On Supporting Users' Reflection during Small Groups Synchronous Collaboration

Meletis Margaritis, Nikolaos Avouris, Georgios Kahrimanis

Human-Computer Interaction Group, University of Patras, GR-26500 Rio Patras, Greece
{margaritis, kahrimanis}@ee.upatras.gr, avouris@upatras.gr,

Abstract. During computer-mediated synchronous collaboration there is need for supporting reflection of the partners involved. In this paper we study techniques for determining the state of an evolving collaborative process, while the activity is in progress, making the users aware of this state. For this reason, a State of Collaboration (SoC) indicator has been defined, which is calculated using a combination of machine-learning and statistical techniques. Subsequently a study was performed during which SoC was presented to a number of groups of collaborating partners engaged in problem-solving activities. It was found that this group awareness mechanism influenced in a significant way the behavior of the groups in which it was used. This study has wider implications to the design of groupware and in particular towards gaining an insight into the effect of group awareness mechanisms on computer-mediated collaborative learning.

Keywords: collaborative problem solving, small group interaction, synchronous collaboration, computer supported collaborative learning, interaction analysis.

1 Introduction

Socially inspired theories, supported by the growing development of network and collaborative technology and increased connectivity, have advanced interest on computer-based collaborative problem solving environments. These theories usually influence our considerations on effectiveness of the collaborative problem solving process, as well as the design of the collaboration-support tools involved. While most research and development of collaboration support technology has been directed towards asynchronous collaboration settings, in which usually large numbers of partners are engaged, there is a growing interest in supporting synchronous interaction in which usually small groups of actors are involved (e.g. 2 to 5 partners). In a recent outlook of the Computer-Supported Collaborative Learning field, Stahl [1] suggests that collaborative learning should be primarily studied at the small group unit of analysis where contributions coming from individual interpretive perspectives are interwoven into group cognition. There seem to be some benefits in this kind of group activity when it is computer-mediated. In cases of problem solving in rich and critical conceptual domains it appears that computer supported collaboration could be significantly effective: For activities aiming at conceptual development, communication in written forms combined with communication through graphical representations, seems to be more effective than face to face interaction alone because it requires a more extensive thinking process [4]. The need to externalize one's own thoughts, in a written or a graphical way, could have significant effects, especially when the learning activity implies rich conceptual knowledge that is under development.

In many of these environments, when actors interact in a synchronous collaborative mode, they work in a shared workspace while they communicate using written dialogue often in combination with gestures in the shared workspace that can take the form of sticky notes and tele-pointer operations. Additional affordances of these environments contribute further towards enriched collaborative experience. For instance the substantiation of communication and interaction, which takes the form of a history log, can be used for supporting supervisor' tasks and actors' reflection and self-awareness. In Computer Supported Collaborative Learning activities, the state of evolving knowledge must be continually displayed by the collaboration participants with each other [6], thus history logs provide a treasury of information directly related to knowledge building.

The paper first presents the Synergo environment, and then describes run-time support features for building awareness at group level, illustrating their usage with the example of some vaidation studies.

2 The Synergo Environment

This section presents the Synergo environment from two standpoints. First, in section 2.1 the diagram building tool for supporting collaboration among students is discussed, followed by an introduction to analysis tools that aid students' reflection.

2.1 Synergo's Diagram Building Tool

Synergo (www.synergo.gr) supports synchronous collaborative building of diagrammatic representations by small groups of students.

The environment has been used in Secondary and Higher education settings for teaching computer science and other subjects. The typical client view of Synergo is shown in Fig. 1, which includes a snapshot of a concept mapping activity. Synergo supports building of different kinds of diagrams. It contains libraries for building flowcharts, entity-relationship diagrams, concept maps, data flow diagrams etc. On the left-hand side column of Fig. 1, libraries of primitive objects are shown. The activity is monitored and logfiles are generated and made available for inspection by the users or supervisors. On the right hand side the group coordination panel and the chat window is shown. Different color codes are used to represent the group members in the chat window, while various attempts have been made to represent the state of the peers during interaction. In the following sections some of these group awareness mechanisms are described.

Examples of use of Synergo in authentic educational conditions include collaborative building of algorithm flowcharts by large number of students in a distance learning course of the Hellenic Open University [7], class activities in the frame of an introductory to computing course in a High school [8], collaborative problem solving of distant groups across two Universities [9].

