Can We Understand van Gogh's Mood? Learning to Infer Affects from Images in Social Networks

Jia Jia[†‡♯], Sen Wu[♯], Xiaohui Wang[†‡♯], Peiyun Hu[♯], Lianhong Cai[†‡♯], Jie Tang[♯]
[†]Key Laboratory of Pervasive Computing, Ministry of Education
[‡]Tsinghua National Laboratory for Information Science and Technology (TNList)
[♯]Department of Computer Science and Technology, Tsinghua University
jjia@tsinghua.edu.cn, ronaldosen@gmail.com, wangxh09@mails.tsinghua.edu.cn, tsinghua.hp@gmail.com,
clh-dcs@tsinghua.edu.cn, jietang@tsinghua.edu.cn

ABSTRACT

Can we understand van Gogh's mood from his artworks? For many years, people have tried to capture van Gogh's affects from his artworks so as to understand the essential meaning behind the images and catch on why van Gogh created these works. In this paper, we study the problem of inferring affects from images in social networks. In particular, we aim to answer: What are the fundamental features that reflect the affects of the authors in images? How the social network information can be leveraged to help detect these affects? We propose a semi-supervised framework to formulate the problem into a factor graph model. Experiments on 20,000 random-download Flickr images show that our method can achieve a precision of 49% with a recall of 24% on inferring authors' affects into 16 categories. Finally, we demonstrate the effectiveness of the proposed method on automatically understanding van Gogh's Mood from his artworks, and inferring the trend of public affects around special event.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval models

General Terms

Algorithms

Keywords

Affects, Factor graph model, Color features

1. INTRODUCTION

Vincent van Gogh was a Post-Impressionist painter whose artworks had a far-reaching influence on 20th-century art. Van Gogh's artworks carry strong affects. For example, his "Wheat fields" is associated with melancholy and extreme loneliness, clearly expressing the artist's state of mind in his final days. Capturing the artist's affects implied in the artworks can immensely help us understand the essential meaning the author wants to express. This has long been viewed as an important issue for studying van Gogh's artworks.

Nowadays, with the rapid development of social networks, e.g. Flickr[1] and instagram[2], everyone becomes an "artist". Images provide a natural way to express one's affects. In addition, social networks offer a platform for the "artists" to share their "artworks" and to influence the affects of their friends. An interesting question is: can we automatically infer authors' affects from those images? Uncovering the authors' affects can benefit many applications. For example, if we can assign affective category to each image, we could leverage the affective information to help image retrieval and personalized recommendation.

The problem is non-trivial and poses a set of unique challenges. First of all, what are the fundamental factors that reflect the authors' affects? Generally, the visual features of an image include color, shape, composition, etc. Among these features, color has been shown to play an important role in image affective analysis [7]. For example, Li-Chen Ou explored the affective information for single color and two-color combinations [4]. Kobayashi defined 16 affective categories [2], and Shin and Kim discussed how to use the image features in particular color features to classify photographic images [5]. Some other work can be found in [3, 6]. However, all of the aforementioned work does not consider how to infer affects from images created by ordinary users in social networks. In addition, when considering this problem in the context of social networks, one has to further deal with the following problems:

- **Network:** Creating and sharing images in social networks is very different from traditional artistic creation. Some users may have a strong influence on their friends' affects. Some affects may quickly spread in the social network. How to leverage the social network information for helping infer affects from images is a challenging issue.
- **Model:** How to design a principled model to automatically uncover the authors' affects? It is unclear how to combine the different pieces of information together into a unified model.

In this paper, we systematically study the problem of inferring affects from images in social networks and propose a semi-supervised framework to formulate the problem into a factor graph model. Experiments on more than 20,000 images randomly downloaded from Flickr show that the proposed method can achieve a precision of 49% and a recall of 24% on inferring authors' affects into Kobayashi's 16 categories, which significantly outperforms the existing method using SVM. Finally, we use case studies to further demonstrate the effectiveness of the proposed method.

MM'12, October 29–November 2, 2012, Nara, Japan.

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

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

2. PROBLEM FORMULATION

In general, our study takes as input an image in social network $G = (V, E)$, where $V = \{p_1, \ldots, p_M\}$ is the set of images with p_i published in time t_i, $|V| = M$ and $E \subset V \times V$ is the edge set. Each edge e_{ij} represents image p_i having a correlation with image p_j(e.g. p_i and p_j uploaded by the same user in a short time or they have the same tag). Our goal is to learn a model that can effectively infer the affects from images. Given this, we can define the user's affects status as follows.

Definition 1. **Affects**: The affective categories of an image p_i is denoted as $y_i \subset \mathcal{Y}$, where $\mathcal{Y}$ is the affective space.

