Journal of Web Engineering, Vol. 8, No.2 (2009) 093-112

WEB USAGE MINING APPROACH TO DETECT STUDENT'S COLLABORATIVE SKILLS

ELENA B. DURÁN

National University of Santiago del Estero, Argentina
eduran@unse.edu.ar

ANALÍA AMANDI

UNICEN University, CONICET, Argentina
amandi@exa.unicen.edu.ar

Received April 7, 2008
Revised October 15, 2008

An effective collaboration in learning environments involves a set of skills that students must learn and cultivate. Detecting the contexts in which students apply these skills facilitates personalized assistance in learning environments during the learning process. This work introduces a method to detect collaborative behavior patterns automatically. It is based on Web Usage Mining techniques and allows us to identify contexts in which collaborative skills are applied. The patterns are discovered using association rules and then are used to update a Collaborative Profile in a Collaborative and Dynamic Student Model. The method was validated with simulation techniques and the results obtained suggest that Web Usage Mining is an effective method for detecting collaborative profiles in distance learning environments.

Key words: Web Usage Mining, association rules, collaborative learning, student model, collaborative profile.
Communicated by: B. White and L. Olsina

1 Introduction

Social science currently offers new conceptions of the learning process. Participation in social practices is a fundamental way of learning. Learning involves becoming attuned to the constraints and resources, the limits and possibilities that are realized in the practices of a community. Learning is promoted by social norms that value the search for understanding and increasing people's opportunities and motivations to interact, receive feedback, and learn [3]. Consequently, Computer-supported Collaborative Learning (CSCL) emerges as a branch of learning sciences concerned with the study of the way in which people can learn with the aid of computers. Collaborative learning involves individuals as group members, but also involves phenomena, such as negotiation and sharing of meanings that are accomplished interactively in group processes [35].

In [34], CSCL is viewed in terms of collaborative knowledge building, group and personal perspectives, mediation by artifacts, and micro-analysis of conversation. The notion of collaborative knowledge building defines a useful paradigm for conceptualizing learning as a social practice, where the emphasis is put on construction and further development of a knowledge object that is shared by the group or "community of learners". Therefore, the focus is not on personal learning by the participants, who are assumed to retain part of what the group discovered, it is on their collaboration skills and their positive experiences of inquiring and intellectual engagement. Accordingly, students may learn effectively in group when they ask questions, explain and justify their opinions, articulate their reasoning, and produce and ponder their own knowledge. Nevertheless, the benefits of collaborative learning are reaped only by means of well articulated learning groups.

Collaboration takes place in a social environment; indeed, another important factor mentioned in [34] is interaction. Social interactions occurring during collaboration promote learning; hence, they are educationally beneficial [4]. For Dillenbourg [8] the effect of collaborative learning depends on the quality of interactions that take place among group members. These interactions, in turn, depend on the different collaborative skills students have, which are often conditioned by the collaborative context in which they are participating. Having collaborative skills is a prerequisite for learners to take part in effective collaboration environments [4]. Empirical evidence demonstrates that people can learn to collaborate. This has been an important motivation in this work, because students can gain those collaborative skills that they still need to develop. Personalized assistance to students in collaborative learning environments has been recognized in recent years. Consequently, this kind of assistance should be made effective so that users accept it. To meet this goal, recording data about students' personal characteristics, knowledge and collaborative skills become completely necessary. Therefore, to obtain an effective personalization in CSCL environments, it is important not only to know what skills a student has developed but also the conditions under which the collaboration has been made effective. Such conditions could be defined by the composition of the group, the features of the task and the context of collaboration [10].

In computer-supported learning systems, personalization is achieved by collecting data about the students working with the system, constructing student models and using these models to adapt different aspects of the systems. Hence, these systems need to record collaborative skills that a student develops in different contexts together with his/her personal characteristics.

In this work, we introduce the notion of contextualized personal collaborative skills inside a student model. The main contribution of this work lies on this part of a student model that we called Collaborative Profile. It gathers individual collaborative skills of a student as the context in which these skills have been used. For example, our collaborative profile could contain that Peter uses the mediate skill in a group composed by peers. In other words, we specify the context in which each skill is used, being this context any combination of contextual parameters in the environments. This information is useful to provide student with support in training in collaborative skills that he/she has poorly developed, support an adequate automatic formation of learning groups based on members' collaborative skills, and adapt the type of collaborative activities to contexts in which the student shows an enhanced development of his/her skills, among other applications.

The second contribution of this paper is a method based on Web Mining techniques to capture collaborative contexts automatically from a computational processing of a log base representing usage behaviour of a learning environment. Thus, the combinations of parameters in which skills have been used are discovered by an automatic analysis. Particularly, we work on Web Usage Mining, which discovers potentially useful and previously unknown information from Web usage data and constructs a model of user's behaviour. Consequently, we define a collaborative profile for each student based on his/her collaborative behaviour patterns. This contribution represents a tool for scaffolding student in collaborative learning environments.

The work is organized as follows. Firstly we discuss some antecedents about the use of Web Usage Mining techniques in e-learning environments. Secondly, we describe the structure of the student model that we proposed. Then, we develop our approach to learn the Collaborative Profile which is part of the student model. Finally, we provide the experimental results and discuss technological consequences of this work.

2 Background

The development of either Web-based educational or e-learning systems has increased extraordinarily in the last years. This has encouraged the application of Data Mining techniques as a tool for improving learning in e-learning systems, defining a field named Educational Data Mining [31].

Some applications of Web Usage Mining techniques in the learning field are conducted for clustering and classification purposes, for example, clustering students by their navigation behavior and clustering pages according to context, type, references, or users' visits [30, 25, 23, 28, 24].

Other applications have the aim of discovering relationships or associations among different web pages visited and discovering navigation patterns [40, 12, 29, 32]. There are also applications of Web Usage Mining (WUM) techniques that aim at analyzing pages visited during a session or in different sessions of a single user [1, 18].

A very important application of Web Mining techniques in learning systems, although scarcely disseminated, is student profile generation. Some Data Mining algorithms are used to discover student profiles from log data [19, 22]. A significant advantage of these systems is that entries are not based on students' subjective descriptions performed by them and therefore are not biased. Student profiles are obtained from patterns in a dynamic way; thus, if profiles remain updated the system performance does not degrade over time.

In the reviewed literature, there are very few antecedents about the application of Web Mining techniques to discover students' profiles and especially in collaborative environments. We can only mention two works [17, 36]. The first one is a data mining application on student group interaction data to identify sequences of activities. In the second one a data mining application is explored to investigate the database generated by the system with the aim of building analytical models that summarize students' interaction patterns. But there are no antecedents about the application of WUM techniques to discover collaborative profiles. So, our work is novel since we discover collaborative skills and their context of usage applying WUM approach.

