Sia, Xin Yun Suzanna

[Web Page] •ssia1@jhu.edu •Github•+14434499249

EDUCATION

PhD Computer Science (NLP), Johns Hopkins University 2018
- Deep latent variable modeling, Topic Modeling, Mutilingual Information Retrieval, Argument and Dialogue. Advisor: Kevin Duh

M.Tech Knowledge Engineering, National University of Singapore Jan 2014 – Dec 2016
- Awarded Honors (Distinction)
- Thesis: An Expert System for Energy Efficient Resource Management. Advisor: Zhu Fangming

BSc. A.I & Psychology, University of Edinburgh Sep 2009 – Jul 2013
- Awarded First Class Honors
- Awarded Best CS Thesis, Video Meeting Search Interfaces. Advisor: Steve Renals

WORK EXPERIENCE

Research Intern, Facebook AI Applied Research Jun 2020 – Oct 2020
- Multimodal Hatespeech Detection, multimodal indexing and retrieval. Advisor: Lambert Mathias

NLP Research Engineer, DSO National Laboratories, Singapore Jun 2015 – Aug 2018
- Defence related NLP projects: Information Retrieval, conversational agents, recommender systems and social computing. Advisor: Chieu Hai Leong

Research Assistant, Singapore Management University Aug 2016 – Dec 2016
- Clustered user sub-groups via collaborative filtering and Variational Inference for Probabilistic Matrix Factorization. Advisor: Jiang Jing

Psychologist, DSO National Laboratories, Singapore Jul 2013 – Jun 2015
- Conducted experiments to evaluate various cognitive and social sensing systems including wearable technology, Microsoft Kinect, and cognitive test batteries. Human factors and usability studies.

Visiting Researcher, Stanford University Mar 2012 – Jul 2012
- Designed metric and implemented the algorithm for automatically scoring a sequence of decisions in an online choice based task. Advisor: Dan Schwartz

AWARDS / SCHOLARSHIPS

Best Poster, NYAS Speech and Dialogue Symposium, (1/60) 2019
PhD Fellowship, DSO National Labs (declined for JHU RAship) 2018
2nd Place, NUS/NUHS-MIT Datathon 2018
3rd Place, 26th Association for Computing Machinery CIKM Analyticup 2017
Kinetic Award, GPA, GPA, DSO National Labs 2016, 2017, 2018
JASSO Scholarship, Japan Government 2016, 2017
Best Computer Science Final Year Project, University of Edinburgh 2013
Best Poster, Lovelace Colloquium (British Universities), British Computing Society 2013
Overseas Undergraduate Scholarship, Defence Science Technology Agency 2009 – 2013

PUBLICATIONS

Sia, S., Dalmia, A., Mielke, S., Tired of Topic Models? Clusters of Pretrained Word Embeddings Make for Fast and Good Topics too! *EMNLP 2020* 2020

Sun, S., Sia, S., Duh, K., **CLIReval: Evaluating Machine Translation as a Cross-Lingual Information Retrieval Task,** *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* 2020

Xie, J., Sia, S., Garcia, P., Povey, D., Khundanpur, S. **Mixture of Speaker-type PLDAs for Children's Speech Diarization** *arXiv:2008.13213* 2020

Sia, S., Jaidka, K., Duh, K., **A semi-supervised hierarchical generative model of argumentation.** *Natural Language, Dialog and Speech Symposium, New York Academy of Sciences.* 2019

Sia, S., Li L.J.A., **Hierarchical Module Classification in Mixed Initiative Conversational Agent System**, *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. ACM, 553–562. 12* [Pdf] 2017

PROJECTS

Infinite PCFGs for social grammars Ongoing

- Hierarchical Dirichlet Process for learning non-parametric grammars. Grammars are evaluated on sentiment classification using easy-first parsing tree LSTMs. [Interim write-up]
- Currently working on extending to word or sequence embeddings.

Patient Conversation Simulator 2017

- Simulated patient conversation for medical education, demonstrated system with explainable response selection at the 26th ACM Conference of International Knowledge Management.

Undergraduate Admissions QA System, National University of Singapore 2017

- Programmed a live chat system for the Undergraduate Computing admissions cycle. Deployed on NUS School of Computing Website and Facebook page. [Twitter]

Computational Propaganda Project Feb 2016 – May 2016

- Developed anomaly detection algorithms for Taiwan case study in the Computational Propaganda Project by Oxford Internet Institute. [Guardian-article]

SERVICE

Reviewer EMNLP19; **Outstanding Reviewer** 2018

Teaching Assistant, CS462/682 Deep Learning, Johns Hopkins 2020

Teaching Assistant (93.3% rating), CS601.465/665 NLP, Johns Hopkins 2019

Student Mentor, Shibaura Institute of Technology, Saitama, Japan, [Report][Program] 2016

SKILLS/TOOLS

Programming: Python, Java, Cython, bash, Octave/Matlab, R, HTML5/Javascript/CSS, Haskell

Research: Git, Scikit-learn, SciPy, NumPy, Pandas, Gensim, Stanford NLP, wordnet/nltk, pymc3, NetworkX, Thulac, Protege, DBPedia, SPARQL, Mallet (Java), EJML (Java), LaTeX

Deep Learning Frameworks: Tensorflow, PyTorch, AllenNLP, Keras

Web: Amazon Web Services, Heroku, NodeJS, ExpressJS, Flask, Scrapy, Selenium, Django

Databases: MongoDB, PostgreSQL, ElasticSearch, Neo4j

Psychology: Amazon Turk, experimental design, participant recruitment, interface design

GRADUATE COURSEWORK

Natural Language Processing, Advanced Bayesian Statistics (Non-parametrics), Linguistic and Sequence Modelling, Distributed Systems, Introduction to Statistics, Information Theory, Randomized Algorithms, Object-Oriented Architecture and Design

ADDITIONAL COURSES / CERTIFICATIONS

Deep Learning Specialization (16 weeks), Coursera/DeepLearning.AI, [Cert] 2018

Machine Learning Specialization (25 weeks), Coursera/University of Washington [Cert] 2017

Machine Learning Summer School, Max Planck Institute of Intelligent Systems, Germany 2017

ITIL Foundation in IT Service Management, AXELOS, UK, [Cert] 2015