We use 16 discrete categories which are proposed by Kobayashi to cover the affective space: "pretty", "casual", "dynamic", "luxurious", "wild", "romantic", "natural", "elegant", "old-fashioned", "dapper", "dignified", "formal", "chic", "clear", "jaunty", "modern". Kobayashi proposed these 16 affective categories by long-term psychology experiments. Each affective category contains a set of related semantic concepts. Examples of some affective categories and their corresponding semantic concepts are shown as follows:

- **Casual:** cheerful, happy, carefree, friendly, humorous, animated, vivid, lovely
- **Modern:** progressive, revolutionary, intellectual, rational, manmade, mechanical
- **Chic:** subtle, quiet, nonchalant, urbane, bleak, cultivated, sober, cerebral
- **Dapper:** quiet, awe-inspiring, gentlemanly, earnest, sound, tidy, neat

Definition 2. **Partially labeled network**: The partially labeled network is denoted as $G = (V^L, V^U, E, \mathbf{X})$, where V^L is a set of labeled affects and V^U is the set of unlabeled affects, $E \subset V \times V$ is the correlations between two affects, $\mathbf{X}$ is an $|V| \times d$ attribute matrix associated with vertices in V with each row corresponding to an edge, each column representing an attribute and an element x_{ij} denoting the value of the j^{th} attribute of vertice p_i. The label of vertice p_i is denoted as $y_i \subset \mathcal{Y}$.

Problem 1. **Learning task:** Given a partially labeled network G, our goal is to learn a predictive function *f* to predict the affective categories of images. Formally, we have

$$f : G = (V^L, V^U, E, \mathbf{X}) \rightarrow \mathcal{A}$$

where $\mathcal{A} = \{A_1, \cdots, A_m\}$ is a set of inferred results among all the affective categories in $\mathcal{Y}$; $A_k \in [0, 1]$ is the probability score indicating whether the corresponding image $p \in V$ represents the affects k.

3. PROPOSED METHOD

3.1 Feature extraction

Image color features contain five dominant colors, color histogram, and so on. The topic on how to accurately describe affects with features is still open. As images from social network are used in this paper, we not only utilize the color features, but also make advantage of the social correlation among images. All the features are summarized in Table 1. Moreover, by comparing the precision of prediction, we evaluate the effects of different features which are shown in Table 3 in Subsection 4.2.

Table 1: Summary of all features.

Type	Name	Short description
Color	Dominant colors	the index of each image is based on the combination of five colors
	HSV	numbering of the HSV feature in the 108 sections, where HSV space is divided by twelve partitions along Hue dimension, and three equal partitions along the other two dimensions
	Saturation, brightness	mean and standard deviation of saturation and brightness
	Pleasure, Arousal, Dominance	one type of affective coordinates calculated by brightness and saturation
	Hue	mean hue, angular dispersion, saturation weighted and without saturation
	Color names	the basic color with top five probability
Social	Uploaded time	the Unix time stamp value when image was uploaded
	Owner ID	the ID of the image owner

3.2 Prediction Model

As the social correlation between images are hard to be modeled by classic classifiers such as SVM, we use a partially-labeled factor graph model(PFG), which was first proposed in [8], for learning and predicting image affects. According to the theory of factor graph model [1], all images uploaded by users can be formalized as variables and observation factor functions in a factor graph. Each image p_i uploaded by users can be mainly described as one affective category which can be mapped as a *affective node* n_i in the PFG model. We denote the labels of affective nodes as $Y = \{y_1, \ldots, y_M\}$ where y_i is a hidden variable associated with n_i. The affects in G are partially labeled, and can be divided into two subset Y^L and Y^U which represent the labeled and unlabeled affects. For each affective node n_i, we define the affective attributes into a vector $\mathbf{x}_i$. Relationships between the images constitute the correlations between hidden variables. Then a factor graph model can be constructed accordingly.

Corresponding to the two intuitions which define two factors are as below:

- **Attribute factor**: $f(y_i, \mathbf{x}_i)$ represents the posterior probability of the affects y_i given by the attribute vector of image n_i
- **Correlation factor**: $g(y_i, N(y_i))$ denotes the correlation among the relationships, where $N(y_i)$ is the set of correlated relationships to y_i.

