

XIANG LORRAINE LI

xiangl@cs.umass.edu ◊ 317-965-5496
<https://people.cs.umass.edu/~xiangl/>

RESEARCH INTEREST

Commonsense knowledge representation & evaluation, natural language processing, machine learning

EDUCATION

University of Massachusetts Amherst Sep.2016 – Aug.2022 (expected)
PhD in Computer Science Amherst, MA
Thesis: Probabilistic Commonsense Knowledge
Advisor: Andrew McCallum
Committee: Yejin Choi (U.Washington), Mohit Iyyer, Laure Thompson, Brian Dillon (Linguistics)

University of Chicago Sep.2014 - Dec.2015
MS in Computer Science Chicago, IL

East China Normal University Sep.2010 – Jun.2014
BE in Software Engineering Shanghai, China

RESEARCH INTERNSHIP EXPERIENCE

DeepMind Feb.2021 – Aug.2021
Mentors: Aida Nematzadeh, Adhiguna Kuncoro, Phil Blunsom
Project: A systematic investigation of commonsense understanding in large language models.

Facebook AI Research May.2020 – Aug.2020
Mentors: Douwe Kiela, Sebastian Riedel
Project: Answering complex open-domain questions with multi-hop dense retrieval.

Bloomberg May.2019 – Aug.2019
Mentor: Gideon Mann, David Rosenberg
Project: Grey-box fuzzing with the guidance of program execution tree.

Google May.2018 – Aug.2018
Mentor: Colin Evans, Chris Waterson
Project: Smoothing the geometry of probabilistic box embeddings.

Google May.2017 – Aug.2017
Mentor: Colin Evans, Chris Waterson
Project: Representation learning and evaluation for hierarchical relations.

ADDITIONAL RESEARCH EXPERIENCE

University of Southern California Jun.2016 - Aug.2016
Mentors: Kevin Knight, Daniel Marcu
Project: Bio-medical event extraction using abstract meaning representation (AMR).

Toyota Technological Institute at Chicago Jun.2015 - Apr.2016
Mentor: Kevin Gimpel
Project: Commonsense knowledge base completion with neural networks.

TEACHING

- Head Teaching Assistant, UMass Amherst** Jan.2021 - May.2021
Course: Advanced Data Science with Projects (COMPSCI 696DS)
- Coordinated with industry mentors from 12 companies (e.g Google, Adobe) on 23 projects.
 - Gave lectures to over 70 students including course overview and research method strategies.
 - Hosted weekly TA hours, and provided feedback for each project throughout the course.
 - Projects results in 8 publications in top-tier conferences and workshops (e.g EMNLP, NeurIPS).
- Guest Lecture, UMass Amherst** Nov.2020
Course: Advanced Natural Language Processing (COMPSCI 685, MS/PhD level)
- Guest Lecture, Johns Hopkins University** Nov.2020
Course: Artificial Intelligence (EN.601.464, BS level)
- Team Mentor, UMass Amherst** 2018, 2019, 2020
Course: Advanced Data Science with Projects (COMPSCI 696DS)
- 2018 with Google project: Hypernym data collection with sentence as context.
 - 2019 with Google project: Probabilistic embeddings on taxonomies in recommendation system.
 - 2020 with IBM project: Natural language inference for question answering.
- Teaching Assistant, East China Normal University** Nov.2013
Course: Android Game Design

PUBLICATIONS

- [1] Xuelu Chen*, Michael Boratko*, Muhao Chen, Shib Sankar Dasgupta, **Xiang Lorraine Li**, Andrew McCallum. “Probabilistic Box Embeddings for Uncertain Knowledge Graph Reasoning.” Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021.
- [2] Anshuman Mishra, Dhruvesh Patel, Aparna Vijayakumar, **Xiang Lorraine Li**, Pavan Kapanipathi, Kartik Talamadupula. “Looking Beyond Short-Premise Natural Language Inference for Downstream Tasks.” Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021.
- [3] Shib Sankar Dasgupta, **Xiang Lorraine Li**, Michael Boratko, Dongxu Zhang, Andrew McCallum. “Box-To-Box Transformations for Modeling Joint Hierarchies.” The Sixth Workshop on Representation Learning for NLP at ACL (ACL WS), 2021.
- [4] Wenhan Xiong*, **Xiang Lorraine Li***, Srinii Iyer, Jingfei Du, Patrick Lewis, William Yang Wang, Yashar Mehdad, Wen-tau Yih, Sebastian Riedel, Douwe Kiela, Barlas Oğuz. “Answering Complex Open-Domain Questions with Multi-Hop Dense Retrieval.” Ninth International Conference on Learning Representations (ICLR), 2021.
- [5] Shib Sankar Dasgupta*, Michael Boratko*, Dongxu Zhang, Luke Vilnis, **Xiang Lorraine Li**, Andrew McCallum. “Improving Local Identifiability in Probabilistic Box Embeddings.” Thirty-fourth Conference on Neural Information Processing Systems (NeurIPS), 2020.
- [6] Anshuman Mishra*, Dhruvesh Patel*, Aparna Vijayakumar*, **Xiang Lorraine Li**, Pavan Kapanipathi, Kartik Talamadupula. “Reading Comprehension as Natural Language Inference: A Semantic Analysis.” The 9th Joint Conference on Lexical and Computational Semantics (*SEM), 2020.

