# 孙晓飞

Stony Brook, NY, 11790
(+1) 3158793236
adoni1203@gmail.com
adoni.github.io

## 教育

2017至今 **博士研究生**, *Stony Brook University*, 美国.

2015–2017 **硕士研究生**, 哈尔滨工业大学, 中国.

2013.9–2014.1 **交换生**, 台湾大学, 中国台湾.

2011–2015 **本科**, 哈尔滨工业大学, 中国.

## 技术能力

编程 Python, C/C++, Prolog, Java, LaTeX.

工具 DyNet, PyTorch, Tensorflow, Scikit-Learn, Pandas, MongoDB, Redis, Selenium, Web Crawler, Web.py.

## Experiences

2017.3–2017.7 **软件工程实习生**, 微众传媒邮箱公司, 北京.

- Build **Smart Choice** System, a recommendation system to suggest goods to sell for WeChat Official Accounts based on their articles.
- Rank a product based on the keywords matched between this product and the articles. Use topic model to avoid the ambiguity of keywords.
- Now it serves as the backend algorithm for Weizoom company.

2015.5–2015.9 **研究实习生**, *HTC*北京研究中心.

- Work on project **User Profiling on E-commerce Websites from Online Review**.
- Use weakly-supervised learning to predict demographic attribute based on demographic mentions.
- For example, if someone said: "`I bought my wife iPhone 6s, and it worked well`". Then the consumer is probable to be a married male.

2015.5–2015.9 **研究实习生**, 北京语言大学.

- Develop **Semantic Cloud Platform**, a system to detect the vicissitude of Chinese semantic in last ten years.
- Learn word embeddings in each year and take the nearest neighbors as the possible semantics. Cluster potential semantics with AP clustering. Regard the clustering results as the multiple semantics of each word.
- Result reveals influences from subculture (new language), big company (like "Apple"), and so on.

## 代表性项目

2016 **Content Enhanced Network Embedding (CENE)**, [Code] .

- Existing methods cannot utilize multi-modal data like text or images, which is common in social networks.
- By joint optimizing loss function in the node-node edge and node-content edge and mapping multi-modal content into the same embedding space, CENE can leverage structure and content information in a social network to improve the embedding accuracy.
- Applications on node classification show that CENE significantly improved the F1 score from 0.72 to 0.79 when comparing to state-of-art models, e.g. node2vec, TriDNR and TADW.

## 发表文章

[1] A General Framework for Content-enhanced Network Representation Learning[J]. **X Sun**, J Guo, X Ding, T Liu. arXiv., [PDF].

[2] Personalized Microtopic Recommendation on Microblogs[J]. Y Li, J Jiang, T Liu, M Qiu, **X Sun**. ACM TIST. [PDF].

[3] Gender Identification on Social Media[C]. **X Sun**, X Ding, T Liu. SMP 2014. [PDF].

## 获奖情况

2015 最佳论文, 第四届全国社会媒体处理大会.

2014 微软小学者.