Haoqiang (Murray) Kang

Tel: +1 2063846726 | Email: haoqik@cs.washington.edu | Website: https://github.com/mk322

Education Background

University of Washington, Seattle, WA 09/2019 - 06/2023
B.S. in Computer Science
Major GPA: 3.86/4.0 GPA in the recent 3 years: 3.92/4.0
Concentration: **Natural Language Processing, Large Language Models, Hallucinations**

Preprints

[1] H. Kang*, T. Blevins*, and L. Zettlemoyer, "Translate to Disambiguate: Zero-shot Multilingual Word Sense Disambiguation with Pretrained Language Models," submitted to EMNLP Conference, 2023. Available: http://arxiv.org/abs/2304.13803v1

Research Experience

Post-Hoc Detection & Mitigation of Long-Form Hallucinations in LLMs 08/2023 - Present

Position: Research Assistant Advisor: Dr. Huaxiu Yao
Purpose: To detect and mitigate the entity-level hallucinations in the long-form generations of LLMs.
Workload:

1. Extracted key entities in a sentence and then constructed atomic facts for each entity.
2. Designed and instituted a comprehensive evidence-retrieval system, followed by a classification process to identify and address both intrinsic and extrinsic hallucinations.
3. Revised "intrinsic hallucinations" using retrieved evidence, and appended user warnings to outputs identified as "extrinsic hallucinations".
4. Boosted factual accuracy of GPT3.5-Turbo in a dataset of 500 human biographies, raising the FactScore from 71.6 to 91.4.

Expected Outcome: To conduct experiments on the Eli5 long-form QA dataset to assess its broader applicability.

Hallucinations in Foundation Language Models in the Finance Domain 08/2023 - Present

Position: Research Assistant Advisor: Dr. Xiao-Yang Liu
Purpose: To investigate the effectiveness of different mitigation strategies of hallucinations for foundation language models into the finance domain.
Workload:

1. Explored the effectiveness of different hallucination detection methods in the task of long-form generations of financial terminologies.
2. Compared retrieval-augmented generation (RAG) techniques against fine-tuning models with knowledge injection from finance texts.

Expected Outcome: To provide foundational research on hallucinations within the finance domain, and draw comparative insights between different mitigation techniques for more reliable financial language modeling.

Multilingual Hallucinations in Large Language Models 05/2023 - 11/2023

Position: Research Assistant Advisor: Prof. Luke Zettlemoyer
Purpose: To scrutinize the phenomenon of factual hallucination in large multilingual language models, with a particular focus on the BLOOMZ and BLOOM model, through an analysis of its pretraining ROOTS corpus.
Workload:

1. Established diverse baselines such as ROUGE and Named Entity Overlap to identify multilingual hallucinations in datasets containing human biographies from Wikipedia across 20 languages.
2. Performed a detailed comparative analysis of each human name in the pretraining ROOTS corpus against the factuality of its generated content.
3. Introduced an innovative cross-lingual knowledge technique to enhance the generation quality in languages with limited resources by knowledge retrieval in high-resource languages.

Expected Outcome: To enhance the understanding of hallucinations in multilingual models and provide strategies to boost their reliability, especially in low-resource languages.

Multilingual Word Sense Disambiguation by Pretrained Language Models 09/2022 - 04/2023

Position: Research Assistant Advisor: Prof. Luke Zettlemoyer

Purpose: To analyze the performance of different language models on their ability to translate words based on a specific in-context usage of the source word in the zero-shot setting.

Workload:

1. Extended the task of word-level translation (WLT) to Contextual Word-Level Translation (C-WLT) and demonstrated that integrating contexts into WLT enhanced the performance of the model.
2. Proposed and demonstrated a zero-shot technique for multilingual Word Sense Disambiguation (WSD) that uses C-WLT as a component, showing its effectiveness in 18 languages, including those with scarce resources or not included in the PLM's pretraining.

Achievement: A first step towards applying the translation skill to the downstream task of WSD by leveraging the cross-lingual knowledge inside the PLM.

Pretraining of LLaMA Language Model from Scratch with multiple GPUs 03/2023 - 06/2023

Course Name: CSE 593 Advanced ML Instructor: Prof. Ludwig Schmidt

Purpose: To develop a multi-GPUs supported pre-training system from scratch using the LLaMA codebase by using the subset of the Pile dataset.

Workload:

1. Established a training pipeline utilizing the Fully Sharded Data Parallelism strategy on Fabric and optimized hyperparameter selection.
2. Performed ablation studies comparing two optimizers, AdamW and Sophia, and evaluated the effects of varying the number of training examples.

Achievement: Successfully trained a small-sized LLaMA model on the Pile dataset, enabling the generation of fluent English sentences.

Investigation on the Factors of In-Context Learning 09/2022 - 12/2022

Course Name: CSE 599 G1 Deep Learning Instructor: Dr. Joseph Redmon

Purpose: To reproduce and validate the statement that input-label mapping, label space, prompt format, and distribution of input text affect the performance differently of the language models in in-context learning, following the same methodology as in the paper (Min et al., 2022).

Workload:

1. Deployed pretrained large language models such as GPT-3, GPT-2, and GPT-J for various text classification tasks using a few-shot in-context learning approach.
2. Analyzed the effects of input-label mapping, the distribution of input text, and label space on in-context learning performance.
3. Conducted experiments with irrelevant and misleading prompt formats during a text classification task to study their impact.

