



PERGAMON

Computers & Industrial Engineering 43 (2002) 661–676

---

---

**computers &  
industrial  
engineering**

---

---

[www.elsevier.com/locate/dsw](http://www.elsevier.com/locate/dsw)

# Feature extraction using rough set theory and genetic algorithms— an application for the simplification of product quality evaluation

Lian-Yin Zhai\*, Li-Pheng Khoo, Sai-Cheong Fok

*School of Mechanical and Production Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore, Singapore 639798*

---

## Abstract

Feature extraction is an important aspect in data mining and knowledge discovery. In this paper an integrated feature extraction approach, which is based on rough set theory and genetic algorithms (GAs), is proposed. Based on this approach, a prototype feature extraction system has been established and illustrated in an application for the simplification of product quality evaluation. The prototype system successfully integrates the capability of rough set theory in handling uncertainty with a robust search engine, which is based on a GA. The results show that it can remarkably reduce the cost and time consumed on product quality evaluation without compromising the overall specifications of the acceptance tests. © 2002 Elsevier Science Ltd. All rights reserved.

*Keywords:* Feature extraction; Rough set; Genetic algorithm; Knowledge extraction

---

## 1. Introduction

Feature extraction from large-scale empirical data is an important research area in data mining. Feature extraction is of great importance to help humans earn necessary knowledge about a specific part of a real or abstract world and further use the knowledge to make sound decisions. With the rapid development of information storage technology, a huge amount of data about a certain object can be stored and kept for later analysis. However, not all the data collected are useful or informative. In the evaluation of the quality of a part/component produced by a manufacturing system, a lot of parameters about the part/component are measured and stored. The operators will have to look through all the data to ascertain the quality of the part/component produced during product quality evaluation. This kind of job is normally monotonous and tedious, and will remarkably reduce productivity. Under such a situation, a feature extraction technique can help discover those important parameters that best describe the

---

\* Corresponding author.

*E-mail addresses:* [mlyzhai@ntu.edu.sg](mailto:mlyzhai@ntu.edu.sg) (L.-Y. Zhai), [mlpkhoo@ntu.edu.sg](mailto:mlpkhoo@ntu.edu.sg) (L.-P. Khoo), [mscfok@ntu.edu.sg](mailto:mscfok@ntu.edu.sg) (S.-C. Fok).

target object and thus can simplify the test procedure. Hence, to some extent, feature extraction can be viewed as a pre-pruning process to choose a small subset of features that is necessary and sufficient to describe target concepts. The importance of feature extraction in a broader sense is not only to reduce the search space, but also to speed up the process of both concept learning and the classification of objects and also to improve the quality of classification (Kira & Rendell, 1992).

In recent decades, many techniques have been developed to deal with feature extraction issues, among which include stepwise backward/forward techniques (James, 1985; Modrzejewski, 1993), dynamic programming (Chang, 1973), branch and bound algorithm (Narendra & Fukunaga, 1977), and so on. These methods have different strengths and drawbacks, based on the specific feature extraction criterion. However, most of these methods may not be applicable for extracting significant features from incomplete/imprecise data. This is an important issue as, in reality, due to various reasons empirical data often have the property of granularity and may be incomplete, imprecise, or even conflicting. For example, in diagnosing a machine in a manufacturing system, the opinions of two engineers can be different, or even contradicting.

The ability to handle imprecise and inconsistent information has become one of the most important requirements for a feature extraction system. Many theories, techniques and algorithms have been developed to deal with the analysis of imprecise or inconsistent data. The most successful ones are based on fuzzy set theory and the Dempster–Shafer theory of evidence. Rough set theory, which was introduced by Pawlak (1982) in the early 1980s, is a new mathematical tool that can be employed to handle uncertainty and vagueness. It focuses on the discovery of patterns in inconsistent data (Pawlak, 1996; Slowinski & Stefanowski, 1989) and can be used as the basis to perform formal reasoning under uncertainty, machine learning and rule discovery (Yao, Wong, & Lin, 1997; Ziarko, 1994). Compared to other approaches in handling uncertainty, rough set theory has its unique advantages (Pawlak, 1996, 1997). It does not require any preliminary or additional information about the empirical training data such as probability distributions in statistics; the basic probability assignment in the Dempster–Shafer theory of evidence; or grades of membership in fuzzy set theory (Pawlak, 1992). Besides, rough set theory is more appropriate in situations where the set of empirical or experimental data is too small to employ standard statistical methods (Pawlak, 1991). In less than two decades, rough set theory has rapidly established itself in many real-life applications such as medical diagnosis (Slowinski, 1992), control algorithm acquisition and process control (Mrozek, 1992) and structural engineering (Arciszewski & Ziarko, 1990). Currently, most work on inductive learning or classification using rough set theory is limited to binary-concept, such as *yes* or *no* in decision making or *positive* or *negative* in the classification of objects.

Concept learning is another important issue in feature extraction. In this aspect, genetic algorithms (GAs) have received much attention from researchers working on machine learning (Goldberg, 1989). Basically, GA-based techniques take advantage of the unique robust search engine of GAs to extract useful information or knowledge from its search space. This paper describes a prototype feature extraction system for simplifying the product quality evaluation process. It is based on a hybrid technique that combines the strengths of rough set theory and GAs. In the following sections, the basic notions of rough set theory and GAs are presented. Details of the feature extraction system and its validation are also described.

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات