Vivek Gupta

Ph.D Applicant in Machine Learning

Contact Information

Microsoft Research India
#9, Lavelle Road
Bangalore, India - 560001

website
keviv9@gmail.com
+91-8765696260

Research Interests

Distributional Semantics, Extreme Classification, Resource Constraint Machine Learning

Current Position

Microsoft Research India (June 2016 - Present)
Research Fellow Advisors: Dr. Praneeth Netrapalli & Dr. Prateek Jain

Education

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

Accepted Publications

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

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

Under Review Publications

Wadbude, R.,**Gupta, V.**, Mekala, D., Karnick, H. *"User Bias Removal in Review Score Prediction"*, Under review in **DAB 2017** (Workshop Paper) **[ArXiv]**[**Poster**]

Dohare, S., Karnick, H., **Gupta, V.**, *"Exploring a full-fledged pipeline for Text Summarization using Abstract Meaning Representation"* Under review in **IJCNLP 2017** (Long Paper) **[ArXiv]**

Wadbude, R., **Gupta, V.**, Rai, P., Natararjan, N.,Karnick, H. *"Leveraging Distributional Semantics for Multi-Label Learning"*, Under review in **NIPS 2017**

Mahajan, D., **Gupta, V.**, Keerthi, S.,Sundararjan, S. *"Efficient Estimation of Generalization Error and Bias-Variance Components of Ensembles"*, Under review in **ICDM 2017** (Long Paper)

Master Thesis

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

- Proposed a new distributional semantics representation for document which outperform other representation methods
- Proposed a novel two-level ensemble approach utilizing (with respect to the taxonomy tree) a path-wise, node-wise and depth-wise classifiers for product classification
- Achieve state of art results on various standard and novel proposed evaluation metrics

Research Experience

Resource Constraint Anomaly Detection (June 2016 - Present)
Advisor: Dr. Prateek Jain and Dr. Praneeth Netrapalli (Microsoft Research India)

- Working on extension of ProtoNN (Resource Constraint Machine Learning) for semi-supervised setting
- Working with company *magma international* on predictive maintenance, investigating time series features, novel evaluation metrics for anomaly detection

Leveraging Distributional Semantics for Multi-Label Learning (June 2016 - May 2017)
Advisor: Dr. Nagararjan Natararjan (Microsoft Research India) Dr. Piyush Rai (IIT Kanpur), Dr. Harish Karnick (IIT Kanpur)

- Design extreme multi-label learning algorithms using distributional semantics with more efficient training procedures
- Extended the approach to naturally incorporate other sources of side-information, in particular 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
- Pointed out some significant problems in the existing evaluation methods making them unsuitable for evaluating summary quality

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
- Showed that fuzzy GMM clustering on word-vectors lead to more coherent topic than LDA and can be used to detect Polysemic words
- Embeddings provide a robust estimation of the query and document language models, thus improving the MAP of language model based retrieval systems

Bayes Optimal Classication for Hierarchy (June 2016 - Present)
Advisor: Dr. 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 $O(log(n))$

Generalization vs Optimization for Deep Learning (June 2016 - Present)
Advisor: Dr. Praneeth Netrapalli, Microsoft Research India

- Working on more efficient techniques (tail-averaging) for non-convex optimization for training deep neural nets
- Investigating the broad question of whether depth of neural nets help in generalization or optimization

Learning bilingual Word Embeddings (Mar 2017 - Present)
Advisor: Raghavendra Udupa, Microsoft Research India

- Working on learning transformations that project Word2Vec embeddings for a pair of languages onto a common embedding space
- We 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
- Evaluating performance on Reuters cross-lingual document classification and word translations tasks

Learning Interpretable Word Embeddings (Mar 2017 - Present)
Advisor: 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 intrepretable on human intrusion task
- Working on use of these embedding for detecting polysemous words and explainable models.

