

Helping Teachers Handle the Flood of Data in Online Student Discussions

Oliver Scheuer and Bruce M. McLaren

Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken, Germany
Oliver.Scheuer@dfki.de, bmclaren@dfki.de

Abstract. E-discussion tools provide students with the opportunity not only to learn about the topic under discussion but to acquire argumentation and collaboration skills and to engage in analytic thinking. However, too often, e-discussions are not fruitful and moderation is needed. We describe our approach, which employs intelligent data analysis techniques, to support teachers as they moderate multiple simultaneous discussions. We have generated six machine-learned classifiers for detecting potentially important discussion characteristics, such as a “reasoned claim” and an “argument-counterargument” sequence. All of our classifiers have achieved satisfactory Kappa values and are integrated in an online classification system. We hypothesize how a teacher might use this information by means of two authentic e-discussion examples. Finally, we discuss ways to bootstrap from these fine-grained classifications to the analysis of more complex patterns of interaction.

Keywords: Educational data mining, Natural Language and Discourse, Architectures, Machine Learning in ITS.

1 Introduction

E-discussion tools provide students with the opportunity not only to learn about the topic under discussion but to acquire argumentation and collaboration skills and to engage in analytic thinking. Tools such as Digalo¹ and Free Styler² [1] allow students to use a shared workspace to present ideas, debate and argue with one another, and ask questions. Visual languages consisting of typed text boxes and links provide additional scaffolds that help students structure the way they think about and discuss a topic. Nevertheless, too often discussions are unfruitful: students misuse the tools for private conversation instead of staying on topic, contributions lack critical reasoning, arguments and questions of other participants are ignored and some students don't participate at all, while others dominate discussions. Thus, there is a need for active help and guidance, be it from a machine tutor or a human teacher / moderator.

Our focus is on helping a teacher moderate a classroom of students using e-discussion tools in which the students comprise multiple discussion groups. The teacher can bring to bear his or her experience and moderation expertise to steer the

¹ <http://dito.ais.fraunhofer.de/digalo/>

² <http://www.collide.info/software>

discussions when problems occur and provide encouragement when discussions are productive. However when multiple e-discussions occur simultaneously, a single teacher may struggle to follow all of the discussions. To direct the teacher's attention to the 'hot spots,' software tools that pre-process, aggregate, and summarize the incoming flood of data could be extremely valuable. In this respect, the task is reminiscent of that faced by systems that monitor power plants and medical patients where vast amounts of raw data are analyzed, filtered and/or condensed to support human decision making.

The ARGUNAUT project follows this approach with the ultimate aim of supporting teachers as they guide multiple, simultaneous e-discussions. Two analytical processes are particularly prominent in ARGUNAUT's analysis of e-discussions: (1) the "Shallow Loop" focuses on surface features that can be computed in a straightforward way (e.g. the total number of contributions per student) and (2) the "Deep Loop" evaluates situations requiring more complex analysis, combining textual, sequential and structural information to classify more abstract aspects of discussion (e.g. on-topic-ness, reasoned claim). The Deep Loop inference mechanism is based on machine-learned classifiers, developed from our corpus of annotated discussions, and is the focus of this paper. Our long-term goal is to use the classifiers also to automate support and feedback for collaborating students in typical intelligent tutoring fashion.

In this paper, we provide an overview of the ARGUNAUT project and an in-depth treatment of the Deep Loop Classification system. We describe the approach we have employed to develop the Deep Loop classifiers and present the quantitative results we've achieved, particularly emphasizing progress made since the work reported in [2]. Finally, we provide specific examples and discuss how a teacher might use the Deep Loop classifiers to identify good and bad discussion situations.

