XINYI(CAROL) ZHENG

· zxycrol@umich.edu · (734) 882-7055 · Ann Arbor, MI

Education

University of Michigan **Ann Arbor, MI**
B.S. Computer Science, B.S. Mathematics *expected May 2020*
GPA: 3.9/4.0

- Coursework: Optimization Methods (grad), Natural Language Processing (grad), Matrix Methods for DSP (grad), Advanced Artificial Intelligence (grad), Coding Theory (grad), Computer Organization, Database Management Systems, Programming Language, Operating Systems, Machine Learning, Data Mining, Conversational AI, Numerical Methods

Shanghai Jiao Tong University, UM-SJTU Joint Institute **Shanghai, China**
Electrial and Computer Engineering *Sept. 2015 - Aug. 2017*

Publications and Manuscripts(* indicates equal contribution)

[1] **"Learning to Answer Complex Questions over Knowledge Bases with Query Composition"**, Nikita Bhutani*, Xinyi Zheng*, HV Jagadish. *In 2019 ACM International Conference on Information and Knowledge Management (CIKM-2019).*
[2] **"What is normal, What is Strange, and What is Missing in a Knowledge Graph: Unified Characterization via Inductive Summarization"**, Caleb Belth, Xinyi Zheng, Jilles Vreeken, Danai Koutra. *In The Web Conference, 2020 (WebConf-2020).*
[4] **"Combining Information from Curated and Extracted Knowledge Bases to Answer Complex Questions"**, Nikita Bhutani*, Xinyi Zheng*, Kun Qian, Yunyao Li, HV Jagadish, *To appear in Annual Conference of the Association for Computational Linguistics, 2020.*
[5] **"Global Table Extractor (GTE): A Framework for Joint Table Identification and Cell Structure Recognition Using Visual Context"**, *https://arxiv.org/abs/2005.00589*
[6] **"Mining Persistent Activity in Continually Evolving Networks "**, Caleb Belth, Xinyi Zheng, Danai Koutra. *To appear in Knowledge Discovery and Data Mining, 2020.*

Industry Experience

Scalable Knowledge Intelligence Group, IBM Research - Almaden **San Jose, CA**
Research Intern, Mentor: Dr. Nancy XR Wang, Manager: Dr. Yunyao Li *May 2019 - Aug. 2019*
Documents in PDF format fail to preserve the formatting of any table the document contains. Table detection and cell structure recognition are crucial tasks as tables contain essential sturctured information. However, it has remained a challenging task with unreliable table extractors in both the commercial and academic space.

- Proposed a systematic framework for vision-guided joint table detection and cell structure recognition, which outperforms previous systems on standard table benchmark.
  (a) Leveraged a cell detection network to guide the training of the table detection network
  (b) Presented a hierarchical network and a novel cluster-based algorithm for cell structure recognition by classifying tables, detecting cells and convert this into structure with spatial clustering
- Designed a method to automatically create ground-truth labels for table recognition and used it to create a new dataset,which provides a large real-world data source with fine-grained cell structure annotation for table related tasks

Research Experience

Michigan Database Research Group **Ann Arbor, MI**
Research Intern, advised by Prof. H.V. Jagadish *Apr. 2018 - May 2019*
Answer Complex Questions over Knowledge Bases with Query Composition
A knowledge base(KB) is a massive collection of structured facts about real-world entities. Powered by a knowledge base question answering(KB-QA) system, people can get precise answers instead of pieces of documents when asking de facto questions in search engines. However, existing KB-QA systems cannot answer complex questions that require joining multiple relations. We built a new KB-QA system which automatically translates a complex, compositional question to the matching query over a knowledge base.

- Proposed a novel *decompose-execute-join* approach to construct complex query patterns from partial queries
- Presented a neural network based semantic matching model that learns to score partial queries using implicit supervision from question-answer pairs
- Paper[1] is to appear in CIKM-2019 as a co-first author.

Joint query extracted and curated knowledge bases to answer complex questions
Curated KBs are often incomplete because they are constructed using closed form information extraction methods on a text corpus. An alternative paradigm is extracted KB that assumes an open form relation. Extracted KBs have higher coverage than curated KBs, but they are harder to query. We aimed to integrate inference over curated and extracted KBs for answering complex questions. Combining information from multiple sources offers two benefits: evidence scattered across multiple KBs can be aggregated, and evidence from different KBs can be used to complement each other.

- Proposed a novel KB-QA system that combines information from curated and extracted knowledge bases to answer complex questions. To leverage multiple KBs, we constructed query patterns for complex questions using simple queries each targeting a specific KB
- Presented a neural-network based model that aligns diverse relation forms from multiple KBs for collective inference

Graph Exploration and Mining at Scale (GEMS) lab **Ann Arbor, MI**
Research Intern, advised by Prof. Danai Koutra *Aug. 2019 - Present*
Knowledge graphs(KGs) often contain errorneous or incomplete information, Research in KG *refinement* attempt to resolve this issue. However, existing approaches are tailored techniques to either detect specific types of errors or complete a KG. We studied the problem of unified KG refinement.

- Rather than targeting a specific refinement task, we unified various KG refinement tasks by joining the problems of refinement and unsupervised summarization
- Proposed to summarize rules of KGs in the form of labeled, rooted graphs describing *arbitrary* graph structure centered around a node
- Constructed novel experimental setup for evaluating knowledge graph patterns

Teaching Experience

UM-SJTU Joint Institute **Shanghai, China**
Teaching Assistant, VV255(Honors Calculus III), Instructor: Dr. Jing Liu *May 2017 - Aug. 2017*

- Graded > 200 students homework, and developed homework problem and solutions every week
- Taught a discussion session and held office hours every week

Service

Discover-CS Coffee Mentor *Ann Arbor, MI* *Fall 2018, Fall 2019*
Database Reading Group Organizer *Ann Arbor, MI* *Winter 2019*

Voluntary Experience

Vice President, UM-SJTU Student and Alumni Association *May 2018 - May 2019*
Outreach Director, Michigan China Forum *Oct. 2017 - Mar. 2018*
President of Liason Department, Alumni Association of No.2 High School of ECNU *Sept. 2015 - Sept. 2016*

Skills

Programming Languages: Python, Java, C/C++, Matlab, Julia, SQL
Framework: Pytorch, Keras, Tensorflow, Hadoop, Flask, CoreNLP, Mallet, ElasticSearch, NLTK
Tools: Git, Screen, LaTeX, Virtuoso

Honors and Awards

CRA Outstanding Undergraduate Research Award Honorable Mention *Dec. 2019*
2019 CRA-W GHC Research Scholar *Orlando, FL* *Oct. 2019*
3^{rd} place, Bloomberg Coding Challenge *Ann Arbor, MI* *Nov. 2017*
Dean's Honor List, University of Michigan *Ann Arbor, MI* *Every Semester*