Emotionally Representative Image Discovery for Social Events

Yun Yang[1], Peng Cui[1], H. Vicky Zhao[2], Wenwu Zhu[1], Shiqiang Yang[1]
[1]Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science, Tsinghua University. Beijing, China
[2]ECE Department, University of Alberta, Edmonton, AB Canada T6G 2V4
yangyun11@mails.tsinghua.edu.cn, vzhao@ece.ualberta.ca,
{cuip, wwzhu, yangshq}@tsinghua.edu.cn

ABSTRACT

With the emerging social networks, images have become a major medium for emotion delivery in social events due to their infectious and vivid characteristics. Discovering the emotionally representative images can help people intuitively understand the emotional aspects of social events. Prior works focus on finding the most visually representative images for the target queries or social events. However, the emotionally representative image should not only visually relevant with the social event, but also has a strong emotional appeal among people. In this paper, we propose an emotionally representative image discovery framework by jointly considering textual, visual and social factors. In particular, we build a hybrid link graph for images of each social event, where the weight of each link is measured by textual emotion information, visual similarity and social similarity. Then we propose the **V**isual-**S**ocial-**T**extual Rank (*VSTRank*) algorithm to calculate the importance score for each image, so that the emotionally representative images can be discovered under the constraint of textual, visual and social representativeness. To evaluate the effectiveness of our approach, we conduct a series of experiments with 15 social events extracted from real social media dataset, and evaluate the proposed method with both quantitative criterions and user study.

Categories and Subject Descriptors

J.4 [**Computer Application**]: Social and Behavioral Sciences

Keywords

Emotionally representative image, textual emotion information, visual similarity, social similarity

1. INTRODUCTION

Nowadays the social networks, as major platforms for communication and information exchange, provide a rich repository of people's opinions and emotions about a vast spectrum of social events [28]. As the major medium appeared in social networks, images have played important roles in delivering emotions of social events due to their infectious and vivid characteristics. Different from texts, images have shown variety of styles in social networks, photographic records, image morphing or cartoonlization, which help describe the event in a simple and clear way. For example, Fig. 1 shows the positive and negative representative images about three different topics, from which people can intuitively understand the emotional aspects of the social events. However, in recent years, with the increasing use of digital photography by the general public, the number of images has exploded into yet unseen numbers. These rapidly growing digital repositories create an urgent need for effective ways of discovering the emotionally representative images to make social events clear at a glance from emotional perspective.

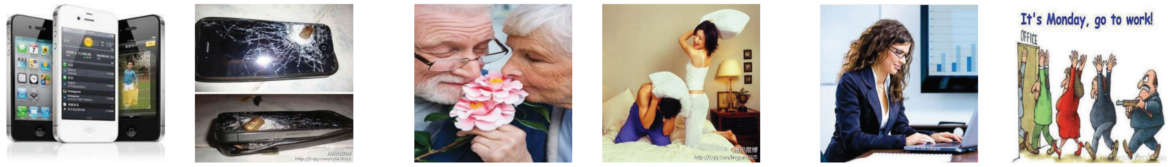
So far textual sentiment analysis has been well developed in areas including opinion mining [22], stock market prediction [2] and event analysis [14]. In contrast, sentiment analysis from visual perspective is still in its infancy. Many works have contributed to the development of visual sentiment analysis, including the analysis of aesthetics [5, 18], interestingness [11] and affect or emotions [3, 12, 17] of images. However, these works basically aim to find the sentiment of the target media, much like a binary classification. Different images, even belonging to the same emotion category, have different infections to people. The ***representative images*** (the images agreed by most people to convey a certain emotion for a certain event) are the most effective ones when conveying the emotions. Prior works focus on finding the most visually representative images [13] for the target queries or social events. However, the emotionally representative image should not only be visually relevant with the social event, but also has a strong emotional appeal among people.

In the context of social networks, the work of discovering the emotionally representative images is challenging due to the following reasons:

- **Low interpretability of visual features.** Image semantics includes several levels. Emotional semantics lies on the highest level of abstract semantics. However, visual features themselves, as the representation of images, cannot well convey the emotional semantics of users due to the limitation of low interpretability.

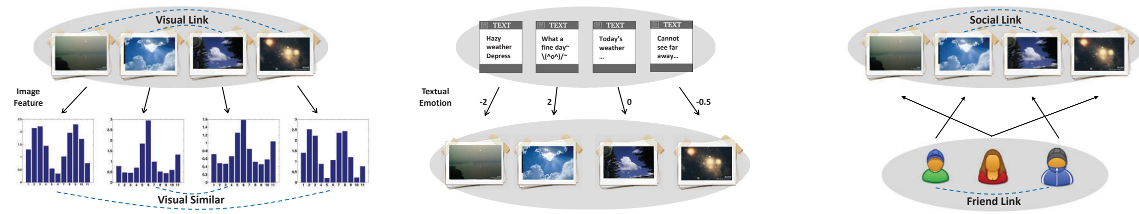
ICMR'14, April 01 - 04, 2014, Glasgow, United Kingdom.

http://dx.doi.org/10.1145/2578726.2578749.



(a) iphone (b) marriage (c) work

Figure 1: The positive (left) and negative (right) representative images about three social events.



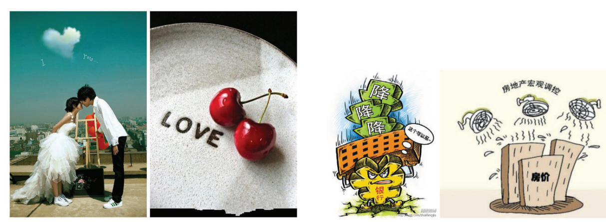
(a) (b) (c)

Figure 2: Images in social networks are connected with different kinds of links. (a), the visual links which connect images with similar visual features. (b), the textual emotion links between images and their corresponding texts. (c), the same user who post two images or users who are friends link the images in social perspective.

- **The variety of styles of images in social networks.** In social networks, the variety of styles of images makes semantic gap problem more challenging since more and more emotionally similar images appear less similar in visual content.

As Fig. 2 shows, images belong to a hybrid link graph, where they have more than one links in social networks. Besides the visual links (Fig. 2 (a)) which connect images with similar visual features, the textual emotion links (Fig. 2 (b)) between images and their corresponding texts can represent the images' textual sentiment, which help address the low interpretability problem of visual features. In the meantime, though images in social networks appear in a variety of styles, users who post or share images link them together in social perspective (Fig. 2 (c)). Generally, users who are friends in social networks usually hold similar emotion towards the same social event. Therefore, their posted images in social networks are likely to convey similar emotions. Fig. 3 shows two sets of images with social links. In Fig. 3 (a), the two images are less similar in visual content, however, they express the same emotion toward "love". In Fig. 3 (b), the two cartoons both convey the negative attitude about "the price of the house". Even though these images have different visual contents, we can link them via a social links, which help bridge the semantic gap between images and users.

To address the above challenges in discovering emotionally representative images in social networks, we propose an emotionally representative image discovery framework by jointly considering textual, visual and social factors. In particular, we build a hybrid link graph for images of each social event, where the weight of each link is measured by textual emotion information, visual similarity and social similarity. Then we propose the **V**isual-**S**ocial-**T**extual Rank (*VSTRank*) algorithm to calculate the importance score of each image, so that the emotionally representative images can be discovered under the constraint of textual, visual and social representativeness.



(a) (b)

Figure 3: Images with social link, which are visually less similar, but share the same emotion.

The contributions of our work can be summarized as follows:

- To the best of our knowledge, we are the first to address the problem of discovering emotionally representative images for social events in social networks, which help people intuitively understand the emotional aspects of social events.

- We propose a novel emotionally representative image discovery framework, which identifies emotionally representative images under the textual, visual and social representativeness constraints.

- The *VSTRank* algorithm is proposed to rank the images for the social events from the emotional aspect. By performing the simple yet efficient learning method, we can quantify the different roles visual, social and textual information play in discovering the emotionally representative images for social events.

The remainder of the paper is organized as follows. Related works are summarized in Section 2. In section 3, our proposed emotionally representative image discovery framework will be discussed in details, including representative image discovery methods from textual, visual and social aspects, respectively, and the *VSTRank* algorithm to discovery the emotionally representative images under the constraint of the three factors. We evaluate the performance of our proposed framework and compare with the state-of-the-art methods in Section 4. Finally, conclusion is drawn in Section 5.

2. RELATED WORK

Since our work involves sentiment analysis and representative image discovery, we review related work in these two areas.

