UNDERSTANDING THE EMOTIONS BEHIND SOCIAL IMAGES: INFERRING WITH USER DEMOGRAPHICS

Boya Wu[1,2,3], *Jia Jia*[1,2,3], *Yang Yang*[1], *Peijun Zhao*[1,2,3], *Jie Tang*[1]

[1]Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
[2]Key Laboratory of Pervasive Computing, Ministry of Education
[3]Tsinghua National Laboratory for Information Science and Technology (TNList)
stella.1991@163.com,jjia@mail.tsinghua.edu.cn,sherlockbourne@gmail.com
421769833@qq.com,jietang@tsinghua.edu.cn

ABSTRACT

Understanding the essential emotions behind social images is of vital importance: it can benefit many applications such as image retrieval and personalized recommendation. While previous related research mostly focuses on the image visual features, in this paper, we aim to tackle this problem by "***linking inferring with users' demographics***". Specifically, we propose a partially-labeled factor graph model named **D-FGM**, to predict the emotions embedded in social images not only by the image visual features, but also by the information of users' demographics. We investigate whether users' demographics like gender, marital status and occupation are related to emotions of social images, and then leverage the uncovered patterns into modeling as different factors. Experiments on a data set from the world's largest image sharing website Flickr[1] confirm the accuracy of the proposed model. The effectiveness of the users' demographics factors is also verified by the factor contribution analysis, which reveals some interesting behavioral phenomena as well.

Index Terms— Emotion, image, users' demographics

1. INTRODUCTION

Emotion stimulates the mind 3,000 times faster than rational thoughts [1]. With the rapid development of social networks, people get used to sharing their emotional experiences on these platforms. As a natural way to express our feelings, images are uploaded and shared on social networks. We define these images as "**Social Images**". Our preliminary statistics indicate that 38% of the images on the world's largest image social network Flickr are explicitly annotated with either positive or negative emotions by their uploaders. Understanding the essential emotions behind social images is of vital significance. It can benefit many applications, such as image retrieval and personalized recommendation.

When it comes to inferring emotions from social images, previous related research mainly focuses on the image visual features, which means enhancing the emotion inferring performance by extracting the effective visual features and choosing their proper combinations. J.Machajdik and A.Hanbury [2] investigate four categories of low-level features like wavelet textures and GLCM-features. S.Zhao [3] et al. explore principles-of-art features for image emotion recognition. Similar works can be found in [4], [5].

Recently, the research on social networks has verified that the users' demographics are associated with the users' behaviors. Y.Dong et al. [6] discover that people of different ages have different social strategies to maintain their social connections. H.Huang et al. [7] uncover how users' demographics influence the formation of closed triads on social networks. Moreover, the behavioral research has proved that the human perception of emotions varies according to their personal attributes. A.Fischer et al. [8] point out that there is a gender difference in the perception of emotions, namely that men report more powerful emotions (e.g., anger), whereas women report more powerless emotions (e.g., sadness, fear). However, can the user's demographics be leveraged to help infer the emotions from social images is still largely undeveloped. The problem is non-trivial and has several challenges. First, though a few literatures demonstrate the existence of the correlation between the users' demographics and the perception of emotions, it is still unclear whether the correlation exists on image social networks. Second, how to model the users' demographics and other information in a joint framework? Third, how to validate the effectiveness of the proposed model on a real-world image social network?

To address these challenges, first we investigate whether users' demographics like gender, marital status and occupation are related to the emotions of social images. Then we leverage the uncovered patterns into modeling as different factors. Specifically, we propose a partially-labeled factor graph model named **D-FGM**, to infer emotions from social images not only by the visual features, but also by the information of users' demographics. As for experiments, we construct a library of millions of images and users (2,060,353

[1]http://www.flickr.com/

images and 1,255,478 users) from the world's largest image sharing website Flickr. The experimental results confirm the accuracy of the proposed model, e.g., achieving 19.4% improvement compared with SVM (Support Vector Machine) under the evaluation of F1-Measure. The effectiveness of the users' demographics factors is also demonstrated by the factor contribution analysis, which reveals some interesting behavioral phenomena. For example, in terms of *sadness*, the image emotion is mainly determined by the visual features. Interestingly however, when it comes to *disgust* and *surprise*, males and females have different emotion perception; when it comes to *fear*, whether the user is single or taken makes differences; and when it comes to *happiness* and *anger*,the perception of these emotions is associated with the user's occupation.

