

TOWARDS A TIME SERIES APPROACH FOR THE CLASSIFICATION AND EVALUATION OF COLLABORATIVE ACTIVITIES

Irene-Angelica CHOUNTA, Nikolaos AVOURIS

HCI Group

University of Patras

Ypatias Str.

26 500 Rion-Patras, Greece

e-mail: {houren, avouris}@upatras.gr

Abstract. The analysis and evaluation of computer-supported collaborative activities is a complex and tedious task. However, it is necessary in order to support collaborative scenarios, to scaffold the collaborative knowledge building and to evaluate the learning outcome. Various automated techniques have been proposed to minimize the workload of human evaluators and speed up the process. In this study, we propose a memory based learning model for the analysis, classification and evaluation of collaborative activities that makes use of time series techniques along with logfile analysis. We argue that the classification of collaborative sessions, with respect to their time series attributes, may be related to their qualitative aspects. Based on this rationale, we explore the use of the model under various settings. The results of the model are compared to assessments made by expert evaluators using a rating scheme. Correlation and error analyses are further conducted.

Keywords: time series, collaboration, logfile analysis, evaluation, computer supported collaborative learning

1 INTRODUCTION

The analysis of collaborative activities is a research area of great importance in the field of Computer Supported Collaborative Learning (CSCL). It provides the means to understand the mechanisms of collaboration and the background to develop new

frameworks to enhance joint activities. The analysis and evaluation of collaborative activities is usually carried out by human evaluators and it is a tedious and time consuming task due to the complex nature of collaboration. Various data representation methods and evaluation methodologies, both qualitative and quantitative, have been proposed to support the task of collaboration analysis. In addition, automated techniques have been implemented to provide assessments of the quality of joint activities, mostly based on statistics and event metrics computed from recorded logfiles. However, these automated approaches also entail drawbacks, since it is argued that the lack of a qualitative approach undermines the depth of the analysis [1]. Moreover, the use of automated metrics usually involves complex statistical methods and mathematical models that are difficult to adapt in new settings.

The main goal of this study is to propose a method to classify and evaluate collaborative activities by matching new collaborative sessions to an existing set of reference cases. The method uses the events produced by user activity and recorded into logfiles. This accelerates the task of analysis and lowers the effort of human evaluators. The collaboration process is analyzed through the interactions between users as well as through the interaction of users with the groupware application. The classification is carried out by a memory-based learning model that is easy to implement, adapt and interpret. The model makes use of time series techniques and a set of previously evaluated activities to classify the new ones. Thus it is possible to analyze and evaluate qualitative aspects of collaboration to a satisfactory extent. The way a collaborative activity unfolds in time carries important information regarding the quality of collaboration and the activity itself. This indicates the need of a representation and analysis method that exploits the aspect of time as the main dimension of analysis [2]. The use of such an analysis schema allows the detection and capturing of undesirable phenomena or bottlenecks. The proposed model does not require ending of the activity in order to be applied, as most automated analysis techniques do, since it relies on the change of activity in time and not on the overall outcome. Therefore, the model could be potentially used in real time to evaluate joint activities and to provide online feedback to learners.

In order to further explore this notion, we designed and carried out a study to explore the use of the proposed memory-based learning model. A rating schema was used to define and assess the collaboration quality of collaborative activities. The results of the model were compared to the ratings of collaboration quality obtained by using the schema. Correlation analysis and error measures were computed to verify the results of the study.

The article is organized as follows. In Section 2 we provide a brief review regarding previous studies on analysis of collaborative activities and time series analysis. In Section 3, the methodology followed in the study is described. In Section 4, we review the structure and the use of the classification model as well as techniques used in the classification setting. In Section 5 we present the results of the model. Further analysis of the results and findings are discussed in Section 6. In Section 7 the conclusions, further improvements and future work are presented.

2 LITERATURE OVERVIEW

A wide range of analysis methodologies has been proposed, focusing not only on the social aspect of collaboration or on group dynamics but also on the way computers support collaborative activities [3]. The use of computers gives the opportunity to record ongoing activities and allows their post-examination through logfile analysis. The groupware applications that facilitate computer supported collaboration provide the capability of automatically recording the activity into structured logfiles. In early studies simple metrics such as the sum of messages or the average number of words [4, 5] were used to assess a collaborative practice. In later studies similar metrics describing the density of the activity in combination with qualitative analysis were exploited to analyze the interaction of students in a CSCL setting [6]. Social Network Analysis (SNA) techniques use similar metrics to represent interactions between users of a network [7]. SNA techniques have also been used in CSCL to analyze the practices of learners [8, 9]. However, it is argued that such kind of metrics and analysis schemas may provide false assessments regarding collaboration quality and outcome [1]. The main disadvantage of these metrics is that they rely heavily on the overall volume of the activity and do not take into account the user interactions within the collaborative context and the contribution of individuals to the group progress. More sophisticated metrics that aimed to describe interaction rather than the volume of activity alone, were further introduced. Various frameworks use annotated events to construct activity metrics [10, 11, 12] or the notion of symmetry to describe interaction [13, 14]. Other frameworks propose more complicated metrics with respect to the spatio-temporal characteristics of user activity [15].

Time is usually a key factor of analysis. The way a collaborative process unfolds in time may offer an indication of the collaboration quality or the activity outcome. Long periods of inactivity, uneven distribution in time and conflicts due to simultaneous user activity, can be indications of problematic situations. With this rationale a number of event metrics have been proposed [16, 17]. However, and besides individual studies, no systematic use of time series techniques has been attempted. The use of time series will potentially:

1. provide a descriptive way to represent the unfolding of collaboration practice,
2. give the opportunity to explore analysis models and techniques widely used in other research fields over the past decades,
3. reveal richer information on the kind of interaction that takes place.

The conclusion of an activity is not required for the application of time series analysis since the method does not rely on overall statistics. Therefore it could be proven useful for the real-time evaluation of a collaborative activity and the support of collaborators through automated feedback.

Any sequence of observations recorded at successive time intervals can be characterized as time series. These observations can be single or aggregated events within certain time frames. Time series find application in numerous fields such as

engineering, economics, biology or environmental studies. The main objectives for using time series analysis are the efficient description of data, the understanding and interpretation of concurrent relationships that occur within a certain setting and the forecasting of future values from past or current ones. Various methodologies have been proposed to model and support time series analysis, such as Hidden Markov Models (HMM), ARIMA and VAR models and Dynamic Time Warping (DTW) [18]. Time series fall into two categories regarding construction and analysis: univariate and multivariate. Univariate time series are those consisting of single observations repeating in time while a multivariate time series is a vector of multiple observations repeated in time that aim to describe a single process. Multivariate time series are preferred, in general, in order to enhance the effectiveness of analysis in terms of better understanding of dynamic relationships as well as to improve the accuracy of a model [19].

Time series have a wide range of application but are rarely met in the analysis of collaborative activities. ARIMA modeling has been used to describe long-term activities of teams working over the internet [20]. However, such time series techniques are difficult to adapt due to the fragmented and complex nature of collaborative activities [2]. To avoid complexity and other constraints imposed by ARIMA and similar modeling techniques, the current study makes use of a memory-based learning model. The Dynamic Time Warping algorithm (DTW) is used to measure the similarity of the time series that represent collaborative activities. The methodology of the study is described in the following section.

3 METHOD OF STUDY

The activities where people work together towards a common goal are subsumed under the term collaborative activities. In this article we deal with collaborative activities in a learning context. The analysis of such activities and the evaluation of their quality is a popular field of research within the CSCL community. Collaborative activities can be represented in various ways, i.e. employing the use of graphs, textual representation or formal language. With the proposed use of time series we aim to describe collaboration efficiently in the dimension of time. Possible interferences and dependencies of concurrent activities can be mapped and further studied. Moreover, we propose a method to classify collaborative sessions in a way that will depict their quality.

