Analysing the Moodle e-learning platform through subgroup discovery algorithms based on evolutionary fuzzy systems

C. J. Carmona
Languages and Computer Technology Systems, Department of Civil Engineering, University of Burgos, 09006, Burgos (Spain)
D. Elizondo
School of Computer Science and Informatics, De Montfort University, The Gateway, Leicester, LE1 9BH, (United Kingdom)

Abstract

Nowadays, there is a increasing in the use of learning management systems from the universities. This type of systems are also known under other different terms as course management systems or learning content management systems. Specifically, these systems are e-learning platforms offering different facilities for information sharing and communication between the participants in the e-learning process.

This contribution presents an experimental study with several subgroup discovery algorithms based on evolutionary fuzzy systems using data from a web-based education system. The main objective of this contribution is to extract unusual subgroups to describe possible relationships between the use of the e-learning platform and marks obtained by the students. The results obtained by the best performing algorithm, NMEEF-SD, are also presented. The most representative results obtained by this algorithm are summarised in order to obtain knowledge that can allow teachers to take actions to improve

student performance.

1. Introduction

On-site education systems allow teachers to directly analyse the situation of the courses, using the interation with students to improve course content and thus achieve better grades for students. However, in learning management systems (LMS) this information can not be directly analised with respect to the use of the resources, activities, quizzes and so on, and teachers must rely on certain techniques like educational data mining.

Educational data mining is an emerging interdisciplinary research area that deals with the development of methods to explore data from educational contexts [35]. It is concerned with the development of mining methods to explore the unique types of data in educational environments and, using these methods, to better understand students and learning systems. A data mining algorithm can discover knowledge using different representation models and techniques from two different perspectives: predictive induction, whose objective is the discovery of knowledge for classification or prediction [27]; or descriptive induction, whose main objective is the extraction of interesting knowledge from data. In this area, attention can be drawn to the discovery of association rules following an unsupervised learning model [1], and other approaches to non-classificatory induction.

This contribution employs Subgroup Discovery (SD) [16, 6] algorithms based on evolutionary fuzzy systems (EFSs) in order to analyse the Moodle platform from the University of Cordoba. SD is a descriptive inductive learning area in which, given a set of data and a property of interest to the user, an attempt is made to locate subgroups which are statistically "most interesting" for the user. A subgroup is interesting if it has an unusual statistical distribution with respect to the property of interest. The objective is to discover interesting properties of subgroups by obtaining simple rules, which are highly significant and with high support. The experimental study is performed with several SD algorithms to show the quality of the powerful NMEEF-SD. This study, joined with an analysis from the point of view of the teacher, is also presented with the aim of improving the e-learning courses.

To do so, the paper is organised as follows: learning management systems are presented in section 2. SD task and EFSs used for SD are presented in section 3. In section 4 the complete experimental study is presented. Finally, conclusions are outlined.

2. Learning management systems

Nowadays, several systems have been developed for online education, most of them using web-based platforms. These web-based educational systems can be classified in different types such as adaptive and intelligent web based educational systems, particular web-based courses and learning management systems. In this paper, we focus on learning management systems (LMSs), also known under other different terms as course management systems or learning content management systems. LMSs are e-learning platforms offering different facilities for information sharing and communication between the participants in the e-learning process. These systems allow the distribution of information to students but also facilitates the task of the educators when producing content material, preparing assignments, managing distance classes, engaging in discussions and enabling collaborative learning with file storage areas, chats, forums or news services. There are both commercial LMSs (eg. Blackboard, Virtual-U, WebCT, or TopClass) and free LMSs (Moodle, Ilias, Claroline, or ATutor) [31].

It is common for LMSs using a relational database to store the students activities and usage information instead of using data log files. However, when the number of students is high it becomes complicated for a tutor to manage the information stored, even using the reporting tools offered by some of the platforms. In this situation, the great amount of information makes it difficult to extract useful information to improve the learning process. Therefore, some researchers have recently proposed using data mining techniques in order to help the tutor in this task.

Data mining techniques allow to identify patterns in the information related to the use of the platform which can be analysed not only to interpret the students' activity on the platform but also to get more objective feedback for instruction and more knowledge about how the students learn on the LMS [35]. In fact, some data mining techniques have already been used in e-learning problems. Thus, a grammar-based genetic programming with multi-objective optimization techniques was performed in [36] in order to provide a feedback to courseware authors. In [39] a clustering task was used to discover patterns reflecting user behavior. In addition, in [28] the sequencing capabilities of the SCORM standard to include the concept of recommended itinerary, by combining educators expertise with learned experience acquired by system usage analysis was presented.

