Chapter 12
Validation of Clustering Techniques for Student Grouping in Intelligent E-learning Systems

Danuta Zakrzewska
Technical University of Lodz, Poland

ABSTRACT
An intelligent e-learning system should be enhanced with personalization features that enable it to be tailored to different students' needs. The individual requirements of learners may depend on their characteristic traits, such as dominant learning styles. Finding groups of students with similar preferences can help when systems are being adjusted for individual requirements. The performance of personalized educational systems is dependant upon the number and quality of student clusters obtained. In this chapter the application of clustering techniques for grouping students according to their learning style preferences is considered. Such groups are evaluated by disparate validation criteria and the usage of different validation techniques is discussed. Experiments were conducted for different sets of real and artificially generated data on students' learning styles and the indices: Dunn's Index, Davies-Bouldin Index, SD Validity Index as well as the S_Dbw Validity Index are compared. From the experiment results some indications concerning the best validating criteria, as well as optimal clustering schema, are presented.

INTRODUCTION
In designing intelligent web based educational systems different student needs and preferences should be taken into consideration. Personalization of a system usually results in an increase in its effectiveness, which can be measured by the degree to which learning outcomes are achieved. However, taking into account the individual requirements of each learner and adjusting the system to their needs may be very costly in cases of large numbers of students. Finding groups of learners with similar preferences seems to be a good solution which allows for differentiating the system in compliance with the needs of group members. Learner preferences depend on characteristic traits such as cognitive features, including dominant learning styles. Students modeled by
learning style dimensions can be assigned into groups by using unsupervised classification. The performance of educational software depends on the quality of obtained clusters, which is connected with the degree of similarity between students who belong to the same groups. The proper choice of clustering algorithm is crucial for receiving the expected results. Evaluation of the applied clustering technique can be performed through validation of the obtained student clusters taking into account the resemblances between their members.

The aim of the chapter is to compare different validation techniques of cluster analysis for grouping students characterized by their dominant learning styles. Different validation criteria applied for disparate clustering algorithms are considered. On the basis of experiments which use real student data as well as artificially generated datasets, certain indications concerning the most appropriate validating criteria and the best clustering schema are presented.

The chapter is organized as follows. A research review concerning intelligent e-learning systems, student grouping as well as cluster analysis and validation techniques is presented in the next section. Following this, the clustering techniques and validity criteria that have been applied for student grouping are described in more detail. In the following section student models as well as problems connected with the application of unsupervised classification for finding student groups are presented. Then, on the basis of experiment results, some indications are given concerning the best validity criteria as well as the best clustering methods. Finally, future research directions and a conclusion are presented.

BACKGROUND

Related Work

Personalization features of e-learning systems enable educational processes to be differentiated and software to be tailored to learner needs. An individual approach to student preferences can improve the performance of a system and can increase its effectiveness at achieving the assumed learning outcomes. Personalization holds great potential to improve people’s ability to interact with information (Karger & Quan, 2004). The efficiency of personalized e-learning systems depends on the quality of individual student models, determined by their characteristic features. Brusilovsky (2001) called these ‘individual traits’, traits which are stable and usually extracted by specially designed psychological tests. However, there is little agreement on which features can and should be used in Web based learning systems (Brusilovsky, 2001). Taking the cognitive features mentioned by Brusilovsky (2001), many researchers have indicated particular personal learning styles which can determine characteristics of individual teaching paths. An overview of the research concerning building adaptive systems by using learning styles was presented by Stash, Cristea and De Bra (2004). The most frequently applied approach for student profiling consists of assigning them to previously defined groups, without the possibility of updating this assignment later (Gilbert & Han, 1999; Saarikoski et al., 2001; Triantafillou, Pomportsis & Georgiadou, 2002).

Application of cluster analysis allows student groups to be dynamically created, according to their individual characteristics, in an unsupervised way. A broad review of intelligent e-learning systems which use clustering techniques is presented by Romero and Ventura (2007). Different data mining techniques which can be applied in intelligent educational software can be also found in Romero and Ventura (2006). Methods used for student grouping that should be mentioned are descriptive statistics for building groups in collaborative learning environments, which were proposed by Alfonseca et al. (2006), or fuzzy logic for finding student models on the basis of their learning activities and interaction history (Xu, Wang & Su, 2002). Garcia et al. (2007), in turn,