McLaren, B. M., Scheuer, O., DeLaat, M., Hever, R., DeGroot, R., & Rosé, C. P. (2007). Using Machine Learning Techniques to Analyze and Support Mediation of Student E-Discussions. In the *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (AIED 2007), 331-338.

Using Machine Learning Techniques to Analyze and Support Mediation of Student E-Discussions

Bruce M. MCLAREN[1], Oliver SCHEUER[1], Maarten DE LAAT[2], Rakheli HEVER[3], Reuma DE GROOT[3], and Carolyn P. ROSÉ[4]

[1] Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken, Germany
[2] Exeter University, United Kingdom
[3] The Hebrew University of Jerusalem, Israel
[4] Carnegie Mellon University, Pittsburgh, USA
{Bruce.McLaren, Oliver.Scheuer}@dfki.de; M.F.DeLaat@exeter.ac.uk; mt.scopus@gmail.com; msruma@mscc.huji.ac.il; cprose+@cs.cmu.edu

Abstract. Students are starting to use networked visual argumentation tools to discuss, debate, and argue with one another about topics presented by a teacher. However, this development gives rise to an emergent issue for teachers: how do they support students during these e-discussions? The ARGUNAUT system aims to provide the teacher (or moderator) with tools that will facilitate effective moderation of several simultaneous e-discussions. Awareness Indicators, provided as part of a moderator's user interface, help monitor the progress of discussions on several dimensions (e.g., critical reasoning). In this paper we discuss preliminary steps taken in using machine learning techniques to support the Awareness Indicators. Focusing on individual contributions (single objects containing textual content, contributed in the visual workspace by students) and sequences of two linked contributions (two objects, the connection between them, and the students' textual contributions), we have run a series of machine learning experiments in an attempt to train classifiers to recognize important student actions, such as using critical reasoning and raising and answering questions. The initial results presented in this paper are encouraging, but we are only at the beginning of our analysis.

1. Introduction

It is becoming increasingly common for students to use computer-based tools to discuss, debate, and argue with one another about topics presented in a classroom. Such collaborative software tools are designed to allow students to work on separate computers but communicate in synchronous fashion, contributing to an evolving discussion through a shared "workspace." In Israel, England, and the Netherlands, for instance, we are working with and collecting data from over 15 classrooms that are using the tool Digalo (http://dito.ais.fraunhofer.de/digalo/) to engage students in e-discussions. We also have plans to collect data from the use of another collaborative tool, Cool Modes (www.collide.info/software). These tools share a common model in supporting e-discussions: a graphical environment with drag-and-drop widgets that students can use to express their ideas, questions, and arguments in visual fashion.

Of course, simply providing students with computer-based tools, such as Digalo and Cool Modes, will not lead to fruitful discussion and collaboration. Evidence from the computer-supported collaboration literature suggests that fruitful collaboration does not occur spontaneously [1]. One approach that has been used and continues to be investigated by many researchers is the notion of providing a "script" to support collaboration (e.g., [2, 3]). Another approach is to provide a software agent that can coach and/or tutor the collaborating students [4]. A third approach is to include an artificial student in the collaboration whose responsibility it is to provide student-like contributions and peer coaching [5].

Yet another approach, taken within the ARGUNAUT project discussed in the current paper, is to assist the human moderator, or teacher, of an e-discussion in keeping students on topic, correcting misstatements, and generally guiding the students toward fruitful discussion and collaboration [6]. Such an approach has the advantage of leveraging the knowledge and skills of perhaps the premier resource in supporting learning and collaboration: the teacher. To the extent a teacher can provide personalized attention to individual conversations, and individuals in those conversations, he or she is *tutoring* the students, widely considered to be the most effective way of providing instruction.

However, the cognitive load of moderating multiple simultaneous e-discussions may be too high on teachers. The ARGUNAUT system aims at relieving this by providing the teachers with feedback regarding important elements of each discussion, explicitly focusing their attention on events or situations requiring their intervention.

An approach we are pursuing is to use machine-learning techniques [7] to evaluate past e-discussions and use the results to provide "Awareness Indicators" for teachers and moderators in the context of new e-discussions. Pedagogical experts on our team have annotated components of past discussions as to whether, for instance, students applied critical reasoning, were on topic, or were engaged in raising and answering questions. We then used these annotations to train machine-learning classifiers for the purpose of identifying these meaningful characteristics in new e-discussions. The annotations and subsequent classifications are based on structural, process-oriented, and textual elements of the student contributions.

