A college degree recommendation model

Milton Rafael Maridueña-Arroyave & Ailyn Febles-Estrada

Received: November 25th, 2015. Received in revised form: May 20th, 2016. Accepted: June 2nd, 2016

Abstract
Selecting a profession suitable to students’ expectations implies taking into account multiple factors. Despite its usefulness and high impact, there are shortcomings in current university major recommendation models. Among these limitations are the lack of flexible models, the dependence on historical information and the inadequate weighting of the factors involved. In this paper, a new college degree recommendation model based on psychological student profiling and the analytical hierarchical process is presented. It includes database construction, student profiling, college degree information filtering and recommendation generation. Its implementation made it possible to improve reliability in the recommendation process of college degree. A case study is shown to demonstrate the model applicability.

Keywords: recommender systems, college degree recommendation, AHP, student profile.

1. Introduction
Selecting a future career is a complex decision process involving preferences, aptitudes, interests and qualities. Current process based solely on multicriteria decision models allows to handle only a limited number of options (college degrees) [1]. Recommendation models are more adequate due to the relative easiness to take into account users profiles and expectations [2]. Despite the high impact and usefulness of recommending a college degree, there is a group of limitations such as:

- Current models are based mainly on collaborative filtering [3-5] or data mining, like association rules and decision trees [6-8], nevertheless very frequently, there is a lack of historical information making impossible to use these approaches. For example when dealing with new students, they do not have information about them, and they are then unable to generate recommendations.
- Another shortcoming is that current approaches are based solely on specific subject recommendation, not on whole college degrees.
- Similarity calculation is based in weighted averaging of features. This operator does not take into account interaction like compensation, orness and bipolarity [9-11].
- Models lack dealing with the psychological profile of students [12] to reach a more reliable recommendation.
- In this paper a new model of college degree recommendation is presented using a flexible similarity calculation based on weights obtained from the analytic hierarchy process (AHP), a hierarchical aggregation
process using the weighted power mean [13] and the student’s psychological profiling.

The outline of this paper is as follows: Section II is dedicated to recommendation models, Section III to AHP. The proposed framework is presented in Section IV. A case study is discussed in Section V. The paper closes with concluding remarks, and the discussion of future work in Section VI.

2. Recommendation models

Recommendation systems are useful in decision making process providing the user with a group of options hoping to meet expectations [2]. Based on the information they use and the algorithms used to generate the recommendations, we can distinguish the following techniques [14, 15]:

- Collaborative Filtering Recommender Systems: they use users’ ratings to recommend items to a specific user. They aggregate preferences of the other users’ preferences to generate new recommendations.
- Content-based Recommender Systems: They learn a user profile based on the features of the items that the user had liked. The user profile could be completed based on psychologic user profiling techniques.
- Knowledge Based Recommender Systems: these systems use the knowledge about users’ necessities to infer recommendations. They use cased based reasoning techniques frequently.
- Utility Based Recommender Systems: they make recommendations by computing a utility value.

In the specific case of the systems for vocational guidance, existing proposals rely fundamentally on collaborative filtering approaches [3-5] or data mining techniques [6-8]. But often there is not historical information which makes it impossible to adopt these approaches. Within these systems the Degree Compass System of Austin Peay State University [16] stands out. However, this system shares a common limitation with the rest of the systems studied related to focusing only in the recommendation of specific courses rather than college degrees entirely.

It is possible to improve the reliability of the recommendations obtaining a student profile based on their psychological traits [17]. This profile allows developing recommendations based on content given the similarity of shared characteristics between the object to be recommended and the student profile [12].

3. Analytic Hierarchy Process

The Analytic Hierarchy Process is a technique created by Tom Saaty [18] for making complex decision based on mathematics and psychology. The steps for implementing the AHP proposed model are:

1. Decompose the problem into a hierarchy of goal, criteria, sub-criteria and alternatives.
2. Collect data from experts or decision-makers corresponding to the hierarchic structure, in the pairwise comparison of alternatives on a qualitative scale.

3. Assign a weight to criteria and sub-criteria.
4. Calculate the score for each of the alternatives through pairwise comparison.

AHP can be used in addition to the group assessment [19]. In this case to obtain the final value, the weighted geometric mean [20] is used. The weighting could give different weights to the criteria of the specialists taking into account various factors such as authority, expertise, effort, etc.

