World Journal on Educational Technology

2 (2009) 118-139

www.world-education-center.org/index.php/wjet

Educational data mining: a sample of review and study case

Alejandro Peña [a, b, c]*, Rafael Domínguez [c], José de Jesús Medel [c]

Researchers of WOLNM [a], ESIME-Z [b] & CIC [c] – [b, c] National Polytechnic Institute, DF, Mexico

Received May 18, 2009; revised July 29, 2009; accepted September 28, 2009

Abstract

The aim of this work is to encourage the research in a novel merged field: Educational data mining (EDM). Thereby, two subjects are outlined: The first one corresponds to a review of data mining (DM) methods and EDM applications. The second topic represents an EDM study case. As a result of the application of DM in Web-based Education Systems (WBES), stratified groups of students were found during a trial. Such groups reveal key attributes of volunteers that deserted or remained during a WBES experiment. This kind of discovered knowledge inspires the statement of correlational hypothesis to set relations between attributes and behavioral patterns of WBES users. We concluded that: When EDM findings are taken into account for designing and managing WBES, the learning objectives are improved.

Keywords: data mining; educational data mining; Web-based Education Systems

©2009 Academic World Education & Research Center. All rights reserved.

1. INTRODUCTION

* Alejandro Peña. Tel.: +52 [55] 5694 0916; fax: +52 [55] 5694 0916
E-mail address: apenaa@ipn.mx, www.wolnm.org/apa

EDM reveals a joint research and application line from WBES and DM in the field of education. WBES aim at providing educational services by means of adaptive and intelligent paradigms (Brusilovsky, 2003). DM is the process devoted to outcome useful and comprehensible knowledge, which was previously unknown, from large information repositories Witten and Frank (2000). Thus, EDM pursues to explain and predict patterns of attributes, behaviors and outcomes that respectively occur and probably happen during WBES sessions. EDM represents a research line whose baseline comes from two interdisciplinary sources.

WBES roots are the earliest computer assisted instruction systems, which started to be built in the 60´s to provide the whole teaching cycle to students (Uhr, 1969). During the 70's, intelligent tutoring systems (ITS) emerged to assist students for solving problems (Burton and Brown, 1979). Since them, several kinds of computer assisted education systems (CAE) have appeared, such as: hypermedia systems (Mathé and Chen, 1996), intelligent learning environments (Brusilovsky, 1995), collaborative learning systems (Ayala and Yano, 1998), and instructional management systems (Instructional Management Systems Global Learning Consortium Inc. [IMS], 2009). Later on, CAE approaches migrated to Internet in order to spread their services world wide. As a result, WBES appeared as a paradigm that meets CAE systems with the support of the artificial intelligence, pedagogy, cognitive, psychology and communication sciences under a holistic view to be deployed on the Web platform. In spite of its short existence, WBES have evolved through several generations as follows: The first one provided static content, while the second generation privileges the user-system interaction. The third generation includes the streaming of lectures in real time, whereas the forth one pursues the personalization of educational services (Sheremetov and Uskov, 2002).

As regards DM, its background essentially is composed of information systems and the statistic. Such underlying elements correspond respectively to the exploitation of information repositories (e.g. files, data bases, data warehouses, online analytical processing, hypertext, hypermedia, Web content…) and the logistics devoted to analyse information in order to outcome new knowledge (Gill and Rao, 1996). Based on this platform, DM represents the top level of the business intelligence hierarchy and the research line that joints several disciplines (e.g. mathematics, artificial intelligence, data bases, decisions theory, information and communication technologies …) (Vit, Luckevic & Misner, 2002).

Essentially, DM tailors two kinds of *models*: descriptive and predictive. *DM descriptive models* are oriented to understand the available information and explain events. *DM predictive models* are devoted to anticipate revealing classifications and behaviors. According to the nature of the problem to be solved, a type of *DM task* is used to deal with it. So classification and regression are DM predictive tasks that respectively assign label values and compute numerical attributes that are unknown for new instances. Whereas, clustering, correlation, and association rules are DM descriptive tasks, which respectively focus on: groups of instances with specific characteristics, similitude degrees between values of a couple of variables, and relationships between attributes that outcome a given result. In addition, these tasks are achieved by *DM techniques* like: algebraic-statistic, Kernel-based, Bayesian, decision trees, rules induction, neural networks, cases based learning, stochastic, frequency counts, relational-declarative-structural techniques, and chaos-density distance based (Hernández, RamÍrez & Ferri, 2008).

Due to the nature of this paper being a mixture of an EDM review and a study case, and its aim is to promote research in this field, the organization of this paper is as follows: In the second section, a profile of several DM techniques is stated in order to set a theoretical background. A sample of EDM research and practical applications is outlined in the third section to show the diversity of EDM approaches. In section four, an EDM study case is pointed out to reveal the characteristics of a descriptive model based on clustering tasks. Moreover, the results fulfilled by descriptive statistics techniques are discussed. Finally, in the conclusions section some comments are given, an analysis of the EDM application is presented and further research is established.

2. METHODS: DATA MINING TECHNIQUES AND ALGORITHMS

In this section, a profile for several DM techniques is given. The description includes the DM tasks that are the object of application for each DM technique. In addition, the DM algorithms that are suitable for implementing a DM technique are outlined. Thereby, this section is organized into three types of DM techniques as follows (Hernández et al., 2008; Romero, Ventura, Espejo & Hervas, 2008; Ayers, Nugent & Dean, 2009).

2.1 Classical Mathematical Data Mining Techniques

One of the most common DM techniques is algebraic and statistics. They are stated as: linear and no linear functions, algebraic formula, distributions, and central tendency measures. This type of DM techniques is useful for regression, classification and clustering DM tasks. Many of these techniques are parametric due to they outcome a pattern based on a given model. Such pattern holds parameters that correspond to the coefficients of the reference model. Some parametric algorithms are: linear regression (Draper and Smith, 1998), logarithmic regression (Gwyn and Angus, 1980), logistic regression (Christensen, 1997), Fisher linear discriminant analysis (Liu and Wechsler, 2002), linear discriminant analysis (Minaei-Bidgoli and Punch, 2003), and least mean square quadratic (Toumazou, 1996). Whereas, statistic instances of non-parametric algorithms are: Chi-square (Satorra and Bentler, 2001), correlation (Snedecor and Cochran, 1980) and independence coefficients (Mahbub Morshed, 2007) for cross tabulation, rank correlation coefficient of Sperman (Lyerly, 1952), and tau rank correlation coefficient of Kendall (J. Cohen, P. Cohen, West & Aiken, 2003; Menard, 2001; Nishishato, 2004).