3 Collaborative Student Model

In this section we introduce the structure of our Collaborative Student Model with three components, each one recording a different category of related data with collaboration. These components are illustrated in Figure 1. Following, we present each component: the individual profile and the group profile give a work context to our proposal, and the collaborative profile concentrates our ideas in contextualized collaborative skills.

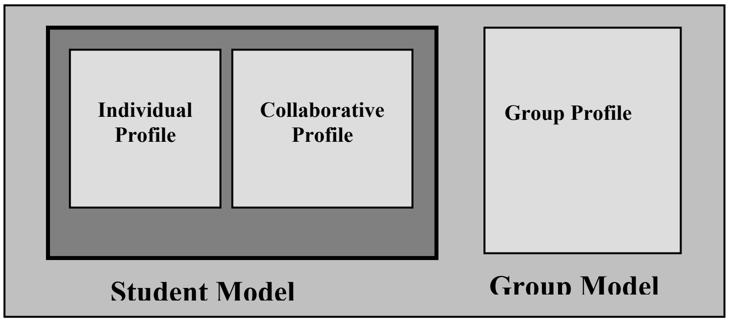

Figure 1: Collaborative Student Model components

Individual Profile: It contains three data categories: demographic data (user identification, first name, last name, date of birth, sex and nationality), knowledge domain data (level of the student's knowledge about a domain); and data about individual personality characteristics (learning style). Learning styles of Felder and Silverman [11] are used for experimental purpose because our students are from computer science, but it is possible to use another learning style model. Four dimensions of Felder and Silverman's model are considered: Perception (sensory, intuitive, sensory_intuitive), Input (visual, auditory, visual_auditory), Processing (active, reflective, active_reflective), and Understanding (sequential, global, sequential_global). The content of Individual Profile is presented in Figure 2.

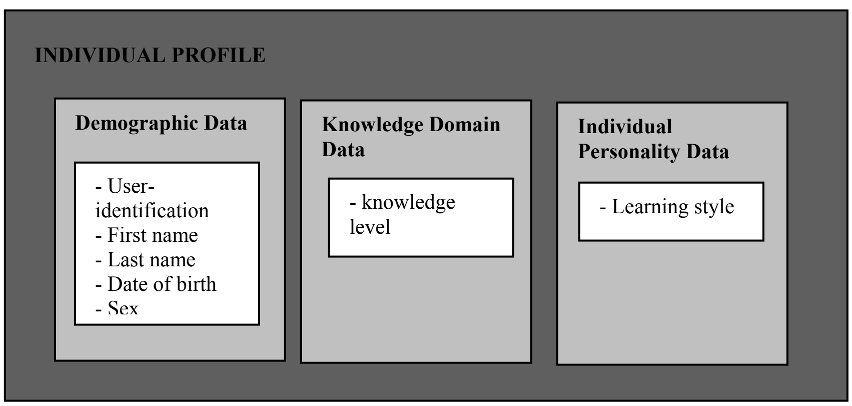

Figure 2: Content of Individual Profile

Example of Individual Profile:

Individual_Profile (pjohn) = (John;Smith;June 04,1978;M;canadian; sensory_visual_active_sequential; high)

The example describes the Individual Profile of John Smith, a Canadian man that was born on June 4 1978 and whose identification in the system is "pjohn". He has a learning style defined as (sensory, visual, active, sequential), and a high level of knowledge about the domain.

Group Profile: The following data are considered in this profile: code that identifies the group (*group-id*), students that compose the group (*members)*, *type of group* that indicates if the group is composed only of peers or peers and a tutor; *conflicts* that indicate occurrence of conflicts among group members and it is considered three types of conflicts [16]: task conflicts, interpersonal conflicts, and process conflicts; *contract* that indicates if there is a contract (document that a group elaborates before beginning the work, and it allows the students to establish their own “commitment rules”) among group members; division of work that indicates if group members work together (without_work_division) or if they divide the task into sub-tasks and solve them individually to finally put results together (with_work_division); *roles* that indicate the role played by each student in the group (Belbin’s classification (Table 2) is followed [2]). It considers the following roles: Shaper (IS), Implementer (ID), Completer/Finisher (FI), Coordinator (CO), Resource Investigator (IR), Team worker (CH), Plant (CE) Monitor/Evaluator (ME), and Specialist (ES). In Figure 3 the content of Group Profile is presented.

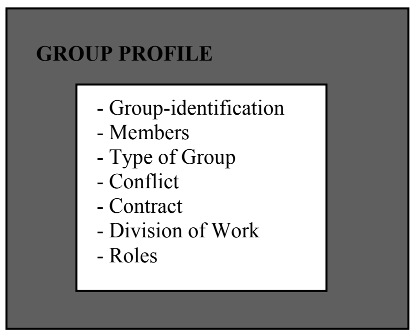

Figure 3: Content of Group Profile

Example of Group Profile:

Group_Profile (G1) = (pJohn, psusa, ppeter; only_peers; pJohn/CE, psusa/IR, ppeter/ID; task_conflict; with_contract; without_work_division)

The example describes the Group Profile of G1 group, whose members are students identified as: “pjohn”, “psusa” and “ppeter”. Only peers participate in the group. Also, in this group student ”pjohn” plays Plant (CE) role, student “psusa” plays Resource Investigator (RI) role, and student “ppeter” plays Implementer (ID) role. There are task conflicts in the group, there is a contract among members, and there is no division of work.

Collaborative Profile: it consists of a set of inputs, each one represented by a collaborative skill and a context in which it appears. The collaborative skills considered in this work are those included in Collaborative Learning Conversation Skill Taxonomy [33], which has been designed to facilitate the recognition of dialogue during active learning. Soller’s taxonomy breaks down each type of learning dialogue skill (active learning, conversation, and creative conflict) into subskills (for instance: request, inform, acknowledge), and attributes (e.g., suggest, rephrase). Each attribute is matched with a short introductory phrase, or sentence opener, which conveys the appropriate dialogue intention. Castelfranchi’s proposal [5] is also considered to enrich the taxonomy (see Table 1), mainly by incorporating *delegate* skill, which is considered very important in harmonic operation of learning groups.

