Trace-based Collaborative Learning System

Yacine Lafifi¹, Noureddine Gouasmi¹, Khaled Halimi¹, Wassila Herkas², Nassima Salhi¹ and Assia Ghodbani¹
¹Department of Computer Science, University of Guelma, Algeria
²Department of Psychology, University of Guelma, Algeria

The users of any educational software may leave traces which concern all their activities. In collaborative learning context, these traces are very voluminous and very heterogeneous. They are the results of various interactions between the actors themselves, and between the actors and the system. Hence, first, they must be collected and filtered. Then, these traces must be analyzed to help or support the actors. In fact, the analysis process has a positive effect on the role of human actors. It helps the tutor in his task of monitoring learners, the teacher (author) in his task of creating educational courses and the learner in his task of viewing his previous actions with more details. It is this context that defines our research work, which is focused on implementing a collaborative learning system based on traces called SYCATA (which is the French acronym of: SYstème pour la Collecte et l’Analyse des Traces d’Apprentissage collaboratif). This system collects all traces of actors’ activities (especially learners) and groups them into five categories. It offers a multitude of forms (graphical, numerical or mixed) to show these traces to the tutors and the authors. Some traces may be viewed by learners to promote their pedagogical activities and raise their awareness. SYCATA was implemented and experimented with a sample of university students where good results were obtained.

Keywords: CSCL, trace, collaborative learning, trace visualization, trace analysis, indicator

1. Introduction

Any activity carried out using software leads the user to leaving traces in a file, where it is sometimes possible to infer interesting facts. Since educational software has existed, the researchers have collected and analyzed the traces left when students use the software, either to follow the work or to improve implemented systems or to understand phenomena associated with their use [5].

In the last few years, trace analysis has been developed and new research fields have emerged concerning the nature and especially the structure of data, models, methods and techniques to analyze and visualize indicators in order to foster learning [6]. These traces are often used for re-learning scenarios [16], to help understanding the user’s behaviour [22] or his learning style, to support the creation of intelligent help tutorials [14], to monitor the learners or to modify their profiles.

These traces can be used in collaborative learning context to facilitate the process of the learners monitoring. In fact, in Computer-supported Collaborative Learning (CSCL), other than the cognitive aspect, learners monitoring must take into account their behavioural aspect. In other words, the interactions between learners should be analyzed to know the content of the exchanges and their nature. This analysis allows the tutor to know the deadlocks in the functioning of each group of learners, to follow the progress of the behavioural profile of each participant and/or each group, to motivate isolated learners and lazy groups, etc. [19]. All these advantages allow one adequate assistance and evaluation of each learner and/or group. This implies that the tutor must have a representation, as accurate as possible, of the activities done by the learner (these activities may be individual or collaborative, cognitive and/or behavioural) and should view these activities to facilitate their interpretation.

In this work, we propose a taxonomy of traces that can be used in collaborative learning context. These traces can be used to assist the hu-
man actors in CSCL environments and provide feedbacks on various activities/tasks of learners. The main objective is to give each actor the means for traces perception and visualization enabling him to carry out his role. The perception of what happens during the learning activity plays a crucial role in enabling tutors to make decisions they deem relevant and carry out their tasks. Furthermore, some collected traces may be used by authors to see the relevance of courses and proposed learning scenarios.

To use them effectively, first of all, these traces must be collected. Then, to be used or interpreted, they must be analyzed. But these traces are very heterogeneous and very large because they concern all the activities done by the learners throughout the period of the system use. That makes the task of viewing these traces very tedious, especially for tutors, teachers and even learners. The difficulty is not limited to this; it also concerns the exploitation and analysis of the large amount of data.

Several questions can be asked: What is the relevant information for each actor? When and how to collect the traces? How will these traces be viewed by the actors concerned? The authors are interested in activities related to learning and assessment processes, the tutors are much more interested in the interactions between learners and their behaviours, while the administrator focuses on the traces related to the available communication tools as well as on the frequency of accesses to different resources of the system, etc. In what form will the traces of a particular learner be presented to his tutor? This form must be understandable and help him to make good decisions according to the cognitive and behavioural profiles of the learner. And, finally, how to analyze the traces?

