International Journal on Artificial Intelligence Tools
© World Scientific Publishing Company

IDENTIFYING IMPORTANT SEGMENTS IN VIDEOS: A COLLECTIVE INTELLIGENCE APPROACH

IOANNIS KARYDIS
MARKOS AVLONITIS
KONSTANTINOS CHORIANOPOULOS
SPYROS SIOUTAS
Department of Informatics, Ionian University
49100, Kerkyra, Greece
{*karydis,avlon,choco,sioutas*}*@ionio.gr*

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)

This work studies collective intelligence behavior of Web users that share and watch video content. Accordingly, it is proposed that the aggregated users' video activity exhibits characteristic patterns. Such patterns may be used in order to infer important video scenes leading thus to collective intelligence concerning the video content. To this end, experimentation is based on users' interactions (e.g., pause, seek/scrub) that have been gathered in a controlled user experiment with information-rich videos. Collective information seeking behavior is then modeled by means of the corresponding probability distribution function. Thus, it is argued that the bell-shaped reference patterns are shown to significantly correlate with predefined scenes of interest for each video, as annotated by the users. In this way, the observed collective intelligence may be used to provide a video-segment detection tool that identifies the importance of video scenes. Accordingly, both a stochastic and a pattern matching approach are applied on the users' interactions information. The results received indicate increased accuracy in identifying the areas selected by users as having high importance information. In practice, the proposed techniques might improve both navigation within videos on the web as well as video search results with personalised video thumbnails.

Keywords: Video; important-segment detection; Semantics; Web; User-based; Interaction; User activity; Signal processing.

1. Introduction

The Web has become one of the prominent media for sharing and watching video content [1] and as the volume of available content increases rapidly [2] video retrieval has already become a very important issue [3]. The identification of salient features in the content of a video offers information that will subsequently be used for analysis, indexing and retrieval of videos based on their content. Though, despite providing important information for the purposes of video retrieval, content-based techniques do not take into consideration the video-viewing pattern of the user that also includes valuable contextual/semantic information [4].

The aforementioned domination of the Web as a means for streaming video-watching offers the unique opportunity of monitoring the user's interaction with the video-player and thus inducing new and useful information concerning the viewing-pattern of a user as well as the content of the video. The user–video-player interaction, for example the press of the play, pause or move backwards buttons, provides information on scene viewing which has been shown [5] to relate to the emotive energy of the scene and thus to a wealth of semantic information.

The present study aims in harnessing such video-viewing interactions in order to identify high semantic value video intervals that may subsequently be used in video scene selection for the purposes of representing the video through a thumbnail. To this end, a novel stochastic method is proposed in order to reveal the emerging collective behavior of users watching a specific video by means of the notion of characteristic bell-like patterns emerging in the corresponding users' activity distribution. Accordingly, users' activity distribution (Figure 1) is constructed from the number of the interactions with the corresponding buttons, such as play or pause, of the video-player. Moreover, it is shown that this collective behavior can be employed to infer the most important scenes of a video which can then be used to automatically generate thumbnails, or even implement a summarisation feature, thus leading to collective intelligence [6].

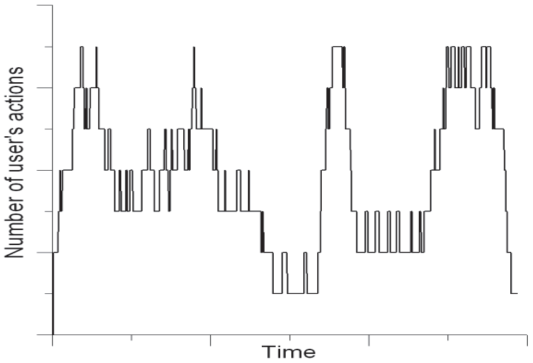


Fig. 1. The users' activity distribution: Video-player button clicks vs. time of click.

The rest of the paper is organised as follows. Section 2 describes related work while Section 3 presents two preliminarily approaches, a stochastic and a pattern matching, that detect patterns of collective behavior and reveal the behavior of the group exhibiting judgment on the importance of video scenes. Next, Section 4 details the setup of the experiment from which data where collected and presents results of the experiments conducted. Finally, the paper is concluded with a discussion in Section 5.

2. Related Work

Research concerning video summarisation and, more generally, important scene selection in videos has mostly been based on content-based methodologies[a]. Nevertheless, as previously mentioned, such content-based methods often fail to capture high-level semantics that adhere to non-specialist users' navigation to videos [4].

In addition to video content, research has also been carried on the users' actions concerning their viewing and searching processes. Yu et al. [7] proposed that users unintentionally show their understanding of the video content through their interaction with the viewing system. Their developed algorithm, *ShotRank*, is computed through a link analysis algorithm that utilises the voting of users on the subjective significance and "interestingness" of each shot. Moreover, in addition to user browsing log mining, *ShotRank* is also taking into consideration low-level content video analysis.

In their work [8], Syeda-Mahmood and Ponceleon, presented *MediaMiner*, a client-server-based media playing and data-mining system aiming at tracking video browsing behavior of users in order to generate fast video previews. In *MediaMiner*, users' interaction with video is recorded at the client side while gathered information is returned to the server for continuous learning and estimation of browsing states. Modeling users' states transition, while browsing through videos, is done with a Hidden Markov Model. *MediaMiner* features common video-browsing interaction buttons (e.g. play and pause) as well as random seek into the video via a slider bar, fast/slow forward and fast/slow backward.

Gkonela and Chorianopoulos [4], presented a user-centric approach, titled *VideoSkip*, wherein by analysis of implicit users' interactions with a web video player (e.g. pause, play, thirty-seconds skip or rewind) semantic information about the events within a video are inferred. Using the simple heuristic concerning the local maxima identification on the accumulated information collected from user-activity, *VideoSkip* has been able to effectively detect the same video-events, as indicated by ground-truth manually annotated by the author of the videos.

