

# BEI YU

## *Curriculum Vitae*

### CONTACT

School of Information Studies  
320 Hinds Hall  
Syracuse University  
Syracuse, NY 13244-1190

Email: [byu@syr.edu](mailto:byu@syr.edu)  
Web: <http://beiyu.syr.edu>  
Phone: (315) 443-3614  
FAX: (315) 443-6886

### APPOINTMENTS

- 2015-present Associate Professor, School of Information Studies, Syracuse University  
Faculty Lead of the CAS Data Science Program
- 2009-2015 Assistant Professor, School of Information Studies, Syracuse University
- 2006-2009 Postdoctoral Fellow, Kellogg School of Management, Northwestern University  
Supervisor: Daniel Diermeier

### EDUCATION

- 2006 University of Illinois at Urbana-Champaign Ph.D. in Library and Information Science  
Dissertation: An evaluation of text classification methods for literary study  
Advisors: John Unsworth and Linda Smith
- 1999 Institute of Computing Technology M.E. in Computer Science  
Chinese Academy of Sciences
- 1996 University of Science and Technology of China B.E. in Computer Science

### RESEARCH INTERESTS

#### General Areas

My primary research interest is in the area of applied natural language processing (applied NLP).

My research focuses on using machine learning, data mining, and language technologies to study long-standing questions in social sciences, humanities, and library and information science.

Specifically, I conduct computational analysis of large amounts of text data to discover linguistic patterns that characterize people's opinions, sentiments, and language styles, and develop prediction models accordingly.

My current work, funded by an IMLS Early Career Award, focuses on developing an automated citation context analysis tool to help researchers retrieve and summarize academic opinions in literature.

## Selected Specific Topics

- Academic opinions in literature
- Political opinions in Congress
- Consumer opinions on social media
- Language, gender, authorship, and stylistics
- Crowdsourcing data collection

## ACADEMIC AWARDS AND HONORS

2014	Early Career Award	Institute of Museum and Library Services (IMLS)
2011	Best Paper Award	SIGIR 2011 Workshop on Crowd-sourcing for Information Retrieval
2006	Vannevar Bush Best Paper Finalist	ACM/IEEE Joint Conference on Digital Libraries

## PUBLICATIONS

### Edited Proceedings

- 2009 **Yu, B.** and Jiang, M. (eds.) (2009). *Proceedings of the First International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion*, ACM, New York.

### Refereed Journal Publications

- 2015 Kwok, L. and Yu, B. (2015). Taxonomy of Facebook messages in business-to-consumer communications: What really works? *Tourism and Hospitality Research* (doi: 10.1177/1467358415600214)
- 2015 Kwok, L., Zhang, F, Huang, Y., **Yu, B.**, Maharabhushanam, P., & Rangan, K. (in press). Documenting business-to-consumer (B2C) communications on Facebook: What have changed among restaurants and consumers? *Worldwide Hospitality and Tourism Themes*.
- 2014 **Yu, B.** (2014). Language and gender in Congressional speech. *Literary and Linguistic Computing*, 29(1): 118-132.
- 2013 **Yu, B.**, Willis, M., Sun, P., and Wang, J. (2013). Crowdsourcing the participatory design and evaluation of medical pictographs. *Journal of Medical Internet Research*, 15(6):e108.
- 2013 Kwok, L. and **Yu, B.** (2013). Spreading social media messages on Facebook: An analysis of restaurant business-to-consumer communications. *Cornell Hospitality Quarterly*, 54(1): 84-94.
- 2012 Diermeier, D., Godbout, J-F., **Yu, B.** and Kaufmann, S. (2012). Language and ideology in Congress. *British Journal of Political Science*, 42(1): 31-55.
- 2008 **Yu, B.**, Kaufmann, S., and Diermeier D. (2008). Classifying party affiliation from political speech. *Journal of Information Technology and Politics* 5(1): 33-48.
- 2008 **Yu, B.** (2008). An evaluation of text classification methods for literary study. *Literary and Linguistic Computing* 23(3): 327-343.

