

A Framework to avoid Scholar Desertion using Artificial Intelligence

Mayra Treviño, Salvador Ibarra, José Castán, Julio Laria, Javier Guzmán

Abstract— Scholar desertion is one of the main problems in the context of the IES (Superior Educative Institutions) in Mexico. Currently, some trends look for aiming with proper solutions providing a balance using Information Technology. In fact, nowadays there are sophisticated computer sciences capable of identifying and prevent the academic desertion by means of using predictive algorithms which allow being aware about the requirements and needs of the students. In this sense, the present paper proposes a new methodology to reduce the students desertion using the advantages of the data mining theory jointly with advanced artificial intelligence techniques such that, fuzzy logic. Such methodology aims to detect in a suitable moment, when a student tends to leave their studies. The design and some relative works are detailed in order to emphasize the relevance of this promising research area. To the end, some conclusions and a discussion is presented to propose news paths for the implementation of the proposed methodology and future work.

Index Terms—Scholar Desertion, Logic Fuzzy, data mining, predictive algorithms.

I. INTRODUCTION

Different factors are commonly implicated when a student most decide if he continues or not with their professional studies. Some of these aspects are related to: economic situation, personal problems, academic level or ignorance of the administrative processes. For avoiding scholar desertion, some IES have developed internal programs such as the creation of the Tutor entity, but the results have not been the expected. In the most of the cases the teens are not disposed to share their ideas, feelings or thinking to their professors due to they do not identify with them. Besides, we also most claim to the professor because not all the guilty must be addressed to the student. In some situation, the tutor is to blame because he not shows interest in the possible arguments of their tutees. On the other hands, there are professors that really are committed with the program, but they not have time to interact with the student due to the workload (i.e., hour in classroom, research, etc) or because of an inadequate induction course that permit tutor learns new skills to approach with their tutees. In this light,

Manuscript received March 22, 2013. This work was supported in part by the Autonomous University of Tamaulipas and the Engineer School at the UAT.

M. Treviño is with the Engineering School at the Autonomous University of Tamaulipas (corresponding author to e-mail: mtrevino@uat.edu.mx).

S. Ibarra, J. Castán, J. Laria and J. Guzmán are with the academic group of Computational Technology at the Engineering School at the Autonomous University of Tamaulipas in the program of the Computer Science.

there are some approaches that promote the tutor entity using information systems in order to present a less human interaction between the tutor and the student. In theory this proposal will be well accepted by the tutees because they would not be face to face with a professor. However, such computational tools have not reach to attract the attention of the student community. Different causes can be addressed to this situation among which can be highlighted: the failures of the system (due to an incorrect administration of the database, imprecise method in the relationship between question/answer), the updates of the system or simply, an inadequate use of the interface.

In this case, if the administrative personal involved in the scholar service could be aware about the constraints and requirements of the students on time, they will be capable to implement a action for helping them. For example, one of the main causes for leaving school is when a student is not capable to accredit a lecture and he do not understand the processes to present the next tests. For thus, the student presents the exam without know that only have a limited number of chances to obtain an acceptable note. This effect is common between novel students and this situation only can be avoided if the student decides ask for advising to the indicated personal. Other interesting approach is related to the creation of a profile for each student throughout its stay in the IES. Such profile could provide with particular information (scholar level, social behavior, level of satisfaction) usefulness to lead the student in an effective way. For example, if we know that a young have economic problems will be able to recommend scholarship programs or maybe include him in a program for searching a half-time job.

According to this, we consider important to mention that scholar desertion implicates problems to IES, especially for those whose financial resource depends on the state. Recently, these resources are controlled and assigned taking into account the number of enrolled students. For instance, when a class does not achieve a particular quantity of schools, the group must be closed. We must consider that such IES have two classes of professors such that: fulltime and halftime teachers. Therefore, the effect of having a less number of students reflects in a minor amount of lectures for being allocated to the halftime professor, reflecting in a saturation of the fulltime professors avoiding one more time that such worker became in a good tutor.

