ABHINAV KUMAR

Contact Information

Email: abhinavkumar.wk@gmail.com
Links: Webpage, Google Scholar, Github

Education

Birla Institute Of Technology and Science, Pilani 2015-2020
M.Sc.(Hons.) Physics GPA: 9.54/10
Thesis: Disentangling Mixtures of Unknown Causal Intervention
Advisor: Dr. Gaurav Sinha, Adobe Research, Bangalore

Birla Institute Of Technology and Science, Pilani 2015-2020
B.E.(Hons.) Computer Science GPA: 9.54/10
Thesis: Fine-Tuning Word Embedding for Domain Adaptation
Advisor: Dr. Partha Talukdar, Indian Institute of Science, Bangalore

Work Experience

Microsoft Research, Bangalore 07/21 - Present
Research Fellow
Generalization and Explainability of ML model with Causal Perspective
Advisor: Dr. Amit Sharma, Dr. Chenhao Tan and Dr. Amit Deshpande

Paypal, Hyderabad 08/20 - 06/21
Software Engineer 1
Backend Service Development for Fraud Detection Platform

Adobe Research, Bangalore 01/20 - 07/20
Research Intern
Root Cause Analysis with Causal Perspective
Advisor: Dr. Gaurav Sinha

Indian Institute of Science *(IISc)*, Bangalore 05/19 - 12/19
Research Intern
Fine-Tuning Word Embedding for Domain Adaptation
Advisor: Dr. Partha Talukdar

Google Summer of Code 05/18 - 08/18
Research Intern, CERN-High Energy Software Foundation
Deep Learning for Particle Detection and Energy Prediction for particle detectors at CERN
Advisor: Dr. Grasseau Gilles and Dr. Florian Beaudett

Center for Astronomy and Astrophysics *(IUCAA)*, Pune 05/17 - 07/17
Research Intern
Efficient Computation of Gravitational Potential in N-body simulation
Advisor: Dr. Kanak Saha

Publications

1. Abhinav Kumar, Chenhao Tan, Amit Sharma. "**Probing Classifiers are Unreliable for Concept Removal and Detection**". To appear in 36th Conference on Neural Information Processing Systems (Paper , NeurIPS 2022).
2. Abhinav Kumar, Gaurav Sinha. "**Disentangling mixtures of unknown causal interventions**". Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence (Paper , UAI 2021) [**Oral**, 6% **acceptance rate**].

3. Gilles Grasseau, Abhinav Kumar, Andrea Sartirana, Artur Lobanov and Florian Beaudette. "**A deep neural network method for analyzing the CMS High Granularity Calorimeter (HGCAL) events**". 24th International Conference on Computing in High Energy and Nuclear Physics (Paper , CHEP 2019).

Selected Research Projects

Unreliability of Probing Classifier 07/21 - 05/22
Advisors: Dr. Amit Sharma and Dr. Chenhao Tan

1. Theoretically proved that latent space based concept detection and removal methods like Null-Space removal (INLP) and Adversarial Removal will fail even under favourable settings.
2. We show that using these methods could be counter-productive i.e they are unable to remove the attributes entirely, and in the worst case may end up corrupting or destroying all task-relevant features.
3. Validated the theoretical observation on three real-world NLP task: Multi-NLI, Twitter sentiment detection and Twitter mention detection.
4. This work was accepted at NeurIPS 2022.

Disentangling Mixture of Unknown Causal Intervention 01/20 - 04/21
Advisor: Dr. Gaurav Sinha, Adobe Research, Bangalore

1. Theoretically proved that, in general, identifying individual constituents given a mixture of interventions is impossible.
2. Gave sufficient condition under which we could provably identify all the unknown intervention targets constituting the mixture.
3. Our identifiability proof gave an efficient algorithm to recover these unknown intervention targets from the exponentially large search space of possible targets.
4. This work was published at UAI 2021 as an *Oral paper* with acceptance rate of 6%.

Fine-Tuning Word Embedding for Domain Adaptation 05/19 - 12/20
Advisor: Dr. Partha Talukdar, Indian Institute of Science, Bangalore

1. Proposed a new regularization scheme based on drift in sense distribution of words between the source and target domain.
2. Characterized sense drift of a word by measuring JS-divergence between the sense distribution of a word between source and target domain and sense distribution of words in both the domain were derived using existing Word Sense Disambiguation tool.
3. Our proposed regularization score preforms equivalent or better than previous work on Stack Exchange Duplicate Question Detection task, 20 Newsgroup Topic classification task and Ohsumed Medical classification task.

Relevant Coursework

Computer Science: Discrete Mathematics, Data Structure and Algorithms, Machine Learning, Data Mining, Information Retrieval, Theory of Computation, Logic in Computer Science, Compilers Construction

Mathematics: Multivariate Calculus, Probability and Statistics, Linear Algebra, Complex Variables and Calculus, Differential Equations

Physics: Statistical Mechanics, Math Methods of Physics, Quantum Mechanics, Non-Linear Dynamics and Chaos, Quantum Information and Computing, Electromagnetic Theory

Online: Probabilistic Graphical Models, Deep Learning Specialization

Skills

Programming Languages: Python, C, C++, Java, Matlab, Fortran
Tools and Systems: Tensorflow/Pytorch, Linux, Git

Academic Service

Peer Review: CODS-COMAD'23

Volunteering Roles

CausalML Reading Group, *Co-organiser* 10/21 - Present
Microsoft Research, Bangalore

- Co-organise weekly meetups to discuss recent trends and paper broadly in Domain Generalization and Interpretability with causal perspective.

Honors and Awards

BITS Merit Scholarship 2015 - 2020

- Recipient of university's merit scholarship awarded to top 2% students based on their academic performance.

References

1. Dr. Amit Sharma
2. Dr. Gaurav Sinha
3. Dr. Chenhao Tan