

Himanshu Jain

himanshu.j689@gmail.com , PhD Student,

Indian Institute of Technology, Delhi

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

Research Interests

Extreme classification, recommender systems, tagging, advertiser bid phrase suggestion & performance evaluation

Academic Profile

2014 - Present	PhD, Computer Science and Engineering Advisor: Manik Varma, Researcher, Microsoft Research Indian Institute of Technology, Delhi	CGPA: 8.6/10.0
2007 - 2012	Bachelor and Master of Technology, Mechanical Engineering Indian Institute of Technology, Kanpur	CGPA: BTech. - 8.9/10 MTech. - 9.7/10

Publications

- H. Jain, Y. Prabhu and M. Varma. **Extreme multi-label loss functions for recommendation, tagging, ranking & other missing label applications.** *KDD 2016*. pdf
- K. Bhatia, H. Jain, P. Kar, M. Varma and P. Jain. **Sparse local embeddings for extreme multi-label classification.** *NIPS 2015*. pdf

Academic Achievements

- Awarded **Google PhD Fellowship** in Machine Learning, 2014 (given to 52 PhD students across the globe)
- Awarded Certificate of Merit for **Academic Excellence** by IIT Kanpur for academic year 2009-10 & 2010-11.
- Selected among 3 students for **Summer Undergraduate Research Grant for Excellence – 2010** for research at Ecole Centrale Paris (France).
- Awarded the **Best Project** in **Summer Undergraduate Research Grant for Excellence – 2009** among more than 50 participants from IIT Kanpur, various NITs, Ecole Centrale Paris & Rice University.
- Ranked among the **top 1%** in **National Standard Examination in Physics** (2007).

Work Experience

Jan'14 – Present	Teaching Assistant for Machine Learning, Artificial Intelligence & Introduction 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

Projects

- PhD Thesis: **Extreme multi-label loss functions for recommendation, tagging, ranking & other missing label applications** (2015 - Present)
 - 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 click-through rate** over the existing microsoft system
- **Sparse local embeddings for extreme multi-label classification** (2014 - 2015)
 - Developed an embedding based method – SLEEC, to solve extreme multi-label learning problems
 - Current **state-of-the-art** among embedding based methods and the only embedding based method that can scale to datasets with millions of labels
- **Automatic Summary Generation from Cricket Commentary** (Fall 2014)
 - 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** (Fall 2014)
 - 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	Computer Vision
Natural Language Processing	Advanced Distributed Systems	Mathematical Programming	

Technical Skills

Python, C, C++, HTML, MATLAB, L^AT_EX