Enabling Automatic Just-in-time Evaluation of In-class Discussions in On-line Collaborative Learning Practices

Santi Caballé[1], Àgata Lapedriza[1], David Masip[1], Fatos Xhafa[2], Ajith Abraham[3]
[1] Open University of Catalonia, Department of Computer Science,
Multimedia and Telecommunication Rambla de Poblenou, 156,
08018 Barcelona, Spain
{scaballe, alapedriza, dmasip}@uoc.edu

[2] Polytechnic University of Catalonia,
Department of Languages and Informatic Systems
Jordi Girona, 1-3, 08034 Barcelona, Spain
fatos@lsi.upc.edu

[3] Center of Excellence for Quantifiable Quality of Service,
Norwegian University of Science and Technology
O.S. Bragstads plass 2E7491 Trondheim, Norway
ajith.abraham@ieee.org

ABSTRACT: *Learning by discussion when applied to on-line collaborative learning settings can provide significant benefits for students in education in general. Indeed, the discussion process plays an important social task in collaborative learning practices. Participants can discuss about the activity being performed, collaborate with each other through the exchange of ideas that may arise, propose new resolution mechanisms, justify and refine their own contributions, and as a result, acquire new knowledge. Considering these benefits, current educational organizations incorporate on-line discussions into web-based courses as part of the very rationale of their pedagogical models. However, in-class collaborative assignments are usually greatly participated and contributed, which makes the monitoring and assessment tasks by tutors and moderators time-consuming, tedious and error-prone. Specially hard if not impossible by human tutors is to manually deal with the sequences of hundreds of contributions making up the discussion threads and the relations between these contributions. Consequently, tutoring tasks during on-line discussions usually restrict to offer evaluation results of the contributing effort and quality after the collaborative learning activity takes place and thus neglect the essential issue of constantly considering the process of knowledge building while it is still being performed. In this paper, we propose a multidimensional model based on data analysis from online collaborative discussion interactions that provides a first step towards an automatic evaluation in just-in-time fashion. The context of this study is a real on-line discussion experience that took place at the Open University of Catalonia.*

1. Introduction

In online collaborative learning environments (Dillenbourg, 1999) the discussion process forms an important social task, where participants can think about the activity being performed, collaborate with each other through the exchange of ideas arising, propose new resolution mechanisms, and justify and refine their own contributions and thus acquire new knowledge. In particular, a complete discussion and reasoning process is based on three types of generic contributions, namely specification, elaboration and consensus (Salomon, 1993). Specification occurs during the initial stage of the process carried out by the tutor or group coordinator who contributes by defining the group activity and its objectives (i.e. statement of the problem) and the way to structure it in sub-activities. Elaboration refers to the contributions of participants (mostly students) in which a proposal, idea or plan to reach a solution is presented. The other participants can elaborate on this proposal through different types of participation such as questions, comments, explanations and agree/disagree statements. Finally, when a correct proposal of solution is achieved, the consensus contributions take part in its approval (this includes different consensus models such as voting); when a solution is accepted the discussion terminates.

Indeed, learning by discussion when applied to collaborative learning scenarios can provide significant benefits for students in collaborative learning, and in education in general. This view is especially relevant in the context of the Bologna Process and the current shifting from a traditional educational paradigm (centered on the figure of a masterful instructor) to an emergent educational paradigm which considers students as active and central actors in their learning process. In this new paradigm students learn, with the help of instructors, technology and other students, what they will potentially need in order to develop their future academic or professional activities (Kulesza & Reinalda, 2006). Considering these benefits, current educational organizations incorporate in-class online student discussions into web-based courses as part of the very rationale of their pedagogical models.

Following this increasing interest, current online collaborative learning applications are incorporating advanced interactive support to on-line discussions resulting in the generation of large

amounts of interaction data, which include complex issues of the collaborative work and learning process (e.g., group well-being (McGrath, 1991) as well as self, peer and group activity evaluation (Daradoumis, et al., 2006)). As a consequence, manual monitoring and evaluation of large on-line discussion processes, typically carried out by tutors and moderators, become tedious, error-prone, and highly unreliable. Moreover, since the evaluation process is done after the completion of the learning activity, it has less impact on it (McDonald, 2003). Indeed, the lack of constantly feeding back immediate evaluation from the tutor on the dynamics and performance of the collaborative activity may negatively impact on participant's motivation, emotional state and problem-solving abilities, and as a result diminish the performance and acquisition of knowledge (Zumback et al., 2003).

