Xinya Du

Postdoctoral Research Associate of Computer Science, UIUC
Webpage: https://xinyadu.github.io Email: xinyadu2@illinois.edu

Research Areas Natural Language Processing, Computational Linguistics, Machine Learning

Education

Cornell University, Ithaca, NY, USA
Ph.D. in Computer Science, Advisor: Claire Cardie 2016 – 2021
M.S degree granted in 2019

Shanghai Jiao Tong University, Shanghai, China
B.E. in Computer Science and Engineering, GPA Rank: Top 1% 2012 – 2016

Professional Experience

University of Illinois at Urbana-Champaign, Department of CS Sep 2021 – Present
Postdoctoral Research Associate
Supervisor: Heng Ji

Cornell University, Department of CS Aug 2016 – Aug 2021
PhD Research Assistant
Supervisor: Claire Cardie

Google AI, Research Intern May 2020 – Aug 2020
Topic: QA-Driven Zero-shot Slot Filling with Weak Supervision Pretraining
Supervisors: Qi Li, Luheng He

Allen Institute for Artificial Intelligence, Research Intern Sep 2018 – Dec 2018
Topic: Improving Procedural Paragraph Understanding by Leveraging Consistency Contraints
Supervisors: Bhavana Dalvi, Peter Clark

Microsoft Research Redmond, Research Intern May 2018 – Aug 2018
Topic: Leveraging web-structured knowledge for question answering
Supervisors: Paul Bennett, Ahmed Hassan Awadallah

Cornell University, Undergrad Research Intern Aug 2015 – Dec 2015
Topic: Researched on deep learning techniques for fine-grained opinion extraction
Supervisor: Claire Cardie

Publications

Dynamic Global Memory for Document-level Argument Extraction
Xinya Du, Sha Li, and Heng Ji
Proceedings of ACL, 2022

Automatic Error Analysis for Document-level Information Extraction
Aliva Das, **Xinya Du**, Barry Wang, Jiayuan Gu, Kejian Shi, Thomas Porter, and Claire Cardie
Proceedings of ACL, 2022

Template Filling with Generative Transformers
Xinya Du, Alexander M. Rush, and Claire Cardie
Proceedings of NAACL, 2021

GRIT: Generative Role-filler Transformers for Document-level Event Entity Extraction
Xinya Du, Alexander M. Rush, and Claire Cardie
Proceedings of EACL, 2021

Few-shot Intent Classification and Slot Filling with Retrieved Examples
Dian Yu, Luheng He, Yuan Zhang, **Xinya Du**, Panupong Pasupat and Qi Li
Proceedings of NAACL, 2021

QA-Driven Zero-shot Slot Filling with Weak Supervision Pretraining
Xinya Du, Luheng He, Qi Li, Dian Yu, Panupong Pasupat and Yuan Zhang
Proceedings of ACL, 2021

Event Extraction by Answering (Almost) Natural Questions
Xinya Du and Claire Cardie
Proceedings of EMNLP, 2020

Improving Event Duration Prediction via Time-aware Pre-training
Zonglin Yang, **Xinya Du**, Alexander M. Rush and Claire Cardie
Proceedings of Findings of EMNLP, 2020

Document-Level Event Role Filler Extraction using Multi-Granularity Contextualized Encoding
Xinya Du and Claire Cardie
Proceedings of ACL, 2020

Leveraging Structured Metadata for Improving Question Answering on the Web
Xinya Du, Adam Fourney, Robert Sim, Claire Cardie, Paul Bennett and Ahmed Hassan Awadallah
Proceedings of AACL, 2020

Be Consistent! Improving Procedural Text Comprehension using Label Consistency
Xinya Du, Bhavana Dalvi, Niket Tandon, Antoine Bosselut, Wen-tau Yih, Peter Clark, Claire Cardie
Proceedings of 17th North American Chapter of the Association for Computational Linguistics (NAACL), 2019

Harvesting Paragraph-Level Question-Answer Pairs from Wikipedia
Xinya Du and Claire Cardie
Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018

Identifying Where to Focus in Reading Comprehension for Neural Question Generation
Xinya Du and Claire Cardie
Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2017

Learning to Ask: Neural Question Generation for Reading Comprehension
Xinya Du, Junru Shao and Claire Cardie
Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL), 2017
– **Featured in *New Scientist* [Link], *TechRepublic***
– **The first paper to propose end-to-end learning for natural question generation**, inspiring work in QA (Alberti 2019), Summarization (Wang 2020), and Dialogue Systems (Laban 2020).

