

Sentiment Analysis on Multi-View Social Data

Teng Niu¹, Shiai Zhu¹(✉), Lei Pang², and Abdulmotaleb El Saddik¹

¹ MCRLab, University of Ottawa, Ottawa, Canada
{tniu085,elsaddik}@uottawa.ca, zshiai@gmail.com

² Department of Computer Science, City University of Hong Kong,
Kowloon Tong, Hong Kong
leipang3-c@my.cityu.edu.hk

Abstract. There is an increasing interest in understanding users' attitude or sentiment towards a specific topic (e.g., a brand) from the large repository of opinion-rich data on the Web. While great efforts have been devoted on the single media, either text or image, little attempts are paid for the joint analysis of multi-view data which is becoming a prevalent form in the social media. For example, paired with a short textual message on Twitter, an image is attached. To prompt the research on this interesting and important problem, we introduce a multi-view sentiment analysis dataset (MVSA) including a set of image-text pairs with manual annotations collected from Twitter. The dataset can be utilized as a valuable benchmark for both single-view and multi-view sentiment analysis. With this dataset, many state-of-the-art approaches are evaluated. More importantly, the effectiveness of the correlation between different views is also studied using the widely used fusion strategies and an advanced multi-view feature extraction method. Results of these comprehensive experiments indicate that the performance can be boosted by jointly considering the textual and visual views.

Keywords: Sentiment analysis · Multi-View data · Social media

1 Introduction

Sentiment analysis aims at predicting the attitude or opinion towards an event, topic or product from the user-contributed data (e.g., text, image or video). Many applications can benefit from the automatic identification of sentiment. For example, opinions contained in consumers' comments on the Web can be utilized by companies to improve their products or adjust marketing strategies. Politicians can also understand voters' reactions about their campaigns.

Previous works on sentiment analysis mainly focus on the data collected from review-aggregation resources, such as Internet-based retailers and forum Websites, where feedbacks from Web users are usually stereotyped texts or qualitative scores. Both the lexicon-based approaches which aggregate the sentiments of words and the statistical learning approaches are extensively investigated. While the structured and accurate data can better reflect the users' sentiment,

the collection is relatively small and only covers limited number of entities due to the fact that it is hard to involve users in the data collection. For example, customers may not comment their purchased products on the Web.

The popularity of social media, which is a convenient platform for sharing messages, introduces explosively increasing opinion-rich data freely available on the Web. Thus sentiment analysis on social data has received significant research attention recently. However, understanding social data contributed from uncontrolled Web users is challenging. Firstly, Twitter messages are usually short, and thus lack of strong evidences indicating the users' sentiment. Secondly, the posted messages are almost free-form texts having quite different styles, contents and vocabularies (e.g., abbreviations). Standard lexicon-based methods and statistical learning approaches are faced with the difficulty of sparse representations.

Besides the texts, portable devices integrated with camera have made the acquisition and dissemination of image/video much more convenient. The attractive visual data has been found to be more repostable than text messages [4]. Thus understanding affect in visual content is becoming an active research area. Inspired by the semantic concept detection, a widely used computational framework includes feature extraction and statistical modeling. Various visual features are investigated in the literature, ranging from handcraft low-level features (e.g., GIST and SIFT) [17] to middle-level features [7, 15]. However, sentiment analysis of visual instance is challenging, since affective expression (e.g., "sentiment") is more abstract than the general concepts (e.g., "dog").

While significant progress has been made on sentiment analysis of social data, most of the efforts are devoted on single media type. However, social data usually has multiple views. For example, over 99% of the images posted on Twitter are accompanied with texts [3]. In [17], a corpus analysis indicates that image and text in a message are positively correlated. The challenges in single view analysis are expected to be addressed by jointly considering the multiple views in the social data. To the best of our knowledge, little attempts are paid to the sentiment analysis of multi-view data. One common obstacle is the lack of annotated data for model learning and performance evaluation. To prompt research on this interesting and important problem, we present a multi-view benchmark (MVSA) with rigorous manual annotations. A set of features extracted from single view and multiple views is also provided. In addition, we provide a good baseline using several state-of-the-art sentiment analysis methods and multi-view learning approaches. The main contribution of this work is in establishing a benchmark for sentiment analysis in multi-view social data. In addition, insightful discussions are provided for better understanding the limitations of existing solutions and the potentiality of exploring the correlations between different views through a comprehensive set of experiments.