Intensive and successful research from the interaction analysis field has been achieved over the last years to facilitate the management by computers of the large amounts of interaction data from online discussions. Current efforts (De Weber et al, 2006; Soller, 2001; Pena-Shaff & Nicholls, 2004; Strijbos et al., 2006; Schire, 2006) aim to alleviate manual procedures while considering relevant aspects of the collaboration, such as how all participants are actually performing during the discussion and the dynamics of each participant with respect to the group. To this end, two levels of interaction analysis are considered, quantitative and qualitative level. Quantitative indicators measure the participants' performance and dynamics (e.g., number of contributions written and read by each participant) as relevant information to model the group functioning and task performance (Daradoumis, et al., 2006). Qualitative information has been also considered valuable to complete the labored task of interaction analysis and evaluation of contributions (Strijbos et al., 2006; Schire, 2006).

In previous research (Caballé et al., 2008), we reported on real experiences of learn-by-discussion fully centred in students and were supported by means of an ad hoc sophisticated knowledge-based web-based discussion bulletin board. In these experiences the lecturer was left as a supportive actor who no longer interfered with the collaboration at his convenience but provided adequate scaffold instead in order to enhance and improve knowledge building as a constructive process among learners. The research goal included the provision of relevant knowledge about the collaboration based on information captured from the actions performed by participants during the collaborative process. The ultimate goal was to extract relevant knowledge in order to provide learners and tutors with efficient awareness, feedback as regards learners' performance and collaboration.

In this paper, we take these entire approaches one step further and also provide an innovative process for just-in-time monitoring and assessment of online discussions by means of interaction data analysis techniques. This process is based on those elements that contribute to the understanding of the nature of the collaborative interactions, such as the students' passivity, proactivity, reactivity as well as the effectiveness and impact of their contributions to the overall goal of the discussion. The knowledge extracted from the interaction analysis is then incorporated into an ad hoc discussion system that implements many of the approaches described so far and the first results drawn from the real collaborative learning show very promising benefits for students and tutors in our real learning context of Open University of Catalonia (UOC)[1] and in education in general.

Finally, a further innovation of this process is to incorporate a machine-learning approach to automatically qualify the exchange type of interactions. The idea is to learn the relationship between a set of discussion contributions types and the perceived intention of their authors. From the literature, the automatic evaluation of online discussion contributions has been little investigated, to the best of our knowledge. Quite a few research studies, such as Weimer, et al., 2007, and especially McDonald, 2003 and Zumbach et al., 2003, show a first step towards this direction by combining several quantitative analysis and modeling the threaded discussions. Some relevant references in this field, (Kim et al., 2006; Kim et al., 2007), propose several techniques for assessing discussion contributions automatically by means of quantitative indicators (such as total of posts and post length) and mining discussion text. The latter is achieved by modeling discussion threads as a sequence of speech acts and using relational dialogue rules to identify dependencies among the messages. However, since the evaluation process is done after the completion of the learning activity, it has less impact on the learning process since there exist no opportunities for timely real-time scaffolding at the moment when it is needed. On the other hand, (Weimer et al., 2007) propose a machine learning approach based on a small set of intrinsic text features, such as syntactic, lexical, and quantitative, to automatically rate posts in a binary fashion (i.e., good/bad). Although this is an innovative approach it has not been sufficiently exploited so far.

The paper is organized as follows. We propose in Section 2 a model for collecting and managing interaction in a discussion process based on both speech act analysis (Martin, 1992; Clark & Schaefer, 1989), and a machine learning approach (Witten & Frank, 2005) to collect reliable data. The information captured by this model is then turned in Section 3 into a multidimensional framework of knowledge used to monitor and assess participation behavior, knowledge building and performance. Section 4 provides analytical data discussion based on the results of an experience carried out at the UOC. The paper concludes in Section 5 summarizing the main ideas and outlining ongoing and future work.

2. Aims and theoretical background

The model proposed in this paper is based on the integration of several models and methods: the Negotiation Linguistic Exchange Model (Martin, 1992); a model of Discourse Contributions (Clark & Schaefer, 1989); the types of learning actions underlying a participant turn (Self, 1994), and a machine-learning approach (Witten & Frank, 2005).

In particular, this section examines how the building and distribution of knowledge is manifested in the context of student-student interaction and how it can be studied in a virtual learning environment. This involves the definition of appropriate collaborative learning situations and the distinction of two levels of student interaction, the discourse and the action level. At the discourse level, the essential element is the interaction among peers (participants need to interact with each other to plan an activity, distribute tasks, explain, clarify, give information and opinions, elicit information, evaluate and contribute to the resolution of problematic issues, and so on). At the action level, task objects (e.g., documents, graphics) are created and manipulated. This approach focuses more at the analysis of the discourse level by seeing discourse as a medium and means through which the building and distribution of cognition is effected.

