User Modeling and User-Adapted Interaction manuscript No.
(will be inserted by the editor)

Data-Informed Design Parameters for Adaptive Collaborative Scripting in Across-Spaces Learning Situations

Ishari Amarasinghe · Davinia Hernández-Leo · Anders Jonsson

Received: date / Accepted: date

Abstract This study presents how predictive analytics can be used to inform the formulation of adaptive collaborative learning groups in the context of Computer Supported Collaborative Learning (CSCL) considering across-spaces learning situations. During the study we have collected data from different learning spaces which depicted both individual and collaborative learning activity engagement of students in two different learning contexts (namely the classroom learning and distance learning context) and attempted to predict individual students future collaborative learning activity participation in a pyramid-based collaborative learning activity using supervised machine learning techniques. We conducted experimental case studies in the classroom and in distance learning settings, in which real-time predictions of students future collaborative learning activity participation were used to formulate adaptive collaborative learner groups. Findings of the case studies showed that the data collected from across-spaces learning scenarios is informative when predicting future collaborative learning activity participation of students hence facilitating the formulation of adaptive collaborative group configurations that adapt to the activity participation differences of students in real-time. Limitations of the proposed approach and future research direction are illustrated.

Keywords Computer Supported Collaborative Learning (CSCL) · Adaptive Collaborative Scripting · Collaborative Learning Flow Patterns (CLFP) · Supervised Machine Learning · Prediction Algorithms

Ishari Amarasinghe (✉) · Davinia Hernández-Leo · Anders Jonsson
ICT Department, Universitat Pompeu Fabra, Roc Boronat, 138, 08018, Barcelona, Spain
Tel.: +34-65-8036558
E-mail: ishari.amarasinghe@upf.edu

Davinia Hernández-Leo
E-mail: davinia.hernandez-leo@upf.edu

Anders Jonsson
E-mail: anders.jonsson@upf.edu

This paper or a similar version is not currently under review by a journal or conference. This paper is void of plagiarism or self-plagiarism as defined by the Committee on Publication Ethics and Springer Guidelines.

1 Introduction

Technological advancements have caused a multiplicity of learning spaces, creating learning opportunities towards students beyond the physical classroom spaces defined by the formal educational context (Ellis and Goodyear, 2018; Kloos et al, 2012). With the increased availability of diverse digital learning spaces students learn, interact, share knowledge and engage in productive discussions with peers leaving behind a vast amount of digital data traces. Retrieving meaningful information combining trace data emerged from multiple sources is challenging and requires specialized knowledge, despite the fact that the analysis and interpretation of this data can provide meaningful insights to design and implement pedagogically meaningful learning activities in different learning spaces (Amarasinghe et al, 2017; Prieto et al, 2017; Martinez-Maldonado et al, 2017; Hernández-Leo et al, 2012; Tsovaltzi et al, 2015).

In the past few decades, Computer Supported Collaborative Learning (CSCL) emerged as a branch of the learning sciences, focusing on how people learn together with the help of computers (Stahl et al, 2006). In contrast to individual learning, CSCL is characterized by social learning phenomena, in which learning occurs socially through group interactions among students (Roschelle and Teasley, 1995). It has been shown that working in groups increase students' learning and pro-social attitudes while solving problems with others, agreeing or disagreeing to different points of views at the same time giving or by receiving help from peers (Fall et al, 2000). In CSCL social interactions among students are being effectively mediated using computers, facilitating synchronous or asynchronous learning in the classroom and distance learning environments. Nonetheless, interactions observed in such learning settings are much more complex than that of the individual learning (Cen et al, 2016) which makes it challenging to conduct fruitful collaborative learning activities in both synchronous and asynchronous modes of collaboration.

In the domain of collaborative learning, *scripts* aim to promote productive interactions among groups of learners by shaping the way they interact with each other (Dillenbourg and Tchounikine, 2007; Kobbe et al, 2007). Using different techniques (e.g., defining the activity sequence, role allocation etc.) scripts attempt to increase the probability of productive student-student and student-teacher learning interactions that would occur rarely or not at all in spontaneous collaboration (Dillenbourg and Tchounikine, 2007; Demetriadis and Karakostas, 2008; Kobbe et al, 2007). In CSCL, collaborative learning scripts have been operationalized using computers formulating CSCL scripts as it facilitates the mediation of collaboration (partly or totally) among distance

and co-present learners (Dillenbourg and Tchounikine, 2007; Demetriadis and Karakostas, 2008; Kobbe et al, 2007; Villasclaras-Fernández et al, 2009).

Nonetheless, research has shown that the static support provided by scripts is not responsive to what is occurring in the actual collaborative learning environment (Kumar et al, 2007). It has been argued that adaptive collaborative scripting, in which collaborative interactions are modeled as they occur (Walker et al, 2009) in an adaptive mode can considerably improve the collaborative learning experience (Demetriadis and Karakostas, 2008). When considering the across-spaces learning scenarios, adapted scripted collaboration becomes challenging since the actions of students in previous activities carried out in diverse spaces or with different technologies is relevant for the planning of following up activities in a new space (Hernández-Leo et al, 2012).

From a learning analytics perspective, fine-grained learning analytics techniques can be employed to interpret data captured across different learning spaces in different modalities (Martinez-Maldonado et al, 2017). Meaningful insights gained from learning analytics can be used to identify relevant adaptive script features hence facilitating the formulation of adaptive collaboration scripts in across-spaces learning situations in real-time (Amarasinghe et al, 2017). Towards this end, the focus of our work is on investigating how predictive analytics can support the formulation of adaptive collaborative scripts in cross-context learning situations. Predictive analytics is described as a subset of data science that facilitates to uncover relationships and patterns within large volumes of data that can be used to make predictions about future events (Waller and Fawcett, 2013; Nyce and Cpcu, 2007). Within this study, predictive analytics have been used to predict future collaborative learning activity participation of students, to facilitate the formulation of collaborative learner groups that adapt to the activity participation differences of students. We have collected data from different learning spaces and used supervised machine learning techniques for prediction purposes. The main research question addressed in this study is the following: Can participation prediction be used to inform decisions for adaptive collaborative scripts in across-spaces learning situations? The main research question composed of the following sub research questions: (i) How to use supervised machine learning techniques to predict future collaborative learning activity participation of students based on data collected from across-spaces learning situations? (ii) How an estimate of future collaborative learning activity participation of students can be incorporated into CSCL scripts in real-time to facilitate the formulation of adaptive collaborative learning scripts?

The rest of this paper is structured as follows. Section 2 presents relevant literature considering adaptive collaborative scripting, its association with across-spaces learning scenarios and different learning analytic techniques that have been deployed in previous studies to support collaborative learning. Section 3 illustrates the proposed approach, along with data collection methods in different learning contexts, feature generation and model selection in detail. Section 4 presents case studies that demonstrate the applicability of the suggested intervention in formulating adaptive collaborative scripts in real-world

collaborative learning sessions along with the lessons learned and the limitations of the proposed approach. The final section provides concluding remarks followed by future research directions.

2 Background

2.1 Adaptive collaboration scripts

CSCL scripts aim to facilitate productive interactions among distant or co-present learners as free collaboration fails often to trigger productive group interactions (Dillenbourg and Tchounikine, 2007). Scripts are based on the scripted cooperation approach and provide a method for structured collaboration which intends to achieve higher levels of cognitive processing and better learning outcomes (Demetriadis and Karakostas, 2008). Scripts provide instructions "for small groups of learners on what activities need to be executed, when and by whom they need to be executed in order to foster individual knowledge acquisition" (Weinberger et al, 2007). Many studies have reported the effectiveness of using collaborative scripts towards achieving benefits of collaboration (Rummel and Spada, 2007; Kollar et al, 2006).

