shawntan

♂ Tan Jing Shan, Shawn
✉ 5 Jalan Asas, Singapore 678763
✓ DOB: 1987/12/17
(+65) 93877026

online

shawn@wtf.sg
shawntan@github
tanshawn@linkedin
gscholar@57Nf7EYAAAAJ

languages

english (proficient)
mandarin

programming

♥ Python, JavaScript, Java, Ruby, C, Perl

technical skills

HTML & CSS, LaTeX, Linux System Administration

about

I'm a graduating PhD student at Montreal Institute of Learning Algorithms (MILA). I focus on natural language processing and structured prediction, primarily with the goal of compositional generalisation.

research interests: natural language processing, compositional generalisation, graphical models, amortised variational inference, structured prediction

education

2018-present **PhD in Computer Science** — University of Montreal
Focus on compositional generalisation through inducing latent structure in neural models. Initial work on Ordered Neurons showed that with language modelling objectives, grammatical structures can be induced (won Best Paper at ICLR 2019). Following work focused on encoding and decoding with tree-structured inductive biases. Used VAEs and normalising flows for related structured generation tasks (e.g. molecule generation)

2016-2018 **Masters in Computer Science** — University of Montreal
Studied VAE methods for few-shot adaptation of face images to identity (CelebA). Worked on latent variable natural language generation for langauge modelling and paraphrase generation. Wrote thesis on Latent Variable Language Models, with several baselines on sentence-level language modelling on Penn TreeBank.
Thesis: Latent Variable Language Models

2009-2012 **Bachelors of Computing (BComp)** — National University of Singapore
Major in Computer Science - Upper 2nd Class Honours
Special Programme in Computing (Turing Programme)
Focus Area: Artificial Intelligence

2004-2007 **Diploma in Information Technology** — Ngee Ann Polytechnic, Singapore
Specialisation in Software Engineering - Diploma with Merit

experience

06-09 2022 **MIT-IBM Watson AI Lab** — Research Intern
Project to scale up Universal Transformers (UTs). UT share parameters across layers and have better compositional generalisation properties. Theoretically, an arbitrary depth UT can be Turing-complete. Project focuses on overcoming UT scaling challenges (memory footprint), and multi-GPU / node training.

2014-2016 **Speech Recognition Group, National University of Singapore** — Research Assistant
Working on the machine learning aspects of speech recognition using neural networks, with applications to speaker adaptation and noise robustness.

2013-2014 **Semantics3** — Software/Data Engineer
The world's largest database of product and pricing information. YCombinator-backed (Winter 2013), Acquired by Hearst Magazines 2019

- Classification of products into a hierarchical taxonomy using a cascading naive bayes model built using the Lucene index.
- Name equivalency model used as part of the disambiguation process to determine if two product names are referring to the same thing. Experimented with using conditional random fields for attribute extraction from product names.

awards

ICLR 2019 **Ordered Neurons: Integrating Tree Structures into Recurrent Neural Networks** Best paper
Proposed an inductive bias by ordering the neurons; a vector of master input and forget gates ensures that when a given neuron is updated, all the neurons that follow it in the ordering are also updated. Good performance on inducing grammar from unlabeled text.

talks

2024 **DSO National Laboratories: The New XOR problem** Invited talk
Presentation on Transformer limitations to DSO National laboratories.

2022 **ACL 2022: Unsupervised Dependency Graph Network** Oral presentation
Awarded oral presentation for Unsupervised Dependency Graph Network paper at ACL 2022, Dublin

2019 **Nanyang Technological University: Ordered Neurons** Invited talk
Ordered Neurons presentation to NTU School of Computer Science and Engineering

teaching

winter 2021 **Deep learning project** Project mentor
Supervising and mentoring students through a semester-long deep learning project.

winter 2018, 2019 **Representation learning** Teaching Assistant
Crafting, disseminating, and correcting problem sets for practical and theoretical concepts for deep learning. Weekly sessions of office hours for students who need help with the material.

community involvement

NeurIPS 2021 **Workshop Organiser** Advances in Programing Languages And NeuroSymbolic Systems
Facilitated logistics for online workshop, interfacing NeurIPS system , OpenReview, and GatherTown. Moderated an invited talk session.

