Solving Course Selection Problem by a Combination of Correlation Analysis and Analytic Hierarchy Process

Mohammed Al-Sarem
Department of Information Science, Taibah University, Medina, Saudi Arabia

Article Info

Article history:

Received Apr 4, 2017
Revised Jun 22, 2017
Accepted Jul 10, 2017

Keyword:

Course selection
Student preferences
Correlation analysis,
AHP method

ABSTRACT

In the universities where students have a chance to select and enroll in a particular course, they require special support to avoid the wrong combination of courses that might lead to delay their study. Analysis shows that the students' selection is mainly influenced by list of factors which we categorized them into three groups of concern: course factors, social factors, and individual factors. This paper proposed a two-phased model where the most correlated courses are generated and prioritized based on the student preferences. At this end, we have applied the multi-criteria analytic hierarchy process (MC-AHP) in order to generate the optimum set of courses from the available courses pool. To validate the model, we applied it to the data from students of the Information System Department at Taibah University, Kingdom of Saudi Arabia.

Corresponding Author:

Mohammed Al-Sarem,
Department of Information System, Taibah University,
PO box 344, Medina, KSA.
Email: mohsarem@gmail.com

1. INTRODUCTION

Course enrollment (CE) is one of the main administrative task that students faces each semester. Often, the CE process starts few week before the start of the term itself and ends a week after the start of the courses. During this period, the need to support students during selection and registration courses is increase. At a time not so long ago, students were responsible for their own choices and the faculty advisor had primarily become assisting students with the transition from high school to college [1]. Nowadays, situation is extended to include guiding students to select courses, to register in each semester, and to fulfill the degree requirement. Generally, the students aim to finish their study as soon as they can taking as many courses as possible even if this affects negatively on their performance. From this end, colleges and universities began to implement so-called academic advising affairs [2]. The academic advisory process is known as “process in which advisor and advisee enter a dynamic relationship respectful of the student's concerns” [3]. Faculty academic advising has a significant impact on a student’s academic success.

The academic advisor is responsible for: i) helping students in adaptation with specialization; ii) following-up to the level of students each semester; iii) encouraging and drawing a good study plan that ensures the improvement students' educational level; vi) determining which courses that may delay student’s graduation at the specified time; finally, v) helping students to correctly register their plan of study according to the rules of deanship of admission and registration [2]. Current work discusses the influential factors that drive students' selection. It suggests to combine the correlation analysis with the multi-criteria hierarchy analytic method. The proposed model aims to present a framework for the future e-academic advisory system. The work is organized as the follow: Section 2 presents the formulation of the course selection problem. Section 3 presents related works and the methods have been applied to solve such problem. Section 4 discusses influential factors that might drive students decision making process. Section 5 presents the work methodology, the used methods, application domain, and the data description. Section 6 describes the experiential part of this work. It presents an illustrative example showing how the model should to work; and finally, Section 7 outlines the outlines the conclusions of this research and the future work.

2. PROBLEM DESCRIPTION and FORMULATION

Let C be the set of all courses to be taught during the study plan $\mathcal{L}$ (academic curriculum plan) for obtaining a university degree. Each course $\mathrm{c} \in \mathrm{C}$ gives a number of credits $\mathrm{r_C} \in \mathbb{Z}^+$ and are might be in prerequisite relation (might find courses without any prerequisite such university required courses). The study plan $\mathcal{L}$ is divided into academic years, and each academic year is divided into semesters. Each semesters, students are faced with selecting list of courses $\mathrm{c} \in \mathrm{C_t}$ where $\mathrm{C_t}$ is list of available courses at an academic semesters in case they satisfy the courses perquisites, $\mathrm{C_t} \in \mathrm{C}$. The prerequisites are formalized as a directed acyclic graph $\mathrm{D} = (\mathrm{V_c}, \mathrm{A})$, where $\mathrm{V_c}$ represents a course, and each arc $(\mathrm{i}, \mathrm{j}) \in \mathrm{A}$ represents a precedence relation between the course i and j in case the $\mathrm{j} - \mathrm{th}$ course cannot master without taught the $\mathrm{i} - \mathrm{th}$ course. Let also $\mathcal{L}(\mathrm{C_i}) = (\mathrm{h}, \mathrm{c_t}, \mathrm{N})$ represent impact of a ith course on the study plan $\mathcal{L}$, where h is hierarchical level of $\mathrm{C_i}$, $\mathrm{c_t}$ is opened course in the next semesters $\mathrm{t} + \mathrm{i}$ and $\mathrm{i} = 1,2,3, \ldots, \mathrm{n}$, and N is the total opened courses in the study plan $\mathcal{L}$.

Let also the ith course $\mathrm{c_i}$ is taught by different instructors T at different time. Each semesters has an allowed academic load λ. The academic load is determined based on the student performance (the average grade point GPA) at the semester $\mathrm{t} - 1$. Let λ obeys the following regulations:

- if student's GPA at the semester $t - 1$ is less than predefined threshold θ, only the minimum course load λ_0 (a value of academic credits per semester required to consider a student as full time) is allowed to register at the semester otherwise up to the maximum course load λ_1.
- if student is expected to graduate and still at least a quite little hours to accomplish his/her study, the course load (extend course load λ_e) is extend and students are allowed to register more hours at the semester.

However, in real educational realm, in order to avoid the second above scenario, the academic workload per semester need to balance keeping the prerequisite conditions. In addition to that, if courses $\mathrm{c_i}, \mathrm{c_j}$ and $\mathrm{c_i}, \mathrm{c_j} \in \mathrm{C}$ are in prerequisite relation, then it is better if a course $\mathrm{c_i}$ is followed as close as possible by a course $\mathrm{c_j}$ [4]. Based on the aforementioned formulation, the course selection problem CSP ,now, is formulated as follows: *Finding these courses per semester that are, on one hand, meet student's preferences. On the other hand, maximize his/her graduation final grades.*

Practice shows that personalizing students' study plan according to their preferences leads to enhance their learning performance. However, with a lot of opportunities to compose the university curricula, restrictions, prerequisites and sometimes the university's rule, students may not be able to select course set that meet their needs and preferences. Furthermore, if they do not know in advance, which performance skills are challenged in the particular course, they may select/enroll in courses that are not adequate, at least, at a particular term. We defend on the idea that, providing students with suitable courses set leads to maximize their final GPA.

The course selection is also affected by other factors: instructor's reputation who give the course[5], the course difficulty [6], GPA value for the course [7], course time scheduling [8], market demand [9], peers' advices, and existing friends in a particular group/section (see Section 4).

