

EDUCATION

- **Nanyang Technological University (NTU)** Singapore
Ph.D. in Computer Science Jan. 2018 – May. 2022
Thesis: Neural Machine Translation with Limited Resources [[Outstanding PhD Thesis](#)]
- **Bangladesh University of Engineering and Technology (BUET)** Dhaka, Bangladesh
Bachelor of Science in Computer Science & Engineering; CGPA: 3.78/4.00 Mar. 2009 – Jul. 2014

PROFESSIONAL EMPLOYMENTS

- **Huawei Research Center** Singapore
Research Scientist Jun. 2022 – Present
- **Nanyang Technological University** Singapore
Graduate Research Student Jan. 2018 – May. 2022
Advisor: [Shafiq Joty](#)
- **Facebook AI Research** Menlo Park, California, United States
Research Intern May 2021 – Sept. 2021
Manager: [Philipp Koehn](#)

RESEARCH PROJECTS

- **Curriculum Training for NMT**: Typical NMT training places all the available corpora into a single pool and samples training examples randomly from it. In this project, we propose a two-stage curriculum training framework for NMT where we leverage the prediction scores of the emerging NMT model for the data selection curriculum.
Publications: [EMNLP-2022](#)
- **BiText Vicinity for Low-Resource NMT**: Our idea here is to exploit the vicinal samples of the given bitext. We intend to propose a new back-translation system which is used to translate the vicinal samples by leveraging the existing relational knowledge as a guide.
Publications: [ACL-IJCNLP-2021](#)
- **Data Augmentation Strategy for Cross-Lingual NLP**: Here we target to solve cross-lingual adaptation problems from a source (language) distribution to an unknown target (language) task distribution assuming it has no training labels in the target language task. We intend to propose a novel data augmentation method for self-supervised learning in zero-resource transfer learning scenarios.
Publications: [ACL-IJCNLP-2021](#)
- **Bilingual Lexicon Induction with Limited/No Supervision**: Given monolingual word embeddings of two languages. We want to find out bilingual-lexicon induction without any supervision or using a small seed dictionary.
Publications: [NAACL-HLT-2019](#), [CL-Journal-2020](#), [EMNLP-2020](#)
- **Coherence Modeling**: We propose novel approaches to solve the coherence modeling problem. We also benchmark the models on both synthetic tasks and downstream applications.
Publications: [ACL-2018](#), [EMNLP-IJCNLP-2019](#), [EACL-2021](#)
- **Speech Acts Recognition in Asynchronous Conversations**: We want to identify the different speech acts classes in the asynchronous conversational analysis.
Publications: [CL-Journal-2018](#), [NAACL-HLT-2019](#)

PUBLICATIONS

- **Tasnim Mohiuddin**, Philipp Koehn, Vishrav Chaudhary, James Cross, Shruti Bhosale, and Shafiq Joty, “Data Selection Curriculum for Neural Machine Translation”, In Findings of the 2022 Conference on Empirical Methods in Natural Language Processing (**EMNLP-22**), Abu Dhabi, UAE. [[PDF](#)]
- M Saiful Bari, **Tasnim Mohiuddin** (*Equal Contributions*), and Shafiq Joty, “UXLA: A Robust Unsupervised Data Augmentation Framework for Cross-Lingual NLP”, In Proceedings of the Annual Meeting of the Association for Computational Linguistics (**ACL-21**). [[PDF](#)]

- **Tasnim Mohiuddin**, M Saiful Bari, and Shafiq Joty, “AugVic: Exploiting BiText Vicinity for Low-Resource NMT”, In Findings of the Annual Meeting of the Association for Computational Linguistics (**ACL-21**). [\[PDF\]](#)
- **Tasnim Mohiuddin**, Prathyusha Jwalapuram, Xiang Lin, and Shafiq Joty, “Rethinking Coherence Modeling: Synthetic vs. Downstream Tasks”, In Proceedings of the 16th conference of the European Chapter of the Association for Computational Linguistics (**EACL-21**). [\[PDF\]](#)
- **Tasnim Mohiuddin**, M Saiful Bari, and Shafiq Joty, “LNMap: Departures from Isomorphic Assumption in Bilingual Lexicon Induction Through Non-Linear Mapping in Latent Space”, In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (**EMNLP-20**). [\[PDF\]](#) [\[Code\]](#)
- **Tasnim Mohiuddin** and Shafiq Joty, “Unsupervised Word Translation with Adversarial Autoencoder”, **Computational Linguistics**, **46:2**, (*presented at ACL-2020*), June 2020. [\[PDF\]](#)
- **Tasnim Mohiuddin** and Shafiq Joty, “Revisiting Adversarial Autoencoder for Unsupervised Word Translation with Cycle Consistency and Improved Training”, In Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL-19**), Minneapolis, USA. [\[PDF\]](#) [\[Code\]](#)
- **Tasnim Mohiuddin**, Thanh-Tung Nguyen, and Shafiq Joty, “Adaptation of Hierarchical Structured Models for Speech Act Recognition in Asynchronous Conversation”, In Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL-19**), Minneapolis, USA. [\[PDF\]](#) [\[Code\]](#)
- Han-Cheol Moon, **Tasnim Mohiuddin** (*Equal Contributions*), and Shafiq Joty, “A Unified Neural Coherence Model”, In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (**EMNLP-19**), Hong Kong, China. [\[PDF\]](#) [\[Code\]](#)
- **Tasnim Mohiuddin**, Shafiq Joty, and Dat Nguyen, “Coherence Modeling of Asynchronous Conversations: A Neural Entity Grid Approach”, In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (**ACL-18**), Melbourne, Australia. [\[PDF\]](#) [\[Code\]](#)
- Shafiq Joty and **Tasnim Mohiuddin**, “Modeling Speech Acts in Asynchronous Conversations: A Neural-CRF Approach.”, **Computational Linguistics** (Special Issue on Language in Social Media), **44:4**, 859 - 894. [\[PDF\]](#) [\[Code\]](#)

PROGRAMMING SKILLS

- **Languages:** Python, C, C++
- **Frameworks:** PyTorch, TensorFlow

SERVICES

- **Reviewer**
ACL, EMNLP, NAACL, AACL, SIGDIAL
- **Teaching Assistant**
Deep Learning for Natural Language Processing: From Theory to Practice [Spring, 2020]
- **Co-Curator**
TEDxNTU
- **Director Of Operations and Logistics**
Graduate Student Council, SCSE-NTU