EXTRACTION OF LOGICAL RULES TO DESCRIBE STUDENTS’ LEARNING BEHAVIOR

Félix Castro & Ángela Nebot
Departament de LSI
Universitat Politècnica de Catalunya
Campus Nord
Jordi Girona Salgado 1-3
Barcelona 08034, Spain
fcastro@lsi.upc.edu, angela@lsi.upc.edu

Francisco Mugica
Instituto Latinoamericano de la Comunicación Educativa (ILCE)
Calle del Puente 45
México D.F., México
fmugica@ilce.edu.mx

ABSTRACT. E-learning offers a new context for education where large amounts of information describing the continuum of the teaching-learning interactions are endlessly generated and ubiquitously available. But raw information by itself may be of no help to any of the e-learning actors. The use of Data Mining methods to extract knowledge from this information can, therefore, be an adequate approach to follow in order to use the obtained knowledge to fit the educational proposal to the students’ needs and requirements. In this brief study we use an extension of Fuzzy Inductive Reasoning methodology to extract comprehensible, actionable and reasonable set of rules describing the students’ learning behavior. The obtained rules can be used to improve the system understanding and to provide valuable information to tutors about the course performance. The extraction rules model presented in this research is applied to a real virtual campus graduate course of the Center of Studies in Communication and Educational Technologies (CECTE, as Spanish acronym).

KEY WORDS: Virtual campus; e-learning; Fuzzy inductive reasoning; Rules extraction.

1 Introduction

Any e-learning system is, by its own nature, likely to generate large amounts of information describing the continuum of the teaching-learning interactions almost in real time. All this information, gathered from diverse and usually heterogeneous sources, may be of no help by itself to any of the e-learning actors in its raw form. Actually, an excess of such information can become a liability for e-learning tutors and managers unless it is processed according to reasonable goals. Data Mining can provide the adequate tools for such processing, obtaining actionable patterns from large data repositories. The use of Data Mining methods to extract knowledge from the e-learning system available information can, therefore, be an adequate approach to follow in order to use the obtained knowledge to fit the educational proposal to the students’ needs and requirements.

Virtual campus environments, such as the one that is subject of this case study, are fastly becoming a mainstream alternative to traditional distance higher education. The Internet medium they use to convey content, also allows the gathering of information on students’ online behavior. Here, we focus on e-learning systems improvement through the analysis of the data generated by the virtual campus students, aiming to discover their system usage patterns. Most of this research uses Soft Computing techniques to analyze the available data. In turn, we can distinguish diverse Soft Computing-based approaches to e-learning process analysis, i.e. methods to classify students’ based on their usage patterns on a web-based course [1, 2, 3], methods oriented towards system personalization [4]. For instance, a neural network model is proposed in [5] to recommend an adequate navigation strategy for the user. A methodology to improve the performance of developed courses through adaptation, using Evolutionary algorithms, is presented in [6]. And, finally, methods that allow automatic detection of atypical students’ behavior such as the Bayesian predictive distribution model to detect irregular learning proposed in [7], and the Generative Topographic Mapping model to detect atypical behavior on the grouping structure of the users of a real virtual campus, presented in [8].

Research concerned with the analysis of data generated by the use of e-learning systems is still scarce, and there is a lack of standard methods and guidelines to address some open problems in distance education. Extraction rules methods applied to e-learning have been investigated in the areas of learning recommendation systems [9, 10], learning material organization [11] and student learning assessments [12]. Association rule mining, inter-session and intra-session frequent pattern mining, were applied in [9, 10] to extract useful patterns that might help educators, educational managers and Web masters to evaluate and interpret on-line course activities. A similar approach can be found in [13], where contrast rules, defined as sets of conjunctive rules describing patterns of performance disparity between groups of students, were used.

Artificial neural networks for the prediction of the students’ final marks were applied in [12] and the predictions made by the networks were interpreted using orthogonal search-based rule extraction (OSRE), in order
to discover interesting behavioral rules in student usage information. In [14] an a priori algorithm for association rules was applied to capture relationships among URL references based on the navigational patterns of students.