Two different areas can be established with respect to the data mining techniques: predictive, in which the main objective is to find a model to classify new instances with respect to an interest variable, and descriptive, whose objective is to describe relationships between all the variables. However, there are some data mining techniques that are somehow halfway between prediction and description. This is the case with the subgroup discovery task, aimed at searching unusual relationships between data with respect to an interest variable. Actually, this technique was applied in [34] to search subgroups in an e-learning platform through evolutionary algorithms.

3. Subgroup discovery

3.1. Main properties

The concept of SD was initially introduced by Kloesgen [20] and Wrobel [40], and more formally defined by Siebes [38] but using the name Data Surveying for the discovery of interesting subgroups. It can be defined as [41]:

> *"In SD, we assume we are given a so-called population of individuals (objects, customer, ...) and a property of those individuals we are interested in. The task of SD is then to discover the subgroups of the population that are statistically "most interesting", i.e., are as large as possible and have the most unusual statistical (distributional) characteristics with respect to the property of interest."*

SD attempts to search relations between different properties or variables of a set with respect to a target variable. Due to the fact that SD is focused in the extraction of relations with interesting characteristics, it is not necessary to obtain complete but partial relations. These relations are described in the form of individual rules.

Then, a rule (R), which consists of an induced subgroup description, can be formally defined as [11, 24]:

$$R : Cond \;\rightarrow\; Target_{value}$$

where $Target_{value}$ is a value for the variable of interest (target variable) for the SD task (which also appears as $Class$ in the literature), and $Cond$ is commonly a conjunction of features (attribute-value pairs) which is able to describe an unusual statistical distribution with respect to the $Target_{value}$.

As an example, let D be a dataset with three variables $Age = \{Less\,than\,25,\ 25\,to\,60, More\,than\,60\}$, $Sex = \{M,\ F\}$ and $Country = \{Spain,\ USA,\ France,\ German\}$, and a variable of interest target variable $Money = \{Poor,\ Normal,\ Rich\}$. Some possible rules containing subgroup descriptions are:

$R_1 :\ (Age = Less\,than\,25\ AND\ Country = German) \rightarrow\ Money = Rich$

$R_2 :\ (Age = More\ than\ 60\ AND\ Sex = F) \rightarrow\ Money = Normal$

where rule R_1 represents a subgroup of German people with less than 25 years old for which the probability of being rich is unusually high with respect to the rest of the population, and rule R_2 represents that women with more than 60 years old are more likely to have a normal economy than the rest of the population.

SD is somewhere halfway between predictive and descriptive induction, and its goal is to generate in a single and interpretable way subgroups to describe relations between independent variables and a certain value of the target variable. The algorithms for this task must generate subgroups for each value of the target variable. Therefore, an execution for each value of the variable must be performed.

A rule for SD is represented in Fig. 1, where two values for the target variable can be found ($Target_{value} = x$ and $Target_{value} = o$). In this representation a subgroup for the first value of the target variable can be observed, where the rule attempts to cover a high number of objects with a single function: a circle. As can be observed the subgroup does not cover all the examples for the target value x even the examples covered are not positive in all the cases, but the form of this function is uniform and very interpretable with respect others. In this way the algorithm achieves a reduction of the complexity. Furthermore, the true positive rate for the value of the target variable is high, with a value of 75%.

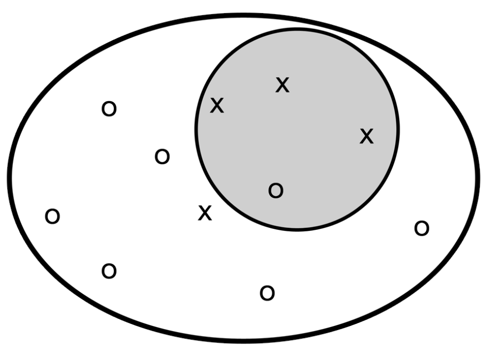

Figure 1: Representation of a subgroup discovery rule with respect to a value (x) of the target variable

The SD task is differentiated from classification techniques basically because SD attempts to describe knowledge for the data while a classifier attempts to predict it. Furthermore, the model obtained by a SD algorithm is usually simple and interpretable, while that obtained by a classifier is complex and precise.

On the other hand, the main elements within the SD approaches can be observed below [3]:

- *Type of the target variable.* Different types for the variable can be found: binary, nominal or numeric. For each one different analyses can be applied considering the target variable as a dimension of the reality to study.

