Chapter 1
Introduction to the Experimental Design in the Data Mining Tool KEEL

J. Alcalá-Fdez
University of Granada, Spain

I. Robles
University of Granada, Spain

F. Herrera
University of Granada, Spain

S. García
University of Jaén, Spain

M.J. del Jesus
University of Jaén, Spain

L. Sánchez
University of Oviedo, Spain

E. Bernadó-Mansilla
University Ramon Llull, Spain

A. Peregrín
University of Huelva, Spain

S. Ventura
University of Córdoba, Spain

ABSTRACT

KEEL is a Data Mining software tool to assess the behaviour of evolutionary learning algorithms in particular and soft computing algorithms in general for different kinds of Data Mining problems including as regression, classification, clustering, pattern mining and so on. It allows us to perform a complete analysis of some learning model in comparison to existing ones, including a statistical test module for comparison. In this chapter the authors will provide a complete description of KEEL, the kind of problems and algorithms implemented, and they will present a case of study for showing the experimental design and statistical analysis that they can do with KEEL.

DOI: 10.4018/978-1-61520-757-2.ch001

INTRODUCTION

Data Mining (DM) is the process for automatic discovery of high level knowledge by obtaining information from real world, large and complex data sets (Ham & Kamber, 2006). This idea of automatically discovering knowledge from databases is a very attractive and challenging task, both for academia and industry. Hence, there has been a growing interest in DM in several Artificial Intelligence (AI)-related areas, including Evolutionary Algorithms (EAs) (Eiben & Smith, 2003).

EAs are optimization algorithms based on natural evolution and genetic processes. Nowadays in AI, they are considered as one of the most successful search techniques for complex problems and they have proved to be an important technique for learning and Knowledge Extraction. This makes them also a promising tool in DM (Cordón, Herrera, Hoffmann, & Magdalena, 2001; Freitas, 2002; Jain & Ghosh, 2005; Grefenstette, 1994; Pal & Wang, 1996; Wong & Leung, 2000). The main motivation for applying EAs to Knowledge Extraction tasks is that they are robust and adaptive search methods that perform a global search in place of candidate solutions (for instance, rules or other forms of knowledge representation).

The use of EAs in problem solving is a widespread practice (Stejic, Takama, & Hirota, 2007; Mucientes, Moreno, Bugarín, & Barro, 2006; Romero, Ventura, & Bra, 2004), however, their use requires a certain programming expertise along with considerable time and effort to write a computer program for implementing the often sophisticated algorithm according to user needs. This work can be tedious and needs to be done before users can start focusing their attention on the issues that they should be really working on. In the last few years, many software tools have been developed to reduce this task. Although a lot of them are commercially distributed (some of the leading commercial software are mining suites such as SPSS Clementine [1], Oracle Data Mining [2] and KnowledgeSTUDIO [3]), a few are available as open source software (we recommend visiting the KDnuggets software directory [4] and The-Data-Mine site [5]). Open source tools can play an important role as is pointed out in (Sonnenburg et al., 2007).

In this chapter, we provide a complete description of a non-commercial Java software tool named KEEL (Knowledge Extraction based on Evolutionary Learning) [6]. This tool empowers the user to assess the behaviour of evolutionary learning for different kinds of DM problems: regression, classification, clustering, pattern mining, etc. This tool can offer several advantages:

- It reduces programming work. It includes a library with evolutionary learning algorithms based on different paradigms (Pittsburgh, Michigan and IRL) and simplifies the integration of evolutionary learning algorithms with different preprocessing techniques. It can alleviate researchers from the mere "technical work" of programming and enable them to focus more on the analysis of their new learning models in comparison with the existing ones.
- It extends the range of possible users applying evolutionary learning algorithms. An extensive library of EAs together with easy-to-use software considerably reduce the level of knowledge and experience required by researchers in evolutionary computation. As a result researchers with less knowledge, when using this framework, would be able to apply successfully these algorithms to their problems.
- Due to the use of a strict object-oriented approach for the library and software tool, these can be used on any machine with Java. As a result, any researcher can use KEEL on his machine, independently of the operating system.