

JOHN CHEN

johnc@cs.toronto.edu
johntiger1.github.io

EDUCATION

UNIVERSITY OF TORONTO. DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE. FOCUS AREA IN FAIRNESS FOR CLINICAL NATURAL LANGUAGE PROCESSING. ADVISED BY DR. FRANK RUDZICZ. 4.0 GPA (A+ AVERAGE). SEPT 2019-ONGOING. TWO PAPERS PUBLISHED AT EMNLP2020; ONE BEST PAPER (FIRST AUTHOR), ONE ORAL.

UNIVERSITY OF TORONTO. BACHELOR OF SCIENCE. SPECIALIST IN COMPUTER SCIENCE. CUMULATIVE GPA: 3.98/4.00. FOCUS AREA IN ARTIFICIAL INTELLIGENCE. SEPT 2014-MAY 2019.

LIST OF RESEARCH PUBLICATIONS

ACCEPTED / PUBLISHED

1. **John Chen**, Ian Berlot-Attwell, Safwan Hossain, Xindi Wang and Frank Rudzicz. "Analyzing Text Specific vs Blackbox Fairness Algorithms in Multimodal Clinical NLP". To appear in *Proceedings of 3rd Clinical Natural Language Processing Workshop, Empirical Methods in Natural Language Processing Conference, November 2020. Best Paper award (Selectivity: 3 awarded/48 submissions)*.

Abstract:

Clinical machine learning is increasingly multimodal, collected in both structured tabular formats and unstructured forms such as freetext. We propose a novel task of exploring fairness on a multimodal clinical dataset, adopting equalized odds for the downstream medical prediction tasks. To this end, we investigate a modality-agnostic fairness algorithm - equalized odds post processing - and compare it to a text-specific fairness algorithm: debiased clinical word embeddings. Despite the fact that debiased word embeddings do not explicitly address equalized odds of protected groups, we show that a text-specific approach to fairness may simultaneously achieve a good balance of performance and classical notions of fairness. We hope that our paper inspires future contributions at the critical intersection of clinical NLP and fairness. The full source code is available here: https://github.com/johntiger1/multimodal_fairness

2. Rohan Bhambhoria, Luna Feng, Dawn Sepehr, **John Chen**, Conner Cowling, Sedef Akinli Kocak, Elham Dolatabadi. "A Smart System to Generate and Validate Question Answer Pairs for COVID-19 Literature". To appear in *Proceedings of 1st Scholarly Document Processing Workshop, Empirical Methods in Natural Language Processing Conference, November 2020. Oral Presentation (Selectivity: 9 oral / 20 accepted papers)*

Abstract:

Automatically generating question answer (QA) pairs from the rapidly growing coronavirus-related literature is of great value to the medical community. Creating high quality QA pairs would allow researchers to build models to address scientific queries for answers which are not readily available in support of the ongoing fight against the pandemic. QA pair generation is, however, a very tedious and time consuming task requiring domain expertise for annotation and

evaluation. In this paper we present our contribution in addressing some of the challenges of building a QA system without gold data. We first present a method to create QA pairs from a large semi-structured dataset through the use of transformer and rule-based models. Next, we propose a means of engaging subject matter experts (SMEs) for annotating the QA pairs through the usage of a web application. Finally, we demonstrate some experiments showcasing the effectiveness of leveraging active learning in designing a high performing model with a substantially lower annotation effort from the domain experts.

IN PREPARATION / IN SUBMISSION

3. “Exploring Doctor-Patient Style Transfer in Text using Adversarial Methods”. **John Chen** and Frank Rudzicz. In submission to *2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

Abstract:

Style transfer in text is an emerging field in the natural language processing domain. However, existing approaches have focused on basic sentence-length style transfer and have generally avoided explicitly defining both style and the style transfer task. This paper is the first to apply the task of style transfer to the biomedical domain, on real-life data collected from MedHelp and the MIMIC-III dataset. Our goal is to improve healthcare outcomes via analysis of stylistic properties of doctors and patients, with the ultimate goal of constructing a robust mapping function between the latent space of concepts and the natural language used by either interlocutor. We present qualitative results, confirming that doctors use more detached and clinical language, while patients are more personal and less technical with their descriptions. We also demonstrate the inadequacy of current methods on nuanced stylistic differences between real-life corpora. We use these negative results to posit an alternate formulation of the style transfer problem, including a separation between sentence-level and document-level style characteristics, paving the way for future work in this rich domain.