2.1 Sentiment Analysis

Recent years have witnessed an increasing number of research works on sentiment analysis from textual [22] and visual [3, 12, 17] aspects. The basic idea of these methods is to build a sophisticated feature space, which can effectively represent the sentiment status of the texts and images. One of the representative ways to perform textual sentiment analysis is the lexicon-based method [6, 26], which relies on the sentiment dictionary to determine the general sentiment polarity of a given document. In the meantime, supervised learning methods with star ratings [23], distant supervision [9], social relations [10] in social networks have also become the polarity signals in sentiment analysis area. Semantic and concept learning approaches based on visual features [7, 17, 21] analyze image sentiments without employing textual information. Information from multiple domains, including textual and visual information, were preliminary combined together in [29]. While, all these works are discovering what emotions the media are conveying, which is different to our target, discovering the emotionally representative images for social events.

2.2 Representative Image Discovery

Recent years, there is an increasing research and commercial interests in building effective search mechanisms for representative image discovery.

The work in [25] introduces the methodology for automatically ranking and classifying photos according to their attractiveness for folksonomy members. In work [8], various visual features are investigated on their applicability to attractiveness estimation of web images for attractive images ranking. Besides attractiveness image discovery, [13] casts the image-ranking problem into the task of identifying "authority" nodes on an inferred visual similarity graph and analyzes the visual link structures among images to discovery the representative images as the ones answer the image-queries well. [16] improves the relevance between returned representative images and user intentions with social and visual factors in social media platforms. [19, 24] use the visual features and social streams information to predict the popular video shots or videos. [30] searches the social media "entities" with a unified, multi-level and correlative entity graph and uses it into various applications. Above works focus on discovering the visually representative images, whether attractive or most relevant to the queries or the videos. However, discovering the emotionally representative images should not only be visually representative, but also have strong emotional appeal among people.

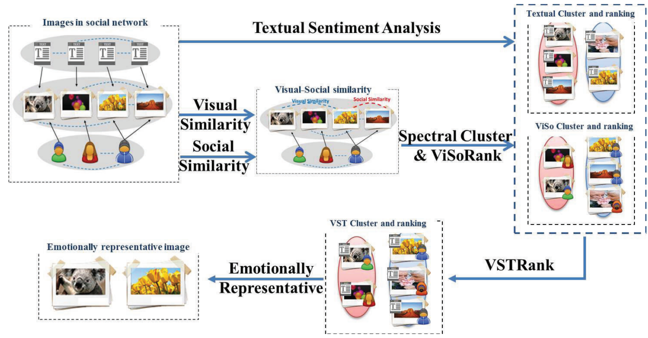


Figure 4: Framework of emotionally representative image discovery for social events.

3. EMOTIONALLY REPRESENTATIVE IMAGE DISCOVERY

In this section, we will present our proposed emotionally representative image discovery framework. Then we will discuss in details how to cluster and rank images based on textual, visual and social factors.

3.1 Framework

As Fig. 2 shows, images in social networks are not only visually linked together, their corresponding textual information conveys emotions, which link the images together. Besides, users who post the microblogs link the images in social perspective. All of these generate a hybrid link graph for images of each social event. Given this hybrid graph, we propose the emotionally representative image discovery framework in which textual, visual and social factors are all considered.

Fig. 4 gives the basic idea of our proposed framework. First, we use lexicon-based sentiment method [6] to analyze the corresponding textual sentiment, which are used to cluster and rank images from textual aspect. Then, to link images in social networks that have similar emotions but different visual contents, we combine the social links with visual similarity between images, constructing a visual-social similarity matrix that quantifies images' similarities from both visual and social perspectives. Spectral clustering [27] is then used to cluster images into two categories, one is positive and the other is negative. Inside each category, we propose ViSoRank algorithm to identify representative images on the inferred visual-social similarity graph. Finally, with the cluster and ranking results from textual and visual-social perspectives, we propose the *VSTRank* algorithm to combine them together to discover the emotionally representative images for social events.

3.2 Image Cluster and Ranking Based on Textual Emotions

There have been extensive researches on textual emotion analysis. Online sentiment dictionary [6] has been widely used in textual sentiment analysis. In our work, we use the dictionary in [1] with predefined emotion polarity of each word w, with which image i_k's corresponding texts T_{i_k}'s emotions can be mined:

$$E(T_{i_k}) = \sum_{w \in T_{i_k}} E(w) \quad (1)$$

where $E(w)$ is the predefined emotion polarity with weight for word w in the sentiment dictionary.