Given a network $G = (V^L, V^U, E, \mathbf{X})$, we can define the joint distribution over *Y* as

$$P(Y|G) = \prod_i f(y_i, \mathbf{x}_i) g(y_i, N(y_i)) \qquad (1)$$

The two factors can be instantiated in different ways. In this paper, we give a general definition for them. For attribute factor $f(y_i, \mathbf{x}_i)$, we define it using a exponential-linear function:

$$f(y_i, \mathbf{x}_i) = \frac{1}{Z_\alpha} \exp\left\{\alpha^T \cdot \mathbf{x}_i\right\} \qquad (2)$$

where α is a weighting vector of $\mathbf{X}$; Z_α is a normalization factor.

The correlation factor can be naturally modeled in a Markov random field. Thus, by the fundamental theorem of random fields, the definition of correlation can be defined as:

$$g(y_i, N(y_i)) = \frac{1}{Z_\beta} \exp\left\{\sum_{y_j \in N(y_i)} \beta_{ij} \cdot h_{ij}(y_i, y_j)\right\} \qquad (3)$$

where $h_{ij}(y_i, y_j)$ is a feature function that captures the correlation between affective nodes n_i and n_j; β is the weight of this function; and Z_β is also a normalization factor.

Finally, we can redefine the following joint distribution

$$P(Y|G) = \frac{1}{Z}\exp\left\{\sum_{y_i \in Y}\left[\alpha^T \cdot \mathbf{x}_i + \sum_{y_j \in N(y_i)} \beta_{ij} h_{ij}(y_i, y_j)\right]\right\} \quad (4)$$

where $Z = Z_\alpha Z_\beta$ is a normalization factor.

Learning the predictive model is to estimate a parameters configuration $\theta = (\{\alpha\}, \{\beta\})$ from the partially-labeled data set, and to maximize the log-likelihood objective function $\mathcal{O} = \log P(Y|G)$, i.e. $\theta^* = \arg\max \mathcal{O}(\theta)$.

3.3 Model learning

Now, we turn to address the problem of estimating the remaining free θ and inferring image affects once the parameter values have been learnt. Specifically, we first write the gradient of each parameter with regard to the objective function:

$$\frac{\partial \mathcal{O}(\theta)}{\partial \alpha_j} = \mathbb{E}[f_j(y_{ij}, x_{ij})] - \mathbb{E}_{P_{\alpha_j}(y_i|x_{ij},G)}[f_j(y_{ij}, x_{ij})] \quad (5)$$

where $\mathbb{E}[f_j(y_{ij}, x_{ij})]$ is the expectation of feature function $f_j(y_{ij}, x_{ij})$ given by the data distribution and $\mathbb{E}_{P_{\alpha_j}(y_i|x_{ij},G)}[f_j(y_{ij}, x_{ij})]$ is the expectation of feature function $f_j(y_{ij}, x_{ij})$ under the distribution $P_{\alpha_j}(y_i|x_{ij}, G)$ given by the estimated model. Similar gradients can be derived for parameter β. Then we update the parameters by $\theta_j^{new} = \theta_j^{old} + \eta \cdot \frac{\mathcal{O}(\theta)}{\partial\theta}$.

Given the observed value $\mathbf{x}$ and the learned parameters θ, the inference task is to find the most likely $\mathbf{y}$, as follows

$$\mathbf{y} = \arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}, \theta) \quad (6)$$

Finally, after the learning process, all unlabeled affects in the factor graph will be assigned with the label which can produce the maximal probability.

4. EXPERIMENTAL RESULTS

In this section, we first describe our experimental setup, then present the performance of the proposed method and the comparison methods. Next, we give several analysis and discussions. Finally, we present some qualitative case studies to further demonstrate the effectiveness of the proposed method.

4.1 Experimental Setup

Dataset. We evaluate the performance of affective prediction on a large image dataset which are downloaded from Flickr. The dataset contains 23,257 randomly downloaded images spanning from 2004 to 2012. More specifically, we use Kobayashi's 16 affective categories as keywords to search images in Flickr. If an image's labels or the author's comments contain one affective category, we say the image is associated with the affect. We also extract user relationships.

Evaluation Metrics. In the evaluation, we perform five-fold cross validation. We quantitatively evaluate the performance of inferring affects in term of *Precision*, *Recall*, *F1-Measure*.

One objective of the evaluation is to justify that the social network information can help infer authors' affects from images. For this purpose, we define a baseline method using Support Vector Machine (SVM), a widely used classification model. In SVM, we considered all attribute features $\mathbf{x}_i$ defined in our model, but did not consider the correlation feature (defined based on the social network information).

Table 2: Performance of affects mining with different methods on Flickr data set(%).