The structure of a long interaction is constructed cooperatively by using the exchange as the basic unit for communicating

knowledge. Following Martin, 1992, we consider three general exchange structure categories: *give-information* exchange, *elicit-information* exchange and *raise-an-issue* exchange, which consist of different types of moves and describe a generic discourse goal. More specifically, the goal of the actor who initiates the *give-information* exchange is to inform his/her partners about a certain situation with the aim to change the partners' mental states. Informing includes moves that explain, give an opinion, describe or remind a situation in different ways. The actor goal of the second exchange is to elicit the partners' state of mind (knowledge, beliefs, attitude, desire or abilities) of a situation, in which the actor is not aware or certain about. The actor goal of the third exchange is to raise an issue (a problem or question) to be resolved by the participants, which causes to explore their state of mind (knowledge, beliefs, etc.).

According to Martin, 1992, there is a move that constitutes the "obligatory move" of the exchange, since it either carries or indicates completion of the discourse goal for which the exchange is initiated. According to Clark & Schaefer, 1989, each move is seen as a contribution to discourse. This means that in a cooperative conversation, contributions are regarded as collective acts performed by the participants working together, resulting in units of conversation - typically turns (moves) - that aim to make a success of the discourse they compose. Yet, not all moves contribute in the same way toward the successful completion of the exchange. According to (Self, 1994), some moves have a pure contributing function toward the realization of the obligatory move of the exchange. In fact, without the presence of those moves, the obligatory move cannot be realized; thus, those moves really contribute toward the realization of the obligatory move. Consequently, it is stated that successful realization of the obligatory move conveys evidence of (initial) success of the exchange. In contrast, other moves have a rather supporting function (provide evidence of support) toward the definite completion of the obligatory move and consequently of the exchange. This is the case of the follow-up moves of the three exchanges. Supporting moves are optional, so they may not be realized. In such a case, they convey an implicit support toward the obligatory move, that is, toward the definitive completion of the exchange.

In general, the three types of exchanges represent standard discourse structures for handling information and suggest a certain type of knowledge building, as a result of giving and eliciting information or working out a solution on an issue set up. These discursive structures enable the participants to take turns, share information, exchange views, monitor the work done and plan ahead. Most importantly, they provide a means to represent and operationalize the cognitive product at individual level, that is, the way the reasoning process is distributed over the participants as it is shared in a collaborative discourse.

Consequently, interaction analysis takes into account both the way the interaction is structured and the types of contributions which are explicitly defined and expressed (see Table 1). For instance, in a set-up-an-issue exchange, a solution move may not be sufficiently complete and thus has to be further elaborated, corrected or extended. To that end, another participant has the option to provide an extend-solution move which completes the initial solution. A complete set of categories or types of contributions and the context of moves where they are found is presented in Table 1. The analysis of these interactions yields very useful conclusions on aspects such as individual and group working, dynamics, performance and success, which

Exchange moves	Exchange categories
support	Greeting
	Encouragement
	Motivation
request	REQUEST-Information
	REQUEST -Elaboration
	REQUEST -Clarification
	REQUEST -Justification
	REQUEST -Opinion
	REQUEST –Illustration
inform	INFORM-Extend
	INFORM-Lead
	INFORM-Suggest
	INFORM-Elaboration
	INFORM-Explain/Clarification
	INFORM-Justify
	INFORM-State
	INFORM-Agree
	INFORM-Disagree
set-up-an-issue	PROBLEM-Statement
provide-solution	PROBLEM-Solution
consent-solution	PROBLEM-Extend solution
	PROBLEM-Assent solution

Table 1. List of the exchange moves and exchange categories to classify a discussion contribution.

allows for obtaining a global account of the progress of the individual and group work and thus to assess whole learning process much better.

To satisfy course evaluation requirements, discourse contributions also need to be evaluated as effectively as possible in terms of quality and usefulness. Evaluation of hundreds of contributions and the relations among them in a multi-member discussion can be a tedious task for tutors and should be adequately supported. Moreover, self and peer evaluation should be also encouraged and facilitated by intuitive means. To this end, following Weimer et al., 2007, in order to automatically qualify the exchange type of interactions, a machine-learning approach is proposed. To this end, state-of-the-art classification algorithms can be used so as to learn the relation between a set of types of interaction and the perceived intention of the authors of these interactions. Similarly, peer manual evaluation could be also replaced with an automatic rating system.

To sum up, a complete dialogue model of asynchronous discourse is to be provided, which is capable of capturing, analyzing and evaluating both the process and the result of the building and distribution of knowledge. This model should be mainly defined in terms of types and structure of student-student interaction.