- *Description language.* The representation of the subgroups must be suitable for obtaining interesting rules. These rules must be simple and therefore are represented as attribute-value pairs in conjunctive or disjunctive normal form in general. Furthermore, the values of the variables can be represented as positive and/or negative, through fuzzy logic, or through the use of inequality or equality and so on.

- *Quality measures.* These are a key factor for the extraction of knowledge because the interest obtained depends directly on them. Furthermore, quality measures provide the expert with the importance and interest of the subgroups obtained. Different quality measures have been presented in the specialised bibliography [12, 20, 21, 24], but there is no consensus about which are the most suitable for use in SD.

- *Search strategy.* This is very important, since the dimension of the search space has an exponential relation to the number of features and values considered. Different strategies have been used up to the moment, for example beam search, evolutionary algorithms, search in multi-relational spaces, etc.

A wide number of quality measures have been presented in the SD literature both to guide the search process in order to find the best SD rules and to measure the quality of the SD rule set finally obtained [20, 26]. The most common quality measures used in SD can be classified by their main objective such as:

- Complexity measures, related to the interpretability of the subgroups, i.e. to the simplicity of the knowledge extracted.
- Generality measures, used to quantify the quality of individual rules according to the individual patterns of interest covered.
- Precision measures, showing the precision of the subgroups.
- Interest measures, intended for selecting and ranking patterns according to their potential interest to the user.
- Hybrid, that attempt to obtain a good trade-off between different objectives.

Table 1 summarises the *Quality measures* most used in SD [16] and their main characteristics.

Table 1: Classification of the quality measures used in subgroup discovery

Quality measure	*C*	*G*	*P*	*I*
Number of rules	X			
Number of variables	X			
Coverage [26]		X		
Support [26]		X		
Confidence [2]			X	
Precision measure Q_c [11]			X	
Precision measure Q_g [20]			X	
Q_g-Weight [11]			X	
Interest [29]				X
Novelty [40]				X
Lift [4]				X
Significance [20]				X
Sensitivity [20]		X	X	
False Alarm [11]		X	X	
Specificity [20]		X	X	
Unusualness [25]		X	X	X
Piatetstky-Shapiro [14]		X	X	X

C=Complexity, G=Generality, P=Precision and I=Interest

According to the SD concept the obtaining of interesting, simple and interpretable subgroups, covering the majority of the examples of the interest property (target variable) is desirable. Considering this definition and the analysis of the different quality measures used in the literature, we propose three guidelines in order to establish the type of measure more suitable, to guide the search process and to analyse the quality of the subgroups obtained by any SD algorithm:

- *Interpretability.* A SD proposal must obtain few rules containing a low number of variables in the antecedent part in order to help to the experts to understand and use the extracted knowledge, i.e. simple and interpretable subgroups are preferred in SD task.

- *Relation sensitivity-confidence.* A SD algorithm must obtain results with a good precision, where the majority examples covered belong to the target variable, i.e. it must achieve the best possible relation between sensitivity and confidence. Both quality measures are primordial in order to provide subgroups to the experts that cover the higher number of described correctly examples. It is difficult for the algorithms to obtain this compromise due to the loss suffered by a measure when trying to increase the other.

- *Novelty.* A SD model must contribute novel knowledge, providing the experts with information in order to describe unusual and interesting behaviour within the data. This objective could be measured with a wide number of quality measures as novelty, interest or significance, among others. Nevertheless, it is important to highlight the utility of the unusualness to measure this objective because it contributes with generality and confidence to the problem. Moreover, this quality measure is widely used in the specialised bibliography.

It can be considered that the main purpose of a SD algorithm is to find a good trade-off between these three guidelines because this lead to the obtaining of good results in a wide number of quality measures and not only in the ones used in the search process.

3.2. Evolutionary fuzzy systems in subgroup discovery

Computational Intelligence techniques such as artificial neural networks [33], fuzzy logic [42], and genetic algorithms [17, 13] are popular research subjects, since they can deal with complex engineering problems which are difficult to solve by classical methods [22].

Hybrid approaches have attracted considerable attention in the Computational Intelligence community. One of the most popular approaches is the hybridization between fuzzy logic and GAs leading to genetic fuzzy systems [7]. A GFS is basically a fuzzy system augmented by a learning process based on evolutionary computation, which includes genetic algorithms, genetic programming, and evolutionary strategies, among other evolutionary algorithms (EAs) [9]. This concepts is extended to the EFSs [15].