With the mined images' corresponding textual emotions in one social event, we classify the images and rank them based on the textual emotion polarities and their emotion values. Hence, we cluster the images into two categories and rank them within each group.

$$G_{T_p} = \{i_{T_1}, i_{T_2}, ..., i_{T_m}\}$$

$$G_{T_n} = \{i_{T_n}, i_{T_{n-1}}, ..., i_{T_{m+1}}\}$$

where i_{T_k} is the textual information T_k's corresponding image. In the positive group G_{T_p}, for $\forall i_k \in G_{T_p}$, we have $E(T_{i_k}) > 0$. Besides, for image i_{T_k}'s rank $TR(i_{T_k})$ and image i_{T_j}'s rank $TR(i_{T_j})$, if $TR(i_{T_k}) < TR(i_{T_j})$, we have $E(T_{i_k}) > E(T_{i_j})$, which means the higher the image ranks, the larger the corresponding textual emotion value is. While for the negative group, images' corresponding textual emotion value is negative and the higher the image ranks, the smaller the corresponding textual emotion value is.

3.3 ViSoRank

In this section, we will cluster images and discover the representative ones based on their visual and social information for a social event.

3.3.1 Visual Similarity

Images belonging to the same category are supposed to have some similarities in visual aspect. For example, most images with dark colors give people the depression feeling, while bright colors give people positive feelings. Besides colors, shapes, composition also play important roles in image affective analysis.

Following the feature design in state-of-the-art visual classification system such as Object-Bank [15], we include several visual features: the Color Histogram extracted from the RGB color channels, the GIST descriptor [20] that has been shown to be useful for detecting scenes, the Local Binary Pattern (LBP) descriptor for detecting textures and faces and a Bag-of-Words quantized descriptor. With these extracted features, each image i_k can be presented as a vector Im_{i_k} that describes its visual content.

Traditional Euclidean distance is used to measure two images' visual distance. To smooth the distance between two images, exponential decay model is used to measure the similarity of two images' visual contents.

$$S_v(i_k, i_m) = e^{-distance(Im_{i_k}, Im_{i_m})} \quad (2)$$

Based on Eqs. (2), the larger the distance of two images is, the less similarity they appear.

3.3.2 Social Similarity

To link images from social aspect, for user U_{i_k} who posts or shares image i_k and user U_{i_m} who posts or shares image i_m, we call that it has a social link from i_k to i_m if Ui_k is a follower of Ui_m or the two images are shared or posted by the same user, which means image i_k has a probability to be emotionally similar to image i_m from social aspect. An image may have more than one social links to other images. Thus, we infer that the image shares the same probability to all the linked images.

As the fact that different users have different social strengths [31] or social influence [4]. We quantify image i_k and i_m's social similarity $S_s(i_k, i_m)$ as:

$$S_s(i_k, i_m) = \begin{cases} \frac{1}{|SL(i_k)|}, & i_m \in SL(i_k) \\ 0, & \text{otherwise} \end{cases}. \quad (3)$$

where $SL(i_k)$ is the image set to which image i_k is socially linked.

3.3.3 Spectral Clustering and ViSoRank

To present two images' similarity more comprehensively, we combine their visual similarity and social similarity together, we called it the *ViSo similarity*:

$$S(i_k, i_m) = S_v(i_k, i_m) + S_s(i_k, i_m) \quad (4)$$

In this way, we construct a similarity graph $G = (V, E)$, where each vertex v_i in this graph represents an image i. The edge between two images i_k, i_m is weighted by the ViSo similarity $S(i_k, i_m)$. We use spectral clustering [27] to find a partition of the graph such that the edges between different groups have very low weights (which means that images in different clusters are dissimilar from each other) and the edges within a group have high weights (which means that images within the same cluster are similar to each other).

Similar to VisualRank [13], we propose a *ViSoRank* algorithm that employs the Random Walk to rank images based on the ViSo similarities (visual similarities and social similarities) among the images. The basic idea of ViSoRank is that, images conveying the same emotion for social events have high probabilities to be visually and socially similar to each other. In particular, if image i_k has a large ViSo similarity to image i_m, then there is a high probability that the two images belong to the same emotion category. Intuitively, images with large in-degrees from those belonging to the same category will be the representative one for the category. At the same time, the outgoing links from the "representative" image are important as the image itself is important in the category.

ViSoRank (VSR) is iteratively defined as following:

$$VSR^{t+1} = dS^* \times VSR^t + (1-d)p \quad (5)$$

where S^* is the column normalized adjacency matrix S, where $S(i_k, i_m)$ measures the ViSo similarity between image i_k and i_m. In each ranking step, ranks of images are refined using their ViSo similarities. Following [13], p is the damping vector and d is the damping factor.