3. RELATED WORKS

During the registration period, at an academic institution, commonly students should determine which courses will be taking or dropping within available registration system. This process provides the teaching staff and administration with clear vision about students' preferences, required class lists, and their number in each class. However, the situation, in reality, is on the opposite. The timetable committee constructs the whole time tables and then asks students to choose from the available course lists. Students, in this case, need to consult their academic advisors before access the system. In case of unavailability of the advisor or laziness to seek advice, these may cause to delay the registration process or the students make decisions depending on their own experience and the available information [6].

Indeed, the described above problem can be tackled several ways. Just as examples, we can mentioned the following approach: constraint programming (CP) [10], integer linear programming (ILP) problem [11], [12], hybrid techniques based on genetic algorithms and constraint programming [13], [14], integer programming and hybrid local search method [4], generalized quadratic assignment problem [15], and ant colony optimization meta-heuristic model [16].

In this work, we present the CSP as multi-criteria based decision problem (MCDP). Gunadhi et al., [16] proposed a decision model for course advising system on student’s need to know “what to do” and “how to do it”. At the core of the system lies the curriculum generator which customizes the study plan to each individual's needs and produces a schedule for the courses chosen. Customizing the study plan is depend on the course selection criterion. Some systems allow students to request only courses for which they have appropriate prerequisites and co-requisites [17]. In the others, the courses are suggested based on balancing the course load, frequency of the course offering, shortening the path length to graduation, students' preferences and their progress in the program [18], [19].

Current academic systems provide information about available courses and professors who will teach them, sections, number of students in each section, and schedule. However, information about students' previous progress from current/past enrollment is usually ignored even though such information are priceless treasure in finding interdependent courses. In this direction, the educational data mining methods have been successfully applied. Association rules e.g., are used as a way to seek dependency among courses of a curriculum plan [20], [21], [22]. The course characteristics similarities of former students' study were used in optimizing curricula of current students [7], [23].

4. INFLUENTIAL FACTORS on STUDENTS SELECTION

In the universities where students have a chance to select and enroll in a particular course, selecting the optimum set of courses from the available courses pool is a high risk decision-making situation because the cumulative impact will effect negatively/ positively on the students' performance progress, their expected graduate date and the final GPA as well as their career direction and future employment opportunities. As mentioned before, course selection process is influenced several factors. Analysis the research literature and the conducted questionnaire, we summarize these factors into three main groups of concerns: (i) *course factors*, (ii) *social factors*, and (iii) *individual factors*. Indeed, these groups is decomposed into sub-groups which influence on the whole decision-making process. Since different courses are selected with different preferences and objectives, the decision process must take all these factors concurrently (see Figure 1 below).

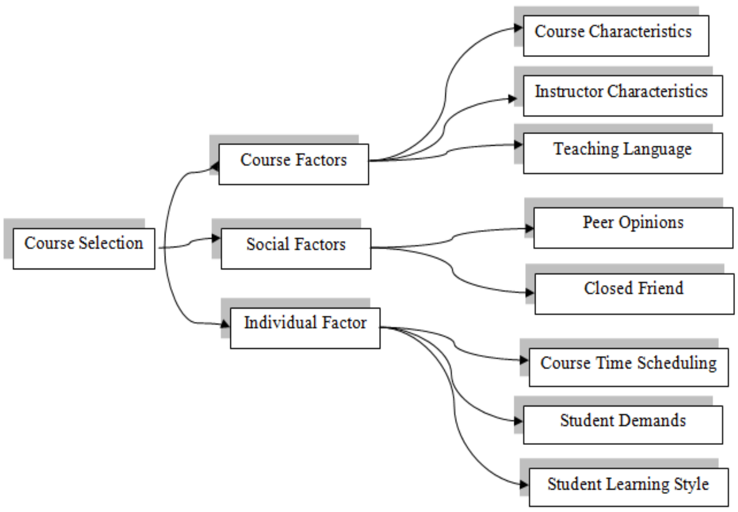

Figure 1. Infuential factors on students' course selection

Next, we discuss the impact of these factors on students' decision-making process and show how will they engage in the proposed approach. Table 1 gives a brief description of decision attributes that are used for driving the selection process.

Table 1. Description of criteria and decision attributes used for selecting a course

Criteria	Decision attributes	Refers to:
Course factors	Course characteristics	Course credit hours, Distance between a course and its prerequisites, Student competence for a given course
	Instructor characteristics	Personal instructor characteristics, Instructor assessment approach, Instructor lecturing style
Social factors	Teaching language	Course teaching language
	Peer opinions	Peers' feedback
	Closed Friend	Existing in the class a closed friend
	Campus Location	Location of the class room, Campus location
Individual factors	Course time scheduling	Time when student attend the class
	Student demands	Student's interest in a course, job opportunities, Local labor.
	Learning style	A way or an approach a student follows in the course of learning.

4.1. Course Characteristics
- Course Characteristics

The questionnaire results show that students' choices regarding course characteristics are depend mainly on the difficulty of the course, course weight (course credit hour), distance between a course c_i and its prerequisites, and student competence for a given course.

Difficulty- refers to complexity level of a course taking in consideration the grades of every student who passed that course successfully to the grades of all students who follow the same curriculum plan. Logically, a course with high $diff(C_i)$ is considered as difficult course, otherwise it is easy.

$$diff(c_i) = 1 - \left(\frac{\sum_{k\in C_i} \sum_{i=1}^{k} g'_{i,k}}{\sum_{k\in C_i} \sum_{i=1}^{k} g_{i,k}} * \frac{m}{n}\right) \tag{1}$$

where, C_i- is the ith course in the curriculum plan, $g'_{i,k}$- is GPA of a student who passed a course C_i successfully from the first attempt, $g_{i,k}$- is GPA of the student who take the course C_i, m- number of student who passed the course C_i from the first attempt, and n- is number of students who follow the same curriculum plan and take the course C_i.

Distance between two courses C_i and C_j taught by a student s is defined as the Euclidean distance of the hierarchical level h at where the courses C_i and $C_j$are being taught.

$$dis\ (C_i, C_j) = \begin{cases} \sqrt[2]{\left(C_j^h - C_i^h\right)^2 + \left(C_{S_j}^h - C_{S_i}^h\right)^2}, & C_i \xrightarrow{prerequisite} C_j \\ 1 & , otherwise \end{cases} \tag{2}$$

where, C_i^h and C_j^h- is the hierarchical level h at where the courses C_i and $C_j^h$are being taught respectively, $C_{S_i}^h$ and $C_{S_j}^h$ - is the academic semester where courses C_i and C_j^h are being taught.

Competence represents student's ability to study a course based on the grades he has obtained in the prerequisites.

$$Competence(c_i^s) = \begin{cases} 1 & , C_i\ has\ not\ prerequisite \\ \sum_{j=1, j\in \mathcal{L}}^{k} n_{C_j}^S * diff(c_i) * dis(C_i, C_j) \Big/ W_i * g_i^s, & otherwise \end{cases} \tag{3}$$

where, g_i^s- is the current GPA grades of student $s, diff(c_i)$ - is difficulty of the course c_i, $dis(C_i, C_j)$- is distance between a course C_j(prerequisite course) and course C_i, W_i- is credit hours of course $C_i$and, n_i^S- is number of attempts student S was enrolled in course C_j.