Fuzzy systems are one of the most important areas for the application of the Fuzzy Set Theory [43, 44]. Usually it is considered a model structure in the form of fuzzy rule based systems (FRBSs). FRBSs constitute an extension to classical rule-based systems, because they deal with "IF-THEN" rules, whose antecedents and consequents are composed of fuzzy logic statements, instead of classical ones. They have demonstrated their ability for control problems [30], modelling [32], classification or data mining [23] in a huge number of applications.

The automatic definition of an FRBS can be seen as an optimization or search problem, and EAs are a well known and widely used global search technique with the ability to explore a large search space for suitable solutions only requiring a performance measure. In addition to their ability to find near optimal solutions in complex search spaces, the generic code structure and independent performance features of EAs make them suitable candidates to incorporate a priori knowledge. In the case of FRBSs, this a priori knowledge may be in the form of linguistic variables, fuzzy membership function parameters, fuzzy rules, number of rules, etc. These capabilities extended the use of GAs in the development of a wide range of approaches for designing FRBSs over the last few years.

The SD is focused on the genetic rule learning where most of the approaches proposed to automatically learn the knowledge base from numerical information have focused on the rule base learning, using a predefined data base. The usual way to define this DB involves choosing a number of linguistic terms for each linguistic variable (an odd number between 3 and 9, which is usually the same for all the variables) and setting the values of the system parameters by an uniform distribution of the linguistic terms into the variable universe of discourse.

The main EFSs for SD presented throughout the literature, as far as we know, are described below:

3.2.1. SDIGA

SDIGA[19] is an evolutionary fuzzy system [15] because it uses a knowledge representation fuzzy rules and evolutionary computation as a learning process. It is interesting to remark that SDIGA searches for rules for each value of the target variable, i.e. the consequent is not represented in the chromosome but is fixed.

This algorithm follows the IRL approach where the solution of each iteration is the best individual obtained and the global solution is formed by the best individuals obtained in the different runs. The representation of the individuals is performed through the "*Chromosome = Rule*" approach and the core of SDIGA is an EA using a post-processing step based on a local search. This hybrid algorithm extracts one simple and interpretable fuzzy rule with an adequate level of support and confidence. The algorithm model can use fuzzy canonical or DNF rules with a predefined set of linguistic labels.

This algorithm is included in an iterative process for the extraction of different rules. In this way, algorithm marks examples cover for rules to prevent a new rule being obtained which covers exactly the same examples in the following runs. Algorithm is obtaining rules while the generated rules reach a minimum level of confidence and give information on areas of the search space in which there are examples not described by the rules generated in previous iterations. The rule is improved in a post-processing phase throughout a hill-climbing process, which modifies the rule in order to increase the degree of support.

The fitness is an aggregation function where the selection of the quality measures like coverage, significance, unusualness, accuracy, sensitivity, crisp support, fuzzy support, crisp confidence and fuzzy confidence is determined by the user. The number of objectives within the weighted aggregation function are between 1 and 3.

3.2.2. MESDIF

MESDIF [18] is a multiobjective EA is an evolutionary fuzzy system based on the SPEA2 approach [45]. It applies the concepts of elitism in the rule selection (using a secondary or elite population) and the search for optimal solutions in the Pareto front. In order to preserve the diversity at a phenotypic level the algorithm uses a niches technique which considers the proximity in values of the objectives and an additional objective based on novelty to promote rules which give information on examples not described by other rules of the population.

The rule induction process obtains rules with high predictive accuracy and which are comprehensible and interesting. In this proposal, the user can choose between a wide number of quality measures (coverage, significance, unusualness, accuracy, sensitivity, support and confidence) to maximise all the defined objectives.

One of the most important aspects of MESDIF is the obtention of results for all the values of the target variable. It returns the individuals of the elite population for each value, whose size is defined by the user.

The algorithm uses the "*Chromosome = Rule*" approach. The multiobjective EA discovers fuzzy rules whose consequent is prefixed to one of the possible values of the target feature. Therefore, all the individuals of the population are associated with the same value of the target variable, and so the chromosome only represents the antecedent of the rule.

3.2.3. NMEEFSD

NMEEF-SD [5] is a multiobjective evolutionary fuzzy system based on NSGA-II [8]. NMEEF-SD codifies each candidate solution according to the "*Chromosome = Rule*" approach, where only the antecedent is represented in the chromosome and the consequent is prefixed to one of the possible values of the target feature in the evolution. Therefore, the algorithm must be executed as many times as the number of different values the target variable contains. With respect to the representation of the rules NMEEF-SD can use canonical or DNF rules.

