Zhifei Zhang

Curriculum Vitæ

Department of Computer Science and Technology
Tongji University, 4800 Cao'an Highway
Shanghai, China, 201804
✉ *zhifei.zzh@gmail.com*
tjzhifei.github.io

Education

09/2008–Present **Department of Computer Science and Technology, Tongji University**.
Ph.D. candidate in Pattern Recognition and Intelligent Systems
– Supervisor: Prof. Duoqian Miao

09/2004–07/2008 **Department of Computer Science and Technology, Tongji University**.
B.Sc. in Computer Science and Technology
– GPA: 4.71/5.00 Rank: 1/138

Research Interests

Natural Language Processing: Text Classification, Sentiment Analysis

Machine Learning: Topic Model, Multi-Label Learning, Rough Sets

Research Experience

05/2012–Present **Multi-granularity Topic-Sentiment Mixture Model on Social Media Texts**.
I host this project supported by the Fundamental Research Funds for the Central Universities. Social media texts contain colorful opinions and cover many topics. An object can be represented by hierarchical attributes. The project aims to discover rich sentiment information about one object within the social media. Due to the short length of social media texts, we apply latent Dirichlet allocation in topic extraction and similarity measure, which is described in [1]. More work are ongoing.

06/2012–08/2012 **NLP&CC 2012 Weibo Sentiment Analysis Evaluation**.
We accomplish a sentiment analysis system with 70.4% precision, 56.2% recall and 62.5% F1 in opinion detection task; 69.1% precision, 38.9% recall and 49.8% F1 in polarity classification task. We fuse many public sentiment dictionaries, one of which is collected by myself. We adopt lexicon-based methods for subjectivity identification. We use three elements, i.e., opinion words, negative words and punctuation marks, for sentiment classification.

12/2010–06/2011 **Opinion Mining for Hotel Reviews**.
We implement a crawler to fetch user reviews about hotels from ctrip.com and construct a hotel review dataset. We measure semantic orientation based on reference concepts using HowNet, which outperforms the word-based method. We establish several supervised and unsupervised algorithms for sentiment classification. We propose an ensemble method based on behavior-knowledge-space and achieve good results. Besides, we develop a visualization system that illustrates opinions of 56 hotels in Shanghai with the help of Google Map. These work are described in [5] and [6].

11/2009–07/2010 **Multi-label Classification for Web Search Results**.

We build a visualization system for web search results based on multi-label classification. For lack of Chinese multi-label corpora, we construct one from SogouC corpus manually. We propose a Multi-Label Naive Bayesian (MLNB) classification algorithm and implement a prototype system named TJ-MLWC. The system acts as an intermediate layer between users and an Internet search engine, which allows the search results of a query displaying by one or more categories. We conduct further research on multi-label learning and propose a Dual-Weighted Multi-Label KNN (DW-ML-KNN) algorithm. The algorithm achieves a satisfying performance especially when dealing unbalanced data. These work are described in [3] and [7].

03/2008–09/2008 **Chinese Text Classification System**.

We build a Chinese text classification system, which includes five functional modules: preprocessing (including word segmentation implemented by ICTCLAS), feature selection, text representation using VSM, classifiers (KNN, SVM implemented by LibSVM) and performance evaluation. We focus on feature selection algorithms and propose a new feature selection algorithm based on rough set theory, named IQR (Improved Quick Reduction). IQR achieves better performance than CHI. This work is described in [2].

Skills

Programming Language: C++, Java
Python, MATLAB

Language: Chinese (native), English (fluent)

Awards & Honors

12/2012 **National Graduate Scholarship of China**.
Awarded to the top 5 of the department

11/2012 **Academic Scholarship for Doctoral Candidates of Tongji University**.
Awarded to the top 15 of the university

12/2011 **Excellent Graduate Student Leader Scholarship of Tongji University**.

12/2010 **Second Prize of National Graduate Mathematical Contest in Modeling**.

12/2009 **Second Prize of National Graduate Mathematical Contest in Modeling**.

12/2009 **Excellent Student of Tongji University**.

05/2008 **Excellent Graduate of Shanghai (Bachlor)**.
Awarded to the top 5% of the department

10/2008 **Excellent Bachelor Thesis Award of Tongji University**.
Awarded to the top 2 of the department

12/2007 **Top-Class Scholarship of Tongji University for Undergraduates**.
Awarded to the top 10 of the university

12/2006 **First-Class Scholarship of Tongji University for Undergraduates**.
Awarded to the top 5% of the department

12/2005 **First-Class Scholarship of Tongji University for Undergraduates**.
Awarded to the top 5% of the department

Activities

Presentations

12/2012 The 8th International Conference on Advanced Data Mining and Applications (ADMA 2012), Nanjing, China

11/2012 The 6th Youth Conference of Computational Linguistics (YCCL 2012), Shanghai, China

09/2012 The 8th International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE 2012), Hefei, China

10/2011 The 7th Chinese Conference on Information Retrieval (CCIR 2011), Jinan, China

Services

12/2012 Session chair of ADMA 2012, Nanjing

11/2012 Session chair of YCCL 2012, Shanghai

11/2011 Web chair of UGrC 2011, Shanghai

05/2010 Volunteer of World Expo 2010, Shanghai

Membership

IEEE Student Member, ACM Student Member
CCF Student Member, CAAI Student Member

Publication

[1] **Zhifei Zhang**, Duoqian Miao, and Can Gao. Short Text Classification Using Latent Dirichlet Allocation. *Journal of Computer Applications*, 2013. (in Chinese)

[2] **Zhifei Zhang**, and Duoqian Miao. Feature Selection for Text Categorization Based on Rough Sets. *CAAI Transactions on Intelligent Systems*, 2009, 4(5): 453–457. (in Chinese)

[3] Zhihua Wei, Hongyun Zhang, **Zhifei Zhang**, Wen Li, and Duoqian Miao. A Naive Bayesian Multi-label Classification Algorithm with Application to Visualize Text Search Results. *International Journal of Advanced Intelligence*, 2011, 3(2): 173–188.

[4] Duoqian Miao, Can Gao, Nan Zhang, and **Zhifei Zhang**. Diverse Reduct Subspaces Based Co-Ttraining for Partially Labeled Data. *International Journal of Approximate Reasoning*, 2011, 52 (8): 1103–1117.

[5] Yuefeng Chen, Duoqian Miao, Wen Li, and **Zhifei Zhang**. Semantic Orientation Computing Based on Concepts. *CAAI Transactions on Intelligent Systems*, 2011, 6(6): 489–494. (in Chinese)

[6] **Zhifei Zhang**, Duoqian Miao, Zhihua Wei, and Lei Wang. Document-level Sentiment Classification Based on Behavior-Knowledge Space Method. In: *Proceedings of the 8th International Conference on Advanced Data Mining and Applications (ADMA 2012)*, LNAI 7713, pp. 330–339.

[7] Duoqian Miao, **Zhifei Zhang**, Zhihua Wei, and Chunyan Wang. DW-ML-kNN: A Dual Weighted Multi-label kNN Algorithm. In: *Proceedings of the 8th Conference on Natural Language Processing and Knowledge Engineering (NLP-KE 2012)*, pp. 519–529.