Recognizing Creative Thinking in Graphical e-Discussions Using Artificial Intelligence Graph-Matching Techniques

Rupert Wegerif, University of Exeter, Graduate School of Education, St Luke's, Exeter, UK,
r.b.wegerif@exeter.ac.uk
Bruce M. McLaren, Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken, Germany,
bmclaren@dfki.de
Marian Chamrada, University of Exeter, Graduate School of Education, St Luke's, Exeter, UK,
mc286@exeter.ac.uk
Oliver Scheuer, Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken, Germany,
Oliver.Scheuer@dfki.de
Nasser Mansour, University of Exeter, Graduate School of Education, St Luke's, Exeter, UK,
N.Mansour@exeter.ac.uk
Jan Mikšátko, Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken, Germany,
Jan.Miksatko@dfki.de

Abstract: Many approaches to analyzing online argumentation focus on explicit reasoning and overlook the creative emergence of new ideas. The value of a dialogic analytic framework including creative emergence was tested through applying it to the coding and analysis of undergraduate synchronous e-discussions using a graphical interface within the EU funded project ARGUNAUT. Qualitative analysis found that critical reasoning functioned to 'deepen' the graph through unpacking assumptions whilst creative emergence of new perspectives produced 'widening' moves. This distinction between deepening and widening was successfully used as the basis for an artificial intelligence (AI) graph-matching algorithm. Given examples of deepening and widening from real e-discussions, the AI algorithm was able to successfully find other occurrences of such moves within new e-discussions. This supports our claim to distinguish between these two aspects of shared thinking and has the potential to provide awareness indicators as a support for e-moderation.

Introduction

This paper reports on part of a European research project, called ARGUNAUT (De Groot, Drachman, Hever, Schwarz, Hoppe, Harrer, De Laat, Wegerif, McLaren, & Baurens, 2007; Hever, De Groot, De Laat, Harrer, Hoppe, McLaren, & Scheuer, 2007; http://www.argunaut.org/), presenting some findings regarding coding for creative thinking in a way that feeds into online awareness tools supporting the moderation of online discussions. The ARGUNAUT system that has been developed during the project uses the graphical e-discussion environments Digalo (dito.ais.fraunhofer.de/digalo/) and FreeStyler (www.collide.info/software) for students, along with a Moderators Interface (MI) for teachers, which includes a range of awareness indicators and tools for intervention designed to provide summarised information and make the task of moderation easier.

As well as providing awareness of relative participation, types of messages, and the relationships between people through social network diagrams, we also sought to provide awareness indicators for the *quality* of discussions. It is in this context that we used discourse analysis to explore effective reasoning and collaboration. In this paper we focus on research into the evocation and coding of creative thinking in contrast to both critical and dialectical reasoning. We report how this distinction was used by artificial intelligence (AI) graph-matching techniques to identify creative thinking (as well as critical reasoning) in new discussions.

The Dialogic Challenge for Coding

Most schemes applied to analyze online argumentation (e.g. those developed originally by Toulmin, Van Eemeren and Walton, see Andreissen, 2006 for discussion) focus on explicit reasoning in the form of claims, challenges to claims and reasons in support of claims. This approach is good at picking up critical reasoning but ignores more creative forms of shared thinking. As opposed to 'dialectics', which always begins as a theory of argument, Bakhtin's 'dialogic' approach begins with 'living' dialogue (Bakhtin, 1986). For Bakhtin consensus or unanimity or intersubjectivity is not the aim of dialogue; rather, the aim is a deepening and expanding of awareness that Bakhtin refers to as inter-illumination. In dialogues voices interact in unpredictable ways to produce new voices and new perspectives that enable participants to see the topic of the dialogue in a new way (Wegerif, 2007).