As the general objective of NMEEF-SD is to obtain a set of rules, which should be general and accurate, the algorithm includes components which enhance these characteristics. In particular, diversity is enhanced in the population using a new operator to perform a re-initialisation based on coverage, in addition to a niching technique (the crowding distance in the selection operator). On the other hand, in order to promote generalisation, as well as the objectives considered in the evolutionary approach, the algorithm includes operators of biased initialisation and biased mutation. Finally, to ensure accuracy, in addition to the objectives NMEEF-SD returns as its final solution those rules which reach a predetermined confidence threshold.

NMEEF-SD allows to choose between two and three quality measures as objectives of the evolutionary process in order to obtain relevant subgroups, between: coverage, significance, unusualness, accuracy, sensitivity, support and confidence.

4. Experimental study

The experimental study is divided into different sections. Firstly, a experimental set up is presented in Section 4.1 where the main properties of the algorithms and the dataset are summarised. Next, Section 4.2 shows the results obtained for the different EFSs in the experimental study. Finally, a study of the usage data with NMEEF-SD algorithm is shown in Section 4.3 where several of the rules obtained are analysed from the point of view of the teacher with the aim of improving the courses content.

4.1. Experimental framework

Moodle system [10] is one of the most used web-based e-learning systems. In addition, Moodle is an alternative to proprietary commercial online learning solutions, is distributed free under open source licensing and has been installed at universities and institutions all over the world.

Moodle system contains a great deal of detailed information on course content, users, usage, etc., stored in a relational data base keeping detailed logs of all the activities performed by the students. We can use these logs in order to determine which students have been active in the course, what they did, when, or if everyone has done a certain task or spent a required amount of time online within certain activities [37].

In this work, available information corresponding to 5 different courses of the University of Cordoba, involving a total of 293 students, is used. In this experimentation, courses with high student participation have been selected to obtain more general results. Furthermore, there is no a minimum amount of students to obtain any rule. This information has been preprocessed for obtaining a summary table with the most important information related to our objective. Table 2 shows this summary including the activities completed by each student in an e-learning course.

It is important to remark that the mark obtained of the students, they have discretised into different values: fail, pass, good and excellent. In the experimentation could have been used numerical values but it is more representative using these values in order to codify them as the rule consequent.

On the other hand, the EFSs employed in the experimental study use the parameters presented in Table 3. These algorithms are executed five times for each experiment because they are non-deterministic algorithms.

4.2. Results and analysis of the experimental study

SD allows to analyse the possible relation between the usage of complementary activities of a course and the final marks obtained by the students. This is conducted using different algorithms. The final mark is used as the

Table 2: Description of the attributes employed for each student

Name	**Description**	**Type**
course	Identification of the course	Discrete
nAssigment	Number of assignments completed	Continuous
nAssigmentP	Number of assignments passed	Continuous
nAssigmentF	Number of assignments failed	Continuous
nQuizz	Number of quizzes completed	Continuous
nQuizzP	Number of quizzes passed	Continuous
nQuizzF	Number of quizzes failed	Continuous
nMessagesC	Number of messages sent to the chat	Continuous
nMessagesT	Number of messages sent to the teacher	Continuous
nMessagesF	Number of messages sent to the forum	Continuous
nRead	Number of forum messages read	Continuous

Table 3: Parameters used by evolutionary fuzzy systems

Algorithm	Parameters
SDIGA	Population size=100, evaluations=10000, crossover probability=0.60, mutation probability=0.01, minimum confidence=0.6, 0,7, 0,8 and 0,9, representation of the rule=canonical, linguistic labels=3, objective1=sensitivity, objective2=unusualness
MESDIF	Population size=100, evaluations=10000, crossover probability=0.60, mutation probability=0.1, elite population=3, 4 and 5, representation of the rule=canonical, linguistic labels=3, objective1=sensitivity, objective2=unusualness, objective3=confidence
NMEEFSD	Population size=50, evaluations=10000, crossover probability=0.60, mutation probability=0.1, minimum confidence=0.6, 0,7, 0,8 and 0,9, representation of the rule=canonical, linguistic labels=3, objective1=sensitivity, objective2=unusualness

variable to characterise, using the different marks to divide the data into classes and codifying them as values of the consequent of the rules. Therefore, the final purpose is to present the results to the teacher in form of rules in order to allow the use of this knowledge in the decision making concerning the complementary activities of the course. For example, the teacher can decide to promote the use of some type of activities to improve marks, or on the contrary eliminate some activities because they are associated with low marks.

