IJWR

# Knowledge Gap Extraction Based on Learner Interaction with Training Videos

Yosra Bahrani[a], Omid Fatemi*[b]
School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran; Bahrani.yosra@ut.ac.ir*[a], ofatemi@ut.ac.ir[b]

**ABSTRACT**

**In recent years, with the advancement of information technology in education, e-learning quality promotion has received increased attention. Numerous criteria exist for promoting learning quality, such as fitness for purpose, which refers to the extent to which service fits its intended purpose. Multiple purposes are considered in e-learning. One is reducing the knowledge gap between the learner's perception of educational concepts and what should be understood of training concepts. Identifying and calculating the learner's knowledge gap is the first step in reducing the knowledge gap. Consequently, this paper presents a new method for calculating the learner's knowledge gap concerning each concept in the training video content based on the learner's click behavior. The association between the learner's knowledge gap and click behavior was determined by categorizing the learner's click behaviors. Similarly, the Apriori algorithm extracted rules for each behavioral category. The results demonstrated that learning outcome correlated with the learner's click behavior. Therefore, four behavioral rules regarding the compatibility between the knowledge gap and learner's click behavior are presented. Experiments were performed by 52 students enrolled in the micro-processing course at Tehran University's e-Learning Center.**

*Keywords:* ***E-learning quality, Knowledge Gap, Learner's Click Behavior, Apriori Algorithm.***

## 1. Introduction

Due to the advancement of information technology in education, e-learning has received more attention than ever in recent years. Therefore, enhancing the quality of e-learning has become a necessity for all stakeholders involved in this field. Numerous studies have analyzed the quality of e-learning from the perspectives of various stakeholders, including students, teachers, educational designers, and universities, among others. [23, 24, 25]. This research examines the quality of e-learning for the student stakeholders.

To this end, the quality criterion used in this research is based on fitness for purpose, Quality is thus judged in terms of the extent to which the product or service fits its meaning [22]. Several purposes are considered within the e-learning environment. This research investigates one of these, namely the reduction of the knowledge gap among learners. The knowledge gap is "the gap between the learner's perception of educational concepts and what should be understood of training concepts." Based on this, the extent to which the learner's knowledge gap is reduced, the learner's outcome is enhanced, and the quality of e-learning is subsequently improved. Examining and analyzing learner behaviors with the e-learning system, such as participation in the exam and its results, watching educational videos, registering and logging into the system, and learner facial expressions, among others, improve the quality of learning based on educational objectives.

This research investigates the interaction of learner behavior with training videos. The main question of this paper is proposed as "How do we identify and measure the learner's knowledge gap with the educational content based on the learner's behavior during video viewing?"

In training video-based learning, two factors are considered: the content of the training video and the learner's behavior when interacting with the training video. This research analyzes the learner's behavior based on click events.

Click events such as pause, skip back, skip forward, and rate change interactions with a training video are used to extract learner behavior patterns. Then knowledge gap calculated based on each learner behavior pattern .To this end, related work, results, and discussion are presented in sections 2, 3, and 4, respectively, and in the final section, the paper's conclusion is presented.

## 2. Related Work

Extracting knowledge from video content is a complex and important issue. The data mining technique is required to transform raw data of video content into useful information such as the position of concept in the video, duration of concept explanation in the video, etc. A systematic literature review of empirical studies on the application of DM techniques in cardiology is presented in [7]. A system based on different machine learning algorithms to sort unstructured data into a structured format and convert it to a user-friendly format is proposed in [8]. Temporal knowledge propagation for propagating temporal knowledge learned by the video representation network to the image representation network is proposed in [9]. A novel approach to extract the knowledge of a trained deep neural network is introduced in [10]. This shows that automatic knowledge extraction from video content requires a data mining algorithm and is very important for content analysis. In this research, learner's behavior is

doi http://dx.doi.org/10.22133/ijwr.2022.254451.1078
Citation O. Fatemi and Y. Bahrani, "Knowledge gap extraction based on learner interaction with training videos," *International Journal of Web Research*, vol.5,no.1, pp.74-81, 2022.

*Coressponding Author
Article History: Received:3 August 2022; Revised: 10 September 2022 ; Accepted: 20 September 2022

analyzed by assuming that an expert has extracted the video content features.

Learner's behavior is one of the key elements in extracting knowledge gap and is related to it. Lemay and Doleck [3] showed that frequency of video viewing per week was better than watching it once. In addition, if rated videos are regularly evaluated, it will yield a deeper learning than when the evaluation is done after the completion of the video content.