In spite of this, the scholar desertion is not only based on the failure in the tutor program or due to a bad training of the scholar service personal. The economic situation of the people is also one of the causes by which a young does not the opportunity to study a professional career. In this sense, the profile of the student can provide an opportunity to prevent him desertion or to try enrolled younger people.

Actually, there is approaches in-charge of analyzing and

evaluating factors implicated in the low level of scholar achievement but these proposals are only tools to process information obtained from statistic studies. Such results are information that only presents a brief of the problem when it is too much late to do something. Thereby, at the moment none of the works presents in the literature is capable to contribute with a prediction about a possible situation of desertion before it success. It is an important challenge for the IES.

Because of this, the paper argues in a preliminary way, the development of a new methodology to prevent a possible desertion, proposing a framework for a suitable approach based on a predictive algorithm capable to identify three important aspects in the academic live of a student:

- the achievement level
- the level of scholar satisfaction
- the constraints of the students

In particular the methodology promotes data mining reinforced by psychometric profiles and complemented with fuzzy logic. All together constitutes a complete framework with the main idea of providing IES with usefulness systems and totally autonomous. With this, the scholar service department will be capable to act in a suitable way before to losing another student.

II. RELATED WORK

A. About Scholar Desertion

Scholar desertion is one of the main issues that IES in Latin America are trying to study in order to establish a possible situation about their weakness. According to [1] this fact have motivated not only IES but also other level of Educational Institutions to try to understand what happened with their students. This subject remains interesting from different points of view, for examples, the professors are concerned about the number of possible groups in order to preserve their lectures; the administrators are interested in catch the major number of possible students; and the students are worried about getting a good professional training in order to get better opportunities for employment.

We can say that desertion if a whole processes, sometimes slow, growing and reinforcing inside the student, whom manifests in their final decision. For instance, in [Vizcaino, 2005] he following kinds of desertion has been detected:

Open Desertion refers to the abandonment from an academic program in an IES.

Program Desertion means that a particular student leaves their professional program, the IES or even the country to continue their professional formation in other IES or country.

Cover Desertion represents the quit of commitments or ideology associated to the fact of coursing a determined scholar program; such situation is beyond to the academic question, implicating the integral formation.

Premature Desertion is a modality in which the student does not enrolled in the IES even when he has been accepted.

Early Desertion is related to the decision of a student to leave their professional formation within the first four periods (particularly in program with a duration of ten

periods).

Later Desertion is referred to the student abandonment in the last six periods of the program.

In this sense, reference [Rodrigues and Sánchez, 2005] proposes another interesting classification taking into account the moment and the duration of the abandonment such as follows:

Intra-semester Desertion: it occurs when the student recurs to a justified (or not) abandonment throughout the period.

Inter-semiannual Desertion: It is presented when a student refused to a new inscription for the next period.

Temporal or Partial Desertion: It means that a student decides to leave the school for a determined time (it is presumed the return of this student).

Total Desertion: It refers to the definitive abandonment of the academic activity (the student never come back to the school).

Reference [2] proposes a classification based on a social behavior of the students related to:

Total Desertion is the definite abandonment of the professional training.

Discriminate Desertion means that the student have the need to leave him studies caused by a particular situation (i.e., social or economic situation).

School Desertion is related to the situation in which the student decides to change of school.

Program Desertion represents the interest of the student to chance of program inside the same school.

Desertion at the First Period involves that a particular student never feels identifies with the university life.

Accumulate Desertion is the sum of desertions in the same IES.

In some particular cases will be necessary to adopt the concept of "recidivism" in order to express that the event of desertion could happen in "more than one episode" [3]. Based on [3] is possible to argue two more types of desertions:

Academic Desertion manifests the regulatory quite by academic failures, also called "academic mortality"

No academic Institutional Desertion refers to a desertion due to rules motivated by different questions not relative with the academic level, which could be referred to behavioral problems or others.

