Divyat Mahajan

Research Fellow
Microsoft Research India

t-dimaha@microsoft.com | divyatmahajan@gmail.com
Webpage : www.divy.at
Github : www.github.com/divyat09
Google Scholar

Education

Indian Institute of Technology Kanpur *July '14 - June '19*
B.S. in Mathematics and Scientific Computing GPA: 8.6/10 (Overall)
Double Major in Computer Science and Engineering

Research Interests

Primary: Generalization, Robustness & Explainability in Machine Learning, Causal Inference
Others: Probabilistic Models, Bayesian Statistics, Algorithmic Fairness

Research Experience

• **Microsoft Research India - Research Fellow** *July '19 - Present*
Advisor: Dr. Amit Sharma
Projects: Causal Inference, Counterfactual Explanations, Domain Generalization, Privacy Attacks in ML

• **Aalto University - Research Intern** *May '18 - July '18*
Advisor: Prof. Samuel Kaski
Projects: Approximate Bayesian Computation for Cancer Simulation

• **National University of Singapore - Research Intern** *May '17 - July '17*
Advisor: Prof. Wynne Hsu and Prof. Lee Mong Li
Projects: Recommender Systems for Side Effect Prediction

• **New York Office, IIT Kanpur - Research Track Exploration Intern** *June '16 - July '16*
Advisor: Prof. Vincent Ng
Projects: Stance Classification of Tweets in Online Debates

Publications

• **Domain Generalization using Causal Matching** [Link]
Divyat Mahajan, Shruti Tople, Amit Sharma
Workshop on Uncertainty and Robustness in Deep Learning **(ICML '20)**

• **Preserving Causal Constraints in Counterfactual Explanations for Machine Learning Classifiers** [Link]
Divyat Mahajan, Chenhao Tan, Amit Sharma
Workshop on "Do the right thing": machine learning and causal inference for improved decision making **(NeurIPS '19)**

• **Split-Treatment Analysis to Rank Heterogeneous Causal Effects for Prospective Interventions** [Link]
Yanbo Xu, Divyat Mahajan, Liz Manrao, Amit Sharma, Emre Kiciman
Proceedings of the ACM International Conference on Web Search and Data Mining **(WSDM '21)**

• **Does Learning Stable Features Provide Privacy Benefits for Machine Learning Models?** [Link]
Divyat Mahajan, Shruti Tople, Amit Sharma
Workshop on Privacy Preserving Machine Learning **(NeurIPS '20)**

• **Towards Unifying Feature Attribution and Counterfactual Explanations: Different Means to Same End** [Link]
Ramaravind Mothilal, Divyat Mahajan, Chenhao Tan, Amit Sharma
Under Review

• **A Generative Framework for Zero-Shot Learning with Adversarial Domain Adaptation** [Link]
Varun Khare*, Divyat Mahajan*, Homanga Bharadhwaj, Vinay Verma, Piyush Rai (*: Equal Contribution)
Proceedings of the IEEE Winter Conference on Applications of Computer Vision **(WACV '20)**

Software

• **RobustDG - Microsoft** [Github] [Commit History]
Core developer for Microsoft's open-source framework for building robust ML models that generalize to unseen domains

• **DiCE - InterpretML** [Github] [Commit History]
Collaborating on the InterpretML's open-source framework to support feasible counterfactual explanation approaches

Research Projects

Domain Generalization with Causal Matching & Privacy Attacks [Project]

Advisors: Dr. Amit Sharma, Dr. Shruti Tople, Microsoft Research *March '20 - Present*

- Proposed a causal framework for the problem of domain generalization (DG) and used it to prove the incorrectness of the prior DG methods via d-separation principles.
- Developed a novel algorithm (MatchDG) that uses contrastive learning to satisfy the invariance criteria from the causal graph and achieves the state-of-the-art out-of-domain accuracy on the Rotated-MNIST, Fashion-MNIST, PACS, Chest X-ray datasets.
- Established a connection between out-of-distribution generalization and privacy attacks, and used it to demonstrate issues with the state of the art DG algorithms under Membership Inference Attacks.
- Work on MatchDG got accepted at the ICML UDL Workshop 2020, and the work on relationship with privacy attacks got accepted at the NeurIPS PPML Workshop 2020, along with the creation of the Microsoft's open source framework RobustDG.

Feasible Counterfactual Explanations for ML Classifiers [Project]

Advisors: Dr. Amit Sharma, Microsoft Research, Prof. Chenhao Tan, University of Colorado Boulder *July '19 - Present*

- Proposed a causal proximity regularizer using Structural Causal Models (SCM) to address the feasibility of counterfactual (CF) explanations for Machine Learning classifiers.
- Developed a generative framework using variational inference for efficient CF generation and feasibility preservation under different assumptions of the problem setting like access to SCM, User Feedback, etc.
- Accepted for oral spotlight presentation at the NeurIPS CausalML Workshop 2019, and integrated into the open source framework DiCE by InterpretML.