In this paper, we provide an overview of ARGUNAUT, present and discuss the initial results obtained in applying machine learning to a corpus of Digalo data for the purpose of supporting ARGUNAUT Awareness Indicators, and discuss our next steps in attempting to realize fully the potential of machine learning applied to e-discussions.

2. Overview of the ARGUNAUT Project

An e-discussion in ARGUNAUT proceeds by students responding to a given question (e.g., "What is your opinion about experiments on animals?") and to one another's contributions. Students make contributions to the discussion by dragging and dropping shapes corresponding to discussion moves, such as "question" or "claim," filling text into those shapes to express an idea or argument, and linking shapes to other shapes with directed or undirected links, labeled as "supports" or "opposes." To support the moderator of such discussions, the approach we have taken is to provide a "Moderator's Interface" that contains, among other data, a series of "Awareness Indicators" associated with each e-discussion. The Moderator's Interface allows the teacher to oversee all of the e-discussions currently taking place in the classroom.

Each Awareness Indicator of an e-discussion provides summarized information about an important aspect of the discussion. An indicator may be a relatively easily computed aspect of the e-discussion, such as how many times each student has contributed to the conversation, or a more complex analysis component, such as whether the students have been or are currently engaged in critical reasoning. The human moderator is the ultimate judge of when and how to intervene in an e-discussion; the indicators are intended only to call attention to potentially important discussion characteristics.

The architecture we have designed and partially implemented to support this approach is shown in Figure 1. The end user environment, shown on the left side of

Figure 1, is the Digalo or Cool Modes tool used by groups of collaborating students. All actions taken by the students, such as creating a new shape, link, or textual contribution, are logged. The moderator (teacher) uses the Moderator's Interface, shown in the middle of Figure 1, to monitor on-going e-discussions and intervene when appropriate. The Awareness Indicators are based on either "Shallow Loop" analysis, straightforward calculations, such as counting the contributions made by each student, or "Deep Loop" analysis, application of machine-learning classifiers to the logged data. Using the Annotation Tool, the moderator can annotate maps in real-time to provide additional data to train the Deep Loop classifiers. The Deep Loop, shown on the right side of Figure 1, is an off-line process that takes annotated maps, translates them to a form suitable for machine learning, and generates new classifiers that are subsequently used at run time to classify actions of the students in real time.

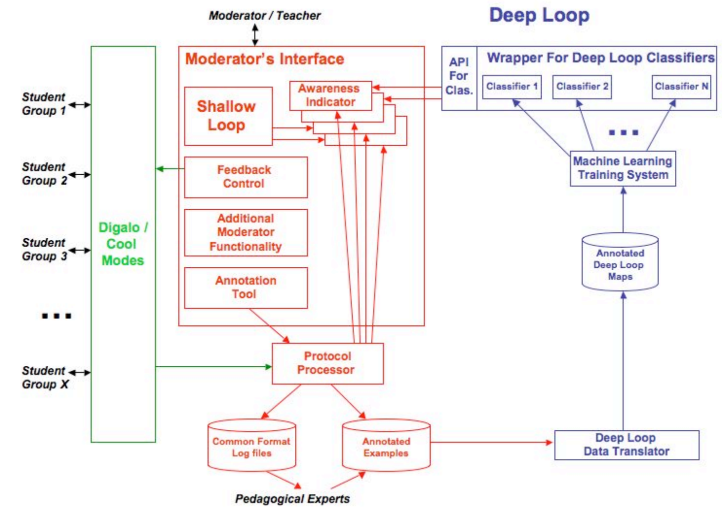

Figure 1: The ARGUNAUT Architecture

3. Annotating Collaborative Data for the Deep Loop

The first step in generating Deep Loop classifiers for the ARGUNAUT system was the manual annotation of existing e-discussion Digalo maps. This corresponds to the process depicted at the bottom of Figure 1, in which human pedagogical experts annotate real maps. The annotated maps are the products of actual Digalo sessions in Israeli junior high school and university classrooms. For the purposes of our initial analyses, the maps were also translated from Hebrew to English.

The experience of our pedagogical experts suggested general criteria for coding, such as participation and responsiveness [8], as well as dialogic analysis [9]. The selection of annotation categories was based on discussion between the pedagogical and technical experts of our team, as well as an iterative analysis of the maps.

