Integration of Automatic Tools for Displaying Interaction Data in Computer Environments for Distance Learning

Aloys Mbala, Christophe Reffay, Thierry Chanier

Integration of Automatic Tools for Displaying Interaction Data in Computer Environments for Distance Learning

Aloys Mbala, Christophe Reffay, Thierry Chanier
Laboratoire d'Informatique de l'Université de Franche-Comté
16, route de Gray, 25030 Besançon Cedex
{mbala,reffay,chanier}@lifc.univ-fcomte.fr

Abstract : Our research concerns distance learning (DL). We are interested with distributed collaborative learning. In this approach, it is important to have indicators permitting the appreciation of durability and the evolution of groups involved. We think that actors responsible for the organisation and the working of groups (tutor for each group and coordinator of the DL session for all groups and its progress in general) can from the types of interactions and their amounts, get revealing elements permitting them to appreciate the state of a group and its evolution. From the analysis of interactions seen during a distance learning experimentation that we led, we show here that the disappearance of a group as we observed could be discerned practically in real time. It justifies for us, the necessity to set up in distance learning environments, agents capable of assisting the coordinator of the training and the tutors in their tasks.

1. Introduction

Current computer environments for distance learning don't have tools which assist learning groups nor support their cohesion. With reduced size groups, involved in a collaborative learning process, this limitation becomes a crucial problem. When one or two members of such groups abandon, the whole group may disappear.

Conversely, the existence in a group of some motivated and dynamic people creates an auspicious group dynamics to the effectiveness of collaborative learning. It is therefore important to be able to form groups or to reorganise them on the fly, in such a manner that every group gets a critical mass, sufficient to generate the beneficial effects expected in this type of pedagogy. In DL, roles of group support and observation often rely on two types of actors : the tutor responsible for his/her group and the coordinator, who has a view on every group and who is responsible of the progress of the whole training course.

We show in this paper that during the training period, one can foresee the evolution of the group from very simple indicators (frequency and number of interactions). It becomes possible to design tools to support group management for the coordinator and the tutor. We analyse interactions recorded during a DL experimentation, named *Simuligne*, and we put in perspective the evolution of interaction data and the behaviour of the groups as we observed. From these intra-group interaction data, it would have been possible to predict the progression of these groups.

This leads us to the conclusion that DL platforms must be endowed with supplementary functionalities permitting to display in real time individual involvement and group behaviour. This new type of environment we call it SIGFAD, a French acronym of '*Soutien des Interactions dans des Groupes de Formation A Distance*'. We introduce a specification of the system, based on the MaSE methodology, a multiagent based approach. MaSE permits, from an initial set of requirements, to develop all steps going from prose specification to an implemented agent system.

2. Theoretical foundations

The management of user groups in DL is fundamental. There exist many reasons for this. The positioning and integration of learners in the group help us to remedy various phenomena which are largely identified as parts of DL, notably : sociological isolation of the learner, loss of motivation. The group constitutes the immediate materialisation of the accompaniment of the learner (this is also valid for other actors, but since the learner is the main issue of every learning enterprise, we'll often restrict to its unique evocation).

Collaborative learning has a natural context in DL, precisely due to the fact that it supposes social interactions among users that create a sociological setting in a manner preventing isolation and maintaining motivation. Some authors used the expression *'distributed collaborative learning'* to refer to the implementation of collaborative learning in DL [11].

Many researchers highlight the fact that in computer-supported distance learning, it is difficult to evaluate the level of interaction and communication among individuals [13]. It is quite surprising, because all user's actions are recorded in log files. These data can be analysed and information about interaction and communication among individuals becomes available. In fact, if researchers made this remark, it simply means that they had no system to analyse log data and to display intra-group interactions. Developing such systems is our current concern.

In general, one can notice an evolution of research on collaborative learning from how individuals do function in a group to the group itself becoming the unit of analysis [8]. This evolution is supported by human and social theories like Activity Theory (AT) that take into consideration persons interacting with each other, with tools such as computers, symbolic or natural languages [1].

