

Scalable algorithms for prediction on a graph

Nicolò Cesa-Bianchi

Università degli Studi di Milano

Thursday, 27 May 2010

12:30pm Room 3-E4-SR03 Via Roentgen 1, Milano

Abstract

Networked data are found in a variety of domains: Web, social networks, biological networks, and many others. In learning tasks, networked data are often represented as a weighted graph whose edge weights reflect the similarity between incident nodes.

In this talk, we consider the problem of classifying in the game-theoretic mistake bound model the nodes of an arbitrary given graph. We characterize the optimal predictive performance in terms of the cutsize of the graph's random spanning tree, and describe a randomized prediction algorithm achieving the optimal performance while running in expected time sublinear in the graph size (on most graphs).

These results are then extended to the active learning model, where training labels are obtained by querying nodes selected by the algorithm. We describe a fast query placement strategy that, in the special case of trees, achieves the optimal number of mistakes when classifying the non-queried nodes.

Joint work with: Claudio Gentile, Fabio Vitale and Giovanni Zappella