Yet, at the same time, CSCL scripts have also been criticized for being overly constrained limiting its modifiability during the script runtime (Dillenbourg and Tchounikine, 2007). Lack of flexibility associated with CSCL scripts and potential risks of over-scripting collaboration has highlighted the requirement towards adaptive collaboration scripts that adjust script parameters during script execution (Demetriadis and Karakostas, 2008). As described in (Demetriadis and Karakostas, 2008) adaptive collaboration scripting "is the idea that collaboration scripts can be adapted during runtime in several of their aspects, to provide learning experiences tailored to individual and group characteristics". However, it is not possible to model any script feature as an adaptation. Intrinsic constraints that preserve the underlying pedagogy of a script are not considered as candidates for adaptation (Dillenbourg and Tchounikine, 2007). For instance, in a Jigsaw script, a constraint that specifies each Jigsaw group requires to consist at least one member from each expert group is an intrinsic constraint that is mandatory to be satisfied and cannot be modeled as an adaptive script parameter. On the other hand, extrinsic constraints are related to the contextual aspects that lead to a particular implementation of the pedagogy. As further illustrated in (Demetriadis and Karakostas, 2008) extrinsic constraints can be further divided into two categories namely "Non-pedagogical" and "Pedagogical" constraints and can be considered as candidates for adaptation. *Non-pedagogical* constraints (constraints that do not possess any pedagogical relevance) e.g., duration of a script phase, can be altered by teachers or students to better accommodate the script to the given learning situation while *Pedagogical* constraints (e.g., increasing the level of support given to avoid learners misconceptions) should be adapted to facilitate a better learning experience. CSCL scripting systems

that embed adaptive scripting techniques have been referred to as Adaptive Collaboration Scripting systems or ACS (Demetriadis and Karakostas, 2008). ACS have been reported to be more effective than non-adaptive collaborative learning systems as ACSs tailor the learning experience to the needs and characteristics of both individuals and learner groups maximizing the benefits from the scripted collaboration (Rummel et al, 2008).

Research has provided evidence that adaptive collaboration support provided in the form of prompts has a beneficial impact on student learning. In (Kumar et al, 2007) adaptive collaborative learning support has been deployed using tutorial dialogue agents. It has been found that the students who gained dynamic support in terms of adaptive prompts have benefited significantly from collaboration when compared to the no support condition. (Walker et al, 2014) have built an ACS to support peer tutoring in high school algebra. The adaptive support has been built into the system (in terms of reflective prompts that appear in the chat), to support peer tutors to provide correct and effective help. Authors have investigated the impact of adaptive support on peer tutor learning and have shown that students in the adaptive support condition learned more than the students in the non-adaptive condition. Further, as the adaptive support increases, the difference between learning gain in the adaptive condition and the non-adaptive conditions became more apparent. In (Baghaei et al, 2007) adaptive support was built into an intelligent tutoring system in which students construct UML class diagrams that satisfy a given set of requirements. Adaptive feedback was provided to groups while collaborating on the design of UML class diagrams in order to guide them towards the correct solution. It has been found out that students who received adaptive feedback while working with the system performed significantly better on the collaborative task. In (Karakostas and Demetriadis, 2011), authors have examined the use of adaptive prompts to enhance domain learning. The ACS implemented in their study monitored students' discussions in order to detect whether students have missed to discuss important subject relevant concepts during their discussions. When a missing concept was detected the system provided a prompt to students showing the missing information. Authors have shown that this mechanism has resulted in improved learning outcomes. In (Demetriadis et al, 2018) authors have proposed the potential use of conversational agents in Massive Open Online Courses (MOOCs) to enhance the MOOC experience of course participants. The study has described how conversational agents can be applied to peer interaction sessions in order to enhance the course engagement of MOOC participants that will help to reduce the overall MOOC dropout rates while facilitating educators to better orchestrate MOOC activities.

However, as emphasized in (Karakostas and Demetriadis, 2011) ACSs are still at an early stage of research and most of the efforts that have implemented adaptive support are strongly related to a particular domain of instruction. Towards this end, the objective of our study is to emphasize the need for implementing adaptive collaborative learning support considering not only learning

that occurs within a specific domain in a particular space, but considering diverse behaviours that occur in cross-context learning situations.

2.2 Cross-context learning and collaboration orchestration

With the increased access to emerging communication technologies, Learning Management Systems (LMS), MOOCs, Virtual Learning Environments (VLEs), Social Networking Sites (SNS), and 3D Virtual Worlds (3DVWs) to name a few, students learn across different digital learning spaces that spread beyond the boundaries of physical spaces defined by traditional classroom environments (Kloos et al, 2012; Martinez-Maldonado et al, 2016; Tsovaltzi et al, 2015). Students engage in different learning activities in different learning spaces and associate different learning communities disregard the place and time in which learning occurs. Such learning scenarios are referred to as across-spaces learning situations, in which learning activities are not restricted or constrained to a single physical or digital environment (Kloos et al, 2012). Across-spaces learning scenarios provide valuable opportunities towards learning, since physical and social interactions that occur in 'real-world', outside the traditional classroom, promote the acquisition of certain skills (Kloos et al, 2012).

Although distinct learning spaces provide a wide variety of learning opportunities towards learners, understanding how learning occurs across-spaces in its totality combining multiple spaces is a complex task (Prieto et al, 2017). This leads to challenges in being able to create interconnected flows between different learning spaces (e.g., formal, informal, virtual spaces) in order to support learners while maintaining smooth transitions across different learning spaces (Kloos et al, 2012). How existing pedagogical strategies e.g., collaborative learning, game-based learning can be effectively utilized considering more complex and dynamic across-spaces learning situations that spreads beyond the traditional classroom walls have been identified as an interesting field worth exploring (Kloos et al, 2012).

In the domain of CSCL, managing learning scenarios while adapting to a number of different parameters both in real-time and across longer scales of time, is referred to as "orchestration" of the collaborative learning activity (Dillenbourg et al, 2011; Tissenbaum and Slotta, 2015). When considering the cross-context learning situations the real-time management or the orchestration of collaboration become more challenging for educators than managing traditional scripts in a single space e.g., classroom, as both macro and micro script parameters now require being adjusted according to learning activities that occurs across-contexts that associates complex technologies (Tissenbaum and Slotta, 2015). Design and execution of complex collaborative scripts in such scenarios demand increased levels of information processing needs of educators and learners (Tissenbaum and Slotta, 2015). In such a context, learning analytics can be effectively utilized to make simplified views on complex across-spaces learning scenarios facilitating educators to make data-informed script

design decisions. These script design decisions can then be used to formulate adaptive collaboration scripts that tailor learning experiences to individual students and group characteristics (Tissenbaum and Slotta, 2015). Further during the execution of the scripts, learning analytics can be used to update the educator on the status of collaboration occurs at different levels (e.g., individual level, group level) by showing which script parameters requires being adjusted (e.g., time) and also by proposing dynamic group re-configurations (e.g., learner dropouts in the middle of the activity) or by highlighting groups that require intervention (Tissenbaum and Slotta, 2015). Apart from formulating adaptive collaboration scripts, the association of intelligent agents and real-time data mining techniques into learning environments have been shown beneficial towards the orchestration of scripted cross-context learning situations (Tissenbaum and Slotta, 2015).

2.3 Collaborative learning and learning analytics

Learning analytics is defined as the "measurement, collection, analysis and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and the environment in which it occurs" (Ferguson, 2012). Recently learning analytics gained a lot of attention as it provides different mechanisms and techniques to better understand learners (Dawson, 2006) while providing insights to improve teaching practices (Dyckhoff et al, 2013; Ferguson, 2012). During recent times, different learning analytics techniques accompanied with data mining and machine learning have been widely adopted in different learning contexts for different purposes as it provides new ways to analyze data on students interactions, engagement, and performances (Coffrin et al, 2014).