Other discriminant non-parametric algorithms come from Kernel-based techniques. Usually, they seek a linear discriminant that maximizes the distance between groups or classes. Such techniques produce kernels to enhance the dimensionality of the data. Normally, Kernel-based techniques are suitable for classification and clustering DM tasks. The main technique is the support vector machines (SVM), whose baseline is the Statistic Learning Theory (Boser, Guyon & Vapnik, 1992; Cortes and Vapnik, 1995). Most of the Kernel-based algorithms correspond to different versions and applications of SVM, such as: maximal margin linear SVM (Thornton, Savvides & Kumar, 2004) and soft margin SVM (Li and Long, 1999).

As regards conditional probability, the Bayesian theorem is taken into account to set many of the DM techniques. They estimate the membership probability of a class or group through conditional probabilities. Thus, Bayesian techniques are applied for classification and clustering DM tasks. In descriptive applications, they pursue to identify independency or relevant relationships between variables of a domain. As regards predictive goals, Bayesian networks are used like classifiers. The most popular Bayesian DM algorithms are: naïve Bayes (Dumais, Platt, Heckerman & Sahami, 1998), maximum verosimility (Sanchez, Casillas, Cordon & del Jesus, 2002), expectation maximum (Dempster, Laird & Rubin, 1997).

Frequency counts estimate the occurrences of a chain of events that simultaneously happen. The approach outcomes contingency tables to examine relationships combined frequency distributions for two or more variables. The cells can be frequency counts or relative frequencies. There are many versions of contingency tables, such as: 2x2 (Zelen, 1971), chi square test (Daya, 2002), multidimensional (Everitt, 1992), probability (Everitt, 1992), and degrees of freedom (Mantel, 1963).

Frequency counts are suitable techniques for classification and clustering DM tasks. Some algorithms that can be applied to DM are: stepping stone (Blum, Song & Venkataraman, 2004), bivariate marginal distribution (Pelikan and Muhlenbein, 1999), and Apriori (Agrawal, Mannila, Srikant, Toivonen & Verkamo, 1996; Jovanoski and Lavrac, 2001; Kavsek and Lavrac, 2006).

Cases based learning is another technique that focuses on the former instances. Thus, any new instance is matched against the older ones, whose class or attribute value is known. Those former instances that are more alike are considered to provide or estimate the objective values. Usually, the similarity is computed by distance or density criteria to identify the nearest old instance. Cases based learning is appropriate for clustering and classification DM tasks. Some case based algorithms for DM applications are the following: two-step (Feldman and van Gemund, 2006), COWBEB (Fisher, 1987), agglomerative clustering (Karypis, Han & Kumar, 1999), K-nearest neighbors (Weinberger and Saul, 2009), K means (Steinbach, Karypis & Kumar, 2000), model-based (Fraley and Raftery, 2000), nearest neighbor (Jain, Murty & Flynn, 1999).

2.2 Rules-based Data Mining Techniques

Decision trees represent patterns of instances through a hierarchical flow of tupples. A tupple reveals an attribute-value. The hierarchical order corresponds to a kind of generalization, where the most frequent values attached to an attribute are explicitly stated. Therefore, attributes that give away the highest degree of generalization are represented at top levels of the hierarchy. Attributes with a lower percentage of generalization are set in descending order. Thus, the tupple allocated at the bottom tier corresponds to the predicted or labeled value for a dependent attribute. Decision trees are used for classification, regression, and clustering DM tasks. Many algorithms are based on the divide and conquer technique, such as: ID3 (Quinlan, 1986), C4.5 (Quinlan, 1993), C5.0 (Kotsiantis, 2007), and CART (Breiman, Friedman, Olshen & Stone, 1984). However, CN2 is an algorithm that takes into account the split and conquers technique (Clark and Niblett, 1989).

Rules induction is a kind of learning technique that outcomes rules from a sample of instances stored in a repository of information. A rule embraces an antecedent and a consequent. The antecedent encompasses a set of conjunctive conditions. The consequent reveals a decision regarding the predicted or assigned value for an attribute. Conditions are ordered in a descending hierarchical way in the antecedent. So the main constraints are stated at the beginning of the rule's antecedent. Thus, conditions devoted to precise the consequence instance are outlined at the end of the antecedent. Rules are worthy of using for clustering and classification DM tasks. Some of the main algorithms devoted to induct rules are XCS –learning classifier system proposed by Wilson (1995), supervised inductive algorithm (Marín and Shen, 2008; Corcoran and Wainwright, 1996), and grammar-based genetic programming algorithm (Koza, 1990).

Relational, declarative and structural techniques depict knowledge models by means of functional and logic languages. They are able to define complex predicates to state rules. Such techniques point out data through structures, sketch relationships patterns, and depict knowledge like rules. These kinds of techniques are suitable for classification and association rules DM tasks. One of their main approachs is the inductive logic programming (ILP) that supports relational DM. It corresponds to the joint venture between the inductive learning and the logic programming (Muggleton, 1991). The application of ILP to DM outcomes methods such as: direct approximation through Progol (Muggleton, 1995), approximation by means of propositionalization (Brandt et al., 2001) through LINUS (Lavrac, Dzeroski & Grobelnik, 1991).