Table 4: Results obtained by SD algorithms in e-learning usage data of the University of Cordoba

Algorithm	**Param**	$\mathbf{n}_r$	$\mathbf{n}_v$	**UNUS**	**SENS**	**CONF**
SDIGA	0.6	5.00	3.10	0.039	0.555	0.421
MESDIF	5	20.00	4.29	0.027	0.499	0.346
NMEEF-SD	0.8	15.40	4.38	0.104	0.716	0.831

Table 4 shows the average results for each SD algorithm for each quality measure: number of rules (n_r), number of variables (n_v), unusualness ($UNUS$), sensitivity ($SENS$) and fuzzy confidence ($CONF$). These quality measures can be analysed in [16]. In addition, the $Param$ column represent the value of the parameter in which the algorithm obtains the best results. As can be observed, the best performance is obtained by NMEEF-SD algorithm. NMEEF-SD not only obtains the best results with respect to the sensitivity-confidence relationship, but also with respect to the unusualness measure. In this sense, NMEEF-SD obtains subgroups with more than 70% of the examples covered and with 83% of examples correctly described. These values show the quality of the subgroups obtained by NMEEF-SD. NMEEF-SD does not obtains the best results with respect to the interpretability (number of rules and number of variables per rule), but the differences are not relevant.

Taking into account the results in table 4, next are analysed the rules obtained by NMEEF-SD, with the aim of bringing new knowledge to the teachers in order to enable them to act to improve the results of their students.

4.3. e-learning usage study with the NMEEF-SD algorithm

Table 5: Rules more representative obtained by NMEEF-SD

$\mathbf{n}_r$	**Rule**
R_1	IF (Course=29) AND (nMessagesT=0) THEN *Fail*
	Sign:25.700 - Unus:0.110 - Sens:0.631 - Conf:0.856
R_2	IF (nAssigment=Low) AND (nQuizz=Low) THEN *Fail*
	Sign:25.829 - Unus:0.107 - Sens:0.765 - Conf:0.836
R_3	IF (nQuizzP=Low) THEN *Fail*
	Sign:6.913 - Unus:0.075 - Sens:0.955 - Conf:0.703
R_4	IF (nAssigment=Normal) AND (nMessagesT=0) THEN *Pass*
	Sign:1.423 - Unus:0.023 - Sens:0.789 - Conf:0.241
R_5	IF (nAssigment=Normal) THEN *Pass*
	Sign:1.141 - Unus:0.018 - Sens:0.824 - Conf:0.231
R_6	IF (nRead=Low) THEN *Pass*
	Sign:0.973 - Unus:0.001 - Sens:0.965 - Conf:0.194
R_7	IF (nQuizzP=High) THEN *Good*
	Sign:29.912 - Unus:0.079 - Sens:0.772 - Conf:0.655
R_8	IF (Course=110) THEN *Good*
	Sign:25.536 - Unus:0.081 - Sens:0.750 - Conf:0.532
R_9	IF (nAssigment=High) AND (nRead=Low) THEN *Excellent*
	Sign:25.309 - Unus:0.005 - Sens:0.750 - Conf:0.076
R_{10}	IF (nQuizzP=High) THEN *Excellent*
	Sign:29.912 - Unus:0.006 - Sens:0.417 - Conf:0.081

NMEEF-SD algorithm returns a comprehensive set of subgroups employing a low number of variables with the highest unusualness. Therefore, in this paper an analysis with the most representative subgroups is performed. In Table 5 the best subgroups for each value of the target variable are presented.

Some key ideas can be highlighted from the results obtained for each one of the marks of the study:

- *Fail.* A set of subgroups with high values in all the quality measures analysed can be observed in this group, where generic rules with low number of variables obtain high values of unusualness and confidence. The information to highlight for the students with the mark *fail* would be a low participation of them, and a low interest for the professors to perform quizzes.

- *Pass.* In this group are obtained subgroups which cover the majority of the students but with confidence very low. This group is very difficult to analyse due to the low instances that the dataset contains for this target value.

- *Good.* In this target value good results are obtained with considerable values in all the quality measures analysed. It would be interesting to note the subgroup R_8 where it could indicate to the professor the indifference with respect to the the assignment performed for the student in the platform for the course. In this way, the subgroups indicates him that he should review the relationships between the course and the activities planned. In addition, there is another subgroup (R_7) with excellent results in all the quality measures for courses with high number of quizzes passed.

