

Knowledge Mining Biological Network Models

Stephen Muggleton

Department of Computing
Imperial College London

Abstract. In this talk we survey work being conducted at the Centre for Integrative Systems Biology at Imperial College on the use of machine learning to build models of biochemical pathways. Within the area of Systems Biology these models provide graph-based descriptions of bio-molecular interactions which describe cellular activities such as gene regulation, metabolism and transcription. One of the key advantages of the approach taken, Inductive Logic Programming, is the availability of background knowledge on existing known biochemical networks from publicly available resources such as KEGG and BioCyc. The topic has clear societal impact owing to its application in Biology and Medicine. Moreover, object descriptions in this domain have an inherently relational structure in the form of spatial and temporal interactions of the molecules involved. The relationships include biochemical reactions in which one set of metabolites is transformed to another mediated by the involvement of an enzyme. Existing genomic information is very incomplete concerning the functions and even the existence of genes and metabolites, leading to the necessity of techniques such as logical abduction to introduce novel functions and invent new objects. Moreover, the development of active learning algorithms has allowed automatic suggestion of new experiments to test novel hypotheses. The approach thus provides support for the overall scientific cycle of hypothesis generation and experimental testing.