

Predicting Online Media Effectiveness Based on Smile Responses Gathered Over the Internet

Daniel McDuff, Rana el Kaliouby, David Demirdjian and Rosalind Picard

Abstract—We present an automated method for classifying “liking” and “desire to view again” based on over 1,500 facial responses to media collected over the Internet. This is a very challenging pattern recognition problem that involves robust detection of smile intensities in uncontrolled settings and classification of naturalistic and spontaneous temporal data with large individual differences. We examine the manifold of responses and analyze the false positives and false negatives that result from classification. 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” and “desire to view again” classifiers was 0.8 and 0.78 respectively when using a challenging leave-one-commercial-out testing regime. The technique could be employed in personalizing video ads that are presented to people whilst they view programming over the Internet or in copy testing of ads to unobtrusively quantify effectiveness.

I. INTRODUCTION

The face has been shown to communicate discriminative valence information with zygomatic muscle activity greater in commercials with a positive emotional tone and corrugator muscle activity greater in commercials with a negative tone [2]. It has been shown that facial expressions can predict variables related to advertising success, with facial responses correlated with recall [6] and ad “zapping” [18]. The Facial Action Coding System (FACS) [4] is a catalogue 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 [22]. 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. Figure 1 shows the framework.

Advertisement likability is a key measure of sales success in marketing [5], [15]. It 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

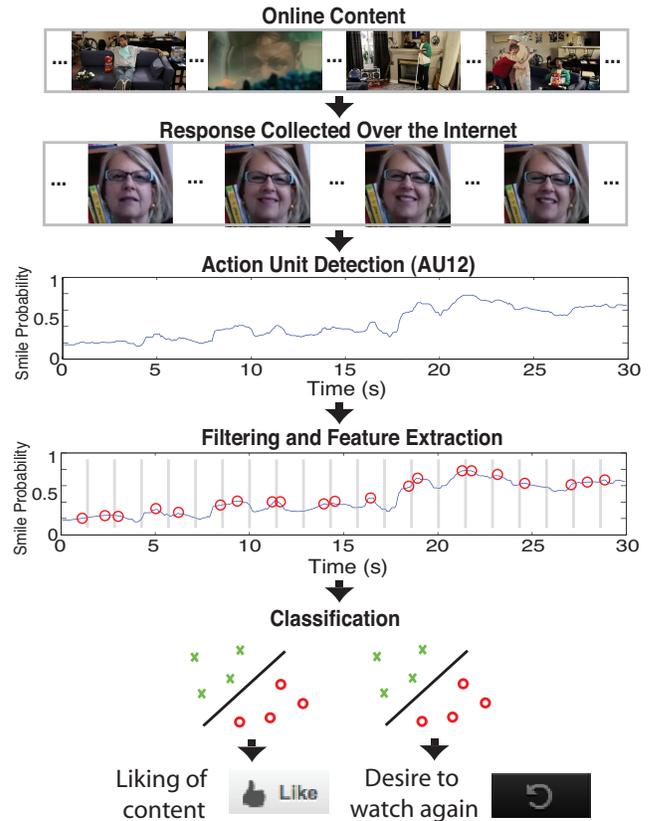


Fig. 1. Framework for classification of content liking and desire to view again based on automatically detected smile responses recorded over the web.

desire to view an advertisement again and desire to view the ad is therefore 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 commercials, smile activity is a good measure of positive advertisement attitude or liking, and this can be measured continuously and unobtrusively from video images [10].

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 [11], [12]. Similar difference in the aggregate responses were observed between individuals who

This work was funded by the Media Lab Consortium Members.

D. McDuff, R. el Kaliouby and R. Picard are with the Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. djmcduff@media.mit.edu, kaliouby@media.mit.edu, picard@media.mit.edu. Picard and Kaliouby also hold positions with Affectiva, Inc.

D. Demirdjian is with CSAIL, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. demirdji@csail.mit.edu

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 dynamics of smile responses are rich and can be used to distinguish between different message judgements associated with them [1], [7] and whether they are posed or naturalistic [19]. 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 [20], [16]. 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.

The main contributions of this paper are; 1) to present the first analysis of individual facial responses to media collected over the Internet, 2) to present results on classification of liking and desire to view again of Internet videos based on the facial responses analyzed over the Internet and 3) to identify conditions under which misclassifications (false positives, false negatives) occurred. The remainder of the paper will discuss the data collection, feature extraction, modeling and results of the work.

II. RELATED WORK

Smile detection is one of the most robust forms of automated facial analysis available. Whitehall et al. [21] present a smile classifier based on images collected over the internet and demonstrate strong performance on this dataset.

Joho et al. [8] 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.

Teixeira et al. [17] showed that inducing affect is important in engaging viewers in online video adverts and to reduce the frequency of “zapping” (skipping the advertisement). They demonstrated that joy was one of the states that stimulated viewer retention in the commercial. Although not directly related, it is intuitive that smiles might play a significant role in a joy response. Again, these studies were performed in a laboratory setting rather than in the wild.