- Instructor Characteristics

Although, the course characteristics have a significant impact on students' enrollment decision, practice shows that the instructor characteristics also play important role on the future decision to enroll in those courses taught by this instructor [5], [24] and on how useful the course can be [25]. Nowadays, majority of universities provide online system to collect students' feedback for all offered courses at the end of academic semester. Often, feedback takes a form of questionnaire or survey which contain a series of items that are ranked on a five-points Likert-scale. The questionnaire/survey items address the question about personal instructor characteristics, course value presented by the instructor, instructor assessment approach, and instructor lecturing style. Researchers, such as, [24], [26] noted that students prefer to take courses with teachers who are enthusiastic, well spoken, knowledgeable, caring, and helpful. Beggs et al., [27] found that the quality of a course presented by the instructor has a large affect on whether a student chooses to enroll in a class. Although questionnaire results show beside the quality of the course, both the instructor assessment approach [28], and instructor lecturing style [24], [26] are critical factors in course enrollment.

- Teaching Language

Several researchers considered language as a significant factor not only in learning process but also in their motivation to learn [29], [30]. According to Coleman [31] the use of a common language allows, on one hand, efficient exchange of ideas, on the other hand, facilitates communication skills.

Nowadays, major of universities present course contents in English even if it is not the official/ native language. The reason behind this choice is that English has a positive impact on modernization, and on the quality of learners' experience [31]. However, students prefer to deal with instructors who share the same native language or with course content that is written in the native language even if they speak and understand English.

4.2. Social Factors

It is obvious that student's preferences are influenced directly or indirectly by peers and friends opinions. Their influences are clear in shaping and molding the course of an individual life [32]. Peer influence is more observable in friendship [33] which is represented as succumbing to the views and opinions of the peers, making a decision based on peer's advice, or just listening to the peer before listening to their teacher and advisors is a form of such influence [34]. Naz et al., [32] found that peer and friends have a positive role in selection of subjects, selection of a class and laboratory.

Analysis the feedback of students of department of Information System at Taibah University (Table 2), the majority of students (57.1%) are agree that their selection is dependent on the received advice from their peers or friends, (55.5%) prefer to enroll in a course if some of their friends are also enrolled in the same course, and (74.6%) indicated that their opinion about instructors are influenced by peers' and friends' opinions. Generally speak, majority of students are agree that their selection is influenced by advice of their peers and friends.

Table 2. Students' preferences respect peer's/friend's opinion

Question	Percentage				
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
My choice of course is mainly depend on advice of my peers/friends	3.2	11.1%	28.6%	31.7%	25.4%
I prefer to enroll in a course if some of my friends are enrolled in it also	7.9%	9.5%	27%	22.2%	33.3%
Enrollment in a course which is taught by an instructor is depend on peers' /friends' opinions about the instructor	3.2%	3.2%	19%	39.7%	34.9%
Total Impact of peer's/friend's advice on course selection	1.3%	4.2%	19.8%	33.2%	41.5%

4.3. Individual Factor

- Course Time schedule

Although student preference respect course time schedule does not play a role in selection process of full-time student, students have made decisions to take a course, or to not take a course, based on the fact of whether or not it fits into their schedule [35], [8].

Table 3. Students' preferences course time schedule

Question	Percentage				
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Choosing the scheduling times of courses have helped me to pass them successfully.	7.9%	12.7%	31.7%	20.6%	27%
Engagement students in scheduling courses time enhances their motivation to study	13.8%	12.7%	20%	33.8%	19.7%
Total Impact of course time schedule on course selection	3.1%	7.2%	23.1%	31.7%	34.9%

Table 3 illustrates that 47.6% of students found that choosing the scheduling times of courses have a positive impact on their study and lead them to pass the courses successfully, and 53.5% of students think that engagement them in scheduling courses time enhances their motivation to study.

- Student Demands

Several studies have considered interest in a course topic or subject as a driving force behind students' enrollment in classes [24], [34], [35]. The interest impact is more evident when students should to make decision to take a course from elective courses available by the collage.

According to [26], student's interest in a course is influenced by numerous factors such as subject matter, topics, and career goals. Enjoyment, job opportunities, and local labor trend are other factors that influences the course selection. Students are attracted to take a course that they think that will increase their chances to get a job.

- Learning Style

Learning style is one of the individual differences that play an important role in learning [36]. In the literature, several definitions can be found which share the same basic idea " the term learning style refers to a way or an approach a student follows in the course of learning". According learning style theory, students' interest in a course is influenced also by their preferred learning style. Adapting course content has been applied intensively in e-learning systems where the learning styles and e-media are integrated together in the design of their applications. Such integration showed a positive results in both learning styles detection and e-learning application [37].

Table 4 presents how the learning style impacts on students' decision. It also presents students' preferences regarding selecting courses. Statistical results emphasize on the fact that during making a selection decision, beside the aforementioned factors, the learning style of a student should take in consideration.

Table 4. Students' preferences respect to learning style

Question	Percentage				
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I prefer to take a course with practical nature before those with theoretical	6.3%	12.7%	36.5%	20.6%	23.8%
I prefer to enroll with maximum allowed workload in an academic level	15.9%	19%	38.1%	11.1%	15.9%
I prefer to postponed university required course to the latest level	23.8%	28.6%	28.6%	9.5%	9.5%
I prefer to finish early university required course as possible as I can	3.2%	6.3%	17.5%	36.5%	36.5%
I prefer to take the course with lowest credit hours firstly, then the highest and so on.	20.6%	20.6%	31.7%	11.1%	15.9%
I think that allowing to take a course from any level, in case I take its prerequisite, help me to success.	7.9%	12.7%	22.2%	31.7%	25.4%
I prefer to follow courses' order as it is in the curriculum plan.	1.6%	3.2%	31.7%	42.9%	20.6%

5. WORK METHODOLOGY

The core of this research is to build a decision model which aim at help and support the students during the enrollment and registration process. The model is two-phased process (Figure 2). The first phase, is similar to those presented in [23] where the most correlated courses are generated. At the second phase, the student preferences are taking in consideration. This preferences are prioritized using multi-criteria analytic hierarchy process (MC-AHP). To understand the research context and the used data, in the next sections, we present a brief explanation of the used methods, the application domain, and the gathered data.