2 Related Work

The most relevant work to ours is by Rosé and colleagues, who have developed the text analysis tool TagHelper [3], also used in our work. Originally, they aimed at freeing corpus analysts from the tedious task of manually coding large amounts of data, rather than analyzing online discussions, which is our goal. In one application [4] they analyzed a corpus of 1,250 coded text segments along multiple dimensions of argumentation in order to derive machine-learned classifiers. Some of the phenomena of interest in their work, like argument-counterargument chains and grounded claims, are quite similar to the categories we are interested in. They achieved acceptable Kappa values of 0.7 or higher for six of seven dimensions. More recently, they developed an approach to providing dynamic support to dyads collaborating on a problem-solving task [5]. Similar to our approach, they perform online analysis of textual communication data, in their case, chat data. In contrast to our approach, their analysis results are not displayed to human teachers but are instead used to trigger automatic interventions in the students' activities. An empirical study showed significant learning benefits in terms of analytical knowledge and conceptual understanding, when dynamic support is provided

Goodman et al. [6] also have developed a machine-learning approach to support collaborative problem solving. Peer groups work together on a problem in the domain of object modelling techniques (OMT). Their collaboration takes place within a shared whiteboard (similar to the shared workspaces in ARGUNAUT) in which diagrams (e.g. class diagrams) have to be constructed. Peers communicate via a text chat with a sentence opener interface; task management is supported by an agenda tool. The system evaluates aspects concerning domain (e.g., domain knowledge of peers), task (e.g., progress in solving the task) and, similar to our objectives, possible problems in the collaboration process (e.g. unanswered questions). The sentence opener interface plays a critical role; it is used to automatically assign a dialogue act classification to each chat contribution. These dialogue acts are used as a meta-level description of the discourse and serve as features for machine-learning analyses, bypassing the complicated task of natural language processing. The provided support is two-fold: Some of the results are displayed immediately to the peers via meters, while direct support is provided by means of an artificial peer agent that verbally interacts with the participants.

3 The ARGUNAUT Approach

In ARGUNAUT students discuss and debate questions within a shared workspace on different networked computers in synchronous fashion [7, 8]. A discussion starts with a shape containing the question to be discussed. Usually, controversial topics are chosen (like experiments on animals, abortion) to allow students to take different positions and to promote a lively exchange of arguments. Students contribute by adding shapes, entering text into the shapes and connecting the shapes by links. Shapes and links are not just simple text boxes and connectors; they have types and comprise a visual language: There are shape types to express claims, arguments, questions, etc, and link types to establish supporting and opposing relations.

A teacher can monitor multiple ongoing discussions in parallel using a tool called the “Moderator’s Interface”. Here, important aspects of the discussion are displayed in the form of “Awareness Indicators”. As discussed above, “shallow indicators” can be computed in a straightforward fashion (e.g. the number of contributions per user); “deep indicators” result from a more sophisticated machine learning-based analysis. Currently, two types of deep indicators are computed: *Shape-level* indicators reflect characteristics of a single contribution (e.g. whether this contribution contains a reasoned claim); *paired-shape indicators* reflect characteristics of two linked contributions (e.g., whether two shapes constitute a contribution-counterargument pair). Six classifiers for deep indicators are currently available: “Reasoned Claim”³ and “Topic Focus” at the shape level, “Question-Answer”, “Contribution followed by Question”, “Contribution followed by Counterargument” and “Contribution followed by Supporting Argument” at the paired-shape-level. (Note that this is four more classifiers than were available in an earlier reporting of our work [2].)

Although the Moderator’s Interface has not yet been used in a real classroom, we anticipate that the combination of shallow and deep indicators will enable teachers to more effectively and efficiently moderate multiple, simultaneous discussions.

³ Note that in previous work, this category was referred to as “Critical Reasoning”.

4 Deep Loop Classifiers

The classifiers were developed using a three-stage process: First, a coding scheme was developed for categories of interest and the data was coded accordingly. Second, the data was translated into a format amenable to standard machine-learning algorithms. Third, experiments with a multitude of machine-learning techniques were carried out in order to derive the most effective classifiers. The resulting classifiers have been integrated into the Classification Web Service to enable a teacher to run classifications online.

