Honglin Yu

School of Computer Science, Australian National University, NICTA
Homepage : http://yuhonglin.github.io
Email: honglin.yu@anu.edu.au
Mobile: +61 406502886

Research Interests Data Mining, Machine Learning, Online Social Network Analysis

Education

Australian National University Aug. 2011 - present
Ph.D. in Computer Science

- Thesis Topic: Popularity Analysis of YouTube Videos
- Work at Machine Learning group of NICTA
- Thesis will be submitted before Aug. 2015

Southeast University, China Aug. 2009 - Jul. 2011
Finished master course in Automation

Southeast University, China Aug. 2005 - June 2009
Bachelor of Automation

- Rank: 4/140
- Best Graduation Thesis (2 out of 140)

Skill Highlights

- Experienced in **Data Science** and **Statistical Analysis**
  - Classification, regression, clustering, time series analysis
  - Data exploration and visualization (matplotlib, R, d3.js)
  - Working on large scale **real data** for >3 years
- Broad knowledge in **Machine Learning**
  - Deep understanding of main-stream algorithms (SVM, Random Forest, ANN etc.)
  - Experienced in carrying out experiments and **feature engineering**
- Firm programming skills, especially in **C/C++** and **python**
  - Also familiar with SQL, Javascript, R, Julia
  - Use Linux everyday for more than 6 years, skilled emacs user
- Experienced in collecting and analyzing large scale Internet data
  - Skilled in implementing various online data **crawlers**
- Domain knowledge related to Computational Social Science Research
  - YouTube video popularity analysis
  - Twitter user network analysis
- Basic controller design

Research Projects Please see http://yuhonglin.github.io/project/ for more details

- Analyzing Twitter-driven YouTube Views
  - Predicting videos' viewcount with Twitter's feed (**Demo**)
  - Contributions: Most of the ideas; All the Software implementation, data collection, analysis and prediction
- Data Mining on YouTube Video Shot Sequences
  - Mining frequent remixed shots from large number of videos
  - Contributions: Part of the ideas and data collection; All the implementation, data preprocessing and analysis
- Exploring the Phases of Popularity Evolution of YouTube Videos
  - Understanding viewcount dynamics by novel time series segmentation techniques
  - Contributions: Most of the ideas; All the software implementation, data collection, analysis and prediction

Software Highlights

- YTcrawl: a YouTube video history viewcount crawler
- SegFit : a time series segmentation algorithm written in C++
- Shotdetect : a video shot detection program written in C++

Media Coverage

- **ANU Reporter** *How the viral video star is born.* August 2015.
  `http://www.anu.edu.au/news/all-news/how-the-viral-video-star-is-born`
- **NCI Research News** *Predicting popularity.* September 2015.
  `http://nci.org.au/2015/09/30/predicting-popularity/`

Papers

- **Honglin Yu**, Lexing Xie and Scott Sanner, Exploring the Popularity Phases of YouTube Videos: Observations, Insights, and Prediction. In proceeding of International AAAI Conference on Web and Social Media (ICWSM) 2015.
- **Honglin Yu**, Lexing Xie and Scott Sanner, Twitter-driven YouTube Views: Beyond Individual Influencers. In proceeding of ACM Multimedia Conference (ACMM) 2014.

Teaching Experience

- Computational Social Science Summer Short Course — Jul. 2013
  Teaching Assistant, Beihang University, China
- Computer Network — 2011
  Teaching Assistant, Southeast University, China

Awards

- Excellent Graduation Dissertation of Southeast University (2 out of 140) — 2009
- Champion of the 3D Soccer Simulation League, Robocup Worldcup — 2008
- National Mathematical Modeling Contest Third Prize of Jiangsu Province — 2008
- Higher Mathematics Competition of Jiangsu Province, Third Prize — 2006

Review Experience

- International World Wide Web Conference (WWW)
- ACM International Conference on Web Search and Data Mining (WSDM)

Language Chinese (native) English (fluent)