Hu et al. [4] analyzed students' video watching behavior in Mooc based on the spark and found that students' video watching behavior had the highest frequency at the two peaks from 8 to 10 am and the other from 7 to 11 pm. In addition, most of the student events (e.g., pause-seek- and speed change) happen at the first 60 seconds of the video. This pattern is done after a fixed amount of 60-300 seconds and is then reduced to 300 seconds. It shows that the most time to focus on the video for each user and the events performed on the video occurs at the beginning time of watching the video. In [5], the time of watching the video includes the events Replay, Download, and Move back, which are called implicit interest indicators. Goulden et al. [6] analyzed clickstream data and examined students' online learning behavior, mainly aiming at examining students' behavioral patterns and discovering the relationships between clickstream behavior and course performance.

Kuo et al. [16] analyzed the records of students who watched videos in two-month period then extracted three view behavioral patterns: adaptive viewer, self-regulating viewer, and infrequent viewer. They showed that infrequent viewer had lower scores than other viewer, which shows that the continuous behavior of students is related to their performance. Song et al. [17] also analyzed the relationship between prior knowledge and learning performance.

Brinton [1] collected information about learner clicks and answers to questions. The aim of this research is to answer the question of how behavioral data can help improve the performance of the learning process. That proposed three main areas: 1) LDA, in which a new method of extracting behavior from the learner click is to watch video and classify users to four groups based on their click behavior patterns. 2) Social Learning Networks (SLN) is presented as a new method for social learning combined the topical and structural aspects. And 3) Integrated and Individualized Courses (IIC) in the third section to create a model based on the personalized behavior of each learner. In this paper, we used the first part (based on user click groups), which deals with extracting the behavioral pattern of learners from the video

Past studies have shown that learners have different patterns that are related to their behavior and performance. Knowledge gap extraction has a great impact on improving the learning process and increasing learner outcomes.

Many of the studies reviewed above have included click analysis, but this article discusses calculating the knowledge gap based on click behavior, and it aims to present how the final score correlates with the learner's clicking behavior.

## 3. Materials And Methods

This paper calculated the learner's knowledge gap based on the learner's click analysis and the training video content. Hence, the proposed method included two sections: 1- Training video content, 2- learner's click behavior, and determining learner's behavioral category and calculating the knowledge gap associated with every category.

Figure 1 shows the block diagram of these three phases for more clarification. Every phase is explained in the following.

Based on Figure 1, the proposed block diagram consists of two primary components: 1- training video content and 2- learner behavior. Each component is explained as follows.

### 3.1 Training Video Content

The training video contains at least one educational concept. An expert extracted the concept map of the training video content. In addition, an expert determined the difficulty level of every concept (Concept difficulty). The concept difficulty refers to learners' difficulty understanding a concept. The training video content phase includes the following four steps:

a) Selecting a training video content and extracting its concepts
b) Extracting the concept map and the interrelationships of the training video concepts
c) Identifying and extracting the key indexes of the video content
d) Identifying the position and time of concepts in the video

Table 1 presents the extracted indexes in step c.

These indicators are extracted to analyze the learners' behavior as they interact with each video content segment. For instance, the indicator of the start and end time of a concept in the video is used to determine the concept's duration. It is possible to analyze the learner's click events during this time period. Did the learner reach the default duration for this concept? Or did the student pass this concept sooner or later than the default duration? These questions show the utility of using this indicators.

The index of root concept and subconcept identifies whether a concept described in the video is the primary concept or a subset of the primary concept. In this case, if the main concept is not correctly understood, the subset concepts

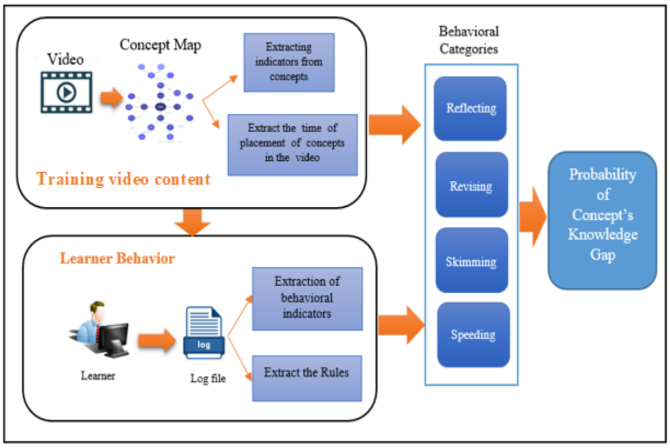



Figure. 1. The proposed block diagram

are also incorrectly understood. A similar interpretation is also possible based on the dependent and independent concepts.

### 3.2 Learner's Behaviour

This research investigates the learner's behavior in interacting with the video. In the learner behavior interaction video, it is possible to observe click behavior and facial behavior. In this research, the learner's click behavior is analyzed, and in a subsequent research where the webcam is active, the learner's facial behavior in interaction with the video will be analyzed. Consequently, this research focuses on the learner's click behavior, its relationship with the learner's content knowledge gap, and the learner's outcome. First, the learner's click behavior while interacting with video content is examined. Then, based on the click behavior, the behavioral categories of the learner are identified, and the learner's knowledge gap regarding each category's behavior is calculated.