III. DATA COLLECTION

A. Framework

Using a web-based framework similar to that described in [citation removed for anonymization] 3,268 videos (2,615,800 frames) were collected of facial responses to three successful Super Bowl commercials: 1. Doritos (“House sitting”, 30 s), 2. Google (“Parisian Love”, 53 s) and 3. Volkswagen (“The Force”, 62 s). All three ads were somewhat amusing and were designed to elicit smile or laughter responses. In addition the VW ad was voted the most

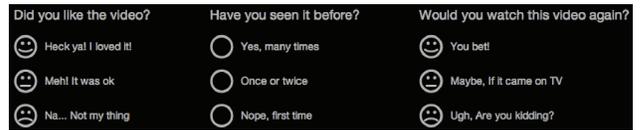


Fig. 2. The self-report questions the viewers were presented with after watching the commercial.

successful ad of the year 2011¹. The responses were collected in natural settings via the Internet and the application was promoted on the Forbes website (citation will be added in camera ready version). In total 6,729 people opted-in and completed the experiment. For the automatic analysis here videos for which it was not possible to identify and track a face in at least 90% of frames were disregarded; this left 4,502 videos (67%). All videos were recorded with a resolution of 320x240 and a frame rate of 15 fps. Participants were aware from the permissions that their camera would be turned on and this may have had an impact on their facial response. However, people responded naturally in a vast majority of cases. McDuff et al. [12] provide more detailed information about the data collection.

Following each commercial, viewers could optionally answer three multiple choice questions: “Did you like the video?” (liking), “Have you seen it before?” (familiarity) and “Would you watch this video again?” (rewatchability). Figure 2 shows a screenshot of the questions asked. In this paper the relationship between the smile responses and the self-report responses for each question is examined. Since viewers were not obligated to complete the responses and the questions “timed out” once the smile response was computed, some participants only answered some of the questions and some none of the questions. On average each question was answered by 47.6% of viewers, which still provides almost 2,400 examples for each question and commercial combination.

B. Smile Detection and Dynamics

To compute the smile probability measure we used a custom algorithm developed by Affectiva. This tracks a region around the mouth using the Nevenvision facial feature tracker² and computes Local Binary Pattern (LBP) [14] features within this region. An ensemble of bagged decision trees is used for classification. The classifier outputs a probability that the expression is a smile. A smile probability value between (0 to 1) is calculated for every frame in which a face was tracked, yielding a one-dimensional smile track for each video. Figure 3 shows an example of one smile track with screenshots of six frames and demonstrates how the smile probability is positively correlated with the intensity of the expression. We refer to the classifier output as the smile intensity from this point on.

The smile classifier was trained on examples from the CK+ and MPL³ databases. We tested the classifier on 3,172

¹<http://www.adweek.com/>

²Licensed from Google, Inc.

³<http://mplab.ucsd.edu>, The MPLab GENKI Database, GENKI-4K Subset

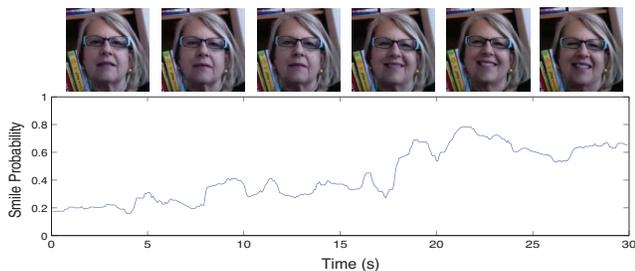


Fig. 3. A smile track with screenshots of the response, demonstrating how greater smile intensity is positively correlated with the probability output from the decision tree classifier.

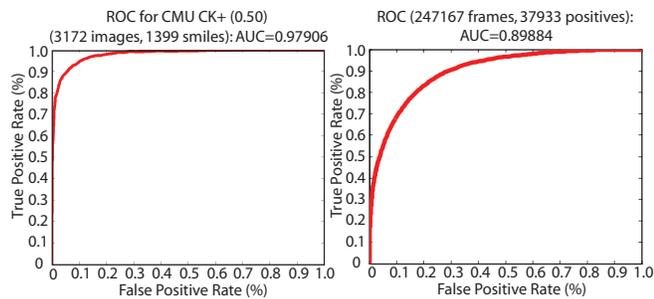


Fig. 4. ROC curves for the smile detection algorithm. ROC curve tested on CK+ database (left), ROC curve for training on CK+ and MPL and testing on webcam data (right).

frames from the CK+ database (the test videos were not included in the training set). The resulting ROC curve is shown in Figure 4 (left), and the area under the curve is 0.979. We also tested how well the smile classifier performs on crowdsourced face videos from a webcam where there is no control on the quality of the resulting face videos (these videos were from a similar but different study to the one described here) 247,167 frames were randomly selected for ground truth labeling. Three labelers labeled every frame of each video and the majority label was taken for each frame. The resulting ROC curve is shown in Figure 4 (right); the area under the curve is 0.899. The performance of the smile classifier degrades with the uncontrolled videos compared to the CK+; however it is still very accurate.