RESEARCH INTERNSHIP

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

- Worked on efficient Estimation of generalization error for ensembles using normality assumption on classification scores
- Worked on efficient prediction of accuracy, 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*

- Developed a machine learned model for deep hierarchal classification of flipkart catalog hierarchy using multilevel classifiers and ensembles methods
- Model outperformed there naive based classifier with 89% with limited features for 900 classes with faster prediction time of 100x

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

- Did an innovation in S-Health Wearable technology on sleep applications by automatically change ambient temperature by using recurrent feedback signal from S-Watch

DEVELOPMENT INTERNSHIP

Flipkart Internet Pvt. Ltd., Bangalore (Dec 2015 - Jan 2015)
Extracting Attribute Value Pairs *Pradhuman Jhala*

- Developed a machine learned multilabel multiclass model for extraction of attribute values from product apparel images
- Successfully extracted attribute values collar ,crew & vneck ,full sleeve & half sleeve for shirts {pointy, non-pointy} for sandals,{baseball, basketball, bracket & ballet} for sport shoes with pair wise accuracy of 74%

Synopsys. Inc., Bangalore (May 2013 - July 2013)
Trend Generation for Dali Coverage *Yogesh Pandey*

- Created interactive visualization of Scenario coverage trend for functional verification project *Dali* using data visualizations technique
- Coded Object Oriented Filter pattern which parse json file and perform analysis on selected properties

SCHOLASTIC ACHIEVEMENTS

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

Selected in top 100 teams all over India in **Internet for all challenge** by Unisys in 2015

Best poster presentation for Internship works at Flipkart Internet Private Ltd. in 2015

Best project award for project *Distributed Classroom* in Topic in Distributed System in 2015

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

Qualified for **Kishore Vaigyanik Protsahan Yojana (KVPY) Scholarship** in 2011

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

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

Selected among the top 1% of students in State, **Indian National Physics Olympiads** in 2010

Travel Grants

Student Volunteer for *EMNLP 2017* Conference in Copenhagen, Denmark

Scholarship for attending *EMNLP 2017* by Microsoft Research India, Bangalore.

Scholarship for attending *COLING 2016* by Flipkart Internet Pvt. Ltd., India.

Scholarship for attending *Workshop on Brain, Computation and Learning, IISC, Bangalore*

Scholarship for attending *IFCAM NMI Summer School, IISC, Bangalore*

Talks & Seminars

Text Categorization using Sparse Composite Document Vectors (June 2017)
CFILT Group, IIT Bombay [**PPT**]

Research as a Student (June 2017)
Think Research Club, VIT, Bombay (Invited Talk) [**PPT**]

Text Classification using Sparse Composite Document Vectors (March 2017)
YAML, Microsoft Research India, Bangalore [**PPT**]

Natural Language Processing in E-Commerce, A Case Study (Feb 2017)
Botathon Event Forge Accelerator, Coimbatore [**PPT**]

Speeding up Stochastic Gradient (Nov 2016)
RF Talk, Microsoft Research India, Bangalore [**PPT**]

Product Classification using Distributional Semantics (August 2016)
Microsoft Research India, Bangalore and IBM Research Lab, India [**PPT**]

Ocular, Vision for Visually Impaired (May 2016)
Erricson Innovation Award, IIT Delhi [**PPT**] [**PIC**]

Parametric R Tree, Indexing of Moving Object (Dec 2015)
Indexing and Searching in Large Databases Seminar [**PPT**]

Deep Classification in Large Scale Web Hierarchy (August 2015)
MLRD, SIGML IIT Kanpur and Flipkart Internet Pvt. Ltd. [**PPT**]

Professional Service

External Reviewer for EMNLP 2017: 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. Organize 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 in Promotion of Work Experience and Research PoWER, Office of Dean of Research and Development IIT Kanpur

Student Mentor in Alumni Contact Program (ACA) under Office of Dean of Research and Development IIT Kanpur

Student Mentorship

Dheeraj Mekala, IIT Kanpur (Jointly with Prof. Harish Karnick and Prof. Puruhottam Kar) on Bayes Optimal Classifier for Hierarchal Classification

Rahula Wadbude, IIT Kanpur (Jointly with Prof. Harish Karnick, Nagararjan Natarajan, Prof. Piyush Rai) on Leveraging Distributional Semantics for Multi-Label Learning (under review NIPS)

Dheeraj Mekala, IIT Kanpur (Jointly with Prof. Harish Karnick) on Sparse Composite Document Vectors (EMNLP 2017) **[Report]**

Rishabh Gupta and Sanjari Srivasatva, IIT Kanpur (Jointly with Prof. Harish Karnick) on Topic Modeling using Independent Component Analysis **[Report]**

Rahul Wadbude and Dheeraj Mekala, IIT Kanpur (Jointly with Prof. Harish Karnick) on User Bias removal in Review Score Prediction (Submitted DAB 2017) **[Report]**