The integration of AHP model with university degrees recommendation allows to assign a weight to each of the factors involved in the suggestion of a college career, doing this more in line with reality and therefore more reliable.

4. Proposed framework

The proposed framework is presented in Fig. 1. It is based mainly on the proposal made by Cordon [15] for recommendation systems based on content/knowledge adapted to the characteristics of the application domain and allowing flexibility in the aggregation of the similarity of the characteristics in the user profile with respect to ideal profiles of the college degree.

4.1. Database creation

A key for a recommendation model is the creation of the database. Each university degree \( a_i \) will be described by a set of characteristics that make up the profile:

\[ C = \{c_1, ..., c_k, ..., c_l\} \] (1)

Each of the features which are reflected in the psychological profile may be composed of sub-features. Each university degree will be described by a vector of features:
There are techniques for generating these profiles automatically or semi-automatically for recommendation systems based on psychological profiles [21]. In this case, an expert or group of experts is suggested.

Having described the university degrees in this set:

$$A = \{a_1, \ldots, a_j, \ldots, a_n\}$$  (3)

Then, it is stored in a database.

4.2. Acquisition of the user profile

The proposed framework presents a fundamental difference with previous proposals, it is focused in the fact that most of this information may be collected by psychological tests and can be supplied by psychologists to advise the student.

The student profile corresponds to his psychological profile. In this activity, this information is stored in the database.

$$P_e = \{p^e_1, \ldots, p^e_k, \ldots, p^e_l\}$$  (4)

This profile will be composed of a set of attributes:

$$C^e = \{c^e_1, \ldots, c^e_k, \ldots, c^e_l\}$$  (5)

Features such as skills and emotional intelligence are included.

4.3. College degree filtering

In this activity, college degrees according to the similarity with the user profile are filtered to find out which are the most appropriate for the student.

The similarity between user profile, $P_e$, and each ideal college degree profile $a_j$ is calculated. For the calculation of the overall similarity a hierarchical aggregation is used taking into account the following factors:

- Degree of simultaneity.
- Relative importance of the inputs (weights).

Aggregation function [22]: $OAG: [0,1]^n \rightarrow [0,1]$ is obtained by a process of hierarchical aggregation. The weighted mean power, (WPM) as in the Logic Scoring of Preference (LSP) method [22] is used. The rth average power is defined as follows:

$$M^r_{\sum_1^n}(a, w) = \left(\sum_{i=1}^n a^t_i w_i\right)^{1/r}$$

where $w_i \in [0,1]$ and $\sum_{i=1}^n w_i = 1$ and r can be selected to achieve desired logical properties. For determining each feature and sub-features weights AHP method [18] is used.

4.4. Recommendation

For the calculation of the similarity measures the following expression is used:

$$s_j = s(P_e, a_j) = OAG \left(V_{sim}(P_e, a_j)\right)$$  (7)

Where

$$V_{sim}(P_e, a_j) = \{sim(p^e_1, v^j_1), \ldots, sim(p^e_k, v^j_k), \ldots, sim(p^e_l, v^j_l)\}$$

is a vector containing the similarity of all user profile attributes regarding the description of the college degree $a_j$.

The similarity measure can be obtained from a distance measurement, if $d(x, y)\in [0, \max]$ then[23] :

$$sim(p^e_k, v^j_l) = 1 - \frac{d(p^e_k, v^j_l)}{max}$$  (8)

In case of ordered lists, such as characterology, interest and professional competencies, Kendal Tau distance is used [24, 25].

4.5. Recommending

In this activity, a set of college degrees that match with the students profiles is suggested. After calculating the similarity between the student profile and each college degree profile in the database each college degree is ordered and is represented with the following similarity vector:

$$S = (s_1, \ldots, s_n)$$  (9)

The best are those that best meet the needs of the student profile (greater similarity).

5. Case study

To show the applicability of the model, a case study at the University of Guayaquil is developed. College degree ideal profiles was acquired from experts taking into account features and sub-features as it is shown in Table 1.

Ideal college degree profiles are obtained in a group of college degree in Health and Information Sciences (Table 2). They are composed by numerical scores (skills, emotional intelligence) and ordered lists (interests, professional competencies, characterology) information.