The classification is carried out by a supervised, machine learning algorithm (near neighbor classification) implemented by a memory-based classification model, we have named TSCMoCA (Time Series Classification Model for Collaborative Activities). For the classification procedure a training and a test set were required. The training set consists of collaborative activities that can be described as ordered pairs (ts_i, CQA_i) , where:

- ts_i is the time series representing the activity i
- CQA_i is an evaluative value of the collaboration quality of activity i .

A test set consists of collaborative activities described by time series as well, but of unknown evaluative value of collaboration quality. The goal is to rate the test set by using the training as a reference set. The method of the study for the classification of collaborative activities with the use of the TSCMoCA model is depicted in Figure 1.

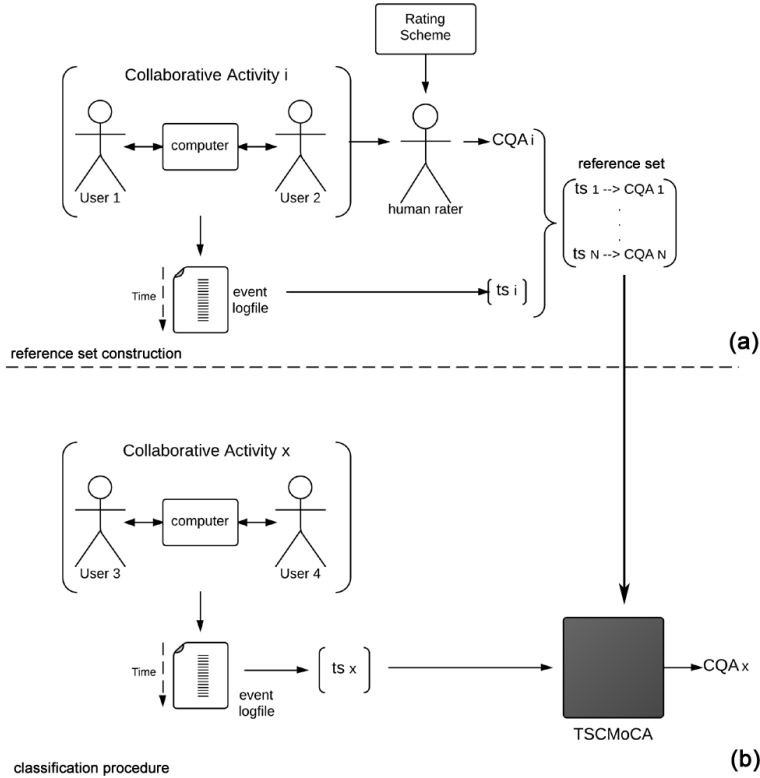


Figure 1. Method of study: Classification of a collaborative activity ts_x using the memory-based model TSCMoCA. Sub-diagram a) describes the construction of the reference set and sub-diagram b) represents the runtime process of the classification

The reference set (also referred as “memory” or training set) was constructed from collaborative activities that took place during a programming course in the Department of Electrical and Computer Engineering. The purpose of the course was the joint creation of algorithm flowcharts. Dyads of students were asked to collaborate to that end. Duration of the activity was about one and a half hour. The activity was supported by a groupware application used in numerous, similar settings. The groupware provides a shared workspace for the creation of flowcharts or any diagrammatic representation and a chat tool that facilitates the communi-

cation between partners (Figure 2), [17]. The activity of each collaborative session is recorded into logfiles according to the OCAF coding scheme [21]. Each event (or action) is added as a new record in the logfile of the activity in the form of XML-like entry, as following:

[< ID >, < timestamp >, < actor >, < event-type >, < attributes >],

where: < ID > is an incremental identifier, unique for each event; < timestamp > represents the time when the event occurred; < actor > the user responsible for the event; < event-type > the type of the event (e.g. chat message, insert/edit/delete/move/resize object in the common workspace); < attributes > a field related to information about the type of event such as the coordinates of an object on the common workspace or the context of a chat message.

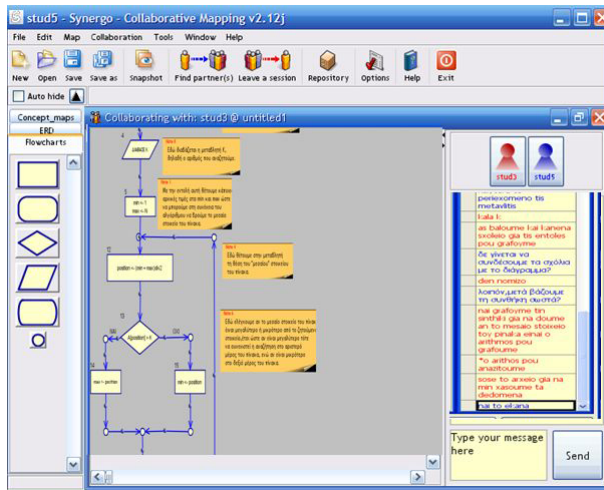


Figure 2. The user interface of the groupware application Synergo, consisting of the libraries toolbar, common workspace for the creation of diagrammatic representations and a chat tool that supports users' communication. Synergo was used to mediate the collaborative activities studied in the article.

The activity is recorded into logfiles in intervals of one second by the relay server. The relay server is also in charge of the communication between the collaborators. This setting can raise synchronization issues due to network delays. Therefore we ensured similar network performance applying for all participants: the users who took part in the activity were participants in laboratory classes of equal size (about 30 users per class); all participants used almost identical computers regarding both software and hardware and they were connected over a high-speed local area network. This setting ensured that the potential delays would be minimized and roughly the same for all users.

The dataset gathered from the experimental procedure consisted of 228 collaborative sessions and was evaluated regarding the quality of collaboration in a previous study [22]. The objective of the original study was the adaptation and application of a rating scheme for the assessment of the quality of collaboration. The rating scheme introduced seven, collaborative dimensions that stand for five, fundamental aspects of collaboration. Table 1 presents the seven collaborative dimensions on which the human evaluators rated the collaborative activities.

The first six dimensions (collaboration flow, sustaining mutual understanding, knowledge exchange, argumentation, structuring the problem solving process, cooperative orientation) describe the collaborative aspects of the activity. The seventh dimension (individual task orientation) takes into account the activity of each user separately. Therefore the dimension of individual task orientation was not included in further analysis. We gain an overall notion of the collaboration quality by rating each one of the six collaborative dimensions. The assessment of the generic dimension of collaboration quality (Collaboration Quality Average – CQA) is computed as the average of the six collaborative dimensions ratings. The rating of the dataset was carried out on a five-point Likert scale ($[-2, +2]$) by two previously trained expert evaluators.

General Aspect of Collaboration	Rating Scheme
Communication	Collaboration Flow (CF)
	Sustaining Mutual Understanding (SMU)
Joint Information Processing	Knowledge Exchange (KE)
	Argumentation (ARG)
Coordination	Structuring the Problem Solving Process (SPSP)
Interpersonal Relationship	Cooperative Orientation (CO)
Motivation	Individual Task Orientation (ITO)

Table 1. Collaborative dimensions representing the aspects of collaboration, as defined by the rating scheme

The ratings were examined regarding the absolute agreement between raters and inter-rater reliability. The scores of the reliability tests were satisfying (ICC ranged between 0.83 and 0.95 and Cronbach's alpha ranged between 0.91 and 0.98) suggesting that the rating scheme is a useful means for obtaining consistent measures of overall collaboration quality.

The final, evaluated dataset of 228 sessions was used in the current study. Each session was split into time frames and the aggregated events per time frame were used for the construction of multivariate time series. Earlier studies revealed that certain event metrics, such as the number of messages partners exchange (COA), the amount of workspace actions (WOA) and the number of role switches in both the dialogue and workspace activity (CALT), (WALT) are significantly and highly correlated with the quality of collaboration [23]. However, these metrics do not capture the change of activity in time. We introduced a new set of metrics that

represent activity change – or difference – between consecutive time frames. The new metrics are represented as (DCOA), (DCALT) for the chat tool and (DWOA), (DWALT) for the common workspace. Table 2 presents the eight events that were used overall.