SKILL	SUBSKILL	ATTRIBUTE	SENTENCE OPENER
Creative Conflict	Mediate	Teacher Mediation	"Let's ask to the teacher"
	Argue	Conciliate	"Both are right in that"
		Agree	"I agree because..."
		Disagree	"I disagree because..."
		Offer Alternative	"Alternatively..."
		Infer	"Therefore...", "So..."
		Suppose	"If...then..."
		Propose exception	"But..."
		Doubt	"I'm not so sure"
Active Learning	Motivate	Encourage	"Very good", "Good Point"
		Reinforce	"That's right"
	Inform	Rephrase	"In others words..."
		Lead	"I think we should..."
		Suggest	"I think..."
		Elaborate	"To elaborate...", "Also..."
		Explain/Clarify	"Let me explain it in this way...."
		Justify	"To justify..."
		Assert	"I'm reasonable sure..."
	Request	Information	"Do you know ...?"
		Elaboration	"Can you tell me more?
		Clarification	"Can you explain why/how...?"
		Justification	"Why do you think that?"
		Opinion	"Do you think...?"
		Illustration	"Please, show me...?"
	Delegate	Confidence	"I am sure you are able to do ..."
		Work	"Could you make it for my?"
Conversation	Acknowledge	Appreciation	"Thank You"
		Accept/Confirm	"OK" "Yes"
		Reject	"No"
	Maintenance	Request Attention	"Excuse me"
		Suggest Action	"Would you please....?"
		Request Confirmation	"Right?" "Is this Ok?"
		Listening	"I see what you're saying"
		Apologize	"Sorry"
	Task	Coordinate Group Process	"Ok. Let's move on", "Are you ready?"
		Request Focus Change	"Let me show you..."
		Summarize Information	"To summarize"
		End Participation	"Goodbye"
		Promote decisions maker	"Please, decide"

Table 1: Enriched Collaborative Learning Conversational Skills Taxonomy

Nowadays research tends to focus on the context of collaborative activity, uses broader definitions of collaboration and investigates it in a wider range of settings [15]. This work is based on Convertino's proposal [6] to define the context and it is extended using three sets of variables: characterization of collaborative situation, the group and the task.

To characterize collaborative situation, the following variables are considered:

Symmetry: two types of symmetry have been considered [9]: action symmetry, when all students in a group can perform the same range of actions and knowledge symmetry, when all members of a group have the same knowledge level.

Role: it refers to the role that a student plays in a group while solving a given task.

Predominant Learning Style: this variable indicates if the student's learning style is the same as or different from the learning styles of the other group members; it also indicates the type of learning style.

A group is characterized by the following variables: *group_type*, *contract, conflict* and *work_division*.

To characterize a task, the collaborative activities proposed by Gouli [13] were used. Then, four levels of collaborative activities have been applied:

Comprehension level: includes process and cognitive skills, which mainly refer to a student's ability to remember and understand things, to infer from facts/processes and to reason their inferences, to identify and specify main components of a construction/concept, to distinguish, classify, compare and relate concepts/ facts/ etc.

Application Level: includes process and cognitive skills, which refer to a student's ability to specify the steps to follow/perform a process, and/or the steps to implement/modify a product according to pre-specified rules/process, or by determining their constituent parts.

Evaluation Level: includes process and cognitive skills, which refer to a student's ability to test the correctness and/or completeness of a "product" and to reason about his/her own opinion.

Creativity Level: includes high level cognitive process and skills, which refer to a student's ability to analyze, elaborate, design and build a "product" by combining several processes/methods, and to plan and manage a project.

Four kinds of tasks are considered in our student model, which are derived from these four levels: comprehension, application, evaluation and creation. The content of Collaborative Profile is presented in Figure 4.

Example of Collaborative Profile

***Collaborative_Profile** (John) = (motivate;*
__context__ (__situation__ (with_action_symmetry,,,);
__group__ (only_peers,with_contract,,without_conflict,,);
__task__ (application));0.7;
argue;
__context__ (__situation__ (,with_knowledge_symmetry,,all_sensory,
all_visual,all_reflective,all_global);
__group__ (only_peers,without_contract,,without_work_division));
__task__ ());0.5)))

Motivate and *argue* are collaborative skills that *John* exhibits in some contexts. These contexts are automatically detected from John's behavior in different work groups. In the example, *John* reveals "*motivate*" skill in a context with the following characteristics (i) a collaborative situation in which all the members may perform the same type of actions in the system (*with_action_symmetry*); (ii) the groups are composed only of student peers (*only_peers*), there is a contract (*with_contract*), and conflicts have not been detected (*without_conflicts*); and (iii) the type of task is an application task (*application*). The value 0.7 is the percentage of appearance of the *motivate* skill in relation to this context considering the total of the student's interventions. *John* exhibits "*argue*" skill in a context with the following characteristics: (i) a collaborative situation in which all members have the same domain knowledge level (*with_knowledge_symmetry*), and *John*'s learning style is the same as the rest of group, which is *sensory_visual_reflective_global*; (ii) previous work groups are characterized by being integrated only by student peers (*only_peers*), they do not have a contract *(without_contract*) and all members work together to solve the task (*without_work_division*).

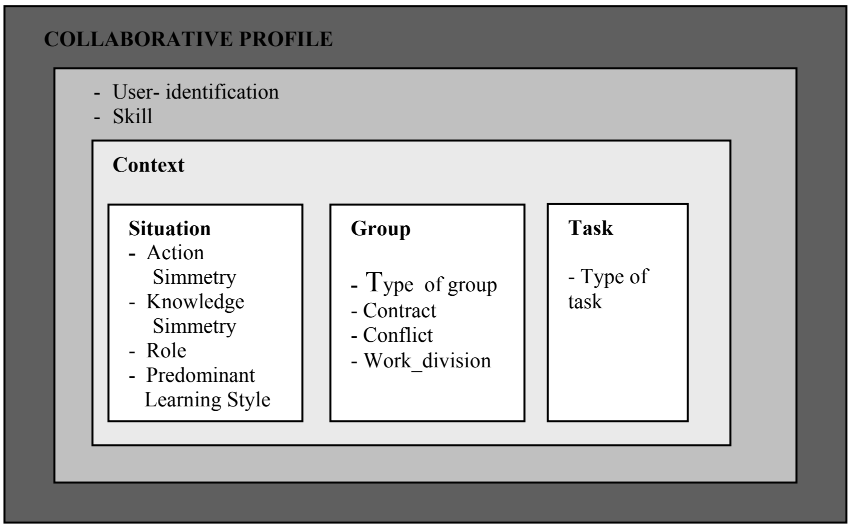

Figure 4: Content of Collaborative Profile

4 The Web Usage Mining approach

Due to the relevance of tools for scaffolding students in learning environments [27], we present a way for supporting an automatic capture of contextualized collaborative skills. Then, to learn the Collaborative Profile, the central idea is to analyze the student's behaviors in previous collaborative activities. In a CSCL environment, the student's behavior is shown through their interactions that are recorded as dialogues. Therefore, it is relevant to analyze these dialogues and to elucidate student intentions, because they reveal the collaborative skills that a student puts into practice. Thus, it is necessary that the student's intervention in the collaborative work is recorded in log files generated by the CSCL system. Consequently, it is necessary to apply a technique that allows us to identify collaborative behavior patterns, based on the analysis of the information that was stored in log files. The approach based on WUM [20] was our choice, and particularly, we apply association rules. It is important to highlight that the WUM approach is not novel; but, up to now it was not applied to discover contexts in collaborative learning environment.