This work, in contrast to the hybrid solution proposed to [7], solely relies on user interaction with the player in order to identify high semantic value video intervals. Contrary to the work in [8], the proposed approach utilises a differentiated methodology than a Hidden Markov Model that does not necessarily require the assumption that the state of "interestingness" of a user is a function of the previous state of the user. Moreover, the proposed approach examines the information received from each button of the application separately, offering thus greater flexibility to the event identification, that the approach adopted in [4].

[a] Interested readers can refer to [3] for an extensive survey.

3. User Activity Modeling

The analysis to follow is based on the idea presented at [9]. Indeed, in order to extract pattern characteristics for each button distribution, i.e. scenes in which users exhibit high interaction with the video-player, three distinct stages (as shown in Table 1), are used.

Table 1. Overview of the user activity modeling and analysis.

Stage	User activity signal processing
1	Smoothness procedure
2	Determination of users' activity aggregates
3	Estimation of pattern characteristics

In the first stage, a simple process is used in order to average out user activity noise in the corresponding distribution. In the context of probability theory, noise removal can be treated with the notion of the moving average [10]: from a curve $S^{exp}(t)$ a new smoother curve $S_T^{exp}(t)$ may be obtained as shown in Equation 1,

$$S_T^{exp}(t) = \frac{1}{T}\int_{t-T/2}^{t+T/2} S^{exp}(t')dt' \qquad (1)$$

where T denotes the averaging "window" in time. The larger the averaging window T, the smoother the curve will be. Schematically, the procedure is depicted in Figure 2. The procedure of noise removal of the experimentally recording distribution is of crucial importance for the following reasons: first, in order to reveal patterns of the corresponding signals (regions of high user's activity), and second in order to estimate local maxima of the corresponding patterns. It must be noted that the optimum size of the averaging window T is entirely defined from the variability of the initial signal. Indeed, T should be large enough in order to average out random fluctuations of the users' activities and small enough in order to avoid distortion of the bell-like localised shape of the users' signal which will in turn show the area of high user activity.

In the second stage, aggregates of users' activity are estimated by means of an arbitrary bell-like reference pattern. Accordingly, it is proposed that there is an aggregate of users' actions if within a specific time interval a bell-like shape of the distribution emerges in the sense that there is high probability that user's actions are concentrated at a specific time interval (the center of the bell) while this probability tends to zero quite symmetrically while moving away from this interval (Figure 3). Without loss of generality, the parameters of the width and height of the Gaussian function are set of the order of the averaging window and half of the number users' actions correspondingly.

During the third step, the estimation if the pattern characteristics takes place, i.e. the number of users' aggregates for the specific signal and moreover their exact

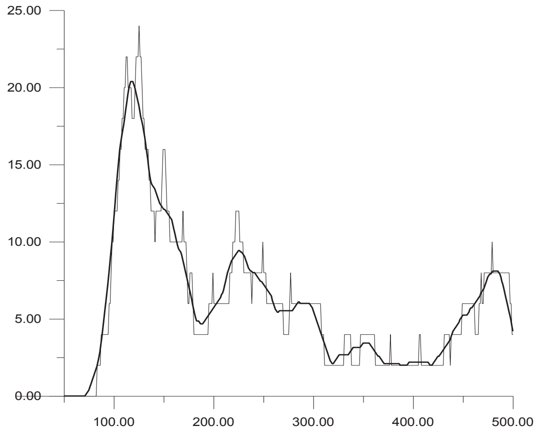


Fig. 2. The user's activity signal is approximated with a smooth signal: The y-axis shows the measured activity of the user while the x-axis shows the time in sec.

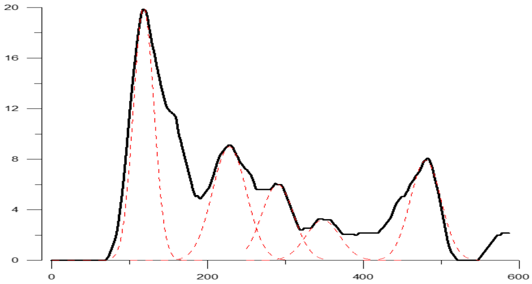


Fig. 3. The users' activity signal is approximated with Gaussian bells in the neighborhood of user activity local maxima: The y-axis shows the measured activity of the user while the x-axis shows the time in sec.

locations in time by employing two different methodologies, a stochastic and a pattern matching:

- In the stochastic approach, the estimation of the exact locations can be done via the estimation of the generalised local maxima. The term generalised local maxima refers to the center of the corresponding bell-like area of the average signal, as the nature of the original signal under examination may cause more than one peaks at the top of the bell due to the micro-fluctuation. It is thus claimed that this is possible by estimating the well known correlation coefficient $r(x, y)$ between the two signals (time series), that is, the average experimental signal and the introduced aforementioned reference bell-like time signal.

  It should be noted that while the height of the reference bell-like pattern does not affect the results of the proposed methodology, the width of the bell D is a parameter that must be treated carefully. In particular, the variability of the average signal determines the order of the width D. Herein, it is proposed that the bell width should be equal to the average half of the widths of the bell-like regions of the signals. This estimation was found optimum in order to avoid overlap between different aggregates.

- In the pattern matching approach, the distance of the reference bell-shaped pattern to the accumulated user interaction signal is measured using 3 different distance measures. Initially, a Scaling and Shifting (translation) invariant Distance (SSD) measure (Equation 2), is adopted from [11]. Accordingly, for two time series x and y, the distance $\hat{d}_{SSD}(x, y)$ between the series is:

$$\hat{d}_{SSD}(x, y) = min_{a,q} \frac{\|x - \alpha y_{(q)}\|}{\|x\|} \tag{2}$$

where $y_{(q)}$ is the result of shifting the signal y by q time units, and $\|\cdot\|$ is the l_2 norm. In this case, and for simplicity, the shifting procedure is done by employing a window the size of which is empirically calculated to minimise the distance, while the scaling coefficient α is adjusted through the maximum signal value in the window context.

