How Do Your Friends on Social Media Disclose Your Emotions?

Yang Yang, Jia Jia, Shumei Zhang, Boya Wu, Qicong Chen, Juanzi Li, Chunxiao Xing, Jie Tang
Department of Computer Science and Technology, Tsinghua University
Tsinghua National Laboratory for Information Science and Technology (TNList)
yyang.thu@gmail.com, {jjia,lijuanzi, xingcx, jietang}@tsinghua.edu.cn

Abstract

Extracting emotions from images has attracted much interest, in particular with the rapid development of social networks. The emotional impact is very important for understanding the intrinsic meanings of images. Despite many studies having been done, most existing methods focus on image content, but ignore the emotion of the user who published the image. One interesting question is: How does social effect correlate with the emotion expressed in an image? Specifically, can we leverage friends interactions (e.g., discussions) related to an image to help extract the emotions? In this paper, we formally formalize the problem and propose a novel emotion learning method by jointly modeling images posted by social users and comments added by their friends. One advantage of the model is that it can distinguish those comments that are closely related to the emotion expression for an image from the other irrelevant ones. Experiments on an open Flickr dataset show that the proposed model can significantly improve (+37.4% by F1) the accuracy for inferring user emotions. More interestingly, we found that half of the improvements are due to interactions between 1.0% of the closest friends.

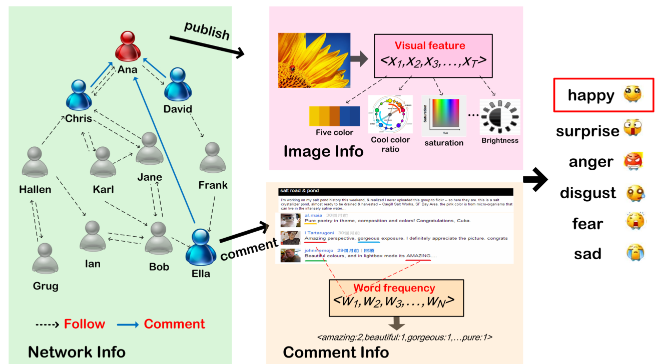


Figure 1: The general idea of the proposed emotion learning method. In the left side of the figure, "Ana" (red colored) publishes an image, and three users (blue colored) leaves comments. We extract the visual features (e.g., five color theme) from the image and emotional words (e.g., "amazing", "gorgeous") appeared in comments. Our goal is to automatically extract emotions from the images by leveraging all the related information (visual features, comments, and friendships).

Introduction

Image is a natural way to express one's emotions. For example, people use colorful images to express their happiness, while gloomy images are used to express sadness. With the rapid development of online social networks, e.g., Flickr [1] and instagram [2], more and more people like to share their daily emotional experiences using these platforms. Our preliminary statistics indicate that more than 38% of the images on Flickr are explicitly annotated with either positive or negative emotions. Understanding the emotional impact of social images can benefit many applications, such as image retrieval and personalized recommendation.

Besides sharing images, in online social networks such as Flickr and Instagram, posting discussions on a shared image is becoming common. For example, on Flickr, when a user publishes an image, on average 7.5 friends will leave comments (when users follow each other on Flickr, we say they are friends). Will such interaction among friends help us extract the hidden emotions from social images? Related studies can be traced back to psychology. Rimé (2005) showed that $88 - 96\%$ of people's emotional experiences are shared and discussed to some extent. Christopher and Rimé (1997) also showed that emotion sharing usually (85%) occurs between close confidants (e.g., family members, close friends, parents, etc.). However, due to the lack of available data, they only studied the problem by interviewing people on a very small scale. Meanwhile, recent research on inferring emotions from social images mainly considers image content, such as color distribution, contrast and saturation. For example, (Shin and Kim 2010) uses the image features, especially color features, to classify photographic images. Ou et al. (2004) explore the affective information for single color and two-color combinations.

In this paper, we aim to study the problem of inferring emotions of images from a new perspective. In particular, when you post an image, how does your friends' discussion (e.g., comments) reveal your emotions? There are several challenges in this problem. First, how to model the image

[1]http://flickr.com, the largest photo sharing website.

[2]http://instagr.am, a newly launched free photo sharing website.