In order to fully explore the significance of various event metrics, three case studies were designed and carried out. Different combinations of events were used for the construction of time series in each case study. The question we aim to answer with this design is whether the combined use of event metrics of different nature (sums, differences, rates) provides more information and leads to improved classification. The three cases are described as follows:

- Case A.** Four metrics describing volume of activity were used for the construction of the time series of the first case (COA), (CALT), (WOA), (WALT).
- Case B.** Four metrics describing changes of activity volume were used for the construction of the time series (DCOA), (DCALT), (DWOA), (DWALT).
- Case C.** The total of the eight metrics of an activity were used for the construction of the time series in the third case (COA), (CALT), (WOA), (WALT), (DCOA), (DCALT), (DWOA), (DWALT).

	Activity Volume	Role Switches
Chat Tool	number of chat messages per time frame (COA)	sum of role switches in chat activity per time frame (CALT)
	difference of chat messages per consequent time frames (DCOA)	difference of role switches in chat activity per consequent time frames (DCALT)
Common Workspace	number of workspace actions per time frame (WOA)	sum of role switches in workspace activity per time frame (WALT)
	difference of workspace actions per consequent time frames (DWOA)	difference of role switches in workspace activity per consequent time frames (DWALT)

Table 2. Event metrics used for the construction of time series

In all three case studies, various time frames were studied. The size of time frames is of great importance when it comes to time series of aggregated events. Valuable information may be lost or concealed if the time frame is too small or too wide [24]. On the other hand, the context and the domain of the research field as well as the duration of the monitored activity should be taken into consideration. It is obvious, for example, that for asynchronous communication as usually employed in message boards different time frames should be used than for the synchronous collaboration of two partners via an online messenger. Therefore, a wide range of time frames should be tested. In a prior study [25], the time frames of 1, 5, 8 and 10

minutes were explored. The analysis pointed out that the best results were obtained for the minimum time frame of 1 minute. In the current setting we have extended the time range by adding the frames of 10, 15, 30 and 45 seconds.

4 METHODOLOGY OF ANALYSIS

The proposed approach uses memory-based learning for the classification of collaborative activities. Memory-based learning is a popular machine learning technique, used in various research fields such as in robotics [26] and language processing [27]. It is commonly used for the classification and evaluation of unknown samples with respect to their nearest match from a set of predefined or pre-evaluated samples. Memory-based learning models that make use of time series have been proposed in various fields [28].

4.1 Memory-Based Learning Classification Model TSCMoCA

As aforementioned, a memory-based model was built to classify collaborative activities by finding the most similar match within a reference set of pre-evaluated activities. The similarity of the activities is measured with respect to their time series.

A memory-based learning model is characterized by three main components:

- The memory, where the reference set of previously evaluated sessions is stored. The dataset of 228 collaborative activities described in Section 3 was used to construct the reference set needed by the model.
- The distance function used to measure the similarity between the reference and the test set samples. The Dynamic Time Warping (DTW) metric was chosen as a distance function. The DTW has been widely used to measure time series similarity and it is further described in Section 4.2
- The number of near neighbors, i.e. the number of similar matches of the under-classification sample. In the present study, the number of near neighbors is set to one.

The memory-based model implements a k-neighbors classification algorithm where k equals to 1 (1NN – Nearest Neighbor). The algorithm is presented in Figure 3. To classify and assign an evaluative value to a new collaborative sample, the most similar match is requested and acquired from the reference set. The operation of the memory-based model is divided into two stages: the classification of the sample and the assessment of the quality of collaboration. The model requires as input multivariate time series ts_x of the collaborative activity x that we wish to classify. The output of the model is the nearest matching neighbor and the quality of collaboration of the activity x (CQA_x).

4.1.1 Classification of a Query Sample

Every new collaborative session that has not been evaluated is named “query sample” and serves as an input to the model. The classification is carried out in regard to the DTW distance between the query sample and each one of the samples of the reference set. To evaluate the collaboration quality of a new session, its time series is compared to the time series of the sessions in the reference set using the DTW algorithm. DTW is further described in Section 4.2. The result is a distance matrix which contains the DTW distances of the query sample and the total of samples placed in the memory. Since the DTW provides a distance metric, the optimal match for the query sample is the one for which the DTW distance is minimum.

In order to assess the results of the model and in the same time to fully take advantage of the large dataset, all 228 collaborative sessions were used for both the training and test set in an iterative procedure where a different session was chosen each time as an entry for the test set, leaving the remaining 227 in the training set and placed back when the next session was chosen. This is a widely used cross-validation technique known as Leave-One-Out.

4.1.2 Evaluation Assessment of the Query Sample

We argue that collaborative activities that unfold in similar ways will also share similar qualitative characteristics. Consequently, collaborative sessions that are described by similar time series will also have a similar evaluative value for the general dimension of collaboration quality (CQA). In case we want to evaluate a new collaborative session X that is represented by time series ts_x , DTW algorithm is used to determine the similarity of the query sample X with each sample in the memory. If DTW distance is minimized for the sample Y of the reference set ($DTW_y = \text{minimum}$), then the value for collaboration quality of the session Y (CQA_Y) is assigned as the collaboration quality for session X ($CQA_X = CQA_Y$). The classification and evaluation algorithm is displayed in Figure 3.

The six dimensions of collaboration (Table 1) are also assessed in the same way as the CQA. The ratings assigned to each dimension range within $[-2, +2]$. In order to evaluate the assessments of the model, the automated ratings were compared to the ratings of human evaluators. From that comparison, correlation and error analyses were carried out. The correlation coefficient and the mean absolute error were studied to assess the validity and accuracy of the model. The results of the experimental procedure are presented in Section 5.

4.2 Dynamic Time Warping

The Dynamic Time Warping algorithm (DTW) is used to measure similarity between two series of events that may vary in time or speed. It was originally used for speech recognition and video processing. The DTW algorithm does not presuppose a stationary time series and is not affected by missing values. It is widely used to

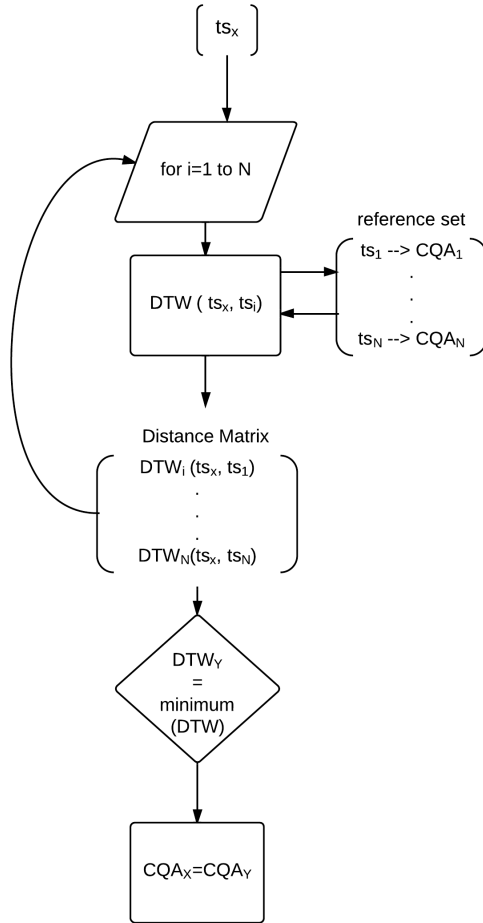


Figure 3. Flowchart of the classification process of a new collaborative session ts_x

measure the similarity of series in various research fields such as sound and video processing as well as time series analysis [29, 30].

