Behavior analysis of students in video classes

Francisco Genivan Silva
*Federal Institute of Education,
Science and Technology - IFRN*
Currais Novos, Brazil
genivan.silva@ifrn.edu.br

Thiago Reis da Silva
*Federal Institute of Education,
Science and Technology – IFMA*
São Raimundo das Mangabeiras,
Brazil
thiago.reis@ifma.edu.br

Alan de Oliveira Santana
*Postgraduate Program in Systems and
Computing – PPgSC
Federal University of Rio Grande do
Norte – UFRN*
Natal, Brazill
alandeoliveirasantana@gmail.com

Eduardo Aranha
*Dept. of Informatics and Applied
Mathematics - DIMAp
Federal University of Rio Grande do Norte
– UFRN*
Natal, Brazil
eduardoaranha@dimap.ufrn.br

Abstract— **This Research-to- Practice full paper presents the result of a study whose core is characterized by the application of data analysis techniques from the use of video lessons. Distance education contributes to the generation of educational data to conduct investigations of educational processes. Methodologies for the automatic extraction of useful information from large volumes of data, especially data mining, have significantly contributed to improvements in the field of education. However, most traditional methods focus entirely on the data or how it is structured, with no significant concern with the educational process as a whole. Also, little attention has been given to information related to the students' behavior during the use of educational resources and media. Therefore, we consider that examining the students' behavior during the execution of educational videos can contribute to a more accurate analysis of the quality of the resources used, the way they learn, and what factors influence their learning. Thus, this research to practice full paper presents conducts a study to investigate how students behave during the use of video classes and seek to highlight the benefits of this type of analysis for education. Results demonstrate that evaluating the educational resource through the analysis of the actions involved can contribute substantially to the activities performed by teachers and content makers.**

Keywords—educational, videos, srudents, behavior, analysis, evaliation

I. INTRODUCTION

Educators have been recording instructional videos for almost as long as this media format has existed [1]. Especially in the last decade, since the emergence of free online video hosting services, the movement for creating and disseminating content has reached a global scale. The extensive use of video as a teaching resource makes us reflect on the methods used to extract and analyze useful information about the relationship between resource, behavior, and learning.

Before addressing this relationship, it is necessary to discuss another issue related to the evaluation of educational resources. First, since previous decades, evaluation is seen as less important than the development processes themselves [2]. Consequently, educators, in general, are not used to being evaluated or frequently evaluate the resources created or used by them. In parallel to this difficulty, those who seek to initiate a process of evaluating educational resources come up against both the lack of knowledge and the distance between the solutions proposed by the academy and the daily practice of these actors in a real environment.

The presentation of these questions does not imply that computing and its multiple dimensions are not advancing or contributing significantly to solving these and other problems. In fact, in recent years there has been a more significant concern with the evaluation of educational resources. Studies such as those by Giannakos, Chorianopoulos and Chrisochoides [3] and Muñoz-Merino *et. al.* [4] bring significant evidence that the evaluation of educational resources provides benefits for teaching and learning.

In the case of video lessons analysis, we find in the literature, mainly, two motivational lines: the segmentation for identification and efficient retrieval of information in the video (where in the video certain information is located) and the analysis of user interaction to understand how the videos are watched (excerpts more and less viewed, for example). In both cases, the main instrument of analysis has been data mining. Grouping of students by clustering, outlier detection, mining association rules, pattern mining, text mining [5]. However, most of these traditional data mining techniques focus on simple sequential data or structures, rather than complete process models with simultaneity patterns [6]. For example, students' actions in different teaching environments generate a heap of data about their own behavior and the environment, but it is not clear how we can create complex models that provide information about the learning process involved. We can know that a student viewed a video and in which parts he stopped, but we were unable to create, through data mining, a visualization model based on this information. It is not possible to define which paths within the video led to a specific pattern of behavior and we were unable to extract a model of viewing the video itself. For these reasons, process mining is proposed as an alternative to carry out this type of assessment. Precisely, the basic idea of process mining is to discover, monitor and improve

real processes (that is, processes that are neither considered nor truncated), extracting knowledge from the event logs automatically registered by the Information Systems [7].

Given the concerns and problems raised, the objective of this work is to expand the investigations through the application and evaluation of process mining techniques applied to the analysis of video classes. The techniques applied use data extracted from the students' behavior when watching video classes and aim to generate useful information for teachers and the academic community.

This work also discusses the value and implications that methods used to analyze data from video classes can bring to the educational process since the recent pieces of research under this bias are still very incipient. The results achieved demonstrate that the techniques employed and the data used contribute to a deeper and more assertive understanding of how students learn from video classes, thus enabling the teacher to reflect on his or her educational material and the video production process.

