Expert Systems with Applications 38 (2011) 755–763

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Clustering and ranking university majors using data mining and AHP algorithms: A case study in Iran

A. Rad [a], B. Naderi [a,*], M. Soltani [b]

[a] Department of Industrial Engineering, Amirkabir University of Technology, Tehran, Iran
[b] Department of Management (EMBA), Islamic Azad University of Bonab, Iran

ARTICLE INFO

Keywords:
Data mining
Clustering
k-means algorithm
Multi-criteria decision making
Analytic hierarchy process
University major ranking problem

ABSTRACT

Although all university majors are prominent, and the necessity of their presence is of no question, they might not have the same priority basis considering different resources and strategies that could be spotted for a country. Their priorities likely change as the time goes by; that is, different majors are desirable at different time. If the government is informed of which majors could tackle today existing problems of world and its country, it surely more esteems those majors. This paper considers the problem of clustering and ranking university majors in Iran. To do so, a model is presented to clarify the procedure. Eight different criteria are determined, and 177 existing university majors are compared on these criteria. First, by *k*-means algorithm, university majors are clustered based on similarities and differences. Then, by AHP algorithm, we rank university majors.

1. Introduction

University major choice is an important decision to make for anybody seeking professional/higher education. It is a decision that will influence the way people look at the world around themselves (Porter & Umbach, 2006). The future occupation of people is closely related to their education. Given this importance, it is always of interest to find the guidance in collaboration with making aforementioned choices about which major to select. It is known that students should draw on available resources to ultimately pick a path that is right for them (Boudarbat, 2008). Nowadays, due to the creation of numerous undergraduate majors, the need for having a more precise approach becomes each time more necessary. Besides individual reasons, governments could be another client of university major choice. They might look for a way to supply their professional labors as one of the most influential factors in its national future. To manage this and to find which majors are of more important in future, they require a systematic approach to have more deep view about majors. For example, they entail to know areas each major affects, how majors can affect, to what extent each major is influential in a given area. Although all university majors are prominent, and the necessity of their presence is of no question, they might not have the same priority basis considering different strategies that could be spotted for a country. Their priorities likely change as the time goes by; that is, different majors are desirable at different time. If the government is informed of which majors could tackle today existing problems of world and its country, it surely more esteems those majors. By more investing on those majors or providing greater grants for those studying the majors, they intend to motivate more talented students to study these majors.

Therefore, with reference to the given explanations, it is a handy contribution to construct a model for such a decision-making process. To this end, we define eight different main specialization groups (or MSG). We first group university majors based on their similarities and differences which are obtained by their magnitude of influence on MSGs. The values of different major group can then be calculated and evaluated to provide useful decisional information for the government to rationally exploit resources. Among available grouping methods, data mining approaches have been attracted more attention. Given different data mining models, clustering is regarded as the art of systematically finding groups in a data set (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). In this paper, to cluster the university majors, we utilize the *k*-means algorithm as the most widely used method that have shown many successes in different applications such as market segmentation, pattern recognition, information retrieval, and so forth (Cheung, 2003; Kuo, Ho, & Hu, 2002). Besides its high performance, it is a very popular approach for clustering because of its simplicity of implementation and fast execution.

Ranking/ordering university majors is a multi-criteria problem; that is, different criteria should be taken into account. For example, one major might be very important for industrial setting while another one is to improve social culture. Armed with this, we apply the analytic hierarchy process (or AHP) as a simple multi-criteria decision making (or MCDM) method for dealing with unstructured, multi-attribute problems. AHP was developed by Saaty (1980,

* Corresponding author.
E-mail address: bahman.naderi@aut.ac.ir (B. Naderi).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.07.029

1989) and widely studied by other authors (Bolloju, 2001; Kablan, 2004; Lipovetsky & Conklin, 2002). It consists of breaking down a complex problem into its components, which are then organized into levels in order to generate a hierarchical structure. The aim of constructing this hierarchy is to determine the impact of the lower level on an upper level, and this is achieved by paired comparisons provided by the decision maker. The hierarchical structure of the AHP model attempts to estimate the impact of each alternative on the overall objective of the hierarchy. Another advantage of the AHP is that it uses a consistency test to filter inconsistent judgments. Taking into account these advantages, many outstanding works have been published based on AHP. They include applications of AHP in different fields, such as planning, selecting a best alternative, ranking alternatives as in our case, resource allocation, resolving conflicts, optimization, etc., as well as numerical extensions of AHP (García-Cascales & Lamata, 2009; Chatzimouratidis & Pilavachi, 2009). An important bibliographic review of MCDM tools was carried out by Steuer (2003). Our objective is to employ an AHP application in the problem of ranking university majors.