In the current setting, the DTW algorithm was used to measure the similarity between multivariate time series that represent collaborative activities. In particular, we made use of the DTW algorithm for the R-statistics software implemented by Giorgino [31]. The DTW algorithm allows the identification of similar patterns between two series X and Y with different phases. The first step of the algorithm is the computation of the cross-distance matrix of X and Y , using a dissimilarity metric such as Euclidean or Manhattan distance. Then the X and Y series are warped and compared. The output of the Dynamic Time Warping algorithm is a distance

measure (DTW distance) per each couple that stands for similarity: the higher the DTW distance, the more dissimilar the time series.

The aforementioned implementation allows the researcher to experiment with various parameters such as the step patterns and dissimilarity methods used, e.g. Euclidean and Manhattan distances. The choice of the most adequate dissimilarity method is highly dependable on the context as well as on the nature of the data. Previous work [25] has shown that the use of Manhattan distance as a dissimilarity metric provides better results than Euclidean distance. It is evident that Manhattan distance performs better than Euclidean in the case of high dimensional data [32, 33]. In the presented experimental setup we make use of multidimensional time series. Therefore it is expected that the Manhattan distance will perform better, in particular as the multi-dimensionality of the time series increases. For the purposes of the study both distance metrics (Euclidean and Manhattan) were used and it was confirmed that the Manhattan distance performs better. In this article we provide the results only for the use of Manhattan distance.

5 RESULTS

The results of the experimental procedure are presented next. The dataset consists of 228 collaborative sessions. A part of the dataset, where outliers were excluded, was used in a preliminary study [25]. As outliers we identify collaborative sessions which appeared to have some kind of anomaly in their logfiles such as large time gaps of inactivity or violent interrupts. In the present study we use the full dataset since large time gaps or breaks may occur due to network issues or other unexpected events but they might also indicate a problematic turn in the activity. Nevertheless, this is still a situation that the orchestrator has to deal with and recover from. Moreover, in case we wish to provide real-time, online evaluation and feedback, such anomalies should be dealt with. For each one of the 228 samples its nearest match was found in the remaining 227 sessions of the data pool. The study was repeated for three cases, regarding the event metrics used for the time series construction (see Section 3), for eight time frames (10, 15, 30, 45, 60, 300, 480 and 600 seconds) and for the Manhattan distance as a dissimilarity function.

Correlation and error analyses were used to gain understanding of both the validity and accuracy of the results. In particular, the ratings of the model for each collaborative session were compared to ratings of human evaluators. The correlation and the estimation error among the two were computed. Correlation is a popular statistical method to describe the association or dependence between two variables. Spearman's Rank Correlation in particular, is a widely known, non parametric measure of dependence that can be used either for ordinal, interval or continuous variables not normally distributed. It is considered that a correlation coefficient of $\rho = 0.1$ indicates a low correlation, a coefficient of $\rho = 0.3$ a medium one and a coefficient of $\rho \simeq 0.5$ or higher, denotes a high correlation [34, 35]. Although this rule of thumb is usually applied when Pearson's correlation is used, many studies use it

as well for the case of Spearman's Rho correlation coefficient ρ [23]. In the present study all correlations were calculated at the significance level of 0.05 ($p = 0.05$).

A variety of error measures is proposed in literature to estimate the accuracy of a model. Such measures are the root mean squared error (RMSE), the mean squared error (MSE), the mean absolute error (MAE) and so on. For the purposes of the study, the root mean squared error (RMSE) and mean absolute error (MAE) in combination were used since they are simple to understand, interpret and they are measured in the same units as the dataset. The RMSE is sensitive to occasional large errors or outliers while this is not the case for MAE [36]. When RMSE and MAE are used together, we gain insight on the variance in errors. The larger the difference (RMSE-MAE), the higher the variance in errors. There is no absolute or safe criterion to define a "good" or acceptable value of the mean absolute error (MAE) or the root mean squared error (RMSE) since they rely heavily on a number of factors such as the data scale and the research field. Therefore this metric should be used in accordance with other metrics and for partial comparison to other models. However, a rule of thumb followed in similar studies, implies that a value of MAE of less than 1 is considered acceptable for the current setting [23].

In order to fully explore the possibilities and effectiveness of the model, various cases were studied. Besides the various time frames, a combination of different event metrics used for the construction of time series, was also explored in three case studies. In Case A, time series were constructed from the sums of main events of the groupware application and were used as a model input. In Case B, the difference of sums of main events per consecutive time frames was used. In Case C, the total of the metrics of both Case A and Case B were used for the construction of time series. In that way we aim to explore which kinds of events (sums or differences) are more influential, describe the activity efficiently and examine whether their combination leads to an improved outcome.

To evaluate the proposed method, the ratings of the model are compared to the ratings of human evaluators. The correlation and error analyses are presented for the three cases. Based on the results, the most appropriate setting is chosen and further analysis is carried out, not only in regard to the generic dimension of collaboration quality (CQA), but also to each one of the six, collaborative dimensions defined by the rating scheme (Table 1). However, we must keep in mind that in order to fully evaluate the quality of a model, there are other factors that should be taken into account: the simplicity of the model, ease of adaptation in various settings and its performance in comparison to other models in a particular field.

5.1 Correlation and Error Analyses of Case Studies

5.1.1 Case A

In the first case study (Case A), the time series used as model input were constructed from event metrics that represent the volume of activity. The assessments of the model significantly correlate with the ratings of the qualitative rating pro-

cedure for all the time frames studied and for Manhattan distance (Table 3). The correlation coefficient ρ takes values within the range $[0.147, 0.297]$, indicating low to medium correlations between the model and human ratings. The lowest correlation is observed for the time frame of 480 seconds and the highest for the 30 seconds frame. The behavior of the correlation coefficient does not follow a specific pattern regarding the time frames. However it appears to maximize for the smaller time frames (Figure 4). The first case study (Case A) appears to have the highest error for the majority of time frames, for all the cases studied. The MAE and RMSE display similar behavior, indicating similar variation in errors for all time frames. The frames of 15 to 60 seconds have the lowest error which increases in the border frames (10, 600 seconds) (Figure 5). For all time frames the mean absolute error (MAE) is greater than 1 which is considered high. Overall, the minimum error and the highest correlation are both observed for the time frame of 30 seconds ($\rho = 0.297$, MAE = 1.050).

5.1.2 Case B

In the second case (Case B), the metrics that were used represent the change of activity volume between consecutive time frames. Statistically significant, low to medium correlations occurred for a number of time frames, mostly the small ones ($\rho \in [0.178, 0.278]$) (Table 3). The correlation coefficient ρ maximizes for the small-sized time frame of 15 seconds and minimizes for time frames of bigger size (Figure 4). Regarding the accuracy of the model, the error increases with the size of a time frame, being however lower than the error in Case A. The MAE and RMSE follow a similar pattern (Table 4). The minimum MAE is observed for the time frame of 15 seconds, likewise the highest correlation ($\rho = 0.278$, MAE = 0.925) while it is less than 1 for most small-sized time frames (15–30 seconds) (Figure 5).

5.1.3 Case C

In the third case (Case C), the time series are constructed from the combination of event metrics used in Case A and Case B. The correlation coefficient ρ takes values within the range $[0.171, 0.365]$. Correlations follow a certain pattern where the lowest correlations occur for the border frames (10 and 600 seconds) and the highest correlation appears for the middle time frame of 60 seconds. The behavior of the correlation coefficient per time frame is displayed in Figure 4. The behavior of MAE and RMSE for Case C is similar to Case B but the error is lower for most time frames (Figure 5). The error is minimized for the 60 seconds time frame in which the correlation coefficient (ρ) also scores the highest value ($\rho = 0.365$, MAE = 0.928).

