Harsh Trivedi

Website: http://harshtrivedi.me **Github:** https://github.com/harshtrivedi

Research Interests

Broad: Natural Language Processing, Machine Learning, Deep Learning and Information Retrieval
Specific: Question Answering and Natural Language Inference Systems

Education

M.S in Computer Science, Stony Brook University, US (**GPA: 4.0/4.0**)   Jan'17- Dec'18
Relevant Courses: Natural Language Processing, Machine Learning, Convex Optimization, Probability and Statistics for Data Science, Computer Vision, Computing with Logic
Thesis: Information Aggregation for Question Answering and Natural Language Inference
Advisor: Professor Niranjan Balasubramanian

B. Tech in Information & Communication Technology,   Aug'12-May'16
Dhirubhai Ambani Institute of Information & Communication Technology, India (**GPA: 8.23/10**)
Relevant Courses: Information Retrieval, Data Mining, Data Structures and Algorithms
Final Project: Automatic Pseudo Relevance Label Generation for Learning to Rank
Advisor: Professor Prasenjit Majumder

Publications

"Controlling Information Aggregation for Complex Question Answering" [Short] (pdf)   2018
European Conference on Information Retrieval **(ECIR 2018)**
H. Kwon, <u>H. Trivedi</u>, P. Jansen, M. Surdeanu, N. Balasubramanian
Inference on Knowledge Graphs built with shallow semantic representations like Open-IE triples are prone semantic drift due to which unrelated information pieces get combined resulting in spurious inferences. I developed a supervised version of Personalized Page Rank which could be explicitly trained to avoid this.

"Noise Correction in Pairwise Document Preferences for Learning to Rank" [Short] (pdf)   2016
Asia Information Retrieval Societies **(AIRS 2016)** <u>H. Trivedi</u>, P. Majumder
Proposed a way to automatically correct errors in training data for Learning-to-Rank with which up to 90% of injected errors could be corrected. Received 439$ ACM SIGIR Travel Grant for presenting at AIRS'16, Beijing.

"Author Masking through Translation" [Workshop] (pdf)   2016
Conference and Labs of the Evaluation Forum **(CLEF 2016)** Y. Keswani, <u>H. Trivedi</u>, P. Mehta, P. Majumder
Trained translation models using europarl parallel corpora with Moses & leveraged it for obfuscating stylo-metric identity of author. Stood 2nd in Author Masking Task of CLEF": PAN-2016.

"A New Approach to Syllabification of Words in Gujarati" [Long] (pdf)   2015
Mining Information & Knowledge Exploration (MIKE 2015) <u>H. Trivedi</u>, A. Patel, P. Majumder
Developed the first large corpus for syllabification of Gujarati language. Built a relatively simple probabilistic model for syllabification of words and combined with CRF based approach to achieve 98.02% syllabic accuracy.

Work Experience

Research Assistant at NLP Lab, Stony Brook University   Jan'17-now
(1) Developed a novel method to convert pre-trained entailment models into multi-sentence QA models.
(2) Developed supervised Personalized Page Rank for solving QA with Knowledge Graphs.
(3) Constructed entailment data-set to evaluate textual entailment models with varying sizes of premise text.

Undergraduate Student Researcher at IRLab - DAIICT, India   May'15-Jun'16
(1) Developed noise correction method for pair-wise Learning-to-Rank (LeToR) algorithms.
(2) Developed feature consensus based method for using pair-wise LeToR in unsupervised fashion.
(3) Developed a new approach Gujarati Syllabification.

Full Stack Developer at Areysun, India   Dec'13-July'14
Built web APIs for Android/Web frontend, built analytics and content generation dashboard for clients and promoters respectively, implemented efficient ephemeral posts using Redis pubsub mechanism. Tools: Ruby on Rails, Javascript, MonogDB, Redis, HTML, CSS, Nginx, AWS

Relevant Projects

Coin betting optimizers for parameter free optimization (link): Coin betting algorithms do not require any learning rates and yet have optimal convergence guarantees for non-smooth convex functions. I implemented two versions of these algorithms in pytorch and did large scale study comparing these to standard optimizers used for deep learning based problems.

Automatic Pseudo Relevance Label Generation for Learning to Rank (link): Learning-to-Rank (LeToR) is a information retrieval paradigm in which instead of ranking documents based on one relevance score, a ranking function is learned from extracted features. I developed a feature consensus method for automatic relevance label generation, training form which LeToR gave better results than training from gold labels.

Visual Question Answering (link): Implemented Neural Module Networks (NMN) for visual question answering in pytorch. It constructs a question specific computation graph from a set of neural modules based on the semantics of the question and answers question about the image. This was the first implementation of NMN in pytorch.

Other Papers Implementations: (1) "Syntax-based alignment of multiple translations: Extracting paraphrases and generating new sentences" (Demo). (2) "Diversification based Feature Selection Methods for Learning to Rank" (link) (3) "Relevance based language models"

Profile Matching for E-Hormony (link): Implemented various Machine Learning algorithms in Matlab from scratch: kernelized ridge regression with SGD, kernelized perceptron, ranksvm, cost-specified regressions, random forest, adaboost to compete for in-class machine learning competition.

AllenNLP (link): Intermittent contributor of AllenNlp, a framework build on pytorch for NLP research.

Awards and Achievements

- Topped Natural Language Processing and Convex Optimization courses at Stony Brook University.
- Stood 1st in In-Class kaggle competition hosted in Machine Learning course at Stony Brook University.
- Ranked 6186 among 1.2M students (top 0.5%), All India Engineering Entrance Examination (AIEEE), 2012.
- 1st prize in National as well as State competition of Mental Arithmetic (ALOHA) [2009].
- 3rd prize in National drawing competition held by Department of Post, India [2008].
- 1st prize in drawing competition held by Forest Department, Gujarat. [2007].

Other Activities

- Volunteered at Tribhuvandas Foundation (NGO by AMUL, India) for spreading cancer awareness, Dec 2013.
- Developed FIRE 2015/16 conference website (link) with content management system for easy maintainability.
- Organized tech meetups and events as a member of Google Developer Group - DAIICT for 2 years, 2013-14.

Technical Skills

Languages:	Python, Ruby, MATLAB, Javascript and HTML
Tools & Technologies:	Pytorch, AllenNlp, Tensorflow, Python NLTK, OpenCv, Panda and Scikit-learn, Weka Ruby on Rails, Flask, MongoDB, Redis, jQuery, AJAX, Git, AWS S3 and EC2

References

- Professor Niranjan Balasubramanian, Stony Brook University — niranjan@cs.stonybrook.edu
- Professor Francesco Orabona, Boston University — francesco@orabona.com
- Professor Ashish Sabharwal, Allen Institute of Artificial Intelligence — ashishs@allenai.org
- Professor Prasenjit Majumder, DA-IICT — p_majumder@daiict.ac.in