

# 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](https://github.com/zdong1) | +1 814 548 6383 | [zdong@uw.edu](mailto: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 *aws SageMaker and Google Cloud*.

## 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

## Research Projects

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

Current June, 2017	<b>Swiss Cell Data Record Project, [5], 2017</b> <ul style="list-style-type: none"><li>&gt; 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</li><li>&gt; Modeled activity space and social interactions with innovated rank-based convergence algorithm</li></ul> <p>R AWS d3 Tensorflow TDN Python SQL Demography Mapping</p>
Current January 2018	<b>Know-Nets : Predictive Question Answering with Reinforced Transfer Graphs, [7,8,9], 2018</b> <ul style="list-style-type: none"><li>&gt; Predictive Graphical Models for Knowledge Graph with Reinforcement Learning, Bayesian Network and Multi-layer bidirectional long short-term memory network (LSTM)</li><li>&gt; Designed end-to-end training of multiple documents across the Document Retriever and Document Reader pipeline for Question &amp; Answer (QA) NLP Tasks</li></ul> <p>GAN Reinforcement Learning SpaCy Python NLP NTLK caffè LSTM</p>
May 2017 January 2016	<b>Neighborhood Sensing : Investigating Crimes with Network Effects, [2,3], 2016</b> <ul style="list-style-type: none"><li>&gt; Scrapped Police data into City Gov APIs and created Integrative <b>R-Shiny</b> Interactive Visualization applications that mapping and modeling the network effects on neighborhood safety</li></ul> <p>R Neighborhood Effects Time Series Social Network</p>
Mar 2018 May 2017	<b>SA-TCN : Self-Attention Temporal Convolutional Nets Framework, [6], 2017</b> <ul style="list-style-type: none"><li>&gt; Designed an self-attention temporal convolutional nets architecture for anomaly detections of time series data, achieved 8.6% improvement accuracy on KDDFA-18 Data (compared to benchmark)</li><li>&gt; Constructed and applied geometric methods to conduct dimensionality reduction</li><li>&gt; Applied the method on NYC taxi ridership and air quality data</li></ul> <p>Bayes Time Series SARIMA Anomalies Python LSTM</p>

## Work Experience

September, 2017 June, 2017	<b>Center for Studies in Demography and Ecology, Summer Research Assistant, Seattle, WA</b> <ul style="list-style-type: none"><li>&gt; Provided 400+ hours of statistical/programming supports to 5 different mini-projects;</li><li>&gt; Need-Based Assistance such as implementing GIS ID Matrix; identification algorithms of fake patient ID in the system with &lt; 0.1% error rate and Viz Project such as <b>leaflet.js</b> on interactive mapping</li></ul> <p>R AWS d3 leaflet Python SQL Demography Mapping</p>
Current Summer 2015	<b>Penn State and Univ. of Washington, Teaching and Research Assistant, —</b> <ul style="list-style-type: none"><li>&gt; Assisted undergraduate-level courses; Shared responsibilities for lectures; Best Rating : <b>4.4/5.0</b></li></ul> <p>Experimental Design Teaching Public Speaking</p>

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

### — — Work in Progress — —

- > [9] Tongshuang Wu, Zhihang Dong, Sicheng Song and Mingrui Zhang “Attention Model Visualizer” [Sample]
- > [8] Zhihang Dong and Tongshuang Wu. “Benchmarking Open Source NoSQL Databases Performance on NLP Queries”. [Sample]
- > [7] Zhihang Dong. “Know-Nets : A Reinforced Framework for Online Knowledge Graph Transfers”. [Sample] *Manuscript Submitted to 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*
- > [6] —. “SA-TCN : Deep Framework for Anomaly Detections in Time Series Data”. [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)*.
- > [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 Intraneighborhood 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]