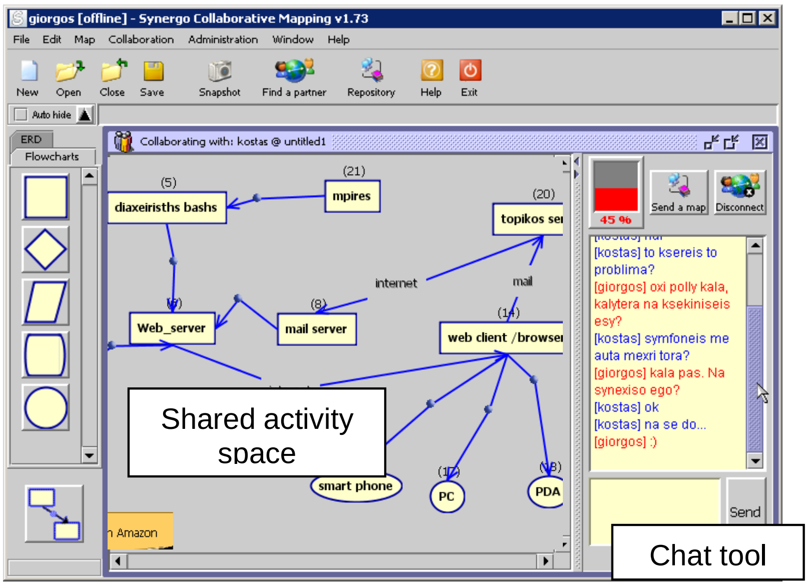

Fig. 1. The Synergo environment: client user interface

2.2 Synergo Analysis Toolkit

An additional feature of Synergo is its inherent support for analysis and supervision of the activity. So a set of analysis and supervision tools is included in the environment, typically in enabled client nodes, called *Teacher nodes*. These are mainly used by the teachers and researchers, while limited versions of the tools may be used in some cases by students as meta-cognitive aids. For instance, the student version of these tools permit playback of the so far activity while problem solving is in progress.

The main functionality of the Analysis tool is the presentation and processing of logfiles which have been produced during group activities. These logfiles contain actions and exchanged messages of group members, in sequential order. An extract of a logfile is shown in Fig. 2. The logfile is based on the same format of the exchanged control and chat messages and is stored in XML form. This file can be viewed, commended and annotated by a researcher using an adequate analysis framework, as discussed by Avouris in [10]. A related functionality of the analysis tool is its capability of post reproduction of the modeling activity, using the logfile, in a step-by-step or continuous way using the playback tool. Further annotation of activity logs through this playback tool can also be done, as discussed in more detail in [11].

The annotated or original history logfiles contain references to the objects involved in the developed activity, by their unique identifier GUID. So if an entity X is used by a logfile L and is not available in the local libraries, the analyst needs to search and download the related entities in order to be able to playback the model and reproduce accurately the activity. In case of missing entities the environment will reproduce them by a default entity with no behavior or iconic representation associated. This decision to disentangle, the logfiles from the often heavy structures associated with model entities is made in order to keep the history logfiles small in size and facilitate their easy exchange and storing. The logfiles can be stored and exchanged in various formats including XML and the tools are based on a database of logfiles, which serve for studies of modeling activities. The format of logfile data is compliant to a proposed model for interoperability of CSCL-related log data described in [2]

1) 00 : 48 : 55	User1	Request Key
2) 00 : 49 : 05	User2	Accept To Give The Key
3) 00 : 49 : 12	User1	Chat "*I asked for the key* "
4) 00 : 49 : 20	User1	Chat "*ok I got i*t"
5) 00 : 49 : 26	User1	Rename Object Ellipse 1 from END USER to END USER #2 (A2412)
6) 00 : 51 : 06	User1	Chat "*Get the key and change all relations with those connected to LANS*"
7) 00 : 52 : 05	User2	Chat "*OK*"
8) 00 : 52 : 08	User2	Request Key
9) 00 : 52 : 13	User1	Accept To Give The Key

Fig. 2. Extract of a history logfile from collaborative problem solving

3. Run Time Support at Group Level: Building Awareness Mechanisms.

One key feature of the presented environment is the support provided at run time to the collaborating partners through a view of the state of evolution of the collaborative activity. Based on the fact that the activity is logged at both the client and the server nodes, some abstract representations of the activity have been defined with the objective to feed them back to the group members in order to increase group awareness and motivate meta-cognitive processes for self-regulation. In this section a mathematical model of collaboration is presented, reflecting the symmetry of participation in dialogue and solution building of the group members. In the following section 4 a new approach, based on data mining of historical data is proposed.

3.1 Modeling Collaborative Activity

In this section the key parameters are described through which collaborative problem solving activity can be modeled in Synergo. In typical problem solving scenarios, dialogue and action are interleaved supporting each-other. So the activity is based on both direct communication acts (e.g. chat messages) and indirect communication through operations in the shared workspace.

This activity can be modeled according to the following four dimensions:

- The time dimension t : (when the action is taking place)
- The actors' dimension: $A = \{A_1, A_2, ..., A_k\}$ (who is acting)
- The objects' dimension: $O = \{O_1, O_2, ..., O_\ell\}$ (the object of action in shared space):
- The typology of events dimension: Ty (what is the type of action).

