

PolyCAFe: Collaboration and Utterance Assessment for Online CSCL Conversations

Mihai Dascalu, Traian Rebedea, Stefan Trausan-Matu, University “Politehnica” of Bucharest,
313 Splaiul Independentei, 060042 Bucharest, Romania

Email: mihai.dascalu@cs.pub.ro, traian.rebedea@cs.pub.ro, stefan.trausan@cs.pub.ro

Gillian Armit, University of Manchester, Higher Cambridge Street, Manchester, M15 6FH, UK,
gillian.armit@manchester.ac.uk

Abstract: Students and members of communities of practice are often using online conversations to enhance their knowledge and skills. Although there have been a lot of efforts to find an efficient method for analyzing this type of discussions, most of the research has been mostly theoretical or has not been transformed into software due to technological limitations. This paper presents PolyCAFe, a system built for providing feedback to students that use chats and forums for solving their learning tasks. Moreover, the system can also be used by tutors to supervise or to support students. Starting with the theoretical fundamental aspects of the system, the paper continues with an insight in the technologies used behind PolyCAFe’s design and ends with an overview of the first validation experiment.

Introduction

Instant messenger (chat) is already used for several years in Computer Supported Collaborative Learning (CSCL) sessions (Stahl, 2006). However, there are very few systems designed for automatically analyzing such conversations, but also capable of generating feedback. The explanation is probably founded on the aspect that Natural Language Processing is required and that the existing technologies in Computational Linguistics are still not mature, especially for analyzing chat conversations, which have many important differences as compared to non-conversational text.

Several CSCL systems were developed for analyzing interactions in conversations (face-to-face or virtual through chat) and forums. Some examples are CORDTRA (Hmelo-Silver, Chernobilsky, & Masto, 2006), COALA (Dowell & Gladisch, 2007), DIGALO and other tools used in the Argonaut system (Harrer, Hever, & Ziebarth, 2007), ColAT (Avouris, Fiotakis, Kahrimanis, & Margaritis, 2007), TATIANA, (Dyke, Lund, & Girardot, 2009), the Scaffold-Argument visualization (Law, Lu, Leng, Yuen, & Lai, 2008), KSV (Teplovs, 2008), VMT-Basilica (Kumar, Chaudhuri, Howley, & Rosé, 2009) and Polyphony (Trausan-Matu, Rebedea, Dragan, & Alexandru, 2007).

Some of the aforementioned systems are using different kinds of argumentation graphs (Toulmin, 1958), more elaborated structures like the contingency graphs (Suthers, Dwyer, Medina, & Vatrupu, 2007) or polyphonic threads visualization in Polyphony. Almost all the systems provide only facilities for manual annotation and input of links and of visualization. No system excepting PolyCAFe, the system presented in this paper, provides complex facilities for chat and forum discussions’ analysis and gives useful feedback for learners and tutors.

In our opinion there are several causes that might explain this situation. The first one is that even if Bakhtin’s dialogism is considered a theoretical model of CSCL (Koschmann, 1999; Stahl, 2006), except our own system we don’t know of any actual implementation based upon it. The second cause is related to the fact that the majority of collaborations in CSCL are based on the exchange of text messages. Thus, another problem arises from the fact that current Natural Language Processing (NLP) systems are far from providing reliable text understanding systems. Moreover, in CSCL chats and forums there are usually more than two participants and the floor might be shared by more than a single participant at a given moment in time, a case which is generally not considered in most NLP theories developed for conversation analysis (Trausan-Matu & Rebedea, 2010).

PolyCAFe is based on an integration of NLP, social network analysis and polyphonic analysis (Trausan-Matu & Rebedea, 2010). Its first version was developed and validated under the LTfLL FP7 project (<http://ltfll-project.org>) and it is currently under change to version 2.0. However, all the results presented in this paper are about PolyCAFe 1.0.

The paper continues with section 2, which presents some concepts of Bakhtin’s dialogism that represent one of the central theoretical bases of PolyCAFe. Section 3 briefly presents the technologies behind the system, its main features, while key results from the first validation experiments with the system are depicted in section 4. The paper ends with several points on transferability and conclusions

The Core Concepts of Bakhtin's Dialogism

Our analysis is centered on three different, but inter-dependent concepts: *utterances* briefly defined as units of analysis, *voices* as distinctive points of view emerging from the ongoing discussion and *echo* as the replication of a certain voice with further implications in the discourse.