The learner's click behavior was extracted from the recorded logs during video watching. Logs contain the user's click information, such as playing, pausing, skipping back, skipping forward, and rate changing.

Before any explanation, the events of a click should be defined precisely, i.e., which behaviors are called pauses or skip backs. Every event of a click is defined in Table 2 based on the explanations and proofs presented in [1].

Each event in [1] has been investigated, and four behavioral categories have been extracted. These four categories, described below according to [1], were employed in this paper.

1. **Reflecting behavior:** It happens when the user pauses to reflect on the video material recurrently. If the reflection time is not so longer than the watching time, the chance of success in the quiz heightens. In contrast, if the pause is extremely short, it denotes a pending confusion.
2. **Revising behavior:** It refers to the recurrent revision of the video content just watched and is associated with a higher chance of success.
3. **Skimming behavior:** It means skipping over the video material rapidly and is connected with a lower chance of success though performed cautiously.
4. **Speeding behavior:** It refers to video watching faster than the default rate and slowdowns occurring at specific times. Several variations are related to diverse effects on the chance of success.

According to these four behavioral categories, the knowledge gap is calculated for every category separately. Before the knowledge gap calculated for every behavioral category, concept difficulty, one of the significant indices attended to in this paper, should be defined.

In this research, Concept difficulty in video content is defined in two ways:

1. The structure of the lesson is subsumed under an expert-decided difficult concept in the concept map.

Table 1. Indicators Extracted from Training Videos

| *Indicator* | *Description* |
|---|---|
| Concept_Start Time | Time to start explaining the concept in the video |
| Concept_End Time | The end time of the concept described in the video |
| Root_Concept | Is there a main concept in the video? (0 and 1) |
| Sub_Concept | Is there a sub-concept in the video? (0 and 1) |
| Concept_Difficulty | The difficulty level of the concept (numerical between zero and one) |
| Concept_Count | The number of concepts in the video content |
| Video_Duration | The video playback time without changing the playback speed |
| Max Time of fast Motion | Maximum video playback time with maximum allowed speed |
| Max Time of Slow Motion | Maximum video playback time with minimum allowed speed |
| Review Concept | Is it a repetitive concept? (1 and 0) |
| Independent Concept | Is the concept independent of the other concepts? (0 and 1) |
| Dependent Concept | Is the concept dependent on the other concepts? (0 and 1) |

Table 2. Click Event Interaction with Video based [1]

| *Event* | *Description* |
|---|---|
| ***Play*** | A play is an event that starts when a click is made for event $E_i$ for which the state $S_i$ is played and continues up to the subsequent click ($E_i$ + 1). Its duration and length equal d = $t_i$ +1 and l = $p_i$+1- $p_i$ , respectively. |
| ***Pause*** | Pause and play events are defined similarly. However, a click in this event makes the state $S_i$ pause, and it is not continued by definition. |
| ***Skip back*** | A skip back (i.e., rewind) is an event that happens when the type $e_i$ = skip and $p'_i > p_i$, where $p'_i$ represents the status of the video player just before the skip. |
| ***Skip forward*** | A skip forward (i.e., fast forward) is an event that is characterized by $S_b$ yet, it catches the case where $p_i > p'_i$ . |
| ***Rate change fast*** | This event takes place when $e_i$= ratechange and the novel rate $r_i$ > 1.0.6. It does not possess any duration or length. |
| ***Rate change fast*** | This event happens when $e_i$ = ratechange and $r_i$ < 1. Similarly, there is no duration or length. |
| ***Rate change default*** | This event arises when $e_i$ = ratechange and $r_i$ = 1. That is to say, the user is turning back to the default rate. |

2. The learner pauses the video to take notes; for example, writing uncertainties about the part and learner cannot establish a regular mental structure of the relationship between this concept and other concepts. Therefore, there is a possibility of difficulty in the concept.

Here, concept difficulty is estimated based on the second definition.

If a false answer is given to the same concept several times in previous quizzes (due to improper design of the concept in the video) or if the video pausing time is between the initiation and termination time of the video concept and 10s longer.(Eq.(1))

$$if\ \left(r_{c_i} = F\right) > 1\ \ or\ \ d_{c_i} \leq d_{pa} \leq d_{c_i} + \alpha\ \ then$$

$$cdf = \frac{\sum_{i=1}^{n} N_{r_{ci=f}}}{N_{q_{ci}}}\ \beta \qquad (1)$$

$$cdf \in \{0,1\}$$

In Eq.(1) :

$r_{c_i}$: Response of concept $c_i$

$d_{c_i}$ : Duration of $c_i$ from start to end times

$d_{pa}$: Duration of pausing time

$\alpha$ : Threshold for a longer period than the start and end of a concept in the video.

