Fei Liu

University of Central Florida

4000 Central Florida Blvd, Orlando, FL 32816

http://www.cs.ucf.edu/~feiliu

U.S. Permanent Resident

Tel: +1 407-823-3183

feiliu@cs.ucf.edu

RESEARCH INTERESTS

Natural language processing, **machine learning**, data mining, text summarization, language generation, social media analysis, information extraction, data analytics and visualization

EDUCATION

Postdoctoral Fellow, **Carnegie Mellon University**, Pittsburgh, Pennsylvania   2013 – 2015

Develop algorithms that combine natural language processing, machine learning, and crowdsourcing to understand website privacy policies. Advisors: Noah A. Smith, Norman Sadeh

Ph.D. in Computer Science, **University of Texas at Dallas**, Richardson, Texas   2007 – 2011

Dissertation: "Identifying the Gist of Conversational Text: Automatic Keyword Extraction and Summarization." Committee: Yang Liu (Chair), Vincent Ng, Carlos Busso, Sanda Harabagiu

B.S./M.S. in Computer Science, **Fudan University**, Shanghai, China   2000 – 2004 / 2004 – 2007

M.S. Thesis: "Extracting Chinese Thematic Terms Based on Association Rules."
Advisors: Xuanjing Huang, Lide Wu

PROFESSIONAL EXPERIENCE

Assistant Professor of Computer Science

University of Central Florida, Orlando, Florida   2015.8 – present

Conduct basic and applied research in the fields of natural language processing, machine learning, summarization, language generation, and social media analytics.

Senior Research Scientist

Bosch Research and Technology Center, Palo Alto, California   2013.1 – 2013.11

Design and develop deep natural language understanding systems for in-car applications, specifically focusing on natural language interaction, spoken language understanding, summarization, social media, and applications of machine learning techniques in these areas.

Research Scientist

Bosch Research and Technology Center, Palo Alto, California   2011.5 – 2012.12

Analyze large-scale social media data using statistical machine learning techniques; develop state-of-the-art text normalization & event summarization systems for reading out Twitter messages and trending news to the car driver via a text-to-speech (TTS) system.

Research Intern

Bosch Research and Technology Center, Palo Alto, California   2010.5 – 2010.12

Develop one of the first social media text normalization systems that automatically translates noisy nonstandard word spellings (e.g., *tmrw*) into standard English words; the system is able to achieve 90% word coverage across four datasets of both SMS and Twitter messages.

HONORS AND AWARDS

- **Best Paper Finalist** at the 25th International World Wide Web Conference (WWW), 2016

- Invited to participate in **MIT Rising Stars in EECS** (An Academic Career Workshop for nearly 60 top EECS women scholars in academia), 2015

- Special Performance Recognition Award, Director of Research, Bosch Research, 2012

- On-Ramps into Academia workshop Travel Grant, National Science Foundation (NSF), 2012

- **Erik Jonsson Distinguished Research Fellowship** ($21,000/year for four years with tuition waiver), University of Texas at Dallas, 2007 – 2011

- Academic Excellence Award, University of Texas at Dallas, 2011

- CRA-W Career Mentoring Workshop Travel Grant, National Science Foundation (NSF), 2011

- The Aage and Margareta Moller Endowed Scholarship, University of Texas at Dallas, 2010

- ACL-IJCNLP Student Travel Grant, COLIPS Fund, 2009

- SIGDIAL Conference Travel Grant, International Speech Communication Association (ISCA), 2008

PUBLICATIONS

Journal Publications

[JP1] Joel R. Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T. Graves, Fei Liu, Aleecia M. McDonald, Thomas B. Norton, Rohan Ramanath, N. Cameron Russell, Norman Sadeh, Florian Schaub. "Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding". *Berkeley Law Technology Journal*, Vol. 30, 2015.

