Predicting Viewer Affective Comments Based on Image Content in Social Media

Yan-Ying Chen[*†§],Tao Chen[§],Winston H. Hsu[*],Hong-Yuan Mark Liao[†],Shih-Fu Chang[§]
[*]National Taiwan University, Taipei, Taiwan
[†]Academia Sinica, Taipei, Taiwan
[§]Columbia University, New York, USA
yanying@cmlab.csie.ntu.edu.tw, taochen@ee.columbia.edu,
winston@csie.ntu.edu.tw, liao@iis.sinica.edu.tw, sfchang@ee.columbia.edu

ABSTRACT

Visual sentiment analysis is getting increasing attention because of the rapidly growing amount of images in online social interactions and several emerging applications such as online propaganda and advertisement. Recent studies have shown promising progress in analyzing visual affect concepts intended by the media content publisher. In contrast, this paper focuses on predicting what viewer affect concepts will be triggered when the image is perceived by the viewers. For example, given an image tagged with "yummy food," the viewers are likely to comment "delicious" and "hungry," which we refer to as *viewer affect concepts (VAC)* in this paper. To the best of our knowledge, this is the first work explicitly distinguishing intended publisher affect concepts and induced viewer affect concepts associated with social visual content, and aiming at understanding their correlations. We present around 400 VACs automatically mined from million-scale real user comments associated with images in social media. Furthermore, we propose an automatic visual based approach to predict VACs by first detecting publisher affect concepts in image content and then applying statistical correlations between such publisher affect concepts and the VACs. We demonstrate major benefits of the proposed methods in several real-world tasks - recommending images to invoke certain target VACs among viewers, increasing the accuracy of predicting VACs by 20.1% and finally developing a social assistant tool that may suggest plausible, content-specific and desirable comments when users view new images.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Visual Sentiment Analysis, Viewer Affect Concept, Comment Assistant, Social Multimedia

ICMR '14, Apr 01-04 2014, Glasgow, United Kingdom ACM 978-1-4503-2782-4/14/04. http://dx.doi.org/10.1145/2578726.2578756

1. INTRODUCTION

Visual content is becoming a major medium for social interaction on the Internet, including the extremely popular platforms, Youtube, Flickr, etc. As indicated in the saying "a picture is worth one thousand words," images and videos can be used to express strong affects or emotions of users. To understand the opinions and sentiment in such online interactions, visual content based sentiment analysis in social multimedia has been proposed in recent research and has been shown to achieve promising results in predicting sentiments expressed in multimedia tweets with photo content [3, 20]. However, these studies but usually do not differentiate *publisher affect* – emotions revealed in visual content from the publishers' perspectives, and *viewer affect* – emotions evoked on the part the audience after viewing the visual content.

Different from the previous work [3, 20], we specifically target what viewer affect concepts will be evoked after the publisher affect concepts expressed in images are viewed. Taking Figure 1 (a) as an example, after viewing the visual content with "yummy food" as the publisher affect concept, the viewers are very likely to respond with a comment "hungry" (viewer affect concept). Understanding the relation between the publisher affect concepts and the evoked viewer affect concepts is very useful for developing new user-centric applications such as affect-adaptive user interfaces, target advertisement, sentiment monitoring, etc. For example, as shown in Figure 1 (f), given an image posting, we may try to predict the likely evoked emotions of the audience even when there are no textual tags assigned to the image (namely visual content based prediction). The results can also be used to develop advanced software agents to interact in the virtual world and generate plausible comments including content relevant affect concepts in response to multimedia content.

The link between image content and subjective emotions it evokes has been addressed in some research on affect [10] and affective content analysis [7]. Meanwhile, from the statistics of the image sharing website Flickr, around 0.2% user comments associated with general images comprise the word "hungry" but the percentage will surge to 14% if we only consider comments associated with images containing visual content "yummy meat." In addition, users are more likely to comment "envious" on the image showing "pretty scene" and "sad" on the image showing "terrible tragedy." The above observations clearly confirm the strong correlation between

