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 **6/102**, GPA **88.31/100**

B.S. Northeastern University (NEU) Sep. 2013 - Jul. 2017
Software College
Ranking **11/272**, GPA **87.35/100**, IELTS 6.5 (L/6.5, S/6.0, R/6.5, W/6.5)

RESEARCH INTRERESTS

- Natural Language Processing and Text Data Mining, Machine Learning
- Text generation and document-level understanding
- Deep Unsupervised learning and deep generative models 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-Encoder
- Familiar with fundamental tasks and models of Natural Language Processing, 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, and our model outperforms state-of-the-art models. Also, clustering results were visualized and interpretable by text topics.
- We are now incorporating knowledge into deep latent-variable models and developing an end-to-end text clustering model in implicit representation of texts.

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

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

- We proposed a 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.
- Results of our experiments over several 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.

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

Joint work with the Beijing Public Transport Group

- We proposed a three-phase clustering strategy for the massive trajectories 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.

- Furthermore, we borrow the idea of text data mining and give a feasible method to mine semantics of clustered trajectories
- This work demonstrated the impact of trajectory clustering on 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 Accepted, Top-tier conference in NLP) A Topic Augmented Text Generation Model: Joint Learning of Semantics and Structural features
  *Hongyin Tang, **Miao Li**, Beihong Jin*
- (ICTAI 2019 Accepted) 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 Postgraduates (<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
- 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

Competitions

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