Chapter 10
Knowledge Discovery for Large Databases in Education Institutes

Robab Saadatdoost
Islamic Azad University, Parand Branch, Iran

Alex Tze Hiang Sim
Universiti Teknologi Malaysia, Malaysia

Hosein Jafarkarimi
Islamic Azad University, Damavand Branch, Iran

Jee Mei Hee
Universiti Teknologi Malaysia, Malaysia

ABSTRACT

This project presents the patterns and relations between attributes of Iran Higher Education data gained from the use of data mining techniques to discover knowledge and use them in decision making system of IHE. Large dataset of IHE is difficult to analysis and display, since they are significant for decision making in IHE. This study utilized the famous data mining software, Weka and SOM to mine and visualize IHE data. In order to discover worthwhile patterns, we used clustering techniques and visualized the results. The selected dataset includes data of five medical university of Tehran as a small data set and Ministry of Science - Research and Technology's universities as a larger data set. Knowledge discovery and visualization are necessary for analyzing of these datasets. Our analysis reveals some knowledge in higher education aspect related to program of study, degree in each program, learning style, study mode and other IHE attributes. This study helps to IHE to discover knowledge in a visualize way; our results can be focused more by experts in higher education field to assess and evaluate more.

INTRODUCTION

Describe the general perspective of the chapter. End by specifically stating the objectives of the chapter.

Nowadays each organization deals with some data about their area and during time it increases, one of these organizations that includes big volume of data is higher education institute that has always held much data about universities, students and teachers. Thus, it is possible for us to discover some worthwhile relations or patterns that can be useful for making decision. For examples, planning the future development of a university and identifying the cluster of students who required more attentions. Management faces many challenges particularly in planning and for this purpose it needs some facts extracted from data. In our rapidly changing world, every year we accumulate data and add it to our data sets so after several years we will have a massive databank, in this environment every year our data volume increases so we need some tools to analysis this data for extracting some valuable outcome from it. Data mining has many techniques that can apply and facilitate analyzing of data.

Data mining has many definitions and almost all of them point to the discovery of patterns, and the analysis of some relations between variables in data. It does not limit to collecting and managing data; it also includes analysis. In this study, we intend to use historical data as the basis of discovering hidden relations. We intend to perform data mining techniques to discover knowledge. There are some examples, such as:

- Mining of statistical data of one university to discover successful students (Venus Shokorniaz & Akbari, 2008).
- Mining on students and discovering groups of students those are available from the data and their relations (Yghini, Akbari, & Sharifi, 2008).

In this project, we applied data mining techniques on data related to Iran Higher Education Institute to discover some relations and patterns that are useful in decision making system of higher education.

We have chosen this topic because of government and management of universities need to plan before an event occurrence. We face huge data and need to analysis them to reach some knowledge. For this purpose, we need some techniques that data mining helps us on this way, data mining has two common techniques that are classification and clustering. In this project we study about these techniques and choose one of them in our project.

Clustering is a data mining technique that is a division of data elements into groups of similar objects without advance knowledge of the group definitions. In addition, it is a tool for data analysis, which solves classification problems. In clustering, there are strong associations between members of each group and according to the type of clustering; Clustering algorithms have 4 types: exclusive, overlapping, Hierarchical, and Probabilistic. We may find some associations between different groups. Some of these associations are strong and some of them are weak. For example exclusive algorithm has weak association and overlapping has strong (Berkhin, 2006). Clustering is a discovery tool that may discover associations and patterns in data which is not previously obvious. In short: clustering attempts to find some groups of elements, based on some similarities (Ong, 2000). One of the cluster analyses is SOM (self-organizing method) that is one of the most important algorithms in data visualization and exploration. Visualization transforms from the invisible to the visible (Alhoniemi et al., 2002, 2003). SOM is a particular type of neural network used in clustering. It maps high dimensional input onto two dimensional.

Related Content

A Hybrid Recommendation Approach for Personalized Retrieval of Research Articles

Olatunji Mumini Omisore (2014). *International Journal of Information Retrieval Research (pp. 42-60).*

www.irma-international.org/article/a-hybrid-recommendation-approach-for-personalized-retrieval-of-research-articles/127363

Visualisation of Blockchain Concepts

Divi Anand, Isha Kaushik, Jasmehar Singh Mann, Ritu Punhaniand Ishika Punhani (2023). *Applying AI-Based IoT Systems to Simulation-Based Information Retrieval (pp. 32-52).*

www.irma-international.org/chapter/visualisation-of-blockchain-concepts/322848

A Study on Web Searching: Overlap and Distance of the Search Engine Results

Shanfeng Zhu, Xiaotie Deng, Qizhi Fangand Weimin Zhang (2004). *Intelligent Agents for Data Mining and Information Retrieval (pp. 208-225).*

www.irma-international.org/chapter/study-web-searching/24165

Mid-Level Image Descriptors

Jean Martinetand Ismail Elsayad (2012). *Intelligent Multimedia Databases and Information Retrieval: Advancing Applications and Technologies (pp. 46-60).*

www.irma-international.org/chapter/mid-level-image-descriptors/59951

Customized Recommendation Mechanism Based on Web Data Mining and Case-Based Reasoning

Jin Sung Kim (2004). *Intelligent Agents for Data Mining and Information Retrieval (pp. 47-64).*

www.irma-international.org/chapter/customized-recommendation-mechanism-based-web/24155