For a deeper and more detailed insight into the data mining methods applied to e-learning systems, the reader is referred to [15].

In this paper we present an extension of the Fuzzy Inductive Reasoning (FIR) methodology that allows the extraction of logical rules from a previous identified FIR model. The main goal of the FIR model is to forecast the final mark of the students, and to determine the most relevant features involved in this process. In the application at hand, the generated set of logical rules describes students’ learning behavior patterns enrolled in a CECTE virtual course.

The remaining of the paper is organized as follows: section 2 presents the basics of the FIR methodology and the rules extraction method. A description of the data collected from the analyzed web-based course is provided in section 3. Results from the experiments are presented and discussed in section 4. Finally, section 5 summarizes the results obtained in this work with some conclusions.

2 The Fuzzy Inductive Reasoning Methodology

The conceptualization of the Fuzzy Inductive Reasoning (FIR) methodology arises from the General Systems Theory proposed by Klir [16]. This modeling and qualitative simulation methodology is based on systems behavior rather than on structural knowledge. It is able to obtain good qualitative relations between the variables that compose the system and to infer the future behavior of that system. It also has the ability to describe systems that cannot easily be described by classical mathematics (e.g. differential equations), i.e. systems for which the underlying physical laws are not well understood. FIR consists of four main processes, namely: fuzzification, qualitative model identification, fuzzy forecast and defuzzification. Fig. 1 describes the structure of the FIR methodology.

The fuzzification (discretization) process converts quantitative data stemming from the system into fuzzy data. The qualitative model identification process is responsible for finding causal and temporal relations between variables and therefore for obtaining the best model that represents the system. A FIR model is composed of a mask (model structure) and a pattern rule base (behavior matrix). Once the FIR model is available, the prediction system can take place using the FIR inference engine. This process is called fuzzy forecast. The FIR inference engine is a specialization of the k-nearest neighbor rule, commonly used in the pattern recognition field. Defuzzification is the inverse process of fuzzification. It allows converting the qualitative predicted output into quantitative values that can then be used as inputs to an external quantitative model.

In order to provide a better understanding of the rule extraction method proposed in this paper, we present a brief explanation of the pattern rule base determination process. The logical rule base is generated from the pattern rule base obtained by FIR methodology.

As previously stated, a FIR model is composed of a mask and a pattern rule base, an example of mask (that is used latter in Fig. 2) is:

<table>
<thead>
<tr>
<th>t</th>
<th>x</th>
<th>u_1</th>
<th>u_2</th>
<th>u_3</th>
<th>u_4</th>
<th>y_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>t - 2\delta t</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>t - \delta t</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>0</td>
<td>-4</td>
<td>0</td>
<td>0</td>
<td>+1</td>
<td></td>
</tr>
</tbody>
</table>

Each negative element in the mask is called a m-input (mask input) and denotes a causal relation with the output. The process of finding the best mask that represents the system under study corresponds to the feature selection process in such a way that the negative elements in the mask are the relevant features selected and give us information of the temporal relation with the output.

The process to obtain the pattern rule base from the mask is illustrated in Fig. 2. The mask can be used to ‘flatten’ dynamic relationships into pseudo-static relationships. The left-hand side of Fig. 2 shows an excerpt of the matrix that stores the class values of the training data set. It shows the numerical rather than the symbolic class values. In the example shown in Fig. 2, all the variables were discretized into three classes, with the exception of variable y/ that was discretized into two classes. The dashed box symbolizes the mask that is shifted downwards along the class value matrix. The round shaded ‘holes’ in the mask denote the positions of the m-inputs, whereas the square shaded ‘hole’ indicates the position of the m-output. The class values are read out from the class value matrix through the ‘holes’ of the mask, and are placed next to each other in the behavior matrix that is shown on the right-hand side of Fig. 2 and that contains the set of pattern rules. Each pattern rule has associated three values, the class value (shown in Fig. 2) and the membership and side values (do not shown in Fig. 2). Notice that the number of pattern rules obtained by FIR corresponds almost to the number of training data.
available. Therefore, the size of the pattern rule base impedes their comprehension and understanding by the educational actors. For a deeper and more detailed insight into the FIR methodology, the reader is referred to [17].