We developed specific annotation schemes for two aspects of the Digalo maps: (1) individual contributions (*the shape level*, i.e., single shapes) and (2) connected contributions of one or more students (the *paired-shapes level,* i.e., two contributions (shapes) with a connecting link) in a map. While the shape-level annotations focus primarily on interpretation of the text within a shape, the paired-shapes level involves analysis of structural, process-oriented, and textual aspects of the shapes. A paired-shape involves the interpretation of two distinct but related pieces of text, the structural relationship between the contributions (a connector), and the order of the contributions. In total, we defined seven annotation variables for the shape level (topic focus, task-management focus, critical reasoning, request for clarification or information, critical evaluation of opinions, summary, and intertextuality) and five annotation variables for the paired-shapes level (question-answer, contribution-counterargument, contribution-supportingargument, contribution followed by question, and qualifier/compromise). Tables 1 and 2 show examples of a few of these variables.

Table 1: Sample of the Shape-Level Coding Scheme

Annotation-variable name	**Explanation / Coding**	**Examples**
Topic focus (TF)	A contribution that focuses on the topic or task. 0. No topic focus content 1. Topic focus content	TF: "Its not nice of human beings to exploit animals for their own needs. I think animals also have rights." Non-TF: "I'm bored."
Critical Reasoning (CR)	An individual contribution that contains critical reasoning or argumentation (i.e., claim + backing). Student provides an explanation or some backing (e.g. evidence) to illustrate a position/opinion. If you can add "because" between two parts of the contribution, it is probably critical reasoning. 0. No critical reasoning 1. Use of critical reasoning	CR: "I am against experiments on animals, because to my opinion it is not fair to use them against their will while they cannot reject." CR: "Here it's not like with humans, as the father disengages from them, and he doesn't see them even in the afternoon, and he doesn't belong to the pack any more"

Table 2: Sample of the Paired-Shapes Coding Scheme

Annotation-variable name	**Explanation / Coding**	**Examples**
Contribution-CounterArgument (CCA)	A contribution in which the 2nd shape opposes the claim/argument raised in the 1st shape and provides reasons or other type of backing for the opposing claim. Typically (but not necessarily) the type of link between the shapes would be "opposition." 0. Not a Contribution-CounterArgument 1. A Contribution-CounterArgument	1st shape text: "Do not separate, the male should be a partner in what happens even after the birth. The offspring is also his and he should take responsibility." 2nd shape text: "But in a situation like this the mother can get pregnant again and so might neglect a group of cubs." Link between shapes is "opposition"
Question-Answer (QA)	A contribution in which the 1st shape is a question, the 2nd shape is an answer to that question. Typically (but not necessarily) the type of link between the shapes would be "other." 0. Not a Question-Answer 1. A Question-Answer	1st shape text: "In the wild does the father separate from the cubs or does he continue to live with them?" 2nd shape: "They all live in a pack" Link between shapes is "other."

At the shape level, a total of 677 shapes in 42 Digalo discussion maps were annotated. Three skilled coders initially applied the coding scheme summarized in Table 1 to a small number of maps. Coding differences were resolved through discussion and decision rules, and the annotation scheme was revised. The three coders then annotated the remainder of the maps, with discussion and decision rules resulting in a "consensus" annotation for every shape of every map. Interrater reliability, using Fleiss' Kappa, was calculated for the second round of annotations, resulting in an acceptable value (or within range of acceptability) for all variables (i.e., close to 0.7).

At the paired-shapes level, 226 paired-shapes in 21 Digalo discussion maps were annotated. Two of our coders applied the coding scheme summarized in Table 2 to five maps, resolved differences, and applied a revised coding scheme to the remainder of the maps. Interrater reliability, using Cohen's Kappa, was calculated, with acceptable values for all variables except qualifier/compromise, which was not used for learning.

4. Use of Machine Learning Techniques in the Deep Loop

Given these manual annotations, our ultimate goal is to train machine-learning classifiers to predict the appearance of these discussion characteristics in the context of new e-discussions and make the resultant classifiers available to the ARGUNAUT Awareness Indicators depicted in Figure 1. Our first step toward this goal was to validate the use of a machine learning approach and to identify the variables that hold the greatest promise of being useful to the Awareness Indicators. It is clearly non-trivial to generate good classifiers, given the complex characteristics of e-discussions, including the shapes chosen by students for contributions, the links created between shapes, and the text provided within a shape. On the other hand, an approach that does reasonably well, with a success rate of, say, 80% to 90% in the most critical situations, may be sufficient for the purposes of moderation.