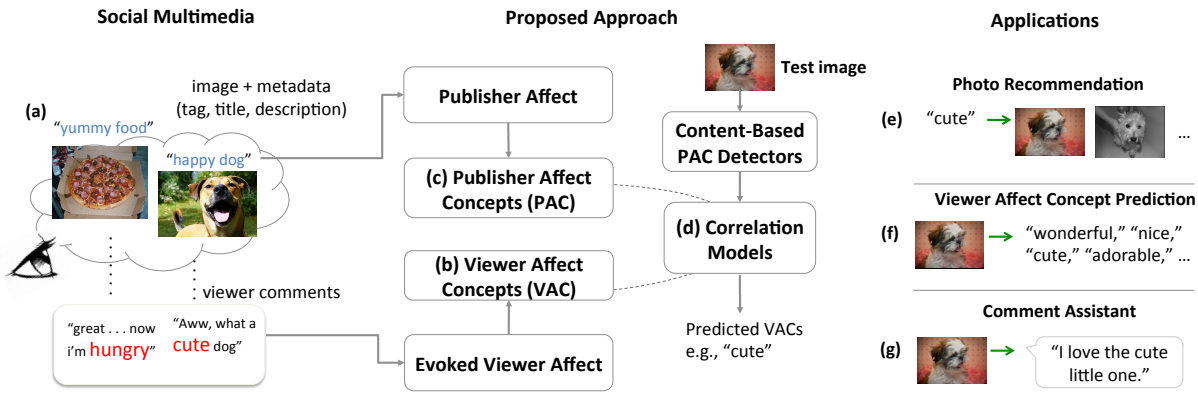

Figure 1: System overview – (a) We discover viewer affect concepts (VAC) from viewer comments associated with social multimedia (Section 3) and (b) represent each comment by VAC. (c) We extract publisher affect concepts (PAC) from publisher provided metadata; meanwhile, each PAC can be automatically detected based on analysis of visual content (Section 4.1). (d) We develop probabilistic models to capture the relations between PACs and VACs of an image. (Section 4.2). We demonstrate utility of the prediction model in three applications: (e) image recommendation (Section 5.2): recommending the optimal images to evoke a specific target emotion in the viewer, (f) viewer affect concept prediction (Section 5.3): predicting the most possible viewer affect concepts that the given image will evoke, and (g) comment assistant (Section 5.4): developing a software robot that can generate a plausible content-relevant comment for each image.

the publisher affect concepts expressed in the image and the affect concepts evoked in the viewer part.

Visual affect has not been addressed much in terms of the relationships between publisher affect and viewer affect. To the best of our knowledge, this paper presents the first work explicitly addressing publisher affect concepts and viewer affect concepts of images, and aiming at understanding their correlations. Furthermore, we propose to predict viewer affect concepts evoked by the publisher affect concepts intended in image content. Two challenges arise in this new framework; firstly, how to construct a rich vocabulary suitable for describing the affect concepts seen in the online social multimedia interaction (Figure 1 (a)). One option is to adopt the existing emotion categories [15] which have also been used for emotional image analysis [18, 12] and affective feedback analysis [1]. However, the affect concept ontology seen in online social interactions, e.g., "cute" and "dirty" in viewer comments may be different from those used in affect concepts intended by the image publishers. In this paper, we expand the basic emotions to a much more comprehensive vocabulary of concepts, called *viewer affect concepts (VAC)*. We propose to discover a large number of VACs (about 400) directly from million-scale real user comments associated with images on Flickr to represent the evoked affect concepts in viewer comments as shown in Figure 1 (b). Specifically, we focus on VACs defined as adjectives that occur frequently in viewer comments and reveal strong sentiment values.

The second challenge is how to model the correlations between publisher affect concepts and viewer affect concepts. We propose to measure such statistical correlations by mining from surrounding metadata of images (i.e., descriptions, title, tags) and their associated viewer feedback (i.e., comments). We develop a Bayes probabilistic model to estimate the conditional probabilities of seeing a VAC given the presence of publisher affect concepts in an image, as shown in Figure 1 (d). Furthermore, the mined correlations are used to predict the VACs by automatically detecting publisher affect concepts from image content (Figure 1 (c)) without needing the metadata tags of an image.

To demonstrate the effectiveness of the proposed approach, we design several interesting applications – recommend best images for each target VAC (Figure 1 (e)), and predict the VACs given a new image (Figure 1 (f)). In addition, we show how VACs may lead to designs of novel agent software that is able to select high quality comments for virtual social interaction (Figure 1 (g)). The results also suggest the potential of using VAC modeling in influencing audience opinions; for example, the automatically selected comments, when perceived as plausible and relevant, may help elicit more favorable responses from the targeted audiences.

The novel contributions of this paper include,

- hundreds of VACs automatically discovered from millions of comments associated with images of strong affective values.
- a novel affect concepts analysis model that explicitly separates the publisher and viewer affect concepts and characterize their probabilistic correlations.
- a higher than 20% accuracy gain in content-based viewer affect concept prediction compared to the baseline by using publisher affect concepts only.
- novel applications enabled by the proposed affect concept correlation model including image recommendation for targeted affect concepts and social agent software with the automated commenting ability.

2. RELATED WORK

Making machine behave like human – not only at the perception level but also the affective level – is of great interest to researchers. Similar motivations have driven recent research in high-level analysis of visual aesthetics [5], interestingness [9] and emotion [12, 7, 18, 16]. These studies

attempted to map low level visual features to high-level affect classes. Despite the promising results, the direct mapping from low level features is quite limited due to the well-known semantic gap and the emotional gap as discussed in [18]. Facing such challenges, recently a new approach advocates the use of mid-level representations, built upon Visual Sentiment Ontology and SentiBank classifiers [3]. It discovers about 3,000 visual concepts related to 8 primary emotions defined at multiple levels in [15]. Each visual sentiment concept is defined as an adjective-noun pair (e.g., "beautiful flower," "cute dog"), which is specifically chosen to combine the detectability of the noun and the strong sentiment value conveyed in adjectives. The notion of mid-level representation was also studied in [20], in which attributes (e.g., metal, rusty) were detected in order to detect high-level affect classes.

