Xuan Wang

CONTACT INFORMATION

Siebel Center 2113A
201 N. Goodwin Avenue
Urbana, IL 61801, USA

Phone: (217)417-9512
E-mail: xwang174@illinois.edu
Homepage: `https://xuanwang91.github.io`

RESEARCH INTERESTS

Natural Language Processing, Data Mining, Biomedical/Chemistry Text Mining, BioNLP

EDUCATION

University of Illinois at Urbana-Champaign, Champaign, IL, USA

Ph.D., Computer Science 2017 - present
Thesis: *Minimum Supervised Text Mining for Literature-based Scientific Discovery*
Advisor: Dr. Jiawei Han

M.S., Statistics 2015 - 2017

M.S., Biochemistry 2013 - 2015

Tsinghua University, Beijing, China

B.S., Biological Science 2009 - 2013

PUBLICATIONS

I have published about 20 research/demo papers in both NLP conferences (e.g., ACL, EMNLP, and NAACL) and biomedical informatics journals (e.g., Bioinformatics) and conferences (e.g., ACM-BCB and IEEE-BIBM). These papers have obtained more than 370 citations. My Google Scholar page: `https://scholar.google.com/citations?user=_IVJi6UAAAAJ&hl=en`.

Journal Papers:

[1] **Xuan Wang**, Yu Zhang, Xiang Ren, Yuhao Zhang, Marinka Zitnik, Jingbo Shang, Curtis Langlotz and Jiawei Han, "Cross-type Biomedical Named Entity Recognition with Deep Multi-Task Learning", **Bioinformatics** 35.10 (2018): 1745-1752.

[2] David Liem, Alexandre, Sanjana Murali, Dibakar Sigdel, Yu Shi, **Xuan Wang**, Jiaming Shen, Howard Choi et al. "Phrase Mining of Textual Data to Analyze Extracellular Matrix Protein Patterns Across Cardiovascular Disease." **American Journal of Physiology-Heart and Circulatory Physiology**, 1;315(4):H910-H924, 2018

Conference Papers:

[3] Haoyu Wang, **Xuan Wang**, Yaqing Wang, Guangxu Xun, Kishlay Jha, and Jing Gao, "InterHG: an Interpretable and Accurate Model for Hypothesis Generation", in Proc. of 2021 IEEE Int. Conf. on Bioinformatics and Biomedicine (**IEEE-BIBM'21**), Dec. 2021, online.

[4] **Xuan Wang**, Vivian Hu, Xiangchen Song, Shweta Garg, Jinfeng Xiao and Jiawei Han, "ChemNER: Fine-Grained Chemistry Named Entity Recognition with Ontology-guided Distant Supervision", in Proc. of 2021 Conf. on Empirical Methods in Natural Language Processing (**EMNLP'21**), Punta Cana, Dominican Republic, Nov. 2021

[5] Yu Meng, Yunyi Zhang, Jiaxin Huang, **Xuan Wang**, Yu Zhang, Heng Ji and Jiawei Han, "Distantly-Supervised Named Entity Recognition with Noise-Robust Learning and Language Model Augmented Self-Training", in Proc. of 2021 Conf. on Empirical Methods in Natural Language Processing (**EMNLP'21**), Punta Cana, Dominican Republic, Nov. 2021

[6] **Xuan Wang**, Xiangchen Song, Bangzheng Li, Kang Zhou, Qi Li, and Jiawei Han, "Fine-Grained Named Entity Recognition with Distant Supervision in COVID-19 Literature", in Proc. of 2020 IEEE Int. Conf. on Bioinformatics and Biomedicine (**IEEE-BIBM'20**), Dec. 2020, online.

[7] **Xuan Wang**, Yingjun Guan, Yu Zhang, Qi Li, and Jiawei Han, "Pattern-enhanced Named Entity Recognition with Distant Supervision", in Proc. of 2020 IEEE Int. Conf. on Big Data (**IEEE-BigData'20**), Dec. 2020, online.

[8] **Xuan Wang**, Yu Zhang, Aabhas Chauhan, Qi Li, and Jiawei Han, "Textual Evidence Mining via Spherical Heterogeneous Information Network Embedding", in Proc. of 2020 IEEE Int. Conf. on Big Data (**IEEE-BigData'20**), Dec. 2020, online.