2.3 Soft Computing Data Mining Techniques

Neural networks are a non-parametric paradigm that provides several strategies and architectures to generalize patterns. They require training data to estimate the appropriate value for the weights of the links that connect couples of neurons. When a stability condition is reached, the topology of the network and the final value attached to its weights' outcome the learned pattern. Test data is also needed to evaluate the accuracy of the results given by a neural network. They are represented through an organization model such as: perceptron (Freund and Schapire, 1998), multilayer (Kurkova, 1992), radial base (Bishop, 1995), Kohonen (Kohonen and Somervuo, 1998), and backpropagation (Nielsen, 1989). They are suitable for classification and clustering DM tasks. Among the diversity of neural networks algorithms for DM we find: multilayer (Moller, 1993), radial basis function neural network (RBFN) (Broomhead and Lowe, 1988), incremental RBFN (Platt, 1991), decremental RBFN (Broomhead and Lowe, 1988), perceptron (MLPerceptron) (Wildrow and Lehr, 1990), hybrid genetic algorithm neural network (Yao, 1999) and neural network evolutionary programming (Martínez, Hervás, Gutiérrez, Martínez & Ventura, 2006).

Stochastic and fuzzy techniques embrace paradigms such as: simulated annealing, evolutionary computation, and fuzzy logic. These three approaches are respectively devoted to: deal with random components for minimizing multivariate functions, seek patterns in a solution space through successive transformations, and represent uncertain knowledge by membership functions (Homaifar and McCormick, 1995). Soft computing is quite useful for classification and associative rules DM tasks. Among its DM algorithms are: hybrid simulated annealing (Sánchez, Couso & Corrales, 2001), parallel simulated annealing (Kliewer and Tschoke, 2000), LogitBoost (Otero and Sánchez, 2005), MaxLogitBoos (Sánchez and Otero, 2007), AdaBoost (del Jesus, Hoffmann, Junco & Sánchez, 2004), hybrid grammar-based genetic programming (McKinney, et al., 2006), grammar-based genetic programming (Leung and Wong, 2002), Wang & Mendel Chi algorithm (Chi, Yan & Pham, 1996).

3. RESULTS: EDUCATIONAL DATA MINING APPLICATIONS

The application of DM methods in the WBES field is quite incipient. It began in the twenty first century as a result of the development of WBES applications. The sources of information to be mined are heterogeneous. They include data bases of the students' profile, log assessments of the user's interaction with the system, evaluation records, background knowledge, educational content, learning objects, student models, tutoring strategies, meta-data, federative teaching services, and many more repositories. Therefore, a sample of EDM applications is shown in this section according to the source of knowledge.

3.1 Student Modeling

Student models represent information about student's characteristics (e.g. student's knowledge, motivation, skills, personality, and learning preferences). An interesting EDM work oriented to student modeling is the comparison of student skill knowledge methods carried out by Ayers et al., (2009). The study analyzes three methods for estimating students' current stage of skill mastery, such as: common conjunctive cognitive diagnosis model, sum-score method, and capability matrix. Therefore, they try to estimate for a given a topic the degree of skill achieved (e.g. complete, partial, none). In addition, the work identifies groups of students with similar skill profiles through three clustering methods: hierarchical agglomerative clustering, K-means and model-based clustering.

Another EDM application for student modeling corresponds to the classification of students fulfilled by Romero et al., (2008). The data source was retrieved from Moodel assessments and the final record accomplished in several courses. Afterwards, they applied discretization and rebalance preprocessing techniques on the former quantitative data to explore better classifier models. The aim of this work is to classify students with similar final marks into several groups according to the activities achieved with WBES. Among the predictive DM methods applied are: statistical classifier, decision tree and rule induction. Moreover, an example of algorithm used respectively for each method is: ADLinear, C45 and CN2.

3.2 Tutoring

Tutoring corresponds to the traditional support that a human tutor offers to students to solve problems of a specific domain. This kind of functionality is fully implemented in ITS and the sequencing module of WBES. Regarding the application of DM in the tutoring field, the work achieved by Barnes, Stamper, Lehman & Croy (2008) uses hints generated from historical data to develop logic proofs. Hints are outcomed by a reinforcement learning technique based on Markov decision processes.

As regards the framework stated by Guo and Zhang (2009), it uses DM algorithms based on evolutionary computation to characterize dynamic learning processes and learning patterns for encouraging students' apprenticeship. The approach supports tutoring and collaboration functionalities to provide content that meet students' accessibility needs and preferences. The framework, also, pursues to match content to students' devices. These kinds of services are valuable for people with special abilities

3.3 Content

Content corresponds to the knowledge domain resources that are tailored to teach a lesson, record the students' behavior, and evaluate students' apprenticeship. This resource is a kind of learning object that contains text, sound, image, video, virtual reality, animation, and many more multimedia options. An example of the DM application to content is given by Pavlik , Cen & Koedinger (2009). They set a transfer model of the knowledge domain of related practice item-types using learning curves. The item-types mean a set of practice items that are alike. Such a model represents the pairwise knowledge component relationships between item-types in the domain. Thus, it maps a Q-matrix whose rows reveal item-types and its columns correspond to knowledge components (e.g. skills, facts, procedures or concepts). In addition, the approach carries out an exhaustive search for pairwise relationships by means of a learning factor transfer algorithm. The model accepts Bayesian information criterion weight test. As a result, it achieves improvements above a probability reference.

Another DM contribution to the content line is the work fulfilled by García, Romero, Ventura & Castro (2009). They built a system to find, share and suggest the suitable modifications to improve the effectiveness of a course and its content. Their approach includes rule mining to discover valuable information through students' assessments like "if-then" recommendation rules. The system holds a collaborative recommender module to share and score the recommendation rules obtained by teachers and specialists in education with common profiles.

3.4 Assessment

The record of the user interaction with a WBES during each session is fulfilled by the assessment module. Based on the information stored, it is possible to supervise the behavior, performance, outcomes, customs, preferences, and many more issues about: who is the student? And what has she/he been doing? As an instance of DM applications to assessment, there is a method for mining multiple-choice assessment data set by Madhyastha and Hunt (2009). The method estimates similarity of the concepts given by multiple choice responses. As an outcome, a similarity matrix shows the distance between concepts in a lower-dimensional space. Such a view reveals the relative difficulty of concepts among the students. In addition, concepts are clustered, and unknown responses in the context of previously identified concepts are acknowledged. The method is used to answer questions related to the similarity of concepts and the difficulty of convincing students to modify an erroneous concept.