Utterances and Discussion Threads

Utterances can be defined as pieces of text whose boundaries are represented by the change of speech subject (Bakhtin, 1986) and embed the central unit of analysis of the discussion. Utterances express both acts of communication and pieces of discourse (Linell, 2009) and direct the path and evolution of the ongoing conversation in terms of future development. Our analysis adheres to Dongs perspective of separating utterances based on turn-taking events between speakers (Dong, 2005).

A central aspect that needs to be addressed is the meaning of each utterance within a given context, in our case a discussion thread derived from the utterance graph. The utterance graph is built upon two types of links between utterances: explicit and implicit ones (Dascalu, Rebedea & Trausan-Matu, 2010). Participants can manually add explicit links during their chat sessions by using a facility of the VMT chat environment (Stahl, 2009) we used. On the other hand, implicit links are automatically identified by means of co-references, repetitions, lexical chains, inter-animation patterns and semantic similarity (Trausan-Matu & Rebedea, 2010). In the resulted directed and acyclic graph, each utterance is a node and the weights of edges are given by the similarity between the utterances, multiplied by the trust assigned to each link. The orientation of each edge follows the timeline of the chat and the evolution of the discussion in time. Starting from the previous graph, a discussion thread can be easily identified as a logical succession of explicitly or implicitly inter-linked utterances. Moreover, the primary extension of each utterance is its inner voice that intertwines with other voices from the same thread or from different ones, but with less strength. A new intervention or a new utterance in terms of units of analysis can be clearly expressed as a voice and the following aspects that need to be addressed include: coherence, degree of interconnection with other previous utterances, relevance within the discourse and future impact in the overall discussion.

Voices

A voice expresses a distinct position, a point of view, even an utterance or an event with further influence in the conversation. All preconditions are met by assuming that each utterance is read or heard, remembered and further discussed, therefore having an impact in the discourse (Trausan-Matu & Rebedea, 2009). Moreover, a voice may be expressed as a perspective or topic (Linell, 2009) of a singular participant or of a group sharing a similar insight on the topical domain. With regards to a single individual, he may adhere, personalize and express several different voices by interacting with other people based on his formal background, education and attitude towards the topic at hand. Therefore, besides internal voices embedding personal perspectives and external voices uttered by other individuals and expressing the influence of others on ones opinion, generalized voices emerge to which a larger group of people consent.

In order to benefit mostly from collaboration, the main goal of the discussion can be defined in terms of voice inter-animation and the aim becomes achieving true polyphony (Bakhtin, 1993). Polyphony is closely related to the musical concept from which it was derived and encapsulates multiple points of view and voices. Dostoevsky's work presents conflicting views, not just various angles and multiple perspectives, not just a single, all-knowing and overwhelming vision common among most writers; all these aspects should also be covered in a truly collaborative conversation.

Echoes

A context is a slice of a discussion thread characterized by high internal cohesion and rather loose coupling with other parts of the conversation. A central voice emerges from a context, brings cognitive and creative significance and by its evolution in time models the unfinalized potential of that specific context (Bakhtin, 1986). The relation is bi-univocal in the sense that a context can encapsulate multiple voices and by merging all perspectives, the context can be defined.

The echo of a specific voice represents its replication in time with enough strength to influence other voices in one or more contexts. Two types of echoes can be identified: individual ones when a participant internalizes a voice and collective echoes when multiple participants react to a voice, enriching the context.

After analyzing all core concepts, two major effects were identified and taken into consideration in our analysis. Firstly, a retrospective, synergic effect, based on overlapping voices from previous utterances and their corresponding echoes, models and influences the current utterance in a given context. Secondly, a prospective effect expresses further implications in the discussion thread with regard to own personal echoes and models the context, highlighting the unfinalizable, dialogic nature of the discussion.

Summing all previous remarks, we can conclude that collaboration is based on voice intertwining and inter-animation and that one of the purposes of our system is to highlight and assess interaction between

participants in a collaborative environment. The next section addresses the technological aspects and the actual implementation of some of the previous concepts centered on the assessment process of utterances.

Technologies and Computational Perspective

Technically, PolyCAFe combines Natural Language Processing (NLP) and Social Network Analysis (SNA). Its main tasks are implicit link detection starting from patterns, repetitions and semantic distances based on WordNet and LSA (Trausan & Rebedea, 2010), utterance evaluation and collaboration analysis based on the utterance graph and the scores for the utterances. As results, the system provides feedback on several distinct levels: for each utterance in the conversation, for each participant and for the conversation as a whole.