[9] Yu Zhang, Yu Meng, Jiaxin Huang, Frank F. Xu, **Xuan Wang** and Jiawei Han, "Minimally Supervised Categorization of Text with Metadata", in Proc. of 2020 ACM SIGIR Int. Conf. on Research and development in Information Retrieval (**SIGIR'20**), Xi'an, China, Jul. 2020

[10] **Xuan Wang**, Yu Zhang, Qi Li, Xiang Ren, Jingbo Shang, and Jiawei Han, "Distantly Supervised Biomedical Named Entity Recognition with Dictionary Expansion", in Proc. of 2019 IEEE Int. Conf. on Bioinformatics and Biomedicine (**IEEE-BIBM'19**), San Diego, CA, Nov. 2019

[11] Yu Zhang, Frank F. Xu, Sha Li, Yu Meng, **Xuan Wang**, Qi Li, and Jiawei Han, "HiGitClass: Keyword-Driven Hierarchical Classification of GitHub Repositories", in Proc. of 2019 Int. Conf. on Data Mining (**ICDM'19**), Beijing, Nov. 2019

[12] **Xuan Wang**, Yu Zhang, Qi Li, Cathy H. Wu and Jiawei Han, "PENNER: Pattern-enhanced Nested Named Entity Recognition in Biomedical Literature", in Proc. of 2018 Int. Conf. on Bioinformatics and Biomedicine (**IEEE-BIBM'18**), Madrid, Spain, Dec. 2018

[13] Qi Li, **Xuan Wang**, Yu Zhang, Fei Ling, Cathy H. Wu and Jiawei Han, "Pattern Discovery for Wide-Window Open Information Extraction in Biomedical Literature", in Proc. of 2018 Int. Conf. on Bioinformatics and Biomedicine (**IEEE-BIBM'18**), Madrid, Spain, Dec. 2018

[14] **Xuan Wang**, Yu Zhang, Qi Li, Yinyin Chen, and Jiawei Han. "Open Information Extraction with Meta-pattern Discovery in Biomedical Literature", in Proc. of 2018 ACM Conf. on Bioinformatics, Computational Biology, and Health Informatics (**ACM-BCB'18**), Washington, DC, Aug. 2018

System Demonstrations:

[15] Qingyun Wang, Manling Li, **Xuan Wang**, Nikolaus Parulian, Guangxing Han, Jiawei Ma, Jingxuan Tu, Ying Lin, et al., "COVID-19 Literature Knowledge Graph Construction and Drug Repurposing Report Generation", in Proc. of 2021 Annual Conf. of the North American Chapter of the Association for Computational Linguistics (**NAACL'21**) (System demo), ***Best Demo Paper Award***, Mexico City, Mexico, Jun. 2021

[16] **Xuan Wang**, Yingjun Guan, Weili Liu, Aabhas Chauhan, Enyi Jiang, Qi Li, David Liem, Dibakar Sigdel, John Caufield, Peipei Ping and Jiawei Han, "EvidenceMiner: Textual Evidence Discovery for Life Sciences", in Proc. of 2020 Annual Conf. of the Association for Computational Linguistics (**ACL'20**) (System demo), Seattle, WA, Jul. 2020

[17] Jingbo Shang, Qi Zhu, Jiaming Shen, **Xuan Wang**, Xiaotao Gu, Lance Kaplan, Timothy Harratty, and Jiawei Han, "AutoNet: Automated Network Construction and Exploration System from Domain-Specific Corpora", in Proc. of 2018 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (**KDD'18**), (demo paper), London, UK, Aug. 2018

[18] Xiang Ren, Jiaming Shen, Meng Qu, **Xuan Wang**, Zeqiu Wu, Qi Zhu, Meng Jiang et al. "Life-iNet: A Structured Network-Based Knowledge Exploration and Analytics System for Life Sciences", in Proc. of 2017 Annual Meeting of the Association for Computational Linguistics (**ACL'17**), (system demo), Vancouver, Canada, Jul. 2017

Conference Tutorials:

[19] **Xuan Wang**, Yu Zhang, Qi Li, Jiawei Han, "Taming Unstructured Big Data: Automated Information Extraction from Massive Text", 2019 IEEE Int. Conf. on Big Data (**IEEE-BigData'19**), Los Angeles, CA, Dec. 2019

Workshop & Abstract Papers:

[20] **Xuan Wang**, Vivian Hu, Xiangchen Song, Qi Li and Jiawei Han, "Textual Evidence Mining in Scientific Literature", TrueFact Workshop: Making a Credible Web for Tomorrow at 2021 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (**TrueFact@KDD'21**), Aug. 2020, online.

[21] **Xuan Wang**, Xiangchen Song, Bangzheng Li, Yingjun Guan and Jiawei Han, "Comprehensive Named Entity Recognition on CORD-19 with Distant or Weak Supervision", 2020 Intelligent Systems for Molecular Biology (**ISMB'20**), Abstracts (oral and poster), Jul. 2020, online.