*cdf* : Concept difficulty

$N_{r_{ci=f}}$ : Number of false responses

$N_{q_{ci}}$: Number of questions

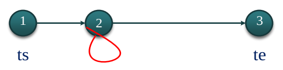
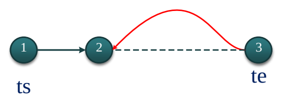
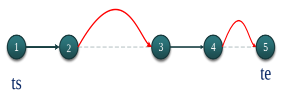
$\beta$ : Weight for difficulty of the concept; this number is the same as the difficulty level between 0 and 1 determined by the expert.

It is worth noting that concept difficulty is estimated according to the first condition, i.e., expert perspective, for a user newly entered into the system. Then, it is calculated based on the second relation after the user enters the system, watches the videos, and answers the questions.

Then, based on the four behavioral categories mentioned in [1], knowledge gap was extracted for each category. Algorithm 1. Describe the process of calculating the knowledge gap of learners while watching videos.

Based on Algorithm 1. , knowledge gap is calculated for each behavior category that shown in Table 3.

Table 3. Knowledge Gap Calculated According to Behavioral Categories of Learners

| | |
|---|---|
| ***Reflective behavioral*** | |
|   | $if\ s_i = paused\,,$<br>$t_{pa} \geq 1,\ d_{pa} < d_{ci}\ then$<br>$KG_{c_i} = cdf(\frac{d_{pa}}{d_{ci}})$ |
| ***Revising Behavior*** | |
|   | $if\ s_i = Sb\,, s_{i-1} = playing\,,$<br>$t_{sc} < t_{sb} < t_{Ec}, t_{sb} > 1\ ,$<br>$if\ s_i = Sb\,, s_{i-1} = paused\,,$<br>$t_{sc} < t_{sb} < t_{Ec}\,, t_{sb} > 1\ then$<br>$KG_{c_i} = cdf \cdot \sqrt{\frac{l}{d_{ci}}} \times \alpha$ |
| ***Skimming Behavior*** | |
|   | $if\ s_i = Sf\,, p_i > p_i'\,,$<br>$t_{sc} < t_{sf} < t_{Ec},$<br>$t_{sf} \geq 1\,,\ t_{ci} = 1\ then$<br>$KG_{c_i} = cdf \cdot\ l(\frac{d_{sf}}{d_{ci}})\alpha \quad , \alpha = 0.01$ |
| ***Speeding Behavior*** | |
|   | $if\ s_i = Rf\ \ , t_{sc} < t_{Rf} < t_{Ec}\,, t_{Rf} \geq 1\,,$<br>$t_{ci} = 1. Rf > avg(Rf)\ then$<br>$KG_{c_i} = cdf \cdot \frac{d'_{ci}}{d_{ci}}$ |

In Table 3, if we consider a general event as a set of the following items: event < $S_i$, $p_i$, $d_i$, $t_i$ > that

$S_i$ : Specifies the status of the video player.

$p_i$ : Specifies the location of the video played.

$d_i$ : Specifies duration elapsed in each event.

$t_i$ : Specifies the time when an event occurred.

Status of the video player includes $S_i$ <Pa, Sb, Sf, Rf> (Pause, Skip back, Skip forward, Rate change forward).

In Table 3, shows four behavioral categories that have been calculated for each category of learner's knowledge gap of the concepts.

***Reflective Behavior***

If the click status within a concept ($c_i$ ) of the video is a pause, the number of pauses between the start time ($t_s$) and the end time ($t_e$ ) of a concept in the video is one or more than one ($t_{pa}>=1$), and the pause duration is less than the duration of explanation of the same concept in the video ($d_{pa} < d_{ci}$), then, as per the above conditions, the knowledge gap for

**Algorithm 1:** calculate knowledge gap of learners while watching video

**Input:** video log file

**Output:** text file

**Parameters:** video log file path, output path

Read count of events like pause, skip back… from video log files.

Calculate concept difficulty for each concept of video.

if count(False response)>1 or pause duration between concept duration and concept duration + α

Then $\quad cdf = \frac{\sum_{i=1}^{n} \text{count of false response}}{\text{count of question of concept}} \cdot \beta$

else

cdf =0

End if

Calculate the knowledge gap for each class of learner behavioral

For ( i=1 to n learners)

if status = paused and time of pause > 1 and pause duration < concept duration

then Reflecting Behavior and $KG_{c_i} = cdf(\frac{\text{pause duration}}{\text{concept duration}})$

else

if status = skip back and time of skip back > 1 and skip back time between concept duration

then Revising Behavior and

$$KG_{c_i} = cdf \cdot \sqrt{\frac{1}{\text{concept duration}}} \times \alpha$$

if status = skip forward and time of sf > 1 and sf duration < concept duration

then Skimming Behavior

and $KG_{c_i} = cdf \cdot 1\,(\frac{\text{skip forward duration}}{\text{concept duration}})\alpha$

else

if status = Rate change forward and time of Rf between concept duration and Rf > avg(Rf)

then Speeding Behavior and $KG_{c_i} = cdf \cdot \frac{d'_{ci}}{d_{ci}}$

end if

end if

end if

end if

end for

Print (learner class, Knowledge gap of learner) this category is calculated as Eq.(2):

$$if\ \ s_i = paused, t_{pa} \geq 1\ , d_{pa} <\ d_{ci}\ \ then$$

$$kG_{ci} = cdf(\frac{d_{pa}}{d_{ci}}) \qquad (2)$$

$$c_i = \{c_1, c_2, \dots, c_n\}$$

In Eq. (2):