To answer all these questions, we have implemented a system for collecting and analyzing a set of traces recorded during the collaborative learning phases. The presentation of this system, called SYCATA (system of collecting and analysing traces of collaborative learning), is the goal of this article.

The article is structured as follows. In Section 2, we present a state-of-the-art on related works where we cite some learning systems based on the traces. Section 3 is devoted to presenting the architecture of SYCATA. Answers to questions raised previously will be presented in Section 4.

Before concluding the article, in Section 5 we give the results of the experiment conducted at an Algerian university.

2. State of the Art

2.1. Trace Definition

Currently there are several viewpoints on what could be the definition of a trace. According to Jermann and his colleagues [15], a trace is an observation or the recording of a learner’s interaction with the system for analysis.

In the same vein, J-P. Pernin [30] defines a trace as an indicator of the actors’ activity in a learning situation, whether it is instrumented or not. In a slightly different way, P-A. Champin [3] speaks about a sequence of states and transitions representing a user’s activity, “the time sequence of objects and operations mobilized by the user when he uses the system is called trace of use” (extracted from [13]).

The researchers of TRAILS project (Personalized and Collaborative Trails of Digital and Non-Digital Learning Objects) see traces of use in hypermedia as a sequence of actions and use them to identify the overall objective of the user. As for Choquet and Iksal [5], they consider as trace any data providing information on a learning session broadening the definition given by the researchers of the TRAILS project.

In our research, we adopted the definition given by Settouti and his colleagues [32]. They define a numerical trace as a “trace of the activity of a user who uses a tool to carry out this activity, saved on a numerical medium”. In this article, the terms “trace” and “numerical trace” are used interchangeably.

2.2. Trace-based Learning Environments

Many learning environments used traces. e Médiathèque [7] is a CSCL system used to view traces in real time. This environment provides for a learner an interactive visualization of his own traces of use during his activities. e Médiathèque aims at improving the reflexive activities in collaborative learning. All user’s interactions with this tool are observed and treated from two models [26]: the former
defines the observable objects (available tools such as Internet browser or instant messaging, but also resources such as texts, images, etc.) and the latter specifies actions carried out by the user (creating, changing, deleting content, etc.) [7].

Another system using traces is DREW (Dialogical Reasoning Educational Web tool) [31]. In DREW, the students exploit traces of activity (especially those of chat sessions) in favour of the reflective process of learning. It’s in a collaborative situation that students are able to retrace their traces, and co-construct a diagram which organizes and prioritizes the elaborated arguments developed during the learning process [7].

Regarding the learners monitoring, we can cite APLUSIX [4], which is a help environment for learning algebra using the learning traces in order to help students solve the exercises. This tool lets students solve exercises and check (by using the traces) whether the solutions are correct and complete.

The tutor in “REFLET” [8] and “FORMID” [11] can see the traces of the learners’ activities. Both tools display each learner’s progress.

In order to analyze the traces, we can mention ColAT (Collaboration Analysis Tool) [1] and LISTEN [28]. ColAT is a learning traces analysis tool independent of any learning system. In the LISTEN project, researchers have developed an intelligent tutor which is capable of listening and interpreting the oral products of a learner in his learning to read English. They developed a set of knowledge on collecting and analyzing traces, and then sought to formalize this know-how as software architecture and methodological recommendations [27].

Some systems are used for the visualization of traces. We can cite “Gismo” (Graphical Interactive Student Monitoring System for Moodle) [24, 25] which outlines the activities of learners using Moodle. Gismo displays traces in a graph by representing the number of accesses to different Moodle resources by different learners.

CourseVis [23] is used for 3D visualization of traces stemming from the WebCT platform. The objective is to make visualizations from the calculation of indicators and statistical measures. CourseVis is used for the tutors to provide a kind of a dashboard of a learning activity (group monitoring) [9].

