

## A Model for Recommending Specialization Courses Based on the Professional Profile of Candidates

Antônio Eduardo Rodrigues de Souza  
Electrical Engineering Post-graduation Program  
Universidade Presbiteriana Mackenzie, UPM  
São Paulo, Brazil  
e-mail: aersouza@gmail.com

Sandra Maria Dotto Stump  
Electrical Engineering Post-graduation Program  
Universidade Presbiteriana Mackenzie, UPM  
São Paulo, Brazil  
e-mail: sstump@mackenzie.br

**Abstract**— The paper studies the candidates' professional profile on choosing a specialization course. A methodology based on the processes Knowledge Discovery in Databases (KDD) and Cross-Industry Standard Process for Data Mining (CRISP-DM) is applied, and proposed a course recommendation model, using a technique of data mining based on decision trees for the discovery of relevant knowledge from database, which will identify the most suitable course to a candidate's profile. In this study, it is expected to be detected the specialization courses which best suits each candidate profile, giving support to academic institution to satisfy candidates needs and reduce the number of dropouts or changes.

**Keywords**—recommender systems; data mining; data filtering techniques; academic counselling

### I. INTRODUCTION

The last decade of the XX century was characterized by intense economic globalization, by the need for continuous and quick modernization of production systems, and by the extreme competitiveness in goods and services markets, requiring a better qualification of manpower. However, in the field of education there was the offer of courses strictly academic, which motivated the need of a more specific qualification for the exercise of certain professions, starting to be required master's degrees or doctorates, creating a growing demand by professionals with highly specialized skills profile and not focused on pure research. Such professionals could not be certainly formed as byproducts of courses targeted to the academic and scientific qualifications, but with a technical and scientific nature [1].

The need, by the companies, of increasingly well qualified professionals to meet the demands required, brings consequences as the imposition of education, specific training and qualifications. Moreover, the search for better job opportunities and salaries, have generated a strong influence on the demand for training programs. The professional has transited, increasingly, between the profession and the acquisition of knowledge. It is known that an undergraduate degree cannot over assure a successful career. Currently, professionals are obliged to seek more and more knowledge to achieve specific abilities and skills, meeting the existing shortage in the labor market. As the

labor market dynamics change quickly, there is a demand for even more specialized professionals.

Specialization courses are sought after by professional options as a way to acquire and update knowledge [2]. Offered by institutions in various areas, targeted audiences with specific interests or general, diversification of courses or the lack of objective information, hinder the understanding of the purpose, the necessary prerequisites and other important factors in the decision to be made. Thus, a poorly chosen option may incur dissatisfaction, frustration and expectations need to change or even chosen option or even cancellation of the course.

In Brazil, the specialization is a post-graduation course. From Latin, *lato sensu* is an expression whose meaning is "broad sense."

Designed to be attended by persons performing other activities simultaneously, specialization consists of a course of professional qualification with a minimum duration of 360 hours. These courses are not evaluated by the Ministry of Education (MEC) and the Coordination of Improvement of Higher Education Personnel (CAPES), but have significant value to the labor market, especially those courses offered by renowned institutions [3].

It is considered that the topic is up to date and relevant to Higher Education Institutions (HEIs) that offer specialized courses, and presenting various types of courses, in order to give opportunity to continue the training of the candidate. It also shows the IES, the importance of offering specialized courses that are aligned both to the skills of the educational institution, as the interests of training and retraining of skilled manpower for the labor market.

In this context, concerns about the quality of information provided to prospective academic specialization courses at university, is an important aspect that should be considered and, moreover, contribute to the proper choice to meet the needs and profile.

The study becomes relevant since options of courses are offered to meet the expectations of candidates.

This paper is organized as follows: the first section presents the problem that is being studied. The background is presented in Section 2. Section 3 presents the proposed model.

## II. BACKGROUND

Advances in science and technology have caused major changes in the global job market, making many professionals, in different areas, hoping to stay in or re-entering the job market, look for continuing education through specialization courses.

Nowadays, the new needs of qualified professionals have shown that the percentage of courses offered has increased to meet market demands. It also shows that quality is searched in the offered courses, which are mainly targeted at preparing more qualified professionals for the job market.

The present study proposes, from historical information of candidates for specialization courses at a private university in the city of São Paulo, to analyze and define behaviour profiles. The research is delimited to candidates who sought specialization courses during the last two semesters.

The database used on this study will serve as a source for data mining, from which will be extracted, in an accurate manner, information that will make courses suggestions for a future candidate. Thus, the objective is to develop a model based on candidate profiles, using artificial intelligence, filtering techniques and data mining, to customize offers of specialization courses. To achieve the objective of this research is intended to perform the following steps: a) Describe by means of literature the main techniques used in Recommender Systems and Data Mining; b) Show how Data Mining can help to identify knowledge in large data volumes; c) Produce an output interface for displaying candidates' data profiles and courses recommendations.