Finally, the system requires the participant to commit certain action to indicate s/he has read a certain contribution, such as send a reply and assent the contribution. The aim is both to provide reliable indicators on the number of contributions read and to promote the discussion's dynamics by increasing the users' interaction with the system.

3. Research methodology

This section presents a methodological approach to validate the previous conceptual model for semi-automatic evaluation of the discussion process. To this end, first, a multi-experiment carried out at the Open University of Catalonia is described. Then, a new interactive discussion tool that was used to collect the experimental data is presented along with the description of a set of indicators we incorporated to measure and ultimately analyze participation behavior, knowledge building, and performance during the discussion.

3.1 Experiences using real learning context

The real context of this study is the virtual learning environment of the Open University of Catalonia (UOC). Given the added value of asynchronous discussion groups, the UOC have incorporated on-line discussions as one of the pillars of its pedagogical model. To this end, great efforts are being made to develop adequate on-line tools to support the essential aspects of the discussion process, which include students' monitoring and evaluation.

Six experiences in all took place at the UOC over the last two academic terms. A total of 730 graduate and undergraduate students from three courses in Computer Science were involved directly or indirectly forming the experiment sample. For each experience, students were equally distributed into two classrooms and participated in the experience with the same rules, at the same time and during the same time (about a fortnight). Students from one classroom were required to use the well-known asynchronous threaded discussion forum offered by the UOC virtual campus while the other group of students used a new discussion tool, which incorporated our model of interaction management. This discussion tool is presented next.

3.2. Data collection through an effective structured discussion forum

All data from these experiences were collected by means of a prototype of an ad hoc web-based structured collaborative learning system, called Discussion Forum (DF). This tool incorporates our conceptual model for interaction management, which gives new opportunities to learn by discussion (Caballé, 2008). For the sake of understanding how the collected data was generated, certain key design aspects of this tool are described here.

3.2.1 Collection of post tagging, assent, and rating

The design of the DF includes certain thematic annotation cards based on the general exchange types identified in Section 2, namely give-information, elicit-information and raise-an-issue. Six exchange moves and quite a few low-level categories (see Table 1 for a complete list) have been identified to qualify each exchange move in the discussion processes occurring at our university though they are not conclusive since more experimentation process has to be undertaken.

In order to avoid unnecessary choice, each context of the discussion process determines a precise and short list of just those categories that are possible in a certain point of the discussion process (e.g., in replying any kind of request, just the cards involving the provision of information are provided to classify the reply). This makes the choice of the appropriate tag shorter and easier (see Figure 1). In addition, the tutor is to examine and assess the quality of all contributions based partially on the tags used by students to categorize them. As a result, students are aware of the potential repercussions of

Figure 1. A list of tags to qualify a contribution

tagging posts incorrectly in order to optimize the evaluation instead of reflecting the true meaning of their posts.

Consequently, DF's users are urged to correctly qualify their contributions before sending a new or reply post. Contributions may also be assented and also evaluated by both the tutor and other participants in terms of content quality and the utility in their progress in the discussion (see Figure 2).

3.2.2 Collecting reliable data

A further innovation for the reliable collection of data is to automate the manual post tagging (see Figure 1) so as to both minimize error-prone of post tagging and release students of unnecessary choice.

To this end, from the six experiences run in the form in-class assignments consisting of online discussions of certain pedagogical issues, we collected as many as 2497 posts. Their authors had already tagged all these posts by using one of the 6 exchange moves presented in Table 1 (i.e., *support*, *set-up-an-issue*, *request*, *inform*, *provide-solution*, *consent-solution*). We then removed 220, which were used just for training purposes. The rest, 2277, were checked and their tags were changed if found wrong according to the real intention of the contribution and thus obtaining a fairly amount of correctly tagged posts. Finally, all posts were classified into the 6 mentioned groups of exchange moves. The distribution was the following:
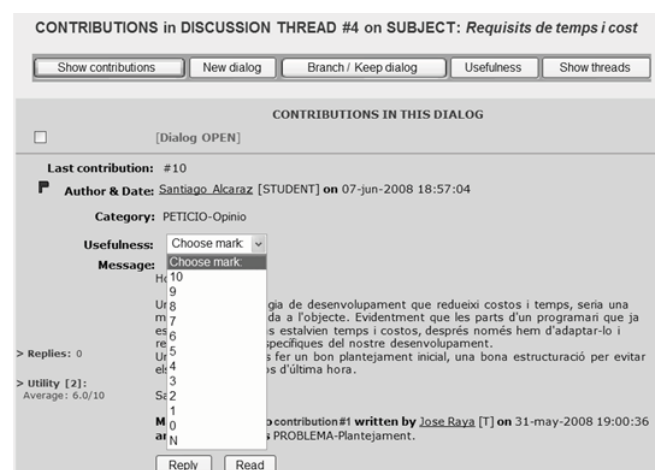

Figure 2. Post rating and assent

support (5.5%); set-up-an-issue (8.2%); request (25.1%); inform (56.9%); provide-solution (3.8%); consent-solution (0.5%).

