Expert Systems with Applications 37 (2010) 6891–6903

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Evaluation of e-learning systems based on fuzzy clustering models and statistical tools

Mofreh A. Hogo *

Electrical Engineering Technology Dept., Benha Higher Institute of Technology, Benha University, Benha, Egypt

ARTICLE INFO

Keywords:
e-Learning
Learner profile
Fuzzy C-means clustering
Kernelized FCM
Log file analyzer

ABSTRACT

This paper introduces a hybridization approach of AI techniques and statistical tools to evaluate and adapt the e-learning systems including e-learners. Learner's profile plays a crucial role in the evaluation process and the recommendations to improve the e-learning process. This work classifies the learners into specific categories based on the learner's profiles; the learners' classes named as regular, workers, casual, bad, and absent. The work extracted the statistical usage patterns that give a clear map describing the data and helping in constructing the e-learning system. The work tries to find the answers of the question how to return the bad students who are away back to be regular ones and find a method to evaluate the e-learners as well as to adapt the content and structure of the e-learning system. The work introduces the application of different fuzzy clustering techniques (FCM and KFCM) to find the learners profiles. Different phases of the work are presented. Analysis of the results and comparison: There is a match with a 78% with the real world behavior and the fuzzy clustering reflects the learners' behavior perfectly. Comparison between FCM and KFCM proved that the KFCM is much better than FCM.

1. Introduction

The design and implementation of web-based education (e-learning) 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 (Romero, González, Ventura, del Jesus, & Herrera, 2009). These systems accumulate a great deal of information; which is very valuable in analyzing students' behavior and assisting teachers 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, 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 and 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 proposed system for evaluating the e-learning systems and e-learners is shown in Fig. 1. The development of this system is our goal in this paper.

The rest of this paper is organized in the following way: Section 2 introduces survey on soft computing in e-learning. Section 3 describes the problems and goals of the work presented in this paper. Section 4 introduces the theoretical review for the fuzzy clustering techniques used. Section 5 introduces the data sets, including data preparation, data cleaning, data normalization, and features selection. Section 6 introduces the statistical analysis tools of web log files. Section 7 presents the experiments design and results analysis. Comparison between the different clustering techniques and the comparison with the real behavior are introduced in Section 8. Section 9 introduces the combination between the results of fuzzy clustering and the log file analyzer results to construct the suggestions and the recommendations for the e-learning system. Finally, the conclusions and further research are outlined in Section 10.

* Tel.: +20 109525805; fax: +20 483667448.
E-mail address: mofreh_hogo@hotmail.com

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.03.032

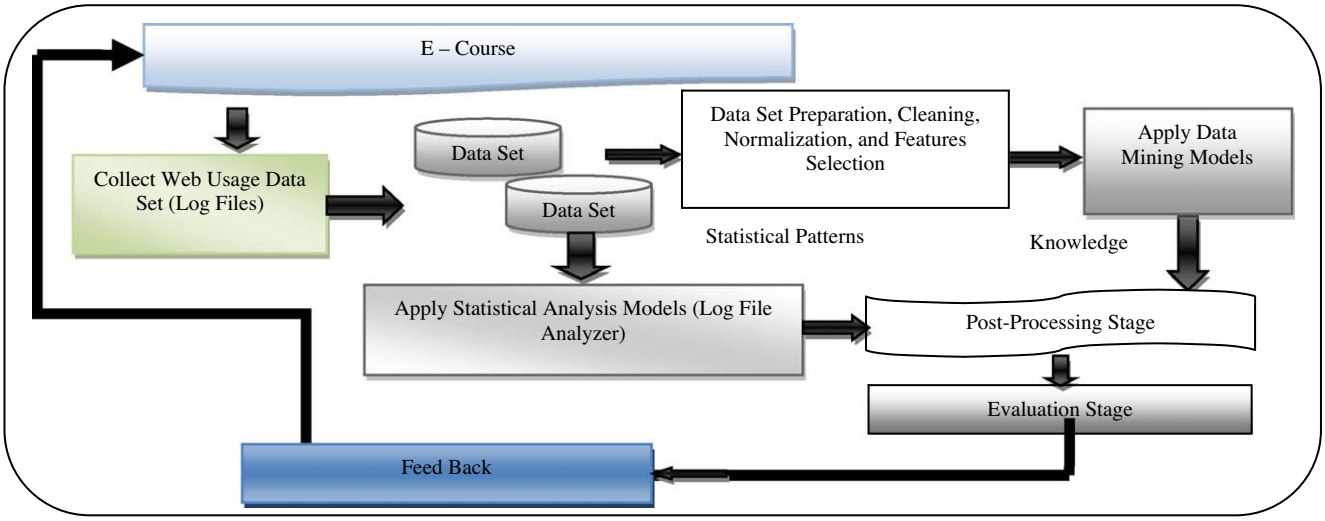

Fig. 1. The proposed system for evaluating and adapting e-learning systems.

