Wei Xu

✉ weixu@cse.ohio-state.edu ☎ +1 614 292 3083 🐦 @cocoweixu
395 Dreese Laboratories, 2015 Neil Avenue Columbus, OH, U.S.A. 43210
http://web.cse.ohio-state.edu/~weixu/

RESEARCH	Natural Language Processing, Machine Learning, Social Media

EXPERIENCE

Assistant Professor, Ohio State University, Columbus, OH — Aug 2016 – present
Department of Computer Science and Engineering

EDUCATION

Ph.D. in Computer Science, New York University, New York, NY — 2014
Advisor: Ralph Grishman; Committee: Satoshi Sekine, Ernest Davis, Bill Dolan (Microsoft Research), Luke Zettlemoyer (University of Washington/Facebook AI)

B.S. and M.S. in Computer Science, Tsinghua University, Beijing, CHINA — 2004/2007

EXPERIENCE

Postdoctoral Researcher, University of Pennsylvania, Philadelphia, PA — Feb 2014 – Aug 2016
Computer Information and Science Department (Advisor: Chris Callison-Burch)

Visiting Graduate Student, University of Washington, Seattle, WA — Jan 2012 – Dec 2013
Computer Science and Engineering Department

Intern, Microsoft Research, Seattle, WA — Jun – Sep 2011
Internet Services Research Center

Consultant, Educational Testing Service, Princeton, NJ — Mar – May, Sep – Dec 2010
Natural Language and Speech Group

Intern, Amazon, Seattle, WA — May – Sep 2010
Catalog Quality

AWARDS, GRANTS, GIFTS

Best Paper Award, COLING, 2018

CrowdFlower AI for Everyone Award, 2018 ($25,000 sole-PI)

Criteo Faculty Research Award, 2018 ($20,000 PI)

NSF CRII Award, 2018 ($175,000 sole-PI) *Robust Intelligence: Learning a Timely Semantic Resource from Social Media Data*

NSF Grant, 2018 – 2021 ($375,732 PI) *Collaborative Research: Automatic Text-Simplification and Reading-Assistance to Support Self-Directed Learning by Deaf and Hard-of-Hearing Computing Workers*

DARPA Grant, 2017 – 2021 ($600,000 co-PI 50%) *Computational Simulation of Online Social Behavior*

NYU MacCracken PhD Fellowship, 2007 – 2012 ($230,000)

Ohio Supercomputer Center Computing Support, 2018

NVIDIA Academic Hardware Gift, 2018

Google/Microsoft Gift for WNUT Workshop, 2019, 2018, 2015

SELECTED PUBLICATIONS

(Note: <u>Underline</u> is used to indicate student advisees at the Ohio State University.)

A Word-Complexity Lexicon and A Neural Readability Ranking Model for Lexical Simplification
<u>Mounica Maddela</u>, Wei Xu
Proceedings of EMNLP 2018, long paper, oral presentation (acceptance rate 25.8%)

Neural Network Models for Paraphrase Identification, Semantic Textual Similarity, Natural Language Inference, and Question Answering
<u>Wuwei Lan</u>, Wei Xu
Proceedings of COLING 2018, long paper (**Best Paper Award**; selection rate 8/888 = 0.90%)

An Annotated Corpus for Machine Reading of Instructions in Wet Lab Protocols
<u>Chaitanya Kulkarni</u>, Wei Xu, Alan Ritter, Raghu Machiraju
NAACL 2018, short paper (acceptance rate 29%)

The Importance of Subword Embeddings in Sentence Pair Modeling
<u>Wuwei Lan</u>, Wei Xu
NAACL 2018, short paper (acceptance rate 29%)

A Continuously Growing Dataset of Sentential Paraphrases
Wuwei Lan, Siyu Qiu, Hua He, Wei Xu
EMNLP 2017, long paper (acceptance rate 25.8%)

A Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter
Jeniya Tabassum, Alan Ritter, Wei Xu
EMNLP 2016, long paper (acceptance rate 26%)

