Dataset on an online collaborative learning situation in a computer networks course

Cristina Villa-Torrano
Universidad de Valladolid
cristina@gsic.uva.es

Pankaj Chejara
Tallinn University
pankajch@tlu.ee

Juan I. Asensio-Pérez
Universidad de Valladolid
juaase@tel.uva.es

Yannis Dimitriadis
Universidad de Valladolid
yannis@tel.uva.es

Miguel L. Bote-Lorenzo
Universidad de Valladolid
migbot@tel.uva.es

Alejandra Martínez-Monés
Universidad de Valladolid
amartine@infor.uva.es

Eduardo Gómez-Sánchez
Universidad de Valladolid
edugom@infor.uva.es

ABSTRACT
This paper presents a dataset of a collaborative learning situation. Students were enrolled in two undergradute courses on computer networks where they were required to carry out a set of learning activities supported by Moodle and an online collaborative environment called CoTrackV2. The data collected includes logs of the writing process of shared documents, logs of the chat messages between the group members, and logs from Moodle with coarser-grained information about course-level interactions. This dataset has been generated with the aim of allowing researchers to study self- and socially-shared regulation in online environments.

Keywords
Computer-Supported Collaborative Learning, Socially-Shared Regulation of Learning, Self-Regulated Learning

1. INTRODUCTION
Academic and work contexts are increasingly demanding the competence of being able to collaborate with peers [5] as one of the 21st Century Skills [6]. In order to have a successful collaboration, many studies show that it is necessary to develop regulatory processes where group members can activate and maintain their cognition, motivation, and emotion towards their common goals [4]. This need is also present in computer science and engineering courses. Moreover, the use of Information Communication and Technology (ICT) tools to support collaboration (leading to Computer-Supported Collaborative Learning settings (CSCL) [3]) enables the collection of traces to model students' behavior while collaborating.

The literature has shown that students' motivation and strategic regulation play a critical role in their success in Science, Technology, Engineering and Mathematics (STEM) courses [2]. For example, in [8] the authors studied how motivation, strategic self-regulation, and creative competency were associated with computational thinking knowledge and skills in introductory computer science courses. They found that student performance and long-term retention were positively correlated with the use of self-regulated strategies. Concerning the motivation, higher pursuit of goals, and positive affect were also correlated with high performers, higher knowledge retention, strategic self-regulation and engagement. Moreover, collaborative activities, especially those including CSCL tools, have been shown to favor knowledge building [7] and also can benefit from socially-shared regulation in order to be successful. However, although there are studies that show that it is necessary to develop regulatory processes while collaborating, further study is needed in STEM courses and, specifically, in computer science courses.

Concerning the latter, we have not found shared datasets enabling the study of regulation in collaborative activities in computer science. This is a challenging issue, because the study of regulation in collaborative learning settings requires the collection and integration of a variety of data sources like, for example, logs of different learning platforms, the communication between group members, and self-reported data. The absence of such a dataset led us to the need of generating one of them. Computer networks are part of the ACM Computer Curricula [1], and we had access to two courses on this topic. Therefore, we generated one dataset related to a learning situation on this subject, designed to fulfill the aforementioned requirements. Further details will be provided in the following sections.

2. CONTEXT AND DATASET

2.1 Description of the learning situation

Table 1: Attributes provided in the document_logs.csv file

Attribute	Description	Example
Timestamp	Timestamp of the action	`14:42:57 17-02-2021`
Author	Student ID	`a.I6ZFAmhSZ4KY2HU1`
Group	Group ID	`1`
Char_bank	Characters added during this action	`How many access points do you have throughout the hotel?`
Source_length	Length of the text before performing the action	`2352`
Operation	Type of operation (>: writing, <: deleting)	`>`
Difference	The difference in number of characters caused by the current action and source length	`56`
Text	Text from the document at the current time	`Have you contacted your Internet service provider, i.e. your operator? [...]`

Table 2: Attributes provided in the chat_logs.csv file

Attribute	Description	Example
Timestamp	Timestamp of the action	`22:15:18 16-02-2021`
Author	Student ID	`a.WCpdVcSKpEcVM13V`
Group	Group ID	`1`
Message	Text message	`OK, let's put a section of definitive questions at the end if you want.`

Table 3: Attributes provided in the table_moodle.csv file

Attribute	Description	Example
Timestamp	Timestamp of the action	`16/02/2021 22:40:00`
User ID	The user ID, it can represent a student or a teacher	`a.WCpdVcSKpEcVM13V`
User involved	If the teacher does an action it can involve other teachers or students. This attribute represents the user ID of the user involved.	`a.I6ZFAmhSZ4KY2HU1`
Event context	The section in Moodle in which the event occurred	`Course: TRAFFIC ENGINEERING IN TELEMATIC NETWORKS (1-211-460-45033-1-2020)`
Component	The type of resource in Moodle	`Questionnaire`
Event name	The name of the event	`Course module viewed`
Description	The description of the action	`The user with id 'a.WCpdVcSKpEcVM13' viewed the 'resource' activity with course module id '974963'`
Source	The source from where Moodle has been accessed and the action has been performed	`web`
IP Address	The IP address from which the action was performed	`83.58.29.136`

The learning situation took place at two undergraduate courses on Computer Networks during 4 days in the spring semester of the academic year 2021 in a European University. There were 33 students, that were grouped into 8 different groups of 4-5 people to carry out an introductory learning situation aimed at challenging their previous knowledge and beliefs about certain computer network topics. Before starting the learning situation, students were asked to fill out an informed consent.