II. PROCESS MINING

The unprecedented accumulation of data in recent years also creates an enormous amount of event data, which are actions carried out in the daily lives of organizations or by system users, such as booking an airline ticket, analyzing a customer's credit, and solving a test in a MOOC.

These events are part of the organizational activities and, therefore, of the business processes. Thus, the increasing inter-organizational competitiveness and the continuous need for evaluations of the organizations' activities demand exceptional care with the monitoring of these processes. The challenge, therefore, is to use event data in a meaningful way, such as checking system processes, identifying constraints, anticipating problems, and discovering insights [8].

Process mining responds to this challenge by representing a bridge between two major areas: (i) the classic process models analysis, which is common in Business Process Management (BPM), and (ii) the data-oriented analysis, such as data mining and machine learning. Process mining has been able to unite these two areas because at the same time that it focuses on processes, it uses real-world data. In data mining, for example, the techniques employed generally do not focus on processes, especially end-to-end processes.

To assess, analyze, or improve educational processes more thoroughly, a new type of analysis was needed, in which the process plays a central role. The Educational Process Mining (EPM) emerged precisely for this. EPM is a relatively new field linked to Educational Data Mining, concerned with discovering, analyzing, and improving educational processes based on the data hidden in the educational systems' records [9, 10].

The event logs extracted from information systems are the starting point for process mining. Despite being able to hold a broader set of components, [11] state that the events contained in the log should have, whenever possible, three minimum characteristics:

- they must refer to an activity, that is, a well-defined step of a process;

- they must refer to a case (process instance); and,

- they must have an ordering, a timestamp, through which we can observe "paths" and precedence/posterior relationships between the observed activities.

In summary, process mining is an essential tool for organizations that need to manage non-trivial operational processes such as educational processes.

Process mining can be subdivided into three major sub-areas or sets of techniques [6]:

- Discovery of process models: it involves the creation of a process model from the reality observed and is usually stored in logs;

- Conformity check: checking an expected or specified model and its equivalence to the reality observed; and,

- Process models extension: the projection of the information extracted from the logs or the addition of a new perspective and extra features to an existing process model.

Each sub-area of process mining has a set of algorithms that support the resolution of varied problems, each of which presents characteristics.

For the discovery of process models, which is an activity that we performed in this work, we can highlight the Heuristic Miner [12] and the Fuzzy Miner [13] algorithms. Both explore the control flow perspective of a process model. For this, they consider the order of events within a case. The base algorithm first creates a dependency graph based on the frequencies of the activities and the number of times that one activity is followed by another. Based on predefined limits, dependencies are added (or not) to the dependency graph, which reveals the "backbone" of the process model. This backbone is used to discover details of the events' behavior.

The most common way to apply any of the PM algorithms is through tools created for this purpose; the most used ones are the PROM and the DISCO [14]. The PROM software allows the use of a dozen algorithms through plugins, while the DISCO is used in the discovery of models using a single algorithm, the Inductive Miner, which is structured based on the Fuzzy Miner.

III. ANALYSIS OF VIDEO CLASSES

If we can understand the students' actions in learning environments or even the act of watching a video class as part of a broader educational process, then these practices can be investigated and analyzed to extract information that will help to understand this process more deeply.

There are several options for techniques used to analyze video classes. Each technique or set of techniques has a greater adherence to a specific objective. For example, the analysis of physiological data helps to assess the students' emotions, while smart video segmentation improves the retrieval of relevant information.

Amongst the techniques documented in related works, we chose to use log analysis, since it is the mechanism that is most

coherent to the intended objective, which is to analyze the students' behavior. To exemplify the application of this technique and how we intend to use it in this work, we can mention studies with similar objectives: Yin *et. al.* used log files to identify the students' behavior patterns from e-book-based learning records and its impacts on learning [15]; and Chen et al, who, in turn, used log files as a database and developed an interactive visual analysis system to help instructors explore the learning sequences in Massive Open Online Courses (MOOCs) and understand the causes and results behind these sequences [16]. With a bias closer to our work, Beatty, Merchant and Albert used a quantitative approach considering the frequency in which students watched video classes to relate this behavior to academic performance [17]. The work of Kim *et al.* used data from the students' interactions in video classes to identify viewing peaks and recommend parts of the video that were considered most relevant [18].

Student interaction data has been used frequently in the latest years to find learning strategies, create groups with similar characteristics, and define rules of association [19]. However, only recently, video classes started to represent a relevant data source. Studies that suggest the use of process mining (PM) in educational data are still very scarce, and those that use data extracted from video classes are almost nonexistent.