Originally, we were interested in twelve discussion categories but after initial experiments we focused our efforts on the six categories described below. There are several reasons why we limited our scope: One category did not have sufficient inter-coder agreement. Other categories were shown to be less promising after some initial machine-learning experiments, partly because of an unbalanced class distribution and too few examples for one class, two problems that are well known for their detrimental effects on machine learning [9, 10].

4.1 Data Description

The first step in building an example corpus was to collect data and code this data with the categories of interest (Deep Loop indicators). The data was collected during real classroom sessions in Israel and the U.K. Because the discussion language in Israel is Hebrew, it was necessary to translate these discussions into English before coding. This was done for experimental purposes only; in the longer term our intention is to use customized versions of TagHelper applied to the language of interest.

After the pedagogical experts on our team agreed on a set of categories, coding instructions were developed, consisting of detailed explanations of when a code applies and additional illustrative examples for further clarification. The final corpus presented here is the product of several coding iterations carried out by our pedagogical experts in Israel and the U.K. To determine the reliability of the coding procedure, inter-rater reliability was computed by means of the Kappa statistic, yielding acceptable values (near or above .70) for all but one category. More details concerning the coding procedure can be found in [2].

The final corpus comprised data of 72 discussions covering ethical questions ('Should we clone humans?') as well as questions of opinion and fact ('How does the use of ICT affect learning experiences?'). In the end, we had 1,260 annotated shapes and approximately 1,000 annotated shape pairs. All but one category (Reasoned Claim) show a clear majority of one class, with proportions ranging between 75 % and 85 % of all instances.

4.2 Machine Learning Experimentation and Results

As a first step, the data was cast in a form suitable for machine learning. We used a data-centric approach by encoding as much information as possible in feature-value form, without considering the specific categories of interest, in hopes that the inference mechanism itself would choose the relevant pieces of information. Shapes and paired shapes were analyzed in terms of structural properties (shape and link types,

incoming and outgoing links), sequential properties (chronological sequence of shapes) and textual properties (textual content of shapes). The textual analysis was done using TagHelper, discussed earlier [3]. We pre-processed the data by reducing terms to their word stems and removing stop words. We extracted unigrams (single terms), bigrams (pairs of consecutive terms), part-of-speech bigrams (two consecutive part-of-speech classes), punctuation marks and text lengths.

The experiments were conducted with RapidMiner (formerly known as “Yale”), a machine learning toolkit offering a wide range of methods for data pre-processing, machine learning and validation [11]. We experimented with a variety of learning algorithms using different feature combinations, estimating the reliability of our classifiers by cross-validating data from one discussion (test set) against the data from the remaining discussions (training set). Because data from one discussion was never in the training and test set at the same time we avoided intra-discussion dependencies and bias. We measured the reliability using Cohen’s Kappa [12] (a criterion more appropriate than the widely used error rate and accuracy measures which are both vulnerable to unbalanced class distributions). A Kappa value of 1.0 signifies a perfect classifier, a Kappa value of 0 means a classifier performing equally well as a trivial classifier that always chooses the majority class and Kappa below 0 means a classifier even worse than the trivial majority voter.

Support Vector Machines (SVM), Boosted Decision Trees and Decision Lists proved to be the most effective machine-learning algorithms. We achieved Kappa values ranging from .60 to .71, which can be interpreted, according to [13], as moderate (1 category) to substantial agreements (5 categories) between the human annotations and machine-learned model⁴. Because these six classifiers performed reasonably well, we integrated them into the classification web service.

5 How the Deep Loop Classifiers Could Help Teachers

In this section, we turn to the practical application of the Deep Loop. As discussed earlier, we have taken the approach of first having pedagogical experts on our project team identify discussion categories of interest, annotate instances of those categories, and then apply machine-learning techniques to create classifiers for those categories. However, at this stage of our project we have received only minimal feedback from teachers on the pedagogical value of the classifiers. The following question then arises: what could a teacher do with the results of the six classifiers described above? We address this question by showing and discussing the Deep Loop classifiers applied in two authentic discussions. We hypothesize how a teacher might interpret actual Deep Loop results to recognize one discussion situation as fruitful and another as requiring support.