However, the aforementioned work on visual sentiment analysis only focuses on the affect concepts expressed by the content publishers, rather than the evoked emotions in the viewer part. For example, a publisher affect concept "yummy food" expressed in the image often triggers VACs like "hungry" and "jealous." Analysis of review comments has been addressed in a broad spectrum of research, including mining opinion features in customer reviews [8], predicting comment ratings [17] and summarizing movie reviews [21]. Most of these studies focus the structures, topics and personalization factors in the viewer comments without analyzing the content of the media being shared. In this paper, we advocate that viewer responses are strongly correlated with the content stimuli themselves, especially for the visual content shared in social media. Thus, a robust VAC prediction system will need to take into account the publisher affect concepts being revealed in the visual content. Analogous to the large concept ontology constructed for the visual sentiment in [3], we believe a large affect concept pool can be mined from the viewer comments. Such viewer affect concepts offer an excellent mid-level abstraction of the viewer emotions and can be used as a suitable platform for mining the correlations between publisher and viewer affects (e.g., "yummy" evokes "hungry," "disastrous" evokes "sad").

In the remainder of this paper, we will discuss viewer affect concept discovery in Section 3 and further introduce the publisher-viewer affect concept correlation model in Section 4. The experiments for three applications, image recommendation, viewer affect concept prediction and automatic commenting assistant, will be shown in Section 5, with conclusions in Section 6.

3. VIEWER AFFECT CONCEPT DISCOVERY

This section presents how and what VACs are mined from viewer comments. We introduce the strategy for crawling observation data, then a post-processing pipeline for cleaning noisy comments and finally the criteria for selecting VACs.

Online user comments represent an excellent resource for mining viewer affect concepts. It offers several advantages: (1) the comments are unfiltered and thus preserving the authentic views, (2) there are often a large volume of comments available for major social media, and (3) the comments are continuously updated and thus useful for investigating trending opinions. Since we are primarily interested in affects related to visual content, we adopt the semi-professional social media platform Flickr to collect the comment data. To ensure we can get data of rich affects, we first search Flickr with 24 keywords (8 primary dimensions plus 3 varying strengths) defined in Plutchik's emotion wheel defined in psychology theories [15]. Search results include images from Flickr that contain metadata (tags, titles, or descriptions) matching the emotion keywords. We then crawl the comments associated with these emotional images as the observation data. The number of comments for each emotion keyword is reported in Table 1, totally around 2 million comments associated with 140,614 images. To balance the impact from each emotion on the mining results, we sample 14,000 comments from each emotion, resulted in 336,000 comments for mining VACs.

Table 1: Our Flickr training corpus for mining viewer affect concepts comprises 2 million comments associated 140,614 images, which are collected by searching Flickr with the 24 emotions defined in psychology.

emotion keywords (# comments)
ecstasy (30,809), joy (97,467), serenity (123,533)
admiration (53,502), trust (78,435), acceptance (97,987)
terror (44,518), fear (103,998), apprehension (14,389)
amazement (153,365), surprise (131,032), distraction (134,154)
grief (73,746), sadness (222,990), pensiveness (25,379)
loathing (35,860), disgust (83,847), boredom (106,120)
rage (64,128), anger (69,077), annoyance (106,254)
vigilance (60,064), anticipation (105,653), interest (222,990)

Table 2: The example VACs of positive and negative sentiment mined from viewer comments.

sentiment polarity	viewer affect concepts (VACs)
positive	beautiful, wonderful, nice, lovely, awesome, amazing, fantastic, cute, excellent, interesting delicious, lucky, attractive, happy, adorable
negative	sad, bad, sorry, scary, dark, angry, creepy, difficult, poor, sick stupid, dangerous, freaky, ugly, disturbing

The crawled photo comments usually contain rich but noisy text with a small portion of subjective terms. According to the prior study of text subjectivity [19, 4], adjectives usually reveal higher subjectivity which are informative indicators about user opinions and emotions. Following this finding, we apply part-of-speech tagging [2] to extract adjectives. To avoid the confusing sentiment orientation, we exclude the adjectives within a certain neighborhood of negation terms like "not" and "no." Additionally, to reduce the influence by spams, we also remove the hyperlinks and HTML tags contained in the comments.

We focus on sentimental and popular terms which are often used to indicate viewer affective responses. Per the first criterion, we measure the sentiment value of each adjective by SentiWordNet [6]. The sentiment value ranges from -1 (negative sentiment) to $+1$ (positive sentiment). We take the absolute value to represent the sentiment strength of a given adjective. To this end, we only keep the adjectives with high sentiment strength (at least 0.125) and high occurrence frequency (at least 20 occurrences). Totally 400 adjectives are selected as viewer affect concepts (VACs). Table 2 presents the example VACs of positive and negative sentiment polarities, respectively.

4. PUBLISHER-VIEWER AFFECT CORRELATION

Given an image, we propose to predict the evoked VACs by (1) detecting publisher affect concepts (PACs) in the image content and (2) utilizing the mined co-occurrences between PACs and VACs. This process considers the PACs as the stimuli and aims at exploring the relationships between the stimuli and evoked VACs.