2. PROBLEM DEFINITION

In this section, we give several necessary definitions and formalize the problem.

Users' demographics: The users' demographics usually refer to the users' personal attributes, which for example, contain the age, gender, location information in [6]. In this paper, we present user v_i's demographics as three vectors $\mathbf{p_i}$: *gender*, *marital status* and *occupation*. The *gender* is defined as male or female. The *marital status* is defined as single or taken. For the *occupation*, by manually screening the users' profiles on Flickr, we pick out 25 main kinds of occupations and classify them into two categories, namely, the artists and the engineers. The "artists" include writer, musician, dancer, etc. and the "engineers" include programmer, scientists, etc.

Image social network: A partially-labeled time-varying image social network can be defined as $G = (V, P, E^t, X^L, X^U)$, where V is the set of $|V| = n$ users, $P = \{p_i\}$ is the set of the users' demographics, $E^t \subset V \times V$ is the friendship among users at time t, X^L represents the labeled images and X^U represents the unlabeled images.

Emotion: The emotion of user v_i at time t is denoted as y_i^t. The emotion of the image $x_{i,j}^t$ uploaded by user v_i at time t is denoted as $y_{i,j}^t$, where j is the index of images uploaded by user v_i.

In this work, we have the following intuition: users' emotions are expressed by the emotions of the images they upload on image social networks, which means $y_i^t = y_{i,j}^t$.

We adopted Ekman's [9] classical theory of basic human emotion categories, namely, *happiness, surprise, anger, disgust, fear and sadness* and denote the emotional space as R.

Based on the above definitions, the learning task of our model is put forward as follows.

Learning task: Given a partially-labeled time-varying image social network $G = (V, P, E^t, X^L, X^U)$, find a function f to predict the emotions from unlabeled images:

$$f : G = (V, P, E^t, X^L, X^U) \to Y \quad (1)$$

where $Y = \{y_{i,j}^t\} \in R$.

3. OBSERVATIONS

Users' demographics have been verified to be associated with users' behaviors in social networks [6], [7]. Wondering whether users' demographics make differences on users' perception of emotions, we conduct a series of observations and present several interesting phenomena we have discovered.

3.1. Data collection

We randomly download 2,060,353 images and 1,255,478 users' profiles from Flickr. To conduct the observations, first we need to know the primary emotion of images. Owing to the massive scale of our data set, manually labeling the emotion for every image is not practical. Herein we adopt a method to label the emotion of images automatically. This method is also used by Xie [10] and Hwang [11]. First we construct word lists for each of the six emotion categories through WordNet[2] and HowNet[3]. Next we compare the image tags written by the uploader with every word list and the image can be labeled with a type of emotion whose word list match the words of the tags most frequently. In this way, 218,816 images are labeled. These images are uploaded by 2,312 users, and each emotional category contains 101189, 21169, 17491, 11571, 37791, 29605 images.

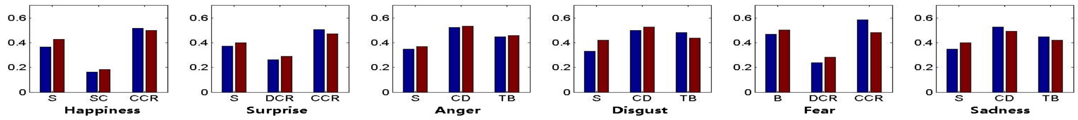
3.2. Observations on users' demographics

Herein we observe the correlation between image emotions and the three parts of the users' demographics respectively.