Finally, JCachesim [2] is a simulation environment of the operation of a computer. It enables learners to observe the functioning of the central processing unit (CPU) and the invisible tasks during the execution of a program and in particular operations of reading/writing. The teachers can follow the learner’s traces by viewing the log file where all interactions are recorded.

3. Architecture of SYCATA

SYCATA is primarily a collaborative learning environment. It has most of the features found in CSCL systems (collaboration support, communication tools, etc.). An important feature of this system is the ability to manage a set of traces representing different interactions between the different actors and the different actions carried out. The originality of our work is the distribution of these traces into five categories (detailed in the following section) facilitating the processes of visualization and analysis. These traces represent all the activities carried out by the learners during their use of the system.

A very important component in SYCATA is an assistant system of the tutor (adviser), which is used to advise him in the task of monitoring the learners. This assistant system uses a basic set of rules initialized by experts in the field (tutorial) and a set of indicators established from collected traces. It provides the tutor with a set of actions to be done for each chosen learner (sending messages, setting an appointment, programming a virtual meeting, etc.).

3.1. Taxonomy of the Proposed Traces

To process these traces, it seems necessary to build traces taxonomy to thereafter define the corresponding treatment to a similar type of traces. Also, traces taxonomy will regroup the different approaches for the processing of traces in the same structure [13].

In literature, we found traces of learning that are used by several researchers, traces of interaction [12, 17, 29, 31, 33] and traces of use [7, 20, 21].
Gwenegan [13] has identified four groups of traces: the footsteps information (personal information and technical information related to the medium used (system version, IP address, . . . )), traces related to the exploitation of a resource (name, reference resources, access number, . . . ), traces associated with the learning activity (the quality of a production, response time, test results . . . ) and the traces attached to the communication activity (number of messages sent and read in a forum, message content, . . . ).

For a better exploitation of the traces collected by our system, we suggest five classes of traces. These are the main classes of traces that provide objective data about the learners and their activities. The proposed classes are:

a. **Learning traces**: traces of access to the content of educational material in SYCATA (learning objects, concepts, educational resources, etc.) and corresponding consultation time (date, duration, etc.).

b. **Assessment traces**: these traces concern data about the assessment tests (exercises, marks, assessment duration, etc.).

c. **Communication traces**: they concern the interaction between the learner and the other actors or between the learners themselves outside the collaborative activities (i.e., there is no explicit request for collaboration).

d. **Collaboration traces**: they concern the collaborative activities such as replies to the forum, acceptance of collaboration request, collaboration tool, collaboration date, content of the collaboration activity, etc.

e. **Traces of use**: they include information on the access to the system (date, frequency, last access . . . ), access to system’s resources (learning objects, tests), access to the collaborators search engine (it’s a tool that allows searching collaborators verifying some criteria [18]) and communication tools (chat, forum, e-mail).

For a better exploitation of the traces collected by our system, we suggest five classes of traces. These are the main classes of traces that provide objective data about the learners and their activities. The proposed classes are:

<table>
<thead>
<tr>
<th>Type of traces</th>
<th>Associated Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning traces</td>
<td></td>
</tr>
<tr>
<td>Assessment traces</td>
<td></td>
</tr>
<tr>
<td>Communication traces</td>
<td></td>
</tr>
<tr>
<td>Collaboration traces</td>
<td></td>
</tr>
<tr>
<td>Traces of use</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Representation of proposed traces.

3.2. Functional Architecture of the System

Figure 1 gives an overview of the functional architecture of SYCATA. It consists of four main interfaces which are associated with each of the following actors: administrator, learner, author (teacher) and tutor. Also, it contains a traces manager which is made up of four components: Traces collector module, Indicators calculator module, Traces visualization system and Assistant system of the tutor.

This architecture contains a database including all necessary information on the system’s actors and the interactions between them. Finally, it includes a web server that provides navigation.

