

Learning from unlabeled data

Tomislav Šmuc, *Rudjer Boskovic Institute*

Abstract

Semi-supervised and transductive learning are hot topics in the machine learning field as the discipline starts to gain importance in knowledge discovery in different fields, especially in biosciences. The availability of huge amounts of data about unlabeled samples and the expense of experimental techniques used for learning sample labels (e.g. for determining the function of genes or proteins in the cell) explains this trend. Both semi-supervised and transductive learning make use of data structures given by both labeled and unlabeled data in the learning phase, with the aim to produce either models with significantly improved predictive capabilities or more reliable label estimates on unlabeled data. A brief overview of the main developments in the field will be given, together with hints for future research within the research program.