Mukala, Buijs and Aalst [20] is one of the rare examples of the application of PM techniques involving video usage data. Even so, the work does not deal with the students' behavior when watching the video; it only considers the action of watching or not the video and when. These works, like other studies examined, show that the use of log files is an analysis strategy that can produce significant results for the field of education. Also, we assessed the ease of access to log files by teachers and those interested in evaluating the students' behavior in educational settings.

In summary, this work proposes the use of process mining to understand the learning processes based on the actions taken by students in video classes, which are recorded in log files. Throughout the work, we provide some indications of useful ideas and guidelines that can inspire intervention measures to improve the evaluation and use of this educational resource.

IV. STUDY PLANNING

The study reported in this paper was organized and executed so that it was possible to observe students in a learning environment aiming to gather and analyze data on their behavior during the use of video classes. In the study, we investigated the application of process mining techniques as a means to analyze the process of learning using video classes.

A. Context

We performed this study in the context of a competition held between students from the first and second years of high school at a public educational institution [21, 22]. This competition involved solving Computational Thinking problems and had the initial adhesion of 53 students.

The competition was called the "Computational Thinking Challenge". The challenge has a start and an end date. Between these dates, questions were released, and videos of their solutions were recorded. Two questions were released each week of the challenge. A student could answer each question multiple times, and the answers and the dates of each attempt were recorded. The questions could only be answered within the period corresponding to each week of the challenge. The answers could only be viewed after the resolution period. There were no restrictions or limitations for watching the videos after their release to the students.

B. Data Collection

To collect the data and generate the log file, the web environment used in the competition was adjusted to capture the actions performed by the students while they were using the system, especially when they interacted with the video. Among the media resources available, play, pause, and seek (forward or backward) events were recorded.

The scheme of the challenge presented earlier allowed the structuring of events and parameters that should form a log file. As a result, we have the fields shown in Table 1 as attributes recorded in each line of the log file.

As can be seen, the actions saved in the log include the identification of the specific video class, the type of action performed, the time of the day when the actions were performed, the duration of each action, and the exact position in the video in which it was performed.

TABLE I.   MAIN FIELDS REGISTERED IN THE LOG FILE.

Field	Description
Server_date	Date the event was held on the server hosting the application.
Server_hour	Time the event was held on the server hosting the application. Along with the date, it forms the event's timestamp.
User_hour	Time the event was held on the user's computer (used to check the consistency of the data saved in the log).
Session	Navigation session, which is used to differentiate events from the same user at different times.
Email	E-mail address of the student who held the event, which is used to identify the user.
Activity	Action of the user on the website or video (where he or she clicked on the page or action performed on the video, such as play, pause, and seek).
Resource	Page accessed or video that generated the activity record.
video_time	The exact time in the video in which the action was performed.

The data from the log generated was pre-processed and integrated with other information, such as the students' performance in school. The result of the pre-processing and the integration was a CSV file which corresponds to the data set that served as the basis for the analysis presented below.

V. Analysis

The application of process mining was only possible because the data set used contains the minimum characteristics that allow the mining to be carried out: they refer to an activity (for example, watching a video), they refer to a case (instance of a process), and they have an order (sequence of actions in time).

In this first analysis, we considered two types of events: the navigation on the website and the visualization, which occurred during the four weeks of the competition. Fifty-three students from the first and second years of high school participated in the study. Only 29 students had active participation, that is, performed any activity apart from the registration. The data collection module generated the log file that was pre-processed generating a consistent data set for the application of the following methods of analysis.

A. Statistics of the Processes

The first step of the analysis is based on extracting metrics from the activities developed by the students during the competition. To perform this analysis, we used the DISCO tool, which received the pre-processed log file as input.

Table 2 illustrates some statistical data of the events captured each week.

TABLE II. Weekly Data Summary

Description	Week 1	Week 2	Week 3	Week 4
Events	1671	950	719	663
Cases	102	53	51	63
Activities	19	24	32	26
Median duration of the cases	2.1 min	3.2 min	2 min	69 sec
Average duration of the cases	59.1 min	89.2 min	5 min	52.4 min
Start	May 4th	May 11th	May 18th	May 25th
End	May 10th	May 17th	May 24th	May 31st

The table shows that 4003 events were recorded throughout the month, which is the total number of actions performed by all the students. All the actions and the time spent on them, such as opening the competition website, viewing the ranking, or watching a video, are being considered in this analysis. We chose to do a general analysis of the students' actions during the competition before answering specific questions about the use of the videos.

Regarding the average time of each case, it is necessary to make some considerations. Each case was registered considering the user's session, which means that when a student accessed a particular content in the application if he or she did not manually end the session by closing the browser, for example, the time kept being computed. In practice, the values related to the average duration of the cases must have been pulled upwards, which justifies the median and the average values being so far apart from each other. This problem affected the way we interpreted each case since it was not possible to infer the time spent by each student in a session considering only the value recorded in the log. Therefore, we consider that the average time spent per student in each session is somewhat closer to the median values.

