2011 11th IEEE International Conference on Advanced Learning Technologies

# Learning Analytics for Communities of Lifelong Learners: a Forum Case

Zinayda Petrushyna, Milos Kravcik, Ralf Klamma
Informatik 5
RWTH University
Aachen, Germany
e-mail: {petrushyna, kravcik, klamma}@dbis.rwth-aachen.de

**This paper describes an experiment investigating interactions in a big forum in order to support students in learning English. Groups of collaborating users form communities that generate a lot of data that can be analyzed. We distinguish different phases of the self-regulated learning process and aim to identify them in learners' activities. Then we attempt to recognize patterns of their behavior and consequently their roles in a community. Based on this analysis we try to explain a success or failure of a community. We conclude that heterogeneity of members helps a learning community to function.**

***Keywords: learning phase, community of practice, learning analytics.***

## I. INTRODUCTION

Scholars have numerous definitions for communities but here we follow the Community of Practice (CoP) approach [4], according to which it is possible to define not only community members but also their common understanding and interests. Even if the community borders are defined in this way, the factors of community survival are not trivial. They may include members, technologies, topics, environment conditions, etc.

Communities of lifelong learners depend on self-regulated learning [1] of their members. It means that they need meta-cognitive skills to organize their learning process: planning of learning activities, performing them, and reflecting on their achievements. For this purpose a higher level of flexibility and control is required, which can be provided by so called Personal Learning Environment (PLE) [2]. The ROLE project (www.role-project.eu) investigates Responsive Open Learning Environments that permit personalization of the entire learning environment and its functionalities, i.e. individualization of its components and their adjustment or replacement by alternative solutions. A crucial feature of these environments is the visualization of the learner's performance and progress to support the reflection phase. This is based on various learning analytics that join different data-gathering tools and analytic techniques.

Our challenge is to clarify how some communities manage to function for a long time and what is the reason of their success? We consider the dimensions mentioned in [4] to analyze the communities: positions of the members, their intentions, opinions, sentiments, and emotions.

In the next paragraph we shortly explain the experiment with the Urch forums that are part of a PLE for a number of students preparing for language tests (TOEFL, GMAT, GRE and others). Based on the data gathered from these forums we have analyzed the communities organized within them and present our findings.

## II. DATA SET

In our experiment we used Urch Forums, formerly called TestMagic. It has over 120,000 threads, over 800,000 posts and over 115,000 members. To analyze the data we used Social Network Analysis, Machine Learning, and Natural Language Processing approaches. We focused on the most relevant sub-forums, namely "Test preparation" and "English tests", which decrease our dataset to 428,514 posts in 67,421 threads (with an average of 6.36 posts per thread) created by 20,942 users. Further, we categorize threads activities based on Figures 1 and 2. The curves show the distribution of posts and users over threads. The posts from October 25th, 2001 to October 27th, 2010 are resulting in a time span of nine years.

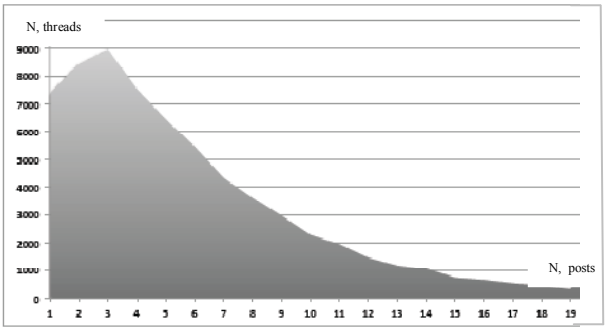


Figure 1. Number of threads which have x posts

Interested in observing the life of communities in the Urch forums, we used a naïve approach to find communities in the forum by looking for cliques, i.e. groups of members tightly connected with each other over a period of time. Based on the data statistics we identify a clique as a group of learners, that have at least 4 users, which occur several times (at least 10) in the whole set of threads. If at least 75% of users of a clique post in a thread, the thread belongs to the clique. The search continues until all threads and users belonging to the clique are defined. As a result, the 12,881 cliques were defined with an average size of 5 users and in

978-0-7695-4346-8/11 $26.00 © 2011 IEEE
DOI 10.1109/ICALT.2011.185

average 14 threads. The active phase of a clique takes in average about 7 weeks.

