Image and Vision Computing 32 (2014) 630–640

Contents lists available at ScienceDirect

Image and Vision Computing

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

Automatic measurement of ad preferences from facial responses gathered over the Internet☆

Daniel McDuff [a,1], Rana El Kaliouby [b], Thibaud Senechal [b], David Demirdjian [c], Rosalind Picard [a]

[a] MIT Media Lab, Cambridge 02139, USA
[b] Affectiva, Waltham 02452, USA
[c] CSAIL, MIT, Cambridge 02139, USA

ARTICLE INFO

Article history:
Received 2 June 2013
Received in revised form 29 October 2013
Accepted 9 January 2014
Available online 17 January 2014

Keywords:
Facial expressions
Preferences
Classification
Crowdsourcing

ABSTRACT

We present an automated method for classifying "liking" and "desire to view again" of online video ads based on 3268 facial responses to media collected over the Internet. The results demonstrate the possibility for an ecologically valid, unobtrusive, evaluation of commercial "liking" and "desire to view again", strong predictors of marketing success, based only on facial responses. The area under the curve for the best "liking" classifier was 0.82 when using a challenging leave-one-commercial-out testing regime (accuracy = 81%). We build on preliminary findings and show that improved smile detection can lead to a reduction in misclassifications. Comparison of the two smile detection algorithms showed that improved smile detection helps correctly classify responses recorded in challenging lighting conditions and those in which the expressions were subtle. Temporal discriminative approaches to classification performed most strongly showing that temporal information about an individual's response is important; it is not just how much a viewer smiles but when they smile. The technique could be employed in personalizing video content that is presented to people while they view videos over the Internet or in copy testing of ads to unobtrusively quantify ad effectiveness.

1. Introduction

The face has been shown to communicate discriminative valence information with zygomatic muscle (AU12) activity greater in ads with a positive emotional tone and corrugator muscle (AU4) activity greater in ads with a negative tone [3]. There is evidence that facial expressions can predict variables related to advertising success, with facial responses correlated with recall [8] and ad "zapping" [26]. The Facial Action Coding System (FACS) [6] is a catalog of 44 unique action units (AUs) that correspond to each independent movement of the face's 27 muscles. Computer vision systems can now reliably code many of these actions automatically [30]. In this paper we show that self-reported video advertisement liking and desire to view again can be accurately predicted from automatically detected spontaneous smile (AU12) responses captured in unconstrained settings over the Internet. Fig. 1 shows the framework we use to automatically classify media preferences.

Advertisement likability is a key measure of sales success in marketing [7,23]. Likability is described as having the dimensions of entertainment, energy, relevance, empathy, irritation and familiarity. However, these metrics are hard to quantify objectively and in many real-life applications self-report measures are impractical to capture (e.g. when people are watching TV). Advertisers wish to increase a viewer's desire to view an advertisement again; thus desire to view the ad is another measure of advertising effectiveness. Knowledge of likability and desire to view again are not only useful in advertisement copy-testing but could also be used to personalize the content viewers are shown when watching TV over the Internet using platforms such as Netflix or Hulu. In the case of humorous ads, smile activity is a good measure of positive advertisement attitude or liking, and this can be measured continuously and unobtrusively from video images [14].

Earlier work has shown that facial responses to content can be collected efficiently over the Internet, and that there are significant differences in the aggregate smile responses of groups that report liking a commercial compared to those that report disliking it [16]. A similar difference in the aggregate responses was observed between individuals who report a desire to watch the content again versus those that report no such desire. However, whether these aggregate trends allow accurate discrimination of liking versus disliking responses on an individual basis was not explored. The first published automated analysis on individual level prediction has shown that it is possible to accurately predict preferences [17]. This paper presents work extending and improving these initial results.

The dynamics of smile responses are rich and can be used to distinguish between different message judgments associated with them [2,9]

☆ This paper has been recommended for acceptance by Stan Sclaroff.

URL: djmcduff@media.mit.edu, http://media.mit.edu/~djmcduff (D. McDuff), kaliouby@affectiva.com (R. El Kaliouby), thibaud.senechal@affectiva.com (T. Senechal), demirdji@csail.mit.edu (D. Demirdjian), picard@media.mit.edu (R. Picard).
[1] Tel.: +1 617 253 8628.

