Towards an Automatic Real-Time Assessment of Online Discussions in Computer-Supported Collaborative Learning Practices

Santi Caballé[1], Fatos Xhafa[2], Ajith Abraham[3]

[1]Open University of Catalonia, Department of Computer Science, Multimedia, and Telecommunication
Barcelona, Spain
scaballe@uoc.edu
[2]Polytechnic University of Catalonia, Department of Languages and Informatic Systems
Barcelona, Spain
fatos@lsi.upc.es
[3]Center of Excellence for Quantifiable Quality of Service, Norwegian University of Science and Technology
Trondheim, Norway
ajith.abraham@ieee.org

Abstract

The discussion process plays an important social task in Computer-Supported Collaborative Learning (CSCL) where participants can discuss about the activity being performed, collaborate with each other through the exchange of ideas that may arise, propose new resolution mechanisms, and justify and refine their own contributions, and as a result acquire new knowledge. Indeed, learning by discussion when applied to collaborative learning scenarios can provide significant benefits for students in collaborative learning, and in education in general. As a result, current educational organizations incorporate in-class online discussions into web-based courses as part of the very rationale of their pedagogical models. However, online discussions as collaborative learning activities are usually greatly participated and contributed, which makes the monitoring and assessment tasks time-consuming, tedious and error-prone. Specially hard if not impossible by humans is to manually deal with the sequences of hundreds of contributions making up the discussion threads and the relations between these contributions. As a result, current assessment in online discussions restricts to offer evaluation results of the content quality of contributions after the completion of the collaborative learning task and neglects the essential issue of constantly assessing the knowledge building as a whole while it is still being generated. In this paper, we propose a multidimensional model based on data analysis from online collaborative discussion interaction that provides a first step towards an automatic assessment in (almost) real time. The context of this study is a real on-line discussion experience that took place at the Open University of Catalonia.

1. Introduction

In CSCL environments [1] 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 [2]. In particular, a complete discussion and reasoning process is based on three types of generic contributions [2], namely specification, elaboration and consensus. 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 [3] 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 [3].

In the context of these new principles and theories, current educational organizations incorporate in-class online student discussions into web-based courses as part of the very rationale of their pedagogical models. One key issue in online discussions is interaction management and analysis to support the provision of relevant and selected knowledge about collaboration [1]. The aim is to support instructors' monitoring and assessment tasks as well as enhance fundamental aspects of the learning process, such as problem-solving abilities by means of supporting peer- and self-evaluation and allowing learners to be aware of the progress of their peers and of their own.

In previous research [4], we reported on real experiences of learn-by-discussion fully student-centered by using a *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 in the collaboration at his convenience but provided adequate scaffold 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, and monitoring as regards learners' performance and collaboration. In this paper we extend the purpose of the provision of information and knowledge to collaborative learning activities for prompt and constant assessment of individual and group performance in online discussions in an automatic fashion.

From the literature, the automatic assessment of online discussion contributions have been, to the best of our knowledge, little investigated. Quite a few research studies, such as [5] and especially [6], [7], show a first step towards this direction by combining several quantitative analysis and modeling the threaded discussions. Some relevant references [6], [7] in this field, 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 assessment 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, [5] 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.

In this paper, we take these entire approaches one step further and also provide an innovative process for real-time assessment of online discussions based on interaction data analysis from online discussions. 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.

To this end, we propose in Section 2 a model for managing interaction in a discussion process based on both speech act analysis [8], [9] and a machine learning approach [10]. The information captured by this model is then turned in Section 3 into a multidimensional framework of knowledge used to assess participation behavior, knowledge building and performance. Section 4 incorporates this framework into a structured discussion forum based on these principles and reports 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. A Model for Managing Interaction in a Discussion Process

The proposed framework is based on an integration of several models and methods: the Negotiation Linguistic Exchange Model [8], a model of Discourse Contributions [9] and a machine-learning approach [10].

In particular, this Section examines how the building

[1] The Open University of Catalonia (UOC) is located in Barcelona, Spain. The UOC offers distance education through the Internet since 1994. About 47,000 students, lecturers and tutors participate in some of the 600 on-line official courses available from 23 official degrees and other PhD and post-graduate programs. The UOC is found at http://www.uoc.edu

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).

The structure of a long interaction is constructed cooperatively by using the exchange as the basic unit for communicating knowledge. Following [8], 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. 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, etc.) of a situation 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.

According to [8], 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. The obligatory move of each of the above exchanges is: the first move of the give-information exchange, the second move of the elicit-information exchange and the third move of the ascertain-information exchange.

According to [9], 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. Some moves have a pure contributing function toward the realization of the obligatory move of the exchange. This is the case of the first move of the elicit-information exchange, as well as of the first and the second moves of the ascertain information 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 [9]. In contrast, the 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.

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.