[22] **Xuan Wang**, Weili Liu, Aabhas Chauhan, Yingjun Guan and Jiawei Han, "Automatic Textual Evidence Mining in COVID-19 Literature", 2020 Intelligent Systems for Molecular Biology (**ISMB'20**), Abstracts (poster), Jul. 2020, online

[23] **Xuan Wang**, Qi Li, Jiaxin Huang, Yu Zhang, Charles Blatti, Mikel Hernaez and Jiawei Han, "ClaimMiner: Query-guided Claim Mining in Biomedical Literature", 2019 Intelligent Systems for Molecular Biology (**ISMB'19**), Abstracts (oral and poster), Jul. 2019, Basel, Switzerland.

[24] Yu Zhang, Xiang Ren, **Xuan Wang**, Qi Li and Jiawei Han. "Organizing Bioinformatics GitHub Repositories with Multidimensional Text Cube", 2019 Intelligent Systems for Molecular Biology (**ISMB'19**), Abstracts (poster), Jul. 2019, Basel, Switzerland.

[25] Jiawei Han, Qi Li, Jiaming Shen, **Xuan Wang**, Jinfeng Xiao and Yu Zhang. "Text Mining for Biomedical Literature-Based Discovery", 2019 Intelligent Systems for Molecular Biology (**ISMB'19**), Abstracts (poster), Jul. 2019, Basel, Switzerland.

Presentations

Invited Talk, UIUC-NLP Seminar — Dec. 3, 2021
Minimum Supervised Text Mininig for Literature-based Scientific Discovery

Conference Talk, EMNLP'21 — Nov., 2021
Fine-Grained Chemistry Named Entity Recognition with Ontology-guided Distant Supervision

Poster, NSF Center for Computational Biotechnology and Genomic Medicine — Nov 11, 2021
Text Mininig for Biomedicine and Healthcare

Conference Talk, TrueFact@KDD'21 — Aug., 2021
Textual Evidence Mining in Scientific Literature

Tutorial, NSF Center for Molecule Maker Lab Institute — Mar. 21 2021
Text Mining for Chemistry Literature-Based Scientific Discovery

Conference Talk, IEEE-BIBM'20 — Dec., 2020
Fine-Grained Named Entity Recognition with Distant Supervision in COVID-19 Literature

Conference Talk, IEEE-BigData'20 — Dec., 2020
Pattern-enhanced Named Entity Recognition with Distant Supervision

Conference Talk, IEEE-BigData'20 — Dec., 2020
Textual Evidence Mining via Spherical Heterogeneous Information Network Embedding

Guest Lecture, Computer Science Department@UCSD — Dec. 1, 2020
Text Mining for Biomedical Literature-Based Scientific Discovery

Conference Talk, ISMB'20 Jul., 2020
Comprehensive Named Entity Recognition on CORD-19 with Distant or Weak Supervision

Conference Talk, ISMB'20 Jul., 2020
Automatic Textual Evidence Mining in COVID-19 Literature

Conference Talk, ACL'20 Jul., 2020
EvidenceMiner: Textual Evidence Discovery for Life Sciences

Invited Talk, School of Information@UT Austin Apr. 23, 2020
Automatic Named Entity Recognition and Evidence Mining in COVID-19 Literature

Tutorial, IEEE-BigData'19 Dec. 11, 2019
Taming Unstructured Big Data: Automated Information Extraction from Massive Text

Conference Talk, IEEE-BIBM'19 Nov., 2019
Distantly Supervised Biomedical Named Entity Recognition with Dictionary Expansion

Conference Talk, ISMB'19 Nov., 2019
ClaimMiner: Query-guided Claim Mining in Biomedical Literature

Conference Talk, IEEE-BIBM'18 Aug., 2018
PENNER: Pattern-enhanced Nested Named Entity Recognition in Biomedical Literature

Conference Talk, ACM-BCB'18 Aug., 2018
Open Information Extraction with Meta-pattern Discovery in Biomedical Literature

Honors and Awards

Best Demo Paper Award, NAACL 2021
Yee Fellowship Award, UIUC 2020 - 2021
Data Mining Research Excellence Award (Honorable Mentioning), UIUC 2019
Student Travel Grant Award, IEEE-BIBM 2019
Biochemistry Student Travel Grant Award, UIUC 2015
Scholarship of Tsinghua Talented Program in Life Sciences, Tsinghua University 2012

Teaching Experiences

Teaching Assistant, UIUC, Champaign, Illinois USA

CS 410 (Text Information Systems): Spring-2019
CS 412 (Introduction to Data Mining): Spring-2017, Summer-2017, Fall-2018
MCB 101 (Introductory Microbiology Laboratory): Fall-2015
MCB 450 (Introductory Biochemistry): Fall-2014, Spring-2015

Mentoring Experiences

*Mentored the research of **three** graduate students and **four** undergraduate students. Among all the students I have mentored, there are **two female students**.*

