Crowdsourcing User Interactions within Web Video through Pulse Modeling

Markos Avlonitis
Ionian University
Corfu, Greece
avlon@ionio.gr

Konstantinos Chorianopoulos
Ionian University
Corfu, Greece
choko@acm.org

David A. Shamma
Yahoo! Research
Santa Clara, CA USA
aymans@acm.org

ABSTRACT

Semantic video research has employed crowdsourcing techniques on social web video data sets such as comments, tags, and annotations, but these data sets require an extra effort on behalf of the user. We propose a pulse modeling method, which analyzes implicit user interactions within web video, such as rewind. In particular, we have modeled the user information seeking behavior as a time series and the semantic regions as a discrete pulse of fixed width. We constructed these pulses from user interactions with a documentary video that has a very rich visual style with too many cuts and camera angles/frames for the same scene. Next, we calculated the correlation coefficient between dynamically detected user pulses at the local maximums and the reference pulse. We have found when people are actively seeking for information in a video, their activity (these pulses) significantly matches the semantics of the video. This proposed pulse analysis method complements previous work in content-based information retrieval and provides an additional user-based dimension for modeling the semantics of a web video.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

General Terms

Algorithms, Measurement, Human Factors.

Keywords

video, interaction, pragmatics, user activity, implicit.

1. PULSE MODELING METHODOLOGY

We approach social video consumption activity as a user activity signal in the temporally linear video playback. This relies on the capture and analyzation of more so implicit user interactions for extracting useful information about a video. Previous research [2] has suggested that implicit interactions between the people and the video-player benefit video summarization. To explore this, we analyzed aggregated user interaction with the video using a stochastic pulse modeling process.

We employed an open data-set [1], which has been created in the context of a controlled user experiment (23 users, approximately 400 user interactions within video), in order to ensure well-defined user-based semantics and noise-free user activity data. In the initialization phase, we consider that every video is associated with four distinct time series of length *k*, where *k* is the number of the duration of the video in seconds. Each series corresponds to the buttons of Play/Pause, GoForward, and GoBackward.

It is our aim to construct a general formalism to treat the statistical properties of the aforementioned discrete signals as well as correlation properties between them. Let us consider *N* user interactions and denote with r the position vectors of those actions in the time domain. The type of the button pushed is labeled by *m*. The discrete system of user's actions can be formally characterized by discrete densities as follows,

$$\rho^m(r)=\sum_j^N \delta^m(r-r_j)$$

which is actually a series of pulses of definite width the centers of which are determined by the vectors *r*.

In the rest of the paper we assume the simplest case of un-correlated button actions while the complete study is postponed for a future paper. Pair correlation functions between pulse signals may be treated as usual with the well-known Pearson correlation coefficient.

Initially, the user activity signal is created as follows: each time the user presses the GoBackward (or the GoForward button), the corresponding moments of the video are incremented by one. In this way, an experimental time series is constructed for each button —a depiction of users' activity patterns over time. In order to extract pattern characteristics for each button array, i.e., video segments with high user activity, the following methodology, consistent of four distinct stages, was used.

In the first stage, we use simple procedure in order to average out user activity noise. In the context of probability theory the noise removal can be treated with the notion of the moving average [3]: from a time series $s^{\exp}$ a new smoother time series $s_T^{\exp}$ may be obtained as:

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

where *T* denotes the averaging window in time. The larger the averaging window *T*, the smoother the signal will be. T should be large enough in order to average out random fluctuations of the user's activities and small enough in order to reveal, and not disturb, the bell-like localized shape of the user's signal which in turn will demonstrate the area of high user activity.

In the second stage, we construct a pulse series from the above constructed user activity smooth signal. The pulse signal is to be compared with the corresponding pulse signal, hereafter called

CrowdMM'12, October 29, 2012, Nara, Japan.
ACM 978-1-4503-1589-0/12/10.

reference pulse, which models the regions of interest of each video as explained in the third stage. Here, in order to construct the pulse signal the exact location of the pulses are defined by means of the generalized local maxima of the experimental smooth signal. It must be noted that while the high of the pulse does not affect our results the width of the pulse *D* is a parameter that must be treated carefully. In particular, the variability of the average signal, determines the order of the pulse width *D*. Here, we propose that the pulse width should be equal to the average half of the widths of the bell-like regions of the signals.

In the third stage we construct the reference pulse signal, which models the regions of interest of each video. For compatibility reasons and without loss of generality the shape of the pulses (width and high) are the same as for the user's activity pulses. On the other hand, the exact locations of the pulses are defined as the center of the corresponding regions of interest as defined initially from the experimental setting.