Affects	Precision		Recall		F1-Measure	
	SVM	PFG	SVM	PFG	SVM	PFG
pretty	1.76	**85.71**	3.61	**14.46**	2.37	**24.74**
casual	15.30	**38.70**	12.81	**28.60**	13.95	**32.89**
dynamic	5.22	**35.42**	**20.87**	14.78	8.35	**20.86**
luxurious	6.72	**63.77**	4.71	**23.04**	5.54	**33.85**
wild	8.00	**41.67**	5.19	**19.48**	6.30	**26.55**
romantic	9.30	**63.75**	1.85	**23.61**	3.09	**34.46**
natural	16.92	**18.47**	24.45	**69.82**	20.00	**29.22**
elegant	7.34	**27.66**	5.63	**27.46**	6.37	**27.56**
old-fashioned	0.00	**79.25**	0.00	**24.28**	0.00	**37.17**
dapper	6.12	**60.00**	5.94	**17.82**	6.03	**27.48**
dignified	0.00	**41.79**	0.00	**18.79**	0.00	**25.93**
formal	2.78	**40.63**	3.39	**22.03**	3.05	**28.57**
chic	7.19	**34.12**	3.89	**28.02**	5.05	**30.77**
clear	2.63	**45.12**	0.68	**25.17**	1.08	**32.31**
jaunty	6.34	**74.29**	6.52	**18.84**	6.43	**30.06**
modern	5.76	**30.30**	3.96	**14.85**	4.69	**19.93**
Average	6.34	**48.79**	6.47	**24.44**	5.77	**28.90**

Table 3: Feature contribution analysis(%).

Feature used	Precision	Recall	F1-Measure
All	48.79	24.44	28.90
Five dominant colors	47.76	22.38	26.62
HSV	39.91	23.05	26.23
Saturation and brightness	42.99	22.31	26.07
Pleasure, Arousal, Dominance	40.52	22.94	26.48
Hue	42.66	22.44	26.13
Color names	45.30	22.60	26.59
Network Correlation	40.12	16.52	22.35
Five dominant colors & Color names	45.94	23.16	26.89
Five dominant colors & Hue & Color names	40.14	19.89	22.26
Except Five dominant colors & HSV	46.77	23.17	27.07
Except Five dominant colors	41.03	18.81	21.15

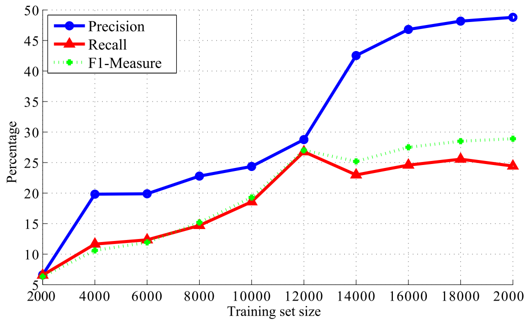

Figure 1: Effect of the size of training set in initialization(%).

4.2 Results and Analysis

Performance comparison. Table 2 shows the performance of the evaluated methods. The proposed method (PFG) shows clearly better performance than the other method. On average, PFG achieves a 17.97–42.45% improvement compared with SVM. The result demonstrates that the social network information is very helpful in our problem. For example, without considering the network information, the SVM cannot correctly predict anyone of the "old-fashioned" category, while our model incorporating the network information improve the accuracy to 37.17%. This confirms the effectiveness of the proposed PFG model.

Feature contribution analysis. We now analyze how different features can help infer image affects. We first use all the features,

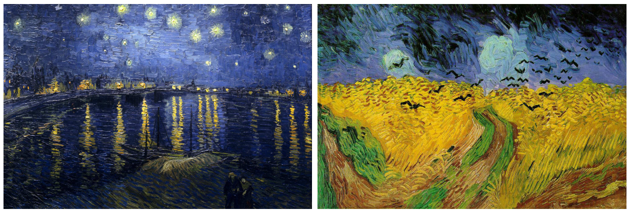

Figure 2: Van Gogh's mood. Left: "Starry Night over the Rhone". The top three predicted categories by our model are casual, modern, and chic. Right: "Wheat field". The top three predicted categories are casual, dapper, chic.

and then remove each kind of features out of the PFG model and evaluate the performance reduction (shown in Table 3). We can see that all the features have contributions to infer affects. Furthermore, HSV feature is most useful in terms of *Precision* which achieves a 8.88% reduction, while saturation and brightness features are the most helpful features in terms of *Recall* and *F1-Measure*. More importantly, the analysis confirms that the network information is one of the most important features. Without considering the network information, the F1-Measure drops 6.55%.

Effect of the size of training set in initialization. Inference accuracy depends on the size of training set for the initialization. A small number might result in high precision but low recall, while a large number might mean higher recall but would hurt the precision. Figure 1 shows how the average performance changes by varying the size of the training set. When the size of the training set is larger than 16,000, *Precision* grows slowly, which indicates the rationality of using 20,000 images as training set in our experiments.