Correlation and error analyses showed that the combined use of event metrics of different nature improves the classification and evaluation results. Overall, the size of the time frame does not appear to have a particular impact in Case A, either in terms of correlation or error. The strongest correlations and minimum errors occur in small sized time frames. However, the behavior of correlation coefficient

CQA			
Time Frame (sec)	ρ_A	ρ_B	ρ_C
10	0.178		0.171
15	0.256	0.278	0.271
30	0.297	0.201	0.304
45	0.223		0.266
60	0.204	0.219	0.365
300	0.259	0.178	0.249
480	0.147		0.243
600	0.221		0.180

Table 3. Correlations ρ between the model and human ratings of Collaboration Quality (CQA) for the three case studies and per various time frames

and mean absolute error on the whole, do not follow a certain pattern. In the second case (Case B), the results of the two rating processes (human and model ratings) did not reveal significant correlations for all time frames. However the ratings correlated stronger and with a minimum MAE for the small time frames (15–60 seconds). In Case B the event metrics portray activity rate. The results indicate that small time frames are more appropriate to depict the change of activity volume in time. Therefore the combined use of metrics from cases A and B is expected to improve the results, especially in the case of small time frames, as confirmed in Case C. The improvement is particularly evident for the small time frames (15–60 seconds). As the results suggest, smaller time frames capture meaningful activity more efficiently than larger ones. In Figure 6, the correlation coefficient and the mean absolute error (MAE) for Case C are displayed. The ratings of human evaluators and the model ratings correlate significantly and with a mean absolute error of less than 1 for the majority of time frames. The best results occurred for the 60 seconds time frame where a high correlation coefficient and a low MAE error are combined ($\rho = 0.365$, $MAE = 0.928$). Border frames show poor results due to little activity or over-densed events which potentially lead to information loss. The correlations and errors studied so far refer to the general dimension of Collaboration Quality (CQA).

Time Frame	MAE_A	RMSE_A	MAE_B	RMSE_B	MAE_C	RMSE_C
10	1.267	1.576			1.003	1.291
15	1.178	1.472	0.925	1.190	0.944	1.218
30	1.050	1.341	0.969	1.174	0.945	1.200
45	1.137	1.440			0.990	1.264
60	1.124	1.430	1.067	1.350	0.928	1.190
300	1.150	1.411	1.240	1.546	1.132	1.416
480	1.177	1.466			1.173	1.495
600	1.178	1.471			1.151	1.425

Table 4. Assessment error (MAE and RMSE) ρ of the model vs. human ratings of Collaboration Quality (CQA) for the three case studies and per various time frames

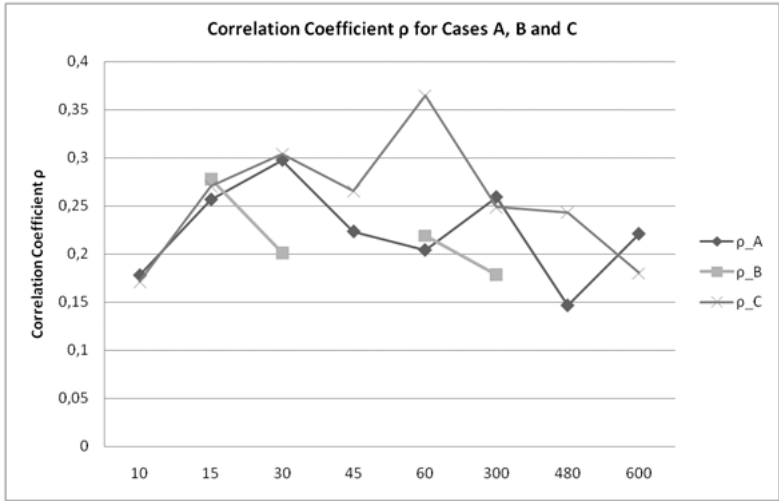


Figure 4. Correlation coefficient ρ for human vs. model ratings of Collaboration Quality (CQA) per case study and various time frame

5.2 Assessments per Dimension of Collaboration

The rating scheme used to assess the quality of collaborative activities provides an opportunity to evaluate collaboration with respect to six fundamental aspects (Table 1). Various issues that affect a collaborative activity may be brought into light, giving the opportunity to the orchestrator to provide specialized, real time feedback. The use of the model, as presented in Case C, for the rating of each collaborative dimension is further studied. The correlation coefficient per dimension and time frame is provided in Table 5.

For most time frames, results indicate low ($\rho \approx 0.1$) to medium ($\rho \approx 0.3$) correlations between ratings of the model and human evaluators. The assessments of the model correlate with the assessments of evaluators on a medium level for the time frame of 60 seconds and for the dimensions that describe communication (CF, SMU) and joint information processing (KE, ARG). For the dimension of Structuring Problem Solving Process (SPSP) that stands for the general aspect of coordination, no statistically significant correlation is observed. For Cooperative Orientation (CO), representing the aspect of interpersonal relationship, a low correlation is observed. Higher correlations occur for the time frames of larger size.

The mean absolute error (MAE) per each dimension of the rating scheme is presented in Table 6. The mean absolute error is higher than 1 for all cases. For the time frame of 60 seconds in particular, the error is minimized for the first three dimensions (CF, SMU, KE). For the dimensions Argumentation (ARG) and Cooperative Orientation (CO), the error for the 60 seconds time frame is the second lowest. We should note, however, that the mean absolute error observed per

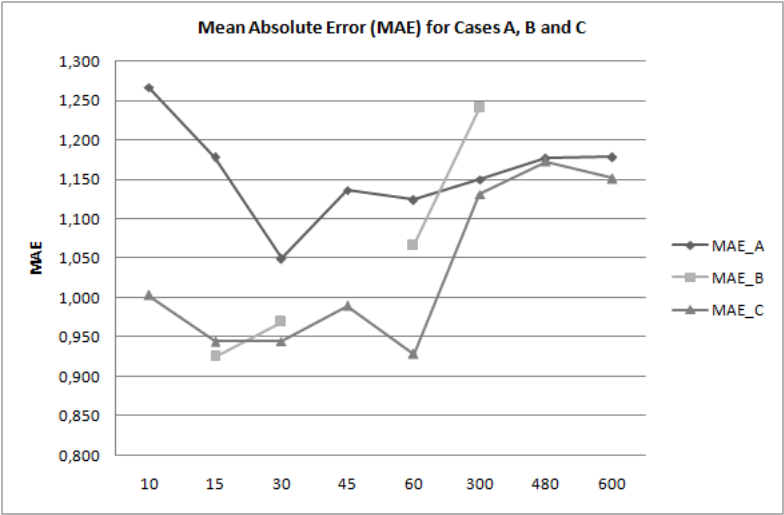


Figure 5. Mean absolute error (MAE) of the model vs. human ratings of Collaboration Quality (CQA) per case study and various time frames

dimension is generally not acceptable and dictates the need for further improvement.

Time Frame	CF	SMU	KE	ARG	SPSP	CO
10	0.268	0.147		0.132	0.132	
15	0.271	0.189	0.283	0.282		0.208
30	0.267	0.150	0.313	0.308		
45	0.276	0.146	0.231	0.194		
60	0.310	0.245	0.327	0.249		0.173
300	0.236	0.145	0.247		0.139	0.294
480	0.257		0.228	0.134	0.201	0.185
600	0.256		0.191			0.214

Table 5. Correlations between the model and human ratings for six collaborative dimensions and per various time frames

6 DISCUSSION

In Section 5, we presented results of the proposed model for the classification and evaluation of the collaboration quality (CQA) of joint activities. Three case studies were carried out. For each case study, time series were constructed by various combinations of event metrics and used as inputs to the model. Numerous time frames were also employed and their effect was explored. Correlation and error

