Methodology of measure of similarity in student video sequence of interactions.

Boniface Mbouzao
Polytechnique Montréal
boniface.mbouzao@polymtl.ca

Michel C. Desmarais
Polytechnique Montréal
michel.desmarais@polymtl.ca

Ian Shrier
McGill University
ian.shrier@mcgill.ca

ABSTRACT
MOOCs make extensive use of videos. Students interact with them by pausing, seeking forward or backward, replaying segments, etc. We can reasonably assume that studens have different patterns of video interactions, but it remains hard to compare student styles of video interactions. Some methods were developed, such as Markov Chain and edit distance between event sequences. However, these methods ignore the time spent between events. This paper proposes a new methodology of comparing video sequences of interaction based both on time spent in each state and the succession of states by computing the distance between the transition matrices of the video interaction sequences. The proposed methodology can measure the level of similarity of interaction between two video sequences of interaction and determine if two different viewers have the same video interaction style through a better distinction between representation of sequences of interactions with video.

Keywords
MOOC, Distance matrix, Level of similarity, Markov Chain, Optimal Maching Distance.

1. INTRODUCTION
In online learning contexts, learner engagement is often measured by their interaction with video. Some researches identified frequent pauses or stops as a sign of less engagement than frequent play [11, 12]. Two common methods used to find the similarity between two video sequences of interactions are the Markov Chain and the measure of distance between sequences of interactions through various algorithms. The main limitation of using Markov Chain to compare video sequences of interactions is that states transitions probabilities do not take into account the time between states. Many sequences can have the same transitions probability matrix but represent different styles and length. By contrast, the distance method of comparing video interaction sequences takes time into account, but it will fail to find a similarity when the activities have a large offset (translation), in spite of similar transitions between activities (Markov Chain). A methodology that can simultaneously take into account the time and transition between activities could help the analysis of video interaction. It could help the analysis of the MOOCs and online teaching systems learning in video intensive environments, and could help to extract meaningful patterns of video interactions that represent a challenging task for the research community (see for eg. [25, 18, 2]).

2. EXITING METHODS AND ISSUES
We first review the basics of the two families of methods and explain in more details their issues, before describing previous work with each approach, and then describe and evaluate the proposed method.

The collection of the data from student interaction with videos is done in a such a way that an event occurs when the student change a state. For example if a student is in "play" state and swish to "pause" then the system of data collection sent and event "pause". By collecting al the events it is then possible to have succession of event of same student with a video and the student time in each state. For example we have those two students interactions as: Student 1: Play (4 seconds) then Pause (4 seconds) and then Play (4 seconds), Student 2: Pause (2 seconds) then Play(8 seconds) and then Pause (2 seconds). The each student spent 12 seconds in total interaction with video. one can the transform those two patterns of interaction into a sequence of interactions. If we consider that a transition last one second then when a student stays 3 seconds in one state, then there is 3 transitions from one that to the same state. And we obtain then those two sequences of the two student as:
Sequence 1: Pl-Pl-Pl-Pl-Pl-Pa-Pa-Pa-Pa-Pa-Pl-Pl-Pl-Pl-Pl
Sequence 2: Pa-Pa-Pa-Pl-Pl-Pl-Pl-Pl-Pl-Pl-Pl-Pl-Pa-Pa-Pa
(Pl=Play video and Pa=Pause video, sequence 1 for student 1 and sequence 2 for student 2)

2.1 Markov Chain method
A Markov chain is specified by a set of states $S = \{s_1, s_2, s_3, ..., s_t\}$. The process starts in one of the state s_i, then move from one state to another s_j with a probability of $p_{i,j}$. The Markov property stipulates that the transition probability is independent states prior to s_i.

A student interaction sequence can be represented as a Markov state transition matrix, where cells represent frequencies of transitions in the sequence, normalized such that row sums are all 1 and thus represent transition probabilities. A measure of distance between matrices can be used as a means of comparison.

The limit of using Markov Chain to compare video interaction sequences lies in the fact that transitions probabilities can be the same for very different sequences. For example, the two sequences above coming from the two students interaction scenarios (sequence 1 and sequence 2) have the same Markov matrix and their transition probability matrix describing their Markov Chain in a 2-states space is:

$$mat_{\text{seq1}} = mat_{\text{seq2}} = \begin{array}{c c} & \begin{array}{cc} \scriptstyle play & \scriptstyle pause \end{array} \\ \begin{array}{c} \scriptstyle play \\ \scriptstyle pause \end{array} & \begin{pmatrix} 8/9 & 1/9 \\ 1/5 & 4/5 \end{pmatrix} \end{array}$$

2.2 Sequence distance method

Longuest Common Prefix algorithm was proposed by Elzinga in 2007 [9] and aims to measure the similarity/distance between sequences. It is based on the length of the longest common prefix (LLCP). For the purpose of this study we tried this algorithm and with a short example of its application one can find out that it is not suitable while comparing the similarity between the entire two sequences.

Longest Common Subsequence algorithm is another one of the metrics considered by Elzinga [9] that is available through the seqdist() function of R and compute the length of the longest subsequence between two sequences.

Optimal Matching distance algorithm (OM) generates distances that are the minimal cost in terms of insertions, deletions and substitutions for transforming one sequence to another. The cost of each deletion or insertion is 1 by default. This algorithm was originally been proposed by Levenshtein (1966) and has been popularized in the social sciences by Forrest (1986) and Abbott (2001). The algorithm implemented in TraMineR is that of Needleman and Wunsch (1970). The result is a substitution cost matrix that gives the distances between all sequences of the list of sequences considered.

