Zhiyuan (Brett) Chen

313 S Halsted Street, Chicago, IL 60661

Homepage: `http://www.cs.uic.edu/~zchen/`

Mobile: (312)532-7291

Email: czyuanacm@gmail.com

PROFESSIONAL PROFILE

– I will join Google full-time early 2016 and work on interesting Machine Learning problems.
– PhD level expert in Machine Learning, Natural Language Processing, Text Mining, and Sentiment Analysis.
– Hands-on experience via internship at Microsoft, Twitter, and Quora; intern offers from Google and LinkedIn.
– Published long research papers in premier international conferences, including KDD, ICML, ACL, and IJCAI.
– 8+ years of programming: proficient in Java, Python, and C++; medals in programming contests.

RELATED EXPERIENCE

Quora Inc. — Mountain View, CA, USA
Data Science Intern — May 2015 - July 2015
Project: Question Quality Classification and Grammar Auto-Correction

- Keywords: Machine Learning, Natural Language Processing, Feature Engineering, Production
- Complexity and Technology: 5K lines of Python code.
- Details: Built a machine learning classifier for a large amount of questions based on a variety of features. Also, applied data mining techniques to improve question quality by auto-correcting grammatical issues.
- Accomplishments: 1) Implemented the entire data pipeline for the project, including extracting data, modeling questions, and evaluating models. 2) Designed fine-grained natural language processing features for model improvements. 3) Integrated the code into production.

Twitter Inc. — Boston, MA, USA
Software Engineering Intern in Media Science Team — May 2014 - Aug 2014
Project: Modeling Twitter Influence on TV Tune in

- Keywords: Machine Learning, Big Data, MapReduce, Time Series Data, Causal Reasoning.
- Complexity and Technology: 1K lines of Scalding (Hadoop, 100TG data), 4K lines of Python code.
- Details: Collected and cleaned a large amount of data (Tweets, TV Ratings, etc) from multiple data sources. Built a model to infer the relationship between Twitter, TV and other media such as News.
- Accomplishments: Designed models and experiments to attest the relationship, with presentations made to several teams at Twitter.

Microsoft Research — Redmond, WA, USA
Research Intern in Internet Services Research Center — May 2012 - Aug 2012
Project: Video Triggering in Bing Search Engine

- Keywords: Information Retrieval, Search Engine, Query Understanding, Production.
- Complexity and Technology: 4K lines of C# code for experiments, 3K lines of C# code for production.
- Details: Modeled the context of Khan Academy educational videos based on domain background. The ranking algorithm was designed by using video metadata and classified user query intent.
- Accomplishments: 1) Dramatically improved precision by more than 30% compared with the previous method; 2) Wrote production code to incorporate the algorithm into Bing search engine.

Microsoft Research Asia — Beijing, China
Research/Engineering Intern in Web Search and Mining Group — Sept 2010 - May 2011
Project: Automatic Wrapper for Web Data Extraction

- Keywords: Web Mining, Website Wrapper, Data Mining, Data Extraction, Production.
- Complexity and Technology: 5K lines of C# code for experiments, 2.5K lines of C++ code for production.
- Details: Designed a new algorithm to speed up the extraction of forum posts from webpages.
- Accomplishments: 1) Significantly reduced the running time from 245 milliseconds to 8.8 milliseconds per page; 2) Wrote production code in the Microsoft back-end system; 3) Got one US Patent (No. 20120124086).

For more projects, please visit my homepage: `http://www.cs.uic.edu/~zchen/projects.html`.

EDUCATION

University of Illinois at Chicago   Chicago, IL, USA
Ph.D. Candidate, Computer Science, Advisor: Professor Bing Liu, GPA: 4.0 / 4.0   Aug 2011 - Present

Dalian University of Technology   Dalian, China
Bachelor of Engineering, Software Engineering, GPA: 3.87 / 4.0, Rank: 3 / 397   Aug 2007 - July 2011

PHD THESIS

- Dissertation Title: Lifelong Machine Learning for Topic Modeling and Classification
- Proposed novel lifelong machine learning techniques to discover coherent topics using big data. The goal is to design computational models that learn as humans do, i.e., retaining the knowledge learned in the past and leveraging it for future learning.
- Other PhD research projects: inferring user intention from forum posts; opinion review spam detection; incorporating domain knowledge into topic modeling; and sentiment analysis on social media.
- Extensive knowledge in machine learning, including classification, clustering, and recommendation.
- Outstanding writing and communication skills with oral presentations in the international conferences.
- Program committee member and reviewer in multiple conferences and journals, peer reviewed 70+ papers.

SELECTED REFEREED PUBLICATIONS (FIRST-AUTHOR PAPERS)

- **Zhiyuan Chen**, Nianzu Ma, and Bing Liu. "Lifelong Learning for Sentiment Classification". *Short Paper*. In Proceedings of **ACL 2015**.
- **Zhiyuan Chen** and Bing Liu. "Mining Topics in Documents: Standing on the Shoulders of Big Data". *Oral Presentation*. In Proceedings of **KDD 2014**.
- **Zhiyuan Chen** and Bing Liu. "Topic Modeling using Topics from Many Domains, Lifelong Learning and Big Data". *Oral Presentation*. In Proceedings of **ICML 2014**.
- **Zhiyuan Chen**, Arjun Mukherjee, and Bing Liu. "Aspect Extraction with Automated Prior Knowledge Learning". *Oral Presentation*. In Proceedings of **ACL 2014**.
- **Zhiyuan Chen**, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. "Exploiting Domain Knowledge in Aspect Extraction". *Oral Presentation*. In Proceedings of **EMNLP 2013**.
- **Zhiyuan Chen**, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. "Discovering Coherent Topics Using General Knowledge". *Oral Presentation*. In Proceedings of **CIKM 2013**.
- **Zhiyuan Chen**, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. "Leveraging Multi-Domain Prior Knowledge in Topic Models". In Proceedings of **IJCAI 2013**.
- **Zhiyuan Chen**, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. "Identifying Intention Posts in Discussion Forums". *Oral Presentation*. In Proceedings of **NAACL-HLT 2013**.

For my full publication list (14 Papers in total), please visit: `http://www.cs.uic.edu/~zchen/publications.html`.

ALGORITHM DESIGN AND PROGRAMMING EXPERIENCE

- 3 Silver Medals and 1 Bronze Medal in **ACM-ICPC** (ACM International Collegiate Programming Contest) Asia Regional Contests. *My Silver medal was the first one in the history of my undergraduate university.*
- Solid foundation in algorithm design and data structures, with 1.5 years experience of tutoring "Algorithms and Data Structures" in the ACM-ICPC group (20+ students).
- Programming Languages: Java (proficient), Python (proficient), C++ (proficient), C (familiar), C# (familiar).
- Familiar with frameworks and techniques for MapReduce: Hadoop, Scalding and Microsoft SCOPE.
- Developed automatic online judge system and website for ACM-ICPC Dalian City Programming Contest.

HONORS AND AWARDS

- Fifty For The Future®, 2015: award 50 exceptional students using innovative technology in Illinois.
- Dean's Scholar Awards, 2015: recognize the most distinguished, top 1% graduate students at UIC.