Li "Harry" Zhang

zharry.com
zharry@seas.upenn.edu
(734)834-7882

RESEARCH INTERESTS

Deep Learning, Natural Language Processing (NLP), Natural Language Understanding, Semantics, Resources and Evaluation for NLP

EDUCATION

University of Pennsylvania, Philadelphia, PA — Aug 2019 – Present
Ph.D. Computer and Information Science — GPA: 3.90/4.00
University of Michigan, Ann Arbor, MI — Sept 2015 – Dec 2018
B.S.E. Computer Science, *summa cum laude* — GPA: 3.82/4.00

PUBLICATIONS

[6] Q. Lyu*, **L. Zhang*** and C. Callison-Burch. *Intent Detection with WikiHow*. In AACL-IJCNLP 2020. (*Equal contribution.)

[5] **L. Zhang**, Q. Lyu and C. Callison-Burch. *Reasoning about Goals, Steps, and Temporal Ordering with WikiHow*. In EMNLP 2020.

[4] **L. Zhang**, H. Zhu, S. Brahma and Y. Li. *Small but Mighty: New Benchmarks for Split and Rephrase*. In EMNLP 2020.

[3] **L. Zhang**, S. R. Wilson and R. Mihalcea. *Multi-Label Transfer Learning for Semantic Similarity*. In *SEM 2019 and presented at NAACL 2019.

[2] **L. Zhang**, S. R. Wilson and R. Mihalcea. *Direct Network Transfer: Transfer Learning of Sentence Embeddings for Semantic Similarity*. In arXiv and presented at IC2S2 2018.

[1] C. Finegan-Dollak, J. K. Kummerfeld, **L. Zhang**, K. R. D. Ramanathan, S. Sadasivam, R. Zhang and D. Radev. *Improving Text-to-SQL Evaluation Methodology.* In ACL 2018.

RESEARCH EXPERIENCE

Goal-Step Inference with WikiHow — Nov 2019 – Present

Propose resources and evaluation for reasoning about goals and steps using wikiHow

Adviser: Prof. Chris Callison-Burch

• Crawled the wikiHow website and released structured and easy-to-consume data.

• Sampled training data and curated benchmarks for inferring the goal from a step, inferring the step from a goal, and the temporal relation between two steps given a goal.

• Showed strong zero- and few-shot performance on various natural language understanding tasks, using our data for pre-training.

• Showed state-of-the-art performance on various intent detection benchmarks.

Split and Rephrase: Evaluation Benchmarks and Metrics — Apr 2019 – Jun 2019

Revamp the evaluation for Split and Rephrase, rewriting long sentences into short ones

Company: IBM Research

• Developed a rule-based model using no training data which performs on par with the current state-of-the-art neural model.

• Released two new crowdsourced benchmarks with improved quality.

• Conducted a case study on the flaws of BLEU score, and the cost-efficiency of using crowd workers to evaluate model performance.

Transfer Learning in Semantic Similarity Oct 2017 – May 2018

Explore transfer learning methods using sentence embeddings in semantic similarity

Adviser: Prof. Rada Mihalcea

• Proposed a new transfer learning method for semantic similarity tasks, achieving state-of-the-art performance on various datasets using various neural networks architectures.

• Compared and analyzed performances of popular transfer learning methods on a collection of mainstream LSTM-based models and semantic similarity datasets.

• Interpreted qualitatively the source of improvement in the domain of human activities.

Multi-label Learning in Semantic Similarity Mar 2017 – Sept 2018

Explore multi-task learning using sentence embeddings in semantic similarity

Adviser: Prof. Rada Mihalcea

• Proposed a modification of LSTM architecture for semantic similarity datasets with multiple relations, achieving state-of-the-art in various dimensions.

• Compared with multi-task learning and single-task learning baselines.

Active Interpretation of Disparate Alternatives Jan 2017 – Feb 2019

Use multi-modal news reports to generate hypotheses about real life events

Adviser: Prof. Rada Mihalcea and Prof. Jia Deng

• Produced knowledge elements using the text from multiple account of the events regarding the Ukrainian-Russian relations.

• Performed keyword extraction and named entity recognition to extract knowledge elements and assign saliency to them.

Natural Language to SQL in Academic Advising Sept 2015 – Apr 2017

Part of the IBM Sapphire project to build a dialogue system for academic advising

Adviser: Prof. Dragomir Radev

• Implemented a named entity recognizer specifically on the academic advising ontology to automatically expand training data by permutating entities.

• Designed over 50 semantically distinct and meaningful advising questions as well as their corresponding SQL queries to be used as training data.

• Contributed in building the Advising dataset parallel to the ATIS and GeoQuery datasets that contains more than 300 entries in the academic advising domain.