This latter dimension leads to interpretation of the activity that takes place. It is assumed that there is an existing analytical framework, which defines this typology *Ty*. If r is the finite number of expected event types, then we define a set $Ty = \{T_1, T_2, ..., T_r\}$ as the analytical framework of the study. *Ty* can be defined by the framework user.

Using the above four dimensions we can describe any given activity as a set of discrete non-trivial events produced by the actors, contained in the logfile. These define an ordered set of m events $E = \{E_1, E_2, ..., E_m\}$. Each one of these events is related to meaningful actions of the actors who interact with objects of set O incrementally contributing to the problem solving activity. Each event is defined as a tuple $E_{i_{tAOT}} = (t_t, A_A, [O_O], [T_T])_i$ where $i \in [1, m]$, t the event timestamp, A the actor who performed the action of the specific event, O an optional parameter referring to the object of the specific action and T an optional parameter which interprets the event according to the analysis framework *Ty*.

This is a useful general model for logging collaborative activities. Every time an event is produced by the actors, this is recorded and a history of such events, i.e. an ordered list of Es can be produced, as a result of such an activity. This record of the activity can be further annotated by including mental or cognitive operators, as interpretations of the recorded activity. This model permits further off-line analysis and interpretation of the activity, while quantitative indices of the activity can be easily produced at run time, given that some of the *Ty* annotations can be produced automatically, by the software itself (e.g. actions of insert, delete, chat, etc.) As a result visualizations of the progress of problem solving can be generated [14], as discussed in the next section.

Synergo permits definition of a typology of generated events *Ty*, and automation of the task of categorization of observed events (e.g. insertion, modification, deletion of primitive objects in the workspace and exchange of text messages). The Synergo environment facilitates the *Ty* definition process, by allowing association of kinds of low level software generated events, to event types. So for instance, all the low level events of type "Change of textual description of concepts" in a concept-mapping tool are associated to the "Modification" type of action, as shown in figure 3. Every time an action is recorded, this is automatically categorized according to the analytical typology defined by the user. Various formal models like OCAF [15] suggest interpretation of exchanged messages (written dialogues during collaboration by distance), or recorded oral utterances (during face to face collaboration), in relation to operations towards "objects" of the activity space, using a language for action approach [16], defining a unifying framework for analysis of dialogue and action. However interpretation of dialogue events at run time is not possible, unless the users themselves classify their exchanged messages through a dialogue annotation scheme. However this approach has not been used in our case, as we considered that it imposes a meta-cognitive load to the users and lucks reliability. Instead in the next section quantitative measures of collaboration are described, using just the automatically classified events.

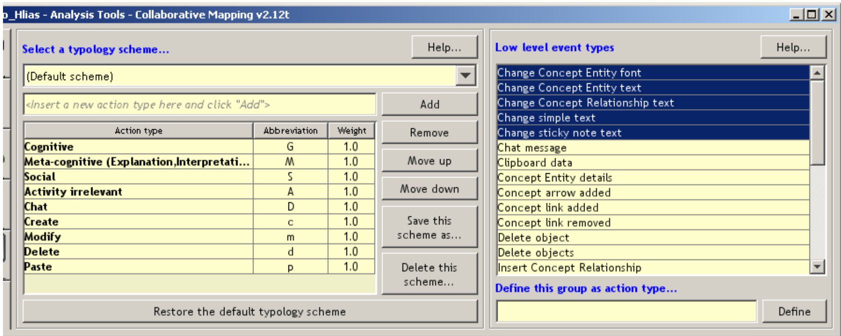

Fig. 3 Definition of an Event Typology scheme Ty: The low level recorded events, generated by the software (right) are related to event types (left).

3.2 Quantitative Indices of Collaboration

Using the model of activity described above, a number of indices, characterizing the state of group activity, have been defined. The objective was to calculate them in order to present them to the group members in a visual form. Some of these indices relate to the density of occurrence of some specific types of event per time interval t_q, e.g. number of exchanged text messages per t_q, number of new objects in the shared space per t_q, etc. These can be calculated at the group level or at the individual partner level.

One other kind of index is related to the degree of symmetry of activity of the group members. This index describes the relative contribution of the group members in a specific type of events.

An example of an empirical general index, called Collaboration Factor is described here. This reflects the symmetry of contribution of actors in the solution, taking into account the relative weights of actors, objects and types of actions.

If we assume that N events of Actor A concern object O, then the contribution of Actor A to object O is measured as:

$AC_{AO} = W(A) \cdot \sum_{i=1}^{N} W(T_i)$ where *W(A)* is the relative weight of actor A and *W(Ti)* is the weight of type *Ti* of event i, that contributed to *O* history.

The history factor *HF* of object *O*, is defined as $HF_O = 1 - \frac{stdev(AC)}{\overline{M}\sqrt{k}}$, where $HF \in [0,1]$ and *M* is the mean value of all actors contributions AC for object O. HF takes value close to 1 when there is symmetrical contribution of all actors in the history of object

O and close to 0 when the object has been discussed and used by small part of the group.
The collaboration factor of object O is defined subsequently, as $CF_O = HF_O \cdot W_O \cdot \frac{L(OE_O)}{m}, \quad CF_O \in [0,1]$

Where W_o the relative weight of object O in the model, $L(OE_O)$ is the length of action events of object O and m the total number of action events in E.
Finally the collaboration factor of the activity CF is defined as the mean value of all components' collaboration factors, including the abstract objects, or objects that were introduced in the solution and later rejected: : $CF = \frac{\sum_{i=1}^{\ell} CF_{O_i}}{\ell}, \quad CF \in [0,1]$

In the formulas of CF defined here, a number of weight factors have been introduced: W(A) is the weight of actor A, Wo is the relative weight of object O and W(T) is the relative weight of type T of event. These factors are defined a priori for a certain kind of activity and reflect the relative importance of the corresponding entities. So for instance in a problem solving activity the learners' contributions are considered more important than those of the tutor, some objects of the problem representation are more important than others (e.g. an entity is more important than an attribute in an Entity-Relation Diagram), while some types of events (e.g. insert a new entity) are more important than others (e.g. modify the description of an existing entity). It should be observed that all dialogue messages were classified as of $T = T_{dialogue}$ without refining further their typology, as already discussed in the previous section.
The Collaboration Factor CF, in addition to the other indices introduced here, like the density of activity of specific type of action events per time unit, can be presented in visual form to the group members in order to support understanding of the collaboration dynamics. An example of use of these indices is included in the following section.

3.3 A Case Study of Calculation and Visualization of Indices of Collaboration

In this section we describe an example of visualization of collaborative activity in the frame of the Synergo tool from a case study. The activity involved building of a *concept map* of an Internet service (an electronic bookshop was the service to be model by the participants in this case) by small groups of students of an undergraduate University course, in the frame of one lab session (45'). We focus on one of these groups made of 4 students in this section. The logfile of the activity of this specific group was studied using Synergo. More details of this study can be found in [11]. First the relative weights of the activity types and the actors were defined. In our case events related to creation and modification of sticky notes are assigned lower weight (0.3), as they are used for administration purposes and were not related to problem solving. The actors were all considered of the same weight W(A)=1, while the objects used (concepts and relations) were also considered of similar importance.

A number of representations were produced using the described model. One possibility was to show the current value of CF at the side of the workspace Also the users can choose to playback the activity and produce in numeric and visual form the evolution of their contribution to the solution and the evolution of the Collaboration Factor CF. This is shown in figure 4(a), and 4(b). In 4(a) the solution is shown with associated history of contribution of various actors to the objects. In 4(b) the evolution of CF with time is shown. This graph provides an indication of the degree of collaboration of the group of the four students as they are building the e-shop concept map. From this graph it seems that while for the first period of the activity the degree of collaboration was high, subsequently the partners became more individualistic, working on parts of the solution, as also shown in the annotated concept map of fig 4(a). Later on towards the end of the session, there is more interaction, the final value was CF=0.073.

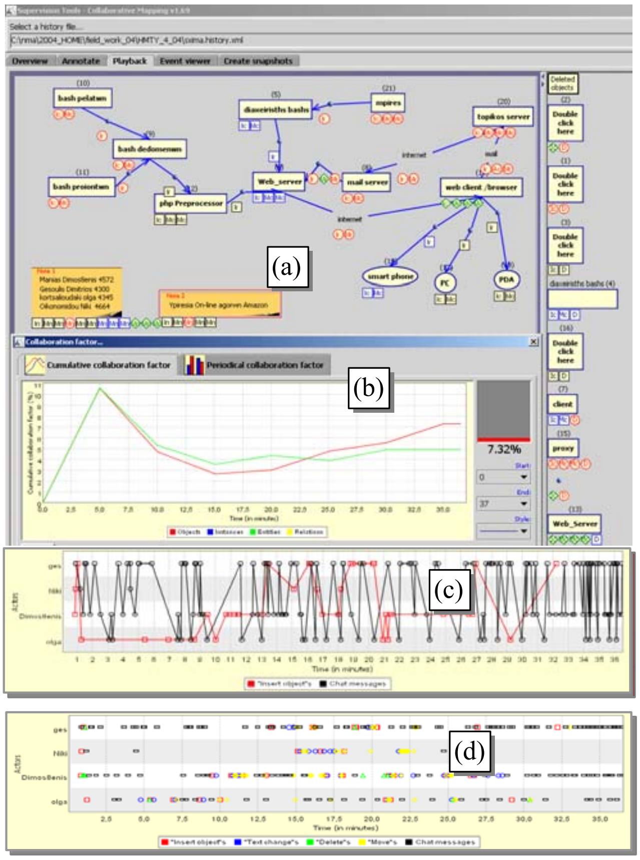

Fig. 4. Visualization of (a) annotated solution (b) Evolution of Collaboration Factor, (c-d) Evolution of Actor activity