[JP2] Fei Liu and Yang Liu. "Towards abstractive speech summarization: Exploring unsupervised and supervised approaches for spoken utterance compression". *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 21, No. 7, pp. 1469–1480, July 2013.

[JP3] Fei Liu, Feifan Liu, and Yang Liu. "A supervised framework for keyword extraction from meeting transcripts". *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 19, No. 3, pp. 538–548, March 2011.

[JP4] Fei Liu, Xuanjing Huang, and Lide Wu. "Approach for extracting thematic terms based on association rules" (in Chinese). *Computer Engineering*, Vol. 34, No. 7, pp. 81–83, 2008.

Peer-Reviewed Conference and Workshop Publications

[CP1] Wencan Luo, Fei Liu, Zitao Liu and Diane Litman. "Automatic Summarization of Student Course Feedback". In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (**NAACL**), San Diego, California, June 2016. [Acceptance rate 182/680=27%, **Oral Presentation**]

[CP2] Shomir Wilson, Florian Schaub, Rohan Ramanath, Norman Sadeh, Fei Liu, Noah A. Smith, and Frederick Liu. "Crowdsourcing Annotations for Websites' Privacy Policies: Can It Really Work?" In *Proceedings of the 25th International World Wide Web Conference* (**WWW**), Montreal, Canada, April 2016. [Acceptance rate 5/727=0.7%, **Oral Presentation**, Best Paper Finalist]

[CP3] Dani Yogatama, Fei Liu, and Noah A. Smith. "Extractive Summarization by Maximizing Semantic Volume". In *Proceedings of the 2015 Conference on Empirical Methods on Natural Language Processing* (**EMNLP**), Lisboa, Portugal, September 2015. [Acceptance rate 312/1300=24%]

[CP4] Fei Liu, Jeffrey Flanigan, Sam Thomson, Norman Sadeh, Noah A. Smith. "Toward Abstractive Summarization Using Semantic Representations". In *Proceedings of the North American Chapter of the Association for Computational Linguistics* (**NAACL**), Denver, Colorado, May 2015. [Acceptance rate 186/714=26%, **Oral Presentation**]

[CP5] Chen Li, Yang Liu, Fei Liu, Lin Zhao, Fuliang Weng. "Improving Multi-document Summarization by Sentence Compression based on Expanded Constituent Parse Tree". In *Proceedings of the 2014 Conference on Empirical Methods on Natural Language Processing* (**EMNLP**), Doha, Qatar, October 2014. [Acceptance rate 225/790=28%]

[CP6] Fei Liu, Rohan Ramanath, Norman Sadeh, Noah A. Smith. "A Step Towards Usable Privacy Policy: Automatic Alignment of Privacy Statements". In *Proceedings of the 25th International Conference on Computational Linguistics* (**COLING**), Dublin, Ireland, August 2014. [Acceptance rate 217/705=31%, **Oral Presentation**]

[CP7] Rohan Ramanath, Fei Liu, Norman Sadeh, Noah A. Smith. "Unsupervised Alignment of Privacy Policies using Hidden Markov Models". In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (**ACL**), Baltimore, Maryland, June 2014. [Acceptance rate 285/1123=25%]

[CP8] Joel R. Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T. Graves, Fei Liu, Aleecia M. McDonald, Thomas B. Norton, Rohan Ramanath, N. Cameron Russell, Norman Sadeh, Florian Schaub. "Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding". In *Proceedings of 2014 TPRC: 42nd Research Conference on Communication, Information and Internet Policy* (**TPRC**), Arlington, Virginia, September 2014.

[CP9] Norman Sadeh, Alessandro Acquisti, Travis D. Breaux, Lorrie Faith Cranor, Aleecia M. McDonald, Joel Reidenberg, Noah A Smith, Fei Liu, N. Cameron Russell, Florian Schaub, Shomir Wilson, James T. Graves, Pedro Giovanni Leon, Rohan Ramanath, Ashwini Rao. "Towards Usable Privacy Policies: Semi-automatically Extracting Data Practices From Websites' Privacy Policies". In *Symposium on Usable Privacy and Security* (**SOUPS**), Menlo Park, California, July 2014.

