Weiqiu You

+1(978)778-0875 | weiqiuy@seas.upenn.edu | fallcat.github.io

Research Interests

Natural language processing, multimodality, reasoning

Education

Ph.D. in Computer Science, University of Pennsylvania, Philadelphia, PA GPA: 3.90/4.00 *Sep. 2020 - Present*
- Advisor: Prof. Mark Yatskar

M.S. in Computer Science, University of Massachusetts, Amherst, MA GPA: 3.90/4.00 *Sep. 2018 – May 2020*
- Advisor: Prof. Mohit Iyyer

B.S. in Computer Science & Mathematics, Gordon College, MA GPA: 3.87/4.00 *Sep. 2014 – May 2018*
- Advisors: Prof. Jonathan Senning, Prof. Russell Bjork
- Honors Thesis: Predict Media Interestingness

Study Abroad: Aquincum Institute of Technology, Budapest, Hungary GPA: 4.00/4.00 *Aug. 2017 – Dec. 2017*

Publication

QuakerBot: A Household Dialog System Powered by Large Language Models *Alexa Prize Taskbot 2022*
Artemis Panagopoulou, Manni Arora, Li Zhang, Dimitri Cugini, **Weiqiu You**, Yue Yang, Liyang Zhou, Yuxuan Wang, Zhaoyi Hou, Alyssa Hwang, Lara Martin, Sherry Shi, Chris Callison-Burch, Mark Yatskar

Macro-Average: Rare Types Are Important Too *NAACL 2021*
Thamme Gowda, **Weiqiu You**, Constantine Lignos, Jonathan May

Hard-Coded Gaussian Attention for Neural Machine Translation *ACL 2020*
Weiqiu You*, Simeng Sun*, Mohit Iyyer ** equal contribution*

Research Experience

Enforcing Consistency in Visual Question Answering *Sep. 2022 – Present*
Enforcing consistency not just during training, but at test time
Advised by Prof. Mark Yatskar
- Generate related questions and enforce consistency of related questions at test time

Zero-Shot Image Classification Using Textual Descriptions *Sep. 2020 – May 2022*
BERT is not necessarily as good as TFIDF; CLIP only remembers names
Advised by Prof. Mark Yatskar
- BERT is worse than TFIDF when used as features for representing textual descriptions for representing classes in zero-shot image classification.
- Discriminative pretraining improves large pretrained models like BERT.
- CLIP only gives high scores for images matching with text that contain the names of the class, while not able to compositionally recognize text describing components of the class.

Cybersecurity Attack Technique Classification with Knowledge Graph & Descriptions *May 2022 – Aug. 2022*
Data augmentation by selecting similar data and making data similar
Advised by Dr. Youngja Park, during internship at IBM Research
- When training a classification model with low-resource settings and extra out-of-distribution data, it helps to augment training with similar data from the OOD data.
- Minimizing distance of in- and out-of- distribution data helps as a pretraining objective.

Procedural Entity Tracking *Jan. 2022 – Oct. 2022*
Knowing entity changes help with reasoning about events in a procedure
Working with Li "Harry" Zhang, Advised by Prof. Chris Callison-Burch
- Annotated data for entities and events.
- Transfer learning from existing entity state tracking dataset.

Amazon Alexa Prize Taskbot Project *Sep. 2021 – May 2022*

Household dialog system powered by large language models

Advised by Prof. Chris Callison-Burch & Prof. Mark Yatskar

- We built a system for dialogs for recipes and household improvement tasks.
- In the system, I worked on building a harm classifier based on zero-shot BART-MNLI.
- Improve intent classification model from feedback from real user interactions.

Qualitative Unsupervised Machine Translation *May 2020 – Aug. 2020*

Unsupervised machine translation models have more problems only detectable by MacroF1

Advised by Prof. Jonathan May, during internship at ISI

- Comparing to supervised neural machine translation (SNMT), unsupervised neural machine translation models (UNMT) have more untranslations, truncations, etc.
- The problems can be detected by MacroF1 but not BLEU because they are on rare classes.

Hard-Coded Gaussian Attention for Neural Machine Translation *Aug. 2019 – May 2020*

We don't need to learn all the heads, but can focus more attention on the local area

Advised by Prof. Mohit Iyyer

- Modified multi-headed attention of Tranformer in NMT to hard-coded Gaussian attention.
- Reduced memory and inference time speed without much BLEU drop.
- Learning one head in cross attention can recover most BLEU.

Key Phrase Extraction on Delivery FAQ Data *Jun. 2018 – Aug. 2018*

Rule-based key phrase system based on result from constituency and dependency parsers

Advised by Dr. Zhongyuan Wang, during internship at Meituan-Dianping Inc.

- Literature review in short text summarization and key phrase extraction.
- Used rule-based methods, statistics based methods, and deep learning models (CSDDM) in Python to successfully extract key phrases for customer QA data.

Internship Experience

Research Intern, IBM Research Yorktown Heights *May 2022 – Aug. 2022*

- Mentor: Dr. Youngja Park
- Project: Cybersecurity attack technique classification with knowledge graph & descriptions

Research Assistant, USC ISI (Information Sciences Institute) *May 2020 – Aug. 2020*

- Mentor: Prof. Jonathan May
- Project: Qualitative unsupervised machine translation

Research Intern, NLP Center, Meituan-Dianping Inc. *Jun. 2018 – Aug. 2018*

- Mentor: Dr. Zhongyuan Wang
- Project: Key phrase extraction on delivery FAQ data

Teaching Experience

Teaching Assistant, Computer and Information Science Department, UPenn *Spring & Fall 2021*

- CIS530 Computational Linguistics

Grader, College of Information and Computer Science, UMass Amherst *Spring 2020*

- COMPSCI685 Advanced Natural Language Processing

Teaching Assistant, Math and Computer Science Department, Gordon College *Jan. 2016 – May 2018*

- CPS222 Data Structures & Algorithms, MAT122 Calculus II, MAT225 Differential Equations
- Calculus and SPSS in Biostatistics Help sessions

Skills

Programming Skills	Python, Java, C, C++, Coq, JavaScript, PyTorch, Numpy, ASP.NET, etc.
Language Skills	Chinese (native), English (fluent), Japanese (intermediate), Hungarian (beginner)

Academic Services

Paper Reviewing	EMNLP (2021, 2022), ACL Rolling Review (2022)