Tanmay Parekh | Curriculum Vitae

✉ tparekh@cs.ucla.edu   •   ⌂ tanmayparekh.github.io

Interests

Information Extraction, Event Extraction, Multilingual Technologies, Controlled Generation

Education

University of California Los Angeles   *Sep '21 - Ongoing*
Doctor of Philosophy (Ph.D.) in Computer Science
GPA: 4.0/4.0

Carnegie Mellon University   *Aug '19 - Jul '21*
Masters of Science (MS) in Language Technologies
GPA: 4.02/4.0

Indian Institute of Technology Bombay   *Jul '14 - May '18*
Bachelor of Technology (B.Tech.) with Honors in Computer Science and Engineering
Minor in Applied Statistics and Informatics
GPA: 9.37/10.0

Selected Publications

Contextual Label Projection for Cross-Lingual Structure Extraction:
Tanmay Parekh, I-Hung Hsu, Kuan-Hao Huang, Nanyun Peng, Kai-Wei Chang
In Proceedings of *NAACL 2024* **(Considered for best paper)**   [paper]

Event Detection from Social Media for Epidemic Preparedness:
Tanmay Parekh, ..., Kuan-Hao Huang, Nanyun Peng, Wei Wang, Kai-Wei Chang
In Proceedings of *NAACL 2024*

GENEVA: Benchmarking Generalizability for Event Argument Extraction with Hundreds of Event Types and Argument Roles:
Tanmay Parekh, I-Hung Hsu, Kuan-Hao Huang, Kai-Wei Chang, Nanyun Peng
In Proceedings of *ACL 2023*   [paper]

A Reevaluation of Event Extraction: Past, Present, and Future Challenges:
Kuan-Hao Huang, I-Hung Hsu, Tanmay Parekh, ..., Kai-Wei Chang, Nanyun Peng
Under review at ARR - Feb 2024

Revisiting Evaluation for Code-switched Generation Models: A Multi-Dataset Perspective:
Rebecca Pattichis, Tanmay Parekh, Nicholas L Chavez, Nanyun Peng
Under review at ARR - Feb 2024   [paper]

Politeness Transfer: Tag and Generate Approach:
Aman Madaan*, Amrith Setlur*, Tanmay Parekh*, ..., Shrimai Prabhumoye (* joint authors)
In Proceedings of *ACL 2020*   [paper]

Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues:
Tanmay Parekh, Emily Ahn, Yulia Tsvetkov, Alan W Black
In Proceedings of *CoNLL 2020*   [paper]

Code-switched Language Models Using Dual RNNs and Same-Source Pretraining:
Saurabh Garg*, Tanmay Parekh* and Preethi Jyothi (* joint authors)
In Proceedings of *EMNLP 2018*   [paper]

Selected Research Projects

Diverse and Domain-aware data generation for Information Extraction
Guide: Prof. Kai-Wei Chang & Prof. Nanyun Peng
○ Utilizing LLMs to generate data for information extraction tasks to train fine-tuned downstream models
○ Exploring reference-based and finetuning based methods to adapt the models to various ontologies across four different domains of news, finance, social media, biomedical, and wikipedia.

Contextual Machine Translation for Label Projection for Cross-Lingual Structure Extraction

Guide: Prof. Kai-Wei Chang & Prof. Nanyun Peng

- Utilized instruction-tuned language models to perform contextual machine translation to solve label projection for structure extraction tasks. We used LLama-2 as the instruction-tuned model.
- Improved translate-train for Event Argument Extraction (EAE) and NER by up to 2 F1 points for 39 languages.

Event Extraction from Social Media for Epidemic Preparedness

Guide: Prof. Kai-Wei Chang & Prof. Wei Wang & Prof. Nanyun Peng

- Created the first wide-coverage ontology and Event Detection dataset SPEED for extracting events for epidemic preparedness from social media. We focused on the COVID-19 pandemic and Twitter as the social media platform.
- Demonstrated the generalization of our framework by predicting epidemic events and providing early epidemic warnings for three unforeseen epidemics of Monkeypox, Dengue, and Zika.

Benchmarking Generalizability for Event Argument Extraction

Guide: Prof. Kai-Wei Chang & Prof. Nanyun Peng

- Created a diverse and comprehensive Event Argument Extraction (EAE) ontology with 100+ event types and argument roles using expert human annotations from a semantic role labeling dataset FrameNet.
- Proposed a dataset GENEVA along with four different benchmarking setups to test the generalizability of models. Benchmarked several EAE models from various families on our dataset.

Tag and Generate Approach for Politeness Transfer

Guide: Prof. Alan W Black & Prof. Graham Neubig

- Introduced a new task of politeness transfer providing a large dataset of nearly 1.4 million instances.
- Designed a tag and generate pipeline that identifies stylistic attributes and subsequently generates a sentence in the target style while preserving most of the source content, outperforming many other state-of-the-art methods.

Towards building Code-Switching Chatbots

Guide: Prof. Alan W Black, Prof. Alexander Rudnicky & Prof. Yulia Tsvetkov

- Proposed a generalized goal-oriented multilingual dialogue framework that elicits code-switching and showed its effectiveness by collecting a code-switched dialogue dataset for Hindi-English.
- Experimented with various agent strategies to study user behavior. Discovered various insights about users' code-switching patterns, personal bias and linguistic accommodation.

Industry Experience

Applied Scientist Intern   *Jun '22 - Sep '22*

Alexa Conversations Team, Amazon

- Utilize Wikipedia to generate minimal pairs for relative characterization of negation in language models.
- Developed an alignment-based fine-tuning algorithm for improving the multilingual alignment of negation in multilingual language models. This algorithm provided zero-shot cross-lingual improvements on downstream tasks.

Applied Scientist   *Jul '18 - Jun '19*

Machine Learning Team, Amazon

- Worked on the identification and extraction of product attributes from the titles of product pages without the use of any human-supervised data, but instead using distant supervision to procure data.
- Modelled the problem as an NER task and developed state-of-the-art baselines. Introduced new regularization techniques and semi-supervised self-training-based techniques to learn in the partially labeled data setting.

Other Internships: Goldman Sachs, Philips Research, Edelweiss, Sportz Interactive

Scholastic Achievements and Grants

- Served as the **Program Chair** for the Socal NLP Symposium 2023
- Received the **UCLA Computer Science Fellowship** providing sponsorship for 2021 − 2022
- Recipient of the **ISCA Student Grant** for attending Interspeech '18 selectively awarded internationally
- Achieved **294th rank** among 1.5 million students in the examination of JEE Mains and **581st rank** among 0.15 million students in the examination of JEE Advanced

Selected Courses

- Neural Networks for NLP
- Natural Language Generation
- Large Scale Machine Learning
- Theoretical Statistics
- Convex Optimization
- Reinforcement Learning
- Multimodal Machine Learning
- Automatic Speech Recognition