Other indices can also be presented, like the density of actors' activity of various types. Also the contribution of each actor in the activity can be visualized. In figure 4(c) and 4(d) the actor contribution of "insert object" events and chat messages is shown. Each line of these diagrams represents one of the four group members. From this picture, it is deduced that the second actor shows relatively low activity.

4. Support for Group Awareness: A Machine Learning Approach.

In addition to the method for calculating and visualizing at run time indices of collaboration, described in the previous section, a new approach that is based on data mining of rich sets of historical logfiles is proposed in this section. The premise of this approach has been that given a rich set of examples of collaborative situations that have been evaluated in terms of the collaborative value of the activity, a module can be trained to be able to classify them accurately enough to be used in the future for classifying other unknown situations. Through this approach it is expected that discourse-related characteristics of the activity that were ignored in the previous case, can be taken into account and a more accurate interpretation of the collaborative progress is made at run time.

First a number of attributes characterizing given segments of previously recorded collaborative activities have to be defined. Then adequate data sets should be selected and effective machine learning algorithms should be used for training classification algorithms, the performance of which subsequently need to be evaluated and to be tested in a typical field study. The described process is a typical data mining approach that has been used often in problems with rich data sets, not solved by analytical or algorithmic approaches [12]. Our problem appears to have these characteristics. So a first attempt to use data sets from previous recorded collaborative problem solving activities in order to train a classifier of collaborative value was made.

We made an assumption that we need to fragment the logfile L of a given activity in consecutive segments L={S1, S2, … Sk}, each one of which containing enough activity in order to be able to establish for the specific segment the quality of collaboration. The fragmentation criterion was established first as a constant time slice t. However we soon discovered that the activity often does not evolve with uniform density with respect to time, so in certain time segments there was enough activity to establish the quality of collaboration factor while in other segments the activity was insufficient. So a second fragmentation criterion was used subsequently: this was the Number of Events (NE) recorded in the events set E. It was decided to set NE=60 which produced a number of segments k for a given logfile. It is obvious that there is a tradeoff between the value of NE and the number of segments k that can be produced from a given logfile of activity, i.e. the higher NE the less number of segments are deduced. In the following, a sensitive analysis was performed in order to establish the effect of the value of NE on the performance of collaboration classifiers.

Given a certain segment Si in which NE events have been included, we need to identify the attributes that would be related to the quality of collaboration. These attributes should be measurable characteristics of the monitored activity, without human intervention, as otherwise deduction of the attribute values would be a tedious process for large data sets.

A set of such attributes was defined and subsequently their predictive power, in terms of the quality of collaboration was tested. The original set of attributes of a given segment of collaborative activity is the following:

- Total number of exchanged dialogue messages (integer)
- Degree of symmetry in participation in dialogue [0..1]
- Number of alternations of speaker in dialogue (integer)
- Average number of words per dialogue message (integer)
- Number of questions in the dialogue - as identified by question mark character (integer)
- Total number of activity events in the shared workspace (integer)
- Degree of symmetry in participation in workspace activity [0..1]
- Number of alternations of actor in workspace activity (integer)
- Degree of symmetry in object modifications in the workspace [0..1]

From these nine (9) attributes, the first five (5) are related to dialogue while the other four (4) are related to activity in the shared workspace. A key attribute is the symmetry of participation of the partners in certain kind of activity, like the dialogue or the modification of the objects of the workspace. A fully symmetrical activity (measured as Symmetry=1) is that in which all partners contribute equally in the activity, while asymmetrical activity (Symmetry=0) is that in which a single partner dominates the activity and collaboration is doubtful.
The effectiveness of this model was tested using logfiles from a number of distinct recorded collaborative activities. The main source of data has been the logfiles of problem solving activities of small groups of students (made of 2 to 3 students) of the Hellenic Open University and of the University of Patras, engaged in building concept maps and flow chart diagrams to given problems, using Synergo. Data from 23 such groups were used. Different segmentation factors NE have been used in these files. The different values of NE and the corresponding different numbers of segments that were created in this study are shown in Table 1.