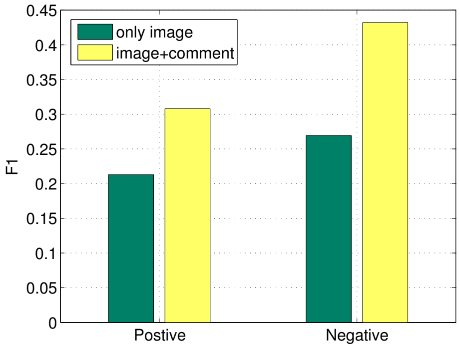


Figure 2: The performance on inferring positive and negative emotions. Two methods are shown here: one only considers image information, and another further considers comment information.

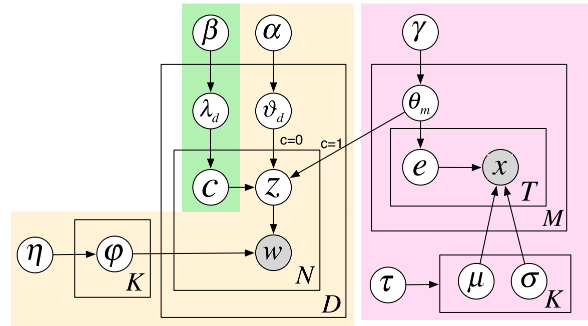


Figure 3: Graphical representation of the proposed model. The purple block can be regarded as a mixture Gaussian, which describes the visual features of images. The yellow block can be seen as a LDA, which describes the comment information. The green block models how likely a comment will be influenced by the relevant image, which combines images and comments together.

information and comment information jointly? Second, different comments reflect the publisher's emotion in different extent. For example, when a user shares an image filled with sadness, most strangers will only comment on the photography skill, while her friends will make comments that comfort the user. How to construct a computational model to learn the association among the implied emotions of different comments? Third, how to validate the proposed model in real online social networks?

To address the above challenges, we propose a novel emotion learning model to integrate both the image content (visual features) and the corresponding comments. Figure 1 clearly demonstrates the framework of the proposed method. More specifically, the proposed model regards the visual features extracted from images as a mixture of Gaussian, and treats the corpus of comments as a mixture of topic models (e.g., LDA (Blei, Ng, and Jordan 2003)). It integrates the two major parts by a cross-sampling process, which will be introduced in detail in Our Approach section. The advantage of the proposed model is that it not only extracts the latent emotions an image implies, but also distinguishes comments from others who really caring about the user.

We further test the proposed model on a real Flickr dataset, which consists of 354,192 images randomly downloaded. Figure 2 shows some interesting experimental results. 1) In the case that only 1% friends give emotional comments, compared with the methods only using image content, our method improves $+44.6\%$ on inferring positive emotions and $+60.4\%$ on inferring negative ones in terms of F1; 2) Positive emotions attract more response compared with negative ones. More detailed results can be found in Experimental Results section.

Emotion Learning Method

Formulation. We are given a set of images M. For each image $m \in M$, we have the user v_m who posts m, and a set of comments D_m which are posted about m. Also, for each comment $d \in D_m$, we know the user v_d who posts d. Our goal is to determine the emotional status of user v_m when she posted the image m.

More precisely, we use a T dimensional vector $\mathbf{x}_m =< x_{m1}, \cdots, x_{mT} >$ ($\forall_t x_{mt} \in \mathbb{R}$) to represent the image m, where each dimension indicates one of m's visual features (e.g., saturation, cool color ratio, etc.). Each comment d is regarded as a N_d-sized bag of words $\mathbf{w}_d$, where each word is chosen from a vocabulary of size W. For users' emotional status, in this work, we mainly consider Ekman's six emotions: $\{$*happiness, surprise, anger, disgust, fear, sadness*$\}$.

The users who has posted either an image or a comment are grouped as a user set V. All comments are denoted as a set D. We incorporate images, comments, and social network information in a single heterogeneous social network.

Definition 1. An **heterogeneous social network** is a directed graph $G =< V, M, D, R >$. The edge set R is the union of four sets: user-image edges $\{(v, m)|v \in V, m \in M\}$, indicating that v posts m; user-comment edges $\{(v, d)|v \in V, d \in D\}$, indicating that v posts d; image-comment edges $\{(m, d)|m \in M, d \in D\}$, indicating that d is posted about m; and user-user edges $\{(u, v)|u \in V, v \in V\}$, indicating that u follows v.

