A SURVEY OF MACHINE LEARNING UNDER DISTRIBUTIONAL SHIFT

by

YONG LIN

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

ABSTRACT

Machine learning models are commonly based on the independent identical distribution (I.I.D.) assumption that the training and testing data are drawn from the same distribution. However, the I.I.D. assumption can be easily violated in the real world applications where the testing distribution $P_{te}(\mathbf{X}, Y)$ differs from the training one $P_{tr}(\mathbf{X}, Y)$. Here, $\mathbf{X}$ is the input and Y is the target of interest. In recent years, enhancing models' robustness under distributional shift attracts great attention from the machine learning community, which is studied under the terminology of Out-of-Distribution Generalization. There are mainly two types of distribution shift considered in the existing literature, a) conditional shift: there exists features whose condition distribution differs, i.e., $P_{te}(Y|\mathbf{X}) \neq P_{tr}(Y|\mathbf{X})$; b) marginal shift: the marginal distribution differs, i.e., $P_{te}(\mathbf{X}) \neq P_{tr}(\mathbf{X})$, while the conditional distribution remains stable, i.e., $P_{te}(Y|\mathbf{X}) = P_{tr}(Y|\mathbf{X})$. The variety of both research direction is enormous and and we will deliberate it to present a comprehensive view, including the well known invariant risk minimization (IRM) and its variants in the conditional shift category and adversarial learning methods in the marginal shift category. We introduce our own contribution to IRM to deal with its catastrophic overfitting when ap-

plied to modern deep neural networks, which significantly boosts IRM's performance. We then introduce our own contribution on extending IRM to the setting where no domain indexes are available. We also discuss the difficulty in the marginal shift problem.