With the aim of focusing on the DM processes, Pechenizkiy, Trčka, Vasilyeva, Aalst & De Bra (2009) stated a DM research line called "Process Mining". The line pursues the development of mining tools and techniques devoted to extract processes-related knowledge from event logs recorded by the system. One EDM application of process mining is devoted to analyze assessment data. The approach analyses assessments from recently organized online multiple choice tests. It, also, demonstrates the use of process discovery, conformance checking and performance analysis techniques.

4. DISCUSSION: EDUCATIONAL DATA MINING STUDY CASE

The application of a DM task to a WBES experiment is outlined in this section. The WBES trial was devoted to enhance the apprenticeship of people through the student centered education paradigm (Peña, 2008). Therefore, several psychological and knowledge domain tests were applied to a universe of 200 participants in order to set a student model. Afterwards, a pre-measure, teaching-learning stimulus, and a post-measure were provided to a sample of 18 individuals. At the end, statistical processes were carried out to estimate and interpret the results. However, this assessment of the WBES experiment reveals a high desertion of volunteers during the trial. Hence, a question is stated in this study case: which attributes characterize participants that deserted and accomplished the experiment? The response to such question is given by the application of a clustering DM task to the data base of volunteers as follows:

4.1 Source Information

The information to be mined is a data base of volunteers. Where, each row corresponds to a profile of the participant. The columns contain personal, academic and labour attributes, such as: id, name, trial-status (e.g. deserted or accomplished), age, group of age (e.g. 16-17, 18-19, 20-24, 25-34...), gender, civil status (e.g. single, married, divorced…), occupation (e.g. name, type and number of occupations, whose instance value could be: professor, academy and 1 respectively), highest academic degree (e.g. Bachelor, Master, Doctoral), Bachelor's record (e.g. status, school type, school prestige, study area, whose possible instance value is: student, public, excellent, medical and social sciences respectively).

4.2 Segementation Algorithm

A segmentation algorithm was used to discover patterns that depict volunteers that abandoned and finished the WBES trial. The algorithm organizes iteratively records from the source information into distinctive groups. The technique clusters items with specific instance values for a set of *input attributes*. The instance values hold high frequencies of occurrence. Thereby, items of a given group share common properties and set a meaningful relationship. However, such properties are different from the instance values attached to the same set of input attributes that are held by items of other clusters. Likewise, the relationship between items reveals different meaning for other groups. What is more, the algorithm estimates how well the clusters depict groupings of the points. Moreover, it tries to redefine the groupings to create clusters that are more suitable to characterize the data. The process ends when the algorithm is not able to further improve the outcomes by redefining the groups. The definition of a cluster model to be accomplished requires three types of attribute: 1) *id*: is the column that uniquely identifies each row; 2) *input*: is a set of one or more attribute whose values are used to outcome the clusters; 3) *predictable*: is an attribute whose values are used as references for further rows that lack such instance values.

4.3 Mining Process

The source information is stored in a data base, where 82 records correspond to women and 118 rows concern men. The clustering models that were designed contain as *id* and *predictable* attributes the volunteer-id and the trial-status respectively. According to the input attributes, the cluster model focuses on a specific study target. Therefore, more than 20 clustering models were outcome by the segmentation algorithm. Among the input attributes that were considered are the following: number of occupations, type of occupations, highest academic degree, Bachelor's school prestige, Bachelor's status, group of age, civil status, Bachelor's study area, Bachelor's school type, and many others. The result for each cluster model is outlined through several elements. One of them is a cluster diagram, where nodes correspond to groups and arcs represent relationships between couples of clusters. Another product is a cluster's profile that is pictured as a bar chart. The third element is a frequency table for the instance values attached to the attributes. An example of such results is pictured in Figures 1 to 3.

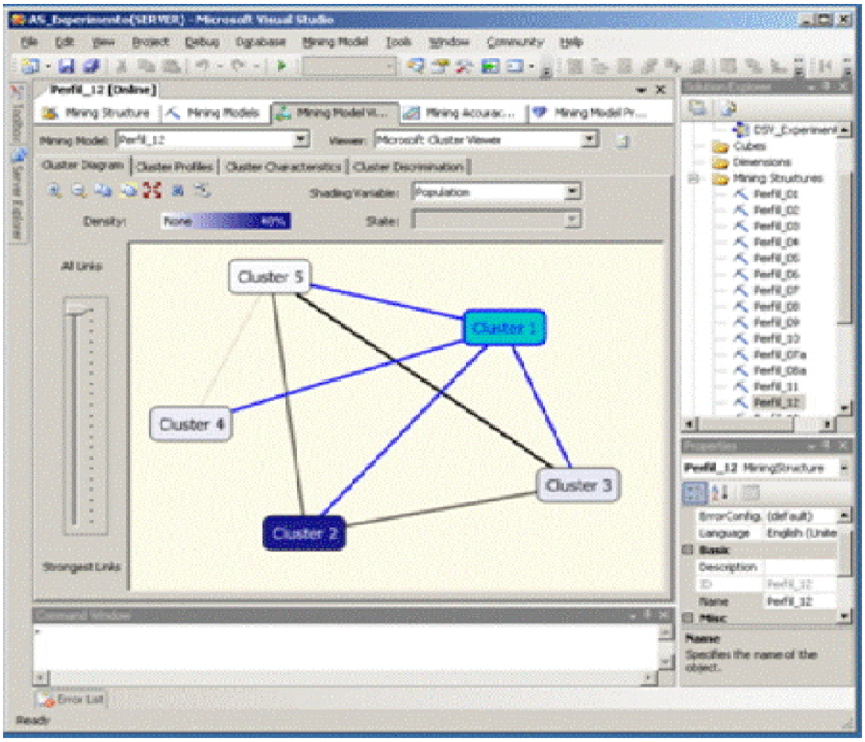

Figure 1. *A cluster diagram for the five groups that correspond to the volunteers' civil status*