0262-8856/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.imavis.2014.01.004

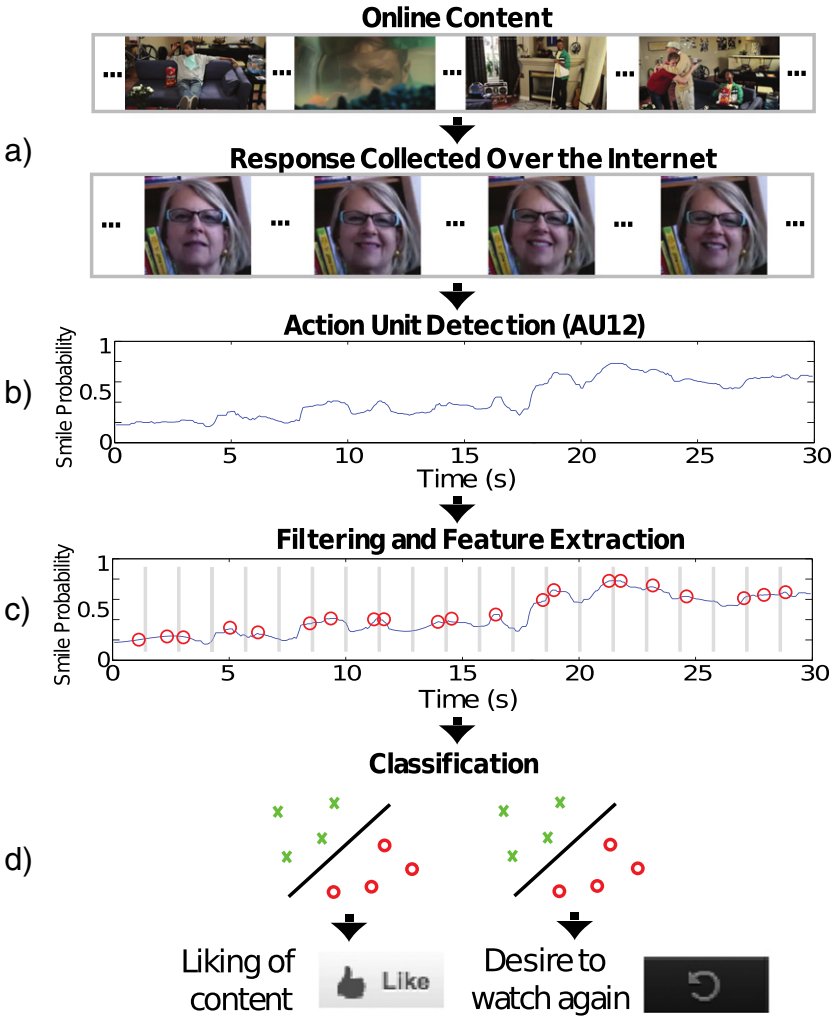

Fig. 1. Framework for classification of content liking and desire to view again based on automatically detected smile responses recorded over the web. a) Facial response to online media captured, b) smiles (AU12) detected automatically, c) temporal features extracted from smoothed smile track, d) features used to predict liking and desire to view again.

and whether they are posed or naturalistic [27]. Discriminative temporal models, in particular, Hidden Conditional Random Fields (HCRFs) and Latent Dynamic Random Fields (LDCRFs) have been shown to perform well in classification of noisy behavioral data [28,24]. In this work we test the power of both static and dynamic models to predict advertisement liking and desire to watch again from an ecologically valid, but challenging and noisy, dataset of automatically detected spontaneous smile responses. Many applications of facial expression recognition for predicting consumer preferences or highlights are tested on data collected within a lab setting which is unlike the environment in which the media is normally viewed [10,31]. In our work we use data collected in a setting much closer to that in which it is normally consumed.

The main contributions of this paper are: 1) to present results on classification of liking and desire to view again of Internet videos based on the facial responses analyzed over the Internet, 2) to identify conditions under which misclassifications (false positives, false negatives) occurred and 3) to show that improved smile detection can reduce the number of misclassified responses considerably. The remainder of the paper will discuss the data collection, feature extraction, modeling and results of the work.

2. Related work

Smile detection is one of the most robust forms of automated facial analysis available. Whitehall et al. [29] present a smile classifier based on images collected over the Internet and demonstrates strong performance on this dataset. A subset of the data was released as the MPL GENKI[2] dataset. Shan [22] demonstrates an accurate and faster smile detector on the MPL GENKI-4K dataset.

Joho et al. [10] showed that it is possible to predict personal highlights in video clips by analyzing facial activity. However, they also noted the considerable amount of individual variation in responses. These experiments were conducted in a laboratory setting and not in a natural context; our work demonstrates the possibility of extending this work to online content and real-world data. Zhao et al. [31] designed a video indexing and recommendation system based on automatically detected expressions of the six basic emotions (sadness, anger, fear, disgust, happiness, surprise). However, this was tested on only a small number of viewers ($n = 10$) in a lab setting.

Teixeira et al. [25] showed that inducing affect is important in engaging viewers in online advertising and is associated with reducing their frequency of "zapping" (skipping the advertisement). They demonstrated that joy, as measured by smile responses, was one of the states that increased viewer retention in the commercial. Again, these studies were performed in a laboratory setting rather than in the wild. Micu and Plummer [19] measured zygomatic major (AU12) activity using facial electromyography (EMG) while people watched TV ads. They

[2] http://mplab.ucsd.edu, The MPLab GENKI Database, GENKI-4K Subset.