Tanmay Parekh | Curriculum Vitae

+1 (412) 708 6498 • tparekh@andrew.cmu.edu
https://tanmayparekh.github.io/

Interests

Multilingual and Code-switching Technologies, Controlled and Interpretable Generation

Education

Carnegie Mellon University *Aug '19 - Jul '21*
Masters of Science in Language Technologies
GPA: 3.96/4.0

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

Publications

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

Politeness Transfer: Tag and Generate Approach:
Aman Madaan*, Amrith Setlur*, Tanmay Parekh*, Barnabas Poczos, Graham Neubig, Yiming Yang, Ruslan Salakhutdinov, Alan Black, Shrimai Prabhumoye (* joint authors)
In Proceedings of *ACL 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]

Dual Language Models for Code Mixed Speech Recognition:
Saurabh Garg, Tanmay Parekh and Preethi Jyothi
In Proceedings of *Interspeech, 2018* (Received ISCA Student Grant) [paper]

Automatic and Accurate Attribute Extraction for E-Commerce:
Tanmay Parekh, Sachin Farfade and Nikhil Rasiwasia
In Proceedings of *AMLC, 2019* (Internal Amazon Machine Learning Conference)

Research Projects

Studying Linguistic Accommodation in Code-Switched Dialogues
Guide: Prof. Alan Black & 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 and linguistic accommodation. Provided a comparitive study of Hinglish with Spanglish.

Tag and Generate Approach for Politeness Transfer
Guide: Prof. Alan 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.

Zero-shot Multi-dimensional Cross-Lingual Transfer Learning
Guide: Prof. Graham Neubig & Prof. Alan Black

- Modeled and explored various strategies to model multiple dimensions (e.g. typology, phonology, morphology) in a zero-shot cross lingual setting for the tasks of NER and POS Tagging
- Extensively analyzed the different fusion strategies for a wide diversity of language families

Tartan: Alexa Socialbot Challenge 2020
Guide: Prof. Alexander Rudnicky

- Developed minibots - self-contained domain-specific chatbots - which are integrated together using a dialog manager. This was the core technology of our submission for the challenge.
- Prototyped a curious bot, which can ask relevant and interesting questions in an unknown conversational context

Accent control in Speech Synthesis

Guide: Prof. Alan Black

- Enriched the input representations for speech synthesis using heterogenous relation graphs (HRGs)
- Utilized auxiliary losses for allophone sequence prediction and accent classification to aid better accent transfer

Language Modelling for Code-Switched Text (Undergraduate Thesis)

Guide: Prof. Preethi Jyothi

- Built a robust framework comprising of Dual Language Models (DLM), wherein we train two complementary n-gram language models and combine them in a probabilistic manner.
- Designed a new DNN architecture comprising of a dual LSTM unit which combines two monolingual specialized LSTM units and improves upon the standard RNNLM architecture

Improving Language Models using Cross-Scripted Text

Guide: Prof. Preethi Jyothi (in collaboration with Microsoft Research India)

- Worked on building a robust language model using additional data in the form of transliterated cross-scripted text to counter limitation of native language data

Product Defect Detection from Reviews

Guide: Prof. Pushpak Bhattacharyya (in collaboration with Accenture)

- Constructed a Defect Ontology (Knowledge Graph) using the existing knowledge of the product in order to capture defect-related phrases and words
- Utilized pattern-based matching to find semantic similarities between a new candidate sentence and the ontology.

Industry Experience

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. Used distant supervision to procure data instead
- 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

Internships

Summer Analyst: Goldman Sachs *Summer '17*

Summer Intern: Philips Research *Summer '16*

Winter Intern: Edelweiss Securities Pvt. Ltd. *Winter '15*

Summer Analyst: Sportz Interactive *Summer '15*

Scholastic Achievements and Grants

- Received sponsorship of $23,000 covering tuition and stipend from DSTA
- Received sponsorship of $100,000 for the proposal for **Alexa SocialBot Challenge 2020**
- 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
- Secured **within 100 ranks** multiple times in nation-wide scholarship examination for mathematics conducted by Institute of Promotion of Mathematics (IPM)
- Received **Letter of Appreciation** from the Education Minister for exemplary performance in grade 12

Selected Courses

- Machine Learning
- Convex Optimization
- Probabilistic Models
- Reinforcement Learning
- Regression Analysis
- Neural Networks for NLP
- Sequence-to-Sequence Models
- Multimodal Machine Learning
- Human Languages for AI
- Automatic Speech Recognition

Leadership and Mentorship

Course Leader *Fall '20*

- Coordinated the course of Introduction to NLP for 27 executives

Teaching Assistant

- Multilingual NLP (11-737) CMU - Fall '20
- Speech Processing (11-692) CMU - Fall '19
- Machine Learning (CS 419) IITB - Spring '18
- Computer Architecture and Lab (CS 305, CS 341) IITB - Autumn '17
- Linear Algebra (MA 106) IITB - Autumn '16
- Calculus (MA 105) IITB - Autumn '15

Department Academic Mentor *'17 - '18*

- Selected in the 20-member team for the Academic Mentorship program for Computer Science Department
- Responsible for guiding a group of 7 students with their academics, curriculum and internships

Data Analyst and Journalism *'16 - '17*

- Part of the 6-member team in *Insight, Student Media Body of IIT Bombay* which is responsible for maintaining the data blog - Datagiri covering articles involving data and visualizations

Extra-Curricular Activities

- Won the **1st Place** in a data analysis competition *23rd Yard, Moneyball* conducted at Techfest, Asia's largest science and technology festival among 50+ shortlisted teams
- Part of the soccer teams which won the **1st place** in *Institute and Department Football League*
- Won the **1st and 3rd places** in High Jump and 100m Sprint respectively in the *Athletics Meet '15*

References

Prof. Alan Black
Language Technologies Institute
Carnegie Mellon University
awb[at]cs.cmu.edu

Prof. Yulia Tsvetkov
Language Technologies Institute
Carnegie Mellon University
ytsvetko[at]cs.cmu.edu

Prof. Graham Neubig
Language Technologies Institute
Carnegie Mellon University
gneubig[at]cs.cmu.edu