As such, a 1-dimensional smile track was computed for each video with length equal to the number of frames of the video. These smile tracks, and the corresponding self-report liking response labels, are used for the analysis and classification in the rest of the paper.

C. Liking and Desire to Watch Again

In this paper the responses to the questions “Did you like the video?” and “Would you watch this video again?” are considered. The answers available for the first question were: “Heck ya! I loved it!” (liking), “Meh! It was ok” (neutral) and “Na... Not my thing” (disliking). The answers available for the second question were: “You bet!” (strong desire), “Maybe, If it came on TV.” (mild desire) and “Ugh, are you kidding?” (weak desire).

For this analysis we consider the binary cases in which the viewer would report liking vs. disliking the commercial

TABLE I

NUMBER (AND PERCENTAGES) OF “LIKING” AND “DESIRE TO VIEW AGAIN” CLASSES FOR EACH OF THE COMMERCIALS.

Ad	Liking		Desire to View Again	
	Like (+ve)	Dislike (-ve)	Strong (+ve)	Weak (-ve)
Doritos	280 (80%)	70 (20%)	91 (73%)	34 (27%)
Google	341 (79%)	92 (21%)	142 (63%)	84 (37%)
VW	711 (95%)	40 (5%)	381 (89%)	46 (11%)
Total	1332 (86%)	212 (14%)	614 (79%)	164 (21%)

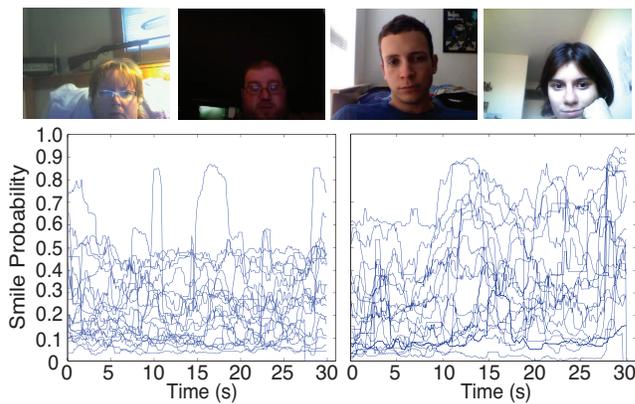


Fig. 5. Top) Examples of frames from the dataset showing the variability in illumination, contrast, pose and scale of the viewer’s faces. Bottom) 20 smile tracks for each of the likability self-report subgroups: disliking (left) and liking (right), showing the variability in responses.

and strong vs. weak desire to view the commercial again. Therefore the neutral and mild responses are not considered in the classification. This is reasonable as the people in these categories do not show a strong feeling towards the content and therefore showing them the content again, or not, will not represent a missed opportunity or negative consequence (misclassification for this group represents a low cost), whereas showing someone the content again who had a very weak desire to see it may be a waste of resources or have a negative impact on a viewer’s perception of the brand. Table I shows the number of samples, class priors, for each of the classes.

IV. FEATURE EXTRACTION

The smile tracks that satisfied the 90% trackable criteria were used. First they were filtered with a low pass FIR filter to smooth the signals. The high frequency 3dB cut-off of the low-pass filter was 0.75 Hz. Secondly, features were extracted from the smile tracks as shown in Figure 6.

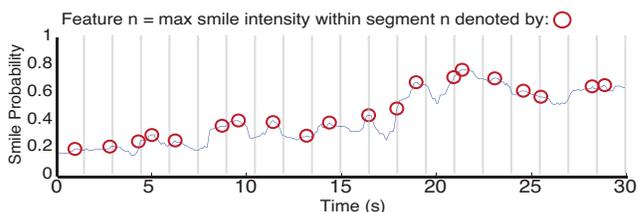


Fig. 6. Feature extraction involved selecting the maximum peaks from the filtered smile track within 20 evenly spaced temporal bins. This formed a feature vector of length 20.

The filtered tracks were divided evenly into 20 segments and the peak smile intensity for each segment calculated to form a feature vector of length 20. This removed any prior information held in the length of the content and will help promote generalizability of the resulting model. The videos have a frame rate of 14fps and therefore the number of frames for each segment was 21 (Doritos), 37 (Google) and 43 (VW). Tests were run with more than 20 features but there was no significant change of performance in results.

Before training and testing classifiers, which is described in Section V, we computed a 2D mapping of the data to get an intuition about the manifold of responses. Figure 7 shows a 2D mapping of the responses computed using Linear Discriminant Analysis (LDA). The reported “disliking” responses are shown in red and the reported “liking” in green. Examples of four of the smile responses are also shown. Gaussian distributions have been fitted to the data along the most discriminative axis. The distributions of the two classes are different with the disliking class characterized by a lower mean intensity and in particular a much lower response towards the end of the commercial. The global gradient, or trend, of the responses is also an important feature. However, it is clear that global features seem insufficient for accurate classification, for instance the mean of example b in Figure 7 is greater than that of example c despite it being a disliking response compared to a liking response.

