Miao Li

State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China

(+86) 188-1070-7818 miaoli.cs@gmail.com https://oaimli.github.io

ACADEMIC BACKGROUND

M.S. University of Chinese Academy of Sciences (UCAS) Aug. 2017 – Jun. 2020
State Key Laboratory of Computer Science,
Institute of Software, Chinese Academy of Sciences (ISCAS)
Exam-exempted postgraduate
Ranking **6th/102**, GPA **88.31/100**, IELTS 6.5 (L/6.5, S/6.0, R/6.5, W/6.5)

B.S. Northeastern University (NEU) Sep. 2013 - Jul. 2017
Software College
Ranking **11th/272**, GPA **87.35/100**

RESEARCH INTRERESTS

- Natural Language Processing and Machine Learning
- Text generation and document-level understanding
- Deep unsupervised learning for Natural Language Processing
- Knowledge guided Natural Language Processing
- Model transparency, interpretability, and controllability for Natural Language Processing

TECHNICAL SKILLS

- Skilled programming in Python, Java, and LaTeX, also familiar with Linux
- Proficient in deep learning programming, especially in Keras and Tensorflow, and programming tool boxes (e.g. Numpy, Scipy, Scikit-learn, NLTK)
- Master at most Machine Learning Models (e.g. SVM, LDA, CRF) and inference methods (e.g. Variational Inference and Monte Carlo Method)
- Understand deep learning well, such as CNN, RNN and high-level deep generative models, especially Variational Auto-Encoders
- Familiar with fundamental tasks and models of Natural Language Processing and Text Mining, like text modeling, information extraction, text generation, etc.

RESEARCH EXPERIENCE

Deep latent-variable models for text clustering Beijing, Apr. 2018 - PRESENT

Granted by National Key R&D Program of China (No. 2017YFC0803300)

- Combining composability of graphical models and flexible modeling capability of deep learning, deep latent-variable models are suitable to text modeling and unsupervised learning tasks.
- We have already proposed a new text clustering model in neural variational inference in Vector Space Model with replacing the prior to Mixture-of-Gaussians, and it was published at ICTAI'2019, and it jointly learns document representations and cluster assignments. Our model outperforms state-of-the-art models. Also, clustering results were visualized and interpretable by text topics.
- We are now incorporating external knowledge, such as knowledge learned by Graph Neural Networks, into deep latent-variable models and developing an end-to-end text clustering model with combination of implicit and explicit representations of texts. Moreover, we are trying to achieve sentiment-aware text clustering, which clusters documents according to their sentiments.

Topic augmented text generation Beijing, Nov. 2018 – May. 2019

Supported by National Key R&D Program of China (No. 2017YFC0803300)

- We proposed a novel text generation model that learns semantics and structural features simultaneously, which captures structural features by a sequential variational autoencoder component and leverages a topic modeling component based on Gaussian distribution to enhance the recognition of text semantics. We made an oral presentation at EMNLP'2019.
- Experimental results over several open datasets demonstrate that our model outperforms several state-of-the-art models in terms of perplexity and topic coherence. Also, the latent representations learned by our model can be used in down-stream tasks and is superior in text classification, and our model can generate texts which hold similar structures but under different topics.

Clustering volume trajectories of buses in Beijing Beijing, Oct. 2016 – Dec. 2017

Joint work with and supported by the Beijing Public Transport Group

- We proposed a three-phase clustering strategy for the massive trajectories (about forty million trajectories per week) in the form of Origin-Destination pairs which were modeled as a sparse graph where the spatial and temporal features as well as the constraints of road networks are integrated into the similarity of trajectories. This work was published at UIC'2018.
- Furthermore, we borrowed the idea of text mining and gave a feasible method to mine semantics of clustered trajectories.
- This work demonstrated the significance of trajectory clustering in evaluating and adjusting public transit operations, and methods we developed are in practical use in Beijing Public Transport Group.

A series of Android development Shenyang, Dec. 2013 – Sep. 2016

As the Manager or developer in charge

- Team management Android system, Diagnostic system for wind power generators in Android, and Pingnan Medical System for communities in Android. I was the major developer of these three Android clients and also in charge of their design.
- Developed many new Android modules (e.g. user-defined muti-picture widget with disk and memory cache, PDF browser without calling other software) and achieved most elusive features in these Apps (dynamic view in a tree structure and offline searching).

PUBLICATIONS

- (EMNLP 2019, Acceptance rate: 684/2877 = 23.8%) A Topic Augmented Text Generation Model: Joint Learning of Semantics and Structural features
  *Hongyin Tang, **Miao Li**, Beihong Jin*
- (ICTAI 2019) A New Effective Neural Variational Model with Mixture-of-Gaussians Prior for Text Clustering
  ***Miao Li**, Hongyin Tang, Beihong Jin, Chengqing Zong*
- (UIC 2018) Clustering Large-Scale Origin-Destination Pairs: A Case Study for Public Transit in Beijing
  ***Miao Li**, Beihong Jin, Hongyin Tang, Fusang Zhang*
- (In progress) Knowledge-aware Generative Embedding with Combination of Explicit and Implicit Representation for Text Clustering
  ***Miao Li**, Hongyin Tang, Beihong Jin*
- Software copyright of a medical system in Android, No.2016SR133938, June 2016

HONORS & AWARDS

- Chinese National Scholarship for Graduates (<Top 2%) Nov. 2018
- The First Prize Scholarship of UCAS (<Top 8%) Oct. 2017&2018
- Excellent Student Cadre of University of Chinese Academy of Sciences, twice Jun. 2018&2019
- Merit Student of University of Chinese Academy of Sciences, twice Jun. 2018&2019
- Outstanding Graduate of Northeastern University June. 2017
- Outstanding Graduate Thesis Award of Northeastern University (Top 1/272) Jul. 2017
- The first prize scholarship of Northeastern University, twice (<Top 4%) Sep. 2015&2016
- Chinese National Encouragement Scholarship, twice (<Top 2%) Oct. 2015&2016
- Excellent Student Cadre of Northeastern University, twice Oct. 2015&2016
- Merit Student of Northeastern University, three times Oct. 2014-2016

CONTESTS

- Zhejiang Lab Cup Global Artificial Intelligence Competition 2018: Zero-shot Learning for Picture Recognition, Ranking 80th/3224 (Top 3%) Sep. 2018
- Honorable Mention in MCM/ICM 2015 Mar. 2015
- The first prize in the "Oracle Cup" Java programming contest in the Northeast of China Oct. 2014

OTHER EXPERIENCE

- Graduate Student Council Chairman of ISCAS Sep. 2018 - PRESENT
- Monitor of Class seven in School of Computer Science in UCAS Sep. 2017 - PRESENT
- Undersecretary of Northeastern University Volunteers Association Nov. 2013 – Nov. 2014
- Volunteer in the 12th Chinese National Game Sep. 2013

HOBBIES & SPECIALTY

Programming Reading Communication Basketball Swimming Hiking