Starting from the description of FIR in last section, we now explain how rule extraction can be implemented as part of this methodology. Figure 3 shows in a schematic way the main phases of the proposed algorithm described next.

The proposed method is an iterative process that compact the pattern rule base obtained by FIR. On the one hand, we aim to obtain interpretable, realistic and efficient rules, describing students’ learning behavior. On the other hand, we want to compact the pattern rule base to speed up the prediction process. In order to get a set of logical rules congruent with the pattern rules previously identified by FIR, the proposed algorithm is based on its initial discretization. The model can be summarized as a set of ordered steps:

1. **Basic compactation.** This is an iterative step that evaluates, one at a time, all the rules in a pattern rule base. The pattern rule base, \( R \), is compacted on the basis of the “knowledge” obtained by FIR. A subset of rules \( R \) can be compacted in the form of a single rule \( r_c \), when all premises \( P \) but one (\( P_a \)), as well as the consequence \( C \) share the same values. Premises, in this context, represent the input features, whereas consequence is the output feature in a rule. If the subset contains all legal values \( LV_a \) of \( P_a \), all these rules can be replaced by a single rule \( r_c \) that has a value of -1 in the premise \( P_a \). When more than one -1 value, \( P_{ai} \), is present in a compacted rule \( r_c \), it is compulsory to evaluate the existence of conflicts by expanding all \( P_{ai} \) to all their legal values \( LV_{ai} \) and comparing the resultant rules \( Xr \) with the original rules \( R \). If conflicts, \( C_f \), exist, the compacted rule \( r_c \) is rejected, and otherwise accepted. In the latter case, the previous subset, \( R \), is replaced by the compacted one \( r_c \). Conflicts occur when one or more extended rules, \( Xr \) have the same values in all its premises, \( P \), but different values in the consequence \( C \).

2. **Improved compactation.** Whereas the previous step only structures the available knowledge and represents it in a more compact form, the improved compaction step extends the knowledge base \( R \) to cases that have not been previously used to build the model: \( R_b \). Thus, whereas step 1 leads to a compacted data base that only contains knowledge, the enhanced algorithm contains undisputed knowledge and uncontested belief. Two options are studied: In the first one, using the compacted rule base \( R’ \) obtained in step 1, all input features \( P \) (premises) are visited once more in all the rules \( r \) that have nonnegative values (not compacted), and their values are replaced by -1. An expansion to all possible full sets of rules \( Xr \) and their comparison with the original rules \( R \) are carried out. If no conflicts \( C_f \) are found, the compacted rule, \( r_c \), is accepted, and otherwise, rejected. The second option is an extension of the basic compaction, where a consistent and reasonable minimal ratio, \( MR \), of the legal values \( LV_a \) should be present in the candidate subset \( R_c \), in order to compact it in the form of a single rule \( r_c \). This latter option seems sensible because, although a reasonable ratio was used to compact \( R_c \) in a single rule \( r_c \), the assumed beliefs are minimal and do not compromise the model previously identified by FIR. Instead, in option 1, beliefs are assumed to be consistent with the original rules; nevertheless, this could compromise the agreement with model identified, specially when the training data are poor and do not describe well all possible behaviors.

![FIR qualitative model identification process](image)

**Fig. 2.** FIR qualitative model identification process.

![Diagram of ordered steps of the rule extraction method](image)

**Fig. 3.** Ordered steps of the rule extraction method.

The obtained set of rules is subjected to a number of refinement steps: removal of duplicate rules and conflicting rules; unification of similar rules; evaluation of the obtained rules and removal of rules with low specificity. Specificity is a standard metric, described latter, that assess the quality of the obtained rules. For
space limitations the refinements steps can not be explained in detail in this paper.