- [7] Michael Boratko*, **Xiang Lorraine Li***, Tim O’Gorman*, Rajarshi Das*, Dan Le, Andrew McCallum. “ProtoQA: A Question Answering Dataset for Prototypical Common-Sense Reasoning.” The 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
- [8] Dhruvesh Patel*, Shib Sankar Dasgupta*, Michael Boratko, **Xiang Li**, Luke Vilnis, Andrew McCallum “Representing joint hierarchies with box embeddings.” Automated Knowledge Base Construction (AKBC), 2020.
- [9] **Xiang Li***, Luke Vilnis*, Dongxu Zhang, Michael Boratko, Andrew McCallum “Smoothing the Geometry of Probabilistic Box Embeddings.”, International Conference on Learning Representations (ICLR) 2019. **Oral presentation. 1.5%**
- [10] Luke Vilnis*, **Xiang Li***, Shikhar Murty, Andrew McCallum “Probabilistic Embedding of Knowledge Graphs with Box Lattice Measures.”, The Annual Meeting of the Association for Computational Linguistics (ACL), 2018.
- [11] **Xiang Li**, Luke Vilnis, Andrew McCallum “Improved Representation Learning for Predicting Commonsense Ontologies.”, Workshop on Deep Structured Prediction, International Conference on Machine Learning (ICML WS), 2017.
- [12] **Xiang Li**, Aynaz Taheri, Lifu Tu, Kevin Gimpel, “Commonsense Knowledge Base Completion.”, The Annual Meeting of the Association for Computational Linguistics (ACL), 2016.
- [13] **Xiang Li**, Xiaoyang Xu, Tanu Malik “Interactive provenance summaries for reproducible science.”, e-Science, 2016 IEEE 12th International Conference on, 355-360. (IEEE), 2016

* Equal Contribution

PUBLICATIONS (PRE-PRINTS)

- [1] Jack W. Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Sarah Henderson, Roman Ring, Susannah Young, Eliza Rutherford, Tom Hennigan, Jacob Menick, Albin Cassirer, Richard Powell, George van den Driessche, Lisa Anne Hendricks, Maribeth Rauh, Po-Sen Huang, Amelia Glaese, Johannes Welbl, Sumanth Dathathri, Safron Huang, Jonathan Uesato, John Mellor, Irina Higgins, Antonia Creswell, Nat McAleese, Amy Wu, Erich Elsen, Siddhant Jayakumar, Elena Buchatskaya, David Budden, Esme Sutherland, Karen Simonyan, Michela Paganini, Laurent Sifre, Lena Martens, **Xiang Lorraine Li**, Adhiguna Kuncoro, Aida Nematzadeh, Elena Gribovskaya, Domenic Donato, Angeliki Lazaridou, Arthur Mensch, Jean-Baptiste Lespiau, Maria Tsimpoukelli, Nikolai Grigorev, Doug Fritz, Thibault Sottiaux, Mantas Pajarskas, Toby Pohlen, Zhitao Gong, Daniel Toyama, Cyprien de Masson d’Autume, Yujia Li, Tayfun Terzi, Vladimir Mikulik, Igor Babuschkin, Aidan Clark, Diego de Las Casas, Aurlia Guy, Chris Jones, James Bradbury, Matthew Johnson, Blake Hechtman, Laura Weidinger, Iason Gabriel, William Isaac, Ed Lockhart, Simon Osindero, Laura Rimell, Chris Dyer, Oriol Vinyals, Kareem Ayoub, Jeff Stanway, Lorraine Bennett, Demis Hassabis, Koray Kavukcuoglu, Geoffrey Irving. “Scaling Language Models: Methods, Analysis & Insights from Training Gopher.”
- [2] **Xiang Lorraine Li**, Adhiguna Kuncoro, Cyprien de Masson d’Autume, Phil Blunsom, Aida Nematzadeh. “A Systematic Investigation of Commonsense Understanding in Large Language Models.”
- [3] Shib Sankar Dasgupta, Michael Boratko, Shriya Atmakuri, **Xiang Lorraine Li**, Dhruvesh Patel, Andrew McCallum. “Word2Box: Learning Word Representation Using Box Embeddings”