The second distance measure used is the Euclidean Distance (ED) measure (Equation 3) that has been shown to be highly effective [12] in many problems, despite its simplicity:

$$d_{ED}(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{3}$$

Finally, the third distance measure utilised is a Complexity-Invariant Distance (CID) measure (Equation 4) for time series as discussed by Batista et al. [13]:

$$d_{CID}(x, y) = ED(x, y) \times CF(x, y) \tag{4}$$

where the two time series x and y are of length n, $ED(x, y)$ is the Euclidean distance (Equation 3), $CF(x, y)$ is the complexity correction factor defined in Equation 5:

$$CF(x, y) = \frac{max(CE(x), CE(y))}{min(CE(x), CE(y))} \tag{5}$$

and $CE(x)$ is a complexity estimate of a time series X, calculated as shown in Equation 6:

$$CE(x) = \sqrt{\sum_{i=1}^{n-1} (x_i - x_{i+1})^2} \tag{6}$$

The aforementioned distance measures produce another time series *dist* that describes the distance of the reference bell-shaped pattern to the accumulated users' interaction signal and thus requires the identification the locations of *dist* where its value is minimal, indicating a close match of the the reference bell-shaped pattern to the accumulated signal. To avoid using a simplistic global cut-off threshold the proposed approach incorporates a local minima peak detection methodology, where a point in *dist* is considered a minimum peak if it has the minimal value, and was exceeded, to the left of the signal, by a value greater by $DELTA$, the peak detection sensitivity value.

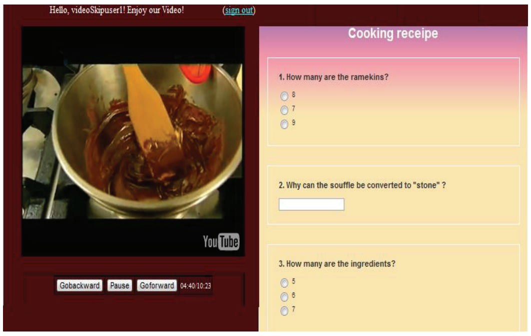
4. Experiments and results

To explore the usefulness of the methodologies presented herein, the dataset collected in [4] is utilised. The goal of the user experiment was to collect activity data from the users but instead of mining real usage data, a controlled experiment was conducted as it provided a clean set of data that was easier to analyse.

The *VideoSkip* platform [14] presentend therein (Figure 4), employs few buttons, in order to be simple in the association of a user's actions with video semantics with common forward and backward buttons modified into *GoBackwards* and *GoForwards.* According to Gkonela and Chorianopoulos [4] the experimental player buttons provide a good trade-off between external validity and collecting a significant data-set of video interactions. *GoBackwards* jumps backwards 30 seconds and its main purpose is to replay the last 30 seconds of the video, while the *GoForward* button jumps forward 30 seconds and its main purpose is to skip insignificant video segments. Therefore, the player provides a subset of the main functionality of a typical VCR device [15]. The selection of videos was based on their degree of visual structure, aiming at videos as much visually unstructured as possible (e.g., lecture, documentary), since content-based algorithms have already been successful with videos that have visually structured scene changes. In particular, *Video A* [16], is a lecture video including typical camera pans and zooms from speaker to projected slides, *Video B* [17], is a documentary including a basic narrative and quick scene changes, *Video C* [18], a lecture video, includes a paper presentation from a local workshop with topic "The acceptance of free laptops, that have been given to secondary education students", while *Video D* [19] is a how-to video including a segment of a cooking TV show for a chocolate soufflé cake.

In order to experimentally replicate users' activity, the experiment designers developed a questionnaire that draws questions from several segments of each video. According to Yu et al. [7] there are segments of a video clip that are commonly interesting to most users, and users might browse the respective parts of the video clip in searching for answers to some interesting questions. Thus, the intuitive assumption of using of these videos in the field (e.g., YouTube) is that when enough user data are available, users' behavior will exhibit similar patterns even if they are not explicitly asked to answer questions.

The experiment took place in a lab with Internet connection, general-purpose computers and headphones. Twenty-three university students (18-35 years old, 13



Hello, videoSkipuser1! Enjoy our Video!

Cooking receipe

1. How many are the ramekins?

○ 8
○ 7
○ 9

2. Why can the souffle be converted to "stone" ?

3. How many are the ingredients?

○ 5
○ 6
○ 7

Fig. 4. The VideoSkip Player has few buttons and questionnaire functionality.

women and 10 men) spent approximately ten minutes to watch each video (buttons were disabled). All students had been attending the Human-Computer Interaction courses at the Department of Informatics at a post- or under-graduate level and received course credit in the respective courses. In order to motivate users to actively browse through the video and answer the respective questions, a time restriction of five minutes was in effect during the experiment. Users were informed that the purpose of the study was to measure their performance in finding the answers to the questions within time constraints.

In the initialisation phase, every video was considered to be associated with four distinct distributions in the time interval of length k, where k is the number of the duration of the video in seconds. Each resulting series corresponds to the frequencies with which the four distinct buttons of Play/Pause, *GoForward* and *GoBackward* were used by the users at specific times. The users' activity distribution was created as follows: each time a user pressed the *GoBackward*/*GoForward* button, the interval matching the last or next, respectively, 30 seconds of the video, were incremented by a unit, meaning that during all these 30 seconds the corresponding button was assumed pushed. The underlying assumption in this case is that the user rewinds a video either because there is something interesting, or because there is something difficult to understand, while the user forwards a video because there is nothing of interest. In this way, a distribution was constructed for each button and for each video, a depiction of users' activity patterns over time.