3 Data from the CECTE virtual campus

The Center of Studies in Communication and Educational Technologies (CECTE: Spanish acronym) is a partially virtual campus, offering postgraduate courses and continuous education (graduate, workshops and specific courses) to Latin-American students. The CECTE is part of the international Latin-American Institute of Educative Communication (ILCE: Spanish acronym), whose main goal is to offer postgraduate courses.

Table 1. Data features collected for the experiment.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Alias</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of the student</td>
<td>AGE</td>
<td>Age of the student.</td>
</tr>
<tr>
<td>Area of expertise</td>
<td>EXP</td>
<td>Area of expertise of the student (mathematics, chemistry, Mexican history, etc.).</td>
</tr>
<tr>
<td>Gender</td>
<td>G</td>
<td>Student’s gender.</td>
</tr>
<tr>
<td>Level of studies</td>
<td>STD</td>
<td>Level of studies (graduate, master, Ph.D., etc.).</td>
</tr>
<tr>
<td>Position of the student</td>
<td>POS</td>
<td>Position of the student as a teacher in his/her school.</td>
</tr>
<tr>
<td>Percentage of the activities performed by the student</td>
<td>ACT</td>
<td>Percentage of the activities performed by the student with respect to the total activities of the course.</td>
</tr>
<tr>
<td>Percentage of session assistance</td>
<td>ASS</td>
<td>Percentage of student’s session assistance with respect to the total number of sessions of the course.</td>
</tr>
<tr>
<td>Average mark of the e-mail</td>
<td>MAIL</td>
<td>Average mark obtained by the student in the activities sent by e-mail.</td>
</tr>
<tr>
<td>Average mark of the co-evaluation</td>
<td>COEV</td>
<td>Average mark of the co-evaluation performed by the student of the class plan of other students.</td>
</tr>
<tr>
<td>Average mark of the forum participation</td>
<td>F</td>
<td>Average mark of the student’s forum participation (referring to topics related to the course).</td>
</tr>
<tr>
<td>Average mark of the forum class plan</td>
<td>FCP</td>
<td>Average mark of the forum class plan (referring only to topics related to the class plan exclusively).</td>
</tr>
<tr>
<td>Average mark of the final class plan</td>
<td>FC</td>
<td>Average mark obtained by the student in his/her final class plan.</td>
</tr>
<tr>
<td>Average mark of the initial class plan</td>
<td>IC</td>
<td>Average mark obtained by the student in his/her initial class plan.</td>
</tr>
<tr>
<td>Average mark of the experience report</td>
<td>ER</td>
<td>Average mark obtained by the student in the experience report.</td>
</tr>
<tr>
<td>Average mark of the work in the branch</td>
<td>BR</td>
<td>Average mark of the work (activities) performed in the branch.</td>
</tr>
<tr>
<td>Final mark</td>
<td>MARK</td>
<td>Final mark obtained by the student in the course.</td>
</tr>
</tbody>
</table>

The most demanded CECTE courses follow a hybrid, semi-presential model, in which students take courses online (WCETCE) but also attend weekly TV sessions through the National System of Educative Television (EDUSAT). Through WCETCE, students can access the course materials and communicate and interact with each other through an e-mail system and a discussion forum. The environment also includes an agenda, a news system, virtual classrooms, a digital library, interactive tutorials, and other related tools.

The tutor is a very relevant actor, as he or she interacts directly with students, assigning learning activities, answering doubts, opening topics in discussion forums, evaluating the activities performed by learners, and verifying that the teaching-learning process be adequate, taking advantage of all the tools provided by WCETCE.

Two novel evaluation topics, not often used in e-learning environments, were incorporated in the course: co-evaluation and experience report. In co-evaluation, the advisor grades how well the student evaluates the class plans of his/her course mates. The experience report is a student description of his/her perception of the course. It can be viewed as a self-evaluation of the student’s own learning process.