**Traces manager**

It allows managing all collected traces (addition, deletion, filtering, etc.) and all the indicators to be used by the visualization system or by the assistant system of the tutor. It includes the traces preservation process and the rights of displaying these traces.
a. Traces collector module
This module allows the collection of learners’ traces and their classification into different classes. It uses some tools and processes to convert data into traces. The collected traces are distributed into five categories (c.f. Section 3.1).

b. Indicators calculator module
It allows calculating the indicators concerning the learners’ activities. The indicators are used during or after the learning activity and represent a variety of information such as the number of accesses to a resource, the time spent on each exercise or the success rate [32]. We consider an indicator as a variable that is calculated from a trace or a set of traces.

In SYCATA, several indicators can be obtained from a single trace. For example, the mark obtained in the assessment activity is considered as a trace. By applying formulas, the first indicator can be the value of cognitive profile of the learner, while the second one can be the number of correct/false answers in a given period.

c. Traces visualization system
It provides the actors with statistical and graphical representation of traces (sectors, stick diagrams, etc.). Visualization is the representation (temporal data) of all collected traces during the learning period of the learner.

The visualization process of these traces is the following. After filtering the traces, the user must choose the viewing period (the current day, any previous day, any past week or also a well determined period). Next, the user must choose the mode of viewing traces: graphical, numerical or mixed. After seeing all the traces,
the user can choose any traces class for more details. For each class, a set of information that characterize the type of traces will be displayed.

d. Tutor’s assistant system

Its role is to assist the human tutor in his task of monitoring the learners. In other words, it can send advice in the form of a message to the learners requiring assistance. These advice are the results of the execution of a series of tutorial rules. The rules are updated by field experts. The assistant system uses a set of indicators (calculated from the traces left by the learners) and basic tutorial rules (established by the experts in the field).

Furthermore, this system resembles an expert system. It has an inference engine which reasons using a database of rules and facts (initialized when registering a new learner) to deduce advice.

4. Research Questions

4.1. What is the relevant information for each actor?

To facilitate the use of traces, each activity of the learner is saved where a set of data are recorded. These data include: the actor owner of the activity, type of activity, starting date, completion date, etc. For each actor, we have provided a set of information to facilitate his tasks. For the teachers, information on consulting concepts and results after the assessment process are offered. As for the tutor, information on any activity of the learner is taken into account.

To sum up, the relevant information concern, first of all, the identity of the learner (name, cognitive profile, etc.) and information on his activities.

4.2. When and how to collect the traces?

The collection of traces is an essential phase before any use of such traces. This phase will collect the traces resulting from the activities of learners. It is done immediately after the completion of an activity by the learner. These traces are recorded in the trace database (see Figure 2).

4.3. How to present the traces to the actors concerned?

One of the difficulties of using traces is their visualization. In fact, registration of any datum from the learner’s activity is very easy, but how will it be reproduced and in what form must it be displayed in order not to lose its importance? There are traces that will be more significant with the graphical mode while others will be more significant with the numerical mode (statistics). In our work and in order to better

Figure 2. Traces visualization process.
reproduce the traces (for maintaining its importance), we offered the following visualization modes: graphical, numerical and mixed.

The visualization is very interesting for:

- **The learner**: to improve his knowledge as well as his cognitive and behavioural profiles. The visualization of his traces promotes his educational activities and constitutes a good reflexive tool.

- **The tutor**: the visualization allows him to know information on the progress of every learner as well as his attitudes to work in regular time intervals. This offers him the opportunity to intervene and provide assistance to the learners who are in trouble.

- **The teacher (author)** can see some traces of the learners concerning learning and assessment processes. So, he can improve the content of the learning objects, the assessment exercises and the pedagogical scenarios.

The tutor can see the traces of all his learners, whereas the learner can only see his own traces. The users of these traces have the opportunity to choose the viewing time of these traces:

- **Traces of the current day**: they can see all traces of the activities done during the current day.

- **Daily traces**: they can see traces of any previous day.