In this similarity graph G, spectral clustering and ViSoRank are used to cluster images and rank them within each emotion category. Finally, we get two categories with ranked images for the social event:

$$G_{VS_p} = \{i_{VS_1}, i_{VS_2}, ..., i_{VS_k}\}$$

$$G_{VS_n} = \{i_{VS_n}, i_{VS_{n-1}}, ..., i_{VS_{k+1}}\}$$

where in the group G_{VS_p} and group G_{VS_p}, the image with the highest rank VSR is the one most similar to other images in the corresponding group in visual and social perspectives.

3.4 Visual-Social-Textual Rank

In the previous two subsections, we use images' corresponding textual information, ViSo similarity between images to make emotional classification and rank the images within each group separately. To discovery the representative images comprehensively from the three aspects, we propose the *Visual-Social-Textual Rank* (*VSTRank*) algorithm, in which images' textual rank and ViSo rank are linearly combined. To measure the roles textual information and ViSo similarity play in the emotional classification, for each social event, we define four parameters: α^0, α^1 measure the weights of textual information for negative and positive emotion classification, respectively, and β^0, β^1 are the weights of ViSo similarity for negative and positive emotion classification, respectively.

For the simplicity, we use the textual information and ViSo similarity's cluster precision to quantize the weights:

$$\alpha^1 = \frac{|\{i|i \in Pos, i \in G_{T_p}\}|}{|G_{T_p}|} \tag{6}$$

where the Pos is the set with images in training set with positive emotions. Similarly, $\alpha^0, \beta^0, \beta^1$ can be calculated in the same way.

The above calculated cluster precision can be seen as the textual information and ViSo similarity's confidences in clustering images for the target social event. For example, if α^1 is relatively high, it means textual information play an important role in cluster images for positive emotions. For the images belonging to the same emotion category from two methods (textual cluster and ViSoRank), we can simply combine the images' ranks from the two methods with corresponding parameters to get the final image representative ranks. While for the images belong to the different categories from two methods, the corresponding parameters are used to decide which category the images should belong to. As for the ranking, the ranks from the misclassified method can be used adversely. For example, the orders from one category can be seen as the inverted orders from the other category. Besides the linear combination, a penalty η is applied to the misclassified result. The whole ranking decision can be summarized as Algorithm 1.

Algorithm 1: *VSTRank* Algorithm

Input: For each social event, four cluster precision parameters $\alpha^1, \alpha^0, \beta^1, \beta^0$; Intra-class normalized ranks for each image using two methods TR, VSR; and misclassification penalty η

Output: Final ranks $VSTR$ within each class

```
for Each social event do
    for Each image i do
        if image i belongs to the same category G_x then
            G_x ← i;
            VSTR_i^x = α^x TR_i^x + β^x VSR_i^x;
        else
            //image i belongs to G_x using textual
            information and G_y using ViSo similarity
            if α^x > β^y then
                G_x ← i;
                VSTR_i^x = α^x TR_i^x + β^y (1 − VSR_i^y) + η;
            else
                G_y ← i;
                VSTR_i^y = α^x (1 − TR_i^x) + β^y VSR_i^y + η;
return rank VSTR;
```

Algorithm 1 gives us the final ranks within each group for each image, which present the images' representative scores by jointly considering textual information, visual similarities and social similarities together. With the ranking scores, we can discover the representative images from positive and negative emotion groups.

4. EXPERIMENTAL RESULTS

In this section, we first describe our experimental setup, and then show experimental results of our proposed algorithm to evaluate its effectiveness with both quantitative criterions and user study.

4.1 Experimental Setup

Dataset: We perform our experiment on real social media network, Tencent Weibo, which is one of the biggest microblogging service in China with more than 780 million users. We collect all the microblogs between November 20 and November 25, 2011, as well as the relationship between the users. 15 hot social events are selected during the 6 days, which include 21,783 microblogs with images and 15,487 users. The detailed social events and their corresponding image numbers can be seen in Table 1.

Ground Truth and Performance Evaluation Metrics: In our dataset, some images are irrelative to their event, which is the noise to our algorithm. We first use visual similarity to make the filtration. Then, for the rest, to train the parameters of our *VSTRank* algorithm, we need the ground truth of the emotions conveyed by the images. Deciding what emotion the image is conveying is relatively objective. Therefore, we let users tag the images' emotions. Each image has 3 people to make the emotional annotation and the majority vote will become the image's emotional tag.

