Rui Meng

Ph.D. Student
University of Pittsburgh

Cell	412-539-6665
Email	rum20@pitt.edu
Webpage	http://memray.me/
LinkedIn	https://linkedin.com/in/memray/
GitHub	https://github.com/memray/

Focus

Natural Language Processing, Data Mining, Machine Learning, Information Retrieval

Education

University of Pittsburgh — **Pittsburgh, PA, USA**
Information Science and Technology - Ph.D. student. GPA: 3.8 — Sep 2015 — Present

Wuhan University — **Wuhan, Hubei, China**
Information Management and System - Master, Bachelor. GPA: 3.7 — Sep 2008 — Jul 2015

Work Experience

EMApp Wireless Technology Co. — **Internship**
Deputy project leader, leading a four-person team to develop a complete social networking app project. — Jan 2012 — Jul 2012

Project Experience

Large-scale Keyphrase Extraction based on Deep Learning

Aiming to detect keyphrases from large academic text corpus, a LSTM-based deep neural network architecture with multiple attention mechanisms is proposed on solving this problem, expecting a huge boost over previous unsupervised methods.

Kaggle: Springleaf Market Response

A large-scale classification task aiming to locate the target users for advertising. More than 290,000 customer records are given, each with nearly 2,000 anonymized features. We conducted highly-complete experiments, including data analysis (statistics and visualization), feature engineering (selection, reduction) and classification (ensemble models).

Automatic Summarization for Yelp Reviews

Summarizing Yelp reviews based on density-based and graph-based extractive summarization methods, along with multiple text representations (BOW, LDA and Word2Vec). A neural-attention-model-based generative method is also explored.

Citation Semantic Classification

Citation performs in different semantic roles in scientific papers. A fine-grined classification is conducted, based on support vector machine with multiple semantic features (n-grams, pos-tagging, linguistic pattern, typed dependency, entity etc.).

Baidu Big Data Competition

A topic classification contest on huge amount of industry query data. We performed a linear SVM along with a feature argumentation on Hadoop.

Real-time Media Monitoring System

Developed a real-time web scrapper and text analyzer on online news and social media. Applied sentiment classification and clustering to detect critical negative mentions.

National Olympiad in Informatics in Provinces

The most influential programming competition for secondary school students in China. Similar to ACM Programming Contest, the NOIP require the contestants to show great IT skills as problem analysis, design of algorithms and data structures, programming and testing. The winners (less 1% of contestants) are recommended and accepted to top universities in China.

Technology Skills

Machine Learning	Theano, Scikit-learn, NLTK, Weka, LIBSVM, CRF++, Mallet, Stanford NLP toolkits
Programming	Python, Java, Linux, Bash, R, C/C++, Web Development
Research Tools	MATLAB, LaTeX

Publications

- Knowledge-based Content Linking for Online Textbooks.
- Automatic Classification of Citation Function with Rich Linguistic Features
- A Framework for Citation Content Annotation: Towards the Bibliographic Relationship between Scientific Literatures
- The Application of Query Performance Prediction to Result Merging of Aggregated Search
- Hot Topics Detection of Web Public Opinion Based on Field-weighted Clustering Algorithm

Awards, Grants & Honors

First Prize of National Olympiad in Informatics in Provinces — **China Computer Federation**, Jan 2008
Awarded to top 1% participants in China for outstanding skills in algorithms and computer programming

The Pacemaker to Graduate Student — **Wuhan University**, Feb 2014
Awarded to top students (1%) for extraordinary leadership

The First-Class Scholarship — **Wuhan University**, Oct 2013
Awarded to Top 5% students

Full Scholarship — **Wuhan University**, Sep 2012
Awarded to top students entering university