Using then OM algorithm for this study could be the best way to find closeness between the sequences where each insertion, deletion and substitution has the same cost which is 1. The limit of using this method is that during computation of distance matrix the overall video style of interaction is not taken in account. For example using the same principle to obtain sequences, the two sequences bellow (sequence 3, sequence 4) have largest distance between them while using OM algorithm while in fact they are both same style of interaction.

Sequence 3: Pl-Pl-Pl-Pa-Pa-Pa-Pl-Pl-Pl-Pa-Pa-Pa
Sequence 4: Pa-Pa-Pa-Pl-Pl-Pl-Pa-Pa-Pa-Pl-Pl-Pl
$distance(sequence3, sequence4) = 12$ which is the largest distance among two sequences of length of interactions of 12 for OM distance. This means that those two sequences are completely dissimilar which is not the case in reality.

Another issue is when two sequences have different length. In this case the distance among them depends on the longest sequence (number of insertions), though very difficult to find similarity. In other hand, this methodology is very efficient for sequences of same range of length of interactions.

In the state of art, the usage of OM distance is to aggregate together the closest sequences through clustering. The major challenge of this method is how to define the number of clusters. In other words, what are the distance thresholds to define closeness between sequences? In general when two sequences are in a same cluster they are considered as similar. Another drawback of this methodology resides in the fact that it is based only on distance between two sequences to measure the level of similarity between them: what is the threshold distance to compute the percentage of similarity among two sequences based on the succession of their transitions?

3. RELATED WORK

3.1 Student sequence of activities

Clustering of students activities based on sequence distances was done by Desmarais et al. [8]. Activities are defined as answer to exercise, browsing through to the exercises, browsing through notes, pausing without any activity in last 5 minutes, answering to an exercise, viewing the score, and the login activity. They studied the sequence of activities and visualized the patterns of study of college math learning.

This classification in terms of activities was exported to specific video interaction. In that sense, a clustering was done for MOOCs by Nan Li et al. [21] to explore the link between video interactions and perceived video difficulty. Here, they used video event instead of activities through the platform to generate each student sequence of interactions.

The high level of dropout in the MOOC system led to many studies to investigate how learners could be classified between those who will stay and those will stop before the end based on the sequence of interaction [19, 31] and how to intervene to help learners to keep on using the system [30]. The purpose was to evaluate the engagement of students [26] and to predict their performances [15]. Some studied the structural configurations in learning activity sequences of students in the MOOC [28, 29]. Some of those studies were able to find learning processes from student behaviours and using pattern mining to classify students according to their skills [10]. In online learning systems, especially in MOOCs, one of the major learning support materials are videos. The study of learners' behaviour during the learning process in terms of their engagement [17], social engagement [4], or their performance [14] can be used to predict dropouts [13] or to analyse demographics [12]. Chen Liang et al. [5, 6, 7] studied the online internet video-on-demand systems to build an accurate model to characterize the distribution of user watching time on a per video basis. They find out that the video watching time can be modeled by a concentration of exponential distribution at the beginning of the videos and truncated power law distribution for the rest of the video as soon as the user watches the video without interruption. They proposed a new model with more detailed alternative to closed queuing network formulation that can help to measure popularity of a video. As concerns videos as a major means of delivering MOOC content, analysing the different type of student video interaction to perceive student video difficulty can potentially help instructors to identify videos that particular students might have trouble with [16] and give them proper help to increase their engagement and prevent dropout [20]. In some courses researchers found that the majority of the students spend almost all their time on watching videos [3, 27]. Guo et al. [11] studied from an empirical point of view how student engagement in a MOOC videos is measured by how long the student watched each video and whether the student attempts to answer post-video assessment problems. The findings are that short, fairly fast videos interspersed with an instructor's talking head or a slide are more engaging that any other MOOC learning resources [24].

3.2 Markov Chain utilization on student sequence of interaction

Mongy et al. [23, 22] clustered video watching behaviors as first order Markov chains in MOOCs in order to extract types of observed behaviors. They asked a domain specialist to analyze their clusters and named them: "fast viewing of video", "viewing of a specific video sequence" and "a complete viewing of the video". They clustered users' behavior by applying the k-means algorithm to produce viewing models that are quite distinct and cover the diversity of observed behaviors.

3.3 Measure of student sequence similarity

In recent years Atapatuu and Falkner [1] explored how video engagement behavior of students in MOOC impacts their success or failure. By analyzing 1.5 million video in-depth, they explored the rational behind the time-wise variation of video interaction. Yet, they focused more on how discourse features correlate with patterns of MOOC videos interactions.

Nirmal Patel and al. [25] investigate on the mining of frequent learning pathways from a large educational dataset. A methodology of pattern mining were introduced using a large graph of sequence pathway and extraction of principal pathways. To obtain this graph they used clustering method following similarity amount student sequences computing Levenshtein distances and using Ward's method. The major challenge of using clustering methods is to define the proper number of clusters. There is not best method yet to find the number of significant cluster in a data. The limit of this methodology is thus link to the limit of the usage of clustering techniques. A pathway that is distributed in various clusters might not be significant in each cluster but while detected all together can be a significant pathway. Thus, there is a need of a methodology that can give all pathway with the number of sequence in each pathway can better help to specify the importance of each pathway. Their methodology does not give how close are the different sequences. A need of a methodology that could give a better measure of similarity between sequences of interaction for such researches.

Klingler et al.[18] clustered temporally based student sequences of interactions in MOOC. The main contribution of their research is to be able to have consistent clusters of sequences. They proposed an evolutionary clustering pipeline that can help to detect the learning evolution over time of learners along with relevant cluster evolution over time. With this method they were able to detect important student behaviors over time and the learning environment properties. Even though they used Markov chain in their pipeline, they are not capturing all types of behavior with number of sequences using each pathway.

