Weiqiu You

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

Research Interests

Machine learning, explainable AI

Education

Ph.D. in Computer Science, University of Pennsylvania GPA: 3.94/4.00 *Sep. 2020 – Present*
- *Advisor: Prof. Eric Wong* *Philadelphia, PA*

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

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

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

Publication

[6] **Weiqiu You**, Helen Qu, Marco Gatti, Bhuvnesh Jain, Eric Wong. *Sum-of-Parts Models: Faithful Attributions for Groups of Features.* Preprint.

[5] Youngja Park, **Weiqiu You**. *A Pretrained Language Model for Cyber Threat Intelligence*. In EMNLP 2023 industry track.

[4] Li Zhang, Hainiu Xu, Yue Yang, Shuyan Zhou, **Weiqiu You**, Manni Arora, Chris Callison-Burch. *Causal Reasoning of Entities and Events in Procedural Texts*. In Findings of EACL 2023.

[3] 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. *QuakerBot: A Household Dialog System Powered by Large Language Models*. In Alexa Prize Taskbot Challenge Preceedings.

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

[1] **Weiqiu You***, Simeng Sun*, Mohit Iyyer. *Hard-Coded Gaussian Attention for Neural Machine Translation*. In ACL 2020. (* equal contribution)

Internship Experience

Research Intern, IBM Research Yorktown Heights *May 2022*
- Mentor: Dr. Youngja Park *– Aug. 2022*
- Project: Augment cybersecurity attack technique classification with class descriptions [5,6]

Research Assistant, USC ISI (Information Sciences Institute) *May 2020*
- Mentor: Prof. Jonathan May *– Aug. 2020*
- Project: Qualitative unsupervised machine translation [2]

Research Intern, NLP Center, Meituan-Dianping Inc. *Jun. 2018*
- Mentor: Dr. Zhongyuan Wang *– Aug. 2018*
- Project: Key phrase extraction on delivery FAQ data

Research Experience

Sum-of-Parts Models: Faithful Attributions for Groups of Features [6] *Mar. 2023*
Overcoming fundamental barriers in feature attribution methods with grouped attributions *– Sep. 2023*
Advised by Prof. Eric Wong
- We prove that feature attributions must incur at least exponentially large error in tests of faithfulness for simple settings. We further show that grouped attributions can overcome this limitation.
- We develop Sum-of-Parts (SOP), a class of models with group-sparse feature attributions that are faithful by construction and are compatible with any backbone architecture.
- We evaluate our approach in standard image benchmarks with interpretability metrics.
- In a case study, we use faithful attributions of SOP from weak lensing maps and uncover novel insights about galaxy formation meaningful to cosmologists.

Visual Topics via Visual Vocabularies *Jun. 2023*
Topic modeling can explain relations in image datasets too *– Sep. 2023*
Advised by Prof. Eric Wong
- We propose visual topic modeling to explain hidden themes in an image dataset. Our methodology derives a visual vocabulary from images as an interface between image datasets and topic modeling algorithms.
- We demonstrate, via experiments and a theoretical example, how visual topics capture relationships in images distinct from what existing dimensionality reduction methods capture.
- We adopt standard topic modeling evaluations from the NLP literature to assess our visual topics. We find our topics to be of good quality according to topic modeling metrics and highly interpretable via human evaluation.

Two-stage Training with Data Augmentation for Cyber Attack Technique Classification [5] *May 2022*
Out-of-distribution data can be more useful when selected and combined better *– Aug. 2022*
Advised by Dr. Youngja Park, during internship at IBM Research
- Simply adding another data source does not necessarily help if the two datasets have different distributions.
- We propose a similarity-based data selection and a two-stage training method for utilizing out-of-distribution data with the same labels to improve classification for low-resource domains.
- We select similar samples from the out-of-distribution data to add to rare classes of in-distribution training data.
- Next, the system first trains the model using augmented training data and then trains more with only the in-distribution data.
- Our method improves Macro-F1 by 5-10 points and keeps Micro-F1 competitive to the baselines on the TRAM dataset for cybersecurity attack technique classification.

Procedural Entity Tracking with Multi-hop Reasoning [4] *Jan. 2022*
Knowing entity changes help with reasoning about events in a procedure *– Oct. 2022*
Led by Li "Harry" Zhang, Advised by Prof. Chris Callison-Burch

- Entities and events are important to natural language reasoning and common in procedural texts.
- We propose a beanchmark on causal reasoning of event plausibility and entity states.
- We show that most language models, including GPT-3, perform close to chance at .35 F1.
- We boost model performance to .59 F1 by creatively representing events as programming languages while prompting language models pretrained on code.
- We inject the causal relations between entities and events as intermediate reasoning steps in our representation.
- We use code-like prompting and chain-of-thought reasoning for multi-hop event reasoning.

Amazon Alexa Prize Taskbot Project [3] *Sep. 2021*
Household dialog system powered by large language models *– May 2022*
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.
- Improved intent classification model from feedback from real user interactions.

Qualitative Unsupervised Machine Translation [2] *May 2020*
Unsupervised machine translation models have more problems only detectable by MacroF1 *– Aug. 2020*
Advised by Prof. Jonathan May, during internship at ISI

- Unsupervised NMT models have more untranslations & truncations than supervised NMT.
- The problems can be detected by MacroF1 but not BLEU because they are on rare classes.
- We propose a new evaluation notion "favorism" to compare how much a model favors one translation over another under a certain metric.
- We showed that SNMT models can be up to 24% better in MacroF1 than UNMT while having similar BLEU.

Hard-Coded Gaussian Attention for Neural Machine Translation [1] *Aug. 2019*
We don't need to learn all the heads, but can focus more attention on the local area *– May 2020*
Advised by Prof. Mohit Iyyer

- Modifying multi-headed attention of Transformer in the encoder in NMT to hard-coded Gaussian attention does not decrease model performance.
- Hard-coded Gaussian attention reduced memory and inference time speed without much BLEU drop.
- Learning one head in cross attention recovers most BLEU while maintaining memory and time efficiency.

Key Phrase Extraction on Delivery FAQ Data *Jun. 2018*
Rule-based key phrase system based on result from constituency and dependency parsers *– Aug. 2018*
Advised by Dr. Zhongyuan Wang, during internship at Meituan-Dianping Inc.

- Used dependency and constituency parsers of Stanford CoreParser for rule-based methods and CDSSM for neural method for key phrase extraction on delivery FAQ data.
- Selected and evaluated relations and types of phrases to use from dependency and constituency parses.
- Selecting key phrases with dependency parser rules obtains more human-preferred results than with CDSSM.

Teaching Experience

Teaching Assistant, Computer and Information Science Department, UPenn *Spring 2021*
- CIS530 Computational Linguistics *& Fall 2021*

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*
- CPS222 Data Structures & Algorithms, MAT122 Calculus II, MAT225 Differential Equations *– May 2018*
- Calculus and SPSS in Biostatistics Help sessions

Skills

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

Academic Services

Paper Reviewing	EMNLP (2021, 2022, 2023), ACL (2023), ACL Rolling Review (2022, 2023)
Event Organization	2022 Fall of CLunch, a weekly NLP research seminar run by PennNLP