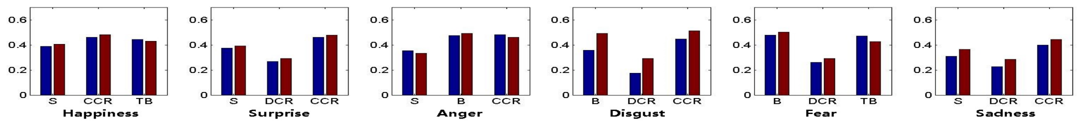
Observation on the gender correlation. First we classify the users into males and females, each containing 363 and 1,670 users. Then we randomly pick out 2,000 images from each emotion category, half uploaded by males and the other half uploaded by females and analyze the distributions of the visual features of the images. Figure 1(a) presents several representative results. We can see that in the case of *disgust*, the distributions of visual features of images uploaded by males and females are different. For instance, the saturation (*S*) of the images uploaded by females is 21.4% lower than the images uploaded by males. It suggests that though both males and females want to express their disgust through images, they tend to use different visual features to convey their feelings. In terms of *surprise*, the cool color ratio (*CCR*) of the images uploaded by females is 19.9% higher than the images uploaded by males, showing that males and females have different ways to express their surprise. The observation results can be concluded that there is a gender difference in the emotion perception of social images.

Observations on the marital status correlation. Similarly, according to the user's marital status, we divide the users into single and taken, each containing 310 and 954 users. We conduct the observations again and the results are visualized in Figure 1(b). The distributions of visual features of images uploaded by single users and taken users are different in *fear* and *sadness*. For example, in terms of *sadness*,

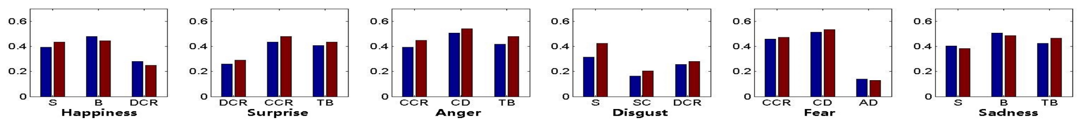
[2]http://wordnet.princeton.edu/
[3]http://www.keenage.com/



(a) The difference of the distributions of the visual features of the images uploaded by females and males.



(b) The difference of the distributions of the visual features of the images uploaded by single users and taken users.



(c) The difference of the distributions of the visual features of the images uploaded by engineers and artists.

Fig. 1. The difference of the distributions of representative visual features of images uploaded by users with different personal attributes, which shows the correlation between image emotions and users' demographics. The features include: S: saturation, SC: saturation contrast, B: brightness DCR: dull color ratio, CCR: cool color ratio, CD: color difference, AD: area difference, TB: texture complexity of background. The values of features are normalized between 0 and 1 over the whole data set.

the saturation (*S*) of the images uploaded by single users is 15.3% lower than the images uploaded by taken users, and in terms of *fear*, the background texture complexity (*TB*) of the images uploaded by single users is 11.0% higher than the images uploaded by taken users. The results show that single users and taken users use different ways to express the same feeling, indicating that their emotion perception of social images differs.

Observations on the occupation correlation. As described in the problem definition section, we carefully select 217 users as "engineers" and 279 users as "artists". We conduct the observations again and Figure 1(c) illustrates the results. In terms of *happiness*, the brightness of the image uploaded by engineers is 7.6% higher than the images uploaded by artists. In terms of *anger*, the cool color ratio of the images uploaded by engineers is (*CCR*) is 11.8% lower the images uploaded by artists. The results suggest that on image social networks, engineers and artists have different emotion perception.

The observation can be summarized as follows:

- Males and Females have different ways to express *disgust* and *surprise*. There is a gender difference in the emotion perception of social images.
- Single users and taken users use different ways to express *fear* and *sadness*, indicating that their emotion perception for these emotions is different.
- Engineers and artists use different ways to convey *happiness* and *anger*, suggesting that the occupation may be related to the users' emotion perception.

4. MODEL

To leverage the above findings to help infer emotions from social images, we propose a factor graph model named **D-FGM** to solve the problem. Our basic idea is to define the correlations using different types of factor functions. In a factor graph model, the objective function is defined based on the joint probability of the factor functions [1], [12], so the problem of emotion model learning is cast as the model parameters learning that maximizes the joint probability.

In our model, four types of correlations can be defined as factor functions.

