VIRAL THAKAR

MACHINE LEARNING RESEARCHER

@viralbthakar
@viralbthakar
viralbthakar@gmail.com
www.viralthakar.com

RESEARCH INTEREST

My primary research areas are machine learning, deep learning and their application to computer vision and natural language processing. Particularly I am interested in solving the learning problems using synthetic data. Recent state-of-the art approaches in machine learning and deep learning, relies on the availability of the labeled datasets with large number of data points. Generation of such datasets is prohibitively expensive and time consuming. In sectors like Finance, Insurance and Healthcare, it is more difficult to collect and build problem specific datasets because of security and privacy restrictions. Comparatively it is easy and economical to create large quantity of labeled synthetic data. Also sometimes it is not possible to collect large quantity of data samples for specific application domains. For example, usually banks have higher number of loan accounts with small businesses compared to medium or large businesses. Similarly insurance companies has higher number of claims and photographic proofs related to cars compared to trucks or any other vehicles. My research is focused to find the techniques to answer a particular question : ***How can we use synthetic or similar data to build models for data scarce problems?*** If we can use such datasets to build models which can solve problems in the real world, it can generate some significant benefits to the industry.

I am interested to research in the field of Domain Adaptation to solve such problems where data collection and annotation is the primary bottleneck. Domain adaptation is the process of building an efficient model when there is a shift between training (source domain - synthetic) and test (target domain - real) data distributions. Particularly I would like to study the unsupervised domain adaptation setup in which the learner tries to transfer the knowledge from a labeled source domain to an unlabeled target domain. Various techniques have been proposed to bridge the gap between the source domain and target domain but recently researchers have started exploring domain-invariant structures and feature representations to perform domain adaptation. These methods are based on the intuition that if a learner can learn a common representation space between the two domains, it can make the two domains appear to have similar distributions, this leads to effective domain adaptation. As per most of the recent methods for domain adaptation, deep neural networks are the primary choice to learn this domain invariant feature representation. The learnt features are expected to be discriminative for the main task but invariant with respect to source and target distributions.

Other than domain adaptation, I am interested in Graph Neural Networks --- GNNs and Self supervised learning. GNNs can be utilized to model non-euclidean data while on other hand self supervised learning techniques utilizes the pseudo labels to train models. During my initial review work I have found that current Domain Adaptation techniques fail for more complex problems. I am interested in developing algorithms where we use GNNs and self supervised learning for domain adaptation in complex problems.