[CP10] Chen Li, Fei Liu, Fuliang Weng, and Yang Liu, "Document Summarization via Guided Sentence Compression". In *Proceedings of the 2013 Conference on Empirical Methods on Natural Language Processing* (**EMNLP**), Seattle, Washington, October 2013. [Acceptance rate 205/772=27%]

[CP11] Chao Shen, Fei Liu, Fuliang Weng, and Tao Li. "A participant-based approach for event summarization using Twitter streams". In *Proceedings of the North American Chapter of the Association for Computational Linguistics* (**NAACL**), Atlanta, Georgia, June 2013. [Acceptance rate 146/455=32%, **Oral Presentation**]

[CP12] Fei Liu, Fuliang Weng, and Xiao Jiang. "A broad-coverage normalization system for social media language". In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics* (**ACL**), Jeju Island, South Korea, July 2012. [Acceptance rate 187/940=20%]

[CP13] Fei Liu, Fuliang Weng, Bingqing Wang, and Yang Liu, "Insertion, deletion, or substitution? Normalizing text messages without pre-categorization nor supervision". In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (**ACL**), Portland, Oregon, June 2011. [Acceptance rate 292/1146=25%, **Oral Presentation**]

[CP14] Fei Liu, Yang Liu, and Fuliang Weng. "Why is 'SXSW' trending? Exploring multiple text sources for Twitter topic summarization". In *Proceedings of the ACL Workshop on Language and Social Media, Portland*, Oregon, June 2011.

[CP15] Feifan Liu, Fei Liu, and Yang Liu. "Learning from Chinese-English parallel data for Chinese tense prediction". In *Proceedings of the International Joint Conference on Natural Language Processing* (**IJCNLP**), Chiang Mai, Thailand, November 2011. [Acceptance rate 176/478=36%]

[CP16] Fei Liu and Yang Liu. "Exploring speaker characteristics for meeting summarization". In *Proceedings of the 11th Annual Conference of the International Speech Communication Association* (**INTERSPEECH**), Makuhari, Japan, September 2010. [Acceptance rate 771/1324=58%]

[CP17] Yang Liu, Shasha Xie, and Fei Liu. "Using n-best recognition output for extractive summarization and keyword extraction in meeting speech". In *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing* (**ICASSP**), Dallas, Texas, March 2010. [Acceptance rate 1356/2855=48%]

[CP18] Fei Liu and Yang Liu. "Using spoken utterance compression for meeting summarization: A pilot study". In *Proceedings of the 2010 IEEE Workshop on Spoken Language Technology* (**IEEE SLT**), Berkeley, California, December 2010.

[CP19] Fei Liu and Yang Liu. "From extractive to abstractive meeting summaries: Can it be done by sentence compression?" In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics* (**ACL**), Singapore, August 2009. [Acceptance rate 214/925=23%, **Oral Presentation**]

[CP20] Feifan Liu, Deana Pennell, Fei Liu, and Yang Liu. "Unsupervised approaches to automatic keyword extraction using meeting transcripts". In *Proceedings of the North American Chapter of the Association for Computational Linguistics* (**NAACL**), Boulder, Colorado, June 2009. [Acceptance rate 146/438=33%, **Oral Presentation**]

[CP21] Fei Liu, Feifan Liu, and Yang Liu. "Automatic keyword extraction for the meeting corpus using supervised approach and bigram expansion". In *Proceedings of the IEEE Workshop on Spoken Language Technology* (**IEEE SLT**), Goa, India, December 2008.

[CP22] Fei Liu and Yang Liu. "What are meeting summaries? An analysis of human extractive summaries in meeting corpus". In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue* (**SIGDIAL**), Columbus, Ohio, June 2008.