2 Related Works

2.1 Sentiment Analysis Datasets

Coming up with the extensive research on text sentiment analysis, there is plenty of datasets available on the Web. These datasets are usually constructed for specific domains. For example, 50,000 movie reviews with annotations for positive and negative sentiment are provided in [8]. Recently, as many researchers turned their attentions to more timely and convenient social data, some corresponding datasets are proposed. A widely used one is STS [5], where training set consists of 1.6 million tweets automatically annotated as positive and negative based on emotioncons (e.g., “:”) or “=”)”, and testing set includes 498 manually labeled tweets. While STS is relatively large, the labels are derived from unreliable emotioncons. In [11], a large dataset including manually labeled 20 K tweets is constructed for the annually organized competition in SemEval challenge. In these datasets, each message is attached with one label. However, each tweet may include mixed sentiments. In STS-Gold [12], both message-level and entity-level sentiments are assigned to 2,206 tweets and 58 entities. Besides the datasets for general sentiment analysis, there are other datasets constructed for specific domains or topics, such as Health Care Reform (HCR) dataset [13] including eight subjects and Sanders dataset¹ for four topics.

Comparing to textual data, very few datasets have been built for sentiment analysis on visual instances. In [18], a total of 1,269 Twitter images (ImgTweet) are labeled for testing their method. In [2], accompanying with the proposed emotional middle-level features (Sentibank), a small dataset including 603 multi-view Twitter messages with manual annotations is provided. Another related large-scale dataset is proposed in [6], which nevertheless is constructed for emotion detection on Web videos. To the best of our knowledge, there are no other datasets dedicatedly designed for multi-view sentiment analysis.

2.2 Sentiment Analysis Approaches

We can roughly categorize existing works on text sentiment analysis into two groups: lexicon-based approaches and statistic learning approaches. The former leverages a set of pre-defined opinion words or phrases, each of which is assigned with a score representing its sentiment. Sentiment polarity is the aggregation of opinion values of terms within a piece of text. Statistic learning approaches usually adopt a variety of supervised learning methods with some dedicated textual features. In addition, sophisticated nature language processing (NLP) techniques are developed to address the problems of syntax, negation and irony. These techniques are discussed in a comprehensive survey [9].

As a new and active research area, visual sentiment analysis adopts a similar framework with general concept detection, where sentiment classifiers are trained on visual features (e.g., “GIST”). In [18], a robust feature using deep

¹ <http://www.sananalytics.com/lab>.

neural networks is introduced into sentiment analysis. Similar to the semantic gap in concept detection, there is also an affective gap in sentiment analysis. To narrow down the gap, middle-level features defined on a set of affective atom concepts [20] or emotional Adjective Noun Pairs [2] are investigated.

While great progress has been made on sentiment analysis of textual or visual data, little effort is paid on the multi-view social data. A straightforward way [2, 6] is to fuse features or prediction results generated from different views. However, it fails to represent the correlations shared by the multiple views, and thus losses important information for sentiment analysis. The most related works on learning cross-view or multi-view representations [16] may be helpful to handle this problem. For example, a joint representation of multi-view data is developed using Deep Boltzmann Machine (DBM) in [14]. However, it is still unclear whether these techniques are able to represent the complex sentiment-related context in the multi-view data.

3 The MVSA Dataset

3.1 Data Collection and Annotation

All the image-text pairs in MVSA were collected from the Twitter, which has over 300 million active users and includes 500 million new tweets per day². We adopted a public streaming Twitter API (Twitter4J)³. In order to collect representative tweets, the stream was filtered by using a vocabulary of 406 emotional words⁴. In specific, only the tweets containing the keywords in the message or hashtags were downloaded. The vocabulary includes ten distinct categories (e.g., happiness and depression) covering almost all the felling of human beings. Some emotional words, such as happy and sad, frequently appear in the tweets. To balance the collected data among different categories, we used the keywords roundly and collected at most 100 tweets for one keyword at each round. In addition, the data collection was daily performed at several time slots within one day. We further extracted the image URLs within the messages to download the paired images. Only the text-image tweets with accessible images were kept for annotation.

Annotating sentiments on large set of image-text pairs is difficult, particularly when uncontrolled Web users may post messages without correlations between the image and text. To facilitate the annotation, we developed an interface. Each time, a image-text pair is shown to an annotator, who will assign one of three sentiments (positive, negative and neutral) to the text and image separately. Note that text and image in a message do not necessarily have same sentiment label. The annotations can be used for generating three subsets of data corresponding to text, image and multi-view respectively. Until now, we have received annotations for 4,869 messages. We only include the tweets that receive same

² <https://about.twitter.com/company>.

³ <http://twitter4j.org/en/>.

⁴ <http://www.sba.pdx.edu/faculty/mblake/448/FeelingsList.pdf>.

Table 1. Statistics of manually annotated datasets for tweet sentiment analysis.

Dataset	#Positive	#Negative	#Neutral	Data type
HCR	541	1,381	470	Text
STS	182	177	139	Text
SemEval	5,349	2,186	6,440	Text
STS-Gold	632	1,402	77	Text
Sanders	570	654	2,503	Text
ImgTweet	769	500	—	Image
Sentibank	470	133	—	Text + image
MVSA	1398	724	470	Text + image

labels on both text and image as the final benchmark dataset. All the data and annotations are released for public⁵. Table 1 lists the details of MVSA and several popular public datasets for tweet sentiment analysis. Comparing to other datasets, MVSA is already the largest dataset for multi-view data analysis. We will keep increasing the dataset by including more up-to-date messages, and the annotations will be regularly released.

3.2 Data Analysis

We have observed that there are inconsistent sentiments represented in user posted image and the corresponding text. This is because that the motivations of posting both text and image may be not always to enhance the sentiment or emotion of users. For example, text showed in Fig. 1(a) is the description of the event in the photo. The two views are visually related, rather than emotionally













(a)		 	MisterTTeaches: Grade 5s helping out other grades during #RAKWeek2015 #prairiewaters #rvsed #caring #pypchat
(b)		 	"I Can't Believe It!": Woman Overjoyed at Sight of Obama in SF #sanfrancisco
(c)		 	We are stunned by the news that Rally Kid Kylie M.@SmileyForKylie passed away last night. Please pray for her family h...
(d)		 	Too Fast and Too Furious? - New Photos and Details!

Fig. 1. Example tweets with both image and text. The top and bottom icons in the middle indicate sentiment (positive, negative or neutral) showed in image and text respectively.

⁵ <http://www.mcrlab.net/research/mvsa-sentiment-analysis-on-multi-view-social-data/>.

related. Another reason is that sentiment expressed in the message is usually affected by the contexts of posting this message. For example, Fig. 1(b) shows a crying woman in the picture, while textual part indicates that the woman was overjoyed at seeing Obama. In contrast, Fig. 1(c) is a photo of smiling kid, however, the fact is that the kid passed away as described in the text. Besides these, image and text can enhance the users’ sentiment. In Fig. 1(d), the negative sentiment of text is weak, which is strengthened by the attached image about a firing car in an accident. In Table 2, we list the percentage of agreements on the labels of text and image. The messages are grouped into ten categories based on the contained emotional words. We can see that only 53.2 % messages show same sentiments in their posted image and text. This poses significant challenges on multi-view sentiment analysis. In addition, users express their feeling about “happiness” (agreement of 64.6 %) more explicitly using happiness words and images.

Table 2. The percentage of messages with same labels in both textual and visual views. The messages are grouped into 10 categories based on the contained emotional keywords.

Category	Anger	Caring	Confusion	Depression	Fear	Happiness
Agreement (%)	47.5	64.6	47.5	48.7	50.9	64.6
Category	Hurt	Inadequateness	Loneliness	Remorse	Overall	
Agreement (%)	48.4	46.1	49.1	51.0	53.2	

We further analyze the accuracy of emotional words for indicating the overall sentiment of a Twitter message. We first manually label the 406 emotional keywords as positive, negative and neutral. The sentiment polarity of each tweet is same as that of the contained emotional keyword. Table 3 shows the results grouped by the 10 categories. We can see that the overall accuracy is only 30.6 %. The performance is especially poor for category “Confusion”, where keywords are ambiguous for expressing human affect. Thus, more advanced technique is needed in sentiment analysis. Again, “happiness” performs much better than other categories.

Table 3. Performance of sentiment analysis using keyword matching method. The 406 emotional keywords are grouped into 10 categories.

Category	Anger	Caring	Confusion	Depression	Fear	Happiness
Accuracy (%)	19.4	50.7	12.3	20.7	18.6	51.4
Category	Hurt	Inadequateness	Loneliness	Remorse	Overall	
Accuracy (%)	25.5	15.2	20.6	19.9	30.6	

4 Predicting Sentiment in Multi-view Data

The state-of-the-art sentiment analysis systems follow the basic pipeline showed in Fig. 2, which is similar to many other recognition problems. The most important component is the feature extraction which converts different type of data into feature vectors. Then a statistical learning approach is conducted to train a classifier which is employed to predict the sentiment polarity of the input testing data. In the following, we will briefly introduce some representative features for different data types. In the experiment, the classifier is learned using the popular linear SVM due to its outstanding performance.

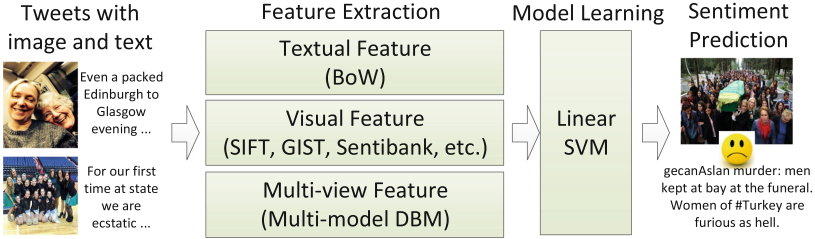


Fig. 2. The pipeline of sentiment analysis on social data.

4.1 Text-Based Approaches

Bag-of-Words (BoW) is a standard textual feature. With a pre-defined vocabulary including a set of terms (individual words or word n-grams), a document is represented as a feature vector, where each element can either be a binary value indicating the appearance of corresponding term or a counting value indicating the term frequency (TF). The vocabulary is usually constructed by selecting the most frequently appeared terms in the corpus. For sentiment analysis, emotion related terms can be also included. In addition, considering that the terms may have different importances, each term can be assigned with a weight such as the inverted document frequency (IDF). There are other variants designed for some specific applications. For text sentiment analysis, we will only test the TF and TF-IDF in the experiment.

Lexicon consists of emotional words which are manually assigned with sentiment scores. Besides the statistical learning approaches with textual features, lexicons can be directly utilized for determining the sentiment of tweet. Usually, a tweet is positive if the overall score is positive, otherwise the tweet is negative. In this paper, we consider two popular lexicons: SentiWordnet⁶ and SentiStrength⁷.

⁶ <http://sentiwordnet.isti.cnr.it/>.

⁷ <http://sentistrength.wlv.ac.uk/>.

4.2 Visual-Based Approaches

In this paper, we consider following visual features.

Color Histogram is the concatenation of three 256 dimensional histograms extracted from RGB color channels respectively.

GIST descriptor, which is helpful for scene classification, is a 320 dimensional feature. It includes the output energies of several filters (3 scales with 8, 8 and 4 orientations respectively) over 4×4 grids of an image.

Local Binary Pattern (LBP) descriptor is a popular texture feature. Each pixel is represented as binary codes (pattern) by comparing its value with that of neighbors. The feature vector is generated by counting the number of different patterns in the image.

Bag-of-Visual-Words (BoVW) borrows the idea of BoW except the words are SIFT descriptors which are computed on densely sampled image patches. Vocabulary includes centers of several groups generated by clustering a set of descriptors. In this paper, the vocabulary and spatial partition follow the same way used in [2].

Classemes [15] is a middle-level feature which consists of the outputs of 2,659 classifiers trained for detecting some semantic concepts (e.g., objects). Each dimension indicates the probability of the appearance of a category. Comparing to low-level features, Classemes represents an images at a higher semantic level.