In the experimentation presented herein, focus has been put on the analysis of the video seeking user behavior, such as *GoBackward* and *GoForward* after the previously described smoothing procedure. An exploratory analysis with time series probabilistic tools, such as variance and noise amplitude, verified what is visually depicted in Figure 5 concerning *Video A*, the lecture video. While the *GoBackward* button signal has a quite regular pattern with a small number of regions with high users' activity, the *GoForward* button signal is characterised by a large number of seemingly random and abnormal local maxima of users' activity. This is due to the experiment design, where there was limited time for information gathering from the respective video and thus, usage of the *GoForward* shows users' tendency to rush through the video in order to remain within the time limit. The use of the Play/Pause buttons has also been considered, but for the current dataset, there were too few interactions. In the following, the preliminary results presented demonstrate the proposed methodologies for detecting patterns of users' activity.

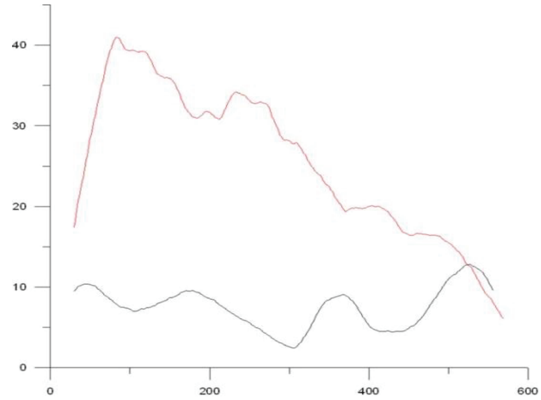


Fig. 5. The GoBackward signal (blue, at the bottom) was compared to the *GoForward* signal (red, at the top), in order to understand which one is closer to the semantics of the video: The y-axis shows the measured activity of the user while the x-axis shows the time in sec.

As far as the stochastic approach is concerned, the analysis of the users' activity distributions was based on an exploration of several alternative averaging window sizes. Results of the proposed modeling methodology for *Video A* are shown in Figure 6, and, in this case, the pulse width D is 60 seconds and the smoothing window T is 60 seconds. The results are depicted by means of pulses instead of the bell shapes in order to compare with the corresponding pulses of the ground-truth designated by the videos' authors. The mapping of between pulses and bells are based on the rule that the pulse width is equal to the width between the two points of the bell where the second derivative changes sign. Similarly, results of the proposed modeling methodology for *Video B* are shown in Figure 7, while in this case, the pulse width D is 50 seconds and the smoothing window T is 40 seconds. The smoothed signals are plotted with the solid black curve. Figure 8 shows results for *Video C* with the pulse width D being 20 seconds and the smoothing window T 30 seconds. Finally, results for *Video D* are shown in Figure 9, in which this case, the pulse width D was 15 seconds and the smoothing window T 20 seconds. Moreover, pulse signals were extracted from the corresponding local maxima indicating time intervals where aggregates were detected according to the definition given in Section 3. These pulses are depicted with the red line. Within the same figures, time intervals that were annotated as ground-truth by the author of the video to contain high semantic value information are also depicted with the blue line.

For the stochastic approach, the correlation of the estimated high-interest intervals and the ground-truth annotated by the author of the video, is visually evident. Cross correlation, between the two intervals, was calculated at 0.673, 0.612, 0.71 and 0.61 correspondingly for videos A, B, C and D, indicating strong correlation between the two pulses.

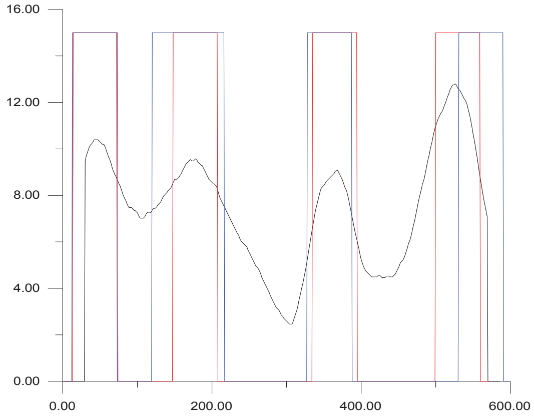


Fig. 6. *Video A*: Cumulative users' interaction vs. time including results from stochastic approach: The y-axis shows the measured activity of the user while the x-axis shows the time in sec.

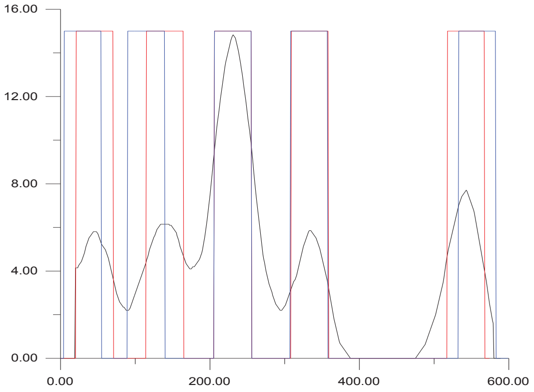


Fig. 7. *Video B*: Cumulative users' interaction vs. time including results from stochastic approach: The y-axis shows the measured activity of the user while the x-axis shows the time in sec.

For the same four videos, the application of the pattern matching approach is examined for each distance measure. The time intervals of each video, where the answers to the questions posed during the experimentation are given, constitute the ground-truth, based on which the classifier's prediction is evaluated.