Optimizing Statistical Machine Translation for Simplification
Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, Chris Callison-Burch
TACL 2016, long paper, oral presentation at ACL 2016

Discovering User Attribute Stylistic Differences via Paraphrasing
Daniel Preotiuc, Wei Xu, Lyle Ungar
AAAI 2016, long paper, oral presentation (acceptance rate 26%)

Results of the WNUT16 Named Entity Recognition Shared Task
Benjamin Strauss, Bethany Toma, Alan Ritter, Marie-Catherine de Marneffe, Wei Xu
COLING 2016 Workshop on Noisy User-generated Text (shared-task overview)

Problems in Current Text Simplification Research: New Data Can Help
Wei Xu, Chris Callison-Burch, Courtney Napoles
TACL 2015, long paper, oral presentation at EMNLP 2015

Cost Optimization for Crowdsourcing Translation
Mingkun Gao, Wei Xu, Chris Callison-Burch
NAACL 2015, long paper (acceptance rate 29%)

SemEval-2015 Task 1: Paraphrase and Semantic Similarity in Twitter
Wei Xu, Chris Callison-Burch, William B. Dolan
SemEval 2015 (shared-task overview)

Extracting Lexically Divergent Paraphrases from Twitter
Wei Xu, Alan Ritter, Chris Callison-Burch, William B. Dolan, Yangfeng Ji
TACL 2014, long paper, oral presentation at NAACL 2015

Infusion of Labeled Data into Distant Supervision for Relation Extraction
Maria Pershina, Bonan Min, Wei Xu, Ralph Grishman
ACL 2014, short paper (acceptance rate 25.2%)

Filling Knowledge Base Gaps for Distant Supervision of Relation Extraction
Wei Xu, Raphael Hoffmann, Le Zhao, Ralph Grishman
ACL 2013, short paper (acceptance rate 24%)

Paraphrasing for Style
Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, Colin Cherry
COLING 2012, long paper (acceptance rate 25%)

Exploiting Syntactic and Distributional Information for Spelling Correction with Web-Scale N-grams Models
Wei Xu, Joel Tetreault, Martin Chodorow, Ralph Grishman, Le Zhao
EMNLP 2011, long paper (acceptance rate 23.7%)

Passage Retrieval for Information Extraction using Distant Supervision
Wei Xu, Ralph Grishman, Le Zhao
IJCNLP 2011, long paper

Who, What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task
Kristen Parton, Kathleen McKeown, Bob Coyne, Mona Diab, Ralph Grishman, Dilek Hakkani-Tür, Mary Harper, Heng Ji, Weiyun Ma, Adam Meyers, Sara Stolbach, Ang Sun, Gokhan Tur, Wei Xu, Sibel Yaman
ACL 2009, long paper, oral presentation (acceptance rate 21%)

Extractive Summarization using Inter- and Intra- Event Relevance
Wenjie Li, Wei Xu, Mingli Wu, Chunfa Yuan, Qin Lu
ACL 2006, long paper, oral presentation (acceptance rate 23%)

Deriving Event Relevance from the Ontology Constructed with Formal Concept Analysis
Wei Xu, Wenjie Li, Mingli Wu, Wei Li, Chunfa Yuan
CICLing 2006, long paper

Service   **Workshop Chair:** ACL (2017)
Publicity Chair: EMNLP (2019), NAACL (2018, 2016)

Area Chair: ACL (2019), NAACL (2019), COLING (2018), EMNLP (2018, 2016)
Best Paper Award Committee: EMNLP (2018)
NSF Review Panelist: CISE (2017)
Journal Reviewer: TACL, JAIR
Program Committee: ACL (2018, 2017, 2015, 2014, 2013), EMNLP (2017, 2015, 2014), NAACL (2018, 2015), WWW (2017, 2016, 2015), AAAI (2016, 2015, 2012), KDD (2015), COLING (2014)
Organizer: Workshop on Noisy User-generated Text (EMNLP 2018 & 2017, COLING 2016, ACL 2015); Shared-task on Paraphrase Identification and Semantic Similarity in Twitter (SemEval 2015)