With our formulation, a straightforward baseline here is to employ a standard machine technology (e.g., SVM) for learning and inference users' emotions, by regarding $\mathbf{x}_m$ and $\mathbf{w}_d$ as input features directly. However, this method lacks of a joint representation of image and comment information. Also, it may easily cause over-fitting problem as $\mathbf{w}_d$ contains much noise (irrelevant words) and the vocabulary size W is huge in practice. To address these problems, we propose an emotion learning method, which bridges the image and comment information by utilizing a latent space.

Overview. Generally, the proposed model consists of three parts: (1) similar with (Elguebaly and Bouguila 2011), it describes visual features of images by a mixture Gaussian, which is shown as the purple part in Figure 3; (2) it describes the comments by a LDA (Blei, Ng, and Jordan 2003) like mixture model, shown as the yellow part in Figure 3; and (3) it bridges the image information and comment information by learning a Bernoulli parameter λ_{d_m} to model how likely

Table 1: Notations in the proposed model.

SYMBOL	DESCRIPTION
e_{mi}	the latent variable indicating the topic (emotion) assigned with the visual feature x_{mt};
θ_m	the parameters of the multinomial distributions over the latent variable e specific to the image m;
μ_{el}, δ_{el}	the parameters of the Gaussian distribution over $x_{\cdot l}$ specific to the latent variable e;
z_{di}	the latent variable indicates the topic assigned with the word w_{di};
ϑ_d	the parameters of the multinomial distributions over the topics z specific to the comment d;
c_{di}	the latent variable indicates whether word w_{di} in comment d is related to emotion expression;
λ_d	the parameter of the Bernoulli distribution over $\mathbf{c}$ specific to comment d;
φ_z	the parameter of the multinomial distribution over w specific to the latent variable z;
α, γ, η	the parameters of the Dirichlet priors to the multinomial distribution ϑ_d, θ_m, and φ_z;
β	the parameter of the Beta prior to Bernoulli distribution λ;
τ	the parameter of the normal-gamma prior to the Gaussian distributions used to generate $\mathbf{x}$.

the author u of a comment d_m will be influenced by the image m when she writes d_m. See Table 1 for the notations in the proposed model.

The model considers social ties in part (3). Particularly, one is more easily to understand her close friends' emotional status, and will more likely to be influenced. Thus, the model gives λ_{d_m} a higher prior when v_m and v_d follow each other. We will introduce this part in detail later.

Generative Process. The generative process of the proposed model consists of two parts: visual feature generation (purple block in Figure 3) and comment generation (green and yellow blocks). First, for each image m, we sample its topic (emotion) distribution θ_m: $\theta_m \sim \mathrm{Dir}(\gamma)$. Next, for each visual feature x_{mt} of m, we sample a latent emotion e_{mt}: $e_{mt} \sim \mathrm{Mult}(\theta_m)$. After that, we generate the feature x_{mt}: $x_{mt} \sim \mathrm{N}(\mu_{e_{mt}t}, \delta_{e_{mt}t})$, where $\mu_{e_{mt}t}, \delta_{e_{mt}t}$ are parameters of the Gaussian distribution and are generated according to a normal-gamma distribution parameterized with τ.

For each comment d, we first generate its topic distribution ϑ_d: $\vartheta_d \sim \mathrm{Dir}(\alpha)$. We also generate the parameter λ_d of a Bernoulli distribution, which indicates how likely the emotion of d will be influenced by the emotion of its corresponding image m (d discusses about m): $\lambda_d \sim \mathrm{Be}(\beta)$. For each word w_{di} of d, we sample a latent variable c_{di}, which indicates whether the user is influenced by the image when she uses this word. When $c_{di} = 1$, we sample a topic z_{di} according to θ_m, otherwise z_{di} is sampled from d's own topic distribution ϑ_d. Finally, we generate the word w_{di}: $w_{di} \sim \mathrm{Mult}(\varphi_{z_{di}})$, where $\varphi_{z_{di}}$ is sampled according to a Dirichlet distribution parameterized with η. The details of the generative process can be found in Algorithm 1.

In practice, to define the value of $\beta = \{\beta_0, \beta_1\}$, we first let $\beta_0 = 1$. When the user u who publishes the image m and the user v who posts the comment d follows each other, we let d's corresponding $\beta_1 = b_1$, otherwise we let $\beta_1 = b_0$ ($b_0 < b_1$). Thus we can control how the social ties influence λ by adjusting b_0 and b_1. Intuitively, larger ratio of $b_1 : b_0$ stands for the prior that close friends are more easily to be influenced by each other.