Looking into the literature, there is no paper published dealing with the major choice as a nationwide problem. They almost tackle the problem as an individual assistance model. These papers usually propose regression models that guide a student to know which major is the best choice regarding her/his personal conditions, characteristics and interests (Porter & Umbach, 2006; Boudarbat, 2008; Berger, 1988; Crampton, Walstrom, & Schambach, 2006). As far as we reviewed, this paper is the first work exploring this problem as a nationwide one, and cluster university majors using a data mining method called *k*-means. Moreover, university majors are ranked by a MCDM method, called AHP algorithm.

The rest of the paper is organized as follows. Section 2 clusters the university majors. Section 3 presents the conceptual model of university majors ranking. Section 4 applies the AHP algorithm to order university majors. Section 5 concludes the paper.

2. University major ranking model

This section presents a conceptual model to describe the decision-making procedure of university major clustering and ranking. In fact, we employ a flow chart (FC) model to show whole procedure. This diagram is to clarify each step of whole procedure regardless of their details. Fig. 1 presents the FC model. The procedure could be divided into three main phases: Data gathering, Data preparation, and Decision making.

In the first phase, the list of existing university majors is solicited from Iranian Ministry of Science Research and Technology. University majors in Iran are presented in five main groups each of which covers an educational background from high schools. These five groups are: (1) Fine art, (2) Mathematics and Physics, (3) Empirical Sciences, (4) Human Sciences, (5) Foreign Languages. Finally, 177 university majors presented in Iran are identified. Then, MSGs are determined. Doing so, this paper intends to consider eight highlighted main specialization groups with due considerations to Iran's own attributes and special areas are needed in order to ease the design process of sustainable development. These eight MSGs were extracted after a review of the literature of the problem and the reports published by the local government for achieving sustainable development, and the validity and reliability of these MSGs have been verified and confirmed by a number of structured interviews. At this stage, additional rules and constraints taken from Iran's strategies and views should be considered as well. Finally, the following eight MSGs are considered as decision criteria:

1. Financial/Economical group
2. Social/Religious group
3. Industrial group
4. Political group
5. Service group
6. Agricultural group
7. Therapy/Health group
8. Environmental/Natural source group

In the second phase, regarding the data gathered in previous phase, two suitable questionnaires are designed. The first one is to compare university majors on their magnitude of influence on above-mentioned MSGs. The second one is to compare the importance/weight of each MSG for today Iran. The questionnaires are sent to several experts (64 experts) whose definition is set in this research as follows: an expert is a person who has at least a Master of Science degree in one of the official university majors along with

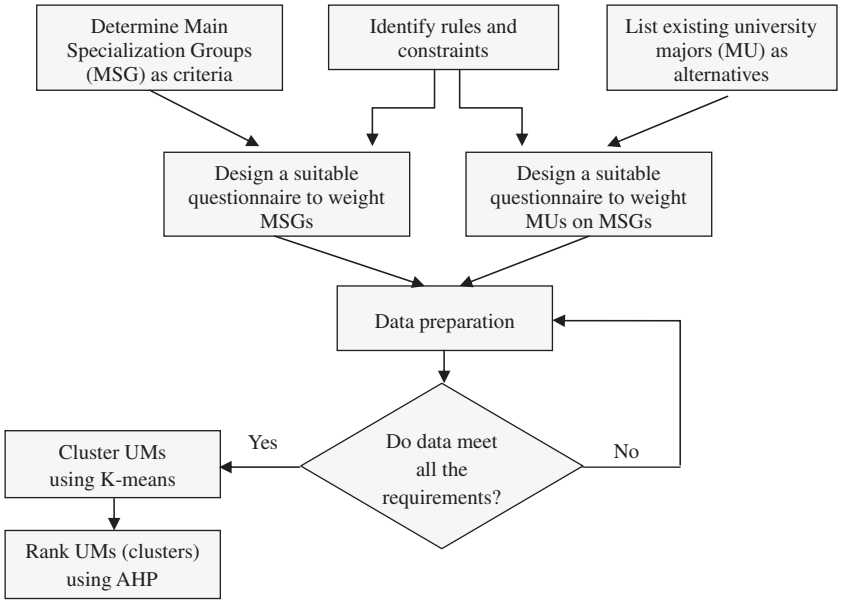

Fig. 1. General model of clustering and ranking university majors.