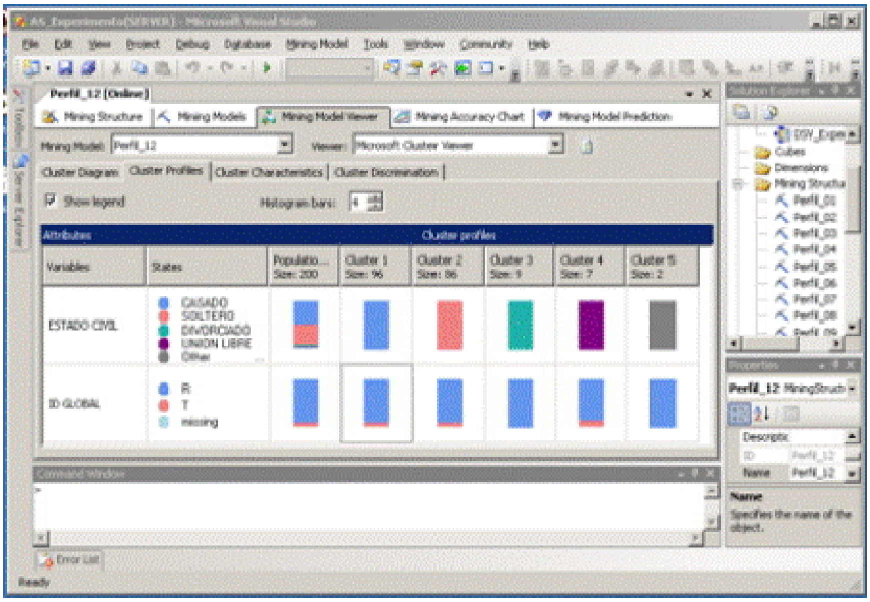

Figure 2. *A cluster's profile for the five groups that correspond to the volunteers' civil status*

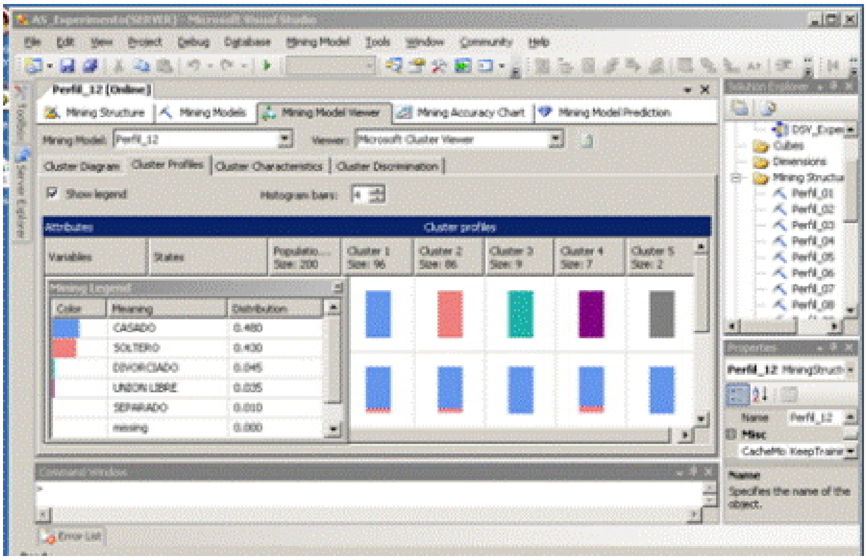

Figure 3. *A frquency table for the five groups that correspond to the volunteers' civil status*

4.4 Evaluation

Once the cluster models were outcome, an evaluation was achieved to identify revealing patterns. Such patterns reveal some highlights to answer the study case question. One of them concerns the number of occupations based on the frequencies stated in Table 1. In that table, there are two occupation attributes: number and type. The first holds two instance values; whereas the second attribute owns nine.

Table 1. *Frequencies for number and types of occupation from the volunteers universe, deserters and reliable subjects*

Attribute / Instance value	Volunteers Universe Absolute and percentage freq.	Set of volunteers, who deserted Absolute and percentage freq.	Set of volunteers, who finshed Absolute and percentage freq.
Ocuppation number:			
- 1	- 82 – 41.0%	- 78 – 95.1%	- 4 – 4.9%
- 2	- 118 – 59.0%	- 104 – 88.1%	- 14 – 11.9%
Ocuppation types			
- Accademy - Office	- 63 – 31.5%	- 57 – 90.1%	- 6 – 9.9%
- Office	- 53 – 26.5%	- 51 – 96.1%	- 2 – 3.9%
- Academy	- 46 – 23.0%	- 43 – 93.6%	- 3 – 6.4%
- Academy – Sports	- 14 – 7.0%	- 11 – 82.1%	- 3 – 17.9%
- Academy – Arts	- 8 – 4.0%	- 6 – 72.5%	- 2 – 27.5%
- Office – Sports	- 8 – 4.0%	- 7 – 82.9%	- 1 – 17.1%
- Office – Arts	- 4 – 2.0%	- 4 – 100%	- 0 – 0%
- Arts	- 3 – 1.5%	- 3 – 100%	- 0 – 0%
- Sports	- 1 - 0.5%	- 1 - 100%	- 0 – 0%

The second column shows frequencies for each instance value from a universe of 200 volunteers. The third and forth columns correspond respectively to participants that deserted and successfully completed the WBES trial. Their frequencies that appear in a given row are complementary, because their total corresponds to the frequency that appears in the same row of the second column.

4.5 Findings

Based on the mining and evaluations outcomes accomplished for every clustering model, several revealing patterns were found to respond to the question stated for the study case. The patterns hold the following attributes: 1) number and type of occupations; 2) highest academic degree; 3) civil status; 4) Bachelor's status, study area, school prestige and type. Based on such findings, a set of correlational hypothesis are stated to characterize volunteers that successfully accomplished the experiment and abandoned the trial as follows:

A high level of trial achievement is carried out when the volunteer:

- develops more than one activity;
- holds a high academic degree;
- studies (or studied) in a Bachelor's school whose prestige is "excellent";
- is actually studying a Bachelor's degree.

A high degree of desertion is outcome when the participant:

- accomplishes just one activity;
- holds a low academic degree;
- is divorced or separated from her/his partner;
- her/his Bachelor's school type is "private".