On the ARGUNAUT project we expanded the dimensions of coding from the traditional single dimension of critical thinking with its focus on claims, counterclaims and reasons (D1) to include the dimension of creative reasoning understood as a sort of dance of perspectives (D2) in which each new perspective or point of view on

a problem is labelled and also the dimension of dialogic engagement which includes not only 'addressivity' (language explicitly addressing the other such as pronouns) and expressions of empathy but also expressions of doubt, changes of mind, 'ventriloquation' (a term from Bakhtin for the presence of another voice within an utterance) and elicitation of the views of others (D3) and finally moderating moves (D4).

To help us code the complex online e-discussions produced in the course of the ARGUNAUT project we developed sequence diagrams, which are visual representations showing both the number and length of sequences of messages and the branching of sequences at different points during the discussion. These sequence diagrams give a visual reference to widening.

This coding scheme and approach to analysis has been tested and further developed through the analysis of over sixty free-form e-discussions created by approximately 100 undergraduates and 12 post-graduate students in the UK. These so-called "discussion maps" were coded with identifying key events such as 'creative widening,' which occurred over several messages. The codes were then used to develop classifiers using artificial intelligence techniques (McLaren, Scheuer, De Laat, Hever, De Groot & Rosé, 2007; Scheuer & McLaren, 2008; Mikšátko & McLaren, 2008) that are able to detect and classify the events automatically and inform the moderator. This is done by a component of the ARGUNAUT system called the "Deep Loop," described below.

An Illustration of Coding for Widening

We began with a very complex Digalo map produced by a group of five undergraduates in response to the question: 'Will the Internet bring the world together or deepen its divisions?' To help the analysis we reorganized the map to a sequence diagram that enabled us to see the critical branching moments more clearly (see Figure 1).
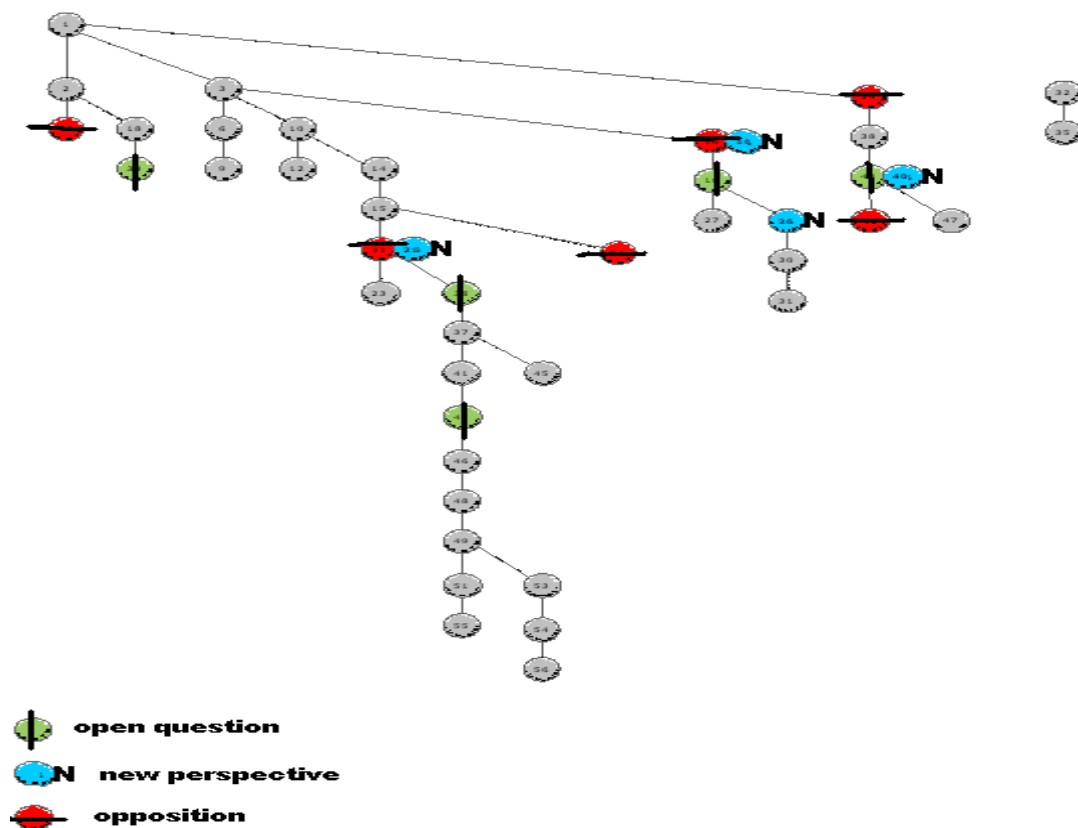

