

Himanshu Jain

Senior PhD Student

Indian Institute of Technology Delhi

email id: hjain@cse.iitd.ac.in

Website: www.cse.iitd.ernet.in/~hjain/

Phone: +918588834589

Education

2014 – Present	PhD Computer Science and Engineering Indian Institute of Technology Delhi Advisor: Prof. Manik Varma	CGPA: 8.6/10.0
2007 – 2012	Bachelor and Master of Technology (Dual Degree) Mechanical Engineering Indian Institute of Technology Kanpur M.Tech. Advisor: Prof. Kalyanmoy Deb	CGPA: B.Tech. - 8.9/10 M.Tech. - 9.7/10

Publications

- H. Jain, V. Balasubramanian, B. Chunduri and M. Varma. **Slice: Scalable linear extreme classifiers trained on 100 million labels for related searches.** In Proceedings of the ACM International Conference on Web Search and Data Mining, Melbourne, Australia, February 2019 (**Best Paper Award, WSDM-2019**). pdf
- H. Jain, Y. Prabhu and M. Varma. **Extreme multi-label loss functions for recommendation, tagging, ranking & other missing label applications.** In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, San Francisco, California, August 2016 (**KDD-2016**). pdf
- K. Bhatia, H. Jain, P. Kar, M. Varma and P. Jain. **Sparse local embeddings for extreme multi-label classification.** In Advances in Neural Information Processing Systems, Montreal, Canada, December 2015 (**NIPS-2015**). pdf

Academic Achievements

2019	Best Paper Award, WSDM-2019
2014 – Present	PhD Fellowship in Machine Learning Google Awarded to 52 PhD students across the globe
2010 – 2011 and 2009 – 2010	Certificate of Merit for Academic Excellence Indian Institute of Technology Kanpur
2010	Summer Undergraduate Research Grant for Excellence for research at École Centrale Paris Awarded to three students from the institute
2010	Best Project Summer Undergraduate Research Grant for Excellence for research at Indian Institute of Technology Kanpur Selected from more than 50 participants from various universities.
2007	Ranked among top 1% National Standard Examination in Physics

Work Experience

July'18 – Oct'18	Internship Google Research, Mountain View Advisor: Nicolas Mayoraz
May'17 – Sept'17	Internship Microsoft Research, Bangalore Advisor: Manik Varma
Jan'14 – Present	Teaching Assistant Machine Learning, Artificial Intelligence & Intro to Computer Science

Aug'12 – Dec'13	Engineer Modelling and Simulation Centre of Excellence, Eaton Technologies, Pune (India) Developed MATLAB Simulink models for vehicles so as to estimate their fuel economy
Aug'11 – Nov'11	Visiting Research Scholar Michigan State University, East Lansing (USA) Advisor: Prof. Kalyanmoy Deb
May'11 – Jul'11	Internship Denmark Technical University, Copenhagen (Denmark) Advisor: Prof. Jesper Hattel
May'10 – Jul'10	Internship École Centrale Paris (France) Advisor: Prof. Regis Cottureau

Research Projects

- **Training 100 million extreme classifiers for recommending related searches**
 - Developed the *Slice* algorithm that reduced the training and prediction times from linear in the number of labels to **logarithmic** while maintaining state-of-the art accuracy.
 - Slice is shown to scale to problems with as large as **100 million labels** and **240 million training points** which is beyond the pale of any existing extreme classification algorithm
 - Used Slice to recommend alternate, reformulated search queries called **related searches**, on Bing. Slice could increase the number of recommendations per query by 33% as compared to Bing algorithms, with **12.62% increase in success rate on tail queries** and 2.62% improvement overall.
- **Extreme multi-label loss functions for recommendation, tagging, ranking & other missing label applications**
 - Proposed a new extreme multi-label loss function, which provides an **unbiased estimate** of the true loss even if the ground truth data is incomplete and also naturally promotes prediction of tail/rare labels
 - Developed the PfastreXML algorithm which optimizes the new loss and efficiently scales to **large datasets** with up to 9 million labels, 70 million points and 2 million dimensions
 - Used PfastreXML's query ranking to serve ads on the Bing search engine and observed a **5% improvement in clickthrough rate** over the existing microsoft system
- **Sparse local embeddings for extreme multi-label classification**
 - Developed an embedding based method – SLEEC, to solve extreme multi-label learning problems
 - Among the **state-of-the-art** embedding based methods for extreme classification.

Course Projects

- **Automatic summary generation from cricket commentary**
 - Created a tool to automatically generate short summary of a given cricket match using its live commentary
 - Awarded best project among more than 20 projects presented
- **Shared file system using the FUSE framework**
 - Created a file system that supported – file creation, modification, appending, deletion & other shell commands.
 - Also implemented a file system cache and ensured that caches across client machines are consistent.

Courses

Probabilistic graphical models	Convex Optimization & Online Learning	Numerical Linear Algebra
Natural Language Processing	Advanced Distributed Systems	Mathematical Programming
Computer Vision		