Using Subgroup Discovery algorithms in R: The **SDR** package

Angel M. Garcia amgv0009@red.ujaen.es

2015-07-12

Abstract

Subgroup Discovery is a data mining task between classification and description, this task is nowadays an relevant task for researchers due to it success in many fields, especially the medical field. The SDR package provide the posibility of using some of the algorithms that exists in the specialized bibliography by means of using datasets provided in the format specified by data mining tool KEEL without any other dependencies in the R console. Also, the package provide a graphical interface to use the algorithms in a more visual way and to do some basics exploratory analysis with datasets in KEEL format.

1   Introduction

Nowadays, a huge amount of information is generated everyday over the Internet. Such information contains implicit knowledge that is very important to many organizations because they give the possibility, for example, of improve their products or services.

In order to achieve this purpose, traditional techniques of knowledge extraction like *OLAP(On-Line Analitical Processing)* are not capable to extract knowledge as well as data mining do. Due to the success of data mining at extracting knowledge, a lot of techniques and methods have been designed. Such methods can be classified into two vast fields according to their final objective:

- Predictive data mining: which objective is to find a value of a predefined variable of interest in new instances that arrive at the system.

- Descriptive data mining: which objective is to find relationships between instances that are at the moment in the system.

This classification doesn