$kG_{ci}$ : Knowledge gap of concept

$cdf$ : concept difficulty that calculated in (1)

$d_{pa}$: duration of video pausing

$d_{ci}$: duration of start to end time of explaining a concept in video

*a) Reflective Behavior Analysis*

When a learner pauses between the start and the end of a concept in a video, it indicates that the concept has been unclear or challenging for them and that they are taking notes or concentrating further on learning. This emphasis on this concept will reduce knowledge gaps and ensure that video quiz questions are correctly answered.

***Revising Behavior***

If the click status within a concept ($c_i$) of the video has been skipped back and the previous status ($S_{i-1}$) is currently playing. Additionally, if the learner skips back between the start and end time of the concept in the video ($t_s < t_{sb} < t_e$ ) and the time of skip back is one or more than one ($t_{sb} >=1$ ), or if the learner's click status is paused and then they skip back, the knowledge gap for this category is calculated as Eq.(3):

$$if\ s_i = sb\ , s_{i-1} = palying,\ t_s <\ t_{sb} < t_e, t_{sb} > 1$$

*or*

$$if\ s_i = sb\ , s_{i-1} = paused,\ t_s <\ t_{sb} < t_e, t_{sb} > 1$$

*then*

$$KG_{ci} = cdf\ \sqrt{\frac{l}{d_{ci}}} \times \alpha \qquad (3)$$

$$\alpha = 0.01$$

In (3):

$kG_{ci}$: knowledge gap of concept

$cdf$: concept difficulty that calculated in (1)

$d_{ci}$: duration of start to end time of explaining a concept in video

$l$: The distance between skip back positions ($p_i$) to the current click position

$$l = |p_i - p_i{}'|$$

$$p_i' = p_{i-1}$$

*b) Revising Behavior Analysis*

When a learner skips between the start and the end of an explanation of a concept or a previous concept, they are attempting to find a connection between the current and previous concepts or to understand the new concept better. Due to the similarity between the video concepts and the skip back to a place in mind, it is expected that this concept will have a smaller knowledge gap than the video concept.

***Skimming Behavior***

If the click status within a concept ($c_i$) of the video skips forward and switches to the next position in the video so that the forward position is between the start and end time of a concept in the video ( $t_s < t_{sf} < t_e$) and the time of skip forward is one or more than one ($t_{sf} >=1$), and the concept is first mentioned in the training video, in other words, it is not a duplicate concept ($t_{ci} =1$), then the knowledge gap for this category is calculated as Eq.(4).

$$if\ s_i = sf\ , p_i >\ p_i', p_i' =\ p_{i-1},\ t_s <\ t_{sf} < t_e, t_{sf} \geq 1, t_{ci} = 1$$

*then*

$$KG_{ci} = cdf \times . l(\frac{d_{sf}}{d_{ci}})\ \times \alpha \qquad (4)$$

$$\alpha = 0.01$$

In Eq.(4):

$kG_{ci}$: knowledge gap of concept

$cdf$: concept difficulty that calculated in (1)

$d_{ci}$: duration of start to end time of explaining a concept in video

$d_{sf}$: duration of skip forward the video.

l: The distance between skip forward positions (pi) to the current click position ($p_i{}'$)

$$l = |p_i - p_i{}'|$$

$$p_i' =\ p_{i-1}$$

*a) Skimming Behavior Analysis*

When a learner skips forward between the start and end time of a concept explanation and switches to the next position to reject a concept in a video so that a new concept appears first in the training video, this indicates that the learner lacked the patience to learn the new concept and rejected this concept. It is expected that the knowledge gap will be greater than that concept and that, in subsequent quizzes, participants will not be able to respond to the questions correctly.

