Learning in Hyperlinked Environments

Marco Gori

Dipartimento di Ingegneria dell'Informazione, University of Siena, Italy
`marco@dii.unisi.it`

Abstract. A remarkable number of important problems in different domains (e.g. web mining, pattern recognition, biology ...) are naturally modeled by functions defined on graphical domains, rather than on traditional vector spaces. Following the recent developments in statistical relational learning, in this talk, I introduce Diffusion Learning Machines (DLM) whose computation is very much related to Web ranking schemes based on link analysis. Using arguments from function approximation theory, I argue that, as a matter of fact, DLM can compute any conceivable ranking function on the Web. The learning is based on a human supervision scheme that takes into account both the content and the links of the pages. I give very promising experimental results on artificial tasks and on the learning of functions used in link analysys, like PageRank, HITS, and TrustRank. Interestingly, the proposed learning mechanism is proven to be effective also when the rank depends jointly on the page content and on the links. Finally, I argue that the propagation of the relationships expressed by the links reduces dramatically the sample complexity with respect to traditional learning machines operating on vector spaces, thus making it reasonable the application to real-world problems on the Web, like spam detection and page classification.