## Table 1. Features included in the student profile

<table>
<thead>
<tr>
<th>Feature</th>
<th>Interests</th>
<th>Skills</th>
<th>Professional Colorado</th>
<th>Emotional intelligence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Verbal</td>
<td>Abstraction</td>
<td>Intrapersonality</td>
<td></td>
</tr>
<tr>
<td>Sub-feature</td>
<td>Stress management</td>
<td>General mood</td>
<td>Interpersonality</td>
<td></td>
</tr>
</tbody>
</table>

Source: The authors.
Table 2.
Ideal college degree profile

<table>
<thead>
<tr>
<th>College degree</th>
<th>SKILLS</th>
<th>INTERESTS</th>
<th>PROFESSIONAL</th>
<th>CHARACTEROLOGY</th>
<th>EMOTIONAL INTELLIGENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VERBAL</td>
<td>ABSTRAC</td>
<td>LOGIC</td>
<td></td>
<td>INTRAPERSONAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>INTERPERSONAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ADAPTABILITY</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>STRESS MANAGEMENT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GENERAL MOOD</td>
</tr>
<tr>
<td>Bibliotecología y Archivología</td>
<td>60</td>
<td>53</td>
<td>40 G C L D A O</td>
<td>PHLEGOMATIC</td>
<td>SANGUINE 90 86 100 116 110</td>
</tr>
<tr>
<td>Odontología</td>
<td>60</td>
<td>60</td>
<td>60 A F E I B O</td>
<td>PASSIONATE</td>
<td>SANGUINE 99 90 90 95 116</td>
</tr>
<tr>
<td>Obstetricia</td>
<td>53</td>
<td>53</td>
<td>47 A G C B F O</td>
<td>PASSIONATE</td>
<td>SANGUINE 99 90 90 116 116</td>
</tr>
<tr>
<td>Enfermería</td>
<td>53</td>
<td>53</td>
<td>47 A N C B M O</td>
<td>PASSIONATE</td>
<td>SANGUINE 99 90 90 116 116</td>
</tr>
</tbody>
</table>

Source: The authors.

Table 3.
Student profile

<table>
<thead>
<tr>
<th>SKILLS</th>
<th>INTERESTS</th>
<th>PROFESSIONAL</th>
<th>CHARACTEROLOGY</th>
<th>EMOTIONAL INTELLIGENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>VERBAL</td>
<td>ABSTRAC</td>
<td>LOGIC</td>
<td></td>
<td>INTRAPERSONAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>INTERPERSONAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ADAPTABILITY</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>STRESS MANAGEMENT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GENERAL MOOD</td>
</tr>
<tr>
<td>33</td>
<td>47</td>
<td>20 A I B I B F</td>
<td>PASSIONATE</td>
<td>PHLEGOMATIC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AMORPHOUS 110 98 112 114 109</td>
</tr>
</tbody>
</table>

Source: The authors.

Table 4.
Feature weights calculation

<table>
<thead>
<tr>
<th>Features</th>
<th>$c_1$</th>
<th>$c_2$</th>
<th>$c_3$</th>
<th>$c_4$</th>
<th>$c_5$</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skills ($c_1$)</td>
<td>1</td>
<td>1/8</td>
<td>1/8</td>
<td>6</td>
<td>1/8</td>
<td>0.0408</td>
</tr>
<tr>
<td>Interest ($c_2$)</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>6</td>
<td>1/8</td>
<td>0.3012</td>
</tr>
<tr>
<td>Professional competencies ($c_3$)</td>
<td>8</td>
<td>1/9</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>0.1543</td>
</tr>
<tr>
<td>Characterology ($c_4$)</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1</td>
<td>1/7</td>
<td>0.0238</td>
</tr>
<tr>
<td>Emotional intelligence ($c_5$)</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Source: The authors.

In the case of interest A, B, C, E, F, G, I, L and N correspond to Science Professionals (health areas), Technology sub-professional (engineering areas), Consumer Economics (business), Job Office (commerce and secretarial), Professional Art (design, general arts), Professional Social Service (related to providing services and care areas), sub-professional technologies (technologies, technical), Communication (use of language as part of the job) and Social Service sub-professionals (personal care) respectively.

In the case of professional skills A, B, D, F, I, M and O correspond to Politics and Law (jurisprudence), Biomedical (medical sciences), Education (educational sciences), Biotechnology (chemical sciences) Oral health (dentistry), Communication and Service (media) and Psychosocial Health (psychology) respectively.

