; gap: 5px;">Bayes Time Series SARIMA Anomalies Python LSTM</div> |

Work Experience

- | | |
|--------------------------------|--|
| December, 2018
August, 2018 | Amazon.com Inc., Applied Scientist Intern, Seattle, WA <ul style="list-style-type: none">➢ Conduct research on NLP at Amazon Lending Group. More details coming as the internship concludes; <div style="display: flex; gap: 5px;">NLP Machine Learning</div> |
| August, 2018
June, 2018 | Deloitte Services LP., Senior Data Scientist Intern, Seattle, WA <ul style="list-style-type: none">➢ Created a protocol for organizational network analysis on employee interaction data using inference, exponential random graph model (simulation) and a predictive modeling using RNN;➢ Built the structure of resume recommendation architecture 3.0 for consulting projects powered by deep learning with pipeline from resume reading to recommendation interface <div style="display: flex; gap: 5px;">NLP Social Network R AI</div> |

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

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

- 2021/6 (est.) Ph.D. — University of Washington, Seattle (CSSS Option, Sociology)
Research Areas : Machine Learning, Graphical Modeling and Geometric Methods, NLP
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

— Work in Progress —

- > [8] Tongshuang Wu, Zhihang Dong, Sicheng Song and Mingrui Zhang “Attention Model Visualizer” [Sample]
- > [7] Zhihang Dong and Tongshuang Wu. “Benchmarking Open Source NoSQL Databases Performance on NLP Queries”. [Sample]
- > [6] —. “st-ANN : Deep Spatial-Temporal Attention Network for Time Series Forecasting”. [Sample] *Manuscript Submitted to IEEE International Conference on Data Mining (ICDM)*

— Conferences —

- > [5] Zhihang Dong, Yen-Chi Chen and Adrian Dobra (2018). “Projecting the Short-term Population Mobility using Cell Data Records”. *Accepted to 2018 Joint Statistical Meetings (JSM)*. [Presentation]
- > [4] Zhihang Dong (2017). “Estimation and Extrapolation of Spatial Trends in Mortality Data using Bayesian APC Modeling”. In : *International Conference on Population Geography*.
- > [3] — (2016). “Theorizing Urban Neighborhoods : Mapping the Interneighborhood and Intra-neighborhood Networks and Criminogenic Factors on Street Crime Victimization”. In : *American Society of Criminology Annual Meeting*.
- > [2] — (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.
- > [1] 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]