For the evaluation of recommendation model, tests will be performed with the database of candidates provided by the university. These candidates answered a questionnaire about goals and interests with respect to the desired course, the time in which they were enrolled.

## III. MODELING THE PROFESSIONAL PROFILE OF CANDIDATES

Will be held, initially, a documentary research, through a literature review, which aims to know the state of the art recommender systems, evaluate the filtering techniques and data mining available, and adapt them to the proposed model.

The work will be developed from the data collection of an exploratory, qualitative approach between those who were candidates for specialization courses at a private university, located in the city of São Paulo, between the years 2006 and 2011.

The university in study conducts semiannual selection of candidates for specialization courses. An amount of approximately three thousand candidates takes part of the process each term. The selection process consists of an electronic form, available on the website of the university, with questions which, besides personal information, allows depth knowledge of academic courses, objective, expectations regarding the content and reasons for choosing the course, for example, professional development, job promotion, etc. The responses of each candidate are recorded

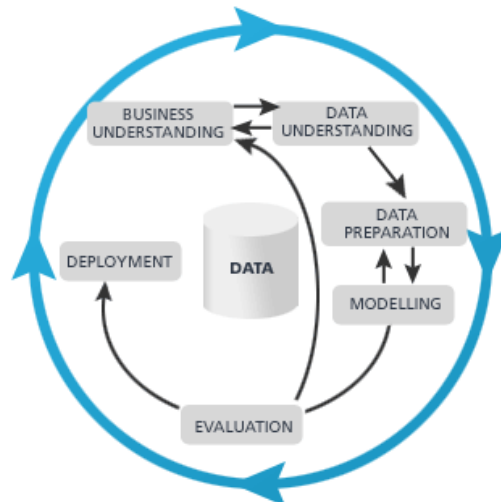


Figure 1. CRISP-DM process steps [4].

in Oracle database. With such information, the coordinator of each course verifies the information stored for each candidate and evaluates whether the chosen course is really best suited to the profile posted. Only at the time of publication of the results of the successful candidates who have knowledge of the course where it will be registered, since the electronic form allows the applicant to select three course options. Classes are formed with maximum students per classroom. If they are not offered various classes of the same course, the successful candidate will be relocated to one of the options selected by him, where there is still vague.

For the identification of knowledge it will be studied CRISP-DM [4], shown in Figure 1, and the KDD process [5], shown in Figure 2. Will be considered an approach based on collaboration, to best suit and adhere the needs of candidates from different areas.

### A. CRISP-DM Process Steps

The first step comprehends the business understanding. All the understanding of the requirements of the candidate is elicited during this step during meetings with the responsible for the process of selections of candidates. The next step comprehends the data understanding, which involves exploratory analysis on the received data. After data comprehension, the next step is the data modelling, which means to create models based on algorithms, to find the

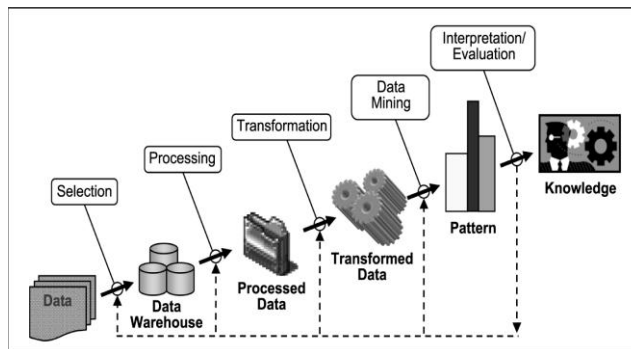


Figure 2. KDD process steps [5].

technique with greater effectiveness. Once the data modelling is finished, performance and gain with the proposed model are evaluated. The last step, installation, comprehends implementation of business rules validated by the pedagogical coordinator of each course in the database to be worked.