Table 2 lists the achieved F1 score while Figures 10, 11 and 12 show the achieved precision, recall, specificity percentage & Matthews Correlation Coefficient (MCC)

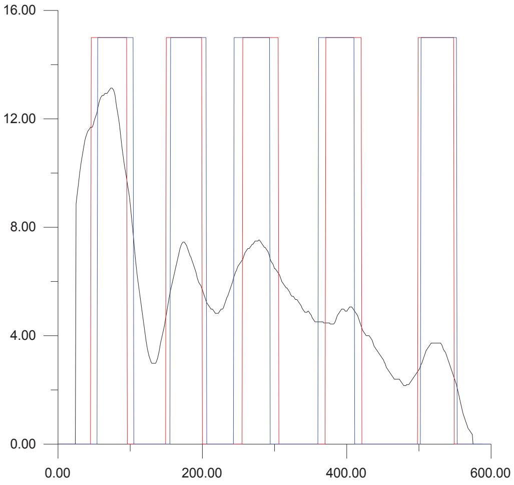


Fig. 8. *Video C*: Cumulative users' interaction vs. time including results from stochastic approach: The y-axis shows the measured activity of the user while the x-axis shows the time in sec.

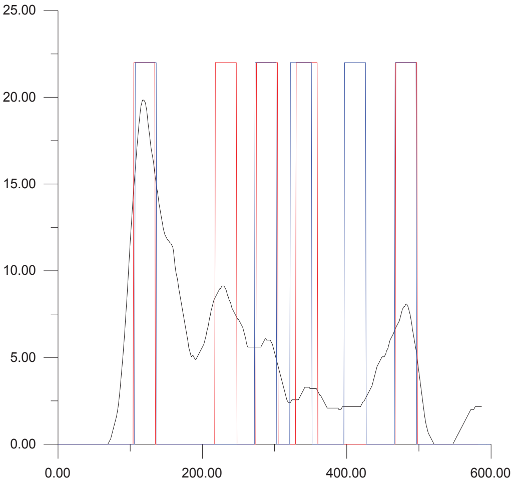


Fig. 9. *Video D*: Cumulative users' interaction vs. time including results from stochastic approach: The y-axis shows the measured activity of the user while the x-axis shows the time in sec.

value for *Video A* for varying peak detection sensitivity values for each of the three distance measures, SSD, ED and CID respectively. It should be noted that the left y-axis of the Figures depict recision, recall and specificity while the right y-axis depicts only the MCC value. As it can be seen in Table 2 and Figure 10, the SSD metric achieves an F1 score of 0.79 in a scale of $[0, 1]$, with 1 being the best value.

Still as the F1 score does not take the true negative rate into account the MCC value has been computed leading to a 0.6 value on a scale of $[-1, 1]$, with 1 implying a perfect prediction. The claim of the ability of Euclidean Distance to be performing relatively high, despite its simplicity, is once again shown in Table 2 and Figure 11 where ED scored an F1 score of 0.72 and an MCC value of 0.42. Finally, the CID measure, shown in Table 2 and Figure 12 was outperformed by the other two measures having scored an F1 score of 0.70 and an MCC value of 0.39.

Table 2. Achieved F1 score for *Video A*

	SSD	ED	CID
F1 score	0.79	0.72	0.70

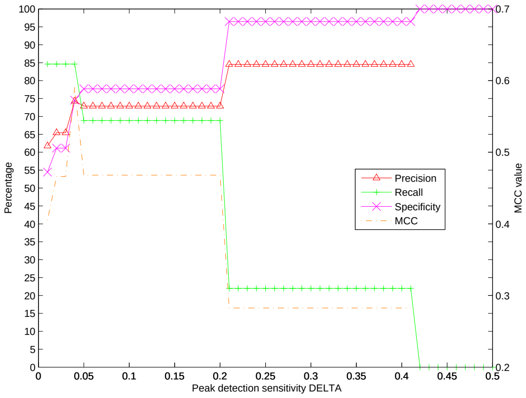


Fig. 10. *Video A*, pattern matching approach, SSD measure: Precision, recall, F1 score, accuracy, specificity percentage & MCC value.

Table 3 lists the achieved F1 score while Figures 13, 14 and 15 show the achieved precision, recall, specificity percentage & MCC value for *Video B* for varying peak detection sensitivity values for each of the three distance measures, SSD, ED and CID respectively. Table 3 and Figure 13 show the SSD metric achieving an F1 score of 0.75 and an MCC value of 0.56. The Euclidean Distance, Table 3 and Figure 14, scored an F1 score of 0.66 and an MCC value of 0.34. Finally, the CID measure for *Video B*, shown in Table 3 and Figure 15, outperformed the ED measure having scored an F1 score of 0.71 and an MCC value of 0.45.

As far as *Video C* is concerned, Table 4 lists the achieved F1 score while Figures 16, 17 and 18 show the achieved precision, recall, specificity percentage & MCC value for varying peak detection sensitivity values for each of the three distance measures, SSD, ED and CID respectively. Table 4 and Figure 16 show the SSD

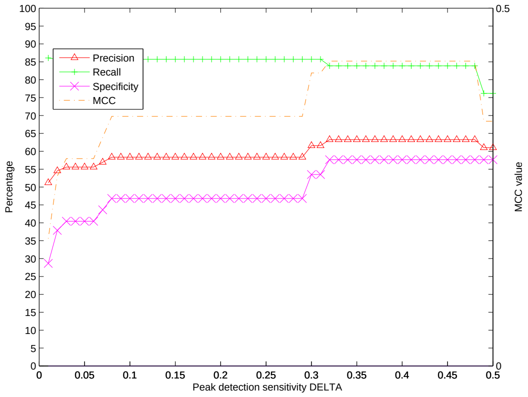


Fig. 11. *Video A*, pattern matching approach, ED measure: Precision, recall, F1 score, accuracy, specificity percentage & MCC value.