Likelihood Definition. According to the generative process, we define the joint probability of a set of images M:

$$P(M, \theta, \mu, \delta; \gamma, \tau) = \prod_{m=1}^{|M|} P(\theta_m|\gamma) \prod_{t=1}^{T} \sum_{e=1}^{K} \theta_{me} P(x_{mt}|\mu_{et}, \delta_{et}) \prod_{e=1}^{K} \prod_{t=1}^{T} P(\mu_{et}, \delta_{et}|\tau) \quad (1)$$

We define the joint probability of a set of comments D as

$$P(D, \vartheta, \lambda, \varphi|\theta; \alpha, \beta, \eta) = \prod_{z=1}^{K} P(\varphi_z|\eta) \prod_{d=1}^{|D|} P(\vartheta_d|\alpha) P(\lambda_d|\beta) \prod_{n=1}^{N_d} \sum_{z=1}^{K} (\lambda_{d0}\vartheta_{dz} + \lambda_{d1}\theta_{m_d z})\varphi_{z w_{dn}} \quad (2)$$

where m_d indicates the image index which comment d discussed about; w_{dn} is the n-th word in comment d.

Finally, we define the likelihood of the proposed model as the product of Eq. 1 and Eq. 2. One advantage of the proposed model is that, by bridging image information and textural information by cross-sampling process, the model is able to differentiate "emotion topics" from irrelevant topics.

To further explain the latent space of the proposed model, in high-level intuition, our latent variables are similar to ones in LDA (Blei, Ng, and Jordan 2003). The difference is that, under each latent topic, LDA represents terms describing the same topic while our model has terms from comments and visual features from images to represent the same emotion.

Learning Algorithm

A variety of algorithms have been used for obtaining parameter estimates of topic models, such as variational method (Wainwright and Jordan 2008) (Jordan 1999). However, variational method suffers from a negative bias in estimating the variance parameters (Jaakkola and Qi 2006). In this paper, we employ Gibbs sampling (Lee 2012) (Resnik and Hardisty 2010) to estimate unknown parameters $\{\theta_M, \theta_D, \lambda, \mu, \delta, \varphi\}$.

In particular, we evaluate (a) the posterior distribution on $\mathbf{e}_m$ for each feature of each image m and then use the sampling results to infer θ_m; (b) the posterior distribution on $\mathbf{z}_d$ for each word in each comment d and use the results to infer ϑ_d. Finally, μ, δ, λ and φ can also be inferred from the sampling results. To the best of our knowledge, few work has studied how to use Gibbs sampling to estimate the parameters of Gaussian distributions, which remains the major challenge that the updating formation for μ and δ is hard to compute. We will introduce how we address this computation challenge in the left part of this section (Eq. 5).

```
Input: the hyper-parameters α, β0, b0, b1, γ, η, and τ, the
       image-based social network G
foreach image m ∈ M do
    foreach visual feature x_mt of m do
        Generate e_mt ∼ Mult(θ_m);
        Generate x_mt ∼ N(x_mt | μ_{e_mt t}, δ_{e_mt t});
    end
    foreach comment d, where a_md ∈ A do
        foreach word w_di of d do
            Generate c_di ∼ Mult(λ_d);
            if c_di = 0 then
                Generate z_di ∼ Mult(θ_d);
            end
            if c_di = 1 then
                Generate z_di ∼ Mult(θ_m);
            end
            Generate w_di ∼ Mult(φ_{z_di})
        end
    end
end
```

Algorithm 1: Probabilistic generative process in the proposed model.

More specifically, we begin with the posterior for sampling the latent variable z and c for each word in comments:

$$P(z_{di}, c_{di}=0|\mathbf{z}_{\neg di}, \mathbf{c}_{\neg di}, \mathbf{w}) = \frac{n^{\neg di}_{z_{di}d}+\alpha}{\sum_z(n^{\neg di}_{zd}+\alpha)} \times \frac{n^{\neg di}_{c_{di}d}+\beta_{c_{di}}}{\sum_c(n^{\neg di}_{cd}+\beta_c)} \times \frac{n^{\neg di}_{z_{di}w_{di}}+\eta}{\sum_w(n^{\neg di}_{z_{di}w}+\eta)} \tag{3}$$

where n_{zd} is the number of times that z has been sampled associated with the document d; n_{cd} is the number of times that c has been sampled in the comment d; n_{zw} is the number of times that word w has been generated by topic z in all comments; $n^{\neg di}$ with the superscript $\neg mi$ denotes a quantity, excluding the current instance. We have a similar formula for the case when $c_{di}=1$, with the only difference that the first term should be replaced by $\frac{n^{\neg di}_{z_{di}m}+\gamma}{\sum_z(n^{\neg di}_{zm}+\gamma)}$.