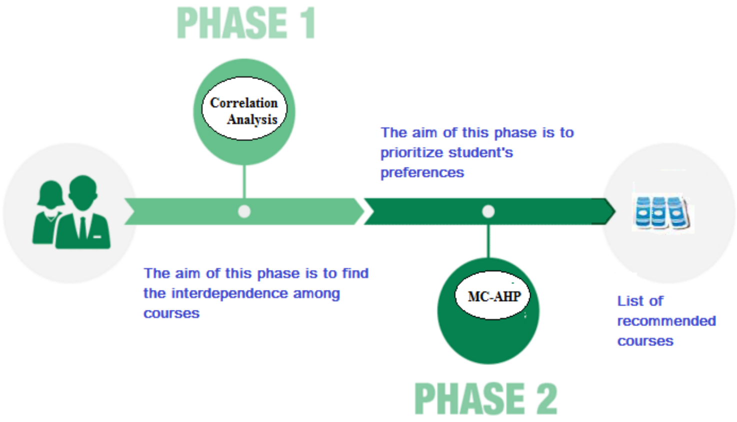

Figure 2. The Research Methodology

5.1. The Used Methods

- Correlation Analysis

Observing relationship among variables is a classical data mining task. Broadly, there are four types of relationship mining: association rule mining, correlation mining, sequential pattern mining, and causal data mining [8]. To help student in making a decision of which course he /she should to take, it is helpful finding positive or negative linear correlations between courses.

Often, to represent the correlation graphically, a scatter diagram is used where the pair of points/data (x, y) is allocated on an orthogonal coordinate system.

The linear correlation coefficient measures the strength of the linear correlation between the two variables; it reflects the consistency of the effect that a change in one variable has on the other. In educational realm, the correlation between two courses C_i and C_j as follows:

$$\mathrm{corr}(C_i, C_j) = \frac{\sum_{i=1}^{k}(g_i^{c_i} - \bar{g}^{c_i})(g_i^{c_j} - \bar{g}^{c_j})}{\sqrt{\sum_{i=1}^{k}(g_i^{c_i} - \bar{g}^{c_i})^2}\sqrt{\sum_{i=1}^{k}(g_i^{c_j} - \bar{g}^{c_j})^2}} \quad (4)$$

$g_i^{c_i}$- is grade points for the ith course
$g_i^{c_j}$- is grade points for the jth course
$\bar{g}^{c_i}$- is average grade point for all students who take the ith course
$\bar{g}^{c_j}$- is average grade point for all students who take the jth course
k- is number of students who take C_i and C_j.

The linear correlation coefficient takes value between −1 and +1:

- $\mathrm{corr}(C_i, C_j) = +1$ reflects a perfect positive linear correlation between both courses C_i and C_j.
- $\mathrm{corr}(C_i, C_j) = -1$ reflects a perfect negative linear correlation between both courses C_i and C_j.
- $\mathrm{corr}(C_i, C_j) = 0$ means that there is NO linear correlation.

if the calculated value is close to +1 or −1, we then suppose that between the two variables there is a linear correlation.

-Multi-criteria Analytic Hierarchy Process

AHP is a well-established decision making technique for dealing with multi-dimensional and often contradictory preferences of individuals [5]. The AHP ranks alternatives in view of criteria and sub-criteria (factors). In AHP, we start firstly with representing the problem with a hierarchal structure which is consists of all factors and alternatives. The hierarchal structure mainly establishes the relationships between the levels of the hierarchy order at which we place the objective (the Goal) at the top of the hierarchy, the criteria and sub-criteria at intermediate levels, and finally the alternatives are placed at the lowest level of the order.

In the second step, a pair-wise comparison judgments are carried out, for each criterion, using a nine points scale (1= equivalent,..., 9= extremely preferred to).

The result of each comparison is a matrix $(n \times n)$, where the diagonal elements a_{ii} are equal to one,$i = 1,2,\ldots,n$, and if $a_{ij} = x$, then $a_{ij} = {}^{1}/_{x}$ where $x \neq 0$.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Next step of the AHP (scoring and weighting) is to compute eigenvectors $u_{j=}(u_1, u_2, \dots, u_n)$ by solving $AW = \lambda_{max}.W$, where λ- is an eigen-value and W- is eigenvector.

The final step of AHP is to perform a consistency check (consistency ratio CR) by dividing the consistency index CI by the random index RI, where the consistency index CI is calculated as follows:$CI = (\lambda_{max} - n)/(n - 1)$, where n is the matrix size and the random index RI which is taken according Table 5.

Table 5. Average random consistency (RI) used in Saaty

Size of matrix	1	2	3	4	5	6	7	8	9	10
Random consistency	0	0	0.58	0.90	1.12	1.24	1.34	1.41	1.45	1.49

The CR is considered acceptable only if it is less than 0.1, otherwise the pair-wise comparison judgments should be reviewed and improved.

5.2. Application Domain

To show how the decision model supports the students during the enrollment and registration process, the experiential part of this work was developed in the context of department of Information System at the Taibah University, Kingdom of Saudi Arabia. Generally, study at Taibah University, as all remains universities in Saudi Arabia, are organized in two regular academic terms by year, plus a summer term which is opened only if there is a quite number of students who failed pass a course in regular terms. The regular terms are spanning four months, whilst the summer term is compressed into two months. Since 2004, the academic program is changed three times. However, number of credit hours is still the same. Each program consists of two parts:

- the preparation period where students spent one academic year at which they took a set of courses that prepare them to their future studies
- the regular period is consists of four academic years. The program consists of 14 credit hours of university requirement courses, 19 credit hours of faculty requirement courses, and 46 credit hours of department requirement courses nine of them are elective courses.

In order to pass a course, the student has to obtain at least 60 points out of 100; otherwise he will be required to attend the course again in the next academic year or in the summer term, in case the number of those students who failed to pass the course is quite enough (the decision is made based on the opinion of the vice dean of the academic affairs at each faculty). The maximum number of attempts to pass a course is depends on the student's GPA. For the student whose the GPA is less than the cut-point (2.5 out of 5) for two sequential academic terms, he will not be able to continue his/her studies. During the enrollment period, students should to register the selected courses including the name of the preferred time and group using the online enrollment system or by assistance the academic advisors. The students is eligible for enrollment a course, only if they passed the prerequisites for the said course, otherwise they are deny to take it.

5.3. Data Description

Since the academic program is changed several times, the historical records contain data from three different curricula, each of them with 42 courses separated through eight regular academic terms and four summer terms, for further details about the total number of students and classes, see Table 6. Due to of modification or changes in the curricula (sometimes, only the prerequisites of a course is changed), we focus only the curricula from 2011 to 2015 namely "new curricula".

Table 6. The number of students in each academic year and average classes to graduate students.

