Mingda Chen

Website: mingdachen.github.io
Email: chenmda@gmail.com

Research Interests

Natural language processing and machine learning

Employment

Research Scientist, 2022.07 – present
Meta AI, NYC, New York, USA

Education

Ph.D. in Computer Science, 2016.09 – 2022.07
Toyota Technological Institute at Chicago, Chicago, IL, USA
Thesis: *Leveraging Natural Supervision for Language Representation Learning and Generation*
Advisor: Kevin Gimpel

M.Sc. in Computer Science, 2016.09 – 2018.09
Toyota Technological Institute at Chicago, Chicago, IL, USA
Advisor: Kevin Gimpel

B.Sc. in Mathematics and Applied Mathematics, 2011.09 – 2015.07
Zhejiang University, Hangzhou, Zhejiang, China

Honors

Google Ph.D. Fellowship in Natural Language Processing, 2020–2022

Publications

See my profiles on google scholar and semantic scholar for a full list of publications

CONFERENCE PAPERS

15. **Mingda Chen**, Jingfei Du, Ramakanth Pasunuru, Todor Mihaylov, Srini Iyer, Veselin Stoyanov, and Zornista Kozareva. Improving in-context few-shot learning via self-supervised training. In *Proceedings of NAACL*, 2022.

14. **Mingda Chen**, Zewei Chu, Sam Wiseman, and Kevin Gimpel. SummScreen: A dataset for abstractive screenplay summarization. In *Proceedings of ACL*, 2022.

13. **Mingda Chen**, Sam Wiseman, and Kevin Gimpel. WikiTableT: A large-scale data-to-text dataset for generating Wikipedia article sections. In *Findings of ACL*, 2021.

12. **Mingda Chen***, Zewei Chu*, Karl Stratos, and Kevin Gimpel. Mining knowledge for natural language inference from Wikipedia categories. In *Findings of EMNLP*, 2020.

11. Zhenzhong Lan, **Mingda Chen**, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. ALBERT: A lite BERT for self-supervised learning of language representations. In *Proceedings of ICLR*, 2020.

10. Zewei Chu, **Mingda Chen**, Jing Chen, Miaosen Wang, Kevin Gimpel, Manaal Faruqui, and Xiance Si. How to ask better questions? A large-scale multi-domain dataset for rewriting ill-formed questions. In *Proceedings of AAAI*, 2020.

9. **Mingda Chen***, Zewei Chu*, Yang Chen, Karl Stratos, and Kevin Gimpel. EntEval: A holistic evaluation benchmark for entity representations. In *Proceedings of EMNLP*, 2019.

8. **Mingda Chen***, Zewei Chu*, and Kevin Gimpel. Evaluation benchmarks and learning criteria for discourse-aware sentence representations. In *Proceedings of EMNLP*, 2019.

7. **Mingda Chen**, Qingming Tang, Sam Wiseman, and Kevin Gimpel. Controllable paraphrase generation with a syntactic exemplar. In *Proceedings of ACL*, 2019.

6. **Mingda Chen**, Qingming Tang, Sam Wiseman, and Kevin Gimpel. A multi-task approach for disentangling syntax and semantics in sentence representations. In *Proceedings of NAACL*, 2019.

5. **Mingda Chen**, Qingming Tang, Karen Livescu, and Kevin Gimpel. Variational sequential labelers for semi-supervised learning. In *Proceedings of EMNLP*, 2018.

4. **Mingda Chen** and Kevin Gimpel. Smaller text classifiers with discriminative cluster embeddings. In *Proceedings of NAACL*, 2018.

WORKSHOP PAPERS AND PREPRINTS

3. **Mingda Chen** and Kevin Gimpel. TVStoryGen: A dataset for generating stories with character descriptions. *arXiv preprint arXiv:2109.08833*, 2021.

2. **Mingda Chen**, Sam Wiseman, and Kevin Gimpel. Exemplar-controllable paraphrasing and translation using bitext. *arXiv preprint arXiv:2010.05856*, 2020.

1. **Mingda Chen** and Kevin Gimpel. Learning probabilistic sentence representations from paraphrases. In *Proceedings of the 5th Workshop on Representation Learning for NLP (RepL4NLP)*, 2020.

Professional Service

Reviewer

- **Workshops**: SRW at NAACL (2021–2022), Creative-Summ at COLING 2022, RepL4NLP at ACL 2021, W-NUT at EMNLP 2020
- **Conferences**: AAAI, ACL, ACL Rolling Review, EMNLP, NAACL, NLPCC
- **Journals**: *Computational Linguistics*, *IEEE Transactions on Audio, Speech and Language Processing (TASLP)*, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*

Teaching Experience

Teaching Assistant

- TTIC 31210: Advanced Natural Language Processing (Graduate), Spring 2019

Internship Experience

Research Intern, Summer 2021
Facebook AI, Menlo Park, CA, USA (remote)
Project: Improving few-shot performance of pretrained language models (Chen et al., 2022b)

Research Intern, Summer 2019
Google AI, Los Angeles, CA, USA
Project: Large-scale language model pretraining (Lan et al., 2020)