In general, we expect a fruitful discussion to be lively, with (close to) equal contributions by all participants. Questions should be answered and claims should be

⁴ There is no universal threshold for an acceptable Kappa value, the decision regarding what is acceptable and what is not depends on domain and application. A more rigorous threshold is given by Krippendorff [14] who recommends a value of .67 as the minimal acceptable inter-rater agreement for content analyses. Given that a teacher who is aware of uncertainties and possible misclassifications will ultimately interpret the provided Deep Loop results, we consider the slightly more generous interpretation as sufficient.

backed by supporting arguments. Contrary positions should be acknowledged and answered with counterarguments. Participants should always be open to persuasion when more compelling arguments than their own are raised. Such quality criteria are supported by the CSCL literature, see for instance [15] for a more detailed account of criteria for assessing the quality of collaboration.

Figure 1 shows a situation that matches much of this description (CASE 1). The question raised in this discussion was whether abortion is ethical when it is known that the child will be born with a serious handicap. The question is introduced through a fictional story that has been read by the students prior to the discussion. The original question posed by the teacher is marked by a double-framed box. The figure shows two students involved in a dialogue where student 1 (solid boxes) takes a position against and student 2 (dashed boxes) in favor of abortion. Student 1 starts by stating his opinion (contra abortion) and backs his position by pointing to the human right to life. Student 2 counters that both parents and child will suffer, that a lot of money will have to be spent, etc. Student 1 gives as a counterargument that, provided enough money for treatments is available, the family can nevertheless live a good life. Student 2 objects that possibly the family does not have enough money. Finally, student 1 closes the thread by integrating both views: An abortion might be acceptable if the family does not have enough money, otherwise not. Regardless of one’s personal position on abortion, this thread exhibits positive discourse characteristics, including:

- (1) both students react to the position of their counterpart resulting in an argument-counterargument chain (3 CCA paired-shape classifications),
- (2) the chain contains a considerable amount of reasoned claims (3 RC shape classifications),
- (3) all student contributions are on-topic (5 TF shape classifications),
- (4) a posed question was answered (1 QA paired-shape classification), and
- (5) the thread ends with an integration of both views.

Figure 2 shows a discussion in which a teacher’s intervention might be helpful (CASE 2). This discussion addresses the same topic as example 1, namely abortion under special circumstances. To the left of the teachers’ assignment (double-framed box) there are contributions in favor of abortion in this situation. In summary, the

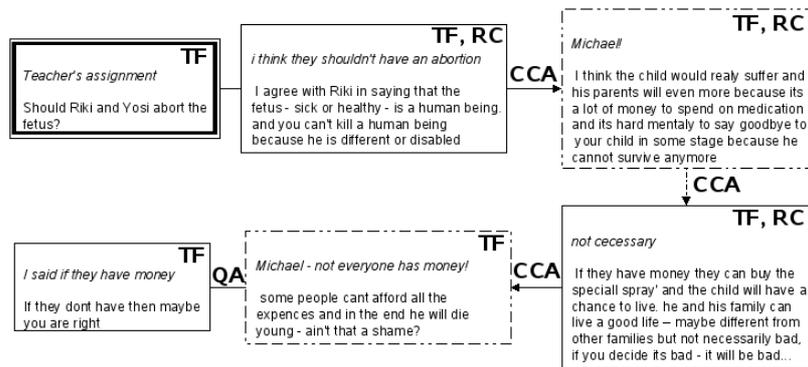


Fig. 1. Positive discussion situation (CASE 1)

main arguments here are that both child and parents will suffer, and that it will be easier for the parents to abort a still unborn fetus than seeing their child die later on. On the right-hand side we see contra-abortion contributions. The main arguments here are based on a human’s right to life and religious convictions. Although here, too, almost all of the contributions are on-topic and the arguments are valid, the discussion suffers from a lack of interaction between participants. The arguments are made in isolation, almost exclusively linked to the original question. Only the three shapes at the lower right show some rudimentary interaction between the participants. Consequently, our classifiers detect only one contribution-counterargument pair in the entire discussion, in contrast to the three contribution-counterargument pairs in CASE 1 (and there we only show a fraction of the entire discussion graph). At this point, a teacher might find it valuable to intervene, encouraging the students to react to one another’s positions and perhaps come to an integration/synthesis of multiple views.