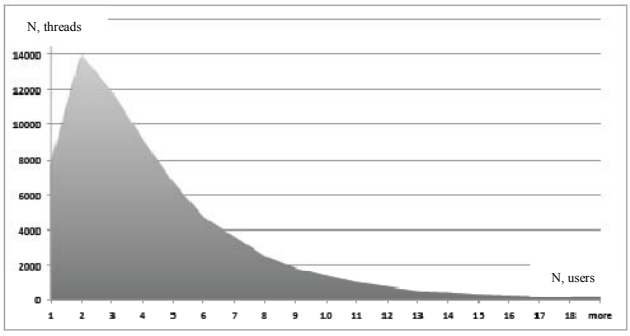


Figure 2. Number of threads which have x users

## III. EXPERIMENTAL SET UP

In our experiment we were interested in defining different learning phases of learners: planning phase (when the sentiment of learners' statement is neutral), learning phase (when learners are involved in longer discussions), and reflecting phase (when learners express emotions talking about their results). We consider various user characteristics devoted to learning activity (starting threads, giving comments, writing posts), help-seeking actions (starting a thread with an expression of particular intents), sentiment characteristics (attitudes toward discussions, communities and community members), and cognitive mechanism characteristics (expressions that will not appear in off-topic or small talk discussions). We tried to identify the goals of learners (intent analysis) and their expressions (sentiment analysis).

Using the text processing toolkit LingPipe (alias-i.com/lingpipe) and the LIWC (www.liwc.net) dictionary, threads are classified according to sentiments, cognitions and intent existence. The classification is performed using Dynamic Language Model Classifier of the LingPipe. The training set for the classifier is near 10% of all examined threads. The verbs like "know", "solve", "analyse" and others appear frequently than other verbs in phrases that are expressing goals.

From our own developments we used the Graphservice tool that detects patterns based on Social Network Analysis. It is based on the PALADIN approach [3] and provides the information about patterns of users in cliques. The Extraction-Transformation-Load service was used for extracting named entities, i.e. people, places, facts and events, out of content. The OpenCalais (www.opencalais.com) application programming interface was used for recognizing the entities in the thread content. After all the steps for data analysis had been performed, the information about each thread was stored in the RDF format, which enables its reuse in future research.

## IV. RESULTS

The evaluation of the experiment was done with the help of 18 volunteers, which were Urch users. The emotional classifier found 18 emotional posts belonging to these users, but only 12 were actually emotional according to users' opinions. The evaluation shows that both the emotional and cognition classifiers succeed in long posts classification but need some refinements in short post classification.

Nevertheless, based on our analysis and considering inaccuracy shown in the evaluation, we can find some interesting facts related to a self-regulated learning circle in a community. The observations of many communities show the need of having different roles in communities in order to support learning processes. Communities include users that are looking for resources and intensively work on them. Such a learning phase can be identified based on user patterns (a questioner or a conversationalist) according to a few sentiment and cognition phrases in messages and some intentional phrases appearing more often than on average. Other users are supporting the learners by providing relevant resources and advices to achieve their learning progress. They are usually answering persons and express more cognition phrases in their posts than other users.

## V. CONCLUSIONS

In this paper we present an approach to exploiting learning analytics in lifelong learner communities, which uses network and data analysis methods. Such a framework is based on modeling and reflection support. In our case study we focused on goal and sentiment mining for self-regulated learners, in order to identify various learning phases. This also outlines a challenge for our future research and development – establishment of learning analytics dashboard and widget collections for learning communities.

## ACKNOWLEDGMENT

The research work described in this paper is partially funded through the ROLE Integrated Project (grant agreement no 231396) of the European Union.

## REFERENCES

[1] A. Nussbaumer, K. Fruhmann, D. Albert, P. Ferdinand, A. Kiefel, U. Kirschenmann, C. Ullrich, S. Wills, A. Mikroyannidis, B. von Der Heiden, M. Werkle, and A. Naeve, "ROLE Deliverable 6.1: Common psycho-pedagogical framework," pp. 1-72, 1.1 edition, 2009.

[2] M. Van Harmelen, "Personal Learning Environments," Proc. International Conference on Advanced Learning Technologies (ICALT), 2006, pp. 815-816.

[3] R. Klamma, M. Spaniol, and D. Denev, "PALADIN: A Pattern Based Approach to Knowledge Discovery in Digital Social Networks," Proc. 6th International Conference on Knowledge Management (I-KNOW 06), Springer, 2006, pp. 457-464.

[4] E. Wenger, Communities of Practice: Learning, Meaning, and Identity. Cambridge University Press, Cambridge, UK, 1998.