Different learning analytics techniques have been used in the domain of CSCL to better understand collaborative interactions, participation behaviours, knowledge building behaviours etc. of students in order to make productive interventions during collaboration interactions. A number of mechanisms such as process mining, sequential mining, data mining, social networking analysis and different machine learning techniques such as predictive analytics, Bayesian networks, and fuzzy logic have been effectively utilized in different studies to address a number of research questions that have covered different aspects of collaboration. For instance, some researchers have used data mining and process mining techniques to analyze data collected in classroom collaborative sessions to distinguish high from low achieving groups (Martinez-Maldonado et al, 2013) while some researchers have used machine learning techniques, i.e., Hidden Markov Models and multidimensional scaling techniques to analyze conversational data collected during collaborative learning activities to detect effective and non-effective knowledge sharing episodes (Soller, 2004). Learning analytics have also been used to make productive interventions during the collaborative construction of written documents (McNely et al, 2012).

With the incorporation of predictive machine learning techniques, some research has attempted to predict group learning performance in collaborative learning sessions as it helps to determine better group-based assessment measurements. For instance, Xing et al (2015) used activity theory to holistically quantify student's participation in CSCL activities, which was then used to build a student performance prediction model, using Genetic Programming. Goode and Caicedo (2014) have analyzed log data collected from a social media website to measure group participation during a collaborative learning task. A model was then proposed to predict team performance in future collaborative learning activities using system-tracked log data. Cen et al (2016) have used supervised machine learning techniques, i.e., classification and regression to predict group performance using data which depicted member interactions. Research has also focused on predicting post-test scores by taking into account pair interactions (Rafferty et al, 2013). Olsen et al (2015) have argued that much of the research on learning predictions have focused on modeling individual learning and much of the work does not attempt to predict student performance as students collaboratively solve problems. In their work Olsen et al (2015) have used a standard logistic regression model, i.e., Additive Factors Models which is widely used for predicting individual student performance in the context of Intelligent Tutoring Systems (ITS) to predict collaborative problem-solving performances of students in an ITS environment.

Based on some research already done in the field it was seen that different learning analytics techniques have been broadly utilized to better understand collaborative group learning processes as well as to predict group learning performances. However, less attention is given to predict individual learners' collaborative learning participation behaviour considering across-spaces learning situations, although such predictions can inform the formulation of adaptive collaborative learning scripts that adapts to diverse individual learning behaviours observed in different learning spaces.

3 Participation Prediction as an Adaptive Collaborative Script Parameter for Pyramid Based Collaborative Learning Scripts

Implementation of tools and techniques to enhance students' engagement in collaborative learning activities has been a research question of interest in the Technology Enhanced Learning (TEL) research community for many years. Formulation of homogeneous or heterogeneous learner groups based on learner's profile details (e.g., preferences, knowledge levels etc.) which were captured using questionnaires or surveys prior to the group formation process is one of the frequently adopted method for criteria-based group formation until recent times. This method has reported being effective at achieving specific objectives in different collaborative learning situations (Spoelstra et al, 2015; Moreno et al, 2012). However, with the increased use of online learning platforms for teaching and learning, recent research has highlighted the feasibility of using data-driven learning analytics techniques to analyze trace data collected

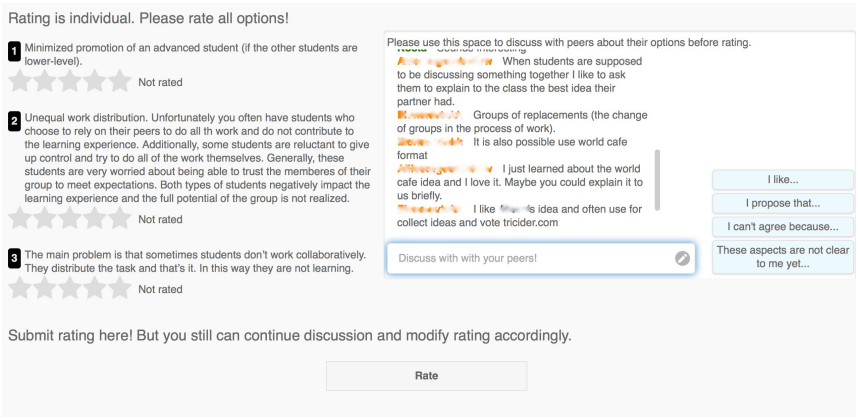

Fig. 1: A screenshot of the PyramidApp showing rating space (left) and the negotiation space (right)

from online learning platforms to formulate meaningful collaborative learning groups. The use of different data-driven techniques to identify team-formation criteria was seen as beneficial to conduct fruitful collaborative sessions in both co-located and distance learning environments (Sanz-Martínez et al, 2017).

In the work presented in this study we propose an adaptive group formation strategy in which an estimation of students' future collaborative learning activity participation was modeled as an adaptive script parameter considering their cross-context learning behaviours. Predicted future activity participation differences of students were used in real-time to formulate heterogeneous groups automatically in a pyramid-based collaborative learning script. A tool called "PyramidApp" was used to operationalize pyramid-based collaborative learning scripts (Manathunga and Hernández-Leo, 2018).

A pyramid flow is initiated with individual students proposing individual answers to a global task. Then, in a second level of the pyramid, individual students are allocated to a number of small collaborative learning groups in which solutions are discussed and rated to agree upon a common proposal. Inbuilt discussion board of the tool provides a negotiation space for participants at group levels to discuss and agree upon the individual options submitted. Once a pyramid activity is designed and published by the educator it becomes accessible to students via a public URL. Activity participants can access the activity by logging to the PyramidApp tool using the given URL. A screenshot of the PyramidApp is shown in Figure 1.

The proposed adaptive group formation strategy is seen vital in a pyramid-based collaborative learning script for many reasons. Firstly, predictions inform the formulation of meaningful group configurations. For instance formulation of heterogeneous groups based on predicted activity participation differences avoid the creation of homogeneous groups that consist only one type of participants e.g., groups consist only inactive participants, yet facilitating the meaningful progression of the collaborative learning activity.

Secondly, as the Pyramid script evolves over time creating increasingly larger groups, an active group i.e., a homogeneous group consist of active participants, collaborating with an inactive group i.e., a homogeneous group consist of inactive participants in the next levels of a pyramid will not result in creating beneficial collaborative learning opportunities for the members of the active group as they cannot build rich pedagogical interactions with members of the inactive group who exhibited little or no interest towards collaboration. Combining these type of homogeneous groups as one big group in the next levels of the pyramid can demotivate members of the active group causing unpleasant learning experiences.

Finally, the formulation of heterogeneous groups based on activity engagement differences of students ensures that every group consists of at least a portion of active participants who will actively contribute to the collaborative learning task at hand. Assigning at least a few active participants in a group can positively influence the inactive participants, as inactive participants get a chance to observe meaningful collaborative interactions and productive communicative acts occur among active participants. Being informed on the positive interactions that occur among active participants can motivate and encourage inactive participants to take part in the pyramid script in the next levels.

3.1 Proposed approach

3.1.1 Formalization of the learning problem and feature representation

The prediction problem addressed in this study was treated as a binary classification problem, in which we attempted to use supervised machine learning techniques to learn a classifier to predict the future collaborative learning activity participation of individual students. The prediction problem addressed in this study can be formulated using mathematical notations as follows.

Given a dataset of observations $S = (x_1, y_1), ..., (x_m, y_m)$ where x_i is a vector specifying various individual student features (extracted from student-platform interaction data) and $y_i \in 0, 1$ representing whether or not a given student will participate in collaborative learning activity, the problem is to learn a classifier to infer value of y_i given x_i. The following sections describes how we collected training data, feature generation and model selection processes adhered in detail.

3.1.2 Data Collection

The training data used in this study were collected from two different learning contexts: (i) Classroom learning context and (ii) Distance learning context. In each learning context, two different learning spaces were used to collect training data that described students' individual and collaborative learning behaviours.