In the context of distributed collaborative learning, the group takes a singular importance. It is the role of tutor to bring and maintain cohesion in the group. The coordinator of the training period has a further more crucial role. He is the one in charge of good progression over the training course and is concerned with the upbringing of various groups. In fact, there should be a minimum quantity of interactions occurring in each group and for this reason, the coordinator may have to reorganise groups, close some or reinforce others. We need a critical mass (in terms of active persons) in every group that can generate sufficient intra-group interactions and productions. Many studies aim at designing systems for supporting group interactions [15], [21].

3. The study

We designed and managed a DL experimentation of ten weeks. Its purpose was to collect data which would be analysed later and give a better understanding of which factors influenced the progression of a DL course. It concerned forty English-speaking learners split up in four small size basic groups A, G, L, N. We call them basic groups because there were other groups (i.e. the M one which gathered all users). The experimentation named *Simuligne* consisted in participating into a global simulation (frequently used by language teachers in intensive face-to-face classes); every basic group had to participate in a competition to designate the French virtual city which could welcome learners of an English university for their summer school. The challenges for every basic group were then to elaborate the candidature (entirely written in French) for the Open City competition. Every basic group had one tutor and a couple of native persons who served as peer companions for learners. Natives were students of Faculty of Letters of *Université de Franche-Comté* (Master in French as Foreign Language). A coordinator helped tutors planned the activities by discussing with them in a designated group restricted to tutors and natives. The group members didn't know themselves before and the whole course was entirely computer-mediated at a distance. Our course has been implemented on WebCT™.

The course was divided in three stages (not taking into account the preliminarily one where everyone introduced him/herself and discovered technical features of the environment). Each stage comprised several activities. Every interaction and communication data occurring during the course were recorded. Tutors and the coordinator had only access to the basic evaluation tools offered by the WebCT platform, which will be referred to as CEDIL (CEDIL stands for Computer Environment for DIstance Learning) in the follow-up.

Let us now give a conceptual description of a Simuligne activity (the reader interested in a more general view of CEDIL used during Simuligne will find it in [3]).

The activity theory (AT) proposed by Leontiev (quoted in [12]) is the most suitable one to explain the structure of a Simuligne activity because it takes into account the group as the basic unit of analysis and gives a central position to the concept of activity. Fjuk et al. [12] argues that AT gives conceptual accounts of work and development, and the role of artefacts (like the computer) within social contexts. Figure 1 presents a model of a Simuligne activity, adapted from the basic structure of activity of Engelström [9]. An activity has an object/goal and is part of a stage; it means that the course has a number of stages, each one composed of a number of activities. An activity is composed of tasks. A subject uses a tool to achieve a task. A subject can be a member of several groups, but he has a unique status in a given group. The usual concepts of 'division of labour' and 'rules' of AT are implemented in our model using status and tasks. We added the notion of time (not present in

Engeström's activity) : an activity possesses an earliest date of beginning and a latest date of end. In other words a subject, with a given status in a group, uses tools (each one linked to a task) for achieving an object of an activity during a length of time.

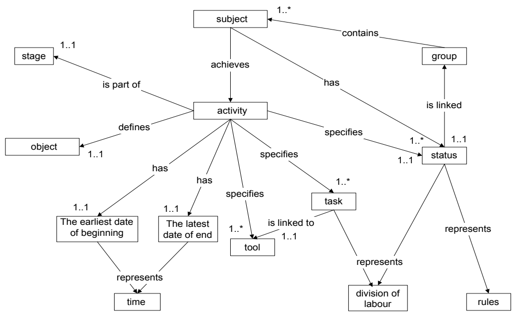

Fig. 1. Conceptual model of a Simuligne activity

The general diagram of CEDIL is shown in figure 2. Every user (login + password) is a member of a course (which we used as group) and have a permanent status in the group. CEDIL offers four user profiles: learner, administrator, designer and pedagogic assistant, giving different rights and tools to each of them. We had to implement each Simuligne status (learner, coordinator, tutor, native) as a CEDIL user profile. We had three status in a basic group: learner implemented as learner CEDIL profile, native also implemented as CEDIL learner profile, tutor implemented as CEDIL designer profile. The coordinator was implemented as CEDIL administrator profile but could also connect herself in every basic group as a silent learner in order to observe what happened in the group. We should also notice the absence of reference to the community (group) in CEDIL; we had to implement our Simuligne groups as CEDIL courses as palliative.