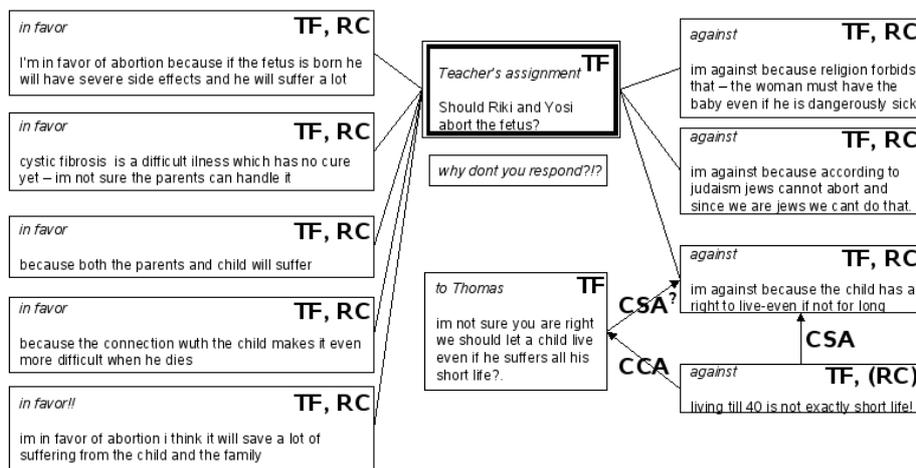


Fig. 2. Discussion situation in which a teacher might want to intervene (CASE 2)

With the current classifiers, a teacher might be able to use the information of Figures 1 and 2 in both a quantitative and a qualitative manner. Quantitative usage is supported in the Moderator’s Interface through display within each ongoing discussion the number of occurrences of each classification type. In this way, a teacher can easily detect that CASE 1 contains a high proportion of counterarguments, whereas CASE 2 has a noticeable lack of this discussion element (there are 12 shapes and only 1 CCA relation), indicating that intervention may be helpful. Qualitative evaluation might also be helpful, as a purely quantitative analysis might point to critical situations but can also be deceiving because dynamic aspects of the discussion are no longer visible. For instance, discussions might be in need of intervention, even with the presence of many ‘reasoned claims’, as exemplified in CASE 2. Inspecting this situation on the level of individual classifications reveals that virtually all of the contributions refer to the original question shape and not to the shapes created by the participants. The students here are only enumerating supportive and opposing arguments

but do not *deepen* their understanding of the space of debate by arguing on arguments and negotiating the meaning of underlying concepts [16].

Furthermore, a qualitative examination – or visual impression – might provide an idea of what is going on without reading the contents. There might be controversial discussion with lots of reasoned claims and argument-counterargument pairs, situations in which students don't critically evaluate their peers' opinions, manifested through a lack of both supportive and opposing arguments, and places in which an accumulation of off-topic contributions suggests a drifting from the discussion. Once such regions are identified, a teacher can read the contributions in more detail and intervene as required.

Our approach enables a teacher to look at a discussion on different levels of detail: at the highest level information is summarized (and maybe also distorted), at the middle level the structure of interaction patterns is preserved but still abstracted from specific content, and at the lowest level the full information is offered without loss of veracity but at the possible cost of detail "overload."

6 Discussion, Conclusions, and Future Work

As we have seen, machine-learned classifiers can compute useful aspects of e-discussions even without an in-depth (i.e., semantic) analysis of natural language. We succeeded in deriving six classifiers that analyze discussion situations in terms of structural, chronological and shallow text characteristics and assign categories like "reasoned claim" and "contribution-counterargument" to the analysis units. The classifiers are integrated within the ARGUNAUT system and allow teachers to analyze simultaneous discussions online. We discussed how a teacher might use the system with two authentic examples. Our initial results are quite encouraging, but there are still open questions.