Following we present an outline of the proposed approach for learning our collaborative profile (see Figure 5) and each stage of the approach is described.

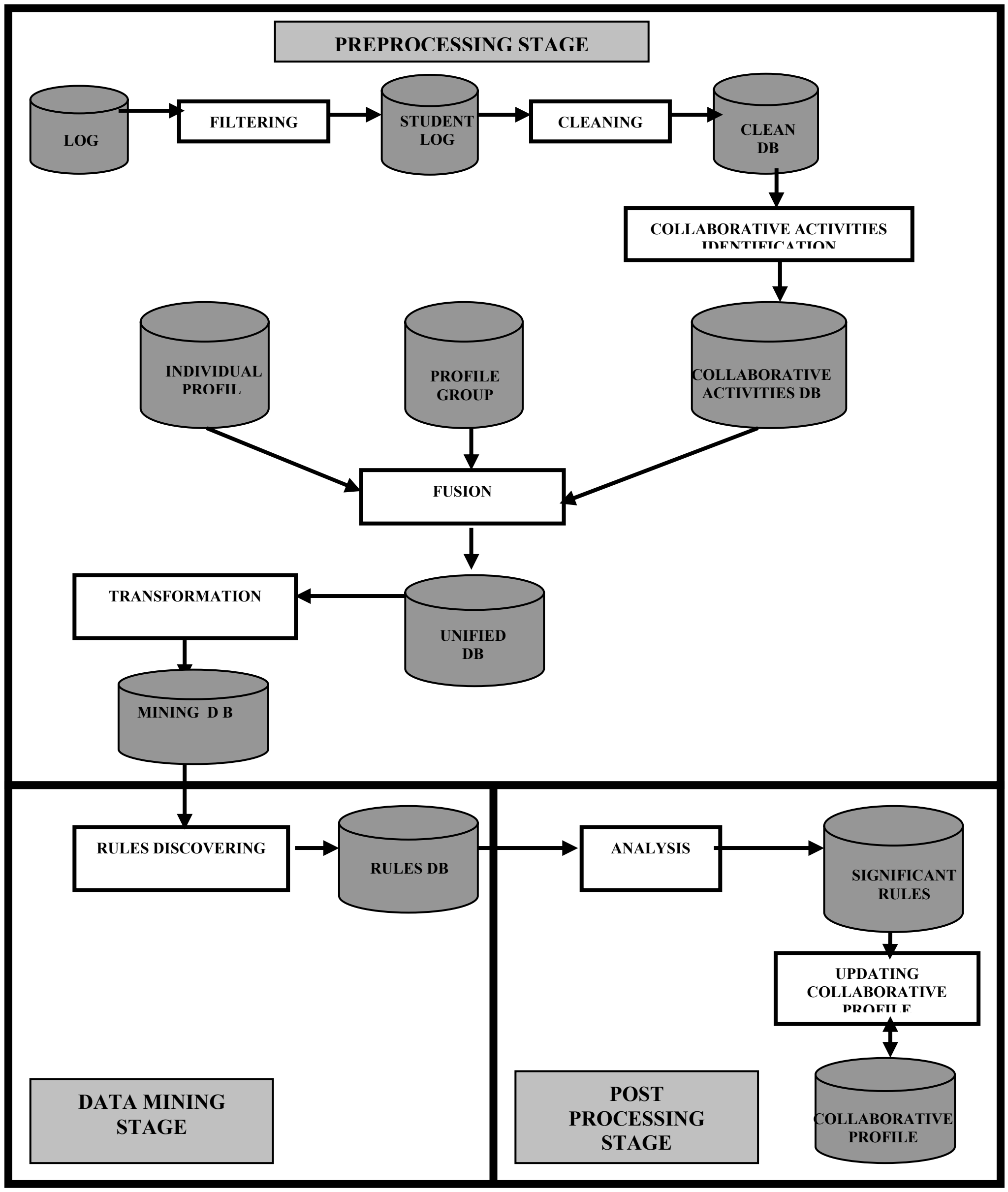

Figure 5: Web Usage Mining Method for learning the Collaborative Profile

4.1 Pre-processing stage

The first step in any WUM process is to identify the source data where the Data Mining task will be applied. In our approach the source data are built from: databases that contain Individual Profiles of all members of the groups in which the student has participated, databases of Group Profiles in which the student took part, and log file of the system that contains, among other things, the interventions of each student during the collaborative work in a dialogue environment (structured, semi-structured or free). The steps that are detailed below were carried out from these sources.

Data Filtering: This step is performed on the log file generated by the system. Since the log file contains the records of all the students interacting in the system, it is necessary to retrieve only those records that belong to the student whose profile is being generated.

Data Cleaning: This step is performed on the output file of the filtering process. First, missing values are completed, noises are eliminated, and inconsistent data are corrected if necessary. The output is a clean database.

Identification of Collaborative Activities: The objective is to select from the clean database those sections that belong to collaborative activities.

Data Fusion: Three data sources (Individual Profile, Group Profile, and the log file) are merged to generate a Unified Data Base (UDB). In the fusion only those data that are considered relevant to discover patterns are taken. The UDB attributes are presented in Table 2

Data Transformation: In this step the data are transformed in an appropriate way to be mined. First, three additional attributes necessary to elucidate collaborative behavior patterns are created: <skill>, <predominant_style> and <student_role>. To create the <skill> attribute, it is necessary to convert the attribute <intervention> of the UDB into a sub-skill of the Enriched Collaborative Learning Conversational Skills Taxonomy (Table 1). The <predominant_style> attribute is derived from <learning_style> attributes of the students' individual profiles that conform the group. Each dimension takes the value "all" plus the learning style of that student whose Collaborative Profile is being built, if the student's learning style coincides with the others group members' learning styles (i.e. all_global); each dimension takes the value "majority" plus the most frequent learning style in the group (i.e. majority_intuitive), if the student's learning style coincides with the learning style of the majority of group members; each dimension takes the value "any_predominant". To create the <student_role> attribute, it is necessary to convert <members_roles> attribute of the UDB. Finally, it is necessary to modify the format of some data to adjust them to the requirements of data mining tools. In this work we used ARTool [7]. Likewise, depending on the algorithm applied, certain specific formats are necessary for the data (for example, the Apriori algorithm only uses nominal data). The output of this step is a Mining Database (see Table 3).

