

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 most interested in **machine learning and data mining**, especially forecasting and translation of high-dimensional spatio-temporal data, such as cell phone data, which often lead to meaningful social computing. Theoretically, I am interested in adaptive learning & geometric methods for machine learning. I am also fascinated about **natural language processing** problems involving question-answering and machine translation in context, which motivates enhanced representation and knowledge graph.

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

> Managed 400G+ sensor data through Spark and cluster computing; Designed spatio-temporal models for complex human mobility; Implemented the topological model for improvement on activity space coverage; Modeling of interactions using tensor deconvolutional network

Geometric Methods | Machine Learning | Human-Computer Interaction

Current
January 2018

Deep Models for Short Text Series, [-], 2018

> In this project, we concern the short text series in temporal contexts. Such series can be those under thread discussion, blogs and e-commerce products. We build a memory-based deep neural network model to identify the effects of earlier information on subsequent formations.

Natural Language Processing

Current
February 2018

Localize Attention Framework with Adjacent Contexts : Applications on Pollution Forecasting, [7], 2018

> Designed a temporal-spatial attention 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

Space-Time Modeling

Work Experience

December, 2018
August, 2018

Amazon.com Inc., Applied Scientist Intern, Seattle, WA

> Worked as NLP scientist on a natural language processing problem related to contextual sentiment extraction and forecasting

NLP | Machine Learning

August, 2018
June, 2018

Deloitte Services LP., Senior Data Scientist Intern, Seattle, WA

> Created a protocol for organizational network analysis on employee interaction data using causal inference, exponential random graph model (simulation) and a predictive modeling using RNN;

> Built the idea of resume recommendation architecture 3.0 for consulting projects powered by deep learning with pipeline from resume reading to recommendation interface

NLP | Social Network | R | AI

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

- > [8] — “Deep Series Model for Sentiment Reinforcement : Modeling the Online Reviews Texts with Neural Networks *Work in Progress*
- > [7] Tongshuang Wu, Zhihang Dong, S. Song and M. Zhang “Interactive Attention Model Explorer for NLP Tasks with Unbalanced Data Sizes” [Sample] *Work in Progress*
- > [6] Zhihang Dong and Tongshuang Wu. “Benchmarking Open Source NoSQL Databases Performance on NLP Queries”. [Sample]
- > [5] Zhihang Dong and . “st-ANN : Deep Spatial-Temporal Attention Network for Time Series Forecasting”. [Sample] *Work in Progress*
- > [4] 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]
- > [3] — (2017). “Estimation and Extrapolation of Spatial Trends in Mortality Data using Bayesian APC Modeling”. Accepted to *International Conference on Population Geography*.
- > [2] — (2016). “Theorizing Urban Neighborhoods : Mapping the Interneighborhood and Intranighborhood Networks and Criminogenic Factors on Street Crime Victimization”. Accepted to *American Society of Criminology Annual Meeting*.
- > [1] Dong, Zhihang (2016). “Modeling Age Homogeneity : Age Homogamy And Marital Happiness Over The Life Course”. *Submitted to Population Association of America Annual Meeting (PAA2019)*.

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 521 : **Design& Analysis of Algorithms** [Link]
- > 10. EE 595 : **Representation Learning** [Link]
- > 11. LING 575 : **Novel NLP Applications** [Link]
- > 12. CSE 599I2 : **Alg. thru Geometric Lens**[Link]

+ Labs & Working Groups

- > **Augmented Intelligence Seminar (AI Sem)**
2018 – Current [Link]
- > **Geometric Data Analysis Reading Group**
2018 – Current [Link]
- > **Working Group for Applied, Bayesian and Computational Statistics**
2016 – Current [Link]
- > **Space-Time Reading Group**
2016 – 2017
- > **Human Factors in GI Science Lab, PSU Geography**
2014 – 2015 [Link]