Computers in Human Behavior 47 (2015) 119–127

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Finding student-centered open learning environments on the internet: Automated dialogue assessment in academic virtual communities of practice

Nicolae Nistor [a,b,c,*], Ştefan Trăuşan-Matu [d], Mihai Dascălu [d], Heather Duttweiler [a], Costin Chiru [d], Beate Baltes [c], George Smeaton [c]

[a] Ludwig-Maximilians-Universität, Faculty of Psychology and Educational Sciences, Leopoldstr. 13, D-80802 München, Germany
[b] Universität der Bundeswehr München, Faculty of Human Sciences, Werner-Heisenberg-Weg 39, D-85577 Neubiberg, Germany
[c] Walden University, Richard W. Riley College of Education and Leadership, 100 Washington Avenue South, Suite 900, Minneapolis, MN 55401, USA
[d] University "Politehnica" Bucharest, Faculty of Automatic Control and Computers, Splaiul Independenţei 313, RO-60042 Bucureşti, Romania

ARTICLE INFO

Article history:
Available online 1 August 2014

Keywords:
Virtual communities of practice
Open-ended learning environments
Natural language processing
Dialogue analysis
Polyphony
Inter-animation

ABSTRACT

Starting from the socio-constructivist concepts of (virtual) community of practice (vCoP) and internet-based argumentative open-ended learning environments, this study proposes and validates two tools for automated dialogue assessment, *ReaderBench* and *Important Moments*, developed on the ground of the polyphonic social knowledge building model. The analyzed corpus was the dialogue produced by an academic vCoP with $N = 179$ community members in 23 months, and consisting of 3685 interventions in 292 text-based discussion threads. The analysis results uncovered significant differences in the discussion threads produced by central and peripheral participants, such that central participants produced more interventions with higher collaborative dialogue quality, and the discussion threads they initiated were longer and involved a larger number of participants. Moreover, based on the automated analysis result, the vCoP participants could be classified in two clusters corresponding to the well-known core-periphery structure of CoPs. These findings are consistent with those revealed by other methods, and suggest that the employed tools are appropriate for identifying virtual communities that are appropriate as open-ended learning environments. Further research and development is needed to deepen quantitative vCoP models and test communication strategies recommended to students in vCoP-based argumentative open-ended learning environments.

1. Introduction

One of the most prominent affordances of educational technologies is that they support authentic learning environments, thus situating learning tasks in the context of real-world situations (Herrington, Reeves, & Oliver, 2014). In particular, student-centered, open learning environments (SCOLEs) are a special case of authentic learning environments in which the individual learner determines his or her learning goals, learning means of both. SCOLEs thus support students' engagement in complex, often ill-structured, open-ended problems (Hannafin, Hill, Land, & Lee, 2014). Internet-based argumentative SCOLEs can be implemented as virtual communities of practice (vCoP), where students may participate in the community discourse and apply domain-specific knowledge and argumentation strategies (Hannafin et al., 2014; Nistor, Schworm, & Werner, 2012). Such approaches are designed to support individual student sense-making using technology tools, resources and scaffolding (Hannafin et al., 2014). Participation in vCoPs fosters intrinsic motivation and the construction of applicable knowledge and skills (see overview in Nistor, 2010), while the construction of formally sound arguments is positively related to deep cognitive elaboration (Stegmann, Wecker, Weinberger, & Fischer, 2012).

In SCOLEs, scaffolding enables participation and mainly includes navigation guidance (Hannafin et al., 2014), which includes instructors' selecting online communities according to the intended discussions. Therefore, there is a need for tools that are capable of searching the internet for online communities (e.g., in discussion forums, blogs, social networks, etc.), assessing

* Corresponding author.
E-mail addresses: nic.nistor@uni-muenchen.de (N. Nistor), stefan.trausan@cs.pub.ro (Ş. Trăuşan-Matu), mihai.dascalu@cs.pub.ro (M. Dascălu), h.duttweiler@campus.lmu.de (H. Duttweiler), costin.chiru@cs.pub.ro (C. Chiru), beate.baltes@waldenu.edu (B. Baltes), george.smeaton@waldenu.edu (G. Smeaton).

http://dx.doi.org/10.1016/j.chb.2014.07.029
0747-5632/

the quality of the community discourse, deciding which community is appropriate for given learning goals and objectives, and recommending the community as a learning environment. Within this framework and based on socio-constructivist approaches (Bakhtin, 1981; Hannafin et al., 2014; Koschmann, 1999; Lave & Wenger, 1991), this study describes and validates two learning analytics tools for automated dialogue assessment in vCoPs.

2. Theoretical background

2.1. Discourse in virtual communities

Communities of practice (CoP) are groups of mutually engaged people sharing goals and practice while constructing appropriate shared knowledge and "ways of doing things" (Lave & Wenger, 1991; Wenger, 1999). The latter includes a communication style and discourse that are characteristic of a given community, and that are tightly related to participation patterns and member identity in that community.