Shirvani et al. [2] extracted latent study patterns from students sequences of interactions. They used two approaches known as hypothesis driven and data-driven. They predefined patterns according to the first student activity in the sequence. They then clustered student according to their first activity to have the Markov chain of each group of student. In some student's sequence of activity data, the first student's event cannot define well the sequence of student activities. In those cases their proposed method could not be suitable in comparing two videos sequence of interactions.

4. PROPOSED METHOD

The proposed method is a combination of two techniques: one based on sequence matrix of transitions and the other using OM distance measure. The combination of results give a full similarity between each pairs of student's sequences of interactions benefiting of advantages from both techniques.

4.0.1 Transition matrix

The video transition matrix of a student s for video i is expressed as:

$$tm_{s,i} = \begin{array}{c} \\ \scriptstyle load \\ \scriptstyle play \\ \scriptstyle pause \\ \scriptstyle seek \\ \scriptstyle stop \end{array} \begin{array}{c} \begin{array}{ccccc} \scriptstyle load & \scriptstyle play & \scriptstyle pause & \scriptstyle seek & \scriptstyle stop \end{array} \\ \begin{pmatrix} a_{1.1} & a_{1.2} & a_{1.3} & a_{1.4} & a_{1.5} \\ a_{2.1} & a_{2.2} & a_{2.3} & a_{2.4} & a_{2.5} \\ a_{3.1} & a_{3.2} & a_{3.3} & a_{3.4} & a_{3.5} \\ a_{4.1} & a_{4.2} & a_{4.3} & a_{4.4} & a_{4.5} \\ a_{5.1} & a_{5.2} & a_{5.3} & a_{5.4} & a_{5.5} \end{pmatrix} \end{array}$$

where $a_{j,k}$ is the number of transitions from event j to event k and $tm_{s,i}$ is the transition matrix of student s interacting with video i.

In the case of video interactions, a new event occurs when a student stop an activity by starting another. For example a student pause a video while watching it. The click to "pause" create a new event which is going from "play" state to "pause" state (changing the state from state "play" to state "pause"). Those kind of change increase by one in the matrix the element represented as $a_{i,j}$ (i and j representing two different states). In the case where no event occurs that means that the student is in the same state: playing video or pausing video or other state, the increase of the matrix element $a_{i,i}$ is done following the maximum number of transitions possibles withing the time spent in that state counting the transition from one state to the same state flowing the idea of equation 2 in the next section. Then each $a_{i,i}$ in the matrix represents each time the number of possible transition that could happened when a student stay in a state i without changing to another state. Each time that a student reach a state i the count of $a_{i,i}$ starts increasing its value by the unit of transitions per unit of time spent in that state until the student click to go to another state j and in this case the matrix element $a_{i,j}$ is increase by one and the count in state j start to the increase of $a_{j,j}$ and so on.

4.1 Distance between two transition matrices

The distance between two student transition matrix is express as:

$$d(tm_{s1,i}, tm_{s2,i}) = \sqrt{\sum_{j=1}^{5}(\sum_{k=1}^{5}(a_{j,k} - b_{j,k})^2)} \quad (1)$$

where

$$tm_{s1,i} = \begin{pmatrix} a_{1.1} & a_{1.2} & a_{1.3} & a_{1.4} & a_{1.5} \\ a_{2.1} & a_{2.2} & a_{2.3} & a_{2.4} & a_{2.5} \\ a_{3.1} & a_{3.2} & a_{3.3} & a_{3.4} & a_{3.5} \\ a_{4.1} & a_{4.2} & a_{4.3} & a_{4.4} & a_{4.5} \\ a_{5.1} & a_{5.2} & a_{5.3} & a_{5.4} & a_{5.5} \end{pmatrix}$$

$$tm_{s2,i} = \begin{pmatrix} b_{1.1} & b_{1.2} & b_{1.3} & b_{1.4} & b_{1.5} \\ b_{2.1} & b_{2.2} & b_{2.3} & b_{2.4} & b_{2.5} \\ b_{3.1} & b_{3.2} & b_{3.3} & b_{3.4} & b_{3.5} \\ b_{4.1} & b_{4.2} & b_{4.3} & b_{4.4} & b_{4.5} \\ b_{5.1} & b_{5.2} & b_{5.3} & b_{5.4} & b_{5.5} \end{pmatrix}$$

Here i specifying the video considered then $s1$ and $s2$ the two students interacting with the same video i.

4.2 Dissimilarity and similarity measure between videos sequences

The maximum distance between two transition matrices for a video is reached when the first matrix to be transformed to the other need to be changed in each transition through either substitution, deletion or insertion (case of OM distance) or never have same transitions (case of matrix of transition).

The measure of the number of transitions of a student sequence s1 of video i of a given according to the time spent in interaction with video and the number of transitions per unit of time that the system collects can be expressed as:

$$T_{s1,i} = L_{s1,i} * N_{t/s} \quad (2)$$

Where $T_{s1.i}$ is number of transitions of student $s1$ interacting with video i , $L_{s1,i}$ is the length of time spent by student $s1$ interacting with video i in second and $N_{t/s}$ is the maximum number of transitions per second collected by the system. In our case, the system collects 3 transitions per second. This is to avoid skipping transition and be able to keep all students' transitions collected by the system.