V. CLASSIFICATION

We compare both static and temporal, generative and discriminative approaches in predicting reported liking and desire to view again based on smile features. For the classification we attempt to correctly classify examples of disliking and liking responses and strong and weak desire responses all of which satisfy the 90% trackable criteria.

A. Models

Naive (Class Priors): We provide a naive baseline for which test labels are predicted randomly but with a distribution that matches the training data class priors. Table I shows the class priors for each of the commercials. The results shown for the class priors model are based on an average performance over 200 sets of predictions.

Naive Bayes: A Naive Bayes (NB) classifier was used to provide a baseline performance. As the Naive Bayes classifier is non-parametric no validation was performed.

Support Vector Machine: Support Vector Machines (SVM) are a static approach to classification and therefore do not explicitly model temporal dynamics. A Radial Basis Function (RBF) kernel was used. During validation the penalty parameter, C , and the RBF kernel parameter, γ , were each varied from 10^k with $k=-3, -2, \dots, 3$. The SVM’s were implemented using libSVM [3].

Hidden Markov Models: An HMM was trained for each of the two classes. This is a generative approach to modeling the data. During validation, the number of hidden states (3, 5 and 7) was varied.

Hidden-state and Latent Dynamic Conditional Random Fields: HCRFs [20] and LDCRFs [13] are discriminative approaches to modeling temporal data. The CRF model and its variants remove the independence assumption made in using Hidden Markov Models (HMMs) and also avoids the label-biasing problem of Maximum Entropy Markov Models (MEMMs) [9]. The dynamics of smiles are significant in distinguishing between their meanings [1]; as such, we hypothesized a potential benefit in explicitly modeling the temporal dynamics of the responses. Song et al. [16] describe a Gaussian temporal-smoothing kernel that improved performance without increasing the computational complexity of inference. This takes a Gaussian-weighted average of observations within a moving window of size N . The HCRF and LDCRF classifiers were tested with Gaussian temporal smoothing window sizes 0, 1, 3 and 5. We found that beyond a window size of 5 performance saturated or began to fall off. During validation, the regularization factor (10^k with $k = -2, -1, \dots, 2$) and number of hidden states (3, 5 and 7) was varied. The HCRF and LDCRF classifiers were implemented using the HCRF toolbox for MATLAB [13].

B. Validation and Training Scheme

It is important that the prediction of success is independent of the particular commercial and that the features generalize across new content. In order to test this we use a challenging leave-one-commercial-out scheme. The data for one commercial was removed for testing and the remaining data was used for training and validation. This was repeated for each of the three commercials. For validation a leave-one-commercial-out methodology was used. The training data set was split and data for one commercial were with-held and validation performed to find the optimal model parameters, this was repeated for both commercials in the training set. The area under the curve (AUC) was maximized in the validation stage to choose parameters. The resulting model was tested against the test data withheld. For the SVM the median parameters selected were $\gamma = 0.1$ and $C = 1$. For the HMM the median number of hidden states selected was 6 hidden states. For the HCRF the median parameters selected were 9 hidden states and regularization factor (RF), $r = 10$. For the LDCRF the median parameters selected were 7 hidden states and RF, $r = 1$.

VI. RESULTS AND DISCUSSION

The Figures 8 a and b show ROC curves for the classifiers tested for predicting liking and desire to view again respectively. Figures 9 and 10 show the area under the ROC curve for each of the classifiers. The accuracy for the classifiers closest to the (0,1) point on the ROC curve is also shown. For both liking and desire to view again the LDCRF classifier with Gaussian smoothing window, ω , of 3 shows the strongest performance with an area under the curve of 0.8. Table III shows the confusion matrix for the LDCRF classifier ($\omega=3$) with the optimal decision threshold based on the point on the ROC curve closest to (0,1). There are a considerable number of false negatives, but in other respects