Automatic posts classification

Using this large data set, we explore the possibility of automatically categorize the posts on the 6 different exchange moves described. Although the design of optimal classifiers is out of the scope of this paper, the proposed methodology would take benefit from a first categorization approach.

Following the similar work of (Weimer et al., 2007), for each post, we constructed a feature vector using the following methodology: (i) First a list with the total words present in all the posts is generated. (ii) From this list, we removed the words that appear only once, in order to mitigate the effects of orthographic errors. (iii) Using the resulting words, we compute the frequency count of each word on each text, obtaining a 16532 dimensional feature vector for each post.

The resulting data lies in a high dimensional subspace, hindering the posterior estimation of the classifiers parameters. In order to mitigate this drawback, a previous dimensionality reduction step has been applied. We used the Principal Component Analysis algorithm (Duda et al., 2008) to extract the first 300 components, which account for the 94% of the data variance.

Using the final 2277 300-dimensional feature vectors, we applied a state-of.the-art SVM classification algorithm to the obtained posts. Briefly, the SVM algorithm (Burges, 1998) learns a binary classifier (two possible classes, a positive one and a negative one) from the training data. This classifier consists of a separating hyperplane that maximises the classification margin. Thus, a new post x is classified in positive or negative class, according the following decision rule type

$$f(x) = \text{sgn}\left(\sum_{i=1}^{N} w_i K(x, x_i) + w_0 \right)$$

where $(x_1,...,x_N)$ are the training samples, $(w_0,...,w_N)$ are the parameters of the classifier, and K denotes a Kernel function (or the dot product in the linear case). In our problem of posts classification, notice that the amount of data available is large and sparse (99.57% of scarcity), being the most part of the frequency counts 0. In this scenario, we opted for a non-linear version of the SVM classifier, based on the application of Radial Basis Function kernels (RBF-SVM) (Müller et al., 2001),

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2} \right)$$

σ being a parameter that will determine the influence area that has the SVM over the data space. The extension of the SVM algorithm to multi-class problems (more than 2 classes) can be carried out by the one-versus-all strategy (Rifkin & Klautau, 2004).

In order to validate the automatic classification procedure, the following protocol has been followed: the total amount of data has been randomly split in a training (90% of the data) and a testing set (the remaining 10%). The amount of Data from the different classes has been balanced in the partitions. We used the training set to learn the RBF-SVM classifier, using a portion of this set to find the optimal sigma and C parameters.

The experimental protocol has been repeated 20 times, and the average accuracy obtained is 61.29% for the 6-class problem (±2.08% confidence interval at 95%). This preliminary result constitutes a promising initial attempt to automatic classification of posts from their content. Nevertheless, we plan as future work to improve this part of the methodology by exploring other classification strategies and data normalization techniques.

3.3 A multidimensional framework to evaluate participation behavior, knowledge building and performance

Based on the previous assumptions, all contributions are recorded in the DF as exchange moves, which are later on analyzed and presented as knowledge to participants either in just-in-time fashion (to guide directly students during the learning activity) or after the task is over (in order to understand the collaborative process). Finally, relevant feedback is provided to the discussants and tutors based on the data collected and the following methodology that identifies and measures relevant dimensions of the discussion process (Figure 3).

Participation behavior (activity) indicators are distinguished into proactive, reactive and supportive. Participants are *proactive* when they take the initiative to open a new exchange of the type give-information, or raise-an-issue. Participants are *reactive* when they reply to moves such as elicit-information, set-up-an issue/problem, or provide-solution. Participants are *supportive* if they give their assent to previous contributions. In that case, a supporting value is defined which is assigned a default numerical value 1 which means that the move fully supports and recognizes the value, contribution and effectiveness of a previous move it refers to. If several supporting moves refer to a particular move M, it implies a broader consensus about the impact of M, which increases M's impact value to 1.

Figure 3. Monitoring information provided to the tutor

Passive participants are considered those who just read others' contributions, as well as the ones who also evaluate the usefulness of these contributions. Passivity becomes an essential indicator for the discussion process' dynamics as it identifies certain important profiles of the participant, such as arrogance (participant who just contributes but does not read the contributions of others) and also promotes reactive attitudes and social grounding skills by engaging the participant in the collaborative process (Dillenbourg, 1999).

