Bhargavi Paranjape

Andrew ID: bvp

Personal Information

E-mail: paranjape.bhargavi@gmail.com, bvp@andrew.cmu.edu
Website: bhargaviparanjape.github.io
Mobile: +91 9800165653

Research Interests

Natural Language Processing, Text & Information Retrieval, Machine Learning, Sociolinguistics, Data Mining

Education

Indian Institute of Technology, Kharagpur, India — July 2012 - May 2016
Bachelors of Technology (Honors) in Computer Science and Engineering
CGPA: 9.54/10

Work Experience

Microsoft Research Lab, Bangalore, India — July 2016-July 2017
Research Fellow
Machine Learning & Natural Language Systems Group — **Mentor**: Raghavendra Udupa

Publications

Chakraborty, A., **Paranjape, B.**, Kakarla, S., Ganguly, N. *"Stop Clickbait: Detecting and Preventing Clickbaits in Online News Media"*
Conference: **ASONAM 2016**
Nominated for Best Paper and awarded **Best Student Paper**

Gupta, V., Mekala, D., **Paranjape, B.**, Karnick, H. *"SCDV : Sparse Composite Document Vectors using soft clustering over distributional representations"*
Conference: **EMNLP 2017** (Full Paper)

Gupta, C., Suggala, A., Gupta, A., Kumar, A., **Paranjape, B.**, Simhadri, H. *"ProtoNN: Compressed and Accurate kNN for Resource-scarce Devices"*
Conference: **ICML 2017** (Full Paper)

Research Experience

Linguistic study of sensationalist headlines used by online news media
Bachelors Thesis Project, IIT Kharagpur — July 2015-May 2016
Guides: Prof. Niloy Ganguly, Abhijnan Chakraborty

- Used semantic and lexical contrasts between ordinary and sensationalist news headlines to create hand-crafted features.
- Achieved a classification accuracy of 93% using SVM and used the model to build a browser extension, *StopClickbait* that detects clickbaity links.
- Augmented the classifier with semantic network generation and pattern matching algorithms that use click history and tune the clickbait classifier for a particular user.

Structured Learning for Pseudo Relevance Feedback(PRF)
Research Project, Microsoft Research India — November 2016-July 2017
Guide: Raghavendra Udupa

- Used Associative Markov Networks(AMNs) for supervised learning of an optimal set of feedback terms for search query expansion. Used rich set of syntactic and semantic features to model term(node) and term-pair(edge) relevance to the query. This is a word-level expansion model.
- Trained a Recurrent CNN based encoder and generator neural model, where the encoder learns to predict query-document relevance score and generator specifies a distribution over the document as candidate rationale to use the document for PRF.

Learning multilingual word embeddings using translation word-alignment
Research Project, Microsoft Research India — March 2017-Present
Guide: Raghavendra Udupa

- We aim to learn transformations that project Word2Vec embeddings for a pair of laguages onto a common embedding space.
- Using word alignments obtained from IBM statistical MT models, we train a model inspired by supervised semantic indexing (SSI) algorithm on samples of positive and negative word translations pairs.

- Evaluating performance on Reuters cross-lingual document classification and word translations tasks.

Task-specific Short Document Expansion
Research Project, Microsoft Research India July-November 2016
Guides: Raghavendra Udupa, Ram Bairi
- Short texts like tweets are expanded for clustering and classification tasks, but they use domain or task specific heuristics and IR systems.
- Developed a robust system for short text expansion that can be used across varied tasks, yet learns to produce expansions that improves a specific task's performance.
- Expansion model uses task-agnostic language features and is trained by SGD to alternately optimize for selecting relevant long document and reducing task loss.

Resource optimized Machine Learning for Internet of Things(IoT)
Research Project, Microsoft Research India July-November 2016
Guides: Raghavendra Udupa, Dr. Prateek Jain
- Edge or fog computing on resource constrained end devices in IoT necessitates compressed yet performant versions of machine learning algorithms like K-Nearest Neighbors.
- Used K-nearest Neighbor models with fewer, low-dimensional class representatives as neighbors and profiled prediction time on Arduino UNO and BBC Microbit boards using these models.
- Optimized prediction on such resource constrained devices by using a CSR-like sparsity representation and learning a lookup table that rapidly computes the exponential kernel function.

Inducing Tag Ontology on StackOverflow
Research Project, IIT Kharagpur July 2015-December 2015
Guide: Prof. Pawan Goyal
- Performed Gaussian Mixture Model clustering of 110,454 user generated tags on StackOverflow to obtain overlapping clusters of 8-10 tags.
- Converted each tag-cluster into a rooted tree based on statistical co-occurrence and entropy of member tags in a corpus of 28 million questions.
- Performed bigram topic modeling over questions tagged with both ends of a tree edge to induce a label set for that tree edge.

Internships

Summer Internship, IIT Bombay *May-July 2014*
Data Acquisition System, Project Ekalavya Prof. D. B. Phatak
- Developed an affordable plug-n-play Sensor Data Acquisition System (DAQ) for science laboratories in public schools and colleges in India
- DAQ uses the 5$ BeagleBone Black that extracts data from attached sensors and displays processed readings on the low-cost Aakash Tablet in an interactive and processable fashion

Summer Internship, Microsoft India Development Center *May-July 2015*
Cloud & Enterprise Division
- Extended the Service Level Agreement (SLA) feature of Microsoft Dynamics CRM 2015 Software
- Offered full time job opportunity at the end of internship.

Awards and Achievements

Awarded the *Kishore Vaigyanik Protsahan Yojana (KVPY) National Fellowship* granted to students with motivation to pursue research careers in Science *2011*
Awarded the Technology Alumni Association Prize for securing highest GPA in the batch among 1,332 students in Freshman Year *2013*
Won the IBM Day Prize for the Browser Extension, *StopClickbait*, that detects sensationalized headlines using machine learning *2016*

Technical Skills

Programming C++,C#, Java, Python
Web Development CSS, JavaScript, Node.js, Django
Packages Apache Nutch, Matlab, OpenCV, Theano, TensorFlow, CNTK

Relevant courses	**Curriculum:** Machine Learning, Natural Language Processing, Information Retrieval, Artificial Intelligence, Social Computing, Probability & Statistics, Linear Algebra **Online:** Deep Leaning(Udacity), Mining Massive Datasets(Coursera), Deep Learning for NLP (Stanford)
Extra curricular activities	• As member of Technology Robotix Society, the Robotics club of IIT Kharagpur, helped organized and participated in the intra-college Robosoccer competition. • As reporter for Scholar's Avenue, the student run fortnightly of IIT Kharagpur, covered several academic and campus issues, including student body elections.