INVITED TALKS	*Learning for Unlimited Human Language*	Dec 2018

Learning for Unlimited Human Language Dec 2018
Peking University, Beijing, China

Learning Large-scale Paraphrases for Natural Language Understanding and Generation Nov 2017 - 2018
Midwest Machine Learning Symposium, Chicago; Facebook, Menlo Park; Twitter, San Francisco; IBM Thomas J. Watson Research Center, New York

How AI Understand Language? Mar 2018
Women in Analytics Conference, Columbus, OH

Can Paraphrase be a Ultimate Solution for NLU and NLG? July 2017
Google Research, New York

Paraphrase ≈ Monolingual Translation Aug 2016
Amazon, Berlin, Germany

Multiple Instance Learning from Unlimited Text Oct 2015 – Dec 2016
Microsoft Research Asia; University of Delaware; University of Edinburgh; University of North Carolina, Chapel Hill; Arizona State University; Vanderbilt University; Imperial College London; University of Waterloo; Indiana University, Bloomington; Washington University, St Louis; Simon Fraser University; University of Alberta; Yale University; University of Maryland

Learning and Generating Paraphrases from Twitter and Beyond Feb 2015
Carnegie Mellon University; Columbia University; Johns Hopkins University

Paraphrases in Twitter Feb 2015
Twitter, San Francisco

Modeling Lexically Divergent Paraphrases in Twitter (and Shakespeare!) Aug 2014 – Mar 2015
City University of New York; IBM Research, Almaden; University of California, Berkeley; University of Texas, Austin; Yahoo!, New York; Carnegie Mellon University; Microsoft Research, Redmond

Students

Current Ph.D. Students: Wuwei Lan, Mounica Maddela, Chao Jiang, Jeniya Tabassum (2/4 are woman)
Current Undergrad Students: Lillian Chow, Sydney Lee, Rita Tong, Raleigh Potluri, Sarah Flanagan, Bohan Zhang (4/6 are woman)
Former Student Advisees: Siyu Qiu, Mingkun Gao, Jim Chen, Ray Lei, Maria Pershina, Wenchao Du, Chaitanya Kulkarni, Piyush Ghai, Pravar Mahajan (2/9 are women)
PhD Thesis Committees: Maria Pershina (New York University - graduated 2016), Kai Cao (New York University - graduated 2017), Sanqiang Zhao (University of Pittsburg - current)

TEACHING

CSE 5522 Artificial Intelligence II: Advanced Techniques Spring 2018, Autumn 2018
Ohio State University, mixed undergraduate and graduate level (teaching evaluation: 4.50/5.00 Spring 2018, 4.85/5.00 Autumn 2018).

CSE 5525 Speech and Language Processing Spring 2017
Ohio State University, mixed undergraduate and graduate level (teaching evaluation 3.80/5.00).

CSE 5539 Social Media and Text Analytics Summer 2016, Autumn 2016, Autumn 2017
http://socialmedia-class.org/
Designed a new course integrated with research, covering from basic to state-of-the-art machine learning algorithms (teaching evaluation: 5.72/6.00 Summer 2015, 4.60/5.00 Autumn 2016, 4.40/5.00 Autumn 2017). The course material I created has been used by other universities (e.g., University of Utah CS 1060, Goergia Tech CS 6452, Iowa State University MIS 546).

Outreach

Panelist: Women in Analytics Conference (2018), CogFest - Cognitive Science Festival (2018)
Mentor: Women and Underrepresented Minorities in NLP Workshop (2017, 2018)
Judge: HackOHIO (2018), OHIO High School Hackathon (2019, 2017)
Instructor: North American Summer School on Logic, Language, and Information (2016)
Organizer: Mid-Atlantic Student Colloquium on Speech, Language and Learning (2016)