Community members differ from each other mainly by their knowledge. The various grades of expertise can be placed on a continuum from novices to experts, and in between these endpoints are the so-called regular members (Kim, 2000; Lave & Wenger, 1991). Furthermore, community members vary in their "community age", from newcomers to senior members. Depending on how intensively a community member is involved in the ongoing activities, his/her participation can be regarded as peripheral or central. This dimension can be also represented by a continuum. Minimal participation is encountered in the cases where a visitor observes the ongoing activity and incidentally interacts with the community members. Peripheral participation allows or even demands observing the activity of more skilled and experienced members. In contrast, central participation involves intensive and challenging activities (Nistor & Fischer, 2012), and decision-making in the most difficult cases. Typically, the participation of a novice is peripheral, while the participation of an expert is more central. Moreover, participation can change over time. Ideally, visitors may enter the community and become novices, regulars and experts (Eberle, Stegmann, & Fischer, in press). At the same time, their participation may increase in terms of intensity and importance, and change from peripheral to central. Additionally, this identity development is correlated with the process of learning and personality development of the community member (Wenger, 1999).

Wenger (1999) describes knowledge construction as the interplay of participation and reification. Through participation, CoP members gain experience and construct knowledge that they reify, i.e., convert into artifacts. On the other hand, artifacts enable further participation. As Nistor and Fischer (2012) demonstrate, CoP members' contribution to artifact development is correlated with their expert status.

Socio-cognitive activity that may be regarded as typical for CoP practice was described by Dewey (1998) as a process consisting of five steps: identifying a problem, exploring the problem, suggesting a solution, judging the solution, and implementing the solution. Dewey referred to this as "critical thinking". Relying on his definition, Weltzer-Ward, Baltes, and Lynn (2009) propose and validate a critical thinking assessment framework (CTAF) as an instrument for content analysis that can be applied to text-based discussions on the internet. This analysis can be performed by human coders and includes four basic dialogue structures: (1) questions that pose a problem or expect an answer within the context of the discussion, (2) claims that present an idea with the intention of furthering discussion, asserting something new, supporting or disagreeing with claims, synthesizing previous claims, or replying to a question, (3) evidence supporting a claim in form of quotations, paraphrased references, personal experience of the discussion participant, opinions, examples, experimental data, or generally accepted theories, and (4) relations describing links between claims and evidences.

2.2. The polyphonic social knowledge building model

Knowledge building in joint discourse is one of the main themes of Computer Supported Collaborative Learning (CSCL) (Koschmann, 1999; Stahl, 2006). In computational linguistics, discourse is usually defined as a coherent group of sentences (Jurafsky & Martin, 2009). The authors of this study share this definition, and extend it in a socio-cultural direction, in line with the dialogistic view of Bakhtin (1981), who considers that discourse analysis should not focus on sentences but on socially-produced utterances (e.g., replies to discussions in form of words, sentences, documents, essays, and even whole books) and on their polyphonic inter-animation (Trăuşan-Matu & Rebedea, 2009).

Bakhtin started from the polyphonic weaving in the musical domain and extended it to analyzing literary discourse, considering that "the voices of others become woven into what we say, write and think" (Bakhtin, 1981; Koschmann, 1999). In this context, a voice is meant not only as the physical, vocal expression of a participant, but rather like a distinct act, a position taken by one or more of the participants and discussed throughout the conversation. Thus, every voice influences the subsequent evolution of the conversation (Trăuşan-Matu, 2010).

Bakhtin's (1981, 1984) dialogism and polyphonic model of discourse can be used to explain and analyze the inter-animation processes that drive knowledge building in collaborative learning sessions. In this vein, a polyphonic model and an associated analysis method were developed (Trăuşan-Matu, 2010) and implemented in several tools. Two of these tools are presented and used in the following.

2.3. The "ReaderBench" tool: automated dialogue analysis based on the polyphonic and social networks perspective

One of the applications of the polyphonic perspective on collaboration aims at voice identification in discourse analysis. First, a *cohesion graph* (Dascălu, Dessus, Trăuşan-Matu, Bianco, and Nardy, 2013; Trăuşan-Matu, Dascalu, & Dessus, 2012) is built, in which utterances are represented as nodes at different levels: the entire discussion thread, posts or interventions, sentences or chat replies (Fig. 1). The edges are cohesive links between utterances, identified considering adjacency pairs, repetitions, lexical chains, speech and argumentation acts (Jurafsky & Martin, 2009), built using a natural language processing pipeline (Manning & Schütze, 1999). The next step is the extraction of topics as key concepts with their corresponding relevance scores reflecting the following factors: (a) *statistical presence* based on information retrieval metrics (term frequency; Manning, Raghavan, & Schütze, 2008), (b) *semantic relatedness* to the entire document, and (c) *overall coverage* and *linkage* with the automatically generated lexical chains (Galley & McKeown, 2003). The initial individual assessment of each analysis element (node in the cohesion graph) is based on its topic coverage and corresponding relevance with respect to the entire forum thread (Dascălu, 2014). Fig. 2 presents the main user interface of the *ReaderBench* tool, employed in the following to process forum discussion threads. The intervention importance scores (reflecting proportionally the presence of topics-key concepts) are displayed in square brackets after each utterance. The automatically identified topics for the thread or for specific participants are presented in the right sidebar.

Building on the cohesion graph and the intervention importance scores, the modeling of the participant interaction covers a deeper qualitative dimension. A second graph is built to model