4.2 Knowledge discovery stage

At this stage the knowledge discovery method and algorithm are selected. In this approach, association rules are used as a method to discover knowledge in collaborative activities. This technique was selected because the main objective of this work is to discover a student's collaborative behavior patterns in different contexts. Therefore, it is necessary to discover interesting relationships that characterized the group, task and collaborative situation where a student puts into practice his/her collaborative skills during the dialogues.

The knowledge discovered during the WUM process consists of interesting patterns. A pattern is considered interesting if it validates a user's hypothesis [14]. In this work we defined four hypotheses, which have been derived from both our experiences as teachers and from some antecedents [6, 38] about the main variables that define our collaborative context.

H1. *A student develops different collaborative skills depending on the collaborative situation in which he/she acts, the group that he/she belongs to and the type of task that he/she must solve in a collaborative manner.*

H2. *A student develops different collaborative skills depending on the collaborative situation in which he/she acts and the characteristics of the group that he/she belongs to.*

H3. *A student develops different collaborative skills depending on the collaborative situation in which he/she acts and the type of task that he/she must solve in a collaborative manner.*

H4. *A student develops different collaborative skills depending on the characteristics of the group that he/she belongs to and the type of task that he/she has to solve in a collaborative manner.*

From these hypotheses, we define a set of meta-patterns to be used as a guide in the process of knowledge patterns discovery. These meta-patterns have the following generic form:

$$Context \rightarrow Skill\ (X,Z) \text{ where } Context = P(X : student\_id, W)\ \mathbf{\Lambda}\ Q_i\ (X,Y_j\ )$$

$$\text{being } i = 1, 2, ..., n \text{ data belonging to our Mining Database}$$

$$\text{being } j = 1, 2, ..., m \text{ attributes of the characteristics } Q_i$$

Where X is the student relation key; P and Q_i are predicate variables that are instantiated with relevant attributes belonging to our Mining Database (e.g., predominant learning style, role that the student plays in the group, group type, existence of contract, etc.); and W, Yj, and Z are object variables that may take values of their predicates to the student X.

ATRIBUTES	SUB-ATRIBUTES	VALUES
student_id		
group_id		
task		comprehension|application|evaluation|creation
intervention	dialogue_type	structure|semiestructure|free
	sentence	
learning_style	Perception	intuitive|sensory|sensory_intuitive
	Input	visual|auditory|visual_auditory
	Processing	active|reflective|active_reflective
	Understanding	sequential|global|sequential_global
type_group		only_peers|peers_with_tutor
conflict		without_conflict|task_conflict|interpersonal_conflict|process_conflict
contract		with_contract|without_contract
work_division		with_work_division|without_work_division
action_symmetry		with_action_symmetry|without_action_symmetry
knowledge_symmetry		with_knowledge_symmetry|without_knowledge_symmetry
members	student_id	
members_roles	student_id	
	role	IS|ID|FI|CO|IR|CH|CE|ME|ES

Table 2: UDB attributes

ATRIBUTES	SUB-ATRIBUTES	VALUES
student_id		
group_id		
Task		comprehension|application|evaluation|creation
skill		mediate|argue|motivate|inform|request|delegate|acknowledge|mantainance|task
type_group		only_peers|peers_with_tutor
conflict		without_conflict|task_conflict|interpersonal_conflict|process_conflict
contract		with_contract|without_contract
work_division		with_work_division|without_work_division
action_symmetry		with_action_symmetry|without_action_symmetry
knowledge_symmetry		with_knowledge_symmetry|without_knowledge_symmetry
predominant_style	predominant_perception	all_intuitive|all_sensory|all_intuitive_sensory|majority_intuitive|majority_sensory|majority_intuitive_sensory
	predominant_input	all_visual|all_auditory|all_visual_auditory|majority_auditory|majority_verbal|majority_visual_auditory
	predominant_processing	all_active|all_reflective|all_active_reflective|majority_active|majority_reflective|majority_active_reflective
	predominant_understanding	all_secuential|all_global|all_sequential_global|majority_secuential|majority_global|majority_sequential_global
student_rol	role	IS|ID|FI|CO|IR|CH|CE|ME|ES

Table 3: Mining Data Base

Then, the association rules search is limited only to those rules that respond to the defined meta-patterns. For a first version of our Collaborative Student Model, and considering the defined hypotheses, the following meta-patterns are defined:

For Hypothesis 1:

MP.1: <situation> (X: *student_id, W*) Λ <group> (X, Y) Λ <task> (X, V) → <skill> (X,Z)

This meta-pattern links a context characterized by a given collaborative situation <situation>, a group <group> with certain characteristics and a given type of task to solve, with the collaborative skill <skill> that the student reveals in that context.

For Hypothesis 2:

MP.2: <situation> (X: student_id, W) Λ <group> (X, Y) → <skill> (X,Z)

This meta-pattern links a context characterized by a given collaborative situation <situation>, and a group <group> with certain characteristics, with the collaborative skill <skill> that the student reveals in that context.

For Hypothesis 3:

MP.3: <situation> (X: student, W) Λ <task> (X, Y) → <skill> (X,Z)

This meta-pattern links a context characterized by a given collaborative situation <situation>, and a determined type of task to solve, with the collaborative skill <skill> that the student exhibits in that context.

For Hypothesis 4:

MP.4: <group> (*X*: student_id, *W*) Λ <task> (*X, Y*) → <skill> (*X,Z*)

This meta-pattern links a context characterized by a group with certain characteristics <group> and a determined type of task to solve, with the collaborative skill <skill> that student exhibits in that context.

4.3 Analysis of discovered knowledge and profile maintenance stage

Given that the objective of this step is to select the most relevant rules, to generate and maintain the student's profile, it is necessary to define convenient criteria to select only the most significant rules. In this work, two interesting measures are used: first the support (percentage of analyzed data tuples in which the pattern appears), and second the number of attributes that the antecedent of the rule has. Those rules that exceed the minimum threshold of support are selected, and if there is more than one, those rules that have the greatest number of attributes in the antecedent are selected to generate or maintain the student's collaborative profile. With each selected rule, an input of collaborative profile is then generated. In profile maintenance, if the input already exists, its support is updated replacing the existing value by the new one.

