Carnegie Mellon University
Computer Science Department

THESIS ORAL

Petar Stojanov

Towards More Efficient and Data-Driven Domain Adaptation

Wednesday, August 18, 2021 – 10:00 a.m. – REMOTE

In recent years with the fast progress made in neural networks research, supervised machine learning approaches have become increasingly powerful in finding flexible functions to predict target variable Y from input features X. However, most of these complex models require a large amount of data to train, and often work under the assumption that the data points are i.i.d. In reality these assumptions are very likely to be violated. A simplified notion of this violation is when the training and the test datasets come from different joint distributions. In this setting, where the training and test datasets are also known as source and target domains respectively, domain adaptation is required to obtain good performance. In particular, when only unlabeled features are observed in the target domain, this setting is referred to as unsupervised domain adaptation, and it will be the main focus of this thesis.

Domain adaptation is a wide sub-field of machine learning with the task of designing algorithms to account for this distributional difference, for the purpose of better prediction performance in the target domain. In this thesis we make use of the data-generating process to address several subproblems of unsupervised domain adaptation (UDA). Namely, we first address the problem of UDA with multiple labeled source domains and an unlabeled target domain under the conditional-target shift setting, and we present an approach to capture the low-dimensional changes of the joint distribution across domains in order to perform prediction in the target domain. Secondly, we introduce an algorithm to reduce the dimensionality of the data when performing domain adaptation under the covariate shift setting. In particular, we make use of the particular properties of the covariate shift setting in order to reduce the dimensionality of the data such that we preserve relevant predictive information about the target variable Y . We further investigate domain adaptation from the perspective of the data-generating process when addressing the problem using neural networks. In particular, we introduce two studies in which we contribute to improving UDA when learning domain-invariant representations, or when performing direct transformation from the source to the target domain.

Thesis Committee:
Jaime G. Carbonell, Co-chair
Kun Zhang, Co-chair
Barnabas Poczos
Eric P. Xing
Aapo Hyvärinen, University of Helsinki