B. Data mining as a detector of scholar desertion.

In recent years, data mining has demonstrated to be a useful and reliable tool capable to be implemented with different proposals. In particular data mining is a subdiscipline of the computer science. The target is to discover knowledge based on data patterns aiming to be valid, novel useful and compressive.

Data mining in education is not a new topic and its study and application has been so relevant in the last years. Such computational technique allows predicting any phenomenon in an educative scenario. According to this, data mining can predicts, with a higher precision, the possibility of the desertion of a student.

In reference [4] the implementation of a process including three modules such that: MS SQL Server to generate a model to store information; SPSS to perform a data

processing; and WEKA software to module a classification of the academic achievement and to detect the desertion pattern of a student.

Besides, in [2] a study to determine the academic achievement and the student abandonment applying KDD (knowledge discovery from Databases), from the store data along the last 15 years is presented. This process was supported by TaryiKDD. This program is a tool of data mining developed by the DCDB laboratory.

For instance, in [3] a paper that shows the usefulness of the TDIDT algorithms to discover a set of particular variables is presented. In particular, these variables outline the behavior of a student before he decides to leave the University. Such process is able to detect the most significant questions involved in the student desertion.

In the background of the IES there are many data about students, used to take decisions concerning to possible problems involved in their academic achievement. In this light, in [6] some data mining techniques have been to look for predicting the scholar desertion implemented in the Technologic University of Izúcar of Matamoros. In particular, an analysis of the socio-economic obtained from the EXANI-II. In specific, two algorithms have been implemented: tree algorithm classification C4.5 and the algorithm k-means. Its first conclusions argue that the implemented data mining methods provide useful information to generate a strategy that determines students with the intention of defect. This is an ad-hoc system developed for their proper students and it is a difficult application to be implemented for other environments. In spite of this, the use of the KDD and the application of the data mining methods could be emulated in the same way.

The development of interfaces or software arises as a needed strategy to solve a determined set of situation according to the proposal of each IES. In [4] is presented a work dedicated to research and propose a methodology that allows identifying students with a higher risk of desertion in an automatic way. To implement this project, the adoption of the CRISP-DM method was required. Such method divided the data mining process in six interactive phases. Neural networks, decision trees and k-means methods were implemented to analyses the student behavior, evaluating factors such that: University selection test, the mean of the obtained calcifications, the gender and the date of the first inscription. Some preliminary results are presented by running experiments on the Engineering School at the University de Las Americas.

The accurate of the models is calculated from a set data. The results indicate that none predictive model was successful because the data provided in the initial input process were not sufficiently reliable.

Another relevant studies aiming to work in this complex but interesting challenge, implement administrative or marketing approaches. For example, [5] affirms that the process of analyzing the variables to profile a deserter student, takes into account different dimensions that are part of the university life.

In this case, [5] presents a research in a lower profile aiming to demographic questions. The main contribution introduces that the acquired information allows interpret the way and the perspective of each student before, during and

after each scholar period. Such information is possible to determine some situation in which a student could be about to leave the school.

In despite of these efforts, the design, development and implementation of artificial intelligence technology in the control of administrative systems are still immature and need to be further studied in order to achieve accuracy and trustworthy applications to build systems regarding to control the desertion and student achievement.

III. THE FRAMEWORK

Artificial intelligence has proven to be a powerful compute science to solve complex and real problems like humans being do. In this sense, the paper proposes a framework capable to predict when students tend to abandon their studies in a suitable way. In particular five phases are needed to reach such predictive methodology (see Fig. 1).

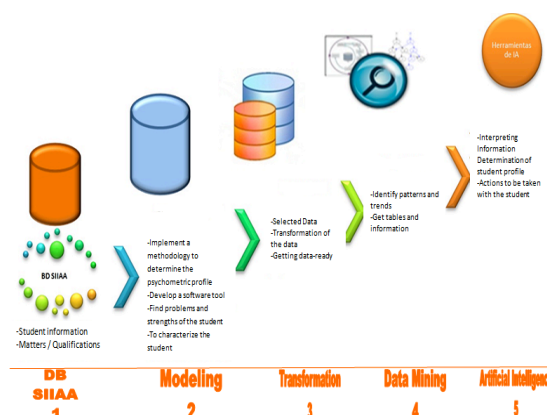


Fig. 1. General scheme of the proposed framework.