Academic Year	Enrolled		Graduated		Average Classes to graduate		Curricula
	Male	Female	Male	Female	Male	Female	
2010/2011	420	600	90	89	10.3	9	Old Curricula
2011/2012	600	676	95	126	10.4	9.8	
2012/2013	484	686	58	150	10.9	10.3	New Curricula
2013/2014	789	698	93	137	10.7	11	
2014/2015	828	674	111	175	10.3	11.1	Developed Curricula

The average classes to graduate students in Table 6 refers to the number of academic terms that students spend to finish their study in case the fail to pass the course from the first attempt. Figure 3 shows the increase in the required classes between both groups (male and female sections).

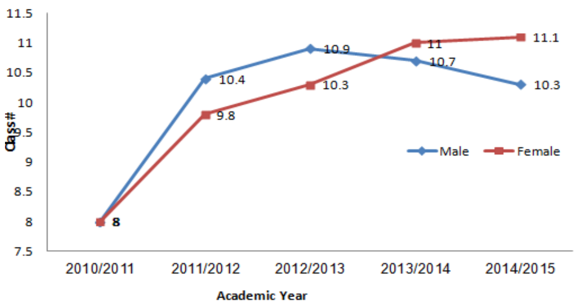

Figure 3. Average number of classes required to graduate students

The main goal of the current research is to give the student (who intends to register on a course) a recommendation based on the gained grades at the previous terms. The correlation analysis is performed based on the final grade of the students. The aim of this step is to link each course with the most correlated courses that may be effected by the selected course. Table 7 shows the used attributes and give a brief description for each of them, whilst Table 10 presents the data type of the attributes and a short statistical summary for each of them. The "Period" attribute refers to the academic term in which student should take a course. It discriminates as follows:

$$\text{Period} = \begin{cases} x \in [1-8], & x - \ is\ a\ regular\ term \\ x \in [9-12], & x - is\ a\ summer\ term \end{cases}$$

Both "Registered Credit hours" and "Gained Credit hours" attributes are used to split the data set in to training and testing set. The highest value of " Registered credit hours" denotes students has a difficulties in finishing his study, whilst the highest value of "Gained credit hours" denotes that the student is near to graduate.

Table 7. The used attributes

Attributes	Description
Course name	Identifier for each course the student is enrolled on
Course code	Identifier for each course in the university system
Course credits	Practical and theoretical workload for each course
Period	Academic term in which student should take the course
Final grade	Result obtained at the end of the term in each course
Student ID	Identifier for each student
GPA	Overview of the student's performance over time
Student gender	The gender of student who took a course
Registered credit hours	Amount of credit hours registered in the university system
Gained credit hours	Amount of credit hours already student passed

Table 8. Statistical summary of the used attributes

Attributes	Data type	Possible Value	Statistical summary	
Course name	String	Mathematics, Physic (1), Programming (1), est.	42 courses	
Course code	String	Math101, Phys101, CS102, est.	42 courses	
Course credits	Discrete	1,2,3,4	Course Credits	Percentage
			1	2%
			2	12%
			3	55%
			4	31%
Period	Discrete	1,2,..., 12.	12 terms	
Final grade	Continuous	⟨0,100⟩	Minimum	38
			Maximum	100
Student ID	String	31#####, 32#####, 33#####	Student ID	Percentage
			31#####	8.4%
			32#####,	43%
			33#####,	48.6%
GPA	Continuous	⟨0.000,5.000⟩	Mean	3.937297
			Standard Deviation	0.552337
			Sample Variance	0.305076
			Minimum	2.06
			Maximum	4.99
Student gender	String	Male, Female	Student gender	Percentage
			Male	56%
			Female	44%
Registered credit hours	Discrete	⟨0,250⟩	Mean	173.0541
			Standard Deviation	11.07285
			Sample Variance	122.6081
			Minimum	153
			Maximum	209
Gained credit hours	Discrete	⟨0,165⟩	Mean	102.797
			Standard Deviation	40.88291
			Sample Variance	1671.412
			Minimum	14
			Maximum	165

6. EXPERIMENTATION AND EVALUATION

Most course recommendation systems use students' personal data and social networking sites to find out what they like or are interested in [38], e.g., proposed to use students' grades in developing a course recommendation system. The system helps students find courses in which they can get high scores. For this purpose, data about the courses which users learned, scores that students received, and the teachers of the courses are collected.

Current work, as mentioned before, suggested a method with two-phased process. At the first phase, the most correlated courses are generated. However, as we check correlation course by course, only those courses (set of courses) that satisfy the following constrains are generated:

- Only courses with highest correlation are generated.
- The total number of credit hours of the generated courses (academic load of student A) should not exceed the maximum course load λ_1.

The Pseudo-Code for generating courses based on correlation analysis is illustrated in Pseudo-Code 1.

Pseudo-Code I: Generating courses based on correlation analysis

GPA_S^{t-1}: GPA of a student S at previous semester $t-1$
λ_1 , λ_0: maximum and minimum course load for a student respectively
c_{t-1}: course at semester $t-1$ that is passed successfully by the student, $c_{t-1} \in \mathcal{L}$
c_{t+1}: course at next semester that student can take, $c_{t+1} \in \mathcal{L}$
S: Final set of recommended courses
$r_c^{\,t+1}$: Credit hours of courses at semester $t+1$

$|S| \leftarrow \emptyset$
IF $GPA_S^{t-1} > \lambda_0$ **THEN**
    $corr(c_{t-1}, c_{t+1})$; # Calculate correlation
END IF
FOR $\forall c \in C^{t-1}$; # Loop for all courses that a student passed them successfully
  $Avrg\_corr[i] \leftarrow \frac{Competence(\, {c^i}_{t+1}) \sum_{j=1}^{n} corr({c^j}_{t-1}, {c^i}_{t+1})}{n}$; # Find average matrix

7: $C^{t+1} \leftarrow \; Select(k, Max(Avrg\_corr(c_{t-1}, c_{t+1}))$; # Select K courses with maximum correlation that are satisfy the university regulation
8: $|S| \leftarrow Package(\lambda_1, C^{t+1}, {r_c}^{t+1})$; # Generate the recommended course package

Pseudo-Code II: $Package(\lambda_1, C^{t+1}, {r_c}^{t+1})$; Generate the recommended course sets
Input: λ_1 : allowed maximum course load for a student
C^{t+1}: Set of all available courses at semester $t+1$
${r_c}^{t+1}$: Credit hours of a course $c \in C^{t+1}$
Output: Set of the recommended courses

1: Set $|R| \leftarrow \emptyset$, $|S| \leftarrow \emptyset$
2: $s \leftarrow GenerateAllCondidateSubSets(C^{t+1})$;
3: Add s to $|S|$
4: **FOR** $i \leftarrow 0$ to s
5: $r[i] \leftarrow CalculateTotalCreditHours({r_c}^{t+1})$;
6: Add $r[i]$ to $|R|$
7: **END FOR**
8: **WHILE** $\boldsymbol{s}$, $r[i] \leq \lambda_1$
9: $Candidate[i] \leftarrow \frac{\sum_{\mathrm{i=1,j=1},c \in C^{t+1}}^{C^{t+1}} ({r_{c_i}}^{t+1}) * \mathrm{corr}(\mathrm{c_i}, \mathrm{c_j})}{r[i]}$
10: $CandidateSet \leftarrow Select(Max(Candidate[i]))$; # Sort the list
12: **END WHILE**
13: **return** $CandidateSet$

Since the recommended courses are generated based on the correlation analysis, the recorded grades can be influenced, as mentioned above, by course content itself or experiencing a particular teacher's style or material of teaching [23]. Assuming that a student S passed a set of courses at semester $\mathrm{t}-1$, and the grades in the database are recorded taking in consideration the aforementioned influential factors.

