Feature Selection using Transductive Support Vector Machine

Zhi-li Wu, Chun-hung Li

Department of Computer Science, Hong Kong Baptist University

{vincent,chli}@comp.hkbu.edu.hk

Abstract

Transductive learning is a useful learning paradigm as the incorporation of labeled data and unlabeled data allows better models to be constructed. This paper presents an integrated approach for feature elimination and scaling using both classical feature scoring methods and Transductive Support Vector Machine (TSVM). The similarities between the support vector machines and some classical feature scoring schemes like Fisher score and odd-ratio are exploited to develop an integrated training method for the TSVM. The initial set of features and their corresponding weights for the TSVM are defined by some classical feature scoring methods and then the TSVM is refined by training with both labelled and unlabelled data. The iterated algorithm eliminates features with small absolute values in weight and then re-scales the remaining features by weights in the TSVM. The algorithm is iterated until an expected balance between generalization capacity and feature relevancy is reached.