Zhilin Yang

BASIC
INFORMATION

Undergraduate
Department of Computer Science
Tsinghua University, Beijing, China

Email: kimiyoung@yeah.net
Phone: +86 188-1030-2101
Home page: http://kimiyoung.github.io

RESEARCH
INTERESTS

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

EDUCATION

Tsinghua University, Beijing, China

B.E., Computer Science, 2011 - 2015 (expected)

- Major GPA **96.8/100**, Overall GPA **94.2/100**, ranked **1/124**
- Ranked **No.1 in 16/36 major courses**, including:
  - **Programming**
    Fundamentals of Programming, Foundation of Object-Oriented Programming, Programming and Training, Service Oriented Software Design and Development
  - **Artificial Intelligence**
    Introduction to Machine Learning, Artificial Neural Network, Pattern Recognition, Principles of Signal Processing
  - **Others**
    Probability and Statistics, Computer Organization, Principles and Techniques of Compilers, Introduction to High Performance Computing

Stanford University, CA

The Undergraduate Visiting Researcher (UGVR) Program, 2014/6 - 2014/8
- Summer research on cancer prediction using machine learning.

HONORS AND
AWARDS

China Computer Federation Outstanding Undergraduate Award ($\sim$1.6%)	2014
Science and Technology Innovation Award (1.1%, awarded by Tsinghua University)	2014
Professional Excellence Award ($\sim$4%, awarded by Tsinghua University)	2014
National Scholarship ($\sim$2%, awarded by the Ministry of Education, China)	2013

PUBLICATIONS

Zhilin Yang, Jie Tang, Yutao Zhang.
Active Learning for Streaming Networked Data.
In Proceedings of ACM International Conference on Information and Knowledge Management, 2014.
CIKM'14, full paper, acceptance rate 21%.

Zhilin Yang, Jie Tang, Bin Xu, Chunxiao Xing.
Active Learning for Networked Data Based on Non-Progressive Diffusion Model.
In Proceedings of ACM International Conference on Web Search and Data Mining, 2014.
WSDM'14, plenary full-length presentation (5%), acceptance rate 18%.

Yang Yang, Jianfei Wang, Yutao Zhang, Wei Chen, Jing Zhang, Honglei Zhuang, **Zhilin Yang**, Bo Ma, Zhanpeng Fang, Sen Wu, Xiaoxiao Li, Debing Liu, Jie Tang.
SAE: Social Analytic Engine for Large Social Networks.
In Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013.
KDD'13, demo paper.

Knowledge Engineering Group, Tsinghua University

Research Assistant   Oct, 2012 - present

Active Learning on Graphical Models
- Designed active learning algorithms based on undirected graphical models for social network mining.
- Built a connection between non-progressive diffusion model and active learning on graphs.
- Proved the hardness, correctness, convergence and approximation ratio of the proposed algorithm.
- Our work is published at WSDM'14. [Link] [PDF] [Code&Data]

Modeling Streaming Networked Data
- Studied active learning with streaming networked data. Proposed a framework for learning streaming networked data with active query strategy and network sampling techniques.
- Studied online learning for graphical models to handle streaming networked data.
- Researched undirected graphical models with dual decomposition to speed up the learning process, in order to handle large amount of streaming data.
- Our work is published at CIKM'14. [Link] [PDF] [Code&Data]

Large Scale Social Network Analysis
- Developed the large scale Social Analytic Engine (SAE) based on parallel vertex programs with integration of knowledge base and expert search systems. Parallellized graphical model learning and inference with dual decomposition.
- Studied emotion prediction on Tencent Weibo, a top 2 largest microblogging service in China. Analysed and modelled over 2TB microblogging and user data on 36 distributed machines. Built graphical models to model social influence and accurately predict user emotions.
- Our work is published at KDD'13. [Link] [PDF] [Public Code]

Healthcare Data Mining, Stanford University

Granted by Stanford Undergraduate Visiting Researcher Program   June, 2014 - Aug, 2014

Acute Myeloid Leukemia Outcome Prediction
- Built a pipeline model with SVM and prior knowledge for predicting cancer outcome with clinical features and protein arrays.
- Proposed using dictionary learning for sparse coding to learn the latent representation of the Reverse Phase Protein Array (RPPA).
- Identified important latent states from the RPPA data, which have both biological implications about signaling pathway regulation and statistical power for remission prediction.
- Extracted Boolean rules by Boolean implications and confidence analysis.
- Proposed a hybrid feature selection algorithm for avoiding mixing up continuous and categorical features and achieved significant improvements over the datasets. [Poster] [Code&Data]

Bioinformatics Group, Tsinghua University

Research Assistant   Sept, 2014 - present

Integrative Analysis and Classification for Liver Cancer
- Studied integrative analysis of genome data for complex cancer classification.
- Studied trans-omics (genome, transcriptome, methylome and proteome) data mining for identifying critical factors implicated in liver cancer.
- Researched feature learning for classification and outcome prediction on the TCGA dataset.
- Researched protein interaction network modeling based on signaling pathways.

Programming Languages and Tools
- C/C++, Java, Python, Ruby, Matlab, R, JavaScript, Pascal, VHDL, LaTeX