Achievement: Successfully replicated and validated the findings of the original paper, while discovering unexpected results concerning the role of prompt formats in in-context learning.

Skin Bias in Oxygen Saturation Measurements 03/2022 - 09/2022

Position: Research Assistant Advisor: Prof. Shwetak N. Patel at Ubicomp Lab
Purpose: To reduce pulse oximeter inaccuracies for patients with darker skin tones.
Workload:

1. Engineered a skin tone sensor comprising a transmittance sensor, an RGB sensor, a LED reflectance, and a flash, enabling the precise measurement of user skin tones.
2. Constructed a machine learning algorithm pipeline that enhances the traditional ratio of ratio models, designed to calibrate raw SpO2 readings from our conventional Red/IR pulse oximeter, thereby accounting for the variability of skin tones.

Under-ice Ocean Currents Reconstruction with Deep Learning 03/2022 – 09/2022

Position: Research Assistant Advisor: Dr. Georgy Manucharyan at Ocean Dynamics Group
Purpose: To reconstruct under-ice ocean velocity vectors using machine learning
Workload:

1. Processed and converted NASIDC datasets of ice concentration, sea ice velocity, ocean velocity, and atmospheric wind velocity from 1978 to 2021 into Ease Grid 1.0 format.
2. Employed linear regression, support vector machine, and multi-layer perceptron as base models to predict deep ocean activity using interpolated datasets.
3. Developed and implemented a hybrid model combining Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) for regression tasks.

Achievement: Constructed an effective model capable of predicting under-ice ocean velocity vectors.

Exploration of Parameter Dependence in Manifold Learning Algorithms 06/2022 – 09/2022

Position: Research Assistant Advisor: Prof. Marina Meila
Purpose: To investigate the influence of parameters on manifold learning algorithms.
Workload:

1. Developed variable density datasets of Swiss rolls and tori in 3D, derived from the transformation of a 2D rectangle.
2. Analyzed the effect of parameters - the ratio of the number of neighbors to the number of points and the radius of neighbors - on 2D and 3D disks using the T-SNE algorithm.
3. Examined the Erdos-Renyi graph and k-regular graph in relation to parameters such as the radius of neighbors and the number of points.

Achievement: Produced visual representations of the varying effects of different parameters in manifold learning algorithms. Repository: https://github.com/mk322/manifold-learning-examples

Facilitating Conference Paper Reviews through an Interactive UI 03/2022 - 06/2022

Position: Research Assistant Advisor: Prof. Marina Meila
Purpose: To develop an intuitive user interface enabling the sorting and organizing of conference paper reviews.
Workload:

1. Implemented a visual display of reviewed papers using Python's Tkinter, arranged on a canvas as per reviewer evaluations.
2. Leveraged graphical attributes to visually represent the numerical ratings and rankings of the papers.
3. Devised a filtering system to present papers based on specific conditions relating to their ratings and rankings.
4. Constructed a menu to manage child windows, each encapsulating textual review information.

Achievement: Released a PyPi package enabling reviewers to visually assess and compare papers on various rating and ranking criteria.

Replication and Evaluation of the SelfExplain Text Classifier 12/2021 - 03/2022

Course Name: CSE 517 Natural Language Processing Advisor: Prof. Noah Smith
Purpose: To reproduce the SelfExplain model (Rajagopal et al.) and test its interpretability and stability.
Workload:

1. Detailed the architecture of the SelfExplain model, highlighting how it integrates locally interpretable, globally interpretable, and linear layers for prediction.
2. Replicated the study's experiments and outcomes using alternative transformers XLNet and RoBERTa, varied hyperparameters, and the SST-5, TREC-6, SUBJ, and CoLA classification datasets from Stanford NLP and DeepAI platforms.
3. Assessed whether the explanations provided by the two interpretable layers met sufficiency, plausibility, and trustability standards.

Achievement: Effectively reproduced the original paper's results and verified its claims through supplementary experiments utilizing new datasets, hyperparameters, and transformer selections.

Debiasing Deep Learning Recommender Systems 07/2021 - 10/2021

Advisor: Prof. David Woodruff, School of Computer Science, Carnegie Mellon University
Purpose: To mitigate bias in deep learning recommender systems.
Workload:

1. Explored various traditional statistical learning methods in recommender systems, such as Collaborative Filtering (CF), Matrix Factorization (MF), Factorization Machine (FM), Field-aware Factorization Machine (FFM), and Gradient Boosting Decision Tree (GBDT).
2. Evaluated the performance of AutoRec and NeuralCF neural models in the recommender system using the MovieLens dataset, which contains over 1 million data points.
3. Introduced and implemented the Inverse Propensity Score (IPS) and doubly robust algorithms to address bias in recommender systems.

Achievement: Successfully reduced the mean squared errors (MSE) of both AutoRec and NeuralCF models by 8% when employing the aforementioned debiasing methods on a real-world dataset.

Skills & Others

- Reviewer, ACL Student Research Workshop 2023
- Programming Skills: Python (proficient), Java, R, and SQL
- Dean's List: Spring 2023, Winter 2023, Autumn 2022, Spring 2022, Winter 2022, Autumn 2021, Summer 2021, Spring 2021, Winter 2021, Spring 2020
- Core Courses: Deep Learning, Natural Language Processing, Machine Learning, Data Structures and Parallelism, Database System, Linear Optimization, Mathematical Reasoning, Applied Linear Algebra and Numerical Analysis, Discrete Mathematical Modeling, Probability.