Yunyan Duan

https://www.linkedin.com/in/yunyan-duan/

Email : yyduan.pku@gmail.com

Mobile : (847)-596-1503

Education

Northwestern University — Evanston, IL

PhD candidate, ***Linguistics*** — *Sep. 2015 – Jun. 2020 (expected)*

- Thesis (prospectus): Word identification and eye movement control in reading as rational decision making
- Advisor: Dr. Klinton Bicknell

Coursera, deeplearning.ai — https://www.coursera.org/

Certificate: ***Deep Learning*** *Specialization* — *Nov. 2018*

- Through 5 courses, developed a profound knowledge of deep learning from its foundations (neural networks) to its advanced techniques and industry applications (convolutional neural networks, recurrent neural networks, etc.).

Stanford Online — https://lagunita.stanford.edu/

Statement of Accomplishment: ***Mining Massive Datasets*** — *Jul. 2017*

- This course covers "big-data" algorithms, including PageRank, stream algorithms, clustering, social-network graph analysis, large-scale machine learning, recommendation systems, computational advertising, etc.

Peking University — Beijing, China

Bachelor of Science, ***Statistics****; Bachelor of Science,* ***Psychology*** *(Double degree)* — *Jul. 2013*

Work Experience

Google — Los Angeles, CA

Analytical Linguist Intern — *Jun. – Sep. 2019*

- Researched and summarized user-task interaction metrics from raters' tracking events to improve rating quality
- Developed an R package and a Python script to automatize standard analyses and visualizations
- Wrote technical documentation for future reference, and presented project to the team

Kellogg School of Management, Northwestern University — Evanston, IL

Research Assistant — *Aug. 2018 – Feb. 2019*

- Part-time research assistant collaborated with professors to help data collection for research projects
- Wrote Python scripts to crawl sports competition results from 5 websites and organized the crawled data
- Wrote Python scripts to extract text data from pdf files and integrate datasets using fuzzy string matching

Baidu — Beijing, China

Data Analyst Intern — *Oct. – Dec. 2014*

- Analyzed conversion rate of a mobile App within different time windows and regions using SQL and R
- Summarized the data into a comprehensive report with tables, graphs, and flow charts

Data Challenge Projects

Architecture highlights in Shanghai — Shanghai library open data challenge

Team lead, back-end developer — *May – Aug. 2019*

- Led a team of 7 to develop a website featuring architectures of historical importance in Shanghai
- Developed back-end code (implemented in Python/Django) to categorize architectures based on text descriptions
- Managed weekly updates, participated in discussion of product design, and prepared final presentation

Word evolution in ancient Chinese poems — Shanghai library open data challenge

Individual project — *May – Aug. 2018*

- Developed a website aiming to help researchers gain insights into word evolution, style change, and social evolution reflected in ancient Chinese poems over a long period of time
- Independently came up with the idea, designed features, developed applications, and wrote documentation
- Implemented website using Python/Django and visualized data using R Shiny

Research Projects

Rational models of eye movements in reading

Research Assistant at Department of Linguistics, Northwestern University, Evanston, IL *Sep. 2015 – present*

- To understand how eye movement control and word identification work in reading, we use rational analysis with computational modeling to study how different sources of information (visual, linguistic, and contextual) act interactively to identify words and influence eye movement decisions.
- Techniques: *Bayesian inference, language models, reinforcement learning*

Inferring sentence comprehension from eye movements in reading

Visiting Student at Department of Brain and Cognitive Sciences, MIT, Cambridge, MA *Aug. 2018 – Dec. 2018*

- Although psycholinguistic experiments suggest a strong eye-mind link, the evidence is mostly based on aggregated data. We predict readers' comprehension of individual sentences from their eye movements using machine learning and neural networks, and evaluate how well the predictive models generalize to new readers and new sentences.
- Techniques: *Machine learning, convolutional neural network (CNN), language models, model selection*

Natural language processing for predicting readmission in pediatric ICU

Research Assistant at Feinberg School of Medicine, Northwestern University, Chicago, IL *Sep. 2016 – Aug. 2017*

- Useful information for predicting readmission after pediatric ICU hospitalization can come from electronic health records, especially social workers' notes. We formalize the idea by extracting text features from this unstructured data and implementing classifiers to predict readmission probability.
- Techniques: *Classification with unbalanced classes, natural language processing (bag-of-words model, sentiment analysis)*

Learning with auxiliary less-noisy labels

Research Intern at Institute of Automation, Chinese Academy of Sciences, Beijing, China *Apr. – Sep. 2014*

- We propose a learning method that considers both noisy labels and auxiliary less-noisy labels available in a small portion of the training data. Our method outperforms traditional classifiers that do not explicitly consider the auxiliary less-noisy labels.
- Techniques: *Matlab, ExpectationMaximization algorithm, logistic regression, crowd-sourcing*