- **Visual features correlation** $f_1(\mathbf{u}_{i,j}^t, y_{i,j}^t)$. It represents the correlation between the visual features $\mathbf{u}_{i,j}^t$ and the image emotion $y_{i,j}^t$.
- **Temporal correlation** $f_2(y_i^{t'}, y_i^t)$. Previous research has verified that there is a strong dependency between one's current emotion and the emotions in the recent past on social networks [1]. This correlation is defined as temporal correlation, which represents the influence of the user's previous emotions in the recent past t' on the current emotion at time t.
- **Social correlation**. Creating and sharing images on image social networks is very different from traditional creation. Some users may have a strong influence on their friends' emotions and some emotions may spread quickly on the social network [5]. The social correlation contains three parts: the correlation between the image emotion and the number of the user's friends $f_3(s_i^t, y_{i,j}^t)$, the correlation between the image emotion

Algorithm 1 The learning and inference algorithm of emotions from social images.

Input:
A partially-labeled time-varying image social network $G = (V, P, E^t, X^L, X^U)$ and the learning ratio λ

Output:
Construct a partially-labeled factor graph.
Initiate parameters $\theta = \{\alpha, \beta, \gamma, \delta, \varepsilon_i, \eta_{i,j}\}$
repeat
  Calculate $E_{(p_\theta(Y|Y^U,G))}S$ using LBP
  Calculate $E_{(p(Y|G))}S$ using LBP
  Calculate the gradient of θ: $E_{(p_\theta(Y|Y^U,G))}S - E_{(p_\theta(Y|G))}S$
  Update θ with the learning ratio λ: $\theta = \theta_0 + \frac{\partial \varnothing}{\partial \theta}\lambda$
until convergence
Get the inference results $Y = y_{i,j}^t,\ y_{i,j}^t \in R$ and the trained parameters $\theta = \{\alpha, \beta, \gamma, \delta, \varepsilon_i, \eta_{i,j}\}$

and the major emotion of the user's friends $f_4(m_i^t, y_{i,j}^t)$ and the correlation between the image emotion and the user's intimacy with friends $f_5(y_i^t, y_j^t, \mu_{i,j}^t)$.

- **Users' demographics correlation** $f_6(\mathbf{p}_i, y_{i,j}^t)$. It denotes the correlation between the image emotion and the users' demographics information, which is formalized as three vectors gender, marital status and occupation.

4.1. The predictive model

The input of the model is an image social network G, and the output of the model is the inference results Y. The correlations described above are instantiated as different factor functions.

(1) **Visual features correlation function**:

$$f_1(\mathbf{u}_{i,j}^t, y_{i,j}^t) = \frac{1}{z_\alpha} \exp\{\alpha^T \cdot \mathbf{u}_{i,j}^t\} \tag{2}$$

where $\mathbf{u_{i,j}^t}$ represents the visual features and $y_{i,j}^t$ represents the emotion of image $x_{i,j}^t$.

(2) **Temporal correlation function**:

$$f_2(y_i^{t'}, y_i^t) = \frac{1}{z_\varepsilon} \exp\{\varepsilon_i \cdot g(y_i^{t'}, y_i^t)\}, t' < t \tag{3}$$

where y_i^t and $y_i^{t'}$ represent the emotion of user v_i at time t and t'. Function $g(y_i^{t'}, y_i^t)$ is used to depict the correlation.

(3) **Social correlation function**:

$$f_3(s_i^t, y_{i,j}^t) = \frac{1}{z_\gamma} \exp\{\gamma^T \cdot s_i^t\} \tag{4}$$

where s_i^t denotes the number of user's friends.

$$f_4(m_i^t, y_{i,j}^t) = \frac{1}{z_\delta} \exp\{\delta^T \cdot m_i^t\} \tag{5}$$

where m_i^t denotes the major emotion of the user's friends.

$$f_5(y_i^t, y_j^t, \mu_{i,j}^t) = \frac{1}{z_\eta} \exp\{\eta_{i,j} \cdot h(y_i^t, y_j^t, \mu_{i,j}^t)\} \tag{6}$$

where y_i^t and y_j^t represents the emotions of user v_i and v_j at time t and $\mu_{i,j}^t$ measures the intimacy between them at time t, which is calculated from their interaction frequency. Function $h(y_i^t, y_j^t, \mu_{i,j}^t)$ is used to depict the correlation.