4.1 Publisher Affect Concepts

We adopt 1,200 sentiment concepts defined in SentiBank [3] as the PACs in image content (Figure 1 (c)). As mentioned earlier, these concepts are explicitly selected based on the typical emotion categories and data mining from images in social media. Each concept combines a sentimental adjective concept and a more detectable noun concept, e.g., "beautiful flower," "stormy clouds." The advantage of adjective-noun pairs is its capability to turn a neutral noun like "dog" into a concept with strong sentiment like "dangerous dog" and make the concept more visually detectable, compared to adjectives only.

The concept ontology spreads over 24 different emotions [15] which capture diverse publisher affects to represent the affect content. SentiBank includes 1200 PACs learned by low-level visual features (color, texture, local interest points, geometric patterns), object detection features (face, car, etc.), and aesthetics-driven features (composition, color smoothness, etc.). According to the experiment results in [3], all of the 1,200 ANP detectors have F-score greater than 0.6 over a controlled testset.

As shown in Figure 1 (c), given an image d_i, we apply SentiBank detectors to estimate the probability of the presence of each publisher affect concept p_k, denoted as $P(p_k|d_i)$. Such detected scores will be used to perform automatic prediction of affect concepts to be described in details later.

Another version of the PAC data use the "ground truth" labels found in the image metadata for the 1,200 PACs. In other words, we detect the presence of each PAC in the title, tags, or description of each image. Such ground truth PAC data will be used in the next section to mine the correlation between PACs and VACs. One potential issue with using such metadata is the false miss error - images without explicit labels of a PAC may still contain content of the PAC. We will address this issue by a smoothing mechanism discussed in Section 4.3.

4.2 Bayes Probabilistic Correlation Model

We apply Bayes probabilistic models and the co-occurrence statistics over a training corpus from Flickr to estimate the correlations between PACs and VACs. Specially, we used the 3 million comments associated with 0.3 million images containing rich PAC keywords crawled from Flickr [1] as the training data. Given a VAC v_j, we compute its occurrences in the training data and its co-occurrences with each PAC p_k over the training data θ. The conditional probability $P(p_k|v_j)$ can then be determined by,

$$P(p_k|v_j;\theta) = \frac{\sum_{i=1}^{|D|} B_{ik} P(v_j|d_i)}{\sum_{i=1}^{|D|} P(v_j|d_i)}, \quad (1)$$

where B_{ik} is a variable indicating the presence/absence of p_k in the publisher provided metadata of image d_i and $|D|$ is the number of images. $P(v_j|d_i)$ is measured by the occurrence counting of v_j in comments associated with images. Given the correlations $P(p_k|v_j;\theta)$, we can measure the likelihood of a given image d_i and a given VAC v_j by multivariate Bernoulli formulation [13].

$$P(d_i|v_j;\theta) = \prod_{k=1}^{|A|} (P(p_k|d_i)P(p_k|v_j;\theta) + (1 - P(p_k|d_i))(1 - P(p_k|v_j;\theta))). \quad (2)$$

A is the set of PACs in SentiBank. $P(p_k|d_i)$ can be measured by using the scores of SentiBank detectors (cf. Section 4.1), which approximate the probability of PAC p_k appearing in image d_i. Here, PACs act as shared attributes between images and VACs, resembling the probabilistic model [13] for content-based recommendation [14].

Based on the above probabilistic model, we can answer the question – what is the possibility that an image will evoke a specific VAC. This is very useful for the application of target advertisement applications - selecting the most possible images that will stimulate the given VAC.

Conversely, we can measure the posterior probability of VACs given a test image d_i by Bayes' rule,

$$P(v_j|d_i;\theta) = \frac{P(v_j|\theta)P(d_i|v_j;\theta)}{P(d_i|\theta)}. \quad (3)$$

$P(v_j|\theta)$ can be determined by the frequency of VAC v_j appearing in the training data and $P(d_i|\theta)$ is assumed equal over images. The above equation is useful for another interesting application – given an image, we can predict the most possible VACs by the posterior probability in Eq. 3. We will demonstrate the performance of these two applications in Section 5.2 and 5.3, respectively.

4.3 Smoothing

In this subsection, we address the issue of the missing associations – unobserved correlations between PACs and VACs. For example, a PAC "muddy dog" will likely trigger the VAC "dirty," but there are no viewer comments comprising this VAC in our data. To deal with such unobserved associations, we propose to add a smoothing factor in the probabilistic model.