Table 1 Different fragmentation criteria for segment creation

# of events (NE) per segment	# of segments of activity
60	*306*
80	*234*
100	*188*
200	*99*

For each one of the segments, manual characterization of the quality of collaboration was qualitatively performed by human evaluators. This was defined using three quality measures: low collaboration (1), medium collaboration (2) and high collaboration (3). Subsequently using these data sets, an attribute selection process was performed in order to establish which of the originally proposed attributes contributed more effectively towards prediction of the quality of collaboration. For attribute selection the Correlation based Feature Selection (CFS) technique was used [13].

As with the most of the feature selection techniques, CFS makes use of a heuristic algorithm along with a gain function to validate the effectiveness of feature subsets. This heuristic rule takes into account the usefulness of the independent features to predict the class feature(s) as well as the level of their correlation. Using this technique in the four data sets defined according to different values of NE, shown in Table 1, we established the most effective predictors, shown in Table 2.

Table 2 shows that the attributes that appear in all data sets are: the number of dialogue messages (2), the number of alternations of speaker in dialogue (4), the average message size (5) and the number of actions in the shared workspace (7). From these four attributes the first three are related to the dialogue and just the fourth one is related to the activity in the shared activity space.

Table 2. Attribute selection using CFS

NE=60	NE=80	NE=100	NE=200
(2) num_chat	(2) num_chat	(2) num_chat	(2) num_chat
(3)symmetry_chat		(3)symmetry_chat	
(4) altern_chat	(4) altern_chat	(4) altern_chat	(4) altern_chat
(5) avg_words	(5) avg_words	(5) avg_words	(5) avg_words
(6) num_quest	(6) num_quest	(6) num_quest	
(7) num_draw	(7) num_draw	(7) num_draw	(7) num_draw

A number of alternative classification algorithms were used for building the classifier of the quality of collaboration (Naïve Bayesian Network, Logistic Regression, Bagging, Decision Trees, Nearest Neighbor). Using the open source data mining environment WEKA [12] we trained a number of these classifiers that belong to different categories and use distinct techniques. It is important to note that Synergo facilitates the export of log file data in the form of tab separated document files that are easy to handle by tools such as WEKA.

Evaluation of the performance of the produced classifiers was performed using a 10-fold cross validation technique, separating our data set in training and testing data. In figure 10 the performance of a set of six classifiers in terms of percentage of correctly classified segments is shown for different values of NE. From this figure it is deduced that the best performance was achieved in the case of fragmentation factor NE=60. For this data set all classifiers achieved success rate of over 85%, with best performance by the Logistic Regression classifier who achieved a performance of 87%. As NE increases, the performance of the classifiers deteriorates with the case of NE=200, as worse case in which the average performance of the six classifiers was just over 80%. If we take in consideration the fact that an additional disadvantage of high values of NE is that it inflicts long waiting times at run time, as a large number of events should be accumulated before a new value of the factor is calculated, the conclusion of this part of the study is that the most effective values of the fragmentation factor NE should be around the lowest value NE=60, while experimentation with even lower values of NE made the task more difficult and increased the number of indecisive segments since the number of events was too low for a clear verdict on collaboration by the human expert.

As a conclusion of this phase of experimentation with building a mechanism for evaluating the quality of collaboration in the frame of our framework at run time, in order to use it as a group awareness mechanism, we discovered that this machine learning approach was effective since the trained classifiers were capable, with accuracy close to 90%, to classify segments of activity in a qualitative way. It should however be observed that this second approach produced a qualitative index of group collaboration (high, medium, low) contrary to the statistical approach that produced a more accurate numerical value.

A final attempt was made to use a combination of the two approaches discussed here, and in section 3. So we built a hybrid collaboration awareness mechanism as a linear combination of normalized values of the collaboration factor (CF) discussed in section 3 and the quality of collaboration factor discussed in section 4. The result was a measure of the state of collaboration (SoC) which was implemented and used in the frame of a case study discussed in the final section of the paper.

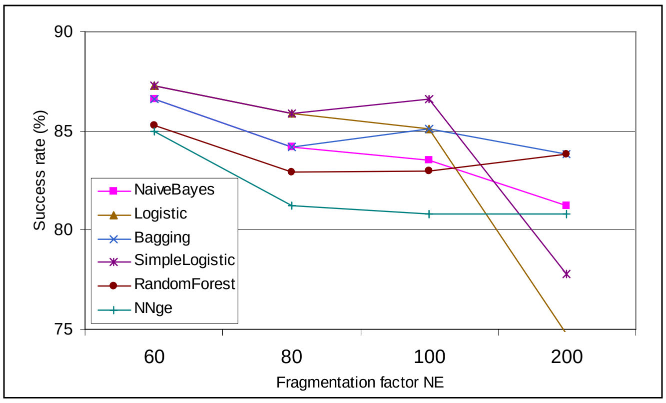

Fig. 5. Performance for different values of the fragmentation factor.

