Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 36 (2009) 1632–1644
Expert Systems with Applications
www.elsevier.com/locate/eswa

Evolutionary algorithms for subgroup discovery in e-learning: A practical application using Moodle data

C. Romero [a,*], P. González [b], S. Ventura [a], M.J. del Jesus [b], F. Herrera [c]

[a] Department of Computer Sciences and Numerical Analysis, University of Córdoba, 14071 Córdoba, Spain
[b] Department of Computer Sciences, University of Jaén, 23071 Jaén, Spain
[c] Department of Computer Sciences and Artificial Intelligence, University of Granada, 18071 Granada, Spain

Abstract

This work describes the application of subgroup discovery using evolutionary algorithms to the usage data of the Moodle course management system, a case study of the University of Cordoba, Spain. The objective is to obtain rules which describe relationships between the student's usage of the different activities and modules provided by this e-learning system and the final marks obtained in the courses. We use an evolutionary algorithm for the induction of fuzzy rules in canonical form and disjunctive normal form. The results obtained by different algorithms for subgroup discovery are compared, showing the suitability of the evolutionary subgroup discovery to this problem.

Keywords: Web-based education; Subgroup discovery; Evolutionary algorithms; Fuzzy rules

1. Introduction

The design and implementation of web-based education systems have grown exponentially in the last years, spurred by the fact that neither students nor teachers are bound to a specific location and that this form of computer-based education is virtually independent of any specific hardware platforms. These systems accumulate a great deal of information which is very valuable in analyzing students' behavior and assisting authors in the detection of possible errors, shortcomings and improvements. However, due to the vast quantities of data these systems can generate daily, it is very difficult to manage manually, and authors demand tools which assist them in this task, preferably on a continuous basis. The use of data mining is a promising area in the achievement of this objective (Romero & Ventura, 2006, 2007).

In the knowledge discovery in databases (KDD) process, the data mining step consists of the automatic extraction of implicit and interesting patterns from large data collections. A list of data mining techniques or tasks includes statistics, clustering, classification, outlier detection, association rule mining, sequential pattern mining, text mining, or subgroup discovery, among others (Klösgen & Zytkow, 2002).

In recent years, researchers have begun to investigate various data mining methods in order to help teachers improve e-learning systems. A review can be seen in (Romero & Ventura, 2007). These methods allow the discovery of new knowledge-based on students' usage data.

Subgroup discovery is a specific method for discovering descriptive rules (Klösgen, 1996; Wrobel, 1997). The objective is to discover characteristics of subgroups with respect to a specific property of interest (represented in the rule consequent). It must be noted that subgroup discovery aims at discovering individual rules (or local patterns of interest), which must be represented in explicit symbolic form and which must be relatively simple in order to be recognized as actionable by potential users. Therefore, the subgroups discovered in data have an explanatory nature and the interpretability for the final user of the

* Corresponding author. Fax: +34 957 218630.
E-mail address: cromero@uco.es (C. Romero).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.11.026

extracted knowledge is a crucial aspect in this field. This task has been applied to different domains: detection of patient groups with risk for atherosclerotic coronary heart disease (Gamberger & Lavrac, 2002b), mining UK traffic data (Kavsek, Lavrac, & Bullas, 2002), personal web pages (Nakada & Kunifuji, 2003), identification of interesting diagnostic patterns to supplement a medical documentation and consultation system (Atzmueller, Puppe, & Buscher, 2004) or marketing problems (del Jesus, González, Herrera, & Mesonero, 2007).

This work proposes the application of subgroup discovery to the usage data of the course management system Moodle at the University of Cordoba, Spain. Moodle is a free open source course management system designed to help educators create effective online learning communities. Moodle has a flexible array of course activities such as forums, chats, quizzes, resources, choices, surveys, or assignments. Our objective is to obtain rules which describe relationships between the student's usage of the different activities and modules provided by this e-learning system and the final score obtained in the courses. These rules can help the teacher to discover beneficial or detrimental relationships between the use of web-based educational resources and the student's learning.

We will focus our attention in the use of a subgroup discovery algorithm-based on the use of genetic algorithms (GAs) called SDIGA (Subgroup Discovery Iterative Genetic Algorithm). SDIGA is an evolutionary model for the extraction of fuzzy rules for the subgroup discovery task. This algorithm is described in detail in (del Jesus et al., 2007). Its main characteristics are presented in this paper.

We compare the results obtained by this algorithm with those obtained by two classical subgroup discovery methods: Apriori-SD (Kavsek & Lavrac, 2006) and CN2-SD (Lavrac, Kavsec, Flach, & Todorovski, 2004). Furthermore, we also use an algorithm for class association rule discovery such as CBA (Classification Based on Association) (Liu, Hsu, & Ma, 1998). We will present an experimental study where SDIGA obtains the best results for our educational mining problem.

This paper is arranged in the following way: Section 2 describes the problem of discovering rules in e-learning and surveys some specific work in the area. Section 3 introduces the subgroup discovery task, the type of rules and quality measures used and the fuzzy evolutionary approach. Section 4 describes the e-learning case study, the experimentation carried out and the analysis of results. Finally, the conclusions and further research are outlined.

2. Rule discovery in learning management systems

Many web-based educational systems with different capabilities and approaches have been developed to deliver online education. There are different types of web-based educational systems: particular web-based courses, learning management systems, and adaptive and intelligent web-based educational systems (Romero & Ventura, 2006). This paper is mostly oriented forwards learning management systems. Different terms are used to denominate these systems: learning management systems (LMS), course management systems or learning content management systems. These systems are e-learning platforms that offer a great variety of channels and workspaces to facilitate information sharing and communication between participants in a course, allow educators to distribute information to students, produce content material, prepare assignments and tests, engage in discussions, manage distance classes and enable collaborative learning with forums, chats, file storage areas or news services. Some examples of commercial LMS are Blackboard, Virtual-U, WebCT, or TopClass among others and some examples of free LMS are Moodle, Ilias, Claroline, or ATutor. (Paulsen, 2003). These systems normally use a relational database to store the large data log of the students' activities and usage information. And although some platforms offer reporting tools, if there are a great number of students and a great amount of information, it becomes difficult for a tutor to extract useful information. Recently, some researchers propose using data mining techniques in order to help the tutor in this task.

Data mining techniques can be applied to analyzing student's usage data in order to identify useful patterns and to evaluate web activity to get more objective feedback for instruction and more knowledge about how the students learn on the LMS (Romero & Ventura, 2007). A data mining algorithm can discover knowledge using different representation models and techniques from two different perspectives:

- Predictive induction, whose objective is the discovery of knowledge for classification or prediction (Michie, Spiegelhalter, & Taylor, 1994). Classification rule discovery (Quinlan, 1993) or clustering (Han, Kamber, & Tung, 2001) are data mining tasks under the predictive induction approach.
- Descriptive induction, whose main objective is the extraction of interesting knowledge from data In this area, attention can be drawn to the discovery of association rules following an unsupervised learning model (Agrawal, Imielinski, & Swami, 1993), subgroup discovery (Klösgen, 1996; Wrobel, 1997) and other approaches to non-classificatory induction.

In the following, some of the most widely used data mining techniques in e-learning are described.

Classification is a supervised process of grouping physical or abstract objects into classes of similar characteristics. It belongs to predictive induction data mining methods. The objective of classification rules (Quinlan, 1993) is to obtain the necessary knowledge to create a classification system. The antecedents of these rules contain requirements (in the form of conditions), which match those objects that belong to the class label identified in the con-