Our approach draws from and is most closely aligned with that of Donmez, Rosé, Stegmann, Weinberger, & Fischer [10]. Using TagHelper, a software tool that leverages machine-learning techniques to classify text, they ran multi-dimensional analyses over a large corpus of textual argumentation data and achieved a 0.7 Cohen's Kappa (or higher) on 6 of 7 dimensions. We are also doing multi-dimensional analysis, as exemplified by the multiple variables of interest at the shape and paired-shapes levels, and also experimented with TagHelper as part of our analysis. On the other hand, our objective is different. Donmez *et al* focused exclusively on textual contributions, while we are interested in structural and chronological data in addition to the text. Also, while Donmez *et al.*'s goal is to automate corpus analysis for coders, we will use our classifiers to provide feedback for e-discussion moderators.

5. Initial Analyses and Results of Machine Learning

For our initial analyses, we experimented both with TagHelper and YALE, freely available software that supports interactive experimentation with a wide range of machine learning algorithms [11]. We ran some preliminary YALE experiments, using the first 21 annotated maps (318 shapes). We derived a basic set of attributes that expressed characteristics of the text and structure of the shapes, including text length, shape type, number of in-links, number of out-links, and number of undirected links. Using a variety of algorithms, including J48 decision tree and PART, and 10-fold cross validation, the critical reasoning variable yielded classifiers that had by far the best results, including a Kappa value of 0.72 (above the standard acceptable level of 0.7) with most other Kappas at or near acceptable (0.6 to 0.7). All of the other shape-level variables yielded classifiers with lower Kappa values, 0 in many cases. A possible reason for the lower Kappa values is the proportionally unbalanced annotations of either positive or negative labels for these other variables, compared to the relatively balanced distribution of positive (47%) and negative (53%) labels for critical reasoning.

We then focused solely on the critical reasoning variable, since it appeared to hold the best prospects for the machine learning approach (and is also a variable of keen interest to our pedagogical experts). We scaled up our analysis to the fully annotated 42 maps (677 shapes) and experimented with deeper aspects of the language within shapes. In addition to the basic set of attributes described above, we derived two additional attributes that represent word-level contributions: *term_evidence_pos*, a count of the number of words used in the text contribution of a shape that are contained in a "positive" term list, and *term_evidence*, the difference between the number of words used in the text contribution that are contained in the "positive" and the "negative" term lists. The two lists consisted of the top 25 words in the "positive" and "negative" categories identified by a word analysis of the entire corpus of 677 shapes, using the Odds Ratio (originally defined, but not named, in [12]). We also applied TagHelper to the data, and it generated over 1600 text-focused attributes, representing terms and parts of speech in the corpus.

Table 3: ML Results for the Critical Reasoning Var. running AdaBoost with DecisionStump over 42 Maps

Attributes	Parameters	True Pos. Rate	True Neg. Rate	Acc.	Kappa
Basic set (baseline)	I = 10 (Default)	84%	74%	79%	**0.57**
Basic set	I = 100 (Tuning)	81%	81%	81%	**0.62**
(+ *term_evidence_pos*)	I = 100 (Tuning)	80%	83%	82%	**0.63**
(+ *term_evidence*)	I = 140 (Tuning)	79%	84%	82%	**0.63**
(+ TagHelper attributes)	Chi-squared att. sel.; I = 60 (Tuning)	82%	87%	85%	**0.68**

A summary of the results of these experiments, again run using 10-fold cross validation, are shown in Table 3. AdaBoost with DecisionStump yielded the best results without parameter tuning, so we focused on this algorithm in our experiments. As can be seen, the baseline Kappa achieved with this algorithm, with no parameter tuning run over the basic set of attributes (i.e., the "Default"), was 0.57. Adding the *term_evidence_pos* and *term_evidence* attributes to the basic set of attributes, as well as applying parameter tuning to the number of AdaBoost iterations (I), yielded improvement over the baseline, but the best Kappa (0.68) was obtained using chi-squared attribute selection over the TagHelper and basic attributes. The difference between the baseline and TagHelper Kappa is significant ($p < 0.003$) using a t-test. Thus, our initial shape-level experiments demonstrate that at least for the critical reasoning variable, we can obtain an acceptable Kappa value with the assistance of TagHelper. These experiments also demonstrate the potential value of text-focused attributes combined with structural attributes when learning over our e-discussion maps, as the use of TagHelper attributes yielded the best results.