Intuitively, some publisher affect concepts share similar semantic or sentimental meaning; for example, "muddy dog" and "dirty dog." More examples can be found in the 1200 publisher affect concepts in SentiBank [3], e.g., "weird cloud" and "strange cloud," "delicious food" and "delicious meat." To this end, we propose to apply collaborative filtering techniques to fill the potential missing associations. The idea is to use matrix factorization to discover the latent factors of the conditional probability ($P(p_k|v_j)$ defined in Eq. 1) and use the optimal factor vectors t_j, s_k for smoothing missing associations between PAC p_k and VAC v_j. The matrix factorization formulation can be expressed as follows,

$$\min_{t,s} \sum_{k,j} (P(p_k|v_j) - {t_j}^T s_k)^2, \quad (4)$$

[1]The training corpus [3] containing the Flickr images and their metadata are downloaded from `http://www.ee.columbia.edu/ln/dvmm/vso/download/flickr_dataset.html`

Note that, we specifically use non-negative matrix factorization [11] to guarantee the smoothed associations are all non-negatives which can fit the calculation in the probabilistic model. The approximated associations between PAC p_k and VAC v_j can then be smoothed as follows,

$$\hat{P}(p_k|v_j) = {t_j}^T s_k. \quad (5)$$

With the smoothed correlations $\hat{P}(p_k|v_j)$, given a VAC v_j, the likelihood with an image d_i is reformulated as,

$$P(d_i|v_j;\theta) = \prod_{k=1}^{|A|}(P(p_k|d_i)\hat{P}(p_k|v_j) \quad (6)$$
$$+(1-P(p_k|d_i))(1-\hat{P}(p_k|v_j))).$$

To avoid floating-point underflow when calculating products of probabilities, all of the computations are conducted in the log-space.

5. APPLICATIONS AND EXPERIMENTS

5.1 Dataset for Mining and Evaluation

This section introduces the dataset for mining PAC-VAC correlations and the additional dataset for evaluation. All the images, publisher provided metadata and comments are crawled from Flickr.

(a) **Dataset for mining correlations between PAC and VAC** comprises comments associated with the images (along with descriptions, tags and titles) of 1200 publisher affect concepts publicly released by SentiBank [3]. Totally, around 3 million comments associated with 0.3 million images are collected as the training data. On the average, an images is commented by 11 comments, and a comment comprises 15.4 words. All the comments are further represented by 400 VACs for mining PAC-VAC correlations. Table 3 reports the example mined PAC-VAC correlations ranked by $P(p_k|v_j)$ (cf. Eq. 1) and filtered by statistical significance value (p-value). PAC and the evoked VACs may be related but not exactly the same, e.g., "hilarious" for "crazy cat," "delicate" for "pretty flower" and "hungry" for "sweet cake." In some cases, their sentiment are even extremely different, e.g., "cute" for "weird dog" and "scary" for "happy halloween." Because PAC may evoke varied VACs, further considering PAC-VAC correlations will benefit understanding viewer affect concepts. We will demonstrate how PAC-VAC correlations benefit viewer-centric applications in the following sections.

(b) **Test image dataset** contains 11,344 images from the public dataset [3] to conduct the experiments for the proposed three applications, image recommendation by viewer concepts (Section 5.2), viewer affect concept prediction (Section 5.3), and automatic commenting by viewer affect concepts (Section 5.4). Note that, the images from the databases (a) and (b) are not overlapped.

5.2 Image Recommendation for Target Affect Concepts

The first application is to recommend the images which are most likely to evoke a target VAC. Given a VAC v_j, the recommendation is conducted by ranking images over the likelihood $P(d_i|v_j)$ measured by Eq. 6. For each VAC, 10 positive images and 20 negative images are randomly selected from the test database (cf. Section 5.1 (b)) for evaluation. The ground truth of VAC for each image is determined by whether the VAC can be found in the comments associated with this image. For example, if the VACs "nice," "cute" and "poor" are found in the comments of an image, then this image will be a positive sample for "nice," "cute" and "poor" VAC image recommendation. The performance is evaluated by average precision (AP) over 400 mined VACs.

Table 3: The significant VACs for example PACs ranked by PAC-VAC correlations. Because PAC may evoke different VACs, further considering PAC-VAC correlations will benefit understanding VACs.

PAC	#1 VAC	#2 VAC	#3 VAC
tiny dog	cute	adorable	little
weird dog	weird	funny	cute
crazy cat	hysterical	crazy	hilarious
cloudy morning	ominous	serene	dramatic
dark woods	mysterious	spooky	moody
powerful waves	dynamic	powerful	sensational
wild water	dangerous	dynamic	wild
terrible accident	terrible	tragic	awful
broken wings	fragile	poignant	poor
bright autumn	bright	delightful	lovely
creepy shadow	creepy	spooky	dark
happy halloween	spooky	festive	scary
pretty flowers	delicate	joyful	lush
fresh leaves	fresh	green	vibrant
wild horse	wild	majestic	healthy
silly girls	sick	funny	cute
mad face	mad	funny	cute
beautiful eyes	expressive	intimate	confident
sweet cake	yummy	hungry	delicious
nutritious food	healthy	yummy	delicious
shiny dress	shiny	sexy	gorgeous
colorful building	colourful	vivid	vibrant
haunted castle	spooky	mysterious	scary

Table 4: Performance of image recommendation for target VACs. Mean Average Precision (MAP) values of the top 100, 200, 300, and entire set of VACs.

