Knowledge Discovery in Higher Educational Big Dataset

Robab Saadatdoost, Department of Information Systems, Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia & Department of Computer and Information Technology, Parand Branch, Islamic Azad University, Parand, Iran

Alex Tze Hiang Sim, Department of Information Systems, Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia

Jee Mei Hee, Department of Educational Foundation, Faculty of Education, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia

Hosein Jafarkarimi, Department of Computer, Damavand Branch, Islamic Azad University, Damavand, Iran

ABSTRACT

This paper seeks to address one of the current issues of large organizations: it is rapid growth of data without any quick way to extract worthwhile and hidden knowledge from considerable huge volume of data. It seems that management of higher education institutes interest to the best method for solving this problem and making a good decisions and strategies. Contrast to the authors’ initial sample consisted of five medical universities of Tehran, in this paper a large sample was chosen because of the expected valuable discovered knowledge. This data was collected from 65 universities of Iran based 18 years. The data is in Persian. The present paper confirms the authors’ previous findings and contributes additional discovered knowledge related to the major group of program with different geographical, the main factor of sharp increase in the number of students and preferred learning style, study mode and programs and considerable growth of female students after 1996. The findings of this study have a number of important implications for future planning of higher education to improve ranking of universities. Another important practical implication is that other researchers can use them in their studies on higher education.

Keywords: Data Mining, Higher Education Institutes, Knowledge Discovery, Management, Weka

INTRODUCTION

One of the most significant current discussions in higher education institutes is gathering data and discovering knowledge to use in educational purposes. Data is an important component in the educational institutes because of hidden knowledge and relations within stored data. Previous studies have reported relationships in databases (Fayyad, Piatetsky-Shapiro, & Smyth, 1996), (Prather et al., 1997; Larose, 2005) Recent developments in knowledge discovery and data mining techniques (Washio & Luo, 2013) heightened the need of organizations to

DOI: 10.4018/ijirr.2013010104
use these techniques to extract knowledge from large volume of data in these organizations (Fayyad et al., 1996). Recently, there has been a rapid growth of data volume (Chen, Yin, Yang & Hung, 2013; Ezhilvathani & Raja, 2013; Govindaraju, Lloyd, Wang, Lin, & Manocha, 2004; Lee, Lu, Ling, & Ko, 1999; Sheth & Shah, 2012; Wang & Shi, 2012), which needs to mine. However, this rapid increase in data volume is having a serious problem with the knowledge discovery (Fayyad et al., 1996; Gama, 2010) for strategic and management values. So far, however, there has been little discussion about knowledge discovery in higher education with high volume of data conducting to worthwhile knowledge for management. The aim of this study was to validate the proposed methodology for knowledge discovery in large data set and extract more knowledge for higher education management. This paper first gives a brief overview of the recent history of knowledge discovery in higher education. In this paper, we applied data mining techniques on data of universities. The data was accumulated from 1988 to 2005 on 65 universities of Iran, Middle East. The essay has been organized based on our previous paper in this area but in the large scale for validating our methodology and extracting more facts about higher education in Iran from 1988 to 2005.

LITERATURE REVIEW

In recent years, there has been a limited literature on data mining in education. Romero did a survey from 1995 to 2005 on educational data mining (Romero & Ventura, 2007). He reported data can come from two types of educational systems: Traditional classroom and distance education (Romero & Ventura, 2007). Besides he pointed educational data mining is an immature research area and it is essential more specialized work.

To determine the use of data mining in educational environment, Vranic, Pintar, et al (2007) presented how data mining techniques can be used in the academic environment to develop some aspects of education quality.

In 2009 Vialardi did a research on the use of data mining techniques for recommendation system used by students for decision making on their academic programs. The main point of this research is on extracting knowledge form students’ performance (Vialardi, Bravo, Shafti, & Ortigosa, 2009). Their work contains data preprocessing and pattern extraction and evaluation to discover patterns that can be used for recommendation systems intended for students.

In other case, Baepler and Murdoch pointed how data mining techniques and their results can be useful to those who are in the education domain (Baepler & Murdoch, 2010). They reported many historians have argued that common data mining techniques which used in higher education are clustering, classification, visualization, and association analysis (Baepler & Murdoch, 2010). Their work is largely expletory and focuses more on the prospective of these analyses.

In 2004, Delavari and et.al published a paper in which they proposed a new model for using data mining technology in higher educational (Delavari, Shirazi, & Beikzadeh, 2004). They reported the lack of deep knowledge at educational system. Discovering the hidden patterns and associations by data mining techniques can decrease this knowledge gap in higher educational systems. Their new model is for the purpose of creating high efficiency and effectiveness in higher education systems and can be used to identify which part of higher education process can be improved by this model and how (Delavari, et al., 2004). Delavari, Beikzadeh et al designed a roadmap draft for the application of data mining in higher education system (Delavari, Beikzadeh, & Amnuaisuk, 2005) and based on the roadmap, HEI decision makers are enabled to detect and enhance the existing processes using appropriate data mining techniques and procedures.

In addition, Mugla University, Turkey has developed an advanced system named MUSKUP (Mugla University Student Knowledge Discovery Unit Program) to analyze
Related Content

Domain Specific Custom Search for Quicker Information Retrieval
www.irma-international.org/article/domain-specific-custom-search-for-quicker-information-retrieval/101358/

On Knowledge-Enhanced Document Clustering
www.irma-international.org/article/knowledge-enhanced-document-clustering/78315/

Attitudes of Librarians in Selected Nigerian Universities towards Digital Libraries in E-Learning
www.irma-international.org/chapter/attitudes-librarians-selected-nigerian-universities/73774/

Hamming Distance based Clustering Algorithm
www.irma-international.org/article/hamming-distance-based-clustering-algorithm/72704/

Querying of Time Series for Big Data Analytics
www.irma-international.org/chapter/querying-of-time-series-for-big-data-analytics/138705/