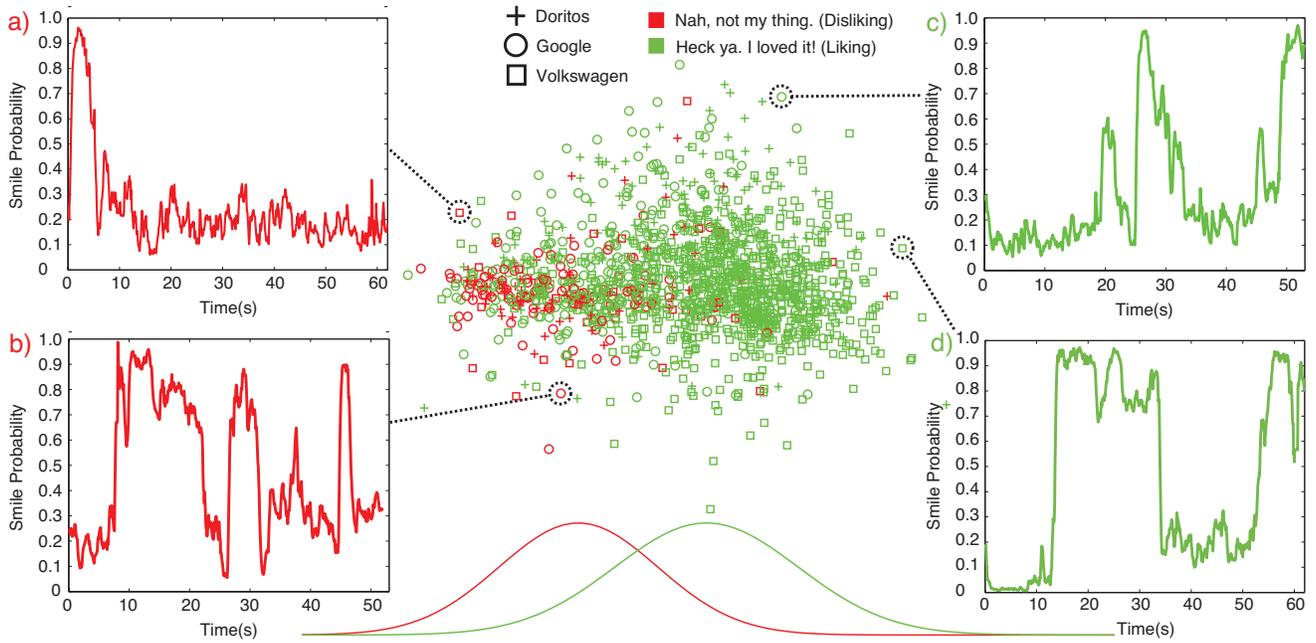


Fig. 7. Linear Discriminant Analysis (LDA) mapping of the smile tracks with labels of reported liking. Smile tracks for those that report disliking (red) and liking (green).

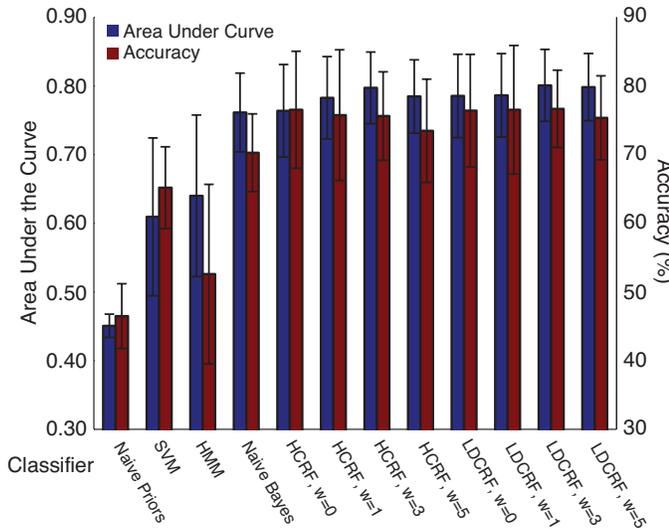


Fig. 9. Bar chart of area under the curve (AUC) and accuracy (%) for predicting liking of content.

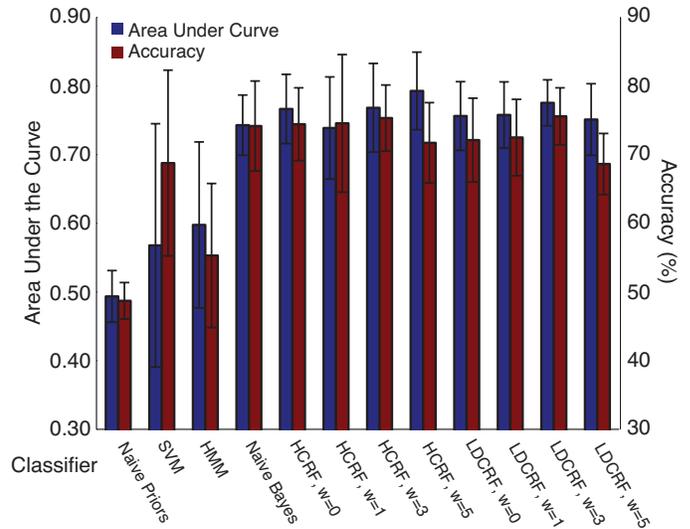


Fig. 10. Bar chart of area under the curve (AUC) and accuracy (%) for predicting desire to view content again.

the performance is very good. We discuss specific details about the false positives and negatives in Section VI-A. It is useful to note here that there are a considerable number of cases where the smile track information seems insufficient to explain why the labels are positive or negative.

A. False Positive and False Negatives

Figure 11 (g-l) shows cases of false positive results that occurred across a majority (all but one) of the classifiers. Figure 11 (m-r) shows cases of false negative results that occurred across a majority (all but one) of the classifiers.

