

Chapter XI

Development of Control Signatures with a Hybrid Data Mining and Genetic Algorithm

Alex Burns

The University of Iowa, USA

Shital Shah

Rush University Medical Center, Health Systems Management, USA

Andrew Kusiak

The University of Iowa, USA

ABSTRACT

This chapter presents a hybrid approach that integrates a genetic algorithm (GA) and data mining to produce control signatures. The control signatures define the best parameter intervals leading to a desired outcome. This hybrid method integrates multiple rule sets generated by a data-mining algorithm with the fitness function of a GA. The solutions of the GA represent intersections among rules providing tight parameter bounds. The integration of intuitive rules provides an explanation for each generated control setting and it provides insights into the decision-making process. The ability to analyze parameter trends and the feasible solutions generated by the GA with respect to the outcomes is another benefit of the proposed hybrid method. The presented approach for deriving control signatures is applicable to various domains, such as energy, medical protocols, manufacturing, airline operations, customer service, and so on. Control signatures were developed and tested for control of a power plant boiler. These signatures discovered insightful relationships among parameters. The results and benefits of the proposed method for the power plant boiler are discussed in the chapter.

INTRODUCTION

Optimizing process controls is imperative in the energy, medical, and service applications. Due

to the increase in the volume of data, decision making in real time is becoming more difficult, and may potentially lead to inefficiencies and hazardous situations. Intelligent control systems

have proven to be effective in optimizing complex processes. Merging computational concepts, such as neural networks, genetic algorithms, data mining, and fuzzy logic, can lead to robust controls (Krishnakumar & Goldberg, 1992; Lee, Perakis, Sevcik, Santoso, Lausterer, & Samad, 2000). Though current intelligent approaches may improve operations, they provide limited insights into the decision-making process.

This chapter describes a hybrid method that integrates data mining (Cios, Pedrycz, & Swiniarski, 1998; Fayyad, Piatetsky-Shapiro, Smyth, Uthurusamy, 1995) and genetic algorithm (GA) (Goldberg, 1989; Holland, 1975; Lawrence, 1987; Michalewicz, 1992) concepts to define robust and explicit parameter set points. The hybrid approach consists of partitioning data, developing classifiers for each data set, and combining and analyzing the classifiers (Mitra, Pal, & Mitra, 2002). These steps lead to the development of control signatures (Kusiak, 2002) that define ranges of parameter settings, producing a desired outcome (decision). Control signatures are helpful in learning the interactions and relationships between the parameters.

Data mining is the process of discovering interesting and previously unknown patterns in data sets (Cios, Pedrycz, & Swiniarski, 1998; Fayyad, Piatetsky-Smapiro, Smyth, & Uthurusamy, 1995). A typical data-mining algorithm, applied to partitioned data sets, generates multiple rule sets that describe parameter relationships. The GA provides a global search mechanism to discover the intersections among the decision rules. The intersections among decision rules could be analyzed through visualization (Kusiak, 2001). As the number of rules increases, the graphical presentation and analysis becomes tedious. Furthermore, this kind of analysis does not provide the additional information leading the tighter parameter bounds, control signatures, and increased prediction accuracy.

In this research, the process data is transformed into multiple knowledge bases that are used as

the foundation of a GA fitness function. The GA mechanism strongly promotes the solutions that are in the feasible region formed by the rule intersections. It not only allows for exclusion of the less reliable (feasible) regions, but also defines the complex commonality among multiple rules. This in turn provides tighter bounds on various parameters. Complex nonlinear applications can be analyzed as the parameter relationships are preserved by the rule sets that are incorporated into the GA fitness function.

The analysis of the control signatures across different outcomes provides information regarding the general parameter trends. Analyzing these trends and the feasible solutions generated by the GA is a way of visualizing the complex relationships and identifying key parameters. The proposed hybrid approach provides increased level of insight and was applied and tested on the data from a power plant boiler. The signatures defined ideal parameter ranges for boiler operations.

METHOD

This section describes a hybrid approach that integrates genetic algorithm (GA) and data mining to define control signatures (Figure 1). The use of GA is novel, due to the incorporation of data-mining output in the fitness function. Data mining defines relationships among parameters, while a standard GA provides the global search mechanism to identify ranges for parameter settings. The basic concepts of data mining and GA that are relevant to the development of control signatures are discussed next.

The hybrid approach requires a data set with known outcomes, but it can handle continuous, discrete, and categorical data. The GA is facilitated by data preprocessing through the normalization of continuous parameters. Discrete and categorical parameters are assigned a value based on their probability of occurrence. All parameter values are transformed back to their respective

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/development-control-signatures-hybrid-data/26140

Related Content

Improving Web Clickstream Analysis: Markov Chains Models and Genmax Algorithms

Paolo Baldini and Paolo Giudici (2008). *Mathematical Methods for Knowledge Discovery and Data Mining* (pp. 233-243).

www.irma-international.org/chapter/improving-web-clickstream-analysis/26143

Heuristic Search-Based Stacking of Classifiers

Agapito Ledezma, Ricardo Alerand Daniel Borrajo (2002). *Heuristic and Optimization for Knowledge Discovery* (pp. 54-67).

www.irma-international.org/chapter/heuristic-search-based-stacking-classifiers/22149

Data Visualization in R

S. R. Mani Sekhar, Siddesh G. M. and Sunilkumar S. Manvi (2019). *Sentiment Analysis and Knowledge Discovery in Contemporary Business* (pp. 205-222).

www.irma-international.org/chapter/data-visualization-in-r/210971

Developing Best Practices for Value Added Research in a Political Context

Valerie Osland Paton, Gerry Dizinno and Roy Mathew (2012). *Cases on Institutional Research Systems* (pp. 339-345).

www.irma-international.org/chapter/developing-best-practices-value-added/60859

Self-Organization as a Perspective for Organizational Learning: A New Role for Learning Practitioners

Robert J. Blomme (2015). *Knowledge Management for Competitive Advantage During Economic Crisis* (pp. 56-68).

www.irma-international.org/chapter/self-organization-as-a-perspective-for-organizational-learning/117842