Lei Xu

PHD Student at MIT EECS

32 Vassar St, Room 32-D712, Cambridge, MA 02139

+1 (617) 909-2935 | leix@mit.edu | www.citr.me | leix28

Research Interests: Machine Learning, Deep Learning, Data Science.

Education

Ph.D, Dept of Electronical Engineering and Computer Science(EECS), MIT GPA: 4.0/4.0 — *Cambridge, MA, USA*

Supervised by Dr. Kalyan Veeramachaneni — *Jun. 2022*

B.E, Dept of Computer Science and Technology, Tsinghua University GPA: 93/100 — *Beijing, China*

Supervised by Prof. Zhiyuan Liu — *Jul. 2017*

Exchange, Computer Science Department, Carnegie Mellon University — *Pittsburgh, PA, USA*

Supervised by Prof. J. Zico Kolter — *Jun. 2016 - Sept. 2016*

Experience

PhD Research Assistant | DAI Lab, MIT — *Cambridge, MA, USA*

Adversarial Attack on Natural Language Classifiers — *Nov. 2019 - present*

- Designed a general sampling framework based on popular pre-trained language model - BERT, which have various application.
- Applied the framework on adversarial attack and successfully identified the non-robustness of state-of-the-art text classifiers.

Synthetic Data Generation use Deep Generative Models — *Nov. 2018 - May. 2019*

- Implemented a open-sourced benchmark library, SDGym, and thoroughly compared existing statistical and deep neural models.
- Designed a conditional generative adversarial network (GAN) model to address challenges in generating synthetic data using GAN.
- Evaluated our model on our benchmark and achieved state-of-the-art performance.

Software Engineer (Hosted by Ji Xue) | Google — *New York, NY, USA*

COVID-19 Recovery Detection — *May 2020 - Aug. 2020*

- Developed an algorithm to monitor the status of each state recovering from the pandemic.

Software Engineer (Hosted by Qingqing Huang) | Google — *Mountain View, CA, USA*

Context-sensitive Matrix Factorization for Recommendation System — *May 2019 - Aug. 2019*

- Designed and implemented a context-sensitive matrix factorization (CFac) method, and achieved significant accuracy improvement.
- Worked closely with my hosts and other folks in the team to apply CFac in the BERT distilling project.

Software Engineer (Hosted by Jiwei Li) | Shannon.AI — *Beijing, China*

Chinese A-Share Stock Question-Answering System — *Jun. 2018 - Aug.2018*

- Developed a module for the main product, a QA system for A-share stocks, with the developing team of 10.

Annual Report Data Mining

- Suggested the roadamp and designed software framework for this research project.
- Implemented a rule-based algorithm to accurately extract production and sales data from companies' annual reports.

Undergrad Research Assistant | THUNLP Lab, Tsinghua University — *Beijing, China*

Nerual Text Summarization — *Nov. 2016 - May. 2017*

- Implemented the sequence-to-sequence summarization model, and trained the model on GPU.
- Open-sourced *THUNLP/TensorFlow-Summarization* repo on GitHub, winning 320 stars so far.
- Improved the model by explicitly incorporating keywords into the loss function, getting significant improvement.

Publications

- **Lei Xu**, Kalyan Veeramachaneni, *Attacking Text Classifiers via Sentence Rewriting Sampler*, Preprint, 2020
- Shubhra Kanti Karmaker Santu, Md. Mahadi Hassan, Micah J. Smith, **Lei Xu** ,ChengXiang Zhai, Kalyan Veeramachaneni, *A Level-wise Taxonomic Perspective on Automated Machine Learning to Date and Beyond: Challenges and Opportunities*, Preprint, 2020
- **Lei Xu**, Maria Skoularidou, Alfredo Cuesta-Infante, Kalyan Veeramachaneni, *Modeling Tabular Data using Conditional GAN*, NeurIPS, 2019.
- Brandon Amos, **Lei Xu**, J. Zico Kolter, *Input Convex Neural Networks*, ICML, 2017.
- Xinxiong Chen*, **Lei Xu***, Zhiyuan Liu, Maosong Sun, Huanbo Luan. *Joint Learning of Character and Word Embeddings*, IJCAI, 2015. (* equal contribution)

Awards/Skills

Awards	2017 Outstanding Graduates of Beijing
Programming	C/C++, Java, Python, Tensorflow, PyTorch, Shell, HTML, Javascript, SQL.
Languages	English (fluent), Chinese (native).