---

## Refereed Conference and Workshop Proceedings

- 2014 **Yu, B.** and Fanelli, D. (2014). Classifying negative findings in biomedical publications. *Proceedings of ACL 2014 Workshop on Biomedical Natural Language Processing*, 19-23.
- 2013 **Yu, B.** (2013). Automated citation sentiment analysis: what can we learn from biomedical researchers. *Proceedings of the American Society for Information Science and Technology*, 50(1): 1-9.
- 2012 **Yu, B.** (2012). Function words for Chinese authorship attribution. *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, Montreal, Canada, June 8, 2012, 45-53.
- 2011 **Yu, B.**, Chen, M., and Kwok, L. (2011). Toward predicting popularity of social marketing messages. *Proceedings of the 2011 International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction (SBP'11)*, University of Maryland, College Park, MD, March 29-31, 2011, *Lecture Notes in Computer Science* 6589, Springer, 317-324.
- 2011 **Yu, B.** and Kwok, L. (2011). Classifying business marketing messages on Facebook. *Proceedings of the SIGIR 2011 Workshop on Internet Advertising*, Beijing, July 28, 2011.
- 2011 Wang, J. and **Yu, B.** (2011). Labeling images with a recall-based image retrieval game. *Proceedings of the SIGIR 2011 Workshop on Crowd-sourcing for Information Retrieval*, July 28, 2011. (**Best Paper Award**)
- 2011 Hu, X. and **Yu, B.** (2011). Exploring the relationship between mood and creativity in rock lyrics. *Proceedings of the 12<sup>th</sup> International Society for Music Information Retrieval Conference (ISMIR'11)*, Miami, Florida, October 24-28, 2011, 789-794.
- 2010 **Yu, B.** and Diermeier, D. (2010). A longitudinal study of language and ideology in Congress. *The 68th Annual National Conference of Midwest Political Science Association*, Chicago, IL, April 2010.
- 2010 Wang, J. and **Yu, B.** (2010). Sentence recall game: a novel tool for collecting data to discover language usage patterns. *Proceedings of the 16<sup>th</sup> ACM SIGKDD Workshop on Human Computation*, Washington, D.C., July 25<sup>th</sup> 2010, 56-59.
- 2008 **Yu, B.**, Kaufmann, S., and Diermeier D. (2008). Exploring the characteristics of opinion expressions for political opinion classification. *Proceedings of the 9<sup>th</sup> Annual International Conference on Digital Government Research*, Montreal, Canada, May 2008, 82-91.
- 2008 Shao, H., **Yu, B.**, and Nadeau, J. (2008). Strangeness-based feature weighting and classification of gene expression profiles. *Proceedings of the 23<sup>rd</sup> Annual ACM Symposium on Applied Computing Bioinformatics Track (SAC'08)*, Fortaleza, Ceara, Brazil, March 2008, 1292-1296.
- 2006 Plaisant, C., Rose, J., **Yu, B.**, Auvil, L., Kuschenbaum, M., Smith, M., Clement, T. and Lord, G. (2006). Exploring erotics in Emily Dickinson's correspondence with text mining and visual interfaces. *Proceedings of the 6<sup>th</sup> ACM/IEEE Joint Conference on Digital Libraries*, Chapel Hill, NC June 2006, 141-150. (**Vannevar Bush Best Paper Finalist**)
- 2004 Zhai, C., Velivelli, A., and **Yu, B.** (2004). A cross-collection mixture model for comparative text mining. *Proceedings of the 10<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, WA, USA, August 2004, 743-748.

- 2002 Wang, J., **Yu, B.**, and Gasser, L. (2002). Concept tree based clustering visualization with shaded similarity matrices. *Proceedings of the 2<sup>nd</sup> IEEE International Conference on Data Mining*, Maebashi City, Japan, December 2002, 697-700.

