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: 9.53/10
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: 9.53/10
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 - 6/21
Software Engineer 1
Backend Service Development for Fraud Detection Service

Adobe Research, Bangalore — 01/20 - 7/20
Research Intern
Root Cause Analysis with Causal Perspective
Advisor: Dr. Gaurav Sinha

Google Summer of Code — 05/18 - 8/18
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 — 01/20 - 7/21
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. Neural network models trained on text data have been found to encode undesired linguistic or sensitive concepts in their latent representation.
2. Latent Space based concept detection and removal methods like Null-Space removal (INLP) and Adversarial Removal, which internally uses probing classifiers, have become indispensable tool for detecting and removing such unreliable concepts.
3. Through an extensive theoretical and empirical analysis, we show that these methods can be counter-productive: they are unable to remove the attributes entirely, and in the worst case may end up corrupting or destroying all task-relevant features.
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. In many real-world scenarios, such as gene knockout experiments, targeted interventions are often accompanied by unknown interventions at off-target sites. Moreover, different units can get randomly exposed to different unknown interventions, thereby creating a mixture of interventions.
2. Identifying different components of this mixture can be very valuable in some applications. We study the problem of identifying all components present in a mixture of interventions on a given causal Bayesian Network.
3. We show that this problem is not identifiable in general and give sufficient condition under which we could provably identify all the unknown intervention targets. Our proof gives an efficient algorithm to recover these 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. Traditional word embeddings like Word2Vec and GloVe are often trained on very large corpus such that several different senses of a word are compressed in one single vector. There are lots of words whose meanings are domain-dependent and these pre-trained vectors instead captures the most dominant sense.
2. The downstream tasks (target domain) in which these embeddings will be used as input could have domain bias on the meaning hence these pretrained word embedding are prone to under-perform. To solve this problem of domain shift, the general strategy involves fine-tuning, projection-based methods and data augmentation on the downstream domain.
3. We propose new regularization scheme based on drift in sense distribution of words between the source and target domain.

Skills

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