5 Experiments

The goal of the experiments is to evaluate the student modeling process, specially the Collaborative Profile modeling process. Then, the following question can be made: "*are the student's collaborative skills successfully detected to be stored in the Collaborative Profile?*"

Following Weibelzahl and Weber's proposal [39], a simulation technique was applied to validate the inference that the system made. This technique has been selected because the evaluation of collaborative systems should be made in current use context [37, 26], but, unfortunately, the field of real work does not often offer all the alternatives. Therefore, an approximation to a more systematic evaluation is to simulate the condition of real use of the system. Simulation represents an advantageous evaluation method because it allows better control and it is more accurate, and also, it minimizes the risk of error and cost [6].

In this work, twenty four hypothetic collaborative students' profiles were defined, and students' interventions in a CSCL environment were simulated with a simulation software to cover different alternatives (different groups, situations and tasks). In the software each characteristic of the context is generated like an aleatory variable and the parameters to simulate them are estimated in base on information provided by teachers with experience in collaborative learning. The simulated data were recorded in Mining Database and they were adapted to process them with the ARTOOL software. From this adapted Mining Database, the WUM process was executed to obtain the association rules that define the collaborative patterns of the students.

To obtain the frequent itemsets, different experiments were made with different minimum support values. The first objective was to obtain rules that have the <skill> attribute as consequent. Then, these rules were post-processed and the uninteresting rules were filtered out (i.e. those rules that have not only the <skill> attribute in the consequent). For example the next rule was eliminated because it has other attributes in the consequent besides to *motivate*:

Antecedent: application, all_actives
Consequent: motivate, all_visual, all_global (Support: 0.39)

Then, redundant rules were eliminated (i.e. those rules that have in the antecedent attributes that are a subset of the antecedent of the other rule. For example in the next rules, the second one was eliminated:

Antecedent: application, without_contract, without_work_division, without_action_symmetry, withou_knowledge_symmetry, all_active, all_global
Consequent: motivate (Support: 0.38)

Antecedent: application, without_contract, without_work_division, without_action_symmetry, all_active
Consequent: motivate (Support: 0.38)

Finally, those rules that fit into the defined meta-patterns were selected. For example, for the student identified like *pjhon*, the following rules were discovered, as they fit into MP1:

Antecedent: application, only_peers, interpersonal_conflict, without_contract, without_action_simmetry, all_sensorial, all_visual, all_active, all_global
Consequent: motivate (Support: 0.39)

Antecedent: aplication, only_peers, without_contract, without_work_division, without_knowledge_simmetry, all_sensorial, all_visual, all_active, all_global, CE
Consequent: argue (Support: 0.32)

Each attribute that is part of the discovered contexts has been qualified considering the percentage of times that it appears in our Mining Database. Therefore, an attribute has a score of 10 if it appears in the 100% of the Mining Database transactions; it has a score of 9 if it appears in a range between 90% and 99% of the Mining Database transactions, and so on. Finally, the patterns discovered were corroborated with hypothetic profiles originally specified. In Table 4 the attributes of the pjhon's hypothetic profile are presented. We indicate with "yes" all the attributes that were discovered by the rules with a minimum support of 0.3, and in the other column the qualification of each attribute.

The total percentage of learning with a support of 0.2 for all simulated students is presented in Figure 6. This graph was constructed considering that a student's profile is learned if his/her skills are discovered and at least 50% of the attributes of context are also discovered. From this graphs we can conclude that the method is efficient to learn collaborative profiles with a support of 0.2.

To complement this result and for a more detailed analysis of the achieved learning, in relation to the attributes scores and the percentage of times that a skill is manifested in the student's intervention, the dispersion graphs (Figure 7 and Figure 8) have been calculated.

SKILL	CONTEXT	SCORE	DISCOVERED ATTRIBUTES
			Support 0.3
Motivate	**Situation**		
	without_action_simmetry	9	YES
	IS	5	
	All_sensorial	10	YES
	All_visual	10	YES
	All_active	10	YES
	All_global	10	YES
	Group		
	only_peers	10	YES
	without_contract	10	YES
	Interpersonal_conflict	10	YES
	Task		
	Application	9	YES
Argumentar	**Situation**		
	Without_knowledge_simmetry	10	YES
	CE	4	
	All_sensorial	10	YES
	All_visual	10	YES
	All_active	10	YES
	All_global	10	YES
	Group		
	only_peers	10	YES
	without_contract	10	YES
	Without_work_division	10	YES
	Task		
	Application	9	YES

Table 4: Discovered attributes for *pjhon's* profile and the scores of the attributes

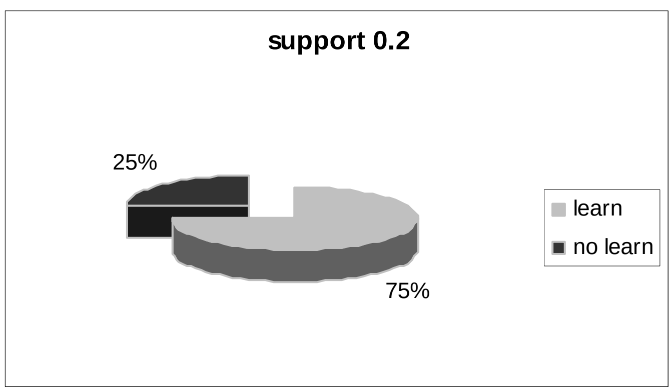

Figure 6: Learning profiles in relation to the total number of simulated students

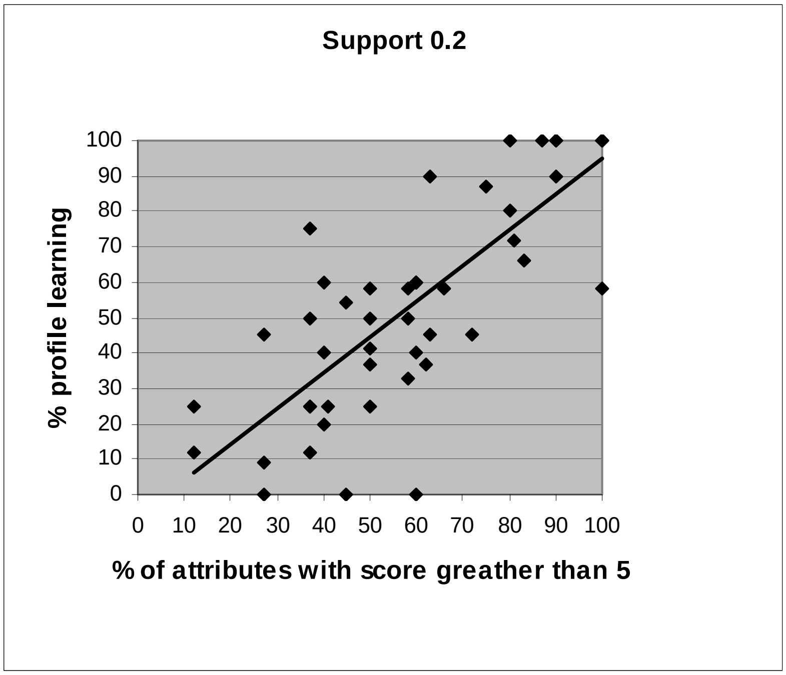

Figure 7: Dispersion of profile learning in relation to the percentage of appearance of a skill in de DB