5 Evaluation study

An evaluation study of the developed group-awareness mechanism was performed next. The objective of the study was to establish the effect of this mechanism to small groups of users of Synergo. In the study thirty-three (33) students of the Electrical & Computer Engineering Department of our University participated. In the context of a laboratory session of the Human-Computer Interaction course they were asked to evaluate collaboratively in small groups the usability of a web-based accommodation booking service of a major international conference. Subsequently they were asked to build a state transition diagram of the typical user interaction with the system, in which to associate usability-related comments. The group members interacted exclusively through the Synergo chat tool and the Synergo shared activity space in which they built the requested diagram. The students were assigned to 11 groups made of 3 students each. Six (6) of these groups were provided with the group awareness collaboration mechanism. The other five (5) groups did not have that facility.

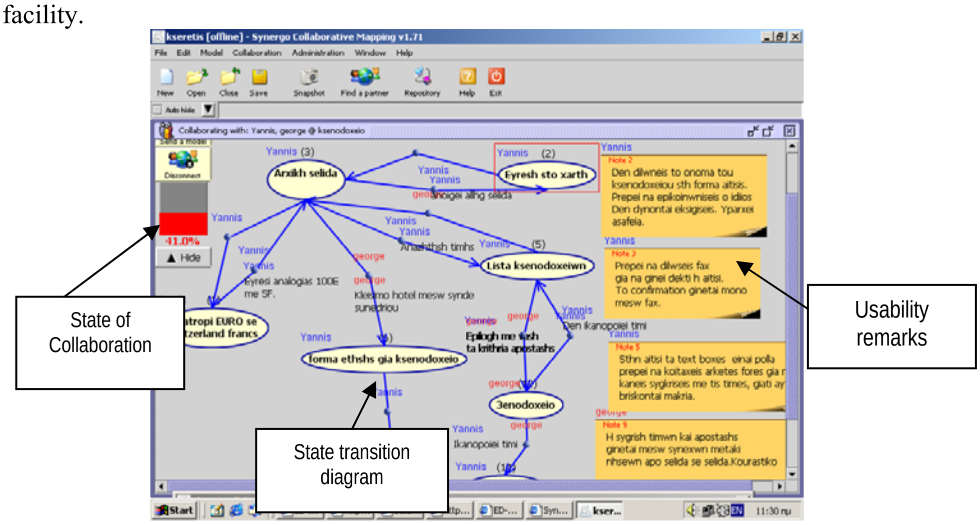

Fig. 6. Typical user workstation during collaborative problem solving.

A comparative qualitative and quantitative evaluation of group interaction of these two sets of groups was preformed. We measured how symmetrical the interaction of the group members were in the two sets. The overall measure combined the degree of symmetry of dialogue events and actions in the shared activity space. This measure took the values shown in Table 3 for the groups of the study.

An extract of a typical solution produced by one of the groups is shown in figure 6.

It was found that there is significant difference between the mean value of the two sets (t-test: $p=0,0423<0,05$). The mean values of collaboration symmetry are 45% for set A and 36% for set B, with standard deviation 0.053 and 0.075 correspondingly. So the activity of the groups of set A who were aware of the collaboration state through the developed group awareness mechanism was more symmetrical.

In addition, by examining more closely the dialogues in groups of set A it was found that in four (4) out of the six (6) groups there was an explicit discussion about the group awareness mechanism.

A side-effect of the group awareness collaboration mechanism was that in some occasions the partners attempted even to affect explicitly the value of this factor, as in the extract of figure 7. However this kind of dialogue events accounted for less than 5% of the overall exchanged messages.

Overall the dialogues were focused in the task and the participation of the partners in the groups of set A was more active and focused than those of set B. The discussion about the group awareness mechanism took place at the beginning and the end of the session in all four occasions and it did not affect the problem solving task. In groups of set B the participation of the partners in the activity was less symmetrical, due often to the existence of partners of limited contribution to the activity. This is an incident observed often in groups with more than two partners in synchronous groupware.

Table 3: Group awareness State of Collaboration factor.

Set A: Groups **with** group-awareness mechanism	**Set B:** Groups **without** group awareness mechanism
38 %	36 %
41 %	47 %
48 %	34 %
49 %	37 %
52 %	26 %
43 %	

```
[U1] Let us start talking about ourselves in order to
increase the bar to 100%
[U2] U3, you should talk!
[U3] hhhm what to say :)
[U1] You see it went up to 42% by just doing that
[U3] hey hey hey
[U3] How do we start drawing?
[U2] As long as you U1 talk, it goes down...
[U1] OK I will shut up then..
```

Fig 7: Example of dialogue extract about the collaboration factor.