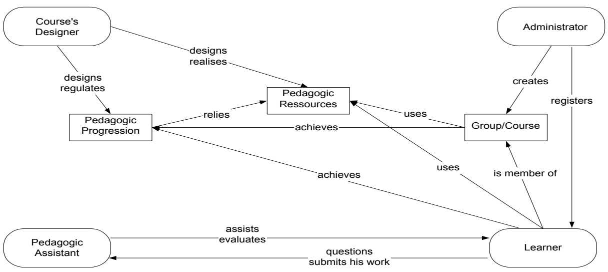

Fig. 2. The General Diagram of CEDIL

During Simuligne, which took place from the 30th of April to the 07th of July 2001, we observed a decrease of activities in the group L to the point that we had close it on May 30. The two remaining active learners were transferred to another group.

We present figures of interaction accumulated weekly in the first four weeks of Simuligne. Tables 1 and 2 respectively present the accrued percentages of the number of times a member accessed his/her group and the duration of his/her online connection. These data give an estimation of the level of activity of the group, because individual activity happened off-line (and are not measured here). From the first week, group L had a relatively low level of activity, especially in terms of connection time.

Basic groups	Week1	Week1+2	Week1+2+3	Week1+2+3+4
A	35,39%	31,21%	30,47%	29,73%
G	22,17%	27,34%	29,01%	30,15%
L	21,75%	20,95%	20,16%	19,44%
N	20,68%	20,50%	20,36%	20,68%
Total	100,00%	100,00%	100,00%	100,00%

Table 1. Accrued percentages of number of entries in CEDIL groups. Week 1 is the first week of stage E1, actual beginning of the simulation, after the completion of the preliminary stage E0.

Basic groups	Week1	Week1+2	Week1+2+3	Week1+2+3+4
A	42,03%	34,99%	33,47%	32,31%
G	18,09%	23,16%	23,93%	25,55%
L	17,94%	17,54%	17,72%	17,27%
N	21,93%	24,31%	24,88%	24,87%
Total	100,00%	100,00%	100,00%	100,00%

Table 2. Accrued percentages of connection time per basic group

Tables 3 and 4 respectively present the accrued percentages of number of intra-group read mails and intra-group posted mails. In CEDIL, groups are insulated in the sense that mails are only available within the group: a user can send/receive mails only to/from a member of his group. This means that the number of mails posted or read partly denotes the vitality of the group. Group L appears again as having a lower level of activity (particularly noticeable when looking at posted mail). On the contrary, group G progressively appears as the most active one whichever table one looks at.

Basic groups	Week1	Week1+2	Week1+2+3	Week1+2+3+4
A	29,15%	23,61%	21,81%	20,92%
G	13,70%	20,69%	25,65%	31,15%
L	19,24%	19,88%	18,84%	17,40%
N	37,90%	35,82%	33,70%	30,53%
Total	100,00%	100,00%	100,00%	100,00%

Table 3. Accrued percentages of number of intra-group read mails

Basic groups	Week1	Week1+2	Week1+2+3	Week1+2+3+4
A	27,62%	24,09%	22,88%	22,24%
G	20,95%	25,41%	29,01%	35,12%
L	17,14%	15,51%	15,09%	14,05%
N	34,29%	34,98%	33,02%	28,60%
Total	100,00%	100,00%	100,00%	100,00%

Table 4. Accrued percentages of number of intra-group posted mails

Tables 5 and 6 present the accrued percentages of number of messages read and posted in the forums respectively. Due to the conception of Simuligne, this is probably the most significant indicator of the level of activities in a group. We should keep in mind that during the period covered by our study (30^{th} April to 30^{th} May), the forum was certainly the tool the most frequently used. Eight forums were opened in every basic group during the period covered by our study. It is not therefore surprising that the group L that died out on May 30 presents some extremely low forum-oriented interaction rates.