- **Weekly traces**: this option allows them to see all traces of the current week.

- **Traces of a given period**: SYCATA offers the users the opportunity to see the traces done during a given period (month, bimonthly, annual, etc.).

Figures 3 and 4 show some traces viewed by the learner (they are the same traces that can be seen by the tutor of this learner).

For each activity, the learner or the tutor can find a detailed plan with all information on each trace, such as:

- Activity designation.

- Starting date / Completion date of this activity.

- Starting time / Completion time of this activity.

- Other information according to the type of each activity. For example, in the case of collaboration traces: receiver of collaboration request, object of collaboration request, date and time of the collaboration request, appointment of collaboration, collaboration tool, duration of collaboration, etc.

To see these details, the tutor/learner can click on the image associated with each traces class (see Table 1).

### 4.4. How to analyze the traces?

To better enjoy the collected traces, several means for their analysis are provided. It’s through the assistant system of the tutor that such traces can be exploited. Indeed, these traces can be transformed into a set of indicators. It is from these indicators that the analysis will be done.

The assistant system is based on the values of calculated indicators to make its advice to tutors. For example, according to the value of the indicator “refusal rate of collaboration”, there will be a piece of advice to be sent to the learners concerned. If this rate is over 80%, a message requesting the learner to collaborate with his colleagues is sent to him. Finally, we remind our readers that each indicator is associated with a set of advice to send to the learners concerned (Figure 5).

This analysis allows to:

- See the progression of cognitive/behavioural profile of each learner during a previously set period (one week, one month, etc.).

- Reconstruct the path of the learner to show his progress in a learning session. This allows explaining the background that has led the learner to perform an action or to obtain a particular outcome [10].

- See the list of isolated learners or too solicited ones.

- Calculate the rate of an activity, frequency of sending messages . . . of each learner.

- See the frequency of using every available tool: search engine of collaborators, communication tools, etc.
Figure 3. Traces visualization in SYCATA.

Figure 4. Visualization of weekly traces with details.
5. Experiment

5.1. Methodology

We conducted an experiment at the computer science department of Guelma University (Algeria). The participants were forty two $(n = 42)$ students from 2nd year licence. These students were divided randomly into two groups: experimental and control. The experimental group consisted of 20 students $(n_1 = 20)$ while the control group consisted of 22 students $(n_2 = 22)$. The subject learnt was: “Languages theory” which is taught in the 4th semester (February – June) of the licence degree (“Information systems” speciality). It was designed by two teachers from the computer science department and included about one hundred concepts and assessment exercises.

The experimental group used SYCATA with all its options and features while the control group used an electronic version of the same course. At the end of the utilization of the system (after two months), all students were subjected to the same final exam (noted on 20 points). In addition, the students in the experimental group were asked to complete a questionnaire (it concerned the functioning of SYCATA and its features). It consisted of 37 multiple choice questions.

5.2. Results and Discussion

Throughout this experiment, we observed that students were very satisfied with this system and its functionalities. The most admired element was the visual appearance of several traces with its different forms (graphical, numerical, etc.). The visualization process attracted the attention of most students who thought that the possibility of showing the learning pathways in detail was very useful to increase the capacity of knowledge memorisation. According to the students, these traces find out the weaknesses and the strengths of each one of them. In addition, these students believe that the tutor’s intervention is very useful to direct or encourage them. They appreciate the content of the messages sent by their tutors.

**Verification of the hypothesis**

In order to know the input of our system, we propose to test the following hypothesis:

Null hypothesis $H_0$: the use of activities traces does not have any effect on the cognitive profiles of learners.