(4) **Users' demographics correlation function**:

$$f_6(\mathbf{p}_i, y_{i,j}^t) = \frac{1}{z_\beta} \exp\{\beta^T \cdot \mathbf{p}_i\} \tag{7}$$

where $\mathbf{p}_i$ denotes the user's demographics information, namely, gender, marital status and occupation.

Given the above factor functions, we define the joint distribution of the model:

$$\begin{aligned}P(Y|G) = \frac{1}{Z} \prod_{x_{i,j}^t} f_1(\mathbf{u}_{i,j}^t, y_{i,j}^t) \prod_{x_{i,j}^t} \prod_{y_i^{t'}} f_2(y_i^{t'}, y_i^t) \\ \prod_{x_{i,j}^t} f_3(s_i^t, y_{i,j}^t) \prod_{x_{i,j}^t} f_4(m_i^t, y_{i,j}^t) \prod_{x_{i,j}^t} \prod_{v_j} f_5(y_i^t, y_j^t, \mu_{i,j}^t) \\ \prod_{x_{i,j}^t} f_6(\mathbf{p}_i, y_{i,j}^t) = \frac{1}{Z} \exp\{\theta^T S\}\end{aligned} \tag{8}$$

where $Z = Z_\alpha Z_\varepsilon Z_\beta Z_\gamma Z_\delta Z_\eta$ is the normalization term, S is the aggregation of factor functions over all nodes, θ denotes all the parameters, i.e., $\theta = \{\alpha, \beta, \gamma, \delta, \varepsilon_i, \eta_{i,j}\}$.

Therefore the target of modeling is to maximize the log-likelihood objective function $\varnothing = \log P(Y|G)$.

4.2. Model learning

Given the model's input and output, next we'll detail the learning process of the model and the algorithm is summarized in Algorithm 1.

The objective function can be rewritten as:

$$\begin{aligned}\varnothing &= \log P(Y|G) = \log \sum_{Y|Y^U} \exp\{\theta^T S\} - \log Z \\ &= \log \sum_{Y|Y^U} \exp\{\theta^T S\} - \log \sum_{Y} \exp\{\theta^T S\}\end{aligned} \tag{9}$$

Thus the gradient of θ can be represented as:

$$\begin{aligned}\frac{\partial \varnothing}{\partial \theta} &= \frac{\partial(\log \sum_{Y|Y^U} \exp\{\theta^T S\} - \log \sum_Y \exp\{\theta^T S\})}{\partial \theta} \\ &= E_{p_\theta(Y|Y^U,G)}S - E_{p_\theta(Y|G)}S\end{aligned} \tag{10}$$

The algorithm updates the parameters by $\theta = \theta_0 + \frac{\partial \varnothing}{\partial \theta} \cdot \lambda$.

5. EXPERIMENTS

5.1. Experimental setup

Data set. The raw data set we employed and the way we establish the ground-truth are described in the observations section. In order to examine the performance of every emotion category, we evenly and randomly pick out 11,500 images from every emotion category and 69,000 images are chosen in total, 60% for training and 40% for testing.

Herein, we adopt the method proposed by Wang [13] to extract the visual features, including the color theme, saturation, brightness, etc. In total we extract 25 features and the effectiveness of these features is confirmed in [12], [5].

Comparison methods. We conduct performance comparison experiments to demonstrate the effectiveness of our

Table 1. The F1-Measure of the emotion inference.

Method	Happiness	Surprise	Anger	Disgust	Fear	Sadness
NB	0.266	0.082	0.058	0.291	0.145	0.275
SVM	0.294	0.129	0.088	0.325	0.233	0.286
FGM	0.433	0.361	0.295	0.447	0.379	0.434
D-FGM	**0.451**	**0.401**	**0.334**	**0.463**	**0.416**	**0.442**

model. Three existing methods, namely, Naive Bayesian (NB), Support Vector Machine (SVM) and traditional factor graph model (FGM) are used for comparison.