The analysis of the time spent in each case offers an idea about the students' engagement during an activity since it does not just show if he or she performed the activity or not. Let us look at the medians, for example. The overall median duration of cases throughout the competition was 2.2 minutes, which means that in less than half of the cases, the session duration was less than this value. We consider this amount of time relatively low, but since the purpose of each session was not assessed separately, we cannot infer other facts from this value. The median values and, probably, the weekly averages were influenced by factors such as student participation and length of the videos. During week 2, the videos released had a longer duration, thus increasing the median value, while week 4 had the lowest student participation, also being the one with the lowest value among the four weeks of the challenge.

Regarding student participation over time, week 1 concentrates 38% of the cases and 42% of the events. That means that almost half of all actions recorded during the competition were carried out in the first week, which shows a high engagement at the beginning of the competition, and a rapid and gradual loss of interest towards the end. The use of continuous rewards, gradient challenges, and medals are elements recognized as gamification factors that help to maintain motivation during a competition. Since these elements were little or not used at all in this study, we can assume that it contributed to the significant decrease in activities over the weeks.

B. Frequency of Activities

One way to deepen the analysis of the metrics presented is to graphically visualize how all the events behave. Studies show that the analysis of the frequency of daily activities in learning environments can be useful to provide relevant information about issues such as dropout and engagement in teaching units [23, 13]. We used the events in the competition log to map how the actions have developed over time. To carry out this activity, we used the PROM tool plugin called Dotted Chart. This plugin gives an overview of the events over time, allowing us to filter and order traces and events according to the purpose of the analysis. Since we sought to analyze the variation in the frequency of activities in a curricular unit similar to a course or discipline, we took as a starting point the fact that the study had a similar structure, with beginning and end, and scheduled dates for tests and material releases.

Figure 1 shows the dot chart generated for the activities that involved performing tasks (sending responses), accessing the ranking (performance), and watching videos (accessing the resource). In the chart, each student session corresponds to a case (an instance of the process), represented by a horizontal line. The chart only has 47 lines, one for each student. This number of students differs from the initial sample because six students did not participate in the competition, despite having registered for it. The dots along each line define a complete

trace, from the beginning to the end of the competition. The colors of the dots differentiate the activities performed, for example, a video watched in a specific week.
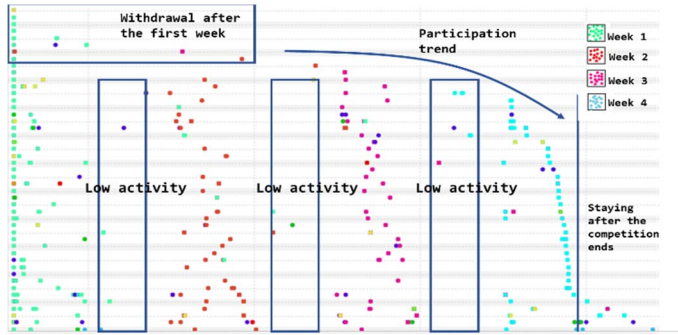

Fig. 1. Dot chart with the students' activities.

The space between the dots shows a higher concentration of activities on the days when the questions were asked. In the chart presented, we used the notion of relative time, in which all the activities start at time 0. Also, we ordered the lines by the duration of the trace.

Although the dot chart presented is specific for this study, if we consider the competition held as a learning unit similar to a course or discipline, the use of this technique allows to have a general view of the students' behavior during its development, making it possible to identify days when students interact more, which activities they are most frequent in, how is their tendency to participate, dropout points, among others. For this reason, we consider it pertinent to use this technique in the analysis of video classes.

C. Views Sequence

The students' behavior in the videos is formally recorded in the log as cases, events, and activities, as demonstrated in the previous analysis. Despite this, certain specific characteristics of their behavior and of the video itself are best seen when we extract a model from these events. This activity - the discovery of models - is a crucial activity and one of the most challenging in process mining [24]. The models usually capture the behavior seen in the log and are more intuitive, facilitating the extraction of useful information.

In this work, during the discovery of the process model, we transform the behavior of students captured in the log into higher-level process models through which we can generate insights and extract relevant information related, for example, to the stages of the learning proces.

To discover the models, we consider only the data related to the students' actions during the use of the videos. Despite being fully applicable to the other type of event collected, the investigation of the navigation on the website is not the focus of this work. The model expected involves the paths taken by the students in the videos, which indicates frequency, precedence, and posteriority of the parts of the video watched.

