

Xiaofei Sun

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

Education

- 2017–Present **Ph.D. of Computer Science**, *Stony Brook University*, United States.
2015–2017 **Master of Computer Science**, *Harbin Institute of Technology*, China.
2013.9–2014.1 **Exchange Study of Computer Science**, *National Taiwan University*, Taiwan.
2011–2015 **Bachelor of Computer Science**, *Harbin Institute of Technology*, China.

Tech Skills

- Programming Python, C/C++, Prolog, Java, \LaTeX .
Tools DyNet, PyTorch, Tensorflow, Scikit-Learn, Pandas, MongoDB, Redis, Selenium, Web Crawler, Web.py.

Experiences

- 2017.3–2017.7 **Software Engineering Intern**, *Weizoom Media Company*, Beijing, China.
 - 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 **Data Science Intern**, *HTC Research & Innovation Center*, Beijing, China.
 - 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 **Data Science Intern**, *Beijing Language and Culture University*, China.
 - 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.

Selected Project

- 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.

Publications

- [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](#)].

Honors & Awards

- 2016 National Scholarship for Graduate, Harbin Institute of Technology, China.
2015 Best Paper Award, *Chinese National Conference on Social Media Processing 2015*.
2014 Microsoft Research Asia Young Fellow Scholarship Award.