Each point in Figure 7 represents the learning percentage of each profile in relation to the percentage of attributes with scores higher than 5 (i.e. the point (28,10) indicates that in a profile with only 28% of attributes with scores higher than 5 only 10% of the attributes are discovered). The tendency of the 24 simulated students indicates that:

- Learning higher than 0% is reached when the profile has at least 10% of attributes with score higher than 5.
- The tendency shows that, a profile learning that varies between 10% and 95%, is achieved for a support of 0.2, for profiles that possess at least 10% of their attributes with a score higher than 5.

Each point in Figure 8 represents the learning percentage of each profile in relation to the times that a student manifests a skill (i.e. the point (40,10) indicates that in a profile with a skill that is manifested 40% of the times only 10% of the attributes are discovered). The tendency of the 24 simulated students indicates that:

- If a skill is manifested in less than 20% of the student's interactions, no context is detected for that skill.
- If a skill is manifested in more than 20% of the student's interactions, there is a tendency to improve profile learning that varies between 20% and 100%.

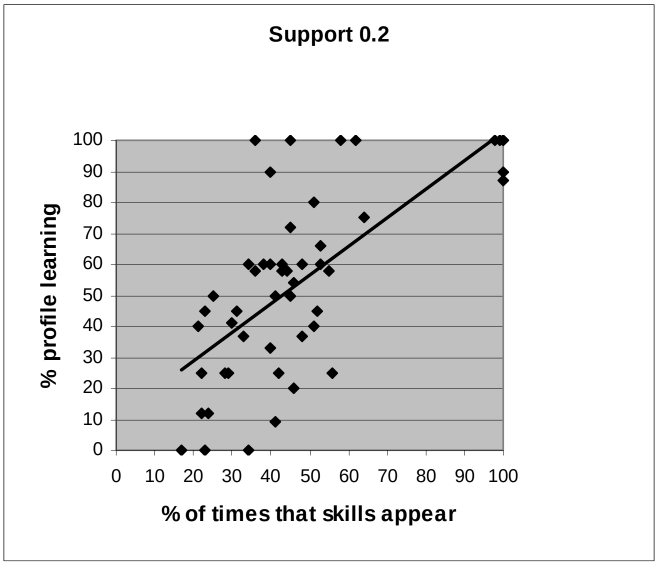

Figure 8: Dispersion of profile learning in relation to the % of attributes with a score higher than 5

To check the hypotheses which were formulated in Section 4.1, the application of the meta-patterns were evaluated. In Figure 9 the results about the application of the meta-patterns are synthesized. In this graph, it is possible to observe that most discovered rules fit into MP2, no rules fit into MP3, few rules fit into MP1 and MP4, and there are some discovered rules that do not fit into the proposed meta-patterns. From these results it is possible to conclude that <task> attribute is not relevant in the context, because it is the common attribute among MP1, MP3 and MP4. In consequence, H2 is the only hypothesis that is accepted and H1, H3 and H4 are rejected.

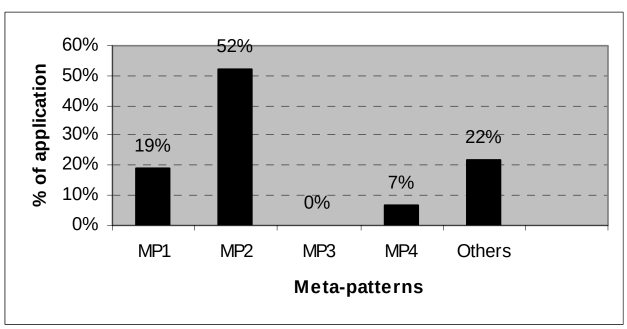

Figure 9: Percentage of meta-patterns application

6 Conclusions

We have introduced a Collaborative Student Model and we have proposed a WUM approach to generate automatically the Collaborative Profile. This approach allows us to detect a student's skills and the contexts where they are revealed for a more effective collaboration in a distance learning environment. The evaluation proved that the approach is efficient to learn collaborative profiles with a support of 0.2 in a 75% of the cases.

In this work we have simulated the student's intervention in collaborative educational systems and we have applied association rules to automatically generate knowledge about students' collaborative patterns. We formulated four hypotheses about the context where the student's skills are manifested, and the experience with simulated students proved that only one of them is true, because the <task> attribute is not relevant in the context. Then, the principal context components are the characteristics of a situation and the characteristics of the group.

References

1. Becker, K., Marquardt, C.G. and Ruiz, D.D. 2004. A Pre-Processing Tool for Web Usage Mining in the Distance Education Domain, Proc. of the International Database Engineering and Applications Symposium (IDEAS'04). 00, 78-87.
2. Belbin, R. 1981. Management Teams. John Wiley (ed.). New York.
3. Bransford, J**.,** Brown A., and Cocking R. 1999. How People Learn: Brain, Mind, Experience, and School. National Academy Press.
4. Burton, M., Brna P. Pilkington R. 2000. Clarissa: A Laboratory for the Modelling of Collaboration. International Journal of Artificial Intelligence in Education, 11, 79-105.
5. Castelfranchi, C., de Rossi, F. and Falcone, R. 1997. Social Attitudes and Ground Social Interaction. In Association for the Advancement of Artificial Intelligence (AAAI) Fall Symposium Series. Socially Intelligent Agents, Massachusetts, November 1997. AAAI Press.
6. Convertino, G., Neale, D.C., Hobby, L., Carroll, J. M. and Rosson, M. B. 2004. A Laboratory Model for Studying Activity Awareness. In Proceedings of NordiCHI 2004. Tampere, Finlandia. ACM Press.
7. Cristofor, L. 2002. ARtool v1.1.2. [Computer Software], Association Rule Mining Algorithms and Tools.
8. Dillenbourg, P. and Tchounikine, P. 2007. Flexibility in macro-scripts for computer-supported collaborative learning. Journal of Computer Assisted Learning, 23(1), 1-13.
9. Dillenbourg, P. 1999. What do you mean by "collaborative learning"?. In P. Dillenbourg (ed.) Collaborative-learning: Cognitive and Computational Approaches. Elsevier: 1-19.
10. Dillenbourg, P., Baker M., Blaye A., and O'Malley C. (1996). The evolution of research on collaborative learning. In E. Spada and P. Reiman (eds.) Learning in Humans and Machine: Towards an interdisciplinary learning science. Elsevier. 189-211.
11. Felder, R. M. and Silverman, L. K. 1988. Learning and teaching styles in engineering education. Journal of Engineering Education, 78(7), 674-681.