Having samples of k students $\mathrm{S} = \{\mathrm{s_1}, \dots, \mathrm{s_k}\}$, who took a course $\mathrm{C_j}$ after a course $\mathrm{C_i}$, and each student $\mathrm{s_i}$ achieved $\mathrm{g_j^{c_j}}$ GPA in course $\mathrm{c_j}$ and $\mathrm{g_i^{c_i}}$ GPA in course $\mathrm{C_i}$as shown in Table 9.

Table 9. Sample of students grades at a semester t

Student ID	Course					
	ISLS101	ARAB101	MATH101	PHYS103	CS101	ENGL103
3151377	95	95	87	70	81	85
3151559	90	98	71	65	71	81
3182286	99	96	93	85	80	72
3182565	95	96	90	85	86	95
3182894	98	95	77	87	75	75
3200079	100	100	98	100	98	90
3200162	100	100	74	93	95	78
3200229	99	100	100	100	95	98

Let a student passed a set of courses as shown in Table 10, and the correlation among all curriculum courses are already calculated.

Table 10. Student grades at current semester

Student ID	Course					
	ISLS101	ARAB101	MATH101	PHYS103	CS101	ENGL103
3200237	95	95	87	70	60	85

Since Eq. (4) assumes that the grades of all courses are given, and because of absence the grades of those courses of the next semester, we propose to use the regression analysis to predict the grades of each courses based on the historical records of all its perquisite courses (see Table 11).

Table 11. Predicted grades for the courses at the next semester

Student ID	Curriculum Plan P													
	Grades at Semester t						Predicted Grades							
	ISLS 101	ARAB 101	MATH 101	PHYS1 03	CS 101	ENGL 103	IS 201	CS 202	ARAB 201	MAT H 102	IS 102	IS 221	CS 112	ENG L 104
3200237	95	95	87	70	60	85	66	86	92	52	85	70	75	75

Table 12 presents the correlation among the already taken courses and those are offered at the next semesters.

Table 12. Example for calculation demonstration.

Student ID	Curriculum Plan P									
			Next Semester $t+1$							
	Current Courses at Semester t		IS 201	CS 202	ARAB 201	MATH 102	IS 102	IS 221	CS 112	ENGL 104
3151377	ISLS101	95	0.04	0.25	0.604	0.331	0.441	0.244	0.221	0.09
	ARAB101	95	0.00	0.27	1.000	0.346	0.210	0.279	0.297	-0.01
	MATH101	87	0.15	0.63	0.526	1.000	0.536	0.447	0.489	0.47
	PHYS103	70	0.29	0.37	0.483	0.625	0.342	0.260	0.358	0.20
	CS101	60	0.34	0.46	0.486	0.640	0.286	0.574	1.00	0.49
	ENGL103	85	0.26	0.49	0.603	0.416	0.379	0.151	0.321	1.00

Table 13 presents average correlation matrix of each generated course. Steps 7-10 allow us to sort and select the top k courses based on its average correlation.

Table 13. Average correlation matrix of the courses

Student ID	Curriculum Plan P									
			Next Semester $t+1$							
	Current Courses at Semester t		IS 201	CS 202	ARAB 201	MATH 102	IS 102	IS 221	CS 112	ENGL 104
3151377	ISLS101	95	0.04	0.25	0.604	0.331	0.441	0.244	0.221	0.09
	ARAB101	95	0.00	0.27	1.000	0.346	0.210	0.279	0.297	-0.01
	MATH101	87	0.15	0.63	0.526	1.000	0.536	0.447	0.489	0.47
	PHYS103	70	0.29	0.37	0.483	0.625	0.342	0.260	0.358	0.20
	CS101	60	0.34	0.46	0.486	0.640	0.286	0.574	1.00	0.49
	ENGL103	85	0.26	0.49	0.603	0.416	0.379	0.151	0.321	1.00
		Average	0.18	0.41	0.617	0.56	0.366	0.33	0.45	0.37

To generate the set of the candidate courses package, first we should to follow the academic regulation. In our case where this research is conducted, the academic regulation determines the maximum course load λ_1 based on the student's GPA as follows:

$$\lambda_1 = \begin{cases} \lambda_1 \in [17,21] & ,\text{student is expected to graduate next semester} \\ \lambda_1 \in [12,17] & ,5 \leq \text{ GPA} \leq 2.8 \\ \lambda_1 = 12 & ,\text{GPA} \geq 2.7 \end{cases}$$

Let that the current GPA of a non-graduated student is 3.21 which means that the student has a right to register courses with total credit hours up to 17. Table 14 shows the combination of all possible sets of the eight courses. The set with the largest average correlation is suggested as a recommended package of courses for the next phase.

Table 14. Combination of all possible sets

Student ID	Curriculum Plan P								λ_1	*Overall Priority Average*	Rank
	ARAB 201	MATH 102	CS 112	CS 202	ENGL 104	IS 102	IS 221	IS 201			
Credit Hours	3	4	4	4	3	3	4	3			
~~3200237~~	ARAB-201, MATH-102, ENGL-104, IS-221, IS-201								17	0.219992771	1
	ARAB-201, MATH-102, ENGL-104, CS-112, IS-201								17	0.197847972	4
	ARAB-201, MATH-102, ENGL-104, CS-202, IS-201								17	0.213518998	2
	…								⋮	⋮	
	ARAB-201, MATH-102, CS-112, CS-202								15	0.191914578	5
	ARAB-201, MATH-102, CS-112, ENGL-104								14	0.207102686	3
	…								⋮	⋮	
	MATH-102, CS-112, CS-202								12	0.161333772	6
	MATH-102, CS-112, IS-102								12	0.16062525	7
	…								⋮		

Let the timetable of these courses are already predefined and each course (in this set) is linked with one/ many instructors in different time (see Figure 4). The next step, as shown in Figure 2, is to prioritize time, instructors, and sections of the courses based on the student preferences. As mentioned previously, the preferences are prioritized using multi-criteria analytic hierarchy process (MC-AHP) (See Section 5.a).

