Ning Miao

Peking University - Beijing, China

miaoning@pku.edu.cn • ningmiao.github.io

Research Interests

My research interests lie in the area of **Machine Learning** and **Natural Language Processing**. I'm devoted to building better models and algorithms, as well as understanding their behaviors theoretically. I also wish to push the limit of current NLP methods, especially text generation models.

Education

Peking University **Beijing, China**

Sept. 2016 - July 2019

- M.S. in Data Science (Statistics)
- TOEFL: R29, L30, S23, W24, total 106
- Advisors: Prof. **Rui Yan**

Peking University **Beijing, China**

Sept. 2012 - July 2016

- B.E. in Mathematics
- Advisors: Prof. **Qingchun Tian** and Prof. **Yuan Yao**

Publications

Improving Maximum Likelihood Training for Text Generation with Density Ratio Estimation

- Yuxuan Song, **Ning Miao**, Hao Zhou, Lei Li.
- Submitted to *The 23rd International Conference on Artificial Intelligence and Statistics* (**AISTATS-20**).

Dispersing Exponential Family Mixture VAEs

- Wenxian Shi, Hao Zhou, **Ning Miao**, Shenjian Zhao, Lei Li.
- Submitted to *The 23rd International Conference on Artificial Intelligence and Statistics* (**AISTATS-20**).

Kernelized Bayesian Softmax for Text Generation

- **Ning Miao**, Hao Zhou, Chengqi Zhao, Wenxian Shi, Lei Li.
- In *Proceedings of The Thirty-third Conference on Neural Information Processing Systems* (**NeurIPS-19**).

Generating Fluent Adversarial Examples for Natural Languages

- Huangzhao Zhang, Hao Zhou, **Ning Miao**, Lei Li.
- In *Proceedings of The 57th Annual Meeting of the Association for Computational Linguistics* (**ACL-19**, short).

CGMH: Constrained Sentence Generation by Metropolis-Hastings Sampling

- **Ning Miao**, Hao Zhou, Lili Mou, Rui Yan, Lei Li.
- In *Proceedings of The Thirty-Third AAAI Conference on Artificial Intelligence* (**AAAI-19** , oral).

Research Experiences

Research on Deep Generation Model for Sequential Data **Mar. 2018 - present**

Research intern at ByteDance AI Lab. Advisor: Dr. Hao Zhou and Dr. Lei Li

- In ByteDance, my works are focused on deep generation models for sequential data, especially texts. Here are some topics that I have explored.
  - **Text Generation under Constraints.**
    We proposed an approach to performing Metropolis-Hastings sampling on text space, which was the first practical method for hard-constrained text generation. We also applied this method to tasks such as unsupervised paraphrase and sentence correction.
  - **Improving Text Decoder by Kernelization**
    We discovered that in text generation models, output embeddings of each word are composed of different

disconnected components and each component has different variances. We introduced a family of kernel functions to explicitly control embedding variances and a hierarchical probabilistic model to represent the disconnected components. This method not only improved the performance of several generation models, but also provided useful visualization for words' properties.

- **Generation by Classification**
  This project aimed to reduce unfluent samples in the current text generation models. We combined a convolutional discriminator with mainstream autoregressive models to improve their ability to discriminate between fluent and unfluent sentences. Experiments showed that this model greatly reduced both the perplexity and reverse-perplexity.

Research on Structured Text Generation **June 2017 - July 2019**
Student at Care Group. Advisors: Prof. Rui Yan

- This research focused on structured text generation, especially tree-shaped generation. We proposed a tree-to-tree model to (1) perform constrained generation and (2) better preserve the syntactical information of source texts. The proposed model achieved (1) higher BLEU scores in the task of generation from keywords as well as (2) higher accuracy in predicting syntactical structures in machine translation.

Project of Fine-Grained Text Sentiment Analysis **Sept. 2016 - July 2017**
Clinical Psychology Center, Peking University

- This project was designed to automatically detect sentiments in texts. By designing special network structures and applying multitask learning, we achieved more than 85% predicting accuracy. And by active learning, we built a high quality dataset in fine-grained sentiment analysis. This system saved more than 90% of clinical psychologists' labor, since only writing samples with high depression score need manual processing.

Classes and Skills

- The following are some representative classes I took during my bachelor and master study.
  - Calculus: Advanced Calculus, Measure Theory, Complex Variables.
  - Algebra: Matrix Theory, Abstract Algebra.
  - Geometry: Analytic Geometry, Point Topology.
  - Advance Probability, Advanced Statistics, Convex Optimization, Random Process.
- Programming Language: Python with NumPy, Tensorflow, NLTK and Matplotlib.