For the posterior to sample the latent variable e, we have

$$P(e_{mt};\mathbf{e}_{\neg mt},\mathbf{x}) = \frac{n^{\neg mt}_{me_{mt}}+\gamma}{\sum_e(n^{\neg mt}_{me}+\gamma)} \times \frac{\Gamma(\tau_2+\frac{n_{e_{mt}t}}{2})}{\Gamma(\tau_2+\frac{n^{\neg mt}_{e_{mt}t}}{2})} \times \frac{\sqrt{\tau_1+n^{\neg mt}_{e_{mt}t}}[\tau_3+\frac{1}{2}(n^{\neg mt}_{e_{mt}t}s^{\neg mt}_{e_{mt}t}+\frac{\tau_1 n^{\neg mt}_{e_{mt}t}(\overline{x}^{\neg mt}_{e_{mt}t}-\tau_0)^2}{\tau_1+n^{\neg mt}_{e_{mt}t}})]^{(\tau_2+\frac{n^{\neg mt}_{e_{mt}t}}{2})}}{\sqrt{\tau_1+n_{e_{mt}t}}[\tau_3+\frac{1}{2}(n_{e_{mi}t}s_{e_{mt}t}+\frac{\tau_1 n_{e_{mt}t}(\overline{x}_{e_{mt}t}-\tau_0)^2}{\tau_1+n_{e_{mt}t}})]^{(\tau_2+\frac{n_{e_{mt}t}}{2})}}$$

where n_{et} is the number of times that the latent variable e has been sampled associated with the t-th visual feature; $\overline{x}_{et}$ and s_{et} is the mean value and the precision of t-th feature associated with the latent variable e respectively; τ is the parameter of the normal-gamma prior to the Gaussian distributions used to sample $\mathbf{x}$. In practice, according to (Murphy 2007), we set τ_0 as the mean of all features, τ_1 as the instance number, τ_2 as the half of the instance number, and τ_3 as the sum of squared deviations of all features. One challenge here is the computation of the gamma function, which costs much time for calculating an exact value. In this work, we use Stirling's formula to approximately calculate the gamma function (Abramowitz and Stegun 1970).

We then estimate the parameters by the sampling results. The updating rule for θ, φ, and λ can be easily deduced with the similar idea with LDA (Heinrich 2005).

$$\theta_{dz} = \frac{n_{zd}+\alpha}{\sum_{z'}(n_{z'd}+\alpha)} \qquad \theta_{me} = \frac{n_{zm}+\gamma}{\sum_{e'}(n_{e'm}+\gamma)}$$
$$\lambda_{dc} = \frac{n_{cd}+\beta_c}{\sum_{c'}n_{c'd}+\beta_{c'}} \qquad \varphi_{zw} = \frac{n_{zw}+\eta}{\sum_{w'}(n_{zw'}+\eta)} \tag{4}$$

The major challenge here is the updating for μ_{et} and δ_{et} as the integration in the exact updating formation is hard to compute. To address this challenge, we approximate μ_{et} and δ_{et} as $E(\mu_{et})$ and $E(\delta_{et})$ respectively, and according to (Bernardo and Smith 2009), we have

$$\mu_{et} \approx E(\mu_{et}) = \frac{\tau_0\tau_1+n_{et}\bar{x}_{et}}{\tau_1+n_{et}}$$
$$\delta_{et} \approx E(\delta_{et}) = \frac{2\tau_2+n_{et}}{2\tau_3+n_{et}s_{et}+\frac{\tau_1 n_{et}(\bar{x}_{et}-\tau_0)^2}{\tau_1+n_{et}}} \tag{5}$$

To infer an image's emotion, one can easily use the emotion distribution of an image (θ_m), which is learned from the train data by the proposed model, as the feature and use a classifier (e.g., SVM (Burges 1998)) to classify images into different emotion categories.

Experimental Results

The dataset, all codes, and visual features used in the experiments are public available [3].

Experimental Setup

We perform our experiments on a large dataset collected from Flickr. In the dataset, we randomly download 354,192 images posted by 4,807 users. For each image, we also collect its tags and all comments. Thus we get 557,177 comments posted by 6,735 users in total. We also record the authors' profiles, including the authors' id, alias and their contact lists. Furthermore, the contact list is a list of ids which the user follows on Flickr, so we are able to figure out the relationship between two users.

