Vivek Gupta

Machine Learning & Natural Language Processing

Contact Information

University of Utah
543 South 900 East, Aprt. C6, Salt Lake City
Utah, United States

http://vgupta123.github.io
keviv9@gmail.com
+1-801-558-7272

Research Interests

Representation Learning, Structured Prediction, Extreme Classification, Resource Constraint Machine Learning, Semi Supervised Learning and Anomaly detection, Ethics in Machine Learning

Current Position

School of Computing, University of Utah (Aug 2018 - Present)
Graduate Research Assistant

Education

University of Utah (Aug 2018 - Present)
PhD, Computer Science, School of Computing

Indian Institute of Technology Kanpur (July 2015 - May 2016)
M. Tech with Honors in Computer Science and Engineering
CGPA: 9.3/10

Indian Institute of Technology Kanpur (July 2011 - July 2015)
B. Tech with Honors in Computer Science and Engineering
CGPA: 7.5/10

Employment

Microsoft Research Lab, India (Oct 2016 - Aug 2018)
Research Fellow, Machine Learning and Natural Language Application Advisors: Dr. Praneeth Netrapalli & Dr. Prateek Jain

Scholastic Achievements

Selected among the top 7 finalist in **Ericson's Innovation Awards, India** in 2016

Secured **All India Rank 183** in **IIT-JEE 2011** amongst more than 500,000 candidates

Qualified for **Kishore Vaigyanik Protsahan Yojana (KVPY) scholarship, 2011**

Selected among top 1% of students in India, **Indian National Physics Olympiads, 2011**

Selected among top 1% of students in India, **Indian National Chemistry Olympiads, 2011**

Publications

Dohare, S., **Gupta, V.**, Karnick, H. *"Unsupervised Semantic Abstractive Summarization"* , **ACL-SRW 2018** (Workshop) **[Paper]**[**Poster**]

Mekala, D.*,**Gupta, V.***, Paranjape, B., Karnick, H. *"Sparse Composite Document Vectors using soft clustering over distributional representations"* , **EMNLP 2017** (Long Oral) **[Paper]** [**PPT**]

Gupta, V., Karnick, H.,Bansal, A., Jhala, P. *"Product Classification in E-Commerce using Distributional Semantics"* , **COLING 2016** (Long Poster) **[Paper]**[**Poster**]

Wadbude, R.,**Gupta, V.**, Mekala, D., Karnick, H., *"User Bias Removal in Review Score Prediction"*, Accepted in **CODS-COMAD 2018** (Long Oral) & **DAB 2017** (Workshop) **[Paper]**[**Poster**]

Gupta, V., Mittal, S.,Bhaumik, S., Roy, R. *"Assisting Humans to Achieve Optimal Sleep by Changing Ambient Temperature"*, **BIBM 2016**, **BHI 2016** & **HI-DS 2016** (Oral) **[Paper]** [**PPT**]

Working Manuscripts

Gupta, V.*, Wadbude, R.*, Rai, P., Natararjan, N.,Karnick, H., Jain, P., *"Distributional Semantics meet Multi-Label Learning"* **[PrePrint]**

Gupta, V.*., Saw, A.*, Talukdar, P., Netrapalli, P. *"Unsupervised Document Representation using Partition Word-Vectors Averaging"*

Gupta, V.*., Saw, A.*, Gupta, H.*, Talukdar, P., Netrapalli, P. *"Revisiting Composite Document Vector From Polysemic Perspective"*

Raunak, V., **Gupta, V.** *"Simple and Effective Dimensionality Reduction for Word Embeddings"*

Master Thesis

Product Categorization in E-Commerce using Distributional Semantics
Advisors: Prof. Harish Karnick (IIT Kanpur) & Pradhuman Jhala (Flipkart.com) [**Thesis**][**PPT**]

- Proposed a novel distributional semantics representation for text documents
- Proposed a two-level ensemble approach utilizing (with respect to the taxonomy tree) a path-wise, node-wise and depth-wise classifiers for product classification

Research Projects

ProtoSSL: Resource Constraint Anomaly Detection (June 2016 - Present)
Advisors: Dr. Prateek Jain(Microsoft Research India), Dr. Nagarajan Natarajan (Microsoft Research India) and Dr. Praneeth Netrapalli (Microsoft Research India)

- Working on a resource constrained machine learning algorithm for semi-supervised learning
- Working on predictive maintenance, investigating time series features, novel evaluation metrics for anomaly detection

Leveraging Distributional Semantics for Multi-Label Learning (June 2016 - May 2017)
Advisors: Dr. Nagarajan Natarajan (Microsoft Research India) Dr. Piyush Rai (IIT Kanpur), Dr. Harish Karnick (IIT Kanpur), Dr. Prateek Jain (Microsoft Research India)

- Design extreme multi-label learning algorithms using distributional semantics with more efficient training procedures based on joint learning
- Extended the approach to naturally incorporate other sources of side-information, particularly the label-label co-occurrence matrix

Text Summarization using Abstract Meaning Representation (June 2016 - May 2017)
Advisor: Dr. Harish Karnick (IIT Kanpur)

- Explored a novel full-fledged pipeline for text summarization with an intermediate step of Abstract Meaning Representation (AMR)
- Proposed method achieves state-of the-art results compared to the other text summarization routines based on AMR

Text Categorization using Sparse Composite Document Vector (June 2016 - May 2017)
Advisor: Dr. Harish Karnick (IIT Kanpur)

- Proposed a novel topic-based document representation which outperforms state-of-the-art models in multi-class and multi-label classification tasks
- Embeddings provide a robust estimation of the query and document language models, thus improving the MAP of language model based retrieval systems

Optimal Hierarchical Classification (June 2016 - Present)
Advisors: Prof. Purushottam Kar (IIT Kanpur) & Prof. Harish Karnick (IIT Kanpur)

- Extended the consistency of hierarchical classification algorithm on asymmetric tree distance loss using calibrated surrogates
- Showed under reasonable assumption over hierarchy that the Bayes optimal classification for this asymmetric loss can be found in $O(log(n))$

Introducing Diversity Priors in Latent Variable Model (Jan 2016 - Apr 2016)
Advisor: Prof. Piyush Rai (IIT Kanpur), Report Link

- Worked on diversity priors regularizes for latent variable models like probabilistic PCA and non-negative matrix factorization
- Introduced log-det divergence, determinant point process and pair-wise Mutual angular regularizes in standard probabilistic models

Learning Bilingual Word Embeddings (Mar 2017 - Present)
Advisor: Dr. Raghavendra Udupa (Microsoft Research India)

- Working on learning transformations that project Word2Vec embeddings for a pair of languages onto a common embedding space

- Used word alignments obtained from IBM statistical MT models to train a model inspired by supervised semantic indexing (SSI) algorithm on samples of positive and negative word translations pairs

Learning Interpretable Word Embeddings (Mar 2017 - Present)
Advisor: Dr. Sundararajan Sellamanickam (Microsoft Research India)

- Working on learning interpretable word embedding by non negative and sparse factorization of PPMI matrix
- Evaluating embedding performance on Google Similarity and Analogy task and embedding interpretability on human intrusion task

Research Internship

Microsoft Research India, Bangalore (May 2016 - Jul 2016)
Estimation of generalization error for ensembles *Dr. Sundararajan Sellamanickam (Principal Applied Scientist)*

- Worked on efficient estimation of generalization error for ensembles using normality assumption on classification scores
- Worked on efficient prediction of ensemble parameters, bias and variance of generalization errors using minimal number of ensembles

Flipkart Internet Pvt. Ltd., Bangalore (May 2015 - July 2015)
Web Scale Product Classification *Pradhuman Jhala (Principal Architect)*

- Developed a model for deep hierarchal classification using multilevel classifiers and ensemble methods
- Model outperformed existing naive bayes classifier in precision@1, precision@1 and prediction time.

Samsung R&D Institute, Bangalore (May 2014 - July 2014)
Mobile and Healthcare Solution Y2014 *Sandip Bhaumik (Group Manager) & Raj Roy (Manager)*

- An innovation in S-Health Wearable technology on sleep applications by automatically changing ambient temperature using recurrent feedback signal from S-Watch (Patented & Published)

Professional Service

Program Committee: Serving on Program Committee of the ACL 2018 Student Research Workshop (SRW).

Meta Reviewer: Served as a reviewer for long and short papers for Empirical Methods in Natural Language Processing 2017

Coordinator: Initiated and managed Special Interest Group in Machine Learning at Computer Science and Engineering Department, IIT Kanpur. Organized regular meetups for discussions and talks on topics in Machine Learning and related fields.

Teaching Assistant: for MLT 2016 - Machine Learning Tool and Technique: Mentored a group of 30 M-Tech students part of an introductory course on Machine Learning. Set up a labeling software for project work, resulting in a new dataset.

Student Secretary & **Student Mentor**: in Promotion of Work Experience and Research PoWER & Alumni Contact Program (ACA) under Office of Dean of Research and Development IIT Kanpur

Student Volunteer: for mentoring and teaching underprivileged students of primary classes from nearby village(Nankari) at Prayas, IIT Kanpur

Talks & Seminars

I had the opportunity to present my work at various places which I thoroughly enjoyed. [**Talks**]

Student Mentorship

I have been extremely lucky to mentor some amazing students. [**Students Mentored**]

Travel Grants & Scholarship

I have been fortunate to receive scholarships/grants at several occasions to support my education and research. [**Travel Grants**]