Rui Dong

111 Park Dr., Apt 54.
Boston, MA, 02215

Phone: 617-992-4676
Email: dongrui@ccs.neu.edu

Education

Northeastern University 2014.09-present
Ph.D. candidate, College of Computer and Information Science

Fudan University 2011.09-2014.07
M.sc., School of Computer Science

East China Normal University 2007.09-2011.07
B.Eng., School of Information Science and Technology

Research Interests

My research interests are in natural language processing and machine learning. I am currently focusing on text error correction and language modeling based on neural network models.

Publication

Published:

Zhipeng Xie, **Rui Dong**, Zhengheng Deng, Zhenying He, Weidong Yang. A Probabilistic Approach to Latent Cluster Analysis. *International Joint Conference on Artificial Intelligence* (*IJCAI 2013*), August 2013, Beijing, China.

Rui Dong, Yizhou Sun, Lu Wang, Yupeng Gu, Yuan Zhong. Weakly-Guided User Stance Prediction via Joint Modeling of Content and Social Interaction. *International Conference on Information and Knowledge Management* (*CIKM 2017*), November 2017, Singapore, Singapore.

Submitted:

Rui Dong, David A. Smith. Multi-Input Attention for Unsupervised OCR Correction. *Association for Computational Linguistics* (*ACL 2018*), July 2018, Melbourne, Australia.

Research Experience

Predictive Typing for Brain-Computer Interface System 2018-present

I am now working with Prof. David A. Smith to design a neural language model for the typing module of a Brian-Computer Interface designed for patients with Lock-In Syndrome.

Optical Character Recognition Post-correction 2017-2018

I worked with Prof. David A. Smith to design an unsupervised model for OCR post-correction that exploits repeated texts in large corpus. An attention-based Seq2Seq model is applied as the correction model and different attention combination strategies are introduced to jointly align, correct and vote among duplicated texts. This work has been submitted to ACL, 2018.

Argument Clustering **2016-2017**

I worked with Prof. Lu Wang to design a deep clustering model to uncover the argumentative facets of different topics according to online discussions. A hierarchical attention-based classification model is applied to learn the representation for each sentence and a neural clustering model is designed to group sentences into different facets.

Understanding User Stance via Comment Board **2015-2016**

I worked with Prof. Yizhou Sun and Prof. Lu Wang to design a unified model to understand users' positions in different topics according to their comments on the news websites. We propose to combine the modeling of the comments and the interactions between the users to predict their stances. This work has been accepted by CIKM, 2017.

User Characteristic Analysis **2014-2015**

I worked with Prof. Yizhou Sun to design a model to discover the users' characteristics according to their behavior recorded in the game dataset VPAL. This project mainly aims to utilize the users' interactions with the dialog boxes in the game to learn their personalities.

Cluster Ensemble **2012-2014**

I worked with Prof. Zhipeng Xie to design a novel algorithm to utilize the clustering results generated by different clustering algorithms to find a consensus clustering result. This work has been accepted by IJCAI, 2013.

Cross-Modal Data Retrieval **2012-2013**

I worked with Prof. Zhipeng Xie to develop new algorithms to improve the search quality for retrieval from heterogeneous datasets. We had implemented a retrieval system and proposed a random walk algorithm to improve the search quality of cross-model retrieval.

Work Experience

Data Scientist Intern at Ancestry.com **2016.06-2016.08**

I worked as a data scientist intern in the Data Science Team at Ancestry.com. I explored the OCR results from historical newspapers and conducted experiments on various applications of the noisy text such as entity recognition and entity relation extraction.

Technical Skills

Python, C, C#, Matlab, Tensorflow