Each one of the phases aforementioned is designed from a different engineering perspective, applying separated metaphors to their future implementation. In particular, this methodology is designed to be adapted to the administrative process of the Engineering Faculty (FI) at the Autonomous University of Tamaulipas (UAT), Mexico. Some of the internal systems and procedures will be used as a complement of this proposal.

Phase 1. DB SHIAA. This process is supported by an information system developed by the UAT. In this phase, the personal of the scholar service department SSD is in-charge to revise and upload the data of each student. In particular, such personal requests to the students to complete a form with personal (i.e., name, gender, age, address, and contact information), academic (i.e., career, previous average, etc) and socio-economic information (i.e., work, economic ingress per capita, etc). This data is obtained in the first inscription and revised and updated in subsequent inscriptions.

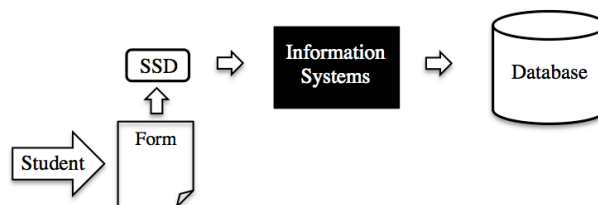


Fig. 2. Data recollection process.

Phase 2. Modeling. The objective of this phase is to perform a profile of each student taking into account the information obtained in the phase 1. In this sense, the implementation of smart user models [8], human value scales [9] or trees of attribute, is a promising idea to model the psychometric and socio-economic profile of each student in a reliable way. Such profiles are indexed in a database. In fact, each profile is classified by means of a number identification to facilitate the administrative processes.



Fig. 3. Generation of student profiles.

Phase 3. Transformation. One of the most transcendental stages of the procedure to avoid scholar desertion is the match between the data and the process of the knowledge discovery. In this sense, this phase is devoted to perform a process to do a data treatment in order to generate a new database with the information ready to be used in the intelligent decision-process.

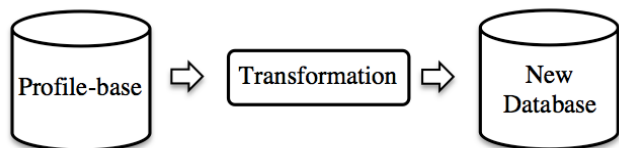


Fig. 4. Resulting database from the KDD.

Phase 4. Data Mining. This phase is devoted to prepare the information from the original database in order to create a new database. To do this, this phase performs a cluster analysis dedicated to group the students based on the profile of the students and the measurement of different variables involved in the previous phase. The objective of this new database is to store the information in a more functional database managing the data for categories. Such characterization allows the framework to generate a suitable prediction to avoid the university abandonment.

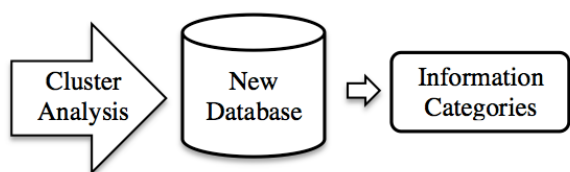


Fig. 5. Information stored in categories.

Phase 5. Artificial Intelligence. The principal proposal of this phase is to provide to the SSD personal with a reliable prediction of the situation of a student and to propose a set of possible actions to do performed in order to avoid the scholar desertion. In particular, fuzzy logic emerges as an ideal method to emulate the analysis developed by humans being. Fuzzy logic is able to examine from different point of solution the variables involved in a particular and special situation.