Cornell Belief and Sentiment System at TAC 2016
Vlad Niculae, Kai Sun, Xilun Chen, Yao Cheng, **Xinya Du**, Esin Durmus, Arzoo Katiyar, Claire Cardie
Text Analysis Conference (TAC), 2016

Awards & Scholarships

CDAC Spotlight Rising Star	2020
Outstanding Graduate of Shanghai Jiao Tong University	2016
National Scholarship (Highest honor for undergraduates in China, awarded to top 1% students.)	2013
National Olympiad in Mathematics in Provinces 2011, The First Prize.	2011

National Olympiad in Informatics in Provinces 2010, The First Prize. 2010

Media Coverage

New Scientist
"Inquisitive bot asks questions to test your understanding" [Link]

TechRepublic
"How researchers trained one AI system to start asking its own questions" [Link]

Mentoring

Zonglin Yang, Cornell CS MEng student → NTU PhD, Spring 2020 – Present
Topic: Commonsense and case-based reasoning for NLP
with publication in Findings of EMNLP 2020

Barry Wang, Cornell CS undergraduate student, Summer 2021 – Present
Topic: Automatic error analysis for information extraction
with publication in ACL 2022, work presented in SciNLP 2021

Aliva Das, Cornell CS undergraduate student, Summer 2021 – Present
Topic: Automatic Error Analysis for information extraction
with publication in ACL 2022, work presented in SciNLP 2021

Maitreyi Chatterjee, Cornell CS undergraduate student Spring 2021
Topic: Applying neural doc-level IE model to scientific domain

Rishi Malhotra, Cornell CS undergraduate student Spring 2021
Topic: Applying neural doc-level IE model to scientific domain

Teaching

Natural Language Processing (TA for Prof. Claire Cardie), Fall 19

Natural Language Processing (TA for Prof. Yoav Artzi), Spring 19

Software Engineering (TA for Prof. William Arms), Spring 17, Spring 18

Introduction to Computing Using Python (TA for Prof. Walker White), Fall 16

Talks

Towards More Informed Extraction of Events from Documents
Rising Stars in Data Science Workshop at University of Chicago, Jan 2021
Tencent AI Research, Nov 2020

Event Extraction by Answering (Almost) Natural Questions
EMNLP 2020 (Online), Nov 2020
UIUC Class Information Extraction and Knowledge Acquisition (Online), Sep 2020

LwLL: Progress on the NLP Front
DARPA site visit, Cornell University (Online), Apr 2020

Harvesting Paragraph-Level Question-Answer Pairs from Wikipedia
ACL 2018, Jul 2018

Services

Program Committee Member

Year 2018: ACL, CoNLL, MRQA, W-NUT
Year 2019: NAACL, ACL, EMNLP, CoNLL, *SEM, W-NUT, IJCAI, AAAI
Year 2020: ACL, IJCAI, EMNLP, AACL
Year 2021: EMNLP, ACL ARR

Journal Reviewer

IEEE Transactions on Knowledge and Data Engineering (TKDE); IEEE Transactions on Audio, Speech and Language Processing (TASLP); ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP); ACM Transactions on Knowledge Discovery from Data (TKDD); AI Communications; Geoscience Frontiers;

Volunteer

ACL 2017, 2018; EMNLP 2017
Cornell CS PhD Visit Day 2019, 2020, 2021
Cornell CS PhD Admission Committee 2021

Code Release

NQG (over 300★) [Link]
Open source repository for neural question generation from sentences (ACL 2017).
Including evaluation scripts and implementations for the models.

EEQA (over 100★) [Link]
Open source repository for using QA techniques for extracting events from text (EMNLP 2020). Including different question templates and implementations for the models.

GRIT: Document Event Entity Extractor (over 50★) [Link]
Open source repository for document-level entity extraction (EACL 2021). Including implementations for our model, evaluation scripts for document-level extraction, and analysis for predictions.

References

Claire Cardie
Professor, Department of Computer Science and Department of Information Science
Cornell University
cardie@cs.cornell.edu

Heng Ji
Professor, Department of Computer Science
University of Illinois at Urbana-Champaign
hengji@illinois.edu

Alexander M. Rush
Associate Professor, Department of Computer Science
Cornell University
arush@cornell.edu

Peter Clark
Senior Research Manager
Allen Institute for Artificial Intelligence (AI2)
peterc@allenai.org

Ahmed Hassan Awadallah
Principal Research Manager, Language and Information Technologies Team
Microsoft Research AI
hassanam@microsoft.com

[*Updated on 2022-02-24*]