***Speeding Behavior***

If the click status within a concept ($c_i$ ) of the video is rate change forward, and the rate change position was between the start and end time of a concept in the video ($t_s < t_{rf} < t_e$ ), and the time of forwarding rate change is one or more than one ($t_{rf} >=1$), and if the concept is first mentioned in the training video, i.e., it is not a duplicate concept ($t_{ci} =1$), then the concept will be included in the rate change forward list. Moreover, if the viewing rate of the concept in the video with the forward rate change operation is faster than the average viewing rate of the learner viewing previous videos, the knowledge gap for this category is calculated as Eq.(5).

$$if\ s_i = RF\ ,\ t_s <\ t_{rf} < t_e,\ t_{rf} \geq 1, t_{ci} = 1, RF > avg(RF)\ then$$

$$KG_{ci} = cdf \times . \frac{d_{ci'}}{d_{ci}} \qquad (5)$$

In Eq.(5):

$kG_{ci}$: knowledge gap of concept

$cdf$: concept difficulty that calculated in (1)

$d_{ci}$: duration of start to end time of explaining a concept in video

$d_{ci}'$: changed duration view concept in video

*a) Speeding Behavior Analysis*

When the learner changes the forward rate between the start and the end of an explanation of a concept, there are two possible interpretations: the learner either has prior knowledge of the concept or lacks the patience to learn the concept. Since the objective here is to identify the knowledge gap, we have stipulated that the concept must be novel. In other words, it has not been covered in previous videos, so it is expected that the learner will have a large knowledge gap regarding this concept and be unable to answer correctly on subsequent concept quizzes.

As described above, four behavior categories of learner analyzed and described. In the next step, to validate the extracted rules, the following association algorithm is used.

## 4. Data Collection and Results

An increasing number of studies analyzed the learner behavior from video interactions [18], [19], [20], [21]. Here, the categories were extracted from learners' behavior based on [1]. Knowledge gap of each category of learners is calculated in the previous section.

First, the training videos were selected from the microprocessor course of Tehran University, key concepts from the video are extracted according to the indicators specified in Table 1 and then evaluated on the students. Data set of this paper was for 52 learners at e-learning center of University of Tehran.

For this section, the extracted rules of the previous section are used for each behavior, and then the rules are obtained using the association rules algorithm, which with the quiz score obtained from the training video concepts shows the same probability of knowledge gap of each category according to four behavioral categories.

### 4.1 Extracted Rules

In this paper, Apriori Association Rule is used to mine frequent pattern of learner behavior interaction with training videos. Since the learner's knowledge gap is calculated based on the two main factors of the learner (click behavior and concept difficulty), its evaluation is analyzed based on the relationship between the final score taken from the concept and the click behavior of the learner. Therefore, three input were imported to this algorithm: 1) learner click event such as pause (pa), skip back (sb), skip forward (sf), and rate change (rc), 2) concept difficulty, and 3) Learner score of training video concept. This determines the behavior of students in interaction with the video, the difficulty level of the observed concept, and scores ultimately received by students. The learner's outcome status is assessed based on the learner's scores, which is labeled by the expert as follows: scores below 40 are considered low, scores between 40 and 70 are middle, and scores above 70 are considered high. The result is found based on the Apriori Association Rule using the data mining tool R (Table IV).

Table 4. Results of Apriori Association Rule Algorithm using the Data Mining Tool R

| ***Best Rule Find*** | ***Support*** | ***Confidence*** | ***Coverage*** | ***Lift*** |
|---|---|---|---|---|
| [1] | 0.23846154 | 0.8732394 | 0.27307692 | 2.162307 |
| [2] | 0.09230769 | 0.8888889 | 0.10384615 | 2.201058 |
| [3] | 0.10769231 | 0.9333333 | 0.11538462 | 1.283951 |
| [4] | 0.08076923 | 0.9130435 | 0.08846154 | 1.256039 |
| [5] | 0.18846154 | 0.9245283 | 0.20384615 | 1.271838 |
| [6] | 0.37307692 | 0.9509804 | 0.39230769 | 1.308227 |

[1]:{event=Pa=No,sb=No,sf=Yes,rc=Yes} => {Score_class=Low}

[2]:{event=Pa=No,sb=No,sf=Yes,rc=Yes,Score_class=High}=>{Cf=Low}

[3]:{Cf=High,Score_class=Middle} =>{event=Pa=Yes,sb=Yes,sf=No,rc=No}

[4]:{Cf=Low,Score_class=Middle} => {event=Pa=Yes,sb=Yes,sf=No,rc=No}

[5]:{Score_class=Middle} => {event=Pa=Yes,sb=Yes,sf=No,rc=No}

[6]:{Score_class=High} => {event=Pa=Yes,sb=Yes,sf=No,rc=No}

Association rule mining (ARM) is one of the most popular tasks of data mining [14]. The most well-known constraint in exploring the rules of dependency are the minimum threshold for support and confidence used to measure the quality of the association rule. Other criteria, such as Lift and Conviction, are also used here.