For training and evaluating the proposed model for inferring emotions, we firstly need to know the primary emotions of the images. Manually acquiring a large labeled image dataset for evaluation is time-consuming and tedious. Thus for fairness and also simplicity, we compare the prediction results by the proposed method with those (emotion) tags (e.g., happy, unhappy) supplied by users. Particularly, we first manually define the word list for each of the six emotion categories based on WordNet [4] and HowNet [5]. Next we compare the adjective words in images' tags with the word lists.

[3] http://arnetminer.org/emotion/

[4] http://wordnet.princeton.edu/

[5] http://www.keenage.com/

Table 2: Performance of emotion inference.

Emotion	Method	Precision	Recall	F1-score	Emotion	Method	Precision	Recall	F1-score
Happiness	SVM	0.242	0.279	0.259	Disgust	SVM	0.192	0.236	0.212
	PFG	0.337	0.312	0.324		PFG	0.309	0.374	0.339
	LDA+SVM	0.333	**0.727**	**0.457**		LDA+SVM	0.223	0.223	0.223
	EL+SVM	**0.367**	0.410	0.388		EL+SVM	**0.331**	**0.432**	**0.374**
Surprise	SVM	0.197	0.037	0.063	Fear	SVM	0.204	0.264	0.230
	PFG	0.349	0.340	0.345		PFG	0.301	**0.408**	0.347
	LDA+SVM	0.218	0.048	0.078		LDA+SVM	0.211	0.225	0.217
	EL+SVM	**0.425**	**0.516**	**0.466**		EL+SVM	**0.371**	0.343	**0.356**
Anger	SVM	0.188	0.105	0.135	Sadness	SVM	0.225	0.365	0.278
	PFG	0.191	0.142	0.163		PFG	0.357	0.286	0.317
	LDA+SVM	0.222	0.109	0.146		LDA+SVM	0.257	0.278	0.267
	EL+SVM	**0.390**	**0.370**	**0.380**		EL+SVM	**0.561**	**0.617**	**0.588**

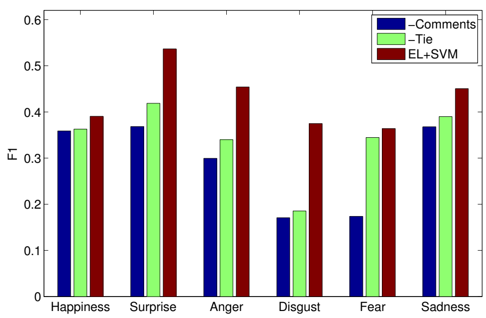


Figure 4: An analysis to study how user comments and social ties help in this problem.

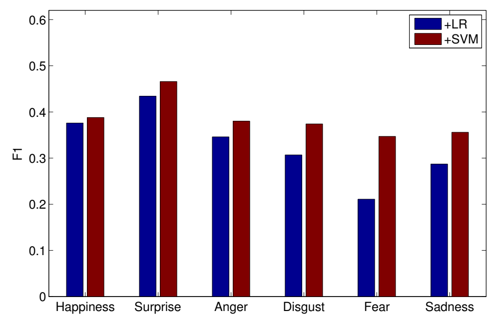


Figure 5: A study to show how different classifiers influence the performance.

The emotion category whose word list has the most same words as the tag words is finally assigned to the image. This has left us six sets of images, consisting of 145,946, 33,854, 22,040, 9,491, 54,637 and 35,935 images each.

Performance Evaluation

We first conduct a performance comparison experiment to demonstrate the effectiveness of our approach.

Evaluation measure. We compare the proposed model with alternative methods in terms of Precision, Recall, and F1-Measure [6]. We conduct a 5-fold cross validation to evaluate each method and report the averaged results.

Comparison methods.

SVM. This method simply regards the visual features of images as inputs and uses a SVM as a classifier. It then uses the trained classifier to infer the emotions. We use LIBSVM (Chang and Lin 2011) in this work.

PFG. This method is used in (Jia et al. 2012) to infer images' emotions. More specifically, it considers both the color features and social correlations among images and utilizes a partially-labeled factor graph model (Tang, Zhuang, and Tang 2011) as a classifier.