For comparison Figure 11 (a-f) shows cases of true posi-

tive results that occurred across a majority (all but one) of the classifiers. In a number of the true positive and false positive cases it is difficult to identify any difference in characteristics of the smile tracks for the positive and negative classes. For instance examples e and k seem to have high mean smile intensities and similar positive trends. For the false negative cases there are a number for which a very low smile probability was detected throughout the content (o and p in particular) but after which the participants reported loving the ad. In some cases (e.g. response p) the smile classifier correctly identified no expression yet the participant reported liking the clip, it seems that the self-report response does not

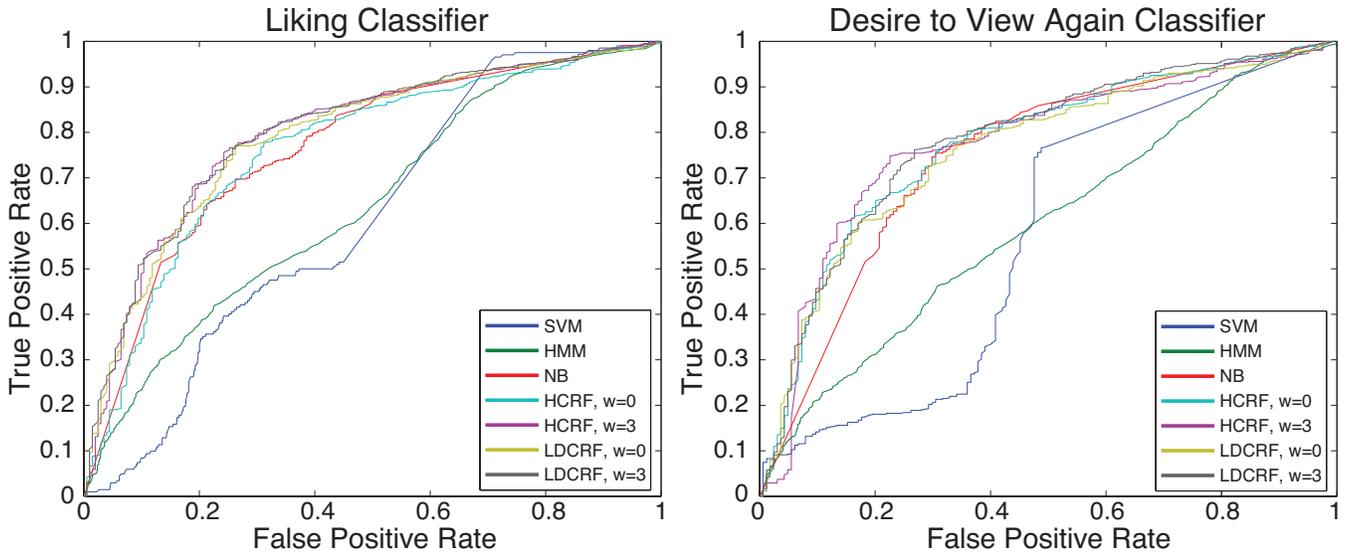


Fig. 8. Receiver Operating Characteristic (ROC) curves for the classifiers on the classification of self-reported liking (left) and desire to view again (right). In both cases the decision threshold was varied. For readability, only the curves for window size, $\omega = 0$ and 3 are shown. The areas under each curve are shown in Table II.

TABLE II

AUC FOR THE ROCs AND AUC FOR THE PRECISION-RECALL CURVES. ACCURACY FOR THE OPTIMAL CLASSIFIER (BASED ON THRESHOLD CLOSEST TO THE ROC (0,1) POINT) IS SHOWN IN BRACKETS.

Classifier	Liking	Desire to view again
SVM	0.61, 0.85 (45%)	0.57, 0.71 (70%)
HMM	0.64, 0.92 (53%)	0.60, 0.85 (55%)
Naive Bayes	0.76, 0.92 (70%)	0.74, 0.91 (74%)
HCRF, $\omega=0$	0.76, 0.94 (77%)	0.77, 0.91 (74%)
HCRF, $\omega=1$	0.78, 0.95 (76%)	0.74, 0.89 (75%)
HCRF, $\omega=3$	0.80, 0.95 (76%)	0.77, 0.90 (75%)
HCRF, $\omega=5$	0.78, 0.95 (73%)	0.79, 0.92 (72%)
LDCRF, $\omega=0$	0.79, 0.95 (76%)	0.76, 0.91 (72%)
LDCRF, $\omega=1$	0.79, 0.95 (77%)	0.76, 0.91 (73%)
LDCRF, $\omega=3$	0.80, 0.96 (77%)	0.78, 0.91 (76%)
LDCRF, $\omega=5$	0.80, 0.96 (75%)	0.75, 0.91 (69%)

TABLE III

CONFUSION MATRICES FOR THE BEST PERFORMING LDCRF CLASSIFIERS. TOP) FOR THE LIKING CLASSIFIER: $\omega = 3$, DECISION THRESHOLD = 0.613. BOTTOM) FOR THE DESIRE TO VIEW AGAIN CLASSIFIER: $\omega = 3$, DECISION THRESHOLD = 0.442