In addition, two complementary behavioral hypotheses are set: a similar level of achievement and desertion is outcome for groups whose volunteers:

- correspond to age groups: 18-19, 20-24, 25-34, 35-44;
- are married or single;
- hold a Bachelor's degree in the area of Physics-Mathematics-Engineering sciences, or Social and Administrative sciences;
- study (or studied) in a Bachelor's school whose prestige is qualified as "high" or "medium".

5. CONCLUSION

The application of DM to the educational field reveals underlying knowledge that is behind the use of WBES. The descriptive findings fulfilled by EDM help to understand learning outcomes, identify students' behavior, characterize groups of students, and explain some patterns. The predictions achieved by EDM encourage the use of proactive strategies in WBES. When the anticipation factor is taken into account in the design and maintenance of WBES, their teaching services are tailored to most of the students. In consequence, such guidelines support the adaptation of WBES in an intelligent way.

In this work a profile of the EDM was introduced with the aim of promoting research and application. Also, a collection of DM tasks, methods and algorithms was outlined to identify a conceptual framework to develop EDM projects. Moreover, the application of DM to the educational systems was stated for several modules of typical WBES in order to define work targets. Likewise, a study case was pointed out to explain the application of the clustering DM task to reveal characteristics of volunteers, who deserted and abandoned a WBES trial.

The findings discovered and hypothesis stated during the trial offer some highlights responding to the question study case. The patterns of attributes depict people, whose behavior reveals high degrees of achievement or desertion. In consequence, we can resume that: individuals that commit high academic degrees, every day develop several activities, and study(ed) their Bachelor in prestigious schools tend to become reliable people. Whereas, people that deal with family issues, just carry out a single duty, and have not achieved postgraduate studies, are disposed to abandon their commitments.

As a future work we set the application of several DM tasks and techniques to find more patterns. Also, the exploration of personality, cognitive and learning preferences attributes is considered. In addition, the inclusion of more volunteers to the experiment is planned to confirm or reject the former outcomes.

ACKNOWLEDGMENTS

First author gives testimony of the strength given by his Father, Brother Jesus and Helper, as part of the research projects of World Outreach Light to the Nations Ministries (WOLNM). Also, the work owns a support from grants given by: CONACYT-SNI 36453, CONACYT 118862, IPN-SIP-EDI: DOPI/3189/08, IPN-SIP-20090068, IPN-COFAA-SIBE.

REFERENCES

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., and Verkamo, A. (1996). Fast discovery of association rules. In: U.M. (Eds. Fayyad et al.) Advances in knowledge discovery and data mining (pp. 307-328). Massachusetts: MIT Press.

Ayers, E., Nugent, R., and Dean, N. (2009). A pilot study on logic proof tutoring using hints generated from historical student data. Paper presented at the 2nd International Conference on Educational Data Mining (EDM'09), Cordoba, Spain.

Ayala, G. and Y. Yano (1998). Collaborative learning environment based on intelligent agents. *Journal for Elsevier, Expert Systems with Applications,* 14(1), 129-137.

Barnes, T., Stamper, J., Lehman, L., and Croy, M. (2008). A pilot study on logic proof tutoring using hints generated from historical student data. Paper presented at the 1st International Conference on Educational Data Mining (EDM'08), Montreal, Canada.

Bishop, C.M. (1995). Neural networks for pattern recognition. Oxford: Oxford University Press.

Blum, A, Song, D., and Venkataraman, S. (2004). Detection of interactive stepping stones: Algorithms and confidence bounds. Paper presented at the 7th International Symposium in Recent Advances in Intrusion Detection, RAID 2004, Sophia Antipolis, France.

Boser, B., Guyon, L., and Vapnik, V. (1992). A training algorithm for optimal margin classifiers. Paper presented at the 5th Annual ACM Workshop on Computational Learning Theory, COLT, Pittsburgh, Pennsylvania, United States.

Brandt, N., Brockhausen, P., de Haas, M., Kietz, J.U., Knobbe, A., Rem, O., and Zücker, R. (2001). Mining multi-relational data. (Technical report, IST Project MiningMart, IST-11993).

Breiman, L., Friedman, J.H., Olshen, R.,and Stone, C. (1984). Classification and regression trees. New York: Chapman & Hall.

Broomhead, D., and Lowe, D. (1988). Multivariable functional interpolation and adaptative networks. *Complex Systems,* 2(1988), 321-355.

Brusilovsky, P. (1995). Towards adaptive learning environments. Paper presented at the Gesellschaft für Informatik und der Schweizer Informatiker Gesellschaft (GISI-95), Zürich, Switzerland.

Brusilovsky, P. (2003). Adaptive and intelligent Web-based education systems. *Journal for Artificial Intelligence in Education*, 13(2003), 156-169.

Burton, R.B., and Brown, J.S. (1979). An investigation of computer coaching for informal learning activities. (Eds. Sleeman and Brown) Intelligent Tutoring Systems (pp. 77-98) New York: Academic Press.

Chi, Z., Yan, H., and Pham, T. (1996). Fuzzy algorithms with applications to image processing and pattern recognition. Singapore: World Scientific.

Christensen, R. (1997). Log-linear models and logistic regression. 2nd Ed. New York: Springer.

Clark, P., and Niblett, T. (1989). The CN2 induction algorithm. *Journal for Machine Learning, Kluwer*, 3(4), 261-283.

Cohen J., Cohen P., West S.G., and Aiken L.S. (2003). Applied multiple regression and correlation analysis for the behavioral sciences. 3rd Ed. Mahwah: SAGE Publications Inc.

Corcoran, A. and Wainwright, R. (1996). Reducing disruption of superior building blocks in genetic algorithms. Paper presented at the ACM symposium on Applied Computing, Pennsylvania, United States.

Cortes, C. Vapnik, V. (1995). Support vector networks. Journal for Machine Learning , 20(3), 273-297.

Daya, S. (2002). Chi-square test for larger (r x c) contingency tables. *Journal for Evidence-based Obstetrics & Gynecology*, 3(2), 59-60.