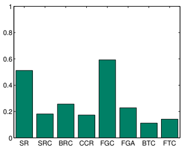
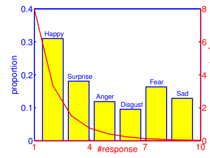
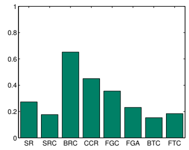
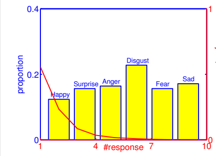
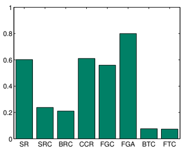
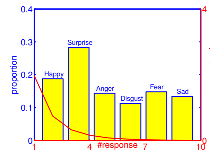
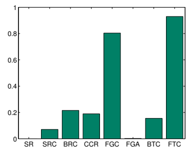
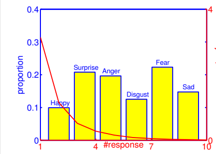
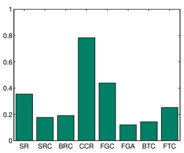
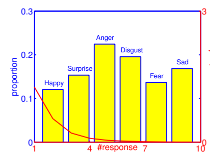
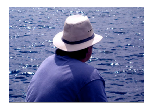
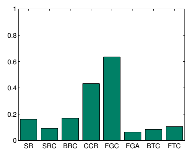
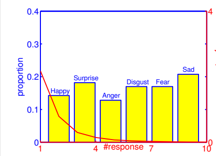
LDA+SVM. This method first uses LDA (Blei, Ng, and Jordan 2003) to extract hidden topics from user comments. It then uses the visual features of images, topic distributions of comments, and relationships between users as features to train SVM as a classifier. We use this method to compare the effectiveness of the joint modeling part of the proposed model and this alternative method.

EL+SVM. This method employs the proposed emotion learning method (EL) to learn the topic distributions of images. It then uses SVM as a classifier. For parameter configuration, we empirically set $K = 6$, $\alpha = 0.1$, $\gamma = 0.1$, $\tau = 0.01$, $b_0 = 1$, and $b_1 = 2$. We demonstrate how different parameters influence the performance in our webpage [5].

Comparison Results. Table 2 shows the experimental results. Overall, EL+SVM outperforms other baseline methods (e.g., +37.4% in terms of F1). SVM only considers visual features and ignores comment and social tie information, which hurts the performance. Although PFG further models the correlations between images, it also does not consider comment information and has worse performance than EL+SVM. LDA+SVM incorporates topic models to bring in comment information. However, it fails to differentiate "emotional topics" from those irrelevant topics as it extracts topics independently with image information. The proposed model naturally bridges these two pieces of information by learning how likely a comment is influenced by the relevant image, and obtains a improvement.

[6] http://en.wikipedia.org/wiki/Information_retrieval

Table 3: Image interpretations. We demonstrate how each visual feature distributes over each category of images by μ_{et} learned by the proposed model. The visual features include saturation (SR), saturation contrast (SRC), bright contrast (BRC), cool color ratio (CCR), figure-ground color difference (FGC), figure-ground area difference (FGA), background texture complexity (BTC), and foreground texture complexity (FTC). We also show a user's friends response after the user publishes different categories of images.

Category	Visual Features	Social Response	Category	Visual Features	Social Response
 Happiness			 Disgust		
 Surprise			 Fear		
 Anger			 Sadness		

Analysis and Discussions

Factor Analysis. We then conduct an experiment to study how comment information and social tie information help in this problem. Figure 4 shows the results. "-Comments" stands for the method using the proposed model but ignoring all comments. As we can see, not considering comment information hurts the performance in all tasks (e.g., -49.6% when infer Fear in term of F1 compared with EL+SVM). "-Tie" in Figure 4 stands for the method which only ignores social tie information and set all β as 1. We see that EL+SVM outperform -Tie especially on inferring Disgust. Disgust is an prototypic emotion encompasses a variety of reaction patterns according to subjective experience of different individuals (Moll et al. 2005). Friends with similar interests and experiences tend to have the same emotion to the same image. Thus social tie information helps more on inferring Disgust.

Classifier Analysis. To study how the proposed model cooperates with different classifiers, we train Logistic Regression and SVM as the classifier respectively and compare their performance. Figure 5 shows the results, from which we see that SVM seems more suitable for this problem, as it outperforms LR based methods in all inference tasks.