Figure 1: Sequence diagram of a Digalo map

This e-discussion map showed us the key moments when new perspectives emerged (the dots with an 'N' next to them) and these coincided with branching moves in the sequence diagram and seemed typically to occur shortly after oppositions (dots with horizontal line) and open questions (dots with vertical line). Focusing on each key incident we were able to pursue qualitative interpretation of the factors leading to the emergence of new perspectives, and we followed these up with Critical Event Recall interviews with the participants. Figure 2 below illustrates a specific example of the emergence of new meaning. In this snippet of e-discussion, the discussants exchange ideas about awareness of other cultures, ethics, and religions. The "new perspective"

emerges when one student suggests that we may "create a divide" by becoming aware of different cultures, ethics, and religions. In the context this is a new and unexpected perspective.

Once we had used this method to code the maps with a breakdown into clusters indicating widening and deepening, i.e., sets of graph nodes indicating the cluster, such as 15, 21, 23, and 36 in Figure 2, this was subjected to computational analysis to see if artificial intelligence techniques could match the patterns and discover new incidents of creativity in new maps.
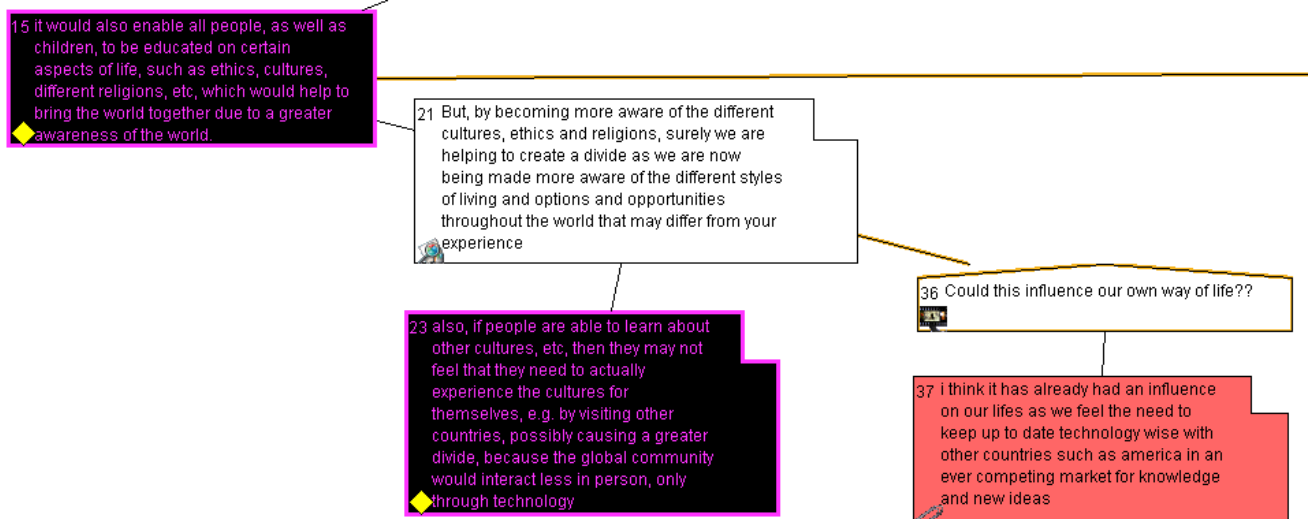