[CP23] Junkuo Cao, Lide Wu, Xuanjing Huang, Yaqian Zhou, and Fei Liu. "Using multiple combined ranker for answering definitional questions". In *Proceedings of the 4th Asia Information Retrieval Symposium* (**AIRS**), Harbin, China, January 2008.

Patent Applications

[PA1] "System and Method for Event Summarization Using Observer Social Media Messages", with Fu-liang Weng, Chao Shen, Lin Zhao. U.S. Patent App. 14/105,233. Filed by Robert Bosch LLC in Dec 2013.

[PA2] "Broad-Coverage Normalization System for Social Media Language", with Fuliang Weng. U.S. Patent Application 13/779,083. Filed by Robert Bosch LLC in Feb 2013. **Issued** in Oct 2015.

[PA3] "Method and System for Text Message Normalization Based on Character Transformation and Web Data", with Fuliang Weng, U.S. Patent Application 13/117,330. Filed by Robert Bosch LLC in May 2011.

Technical Reports and Others

[TR1] Norman Sadeh, Alessandro Acquisti, Travis Breaux, Lorrie Cranor, Aleecia M. McDonald, Joel Reidenberg, Noah A. Smith, **Fei Liu**, N. Cameron Russel, Florian Schaub, Shomir Wilson. "The Usable Privacy Policy Project: Combining Crowdsourcing, Machine Learning and Natural Language Processing to Semi-automatically Answer Those Privacy Questions Users Care About". *Technical Report CMU-ISR-13-119*, Carnegie Mellon University.

[TR2] Rohan Ramanath, Florian Schaub, Shomir Wilson, **Fei Liu**, Norman Sadeh, Noah A Smith. "Identifying Relevant Text Fragments to Help Crowdsource Privacy Policy Annotations". *Second AAAI Conference on Human Computation and Crowdsourcing* (HCOMP 2014), Works-in-Progress Abstract, Pittsburgh, Pennsylvania, 2014

PROFESSIONAL SERVICES

Area Chair

- The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL** 2015), co-chair the Generation and Summarization track with Chris Callison-Burch at the University of Pennsylvania
- The 15th China National Conference on Computational Linguistics (**CCL** 2016), co-chair the Text Categorization and Summarization track with Xiaojun Wan at Peking University

Session Chair

- The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL** 2015)
- The 2015 Conference of the Association for Computational Linguistics (**ACL** 2015)
- The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL** 2016)

Grant Reviewer

- National Science Foundation (NSF), Directorate for Computer & Information Science & Engineering(CISE), Division of Information & Intelligent Systems (IIS), Panel and Ad-hoc, 2016
- Government Grant, Autonomous Province of Bolzano, Italy, 2015

Journal Referee

- Transactions of the Association for Computational Linguistics, 2016
- Computational Linguistics, 2015
- Journal of Artificial Intelligence Research, 2015
- IEEE Transactions on Knowledge and Data Engineering, 2014, 2015
- ACM Transactions on Knowledge Discovery from Data, 2013, 2014
- ACM Transactions on Asian Language Information Processing, 2013, 2014, 2016
- Natural Language Engineering, 2013, 2014
- Speech Communication, 2013
- ACM Transactions on Speech and Language Processing, 2012
- Data & Knowledge Engineering Journal, 2011, 2012
- IEEE Transactions on Audio, Speech and Language Processing, 2011, 2012