The dissimilarity between two sequences of interaction is expressed in terms of percentage as:

$$Dis(tm_{s1,i}, tm_{s2,i}) = \frac{d(tm_{s1,i}, tm_{s2,i})}{T_{s1,i} + T_{s2,i}} \quad (3)$$

Where $Dis(tm_{s1,i}, tm_{s2,i})$ is the level of dissimilarity between $tm_{s1,i}$ and $tm_{s2,i}$ two matrices of two students interactions with video i and $d(tm_{s1,i}, tm_{s2,i})$ is the distance among them. $T_{s1,i}$ and $T_{s2,i}$ are the number of transitions of student $s1$ and student $s2$ sequence of the video i. If $D(tm_{s1,i}, tm_{s2,i})$ is 0 then the two sequences are completely similar and when it is 1 then they are completely dissimilar. Between 0 and 1 shows the percentage of dissimilarity between the two sequences of transitions.

The similarity between two interaction video sequences based on transition matrices with a video i is then expressed as:

$$S_{mat}(tm_{s1,i}, tm_{s2,i}) = 1 - Dis(tm_{s1,i}, tm_{s2,i}) \quad (4)$$

Here $S_{mat}(tm_{s1,i}, tm_{s2,i})$ is the similarity level between sequence of interaction of student $s1$ and student $s2$ of video i using matrix of interactions and $Dis(tm_{s1,i}, tm_{s2,i})$ is the dissimilarity between them.

4.3 Usage of optimal matching (OM) distance

For each pair of sequences, we compute the OM distance to obtain the distance matrix and from there compute the level of similarity among them. The level of similarity between two sequences is compute using OM distance as:

$$S_{om}(seq_{s1,i}, seq_{s2,i}) = 1 - \frac{dist_{om}(seq_{s1,i}, seq_{s2,i})}{max(T_{s1,i}, T_{s1,i})} \quad (5)$$

Where $S_{om}(seq_{s1,i}, seq_{s2,i})$ is the similarity level between sequence of student $s1$ and sequence of student $s2$ of video i and $dist_{om}(seq_{s1,i}, seq_{s2,i})$ is the OM distance between the two sequences and $T_{s1,i}$ and $T_{s2,i}$ are the numbers of transition of the sequence of each student given in equation (2). $max(T_{s1,i}, T_{s1,i})$ is the maximum between the number of transitions of the two student sequences of interactions.

4.4 Combination of results from the two techniques

The last step of this proposed methodology is to combine the two techniques by taking for each pair of sequences the proper level of similarity among the levels given by each techniques. This is to take in account the complementary of those techniques: one can find styles and gives good similarity for sequences of different length and the other gives regularity among sequences and gives good similarity among sequences from same range of length. The final similarity level is then given by:

$$S(seq_{s1,i}, seq_{s2,i}) = Select(S_{om}(seq_{s1,i}, seq_{s2,i}), S_{mat}(tm_{s1,i}, tm_{s2,i})) \quad (6)$$

Where $S(seq_{s1,i}, seq_{s2,i})$ is the level of similarity between sequence of interaction of student $s1$ and sequence of student $s2$ with the video i, $S_{om}(seq_{s1,i}, seq_{s2,i})$ similarity level between the two sequences based on OM distance as expressed in equation (5) and $S_{mat}(tm_{s1,i}, tm_{s2,i})$ similarity level between the two sequences based on sequence matrix as expressed in equation (4).

The function $Select()$ is defined by selecting the proposed similarity when one of the two sequences is less than the half length of the other and select the maximum level of similarity between the proposed method and OM method otherwise. One are takes here the maximum between OM similarity and matrix similarity to avoid the OM drawback of finding dissimilarity between sequences of same range of length but some dis-matching between states as illustrated in section 3.1.2. The flow of the proposed method is illustrated in figure 1 from the sequences to the computation of their similarity level.

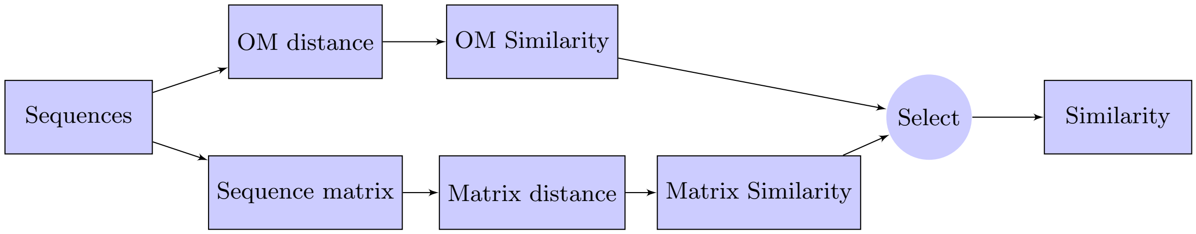


Figure 1: Flow of the proposed method to compute similarity between students' video sequences. "Select" is the selection process between the two technique similarity: when one of the sequence is less than half of the other in length the matrix based similarity is selected other wise the OM based similarity is selected

5. EXPERIMENT

5.1 Method validation

For the validation of the proposed method we compared its capacity of finding level of similarity between sequences and how the proposed method does better than the existing methods. The results then of our proposed method is compared of the results of the Markow chain technique as used by Severin Klingler et al.[18] and the OM based method used for clustering same kind of sequences.

For this purpose we take two main cases: sequences of same lengths of transitions and sequences of different length of transitions. For the same length sequences of interactions, we considered a cyclic same sequence of transitions as illustrated in figure 2. The cycle of transitions is: $Lo - Pl - Pa - Pl - Pa - Se - Pl - St$. The cycle of transition can start anywhere and finish by St for any of the sequence.

The expected level of similarity must be 100% as it is the same sequence following a cycle. The result based on OM distance cannot find that level of similarity as shown in figure 3 (a) compared to the Markov based method in figure 3 (b) (with some exceptions which don't reach the 100% similarity as expected but close enough to be considered as such) and the proposed method in figure 3 (c) (finds perfect match of style by 100% similarity in each case).