Attribute is another middle-level feature, which represents abstract visual aspects (adjectives, e.g., “red” and “striped”), rather than the concrete objects used in Classemes. We adopt the 2,000 dimensional attribute proposed in [19], which is designed for representing category-wise attributes.

SentiBank [2] is an attribute representation dedicatedly designed for human affective computing. It includes 1,200 Adjective Noun Pairs (ANPs), e.g., “cloudy moon” and “beautiful rose”, which are carefully selected from Web data and representative for expressing human affects. SentiBank is intuitively suitable for visual sentiment analysis.

Aesthetic feature is helpful for understanding the visual instance at more abstract level such as “beautiful”. We adopt following aesthetic features used in [1]: dark channel feature, luminosity feature, S3 sharpness, symmetry, low depth of field, white balance, colorfulness, color harmony and eye sensitivity.

4.3 Multi-view Sentiment Analysis

Multi-view analysis makes use of the information extracted from both textual and visual aspects of a tweet. The most straightforward and standard way is either early fusion or late fusion. In early fusion, the features extracted from text and image are concatenated into a single feature vector. Late fusion combines the output scores of two models learned on textual and visual data respectively.

While both early and late fusion are able to boost the performance, the inherent correlations between two views is missing. Recently, multi-model learning method has showed strong performance in multi-view data analysis. In this paper, we adopt the approaches proposed in [14], where a multi-model Deep

Boltzmann Machine (DBM) is trained using textual and visual features as inputs. In [10], it has been showed to be helpful in detecting emotions from Web videos. We use a similar architecture including a visual pathway and a textual pathway. The input of the visual pathway is a 20,651 dimensional feature combining the Dense SIFT, HOG, SSIM, GIST, and LBP. On the other hand, BoW representation is utilized for generating input features for the textual pathway. The joint layer upon the two pathways contains 2,048 hidden units. The multi-model DBM is trained on 827,659 text-image pairs provided by the SentiBank dataset [2]. Details of the architecture can be referred to [10, 14].

5 Experiments

We perform experiments on polarity classification of sentiment by using the data labeled as positive and negative in our constructed dataset. The dataset is randomly split into training and testing set by 50%-50%. We use accuracy and F-score for performance evaluation. F-score is computed on positive class (F-positive) and negative class (F-negative) respectively, and their average is denoted as F-average.

5.1 Results on Textual Messages

We first evaluate the performance of BoW feature using TF and TF-IDF strategies respectively. Figure 3 shows the accuracy with various vocabulary size. The overall trend of performance is that accuracy can be improved with larger vocabulary size. However, TF is less sensitive to the vocabulary size than TF-IDF. We observe that an appropriate size is 2,000 or larger. In the following, we fix the vocabulary size as 4,000. In addition, IDF weighting strategy hurts the performance for various vocabulary size, which is different from the conclusion in general document classification. The reason may be that the IDF assigns large weights to rare words, rather than the words reflecting human feelings. For example, “path” is assigned with larger weight than “good”. This may pose negative impact on the sentiment analysis.

Table 4 further lists the results of lexicon-based approaches. For comparison, the results of TF and TF-IDF are also included. Generally, statistical learning approaches perform better than lexicon-based approaches. This indicates that sentiment analysis on tweets needs more advanced approaches and dedicated designs. However, there are many factors which may influence the performance of statistical learning approaches, such as the imbalance between positive and negative data. Our dataset includes more positive tweets. As a result, performance of negative class (F-negative) is relatively worse than that of positive class (F-positive) for both TF and TF-IDF. This is consistent to the observation in [12]. On the other hand, lexicon-based approaches are employed on each tweet independently. In some cases, it may be helpful to boost the performance of the rare class. Thus, F-negative of SentiStrength is better than that of TF or TF-IDF.