- *Excellent.* In this target value occurs a similar situation like *pass* value, i.e. is a minority class. However, in this set of subgroups there are good values of sensitivity. Furthermore, the same rule (R_{10}) that appears in the target value *good* is obtained, though as can be observed in the table the confidence is lower. A remarkable new information obtained is that students obtain excellent results if the number of forum messages read is low and the number of assignments completed is high. Of course, teachers want their students to use the forum (as it is a valuable tool to improve the students' skills) but perhaps they need to make an effort to educate their students in its proper use.

Acknowledgment

This work was supported by the Spanish Ministry of Economy and Competitiveness under projects TIN2012-33856, and by the Andalusian Research Plan under project TIC-3928.

References

[1] Agrawal, R., Imieliski, T., Swami, A., 1993. Mining association rules between sets of items in large databases, in: Proc. of the 1993 ACM SIGMOD International Conference on Management of Data, ACM Press. pp. 207–216.

[2] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A., 1996. Fast discovery of association rules, in: Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (Eds.), Advances in Knowledge Discovery and data mining. AAAI Press, pp. 307–328.

[3] Atzmueller, M., Puppe, F., Buscher, H.P., 2004. Towards Knowledge-Intensive Subgroup Discovery, in: Proc. of the Lernen - Wissensentdeckung - Adaptivität - Fachgruppe Maschinelles Lernen, pp. 111–117.

[4] Brin, S., Motwani, R., Ullman, J.D., Tsur, S., 1997. Dynamic Itemset Counting and Implication Rules for Market Basket Data, in: Proc. of the 1997 ACM SIGMOD International Conference on Management of Data, ACM Press. pp. 255–264.

[5] Carmona, C.J., González, P., del Jesus, M.J., Herrera, F., 2010. NMEEF-SD: Non-dominated Multi-objective Evolutionary algorithm for Extracting Fuzzy rules in Subgroup Discovery. IEEE Transactions on Fuzzy Systems 18, 958–970.

[6] Carmona, C.J., González, P., del Jesus, M.J., Herrera, F., 2014. Overview on evolutionary subgroup discovery: analysis of the suitability and potential of the search performed by evolutionary algorithms. WIREs Data Mining and Knowledge Discovery 4, 87–103.

[7] Cordón, O., Herrera, F., Hoffmann, F., Magdalena, L., 2001. Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases. World Scientific.

[8] Deb, K., Pratap, A., Agrawal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions Evolutionary Computation 6, 182–197.

[9] Eiben, A.E., Smith, J.E., 2003. Introduction to evolutionary computation. Springer.

[10] Flate, M., 2003. Online education and learning management systems. Global e-learning in a Scandinavian perspective, NKI Gorlaget, Oslo.

[11] Gamberger, D., Lavrac, N., 2002. Expert-Guided Subgroup Discovery: Methodology and Application. Journal Artificial Intelligence Research 17, 501–527.

[12] Gamberger, D., Lavrac, N., 2003. Active subgroup mining: a case study in coronary heart disease risk group detection. Artificial Intelligence in Medicine 28, 27–57.

[13] Goldberg, D.E., 1989. Genetic Algorithms in search, optimization and machine learning. Addison-Wesley Longman Publishing Co., Inc.

[14] Grosskreutz, H., Rueping, S., Wrobel, S., 2008. Tight optimistic estimates for fast subgroup discovery, in: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, pp. 440–456.

[15] Herrera, F., 2008. Genetic fuzzy systems: taxomony, current research trends and prospects. Evolutionary Intelligence 1, 27–46.

[16] Herrera, F., Carmona, C.J., González, P., del Jesus, M.J., 2011. An overview on Subgroup Discovery: Foundations and Applications. Knowledge and Information Systems 29, 495–525.

[17] Holland, J.H., 1975. Adaptation in natural and artificial systems. University of Michigan Press .

[18] del Jesus, M.J., González, P., Herrera, F., 2007a. Multiobjective Genetic Algorithm for Extracting Subgroup Discovery Fuzzy Rules, in: Proc. of the IEEE Symposium on Computational Intelligence in Multicriteria Decision Making, IEEE Press. pp. 50–57.

[19] del Jesus, M.J., González, P., Herrera, F., Mesonero, M., 2007b. Evolutionary Fuzzy Rule Induction Process for Subgroup Discovery: A case study in marketing. IEEE Transactions on Fuzzy Systems 15, 578–592.