The presentation of the feedback to the users is done in simple web widgets that can be used independently or together and can be integrated into most online learning environments (Rebedea et al., 2010). There are widgets for each level of feedback discussed above, plus two helper widgets:

- The *conversation feedback widget* presents statistics about the whole chat: the most frequent synsets, the most relevant concepts that are present in the chat and in the domain specific corpora used for training the LSA, a suggestion of concepts from the semantic space that are semantically similar to the ones discussed in the chat (that might be seen as the concepts that are in the zone of proximal development for the current conversation) and statistics regarding the density of the utterance graph, percent of personal opinions and argument, etc.
- The *participant feedback* offers assessment for each participant on several levels: relevance with regards to the domain corpora, social presence and importance, coherence, etc.
- The *utterance feedback* gives indicators on the value of each post: speech acts and argumentation patterns that are present in the utterance, plus a social and semantic score.
- The *conversation visualization* is a helper widget that contains an intuitive display of the utterance graph and the collaboration graphics as shown in Figure 1.
- The *search conversation widget* provides a mechanism for ranking utterances and participants with regards to a search query and by taking into consideration not just the lexical items, but also the semantic relations and the importance of each utterance as considered by the utterance evaluation process.

Validation Experiment and Results

A first validation experiment has been performed at a Human Computer Interaction course, involving 9 senior (4th year) students and 5 tutors that used PolyCAFe for analyzing the conversations and providing feedback to the students. The experiment was structured in the following way: the students had to document on a given topic using online and printed materials and then they had a debate using VMT chat system in two small groups of 4-5 students. After the debate, they used PolyCAFe's widgets to understand their actual role in the conversation and what could have been improved. This activity was monitored by two tutors that provided help to the students and took notes on how they used the widgets. This activity lasted between 90-120 minutes and was followed by a questionnaire with 32 validation statements with answers on a 5-level Likert scale (1-strongly disagree, to 5-strongly agree). The tutor validation experiment followed a similar scheme. Tutors were asked to provide feedback to a chat conversation using PolyCAFe and to the another without the system. After this step, they were invited to answer a questionnaire with 35 validation statements.

Both validation experiments have been promising, with very encouraging results as can be seen in Table 1. For a better understanding of the collected results, the statements have been divided into five categories: pedagogical effectiveness, efficiency, cognitive load, usability and satisfaction.

Table 1: The validation results for the tutors and student experiments.

Validation statement category	Tutor average score	Tutor percentage agreement	Student average score	Student percentage agreement
Pedagogic effectiveness	4.11	83%	3.94	77%
Efficiency	5.00	100%	4.22	78%
Cognitive load	4.60	100%	3.56	56%
Usability	4.36	93%	4.11	81%
Satisfaction	4.57	91%	3.89	72%
TOTAL	4.53	93%	3.94	73%

For the questionnaires, the tutors have agreed with all but one of the statements with average scores between 3.50-5.00/5.00. It is clear that all the tutors find the system efficient for their task as it helps them

reduce the time needed for providing their own feedback to the students. It is easily noticeable that the students' results are worse for all categories than the ones of the tutors. The lowest score was obtained for cognitive load showing that the users had some problems accommodating to the system on their first use. This high cognitive load might have influenced the results for all other categories. As in the tutor experiment, the highest average score was also reached for effectiveness showing that the system provides feedback that helps them understand better the characteristics of their conversation. However, the lower scores obtained for effectiveness and satisfaction prove that the feedback is not easily interpretable by the students, enabling them to improve their future learning activities. Moreover, the results show that more than a quarter of learners are not satisfied by the system and the main reason for this score was that the students do not trust the results offered by the system as it employs statistical methods and it has also provided some wrong indicators for a few utterances in their conversations. Considering each validation statement in particular, the students agreed with 27 out of the 32 statements with average scores between 3.56-5.00. As it can be noted from the results presented above, there are some serious differences between the results for students and those of the tutors. Another conclusion that can be extracted from these facts is that tutors might have overrated the system due to the fact that it helps them provide feedback more quickly and effectively.

Conclusions and Transferability

In this paper we have introduced PolyCAFe, a system designed for providing feedback and support for learners that use online discussions in their learning activities. Starting from the theoretical underpinning of dialogic and polyphonic theories, the system uses NLP and SNA processing in order to discover implicit relations between utterances and builds an utterance graph that is then used for utterance evaluation and collaboration assessment.

The system has also been used in a formal education context for validating its utility and the validity of the feedback provided to users. The results of the first validation experiment are encouraging, but highlight that tutors are considering PolyCAFe more useful and relevant for their task than the students. The transferability of the system involves three aspects: domain, language and learning task. The system needs all the tools that form the NLP pipe, plus an open-data WordNet or, at least, a dictionary. The domain transferability is mostly concerned with the existence of a serious corpus for training the latent semantic spaces (preferably, made up of online conversations or titles plus abstracts for forums). Any learning task that can be easily expressed in writing and involves open argumentations or problem solving is suitable for being analyzed by PolyCAFe. However, the system is not designed for tasks that make use of scripted collaborations.