For the experiments in this study, a set of 722 students, enrolled in the “Didactic Planning” graduate course, was selected. The course is addressed to second term high school teachers offering specialized subjects, namely Mathematics, Chemistry, Mexican History, Computer Science, English, as well as Reading and Writing, and Ethics and Values workshops. The students are meant to perform a set of activities throughout the course with the main purpose of learning new methods and strategies for planning the classes that they teach. This is the reason why these activities are centred on the so-called “class plan”. A class plan is a document where a set of strategies are suggested to develop a teaching-learning session, taking into account different factors that appear in the educational process, such as students’ characteristics, teaching style, teachers’ experience, etc. The data features available for this study are detailed on Table 1.

4 Experiments: Results and discussion

In a previous work [18], the FIR methodology was used in the same application described here. The goal of this research was twofold. On the one hand, the identification of FIR models capable of predicting student’s performance. On the other hand, the determination of data features with the highest relevance from the student’s performance point of view.

The best mask obtained in this work was the following:

```
T 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -1 0 0 0 0 -2 -3 0 +1
```

Meaning that, from all the features involved (see Table 1), three of them i.e. COEV, IC and ER, were the more relevant when the goal was to predict student’s learning performance. The results presented in [18] were very successful in the prediction of student’s performance (96% of accuracy). However, the pattern rule base obtained had a size of 720 fuzzy pattern rules, a number extremely high to become useful for the easy understanding of student’s learning performance.

In this paper, the interpretability of the prediction results is improved by their description in terms of simple and actionable rules. This is accomplished through the
application of the new extension of FIR that allows the
extraction of logical rules from the already available
pattern rule base. The obtained logical rules are more
comprehensive, readable and provide explanations (not
only assumptions) that may be validated by domain
experts, increasing confidence in the analysis.

The experimental results obtained using the rule
extraction algorithm described in section 2 are presented
in Table 2. The first column of this table presents the
logical rules obtained by the extraction process. The
second and third columns show the specificity, sensitivity
and positive predictive value (PPV) measures for the
training and test data sets, respectively. Specificity is
defined as one minus the ratio of the number of out-of-
class data records that the rule identifies to the total
number of out-of-class data. Sensitivity is the ratio of the
number of in-class data that the rule identifies to the total
number of in-class data. PPV is the ratio of the number of
in-class data that the rule identifies to the total number of
data the rule identifies.

The number of logical rules extracted from the set of
720 pattern rules is 6. Therefore, a huge reduction has
been obtained. A set of 6 rules gives comprehensible
explanation to the educative actors making easiest the
understanding of students learning behavior.

From table 2 it can be seen that the specificity and the
PPV measures have reasonable values for each specific
rule as well as for the whole set of rules (last row of Table
2). However, the sensitivity has very low values in 4 of
the 6 logical rules extracted. Only the sensitivity of rules 2
and 6 (in bold in Table 2) have reasonable values,
showing a very common pattern in the analyzed data.

A high sensitivity value implies a very general rule,
i.e. a high number of students fit in that rule. A small
sensitivity value denotes a very specific rule, i.e. the rule
describes a small set of students. However, it is possible
that this set of students is not represented with any other
rule. Therefore, we consider important to keep the rules
with low sensitivity but high specificity in such a way that
all the students (each one with its own characteristics) are
represented in the rule base. In that way we are assuring
better predictions of the students’ performance. A rule
with a high specificity value indicates that when the
premises (the input features) are fulfilled the output is also
fulfilled. We think that this is the most valuable measure
when the goal is to obtain good predictions of the output
value, i.e. students mark in our case.

In Table 2 we found rules with very low sensitivity
but high specificity values. This is for example the case of
rule 4 with a sensitivity of 0.07 and a specificity of 1. If
we analyze this rule carefully we find out that it
represents 4 students of the 61 that did not pass the
course, i.e. that have a mark value between 0 and 4.9.
These 4 students are the only ones that obtained very low
evaluations on the three features involved, i.e. COEV, ER
and IC. Contrarily, the rule 6 has relatively high values of
specificity and sensitivity measures. In that case, we
obtain a sensitivity of 0.81 because this rule represents
314 students of 384 that have the mark in the same range
(high evaluation). It is interesting to notice that the
distribution of the data available in this study is
unbalanced, i.e. 61 students did not pass the course, 104
passed the course with a low grade and 384 students
obtained high grades. Based on the results obtained and
on our experience we think that in those applications
where the available data are more balanced than in the
current application the extraction algorithm will obtain
rules with higher sensitivity values.