Basic groups	Week1	Week1+2	Week1+2+3	Week1+2+3+4
A	50,65%	42,39%	40,27%	38,16%
G	25,26%	28,27%	26,62%	26,26%
L	12,66%	13,55%	13,69%	13,94%
N	11,43%	15,79%	19,43%	21,64%
Total	100,00%	100,00%	100,00%	100,00%

Table 5. Accrued percentages of number of read messages in forums

Basic groups	Week1	Week1+2	Week1+2+3	Week1+2+3+4
A	57,75%	50,26%	49,46%	47,71%
G	19,01%	23,28%	20,43%	19,82%
L	8,45%	9,26%	9,32%	9,45%
N	14,79%	17,20%	20,79%	23,02%
Total	100,00%	100,00%	100,00%	100,00%

Table 6. Accrued percentages of number of posted messages in forums

The data contained in the previous tables were extracted from the HTTP server log file. This file registers every user's action in a different line containing information like the IP address of the user, his CEDIL login name, the date and the type of action, the size of the file. These data are not directly available in the log files. We extracted the data from raw information contained in log files, saved them in databases and built them from appropriate SQL queries. These manipulations are essential, and current statistical analysis tools of HTTP logs can not provide such interaction data. Furthermore, one needs to go inside several CEDIL files in order to build group interaction data.

We present in the next table, the durations of connection to instruction pages from the beginning of the session to the 30th of May. The durations are not presented weekly as the other data but also reveal that group L paid few attention to instructions. The durations were calculated from data recorded in specific files of CEDIL.

	Number of access	Duration (hh mm ss)	Duration (in seconds)	Percentage
A	626	103h06m03s	371 163	32,26%
G	364	99h18m26s	357 506	31,08%
L	379	56h22m43s	202 963	17,64%
N	390	60h46m40s	218 800	19,02%
Total	1759		1150432	100,00%

Table 7. Connection times to instruction pages

Up till now, we can conclude that information collected in log files can provide, after judicious computer operations, interaction data to the users in such manner that they can know daily and even in real time the situation of the group and predict their evolution. For conciseness reasons, we presented in the previous tables weekly data aggregates, but this could be also done per day and even in smaller time units as an hour. We would want to highlight the fact that if some authors argue that "*in computer-supported distance learning classes, it is often difficult to know to what extent individuals are interacting and how much they communicate with other class members"* [13], it simply denotes the absence of requisite functionalities in DL platforms. The existing platforms need to be coupled to systems that collect interaction information in appropriate places, computerize them and display judicious data on the state and the durability of the groups.

4. SIGFAD, a multiagent system to assist users involved in DL

In section 3, we reported a computer-mediated DL experimentation. In this training course, we had four basic groups and one of them disappeared. In fact, we closed this group because we noted its very weak level of activity and because the minimum number of active persons expected to play a role in the simulation was not attained. The remaining two learners were "sent" to another group. We took this decision (closing the group and transferring learners) after examining the daily journal of the coordinator and the information collected from different log and tracking-actions and processed by simple programs. The kind of information displayed in the previous tables were not accessible and come from a post processing.

We argue that existing DL platforms must be provided with automatic tools that will make it possible for one to know the state of a group, the behaviour of users and predict in real time the evolution of the DL training course. The objective of showing different types of interaction data as done in the previous section is to demonstrate that judicious indicators can be built from data recorded in different files. Computer scientists are challenged to build these automatic indicators and to provide them to users in such a manner that offers a straightforward and immediate usage.

The agent paradigm is a good solution to this challenge. This paradigm is suitable to the metaphor of *personal and intelligent assistant* [16]. Good reviews of the concepts of agent and multiagent systems can be found in [2], [10], [19]. We used the notion of agent to specify a multiagent system, named SIGFAD, which purpose is to provide assistance to users in DL. This section presents the beginning of its specification.

When designing and specifying SIGFAD, we used Multiagent Systems Engineering (MaSE), a methodology for developing heterogeneous multiagent systems. MaSE uses a number of graphically based models to describe system goals, behaviours, agent types and agent communication interfaces. MaSE is also associated with a tool, agentTool, which supports the methodology. For a detailed review of MaSE and the associated agentTool environment, see [5], [6], [22].

The first task when designing agent and multiagent systems is to identify goals and sub-goals. In MaSE, this is made during the Capturing Goals step. This step consists of two sub-steps: identifying goals and structuring them in a Goal Hierarchy Diagram. The Goal Hierarchy Diagram of SIGFAD is shown in figure 3.

