

Applying Rough Set Theory for Student Clustering on Assessment Datasets

Suhirman^{1,2}, Jasni Mohamad Zain²

¹University Technology Yogyakarta
Yogyakarta, Indonesia

²Faculty of Computer System and Software Engineering
Universiti Malaysia Pahang Gambang
Kuantan Pahang, Malaysia
suhirman@uty.ac.id, jasni@ump.edu.my

Abstract - Assessment is the last session of a lecture in college. There are several components that form the basis of calculations on the scoring end. Data clustering under rough set theory can be considered as a technique for data mining. A technique to select a clustering attribute based on rough set theory is presented. Max-Max Roughness is applied to select the clustering attributes. To find meaningful clusters from a dataset, clustering attribute is conducted so that attributes within the clusters made will have a high correlation or high interdependence to each other while the attributes in other clusters are less correlated or more independent. Dataset is taken from a survey of 284 software engineering students. Data are taken on 5th semester students majoring in Information Engineering University of Technology of Yogyakarta Indonesia. Assessment consists of five components, namely two tasks, presentations, midterms and final exams. This assessment was conducted in 2011. The evaluation criteria used range from [0-100]. Student name, age, race, and attendance are not required in this assessment. In this study, we show how to determine the dominant attributes of a set of attributes of an assessment list by using the rough set theory (Max-Max roughness). The results obtained can potentially contribute to give a recommendation in awarding the final grade of a course more quickly and accurately.

IndexTerms - Assessment, Clustering, Rough set theory, Attributes

1. Introduction

Education is the foundation for achieving sustainable development. Concerning with the importance of this kind of education, the key aspect is needed on the measuring achievement levels in higher environmental education [1]. Higher education institutions are overwhelmed with huge amounts of information regarding student's enrollment, number of courses completed, achievement in each course, performance indicators and other data. This has led to an increasingly complex analysis process of the growing volume of data and to the incapability to take decisions regarding curricula reform and restructuring. On the other side, educational data mining is a growing field aiming at discovering knowledge from student's data in order to thoroughly understand the learning process and take appropriate actions to improve the student's performance and the quality of the courses delivery [2].

The techniques of grouping data of an object are similar, both to deal with certain conditions and in conditions of uncertainty. especially with high complexity, speeds and stability will be a problem in itself. The data collection can be described as vague and uncertain. Obtained clusters and applied queries do not necessarily have boundaries. Rough set theory was developed as a mathematical tool for dealing with vagueness and uncertainty. It is successfully applied in various tasks, the selection / attribute extraction, synthesis and classification rules, knowledge discovery, etc. Tolerance rough set model employing a tolerance relation is not an equivalence relationship in the original model of rough sets [3].

Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification. To make learning process more effective, the educational systems deliver content adapted to specific user's needs. Adequate personalization requires the domain of learning to be described explicitly in a particular detail, involving relationships between knowledge elements referred to as concepts [4].

Clustering is a mostly unsupervised procedure and the majority of the clustering algorithms depend on certain assumptions in order to define the subgroups present in a data set. As a consequence, in most applications the resulting clustering scheme requires some sort of evaluation as regards to its validity.

2. Related Works

One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in examination, detection of abnormal values in the result sheets of the students, prediction about