The learning behavior rules obtained were analyzed
and validated by educative experts of CECTE. They
agreed that the obtained results were intuitive, realistic,
and mostly consistent with their own perception of the
CECTE course students’ learning behavior. When the
students obtain high evaluations in two relevant variables
determined by FIR, i.e. COEV and ER, they obtain a high
degree in the course (rule 6 in Table 2). When the COEV
grade is not very high the student can still obtain a good
degree in the course (rule 6 in Table 2). Few students obtain very
low grades, only 61, but there are different circumstances
of students’ bad results. This is the reason way three rules
are necessary to cover these students behaviour (rules 2, 3
and 4 in Table 2). Only one rule (rule 1 in Table 2) is
obtained to define those students that passed the course
with low degrees.

Table 2. Results obtained using the new rule extraction method for both training and test data sets. Spec stands for Specificity, Sens for Sensitivity and PPV for Positive Predictive Value.

<table>
<thead>
<tr>
<th>RULE</th>
<th>TRAIN</th>
<th>TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spec</td>
<td>Sens</td>
</tr>
<tr>
<td>IF 0&lt;=IC&lt;=5.1 AND 4.9&lt;=COEV&lt;=10 THEN 0&lt;=MARK&lt;=7.9</td>
<td>0.98</td>
<td>0.29</td>
</tr>
<tr>
<td>IF 5.1&lt;=IC&lt;=10 AND 0&lt;=ER&lt;=8.1 THEN 0&lt;=MARK&lt;=4.9</td>
<td>0.79</td>
<td>0.67</td>
</tr>
<tr>
<td>IF 0&lt;=COEV&lt;=4.90 AND 8.1&lt;=ER&lt;=10 THEN 0&lt;=MARK&lt;=4.9</td>
<td>1.00</td>
<td>0.11</td>
</tr>
<tr>
<td>IF 0&lt;=IC&lt;=5.1 AND 0&lt;=COEV&lt;=4.90 AND 0&lt;=ER&lt;=8.1 THEN 0&lt;=MARK&lt;=4.9</td>
<td>1.00</td>
<td>0.07</td>
</tr>
<tr>
<td>IF 5.1&lt;=IC&lt;=10 AND 4.90&lt;=COEV&lt;=7.9 THEN 7.9&lt;=MARK&lt;=10</td>
<td>0.95</td>
<td>0.03</td>
</tr>
<tr>
<td>IF 7.90&lt;=COEV&lt;=10 AND 8.1&lt;=ER&lt;=10 THEN 7.90&lt;=MARK&lt;=10</td>
<td>0.75</td>
<td>0.81</td>
</tr>
<tr>
<td>TOTAL RULES</td>
<td>0.93</td>
<td>0.38</td>
</tr>
</tbody>
</table>
5 Conclusion

The possibility of tracking user behavior in virtual campus e-learning environments makes the web mining of the resulting databases possible. This opens new possibilities for the pedagogical and instructional designers who create and organize the learning contents. One of the most interesting options is the personalization of the e-learning process. The characterization of the students’ online behavior would benefit from a method capable of determining the relevance of the features involved in the analyzed data set in terms of the students’ mark prediction. In this study, we have presented a new rule extraction method based on FIR methodology. This method has been used on a data set obtained from a virtual campus real e-learning experience.

The logical rules obtained are easily understandable by experts in an educative domain and they may expose problems with the data itself. This knowledge could be used for real time student personalization guidance, and to help teachers in finding patterns of students’ behavior. For this knowledge to have an intuitive and useful form, results have been described in terms of a set of logical rules describing the diverse levels of the students’ performance.

The experimental results have shown that the extraction rules method presented in this paper was able to obtain comprehensible, actionable and realistic logical rules describing students’ learning behavior patterns.

Acknowledgment

Félix Castro is a research fellow within the PROMEP program of the Mexican Secretary of Public Education.

References