For the estimation of the ranking qualities, generally, users pay more attentions to the images with high ranks. Therefore we only compare the Top 5 images' representative qualities from different methods. Judging whether an image is emotionally representative is subjective. To make our evaluation fair, for each topic, we ask at least 3 people to do the user study. Each image will be given a score ranging from -2 to 2 to quantize the representativeness. 2 means the image is classified into the correct emotion category and also highly representative to the emotion. 0 means we cannot decide what emotion the image is conveying or it is irrelevant to the social event. -2 is opposite to 2, which means the images is representative, while it is misclassified into the opposite emotion category.

4.2 The Performance in Discovering Representative Images

To illustrate the effectiveness of our *VSTRank* algorithm in discovery representative images for social events, besides our proposed *VSTRank* algorithm that jointly considers textual information, visual similarity and social similarity, we also implement the following baseline algorithms for comparison.

Table 1: Social events and corresponding image numbers.

Event	Image No.	Predicament	1089
Weather	3027	Work	1061
Marriage	2344	Relationship	839
Pressure	2253	Challenge	823
Education	2153	Iphone	678
Love	2000	Friendship	670
Life	1888	Reform	667
Government	1820	Economy	471

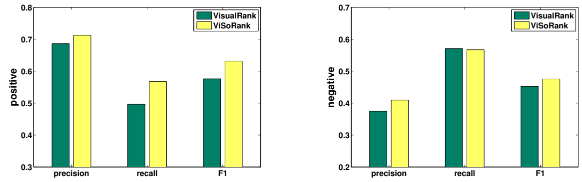
- *VisualRank:* In VisualRank [13], we only use visual similarity between images to classify and rank images from emotional perspective.
- *Textual Sentiment Analysis:* We use traditional lexicon-based method to analyze images' corresponding textual information [26], the images with strong positive or negative emotion from textual information will be discovered as the representative ones.

As our final goal is to discover the representative images, we compare the mean scores of positive and negative representative images from different methods. The results can be seen in Table 2. We can observe that our algorithm, *VSTRank*, achieves the best performance from both positive or negative representative image discovery. With the Top 1's positive score 1.30 and negative score 1.45, we can discover the images which relatively highly represent the emotion of the social event. Besides, in our algorithm, Top 1's score is higher than the average of the Top 3 and Top 5's, which means the Top 1's images are considered to be the most representative image by users. As for the other baseline methods, VisualRank performs better than Textual sentiment analysis in positive image discovery, while the opposite is true for negative image discovery. However, by joint consideration of visual similarity, social similarity and textual information together, our *VSTRank* algorithm achieves the best performance among the three.

4.3 Observations and Insights

In our algorithm, textual information, visual similarity and social similarity are used in our emotionally representative image discovery framework. To gain more insights on how these factors affect the classification and ranking results for different social events, we calculate statistics about VisualRank and textual sentiment analysis mean scores for representative image discovery respectively. For events like "economy", "government" and "reform", the performance of textual sentiment analysis is better than the VisualRank algorithm due to the complicated images and their variety of styles. While for the events concerned more about life, for example, "love" or "marriage", VisualRank on the other hand achieves better performance. In these events, users' emotions can be easily seen through images from the colors, shapes or compositions.

Visual and textual information have their own advantages in classifying and ranking images for different events. In the meantime, the social factor links images with similar emotions but less similar in visual content. To evaluate the performance of social factor in classifying images from emotional aspect. We compare the spectral cluster results



(a) positive cluster (b) negative cluster

Figure 5: Social factor's role in classifying images from emotional aspect.

by purely using visual similarity and by using visual-social similarity between images. The performance of social factor in classifying images from emotional aspect can be seen in Fig. 5. For both types of emotions, social factor helps link the images with similar emotions but visually less similar, which bridges the semantic gap between images and users in a way.

4.4 Representative Cases

Besides the quantitative criterions, we show some typical examples of emotional representative image discovery with three methods to present the effectiveness of our *VSTRank* algorithm in this section.

In Fig. 6, we respectively show "iphone", "work", "life" three social events. For each social event and for each method, we show the Top 5 representative images for both positive and negative emotions. The image with the red box means the image is either irrelevant to the social event or opposite to the specified emotion. Overall, we can observe that our algorithm, *VSTRank*, have fewer wrongly classified images compared to the other two baseline methods.