As happens in data mining, process mining also demands the choice of algorithms that are most suited to the problem. This choice should be guided mainly by the characteristics of the log

to which the algorithms will be applied. Log features such as the heterogeneity of activities, the presence or absence of parallelism, and the repetition of activities need to be taken into account before defining the most appropriate algorithm to be used [25].

During the log analysis, we found two characteristics that guided the choice of discovery algorithms: the log has a wide variety of activities, and there is no parallelism between them within the same case. These two characteristics led to the choice of the Fuzzy Miner algorithm. This algorithm is especially suitable when there are complex and unstructured log data or when we want to simplify the model in an interactive way [13].

In this analysis, each node represents a segment in the video. Whenever a segment was viewed, it was marked as an event. The number of times a segment was seen, along with the pre and post visualization sequence, were used to generate the model using the Fuzzy Miner algorithm. Figure 2 shows the behavior model generated for the first video of the competition.

There are studies and visualization techniques that demonstrate that it is simple to identify where there are sharp peaks and declines in visualization, but we have not identified works in which a method for the "internal" viewing of videos is feasibly presented. Issues such as determining which parts of the video cause viewing spikes or which are the constraints to understanding a video cannot be answered by the graphs extracted from the view count. This type of analysis is performed based on the process models that we present below, through which we complement.
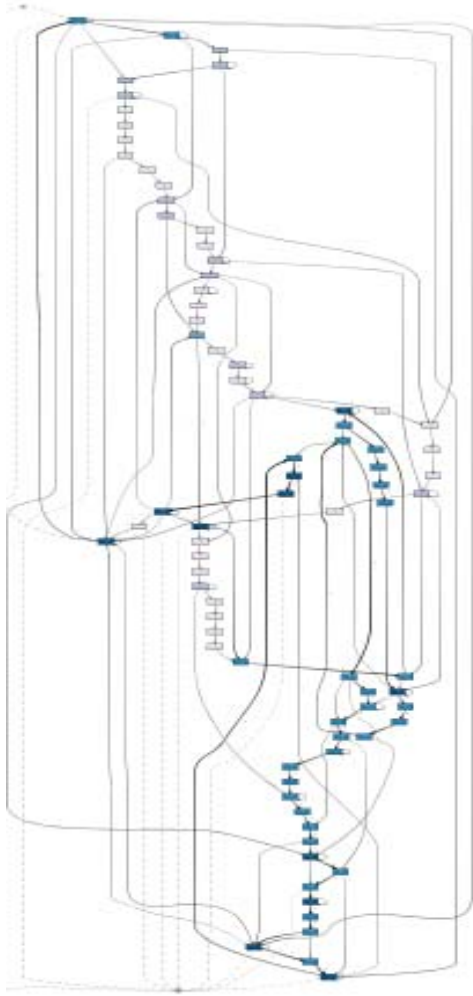
Fig. 2.   Process model generated for video 1.

The model generated presents all the paths taken by the students during the video. Although we have segmented the video into 3-second snippets, the number of activities (segments) and paths makes this model complex and hard to understand. The model generated is similar to many others found in the literature and is known as the spaghetti model. These models present many variants of traces and activities, so it is necessary to use strategies that facilitate their understanding, such as grouping or applying filters on activities [24]. Another strategy is to "cut out" the model, bringing up only the aspects and parts that represent a relevant or distinctive behavior [23]. We chose to use these two strategies to analyze the students' behavior, and we present below what we consider pertinent and indispensable to understand the use of video classes.

To better visualize the segments in the video that present the highest frequency and correlation, we applied a filter on the model that we generated previously, reducing the activities and paths to 10% of the original. Figure 3 shows the new model generated.
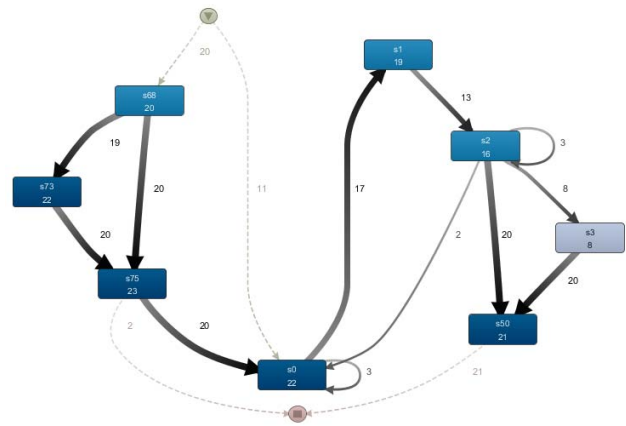

Fig. 3.   Process model reduced to 10%.