Dempster, A.P., Laird, N.M., and, Rubin, D.B. (1997). Maximum likelihood from incomplete data via the EM algorithm. *Journal for Royal Statistic Sociaty*, 39, 1-39.

Draper, N.R., and Smith, H. (1998). Applied regression analysis. 3rd Ed. UK: Wiley Series in Probability and Statistics.

Dumais, S., Platt, J., Heckerman, D., and Sahami, M. (1998). Inductive learning algorithms and representations for text categorization. Paper presented at the seventh international conference on information and knowledge management, Maryland, USA.

Everitt, B. S. (1992). The analysis of contingency tables. 2nd Ed. USA: Chapman & Hall/CRC.

Feldman, A., and Gemund, A. (2006). A two-step hierarchical algorithm for model-based diagnosis. Paper presented at the 21st National Conference on Artificial intelligence, Boston, United States.

Fraley, C., and Raftery, A.E. (2000). Model-based clustering, discriminant analysis, and density estimation. (Technical Report no. 380). University of Washington, U.S. Department of Statistics.

Fisher, D. (1987). Knowledge acquisition via incremental conceptual clustering. *Journal for Machine Learning*, 2(2), 139-172.

Freund, Y., and Schapire, R. E. 1(998). Large margin classification using the perceptron algorithm. Paper presented at the 11th Conference on Computational Learning Theory (COLT' 98), Wisconsin, USA.

García, E., Romero, C., Ventura, S., and Castro, C. (2009). An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering. *Journal for User Model and User Adapted Interaction,* 19(1-2), 99-132.

Gill, H., and Rao, P. (1996). The official guide to datawarehousing. USA: Prentice Hall.

Greenwood, P., and Nikulin, M. (1996). A guide to chi-squared testing. New York: Wiley.

Guo, Q., and Zhang, M. (2009). Implement web learning environment based on data mining. *Elsevier Knowledge Based Systems*, 22 (6), 439-442.

Gwyn, A.E., and Angus, D. (1980). Testing linear versus logarithmic regression models. *Journal for Review of Economic Studies*, 47(1), 275-291.

Hernandez, J., Ramirez, M.J., and Ferri, C. (2008). Introducción a la minería de datos. México: Prentice Hall.

Homaifar, A., and McCormick, E. (1995). Simultaneous esign of membership functions and rule sets for fuzzy controllers using genetic algorithms. *Journal for IEEE Transactions on Fuzzy Systems*, 3(2), 129-139.

IMS (2009). Instructional Management Systems Global Learning Consortium Inc http://www.imsglobal.org. [November 20, 2009].

Jain, A.K., Murty, M.N., and Flynn, P.J. (1999). Data clustering: A review. *Journal for ACM Computing Surveys*, 31(3), 264-323.

Jesus, del., Hoffmann, M., Junco, F.,and Sánchez, L. (2004). Induction of fuzzy-rule-based classifiers with evolutionary boosting algorithms. *Journal for IEEE Transactions on Fuzzy Systems*,12(3), 296-308.

Jovanoski, V., and Lavrac, N. (2001). Classification rule learning with APRIORI-C. Paper presented at the10th Portuguese Conference on Artificial Intelligence, Knowledge Extraction, Multi-agent Systems, Logic Programming and Constraint Solving, Porto, Portugal.

Karypis, G., Han, E.H., and Kumar, V. (1999). Chameleon: Hierarchical clustering using dynamic modeling. *Journal for IEEE Computer*, 32(8), 1-22.

Kavsek, B., and Lavrac, N. (2006). APRIORI-SD: Adapting association rule learning to subgroup discovery. *Journal for Applied Artificial Intelligence*, 20(7), 543-583.

Kliewer, G., and Tschoke, S. (2000). A general parallel simulated annealing library and its application in airline industry. Paper presented at the 14th International Parallel and Distributed Processing Symposium (IPDPS 2000), Cancun, Mexico, 2000.

Kohonen, T., and Somervuo, P. (1998). Self-organizing maps of symbol strings. Journal for Neurocomputing, 21(1-3), 19-30.

Kotsiantis, S.B. (2007). Supervised machine learning: A review of classification techniques. *Journal for Informatica*, 31(3), 249-268,

Koza, J. (1990). Genetic programming: a paradigm for genetically breeding populations of computer programs to solve problems. (Technical Report: CS-TR-90-1314) Stanford University, USA, Computer Science Department.

Kurkova, V. (1992). Kolmogorov's theorem and multilayer neural networks. *Journal for Neural networks*, 5(3), 501-506.

Lavrac, N., Dzeroski, S., and Grobelnik, M. (1991). Learning nonrecursive definitions of relations with LINUS. Paper presented at the 5th European Working Session on Learning, Porto, Portugal.

Leung,K.S., and Wong, M.L. (2002). Data mining using grammar based genetic programming and applications, France: Lavoisier

Li, Y., and Long, P. M. (1999). The relaxed online máximum margin algorithm. Paper presented at Conference on Advances in Neural Information Processing Systems, NIPS, Denver, Colorado, USA, Journal for New York: Springer, 46(1-3), 361-387.

Liu, Ch., and Wechsler, H. (2002). Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *Journal IEEE Transaction Image Processing*, 11(4), 467-476.

Lyerly, S. B. (1952). The average spearman rank correlation coefficient. Journal for New York: Springer, 17(4), 421-428.

Madhyastha, T., and Hunt, E. (2009). Mining diagnostic assessment data for concept similarity. *Journal of Educatinal Data Mining,* 1(1), 1-19.

Mantel, N. (1963). Chi-square tests with one degree of freedom; extensions of the mantel- Haenszel procedure. *Journal of the American Statistical Association*, 58(303), 690-700.

Mahbub Morshed, A.K.M. (2007). Is there really a "border effect?" *Journal for International Money and Finance*, 26(7), 1229-1238.

Marín, J., and Shen, Q.(2008). A fuzzy-XCS classifier system with linguistic hedges. *Journal for Computational Intelligence Research*, 4(4), 329–34.

Martínez, F., Hervás, C., Gutiérrez, P., Martínez, A., and Ventura, S. (2006). Evolutionary product-unit neural networks for classification. Paper presented at the 7th Conference on Intelligent Data Engineering and Automated Learning, Burgos, Spain.