For "iphone" topic, some users like the design or the technology, while some also complain about the fragile screen. In our algorithm, for the positive images, they all convey the positive attitude about the iphone, while for the negative images, we rank the iphone with broken screen into the Top 2. For "work" topic, some are enthusiasm, while some feel boring or even painful. Compared to the two baselines, our algorithm can discover the correct images with specified emotions. For "life" topic, someone loves it while someone feels it hard. With fewer wrongly classified images in *VSTRank*, we can discover the attractive images about "life" for positive attitude and the negative representative images for negative attitude.

Overall, by joint visual similarity, social similarity and textual information, *VSTRank* algorithm can achieve the best performances in image emotion classification and representative image discovery compared to the state-of-the-art methods.

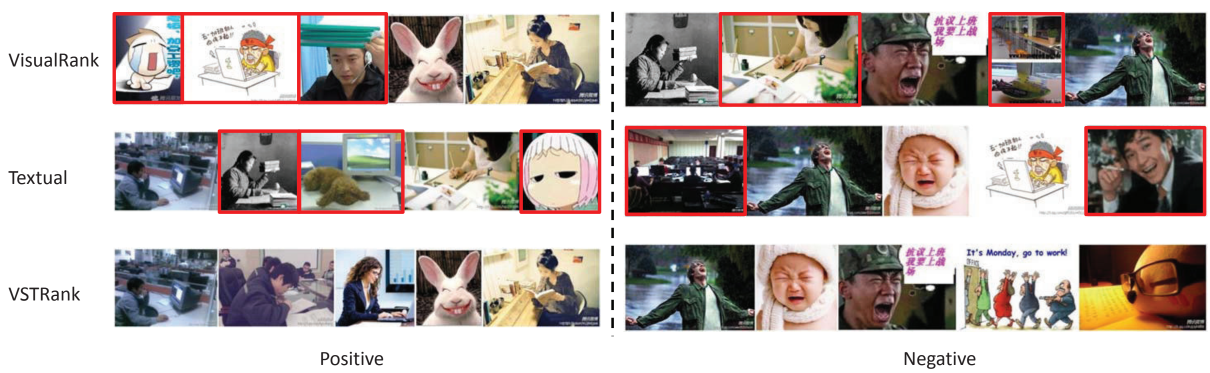
5. CONCLUSION

In this paper, we propose an emotionally representative image discovery framework by jointly considering textual, visual and social factors. Specifically, for each social event, images are linked together by textual information, visual similarity and social similarity in the hybrid link graph. To measure the factors' roles and rank images based on their representativeness, we propose the *VSTRank* algorithm, where the weight of each factor was learned by its classi-

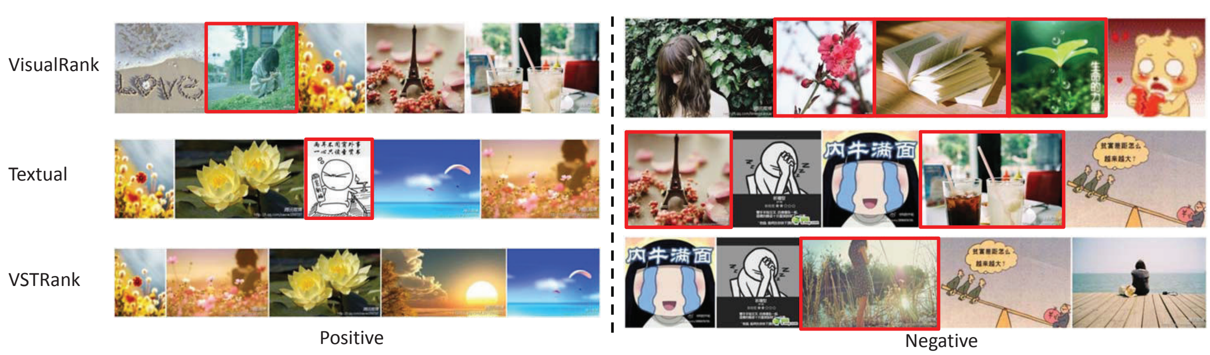
Table 2: Different methods' mean scores for positive and negative representative images.