top VACs	100	200	300	overall
MAP	0.5321	0.4713	0.4284	0.3811

As shown in Table 4, the mean value of the average precision of the 100 most predictable VAC is around 0.5321. Mean AP exceeds 0.42 in the best 300 VACs and decreases to 0.3811 over the entire set of 400 VACs. Figure 2 shows the top five recommended images of 10 sampled VACs sorted by average precision from top to bottom. We found that the most predictable VACs are usually of higher visual content and semantic consistency. For example, top recommended images for "splendid" affect concept are correlated with beautiful scenic views (e.g., rank #1, #2, #3 in Figure 2) while the "festive" images usually display warm color tones. That suggests the viewers usually have common evoked affect concepts for these types of visual content. Moreover, our approach can recommend images containing more diverse semantics in visual content (e.g., "freaky" and "creepy"), because it aims to learn PAC-VAC correlations from a large pool of image content with rich comments (millions).

As discussed in Section 5.1, the comments associated with images are naturally sparse (averagely 11 comments for each image and 15.4 words per comment in our training data) and leads to many missing associations. For example, the top 1

Figure 2: Examples of recommended images for each target view affect concept. The images are ranked by likelihood (Eq. 6) from left to right and the sampled VACs are sorted by average precision shown in parentheses. The most predictable VACs usually have consistent visual content or semantics. For example, the "splendid" images are correlated with scenic views (e.g., rank #1, #2, #3). Conversely, the VACs with less agreement among viewers (e.g., "unusual" and "unique") are less predictable by the proposed approach. Note faces in the images are masked.

Table 5: The performance of viewer affect concept prediction given a new image. The overlap ratio by using our approach (Corr) surpasses the baseline (PAC-only) with 20.1% improvement. Moreover, our approach obtains superior hit rate and the hit rate of the top 3 predicted VACs. That suggests higher consistency of the predicted VAC and the ground truth VACs.

method	PAC-only [3]	Corr
overlap	0.2295	0.4306 (+20.1%)
hit rate	0.4333	0.6231 (+19.0%)
hit rate (3)	0.3106	0.5395 (+22.9%)

and 2 recommended images for "delightful" actually comprise smile, which likely evokes "delightful" affect concept. But because this term was never used in the comments of the images, it was treated as incorrect prediction even though the results should be right upon manual inspection. In general, the VACs without clear concensus among viewers (e.g., "unusual" and "unique") usually are less predictable by the proposed approach.

5.3 Evoked Viewer Affect Concept Prediction

The second application, viewer affect concept prediction, is opposite to the aforementioned image recommendation. Given an image d_i, we aim at predicting the most possible VACs stimulated by this image. We measure the posterior probability of each VAC v_j by the probabilistic model in Eq. 3. The higher posterior probability means the more likely that the VAC v_j will be evoked by the given image d_i. In addition, we compare our method (Corr) with the baseline using PACs [3] only. Given a test image, the baseline method (PAC-only) chooses all the VACs appearing in the comments associated with the training images which comprises the PACs with the highest detection scores in the test image. In contrast, our method (Corr) considers the soft detection scores of all PACs and use the PAC-VAC correlations described in Eq. 3 to rank VACs based on $P(v_j|d_i;\theta)$. The predicted VACs are the VACs with probabilities higher than a threshold. For fair comparisons without being affected by sensitivity of threshold setting, the threshold is set to include the same number of VACs predicted by the baseline method.

The test images are selected from database (b) described in Section 5.1 and each test image has comments comprising at least one VAC. Totally 2,571 test images are evaluated by the two performance metrics, overlap ratio and hit rate. Overlap ratio indicates how many predicted VACs are covered by the ground truth VACs, normalized by the union of predicted VACs and ground truth VACs.

$$overlap = \frac{|\{groundtruthVACs\} \cap \{predictedVACs\}|}{|\{groundtruthVACs\} \cup \{predictedVACs\}|}. \quad (7)$$

As shown in Table 5, the overlap of our approach (Corr) outperforms the baseline approach by 20.1%. The higher overlap indicates higher consistency between the predicted VACs and the ground truth VACs given by real users.

Considering the sparsity in comments, the false positives in the predicted VACs may be simply missing but actually correct. To address such missing label issue, we further evaluate hit rate, that is, the percentage of the test images that have at least one predicted VAC hitting the ground truth VACs. Hit rate is similar to overlap ratio but deemphasizes the penalty of false positives in the predicted VACs. As shown in Table 7, our approach achieves 19.0% improvement in overall hit rate compared to the baseline. The gain is even higher (22.9%) if the hit rate is computed only for the top 3 predicted VACs (hit rate (3)). Some example prediction results are shown in Figure 3 (e.g., "gorgeous," "beautiful" for image (a) and "lovely," "moody," "peaceful" for image (b)). In the next section, we will introduce how to exploit the predicted VACs in generating comments for images, for which subjective evaluation will be used instead

of the aforementioned overlap and hit ratios.