In the classroom context, we extracted data from a Moodle - an open source Learning Management System (LMS) - course (164 cases). In the distance learning context, we collected data from a MOOC course querying the Canvas LMS REST API, which described students learning behaviour in a different digital learning space (230 cases). In both spaces, the data consisted of records that provided insights on individual student-platform interactions details (e.g., course content page views, forum discussion views, assignment submissions, quiz attempts, quiz submissions and forum post submissions).

We conducted collaborative learning activities in the classroom and distance learning contexts to collect training data that depicted students' collaborative learning behaviours. The collaborative learning activities were implemented using PyramidApp. The Pyramid script adopted in both contexts consisted of five phases: (i) an individual phase where students study a learning material and formulate their own answers to a given question related to the material studied (ii) an individual phase where students log in to the PyramidApp and submit individual answers (iii) a small group collaborative phase where students discuss and rate individual answers submitted (iv) a larger group collaborative phase in which students further discuss and rate answers previously selected or best rated in small group levels (v) a debriefing session, where the teacher explained the best rated/ winning answers of each pyramid.

Since the training data collected was originated from different data sources i.e., PyramidApp, Moodle LMS course, MOOC course, data preprocessing became mandatory in order to interpret meaningful information out of raw data. During data preprocessing, for each individual student s, we considered event history up to time t in Moodle LMS course log data and MOOC course log data, given that the student has participated in small group collaboration phase in pyramid script at time t. During pre-processing we had to deal with unstructured data as there had no predefined data model in-place for data gathering from multiple sources in cross-context learning situations. In particular, date-time formats were not consistent and needed to convert them to a common date-time format without losing any important information.

After data preprocessing stage, we built two data sets: (i) a data set merging PyramidApp log data with Moodle LMS course data that described collaborative and individual learning behaviours of a particular set of students (ii) a data set merging PyramidApp log data with MOOC course data that described collaborative and individual learning behaviours of a particular set of students (see Figure 2). The two data sets were later used to train and test machine learning classifiers.

3.1.3 Feature selection

The accuracy of a given classification task depends on the choice of informative and discriminating features that are provided as inputs to the supervised learning algorithm (Cen et al, 2016). In the following, we describe the features used in this study for prediction purposes.

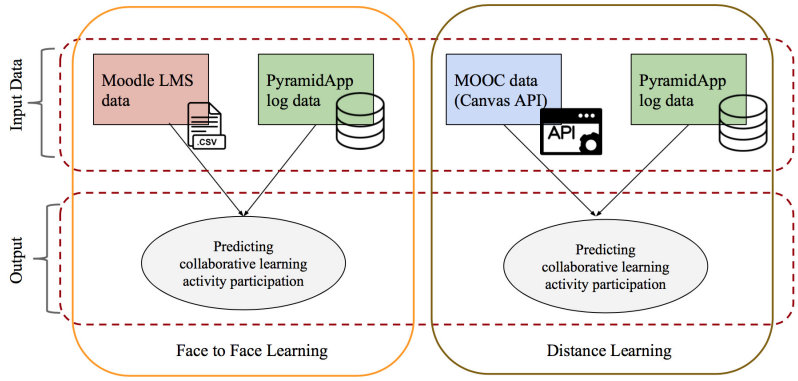

Fig. 2: Heterogeneous data sources

We used a correlation-based approach for feature selection since the removal of irrelevant and redundant features often improves the performance of machine learning algorithms (Yu and Liu, 2003). Based on correlation coefficient values it was observed that in both learning contexts some features positively or negatively correlated with the class, while some features do not have a relationship with class variable (correlation coefficient was zero). Based on the results of the correlation analysis in the classroom context the input vector x_i included seven input features (generated using Moodle LMS log data and PyramidApp Log Data) as mentioned below:

- Total number of course page views before collaborative activity participation
- Total number of forum discussion entry views before collaborative activity participation
- Total number of quiz attempts before collaborative activity participation
- Total number of quiz submissions before collaborative activity participation
- Total number of assignment submitted before collaborative activity participation
- Student's participation in the initial stage of the pyramid activity
- Student's collaborative activity participation (class variable)

In the distance learning context the input vector x_i included ten input features (generated using MOOC course log data and PyramidApp log data) as mentioned below:

- Total number of course page views before collaborative activity participation
- Total number of assignment submitted before collaborative activity participation

- Total number of discussion entries posted before collaborative activity participation
- Total number of quiz submissions before collaborative activity participation
- Total number of quiz attempts before collaborative activity participation
- Total number of quizzes answered correctly
- Total number of quizzes answered incorrectly
- Total quiz score
- Student's participation in the initial stage of the pyramid activity
- Student's collaborative activity participation (class variable)

In both contexts, the class variable y_i was used to specify each individual student's collaborative activity participation. In other words, y_i can take one out of the two values in the classification task in which, 1 depicting 'yes' and 0 depicting 'no' with regard to the small group collaborative phase participation of each individual during pyramid script enactment.

3.1.4 Algorithm implementation and model selection

To predict individual student's collaborative activity participation in Pyramid activities, we explored the applicability of three widely adapted supervised machine learning techniques: Support Vector Machines (SVMs), Feed Forward Neural Networks (FFNNs) and Random Forests (RFs). In the following, we provide a brief explanation of each model.

The SVMs are pioneered by Vapnik (Vapnik, 2013) and have been used to solve both classification and regression problems in different contexts, although it is widely used to solve classification problems. The SVMs construct a hyperplane(s) usually in the high dimensional space, in order to separate two data classes, i.e., positive and negative instances in a given dataset. Intuitively, the maximum-margin hyperplane, which represents the largest margin between two data classes achieves the best possible separation and has been proven to lower the classifier's expected generalization error (Kotsiantis et al, 2007). The FFNN is an artificial neural network which simulates the functionality and behaviour of biological neurons (Hagan et al, 1996). FFNNs typically consist three types of layers: (i) input layer–consists of input nodes, (ii) one or more hidden layers–consist of hidden nodes, and (iii) output layer–consists output nodes. In FFNNs information flow only in one direction through the network from the input layer to the output layer, without forming cycles in the network. During the training phase of the network, network parameters (e.g., weights and biases) requires being adjusted using back-propagation algorithm. Afterward, the trained network can be presented with unseen test data for classification tasks. Finally, RFs is an ensemble learning technique used for classification tasks. In general ensemble learning techniques generate many classifiers and aggregate their results to provide a final classification output. During the training phase of RFs, a number of decision trees are being generated and the mode of the classes output by individual trees is provided as the prediction output (Breiman, 2001).

Table 1: Prediction performance accuracy comparisons of different models using 10-fold cross validation

Learning context	Model	Accuracy score
Classroom	SVMs*	0.82
	NNs	0.81
	RFs	0.79
Distance	SVMs	0.80
	NNs*	0.81
	RFs	0.78

* Best performed model in each learning context

The aforementioned classification algorithms were implemented using scikit-learn machine learning library [1]. Each algorithm was trained separately in both learning contexts, i.e., classroom and distance learning, to determine the best performing classifier. To obtain the best hyper-parameters for each algorithm a grid search was carried out. Each model, i.e., an algorithm with best hyper-parameters, was then evaluated using K-fold cross-validation method given its benefits over train/test split procedure. In particular, we implemented 10-fold cross-validation, which has been shown as a reliable estimate in the literature towards model evaluation (Cen et al, 2016). Table 1 provides the cross-validation accuracy of each model. Based on cross-validation accuracy scores it was seen that SVMs outperformed other models when predicting collaborative activity participation in classroom context while NNs performed slightly better than SVMs in distance learning context.

4 Evaluation: Formulation of adaptive collaborative learning groups in Pyramid scripts in real-time

In the following sections, we present case studies in which we used the prediction outcomes of the best performed models i.e., SVMs in classroom learning context and NNs for distance learning context (see Table 1) to formulate adaptive collaborative learning groups in pyramid scripts in real-time.