### Refereed Posters and Extended Abstracts in Conference Proceedings

- 2015 **Yu, B.** & Zhang, F. *Disciplinary Difference in Citation Opinion Expression*. In Proceedings of the 2015 Association for Computing Machinery iConference. ACM, Newport Beach, CA.
- 2012 Wang, J. and **Yu, B.** (2012). Collecting representative pictures for words: a human computation approach based on Draw Something game. *Proceedings of the AAI Human Computation Workshop*, Toronto, Canada, July 23, 2012, 157-158.
- 2011 **Yu, B.** (2011). The emotional world of health online communities. *Proceedings of the 2011 iConference*, Seattle, WA, February 8-11, 2011, 806-807.
- 2010 **Yu, B.** and Ku, M.-C. (2010). Collecting legacy corpora from social science research for text mining evaluation. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1-2, Pittsburgh, PA, October 22-27, 2010.
- 2010 Chen, M., **Yu, B.**, and Liu, X. (2010). Building folk UMLS: an approach to finding meaning of folk terms in medical domain. *Proceedings of the 2010 iConference*, 415-416, Champaign, IL, February, 2010.
- 2006 **Yu, B.** and Unsworth, J. (2006). Toward discovering potential data mining applications in literary criticism. *Digital Humanities*, Paris, France, July 2006.
- 2005 **Yu, B.**, Mei, Q., and Zhai, C. (2005). English usage comparison between native and non-native English speakers in academic writing. *The 7th Joint International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing (ACH/ALLC)*, University of Victoria, Canada, June 2005.
- 2004 **Yu, B.** (2004). Comparative literary style mining between native and non-native English writers. *The 67<sup>th</sup> Annual Meeting of American Society for Information Science and Technology*, Providence, Rhode Island, November 12-17, 2004.
- 2002 **Yu, B.** and Searsmith, D. (2003). Genre-based in-document content type classification. *The 13<sup>th</sup> ASIST SIGCR Workshop*, Long Beach, CA, October 2003.

### Refereed Book Chapter

- 2009 Godbout, J-F and **Yu, B.** (2009). Speech and legislative extremism in the U.S. Senate. In L. Imbeau (Ed.), *“Do They Walk Like They Talk?” Speech and Action in Policy Processes*. Chapter 11, 185-206, Springer (Studies in Public Choice Series): New York.

### INVITED TALKS

- 2015 Citation Opinion Retrieval and Analysis. Faculty of Education, Hong Kong University, July.
- 2013 Predicting gender and ideology from political speeches. *NxtMedia Conference*. Trondheim, Norway, October.
- 2013 Language, ideology, and gender in Congressional Speech: a computational approach. *NTNU, VTT, and NxtMedia Seminar on Recommendation Technology*. Trondheim, Norway, October.

- 2010 Computational thinking in text mining for social science research. *Center for Advanced Study in the Behavioral Sciences Workshop on Tracking, Transcribing, and Tagging Government: Building Digital Records for Computational Social Science*, Stanford University, June.
- 2009 Language and ideology in Congress. *Information Science Colloquium*, Cornell University, October.
- 2009 Towards automatic polarity classification of mass opinion. *iForum*, School of Information, University of Texas at Austin, February.
- 2007 Text classification for political science research. with Daniel Diermeier and Stefan Kaufmann, *2nd Chicago Colloquium on Digital Humanities and Computer Science*, Evanston, IL, October.