[20] Kloesgen, W., 1996. Explora: A Multipattern and Multistrategy Discovery Assistant, in: Advances in Knowledge Discovery and Data Mining. American Association for Artificial Intelligence, pp. 249–271.

[21] Kloesgen, W., Zytkow, J., 2002. Handbook of Data Mining and Knowledge Discovery. Oxford.

[22] Konar, A., 2005. Computational Intelligence: Principles, Techniques and Applications. Springer-Verlag New York, Inc.

[23] Kuncheva, L., 2000. Fuzzy classifier design. Springer.

[24] Lavrac, N., Cestnik, B., Gamberger, D., Flach, P.A., 2004a. Decision Support Through Subgroup Discovery: Three Case Studies and the Lessons Learned. Machine Learning 57, 115–143.

[25] Lavrac, N., Flach, P.A., Zupan, B., 1999. Rule Evaluation Measures: A Unifying View, in: Proc. of the 9th International Workshop on Inductive Logic Programming, Springer. pp. 174–185.

[26] Lavrac, N., Kavsek, B., Flach, P.A., Todorovski, L., 2004b. Subgroup Discovery with CN2-SD. Journal of Machine Learning Research 5, 153–188.

[27] Michie, D., Spiegelhalter, D.J., Tayloy, C.C., 1994. Machine Learning. Ellis Horwood.

[28] Mor, E., Minguillon, J., 2004. E-learning personalization based on itineraries and long-term navigational behavior, in: Proc. of the 13th international world wide web conference, pp. 264–265.

[29] Noda, E., Freitas, A.A., Lopes, H.S., 1999. Discovering interesting prediction rules wih a genetic algorithm. IEEE Congress on Evolutionary Computation 2, 1322–1329.

[30] Palm, R., Hellendoorn, H., Driankov, D., 1997. Model Based Fuzzy Control. Springer.

[31] Paulsen, M., 2003. Online Education and Learning Management Systems. NKI Forlaget.

[32] Pedrycz, W., 1996. Fuzzy Modelling: Paradigms and Practices. Kluwer Academic Publishers.

[33] Rojas, R., 1996. Neural Networks: A Systematic Introduction. Springer-Verlag New York, Inc.

[34] Romero, C., González, P., Ventura, S., del Jesus, M.J., Herrera, F., 2009. Evolutionary algorithm for subgroup discovery in e-learning: A practical application using Moodle data. Expert Systems with Applications 36, 1632–1644.

[35] Romero, C., Ventura, S., 2010. Educational Data Mining: A Review of the State-of-the-Art. IEEE Tansactions on Systems, Man and Cybernetics, part C: Applications and Reviews 40, 601–618.

[36] Romero, C., Ventura, S., Bra, P.D., 2004. Knowledge discovery with genetic programming for providing feedback to courseware author. User Modeling and User-Adapted Interaction: The Journal of Personalization Research 14, 425–464.

[37] Romero, C., Ventura, S., Salcines, E., 2008. Data mining in course management systems: Moodle case study and tutorial. Computer & Education 51, 368–384.

[38] Siebes, A., 1995. Data Surveying: Foundations of an Inductive Query Language, in: Proc. of the 1st International Conference on Knowledge Discovery and Data Mining, AAAI Press. pp. 269–274.

[39] Talavera, L., Gaudioso, E., 2004. Mining student data to characterize similar behavior groups in unstructured collaboration spaces, in: Proc. 16th European conference on artificial intelligence, pp. 17–23.

[40] Wrobel, S., 1997. An Algorithm for Multi-relational Discovery of Subgroups, in: Proc. of the 1st European Symposium on Principles of Data Mining and Knowledge Discovery, Springer. pp. 78–87.

[41] Wrobel, S., 2001. Inductive logic programming for knowledge discovery in databases. Springer. chapter Relational Data Mining. pp. 74–101.

[42] Yager, R.R., Filev, D.P., 1994. Essentials of Fuzzy Modeling and Control. John Wiley & Sons, Inc.. 1st edition.

[43] Zadeh, L.A., 1965. Fuzzy sets. Information Control 8, 338–353.

[44] Zadeh, L.A., 1975. The concept of a linguistic variable and its applications to approximate reasoning. Parts I, II, III. Information Science 8-9, 199–249,301–357,43–80.

[45] Zitzler, E., Laumanns, M., Thiele, L., 2002. SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization, in: International Congress on Evolutionary Methods for Design Optimization and Control with Applications to Industrial Problems, pp. 95–100.