

Xiang Dai

PhD Candidate at
University of Sydney

Email: dai.xiang.au@gmail.com

Summary

I am a final year PhD student from University of Sydney, working on Natural language processing. My research interests mainly focus on information extraction in the biomedical domain. Some concrete problems I have worked on include: recognizing complex (discontinuous, overlapping) entities; pretraining domain-specific models; data augmentation for low-resource NLP; detecting adverse drug events from social media text.

Education

University of Sydney

Ph.D, Natural Language Processing, 2017 - 2021 (Expected)

Supervisors: Sarvnaz Karimi, Ben Hachey, Cecile Paris, Joachim Gudmundsson

Visiting Researcher at *Saarland University*, Aug 2020 - Jan 2021, hosted by Dr. Dietrich Klakow

Research Intern at *Bosch Center for Artificial Intelligence*, Mar 2020 - Jul 2020, supervised by Dr. Heike Adel

University of Sydney

Master's Degree, Information Technology and Information Technology Management, 2015 - 2017

Supervisors: Joachim Gudmundsson, Michael Horton

Research Assistant at *CSIRO Data61*, Feb 2017 - Jul 2017, supervised by Dr. Sarvnaz Karimi

Summer Scholar at *CSIRO Data61*, Nov 2016 - Feb 2017, supervised by Dr. Sarvnaz Karimi

South China University of Technology

Bachelor's Degree, Computer Science, 2003-2007

Experience

Teaching Assistant at University of Sydney

2017 - 2019

Software Developer at Lanke Technology Co., Ltd.

2010 - 2014

Software Developer at Qiaoqing Technology Co., Ltd.

2008 - 2010

Publications

An Analysis of Simple Data Augmentation for Named Entity Recognition

Xiang Dai, Heike Adel

International Conference on Computational Linguistics (COLING 2020)

Cost-effective Selection of Pretraining Data: A Case Study of Pretraining BERT on Social Media

Xiang Dai, Sarvnaz Karimi, Ben Hachey, Cecile Paris

Findings of Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)

NLNDE at CANTEMIST: Neural Sequence Labeling and Parsing Approaches for Clinical Concept Extraction

Lukas Lange, Xiang Dai, Heike Adel, Jannik Strötgen

Iberian Languages Evaluation Forum (IberLEF 2020)

An Effective Transition-based Model for Discontinuous NER

Xiang Dai, Sarvnaz Karimi, Ben Hachey, Cecile Paris

Annual Meeting of the Association for Computational Linguistics (ACL 2020)

NNE: A Dataset for Nested Named Entity Recognition in English Newswire

Nicky Ringland, Xiang Dai, Ben Hachey, Sarvnaz Karimi, Cecile Paris, James R. Curran

Annual Meeting of the Association for Computational Linguistics (ACL 2019)

Using Similarity Measures to Select Pretraining Data for NER

Xiang Dai, Sarvnaz Karimi, Ben Hachey, Cecile Paris

Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2019)

Shot Or Not: Comparison of NLP Approaches for Vaccination Behaviour Detection

Aditya Joshi, Xiang Dai, Sarvnaz Karimi, Ross Sparks, Cecile Paris, C Raina MacIntyre

EMNLP Workshop on Social Media Mining for Health Applications (SMM4H 2018)

Shared task system paper: our system ranked 1st for task 4 of the SMM4H 2018 shared task among all 17 systems.

Recognizing Complex Entity Mentions: A Review and Future Directions

Xiang Dai

ACL Student Research Workshop (ACL-SRW 2018)

Medication and Adverse Event Extraction from Noisy Text

Xiang Dai, Sarvnaz Karimi, Cecile Paris

Australasian Language Technology Association Workshop (ALTA 2017)

Automatic Diagnosis Coding of Radiology Reports: A Comparison of Deep Learning and Conventional Classification Methods

Sarvnaz Karimi, Xiang Dai, Hamed Hassanzadeh, Anthony Nguyen

ACL Workshop on Biomedical Natural Language Processing (BioNLP 2017)

Talks

2020-09 Challenges and Solutions for Medical Concept Extraction

Spoken Language Systems research group at the Saarland University

2020-07 An Analysis of Simple Data Augmentation for Named Entity Recognition

Natural Language Processing and Semantic Reasoning research group at the Bosch Center for Artificial Intelligence

2020-04 Complex Entity Recognition

Natural Language Processing and Semantic Reasoning research group at the Bosch Center for Artificial Intelligence

2020-02 Complex Entity Recognition

Language and Social Computing team at the CSIRO Data61

Other Academic Activities

Journal reviewing

ACM Transactions on Intelligent Systems and Technology (2018)

International Journal of Medical Informatics (2018)

Language Resources and Evaluation (2020)

Conference PC (reviewing)

ACL 2020

EACL 2021

EMNLP 2019 (secondary reviewer), 2020

NAACL 2021

ALTA 2019, 2020

Student Research Workshop at ACL 2020

Student Research Workshop at AACL-IJCNLP 2020

Workshop on NLP for COVID-19 at EMNLP 2020

Member of the ALTA (Australasian Language Technology Association) executive committee

2019-2020

Honors and Awards

Postgraduate Research Support Scheme

USYD, 2019 and 2020

Engineering and Information Technologies Research Scholarship

Faculty of Engineering and Information Technologies, USYD, 2018 - 2021

Data61 top-up Scholarship

Data61, CSIRO, 2018 - 2021

SIT Summer Scholarships

School of Information Technologies, USYD, November 2016 - February 2017

Postgraduate Coursework High Honour Roll

School of Information Technologies, USYD, April 2016, May 2017 and June 2018

MIT and MITM Half-Fee Scholarship
School of Information Technologies, USYD, April 2016

Last updated: October 21, 2020