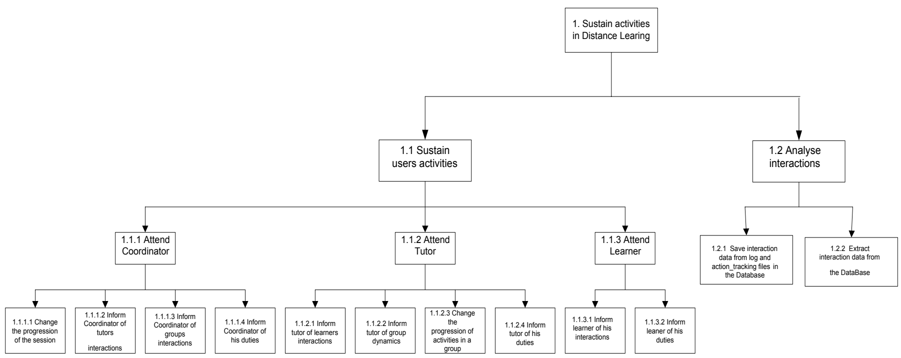

Fig. 3. The Goal Hierarchy Diagram of SIGFAD

At this state of our conception, the main objectives of SIGFAD concern the maintenance tasks of the groups. Even though in the previous sections we insisted on coordinator's and tutors' roles, it is obvious that learner's actions (actions of any group member in general) also influence the durability of the group. The goals of SIGFAD are then threefold, with regards to coordinator, tutors and learners. The coordinator deals with the adjustment of activities (suspend some activities, report others, lengthen or modify time limits). This is done by evaluating tutors activities and group performances. He must also appreciate his own participation to the environment. A tutor has to identify learners' failures in order to dispense weaker learners of certain tasks; he has to appreciate the state of his group, to predict its durability and also to display his own participation. Learners need to know their own participation and what activities have to be achieved in a given period. In SIGFAD a goal called 'analyse interactions' is defined. This goal consists of having access to tracking-actions pages of CEDIL and HTTP server log files, extracting interaction data and saving them in a database, analyse automatically interaction data and display indicators related to the progression and the state of individuals and learning groups.

4. Future Work and Conclusion

Beyond the data analysis presented here, it is necessary to note that the construction of judicious indicators requires to elaborate mathematical models which from the interaction data available, such as those shown in the previous section, will automatically compute variables reflecting the state and the progression of committed groups. We investigate at the moment the analysis social networks which can be usefully applied in DL to build parameters related to the cohesion of the group or the centrality of a given member of the group [18]. Up to now, we exhibited the objectives of SIGFAD. The next stage of our work will consist in specifying our system entirely by achieving all steps of MaSE. These steps include the description of system behaviours, agent types and agent communication interfaces.

The data analysis presented in this paper is certainly partial and we don't argue that agent systems based only on counting messages can support actors in DL. This analysis showed however that automatic tools for displaying users' participation miss in the existing DL platforms. These automatic tools don't exist actually and can't be replaced by current commercial analysis tools of web logs. In order to display on the fly this information, we need tools which will help appreciate the state of the groups, predict their evolution, reorganise them if necessary. These tools, if available, could also permit the appreciation of the quantity and the quality of intra-group interactions and therefore take judicious actions to encourage them. Based on the agent paradigm, we proposed these required functionalities as a multiagent system, named SIGFAD. We presented its Goal Hierarchy Diagram. It is the main output of the Capturing Goal, first step of MaSE, a methodology for designing heterogeneous multiagent systems.

References

1. Baker, M.,. "The roles of models in Artificial Intelligence and Education research : a prospective view". *International Journal of Artificial Intelligence in Education*, 11. (2000) 122-143
2. Bradshaw, J. M. "An Introduction to Software Agents". In *Software Agents*. Ed. J.M. Bradshaw. Menlo Park, Calif. : AAAI Press (1997)
3. Chanier, T. "Créer des communautés d'apprentissage à distance". *Les dossiers de l'Ingénierie Educative, no 36 sur "Les communautés en ligne", octobre*. Centre National de Documentation Pédagogique(CNDP) : Montrouge (2001) 56-59.
4. Charlier, B., Daele, A., Cheffert, J-L., Peeters, R., Lusalusa, S. "Learning collaboratively in a virtual campus : teachers' experiences". ISATT 99. Dublin (1999)
5. Deloach, S., A. "Analysis and Design using MaSE and agentTool*". The 12th Midwest Artificial Intelligence and Cognitive Science Conference* (2001)
6. Deloach, S. A., Wood, M. F., Sparkman, C. H. "Multiagent Systems Engineering". *International Journal on Software Engineering and Knowledge Engineering. World Scientific Publishers. Vol. 11 n° 3* (2001) 231-258.
7. Dillenbourg, P. "What Do You Mean By "Collaborative Learning"". In P. Dillenbourg (Ed.) *Collaborative Learning : Cognitive and Computationnal Approaches*. Amsterdam : Pergamon/Elsevier Science (1999) 1-19.