Conference Program Committee

- Annual Meeting of the Association for Computational Linguistics (**ACL**), 2012 – 2016
- Conference on Empirical Methods on Natural Language Processing (**EMNLP**) 2013, 2015, 2016
- AAAI Conference on Artificial Intelligence (**AAAI**), 2013, 2016
- International Joint Conference on Artificial Intelligence (**IJCAI**) 2013, 2016
- International Conference on Machine Learning (**ICML**), 2015, 2016
- Annual Conference on Neural Information Processing Systems (**NIPS**), 2015, 2016
- ACM International Conference on Information and Knowledge Management (**CIKM**), 2013, 2016
- IEEE International Conference on Data Mining (**ICDM**) PhD Forum, 2015
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**) 2012 – 2015
- International Conference on Acoustics, Speech, and Signal Processing (**ICASSP**) 2012, 2013
- NAACL Student Research Workshop (SRW) 2013, 2015, 2016
- AAAI Spring Symposium: Analyzing Microtext, 2013
- Pacific-Asia Conference on Knowledge Discovery and Data Mining (**PAKDD**), 2013 – 2015
- European Chapter of the Association for Computational Linguistics (**EACL**), 2012
- Asia Information Retrieval Societies Conference (AIRS), 2012
- International Joint Conference on Natural Language Processing (**IJCNLP**), 2011
- ACL Workshop on Automatic Summarization for Different Genres, Media, and Languages, 2011

TEACHING EXPERIENCE

CAP 6640: Natural Language Processing
Instructor, University of Central Florida

- Algorithms and formalisms used in natural language processing.
  Textbook: Speech and Language Processing, 2nd (and 3rd) edition. By Daniel Jurafsky and James H.
  Martin. Taught in Spring 2016 with 37 graduate-level students. Course evaluation: 4.2/5.

CAP 5610: Machine Learning
Instructor, University of Central Florida

- Theory and algorithms for machine learning, deep learning.
  Textbook: Machine Learning: A Probabilistic Perspective. By Kevin P. Murphy.
  Taught in Fall 2015 with 39 graduate-level students. Course evaluation: 4.5/5.

STUDENT ADVISING

Supervised students at University of Central Florida

- Kristjan Arumae (2016.1 – present), Ph.D. student in Computer Science

- Kexin Liao (2016.1 – present), Ph.D. student in Computer Science

- Nicole Fella (2016.5 – 2016.8), B.S. student at Manhattan College, NSF/REU Program on Internet of Things. Project entitled "An Exploratory Study of Language Vagueness in Website Privacy Policies."

- Jongho Cha (2016.1 – 2016.5), M.S. student in Computer Science.
  Current employment: Ministry of National Defense (South Korea).

- Kylie McCarty (2016.1 – 2016.5), B.S. student in Computer Science.
  Project entitled "Social Media Language Generation."

Supervised students at Carnegie Mellon University

- Eric Lee (2014.1 – 2014.5), B.S. student at Carnegie Mellon University.
  Project entitled "Privacy Policy Visualization using JavaScript and D3.js."
  First employment: Google Inc.

- Rohan Ramanath (2013.12 – 2014.7), M.S. student at Carnegie Mellon University.
  Project entitled "Unsupervised Alignment of Privacy Policies."
  First employment: LinkedIn.

- Wencan Luo (2014.10 – present), Ph.D. student at University of Pittsburgh.
  Project entitled "Automatic Summarization of Student Course Feedback."

Supervised students during their internship at Bosch Research

- Chen Li (2013.5 – 2013.8), Ph.D. student at University of Texas at Dallas.
  Project entitled "Automatic Text Summarization and Applications." First employment: Microsoft

- Chao Shen (2012.5 – 2012.8, 2013.5 – 2013.8), Ph.D. student at Florida International University.
  Project entitled "Twitter event summarization research and system development."
  First employment: Amazon, now at Google Inc.

- Xiao Jiang (2011.10 – 2012.8), M.S. student at Cambridge University.
  Project entitled "Twitter text normalization and system prototyping." First employment: Mobvoi Inc.

Ph.D. Thesis Committee

- Wencan Luo, Ph.D. Student, Computer Science, University of Pittsburgh

- Xuhong Zhang, Ph.D. Student, Electrical and Computer Engineering, University of Central Florida

- Yicong Tian, Ph.D. Student, Computer Science, University of Central Florida