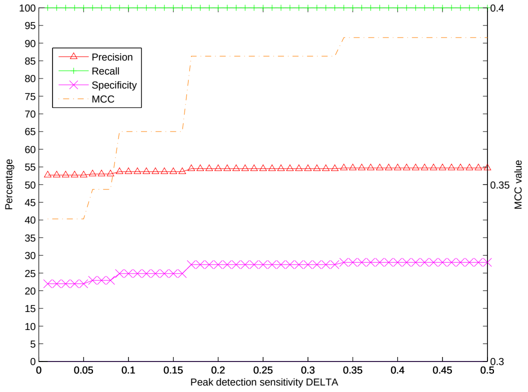


Fig. 12. *Video A*, pattern matching approach, CID measure: Precision, recall, F1 score, accuracy, specificity percentage & MCC value.

Table 3. Achieved F1 score for *Video B*

	SSD	ED	CID
F1 score	0.75	0.66	0.71

metric achieving an F1 score of 0.78 and an MCC value of 0.82. The Euclidean Distance, Table 4 and Figure 17, scored an F1 score of 0.71 and an MCC value of 0.48. Finally, the CID measure for *Video C*, shown in Table 4 and Figure 18, performed similarly to the ED measure having scored an F1 score of 0.70 and an MCC value of 0.47.

Table 5 lists the achieved F1 score while Figures 19, 20 and 21 show the achieved precision, recall, specificity percentage & MCC value for *Video D* for varying peak

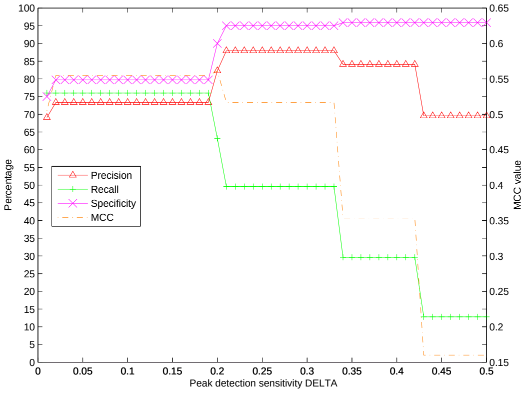


Fig. 13. *Video B*, pattern matching approach, SSD measure: Precision, recall, F1 score, accuracy, specificity percentage & MCC value.

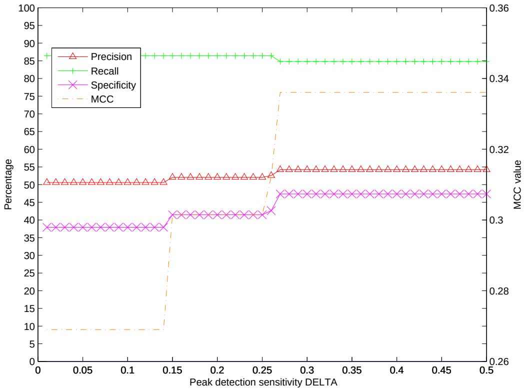


Fig. 14. *Video B*, pattern matching approach, ED measure: Precision, recall, F1 score, accuracy, specificity percentage & MCC value.

Table 4. Achieved F1 score for *Video C*

	SSD	ED	CID
F1 score	0.78	0.71	0.70

detection sensitivity values for each of the three distance measures, SSD, ED and CID respectively. Table 5 and Figure 19 show the SSD metric achieving an F1 score of 0.66 and an MCC value of 0.56. The Euclidean Distance, Table 5 and Figure 20, scored an F1 score of 0.53 and an MCC value of 0.34. Finally, the CID measure for *Video D*, shown in Table 5 and Figure 21, performed similarly to the ED measure having scored an F1 score of 0.54 and an MCC value of 0.40.

It is obvious in almost all of the previously shown results that the achieved recall

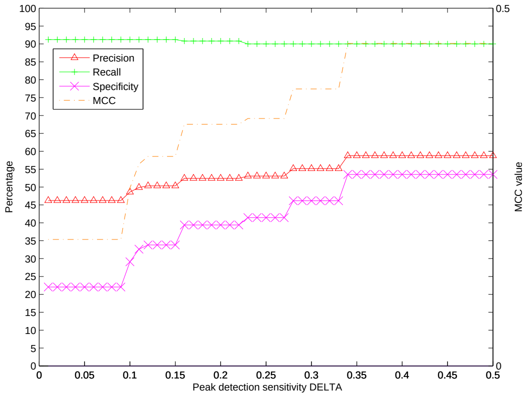


Fig. 15. *Video B*, pattern matching approach, CID measure: Precision, recall, F1 score, accuracy, specificity percentage & MCC value.

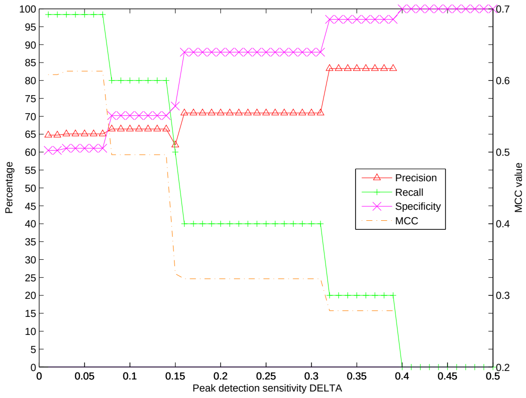


Fig. 16. *Video C*, pattern matching approach, SSD measure: Precision, recall, F1 score, accuracy, specificity percentage & MCC value.

Table 5. Achieved F1 score for *Video D*

	SSD	ED	CID
F1 score	0.66	0.53	0.54

is much higher that the precision. This fact can be attributed to the capability of the proposed methodology in successfully retrieving most of the time instances that were designated by users as points of interest. On the other hand, the precision is lower since the retrieved time intervals indicated more time instances as interesting in contrast to the time instances that were designated by users as points of interest. Accordingly, the F1 score measurement acts as a weighted average of precision and recall in order to compensate for shortcomings of both measurements.

