Ning Miao

Peking University - Beijing, China

miaoning@pku.edu.cn • www.ningmiao.space

Research Interests

- **Machine Learning** and **Natural Language Processing**.
- Committed to building better models and algorithms, as well as understanding their behaviors from a mathematical perspective.
- Currently working on Bayes methods and deep generative models for structured data, such as texts.

Education

Peking University **Beijing, China**

Sept. 2016 - July 2019

- M.S. in Data Science (Statistics)
- Advisors: Prof. **Rui Yan**
- Notable Courses: Advanced Algebra, Functions of Real Variables, Topology, Stochastic Process.

Peking University **Beijing, China**

Sept. 2012 - July 2016

- B.E. in Mathematics
- Advisors: Prof. **Qingchun Tian** and Prof. **Yuan Yao**
- Notable Courses: Advanced Probability and Statistics, Bayesian Theory.

Publications

Improving Maximum Likelihood Training for Text Generation with Density Ratio Estimation

- Yuxuan Song, **Ning Miao**, Hao Zhou, Lei Li.
- In *Proceedings of 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**

Intern and full-time researcher at ByteDance AI Lab. Advisor: Dr. Hao Zhou and Dr. Lei Li

- At ByteDance, my work focuses on deep generation models for sequential data, especially texts. Below are some topics I have explored:
  - **Text Generation under Constraints.**
    Traditional autoregressive models cannot generate sentences under hard constraints. Proposed non-autoregressive model based on Metropolis-Hastings sampling, which generates samples by jumping between neighboring sentences. Experiments confirmed it as the most practical method for generating fluent samples under hard constraints. This work was published at AAAI-19.
  - **Improving Text Decoder by Kernelization.**
    Discovered that output embeddings of each word are not regularly distributed. On the contrary, they are composed of several disconnected components, and each component has a different variance. To

model output embeddings more accurately, we introduced a family of kernel functions to explicitly control embedding variances and a hierarchical probabilistic model to represent disconnected components. This method not only improved the performance of text generation models but provided useful visualization for properties of words. This work was accepted by NeurIPS-19.

- **Tailor for Language Models.**
  Explained why language models tend to over-estimate the probabilities of some patterns while under-estimating the probabilities of other patterns. Designed a probabilistic tailor to prune generation probabilities of language models by efficient sequential sampling methods. Experiments showed that this method could greatly improve sample quality without losing diversity. This work was submitted to ACL-20.

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

- Aimed to build an automatic intensity detector of six different text sentiments. Achieved more than 85% prediction accuracy using multitask learning and active learning. Detector saved more than 90% of clinical psychologists' labor, since only writing samples with high depression scores required manual processing.

Skills and Languages

- Skills: Python with NumPy, Tensorflow, NLTK and Matplotlib; LaTeX.
- Languages: English (TOEFL: 106); Chinese.