The circle with the arrow pointing down shown at the top of the model is known as the starting point. It is an artificial event created by the Disco tool to indicate where the traces start. The color of the rectangles in the model, as well as the width and colors of the arrows, indicate a greater or lesser frequency value of the activities and the links between them. At the top of the rectangles, there is an indication of the video segment (s0, s1, s2), and the number below the segment represents the value of the absolute frequency in which it was viewed. The video analyzed has 240 seconds and was divided into 80 segments. The model generated presents the most viewed segments and from where and to where students jumped in the video.

We can extract some statements from this model. As expected, the initial segments are among the 10% most frequently watched. However, the loss of interest in the video occurs very quickly, starting from s3, as we can note from the following segments, in which frequency decreases and disappears from the model. An interesting aspect to be noted concerns the final segments, which, besides being the most viewed ones, appear next to the starting point. Since the model was reduced, the link between the starting point and segment s68 does not indicate that students started the video at the end, but that they jumped from different parts of the video to that point. Since the segments from which they started are different, their frequency is low, and they are also omitted from the model.

Another interesting feature is related to segment s33. If we restore the model to 100% of the events, we can see this segment as shown in Figure 4.

Segment 33 has a considerable frequency - it was viewed 26 times. The difference is that its viewing does not come only from the previous segment, s32, but from other segments spread throughout the video - in nine times the viewer came from s22; in six times, the viewer came from s79, and in eight times the viewer came from s33 itself. It is not our intention to investigate why this phenomenon happened in this video but to demonstrate that this type of particular behavior characteristic in the video can be identified using the process model generated. The indicated segment, for example, shows a critical point in the video that needs the attention of the teacher who uses or who created it and reinforces that the discovery of models proves to be a powerful technique in the extraction of useful information from the process of using video classes.
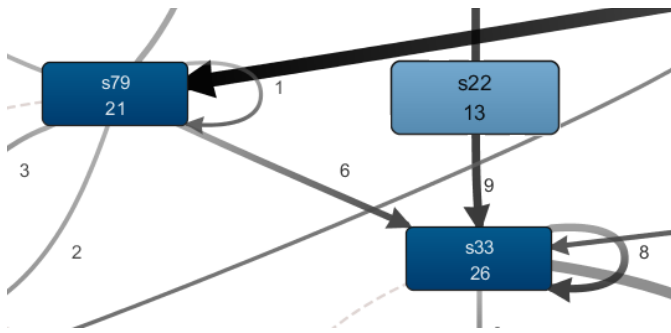
Fig. 4. Cutout of the model with zoom.

D. Relationship Between Views Sequences and Students' Performance

In addition to the evaluation based on video analysis, we can also use model discovery to investigate video-based learning strategies and their relationship to students' performance. For this, we divided the students' data set in two using a performance classification: satisfactory and unsatisfactory. Figure 5 shows a section of the process model generated for the students' behavior in a video. On the left side (Figure 5A), there is a model for students with satisfactory performance and on the right side (Figure 5B), the model for students with unsatisfactory performance.
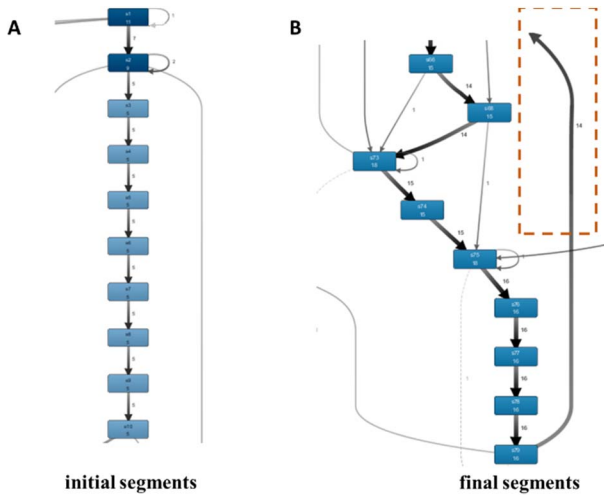

Fig. 5. Process models by performance. (A) satisfactory and (B) unsatisfactory.

The content of the videos used in the study was divided into two parts: the explanation of the problem and the resolution of the problem. Based on this characteristic of the video and the analysis of the models, we can make the following observations. Students have different behavioral models. While students with satisfactory performance viewed the initial segments more frequently and continuously, students with unsatisfactory performance watched the final segments more frequently. Also, students with lower performance returned more frequently to intermediate sections of the video (indicated by the orange dash), such as segments s22 and s33. That is, while students with a good performance show greater interest in the problem's explanation, regular students jump to the resolution, even if, later, they have to return to the explanation.

