Semi-Supervised Single- and Multi-Domain Regression with Multi-Domain Training

Tomer Michaeli

We address the problems of multi-domain and single-domain regression based on distinct and unpaired labeled training sets for each of the domains and a large unlabeled training set from all domains. We formulate this type of problems as Bayesian estimation with partial knowledge of statistical relations. We propose a worst-case design strategy and study the resulting estimators. Our analysis explicitly accounts for the cardinality of the labeled sets and includes the special cases in which one of the labeled sets is very large or, in the other extreme, completely missing. We demonstrate our estimators in the context of removing expressions from facial images, audio-visual word recognition and image zooming and provide comparisons to several recently proposed multi-modal learning algorithms.
Besides learning from examples, the framework we propose is also suited for treating signal and image processing situations in which a prior on the underlying signal and measurements is assumed known. In these scenarios, the expression for the minimum mean square error estimator is often computationally demanding. By considering estimators which are optimal with respect to partial knowledge of the prior, our approach can be used to design solutions that are much easier to apply. We analyze this methodology in the context of recovering a signal, which is known to be sparse in a unitary dictionary, from noisy observations of it and of a filtered version of it. We demonstrate the usefulness of this approach in image enhancement and in target tracking.