Figure 3 shows the architecture adapted for this purpose. As can be seen in Figure 3 the prediction output (which differentiated active vs. inactive participants) was associated with the other learning design parameters (e.g., group size, time allocation, number of pyramid levels) of the Pyramid script during the activity design stage. Heterogeneous groups were then formulated in real-time during small group collaboration stage of the Pyramid script automatically.

[1] http://scikit-learn.org

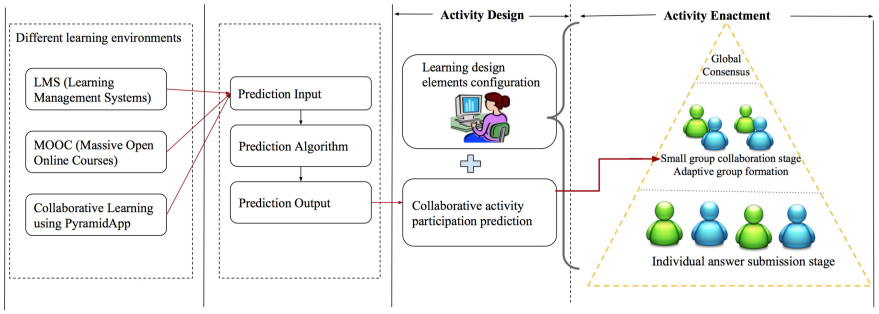

Fig. 3: Pipeline-Integrating prediction results as collaborative script parameters

Table 2: PyramidApp design parameters for classroom activities

Design Parameter	Value
No. of Pyramid Levels	3 (e.g., initial answer submission stage, small group collaboration stage and large group collaboration stage)
Minimum students per Pyramid	6
Small group size	3
Initial answer submission time	5 mins.
Rating submission time	4 mins.

4.1 Case studies

4.1.1 Collaborative learning activities in classroom context

We carried out collaborative learning activities in four undergraduate classes in January 2018. First year undergraduate engineering students who were enrolled in *Computer Organization* course participated, with informed consent, in the collaborative learning activities. Prior to the activity enactment, we did a demonstration explaining the flow of the activity.

Design elements associated with pyramid activities conducted in classroom sessions are shown in Table 2. Based on design configurations of the PyramidApp, i.e., the minimum number of learners per Pyramid, a number of pyramids were instantiated allocating participants to Pyramids who logged into the system at different times. Further details about the implementation of this tool can be found in (Manathunga and Hernández-Leo, 2018). The task given to students was related to a programming problem, in which the students were asked to collaboratively decide the best answer to the given programming problem. Predicted future collaborative learning activity participation of each student (obtained from trained SVM model) was incorporated to formulate heterogeneous groups automatically in the small group collaboration level of the Pyramid script.

$$\text{Overall accuracy} = \frac{\text{Correctly predicted active participants} + \text{Correctly predicted inactive participants}}{\text{Total activity participants}} \quad (1)$$

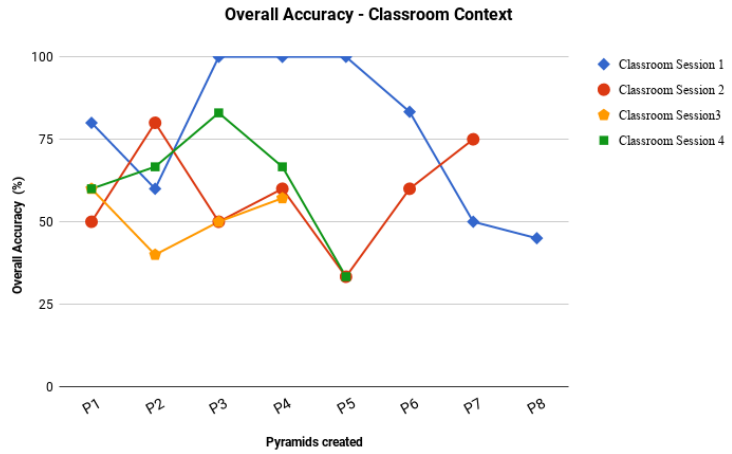

Fig. 4: Overall accuracy of prediction in classroom context

4.1.2 Results

We adopted a similar decision scheme proposed in (Lykourentzou et al, 2009) to evaluate the prediction accuracy of the machine learning models during case studies conducted in the real-world context. We modified their decision scheme to match with the specific prediction problem we are interested in, although the original work was related to dropout prediction in an e-learning system. Following paragraphs describe the decision scheme adapted and the interpretation of the results.

The overall accuracy criterion (see equation 1) measures on average the proportion of accurately predicted active and inactive participants given the total number of activity participants. Figure 4 depicts the overall accuracy results of the machine learning model. The vertical axis in Figure 4 presents the overall accuracy and the horizontal axis represents each pyramid starting from P1 which refers to the first pyramid and so on in each classroom session.

Based on the overall prediction accuracy results, it was observed that in many pyramids the classifier has achieved an overall prediction accuracy which was above 50%. However in P8 generated in classroom session 1, P5 generated in classroom session 2, P2 generated in classroom session 3 and in P5 generated in classroom session 4, the overall accuracy has dropped below 50%, which is much less than the performance accuracy score reported during 10-fold cross-validation for SVM classifier which was 0.82 (see Table 1).

Table 3: PyramidApp design parameters for MOOC activities

Design Parameter	Value
No. of Pyramid Levels	3 (e.g., initial answer submission level, small group level and large group level)
Minimum students per Pyramid	15
Small group size	5
initial answer submission time	1 day
Rating submission time	1 day

4.1.3 Collaborative learning activities in MOOC context

We conducted two Pyramid collaborative learning activities in a MOOC course named *Concepts and Practice of Responsible Research and Innovation* in February 2018. The first pyramid activity asked course participants to discuss which responsible research and innovation practices are easier to implement while in the second activity students were asked to discuss which responsible research and innovation practices are difficult to implement. Design parameters associated with the Pyramid activities are given in Table 3. Participants were informed that the activity was voluntary and that activity participation was part of a research experience and responses collected will be treated anonymously.

4.1.4 Results

In contrast to the classroom pyramid activities presented earlier (see Sect. 4.1.1) in which we formulated adaptive collaborative groups based on prediction results, within the MOOC context we were unable to do the same due to the poor performance of the trained NNs classifier. The predicted outcome of the classifier was 0 for all the students which indicated that none of the students will participate in the collaborative learning activity. Training data sets that constituted a limited number of samples that are also imbalanced with regard to the target class may have caused the aforementioned issue. Hence, we attempted to improve the classifier's performance by using normalized features and by introducing new features which were calculated based on percentile ranks (see below) that have been reported to enhance the performance of the classifier accuracy in previous studies conducted in the field (Taylor et al, 2014). We have used the same training data set described in section 3.1.2 to recalculate the features to train and test the NNs classifiers using 10-fold cross validation.