In this kind of situations, the proposed method with the Markov based similarity methods are performing better than OM based method (because of the dis-matching disposition of state) in finding similarity between two cyclic same sequences of interactions.

The second validation of the proposed method is to compare its capacities to a Markov based method in different length sequences with known similarities. For this purpose, we considered four sequences of same transitions levels as shown in figure 4. In this case, the percentage of transition between states are the same, but the time spent in each state are different from one sequence to another. The expected level of similarity depends here on the lengths of each sequence as the succession of states are the same for all four sequences. We should have then as result a progressive increase in level of similarity from the shortest sequence to the longest.

The result from Markov Chain based method as in figure 5(a) could not find the different levels of similarity as the percentage of transition between the states is preserve with different sequences lengths. The proposed method performs

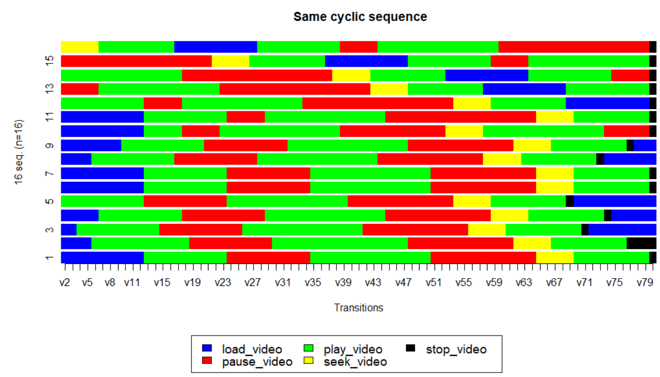
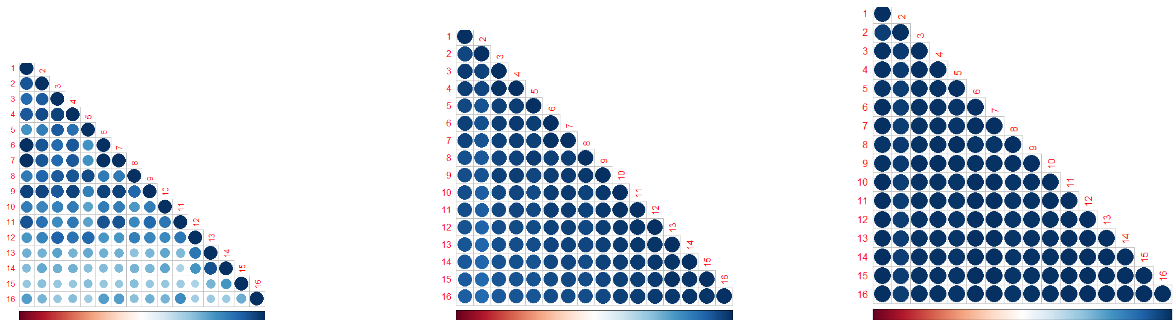


Figure 2: Same cyclic sequence of transitions. The cycle starts and follow the same pattern of transition to close the cycle.

better as shown in figure 5(b) because it is based on number of transitions rather than probability of transition as Markov Chain is.

5.2 Experiments based on real data

The validation of the proposed method on real data is to be able to show that the proposed method has better capacity to identify each sequence of interaction in its uniqueness. It is important that each different sequence of interaction could unique in its representation and so can be predictable. For this reason, for the experiment in this part, we choose three well known classifiers such as support vector machine (SVM), boosted tree (GBM) and k nearest neighbor (KNN) for each method of representation of sequence of interactions to predict first the student and then video to which sequence of interaction belongs. If a specific representation of student sequence of interaction is predictable in terms of which video and student that interact with the video, that means that the representation is able to better distinguish different types of interaction and even showing the specificity of a video in the way that students interact with it. For this experiment, we use the tree (3) different representations (OM representation, classic Markov chain represen-



(a) OM (b) Markov (c) Proposed

Figure 3: Result of similarity: (a) Similarity based on OM distance cannot recognize the fact that those sequences are same cyclic sequence. (b) Similarity based on Markov Chain can recognize the fact that those sequences are same cyclic sequence with some exceptions that not reach the complete 100%. (c) Proposed similarity can recognize the fact that those sequences are same cyclic sequence.

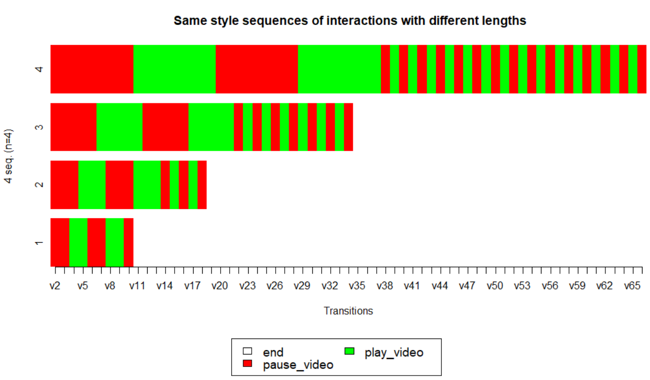


Figure 4: Same sequence of states sequence with different lengths.