7 Conclusions

Building group reflection mechanisms for groupware systems, like the State of Collaboration (SoC) factor for the Synergo environment discussed in this paper, presents difficulties, since these factors are calculated from many diverse indices who are produced by the dispersed activity of the collaborating community. Use of just statistical aggregate measures is an approach that has been used effectively in the past, however there is an increasing need to capture the semantics of the evolving collaborative process in order to feed them back to the group of the partners providing them with more realistic group awareness view. Use of machine learning techniques for this purpose presents great advantages, since these techniques often require much less data and use less processing power than the statistical techniques, while they are more flexible in providing qualitative measures of the state of collaboration. However in order for such techniques to be proven effective, a tedious modeling phase should proceed followed by a careful training phase of the algorithms. In addition, rich data sets which depict many examples of collaborative or antagonistic situations should be used during the data mining process.

An overall conclusion of the study is that group awareness seems to play a significant role in the group activity, as it is easy to interpret, not requiring high cognitive load and focusing ability of the partners concerned, as is the case with individual partners' awareness mechanisms. Through a single graphic measure or a plot represents vividly the state of the group. The result in our case study was this mechanism to cause higher degree of involvement of the individual partners and lead to improved collaboration.

Acknowledgments. This research has been funded by the Hellenic Ministry of Education, program Herakleitos on Methods & Tools for CSCL and the European Research Team CAViCoLA in the Kaleidoscope NoE (Contract No. 507838).

References

1. Stahl, G. Rediscovering CSCL. In Koschmann, T., Hall, R., Miyake, N. (Eds.), CSCL2: Carrying Forward the Conversation, Lawrence Erlbaum Associates, Hillsdale, NJ (2001)
2. Kahrimanis G., A. Papasalouros, N. Avouris, S. Retalis, A Model for Interoperability in Computer Supported Collaborative Learning, Proc. ICALT 2006 - 6th IEEE International Conference on Advanced Learning Technologies. July 5-7, 2006 – Kerkrade , Netherlands
3. Stahl, G. Group cognition in computer-assisted collaborative learning, Journal of Computer Assisted Learning (2005) 79-90.
4. Dewan, P. Architectures for Collaborative Applications, Chapter 7 in Computer Supported Cooperative Work, Edited by Beaudouin-Lafon, pp.. 169-193, 1999 JohnWiley & Sons Ltd..
5. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. 3rd edn. Springer-Verlag, Berlin Heidelberg New York (1996)
6. Dimitracopoulou, A., Komis, V., Design principles for the support of modelling and collaboration in a technology based learning environment", Int. J. Continuing Engineering Education and Lifelong Learning, 15, (1/2), (2005), 30-55.
7. Xenos, M., Avouris, N., Komis, V., Stavrinoudis D, Margaritis, M. Synchronous Collaboration in Distance Education: A Case Study on a CS Course, in Proc. IEEE ICALT 2004, Joensuu, FI, (2004)
8. Voyiatzaki, E., Christakoudis, C., Margaritis, M., Avouris, N., Algorithms Teaching in Secondary Education: A collaborative Approach, in Proc. ED-Media 2004, pp. 2781-2789, Lugano, June 2004.
9. Harrer A., Kahrimanis G., Zeini S., Bollen L., Avouris N., (2006). Is there a way to e-Bologna? Cross-National Collaborative Activities in University Courses, 1st European Conference on Technology Enhanced Learning, Crete, Greece, October 1-4, 2006
10. Avouris, N., Margaritis, M., Komis, V. Modelling interaction during small-groups synchronous problem-solving activities: The Synergo approach, in Proc. ITS 2004, Maceio.
11. Avouris, N., Komis, V., Fiotakis, G., Margaritis, M., Tselios, N. "Tools for Interaction and Collaboration Analysis of learning activities", Proc. CBLIS 2003, Nicosia, Cyprus.
12. Witten, I. H., Frank, E. "Data Mining: Practical Machine-Learning Tools", Academic Press, San Diego, CA, (2000).
13. Hall, M. A., "Correlation-based Feature Selection for Machine Learning". Ph.D. diss. Dept. of Computer Science, Waikato Univ., (1998).
14. Margaritis, M., Avouris, N., Komis, V. Methods and Tools for representation of Collaborative Learning activities. Proc. ETPE 2004, September 2004, Athens.
15. Avouris, N., Dimitracopoulou, A., Komis, V. On analysis of collaborative problem solving: An object-oriented approach, Computers in Human Behavior, 19, (2), 2003, pp. 147-167.
16. Winograd, T., A Language/Action Perspective on the Design of Cooperative Work, Human-Computer Interaction 3:1 (1987-88), 3-30.