One crucial question is how far our classifiers generalize beyond the training corpus. Clearly, we cannot claim that our data sample has been drawn randomly from the population of *all* possible discussions. Although we have collected a considerable amount of discussion data, the number of covered topics is still somewhat restricted. Especially our 'topic focus' classifier might incorporate idiosyncrasies from the topics being covered in the training corpus and suffer when applied to different topics. Other categories, like "reasoned claim", "question-answer" and "contribution-counterargument", are largely topic-independent and may not be vulnerable to such dangers.

The real measure of success for any computer program is successful use by its intended audience working on tasks for which the program is intended for. Consequently, the logical next step is to test the Deep Loop classifiers 'in the wild' and to collect feedback from teachers charged with monitoring multiple discussions simultaneously. This will help us find answers to the question of which categories help (and which do not), which additional (missing) categories may be of interest, whether the visualization of the classifications is appropriate, and whether the reliability threshold we've adopted ($Kappa > 0.6$) is strict enough for this specific domain of application. Such feedback will help us to move towards a real usable system.

Currently, as demonstrated by the two examples above, the classifications might help a teacher find patterns of good and bad discussion situations (e.g. commission and omission of argument-counterargument chains). An alternative to such qualitative analyses is

for the classifications to be used in a quantitative way to *summarize* discussions along categories of interest. A first step in this direction has already been done: the total number of positive classifications is displayed to the teacher. We envision going much further by providing comparative statistics, in the form of appropriate visualizations (like bar charts or pie charts), showing, for instance, that 40 % of all links in discussion *X* represent an argument-counterargument relation whereas only 20 % a argument-supporting argument relation. A final step might be to infer automatically qualities of the whole discussion by analyzing its profile in terms of shape and paired-shape classifications. One could define a model using rules such as “If more than 30 % of all links define an argument-counterargument relation then the discussion qualifies as *controversial*”. But of course, hand-crafting such a model requires superior human judgment and expertise. Alternatively, one could use inductive inference. Labels would be assigned to complete discussions and a machine-learned model computed that infers discussion-level classifications from shape and paired-shape classifications. One problem with such an approach is to obtain sufficient data: It is questionable whether the 72 discussions we currently have are sufficient. Another difficulty lies in the propagation of errors in a two-stage classification process: Erroneous classifications on the shape and paired-shape level might cause additional noise in the input for the discussion-level classifier and hence, might have a harmful effect on performance.

Given the uncertainties regarding whether it is possible to define or learn a reliable classification model for discussion characteristics, we see potential in another way of bootstrapping from our shape and paired-shape results: Mikšátko and McLaren [17] have developed a graph-matching algorithm, DOCE, that enables teachers to define and find pedagogically interesting *clusters*, i.e. arbitrary large patterns, in on-going discussions. Clusters are defined in terms of examples (e.g. an example for “Deepening discussion with multiple opinions”) that can be used to find similar clusters in other discussions. We expect these example patterns will prove more useful to a teacher than shape and paired-shape indicators in isolation, which might be structurally too limited and fine grained to capture significant interactions between students. Initial evidence discussed in [17] shows that the use of shape- and paired-shape indicators as cluster features play an important role in the graph-matching process.

Acknowledgements

This work would not have been possible without Rakheli Hever, Reuma De Groot, Maarten De Laat, Matthias Krauß, and Adam Giemza, as well as other members of the ARGUNAUT project team. This research was sponsored by the 6th Framework Program of the European Community, Proposal/Contract No. 027728.