Liking	Actual +ve (Liking)	Actual -ve (Disliking)
Predicted +ve	1027	53
Predicted -ve	305	149

Desire	Actual +ve (Strong)	Actual -ve (weak)
Predicted +ve	468	44
Predicted -ve	146	120

necessarily fit with the smile response that one might expect. Frames from response p can be seen in Figure 11. From smile response alone it is unlikely that we could achieve 100% accuracy in predicting liking or desire to view again. In other cases the miss classification is due to noise in the smile classifier prediction. In response i there was no smile activity yet the classifier predicted a relatively high smile probability. This was a dark video which may have been a cause of the error. The latter errors can potentially

be removed by improving the performance of the smile prediction algorithm, however, the former errors raise much more interesting questions about whether self-report is an accurate reflection of feeling towards content or whether people always express their feeling through facial behavior. As the data collection was unconstrained a third possibility is that people participating may have been playing with the system and intentionally recording false data.

VII. CONCLUSIONS AND FUTURE WORK

This paper demonstrates the ability to predict reported liking and desire to watch again, two key measures of advertising effectiveness, automatically from spontaneous smile responses collected over the Internet. Using only temporal information about a person's smile response we can predict success with the area under the ROC curve for the liking and desire to watch again classifiers 0.8 and 0.78 respectively. These results were obtained using a challenging leave-one-commercial out training scheme to ensure generalizability across other amusing video ads. Temporal information was important in the classification and modeling the dynamics using an LDCRF classifier produced the best results. Classification errors highlight interesting examples in which people showed very little facial activity yet reported liking the content. In such a case it may not be possible to predict liking directly from facial activity. Other errors were due to the inaccuracy of the smile classifier under challenging conditions, namely low illumination and poor contrast.

The results of this work are very promising and demonstrate the possibility of automatically evaluating content effectiveness based on facial responses collected under very challenging conditions. Future work will take advantage of a greater number of action units and gestures (such as eyebrow raise (AU1+2), brow lowerer (AU4) and nose wrinkler (AU9)). The ground-truth considered in this study