NeurIPS 2021 **Advances in Programing Languages And NeuroSymbolic Systems** Workshop Organiser
Facilitated logistics for online workshop, interfacing NeurIPS system , OpenReview, and GatherTown. Moderated an invited talk session.

open source

09-12 2016 **Theano** contributor
Migration of some crucial functionality to new CUDA backend, including advanced indexing.

10 2014 **Neural Turing Machines implementation in Theano** owner
Implementation of system described in "Neural Turing Machines." by Alex Graves, Greg Wayne, and Ivo Danihelka.

publications

conference papers

Sparse Universal Transformer
Shawn Tan, Yikang Shen, Zhenfang Chen
Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.

Unsupervised Dependency Graph Network
Yikang Shen, Shawn Tan, Alessandro Sordoni, Peng Li, Jie Zhou, Aaron Courville
Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022, **Oral presentation**.

Learning to Dequantise with Truncated Flows
Shawn Tan*, Chin-wei Huang, Alessandro Sordoni, Aaron Courville
International Conference on Learning Representations (ICLR), 2022.

Explicitly Modeling Syntax in Language Models with Incremental Parsing and a Dynamic Oracle
Yikang Shen, Shawn Tan, Alessandro Sordoni, Siva Reddy, Aaron Courville
Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021.

Recursive Top-Down Production for Sentence Generation with Latent Trees
Shawn Tan*, Yikang Shen*, Timothy J O'Donnell, Alessandro Sordoni, Aaron Courville
Findings of EMNLP, 2020.

Ordered Memory
Yikang Shen*, Shawn Tan*, Arian Hosseini*, Zhouhan Lin, Alessandro Sordoni, Aaron C Courville
Advances in Neural Information Processing Systems, 2019.

Ordered Neurons: Integrating Tree Structures into Recurrent Neural Networks
Yikang Shen*, Shawn Tan*, Alessandro Sordoni, Aaron Courville
International Conference on Learning Representations (ICLR), 2019, **Awarded Best Paper**.

Improving explorability in variational inference with annealed variational objectives
Chin-Wei Huang, Shawn Tan, Alexandre Lacoste, Aaron Courville
Advances in Neural Information Processing Systems, 2018.

Towards implicit complexity control using variable-depth deep neural networks for automatic speech recognition
Shawn Tan, Khe Chai Sim
2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016.

Learning utterance-level normalisation using Variational Autoencoders for robust automatic speech recognition
Shawn Tan, Khe Chai Sim
Spoken Language Technology Workshop (SLT), 2016 IEEE, 2016.

Improving the interpretability of deep neural networks with stimulated learning
Shawn Tan, Khe Chai Sim, Mark Gales
IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2015.

others

ModuleFormer: Learning Modular Large Language Models From Uncurated Data
Yikang Shen, Zheyu Zhang, Tianyou Cao, Shawn Tan, Zhenfang Chen, Chuang Gan
arXiv preprint arXiv:2306.04640

.

Icentia11K: An Unsupervised Representation Learning Dataset for Arrhythmia Subtype Discovery
Shawn Tan, Guillaume Androz, Ahmad Chamseddine, Pierre Fecteau, Aaron Courville, Yoshua Bengio, Joseph Paul Cohen
arXiv preprint arXiv:1910.09570

.

Investigating Biases in Textual Entailment Datasets
Shawn Tan, Yikang Shen, Chin-wei Huang, Aaron Courville
arXiv preprint arXiv:1906.09635

.

Inferring Identity Factors for Grouped Examples
Shawn Tan, Christopher J Pal, Aaron Courville

.

Generating contradictory, neutral, and entailing sentences
Yikang Shen, Shawn Tan, Chin-Wei Huang, Aaron Courville
arXiv preprint arXiv:1803.02710

.

Self-organized hierarchical softmax
Yikang Shen, Shawn Tan, Chrisopher Pal, Aaron Courville
arXiv preprint arXiv:1707.08588

.

reports

Latent variable language models
Shawn Tan
Masters' thesis, 2019.

grabsmart: A User-centric Web Information Extraction System
Shawn Tan
Undergraduate Thesis, 2012.