- Total number of course page views before collaborative activity participation (normalized)
- Total number of assignment submitted before collaborative activity participation (normalized)

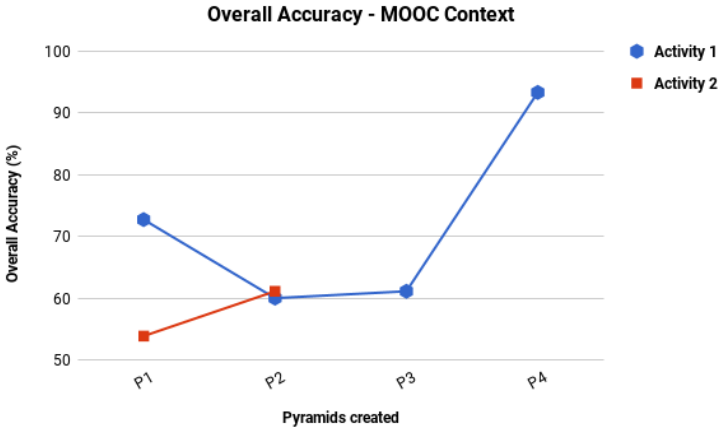

Fig. 5: Overall accuracy of prediction in MOOC context

- Total number of discussion entries posted before collaborative activity participation (normalized)
- Total number of quiz submissions before collaborative activity participation (normalized)
- Total number of quiz attempts (normalized)
- Total number of quizzes answered correctly (normalized)
- Total number of quizzes answered incorrectly (normalized)
- Total quiz score (normalized)
- Total number of course page views before collaborative activity participation as a percentile
- Total number of assignment submitted before collaborative activity participation as a percentile
- Students participation in the initial stage of the pyramid activity
- Students collaborative activity participation (class variable)

4.1.5 Improved classifier performance

At the time of presenting the results of the study, we did not have access to an on-going MOOC to evaluate the performance of the improved NNS classifier in real-time. In Figure 5, we present the overall accuracy of the improved classifier as calculated considering the predicted outcome against the actual collaborative learning activity participation of students within the MOOC collaborative learning activities described in Sect. 4.1.3

When considering the overall prediction accuracy, it was observed that in both activity 1 and activity 2 classifier has achieved relatively higher levels of overall accuracy rates which are above 50%. In particular, during activity 1, in P4 the overall classification accuracy has increased over 90% which shows a good prediction performance. However, in P1 in activity 2, the overall accuracy

has dropped below 60%, which is much less than the overall accuracy observed in other pyramid activities.

4.2 Discussion

Figure 4 and Figure 5 summarize the prediction performance of the machine learning classifiers in predicting future collaborative learning activity participation of students. The overall accuracy criteria was used to measure the proportion of active (students who will participate in the collaborative learning activity) and inactive participants (students who will not participate in the collaborative learning activity) correctly predicted by the SVM and NNs classifiers in classroom and distance learning contexts respectively. A Cohen's kappa measure has been calculated to better elaborate the prediction performance of the classifiers in the two different learning contexts. In the classroom setting it was seen there was no agreement between the instances classified by the machine learning classifier and the data labeled as ground truth ($k = 0.211$, $p > 0.001$). In order to better understand the reason behind the poor performing classifier we have further analyzed the characteristics of the learner's profiles in both training and test datasets in the classroom context. It became evident that in the classroom setting the time frame in which we placed the evaluation studies has affected the classifier performance. The test data did not contain records of quiz taking behaviours of students, due to the fact that by the time we placed evaluation studies in the classroom context, no quiz related activities were posted in the LMS (as it was the beginning of the semester). Being unable to have access to the quiz related data which was seen as the most correlated variable and the fact of being inactive, describes the poor performance of the classifier in the classroom setting. In general, the noisy data in the classroom setting has resulted in a poor performing classifier. A Cohen's kappa measure has also been calculated to evaluate the prediction performance of the improved classifier in the distance learning context. A moderate agreement between the instances classified by the machine learning classifier and the data labeled as ground truth ($k = 0.625$, $p < 0.001$) was observed within this context.

4.3 Limitations of the study

In this study, we have attempted to emphasize the use of predictive analytics to inform the adaptive collaborative scripting in across-spaces learning situations. We have presented how machine learning techniques can be used to obtain an estimate on future collaborative learning activity participation of students based on data collected from different learning spaces that described their individual and collaborative learning behaviours in previous activities. The findings of the present study should be interpreted in light of the following limitations.

One of the major limitations of our study is the use of training data sets that constituted a limited number of samples that are also imbalanced with regard to the target class. A larger and balanced dataset can potentially enhance the model performance creating opportunities to obtain more accurate test results. Although the current accuracy level of predictions is informative to achieve the objective of the study (which is to formulate adaptive collaborative learner groups based on future collaborative activity participation differences of students) more accurate predictions can provide more reliable estimates with increased overall accuracy levels.

On the other hand, the time frame in which we collected training data and the time in which we have positioned the evaluation studies (due to the designs of each real-world learning scenarios) have affected the classifier performance. In the classroom context and distance learning context the training data collected from Moodle LMS and MOOC API respectively depicted student-platform interactions for a period of one week. In the classroom context the educator conducted pyramid activities at the beginning of the course and in the distance learning context, the MOOC course was designed to have collaborative learning activities in the first and second week. As it was mentioned earlier, for each individual student s we considered event history up to time t in Moodle LMS course log data and MOOC course log data, given that the student has participated in small group collaboration phase of pyramid script at time t. Hence, the log data obtained to train classifiers from both Moodle LMS and MOOC platform consisted of records that described individual learners learning behaviour for a short period of time. On the other hand, in the classroom context, the evaluation studies were conducted in another course after three weeks from the course start date. The structure of the course was different from the course which we used to collect training data and consisted of records that described student-platform interactions over a relatively longer duration. In other words, the differences associated with the time frames in which we positioned the evaluation studies in the classroom context and the differences associated with the structure of the course make it difficult to model individual students which resulted in a more difficult prediction task.

Finally, the present study does not evaluate whether the impact of adaptive collaborative scripting is more beneficial to students than non-adapted collaborative scripts. As it was mentioned earlier, the main focus of the study was to evaluate whether predictive analytics can be used to inform the formulation of adaptive collaborative learning groups in the context of CSCL considering across-spaces learning situations and how such predictions can be used to formulate adaptive collaborative learner groups automatically in real-time. However, from a pedagogical perspective, it is important to measure whether the proposed adaptive group formation strategy has created an impact on students. Whether adaptive group configurations has resulted in increased learning gains, other than facilitating to maintain the flow of collaboration across pyramid levels is an important aspect which requires to be further researched.

5 Conclusions and Future Work

In this study, we have presented how predictive analytics inform the formulation of adaptive collaborative group configurations in the context of CSCL. The main contribution of the present study is the use of data collected in cross-context learning situations that exhibited students' prior activities, to predict future collaborative learning activity participation of students in a pyramid-based script. The prediction problem of interest was modeled as a supervised machine learning problem and solved using well-known supervised machine learning techniques, i.e., SVMs and NNs. Each classifier was tested using 10-fold cross-validation to evaluate model performance. During several case studies conducted in two different learning contexts i.e., classroom and distance learning context, we then incorporated the prediction results obtained from machine learning models to formulate adaptive group configurations in pyramid-based collaborative learning sessions.

Findings of the case studies showed that the data collected from across-spaces learning scenarios is informative to automatically classify students that can then allow teachers to make more informed adaptive group configurations adapting to the estimated activity engagement differences of students. Most importantly, the work presented in this article conveys that the learning occurs in one space is informative to learning that occurs in another space, which highlights the interesting connections exist across different learning spaces although understanding the complex interplay between different learning spaces and interpreting the connections that lie across-spaces is a challenging task that requires effort. Nevertheless, it should be pointed out that the present study sheds light on the applicability of learning analytics techniques i.e. predictive analytics to make those connections explicit in a useful manner suggesting that application of sophisticated learning analytic techniques can advance this field of research. We consider the work presented in the study is an important step for the field to begin to use previous behavioural data to understand how to create interventions in later activities. Although the present study lacks a discussion on how the interventions developed using the predictions impact students, we argue that understanding the predictions themselves is important and showing that these can be calculated in real-time even with scare data available that exhibited previous activities of students in cross-context learning situations is an important contribution of our work. As it was described in previous studies (Liaw and Huang, 2000; Northrup, 2001) interactions among participants does not occur automatically, rather intentionally designed collaborative learning activities facilitate interactions. Towards this end, we hope that the proposed adaptive group formation approach that attempts to formulate groups based on activity participation differences of students is a meaningful strategy that will facilitate students to gain benefits of collaboration.