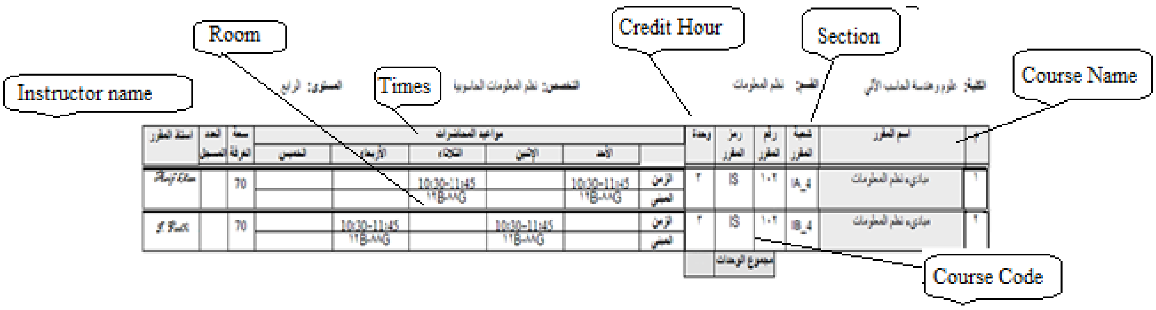

Figure 4. Snapshot of IS-102 schedule

Following the procedure of MC-AHP, there are a total of six pair-wise comparison matrix tables:

1. The pair-wise comparison matrix of the criteria relating to the goal. This is illustrated in Table 15.
2. The pair-wise comparison matrices for the five options (set of the recommended courses) regarding all the "criteria concerned", where the criteria in all levels are connected to the options.

The consistency ratios of all comparisons were less than 0.1, which indicates that the weights used are consistent.

Table 15. Pair-wise comparison matrix of the key criteria with regards to the goal

Criteria	*Course factors*	*Social factors*	*Individual factors*	***Global Priority Vector***
Course factors	1	9	3	0.67
Social factors	1/9	1	1/5	0.06
Individual factors	1/3	5	1	0.27
	$CI = 0.014606$	$CR = 0.025182 \leq 0.1$		

Tables 16 illustrates the pair-wise comparisons of the alternatives for the first offered sections of IS102 course in terms of aforementioned criteria.

Table 16. Pair-wise comparison matrix with regards to the sub-criteria

Criteria	*Course characteristics*	*Instructor characteristics*	*Teaching language*	*Local Priority Vector*	*Priority respect Global Vector*
Course characteristics	1	5	3	0.63	0.4221
Instructor characteristics	1/5	1	1/3	0.11	0.0739
Teaching language	1/3	3	1	0.26	0.1742
	$CI = 0.019357$	$CR = 0.033375 \leq 0.1$			
Criteria	*Peer opinions*	*Friendship*	*Campus Location*	*Priority Vector*	*Priority respect Global Vector*
Peer opinions	1	3	1/3	0.29	0.0174
Friendship	1/3	1	1/3	0.14	0.00084
Campus Location	3	3	1	0.57	0.0342
	$CI = 0.076$	$CR = 0.1$			
Criteria	*Course time scheduling*	*Student demands*	*Learning style*	*Priority Vector*	*Priority respect Global Vector*
Course time scheduling	**1**	**7**	**1**	**0.51**	**0.1372**
Student demands	**1/7**	**1**	**1/3**	**0.10**	**0.027**
Learning style	**1**	**3**	**1**	**0.39**	**0.1053**
	$\mathbf{CI = 0.040474}$	$\mathbf{CR = 0.069784}$			

The output of the MC-AHP algorithm is summarized in the overall priority matrix as shown in Table 17.

Table 17. Overall priority matrix

	Course Name IS-102 (Foundation of Information Systems)				
	Course factors	*Social factors*	*Individual factors*	*Overall Priority*	**Rank**
IA_4	0.6702	0.05244	0.2695	0.992	I
IB_4	0.4763	0.3661	0.109	0.9514	II

The procedure is continuing for all courses that are generated at the first phase. The aim of this step is to prioritize the offered sections in the timetable. Thus, students are provided by a list of courses and its sections taking in consideration their preferences.

7. CONCLUSION and FUTURE WORK

Course enrollment (CE) as administrative task is a repetitive process which faces students each semester. Students, during the enrollment period, often, need to support. At a time not so long ago, students were responsible for their own choices. Since the students aim to finish their study as soon as they can taking as many courses as possible, their choices might affect negatively on their performance. From this end, colleges and universities began to implement so-called academic advising affairs. Although, the faculty academic advising has a significant impact on a student's academic success, several issues may lead to limit this success specially when the ratio of the academic advisors to the students is high.

In this paper, we have presented the course enrollment task as a function of maximization of GPA. For this purpose, we have proposed a two-phased process. The first phase, is similar to those presented in [23] where the most correlated courses are generated. At the second phase, the courses are prioritized based on the student preferences. The students selection is influenced several factors which have been categorized into three main groups of concerns: (i) course factors, (ii) social factors, and (iii) individual factors.

Through this work, we have evaluated our decision model in the context of department of Information System at the Taibah University, Kingdom of Saudi Arabia. Since the collected data were from different curriculum plans, our concern was focused only on one curriculum plan because the change in perquisite courses will cause different recommended courses. In the future, we intend to integrate with timetable system of the admission and registration deanship, and build an unified model that can deal with different curriculum plans. Further analysis should cover more factors that influence the course selection itself. It will be more appropriate to shift the research towards collaborative recommended systems and cluster students with same preferences and analyze their behaviors. The date mining approach would also be very interesting.