Finally, we performed a series of YALE experiments to investigate how well we could train classifiers to identify the paired-shapes variables, such as those shown in Table 2 (e.g., question-answer), using the 21 annotated maps (226 paired shapes) as training data. In this analysis we accounted for all three aspects of the maps – text, structure, and process. However, we did not use TagHelper, as it seemed to have a somewhat less logical application to multiple texts associated with single training instances (i.e., the separate texts associated with each of the shapes in a paired shape). We also tried to see if we could re-use the shape-level annotations as attributes in these learning experiments. At the paired-shapes level, we identified and experimented with three sets of mutually exclusive attributes, as well as combinations of those sets:

1. *Basic attributes that do not represent process or structure:* verticesSameUser, combinedTextLength, diffInTextLength, linkType, timeBetweenFirstModOfEachShape
2. *Attributes that represent ordering (i.e. process) and structure of shapes:* link_v[1,2]_sameUser, textLength[1,2], shape[1,2], linkDirection
3. *Attributes that represent ordering and rely on shape-level annotations (using the human annotations):* vertex[1,2]_TopicFocus, vertex[1,2]_TaskManagementFocus, vertex[1,2]_CriticalReasoning

Table 4 shows the results we achieved for the question-answer variable, which had the most balanced proportion of positive (29%) and negative (71%) examples, using the PART algorithm, 10-fold cross validation, and parameter tuning of the confidence threshold for pruning (C) and the minimum number of examples per leaf (M). The default parameters were used in the baseline case of attribute set 1 on its own. The paired-shapes results for question-answer can be viewed as highly encouraging, even more so than the critical reasoning variable at the shape level, with the highest Kappa achieved 0.87 when applying the PART algorithm to the attribute sets 1, 2, and 3. The difference between the baseline and this Kappa is significant ($p = 0.00001$) using a t-test. Results for the other paired-shapes level variables were generally not as good as the results shown in Table 4, with the problem of too few labels leading to Kappa values of 0 in the worst cases. On the other hand, inclusion of the process-related and shape-level annotation attributes improved the learning results in virtually all cases.

Table 4: ML Results for the Question-Answer Variable running PART algorithm over 21 Maps

Att. Sets	Parameters	True Pos. Rate	True Neg. Rate	Acc.	Kappa
1 (baseline)	C = 0.25, M = 2 (Default)	59%	95%	84%	**0.59**
1	C = 0.3, M = 12 (Tuning)	67%	97%	88%	**0.69**
1 + 2	C = 0.2, M = 2 (Tuning)	88%	95%	93%	**0.83**
1 + 3	C = 0.3, M = 2 (Tuning)	92%	94%	94%	**0.85**
1 + 2 + 3	C = 0.8, M = 2 (Tuning)	92%	96%	95%	**0.87**

6. Discussion

In sum, our initial results, while preliminary, provide promise that we will be able to support moderation in ARGUNAUT, at least for some variables.

As discussed with respect to the shape-level analysis, the addition of text analysis attributes to basic structural attributes improved learning significantly. The benefits that TagHelper provided over learning with structural attributes alone, or learning with structural attributes combined with simple word attributes, shows that the kinds of linguistic features TagHelper can extract are important for distinguishing between different forms of conversational behavior. In fact, it may be that the textual features alone, without the structural features, is sufficient for training the classifiers. Upon one reviewer's suggestion, we tested this and got a Kappa of 0.67, marginally below the best Kappa of 0.68 in Table 3. So an important next step is to investigate more fully the relative contributions of the two types of attributes. Although we have not yet applied TagHelper to the paired-shapes level of analysis – it seemed less a fit for those instances and our initial results were already quite good – we will investigate whether we can improve those results, too, using TagHelper.

The paired-shapes analysis illustrated the importance of learning with structural and process attributes, both of which are critical attribute types derivable from our e-discussion maps. Including these types of attributes with the basic attributes (i.e., adding attribute sets 2 and/or 3 to attribute set 1), along with parameter tuning, led to statistically significant improved learning. In addition, part of our vision for using machine learning is to see how much we can leverage smaller building blocks, for instance the classification of individual shapes, in analyzing and learning from larger portions of the maps. We took an initial step in this direction by including individual shape annotations as attributes in training at the paired-shapes level and this supported the best results of all our experiments.

An issue related to language analysis is the translation that we did from Hebrew to English to perform the experiments described in this paper. While this was a necessary step for initial experimentation, it is an unrealistic approach if the goal is ultimately to classify Hebrew (or other language) maps. We will need to include some form of cross-language analysis in our approach. This is an issue we will investigate with respect to our TagHelper work and collaboration with Carolyn Rosé. In the short-term, however, we will continue to focus on maps translated to or originally created in English.