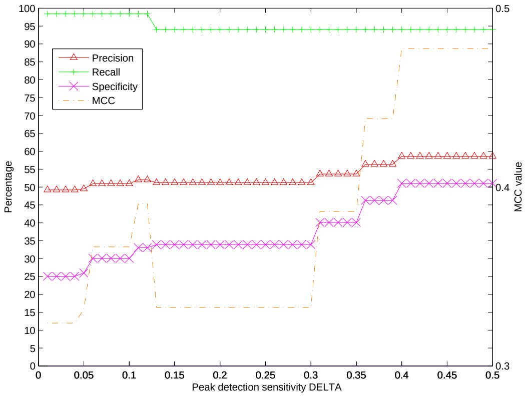


Fig. 17. *Video C*, pattern matching approach, ED measure: Precision, recall, F1 score, accuracy, specificity percentage & MCC value.

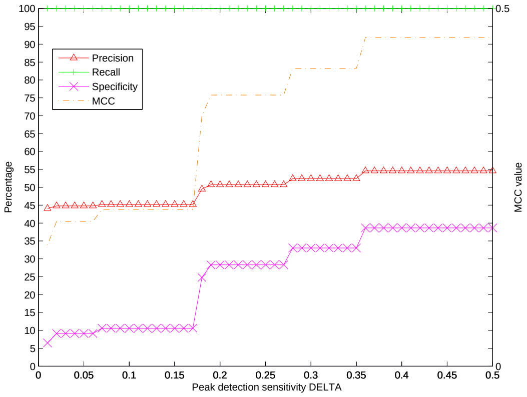


Fig. 18. *Video C*, pattern matching approach, CID measure: Precision, recall, F1 score, accuracy, specificity percentage & MCC value.

As it can be inferred by the results, the Scaling and Shifting (translation) invariant Distance (SSD) measure has been shown to outperform the rest of the distance measures tested in the pattern matching approach. This should not come as surprise as in the matching process of the reference bell-shaped pattern and the accumulated user interaction signal both scaling and shifting of the time-series is required. Nevertheless, the results obtained for the SSD metric show that there is room for further amelioration on the process of similarity distance measurement.

5. Discussion

This research, has proposed two methods that detect collective behavior of users via the detection of aggregates within the corresponding distribution of users' activity. The methodologies proposed are tested on web videos under a controlled experi-

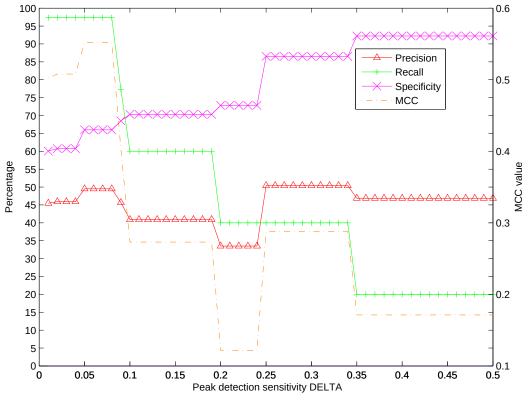


Fig. 19. *Video D*, pattern matching approach, SSD measure: Precision, recall, F1 score, accuracy, specificity percentage & MCC value.

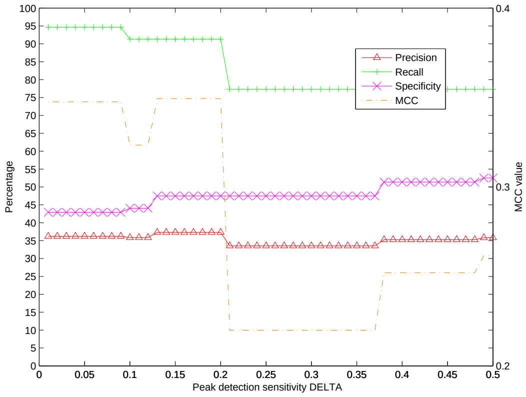


Fig. 20. *Video D*, pattern matching approach, ED measure: Precision, recall, F1 score, accuracy, specificity percentage & MCC value.

ment. Collective intelligence is attributing to the claim of being able to understand the importance of video content from users' interactions with the player. The results of this study can be used to understand and explore collective intelligence in general i.e., how to detect users' collective behavior as well as how the detected collective behavior leads to judgment about the content from which users' activity was gathered. Moreover, collective intelligence may be used as a tool of user-based content analysis having the benefits of continuously adapting to evolving users' preferences, as well as providing additional opportunities for the personalisation of content. For example, users might be able to apply other personalisation techniques, e.g. collaborative filtering, to the user activity data.

In addition, two approaches for aggregates of users' activity estimation have been shown by means of an arbitrary bell-like reference pattern. According to the definition provided, it has been argued that the aggregate of users' actions locally

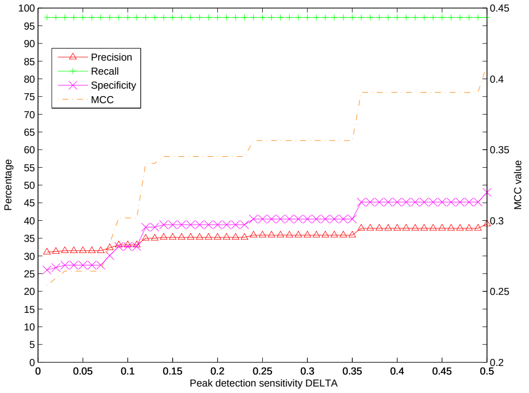


Fig. 21. *Video D*, pattern matching approach, CID measure: Precision, recall, F1 score, accuracy, specificity percentage & MCC value.

coincides, to a large degree, with a bell-like shape of the corresponding distribution. The complete pattern of users' interactions is defined by the exact location of the center of bells of the total number of the bell-like patterns detected. In this way different users' behavior can be mapped to different patterns observed, while these patterns of users' actions can reveal specific judgment about the content for which actions were collected, leading thus to collective intelligence. Indeed, for the case study presented herein, the exact locations of the bell-like patterns detected can be mapped to the most important parts as was shown by experimentation. On the other hand, collective intelligence could reveal new unexpected results, i.e. important intervals of users' behavior that were unexpected.

