

Chapter 12¹

DISCOVERING KNOWLEDGE NUGGETS WITH A GENETIC ALGORITHM

Edgar Noda

*School of Electrical & Comp. Eng.(FEEC)
State University of Campinas (UNICAMP)
Campinas –SP, Brazil
E-mail: edgar@dt.fee.unicamp.br*

Alex A. Freitas²

*Computing Laboratory, University of Kent
Canterbury, Kent, CT2 7NF, UK.
Email: A.A.Freitas@kent.ac.uk
Web: <http://www.cs.kent.ac.uk/people/staff/aaf>*

Abstract: Measuring the quality of a prediction rule is a difficult task, which can involve several criteria. The majority of the rule induction literature focuses on discovering accurate, comprehensible rules. In this chapter we also take these two criteria into account, but we go beyond them in the sense that we aim at discovering rules that are interesting (surprising) for the user. Hence, the search for rules is guided by a rule-evaluation function that considers both the degree of predictive accuracy and the degree of interestingness of candidate rules. The search is performed by two versions of a genetic algorithm (GA) specifically designed to the discovery of interesting rules - or "knowledge nuggets." The algorithm addresses the dependence modeling task (sometimes called "generalized rule induction"), where different rules can predict different goal attributes. This task can be regarded as a generalization of the very well known classification task, where all rules predict the same goal attribute. This chapter also compares the results of the two versions of the GA with the results of a simpler, greedy rule induction algorithm to discover interesting rules.

Key Words: Genetic Algorithms, Rule Interestingness, Prediction, Dependence Modeling.

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. 395-432, 2006.

² Corresponding Author