To verify this hypothesis, we compared the averages of the two groups of students (control and experimental). To determine if the difference is significant between the two averages, we used the $t$-test as the sample size is less than 30. After using the software R [34], which is a free software, we obtained the following results with a confidence level of 95%:

<table>
<thead>
<tr>
<th>Group</th>
<th>$n$</th>
<th>Average</th>
<th>Std Deviation</th>
<th>$t_{score}$</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>20</td>
<td>12.85</td>
<td>2.39</td>
<td>-5.88</td>
<td>0.001</td>
</tr>
<tr>
<td>Control group</td>
<td>22</td>
<td>8.77</td>
<td>1.51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to the $t$-test table, $t_{0.975} = 2.04$, whereas $t_{score} < t_{0.975}$ ($-5.88 < -2.04$), the difference is very significant. Therefore, the null hypothesis $H_0$ is rejected in favor of the alternative hypothesis. So, we can say that the use of activity traces by the learners has a positive effect on the level of learner’s knowledge (cognitive).
Questionnaire results and problems encountered

In the following section we present the responses from students in the experimental group ($n = 20$) to some questions in the questionnaire (composed of 37 questions).

1. How do you see the use of learning traces?
   Need:
   
   ![Graph showing response distribution for need of learning traces](image)

   Usefulness:
   
   ![Graph showing response distribution for usefulness of learning traces](image)

2. How do you see the use of collaboration traces?
   Need:
   
   ![Graph showing response distribution for need of collaboration traces](image)

   Usefulness:
   
   ![Graph showing response distribution for usefulness of collaboration traces](image)

3. How do you see the interventions of your tutor?
   Need:
   
   ![Graph showing response distribution for need of tutor's interventions](image)

   Usefulness:
   
   ![Graph showing response distribution for usefulness of tutor's interventions](image)

4. How do you see the visualization of traces?
   Importance:
   
   ![Graph showing response distribution for importance of trace visualization](image)

   Quality:
   
   ![Graph showing response distribution for quality of trace visualization](image)

Concerning the problems encountered, the learners cited: the lack of opportunity to record the displayed traces and the lack of tools for storing conversations made between the learners themselves. They propose to obtain help for expressing their assistance requests (using a semi-structured interface, for example).
Regarding the tutors, they appreciated the use of this system, especially the traces visualization and the use of the assistant system. However, they noticed some gaps on the graphical representation of some traces, in particular those of communication traces (by sectors) and found the process of updating system setting wizard too long (tutors want to have assistance). They expressed their need for a guide who would provide more explanation regarding their roles and tasks.

6. Conclusion and Future Work

The aim of our research is to develop tools for the collection, analysis, interpretation and reuse of traces in CSCL environments. Currently, we have implemented a system that collects and analyzes the traces of the learners (baptized SYCATA).

In this article, we presented SYCATA, which is a trace-based collaborative learning system. Its main feature is the use of traces of activities produced by the learners during their learning process. So, its main objectives are to foster learning, facilitate the analysis and interpretation of these traces by the tutor, ensure proper use of traces for the learners to guarantee effective monitoring in remote contexts, offer feedback to authors on the behaviour of the learners towards the content and structure of learning objects, etc. As a result of our research, a model for the collection and use of traces is proposed. This model is based on the distribution of these traces into five categories: traces of learning, assessment, communication, collaboration and, finally, traces of use. These traces are presented to tutors in various forms, from simple statistical charts to complex graphics and diagrams. Furthermore, each type of traces is associated with a graphical form (ellipse, rectangle, etc.).

Another very important aspect of this system is the exploitation of these traces. In this context, advice may be sent to learners having difficulties in any of the suggested pedagogical activities: learning, assessment, collaboration and use of tools offered by the environment.

In order to test the input of this system, we have conducted an experiment at the computer science department of Guellma University. The obtained results were very satisfactory. In fact, we have observed that the participants were very enjoyed by using SYCATA and its features. The idea of using traces and viewing them at any moment drew the attention of many learners and tutors. As a result, the use of the system has a positive effect on their cognitive levels.

As perspectives to this work, we currently work on the design of a model for the interpretation of traces produced by learners. Furthermore, we want to study the reuse of these traces. In the last case, a model of “reusable traces” must be conceived. These two projects are the subject of two master theses at the computer science department.

References


Trace-based Collaborative Learning System


[34] R-Project: http://www.r-project.org/