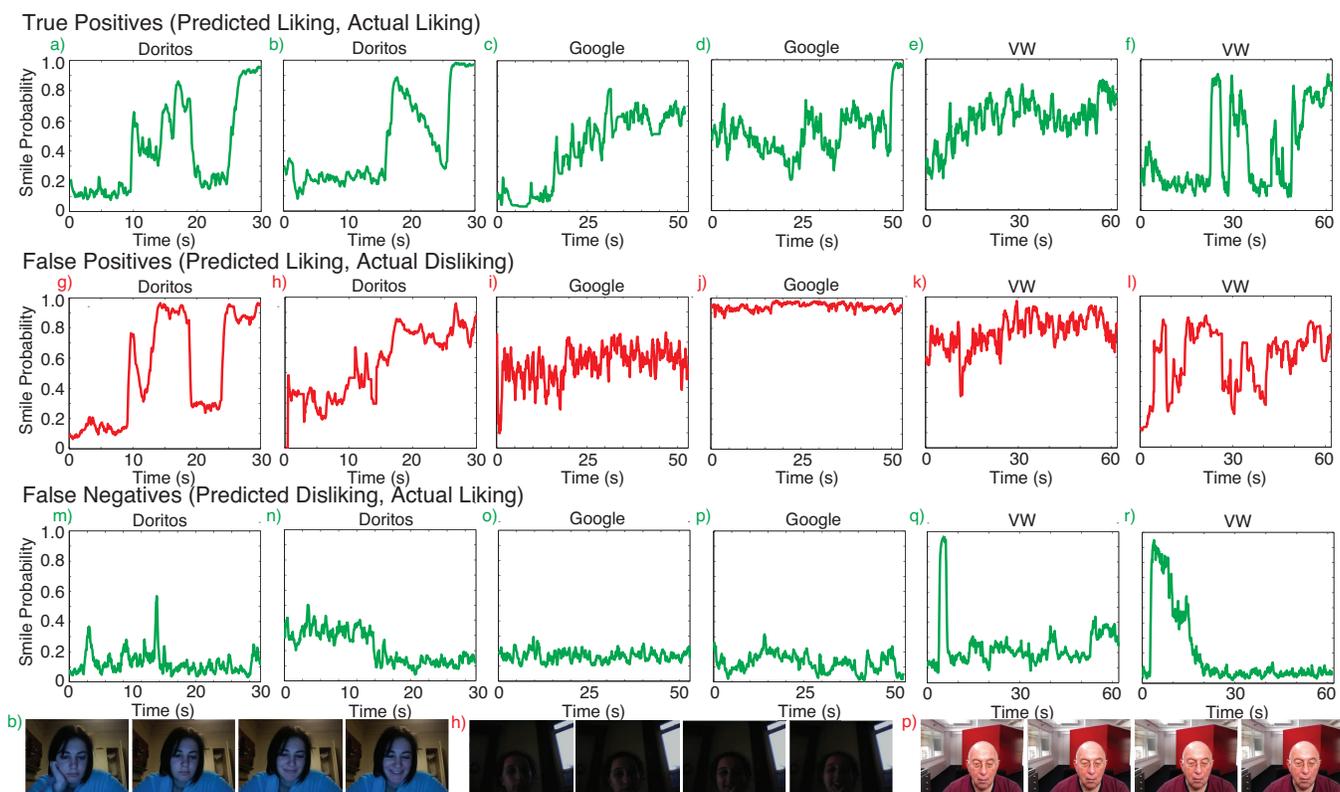


Fig. 11. Examples of true positives (top), false positive (center) and false negatives (bottom) which were consistent across all or all but one of the classifiers. Most of the false negative examples show responses with very low smile intensity despite the viewer reporting liking the commercial. Shown below are frames from examples of TP, FP and FN videos.

was self-reported and future work will consider behavioral measures of success (such as sharing) and sales.

These results represent a significant improvement in performance over a naive method of evaluating whether a viewer should be shown a particular piece of content again. The technique could be used in personalizing video ads that are presented to people whilst they view TV programming or in copy testing of ads to automatically quantify effectiveness.

VIII. ACKNOWLEDGEMENTS

Acknowledgements to be added in the final version. Richard Sadowsky, Oliver Wilder-Smith, Jay Turcot, Zhihong Zeng and Affectiva provided generous support in providing access to the crowdsourcing platform. Brian Staats provided a great front end design for the site. Jon Bruner and Forbes kindly promoted the work on their site. Google provided use of their facial feature tracker. This work was funded by the Media Lab Consortium Members.

REFERENCES

- [1] Z. Ambadar, J. Cohn, and L. Reed. All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of nonverbal behavior*, 33(1):17–34, 2009.
- [2] P. Bolls, A. Lang, and R. Potter. The effects of message valence and listener arousal on attention, memory, and facial muscular responses to radio advertisements. *Communication Research*, 28(5):627, 2001.
- [3] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.
- [4] P. Ekman and W. Friesen. Facial action coding system. 1977.
- [5] R. Haley. The arf copy research validity project: Final report. In *Transcript Proceedings of the Seventh Annual ARF Copy Research Workshop*, 1990.
- [6] R. Hazlett and S. Hazlett. Emotional response to television commercials: Facial emg vs. self-report. *Journal of Advertising Research*, 39:7–24, 1999.
- [7] M. Hoque, L. Morency, and R. Picard. Are you friendly or just polite?—analysis of smiles in spontaneous face-to-face interactions. *Affective Computing and Intelligent Interaction*, pages 135–144, 2011.
- [8] H. Joho, J. Staiano, N. Sebe, and J. Jose. Looking at the viewer: analysing facial activity to detect personal highlights of multimedia contents. *Multimedia Tools and Applications*, pages 1–19, 2011.
- [9] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.
- [10] D. McDuff, R. El Kaliouby, K. Kassam, and R. Picard. Affect valence inference from facial action unit spectrograms. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 17–24. IEEE.
- [11] D. McDuff, R. El Kaliouby, and R. Picard. Crowdsourced data collection of facial responses. In *Proceedings of the 13th international conference on Multimodal Interaction*. ACM.
- [12] D. McDuff, R. El Kaliouby, and R. Picard. Crowdsourcing facial responses to online videos. *Transactions on Affective Computing*, 2012.
- [13] L. Morency, A. Quattoni, and T. Darrell. Latent-dynamic discriminative models for continuous gesture recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. Ieee, 2007.
- [14] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.
- [15] E. Smit, L. Van Meurs, and P. Neijens. Effects of advertising likeability: A 10-year perspective. *Journal of Advertising Research*, 46(1):73, 2006.
- [16] Y. Song, D. Demirdjian, and R. Davis. Multi-signal gesture recognition using temporal smoothing hidden conditional random fields. In

- Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 388–393. IEEE, 2011.
- [17] T. Teixeira, M. Wedel, and R. Pieters. Emotion-induced engagement in internet video ads. *Journal of Marketing Research*, 2010.
- [18] T. Teixeira, M. Wedel, and R. Pieters. Moment-to-moment optimal branding in tv commercials: Preventing avoidance by pulsing. *Marketing Science*, 29(5):783–804, 2010.
- [19] M. Valstar, H. Gunes, and M. Pantic. How to distinguish posed from spontaneous smiles using geometric features. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 38–45. ACM, 2007.
- [20] S. Wang, A. Quattoni, L. Morency, D. Demirdjian, and T. Darrell. Hidden conditional random fields for gesture recognition. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1521–1527. IEEE, 2006.
- [21] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan. Toward practical smile detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(11):2106–2111, 2009.
- [22] Z. Zeng, M. Pantic, G. Roisman, and T. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):39–58, 2009.