To verify that these observations were not just accidental, we performed the model discovery for a second video. The result is shown in Figure 6:
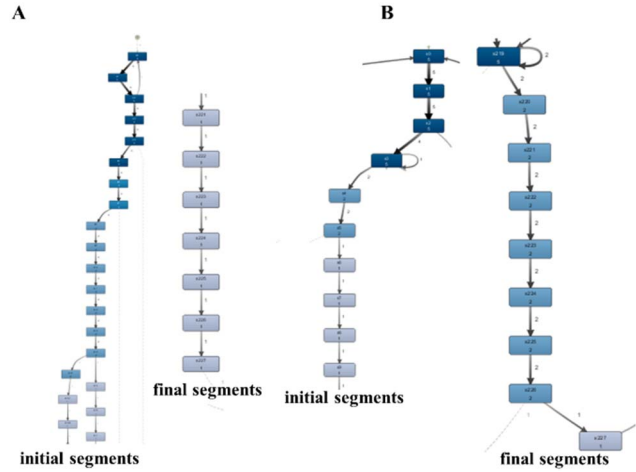

Fig. 6. Process models by performance in video 2. (A) satisfactory and (B) unsatisfactory.

This time we chose to show the two ends of the models, with the initial and final segments for the two groups of students. The discovered models reinforce, even if in a less obvious way, what was previously reported. The model of students with satisfactory performance shows more prolonged engagement in the initial segments of the video, while in the other group, we observed greater engagement in the final segments.

This model discovery-based analysis demonstrates how it is possible to identify patterns of behavior for groups of students with different performances, which represents an important finding when we seek a relationship between the use of video and learning.

As can be seen, the techniques used in this study demonstrate the power of discovering knowledge generated by performing process mining on the students' behavior in video classes. The models discovered enable a global view of the use of videos, which can be understood more easily, both by education stakeholders and by non-specialists in the problem domain.

VI. CONCLUSIONS

As the distance education platforms and services expand, so does the number of video classes and, consequently, the number of views and interactions of the students with this resource. The literature shows that the dynamic analysis of these data allows us to understand better the students' experience and how they learn through video classes, which can significantly improve the teaching and learning process.

This study does not attempt to suggest that the techniques employed replace other types of data analysis, such as data mining, but to present a complementary procedural approach based on activities sequencing that can be replicated by teachers and those interested in technology-mediated education.

When the process under analysis is highly unstructured and contains many exceptions, existing process discovery techniques tend to generate highly complex and

incomprehensible models. In fact, it happened in this work, but this is the nature of educational processes, which are processes that reflect the high diversity of behaviors in the students' habits and learning paths.

Despite this, this work tries to establish a relationship between the process models and the "reality" captured in the form of an event log. Moreover, it was searching for this relationship that we applied some process mining algorithms. The results presented demonstrate the extent and advantages of using data from the students' behavior when watching a video to perform analysis related to educational data. More than that, this study represents a step-in solving problems such as the discovery of hidden and useful information from a large amount of log data automatically collected by learning environments.

The main contribution of this work, therefore, is to demonstrate that process mining contributes significantly to the systematization of the analysis of video classes, bringing evidence that the application of these techniques contributes to the effective understanding of the use of videos and the learning process. Secondly, the use of process mining has also brought a greater understanding of the algorithms used for this purpose, also allowing the expansion of studies in a relatively new field of research.

Future works should focus on the implementation of portable solutions integrated into the learning environments or focus on the development of Educational Process Information Systems, in which process mining techniques can be applied more easily and with greater correction.

REFERENCES

[1] Yin, P. J.; Guo, J. K.; R. Rubin. "How video production affects student engagement: an empirical study of MOOC videos". In: Proceedings of the first ACM conference on Learning @ scale conference. Pages 41-50, 2014.

[2] Albion, P. "Heuristic evaluation of educational multimedia: from theory to practice". ASCILITE 1999: 16th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Responding to Diversity, Brisbane, Australia, December 1999.

[3] Giannakos, M. N.; Chorianopoulos, K.; Chrisochoides, N. "Collecting and Making Sense of Video Learning Analytics". In: IEEE Frontiers in Education Conference (FIE), 2015.

[4] Muñoz-Merino, P. J.; Ruipérez-Valiente, J. A.; Alario-Hoyos, C.; Pérez-Sanagustín, M.; Kloos, C. D. "Precise Effectiveness Strategy for analyzing the effectiveness of students with educational resources and activities in MOOCs". In: Computers in Human Behavior, 2015.

[5] Silva, F. G., "Video Analytics – Systematic Review of the Literature", unpublished.

[6] Rozinat, A.; Aalst, van der, W.M.P. "Conformance checking of processes based on monitoring real behavior". In: Information Systems. V.33, I. 1, p.64-95, 2008.