Avikalp Srivastava, IIT Kharagpur (Jointly with Dr. Sundararajan Sellamanickam) on Interpretable and Efficient Word and Document embedding

Shibhansh Dohare, IIT Kanpur (Jointly with Prof. Harish Karnick) on Text Summarization using Abstract Meaning Representation **[Report]**

Other Projects

S-Voice Commerce Dialog Learning (Jan 2016 - Apr 2016)
Advisor: Prof. Harish Karnick (IIT Kanpur) & Praveen Reddy (Samsung Research)

- Worked on dialog state tracking system and problem of attribute value extraction for Samsung S-Voice Commerce
- Referring to literature for state space models (DSTC Challenge), dependency based wordvec and sequence learning DNN models (LSTM, RNN)

Independence & Randomization in Deep Neural Net (Jan 2016 - Apr 2016)
Advisor: Prof. Harish Karnick(IIT Kanpur) & Gaurush Hiranandani (Adobe Research)

- Worked on theoretic proof of *Improved Classification and Reconstruction using Independence and Randomization* in Deep Neural Networks DICTA 2015
- Evaluated proposed non-linear dimensionality reduction techniques by appropriately scoring topology preservation and information gain perspective

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

- Worked on Diversity Priors Regularizes for Latent Variable Models like Probabilistic PCA and Non-Negative Matrix Factorization
- Introducing log-det divergence, determinantal point process and pair-wise Mutual angular regularizes in standard probabilistic models

Large Scale Multi Kernel Learning (July 2015 - Dec 2015)
Advisor: Prof. Harish Karnick(IIT Kanpur) **[Report][PPT]**

- Showed that Multi Kernel Learning work better than single kernel independent data separability and distribution in space
- Extended toolbox for multi-modal data using different kernels for data attributes from different heterogeneous sources

Forest Cover Type Prediction (Dec 2014 - Apr 2015)
Advisor: Prof. Harish Karnick(IIT Kanpur) **[Report][PPT]**

- Predict forest type of forest from 7 classes using 54 attributes (kaggle data) and was ranked 42 with 82% accuracy on kaggle.
- Used classification algorithms like Random Forests, Extra-Trees Classifier, Auto-encoders, Gradient Boosting Model, K -Nearest Neighbor, Label propagation, Support Vector Machine etc.

Location Based Approximate Query (Dec 2014 - Apr 2015)
Advisor: Prof. Arnab Bhattacharya (IIT Kanpur) **[Report]**

- Modified a Filter Effective Hybrid tree (FEH) for indexing approx geo-location queries using filter tree based on distance metric like Euclidean & Manhattan
- Result out perform in search time by significant margin for R* tree using line extra space. Also developed a Graphical User Interface using Google Map API for demo

Distributed Classroom (PANGO) (Aug 2014 - Dec 2014)
Advisor: Prof. Ratan K Ghosh(IIT Kanpur) **[Report][PPT][Video][JED-I]**

- Developed a server-less Distributed Application for File Sharing, Digital Whiteboard, Online Tests, Live Lecture Slideshare,Video and Document Tagging and automatic Lecture Scribing
- Implemented various distributed protocols like token based leader election, mutual exclusion joining, deadlock prevention, novel distributed file sharing and non-buffering syncing protocol

Image Search Engine-PicBook (Aug 2014 - Dec 2014)
Advisor: Prof. TV. Prabhakar(IIT Kanpur) **[Report][PPT][Video]**

- Developed a collaborative social Image search engine (PICBOOK) using Neo4j graph database which capture collaborative clicks
- Software support automating tagging, re-ranking, graph search queries like *Photos of my friend friends, my photos, female friend photos*, text search, uploads, friends addition

Image Retrieval & Re-Ranking (Aug 2014 - Nov 2014)
Advisor: Prof. Harish Karnick **[Report]**

- Work on a new approach to Image retrieval and ranking using latent semantic indexing and cross view learning. Project was co-mentored by Yahoo India R&D lab
- Learned a latent space for single representation of cross views (text-image) from Latent Semantic Analysis features for text semantic & SIFT, HOG and RGB for image semantics

Extra Curriculars

Active Volunteer for teaching underprivileged student at Prayas IIT Kanpur for 2 years

C Rank National Cadet Corps is the Indian military cadet corps.

Attended Microsoft Research India Summer Workshop on Artificial Social Intelligence 2017

Attended Microsoft Research India Summer School on Internet of Things 2016