

Nearly Optimal Semi-Supervised Learning on Subgraphs

Gayatree Ganu
Rutgers University
gganu@cs.rutgers.edu

Branislav Kveton
Technicolor Labs, Palo Alto
branislav.kveton@technicolor.com

Abstract

The harmonic solution (HS) on a graph is one of the most popular approaches to semi-supervised learning. This is the first paper that studies how to identify highly confident HS predictions on a graph based on the HS on its subgraph. The premise of our method is that the subgraph is much smaller than the graph and therefore the most confident predictions can be identified much faster than computing the HS on the graph. We introduce a class of subgraphs that allow for good approximations, prove bounds on the difference in the HS on the graph and its subgraph, and propose an efficient approach to building the subgraphs. Our solution is evaluated in the domains of handwritten digit recognition, and topic discovery in restaurant and hotel reviews. In all cases, we show that only a small portion of the graph is sufficient to identify highly confident predictions.