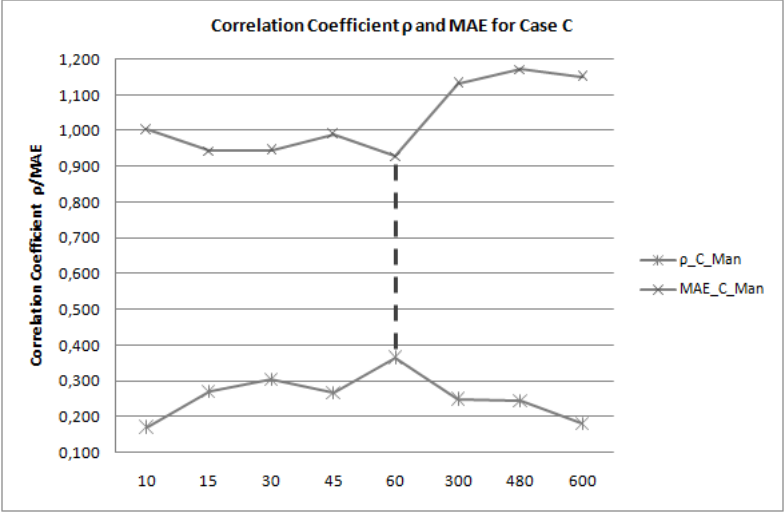


Figure 6. Correlations ρ and mean absolute error of the model vs. human ratings of Collaboration Quality (CQA) for Case C and per time frame

Time Frame	CF	SMU	KE	ARG	SPSP	CO
10	1.180	1.268	1.285	1.386	1.232	
15	1.276	1.351	1.250	1.263		1.118
30	1.149	1.162	1.114	1.171		
45	1.096	1.197	1.224	1.298		
60	1.075	1.114	1.114	1.241		1.145
300	1.237	1.390	1.246		1.360	1.224
480	1.202		1.333	1.504	1.289	1.338
600	1.171		1.325			1.263

Table 6. Mean absolute error (MAE) of the model vs. human ratings per collaborative dimension and various time frames

analyses were carried out to evaluate the results of the model. The results showed that small time frames describe the change of the activity volume in time more efficiently. The strongest correlations and minimum errors occurred in small time frames for Case B. This also applies in Case A even though there was no obvious pattern in the correlation coefficient/error behavior. The combined use of event metrics from Case A and Case B was expected to act additively and to overall improve the model behavior. This was proved in the third case study (Case C), in which the highest correlations and lowest error values occurred. In this case, time series were constructed from eight event metrics representing collaborative activity in the chat tool as well as in the common workspace. The use of small time frames (15–60 seconds) in time series construction further improved the results. Overall,

for the time frame of 60 seconds, the model rated the 66 % of the cases with an error of less than 1 with respect to the human ratings. 91 % of the cases was rated with an error of less than 2 and 99 % was rated with an error of less than 3. While this is a satisfactory outcome, it still leaves room for further improvement.

The results follow a similar pattern in the case of the six collaborative dimensions as well. For the dimensions that cover the general aspects of communication and joint information processing, the best results occur for the time frame of 60 seconds. For the dimensions covering the aspects of coordination and interpersonal relationship, the strongest correlations occur for larger time frames (300, 480 seconds). Most of the event metrics used in time series construction (e.g. number of chat messages, role changes in dialogue, etc.) provide rich information related to communication aspects. This kind of events are found in high density and in small time frames since the chat tool implements instant messaging, with no turn taking or blocking mechanisms and takes advantage of the students' familiarity with similar technologies. Therefore the messages are posted within short time intervals. The aspects of coordination and interpersonal relationship are high-level constructs that build up over longer periods of time. The complicated mechanisms that portray the quality of these constructs, such as efficient management of shared-resources, symmetry and reciprocal activity, require a deeper analysis of user interactions and content analysis [16]. As a result, not only are larger time frames needed to describe these collaborative dimensions but also low-level metrics might not be enough to describe them to a full extent.

A number of related studies within the CSCL context address the temporal nature of collaborative activities and look for patterns that reveal a "cause-effect" or "action-reaction" activity. The notion of activity relevance in time has been previously expressed as temporal proximity and has been used in the creation of contingency graphs for analyzing distributed interaction [37]. It is assumed that the closer any two (or more) actions take place in time, the more relevant they are expected to be. However, it is hard to define what "close" means in terms of time since this is highly dependent on the task, the form of communication and the user characteristics. In similar studies for synchronous collaboration and collaborative puzzle solving, the time frame in which relevant activities took place was experimentally calculated at 25 seconds [15], where for asynchronous interaction over a shared workspace a temporal proximity of 2 minutes is considered evidence of contingency [37]. The results presented here are in agreement with the aforementioned studies since the small time frames of less than one minute present better results. It is therefore safe to assume that the meaningful interaction and significant changes leading to the desirable cognitive effects, mutual knowledge building and eventually successful collaboration, are taking place within the time frames of 30 to 60 seconds in the context of a particular collaborative setting.

The study described in this article was based on a preliminary one that had as a purpose to explore whether time series could be used to describe and analyze collaborative activities [25]. In the present study, a larger dataset was used. Identified outliers were not removed so that the data is representative of real-life scenarios.

Comparing the results of the present study (CS0B) to the preliminary one (CS0A), it was found that for small-sized frames (60 seconds) the correlation coefficient in the current study was improved. On the other hand, the error also increased due to the presence of the outliers in the dataset. For the larger time frames both indices had similar values. In Table 7 and Figure 7, we provide the comparison of the results of the aforementioned studies for Case C and the time frames of (60, 300, 480, 600) seconds.

Time Frame	Correlation Coefficient (ρ)		MAE	
	CS0A	CS0B	CS0A	CS0B
60	0.3	0.365	0.89	0.928
300	0.2	0.249	1.19	1.132
480	0.23	0.243	1.18	1.173
600	0.17	0.180	1.17	1.151

Table 7. Comparison of the results of related studies CS0A (preliminary) and CS0B (current)

The suggested setting can be easily adapted for real-life scenarios, to detect undesirable phenomena and to assist the orchestrator of such activities. The classification model TSCMoCA can be integrated or used in combination with various learning platforms for collaborative activities. The method uses logged activity that is recorded by most applications. The classification algorithm is time-efficient and requires minimum resources. The method requires the existence of a reference set consisting of pre-evaluated activities of similar nature to the ones we wish to classify. The classification algorithm can be used for the post-classification of collaborative activities but also in real time. Real-time classification can provide insights to the orchestrator in order to give feedback and guidance when needed. In Figure 8, the hypothetical use of the model in a real-life setting is displayed. The collaborative activity *CA* is supported by a groupware application that also allows the orchestrator to monitor it. It starts at t_o and at the time instance t_i the orchestrator requests an evaluation of the activity *CA* so far. The model is invoked and provides the orchestrator with an evaluative value $CQA_{CA[ti, to]}$. The orchestrator is then able to decide whether to provide feedback in order to improve the activity or to let it unfold as is.

7 CONCLUSIONS AND FUTURE WORK

This article suggests the use of time series for the automated classification and evaluation of collaborative activities. Based on the findings presented here, we conclude that the quality of collaborative activities may be related to their time series characteristics. The proposed approach involved the evaluation of collaborative learning activities expressed as time series data, based on their similarity to previously

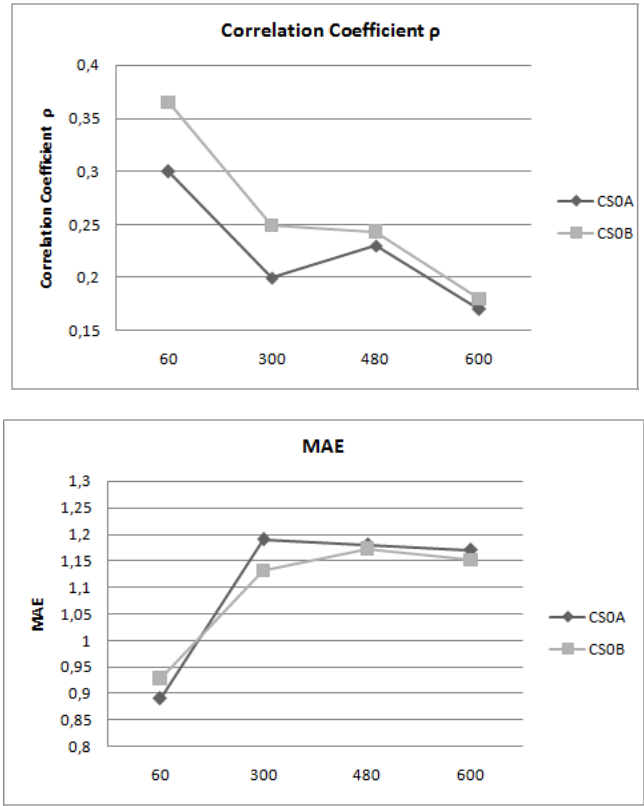


Figure 7. Correlations and mean absolute error of model vs. human ratings for related studies CS0A (preliminary) and CS0B (current)

evaluated ones. The method proposes an innovative way to represent and classify collaborative activities in an automated manner. To that end we introduced the use of time series analysis in a CSCL setting.