The *impact* value is assigned an initial (default) numerical value between 0 and 1 which is modified (increased or decreased) according to the impact (number of reactions received) that the move M has on the dialogue and on the achievement of the current discourse goal and task. If the reaction is positive (the move M is being assented), then M receives a positive one (+1) point. If the reaction is negative (M is not assented) then it receives a negative 0.5 points. The points received by a reaction move depends on the type of learning action underlying the move and take on the default value of the move's impact value. The final value is obtained by the mean value of all moves involved in move M.

The *effectiveness* value of a move is calculated by the mean value of the number of assent moves received. An assent move M is identified and recorded after a participant receives M and consents it. Note that only give-information and raise-an-issue exchange acts can be assented. A negative assent requires a reply move on M to provide further information to reason why M has not been assented, which generates another move in the current discourse.

Finally, tutor and peer *assessment* indicators are to evaluate both the quality of the contribution's content by the lecturer monitoring the discussion process and the usefulness of the contribution by the student participating in the discussion. Both indicators are on the scale 0-10 so as to be accurate in providing mean values of them. Please note that despite being human evaluation, this does not contradict our approach of generating an overall automatic evaluation to individual and group performance on the discussion. However, delayed human evaluations may impede a prompt updated evaluation.

All these quantitative and qualitative indicators are to be weighted adequately according to the specific goals and procedures of each discussion. To that end, a fully customizable environment is necessary to parameterize and adjust each indicator with an appropriate weight by the tutor at any moment of the discussion process.

4. Validation results and interpretation

For the specific purpose of validating the reliability of the automatic evaluation approach, the tutor supervising the discussion was required to both (i) submit through the system a precise assessment on content quality of every contribution posted, which was presented to students as feedback information and (ii) evaluate students' performance manually by the tutor by filling out a spreadsheet that helped score each student's participation according to both the content quality of each of his/her contribution and the purpose and context where the contribution took place (e.g., whether it was a new argumentation or a reply, brought interesting opportunities for further discussion, it was just a greeting-type post, etc.). This second evaluation task could be complemented with extra information on individual and personal behavior in the discussion added by the tutor according to his knowledge and experience in this type of class assignment.

Selected questions	Average of structured responses (0 – 5)	Excerpt of students' comments
Assess in general the new Discussion Forum tool (DF)	3	"It was very useful to know the immediate effect of my participation in the discussion and compare it to the rest of the class"
Evaluate how the DF fostered your active participation	4	"The statistical data and quality evaluation displayed influenced my participation"
Did the DF help you acquire knowledge on the discussion's topic?	4	"The DF should be used to support discussions in other courses, since the standard discussion tool does not provide any evaluation information"
From your experience, compare the DF to the campus' standard discussion tool.	4	

The ultimate aim of this double evaluation process was to compare the manual evaluation performed by the tutor to the semi-automatic evaluation process provided by the system. To this end, each evaluation process resulted in proposing both a final mark for each student and a position list where all students were ranked according to his/her final mark (see first and last columns in the monitoring information depicted in Figure 3). In the automatic evaluation, on the one hand, the system addressed four indicators, namely, activity, passivity, impact and effectiveness, becoming 50% of the automatic evaluation. The rest of the evaluation came from the quality indicator only, which was addressed by the tutor who was in charge of assessing the contributions' content quality (40%), and the peers who assessed the usefulness of others' contributions on average (see also Figure 2). Please note that these percentages may vary according to the type of the discussion and they can be adjusted by the tutor. On the other hand, the manual evaluation process was carried out entirely by the tutor and followed the same evaluation procedure as that performed while using the standard discussion tool of the UOC.

The results of the automatic evaluation were very promising since the tutor in charge of the DF agreed with the final marks proposed by the system in more than 75% of cases. 31 out of 40 students in the DF's rank matched the same position as in the rank appeared in the tutor's spreadsheet. In addition, the tutor reported how the DF alleviated him from the tedious and error prone work of monitoring and assessing the discussion's dynamics and outcomes manually.

From the students' standpoint, the continuous provision of feedback in terms of evaluation information also resulted very promising. Comparing the students' performance using the standard tool at the UOC with the DF, on average 32% of students had improved their qualitative mark by going through the discussion in their threads, 68% kept the same mark, and no mark had dropped. In addition, students reported many benefits from using the new tool. Table 2 shows an extract of the results of the questionnaire addressed to the DF' users.

