

Tutorial on "From ILP to PILP"

Stephen H. Muggleton

Head of Computational Bioinformatics Laboratory
Department of Computing Imperial College London

ABSTRACT:

Inductive Logic Programming (ILP) is the area of Computer Science which deals with the induction of hypothesised predicate definitions from examples and background knowledge. Probabilistic ILP (PILP) extends the ILP framework by making use of probabilistic variants of logic programs to capture background and hypothesised knowledge. ILP and PILP are differentiated from most other forms of Machine Learning (ML) both by their use of an expressive representation language and their ability to make use of logically encoded background knowledge. This has allowed successful applications in areas such as Systems Biology, computational chemistry and Natural Language Processing.

The problem of learning a set of logical clauses from examples and background knowledge has been studied since Reynold's and Plotkin's work in the late 1960's. The research area of ILP has been studied intensively since the early 1990s, while PILP has received increasing amounts of interest over the last decade. This talk will provide an overview of results for learning logic programs within the paradigms of learning-in-the-limit, PAC-learning and Bayesian learning. These results will be related to various settings, implementations and applications used in ILP.

It will be argued that the Bayes' setting has a number of distinct advantages for both ILP and PILP. Bayes' average case results are easier to compare with empirical machine learning performance than results from either PAC or learning-in-the-limit. Broad classes of logic programs are learnable in polynomial time in a Bayes' setting, while corresponding PAC results are largely negative. Bayes' can be used to derive and analyse algorithms for learning from positive only examples for classes of logic program which are unlearnable within both the PAC and learning-in-the-limit framework. It will be shown how a Bayesian approach can be used to analyse the relevance of background knowledge when learning. General results will also be discussed for expected error given a k-bit bounded incompatibility between the teacher's target distribution and the learner's prior.