

Chapter 4¹

DISCOVERING RULES THAT GOVERN MONOTONE PHENOMENA

Vetle I. Torvik * and Evangelos Triantaphyllou **

*: *University of Illinois at Chicago, Dept. of Psychiatry, 1601 W Taylor St, Chicago, IL 60612*
Email: vtorvik@uic.edu; Web: <http://arrowsmith2.psych.uic.edu/torvik>

** : *Louisiana State University, Dept. of Computer Science, 298 Coates Hall, Baton Rouge, LA 70803,* Email: trianta@lsu.edu; Web: <http://www.csc.lsu.edu/trianta>

Abstract: Unlocking the mystery of natural phenomena is a universal objective in scientific research. The rules governing a phenomenon can most often be learned by observing it under a sufficiently large number of conditions that are sufficiently high in resolution. The general knowledge discovery process is not always easy or efficient, and even if knowledge is produced it may be hard to understand, interpret, validate, remember, and use. Monotonicity is a pervasive property in nature: it applies when each predictor variable has a non-negative effect on the phenomenon under study. Due to the monotonicity property, being able to observe the phenomenon under specifically selected conditions may increase the accuracy and completeness of the knowledge at a faster rate than a passive observer who may not receive the pieces relevant to the puzzle soon enough. This scenario can be thought of as learning by successively submitting queries to an oracle which responds with a Boolean value (phenomenon is present or absent). In practice, the oracle may take the shape of a human expert, or it may be the outcome of performing tasks such as running experiments or searching large databases. Our main goal is to pinpoint the queries that minimize the total number of queries used to completely reconstruct all of the underlying rules defined on a given finite set of observable conditions $V = \{0,1\}^n$. We summarize the optimal query selections in the simple form of selection criteria, which are near optimal and only take polynomial time (in the number of conditions) to compute. Extensive unbiased empirical results show that the proposed selection criterion approach is far superior to any of the existing methods. In fact, the average number of queries is reduced exponentially in the number of variables n and more than exponentially in the oracle's error rate.

Key Words: Monotone Boolean Functions, Active Learning, Data Mining.

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. 149-192, 2006.