5. Conclusions and further work

This is an initial effort towards a just-in-time evaluation in on-line discussions. Although it may not be pedagogically appropriate to consider a whole course or curricula, we have shown the feasibility of automating the evaluation of certain in-class assignments, such as online discussions. Overall, the results presented here are not conclusive but they encourage us to undertake more experimentation and especially validation processes on the automatic evaluation approach. Nevertheless, the new discussion tool has been proved to promise significant benefits for students in the context of learn by discussion and collaborative learning in general.

Ongoing and future research directions are going through several perspectives:

- Conceptual. We plan to use other classification strategies as well as Natural Language Processing (NPL) techniques (Vargas-Llosa & Morales, 2005) to improve the automatic categorization of discussion posts.
- Technological. We are exploring the interesting possibilities offered by adding decentralized distributed infrastructure to the prototype of our discussion tool. The gain in performance (Caballé et al., 2007) might help us, for instance, collect more complex information of the collaboration and presented it in real time for evaluation purposes, such as modeling the participants' behavior during the discussion by combining individual and group session and navigation information.
- Application. Next step is to validate our approach at large scale by leveraging the investigations reported here to help tutors and moderators to monitor and evaluate the heterogeneous discussion dynamics found in the different studies and programs of the UOC.

Acknowledgements

References

[1] Dillenbourg, P. (ed.) (1999): Collaborative Learning. Cognitive and Computational Approaches. ElsevierScience Ltd. 1-19.

[2] Salomon, G. (1993). No distribution without individual's cognition: a dynamic interactional view. In G. Solomon (Ed.), Distributed cognitions (pp. 111–138). Cambridge, MA: Cambridge University Press.

[3] Kulesza, E., Reinalda, B. (2006). The Bologna Process - Harmonizing Europe's Higher Education. 2nd edition.

[4] McGrath, J.E. (1991). Time, interaction and performance (TIP): A theory of groups. *Small Group Research*, 22, 147-174. SAGE.

[5] Daradoumis, T., Martínez, A., Xhafa, F (2006). A Layered Framework for Evaluating Online Collaborative Learning Interactions. *International Journal of Human-Computer Studies*, 64(7), 622-635. Elsevier Ltd.

[6] McDonald, J. (2003). Assessing online collaborative learning: process and product. *International Journal on Computers and Education*, 40 (4), 377-391, Elsevier Ltd.

[7] Zumbach, J., Hillers, A., Reimann, P. (2003). Supporting Distributed Problem-Based Learning: The Use of Feedback in Online Learning. In: T. Roberts (Ed.), Online Collaborative Learning: Theory and Practice, 86-103. Idea Group.

[8] De Wever, B., Schellens, T., Valcke, M., Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*. 46(1), 6-28. Academic Press: Elsevier Ltd.

[9] Soller, A (2001). Supporting Social Interaction in an Intelligent Collaborative Learning System. *Int. J. of Artificial Intelligence in Education,* 12, 40-62.

[10] Pena-Shaff, J., Nicholls, C. (2004). Analyzing student interactions and meaning construction in computer bulletin board discussions. *Computers & Education*. 42(3), 243-265. Elsevier Ltd.

[11] Strijbos, J-W., Martens, R., Prins, F., Jochems, W. (2006). Content analysis: What are they talking about? (2006). *Computers & Education*. 46(1), 29-48. Elsevier Ltd.

[12] Schrire, S. (2006). Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. *Computers & Education*. 46(1), 49-70, Academic Press: Elsevier Ltd.

[13] Caballé, S., Daradoumis, Th., Xhafa, F. (2008). Providing an Effective Structured and Context-aware Asynchronous Discussion Forum for Collaborative Knowledge Building. *In Proc. of ED-MEDIA 2008. Vienna, Austria.* AACE Press. ISBN: 1-880094-62-2.

[14] Weimer, M., Gurevych, I., Muhlhauser, M. (2007). Automatically Assessing the Post Quality in Online Discussions on Software. *In Proc. of ACL 2007.*

[15] Kim, J., Shaw, E., Feng, D., Beal C., Hovy, E. (2006). Modeling and Assessing Student Activities in On- Line Discussions, *Proc. of the AAAI Workshop on Educational Data Mining*.

[16] Kim, J., Chern, G., Shaw, E. (2007). Towards automatic assessment of on-line discussions: Analyzing student "speech acts", *American Educational Research Association* (AERA 2007).

[17] Martin, J.R. (1992). English Text: Systems and Structure. Amsterdam: Benjamin Press.

[18] Clark, H., Schaefer, E. (1989). Contributing to discourse. Cognitive Science. 13(2), 259-294.