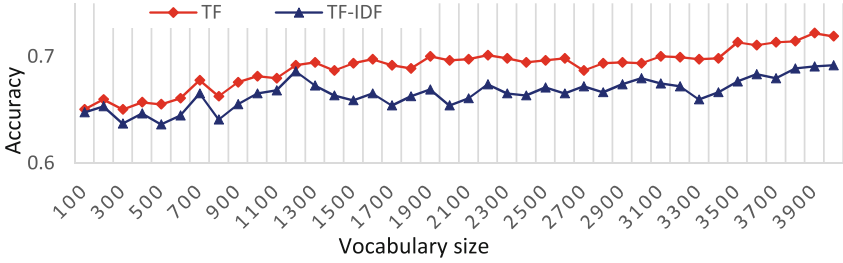


Fig. 3. Accuracy of BoW feature (TF and TF-IDF) on text sentiment analysis using different vocabulary sizes.

Table 4. Performance of different approaches for text sentiment analysis.

Method	Accuracy	F-positive	F-negative	F-average
SentiWordnet	0.603	0.640	0.557	0.598
SentiStrength	0.632	0.628	0.636	0.632
TF	0.719	0.791	0.569	0.680
TF-IDF	0.692	0.767	0.542	0.655

5.2 Results on Images

In addition to the visual features introduced in Sect. 4.2, we further test the early and late fusion of different features. LV-Late and LV-early fuse the four low-level visual features. V-Late and V-early combine all the eight features. The experiment results are showed in Table 5. For the low-level features, BoVW, which is

Table 5. Performance of different visual features for sentiment analysis.

Feature		Accuracy	F-positive	F-negative	F-average
Low-Level	Color histogram	0.652	0.772	0.263	0.518
	GIST	0.647	0.778	0.146	0.462
	LBP	0.653	0.787	0.066	0.426
	BoVW	0.667	0.775	0.354	0.565
Middle-Level	Classemes	0.650	0.747	0.432	0.589
	Attribute	0.680	0.782	0.400	0.591
	SentiBank	0.687	0.791	0.378	0.584
Aesthetic	Aesthetic	0.659	0.785	0.181	0.486
Fusion	LV-Early	0.673	0.780	0.366	0.573
	LV-Late	0.679	0.788	0.338	0.563
	V-Early	0.681	0.780	0.421	0.601
	V-Late	0.683	0.788	0.378	0.583

a preferred feature in many visual recognition tasks, only slightly outperforms others. This is because that visual appearances in a sentiment class is extremely diverse. Local features as SIFT in BoVW may be not representative for sentiment analysis. By combining the four features, which can capture different visual aspects of an image, LV-Late and LV-early further improve the results over each individual low-level feature.

For the middle-level features, both Attribute and SentiBank performs better than Classemes. This may indicate that the overall styles of the images or emotional concepts are more useful than concrete concept detectors defined in Classemes. In addition, Attribute and SentiBank, which model the semantic aspects of visual instances, achieve better performance than the low-level visual features. We can also see that aesthetic feature is an important aspect on visual sentiment analysis. However, fusion of all the features fails to boost the performance, as the result is dominated by some robust features (e.g., “SentiBank”). Furthermore, there is no winner between early fusion and late fusion with respect to the accuracy and F-average. Due to the fact that sentiment is much more abstract and extremely challenging to be represented from visual data, elaborative designs on feature selection and multi-feature fusion strategy are needed.

5.3 Results on Multi-View Data

In this section, we examine the effectiveness of different approaches on multi-view data. T-V-Early and T-V-Late respectively represent the early fusion and late fusion of TF feature and all the eight visual features. M-DBM is the performance of the feature using the outputs of the final joint layer in our learned multi-model DBM. Comparing Table 6 with Tables 4 and 5, we can see that jointly utilizing textual and visual information in the tweets can boost the performance even simply linearly fusing the features. In addition, the correlation between two views can be captured and represented by the multi-model DBM. Thus, M-DBM performs better than the results in Tables 4 and 5. We can also see that M-DBM performs better than T-V-Late and slightly worse than T-V-Early. This may be caused by the domain shift as multi-model DBM is trained on Flickr images. However, M-DBM feature (2,048 dimension) is much more compact than T-V-Early using a 12,655 dimensional feature. In general, Table 6 shows encouraging performances on multi-view sentiment analysis, which is worthy to be further investigated.

Table 6. Performance of different approaches for multi-view sentiment analysis.

