TANMAY PAREKH

@ tparekh@cs.ucla.edu https://tanmayparekh.github.io/ linkedin.com/in/tanmayparekh/

EDUCATION

University of California Los Angeles (UCLA)

Doctor of Philosophy in Computer Science

2021-Ongoing 4.0/4.0

Carnegie Mellon University (CMU)

Master of Science in Language Technologies

2019-2021 4.02/4.0

Indian Institute of Technology Bombay (IITB)

B. Tech with Honors in Computer Science & Engineering

2014-2018 9.37/10

RESEARCH PROJECTS

CONTEXTUAL LABEL PROJECTION FOR CROSS-LINGUAL STRUCTURE EXTRACTION

- Utilized instruction-tuned language models to perform contextual machine translation to tackle cross-lingual label projection
- Improved translate-train for Event Argument Extraction (EAE) by 2-3 F1 points for Arabic and Chinese.

EVENT EXTRACTION FROM SOCIAL MEDIA FOR EPIDEMIC PREPAREDNESS

- Created the first Event Detection dataset SPEED for extracting events for epidemic preparedness from social media
- Demonstrated the generalization of our framework by providing early epidemic warnings for the unseen epidemic of Monkeypox

BENCHMARKING GENERALIZABILITY FOR EVENT ARGUMENT EXTRACTION

- Created a diverse and comprehensive 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

TOWARDS BUILDING CODE-SWITCHING CHATBOTS

- Proposed a generalized goal-oriented multilingual dialogue framework that elicits code-switching
- Experimented with various agent strategies to study user behavior. Discovered various insights about users' code-switching patterns, personal bias, and linguistic accommodation

STYLE TRANSFER FOR POLITENESS

- Introduced the task of politeness transfer and discussed the pecularities involved in the task
- Proposed a tag and generate approach beating the state-of-the-art techniques in automatic and human evaluation

ACHIEVEMENTS & ROLES

- Received the **UCLA Computer Science Fellowship** providing sponsorship for 2021 – 2022
- Received **sponsorship and grants** ranging upto $150,000 from DSTA and Alexa Socialbot Challenge
- Represented CMU at Alexa Socialbot Challenge 2020 and reached the **Semifinals**
- Received **ISCA Student Grant** at Interspeech '18
- **Teaching Assistant** for 8 undergraduate and graduate courses

SELECTED PUBLICATIONS

- **T Parekh**, IH Hsu, KH Huang, KW Chang, N Peng, "Contextual Label Projection for Cross-Lingual Structure Extraction", in *Arxiv*
- **T Parekh**, et. al., "Event Detection from Social Media for Epidemic Preparedness", submitted at *EMNLP 2023*
- **T Parekh**, IH Hsu, KH Huang, KW Chang, N Peng, "GENEVA: Pushing the Limit of Generalizability for Event Argument Extraction with 100+ Event Types", in Proceedings of *ACL 2023*
- **T Parekh**, E Ahn, Y Tsvetkov, AW Black, "Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues", in Proceedings of *CoNLL 2020*
- A Madaan*, A Setlur*, **T Parekh***, et. al., "Politeness Transfer: Tag and Generate Approach", in Proceedings of *ACL 2020*
- S Garg*, **T Parekh*** and P Jyothi, "Code-switched Language Models Using Dual RNNs and Same-Source Pretraining", in Proceedings of *EMNLP 2018*
- S Garg, **T Parekh** and P Jyothi, "Dual Language Models for Code Mixed Speech Recognition", in Proceedings of *Interspeech 2018*

SKILLS

Programming: Python, C++, Bash, R, MATLAB, Java
Frameworks: Pytorch, Tensorflow

INDUSTRY EXPERIENCE

Applied Scientist Intern — Jun '22 – Sep '22

Amazon (Alexa AI Team)

- Explored the alignment of higher order semantics (like negation) across languages in multilingual models
- Showed the effectiveness of our alignment technique on related downstream tasks like sentiment analysis

Applied Scientist — Jul '18 – Jun '19

Amazon (Machine Learning Team)

- Worked on the problem of product attribute extraction from the product page titles without using human supervised data. Modeled it as a NER task
- Proposed use of semi-supervised learning and regularization techniques to learn from partially labeled data

ACADEMIC PROJECTS

CONTROLLED MULTILINGUAL POETRY GENERATION

Extended an existing English poetry generation model SONG to generate controlled sonnets for the language of Spanish.

ACCENT CONTROL FOR SPEECH SYNTHESIS

Enriched input representations for speech synthesis using heterogenous relation graphs (HRGs) for better control of accent