REFERENCES

[1] Winston RB, Ender SC, Miller TK. Academic advising as student development. Developmental approaches to academic advising. New Directions for Student Services. 1982 (17).
[2] Al-Sarem M. Predictive and statistical analyses for academic advisory support. arXiv preprint arXiv:1601.04244. 2015 Nov 7.
[3] O'Banion T. An academic advising model. NaCADA Journal. 1994 Sep; 14(2): 10-6.
[4] Chiarandini M, Di Gaspero L, Gualandi S, Schaerf A. The balanced academic curriculum problem revisited. Journal of Heuristics. 2012 Feb 1; 18(1): 119-48.
[5] Ognjanovic I, Gasevic D, Dawson S. Using institutional data to predict student course selections in higher education. The Internet and Higher Education. 2016 Apr 30; 29: 49-62.
[6] Vialardi C, Chue J, Peche JP, Alvarado G, Vinatea B, Estrella J, Ortigosa Á. A data mining approach to guide students through the enrollment process based on academic performance. User modeling and user-adapted interaction. 2011 Apr 1; 21(1-2): 217-48.
[7] Takada K, Miyazawa Y, Yamamoto Y, Imada Y, Tsuruta S, Knauf R. Curriculum Optimization by Correlation Analysis and Its Validation. InHuman-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data 2013 (pp. 311-318). Springer Berlin Heidelberg.
[8] Baker RS. Data mining for education. International encyclopedia of education. 2010 May; 7(3): 112-8.
[9] Hsia TC, Shie AJ, Chen LC. Course planning of extension education to meet market demand by using data mining techniques–an example of Chinkuo technology university in Taiwan. Expert Systems with Applications. 2008 Jan 31; 34(1): 596-602.
[10] Monette JN, Schaus P, Zampelli S, Deville Y, Dupont P. A CP approach to the balanced academic curriculum problem. InSeventh International Workshop on Symmetry and Constraint Satisfaction Problems 2007 Sep 23 (Vol. 7).
[11] Hnich B, Kiziltan Z, Walsh T. Modelling a balanced academic curriculum problem. In Proceedings of CP-AI-OR-2002 2002 Mar.
[12] Castro C, Crawford B, Monfroy E. A quantitative approach for the design of academic curricula. InSymposium on Human Interface and the Management of Information 2007 Jul 22 (pp. 279-288). Springer Berlin Heidelberg.
[13] Lambert T, Castro C, Monfroy E, Riff MC, Saubion F. Hybridization of genetic algorithms and constraint propagation for the BACP. InInternational Conference on Logic Programming 2005 Oct 2 (pp. 421-423). Springer Berlin Heidelberg.
[14] Lambert T, Castro C, Monfroy E, Saubion F. Solving the balanced academic curriculum problem with an hybridization of genetic algorithm and constraint propagation. Artificial Intelligence and Soft Computing–ICAISC 2006. 2006: 410-9.
[15] Ünal YZ, Uysal Ö. A new mixed integer programming model for curriculum balancing: Application to a Turkish university. European Journal of Operational Research. 2014 Oct 1; 238(1): 339-47.
[16] Rubio JM, Palma W, Rodriguez N, Soto R, Crawford B, Paredes F, Cabrera G. Solving the balanced academic curriculum problem using the ACO metaheuristic. Mathematical Problems in Engineering. 2013 Dec 17; 2013.
[17] Laghari MS, Khuwaja GA. Electrical engineering department advising for course planning. InGlobal Engineering Education Conference (EDUCON), 2012 IEEE 2012 Apr 17 (pp. 1-6). IEEE.
[18] Hashemi RR, Blondin J. SASSY: A petri net based student-driven advising support system. InInformation Technology: New Generations (ITNG), 2010 Seventh International Conference on 2010 Apr 12 (pp. 150-155). IEEE.
[19] Nambiar AN, Dutta AK. Expert system for student advising using JESS. InEducational and Information Technology (ICEIT), 2010 International Conference on 2010 Sep 17 (Vol. 1, pp. V1-312). IEEE.
[20] Mashiloane L. Using association rule mining to find the effect of course selection on academic performance in computer science i. InMining Intelligence and Knowledge Exploration 2014 (pp. 323-332). Springer International Publishing.
[21] Hsieh TC, Wang TI. A mining-based approach on discovering courses pattern for constructing suitable learning path. Expert systems with applications. 2010 Jun 30; 37(6): 4156-67.
[22] Aher SB. EM&AA: An Algorithm for Predicting the Course Selection by Student in e-Learning Using Data Mining Techniques. Journal of the Institution of Engineers (India): Series B. 2014 Jan 1; 95(1): 43-54.
[23] Tsuruta S, Knauf R, Dohi S, Kawabe T, Sakurai Y. An intelligent system for modeling and supporting academic educational processes. InIntelligent and Adaptive Educational-Learning Systems 2013 (pp. 469-496). Springer Berlin Heidelberg.
[24] Curran JM, Rosen DE. Student attitudes toward college courses: An examination of influences and intentions. Journal of Marketing Education. 2006 Aug; 28(2): 135-48.
[25] Wilhelm WB. The relative influence of published teaching evaluations and other instructor attributes on course choice. Journal of Marketing Education. 2004 Apr; 26(1): 17-30.
[26] Babad E, Tayeb A. Experimental analysis of students' course selection. British Journal of Educational Psychology. 2003 Sep 1; 73(3): 373-93.
[27] Beggs JM, Bantham JH, Taylor S. Distinguishing the Factors Influencing College Students'choice of Major. College Student Journal. 2008 Jun 1; 42(2): 381.
[28] Ferrer-Caja E, Weiss MR. Cross-validation of a model of intrinsic motivation with students enrolled in high school elective courses. The Journal of Experimental Education. 2002 Jan 1; 71(1): 41-65.

[29] Barak M, Watted A, Haick H. Motivation to learn in massive open online courses: Examining aspects of language and social engagement. Computers & Education. 2016 Mar 31; 94: 49-60.
[30] Lee A. Virtually Vygotsky: Using Technology to Scaffold Student Learning: By. Technology in Pedagogy. 2014; 20: 1-9.
[31] Coleman JA. English-medium teaching in European higher education. Language teaching. 2006; 39(01): 1-4.
[32] Naz, A., Saeed, G., Khan, W., Khan, N., Sheikh, I., & Khan, N. (2014). Peer and Friends and Career Decision Making: A Critical Analysis. Middle-East Journal of Scientific Research, 22(8), 1193-1197.
[33] Webber DJ, Walton F. Gender-specific peer groups and choice at 16. Research in Post-Compulsory Education. 2006 Mar 1; 11(1): 65-84.
[34] Malgwi CA, Howe MA, Burnaby PA. Influences on students' choice of college major. Journal of Education for Business. 2005 May 1; 80(5): 275-82.
[35] Anderson N, Lankshear C, Timms C, Courtney L. 'Because it's boring, irrelevant and I don't like computers': Why high school girls avoid professionally-oriented ICT subjects. Computers & Education. 2008 May 31; 50(4): 1304-18.
[36] M. Al-Sarem, M.M. Bellafkih and Ramdani. Adaptation Patterns with respect to Learning Styles. Proceeding of 9th International Conference on Intelligent Systems: Theories and applications (SITA'14), Rabat, Morocco.
[37] Truong HM. Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. Computers in Human Behavior. 2016 Feb 29; 55: 1185-93.
[38] Chang PC, Lin CH, Chen MH. A Hybrid Course Recommendation System by Integrating Collaborative Filtering and Artificial Immune Systems. Algorithms. 2016 Jul 22; 9(3): 47.

BIOGRAPHY OF AUTHOR

Mohammed Al-Sarem is an assistant professor of information science at the Taibah University, Al Madinah Al Monawarah, KSA. He received the PhD in Informatics from Hassan II University, Mohammadia, Morocco in 2014. His research interests center on E-learning, educational data mining, Arabic text mining, and intelligent and adaptive systems. He published several research papers and participated in several local/international conferences