The situation was designed following the so-called pyramid or snowfall pattern, where the students had to first carry out the proposed activities individually and then in groups (thus fostering the agreement among the group members in order to submit a common solution). The different activities were carried out during 4 two-hour face-to-face sessions.

The learning situation was based on the following scenario: A hotel owner (role played by the teacher) goes to a team of telco engineers (role played by the students) to ask them to solve his problem: the internet connection is not working properly; the internet access is very slow and sometimes does not work at all. The hotel owner and the telco engineers agree to an interview in a few days. In order for the telco engineers to think about the problem, the hotel owner sends them a diagram of the current network. The different activities that students needed to complete were:

- Questions_ind (individual): Thinking of questions to ask the hotel owner to find out more about his network.

- Questions_group (in groups of 4-5 students): Agreeing on 7 final questions to ask the hotel owner.

- Questions_class (whole class): Asking the hotel owner about his network. For this task, there was a spokesperson in each group. The teacher, playing the role of the hotel owner, answered those questions posed by the groups.

- Diagnosis_ind (individual): Proposing a solution to the hotel's Internet access problem.

- Diagnosis_group (in groups of 4-5 students): Agreeing on a final proposal with the rest of the group members.

- Diagnosis_class (whole class): Creating a concept map of the technical concepts that emerged during the whole situation.

Students had to work through an online collaborative environment called CoTrackV2[1]. This environment offered the possibility to write documents collaboratively and had a built-in chat so that the different members of the group could communicate. In addition, students used Moodle to submit individual assignments, to visit subject-related content and to access the link to the CoTrackV2 sessions, so we were able to obtain traces of the content visited by the students, the writing process and the chat messages. Besides these traces, at the end of the learning situation, the students answered a questionnaire related to group regulation.

2.2 Description of the dataset

The dataset collected[2] is based on the logs of two different tools: CoTrackV2 and Moodle. The data obtained by CoTrackV2 is divided into 2 files: 1) `document_logs.csv`, with actions from the writing process of the shared documents; and 2) `chat_logs.csv`, that contains the logs of the communication between the group members. The attributes in the two files are presented in Table 1, and Table 2, respectively. Regarding the data obtained through Moodle, we have 2 files: 1) `moodle_logs.csv`, with the logs of the contents visited by the students. Details are given in Table 3[3]; and 2) `individual_submissions.csv`, where the individual submissions for activities 1 and 4 are collected, containing the timestamp of the submission, the id of the student submitting the solution and the solution itself. Besides these files, we have two others: 1) a file containing the learning design, including the start time and the name of the tasks; and 2) a file containing the students' answers to the final questionnaire. All files provided have been properly anonymized.

[1]CotrackV2 website: https://www.cotrack.website/

[2]The dataset will be available at https://zenodo.org/record/5033198#.YNsQv-gzaUk

[3]The examples presented in the different tables have been translated into English for a better understanding, but the dataset is in Spanish.

3. ANALYSIS

The dataset we have generated may allow researchers to answer different research questions related to Socially-Shared Regulation of Learning (SSRL). For example, the research questions that have guided the design of this learning situation are the following: 1) How do self- and socially-shared regulation processes occur in groups that complete group activities with different levels of success?; 2) Are there different patterns of regulation associated with the performance of groups when solving activities? To answer these questions, we want to analyse the data from a temporal perspective using different techniques, like process mining (e.g.: Heuristic Miner or Fuzzy Miner algorithms), Markov models (e.g.: pMiner algorithm), social network analysis (temporal networks) and epistemic network analysis. Beforehand, we want to identify SSRL features that allow us to map low-level data to higher-level constructs. After that, we could make use of the techniques mentioned above and compare the results of the different approaches. Beyond detecting the different processes, we would like to build predictive models with the identified features. However, at this stage of the research, it would be very beneficial to get feedback from the community to better guide the analysis.

4. ACKNOWLEDGMENTS

This research is partially funded by the European Regional Development Fund and the Spanish National Research Agency under project grant TIN2017-85179-C3-2-R.

5. REFERENCES

[1] ACM and IEEE-CS. Computing Curricula 2020–CC 2020: Paradigms for Global Computing Education, A Computing Curricula Series Report, 2020.

[2] N. R. Council. *How students learn: History, mathematics, and science in the classroom.* National Academies Press, 2004.

[3] P. Dillenbourg, S. Järvelä, and F. Fischer. The evolution of research on computer-supported collaborative learning. In *Technology-enhanced learning*, pages 3–19. Springer, 2009.

[4] S. Järvelä, D. Gašević, T. Seppänen, M. Pechenizkiy, and P. A. Kirschner. Bridging learning sciences, machine learning and affective computing for understanding cognition and affect in collaborative learning. *British Journal of Educational Technology*, 51(6):2391–2406, 2020.

[5] J. Malmberg, S. Järvelä, H. Järvenoja, and E. Panadero. Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high-and low-performing groups. *Computers in Human Behavior*, 52:562–572, 2015.

[6] A. J. Rotherham and D. Willingham. 21st century. *Educational leadership*, 67(1):16–21, 2009.

[7] M. Scardamalia and C. Bereiter. Knowledge building. *The Cambridge*, 2006.

[8] D. F. Shell, M. P. Hazley, L.-K. Soh, E. Ingraham, and S. Ramsay. Associations of students' creativity, motivation, and self-regulation with learning and achievement in college computer science courses. In *2013 IEEE Frontiers in Education Conference (FIE)*, pages 1637–1643. IEEE, 2013.