Information Sciences 298 (2015) 315–329

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Early prediction in collective intelligence on video users' activity

Markos Avlonitis, Ioannis Karydis *, Spyros Sioutas

Dept. of Informatics, Ionian University, Greece

ARTICLE INFO

Article history:
Received 10 January 2014
Received in revised form 23 November 2014
Accepted 26 November 2014
Available online 17 December 2014

Keywords:
Collective intelligence
Early prediction
Scene popularity
Stochastic patterns
Neural networks

ABSTRACT

The huge volume of available video content calls for methods that offer insight to the content without necessitating burdensome users' extra effort or being applicable to specific types or conditions. Based on experimentation on collective users' interactions on a controlled user-experiment, this work analyses the results collected following the argument that bell-shaped reference patterns are shown to significantly correlate with scenes of interest for each video, as designated by the viewers. Though, in order to ensure the correlation of results to bell-shaped reference patterns, aggregation of a number of users' interactions is required. In order to overcome such an impediment and adhere to a real-case cold start scenario, we propose a stochastic transformation of the aggregated users' interaction signal into a space defined by its correlation to the bell-shaped reference patterns that is shown to offer significant amelioration as to the percentage of users' interaction required in order to achieve comparable results to the original users' interaction space. Moreover, to ensure further the realistic character of the proposed scenario, given an amount of already collected users' interaction, the interaction of new users' is shown to be predictable using neural network time series prediction and modeling methods. The results received indicate increased accuracy on how one can predict the most important scenes from low quantity early data of users' interactions as well as future interaction of unique users. In practice, the proposed techniques might improve both navigation within videos on the web as well as video search results with personalised video thumbnails.

1. Introduction

Nowadays, video content consumption and creation is easier than ever. On one side, widespread penetration of fast and highly interactive internet allows for an ever increasing number of users enjoying video content while on the other hand affordable storage as well as high-quality capturing devices have made creating such content an ubiquitous process with unprecedentedly high demand. The most popular web streaming video content service, YouTube [16], serves more than 1 billion unique users per month, while storing 100 h of video every minute [17]. Accordingly, being able to make sense of the available content in a computerised manner, that is, being able to extract new and interesting information that is otherwise very difficult to be done due to the sheer volume of data, is of paramount importance.

* Corresponding author.
E-mail addresses: avlon@ionio.gr (M. Avlonitis), karydis@ionio.gr (I. Karydis), sioutas@ionio.gr (S. Sioutas).

http://dx.doi.org/10.1016/j.ins.2014.11.039
0020-0255/

Traditional content-based methodologies for the aforementioned data mining processes examine the actual content of each video in order to extract information. Nevertheless, their performance and capabilities fall short in certain occasions and thus research has recently focused on contextual or user-based semantics.[1] Such semantics rely on a broad spectrum of interactive behaviour and "social activities" users exhibit and perform in relation to video content consumption such as sharing with others, assigning comments/tags, producing replies by means of other videos or even just expressing their preference/rating on the content. Rich as these "social metadata" may be, they have also been critiqued [2] as offering extra burden in the usual content consumption process that mainly includes viewing and browsing, by necessitating extra user effort, leading to the long-tail effect as to their existence. Thus, "social metadata" aside, research [8,11] has examined the interaction information during the core processes of video content consumption, i.e. during viewing and browsing.

The previously mentioned increased interactivity Web 2.0 offered for the consumption of video content additionally assisted, through web-oriented architectures, in exposing content providers' functionality that other applications leverage and integrate in order to provide a set of much richer applications. In order to enhance the effectiveness of these applications, there is need for extensive studies of large users' interaction data. To this end, the design and implementation of controlled users' experiments has gained increasing attention. Indeed, controlled experiments provide sets of data of (almost) any desirable size under controlled conditions giving thus the possibility to study specific users' interaction properties for specific Web content. Accordingly, Gkonela and Chorianopoulos [8] utilising the SocialSkip platform [4,12] collected a pioneering user-based interaction dataset by conducting a controlled experiment during video content consumption providing a clean set of data that was easier to analyse. The platform integrated custom interface videos from YouTube with querying form functionalities of the Google Docs API in order to create an environment that would allow for video content consumption as well as user querying in order to accumulate data. Using a modified version of the SocialSkip platform, Spiridonidou et al. [13] conducted a controlled user-experiment designed and implemented in order to achieve high degree of realism to the typical contemporary scenario of web video-streaming services. In this way, the collective behaviour of Web users watching the video content emerged by means of characteristic patterns in their activity leading to *collective intelligence* as to the importance of video content solely from users' interactions with the video player.

The dataset introduced at [8] was based on a set of restrictive assumptions as to its generality and thus the experiment conducted in [13] adhered to the following more real-use principles:

- the viewing interface included the controls found in YouTube in order to simulate realistic user video content consumption,
- content viewing did not have any time limit, again in order to simulate realistic user video content consumption,
- the questionnaire requested free-text replies in order to ensure that users declared the scenes they thought of as most important without interference or guidance,
- the significance of scenes was not predefined, allowing for true collective intelligence.

Nevertheless, the dataset collected from the experiment conducted in [13] did not undergo thorough testing in order to show that bell-shaped reference patterns significantly correlate with user defined scenes of interest in the video content.

Moreover, the existence of the aforementioned users' interaction signal can provide the basis for new series of metrics in order to study new characteristics of users' interactions as well as describe the content on which the interaction took place. Indeed, the aggregated users' interaction signal is useful (e.g. for identification of most important scenes), but requires a certain amount of users having interacted with the content interface in order to provide for any conclusions. Accordingly, real-use scenarios that refer to content with little consumption/interaction would not be able to benefit the methodologies' capabilities. It is thus required to devise methodologies that can predict the aggregated users' interaction signal of the necessary quantity of users from a set of low quantity early data of users' interactions.

In the same direction of early interaction prediction, aiming now not at the previously mentioned aggregated users' interaction but for a single user's interaction, such methods could be valuable in numerous cases such as: to model the specific user and his/her interaction pattern, to enhance and customise associated supportive content of the actual consumed content such as advertisements or other entities of interactive environments as well as to adjust content delivery ranging from parameters of the quality of scenes predicted to be interesting to customised content thumbnailing/summarisation to customised recommendations. Accordingly, methods are required that predict a user's interaction based on other users' interaction on the same content as well as a potential small amount of the specific user's initial interaction.

1.1. Motivation and contribution

Bearing in mind the aforementioned lack of analysis on the correlation of the aggregated users' interaction signal of the datased collected in [13] to bell-shaped reference patterns as well as the requirements of small user quantity early interaction prediction both on the aggregated users' as well as on a single user's interactivity level, the contribution of this work is summarised as follows:

[1] For a detailed discussion about the complementary character of the two approaches see [2].