Image Interpretations. Compared with traditional methods, a more clear interpretation of the correlations between images and emotions is one of the advantages of the proposed model. Table 3 demonstrates how each visual feature $\mathbf{x}_t$ distributes over different emotions by the learned μ_t of the proposed model. For example, in the Happiness category, images tend to have high saturation and high bright contrast, which both bring out a sense of peace and joy. On the contrary, images in Sadness category tend to have lower saturation and saturation contrast, which both convey a sense of dullness and obscurity. Sad images also have low texture complexity, which gives a feeling of pithiness and coherence.

Social Response column of Table 3 elucidate how a user's friends response after the user publishes a certain category of images. Specifically, the red x-axises indicate the number of comments a friend leaves. And the red y-axises denote the number of friends who leave certain number of comments. Bars colored by yellow show the proportion of a user's friends with different emotions when leaving the comments. From the figure, we see an interesting phenomenon: the positive emotion (happiness) attracts more response (+4.4 times), and more easily to influences others to have the same emotion. On the contrast, when a user publishes a sad image, her friends' emotions distribute more uniformly, which implies sadness has less influence compared with happiness.

Conclusion

Can friends' interactions help us better understand one's emotions? In this paper, we propose a novel emotion learning method, which models the comment information and visual features of images simultaneously by learning a latent space to bridge these two pieces of information. It provides a new viewpoint for us to better understand how emotion differs from each other. Experiments on a Flickr dataset demonstrates that our model improves the performance on inferring users' emotions $+37.4\%$.

Acknowledgements. The work is supported by the National High-tech R&D Program (No. 2014AA015103), National Basic Research Program of China (No. 2014CB340500, No. 2012CB316006, No. 2011CB302201), Natural Science Foundation of China (No. 61222212, No. 61370023 and No. 61035004), and NSFC-ANR (No. 61261130588).

References

Abramowitz, M., and Stegun, I. 1970. Handbook of mathematical functions.

Bernardo, J. M., and Smith, A. F. 2009. *Bayesian theory*, volume 405. Wiley. com.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *JMLR* 3:993–1022.

Burges, C. J. 1998. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery* 2(2):121–167.

Chang, C.-C., and Lin, C.-J. 2011. LIBSVM: A library for support vector machines. *ACM TIST* 2:27:1–27:27.

Christophe, V., and Rimé, B. 1997. Exposure to the social sharing of emotion: Emotional impact, listener responses and secondary social sharing. *European Journal of Social Psychology* 27(1):37–54.

Elguebaly, T., and Bouguila, N. 2011. Bayesian learning of finite generalized gaussian mixture models on images. *Signal Processing* 91(4):801–820.

Heinrich, G. 2005. Parameter estimation for text analysis. *Web: http://www. arbylon. net/publications/text-est. pdf*.

Jaakkola, T. S., and Qi, Y. 2006. Parameter expanded variational bayesian methods. In *NIPS'06*, 1097–1104.

Jia, J.; Wu, S.; Wang, X.; Hu, P.; Cai, L.; and Tang, J. 2012. Can we understand van gogh's mood?: learning to infer affects from images in social networks. In *ACM Multimedia 2012*, 857–860.

Jordan, M. I. 1999. An introduction to variational methods for graphical models. In *Machine Learning*, 183–233.

Lee, P. M. 2012. *Bayesian statistics: an introduction*. John Wiley & Sons.

Moll, J.; de Oliveira-Souza, R.; Moll, F. T.; Ignácio, F. A.; Bramati, I. E.; Caparelli-Dáquer, E. M.; and Eslinger, P. J. 2005. The moral affiliations of disgust: A functional mri study. *Cognitive and behavioral neurology* 18(1):68–78.

Murphy, K. P. 2007. Conjugate bayesian analysis of the gaussian distribution. *def* 1($2\sigma 2$):16.

Ou, L.-C.; Luo, M. R.; Woodcock, A.; and Wright, A. 2004. A study of colour emotion and colour preference. part i: Colour emotions for single colours. *Color Research & Application* 29(3):232–240.

Resnik, P., and Hardisty, E. 2010. Gibbs sampling for the uninitiated. Technical report, DTIC Document.

Rimé, B. 2005. *Le partage social des émotions*. Presses universitaires de France.

Shin, Y., and Kim, E. Y. 2010. Affective prediction in photographic images using probabilistic affective model. In *CIVR'10*, 390–397.

Tang, W.; Zhuang, H.; and Tang, J. 2011. Learning to infer social ties in large networks. In *ECML/PKDD (3)'11*, 381–397.

Wainwright, M., and Jordan, M. 2008. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning* 1(1-2):1–305.