Moreover, some of the lessons learned and observations captured while conducting evaluation studies in real-world context are interesting to be summarized in the conclusions. For instance, when conducting evaluation studies in the classroom context we realized not only the features extracted from log

data but also features that describe learner's cognitive-affective states such emotions, moods, feelings, which could be captured in the physical space using sensory inputs can provide useful information to generate fine-grained predictive models as those states can vastly dominate learning activity participation of students. Incorporation of such relevant data that further describe learner's behaviours in different perspectives in different modalities may enhance the model performance. On the other hand, the technological tools that used to enable and structure collaborative learning session alone may not necessarily result in productive learning activity gains. Interactions that occur among students physically in the classroom require to be continuously reinforced by the educator in order to maintain students attention towards the learning activity which adds to the "orchestration load" of the educator (Prieto et al, 2018). We have observed in several instances that students missed the participation in different levels of the pyramid script as they speak with the colleagues sitting next to them or due to lack of attention towards collaborative learning task e.g., checking notifications on their mobile phones. Although some of these students might have been classified as active participants who would contribute to the collaborative learning task (based on the behaviour they have exhibited in the Moodle space), it was observed that the classroom behaviour of students cannot be fully described alone using log data, which highlighted the need for incorporating physiological, behavioural and subjective data that better describe learners behavior in real classroom settings (Prieto et al, 2018). For instance, the NISPI framework suggested in (Cukurova et al, 2018) provides a good understanding of how physiological measures can be used to identify Collaborative Problem Solving (CPS) competence levels of students. As described in (Cukurova et al, 2018) hand position and heads direction data provide useful information to predict CPS competency levels of students. The applicability of such physiological measures to predict the quality of collaboration among groups of students are presented in (Spikol et al, 2018). (Grover et al, 2016) have provided evidence that physiological measures such as screen pointing, leaning forward, joint attention (looking at screen), taking the mouse (with or without consent) and synchrony in body position are useful features in predicting the level of collaboration in pair programming context.

On the other hand, in the distance learning context, it was observed that the two different MOOCs that we used to collect data and to position evaluation studies are different in nature which can cause a significant effect on the accuracy of the prediction results. Training data was collected from a MOOC designed for secondary and higher education teachers while evaluation studies were placed in a MOOC which was designed for a research-oriented audience. We have observed that the engagement of MOOC students in collaborative learning activities varied drastically in the two MOOC contexts. Lack of contextual information presented when training machine learning models can also affect the accuracy of the real-time prediction. In the future, we plan to consider these lessons learned, to extend the data sources considered, the experimentation in diverse contexts, the evaluation of its impact in terms of

learning gains, and the provision of orchestration dashboards for teachers to monitor and regulate the adaptive scripts.

Acknowledgements This work has been partially funded by FEDER, the National Research Agency of the Spanish Ministry of Science, Innovations and Universities MDM-2015-0502, TIN2014-53199-C3-3-R, TIN2017-85179-C3-3-R and la Caixa Foundation (CoT project, 100010434). DHL is a Serra Hnter Fellow.

References

Amarasinghe I, Hernández-Leo D, Jonsson A (2017) Towards data-informed group formation support across learning spaces. In: International Workshop on Learning Analytics across-spaces (Cross-LAK), 7th International Conference on Learning Analytics & Knowledge (LAK'17)

Baghaei N, Mitrovic A, Irwin W (2007) Supporting collaborative learning and problem-solving in a constraint-based cscl environment for uml class diagrams. International Journal of Computer-Supported Collaborative Learning 2(2-3):159–190

Breiman L (2001) Random forests. Machine learning 45(1):5–32

Cen L, Ruta D, Powell L, Hirsch B, Ng J (2016) Quantitative approach to collaborative learning: performance prediction, individual assessment, and group composition. International Journal of Computer-Supported Collaborative Learning 11(2):187–225

Coffrin C, Corrin L, de Barba P, Kennedy G (2014) Visualizing patterns of student engagement and performance in moocs. In: 4th International Conference on Learning Analytics and Knowledge, pp 83–92

Cukurova M, Luckin R, Millán E, Mavrikis M (2018) The nispi framework: Analysing collaborative problem-solving from students' physical interactions. Computers & Education 116:93–109

Dawson S (2006) Study of the relationship between student communication interaction and sense of community. Internet and Higher Education 9(3):153–162

Demetriadis S, Karakostas A (2008) Adaptive collaboration scripting: A conceptual framework and a design case study. In: International conference on complex, intelligent and software intensive systems, IEEE, pp 487–492

Demetriadis S, Karakostas A, Tsiatsos T, Caballé S, Dimitriadis Y, Weinberger A, Papadopoulos PM, Palaigeorgiou G, Tsimpanis C, Hodges M (2018) Towards integrating conversational agents and learning analytics in moocs. In: International Conference on Emerging Internetworking, Data & Web Technologies, Springer, pp 1061–1072

Dillenbourg P, Tchounikine P (2007) Flexibility in macro-scripts for computer-supported collaborative learning. Journal of computer assisted learning 23:1–13

Dillenbourg P, Zufferey G, Alavi H, Jermann P, Do-Lenh S, Bonnard Q, Cuendet S, Kaplan F (2011) Classroom orchestration: The third circle of usability. CSCL2011 Proceedings 1:510–517

Dyckhoff A, Lukarov V, Muslim A, Chatti M, Schroeder U (2013) Supporting action research with learning analytics. In: 3rd International Conference on Learning Analytics and Knowledge, pp 220–229

Ellis RA, Goodyear P (2018) Spaces of Teaching and Learning: Integrating Perspectives on Research and Practice. Springer

Fall R, Webb NM, Chudowsky N (2000) Group discussion and large-scale language arts assessment: Effects on students' comprehension. American Educational Research Journal 37(4):911–941

Ferguson R (2012) Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning 4(5-6):304–317

Goode W, Caicedo G (2014) Online collaboration: Individual involvement used to predict team performance. In: Zaphiris P., Ioannou A. (eds) International Conference on Learning and Collaboration Technologies, Springer, pp 408–416

Grover S, Bienkowski M, Tamrakar A, Siddiquie B, Salter D, Divakaran A (2016) Multimodal analytics to study collaborative problem solving in pair programming. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, ACM, pp 516–517

Hagan MT, Demuth HB, Beale MH, De Jess O (1996) Neural network design, vol 20. Pws Pub. Boston

Hernández-Leo D, Nieves R, Arroyo E, Rosales A, Melero Gallardo J, Blat J (2012) SOS: Orchestrating collaborative activities across digital and physical spaces using wearable signaling devices. Journal of Universal Computer Science 18(15):2165–2186

Karakostas A, Demetriadis S (2011) Enhancing collaborative learning through dynamic forms of support: the impact of an adaptive domain-specific support strategy. Journal of Computer Assisted Learning 27(3):243–258

Kloos CD, Hernández-Leo D, Asensio-Pérez JI (2012) Technology for learning across physical and virtual spaces: J. ucs special issue. Journal of Universal Computer Science 18(15):2093–2096

Kobbe L, Weinberger A, Dillenbourg P, Harrer A, Hämäläinen R, Häkkinen P, Fischer F (2007) Specifying computer-supported collaboration scripts. International Journal of Computer-Supported Collaborative Learning 2(2-3):211–224

Kollar I, Fischer F, Hesse FW (2006) Collaboration scripts–a conceptual analysis. Educational Psychology Review 18(2):159–185

Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering 160:3–24

Kumar R, Rosé CP, Wang YC, Joshi M, Robinson A (2007) Tutorial dialogue as adaptive collaborative learning support. In: Luckin R. and Kenneth, R and Greer Jim E (eds.) International Conference on Artificial Intelligence in Education, pp 383–390

Liaw Ss, Huang Hm (2000) Enhancing interactivity in web-based instruction: A review of the literature. Educational Technology 40(3):41–45