PROFESSIONAL SERVICES

Workshop Organizer:

The 7th workshop on Representation Learning for NLP at ACL 2022.
Workshop on Commonsense Representation and Reasoning at ACL 2022.
workshop on Commonsense Reasoning and Knowledge Bases at AKBC 2021.

Reviewer:

ACL Rolling Review: 2021—present
NAACL: 2019, 2021
ACL: 2019, 2020, 2021
AKBC: 2019, 2020, 2021
EMNLP: 2019, 2020, 2021
AAAI: 2020, 2021
AAACL: 2020
CoNLL: 2020
Sets & Partitions Workshop at NeurIPS: 2019
The Workshop on Representation Learning for NLP at ACL: 2020, 2021
*The 10th Joint Conference on Lexical and Computational Semantics (*SEM)*: 2021

Student Volunteer:

NAACL 2016, NeurIPS 2017

OUTREACH

Mentor for Underrepresented CS PhD Applicants, UMass Amherst, 2020 2021

Mentored four PhD applicants in the 2020 and 2021 PhD application season. I provided feedback on their application materials (CV, SoP etc) and a one-on-one call to answer any questions.

Mentor for Junior TA, UMass Amherst, 2020

As the senior TA, I provided communication, teaching, and grading guidance to junior TA.

Panelist in CS Woman PhD/PostDoc Panel, UMass Amherst, 2019

Joined CS woman panel discussion for master students at CICS, UMass Amherst about PhD program.

PhD Candidate Host, UMass Amherst, 2018 2019 2020

Volunteered to be the graduate student Candidate Friday host in 2018, 2019, 2020. I was the primary contact for the candidates and provided housing for them during their visit.

High School Tutor, East China Normal University, 2010

Tutored two high school students for a math class.

RESEARCH MENTORING

I approached students with project proposals outlined by myself, interviewed them when there were multiple student candidates, and selected the most suitable one for the project. Once the mentor-mentee relationship is established, I met with students at least once a week (mostly twice a week) to resolve any project questions. I also assisted students with their course projects and any other questions they encountered.

Shib Sankar Dasgupta

PhD Student

Project: Probabilistic box embeddings and its applications.

2019-present
UMass Amherst

Parin Rajesh Jhaveri Master Student <i>Project:</i> Injecting commonsense knowledge for text-adventure games.	2021-present <i>UMass Amherst</i>
Pranay Kumar Yelugam Master Student <i>Project:</i> Generative evaluation of commonsense question answering task.	2021-present <i>UMass Amherst</i>
Nalini Singh Master Student <i>Project:</i> Data collection task design for commonsense frame completion.	2021 <i>UMass Amherst</i>
Melnita Dabre Master Student <i>Project:</i> Re-ranking answers with T5 for commonsense question answering.	2020 <i>UMass Amherst</i>
Wenlong Zhao PhD Student <i>Project:</i> Language model with hierarchical semantics.	2020 <i>UMass Amherst</i>
Rachel Bialik Undergraduate Student <i>Project:</i> Representing AMR graphs with probabilistic box embeddings.	2020-2021 <i>UMass Amherst</i>
Eunjeong Hwang Master Student <i>Project:</i> Answer knowledge base queries with probabilistic box embeddings.	2019-2020 <i>UMass Amherst</i>
Anshuman Mishra Master Student <i>Project:</i> Looking beyond short-premise natural language inference for downstream tasks.	2019-2020 <i>UMass Amherst</i>
Varun Iyer Undergraduate honored thesis <i>Project:</i> Fine-grained entity typing with probabilistic box embeddings.	2019 <i>UMass Amherst</i>
Evan Rourke Master Student <i>Project:</i> Probabilistic box embeddings for taxonomy alignment.	2019 <i>UMass Amherst</i>
Vasishtha Jayapati Master Student <i>Project:</i> Unsupervised training for probabilistic box embedding.	2019 <i>UMass Amherst</i>