## GRANTS

- 2014-2017 **Principal Investigator** on IMLS Early Career Grant RE-07-14-0015, \$386,025, for “Citation Opinion Retrieval and Analysis (CORA): An Automated Plug-in Tool for Digital Libraries”.
- 2012 **Investigator** on SU-ADVANCE Cross-Sector Opportunity Grant, \$4,000, for exploring cross-sector collaboration opportunities.
- 2011-2012 **Principal Investigator** on OCLC/ALISE Library and Information Science Research Grant, \$14,996, for “Text Classification of Digital Reference Interviews: an Investigation of Information-seeking Behavior in the Social Web Environment”, in support of Keisuke Inoue’s doctoral dissertation research.
- 2010-2011 **Co-Principal Investigator** on Harrah Hospitality Research Center Grant, \$32,333, for “Typology of Social Media Marketing Messages: a Social Networking Website Perspective”, with Linchi Kwok.

## PROFESSIONAL ACTIVITIES

### Workshop Chair

- 2015 Co-organizer of the Third International Workshop on News Recommendation and Analytics (**INRA’15**), affiliated with the 9<sup>th</sup> ACM Conference on Recommender Systems (**RecSys’15**), Vienna, Austria, September 20.
- 2009 Co-chair of the First International Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement (**TSA’09**), affiliated with the 18<sup>th</sup> ACM Conference on Information and Knowledge Management (**CIKM’09**), Hong Kong, November 9.

### Conference Program Committee Members

- 2015 **ACL’15**: 53<sup>rd</sup> Annual Meeting of the Association for Computational Linguistics
- SBP’15**: International Social Computing, Behavioral-Cultural Modeling and Prediction Conference
- IEEE BigDataService’15**: 1<sup>st</sup> IEEE International Conference on Big Data Computing Service and Applications

- CLFL'15**: Third Workshop on Computational Linguistics for Literature  
**CBRecSys'15** Second Workshop on New Trends in Recommender System  
**INRA'15** Third International Workshop on News Recommendation and Analytics
- 2014 **ACL'14**: 52<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics  
**SBP'14**: International Social Computing, Behavioral-Cultural Modeling and Prediction Conference  
**NRA'14**: First International Workshop on News Recommendation and Analytics  
**CLFL'14**: Third Workshop on Computational Linguistics for Literature
- 2013 **ACL'13**: 51<sup>st</sup> Annual Meeting of the Association for Computational Linguistics  
**SBP'13**: International Social Computing, Behavioral-Cultural Modeling and Prediction Conference  
**IJCNLP'13**: Sixth International Joint Conference on Natural Language Processing  
**CLFL'13**: Second Workshop on Computational Linguistics for Literature
- 2012 **SBP'12**: International Social Computing, Behavioral-Cultural Modeling and Prediction Conference
- 2011 **IJCNLP'11**: Fifth International Joint Conference on Natural Language Processing  
**WWW'11**: 20<sup>th</sup> International World Wide Web Conference
- 2010 **COLING'10**: 23<sup>rd</sup> International Conference on Computational Linguistics  
**EMNLP'10**: Conference on Empirical Methods in Natural Language Processing  
**OMBI'10**: First International Workshop on Opinion Mining and Business Intelligence  
**TFDOM'10**: International Workshop on Topic Feature Discovery and Opinion Mining

### Journal Reviewer

ACM Transactions on Information Systems (**ACM TOIS**)  
Canadian Journal of Information and Library Science (**CJILS**)  
Electronic Library Journal (**ELJ**)  
Journal of Data and Knowledge Engineering (**DKE**)  
Journal of Decision Support Systems (**DSS**)  
Journal of Information Processing and Management (**IP&M**)  
Journal of the American Society for Information Science and Technology (**JASIST**)  
Journal of Information Technology and Politics (**JITP**)  
Journal of Medical Internet Research (**JMIR**)  
Literary and Linguistic Computing (**LLC**)

### Ad hoc Conference Reviewer

Medicine 2.0 Conference (2014, 2013)  
First International Workshop on Emotion and Sentiment in Social and Expressive Media (2013)  
iConference (2011, 2010)  
Digital Humanities Conference (2011, 2010, 2009, 2008)

### Professional Membership

Association for Computing Machinery (ACM)  
Association for Computational Linguistics (ACL)  
Association for Information Science and Technology (ASIS&T)