The support index represents the ratio of the number of solutions with material at point x to the total (XUY) [15], as shown in Eq.(6).

$$Support(X,Y) = \frac{P(X \cup Y)}{n} \quad (6)$$

Confidence is the ratio of the number of transaction XUY to the number of transaction X [15], as shown in Eq.(7).

$$Confidence(X \rightarrow Y) = \frac{P(X \cup Y)}{P(X)} \quad (7)$$

*a) Results of Rules of Table 4:*

**Rule [1]:** Rule 1 determines that a concept's score is low if the learner has skipped forward an event, changed the start and end time duration of a concept in the video, indicating that the learner ignored the concept and that the concept's knowledge gap is high.

**Rule [2]:** Rule 2 determines that a learner has skipped forward an event or a rate change has occurred in the video's start and end time duration, resulting in a high score for that concept. This indicates that the difficulty level of the concept is low or that the learner is aware that it is promptly skipped, and the knowledge gap for this concept is low.

**Rule [3, 4]:** Rules 3 and 4 state that if the learner's score is middle and the concept's difficulty is either high or low (the difficulty level is not relevant), the learner has either paused or skipped back the event. This indicates that when the learner is focused on the concept, they will receive a high score, and the knowledge gap for that concept will be small.

**Rule [5, 6]:** According to Rules 5 and 6, if the learner's score is middle or high, they have either paused or skipped the event. This indicates that the learner paused or skipped back to the concept, in which case they will receive a high score and have a low knowledge gap for that concept.

## 5. Conclusion

This research calculates the knowledge gap of learners based on two factors: 1) training videos and 2) learners' interaction with training video concepts. After analyzing the learning events in interaction with the video, the association rules between the learner's behavior and the concept's score were extracted. It was determined that: 1) if the concept was quickly jumped and ignored by the learner, then the knowledge gap of that concept was high; 2) if the concept difficulty level was low or the learner already knew the concept, it was rapidly skipped, and knowledge gap of that concept was low; 3) if the learner were focused on a particular concept, a good score would be obtained for that concept, and knowledge gap for that concept is low; and 4) if the learner paused or skipped back, a good grade will be obtained for that concept, and the knowledge gap for that concept is low.

According to these rules, the knowledge gap directly results from the learner's interaction with training videos. This proposed method will be expanded to include additional learner activities, thereby enhancing the calculation of the knowledge gap. In the future, we intend to derive additional characteristics from calculating the knowledge gap and improving the computational formula. In addition, suggest resources to reducing the learner's knowledge gap further. Future plans also include incorporating the analysis of the learner's facial behavior into calculating the knowledge gap and improving the calculation of the knowledge gap.

***Funding***

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

***Authors' contributions***

YB: Study design, acquisition of data, interpretation of the results, statistical analysis, drafting the manuscript.
OF: Supervision, discussion of the result, revision of the manuscript.

***Conflict of interest***

The authors declare that there is no conflict of interest.

***Acknowledgements***

This research was conducted at the University of Tehran and the authors of this paper would like to thank the Tehran e-Learning Center for providing their data.

## References

[1] C. G. Brinton, "Technology and Pedagogy: Using Big Data to Enhance Student Learning", ProQuest LLC. 2016.

[2] J. Lin, C. Liu, Y. Li, L. Cui, R. Wang, X. Lu, Y. Zhang, and J. Lian. "Automatic Knowledge Discovery in Lecturing Videos via Deep Representation", *IEEE Access*, vol. 7, pp. 33957-63, 2019.

[3] D. J. Lemay and T. Doleck, "Grade prediction of weekly assignments in MOOCS: mining video-viewing behavior", *Education and Information Technologies*, vol. 25, no. 2, pp. 1333-1342, 2020.

[4] H. Hu, G. Zhang, W. Gao, and M. Wang, "Big data analytics for MOOC video watching behavior based on Spark", *Neural Computing and Applications*, vol. 32, no. 11, pp. 6481-6489, 2020.

[5] N. Belarbi, N. Chafiq, M. Talbi, A. Namir, and E. Benlahmar, "User profiling in a SPOC: A method based on user video clickstream analysis", *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, no. 1, pp. 110-124, 2019.

[6] M. C. Goulden, E. Gronda, Y. Yang, Z. Zhang, J. Tao, C. Wang, X. Duan, G. A. Ambrose, K. Abbott, and P. Miller, "CCVis: Visual analytics of student online learning behaviors using course clickstream data", *IS&T International Symposium on Electronic Imaging Conference*, 2019.

[7] I. Kadi, A. Idri, and J. L. Fernandez-Aleman, "Knowledge discovery in cardiology: A systematic literature review", *International journal of medical informatics*, vol. 97, pp. 12-32, 2017.

[8] H. Patil, V. Gaikwad, D. Pawar, and M. Nikam, "Intelligence Extraction Using Machine Learning Technics", *Intelligence*. Vol. 6, no. 6, pp. 35-37, 2019.

[9] X. Gu, B. Ma, H. Chang, S. Shan, and X. Chen, "Temporal knowledge propagation for image-to-video person re-identification*", In Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 9647-9656.

