

Chapter 5 ¹

LEARNING LOGIC FORMULAS AND RELATED ERROR DISTRIBUTIONS

Giovanni Felici*, Fushing Sun**, and Klaus Truemper***

* *Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti"*

*Consiglio Nazionale delle Ricerche
Viale Manzoni 30, 00185 Roma, Italy
Email: felici@iasi.cnr.it*

** *Computer Science Department, Ball State University*

*Muncie, IN 47306
Email: fsun@cs.bsu.edu*

*** *Department of Computer Science*

*University of Texas at Dallas
Richardson, TX 75083-0688, U.S.A.
Email: truemper@utdallas.edu*

Abstract: This chapter describes a method for learning logic formulas that correctly classify the records of a given data set consisting of two classes. The method derives from given training data certain minimum cost satisfiability problems, solves these problems, and deduces from the solutions the desired logic formulas. There are at least two ways in which the results may be employed. First, one may use the logic formulas directly as rules in application programs. Second, one may construct vote-based rules, where the formulas produce votes and where the votes are combined to a vote-total. The latter approach allows for assessment and even control of prediction errors, as follows: Once the method has produced the logic formulas, it computes from the training data estimated distributions for the vote-totals without use of any additional data. From these distributions the method estimates probabilities for prediction errors. That information supports assessment and control of errors. Uses of the method include data mining, knowledge acquisition in expert systems, and identification of critical characteristics for recognition systems. Computational tests indicate that the method is fast and effective.

Keywords: Inductive Inference, Supervised Learning, Logic Programming, Minimum Cost Satisfiability Problem, Learning vote distributions and Error Probabilities

¹ Triantaphyllou, E. and G. Felici (Eds.), **Data Mining and Knowledge Discovery Approaches based on Rule Induction Techniques**, Massive Computing Series, Springer, Heidelberg, Germany, pp. 193-226, 2006.