Later, the psychologist obtained a student profile which is shown in Table 3, based on observation and psychological tests.

Using the AHP method the following weights structure (Table 4) was obtained. These are translated into weight vector associated with the features $V = (0.0408, 0.3012, 0.1543, 0.0238, 0.48)$. In this case, equal weight to the sub-attributes are set.

Then, the aggregation structure is obtained (Fig. 2). Aggregation operators that reflect simultaneity as established LSP [26, 27] were used.

These operators reflect specific requirements and logic conditions, such as simultaneity and replaceability.

Then, the aggregation structure is obtained (Fig. 2). Aggregation operators that reflect simultaneity as established LSP [26, 27] were used.

These operators reflect specific requirements and logic conditions, such as simultaneity and replaceability.

Inputs | Operators | Block ID | Operator | Block ID |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>0.33</td>
<td>C-</td>
<td>Skills</td>
<td>0.0408</td>
</tr>
<tr>
<td>Abstract</td>
<td>0.33</td>
<td>C-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logic</td>
<td>0.33</td>
<td>C-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest</td>
<td>0.3012</td>
<td>C-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional competencies</td>
<td>0.1543</td>
<td>C-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Characterology</td>
<td>0.0238</td>
<td>C-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional intelligence</td>
<td>0.48</td>
<td>C-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Components of the similarity calculation aggregation structure
Source: The authors.
Table 5.
Similarity between the ideal college degree and the student’s profiles.

<table>
<thead>
<tr>
<th>Bibliotecología</th>
<th>Odontología</th>
<th>Obstetricia</th>
<th>Enfermería</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$a_2$</td>
<td>$a_3$</td>
<td>$a_4$</td>
</tr>
<tr>
<td>0,651</td>
<td>0,877</td>
<td>0,815</td>
<td>0,822</td>
</tr>
</tbody>
</table>

Source: The authors.

The similarity of the ideal profile to different college degrees gives the following result.

In the phase of recommendation, those college degrees that come closest to student profile will be recommended. An ordering based on this comparison is:

$$\{a_2, a_4, a_3, a_1\}$$

If the system recommend the three college degrees more similar to the student profile, they would be the following: Odontology, Nursery and Obstetrics; which coincide with the actual recommendations given by the department of student welfare.

6. Conclusions

Despite the impact along life of deciding what career to pursue, shortcomings persist in treating recommendation process of college degrees. This paper presents a model for recommendation of college degrees following the content-based approach. It is based on the psychological student profiling and the database of ideal college degree profiles.

The AHP method allows a correct weighting of different factors involved. Additionally, the LSP method of aggregation operators permits to reflect simultaneity and replaceability in the process. The previous elements and the inclusion of the psychological profiling of students allows to reach a more reliable recommendation.

Future work will be related to the inclusion of context information in the model creation of the database from multiple experts, as well as obtaining the weights of the features using group assessments. Other areas of future work will be related to the management of heterogeneous information and the development of a software tool.

References

[22] Pérez-Teruel, K., Leyva-Vázquez, M. and Estrada-Senti, V., Mental models consensus process using fuzzy cognitive maps and computing with words. Ingeniería y Universidad, 19, pp. 173-188. 2015. DOI:10.1007/s10479-014-1576-8

M.R. Maridueña-Arroyave, es Dr. (Ph.D.) en Ciencias Pedagógicas, Candidato a Dr. (Ph.D.) en Ciencias Informáticas, MSc. en Docencia Universitaria e Investigación Educativa, Candidato a MSc. en Investigación...
ORCID: 0000-0002-8876-1896

A. Febles-Estrada, es Lic. en Ciencias de la Computación de la Universidad de La Habana, Cuba. Dra. en Ciencias Técnicas (Ph.D.) del Instituto Superior Politécnico José Antonio Echeverría. Profesora de la Universidad de las Ciencias Informáticas, La Habana, Cuba.
ORCID: 0000-0002-9678-2230

Área Curricular de Ingeniería de Sistemas e Informática

Oferta de Posgrados

Especialización en Sistemas
Especialización en Mercados de Energía
Maestría en Ingeniería - Ingeniería de Sistemas
Doctorado en Ingeniería- Sistema e Informática

Mayor información:
E-mail: acsei_med@unal.edu.co
Teléfono: (57-4) 425 5365