Figure 2: A cluster of shapes around the emergence of a new perspective

A Computational Model to Explore Deepening and Widening

As part of what we call the "Deep Loop" of the ARGUNAUT system, we have developed a computational model called **DOCE** (**D**etection **o**f **C**lusters by **E**xample) (Mikšátko & McLaren, 2008) that allows us to identify places in e-discussions in which students may be deepening or widening the conversation, as well as other types of complex conversational moves. DOCE is one of a number of tools that we developed to assist a teacher in monitoring the on-going simultaneous e-discussions of several groups of collaborating students. The students use the collaborative software tools Digalo or FreeStyler to communicate with one another, with each student working on his or her own computer, while a tool called the "Moderator's Interface" provides the teacher with a variety of important views of the on-going discussions. One of the "views" provided to the teachers is a set of *alerts* that point to critical aspects of the conversation, such as whether students are staying on topic and supporting their claims with good justifications. Some alerts are supported by relatively simple calculations (e.g., how often each student has contributed to the conversation, whether students use swear words), some by machine-learned classifiers (McLaren *et al*, 2007; Scheuer & McLaren, 2008), and some by the DOCE algorithm (Mikšátko & McLaren, 2008).

In particular, the DOCE algorithm identifies *clusters* of contributions, for example, several contributions made by different students that indicate deepening or widening of a conversation. DOCE is based on the idea of using cluster *examples* to find similar clusters in new discussions, inspired by the subfield of artificial intelligence known as case-based reasoning (Kolodner, 1993; McLaren, 2003). DOCE operates by a researcher or teacher selecting a cluster in an existing e-discussion that exemplifies an interesting pattern (e.g. connected individual contributions that provide a good example of deepening). The example cluster (also called a "model graph" in the following text) is then used as a search query for similar clusters across other discussion maps (called "input graphs"). The algorithm uses both structural features (e.g., the pre-specified types of contributions made by students – for instance, "claim" or "question" – and types of links between contributions – for instance, "supporting" or "opposing") and textual features (i.e., the text provided by the students, unigrams, bigrams, and syntactic structures from that text) of the discussion map to find similar clusters. The output of the algorithm is a list of matching clusters in the discussion map(s), sorted according to a similarity rating, as is done by web search engines, such as Google. DOCE can be used as a tool to help researchers find and analyze clusters, such as examples of deepening or widening or it can be used as a "live" classifier of clusters – characteristic example(s) representing a cluster of a particular type are stored in the database and used later as queries for automated cluster detection. Details about the underlying DOCE algorithm are provided in (Mikšátko & McLaren, 2008).

An Experiment to Test the Effectiveness of the Computational Model

We took hand-annotated examples of deepening and widening (annotated by the members of the Exeter team on the co-author list) from actual classroom discussion maps, and tested whether DOCE was able to use those examples to find the *other* examples of deepening and widening in our data set. More specifically, we took 30 annotated examples of both deepening and widening from 14 distinct discussion maps, and did the following:

For each annotated example, we ran DOCE with that annotation as the model graph against all of the other 13 discussion maps (i.e., as "input graphs", as discussed above):

- We considered a *relevant match* to be 70% overlap, e.g., the following model graph and found cluster in an input graph would constitute a relevant match, since there is a 75% node overlap (bold-faced nodes overlap): Model Graph (Node1, **Node3**, **Node4**, **Node5**); Cluster in Input Graph (**Node3, Node4, Node5**, Node6)
- We varied parameters, such as the number (N) of clusters that were returned by DOCE and the relative impact of structural and textual properties on the similarity score of cluster pairs (e.g., Is it more important that texts or shape types are similar?).
- We evaluated recall, precision, and recall+precision on each run of DOCE. These are metrics typically used in information retrieval and were calculated as follows: *Recall* represents the number of relevant matches in the Top N divided by the count of annotations in the searched map (value between 0 and 1.0) whereas *Precision* is the number of relevant matches in the Top N divided by N (value between 0 and 1.0).