[10] J. Yoo, M. Cho, T. Kim, and U. Kang, "Knowledge extraction with no observable data", *In Advances in Neural Information Processing Systems*, 2019, pp. 2705-2714.

[11] R. H. Kay, "Exploring the use of video podcasts in education: A comprehensive review of the literature", *Computers in Human Behavior*, vol. 28, no. 3, pp. 820–831, 2012. doi:10.1016/j.chb.2012.01.011,2012.

[12] M. N. Giannakos, L. Jaccheri, and J. Krogstie, "Exploring the relationship between video lecture usage patterns and students' attitudes", *British Journal of Educational Technology*, vol. 47, no. 6, pp. 1259–1275, 2016. doi:10.1111/bjet.12313,2016.

[13] L. Y. Li, "Effect of prior knowledge on attitudes, behavior, and learning performance in video lecture viewing", *International Journal of Human–Computer Interaction*, vol. 35, no. 4-5, pp. 415-26, 2019.

[14] E. Varol Altay and B. Alatas, "Performance analysis of multi-objective artificial intelligence optimization algorithms in numerical association rule mining", *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 8, pp. 3449-3469, 2020.

[15] C. Zhang, and S. Zhang (eds.), *Association rule mining: models and algorithms*, Springer, 2003.

[16] Y. Y. Kuo, J. Luo, and J. Brielmaier, "Investigating students' use of lecture videos in online courses: A case research for understanding learning behaviors via data mining", In F. Li, R. Klamma, M. Laanpere, J. Zhang, B. Manjón, R. Lau (eds), *Advances in web-based learning–ICWL 2015. Lecture Notes in Computer Science (vol. 9412).* Springer, Cham, 2015.

[17] H. S. Song, A. L. Kalet, and J. L. Plass, "Interplay of prior knowledge, self-regulation and motivation in complex multimedia learning environments", *Journal of Computer Assisted Learning*, vol. 32, no. 1, pp. 31–50, 2016. doi:10.1111/jcal.12117

[18] T. Sinha, P. Jermann, N. Li, and P. Dillenbourg, "Your Click Decides your Fate: Inferring Information Processing and Attrition Behavior from MOOC Video Clickstream Interactions", *In Conference on Empirical Methods in Natural Language Processing (EMNLP)*,2014, pp. 3–14.

[19] C. G Brinton, S. Buccapatnam, M. Chiang, and H. V. Poor, "Mining MOOC Clickstreams: Video-Watching Behavior vs In-Video Quiz Performance", *IEEE Transactions on Signal Processing*, vol. 8, no. 1, pp. 136–148, 2016.

[20] C. G. Brinton, and M. Chiang, "MOOC Performance Prediction via Clickstream Data and Social Learning Networks", *IEEE Conference on Computer Communications (INFOCOM)*, 2015, pp. 2299–2307.

[21] D. A. Filvà, M. A. Forment, F. J. García-Peñalvo, D. F. Escudero, and M. J. Casañ, "Clickstream for learning analytics to assess students' behavior with Scratch", *Future Generation Computer Systems*, vol. 93, pp. 673-86, 2019.

[22] L. Harvey and D. Green, "Defining quality", *Assessment & evaluation in higher education*, vol. 18, no. 1, pp. 9-34, 1993.

[23] I. Jung, "The dimensions of e-learning quality: from the learner's perspective", *Educational Technology Research and Development*, vol. 59, no. 4, 445-464, 2011.

[24] C. McNaught, *Quality assurance for online courses: From policy to process to improvement?*, ERIC Clearinghouse, 2001, pp. 435-442.

[25] R. Phipps and J. Merisotis, *Quality on the line: Benchmarks for success in internet-based distance education*, ERIC, 2000.

**Yosra Bahrani** received the M.S. degrees in computer engineering from Amirkabir University of Tehran, Iran, in 2015. She is currently working toward the Ph.D. degree in computer engineering from the Department of Electrical and Computer Engineering, University of Tehran, Iran. Her current research interests include eLearning, Learning analytics, Analysis of learner behavior and big data. She can be reached by sending email to Bahrani.yosra@ut.ac.ir.

**Dr. Omid Fatemi** received his Ph.D. from University of Ottawa in 2000. He subsequently joined University of Tehran where he is currently presenting courses in information technology, advanced computer architecture, visual domain processing, digital VLSI design, microprocessor and embedded systems and e-Learning. He has published several papers in these areas and has supervised more than 80 graduate students. He is the founder of System Level Digital VLSI Design laboratory and also the co-founder of Multimedia Processing Laboratory. He is currently directing eLearning Center and Center of Information Technology in University of Tehran. These centers are in charge of Digital Transformation project in the University. His research interests include system level digital vlsi design, IT governance, eLearning, learning analytics, big data and cloud computing. He can be reached by sending email to omid@fatemi.net.