Peng Xu
(780)6558396
https://github.com/billy-inn

April 30, 2021
pxu4@ualberta.ca
Senior Machine Learning Researcher

Research Interests

- Natural Language Processing
  - Semantic Parsing
  - Language Modelling
  - Information Extraction
  - Text Generation
- Machine Learning
  - Generative Modelling
  - Transformer Optimization

Professional and Research Experience

- **Borealis AI** Toronto
  Senior Machine Learning Researcher *2018.7-Present*
  - Implement and improve the state of the art semantic parsing models to build a Text-to-SQL system, which has been successfully applied to different banking business, including direct investing and commercial banking. Demo: https://turing.borealisai.com.
  - Published five first-authored papers at ACL 2019, AISTATS 2020, ICML 2020, ACL 2021 on topics of coherence modelling, language modelling, text generation and semantic parsing.
  - Mentored three internship research projects on semantic parsing.
  - Filed four patent applications based on the above-mentioned research work.
- **Shannon AI** Beijing
  NLP Algorithm Intern *2018.5*
  - Develop intelligent financial question answering system
  - Build the pipeline to automatically extract information from company announcements
- **DBpedia**
  Active Contributor and Mentor for Google Summer of Code (GSoC) *2016.4-2018.9*
  - Inferred infobox template class mappings from Wikipedia and WikiData in `python` in 2016
  - Mentored two GSoC projects about entity embeddings based on DBpedia in 2017
  - Mentored one GSoC project about embeddings for Out-Of-Vocabulary resources in 2018
- **University of Alberta** Edmonton, AB
  Research Assistant under Prof. Denilson Barbosa *2016.5-2018.7*
  - Research on fine-grained entity type classification (FETC) with deep neural models
  - Research on relation extraction (RE) with deep neural models
  - Research on incorporating knowledge base (KB) information to facilitate neural information extraction, mainly on task of FETC and RE

- One long oral paper on FETC accepted at NAACL 2018, one short paper on RE accepted at NAACL 2019, and master thesis on neural information extraction awarded runner up for the Outstanding Master Thesis Award

- **Tsinghua University** Beijing
  Research Assistant under Prof. Jie Tang *2014.9-2015.6*
  - Develop **Arnetminer II** which is a website that offers comprehensive search and mining services for academic community in `python` and `scala`

- **Tsinghua University** Beijing
  Research Assistant under Prof. Zhihu Du *2014 Summer*
  - Research on efficient partical-mesh spreading in `C` and `CUDA`

Education

- **University of Alberta** Edmonton, AB
  M.Sc. Computer Science *2016.1 - 2018.7*
  - GPA: 4.0/4.0
  - Supervisor: Prof. Denilson Barbosa

- **Beijing University of Posts and Telecommunications** Beijing
  B.Eng. Electronic Information Engineering *2011.9 - 2015.6*
  - Major: 89.62/100, Overall: 87.06/100
  - Thesis Advisor: Prof. Jie Tang

Publications

(* indicates equal contribution)

1. *Optimizing Deeper Transformers on Small Datasets*
   The 59th Annual Meeting of the Association for Computational Linguistics (ACL), 2021, *long paper*
   Peng Xu, Dhruv Kumar, Wei Yang, Wenjie Zi, Keyi Tang, Chenyang Huang, Jackie Chi Kit Cheung, Simon J.D. Prince, Yanshuai Cao

2. *TURING: an Accurate and Interpretable Multi-Hypothesis Cross-Domain Natural Language Database Interface*
   The 59th Annual Meeting of the Association for Computational Linguistics (ACL), 2021, *demonstration track*
   Peng Xu*, Wenjie Zi*, Hamidreza Shahidi, Akos Kadar, Keyi Tang, Wei Yang, Jawad Ateeq, Harsh Barot, Meidan Alon, Yanshuai Cao

3. *On Variational Learning of Controllable Representations for Text without Supervision*
   The 37th International Conference on Machine Learning (ICML), 2020
   Peng Xu, Yanshuai Cao, Jackie Chi Kit Cheung

4. *Better Long-Range Dependency By Bootstrapping A Mutual Information Regularizer*
   The 23rd International Conference on Artificial Intelligence and Statistics (AISTATS), 2020
   Yanshuai Cao* and **Peng Xu***

5. *A Cross-Domain Transferable Neural Coherence Model*
   The 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019, *long paper*
   Peng Xu, Hamidreza Saghir, Jin Sung Kang, Teng Long, Avishek Joey Bose, Yanshuai Cao, Jackie Chi Kit Cheung

6. *Connecting Language and Knowledge with Heterogeneous Representations for Neural Relation Extraction*
   The 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019, *short paper*
   Peng Xu, Denilson Barbosa

7. *Neural Fine-Grained Entity Type Classification with Hierarchy-Aware Loss*
   The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2018, *long oral paper*
   Peng Xu, Denilson Barbosa

8. *Matching Resumes to Job Descriptions with Stacked Models*
   The 31st Canadian Conference on Artificial Intelligence, 2018.
   Peng Xu, Denilson Barbosa

9. *Investigations on Knowledge Base Embedding for Relation Prediction and Extraction*
   arXiv preprint, 2018
   Peng Xu, Denilson Barbosa

10. *Efficient Particle-Mesh Spreading on GPUs*
    The 15th International Conference on Computational Science (ICCS), 2015
    Xiangyu Guo, Xing Liu, **Peng Xu**, Zhihu Du, Edmond Chow

Thesis

- *Towards Neural Information Extraction without Manual Annotated Data*
  M.Sc. Thesis, **runner up for the Outstanding Master Thesis Award**

Patents Filed

1. *System and Method for Cross-domain Transferable Neural Coherence Model*
   US Patent App. 16/669,741.

2. *System and Method for Machine Learning with Long-Range Dependency*
   US Patent App. 16/809,267.

3. *System and Method for Controllable Machine Text Generation Architecture*
   US Patent App. 16/881,843.

Teaching

- **Machine Learning** — University of Alberta
  Teaching Assistant — *2016 Fall, 2017 Fall*
- **Introduction to the Foundations of Computation II** — University of Alberta
  Teaching Assistant — *2018 Winter*
- **Introduction to the Foundations of Computation I** — University of Alberta
  Teaching Assistant — *2017 Winter*

Awards & Honors

ATEC: NLP for Financial Intelligence (hosted by Alibaba), 10th out of 443 teams 2018
Programming Contest of *BUPT*, Gold Medalist 2013
ACM-ICPC Hunan Invitational Programming Contest, Bronze Medalist 2013
Third Grade Scholarship for Undergraduates 2013-14
Second Grade Scholarship for Undergraduates 2012-13
Third Grade Scholarship for Undergraduates 2011-12

Skills

- Professional in Python (PyTorch, Scikit-learn, XGBoost)
- Familiar with LaTeX, Bash, Git, Vim, Tmux, Slurm
- Limited previous experience in:
  - Programming Language: C/C++, CUDA, Scala, R, Matlab(Octave)
  - Python Framework: Tensorflow, Keras, Gensim
  - Production Deployment: Docker, MySql, Sqlite