Our initial analysis uncovered other issues to address. For instance, the lack of positive or negative labels for a number of variables, both at the shape and paired-shapes level, led to poor results for those variables. We plan to address this in two ways. First, we will find and annotate more examples of the minority labels in our large repository of Digalo maps. Second, we will investigate the use of cost-sensitive machine learning techniques, an approach sometimes used to deal with imbalanced data sets [13, 14]. Since in many cases the minority label is the one our moderators will be most interested in (e.g., it is more important to correctly identify when students are off than on topic), such an approach makes sense in our case.

Finally, we simply need to do more extensive machine learning experimentation to verify the viability of this approach. The current sample of data is relatively small and focuses on a limited set of e-discussions. Annotating and training on a larger corpus will provide a clearer picture of whether this direction is fruitful and will allow us to generalize the classifiers.

7. Conclusion

The ARGUNAUT system aims to provide a moderator (or teacher) with tools that will allow him or her to moderate the simultaneous e-discussions of many groups of students as they discuss, debate, and argue difficult topics using a visual argumentation software tool. Awareness Indicators, provided as part of a Moderator's Interface, are intended to alert the moderator of important events in the e-discussion, such as students not using critical reasoning in their contributions. In this paper we have discussed preliminary steps taken in using machine learning techniques to support the Awareness Indicators. Our preliminary results are quite encouraging, but there is still much work to be done. We are only in the first year of a three-year project.

References

[1] Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1995). The evolution of research on collaborative learning. In P. Reimann & H. Spada (Eds.), Learning in humans and machines: Towards an interdisciplinary learning science, 189-211. Oxford: Elsevier/Pergamon.

[2] O'Donnell, A. M. (1999). Structuring dyadic interaction through scripted cooperation. In A. M. O'Donnell & A. King (Eds.), Cognitive perspectives on peer learning, 179-196. Mahwah, NJ: Erlbaum.

[3] Diziol, D., Rummel, N., Spada, H. & McLaren, B. M. (in press). Promoting Learning in Mathematics: Script Support for Collaborative Problem Solving with the Cognitive Tutor Algebra. To appear in *Computer Supported Collaborative Learning* (*CSCL-07*).

[4] Constantino-Gonzalez & Suthers (2002). Coaching Collaboration in a Computer-Mediated Learning Environment. In the *Proceedings of Computer Supported Collaborative Learning (CSCL-02)*.

[5] Vizcaíno, A. (2005). A Simulated Student Can Improve Collaborative Learning. *Journal of Artificial Intelligence in Education*, 15, 3-40.

[6] De Groot, R., Drachman, R., Hever, R., Schwarz, B., Hoppe, U., Harrer, A., De Laat, M., Wegerif, R. McLaren, B. M. & Baurens, B. (in press). Computer-Supported Moderation of E-Discussions: the ARGUNAUT Approach. To appear in *Computer Supported Collaborative Learning* (*CSCL-07*).

[7] Witten, I. H. & Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques, Second Edition, Morgan Kaufmann.

[8] Schwarz, B. B. & Glassner, A. (in press). The role of CSCL Argumentative Environments for Broadening and Deepening Understanding of the Space of Debate. In R. Saljo (Ed.), Information Technologies and Transformation of Knowledge.

[9] Wegerif, R. (2006). A dialogic understanding of the relationship between CSCL and teaching thinking skills. *Int'l Journal of Computer Supported Collaborative Learning*, 1(1), 143-157.

[10] Donmez, P., Rosé, C. P., Stegmann, K., Weinberger, A., & Fischer, F. (2005). Supporting CSCL with Automatic Corpus Analysis Technology. In the *Proc. of Computer Supported Collaborative Learning* (*CSCL-05*), 1-10.

[11] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M. & Euler, T. (2006). YALE: Rapid Prototyping for Complex Data Mining Tasks. In *Proc. of the 12th ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining (KDD 2006)*, 935-940. ACM Press.

[12] Van Rijsbergen, C. J. (1979). Information Retrieval. Butterworths, London, 2nd edition.

[13] Zadrozny, B., Langford, J. & Abe, N. (2003). A Simple Method for Cost-Sensitive Learning. IBM Technical Report RC22666, December 2002.

[14] Japkowicz, N. & Stephen, S. (2002). The Class Imbalance Problem: A Systematic Study, *Intelligent Data Analysis*, 6 (5), 429-450.