5.4 Automatic Commenting Assistant

We propose a novel application – given an image, automatically recommend comments containing the most likely VACs predicted based on image content. Automatic commenting is an emerging function in social media [2], aiming at generating comments for a given post, e.g., tweets or blogs, by observing the topics and opinions appearing in the content. However, commenting image has never been addressed because of the difficulty in understanding visual semantics and visual affects. Intuitively, commenting behavior is strongly influenced by viewer affect concepts. This motivates us to study automatically commenting images by the proposed viewer affect concept prediction.

The proposed method (Corr) considers the PACs detected from the visual content and the PAC-VAC correlations captured by the Bayesian probabilistic model described in Section 4.2. First, we detect the PACs in the test image and construct a candidate comment pool by extracting comments of images in the training set that contain similar PACs (the top 3 detected PACs with the highest $P(p_k|d_i)$) in the visual content. Each comment is represented by bag-of-viewer-affect-concepts as a vector C_l, indicating the presence of each VAC in that comment. Meanwhile, the test image is represented by a vector V_i consisting of the posterior probability $P(v_j|d_i)$ (cf. Eq. 3) of each VAC given the test image, d_i. The relevance between a comment and the test image is measured by their inner product $s_{li} = C_l \cdot V_i$. Finally, we select the comment with the highest relevance score s_{li} from the candidate comment pool for automatic commenting. Note that, the images, which are used to extract comments in the candidate pool, do not overlap with the test image set. We compare our method with the two baselines (1) PAC-only: selecting one of the comments associated with another image having the most similar PAC to that of the test image and (2) Random: randomly selecting a comment from the comments of training images.

We conduct user study to evaluate the automatic commenting quality in terms of (1) plausibility, (2) specificity to the image content and (3) whether it is liked by users. Totally, 30 users are involved in this experiment. Each automatic comment is evaluated by three different users to avoid potential user bias. Each user is asked to evaluate 40 automatic comment, each is generated for a test image. The users are asked to rate the comment in three different dimensions (score from 1 to 3 in each dimension), **Plausibility**: how plausible the comment given the specific image content; **Specificity**: how specific the comment is to the image content; **Like**: how much does the user like the comment. Totally, 400 image-comment pairs are included in this investigation.

As shown in Figure 4, the most gain appears in plausibility where our method significantly outperforms the other two baselines (PAC-only) and (Random) by 35% and 56% (relative improvement), respectively. Additionally, the proposed approach also clearly improves specificity of the generated comments to the visual content in the image. For example, comments containing the affect concept "cute" are selected by our methods for images containing "dog," "kid." Our method (Corr) produces comments that are more liked

[2]More details regarding commenting bot is introduced in `http://en.wikipedia.org/wiki/Twitterbot`

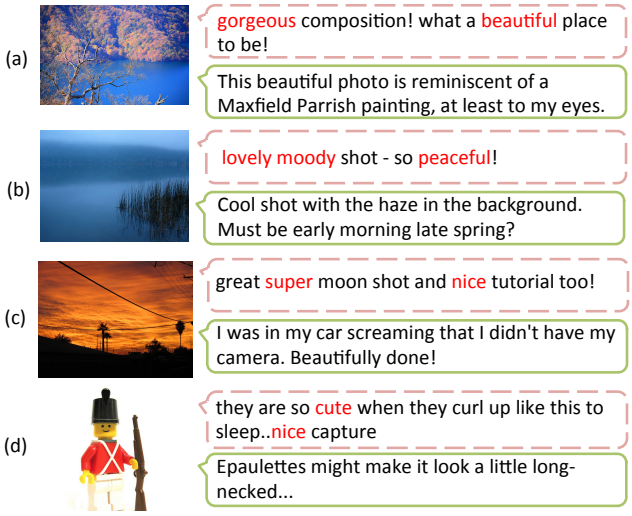

Figure 3: Example results of VAC prediction and automatic comment selection. The red words are the VACs predicted by our model given the test image on the left. Our method also automatically selects a comment from a large pool that contains the predicted viewer affect concepts and most relevant to the image content. In (a) and (b), the automatically selected comments (dashed) looks plausible and specific to the given image even compared to the original comments (solid). However, if automatic comments mention incorrect objects (e.g., "moon" in (c)) or actions ("sleep" in (d)), users can easily tell the faked ones from the originals.

by users. The potential reasons are, (1) our methods tend to include viewer affect concepts that comprise more emotional words and thus evoke stronger responses from the subjects; (2) our method uses the correlation model that tries to learn the popular commenting behavior discovered from real comments in social multimedia, as described in Section 4.2. Overall, commenting by our method has the quality closest to original real comment. Figure 3 (a) and (b) shows a few plausible and content relevant fake comments (dashed) automatically generated by the proposed commenting robot. One additional finding is if selected comments mention incorrect objects ("moon" in (c)) or actions ("sleep" in (d)) in the given image, users can easily distinguish them from the real ones. This points out interesting future refinement by incorporating object detection in the automatic commenting process.

