

Virtual Communities of Practice in Academia: Automated Analysis of Collaboration Based on the Social Knowledge-Building Model

Nicolae Nistor^{1,2,*}, Mihai Dascalu³, Stefan Trausan-Matu³,
Dan Mihaila³, Beate Baltes², and George Smeaton²

¹ Ludwig-Maximilians-Universität, Faculty of Psychology and Educational Sciences
Leopoldstr. 13, D-80802 München, Germany

nic.nistor@uni-muenchen.de

² Walden University, Richard W. Riley College of Education and Leadership
100 Washington Avenue South, Suite 900, Minneapolis, MN 55401, USA
{beate.baltes, george.smeaton}@waldenu.edu

³ University Politehnica of Bucharest, Faculty of Automatic Control and Computers
Splaiul Independentei 313, RO-60042 Bucuresti, Romania

{mihai.dascalu, stefan.trausan, dan.mihaila}@cs.pub.ro

1 Rationale

Participation in communities of practice (CoPs) [1], either in face-to-face, or in technology-based environments, leads to the accumulation of experience, stimulates the social construction of knowledge and the development of expertise. An overview over the occurring complex long-term processes may provide means to foster collaboration and social knowledge construction in CoP. In technology-based environments, however, this requires valid and reliable automatic analysis procedures of the community dialogue. In previous work [2], social knowledge-building was analyzed as a cohesion-based collaboration assessment model applied to individual chats. This study extends the procedure after aggregating discussion threads from asynchronous discussion forums and performs a validation by comparison with the results of a manual content analysis based on a critical thinking framework [3]. The developed tool makes headway on applying learning analytics in field research on virtual CoPs.

2 Tool Development

For the automated analysis of dialogue in asynchronous, text-based discussion forums, the messages were counted along with their authors' identifiers, then social network analysis was applied to extract betweenness centrality. As a representation of the underlying discourse, a cohesion graph was built for each forum discussion thread [2]. This was a multilayered graph with three types of nodes: (1) a central node, the overall discussion thread, (2) participant interventions and (3) sentences, the main

* Corresponding author.

units of analysis as forum posts usually integrate multiple phrases, on which the initial natural language processing pipe was applied [2]. As specificity of the previous graph, the cohesion expressing the strength of the links was viewed in terms of: (1) the *distance* between interventions; (2) *lexical proximity* reflected in identical lemmas and semantic distances within WordNet; (3) *semantic similarity* from Latent Semantic Analysis (LSA) vector spaces and Latent Dirichlet Allocation (LDA) topic models [2]. Links within the graph were established through the explicit referencing facility made available in the forum environment (the “reply to” links) or as automatically identified cohesive links. Afterwards, each intervention was automatically scored and the values are later on cumulated per vCoP member. The *importance score* integrates both a quantitative and a qualitative dimension. From a quantitative perspective, different Information Retrieval techniques were used, whereas quality was reflected in the topics coverage and in the cohesive links that were used to augment the local importance of each intervention. Eventually, all discussion threads were automatically aggregated, and a global social network [2] was built with all the involved vCoP members. On the other hand, *collaboration* as a complement to *participation* is assessed as a social knowledge-building effect reflected in the information transfer between different participants, by using the previous utterance scores and the corresponding cohesive links [2].

3 Tool Validation

The validation study was conducted in an academic vCoP emerging at a US American online university. The entire vCoP consisted of $N = \text{approx. } 500$ faculty members. During 23 months, they made a total of 7370 interventions. Applying the automated content analysis procedure, the overall social knowledge-building scores varied between 0 and 5927 ($M = 229.83$, $SD = 576.36$). To validate the automated analysis, a sample of 414 interventions made by $n = 15$ vCoP participants was manually analyzed applying the critical thinking framework developed and validated for assessing critical thinking in online discussions by Weltzer-Ward, Baltes and Knight-Lynn [3]. A strong bivariate correlation of $r = .72$ ($p < .001$) was found between the manually determined messages’ relevance and the automatically determined importance scores, which suggests that the automated dialogue analysis provides results similar to the manual analysis based on the critical thinking framework.

Acknowledgment. This research was partially supported by the 264207 ERIC–Empowering Romanian Research on Intelligent Information Technologies/FP7-REGPOT-2010-1 project.

References

1. Lave, J., Wenger, E.: *Situated learning: Legitimate peripheral participation*. Cambridge University Press, Cambridge (1991)
2. Dascalu, M., Trausan-Matu, S., Dessus, P.: Cohesion-based analysis of CSCL conversations: Holistic and individual perspectives. In: CSCL 2013, Madison, USA (2013)
3. Weltzer-Ward, L., Baltes, B., Knight Lynn, L.: Assessing quality of critical thought in online discussion. *Campus-Wide Information Systems* 26(3), 168–177 (2009)