(a) Markov (b) Proposed

Figure 5: Result of similarity: (a) similarity based on Markov Chain cannot recognize the duration in each state as the probability of transition between the states are preserve over the length of each sequence. All the sequences as considered as same even they are not quite the same based on duration in each state. (b) proposed similarity can recognize the fact that those sequences are same but the level of similarity is based on the time spent in each state. The progressive increase of similarity level is shown in the result.

tation and the proposed representation) of each student interaction with video and compared the predictive power of each of them in terms of predicting student and video. In others words, can we predict knowing an representation of a sequence of interaction to which student the interaction is and then which video that the sequence of interaction is the interaction with. The experiment is done on nine (9) different videos with 4800 students that interacted with each of the nine (9) videos. Same students and same videos were considered for prediction for the tree different methods of representation. The metrics that we use to compare the three methods of representation of student video interaction in their prediction power on distinguishing each video interaction are:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$BalanceAccuracy = \frac{TPR + TNR}{2} \quad (8)$$

$$F_1 = \frac{2.Precision.Recall}{Precision + Recall} \quad (9)$$

Where TP = True Positive, TN= True Negative, FP=False Positive, FN= False Negative, $TPR = \frac{TP}{TP+TN}$=True Positive Rate, $TNR = \frac{TN}{TN+FP}$ = True Negative Rate, $Recall = \frac{TP}{TP+FP}$ and $Precision = \frac{TP}{TP+FN}$.

6. RESULTS

The results show that the proposed method through the level of similarity can tell how similar two arbitrary sequences are. Through the tests of validation on synthetic data, the proposed method is giving better results than the two existing other methods as one can see through figures 3 and 5. In fact, the expected level of similarity is given by the proposed method. For the same sequence represented as a cyclic sequence of interaction with various way of representation show in figure 2 the expected degree of similarity 100% but only the proposed method give us the closest results to the expected one as show in figure 3. One can see in this figure also that the Markov chain based similarity is the second best estimation of similarity after the proposed method

Predictions:	45 records, 5 students to predict.								
Approach:	SVM			GBM			KNN		
Method:	OM	Mar.	Prop.	OM	Mar.	Prop.	OM	Mar.	Prop.
Accuracy	0.60	0.00	0.80	0.40	0.00	1.00	0.20	0.22	1.00
Balanced Acc.	0.75	0.38	0.88	0.63	0.38	1.00	0.50	0.38	1.00
F_1	0.75	0.00	0.89	0.57	0.00	1.00	0.33	0.36	1.00
Kappa	0.50	-0.25	0.75	0.50	-0.25	1.00	0.00	0.13	1.00
Predictions:	108 records, 12 students to predict.								
Accuracy	0.58	0.18	0.67	0.42	0.36	0.42	0.11	0.00	0.40
Balanced Acc.	0.77	0.50	0.86	0.68	0.60	0.67	0.56	0.45	0.67
F_1	0.73	0.20	0.78	0.59	0.50	0.63	0.20	0.00	0.67
Kappa	0.55	0.10	0.63	0.36	0.30	0.35	0.00	-0.10	0.33

Table 1: Results of Twenty fold cross validation 400 runs of student prediction of 5 and 12 students using three different methods of representation of student interaction with videos showing that the proposed representation technique is performing better than others.

Predictions:	45 records, 9 videos to predict.								
Approach:	SVM			GBM			KNN		
Method:	OM	Mar.	Prop.	OM	Mar.	Prop.	OM	Mar.	Prop.
Accuracy	0.11	0.33	0.56	0.33	0.56	0.56	0.22	0.22	0.33
Balanced Acc.	0.50	0.62	0.75	0.63	0.75	0.56	0.46	0.57	0.63
F_1	0.20	0.50	0.72	0.50	0.36	0.72	0.36	0.36	0.50
Kappa	0.00	0.25	0.50	0.25	0.50	0.50	0.13	0.13	0.25
Predictions:	108 records, 9 videos to predict.								
Accuracy	0.22	0.11	0.56	0.11	0.33	0.56	0.22	0.11	0.22
Balanced Acc.	0.56	0.50	0.75	0.50	0.63	0.75	0.57	0.50	0.57
F_1	0.36	0.20	0.61	0.20	0.50	0.61	0.36	0.20	0.36
Kappa	0.13	0.00	0.50	0.00	0.25	0.50	0.13	0.00	0.13

Table 2: Results of Twenty fold cross validation 400 runs of video prediction of 45 and 108 students interactions records using three different methods of representation of student interaction with videos showing that the proposed representation perform better than others.

based one. When we considered a same sequence of states with different length of time as shown in figure 4, the expected results of similarity is an progressive increase of level of similarity according to the length of the sequence. The classic Markov chain based method could not find that the length of sequences are different when the proposed method is able to find it well as it is in figure 5.

When considering in the real data, the capacity of each method of representation of video interaction to identify uniquely each sequence of interaction the proposed method to its predictability power in terms of student sequence and video shows better results than existing ones as the table 1 show the complete performance parameters on student prediction using SVM, GBM and KNN on predicting five (5) students and twelve (12) students with nine (9) records of each student (where eight (8) records are for training and predicting one record of each student). The results show that the proposed method is performing better than the two others. The full comparison between accuracy of the tree methods is in figure 6 in the second line for the student predictions with various number of students to predict. These results show in each case that the proposed method is more accurate than the others method representations.

In predicting video, the complete results for forty five (45) records from five (5) different students and hundred and eight (108) records from twelve (12) students in predicting the nine (9) videos shown in table 2 demonstrate that the proposed method is better also on recognizing video than the two others methods of presentation of student interaction with video. The complete performance comparison on predicting video is in figure 6 in the first line of the figure. The results show also here that the proposed method in better on distinguishing video from one to another. In both cases, predicting videos or students, the proposed method performed in terms of balanced accuracy better than existing methods. These results show that the proposed method has a better way of representing a student video interaction with videos and so can be used for comparing two different interactions with video.