In the present study we employed a memory-based learning model for the classification of collaborative activities with the use of a reference set. Three case studies were carried out to evaluate the performance of the model. In each case study, a different combination of event metrics (regarding the nature of the events) was used for the time series construction (Table 2). The ratings of the model for the general aspect of collaboration quality (CQA) as defined by Kahrmanis et al. [22], were compared to the assessments of human evaluators. The analysis of the results revealed a significant positive correlation between the ratings of the model and the ratings of evaluators.

The best results in terms of correlation and error were observed when activity metrics of different nature (sums, differences and rates) were used for the time series

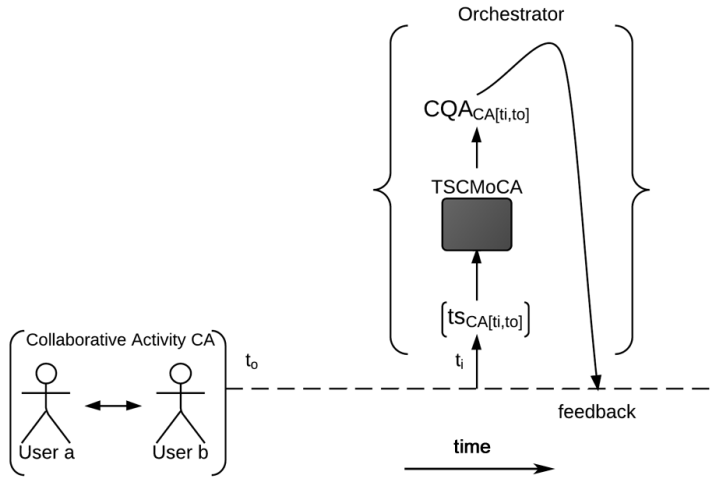


Figure 8. The TSCMoCA model integrated in a real-life setting

construction (Case C). The use of a variety of event metrics and multidimensional time series improved the performance of the model. In addition, the model performed more efficiently for the small sized time frames for all the cases studied. In the case of large time frames, important information regarding users interaction may be lost due to the density of the activity. Likewise, the smallest time frames may fail to include the volume of activity required. This finding indicates that meaningful interaction takes place within small time frames and can be captured adequately using time series techniques. This comes in agreement with other studies [15, 37].

The proposed automated classification and evaluation schema will not only minimize the efforts of human evaluators but could also be used to provide real-time feedback to users. Therefore, the method was also used to evaluate the quality of collaborative sessions with respect to certain aspects of collaboration. The assessment of an activity on various collaborative dimensions could make possible the identification of specific issues. That would further allow the orchestrator to provide users with detailed, personalized feedback when and if needed. The results of the model were satisfactory for the collaborative aspects of communication and joint information processing. More complicated constructs, such as coordination and interpersonal relationship either require larger time frames or low-level activity metrics and are not enough to describe effectively high-level collaborative practices. To that end, future work has to be done to improve the efficiency of the proposed model. The use of advanced classification algorithms, such as the K-Nearest-Neighbor algorithm (where $K > 1$) is expected to improve the results of the model in terms of accuracy.

The method is solely based on activity metrics extracted from logfiles. The content of user activity is expected to provide additional information regarding the quality of collaboration and therefore the use of content and conversation analysis methods can be further explored. Time series also provide the means for visual representation of data. However, apart from traditional visualization techniques, a rich set of methods from graph theory and social network analysis can be applied to time series data [38]. The graphical description of a collaborative activity as a dynamically evolving social network could provide further insight with respect to the sequential nature of the activity and possible patterns of interaction that relate to collaboration quality.

REFERENCES

- [1] STAHL, G.: Rediscovering CSCL. In: Koschmann, T., Hall, R., Miyake, N. (Eds.): CSCL2: Carrying Forward the Conversation. Lawrence Erlbaum Associate, Mahwah, NJ, 2002, pp. 169–181.
- [2] REIMANN, P.: Time Is Precious: Variable- and Event-Centred Approaches to Process Analysis in CSCL Research. *Computer-Supported Collaborative Learning*, Vol. 4, 2009, pp. 239–257.
- [3] STAHL, G.—KOSCHMANN, T.—SUTHERS, D.: Computer-Supported Collaborative Learning: An Historical Perspective. In: Sawyer, R.K. (Ed.): *Cambridge Handbook of the Learning Sciences*. Cambridge University Press, Cambridge, UK, 2006, pp. 409–426.
- [4] HARASIM, L.: Collaborating in Cyberspace: Using Computer Conferences as a Group Learning Environment. *Interactive Learning Environments*, Vol. 3, 1993, No. 2, pp. 119–130.
- [5] BENBUNAN-FICH, R.—HILTZ, S.R.: Impacts of Asynchronous Learning Networks on Individual and Group Problem Solving: A Field Experiment. *Group Decision and Negotiation*, Vol. 8, 1999, No. 5, pp. 409–426.
- [6] LIPPONEN, L.—RAHIKAINEN, M.—HAKKARAINEN, K.—PALONEN, T.: Effective Participation and Discourse Through a Computer Network: Investigating Elementary Students' Computer Supported Interaction. *Journal of Educational Computing Research*, Vol. 27, 2003, No. 4, pp. 355–384.
- [7] SCOTT, J.: *Social Network Analysis: A Handbook*. Sage Publications, 2000.
- [8] MARTÍNEZ, A.—DIMITRIADIS, Y.—RUBIA, B.—GÓMEZ, E.—DE LA FUENTE, P.: Combining Qualitative Evaluation and Social Network Analysis for the Study of Classroom Social Interactions. *Computers & Education*, Vol. 41, 2003, No. 4, pp. 353–368.
- [9] REFFAY, C.—CHANIER, T.: How Social Network Analysis Can Help to Measure Cohesion in Collaborative Distance-Learning. *Proceedings of International Conference on Computer Supported Collaborative Learning (CSCL 2003)*, 2003, pp. 343–352.