Yingjun Guan (Ph.D. student, I-School@UIUC, Co-authored [7, 15, 16])
Aabhas Chauhan (M.S student, CS@UIUC, Co-authored [8, 15, 16], Current: Amazon)
Vivian Hu (M.S. student, CS@UIUC, Co-authored [4], Current: Google)
Enyi Jiang (Undergrad, ECE@UIUC, Co-authored [15, 16], Current: Ph.D. student, ECE@UIUC)
Weili Liu (Undergrad, CS@UIUC, Co-authored [15, 16], Current: M.S. student, CS@UC Berkeley)
Xiangchen Song (Undergrad, CS@UIUC, Co-authored [4, 6, 15], Current: Ph.D student, CS@CMU)
Bangzheng Li (Undergrad, CS@UIUC, Co-authored [6, 15])

Research Grant Experiences

*Assisted in the preparation of proposals for the following **two** research grants:*

NSF-IIS: Medium: Collaborative Research: Mining and Leveraging Knowledge Hypercubes for Complex Applications (Total Awarded: $400,000) Oct., 2020 – Sep., 2023

NSF-IIS: Medium: Collaborative Research: StructNet: Constructing and Mining Structure-Rich Information Networks for Scientific Research (Total Awarded: $411,730) Jul., 2017 – Jun., 2022

External Collaborations

UC Davis Heart Failure, Heart Transplantation and Mechanical Circulatory Support Program, UC Davis Medical Center, Bethesda, MD, USA
Collaborations on precision medicine development for heart diseases with Dr. David Liem
Using text mining to identify target proteins for six main groups of heart diseases

National Center for Biotechnology Information (NCBI), National Institutes of Health (NIH), Bethesda, MD, USA
Regular discussions on biomedical text mining with Dr. Zhiyong Lu

IBM Research - AI, Yorktown Heights, NY, USA

Summer Intern Jun. - Aug., 2020
Project: *Transformer-based Link Prediction over Open Knowledge Graphs*
Advisor: Dr. Alfio M. Gliozzo

Eli Lilly and Company (Pharmacutical Company), Indianapolis, IN, USA
Regular discussions on clinical text mining with Dr. Sean Liu

Professional Experiences

Session Chair
IEEE-BigData 2019 Conference

Journal Reviewer
Bioinformatics, Journal of the American Medical Informatics Association (JAMIA)

Conference Reviewer
KDD'21, TheWebConf'21, IJCAI'21, KDD'20, KDD'19, ICDM'19, AAAI'19, KDD'18, AAAI'18

Open-Source Contributions

CORD-NER (`https://xuanwang91.github.io/2020-03-20-cord19-ner/`) is a fine-grained named entity annotated dataset on the COVID-19 Open Research Dataset Challenge corpus. CORD-NER covers 75 biomedical entity types related to COVID-19. It has been downloaded more than 350 times by users world-wide in the past year.

ChemNER (`https://github.com/xuanwang91/ChemNER`) is a recently released fine-grained chemistry named entity recognition dataset (published in EMNLP'21 [4]). ChemNER covers 62 chemistry types and can be used to benchmark distantly supervised NER methods for the fine-grained chemistry named entity recognition task.

EvidenceMiner (`https://evidenceminer.com/`) is a web-based system that allows a user's query as a natural language statement or an inquired relationship at the meta-symbol level (e.g., DRUG treat DISEASE) and automatically retrieves textual evidence from a background corpora of COVID-19. EvidenceMiner has users (including biomedical and clinical researchers) from various universities and institutions (e.g, UCLA Medical School, UC Davis Medical School, Army Research Lab, etc.).

SciContrast (`https://scicontrast.firebaseapp.com/`) is a web-based system for comparative knowledge discovery in scientific literature. SciContrast saves researchers' time and efforts by providing a focused list of prioritized candidates for researchers to explore before they conduct any expensive wet-lab experiments.

References

Jiawei Han (Advisor)
Michael Aiken Chair Professor
Department of Computer Science, UIUC
201 N. Goodwin Avenue, SC 2132
Urbana, IL 61801, USA
E-mail: hanj@illinois.edu

Heng Ji
Professor, Amazon Scholar
Department of Computer Science, UIUC
201 N. Goodwin Avenue, SC 3318
Urbana, IL 61801, USA
E-mail: hengji@illinois.edu

Zhiyong Lu
Deputy Director, Literature Search, NCBI
Senior Investigator, NLM/NIH
8600 Rockville Pike, Building 38A, Room 1003A
Bethesda, MD 20894, USA
E-mail: luzh@ncbi.nlm.nih.gov

David A. Liem, MD, PhD
Associate Director
UC Davis Heart Failure Program
UC Davis Medical Center
4860 Y Street, Suite 2820
Sacramento, CA 95817, USA
Email: daliem@ucdavis.edu