In another evaluation scheme, we focus on plausibility of the faked comments. Each test includes an image, one original comment and the fake comments selected by the proposed method and the baseline (Random). User is asked to decide which one of the four comments is most plausible given the specific image. Comments generated by content-aware method can confuse the users in 28% of times, while the real comment was considered to be most plausible in 61% of times. This is quite encouraging given the fact that our method is completely content-based, namely the prediction is purely based on analysis of the image content and the affect concept correlation model. No textual metadata of the image was used. It is also interesting that 11% of randomly selected comments are judged to be more plausi-

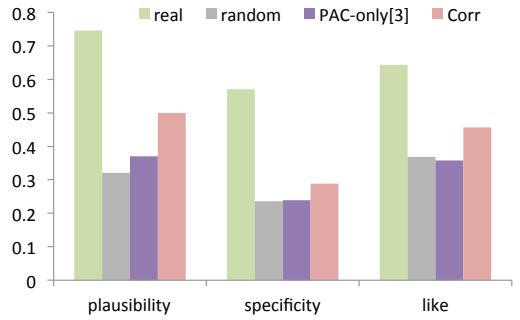

Figure 4: Subjective quality evaluation of automatic commenting for image content. The proposed approach (Corr) shows superior quality (plausibility, specificity and like) compared to the baselines (PAC-only) and (Random). The most gain appears in plausibility, outperforming (PAC-only) and (Random) by 35% and 56% (relative gain), respectively.

ble than the original real comment. However, as discussed earlier, such random comments tend to have poor quality in terms of content specificity.

6. CONCLUSIONS

In this paper, we study visual affect concepts in the two explicit aspects, publisher affect concepts and viewer affect concepts, and aim at analyzing their correlations – what viewer affect concepts will be evoked when a specific publisher affect concept is expressed in the image content. For this purpose, we propose to discover hundreds of viewer affect concepts from a million-scale comment sets crawled from social multimedia. Furthermore, we predict the viewer affect concepts by detecting the publisher affect concepts in image content and the probabilistic correlations between such affect concepts and viewer affect concepts mined from social multimedia. Extensive experiments confirm exciting utilities of our proposed methods in the three applications, image recommendation, viewer affect concept prediction and image commenting robot. Future directions include incorporation of the viewer profiles in predicting the likely response affects, and extension of the methods to other domains.

7. ACKNOWLEDGMENTS

Research was sponsored in part by the U.S. Defense Advanced Research Projects Agency (DARPA) under the Social Media in Strategic Communication (SMISC) program, Agreement Number W911NF-12-C-0028. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Defense Advanced Research Projects Agency or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

The work is supported by NSC Study Abroad Program grants 102-2917-I-002-078.

8. REFERENCES

[1] I. Arapakis, J. M. Jose, and P. D. Gray. Affective feedback: An investigation into the role of emotions in the information seeking process. In *SIGIR*, 2008.

[2] Bird, Steven, E. Loper, and E. Klein. Natural language processing with python. 2009.

[3] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *ACM Multimedia*, 2013.

[4] R. F. Bruce and J. M. Wiebe. Recognizing subjectivity: A case study of manual tagging. *Natural Language Engineering*, 1999.

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *ECCV*, 2006.

[6] A. Esuli and F. Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *LREC*, 2006.

[7] A. Hanjalic. Extracting moods from pictures and sounds: Towards truly personalized tv. *IEEE Signal Processing Magazine*, 2006.

[8] M. Hu and B. Liu. Mining opinion features in customer reviews. In *AAAI*, 2004.

[9] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *CVPR*, 2011.

[10] P. Lang, M. Bradley, and B. Cuthbert. International affective picture system (iaps): Affective ratings of pictures and instruction manual. *Technical Report A-8. University of Florida, Gainesville, FL*, 2008.

[11] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, 2001.

[12] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *ACM Multimedia*, 2010.

[13] A. McCallum and K. Nigam. A comparison of event models for naive bayes text classification. In *AAAI Workshop on Learning for Text Categorization*, 1998.

[14] M. J. Pazzani and D. Billsus. Content-based recommendation systems. In *The Adaptive Web: Methods and Strategies of Web Personalization. Volume 4321 of Lecture Notes in Computer Science*, 2007.

[15] R. Plutchik. Emotion: A psychoevolutionary synthesis. *Harper & Row, Publishers*, 1980.

[16] N. Sebe, I. Cohen, T. Gevers, and T. S. Huang. Emotion recognition based on joint visual and audio cues. In *ICPR*, 2006.

[17] S. Siersdorfer, S. Chelaru, W. Nejdl, and J. San Pedro. How useful are your comments?: Analyzing and predicting youtube comments and comment ratings. In *WWW*, 2010.

[18] W. Wang and Q. He. A survey on emotional semantic image retrieval. In *ICIP*, 2008.

[19] J. M. Wiebe, R. F. Bruce, and T. P. O'Hara. Development and use of a gold-standard data set for subjectivity classifications. In *ACL*, 1999.

[20] J. Yuan, Q. You, S. McDonough, and J. Luo. Sentribute: Image sentiment analysis from a mid-level perspetive. In *Workshop on Sentiment Discovery and Opinion Mining*, 2013.

[21] L. Zhuang, F. Jing, and X.-Y. Zhu. Movie review mining and summarization. In *CIKM*, 2006.