NB: Naive Bayesian is a widely used classifier and achieves good performance [2]. It is also used as the baseline method in [1]. We use the Naive Bayesian tool provided by MATLAB[4] .

SVM: SVM is a frequently-used method in many classification problems. The method is also used as the baseline method in [1], [5]. Herein we use LIBSVM design by Chang and Lin [5].

FGM: This method is used in [5] to infer emotions of images. A partially-labeled factor graph model is utilized as a classifier.

Evaluation metrics. We compare the performance of our proposed model with three baseline methods in terms of precision, recall and F1-Measure. These evaluation metrics are widely used in the classification problems [12].

5.2. Experimental results

Due to the limit of the length of the paper, herein we just exhibit the F1-Measure in Table 1 on the behalf of the evaluation metrics. As shown in the table, our model significantly enhances the performance. The average F1-Measure reaches 0.420, increased by 23.4% compared with Naive Bayesian, 19.4% compared with SVM and 2.4% compared with FGM.

NB and SVM are only capable of handling *vectors*. In this problem the vectors contain the visual features, the users' demographics and parts of the social attributes (the number of the user's friends and the major emotion of the user's friends). However, these two models cannot handle the correlations between images, which are instantiated as *edges* in FGM and D-FGM. As a result they let go of the temporal correlation and the intimacy with the user's friends. As for FGM, it can model the vectors and edges jointly. However, all the edges are of the same weight in FGM, so though the model can take the correlations between images into consideration, it still cannot model the differences between edges. This drawback hurts the performance and lets go of some important attributes, such as the user's intimacy with friends, where the intimacy is modeled as the weight of the edges. On the contrary, the proposed D-FGM can model the vectors and the weighted edges together, so it better depicts the image social network and achieves the best performance.

[4] A widely used software developed by MathWorks, Inc.

[5] A library for support vector machines.

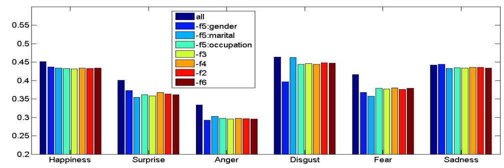


Fig. 2. F1-Measure of different factor combinations.

5.3. Factor contribution analysis

In our work, we utilize the information of the users' demographics and introduce them into a factor graph model as factor functions. Wondering whether these factors benefit the inference, we investigate the contribution of every factor in the model. Every time we take each of the factors out of the primitive model and examine the performance while the other factors remain the same.

The experimental results evaluated by F1-Measure are visualized in Figure 2. The model involving all factors achieves the best performance in all emotion categories, which validates the effectiveness of the factors. Other interesting results are summarized as follows.

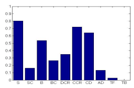
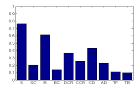
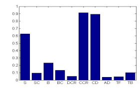
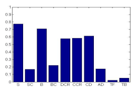
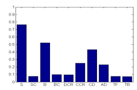
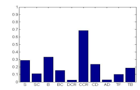
- When inferring *disgust* and *surprise*, the gender information benefits the inference remarkably (+6.7% for *disgust* and +2.8% for *surprise*).
- When inferring *happiness* and *anger*, the occupation information really matters. The F1-Measure increases by 1.9% when inferring *happiness* and 3.7% when inferring *anger*.
- When inferring *fear*, the marital status information is very useful by showing 5.9% improvement.
- However, interestingly, when inferring *sadness*, the gender and occupation information makes little help, and the marital status information helps slightly (+1.0%), which indicates that the perception of *sadness* is mainly determined by the visual features.

The results also correspond to the observations we described before, which verifies the rationality of introducing the users' demographics into the modeling of inferring emotions from social images.

5.4. Case study

In the above investigation we discover that different users' demographics result in different emotion perception of images. Table 2 details the analysis by reporting the labeled emotion, the visual features and the users' demographics of several images. Two images on the left depict the same scene and their visual features are quite alike. However, we find out that the image on the top is uploaded by a female named *bekahpaige* on Sept, 6th, 2003, who labels this image as *happiness* and the image on the bottom is uploaded by a male named *54rf* on Apr, 17th, 2009, who labels this image as *surprise*. The gender difference in human emotion perception is verified by the behavioral study [8]. The difference can be explained that

Table 2. Different users' demographics result in different emotion perception of images.