• Presented in 2016 Michigan Research Community poster symposium.

Text Clustering Based on Humor in Cartoons Jan 2016 – Apr 2016

Dataset from caption submissions for the cartoon section of New Yorker magazine

Adviser: Prof. Dragomir Radev

• Restarted and oversaw the project with under-documented codebase.

• Rewrote Perl scripts in Python using state-of-the-art machine learning APIs.

• Experimented with text embeddings such as word2vec and Skip-Thought to compare performances in multiple clustering algorithms such as the Louvain algorithm.

ACL Anthology Network Sept 2016 – Dec 2016

A power taxonomy of papers from top NLP conferences

Adviser: Prof. Dragomir Radev

- Implemented distance metrics between papers classified into the same category.
- Fixed display issues on the front end and did QA on the database.
- Presented in 2016 University of Michigan NLP workshop.

ACADEMIC SERVICES

Paper Reviewing

- International Conference on Computational Linguistics (COLING) 2020
- Computer Speech and Language (CSL) journal. 2018

TEACHING EXPERIENCE

Teaching Assistant — Computational Linguistics Jan 2020 – Dec 2020

CIS 530: The graduate level NLP course University of Pennsylvania

- Held weekly office hours and answered questions online for students.
- Helped design course contents such as homework and quizzes.
- Gave supplementary lectures.

Teaching Assistant — Natural Language Processing Sept 2018 – Dec 2018

EECS 595: The graduate level NLP course University of Michigan

- Held weekly office hours and answered questions online for 135 students.
- Helped design course contents.

Teaching Assistant — Programming and Data Structures Sept 2016 – Apr 2017

EECS 280: An introductory programming course University of Michigan

- Led weekly lab sessions with more than 30 students to review course materials and guide them through hands-on coding challenges.
- Held weekly office hours and answered questions online for over 1,000 students.
- Helped design course contents, projects and exams.

Tutor — Elementary Chemistry Sept 2016 – Dec 2016

Science Learning Center University of Michigan

- Hosted weekly walk-in tutoring for an introductory chemistry course.
- Compiled review materials to help students prepare for exams.

INDUSTRY EXPERIENCE

Split and Rephrase: Evaluation Benchmarks and Metrics Apr 2019 – Jun 2019

IBM Research San Jose, CA

- Conducted NLP research; see more in Research Experience.

Collaborative Filtering Based Recommender System May 2017 – Aug 2017

Goldman Sachs Group, Inc. Jersey City, NJ

Worked in GS App Store, the firm's software management and delivery platform

- Developed a highly scalable recommender system using collaborative filtering to suggest personalized app recommendations for each user.
- Designed features such as "frequently installed together", "users who installed this also installed", "trending apps" and "top apps by business unit".

• Implemented end-to-end interfaces using C#, JavaScript, AngularJS and Elasticsearch.

COURSES

Natural Language Processing(A+), Directed Research (A+), Information Retrieval(A), Machine Learning(A), Artificial Intelligence (A), Computer Security(A), Multivariate Calculus(A+), Probability and Statistics (A-), Matrix Algebra (A-)

HONORS

Merit-Based Scholarship of $2,000, UM Engineering Class of 1935	2017 – 2018
James B. Angell Scholar, University of Michigan	2017
University Honors of all semesters, University of Michigan	2015 – 2018
Dean's Honor List of all semesters, UM College of Engineering	2015 – 2018

SKILLS

Programming Skills

Python, C++, SQL, Elasticsearch, MATLAB, C#, JavaScript…

Language Skills

Chinese (native), English (fully proficient), French (conversational)

TEST SCORES

GRE (May 2017): Verbal 162 91%, Quantitative 170 97%, Analytical Writing 4.5 82%

SAT (Dec 2015): Reading 750, Writing 790, Math 800

TOEFL (Dec 2013): Reading 28, Listening 29, Speaking 29, Writing 28

ACTIVITIES

Artificial Intelligence Labs

Member, Language and Information Technologies Research Group Oct 2017 – Present

Member, Computational Linguistics & Information Retrieval Lab Oct 2015 – May 2017

Billiards and Pool

1st Place out of 12, 9-Ball Scotch Doubles, 1st Midwest Invitational Feb 2018

7th Place out of 32, 9-Ball Scotch Doubles, 14th UM Team Pool Championship Nov 2017

Captain, Michigan Billiards Team Sept 2017 – Present

Membership Chair, Michigan Billiards Club Sept 2016 – Present

E-Sport

Top 32 out of 828 teams, TESPA Hearthstone Collegiate Championship Apr 2018

#29 out of approx. 5 million in Ranked Play, Hearthstone, North America Mar 2018

Member, Michigan Hearthstone Team Sept 2016 – Present