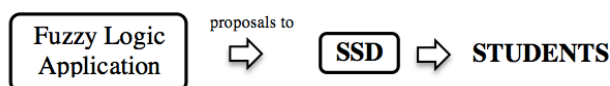


Fig. 6. Generation of possible solutions.

IV. FINAL REMARKS AND DISCUSSION

Scholar desertion is a latent problem in our times. Along this paper, we preset different approaches aiming to avoid such disturbing situation from distinct points of view. In particular, the paper argues the need of performing a detailed study of several variables involved in the situation of the students. The main discussion arises as a consequence of the information provided by the student: how accurate is the information?, the requested information is enough?, the information is well categorized?, how to ensure that the information is reliable?, etc.

In short, the conclusions of this preliminary work allow seeing how academic desertion can be explained from analyzing multiple factors, among which stand: personal, familiar, economics, institutional and vocational issues. In this sense, the proposed framework appears as a proper paradigm to avoid the scholar desertion in a reliable and successful way. Each one of the phases is dedicated to solve a particular situation aiming to provide alternatives to each particular student in an autonomous way. The relevance of the information treatment is one of the main contributions here highlighting.

The knowledge discovery in the databases permits to detect a common denominator among the students that decides to leave the school. Therefore, the generation of a profile-database is quite relevant in order to detect a possible abandonment in a timely manner. The transformation and the clustering phases remains transcendental to store the information ready to the implementation of the fuzzy logic theory. Finally, the paper concluded that the development of the framework combining the use of data mining with fuzzy logic represents a novel and useful approach to reduce the school desertion rate and to improve academic achievement.

REFERENCES

- [1] J. Maya, Identificación y Análisis de Factores Incidentes en el Rerito Voluntario de Estudiantes de Pregrado en el Politécnico Colombiano Jaime Isaza Cadavie – Período 02-2004 a 01-2006, Tesis de Maestría para optar al título de Magister en Ingeniería Administrativa, Universidad Nacional de Colombia, 2008
- [2] R. Timarán, Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos, Conferencia Iberoamericana de Sistemas, Cibernética e Informativa, Orlando, Florida 2009.
- [3] G. Villamizar and L. Pérez, Identification of Motivational and Sociodemographic Factors in Deserts Phycology Students at University Pontificia Boliviana, *Psicogente*, 14 (25): pp. 132-150. Junio, 2011. Universidad Simón Bolívar. Barranquilla, Colombia. ISSN 0124-0137 EISSN 2027-212X <http://www.unisimonbolivar.edu.co/rdigital/psicogente/index.php/psicogente>
- [4] E. S. Fischer, Modelo para la Automatización del Proceso de Determinación de Riesgo de Deserción en Estudiantes Universitarios, Tesis de Magíster, Santiago de Chile, 2012
- [5] L. Pinzon, Applying data mining to education marketing, In *Notas D Marketing*, Vol. 1, Enero 2011.
- [6] E. Pocerl, G. Dapozo and M. López, Modelos predictivos y técnicas de minería de datos para la identificación de factores asociados al rendimiento académico de alumnos universitarios, Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura Universidad Nacional del Nordeste. 9 de Julio No 1449. CP 3400. Corrientes. Argentina, 2010.

- [7] Abril, E., Román, R., Cubillas, M. J. y Moreno, I. ¿Deserción o autoexclusión? Un análisis de las causas de abandono escolar en estudiantes de educación media superior en Sonora, México. *Revista Electrónica de Investigación Educativa*, 10 (1).
- [8] G. González, B. López and J.L.I. de la Rosa, Managing emotions in smart user models for recommender systems, in Proceedings of the 6th International Conference on Enterprise Information Systems. 187–194.
- [9] S. Schwarts, G. Melech, A. Lehmann, S. Burgess, M. Harris and V. Owens, Extending the Cross-Cultural Validity of the Theory of Basic Human Values with a Different Method of Measurement, *Journal of Cross-cultural Psychology*, September 2001.