References

- Avouris, N., Fiotakis, G., Kahrmanis, G., & Margaritis, M. (2007). Beyond Logging of Fingertip Actions: Analysis of Collaborative Learning Using Multiple Sources of Data. *Journal of Interactive Learning Research*, 18, 231–250.
- Bakhtin, M.M. (1986). Speech genres and other late essays, *University of Texas*, Austin.
- Bakhtin, M.M. (1993). Problems of Dostoevsky's Poetics, Edited and translated by Caryl Emerson, *University of Minnesota Press*, Minneapolis.
- Dascalu, M., Rebedea, T. & Trausan-Matu, S. (2010). A Deep Insight in Chat Analysis: Collaboration, Evolution and Evaluation, Summarization and Search, *AIMSA 2010*, LNAI 6304, 191-200, Springer.
- Dascalu, M., Trausan-Matu & S. Dessus, Ph. (2010). Utterances Assessment and Summarization in Chat Conversations, *Special Issue on Natural Language Processing and its Applications*, Mexico, 46.
- Dong, A. (2005). The latent semantic approach to studying design team communication, *Design Studies*, 26(5): 445-461.
- Dowell, J., & Gladisch, T. (2007). Design of argument diagramming for case-based group learning. *ACM International Conference Proceeding Series; Vol. 250*, 99–105.
- Dyke, G., Lund, K., & Girardot, J.-J. (2009). Tatiana: an environment to support the CSCL analysis process. *Computer Support for Collaborative Learning*, 58–67.
- Harrer, A., Hever, R., & Ziebarth, S. (2007). Empowering researchers to detect interaction patterns in e-collaboration. *Frontiers in Artificial Intelligence and Applications; Vol. 158*, 503–510.
- Hmelo-Silver, C. E., Chernobilsky, E., & Masto, O. (2006). *Representations for Analyzing Tool-mediated Collaborative Learning*. Proceedings of ICLS 2006, New Brunswick.
- Koschmann, T. (1999). *Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration*. In CSCL'99, Palo Alto.
- Kumar, R., Chaudhuri, S., Howley, I., & Rosé, C. P. (2009). VMT-Basilica: an environment for rapid prototyping of collaborative learning environments with dynamic support. *Computer Support for Collaborative Learning*, 192–194.
- Law, N., Lu, J., Leng, J., Yuen, J., & Lai, M. (2008). *Understanding Knowledge Building from Multiple Perspectives*. Proceedings of Workshop on Interaction Analysis, ICLS 2008.
- Linell, P. (2009). Rethinking language, mind, and world dialogically, *Information Age Publications*, Greenwich.

- Rebedea, T., Dascalu, M., Trausan-Matu, S., Banica, D., Gartner, A., Chiru, C. & Mihaila, D. (2010). *Overview and preliminary results of using PolyCAFe for collaboration analysis and feedback generation*. Proceedings of EC-TEL 2010, LNCS 6383, 420-425, Springer.
- Stahl, G.: (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge*, MIT Press.
- Stahl, G. (Ed.). (2009) *Studying Virtual Math Teams*, Springer, Boston.
- Suthers, D., Dwyer, N., Medina, R., & Vatrpu, R. (2007). *A framework for eclectic analysis of collaborative interaction*. Proceedings of Computer Support for Collaborative Learning 2007.
- Teplövs, C. (2008). *The Knowledge Space Visualizer: A Tool for Visualizing Online Discourse*. Proceedings of the International Conference of the Learning Sciences 2008.
- Trausan-Matu, S., & Rebedea, T. (2009). Polyphonic inter-animation of voices in VMT. In Stahl, G. (ed.), *Studying virtual math teams*, pp. 451-473, Springer, New York.
- Trausan-Matu, S., & Rebedea, T. (2010). A Polyphonic Model and System for Inter-animation Analysis in Chat Conversations with Multiple Participants. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing*, Vol. 6008, pp. 354-363, Springer.
- Toulmin, S. (1958). *The Uses of Arguments*: Cambridge Univ. Press.

Acknowledgments

The research presented in this paper was partially funded by the projects no. 212578, FP7-ICT LTfLL (<http://www.ltfll-project.org/>) and project no. 264207, FP7-REGPOT-2010-1 ERRIC (<http://erric.eu>).