Mathé, N., and J. Chen (1996). User-centred, indexing for adaptive information access. *Journal for User Modeling and User-Adapted Interaction,* 6(2-3), 225-261.

McKinney, B.A., Crowe, J.E., Voss, H.U., Crooke, P.S., N. Barney, N., and Moore, J.H. (2006). Hybrid grammar-based approach to nonlinear dynamical system identification from biological time series, Journal for APS Physical. Review E, 73(2), 1-7.

Menard S. (2001). Applied logistic regression analysis. 2nd Edition, Mahwah: SAGE University Series on Applicative Applications in the Social Sciences.

Nielsen, R.H. (1989). Theory of the backpropagation neural network. Paper presented at the International Joint Conference on Neural Networks, Washington DC, USA.

Minaei-Bidgoli, B., and Punch, W. (2003). Using genetic algorithms for data mining optimization in an educational Web-based system. (Eds. E. Cantú-Paz et al.). Genetic and Evolutionary Computation, Part II. LNCS 2724 (pp.2252–2263). Berlin Heidelberg: Springer-Verlag

Moller, M. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Journal for Neural Networks*, 6(4), 525-533.

Muggleton, S. (1991). Inductive logic programming. *Journal for New Generation Computing*, 8(4), 295-318.

Muggleton, S. (1995). Inverse Entailment and Progol. *Journal for New Generation Computing,* 13(3) 245-286.

Nishishato S. (2004). Dual scaling. 1st Ed. Mahwah: SAGE University Series on Applicative Applications in the Social Sciences.

Otero, J., and Sánchez, L. (2005). Induction of descriptive fuzzy classifiers with the logitboost algorithm. *Journal for Soft Computing*, 10(9), 825-835.

Pavlik, P., Cen, H., and Koedinger, K. (2009). Learning factors transfer analysis: Using learning curve analysis to automatically generate domain models. Paper presented at the 2nd International Conference on Educational Data Mining (EDM'09), Cordoba, Spain.

Platt, J. (1991). A resource allocating network for function interpolation. *Journal for Neural Computation*, 3(2), 213-225.

Pechenizkiy, M., Trčka, N., Vasilyeva, E., Aalst, W., and De Bra, P. (2009). Process mining online assessment data. Paper presented at the 2nd International Conference on Educational Data Mining (EDM'09), Cordoba, Spain.

Pelikan, M., and Muhlenbein, H. (1999). The bivariate marginal distribution algorithm. (Eds. R. Roy, et al.). Advances in Soft Computing-Engineering Design and Manufacturing (pp. 521-535). London: Springer.

Peña, A. (2008). A student model based on cognitive maps. National Polytechnic Institute, Center of Computer Research, PhD thesis.

Quinlan, J., (1986). Induction of decision trees. *Journal for Machine Learning*, 1(1), 81-106.

Quinlan, J. (1993). C4.5: Programs for machine learning. San Francisco: Morgan Kaufman.

Romero, C., Ventura, S., Espejo, P., and Hervas, C (2008). Data mining algorithms to classify students. Paper presented at 1st International Conference on Educational Data Mining (EDM'08), Montreal, Canada.

Toumazou, Ch. (1996). Circuits and systems tutorials. New York: IEEE Press.

Sánchez, L., Couso, I., and Corrales, J.(2001). Combining GP operators with SA search to evolve fuzzy rule based classifiers. *Journal of Information Sciences*, 136(1-4), 175-191.

Sánchez, L., and Otero, J. (2007). Boosting fuzzy rules in classification problems under single-winner Inference. *Journal for Intelligent Systems*, 22(9), 1021-1034.

Sanchez, L., Casillas, J., Cordon, O., and del Jesus, M.J. (2002). Some relationships between fuzzy and random classi6ers and models. *Journal for Approximate Reasoning*, 29(2), 175–213.

Satorra, A., and Bentler, P. M. (2001). A scaled difference chi-square test statistic for momento structure analysis. *Journal for Psycometrika*, 66(4), 507-514.

Snedecor, G.W., and Cochran, W.G. (1980). The sample correlation coefficient and Properties. In Statistical Methods. 7th Ed. Ames: Iowa State Press, 175-178.

Steinbach, M., Karypis, G., and Kumar, V. (2000). A comparison of document clustering techniques. Paper presented at Knowledge Discovery and Data Mining (KDD) Workshop on Text Mining, Ljubljana. Slovenia.

Sheremetov, L. y V. Uskov (2002). Hacia la nueva generación de sistemas de aprendizaje basado en la Web. Revista Computación y Sistemas, 584), 256-267.

Thornton, J., Savvides, M., and Vijaya Kumar, B.V.K. (2004). Linear shift-invariant maximum margin SVM correlation filter. Paper presented at the Conference Intelligent Sensors, Sensor Networks and Information Processing, Melbourne, Australia.

Uhr, I. (1969). Teaching machine programs that generate problems as a function of interaction with students. Paper presented at 24th National Conference, 125-134.

Vit, E., Luckevic, M., and Misner, S. (2002). Making better business intelligence decisions faster. Redmon, Washington: Microsfot Press.

Yao, X. (1999). Evolving artificial neural networks. Journal for IEEE, 87(9), 1423-1447.

Weinberger, K.Q., and Saul, L.K. (2009). Distance metric learning for large margin nearest neighbor classification. *Journal for Machine Learning Research*, 10, 207-244.

Wildrow, B., and Lehr, M. (1990). 30 Years of adaptive neural networks: Perceptron, Madaline, and Backpropagation. Journal for IEEE, 78(9), 1415-1442.

Wilson, S.W. (1995). Classifier fitness based on accuracy. *Journal for Evolutionary Computation*, 3(2), 149-176.

Witten. I. H., and Frank, E. (2005). Data mining: Practical machine learning tools and techniques. 2nd Ed. San Francisco: Morgan Kaufmann.

Zelen, M. (1971). The analysis of several 2 x 2 contingency tables. *Journal for Biom*etrika, 58(1), 129-137.