References

1. Zeini, S., Malzahn, N., Hoppe, U.: Kooperationswerkzeuge im Kontext virtualisierter Arbeit. Virtuelle Organisationen und neue Medien 2004. Gemeinschaften in neuen Medien GeNeMe (2004)
2. McLaren, B.M., Scheuer, O., De Laat, M., Hever, R., De Groot, R., Rosé, C.: Using Machine Learning Techniques to Analyze and Support Mediation of Student E-Discussions. In: Luckin, R., Koedinger, K.R., Greer, J. (eds.) Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED 2007), pp. 331–340. IOS Press (2007)

3. Rosé, C., Wang, Y.C., Cui, Y., Arguello, J., Fischer, F., Weinberger, A., Stegmann, K.: Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning. *International Journal of Computer-Supported Collaborative Learning* (in press)
4. Dönmez, P., Rosé, C., Stegmann, K., Weinberger, A., Fischer, F.: Supporting CSCL with Automatic Corpus Analysis Technology. In: Koschmann, T., Suthers, D.D., Chan, T.-W. (eds.) *Proceedings of the Conference on Computer Supported Collaborative Learning 2005 (CSCL 2005)*, pp. 125–134. Lawrence Erlbaum (2005)
5. Kumar, R., Rosé, C., Wang, Y.C., Joshi, M., Robinson, A.: Tutorial Dialogue as Adaptive Collaborative Learning Support. In: Luckin, R., Koedinger, K.R., Greer, J. (eds.) *Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED 2007)*, pp. 383–390. IOS Press (2007)
6. Goodman, B., Linton, F., Gaimari, R., Hitzeman, J., Ross, H., Zarrella, G.: Using Dialogue Features to Predict Trouble During Collaborative Learning. *User Modeling and User-Adapted Interaction* 15, 85–134 (2005)
7. De Groot, R., Drachman, R., Hever, R., Schwarz, B., Hoppe, U., Harrer, A., De Laat, M., Wegerif, R., McLaren, B.M., Baurens, B.: Computer Supported Moderation of E-discussions: the ARGUNAUT Approach. In: Chinn, C., Erkens, G., Puntambekar, S. (eds.) *Proceedings of the Conference on Computer Supported Collaborative Learning 2007 (CSCL 2007)*, pp. 165–167 (2007)
8. Hever, R., De Groot, R., de Laat, M., Harrer, A., Hoppe, U., McLaren, B.M., Scheuer, O.: Combining Structural, Process-Oriented and Textual Elements to Generate Awareness Indicators for Graphical E-discussions. In: Chinn, C., Erkens, G., Puntambekar, S. (eds.) *Proceedings of the Conference on Computer Supported Collaborative Learning 2007 (CSCL 2007)*, pp. 286–288 (2007)
9. Japkowicz, N., Stephen, S.: The Class Imbalance Problem - A Systematic Study. *Intelligent Data Analysis* 6, 429–450 (2002)
10. Weiss, G.M.: Mining with Rarity: A Unifying Framework. In: *SIGKDD Explor. Newsletter*, vol. 6, pp. 7–19. ACM Press (2004)
11. Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T.: YALE: Rapid Prototyping for Complex Data Mining tasks. In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006)*, pp. 935–940. ACM Press (2006)
12. Cohen, J.: A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20(1), 37–46 (1960)
13. Landis, J.R., Koch, G.G.: The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 159–174 (1977)
14. Krippendorff, K.: *Content Analysis: An Introduction to its Methodology*. Sage Publications (1980)
15. Meier, A., Spada, H., Rummel, N.: A Rating Scheme for Assessing the Quality of Computer-Supported Collaboration Processes. *International Journal of Computer-Supported Collaborative Learning* 2, 63–86 (2007)
16. Baker, M., Andriessen, J., Lund, K., van Amelsvoort, M., Quignard, M.: Rainbow: A Framework for Analyzing Computer-Mediated Pedagogical Debates. *International Journal of Computer-Supported Collaborative Learning* 2, 315–357 (2007)
17. Mikšátko, J., McLaren, B.M.: What's in a Cluster? Automatically Detecting Interesting Interactions in Student E-Discussions. In: *Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS 2008)*. Springer (2008)