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

A further innovation of this model is to incorporate a machine-learning approach to learn the relation between a set of types of contributions and the perceived intention of the authors of these contributions. The ultimate aim is to automate the manual post tagging (see Figure 1) so as to both minimize wrong post tagging and release students of unnecessary choice. To this end, 1241 individual posts to a online discussion forum were tagged by their authors using one of the 6 exchange moves presented in Table 1. We removed 220 which were used just for training purposes. The rest, 1021, 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: *support* (16.3%); *request* (31.1%); *inform* (36%); *set-up-an-issue* (5.1%); *provide-solution* (9.8%); *consent-solution* (1.6%). Following the similar work of [5], for each post, we compiled a category vector, and category values were normalized to the range [0.0, …, 1.0]. Based on [10], we used state-of-the-art classification algorithms so as to learn the real intention of a contribution. This is a very initial attempt and more validation process needs to be undertaken.

To satisfy course assessment 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 assessment should be also encouraged and facilitated by intuitive means. Following [5], peer evaluation could be also replaced with an automatic rating system.

Then, a 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. A multidimensional Framework to Assess Participation Behavior, Knowledge Building and Performance

Participation behavior 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.

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 [1].

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 assessment to individual and group performance on the discussion. However, delayed human evaluations may impede a prompt updated assessment.

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. Providing Updated Assessment to Real Online Discussions

A prototype of a web-based structured discussion forum system, called Discussion Forum (DF) [4], was developed to validate the approach. We report here this novel experience that gives new opportunities to learning by discussion, and is applied to meet new pedagogical needs. To this end, certain details of the design of this application regarding the assessment process and presentation are presented. Finally, in order to evaluate this prototype, and most importantly, the provision of an automatic updated assessment of the online discussion, the experimental results are presented and discussed.

4.1 An effective structured discussion forum

The design of the DF includes certain thematic annotation tags based on the low-level exchange categories identified in Section 2, such as *information-clarification* and *request-opinion* (see Table 1 for a list of all categories), which qualify each contribution and as a result structure the discussion process. 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 much shorter and easier and no error-prone (see Figure 1). Please note that in this early version of the prototype users were urged to qualify their posts. In next iterations of this application, it is planned to incorporate an automatic tagging system based on the machine-learning approach presented in Section 2.

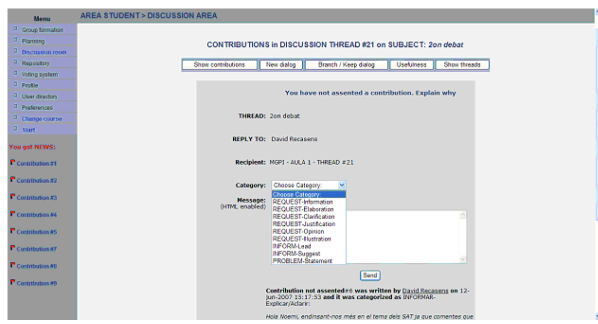

Figure 1. The specific list of tags for a reply to a post categorized as INFORM-Explain

In addition, as part of the design, the tutor is to examine and assess the quality of all contributions based partially on the tags used by students to categorize them, and as a result students are aware of the potential repercussions of tagging posts incorrectly in order to optimize the assessment instead of reflecting the true meaning of their posts.

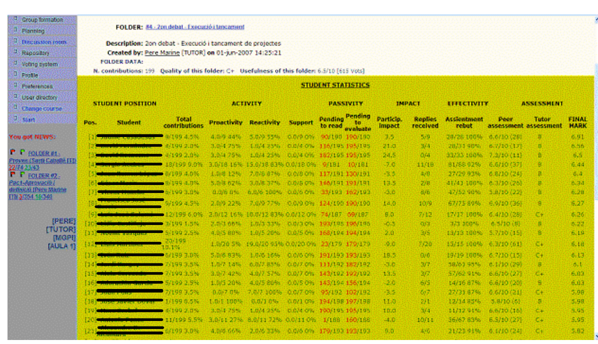

Figure 2. Monitoring information provided to the tutor

4.2 Validation of the approach

In order to evaluate the prototype of the DF, 40 graduated students enrolled in the course Methodology and Management of Computer Science Projects at the UOC were involved in a pilot experience. Students were required to use the new DF outside the campus to participate in a two-week class assignment consisting of an online discussion about the issue: *project management requirements vs. product requirements*.

For the specific purpose of validating the reliability of the automatic assessment approach, the tutor supervising the discussion was required to both (1) submit a precise assessment on content quality of every contribution posted, which was presented to students as feedback information and (2) 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.

The ultimate aim of this double evaluation process was to compare the manual evaluation performed by the tutor to the semi-automatic assessment 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 2). In the automatic evaluation, on the one hand, the system addressed four indicators, namely, activity, passivity, impact and effectiveness, becoming 50% of 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 assessment procedure as that performed while using the standard discussion tool of the UOC.

The results of the automatic assessment 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 alleviates him from the tedious and error prone work of tracking and assessing the discussion's dynamics and outcomes manually.

5. Conclusions and Ongoing Work

This is an initial effort towards an automatic assessment in on-line discussions. Although it may not be pedagogically appropriate to automate a whole course or curricula, we have shown the feasibility of automating the assessment 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 assessment approach.

Ongoing work aims at incorporating the automatic post tagging and rating system introduced in Section 2 so as to obtain a more reliable assessment process as well as to release students from unnecessary choice.

References

[1] Dillenbourg, P. (ed.) (1999): Collaborative Learning. Cognitive and Computational Approaches. Elsevier Science 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] 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.

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

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

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

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

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

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