**B. KDD Process Steps**

First there will be the understanding and definition of a domain. Then will be selected within the domain established, the data on which the discovery is performed. Among the acquired data will be selected for analysis the factors that can be identified as potential influencers in choosing a course, such as interests and professional goals: Applicants may direct the choice in their training or opt for a more diversified; chance continuity of learning: candidates may tend to choose courses that enable continuous learning, knowledge and difficulty: where the complexity of the content and the selection criteria may influence the choice; Lack of information: applicants may not be aware of the content and target audience of the course; Location: a campus closest to the address of residence or work can contribute positively or negatively to choice. Besides these factors, the information will be considered professional profiles of the candidates listed on the registration form. The professional profile is a set of characteristics that need to be found in a candidate so that it can occupy a certain position, and can be divided into technical knowledge (education, training area, languages, work experience, computer) and behavioural profile (communication, interpersonal relationships, judgment, attitude, ethics) [6]. Due to the subjectivity and difficulty encountered in measuring values that express the behavioural profile will be considered in this work only the technical knowledge, as shown in the table below:

**TABLE 1. ALL VARIABLES (ATTRIBUTES) USED ON THE STUDY**

Variable	Description	Example
Graduation	Graduation course name	Administration, Architecture and Urban Planning, Computer Science
Degree	Received degree title	Bachelor, technologist, doctor
Post-graduation	Post-graduation course name	Administration, Architecture and Urban Planning, Computer Science
Professional activity	Activity name	Systems Analyst, Support Analyst, Architect, Controller, Journalist, Secretary
Position	Position name	Advisor, Analyst, Assistant Supervisor, Coordinator, Manager
Experience (years)	Experience time on the position	Less than 1 year, 1 to 3 years, over 3 years
English idiom	Knowledge level of English language	None, basic, intermediate, advanced, fluent

As a way of preparing for the next step, these data must be cleaned and processed. This cleanup includes removing noises, which are data errors or outliers, the adequacy of values that are out of context, the inclusion of missing values, selection and summary of variables to be used (see Table 2 and Table 3). Missing values, that is the absence of information contained in the records, are entered using a global constant, such as the average of each attribute or the average of all the variables of the same class. In case of nominal attributes a dominant subset will be used, whereas in case of non-nominal attributes will be used the mode of each class, which is the value which most frequently occurs in a data set. It will be selected the variables and eliminated some unnecessary variables after checking the information gain of each variable. Courses no longer offered by the university, and courses whose identification codes have changed, will be replaced by courses or equivalent codes in addition to eliminating redundant data, generating greater reliability.

**TABLE 2. DATA CLEANSING SAMPLE**

Received data	Cleaned and processed data
AdministraA?A?o	Administração
AutA?nomo	Autônomo
CiA?ncias	Ciências
SupervisA?o	Supervisão
TecnA?logo	Tecnólogo

**TABLE 3. DATA TRANSFORMATION SAMPLE**

Professional activities (Non transformed data)	Professional activities (Transformed data)
ADM, administrador de empresas, administração, administrador de empresas, administradora de empresas	Administração de Empresas
Adm./Financeiro, administrativo e financeiro, administrativo financeiro	Administração e Finanças
Advogada, advogado, advogada autônoma, advogado júnior, jurídico, jurídica, jurista	Advocacia
Analista de sistemas junior, analista de sistema, analista de sistemas informática, analista de sistemas, analista sistema, analista sistemas	Análise de Sistemas

Still in this step, it will be defined the techniques and data mining algorithms to be used, the domain selected to be then processed according to the technical characteristics of the algorithms. After this step, the data will be submitted to data mining itself.

To ensure a satisfactory number of elements of analysis it will be evaluated data from several previous semesters. The data will be received in Microsoft Excel spreadsheets, because of its easiness of being transformed into one of the formats accepted by the tool used in this step.

This stage also includes the use of classification rules. The classification will be used to identify candidate profiles,

which also represent the choice of courses. The main mining technique and algorithm used in this step are: Decision Tree - hierarchical data, based on stages of decision (nodes) and the separation of classes and subsets. Major current algorithms: CART, CHAID, C5.0. For the proposal, it is intended to use association rules to classify data due to its better accuracy in recommending courses [7].

Will take part of the model the 56 courses offered semi-annually by the university, but the main focus of the analysis will be the courses that generate a greater number of dropouts or substitutions, and therefore, in addition to modification of various administrative procedures, can be the cause for discouragement, frustration and dissatisfaction from candidates.

For the information related to occupational profiles can be tested, it will be used professional profiles among exact, humanities and social areas, such as Business Administrator, Controller, Tax Attorney, Educator, Financial Analyst, Journalist, Human Resources Analyst, Psychologist, Analyst Systems, Advertising, Support Analyst, Secretary.

With the prepared database, it will be held the mining and machine learning steps. The tool that will be used is WEKA (Waikato Environment for Knowledge Analysis), developed by the University of Waikato, New Zealand. This tool was chosen because of its public domain, have been developed in the Java language, and working with various data mining techniques such as association rules, clustering, classification, and different algorithms.

As a result generated, the knowledge gained will be examined by a specialist, such as the educational coordinator of each course, to improve understanding of the knowledge discovered by the mining algorithm. If mining results are not satisfactory, several process steps can be carried back.