Lykourentzou I, Giannoukos I, Nikolopoulos V, Mpardis G, Loumos V (2009) Dropout prediction in e-learning courses through the combination of machine learning techniques. Computers & Education 53(3):950–965

Manathunga K, Hernández-Leo D (2018) Authoring and enactment of mobile pyramid-based collaborative learning activities. British Journal of Educational Technology 49(2):262–275

Martinez-Maldonado R, Yacef K, Kay J (2013) Data mining in the classroom: Discovering groups' strategies at a multi-tabletop environment. In: International Conference on Educational Data Mining, pp 121–128

Martinez-Maldonado R, Pardo A, Hernández-Leo D (2016) Introduction to cross lak 2016: Learning analytics across spaces. In: First International Workshop on Learning Analytics Across Physical and Digital Spaces co-located with 6th International Conference on Learning Analytics & Knowledge (LAK 2016), CEUR

Martinez-Maldonado R, Hernandez-Leo D, Pardo A, Ogata H (2017) 2nd cross-lak: learning analytics across physical and digital spaces. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference, ACM, pp 510–511

McNely BJ, Gestwicki P, Hill JH, Parli-Horne P, Johnson E (2012) Learning analytics for collaborative writing: a prototype and case study. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, ACM, pp 222–225

Moreno J, Ovalle DA, Vicari RM (2012) A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics. Computers & Education 58(1):560–569

Northrup P (2001) A framework for designing interactivity into web-based instruction. Educational Technology 41(2):31–39

Nyce C, Cpcu A (2007) Predictive analytics white paper. American Institute for CPCU Insurance Institute of America pp 9–10

Olsen JK, Aleven V, Rummel N (2015) Predicting student performance in a collaborative learning environment. In: International Conference on Educational Data Mining, ERIC, pp 211–217

Prieto LP, Martínez-Maldonado R, Spikol D, Hernández-Leo D, Rodríguez-Triana MJ, Ochoa X (2017) Joint proceedings of the sixth multimodal learning analytics (MMLA) workshop and the second cross-lak workshop. In: CEUR Workshop Proceedings

Prieto LP, Sharma K, Kidzinski Ł, Dillenbourg P (2018) Orchestration load indicators and patterns: In-the-wild studies using mobile eye-tracking. IEEE Transactions on Learning Technologies 11(2):216–229

Rafferty A, Davenport J, Brunskill E (2013) Estimating student knowledge from paired interaction data. In: International Conference on Educational Data Mining, pp 260–263

Roschelle J, Teasley SD (1995) The construction of shared knowledge in collaborative problem solving. In: O'Malley, C.E, (eds.) Computer supported collaborative learning, Springer, pp 69–97

Rummel N, Spada H (2007) Can people learn computer-mediated collaboration by following a script? In: Scripting computer-supported collaborative learning, Springer, pp 39–55

Rummel N, Weinberger A, Wecker C, Fischer F, Meier A, Voyiatzaki E, Kahrimanis G, Spada H, Avouris N, Walker E, et al (2008) New challenges in cscl: Towards adaptive script support. In: Proceedings of the 8th international conference on International conference for the learning sciences-Volume 3, International Society of the Learning Sciences, pp 338–345

Sanz-Martínez L, Martínez-Monés A, Bote-Lorenzo ML, Muñoz-Cristóbal JA, Dimitriadis Y (2017) Automatic group formation in a mooc based on students activity criteria. In: European Conference on Technology Enhanced Learning, Springer, pp 179–193

Soller A (2004) Computational modeling and analysis of knowledge sharing in collaborative distance learning. User Modeling and User-Adapted Interaction 14(4):351–381

Spikol D, Ruffaldi E, Dabisias G, Cukurova M (2018) Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. Journal of Computer Assisted Learning 34(4):366–377

Spoelstra H, Van Rosmalen P, Houtmans T, Sloep P (2015) Team formation instruments to enhance learner interactions in open learning environments. Computers in human behavior 45:11–20

Stahl G, Koschmann T, Suthers D (2006) Computer-supported collaborative learning: An historical perspective. Cambridge handbook of the learning sciences 2006:409–426

Taylor C, Veeramachaneni K, O'Reilly UM (2014) Likely to stop? predicting stopout in massive open online courses. arXiv preprint arXiv:14083382

Tissenbaum M, Slotta JD (2015) Scripting and orchestration of learning across contexts: A role for intelligent agents and data mining. In: Seamless Learning in the Age of Mobile Connectivity, Springer, pp 223–257

Tsovaltzi D, Judele R, Puhl T, Weinberger A (2015) Scripts, individual preparation and group awareness support in the service of learning in facebook: How does cscl compare to social networking sites? Computers in Human Behavior 53:577–592

Vapnik V (2013) The nature of statistical learning theory. Springer-Verlag, New York

Villasclaras-Fernández ED, Hernández-Gonzalo JA, Hernández Leo D, Asensio-Pérez JI, Dimitriadis Y, Martínez-Monés A (2009) Instancecollage: A tool for the particularization of collaborative ims-ld scripts. Journal of Educational Technology & Society 12(4):56–70

Walker E, Rummel N, Koedinger KR (2009) CTRL: A research framework for providing adaptive collaborative learning support. User Modeling and User-Adapted Interaction 19(5):387–431

Walker E, Rummel N, Koedinger KR (2014) Adaptive intelligent support to improve peer tutoring in algebra. International Journal of Artificial Intelligence in Education 24(1):33–61

Waller MA, Fawcett SE (2013) Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. Journal of Business Logistics 34(2):77–84

Weinberger A, Stegmann K, Fischer F, Mandl H (2007) Scripting argumentative knowledge construction in computer-supported learning environments. In: Scripting computer-supported collaborative learning, Springer, pp 191–211

Xing W, Guo R, Petakovic E, Goggins S (2015) Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. Computers in Human Behavior 47:168–181

Yu L, Liu H (2003) Feature selection for high-dimensional data: A fast correlation-based filter solution. In: Proceedings of the 20th international conference on machine learning (ICML-03), pp 856–863

Ishari Amarasinghe is a Ph.D. candidate in the Information and Communications Technologies Department of Universitat Pompeu Fabra (UPF), Barcelona. She received her Master's in Intelligent interactive Systems from the UPF. She received her BSc in Information and Communication Technologies (First Class) from the University of Colombo School of Computing, Sri Lanka. She has won many awards as a student and some of them include gold medal for the best final year research for her bachelor thesis and a student scholarship from the Google Anita Borg Institute to attend the Grace Hopper Celebrations of women in computing. Her bachelor thesis and Master thesis work sparked her interest in interdisciplinary research. Her research interests are framed in the interdisciplinary intersection of Machine Learning, Human-Computer Interaction and Education; with an emphasis on Computer Supported Collaborative Learning.

Prof. Davinia Hernández-Leo received the MS and Ph.D. degrees in telecommunications engineering from the University of Valladolid, in 2003 and 2007, respectively. She is currently Associate Professor, Serra Húnter Fellow and head of the Interactive and Distributed Technologies Group (TIDE) at the ICT Department of Universitat Pompeu Fabra, Barcelona. She is also Vice-Principal of the UPF Polytechnic School and the director of its Teaching Quality and Innovation Unit since 2008. Her research interests research lies at the intersection of human-computer interaction, network and computer applications, and learning sciences. In particular, they are framed in the area of Learning Technologies, with emphasis in learning design technologies, CSCL, community platforms, data analytics, and architectures and devices for learning. She is currently Vice-President of the European Association of Technology-Enhanced Learning and Associate Editor of IEEE Transactions on Learning Technologies.

Prof. Anders Jonsson is the head of the Artificial Intelligence and Machine Learning research group at Universitat Pompeu Fabra. He has authored

more than 50 publications in peer-reviewed international conferences and journals. He has also been an investigator on several European Union projects, notably APIDIS and SpaceBook, and he is currently the coordinator of the CHIST-ERA project DELTA on lifelong reinforcement learning.