Results on the Effectiveness of the Computational Model

The results of our experiment are summarized in figures 3 and 4. Note, first of all, that the *best* results for deepening and widening are quite reasonable (the middle bar for recall, precision, and recall+precision in each of the figures), especially for recall, the metric we consider most important. By "best" result, we mean the human-annotated cluster that led to the best recall and precision values when used as a model graph to DOCE. For instance, notice that the best deepening model graph (the middle bar in each of the first two sets of three metrics in Figure 3) led to a recall of 0.80 and precision of 0.52. The *average* results, calculated across *all* of the annotated clusters (the leftmost bar for recall, precision, and recall+precision in each of the figures), are not good (e.g., the 0.42 recall and 0.27 precision in Figure 3 are very poor). However, focusing on the best results is more important because, by the nature of the DOCE algorithm, only the *best* examples of deepening and widening will subsequently be used as model graphs to DOCE. That is, once one finds the best model for a particular cluster type – or the best set of models – that model (or models) will then be used as a "search probe" for all subsequent searches.
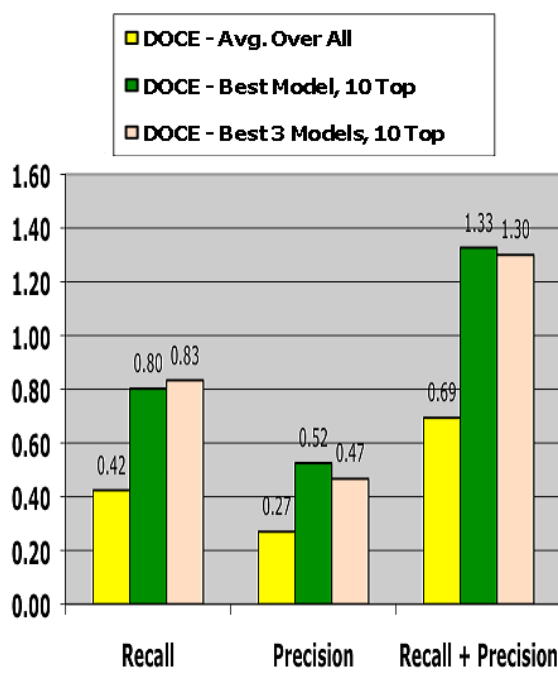

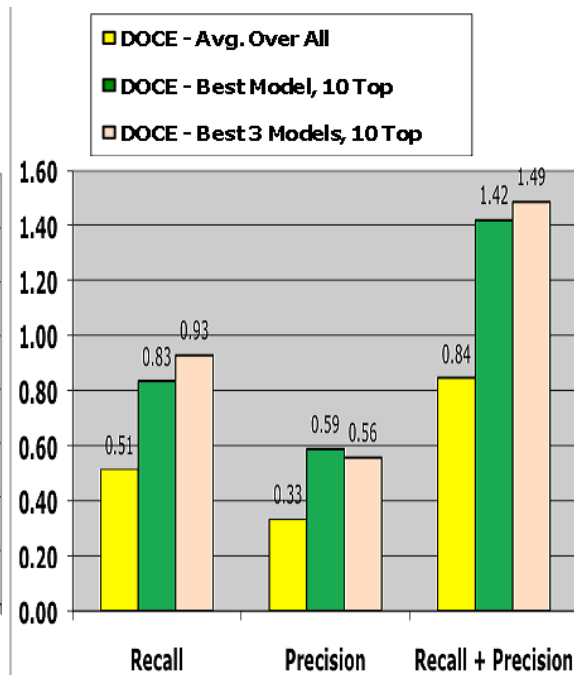**Figure 3**: Results of DOCE on the deepening clusters   **Figure 4**: Results of DOCE on widening clusters