7. CONCLUSIONS

The proposed methodology aims to fill out a methodological gap on representing and comparing video sequences of interaction methods. This proposed method overcomes the drawbacks of the previous methods based on classic Markov Chain and sequence of interactions known as Optimal Matching distance based (OM). The main contribution of this proposed method is the fact that it takes in account the time

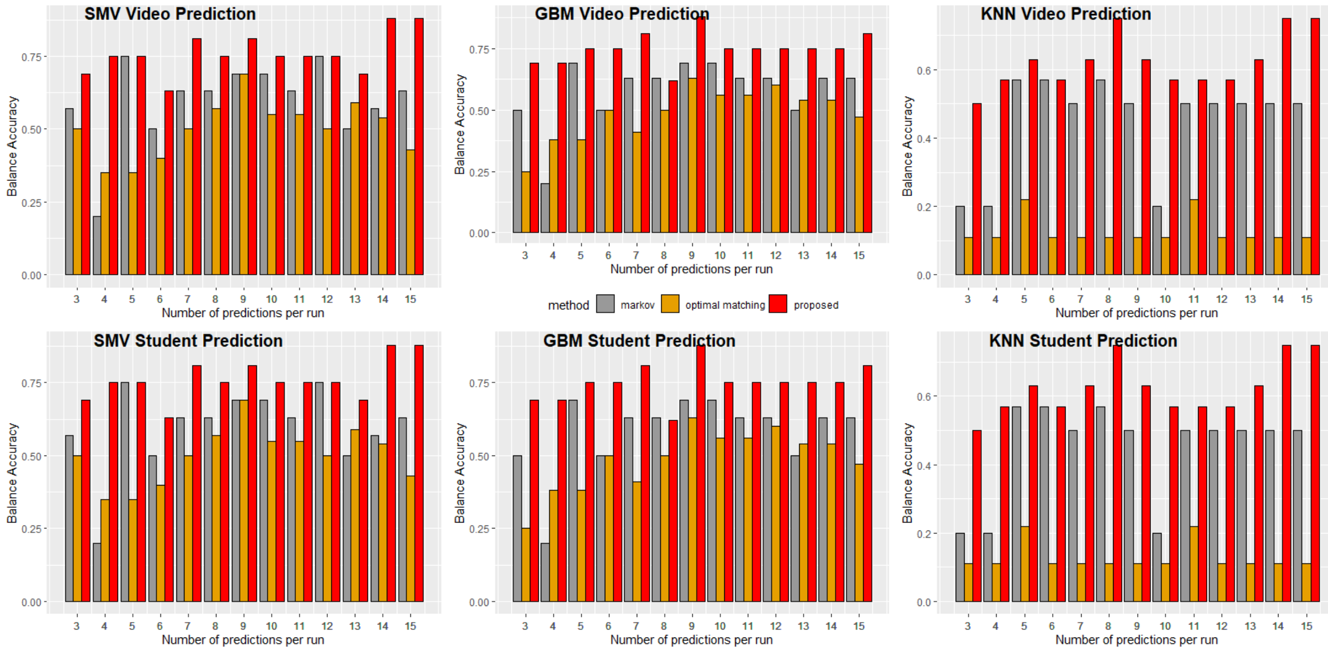


Figure 6: Results of comparison of the proposed representation method versus others method of prediction of video (first line) and student (second line) from students' interactions.

spent in each state and the general style of succession of states. This give a new tool to researchers who want to compared video viewers interaction and find eventually video style of interaction.

This proposed method combine two styles of representation of video sequence of interaction and compute similarity based on advantage of each style of representation. The OM based similarity is generally good on same range length of interactions sequences and the matrix of interaction based representation does better on sequences of different range of length.

The proposed method is compared to the state of art methods such as Markov Chain based similarity and the OM based similarity and the results show it has a better results in computing level of similarity between sequences of interaction taking in account some particularities of sequences of interaction such as cyclic sequences of interaction or same sequence with different length of time spent in each state.

The proposed method by giving the level of similarity in terms of percentage open the possibility to better compare two sequences of interaction with same or different video (normalized representation of the matrix of interaction). This proposed method can better address the similarity between sequences by giving even in the cluster based strategies as previous methods used to do for finding similar sequences of interactions, the level of closeness between sequences in the same cluster the level of similarity among them.

This proposed method is also able to better represent a sequence of interaction when doing classification tasks as the results show. In fact, proposed method has a better performance in predicting student sequence of interaction and prediction video when having a representation of a video sequence of interaction. These results show that the proposed representation of student video interaction is then better than the state of art representation in terms of specifying a video sequence of interaction.

8. FUTURE WORK

The use of this proposed methodology open new avenues for research. Is each video viewer has a specific video interaction style? Using this proposed method could help to find out if there is consistency in video interaction for each viewer by the level of similarity of their sequences. One can also use this method to find out if each video style impose a specific interaction style to the video viewers. An other avenue that this proposed method opens is the link that one can find between video style of interaction and student performance. Could one find the different style of interaction that could lead to failure or success in an online course? Our future investigations will eventually answer to those important issues for the video based online courses.

References

[1] Thushari Atapattu and Katrina E Falkner. Discourse analysis to improve the effective engagement of mooc videos. In *LAK*, pages 580–581, 2017.

[2] Mina Shirvani Boroujeni and Pierre Dillenbourg. Discovery and temporal analysis of latent study patterns in mooc interaction sequences. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, pages 206–215. ACM, 2018.

[3] Lori Breslow, David E Pritchard, Jennifer DeBoer, Glenda S Stump, Andrew D Ho, and Daniel T Seaton.

Studying learning in the worldwide classroom: Research into edx's first mooc. *Research & Practice in Assessment*, 8, 2013.