- [10] KAHRIMANIS, G.—CHOUNTA, I. A.—AVOURIS, N.: Study of Correlations between Logfile-Based Metrics of Interaction and the Quality of Synchronous Collaboration. *International Reports on Socio-Informatics (IRSI)*, Vol. 7, 2010, No. 1, pp. 24–31.
- [11] HOPPE, H. U.—ENGLER, J.—WEINBRENNER, S.: The Impact of Structural Characteristics of Concept Maps on Automatic Quality Measurement. *International Conference of the Learning Sciences (ICLS 2012)*, Vol. 1, 2012, pp. 291–298.
- [12] BARROS, B.—VERDEJO, M. F.: Analysing Student Interaction Processes in Order to Improve Collaboration. The DEGREE Approach. *International Journal of Artificial Intelligence in Education*, Vol. 11, 2000, No. 3, pp. 221–241.
- [13] MARSHALL, P.—HORNECKER, E.—MORRIS, R.—DALTON, S.—ROGERS, Y.: When the Fingers Do the Talking: A Study of Group Participation for Different Kinds of Shareable Surfaces. *Proceedings of TABLETOP '08*. IEEE Computer Society, 2008.
- [14] BUISINE, S.—BESACIER, G.—AOUSSAT, A.—VERNIER, F.: How Do Interactive Tabletop Systems Influence Collaboration? *Computers in Human Behavior*, 2011.
- [15] SCHUMMER, T.—STRIJBOS, J. W.—BERKEL, T.: A New Direction for Log File Analysis in CSCL: Experiences with a Spatio-Temporal Metric. In: Koschmann, T., Suthers, D., Chan, T. W. (Eds.): *Computer Supported Collaborative Learning 2005: The next 10 years!* Erlbaum, Mahwah 2005, pp. 567–576.
- [16] MEIER, A.—SPADA, H.—RUMMEL, N.: A Rating Scheme for Assessing the Quality of Computer-Supported Collaboration Processes. *International Journal of Computer-Supported Collaborative Learning*, Vol. 2, 2007, No. 1, pp. 63–86.
- [17] AVOURIS, N.—MARGARITIS, M.—KOMIS, V.: Modelling Interaction during Small-Group Synchronous Problem Solving Activities: The Synergo Approach. *2nd International Workshop on Designing Computational Models of Collaborative Learning Interaction (ITS 2004)*, 2004, pp. 13–18.
- [18] CHATFIELD, C.: *The Analysis of Time Series: An Introduction*. Chapman & Hall, CRC Texts in Statistical Science, 2003.
- [19] BOX, G. E. P.—JENKINS, G. M.—REINSEL, G. C.: *Time Series Analysis: Forecasting and Control*. Wiley, 2011.
- [20] VASILEIADOU, E.: Stabilisation Operationalised: Using Time Series Analysis to Understand the Dynamics of Research Collaboration. *Journal of Informetrics*, Vol. 3, 2009, No. 1, pp. 36–48.
- [21] AVOURIS, N.—DIMITRACOPOULOU, A.—KOMIS, V.: On Analysis of Collaborative Problem Solving: An Object-Oriented Approach. *Computers in Human Behavior*, Vol. 19, 2003, No. 2, pp. 47–167.
- [22] KAHRIMANIS, G.—MEIER, A.—CHOUNTA, I. A.—VOYIATZAKI, E.—SPADA, H.—RUMMEL, N.—AVOURIS, N.: Assessing Collaboration Quality in Synchronous CSCL Problem-Solving Activities: Adaptation and Empirical Evaluation of a Rating Scheme. *Learning in the Synergy of Multiple Disciplines. 4th European Conference on Technology Enhanced Learning (EC-TEL 2009)*, Springer-Verlag, 2009, pp. 267–272.
- [23] KAHRIMANIS, G.—CHOUNTA, I. A.—AVOURIS, N.: Study of Correlations between Logfile-Based Metrics of Interaction and the Quality of Synchronous Collaboration. *International Reports on Socio-Informatics (IRSI)*, Vol. 7, 2010, No. 1, pp. 24–31.

- [24] SHELLMAN, S. M.: Time Series Intervals and Statistical Inference: The Effects of Temporal Aggregation on Event Data Analysis. *Political Analysis*, Vol. 12, 2004, No. 1, pp. 97–104.
- [25] CHOUNTA, I. A.—AVOURIS, N.: Time Series Analysis of Collaborative Activities. In: Herskovic, V., Hoppe, H., Jansen, M., Ziegler, J. (Eds.): *Collaboration and Technology*. Lecture Notes in Computer Science, Vol. 7493, 2012, pp. 145–152.
- [26] SCHAAL, S.—ATKESON, C. G.: Robot Juggling: Implementation of Memory-Based Learning. *Control Systems Magazine, IEEE*, Vol. 14, 1994, No. 1, pp. 57–71.
- [27] DAELEMANS, W.—VAN DEN BOSCH, A.: *Memory-Based Language Processing*. Cambridge University Press, 2005.
- [28] DENG, K.—MOORE, A. W.—NECHYBA, M. C.: Learning to Recognize Time Series: Combining Arma Models with Memory-Based Learning. *IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA 1997)*, Vol. 1, 1997, pp. 246–251.
- [29] WARREN LIAO, T. Q.: Clustering of Time Series Data – A Survey. *Pattern Recognition*, Vol. 38, 2005, No. 11, pp. 1857–1874.
- [30] CHAO, S.—WONG, F.—LAM, H. L.—VAI, M. I.: Blind Biosignal Classification Framework Based on DTW Algorithm. *International Conference on Machine Learning and Cybernetics (ICMLC 2011)*, Vol. 1973, 2011, pp. 1684–1689.
- [31] GIORGINO, T.: Computing and Visualizing Dynamic Time Warping Alignments in R: The DTW Package. *Journal of Statistical Software*, Vol. 31, 2009, No. 7, pp. 1–24.
- [32] AGGARWAL, C.—HINNEBURG, A.—KEIM, D.: On the Surprising Behavior of Distance Metrics in High Dimensional Space. *Proceedings of the International Conference on Database Theory (ICDT 2001)*, Lecture Notes in Computer Science, Springer-Verlag, Vol. 1973, 2001, pp. 420–434.
- [33] VADIVEL, A.—MAJUMDAR, A.—SURAL, S.: Performance Comparison of Distance Metrics in Content-Based Image Retrieval Applications. *Proceedings of International Conference on Information Technology, Bhubaneswar, India, 2003*, pp. 159–164.
- [34] COHEN, J.: *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed., Lawrence Erlbaum, New Jersey, 1988.
- [35] HOPKINS, W.: A New View of Statistics: A Scale of Magnitudes for Effect Statistics. Available on: www.sportsci.org/resource/stats/effectmag.html#cohen.
- [36] HYNDMAN, R. J.—KOEHLER, A. B.: Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting*, Vol. 22, 2006, No. 4, pp. 679–688.
- [37] SUTHERS, D. D.—DWYER, N.—MEDINA, R.—VATRAPU, R.: A Framework for Conceptualizing, Representing, and Analyzing Distributed Interaction. *International Journal of Computer-Supported Collaborative Learning*, Vol. 5, 2010, No. 1, pp. 5–42.
- [38] HECKING, T.—GÖHNERT, T.—ZEINI, S.—HOPPE, U.: Task and Time Aware Community Detection in Dynamically Evolving Social Networks. *Procedia Computer Science*, Vol. 18, 2013, pp. 2066–2075.



Irene-Angelica CHOUNTA holds a post-doctoral researcher position in the Collide research group of the University of Duisburg-Essen in Germany since August 2014. She studied Electrical and Computer Engineering and received her Ph.D. from the University of Patras, Greece. During her Ph.D. she focused on human-computer interaction, groupware evaluation, technology enhanced and computer-supported collaborative learning. Her research interests mainly focus on learning analytics for monitoring, mirroring and guiding learning activities, group formation and group dynamics in collaborative and cooperative settings.



Nikolaos AVOURIS is Professor of Human-Computer Interaction in the Department of Electrical and Computer Engineering, University of Patras, Greece, and has over 25 years of experience in industry and academia in Greece, UK and Italy. He serves in the editorial board of many journals and has been keynote speaker and program chair of international conferences. He is the founder and head of the ITLab Human-Computer Interaction (HCI) Group of the University of Patras (<http://hci.ece.upatras.gr/>) that is active since 1994. His research mainly focuses on human-computer interaction, software usability

and accessibility, interactive systems design and evaluation, internet and agent-based technology, computer-supported collaborative learning and work, intelligent interfaces and operator support systems, data visualisation, knowledge discovery in data, context-aware computing and usability of mobile devices. Application fields are education, environmental and various industrial areas. He has published 6 books and over 200 refereed journal and conference papers.