12. Freyberger, J., Hefferman, N.T. and Ruiz, C. 2004. Using Associations Rules to Guide a Search for Best Fitting Transfer Models of Student Learning, in International Conf. on Intelligent Tutoring Systems.

13. Gouli, E., Gogoulou, A., Grigoriadou, M. and Samarakou, M. 2003. Towards the Development of an Adaptive Communication Tool Promoting Cognitive and Communications Skills. In Proceedings of the 11th International Conference PEG 2003: Powerful ICT Tools for Teaching and Learning. June-July, 2003, St. Petersburgo, Rusia.

14. Han, J. and Kamber, M. 2001. Data Mining. Concepts and Techniques. Morgan Kaufman Publishers, San Francisco, USA.

15. Hoppe, H.U., Joiner, R., Milrad, M. and Sharples, M. 2003. Guest editorial: Wireless and Mobile Technologies in Education. Journal of Computer Assisted Learning,19(3), 255-259, September 2003.

16. Jehn, K. A. and Mannix, E. A. 2001. The dynamic nature of conflict: a longitudinal study of intragroup conflict and group performance. Academy of Management Journal 44(2), 238-252.

17. Kay, J., Maisonneuve, N., Yacef, K., and Zaïane, O. 2006. Mining patterns of events in students´ teamwork data, in Proc. of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems, Taiwan, 45-52.

18. Kristofic, A. and Bieliková, M. 2005. Improving Adaptation in Web-based Educational Hypermedia by means of Knowledge Discovery, in Sixteenth ACM Conference on Hypertext and Hypermedia (HT'2005), September 2005, Salzburgo, Austria, 184-192.

19. Mavrikis M. 2008. Data-driven modelling of students' interactions in an ILE, in First International Conference on Educational Data Mining (EDM'08), Canada.

20. Mobasher, B. 2004. Web Usage Mining and Personalization. In Practical Handbook of Internet Computing. Munindar P. Singh (ed.), CRC Press.

21. Mobasher, B., Cooley, R. and Srivastava, J. 2000. Automatic Personalization Based on Web Usage Mining, Communications of the ACM, 43(8), 142-151.

22. Mor, E. and Minguillón, J. 2004. E-Learning Personalization based on Itineraries and Long-term Navigational Behavior Can Museum Exhibits Support Personalized Learning in Collaborative Classroom Activities By Using Argumented, in World Wide Web Conference, New York, 264-265.

23. Nasraoui, O., Krishnapuram, R., Frigui, H. and Joshi, A. 2000. Extracting Web User Profiles Using Relational Competitive Fuzzy Clustering. Artificial Intelligence Tools, 9(4), 509-526.

24. Ng Cheong Vee, M-H, Meyer, B. and Mannock, K.L. 2006. Understanding novice errors paths in Object-Oriented programming through log analysis, in 8th International Conference on Intelligent Tutoring Systems, Jhongli, Taiwan, June 2006.

25. Paliouras, G., Papatheodorou, C., Karkaletsis, V. and Spyropoulos, C.D. 2000. Clustering the Users of Large Web Sites Into Commmunities, in Proceedings of 17th International Conference on Machine Learning (ICML), Morgan Kaukmann, San Francisco, CA, 719-726.

26. Prinz, W., Mark, G. and Pankoke-Babatz, P. 1998. Designing Groupware for Congruency in Use, in Proc. of Conf. on Computer Supported Cooperative Work (CSCW'98), ACM Press, 373-382.

27. Puntambekar, S. 2005. Tools for Scaffolding Students in a Complex Learning Environment: What Have We Gained and What Have We Missed?. Educational Psychologist, 40(1), 1-12.

28. Ramadhan, H.A., Fiaidhi, J.A. and Ali, J.M.H. 2005. The Aplication of Web-Mining to Theme-Based Recommender Systems, in International Journal of Instructional Technology and Distance Learning. May 2005, 2(5), 9-20.

29. Retalis, S., Papasalouros, A., Psaromiligkos, Y., Siscos, S. and Kargidis. T. 2006. Towards Networked Learning Analytics – A concept and a tool, Networked Learning 2006.

30. Romero, C., Ventura S., Espejo P., and Hervas C. 2008. Data Mining Algorithms to Classify Students, in First International Conference on Educational Data Mining (EDM'08), Canada.

31. Romero, C. and Ventura, S. 2007. Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, 33(1), 135-146.

32. Romero, C., Ventura, S., and Bra, P. de. 2005. Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Author. User Modeling and User Adapted Interaction, 14(5).

33. Soller, A. 2001. Supporting Social Interaction in an Intelligent Collaborative Learning System. International Journal of Artificial Intelligence in Education, 12, 40-62.

34. Stahl, G. 2002. Contributions to a theoretical framework for CSCL, in Proceedings of *Computer Supported Collaborative Learning (CSCL 2002)*. Boulder, CO.

35. Stahl, G., Koschmann T., and Suthers D. 2006. Computer-supported collaborative learning: An historical perspective. Based on a chapter in: R. K. Sawyer (ed.). *Cambridge Handbook of the Learning Sciences*. Cambridge, UK: Cambridge University Press.

36. Talavera, L. and Gaudioso, E. 2004. Mining Student Data to Characterize Similar Behavior Groups in Unstructured Collaboration Spaces, in Workshop on Artificial Intelligence in CSCL. 16th European Conference on Artificial Intelligence (ECAI'04), 17-23.

37. Twidale, M., Randal, D. and Bentley, R. 1994. Situated Evaluation for Cooperative Systems. In Proceedings of Conference on Computer Supported Cooperative Work (CSCW'94), ACM Press: 22-26.

38. Vonderwell S. and Zachariah S.. Factors that Influence Participation in Online Learning. *Journal of Research on Technology in Education*. ISTE, 38 (2), 213-230.

39. Weibelzahl, S. and Weber, G. 2003. Evaluating the Inference Mechanism of Adaptive Learning Systems. In Brusilovsky, P., Corbett, A. and de Rosis, F. (Eds.) User Modeling: Proc. of the Ninth International Conf.. Lecture Notes in Computer Science, 154-168, Berlin: Springer.

40. Zaïane, O. 2001. Web Usage Mining for a Better Web-Based Learning Environment. In Proceedings of Conference on Advanced Technology for Education, 60-64, Banff, Alberta.