Vihari Piratla

Research Associate, University of Cambridge

vp421@cam.ac.uk

Research Focus

I research challenges related to the deployment of AI systems. My recent efforts focused on expressive forms of human supervision for well-defining the task and explanation methods to expose misalignment.
Research Areas: Reliable and Safe ML systems, Trustworthy ML, AI Alignment.

Education

- **Indian Institute of Technology, Bombay** *June 2017 - July 2022*
  M.Tech. + Ph.D. Dual Degree, Department of Computer Science GPA: 9.75/10
  Advisors: Prof. Sunita Sarawagi and Prof. Soumen Chakrabarti
  Thesis: *Robustness, Evaluation and Adaptation of Machine Learning Models in the Wild* [PDF]

- **Indian Institute of Technology, Mandi** *2010 - 2014*
  B.Tech. Department of Computer Science

Professional Experience

- Research Associate Aug'22 - *Present*
  Mentor: Dr Adrian Weller *ML Group, University of Cambridge*
  Trustworthy Machine Learning.
  - By-fellow, Churchill College Feb'23-Present
  - Postdoctoral Research Associate, Trinity Hall College Sept'22-Jan'23

- Research Intern Aug'19 - Nov '19
  Mentor: Dr Praneeth Netrapalli *Microsft Research, India*
  Worked on algorithms that enable efficient transfer and generalization to new domains.
  ☆ Work done during the internship was published at ICML 2020.

- Project Staff Oct'16 - July'17
  Mentor: Prof. Sunita Sarawagi *IIT Bombay*
  Developed a smart complaint management system.

- Research Member Staff June'14 - Feb'16
  Mentor: Dr Sudheendra Hangal *Amuse Labs, Stanford University Libraries*
  Contributed the following language processing features to an open-source digital archival project called ePADD. (1) A fine-grained entity recognizer robust to domain shifts using binomial mixture models. (2) Cross document co-referencing and entity linking using context cues.

Scholastic Achievements

- Best thesis awards from the department (Vashee Award), institution (Naik & Rastogi Excellence in Ph. D. Thesis), and a national competition (ACM IKDD Doctoral Dissertation In Data Science Award).
- One of the sixteen global recipients of Google PhD fellowship in Machine Learning in 2020.
- Selected to receive Prime Minister's Fellowship for Doctoral Research, 2021 (declined for clerical reasons).
- Travel grant awards to present at ICLR 2018, ACL 2019 from Google and Microsoft.
- Department Rank one among 100 students in the M.Tech. computer science batch of 2019.
- Represented IIT Mandi, as a member of a team, in ACM International Collegiate Programming Contest 2012-13 Kharagpur and 2013-14 Kanpur regional.
- Qualified for the Indian National Mathematical Olympiad after clearing Regional Mathematical Olympiad'09 with a state rank of 26.

Selected Publications

Please see Google Scholar page 🎓 for a full list of publications.

1. Use Perturbations when Learning from Explanations
   J Heo*, **V Piratla***, M Wicker, A Weller [code][slides]
   Neural Information Processing Systems (NeurIPS) 2023.

2. Focus on the Common Good: Group Distributional Robustness Follows
   V Piratla, P Netrapalli, S Sarawagi [code][talk][slides]
   International Conference on Learning Representations (ICLR) 2022.

3. Active Assessment of Prediction Services as Accuracy Surface Over Attribute Combinations
   V Piratla, S Chakrabarti, S Sarawagi [code][talk][slides]
   Neural Information Processing Systems (NeurIPS) 2021.
   ☆ Recognised as an AI Game Changer by NASSCOM in ML Fundamentals category.

4. Efficient Domain Generalization via Common-Specific Low-Rank Decomposition
   V Piratla, P Netrapalli, S Sarawagi [code][talk][slides]
   International Conference on Machine Learning (ICML) 2020.

5. Topic Sensitive Attention on Generic Corpora Corrects Sense Bias in Pretrained Embeddings
   V Piratla, S Sarawagi, S Chakrabarti. [code][talk][slides]
   Annual Meeting of the Association for Computational Linguistics (ACL) 2019 (Oral).

6. Generalizing Across Domains via Cross-Gradient Training
   S Shankar*, **V Piratla***, S Chakrabarti, S Chaudhuri, P Jyothi, S Sarawagi [code][talk][slides]
   International Conference on Learning Representations (ICLR) 2018.

** Shared first author*

Talks

- Are we teaching machines right? Role of supervision in training ML models. [slides]
  Presented at IIT Bombay, Google Research India, University of Cambridge in 2023
- Research Challenges when scaling to millions of users through Prediction Service APIs [talk][slides]
  Presented at Trust ML Rising Star Spotlights Series

Development Experience

ePADD: Digital Archival Project Aug'14 – Feb'16
Stanford University Libraries

- ePADD is an open-source project to develop tools for collecting and processing of digital archives.
- I contributed tens of thousands of lines of code for smooth functioning of the application across various operating systems, browsers, compute hardware, and for processing archives that are several gigabytes large.

☆ The features I contributed continue to be a big part of the project to this day: browse here.

Technical Skills

- **Programming & Scripting**: Java, Python, C, Shell Scripting, PERL
- **Technologies** : Spring, HTML, CSS, JavaScript, Lucene
- **Programming Libraries**: PyTorch, GPyTorch, Tensorflow, NumPy, Pyro