It is our aim to examine whether the two signals (user activity and reference pulses) are correlated, e.g., whether the patterns revealed from the user's activity are correlated with objective regions of interest of each video. In order to check this hypothesis the cross correlation coefficient was used which estimates the degree to which two series are correlated (e.g. Vanmarcke [3]).

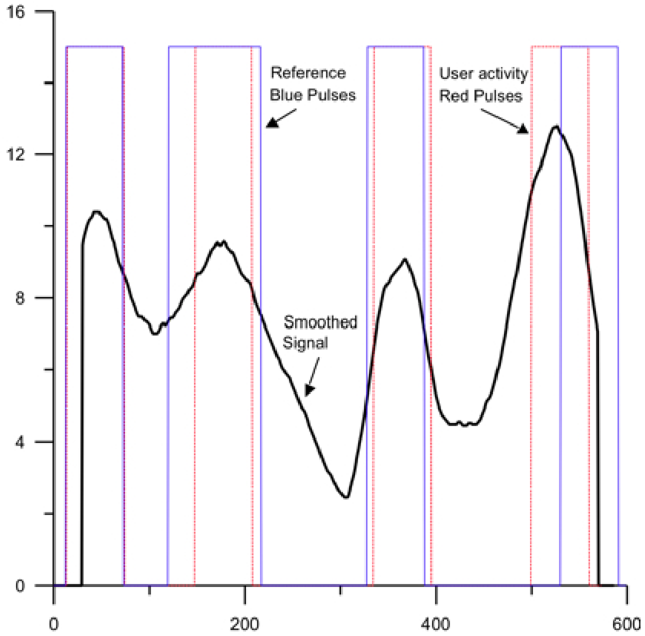

Figure 1 The video http: //www.youtube.com/watch?v=GOQfIXxbjlE is a documentary. The pulse width D is 50 seconds and the smoothing window T is 40 seconds. The pulse modeling is reported with respect to the center of each pulse.

Results. In the following, we present the results of the GoBackward signal analysis for one video. The analysis of the user activity signal was based on an exploration of several alternative averaging window sizes. The results of the pulse modeling methodology are depicted in figure 1. The smoothed signal is plotted with the solid black curve. The pulse signals were extracted from the corresponding local maxima are depicted with the red discontinued pulse signal while the pulse signals that model the regions of interest are depicted with the blue solid pulse. Although the correlation of the constructed pulse signals is visually evident in the graph (Figure 1), we also employed the cross correlation coefficient in order to establish the respective quantitative measures. Indeed, the cross correlation coefficients that we estimated were 0.67, indicating strong correlation between the two signals (reference and user signal). Thus, the pulse modeling process has identified the video scenes with high accuracy. In Figure 1 the video scenes (S1. . . S5) detected by the algorithm (user activity pulse modeling) are compared to the reference video scenes.

Conclusions. In summary, the above results demonstrate the efficacy of this approach and provide a small set of parameters (video browsing actions, averaging window duration *T*, pulse width *D*) that need to be further explored. Crowdsourcing of user interactions within web video should be suitable for a growing number of videos on the web contributed by schools and hobbyists such as lecture and how-to videos. Moreover, users browse web video in established ways that require no special buttons, besides play, pause and skip.

The research methodology provides several opportunities for a new layer of signal processing research on video content; one that is concerned with patterns of user activity instead of video contents ones. The idea to interpret user's actions as a time-based signal (sum of discrete pulses as was mentioned before) is common to other fields. Actually what is common is the existence of different populations (here different types of buttons) of discrete nature (discrete user's actions) and their patterning or morphogenesis in the corresponding space (here patterning of user's actions within the video duration). Note that since populations are discrete in nature the corresponding emerged patterns are also discrete thus resulting to theoretical models by means of pulses of definite width. Therefore, further research should consider the re-appropriation of signal processing methods for the purpose of user activity understanding within web video.

Acknowledgements. The work reported in this paper has been partly supported by project CULT (http://cult.di.ionio.gr). CULT (MC-ERG-2008-230894) is a Marie Curie project of the European Commission (EC) under the 7th Framework Program (FP7).

2. REFERENCES

[1] C. Gkonela, K. Chorianopoulos. VideoSkip: event detection in social web videos with an implicit user heuristic. Multimedia Tools and Applications, http://dx.doi.org/10.1007%2fs11042-012-1016-1, 2012.

[2] D. A. Shamma, R. Shaw, P. L. Shafton, and Y. Liu. Watch what i watch: using community activity to understand content. In MIR '07: Proceedings of the international workshop on Workshop on multimedia information retrieval, pages 275–284, New York, NY, USA, 2007. ACM.

[3] E. Vanmarcke. Random fields: Analysis and Synthesis. Cambridge, MA: MIT Press, 198