Emotion	Positive			Negative		
Method	VisualRank	Textual	*VSTRank*	VisualRank	Textual	*VSTRank*
Top 1	0.875	0.70	**1.30**	0.325	0.775	**1.45**
Top 3	0.908	0.658	**1.175**	0.40	0.558	**0.97**
Top 5	0.91	0.65	**1.165**	0.265	0.405	**0.85**

(a) iphone



(b) Work



(c) Life

Figure 6: Show case.

fication precision confidence. With the *VSTRank* algorithm, the emotionally representative images can be discovered under the constraint of textual, visual and social representativeness. We conduct a set of comprehensive experiments to validate the effectiveness of our approach and evaluate the results from both quantitative criterions and user study. The experimental results show that our algorithm can achieve the best performance in representative image discovery compared to the state-of-the-art methods.

6. ACKNOWLEDGMENTS

This work is supported by National Natural Science Foundation of China, No. 61370022, No. 61303075 and No. 61210008; International Science and Technology Cooperation Program of China, No. 2013DFG12870; National Program on Key Basic Research Project, No. 2011CB302206. Thanks for the support of NExT Research Center funded by MDA, Singapore, under the research grant, WBS:R-252-300-001-490.

7. REFERENCES

[1] http://www.datatang.com/data/43216.

[2] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

[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 *Proceedings of the 21st ACM international conference on Multimedia*, pages 223–232. ACM, 2013.

[4] P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun. Who should share what?: item-level social influence prediction for users and posts ranking. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 185–194. ACM, 2011.

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *Computer Vision–ECCV 2006*, pages 288–301. Springer, 2006.

[6] A. Esuli and F. Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422, 2006.

[7] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1778–1785. IEEE, 2009.

[8] B. Geng, L. Yang, C. Xu, X.-S. Hua, and S. Li. The role of attractiveness in web image search. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 63–72. ACM, 2011.

[9] A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, pages 1–12, 2009.

[10] X. Hu, L. Tang, J. Tang, and H. Liu. Exploiting social relations for sentiment analysis in microblogging. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 537–546. ACM, 2013.

[11] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 145–152. IEEE, 2011.

[12] 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 *Proceedings of the 20th ACM international conference on Multimedia*, pages 857–860. ACM, 2012.

[13] Y. Jing and S. Baluja. Visualrank: Applying pagerank to large-scale image search. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(11):1877–1890, 2008.

[14] E. Kim, S. Gilbert, M. J. Edwards, and E. Graeff. Detecting sadness in 140 characters: Sentiment analysis and mourning michael jackson on twitter. *Web Ecology*, 3, 2009.

[15] L.-J. Li, H. Su, L. Fei-Fei, and E. P. Xing. Object bank: A high-level image representation for scene classification & semantic feature sparsification. In *Advances in neural information processing systems*, pages 1378–1386, 2010.

[16] S. Liu, P. Cui, H. Luan, W. Zhu, S. Yang, and Q. Tian. Social-oriented visual image search. *Computer Vision and Image Understanding*, 118:30–39, 2014.

[17] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the international conference on Multimedia*, pages 83–92. ACM, 2010.

[18] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1784–1791. IEEE, 2011.

[19] D. H. Nga and K. Yanai. Large-scale web video shot ranking based on visual features and tag co-occurrence. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 525–528. ACM, 2013.

[20] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175, 2001.

[21] V. Ordonez, G. Kulkarni, and T. L. Berg. Im2text: Describing images using 1 million captioned photographs. In *Advances in Neural Information Processing Systems*, pages 1143–1151, 2011.

[22] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

[23] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics, 2002.

[24] S. Roy, T. Mei, W. Zeng, and S. Li. Towards cross-domain learning for social video popularity prediction. 2013.

[25] J. San Pedro and S. Siersdorfer. Ranking and classifying attractiveness of photos in folksonomies. In *Proceedings of the 18th international conference on World wide web*, pages 771–780. ACM, 2009.

[26] H. Takamura, T. Inui, and M. Okumura. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 133–140. Association for Computational Linguistics, 2005.

[27] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[28] Z. Wang, P. Cui, L. Xie, H. Chen, W. Zhu, and S. Yang. Analyzing social media via event facets. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 1359–1360. ACM, 2012.

[29] Y. Yang, P. Cui, W. Zhu, and S. Yang. User interest and social influence based emotion prediction for individuals. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 785–788. ACM, 2013.

[30] T. Yao, Y. Liu, C.-W. Ngo, and T. Mei. Unified entity search in social media community. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1457–1466. International World Wide Web Conferences Steering Committee, 2013.

[31] J. Zhuang, T. Mei, S. C. Hoi, X.-S. Hua, and S. Li. Modeling social strength in social media community via kernel-based learning. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 113–122. ACM, 2011.