2. Survey on soft computing in e-learning

2.1. Artificial Neural Networks and Evolutionary Computation Models (Minaei-Bidgoli & Punch, 2003; Mizue & Toshio, 2001; Teng, Lin, Cheng, & Heh, 2004)

A navigation support system based on an Artificial Neural Networks (more precisely, a Multi-Layer Perceptron, or MLP) was put forward in Mizue and Toshio (2001) to decide on the appropriate navigation strategies. The Neural Network was used as a navigation strategy decision module in the system. Evaluation has validated the knowledge learned by the Neural Network and the level of effectiveness of the navigation strategy. In Minaei-Bidgoli and Punch (2003), Teng et al. (2004), evolutionary algorithms were used to evaluate the students' learning behavior. A combination of multiple classifiers (CMC), for the classification of students and the prediction of their final grades, based on features extracted from logged data in an education web-based system, was described in Minaei-Bidgoli and Punch (2003). The classification and prediction accuracies are improved through the weighting of the data feature vectors using a Genetic Algorithm. In Teng et al. (2004), we find a random code generation and mutation process suggested as a method to examine the comprehension ability of students.

2.2. Graph and/or tree theory Carchiolo et al., 2003, Chang & Wang, 2001, Chang, Hung, & Shih, 2003, Grieser et al., 2002, Jantke et al., 2004, Liang et al., 2000, Licchelli et al., 2004, Tsai et al.., 2001, Wang et al., 2002

The student's needs and capabilities and on the teacher's profile, were described in Carchiolo et al., (2003). Personalized learning paths in the courses were modeled using graph theory. In Liang, Ziarco, and Maguire (2000), Licchelli, Basile, Di Mauro, and Esposito (2004), DT as classification models were applied. A discussion of the implementation of the Distance Learning Algorithm (DLA), which uses Rough Set theory to find general decision rules, was presented by Liang et al. (2000): a DT was used to adequate the original algorithm to distance learning issues. On the basis of the obtained results, the instructor might consider the reorganization of the course materials. System architecture for mining learners' on-line behavior patterns was put forward in Chang and Wang (2001). A framework for the integration of traditional web log mining algorithms with pedagogical meanings of web pages was presented. The approach is based on the definition of an e-learning system concept hierarchy and the sequential patterns of the pages shown to users. An e-learning model for the personalization of courses, based both on Also in Licchelli et al. (2004), an automatic tool, based on the students' learning performance and communication preferences, for the generation and discovery of simple student models was described, with the ultimate goal of creating a personalized education environment. The approach was based on the PART algorithm, which produces rules from pruned partial DTs. In Wang, Bao, Yu, and Wang (2002), a tool that can help trace deficiencies in students' understanding was presented. It resorts to a tree abstract data type (ADT), built from the concepts covered in a lab, lecture, or course. Once the tree ADT is created, each node can be associated with different entities such as student performance, class performance, or lab development. Using this tool, a teacher could help students by discovering concepts that needed additional coverage, while students might discover concepts for which they would need to spend additional working time. A tool to perform a quantitative analysis based on students' learning performance was introduced in Chang, Hung, and Shih (2003). In Grieser et al. (2002), Jantke, Grieser, and Lange (2004), Tsai, Tseng, and Lin (2001), personalized web-based learning systems were defined, applying web usage mining techniques to personalized recommendation services. The approach is based on a web page classification method, which uses attribute-oriented induction according to related domain knowledge shown by a concept hierarchy tree.

2.3. Association rules for classification

These techniques were applied in the areas of learning recommendation systems (Chu, Chang, & Hsia, 2003; Yoo, Yoo, Lance, & Hankins, 2006; Zaïane, Luo, 2001), learning material organization (Traynor & Gibson, 2005), student learning assessments (Hwang, Hsiao, & Tseng, 2003; Kumar, 2003; Minaei-Bidgoli, Tan, & Punch, 2004; Resende & Pires, 2001, 2002), course adaptation to the students' behavior (Costabile, De Angeli, Roselli, Lanzilotti, & Plantamura, 2003; Hsu, Chen, & Tai, 2003; Markellou, Mousourouli, Spiros, & Tsakalidis, 2005), and evaluation of educational web sites (Dos Dos Dos Santos & Becker, 2003). Data mining techniques such as association rule mining, and intersession and intra-session frequent pattern mining, were applied in Yoo et al. (2006), Zaïane and Luo (2001) to extract useful patterns that might help educators, educational managers, and web masters to evaluate and interpret on-line course activities. A similar approach can be found in