[19] Self, J.A. (2004). Dormobilea: vehicle for metacognition. In: Chan, T.W. and Self, J.A. (Eds.), Emerging Computer Technologies in Education, 1-20. AACE Press.

[20] Witten, I. H., Frank, E. (2005). Data mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco, CA. 2nd Edition.

[21] Caballé, S., Daradoumis, Th., Xhafa, F. Providing an Effective Structured and Context-aware Asynchronous Discussion Forum for Collaborative Knowledge Building (2008). *In Proc. of ED-MEDIA 2008.* Vienna, Austria. AACE Press.

[22] Caballé, S. (2008). A Computational Model for the Construction of Knowledge-based Collaborative Learning Distributed Applications. *PhD thesis.* Open University of Catalonia, Barcelona, Spain.

[23] Duda, R., Hart, P., Stork, D. (2001). Pattern Classification. 2nd. Edition. New York.

[24] Burges, C. J. C. A. (1998). Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery*. 2(2), 121-167.

[25] Müller, K. R., Mika, S., Rätsch, G., Tsuda, K., Schölkopf, B. (2001). An Introduction to Kernel-Based Learning Algorithms. *IEEE Trans. Neural Networks,* 12 (2), 181-201.

[26] Rifkin, R., Klautau, A. (2004). In Defense of One-Vs-All Classification (2004). *Journal of Machine Learning Research*, 5, 101-141.

[27] Caballé, S., Xhafa, F., Daradoumis, T. (2007). A Service-oriented Platform for the Enhancement and Effectiveness of the Collaborative Learning Process in Distributed Environments. LNCS 4804. Springer-Verlag.

[28] Maria Vargas-Vera, M., Moreales, E. (2005). Automatic extraction of knowledge from student essays. *International Journal of Knowledge and Learning,* 1(4), 318-331. Inderscience En-terprises Ltd.

Author Biographies

Santi Caballé has a PhD, Masters and Bachelors in Computer Science from the Open University of Catalonia. Since 2003, he has been an Assistant Professor at the Open University of Catalonia teaching a variety of courses in Computer Science in the areas of Information Systems, Software Engineering and Collaborative Learning. Since 2006, he has been working as an Associate Professor of the department of Computer Science, Multimedia and Telecommunication at the Open University of Catalonia where he coordinates several online courses in the area of Software Engineering. His research focuses on e-learning, software engineering, network technologies, distributed learning, computer-supported collaborative learning, interaction analysis, and grid technologies. His home page is located at http://cv.uoc.edu/~scaballe.

Àgata Lapedriza received the Ph.D. degree in Computer Science from the Universitat Autònoma de Barcelona, Spain, in 2009, for her work on Multitask Learning techniques for Automatic Face Classification. Since 2008 she has been with the Universitat Oberta de Catalunya, Barcelona, as an Assistant Professor. She has authored several scientific publications and her research interests are related to Statistical Pattern Recognition, Machine Learning and Face Classification. Her email address is alapedriza@uoc.edu.

David Masip received the Ph.D. degree in computer science at the Computer Vision Center (CVC), in the Autonomous University of Barcelona (UAB), Spain (2005). In 2007 he became Associate Professor in the Computer Science, Multimedia and Telecommunication Department from the Open University of Catalonia (UOC). His research interests concern the development of algorithms of feature extraction, pattern recognition, face classification and machine learning. He has authored more than 30 scientific papers and one book. His email address is dmasipr@uoc.edu.

Fatos Xhafa received his PhD in Computer Science from the Polytechnic University of Catalonia (Barcelona, Spain) in 1998. He joined the Department of Languages and Informatics Systems of the Polytechnic University of Catalonia as an Assistant Professor in 1996 and is currently an Associate Professor. He is member of the ALBCOM Research Group of this department. His research is supported by several research projects from Spain, European Union and NSF/USA. He has published in leading international journals and conferences. He serves in the editorial board of several leading journals and has served as Co-chair / PC member for many conferences and workshops. His email address is fatos@lsi.upc.edu.

Ajith Abraham received Ph.D. degree from Monash University, Melbourne Australia and a Master of Science Degree from Nanyang technological University, Singapore. His research and development experience includes over 17 years in the Industry and Academia spanning different continents in Australia, America, Asia and Europe. He works in a multi-disciplinary environment involving computational intelligence, network security, sensor networks, e-commerce, Web intelligence, Web services, computational grids, data mining and applied to various real world problems. He has authored/co-authored over 350 refereed journal/conference papers and book chapters and some of the papers have also won best paper awards at international conferences and also received several citations. He is a senior member of IEEE, IEEE Computer Society, IEE (UK), ACM etc. More information at: http://www.softcomputing.net