[7] Aalst, van der, W. M. et al., "Process mining manifesto," International Conference on Business Process Management Workshops (BPM 2011), pp. 169–194, 2011

[8] Aalst Van Der, W.M.P. "Process Mining: Overview and Opportunities". In: ACM Transactions on Management Information Systems (TMIS). V.3 I.2, 2012.

[9] Trcka, N.; Pechenizkiy, M. "From local patterns to global models: Towards domain driven educational process mining". In: Intelligent Systems Design and Applications, ISDA'09. Ninth International Conference on. IEEE, pp. 1114–1119, 2009

[10] Awatef, H.; Gueni, B.; Fhima, M.; David, S. "Process mining in the education domain". In: International Journal on Advances in Intelligent Systems, vol. 8, no. 1–2, 2015.

[11] Trčka, N.; Pechenizkiy, M.; Aalst, van der, W. M. P. "Process mining from educational data (Chapter 9)". Handbook of Educational Data Mining, CRC Press, pp. 123-142, 2010.

[12] Weijters; A. J. M. M.; Aalst, van der, W. M. P.; Medeiros, A. K. A. "Process mining with the heuristics miner-algorithm". In: Technische Universiteit Eindhoven Technical Report WP; 166: 1-34. 2006

[13] Günther, C. W.; Aalst, van der, W. M. P. "Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics". Alonso G., Dadam P., Rosemann M. (eds) Business Process Management. BPM 2007. Lecture Notes in Computer Science, vol 4714. Springer, Berlin, Heidelberg, 2007

[14] Ghazal, M. A.; Ibrahim, O.; Salama, M. A. "Educational Process Mining: A Systematic Literature Review". 2017 European Conference on Electrical Engineering and Computer Science (EECS), pp. 198-203, Bern, 2017

[15] Yin, C.; Uosaki, N.; Chu, H.-C.; Hwang, G.-J.; Liu, G.-Z.; Hwang, J.-J.; Hatono, I.; Kumamoto, E.; Tabata, Y. "Learning Behavioral Pattern Analysis based on Students' Logs in Reading Digital Books". In: International Conference on Computers in Education. New Zealand: Asia-Pacific Society for Computers in Education, 2017.

[16] Chen, Q.; Yue, X.; Plantaz, X.; Chen, Y.; Shi, C.; Pong, T.-C.; Qu, H. "ViSeq: Visual Analytics of Learning Sequence in Massive Open Online Courses". In: IEEE transactions on visualization and computer graphics, 2018, PP. 10.1109/TVCG.2018.2872961.

[17] Beatty, B.; Merchant, Z.; Albert, M. "Analysis of Student Use of Video in a Flipped Classroom", In: TechTrends, 2017.

[18] Kim, J.; Guo, P.; Cai, C.; Li, S. W.; Gajos, K. Z.; Miller, R. C. "Data-driven interaction techniques for improving navigation of educational videos". In: Symposium on User Interface Software and Technology - UIST'14. 2014

[19] Romero, C.; Ventura, S.; Pechenizkiy, M.; Baker, R. "Handbook of Educational Data Mining". Boca Raton, FL: CRC Press, Taylor&Francis, 2010.

[20] Mukala, P.; Buijs, J. C. A. M.; Aalst Van Der, W. M. P. "Exploring students' learning behaviour in MOOCs using process mining techniques". In: BPMcenter. 26 p. (BPM reports), 2015.

[21] Silva, T. R.; Aranha, E. H. S. "Online game-based programming learning for high school students — A case study" in: 2015 IEEE Frontiers in Education Conference (FIE), El Paso, TX, 2015, pp. 1-8, doi: 10.1109/FIE.2015.7344131.

[22] Regalado, M. R.; Aranha, E. H. S.; Silva, T. R. "Gamifying an online approach for promoting game development Learning and contest: An experience report". In: 2016 IEEE Frontiers in Education Conference (FIE), Erie, PA, USA, 2016, pp. 1-8, doi: 10.1109/FIE.2016.7757441.

[23] Aalst, van der, W. M. P.; Guo, S.; Gorissen, P. "Comparative Process Mining in Education: An Approach Based on Process Cubes". In: IFIP International Federation for Information Processing, SIMPDA 2013, LNBIP 203, pp. 110–134. 2015. DOI: 10.1007/978-3-662-46436-6 6

[24] W.M.P. Aalst, van der. "Process Mining: Discovering and Improving Spaghetti and Lasagna Processes". IEEE Symposium on Computational Intelligence and Data Mining (CIDM), 2011

[25] E. Gupta. "Process Mining Algorithms". In: International Journal of Advance Research in Science and Engineering 3(11), .401-412, 2014.