





# ZHE YU

## Resume

### CONTACT

 Raleigh, North Carolina  
 +1 (919) 949-1382  
 [zyu9@ncsu.edu](mailto:zyu9@ncsu.edu)  
 <http://azhe825.github.io>

*I believe the future of AI is not replacing humans, but, rather, better supporting humans with automated intelligences. Hence, my carrier goal is the creation of "human in the loop" machine learning environments.*

## Research Experience

*Current, from August 2015*

**PhD Scholar**, The RAISE Lab, Department of Computer Science, NC State University.

*Working as a member of RAISE Lab, following the instruction of Dr. Menzies, my primary research is to apply machine learning algorithms to support human retrieving desired information from big data with less effort. (A "finding needles in a haystack" problem.)*

- Developed a method called FASTREAD which reduces the review cost. FASTREAD outperforms the state-of-the-art methods in litigation and medicine.*
- A tool has been developed to implement FASTREAD, which can be found at <https://github.com/fastread/src>.*
- Same idea applied to solve software security vulnerability prediction problem and test case prioritization problem.*

*August 2015 – August 2017*

**Graduate Research Assistant**, The SeBIG Lab, LexisNexis and NC State University.

*Member of a new lab, called "SeBig" (Software Engineering for Big Data), established as joint research collaboration between LexisNexis and NC State. Currently working on test case optimization.*

*March 2014 – August 2014*

**Research Intern**, Department of Computer Science, Shanghai Jiaotong University

*Worked as a research intern under Dr. Yuan. Conducted several experiments on stock index futures data. Established a feature selection scheme with low-rank approximation and sparse representation.*

*February 2011 – March 2014*

**Graduate Research Assistant**, The RCIR Lab, Shanghai Jiaotong University.

*Worked as a research assistant under Dr. Su. My research focuses on the disturbance observer (DOB) based control on multi-variable plants.*

- Established a sufficient condition for the closed-loop robust stability of a DOB-based multi-variable control system.*
- Proposed a systematic design procedure of the multi-variable disturbance observer.*
- Validated the efficacy of control method through experiments on a quadrotor system.*

## Education

Aug 2015 – Now **PhD in Computer Science**, NC State University. **Current GPA: 4.0.**  
Sep 2011 – Mar 2014 **M.S. in Control Science and Engineering**, Shanghai Jiaotong University, CHINA.

## Technical Skills

### Programming Languages

*Experienced++* Python, C++, and JS  
*Experienced* Java, Scala

## Publications

- Submitted **FAST<sup>2</sup>: Better Automated Support for Finding Relevant SE Research Papers.** *arXiv preprint* [arXiv:1705.05420](https://arxiv.org/abs/1705.05420) (2018).
- Accepted **Finding Better Active Learners for Faster Literature Reviews.** *Empirical Software Engineering* [arXiv:1612.03224](https://arxiv.org/abs/1612.03224) (2018).
- May 2016 **The BigSE project: lessons learned from validating industrial text mining.** *In Proceedings of the 2nd International Workshop on BIGDSE*, pp. 65-71.
- Nov 2014 **Disturbance Observer Based Control for Linear Multi-variable Systems with Uncertainties.** *Acta Automatica Sinica*, 40(11): 2643-2651, CHINA

## Presentations

- Oct 2017 **Needle in a Haystack (Advanced text mining with ECL).** [2017 HPCC Systems Summit Community Day](#)
- Sep 2017 **Data Balancing for Technologically Assisted Reviews: Undersampling or Reweighting.** *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum* Dublin, Ireland, September 11-14, 2017.
- May 2016 **The BigSE project: lessons learned from validating industrial text mining.** *In Proceedings of the 2nd International Workshop on BIGDSE*, pp. 65-71.

## Working Experience

*May 2018 – Aug 2018*

**Software Engineering Intern, Google, Mountain View**

Image and Caption Embedding Distance

- Found representative images for each entity, by averaging their starburst embeddings.*
- Trained a dual encode model, between entity and image starburst.*
- Designed and tried different metrics to evaluate the model performance.*
- Added a feature to dual encoder framework to support dense feature.*

*May 2017 – Aug 2017*

**Software Engineering Intern, LexisNexis, Raleigh**

- Constructed the software architecture for Python tasks on Amazon Web Service.*
- Got great experiences on AWS Lambda function, S3, EMR, Apache Spark, Livy, Flask.*
- Implemented several machine learning algorithms, e.g. paragraph vector, latent dirichlet allocation, named entity recognition.*

*May 2016 – Aug 2016*

**Software Engineering Intern, LexisNexis, Raleigh**

- Created a sandbox for prototyping new DiscoveryIQ features. (Python + JS + Elasticsearch)*
- Developed new feature, which is called "Open the blackbox", of DiscoveryIQ.*
- Incorporate new feature into current DiscoveryIQ product. (Scala + Spark).*