[4] Christopher G Brinton, Mung Chiang, Shaili Jain, Henry Lam, Zhenming Liu, and Felix Ming Fai Wong. Learning about social learning in moocs: From statistical analysis to generative model. *IEEE transactions on Learning Technologies*, 7(4):346–359, 2014.

[5] Liang Chen, Yipeng Zhou, and Dah Ming Chiu. Video browsing-a study of user behavior in online vod services. In *2013 22nd International Conference on Computer Communication and Networks (ICCCN)*, pages 1–7. IEEE, 2013.

[6] Liang Chen, Yipeng Zhou, and Dah Ming Chiu. A study of user behavior in online vod services. *Computer Communications*, 46:66–75, 2014.

[7] Yishuai Chen, Yong Liu, Baoxian Zhang, and Wei Zhu. On distribution of user movie watching time in a large-scale video streaming system. In *2014 IEEE International Conference on Communications (ICC)*, pages 1825–1830. IEEE, 2014.

[8] Michel Desmarais and François Lemieux. Clustering and visualizing study state sequences. In *Educational Data Mining 2013*, 2013.

[9] Cees H Elzinga. Sequence analysis: Metric representations of categorical time series. *Sociological methods and research*, 2006.

[10] Bruno Emond and Scott Buffett. Analyzing student inquiry data using process discovery and sequence classification. *International Educational Data Mining Society*, 2015.

[11] Philip J Guo, Juho Kim, and Rob 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. ACM, 2014.

[12] Philip J Guo and Katharina Reinecke. Demographic differences in how students navigate through moocs. In *Proceedings of the first ACM conference on Learning@ scale conference*, pages 21–30. ACM, 2014.

[13] Sherif Halawa, Daniel Greene, and John Mitchell. Dropout prediction in moocs using learner activity features. *Experiences and best practices in and around MOOCs*, 7, 2014.

[14] Suhang Jiang, Adrienne Williams, Katerina Schenke, Mark Warschauer, and Diane O'dowd. Predicting mooc performance with week 1 behavior. In *Educational Data Mining 2014*, 2014.

[15] Williams Jiang, Warschauer Schenke, and O'dowd. Predicting mooc performance with week 1 behavior. In *Educational Data Mining*, volume EDM 2014, page 3, 2014.

[16] Junzo Kamahara, Takashi Nagamatsu, Yohei Kaieda, and Yutaka Ishii. Behavioral analysis using cumulative playback time for identifying task hardship of instruction video. In *2010 5th International Conference on Future Information Technology*, pages 1–6. IEEE, 2010.

[17] René F Kizilcec, Chris Piech, and Emily Schneider. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge*, pages 170–179. ACM, 2013.

[18] Severin Klingler, Tanja Käser, Barbara Solenthaler, and Markus H Gross. Temporally coherent clustering of student data. In *EDM*, pages 102–109, 2016.

[19] Marius Kloft, Felix Stiehler, Zhilin Zheng, and Niels Pinkwart. Predicting mooc dropout over weeks using machine learning methods. In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, pages 60–65, 2014.

[20] Nan Li, Lukasz Kidzinski, Patrick Jermann, and Pierre Dillenbourg. How do in-video interactions reflect perceived video difficulty? In *Proceedings of the European MOOCs Stakeholder Summit 2015*, number EPFL-CONF-207968, pages 112–121. PAU Education, 2015.

[21] Nan Li, Łukasz Kidziński, Patrick Jermann, and Pierre Dillenbourg. Mooc video interaction patterns: What do they tell us? In *Design for teaching and learning in a networked world*, pages 197–210. Springer, 2015.

[22] Sylvain Mongy, Fatma Bouali, and Chabane Djeraba. Analyzing user's behavior on a video database. In *Multimedia data mining and knowledge discovery*, pages 458–471. Springer, 2007.

[23] Sylvain Mongy, Chabane Djeraba, and Dan A Simovici. On clustering users' behaviors in video sessions. In *DMIN*, pages 99–103, 2007.

[24] Ozlem Ozan and Yasin Ozarslan. Video lecture watching behaviors of learners in online courses. *Educational Media International*, pages 1–15, 2016.

[25] Nirmal Patel, Collin Sellman, and Derek Lomas. Mining frequent learning pathways from a large educational dataset. *arXiv preprint arXiv:1705.11125*, 2017.

[26] Arti Ramesh, Dan Goldwasser, Bert Huang, Hal Daumé III, and Lise Getoor. Modeling learner engagement in moocs using probabilistic soft logic. In *NIPS Workshop on Data Driven Education*, volume 21, page 62, 2013.

[27] Daniel T Seaton, Yoav Bergner, Isaac Chuang, Piotr Mitros, and David E Pritchard. Who does what in a massive open online course? *Communications of the ACM*, 57(4):58–65, 2014.

[28] Tanmay Sinha, Patrick Jermann, Nan Li, and Pierre Dillenbourg. Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions. *arXiv preprint arXiv:1407.7131*, 2014.

[29] Tanmay Sinha, Nan Li, Patrick Jermann, and Pierre Dillenbourg. Capturing" attrition intensifying" structural traits from didactic interaction sequences of mooc learners. *arXiv preprint arXiv:1409.5887*, 2014.

[30] Jacob Whitehill, Joseph Jay Williams, Glenn Lopez, Cody Austun Coleman, and Justin Reich. Beyond prediction: First steps toward automatic intervention in mooc student stopout. *Available at SSRN 2611750*, 2015.

[31] Diyi Yang, Tanmay Sinha, David Adamson, and Carolyn Penstein Rosé. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-driven education workshop*, volume 11, page 14, 2013.