We also tested whether *combining* the results of multiple runs of DOCE might further improve the results. That is, we wanted to answer the question: Can multiple, high-quality clusters lead to even better results than single "best" clusters in retrieving relevant clusters? We implemented this combination by ranking the results according to the average relevance scores of the three single-best models. The third bar in each set of three bars in Figures 3 and 4 depicts these results. Notice that for the deepening cluster results shown in Figure 3 the combination approach did marginally worse (i.e., recall+precision = 1.30 for the combination approach vs. 1.33 for the single best model), but for the widening clusters shown in Figure 4, the combination approach did a bit better (i.e., recall+precision = 1.49 for the combination approach vs. 1.42 for the single best model).

Discussion

These results, while preliminary, are very encouraging. It appears that the DOCE algorithm is reasonably capable of finding examples of the creative widening of a conversation, given prior, annotated examples of such reasoning in earlier discussions. Furthermore, as long as the researcher or teacher is careful not to use too-large model graphs against too-large discussion maps, the DOCE algorithm runs in a practical amount of time. Thus, the DOCE algorithm is a tool that either a researcher or a teacher can use to pinpoint and evaluate indicators of creative reasoning in the context of real e-discussions. The rigor imposed upon qualitative discourse analysis by the use of DOCE algorithm has already proved useful in iteratively refining the essential nature of creative widening in e-discussions. This technique has the potential to inform moderators when widening associated with creative thinking and deepening associated with critical reasoning is occurring in maps, as well as when it is not occurring and thus indicating an opportunity for a teacher to intervene. From analysis of the maps and from critical event recall interviews with participants, it appears that both oppositions and open questions serve to open up reflective spaces in which new insights emerge.

References

Andriessen, J. (2006). Arguing to learn. In: K. Sawyer (Ed.) *Handbook of the Learning Sciences*. Cambridge: Cambridge University Press.

Bakhtin, M. (1986). *Speech genres and other late essays*. Austin: University of Texas.

De Groot, R., Drachman, R., Hever, R., Schwarz, B., Hoppe, U., Harrer, A., De Laat, M., Wegerif, R., McLaren, B.M., & Baurens, B. (2007). Computer supported moderation of e-discussions: the ARGUNAUT approach. In the *Proceedings of the Conference on Computer Supported Collaborative Learning (CSCL-07)*. (p. 165-167).

Hever, R., De Groot, R., De Laat, M., Harrer, A., Hoppe, U., McLaren, B.M., & Scheuer, O. (2007). Combining structural, process-oriented and textual elements to generate awareness indicators for graphical e-discussions. In the *Proceedings of the Conference on Computer Supported Collaborative Learning (CSCL-07)*. (p. 286-288).

Kolodner, J. (1993). *Case-based reasoning*, Morgan Kaufmann Publishers, San Francisco.

McLaren, B.M. (2003). Extensionally defining principles and cases in ethics: An AI model. In: *Artificial Intelligence*, vol. 150, pp. 145-181.

McLaren, B.M., Scheuer, O., De Laat, M., Hever, R., De Groot, 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)*, IOS Press, (p. 141-147).

Mikšátko, J. & McLaren, B.M. (2008). What's in a cluster? Automatically detecting interesting interactions in student e-discussions. In B. Woolf, E. Aimeur, R. Nkambou, S. Lajoie (Eds), *Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS-08)*, Lecture Notes in Computer Science, 5091 (pp. 333-342). Berlin: Springer.

Scheuer, O. & McLaren, B.M. (2008). Helping teachers handle the flood of data in online student discussions. In B. Woolf, E. Aimeur, R. Nkambou, S. Lajoie (Eds), *Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS-08)*, Lecture Notes in Computer Science, 5091 (pp. 323-332). Berlin: Springer.

Wegerif. R. (2007) *Dialogic, educational and technology: Expanding the space of learning*. New York: Springer-Verlag.

Acknowledgments