Method	Accuracy	F-positive	F-negative	F-average
T-V-Early	0.752	0.825	0.572	0.699
T-V-Late	0.725	0.818	0.451	0.635
M-DBM	0.747	0.825	0.535	0.680

6 Conclusion and Future Work

We have introduced a new dataset called MVSA consisting of multi-view tweets for sentiment analysis. To the best of our knowledge, it is already the largest dataset dedicatedly constructed for multi-view sentiment analysis. While current dataset only includes several thousands of annotations, we will continuously update the data and regularly release the annotations on the most up-to-date tweets in the future. With this dataset, we have provided a pipeline for sentiment analysis on single-view and multi-view data. Extensive experiments are conducted to evaluate different features extracted from single view and multiple views, as well as their different combinations. Encouraging results suggest the joint analysis of both textual and visual information. Our result can be utilized as a baseline.

Besides the sentiment analysis, our dataset can also be used for investigating other interesting open issues, such as the subjective tweets detection. In addition, we have showed that there are many inconsistent labels between the text view and image view in the collected tweets. This suggests that future research should pay particular attention on the differentiation of emotional context with other contexts between two views, so that we can appropriately leverage the information from two views.

References

1. Bhattacharya, S., Nojavanasghari, B., Chen, T., Liu, D., Chang, S.F., Shah, M.: Towards a comprehensive computational model for aesthetic assessment of videos. In: ACM MM (2013)
2. Borth, D., Ji, R., Chen, T., Breuel, T., Chang, S.F.: Large-scale visual sentiment ontology and detectors using adjective noun pairs. In: ACM MM (2013)
3. Chen, T., Lu, D., Kan, M.Y., Cui, P.: Understanding and classifying image tweets. In: ACM MM (2013)
4. Chen, T., SalahEldeen, H.M., He, X., Kan, M.Y., Lu, D.: VELDA: Relating an image tweets text and images. In: AAAI (2015)
5. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. *Processing* **150**(12), 1–6 (2009)
6. Jiang, Y., Xu, B., Xue, X.: Predicting emotions in user-generated videos. In: AAAI (2014)
7. Li, L.J., Su, H., Xing, E.P., Li, F.F.: Object bank: a high-level image representation for scene classification and semantic feature sparsification. In: NIPS (2010)
8. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C.: Learning word vectors for sentiment analysis. In: ACL (2011)
9. Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.* **2**(1–2), 1–135 (2007)
10. Pang, L., Ngo, C.W.: Multimodal learning with deep boltzmann machine for emotion prediction in user generated videos. In: ICMR (2015)
11. Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S.M., Ritter, A., Stoyanov, V.: SemEval-2015 Task 10: sentiment analysis in twitter. In: SemEval 2015 Workshop (2015)

12. Saif, H., Fernandez, M., He, Y., Alani, H.: Evaluation datasets for twitter sentiment analysis: a survey and a new dataset, the STS-Gold. In: ESSEM Workshop (2013)
13. Speriosu, M., Sudan, N., Upadhyay, S., Baldrige, J.: Twitter polarity classification with label propagation over lexical links and the follower graph. In: EMNLP Workshop (2011)
14. Srivastava, N., Salakhutdinov, R.: Multimodal learning with deep boltzmann machines. *J. Mach. Learn. Res.* **15**(1), 2949–2980 (2014)
15. Torresani, L., Szummer, M., Fitzgibbon, A.: Efficient object category recognition using classemes. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 776–789. Springer, Heidelberg (2010)
16. Xie, W., Peng, Y., Xiao, J.: Cross-view feature learning for scalable social image analysis. In: AAAI (2014)
17. You, Q., Luo, J.: Towards social imagematics: sentiment analysis in social multi-media. In: MDMKDD (2013)
18. You, Q., Luo, J., Jin, H., Yang, J.: Robust image sentiment analysis using progressively trained and domain transferred deep networks. In: AAAI (2015)
19. Yu, F., Cao, L., Feris, R., Smith, J., Chang, S.F.: Designing category-level attributes for discriminative visual recognition. In: CVPR (2013)
20. Yuan, J., Mcdonough, S., You, Q., Luo, J.: Stribute: image sentiment analysis from a mid-level perspective. In: WISDOM (2013)