4. “Extending local superlinear convergence results of some Quasi Newton methods under noise and random switching”. Ian Berlot-Attwell and **John Chen**. In preparation for submission to *International Conference on Machine Learning 2021*.

Abstract:

In this paper, we prove convergence results for Quasi Newton optimization methods under certain stochastic conditions. We prove that the addition of bounded random noise does not affect known convergence results, and that randomly switching between Quasi-Newton update rules also maintains the overall convergence properties. We also provide experimental results that support the validity of our proof. Experimentally, we also discovered that introducing random switching of the QN-update rule (between Greenstadt update and BFGS update) can improve convergence and stability.

5. “Generative Visual Question Answering using Cross-Modal Visual-Linguistic Embeddings”. **John Chen** and Yushi Guan. In submission to *Computer Vision and Pattern Recognition 2021*.

Abstract:

In this paper, we investigate the performance of several recent state of the art vocabulary-based Visual Question Answering models on the task of answer generation. Answer generation is a generalization of vocab-based QA, where the model must generate the answer token by token, and relaxes the strong assumptions made in the vocab-based formulation. We experiment with adding two different types of generative decoder heads to two state of the art vocab-based models, VL-BERT and LXMERT. In both cases, the decoder generator is not able to match the performance of the vocab-based original, although performance can remain competitive. Our findings indicate that simpler decoder heads can be more

performant, demonstrating the importance of carefully crafting a training procedure or doing curriculum learning to jointly optimize the massive models. Finally, we confirm the intuition that answer generation is more challenging than answer selection and indicate that much progress remains to be made on the search for a general question answer system that is truly able to understand the question and express the answer fluently.

6. "Reinforcement Learning for Query Synthesis for Pool-Based Active Learning". **John Chen** and Jieyu Lin. In preparation for submission to *Conference on Neural Information Processing Systems 2021*.

Abstract:

A thorny and universal open problem that continues to plague deep learning is the scarcity of labelled data. Despite copious amounts of unlabelled data (e.g. images on the Internet), labelling is a procedure that is costly, labour-intensive, time-consuming, and not scalable. Pool-based active learning thus deals with picking good data points from the unlabelled dataset to label. In batch-mode active learning, this discrete choice of k unlabelled data points is called a query. Recent work such as Variational Adversarial Active Learning (VAAL) has improved performance by learning good representations for the active learning problem. However, we found that both VAAL and simple heuristic query functions such as uncertainty sampling are still unequipped to deal with the interaction effects between utility of data points within a batch query. In this paper, we investigate the correlated query problem in batch mode active learning, and furthermore propose learning the heuristic query selection function using reinforcement learning. Specifically, we explore adding class-conditional labels to VAAL, and propose two novel architectures (deep Q network and policy gradient network) for determining a good querying policy. We show that addressing these fundamental issues can improve performance compared to existing work and baselines.

List of Invited Research Presentations:

- "Refined Online Bayesian Ontology Query Algorithm." *Canadian Undergraduate Computer Science Conference 2017*. June 2017. Toronto, Canada.
 - Invited to present my summer research project with Professor Michael Brudno at SickKids Hospital at national undergrad computer science conference.
- "Fairness in multimodal clinical machine learning." *University of Toronto Graduate Visit Day*. March 2020. Toronto, Canada.
 - Selected to represent Natural Language Processing and Computational Linguistics subdisciplines to prospective students. Attendance of 50+ (in-person and remote).
- "Fairness vs Performance in Machine Learning on Text and Numbers." *Lyft Lightning Talk*. June 2020. San Francisco, California (remote).
 - Invited to give a research presentation to Lyft machine learning team for dynamic pricing.
- "Exploring the Fairness-Performance Trade-off When Combining Clinical Notes with Structured Medical Data." *Vector Institute Thursday Seminar Series*. June 2020. Toronto, Canada (remote).
 - Invited to give a research presentation on my thesis work at the Vector Institute.
- "Fairness in Machine Learning: Primer and Research Directions." *CSC384: Intro to Artificial Intelligence, Guest Lecture*. August 2020. Toronto, Canada (remote).
 - Invited to give a guest presentation to students enrolled in artificial intelligence course, introducing foundational concepts in fairness in machine learning as well as my current research in the domain of clinical multimodal machine learning.

- “Analyzing Text Specific vs Blackbox Fairness Algorithms in Multimodal Clinical NLP”. *3rd Clinical Natural Language Processing Workshop, Empirical Methods in Natural Language Processing Conference*. November 2020. Punta Cana, Dominican Republic (remote).
 - “**Best Paper**” presentation for paper accepted at workshop

ADDITIONAL RESEARCH ACTIVITIES

INDUSTRY RESEARCH:

- Graduate Applied Scientist Intern at Amazon Alexa - Entity Resolution. Toronto, Canada. Entity Resolution Self-Service Team (LoBERT). Fall 2020.
 - Mentor: Dr. Majid Laali
 - Project: Natural Language Processing research into utilizing ASR n-best hypotheses to improve Entity Resolution for Amazon Alexa. Designed novel neural network architecture that achieves 15% relative performance improvement compared to production model. Model is currently being put into production for Smart Home devices. Preparing industry research paper for submission to Association of Computational Linguistics (ACL) 2021 conference.

REVIEWING SERVICE:

- Reviewer for EMNLP 2020 (Multimodal and Question Answering Track)
- Reviewer for NAACL 2020

EDITOR SERVICE:

- Editor-in-Chief of 3rd Annual Edition of Review of Undergraduate Computer Science (RUCS)
 - RUCS is a peer-reviewed journal and club for undergraduate computer science research
 - Expanded journal submission pool beyond University of Toronto; received extended abstracts from MIT and UBC for first time ever
 - Managed multidisciplinary team to produce 3rd annual edition, available at: http://rucss.ca/archive_2017
 - Hosted first ever Undergraduate Research in Computer Science Conference attended by 100+ students and with faculty keynotes by Dr. Roger Grosse and Dr. Ishtiaque Ahmed

MENTORSHIP:

- Invited to represent industry and academic studentships as a student panelist at the 2nd AI Squared Conference (November 2019). Attended by ~100 students.
- Invited to serve as Graduate Research Mentor at RUCS Research-a-thon (March 2020)

DEPARTMENTAL SERVICE:

- Invited to represent MSc. program as the MSc. panelist at *Graduate Programs Information Session*, University of Toronto, October 2020 (remote). Attended by 90+ students.
- Host and Moderator and Panelist of *Graduate Research Info Session Roundtable*, University of Toronto (July 2020). Interviewed 6+ PhD, MSc. and MScAc students. Attended by 35+ students, many of them funded by NSERC or UTEA USRA.

IMPROVING EQUITY AND INCLUSION IN RESEARCH:

- Director of Communications and Research Evangelist at Review of Undergraduate Computer Science
 - Mentored new editor-in-chief to continue club emphasis on improving diversity and inclusion in undergraduate research, especially amongst younger students and underrepresented minorities
 - Provide communications support and hosting / moderation at multiple events, each with average attendance of 30+ people:
 - September 2020: Virtual CS Frosh
 - March 2020: Research-a-thon
 - Jan 2020: Undergrad CS Research Social
 - September 2019: CS Frosh
 - March 2019: Research-a-thon talk
 - Jan. 2019: Undergrad CS Research Social
 - Increased research accessibility by contributing to Undergraduate Research Guide, which has been viewed 1000+ times
 - <http://rucs.ca/assets/info/Undergraduate-Research-Guide.pdf>