Ranking Causal Effects for Prospective Interventions [Paper]

Advisors: Dr. Amit Sharma, Dr. Emre Kiciman, Microsoft Research *December '20 - October '20*

- Worked on the sensitivity analysis of methods for heterogeneous causal effect estimation of novel treatments.
- Developed a technique to capture the sensitivity of a model by generating unobserved confounders and implemented it on a large real-world software dataset for model selection under the proposed Split Treatment framework.
- Accepted at the proceedings of the conference WSDM 2021.

Generative Zero Shot Learning with Adversarial Domain Adaptation [Project]

Advisors: Prof. Piyush Rai, IIT Kanpur *February '18 - November '18*

- Addressed the issue of domain shift between the training and test classes in zero shot learning (ZSL) using adversarial domain adaptation.
- Developed a generative framework to efficiently estimate class data distributions and proposed a regularizer based on Cycle GAN that improves its robustness against domain shifts.
- Accepted at the proceedings of the conference WACV 2020, with results better than many state-of-the-art ZSL models on various benchmark datasets.

Relevant Coursework (*: Received Best (A) Grade)

Machine Learning:	Machine Learning Techniques*, Topics in Probabilistic Modelling and Inference* Visual Recognition*, Probabilistic Machine Learning
Statistics	Probability and Statistics*, Applied Stochastic Process, Statistical Inference
Mathematics	Linear Algebra, Math-I*, Math-II*, Mathematical Logic*, Topology*, Real Analysis, Complex Analysis, Abstract Algebra, Numerical Computation*
Algorithms & Theory	Data Structure and Algorithm, Algorithms II, Theory of Computation, Quantum Computing

Awards & Achievements

- Selected for the **Oral Spotlight** presentation at the NeurIPS CausalML Workshop 2019
- Received the **Academic Excellence Award**, IIT Kanpur for the academic session 2017-2018
- Received Certificate of Achievement for Rank 33 in **ACM ICPC 2017** Asia Gwalior Online Programming Round
- Secured All India Rank 1940 in JEE-Advanced 2014 out of 150,000 students with percentile 98.71
- Obtained merit with Rank 13 in Matriculation Examination (2012) and Rank 36 in Senior Secondary Examination (2014) among applicants from the H.P. state

Other Relevant Projects

Visual Program Synthesis [Presentation] [Report]
Advisors: Prof. Vinay Namboodiri, IIT Kanpur *January '18-December '18*

- Proposed a Deep Generative model using Conditional DC GAN and LSTM for program synthesis of Logo language from the input textual specification.
- Trained the complete pipeline end to end using TensorFlow on a self created-dataset of geometrical objects, and generated syntactically and semantically correct Logo program.

Interpretable Hierarchical Reinforcement Learning [Report]
Advisors: Prof. Vinay Namboodiri, IIT Kanpur *January '19 - May '19*

- Worked on designing interpretable Reinforcement Learning solutions that can generalize to unseen goals during evaluation using meta learning and information maximization.
- Experimented with the frozen lake environment and showed our model can generalize over all the goals by training on only 8 goals, while maintaining interpretability in the lower/action policies.

Approximate Bayesian Computation for Cancer Simulation [Presentation]
Advisors: Prof. Samuel Kaski, Aalto University *May '18 - July '18*

- Worked on the inference of a complex stochastic simulator based model with an intractable likelihood, that represents the cancer treatment process of a patient.
- Performed exploratory data analysis to determine informative statistics for dimensionality reduction and inference of the parameters with the bayesian optimisation framework for Likelihood Free Inference.

Recommender Systems for Side Effect Prediction [Code]
Advisors: Prof. Wynne Hsu and Prof. Lee Mong Li, National University of Singapore *May '17 - July '17*

- Worked on building a recommender system that predicts effectiveness and side effects for the usage of a drug on a patient.
- Created, Preprocessed a dataset and performed baseline evaluations using matrix factorization algorithms, regression models, along with Deep Learning models to learn better latent features.

Stance Classification of Tweets in Online Debates [Project]
Advisors: Prof. Vincent Ng, New York Office, IIT Kanpur *June '16 - July '16*

- Worked on predicting stance for tweets against a target, a task in International Workshop on Semantic Evaluation 2016.
- Developed a model for tweets that do not express opinion directly about the main target and implemented it using sklearn library.

Technical Skills

Programming Languages C/C++, Python, Bash
Software and Utilities Git, Docker, LaTeX, PyTorch, TensorFlow, Sklearn, ELFI, DoWhy

Extra Curricular Activities

- (2020) Managing the **Machine Learning Reading Group** at Microsoft Research India
- (2018) **Project Mentor** for the course Machine Learning Techniques (CS771A) offered by Prof. Piyush Rai at IIT Kanpur
- (2018) Mentored 5 students for a project on Recommender Systems under Association of Computing Activities, IITK
- (2017) Managed a team of 5 members to publish newsletter Alpha under Statmatics, mathematics society of IITK
- (2016) Volunteered in Blood Connect, NGO working to provide a solution for the shortage of blood in India
- (2015) Worked in National Service Scheme at IIT Kanpur to provide better education to underpriviledged children