Image & Emotion tags written by the uploader	**Visual features**	**User's demographics**	**Image & Emotion tags written by the uploader**	**Visual features**	**User's demographics**	**Image & Emotion tags written by the uploader**	**Visual features**	**User's demographics**
 Happiness		**Female** Single Artist	 Sadness		Male **Single** Engineer	 Happiness		Male **Artist**
 Surprise		**Male** Single Artist	 Happiness		Male **Taken** Engineer	 Disgust		Male **Engineer**

males are likely to be less aware of their surroundings, and thus the rare occasion of noticing the daily sunrise fills the male's heart with surprise, while the more observant female is simply happy with the pleasant phenomenon. Similarly, two images in the middle both capture the blossom of flowers, but the top one expresses *sadness* by a single user *akshaydavis* on Apr, 20th, 2008 and the bottom one conveys *happiness* by a taken user *davidhelan* on Jul, 20th, 2005, indicating that single users and taken users have different emotion perception. The images on the right demonstrate the different emotion perception between engineers and artists.

6. CONCLUSIONS

In this paper, we study the problem of "link inferring with users' demographics" for understanding the emotions behind social images. First we investigate whether users' demographics relate to image emotions on social networks. Then by introducing these patterns as factor functions into modeling, we propose a factor graph model called **D-FGM** which can not only infer emotions from social images by the visual features, but also by the users' demographics. Experiments on the world's largest image sharing website Flickr validate the effectiveness of our model.

7. ACKNOWLEDGEMENTS

This work is supported by the National Basic Research Program of China (2012CB316401), National Natural, and Science Foundation of China (61370023). This work is partially supported by the National Basic Research Program of China (2011CB302201), and the National High Technology Research and Development Program("863"Program) of China (2012AA011602). We would also like to thank Microsoft Research Asia-Tsinghua Univertity Joint Laboratory: FY14-RES-SPONSOR-111 for its support.

8. REFERENCES

[1] J. Tang, Y. Zhang, J. Sun, J. Rao, W. Yu, Y. Chen, and ACM Fong, "Quantitative study of individual emotional states in social networks," *IEEE Transactions on Affective Computing (TAC)*, vol. 3, pp. 132–144, 2012.

[2] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *ACM Multimedia*, 2010, pp. 83–92.

[3] S. Zhao, Y. Gao, X. Jiang, H. Yao, T. Chua, and X. Sun, "Exploring principles-of-art features for image emotion recognition," in *ACM Multimedia*, 2014, pp. 47–56.

[4] Y. Shin and E.Y. Kim, "Affective prediction in photographic images using probabilistic affective model," in *ACM International Conference on Image and Video Retrieval*, 2010, pp. 390–397.

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

[6] Y. Dong, Y. Yang, J. Tang, Y. Yang, and N.V. Chawla, "Inferring user demographics and social strategies in mobile social networks," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014, pp. 15–24.

[7] H. Huang, J. Tang, S. Wu, L. Liu, and X. Fu, "Mining triadic closure patterns in social networks," in *World Wide Web Conference (WWW)*, 2014, pp. 499–504.

[8] A.H. Fischer, P.M. Rodriguez Mosquera, A.E. van Vianen, and A.S. Manstead, "Gender and culture differences in emotion," *Emotion*, vol. 4, pp. 87–94, Mar. 2004.

[9] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[10] L. Xie, "Picture tags and world knowledge," in *ACM Multimedia*, 2013.

[11] S. J. Hwang, "Discriminative object categorization with external semantic knowledge," 2013, Ph.D. Dissertation, The University of Texas at Austin.

[12] Y. Yang, J. Jia, S. Zhang, B. Wu, Q. Chen, J. Li, C. Xing, and J. Tang, "How do your friends on social media disclose your emotions?," in *AAAI Conference on Artificial Intelligence (AAAI)*, 2014, pp. 306–312.

[13] X. Wang, J. Jia, J. Yin, and L. Cai, "Interpretable aesthetic features for affective image classification," in *ICIP*, 2013, pp. 3230–3234.