TANMAY PAREKH

@ tparekh@cs.ucla.edu +1(412)-708-6498 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

BENCHMARKING GENERALIZABILITY FOR EVENT ARGUMENT EXTRACTION

- Introduced a new dataset GENEVA for covering a wide range of event and argument role types and four benchmarking test suites for evaluating the generalizability of EAE models
- Proposed an automated prompting model AutoDEGREE which sets a new benchmark on these test suites

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

TAG AND GENERATE APPROACH FOR POLITENESS TRANSFER

- 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

TARTAN: ALEXA SOCIALBOT CHALLENGE 2020

- Developed and maintained minibots - self-contained domain-specific chatbots - which are integrated together using a global dialog manager
- Prototyped a curious bot, which can ask relevant and interesting questions in an unknown conversational context

LANGUAGE MODELLING FOR CODE-SWITCHED TEXT

- Built Dual Language Models (DLM) training two complementary n-gram LMs and combining them in a probabilistic manner
- Designed a new DNN architecture comprising of a dual LSTM unit combining two monolingual LSTM units and out-performed the standard RNNLM baseline

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 6 undergraduate and graduate courses

PUBLICATIONS

- **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 *arXiV 2022*
- **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*
- A Setlur*, A Madaan*, **T Parekh***, Y Yang, AW Black, "Towards using Heterogeneous Relation Graphs for end-to-end TTS", in Proceedings of *ASRU 2021*
- 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*
- **T Parekh**, S Farfade and N Rasiwasia, "Automatic and Accurate Attribute Extraction for E-Commerce", in Proceedings of *Amazon Machine Learning Conference 2019*

SKILLS

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

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