8. Dillenbourg, P., Baker, M., Blaye, A., O'Malley, C. "The evolution of research on collaborative learning". In E. Spada & P. Reiman (Eds*) Learning in Humans and Machine : Towards an interdisciplinary learning science*. Oxford. Elsevier (1996) 189-211
9. Engelström, Y. *Learning by expanding. An activity-theoretical approach to developmental research*. Orienta-Konsultit Oy, Helsinki (1987)
10. Ferber, J., *Les systèmes multi-agents : vers une intelligence collective*. Intereditions (1997)
11. Fjük, A. *Computer Support for Distributed Collaborative Learning. Exploring a Complex Problem Area*. Dr. Scient. Thesis5. Department of Informatics. University of Oslo (1998)
12. Fjük, A., Nurminen, M. I., Smordal, O. "Taking Articulation Work Seriously – an Activity Theoretical Approach". *TUCS Technical Report N° 120*. 16 p. (1997)
13. Haythornthwaite, C. "Networks of Information Sharing among Computer-Supported Distance Learners". *Proceedings of the Third International Conference on Computer-Supported Collaborative Learning* (1999) 218-222.
14. Jermann, P., Soller, A., Muehlenbrock, M. "From Mirroring to Guiding : A Review of State of the Art Technology for Supporting Collaborative Learning". *Proceedings of the First European Conference on Computer-Supported Collaborative Learning* (2001)
15. Kusunoki, F., Sugimoto, M., Hashizume, H. "A System for Supporting Group Learning that Enhances Interactions". *Proceedings of the Third International Conference on Computer-Supported Collaborative Learning* (1999) 323-327.
16. Maes, P. "Agents that Reduce Work and Information Overload". In *Software Agents*, ed. J.M. Bradshaw. Menlo Park, Californie, AAAI Press (1997)
17. Nurmela, K., Lehtinen, E., Palonen, T. "Evaluating CSCL Log Files by Social Network Analysis". *Proceedings of the Third International Conference on Computer-Supported Collaborative Learning* (1999) 434-442.
18. Reffay, C., Chanier, T. "Social Network Analysis used for modelling collaboration in distance learning groups" (2002) This volume.
19. Weiss, G., Dillenbourg, P. "What is 'multi' in multi-agent learning". In P. Dillenbourg (Ed.) *Collaborative Learning : Cognitive and Computationnal Approaches*. Amsterdam : Pergamon/Elsevier Science (1999)
20. Wertsch, J.V.. "A socio-cultural approach to socially shared cognition". In Resnick, L., Levine, J., Teasley, S. (Eds). *Perspectives on Socially Shared Cognition*. Hyattsville, MD: American Psychological Association (1991) 85-100
21. Whatley, J., Staniford, G., Beer, M., Scown, P. "Intelligent Agents to Support Students Working in Group Online". *Journal of Interactive Learning Research*. vol. 10, N° 3/4. AACE, Charlottesville (1999)
22. Wood, M. F., Deloach, S. A. "An overview of the Multiagents Systems Engineering Methodology". *Proceedings of the First International Workshop on Agent-Oriented Software Engineering*. Ciancarini P., Wooldridge, M. (Eds). Lectures Notes in Computer Science, Vol. 1957. Springer Verlag. Berlin (2000)
23. Wooldridge, M. "Intelligent Agents". In G. Weiss (Ed). *Multiagent Systems*, The MIT Press. (1999)

Acknowledgment

Simuligne was born in a trans-disciplinary research project named ICOGAD, which partners are: the department of Language Learning at The Open University, UK, the Computer Science Laboratory of the Université de Franche-Comté, and the Psychology Laboratory of Université de Nancy 2, France. Special thanks to the French Minister of Research (MRT) and its cognitive science programme (Programme Cognitique 2000) which supports the ICOGAD project.