

# Shang Gao

Phone: 470 301 7095

Email: emailshang@gmail.com

Github: <https://github.com/iamshang1/Projects>

=====

## Oak Ridge National Laboratory – Graduate Research Assistant, Oak Ridge, TN

Jan 2017 – Present

- Develop and deploy deep learning architectures for automated information extraction from cancer pathology reports; techniques include CNN, RNN, and GAN-based approaches
- Developed new state-of-the-art text classification model based on neural self-attention that achieves better performance and runs up to 10x faster than the previous state-of-the-art method
- Develop and apply multi-task, semi-supervised, and uncertainty quantification methods for deep learning models
- Scale deep learning algorithms across multiple GPUs and nodes on Oak Ridge supercomputer clusters

## University of Georgia – Graduate Research Assistant, Athens GA

Aug 2016 – Dec 2016

- Work with interdisciplinary team on human activity recognition project that attempts to classify activity type based on hip-worn accelerometer device
- Developed convolutional-LSTM model that achieves competitive performance on human activity recognition tasks without requiring manual engineering of features

## Noble Systems – Technical Writer, Atlanta GA

Mar 2012 – Jul 2016

- Produce customer-facing online training for a wide range of contact center products, including campaign management software, IVR scripting interfaces, and more
- Maintain and develop structure, templates, procedures, and single-sourcing guidelines for internal, VAR, and customer knowledge bases—content includes product technical specifications, client connectivity information, troubleshooting and configuration guides, database reference tables, and best practices
- Troubleshoot all technical problems related to internal and customer knowledge bases, including issues with HTML/CSS formatting, Team Foundation Server version control, and nightly auto-build and publishing process

=====

## Ph.D. Data Science – University of Tennessee, Knoxville c/o 2021

- Research focus: deep learning for natural language processing

## B.S. Economics – Duke University c/o 2009

- Major in Economics, minor in Film, second minor in Markets and Management

=====

## Papers and Publications

- Alawad M, et. al. {2018} *Multi-task Convolutional Neural Networks for Information Extraction from Cancer Pathology Reports*. In preparation.
- Catanho, M et. al. {2018}. *Discovering sequence coevolution signatures with hierarchical attention networks*. In preparation.
- Gao S, et. al. {2018}. *Neural Architectures for Extractive Text-Based Question Answering*. Under Review in Computing Surveys.
- Gao S, et. al. {2018}. *Hierarchical Convolutional Attention Networks for Text Classification*. Representation Learning for Natural Language Processing, Proceedings of ACL 2018.
- Bhowmik D, et. al. {2018}. *Deep clustering of protein folding simulations*. BMC Bioinformatics.
- Gao S, et. al. {2017}. *Hierarchical attention networks for information extraction from cancer pathology reports*. Journal of the American Medical Informatics Association.

## Awards and Achievements

- 2018 Bredesen Center Entrepreneurship Award
- 2017 Smoky Mountain Data Science Challenge Most Novel Solution
- 2016 Patent for Utilizing Predictive Models to Improve Predictive Dialer Pacing Capabilities (US9723144B1)

## Programming Languages and Skills

- I am proficient in Python, Numpy, Pandas, SciKitLearn, and TensorFlow
- I am comfortable working in Linux, SQL, Java, R, Matlab, PyTorch, PySpark, and MPI
- I can speak conversational Chinese