Publications

1. **Duan, Y.**, & Bicknell, K. (2019). A rational model of word skipping in reading: Ideal integration of visual and linguistic information. In *Proceedings of the 41th Annual Conference of the Cognitive Science Society*: 275-281. ***Winner of best Computational Modeling paper in Perception & Action.***
2. **Duan, Y.**, & Bicknell, K. (2017). Refixations gather new visual information rationally. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*: 301-306.
3. Yu, H., **Duan, Y.**, & Zhou, X. (2017). Guilt in the eyes: Eye movement and physiological evidence for guilt-induced social avoidance. *Journal of Experimental Social Psychology, 71*, 128-137.
4. ***Duan, Y.**, & *Wu, O. (2016). Learning with auxiliary less-noisy labels. *IEEE Transactions on Neural Networks and Learning System*, *28(7)*, 1716-1721. (* indicates equal contributions.)
5. Luo, Y., **Duan, Y.**, & Zhou, X. (2015). Processing rhythmic pattern during Chinese sentence reading: an eye movement study. *Frontiers in Psychology, 6*, 1881.
6. Wang, L., **Duan, Y.**, Theeuwes, J., & Zhou, X. (2014). Reward breaks through the inhibitory region around attentional focus. *Journal of Vision, 14(12):2*, 1-7.

Conference Presentations (excluding those with proceedings)

1. Sheng, Y., **Duan, Y.**, & Wu, F. Corpus-based Analysis of Complement Coercion in Mandarin Chinese. Poster presentation at the 25th annual conference on Architecture and Mechanisms for Language Processing (AMLaP), Moscow, Russia, 6-8 September 2019.
2. **Duan, Y.**, Berzak, Y., Bicknell, K., & Levy, R. Inferring sentence comprehension from eye movements in reading. Poster presentation at the 32nd annual CUNY Conference on Human Sentence Processing, University of Colorado, Boulder, Colorado, 29-31 March 2019.
3. **Duan, Y.**, & Bicknell, K. (2019). A rational model of word skipping in reading: ideal integration of visual and linguistic information. Poster presentation at the 32nd annual CUNY Conference on Human Sentence Processing, University of Colorado, Boulder, Colorado, 29-31 March 2019.

4. **Duan, Y.**, & Bicknell, K. (2016). Word identification in reading is constructive: Refixations seek new visual information. Poster presentation at the 22nd annual conference on Architecture and Mechanisms for Language Processing (AMLaP), Bilbao, Spain, 1-3 September 2016.
5. Hu, J., Liu, J., **Duan, Y.**, Zhao, C., Gong, X., Xiang, Y., Jiang, C., & Zhou, X. (2014). Resting-state functional connectivity indexes emotion recognition bias. Poster presentation at the 20th Annual Meeting of the Organization for Human Brain Mapping (OHBM), Hamburg, Germany, 8-12 June 2014.
6. **Duan, Y.**, Yu, H., & Zhou, X. (2014). Avoiding eyes reveals guilty heart: An eye movement study on interpersonal guilt. Poster presentation at the 6th Chinese International Conference on Eye Movements (CICEM), Beijing, China, 5-9 May 2014.
7. *Feng, W., ***Duan, Y.**, Luo, Y., & Zhou, X. (2013). When language hurts you: Aggression provoked by rhetorical questions. Poster presentation at the 1st Brain Research Symposium by PKU-IDG/McGovern Institute, Beijing, China, 20-21 August 2013.

Awards & Fellowships

Successful Participants of Mathematical Contest in Modeling (MCM), 2013

First-class Award in Beijing district in China Undergraduate Mathematical Contest in Modeling, 2011

Hui-Chun Chin and Tsung-Dao Lee Chinese Undergraduate Research Endowment (CURE), 2011

Second-class Freshman Scholarship, Peking University, 2009

Skills

Programming languages: Proficient in Python, R, MATLAB; Intermediate in SQL, HTML/CSS, C/C++

Expertise: Machine learning, Data mining, Natural language processing, Text analysis

Other software/software packages: LaTeX, Stanford CoreNLP, Selenium, R Shiny, Django, Praat

Statistics: Linear-mixed models, logistic regression, cluster analysis, Bayesian inference, etc.

Courses & Training

Courses:

- **Linguistics:** Bayesian inference for language scientists; Introduction to computational linguistics; Fundamentals of syntax/phonology/meaning
- **Math/Statistics:** Linear algebra; Mathematical statistics; Probability theory; Mathematical modeling
- **Computer science:** Data structure; Algorithms; Artificial intelligence; Machine learning; Deep learning and reinforcement learning
- **Psychology/Neuroscience:** Fundamentals of neuroscience; Cognitive neuroscience; Functional anatomy of central nervous system; Computational vision, Sensation and perception

Experimental techniques:

- **Eye-tracking:** Design and run eye-tracking experiments in Experiment Builder and analyze data in Data Viewer. Conduct area-of-interest analysis, scan-path analysis, and computational modeling
- **fMRI and ERP:** Experience with data collection. Analyze functional MRI data using SPM. Carry out functional connectivity analysis on resting-state fMRI data using DPARSF and REST

Other training:

- **Kavli Summer Institute in Cognitive Neuroscience**: Lectures and lab sessions on computational perspectives in cognitive neuroscience research on language processing and decision making *(Jul. 2017, UC Santa Barbara)*
- **Winter School on Computational Neuroscience**: A series of introductory courses to models of individual neurons, neural circuits and networks well-known in the field of computational neuroscience *(Dec. 2012, Shanghai Jiao Tong University)*