Pratik Joshi

in linkedin.com/in/pratik-joshi-645305150 @ pratikmjoshi123@gmail.com
📍 Bangalore, India github.com/pratikmjoshi 🔗 pratikmjoshi.github.io

Education

Jun' 2019
Aug' 2015
Birla Institute of Technology and Science, Pilani, Goa, India
B.E. (Honors) Computer Science, CGPA : 8.44/10.00 *(Final year GPA : 9.82/10.00)*

Experience

Present
Aug' 2019
Microsoft Research, Bangalore, India
Research Intern | Advisors : Dr. Kalika Bali, Dr. Monojit Choudhury

- Currently exploring structured learning in natural language inference tasks, and different ways to fuse external knowledge into models. Additionally investigating various logical and linguistic phenomena in NLI datasets to potentially probe the deficiencies in current state-of-the-art pre-trained models.
- Analyzed conversational data and study feedback of 1.5 year-long conducted study at Microsoft exploring the efficacy of multilingual chatbots and appropriate language mixing policies. [CSCW'20]
- Conducted quantitative analysis of language resource disparity and language technology support for large number of languages. Created data model of publication and citation networks, used vanilla metrics as well as a novel entity embeddings method to classify and map out progress of multilingual systems over the years. [ACL'20]
- Analyzed conversational data and study feedback of 1.5 year-long conducted study at Microsoft exploring the efficacy of multilingual chatbots and appropriate language mixing policies. [CSCW'20]
- Quanititatively assessed quality of speech data collected by Project Karya, a crowdsourcing platform for social benefit. Used neural speech alignment pipelines to inspect benefit of using the data for training speech engines in Microsoft products. [LREC'20]
- Carried out a survey of low-resource technologies [ICON'19], conducted user studies and analyzed results for a code-mixed speech annotation tool [AnnoNLP@EMNLP'19].

Dec' 2018
Jul' 2018
Microsoft Research, Bangalore, India
Research Intern | Advisors : Dr. Navin Goyal, Dr. Monojit Choudhury

- Constructed neural and rule-based techniques which show decisively better results than the existing models for parsing natural language to regular expressions. Performed competitively with the state-of-the-art on this problem using multi-task learning across 4 related semantic tasks.
- Devised efficient data collection techniques for potential benchmark semantic parsing dataset for regex.
- Investigated semantic parsing failure points by conducting questionnaires to explore real world data, as well as analyzing existing semantic parsing datasets. Work done as part of undergraduate thesis.

Jul' 2017
May 2017
White Data Systems India (i-Loads), Chennai, India
Software Engineering Intern | Manager : Gokulan Jayaram

- Created voice interface application for truck drivers. Used pocketsphinx-android from CMUSphinx to power voice recognition. Customized to recognize key commands in English, Hindi, and Indian-English.
- Created capabilities like an in-built support system, location and route support, and verbal data entry.
- Presented prototype to I-Loads administration (CEO,CFO,CTO) and tech team.

Publications

- **The State and Fate of Linguistic Diversity and Inclusion in the NLP World**
  2020 Annual Conference of the Association for Computational Linguistics (ACL'20)
  Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, Monojit Choudhury [pdf] [website]
- **Unsung Challenges of Building and Deploying Language Technologies for Low Resource Language Communities**
  2019 International Conference on Natural Language Processing (ICON'19)
  Pratik Joshi, Christain Barnes, Sebastin Santy, Simran Khanuja, Sanket Shah, Anirudh Srinivasan, Satwik Bhattamishra, Sunayana Sitaram, Monojit Choudhury and Kalika Bali [pdf]
- **CoSSAT : Code-Switched Speech Annotation Tool**
  AnnoNLP Workshop : Empirical Methods in Natural Language Processing, 2019 (EMNLP'19)
  Sanket Shah, Pratik Joshi, Sebastin Santy and Sunayana Sitaram [pdf]
- **Do Multilingual Users Prefer Chat-bots that Code-mix ? Let's Nudge and Find Out !**

ACM Conference on Computer Supported Cooperative Work and Social Computing, 2020 (CSCW'20)
Anshul Bawa, Pranav Khadpe, Pratik Joshi, Kalika Bali and Monojit Choudhury

- Crowdsourcing Speech Data for Low-Resource Languages from Low-Income Workers
  12th International Conference on Language Resources and Evaluation, 2020 (LREC'20)
  Basil Abraham, Danish Goel, Divya Siddarth, Kalika Bali, Manu Chopra, Monojit Choudhury, Pratik Joshi, Preethi Jyoti, Sunayana Sitaram and Vivek Seshadri (Alphabetically Ordered) [pdf]

Selected Projects

Semantic Question-Answering System — Jan' 2019 - Jun' 2019

Advisor : Prof. Ashwin Srinivasan

- Created Prolog implementation of the paper "Operation of a Semantic Question-Answering System" (Raphael 1963).
- Implemented numerical reasoning, positional inference, commonsense reasoning by manipulating relational entity graphs.
- Proposed extensions of the system to accommodate a knowledge-graph learner (Hixon et al.). [code]

Prolog | Python

Multi-Document Abstractive Summarization — Jan' 2018 - Jun' 2018

Advisor : Prof. Rajendra Roul

- Investigated different neural architectures for the summarization of the DUC corpus. Studied various encoder-decoder architectures, attention models and pointer networks.
- Attempted to break down the individual components (such as encoder-decoder architecture, attention, time-distributed output) of the various papers by creating implementations for simpler scenarios like machine translation. [code]

Python | Pytorch | Keras

Analysis of Methods in Sentiment Classification — Aug' 2017 - Dec' 2017

ML Nanodegree Final Project

- Used different techniques for sentiment classification of Stanford Large Movie Review Dataset.
- Experimented with different preprocessing methods, implemented various models (LSTM,CNN,SVM) and used different embeddings (GloVe,Word2Vec). [code]

Python | Keras | NLTK | Gensim

Skills

Languages	Python, Prolog, Java, C++, C, MySQL, HTML, QBasic
Frameworks	PyTorch,NLTK,Keras,Tensorflow,Seaborn,Matplotlib,Scikit-learn,Pandas, Numpy,Scipy,Gensim,Open-NMT
Tools	Visual Studio, Git, Eclipse, Android Studio, Powershell
Courses	**Course Rank 1 :** Artificial Intelligence **Class Top 5% :** Machine Learning, Data Mining, Neural Networks, Computer Vision (Reading Course), Operations Research, Theoretical Neuroscience **Others :**Data Structures and Algorithms, Design and Analysis of Algorithms, Object Oriented Programming, Probability and Statistics, Operating Systems, Compilers

Teaching and Mentoring

May 2018
Jan' 2018

Machine Learning (BITS F464) | Birla Institute of Technology and Science Pilani, Goa, India
Teaching Assistant | Course Instructor : Prof. Ashwin Srinivasan

- In charge of the design and supervision of the lab component of course.
- Taught implementation of concepts such as regression, classification, clustering, hyperparameter optimization, ensemble learning. Introduced students to libraries such as Scikit-Learn, Numpy, Pandas, Keras.
- Created assignments and final evaluative project for students, in Python. [course material]

Courses and Other Initiatives

- Udacity ML Nanodegree
  6-month certified course covering topics in supervised learning, unsupervised learning, reinforcement learning, and deep learning.
- General Support Application for Spanish Farmers
  Created app for KiKi, and NGO which provides farmers in Spain with various capabilities, such as push notifications for weather alerts and farming expos in the area, user detail registration, synchronized online and offline detail storage, crop health data collection, and video tutorials. [code]