4.3 Demonstrations

Inferring van Gogh's mood. We choose two typical van Gogh's paintings: "Starry Night over the Rhone" (Figure 2 left) and "Wheatfield with Crows" (Figure 2 right). We use our proposed model to infer van Gogh's mood from these two paintings. The results reflect the common affective cognition on these two paintings. For the first painting, the top three prediction categories are casual (probability: 19.3%), modern (15.03%), and chic (10.94%). The comments from *vangoghgallery* on this painting is quiet, rational, and stylish, which are all included in the semantic concepts of these three categories. For the second painting, the top three prediction categories are casual (19.09%), dapper (13.17%), chic (12.33%). The semantic concepts in casual and chic seem to contradict with the semantic concepts in dapper, such as carefree vs. awe-inspiring, or animated vs. bleak. This is the last painting before van Gogh's death, the comments from *vangoghgallery* on this painting is heavy, gloom, and insecure. These results indicate that our prediction results are almost consistent with the comments.

Inferring affects around special events. Here we download images from Flickr around Thanksgiving 2011, and use our model to predict the affective category of each image. Figure 3 shows affective distributions before and during Thanksgiving, with each containing 7,354 and 7,132 images respectively. We visualize the prediction results in Kobayashi's Color Image Scale. For each category, we use the typical five dominant colors to colorize its corresponding circle. And the area of a circle indicates the total number of images belong to this category. Before Thanksgiving, the public affects distribute normally in each category. During Thanksgiving Holiday, the public affects significantly concentrate on casual, the semantic concepts of which are cheerful, happy, free, friendly, enjoyable, etc. That is a quite interesting but rational result.

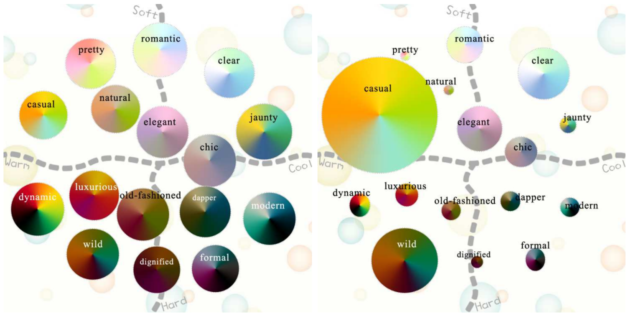

Figure 3: Affects around Thanksgiving 2011 inferred by images on Flickr. Left: affective distribution before Thanksgiving; Right: affective distribution after Thanksgiving. During Thanksgiving Holiday, the detected public affects are mainly on casual, which indicate cheerful, happy, free, and friendly.

5. CONCLUSION

In this paper, we study the problem of inferring affects from images in social networks. We first experimentally analyze features of color compositions that reflect the human's affects. And then, we propose a partially-labeled factor graph (PFG) model for inferring affective information on large scale images in social networks. Experiments demonstrate the effectiveness of the proposed model. As to the future work, we are planning to incorporate other features such as shapes for further improving the inferring accuracy.

Acknowledgements. This work is supported by the National Basic Research Program (973 Program) of China(2011CB302201), and Science Foundation of China (61003094, 90820304, 61073073), and also funded by Tsinghua National Laboratory for Information Science and Technology(TNList)Cross-discipline Foundation. Jie Tang is supported by Chinese National Key Foundation Research (No. 60933013) and National Significant Science and Technology Projects of China under Grant(No.2011ZX01042-001-002).

6. REFERENCES

[1] B. Frey and D. Dueck. Mixture modeling by affinity propagation. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *NIPS*, pages 379–386, 2006.

[2] S. Kobayashi. *Art of Color Combinations*. Kodansha International, 1995.

[3] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *Multimedia*, pages 83–92. ACM, 2010.

[4] L. Ou, M. Luo, A. Woodcock, and A. Wright. A study of colour emotion and colour preference. *Color Research & Application*, 29(3&4):232–240&292–298, 2004.

[5] Y. Shin and E. Kim. Affective prediction in photographic images using probabilistic affective model. In *ICIVR*, pages 390–397. ACM, 2010.

[6] M. Solli and R. Lenz. Color semantics for image indexing. In *ECCGIV*, 2010.

[7] J. Tanaka, D. Weiskopf, and P. Williams. The role of color in high-level vision. *Trends in cognitive sciences*, 5(5):211–215, 2001.

[8] W. Tang, H. Zhuang, and J. Tang. Learning to infer social ties in large networks. In *ECML/PKDD'11*, pages 381–397, 2011.