Yao Lu

https://yaolu.me

Room 3519, Davis Center
University of Waterloo, ON, Canada
+1 519-580-5528 | yao.lu@uwaterloo.ca

EDUCATION and EXPERIENCES

University of Waterloo **Waterloo, Canada**
Master of Mathematics. in *Computer Science (thesis)* *Sept.2018 - August.2020*

- Related Courses: Optimization, Theory of Deep Learning, Data-Intensive Distributed Computing
- TA-ed Courses: CS136 (Algorithm Design), CS431/451 (Distributed Computing for Text Processing)

Montreal Institure of Learning Algorithm (MILA) **Montreal, Canada**
Research Intern *Sept.2019 - May.2020*

- Multi-Document summarization and knowledge graph extraction

University of Alberta **Edmonton, Canada**
Research Intern at Dept. of Computing Science and Amii *Oct.2017 - June.2018*

Tongji University **Shanghai, China**
B.Eng. in *Material Science and Engineering* *Sept.2012 - Jun.2017*

RESEARCH EXPERIENCES

Interest: Natural Language Processing

PLM2KG: Extracting Knowledge Graph from Pretrained Language Models **Mila**
Collaborate with Prof. Laurent Charlin, MILA *Oct.2019 - Now*

- Current methods to measure the knowledge in PLMs require human-curated templates to express existing semantic triples as natural language sentences, whose perplexity is then computed. Our approach bridges the gap between discrete triples and human language, without any additional manual effort needed for template definition.
- Our model consists of two parts: nested span-based entity/relation recognition, knowledge graph (KG) and language model (LM) based link predictor. The model takes a document as input and tries to identify potential text spans for entities and relations. The model then uses the output of PLMs over the extracted spans as the input to the link predictor. Being KG-based, the link predictor represents a mapping function from the PLM's latent space to the knowledge graph embedding space. For the LM link predictor, we adopt the assumption that templates are only conditioned on the relation itself. Ranking the probabilities of triplets of same relation type is the same as ranking the perplexity of sentences composed with mask token. With the regularization of language model, our model can perform much better than previous methods in terms of novel facts discovery.

Disentangled Learning of Syntax and Semantics **U Waterloo**
Collaborate with Prof. Lili Mou and Prof.Olga Vechtomova, University of Waterloo *May.2019 - Now*

- Variational autoencoders (VAE) have been widely used for sequence modeling, due to the effective regularization of latent space. However, vanilla VAEs fail to modeling syntactic structure well, which restricts its further usage for controlled generation.We propose an approach to regularize the latent space to better model syntax and semantics. Similar to vanilla VAEs, we adopt two independent latent variables to represent syntactic and semantic information, respectively. To improve the disentanglement of syntax and semantics, we use an auxiliary objective, namely reconstructing the constituency tree and bag-of-words, to regularize the latent space. With the joint training of all objectives, this system can perform quite well on syntax style transfer.

Distilling Task-Specific Knowledge from BERT into Simple Neural Networks **U Waterloo**
Collaborate with Dr. Lili Mou and Prof. Jimmy Lin, University of Waterloo *Mar.2019 - Jun.2019*

- In the natural language processing literature, neural networks are becoming increasingly deeper and complex. The recent poster child of this trend is the deep language representation model, which includes BERT, ELMo, and GPT. These developments have led to the conviction that previous-generation, shallower neural networks for language understanding are obsolete. However, we demonstrate that rudimentary, lightweight neural networks can still be made competitive without architecture changes, external training data, or additional input features. We propose to distill knowledge from BERT, a state-of-the-art language representation model, into *a single-layer BiLSTM*. Across multiple datasets in paraphrasing, natural language inference, and sentiment classification, we achieve comparable results with ELMo, while using *100 times fewer parameters* and *15 times less inference time*.

MATS: a multi-task learning framework for abstractive summarization **Amii & U Alberta**
Collaborate with Prof. Min Yang, CAS and Prof. Randy Goebel, University of Alberta *Oct.2017 - Feb.2018*

- We propose a Multi-task learning approach for Abstractive Text Summarization (MATS). Specifically, MATS consists of three key components: (i) a text categorization model that learns rich category-specific text representations using a LSTM encoder; (ii) a syntax labeling model that learns to improve the syntax-aware LSTM decoder; and (iii) an abstractive text summarization model that shares its encoder and decoder with the text categorization task and the syntax labeling task, respectively. In particular, the abstractive summarization model enjoys benefit from the additional text categorization and syntax knowledge. Our experimental results show that MATS outperforms the competitors.

PUBLICATIONS

- **Yao Lu**, Zhaocheng Zhu, Jian Tang and Laurent Charlin *PLM2KG: Extracting Knowledge Graph from Pretrained Language Models.* [draft, work in progress]
- **Yao Lu**, Lili Mou and Olga Vechtomova *Disentangled learning of syntax and semantics.* [Master's thesis topic, work in progress]
- **Yao Lu***, Raphael Tang*, Linqing Liu*, Lili Mou, Olga Vechtomova, Jimmy Lin *Distilling Task-Specific Knowledge from BERT into Simple Neural Networks* arXiv:1903.12136 **(equal contribution, flip order)**
- **Yao Lu**, Linqing Liu, Zhile Jiang, Min Yang and Randy Goebel *A Multi-task Learning Framework for Abstractive Text Summarization* AAAI 2019
- **Yao Lu**, Yusheng Ding, Qingyang Xiao, Jianwei Lu and Tianwei Yu *Detecting Differential Consistence Genes* To Appear in BMC Bioinformatics
- Linqing Liu, **Yao Lu**, Min Yang, Qiang Qu and Jia Zhu *Generative Adversarial Network for Abstractive Text Summarization* AAAI 2018
- Linqing Liu, **Yao Lu**, Ye Luo, Renxian Zhang, Laurent Itti, and Jianwei Lu. *Detecting" Smart" Spammers On Social Network: A Topic Model Approach.* NAACL 2016

Open Source Contribution

Differential network local consistency (DNLC) **R Package**
*Authors: **Yao Lu**, Yusheng Ding, Linqing Liu and Tianwei Yu* *Dec.2016*

- Project about detection local consistency. Package available in R-CRAN (Package Link)

PROJECTS AND COMPETITIONS

Beauty of Programming Competition **Microsoft Research Asia**
May.2016

- Top 3/20,000 among students from 150 universities. Organizer: Microsoft and IEEE

TECHNICAL STRENGTH

Programming Languages Python, C/C++, R, Matlab, LaTeX, Shell Script
Packages&Tools PyTorch, Tensorflow