The aim is to analyze comparatively the number of changes in previous years and those requested by implementing them in the proposed model. In sequence will be evaluated, by the number of candidates who changed course or dropout, the assertiveness of the recommendation by the proposed model, in comparison to data collected by the enrolment system.

The last phase includes the development of an online report, for easy viewing of results, where it intends to use the Java language to implement it, due to its high performance in a web environment. This information will allow the university staff to identify which professional profiles are best suited to specific courses, considering, also, possible relocations.

Based on different candidates' profile, Figure 3 shows part of the report that will be used to support the courses coordinators on course recommendation. It also shows which courses best suits to each candidate profile, according to the discovered model.

Specialization Course Recommendation Based on Candidates Professional Profile			
Graduation / Degree / Post-graduation	Professional activity / Position / Experience (years)	English	Intended course
Computer Science	Systems Analyst	Fluent	Project Management
Bachelor	Manager		
	Systems Engineering	1 to 3	
<b>Recommended course: Project Management</b>			
Computer Science	Systems Analyst	Basic	Project Management
Bachelor	Analyst		
none		1 to 3	
<b>Recommended course: Systems Project and Development</b>			
Administration	Project Analyst	Advanced	Project Management
Bachelor	Supervisor		
	Software Engineering	Over 3	
<b>Recommended course: IT Governance</b>			

Figure 3. Output report prototype.

#### IV. CONCLUSION

This ongoing study has introduced another approach for recommending courses based on candidates' profile, taking into consideration the career that a candidate is following. The use of professional profiles for recommending specialization courses can provide a better qualitative recommendation, matching the professional career and academic objectives.

This paper reports, based on groups of courses that have a high number of dropouts or changes, that it is possible to extract relevant knowledge from the professional data of applicants for specialization courses in any educational institution, and use this knowledge for better planning of course offerings and classes' sizes, as well as provide support to academic advising.

With the contribution of artificial intelligence by means of algorithms and techniques of data mining, the studies developed on this work provided a better understanding of the techniques and concepts used in knowledge discovery, relevant to a selection process of candidates for specialized courses. With the knowledge gained was possible to develop a model of professional activities, with their specialization courses recommended by the applied algorithm. This model allowed us to identify the most relevant courses to professional profiles participants in the selection process.

The suggested model can help to define the table of courses, leading the university to rethink whether a particular course deserves to be offered in subsequent semesters. Still, the presented result, besides providing support to the coordinators of courses in the selection of candidates, also serves to alert the institution about the lack of understanding, from candidate point of view, about the contents presented by the offered course, so that the university could take action to improve disclosure of their courses.

The generated model can support decision-making of any academic institution which intends to improve their academic counselling and reduce the number of students dissatisfied with the course, avoiding dropouts or course changes.

#### ACKNOWLEDGMENT

The authors of this paper thanks Mackenzie Research Funding (MackPesquisa) for sponsoring this research.

#### REFERENCES

- [1] A. C. Giuliani, A. F. N. Netto, M. C. Ponchio, M. S. Neto, and C. M. Batista, "MBAs, mestrados acadêmicos, mestrados profissionais e doutorados em administração: suas contribuições para o ensino e a pesquisa," *Revista de Administração da UNIMEP*, vol. 5, Jan. 2007, pp. 52-73.
- [2] S. M. C. Silva, "A pós-graduação no contexto atual: uma exigência do mercado de trabalho," [http://artigos.netsaber.com.br/resumo\\_artigo\\_23229/artigo\\_sobre\\_a\\_pos-graduacao\\_no\\_contexto\\_atual:\\_uma\\_exigencia\\_do\\_mercado\\_de\\_trabalho](http://artigos.netsaber.com.br/resumo_artigo_23229/artigo_sobre_a_pos-graduacao_no_contexto_atual:_uma_exigencia_do_mercado_de_trabalho), [retrieved: 7 February, 2013].
- [3] Brasil, Diário Oficial da União, Resolução CNE/CEN 1/2001, Apr. 9, 2001.
- [4] P. Chapman et al., "CRISP-DM 1.0 - Step-by-step data mining guide," The CRISP-DM Consortium, 2000.
- [5] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, "From data mining to knowledge discovery: an overview," AAAI Press, 1996, pp. 1-34.
- [6] S. A. Araújo, D. Santos, and M. A. Bonaldo, "Redes neurais artificiais aplicadas em análise de perfis profissionais," *Proc. CBComp*, Oct. 2004, pp. 280-284.
- [7] N. Bendakir and E. Aïmeur. "Using association rules for course recommendation," *Proc. AAAI Workshop on Educational Data Mining*, Jul. 2006.