In a more general fashion, the proposed methodology may treat general users' interactions for a specific (on line) content, by interpreting these interactions as an explicit time series. This could be a time series of clicks or plays of a video on YouTube, the number of times an article on a newspaper website was read, or even the number of times that a hash tag in Twitter was used.

Thus, the proposed methodology can be applied for the detection of patterns emerging in the temporal variation of the corresponding time series indicating the importance of a segment of content at a specific time interval of its duration.

One may formally define this either as a problem of time series correlation based on the correlation between the shape of the (experimentally collected) time series with the shape of a reference time series indicating local maximization of users' activity or as pattern matching of time-series wherein different similarity measures can be utilised for the detection of local minima in the distance. In both cases a Gaussian function can be chosen as the optimum function for the reference time series.

Given that online content has large variation during its duration, i.e. users' actions occur at arbitrary times and with very different time intervals, a further extension of the proposed methodology is needed in order to adapt a time series metric that is invariant to scaling and shifting, i.e. to be able not only to detect the exact location of the local maxima of user's popularity but also to estimate the corresponding absolute importance as well as the corresponding time interval over which the specific piece of content was important enough.

To this end, it is possible, based to the proposed approach, to build a scale free similarity metric introducing the notion of the aforementioned reference bell-like time signal. Indeed, the final result of this extended algorithm would be the estimation of the maximum correlation coefficient in terms of the optimum time moment and optimum bell width.

In any case, the scope of this work is to report the large area in which collective intelligent can be applicable, to provide some initial results as to how one can treat these phenomena as well as how to detect and define patterns which interpret collective intelligence. It is thus the aim of the authors to evolve the methodology presented herein and explore its applicability to more complex cases where interactions between users as well interactions of users with their environment come into play.

References

1. Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proc. ACM SIGCOMM Conference on Internet Measurement*, pages 1–14, 2007.
2. YouTube. Statistics, 2012. http://www.youtube.com/t/press_statistics.
3. P. Geetha and Vasumathi Narayanan. A survey of content-based video retrieval. *Journal of Computer Science*, 4(6):474–486, 2008.
4. Chrysoula Gkonela and Konstantinos Chorianopoulos. Videoskip: event detection in social web videos with an implicit user heuristic. *Multimedia Tools and Applications*, pages 1–14, 2012.
5. David A. Shamma, Ryan Shaw, Peter L. Shafton, and Yiming Liu. Watch what i watch: using community activity to understand content. In *Proc. of International Workshop on Multimedia Information Retrieval*, pages 275–284, 2007.
6. S. Diplaris, A. Sonnenbichler, T. Kaczanowski, Ph. Mylonas, A. Scherp, M. Janik, S. Papadopoulos, M. Ovelgonne, and Y. Kompatsiaris. *Emerging Collective Intelligence for personal, organisational and social use*, pages 527–573. Springer, 2011.
7. Bin Yu, Wei-Ying Ma, Klara Nahrstedt, and Hong-Jiang Zhang. Video summarization based on user log enhanced link analysis. In *Proc. of ACM International Conference on Multimedia*, pages 382–391, 2003.
8. Tanveer Syeda-Mahmood and Dulce Ponceleon. Learning video browsing behavior and its application in the generation of video previews. In *Proc. of ACM International Conference on Multimedia*, pages 119–128, 2001.
9. Markos Avlonitis, Konstantinos Chorianopoulos, and David Ayman Shamma. Crowdsourcing user interactions within web video through pulse modeling. In *Proceedings of the ACM multimedia 2012 workshop on Crowdsourcing for multimedia*, Proc. ACM Multimedia Workshop on Crowdsourcing for multimedia, pages 19–20, 2012.
10. Erik Vanmarcke. *Random fields, analysis and synthesis.* MIT Press, 1983.
11. K. K. W. Chu and M. H. Wong. Fast time-series searching with scaling and shifting. In *PODS*, pages 237–248, 1999.

12. Hui Ding, Goce Trajcevski, Peter Scheuermann, Xiaoyue Wang, and Eamonn Keogh. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proc. VLDB Endowment*, 1(2):1542–1552, 2008.
13. Gustavo E. A. P. A. Batista, Xiaoyue Wang, and Eamonn J. Keogh. A complexity-invariant distance measure for time series. In *Proc. SIAM Conference on Data Mining*, pages 699–710, 2011.
14. Konstantinos Pardalis and Konstantinos Chorianopoulos. Socialskip: User-based video analytics, 2013. https://code.google.com/p/socialskip/.
15. Chris Crockford and Harry Agius. An empirical investigation into user navigation of digital video using the vcr-like control set. *Int. J. Hum.-Comput. Stud.*, 64(4):340–355, 2006.
16. xrgk. Multiple input devices, 2010. http://www.youtube.com/watch?v=8LebAtvulIY.
17. Mega tv. Protagonists tv series - use of internet by young people, 2010. http://goo.gl/98Nr0.
18. xrgk. Acceptance of portable computers in students and teachers of the 1st grade of junior high school, 2007. http://www.youtube.com/watch?v=Z09ythJT9Wk.
19. ERT. Chocolate soufflé cake, 2010. http://www.youtube.com/watch?v=LzkYvtqlT5I.