



Sentiment Analysis for Social Sensor

Xiaoyu Zhu, Tian Gan^(✉), Xuemeng Song, and Zhumin Chen

School of Computer Science and Technology, Shandong University, Jinan, China
xiaoyu_lorraine@163.com, {gantian, songxuemeng, chenzhumin}@sdu.edu.cn

Abstract. Every day, a huge number of information is posted on social media platforms online. In a sense, the social media platform serves as a hybrid sensor system, where people communicate information through the system just like in a sensor world (we call it Social Sensor): they observe the events, and they report it. However, the information from the social sensors (like Facebook, Twitter, Instagram) typically is in the form of multimedia (text, image, video, etc.), thus coping with such information and mining useful knowledge from it will be an increasingly difficult task. In this paper, we first crawl social video data (text and video) from social sensor Twitter for a specific social event. Then the textual, acoustic, and visual features are extracted from these data. At last, we classify the social videos into subjective and objective videos by merging these different modality affective information. Preliminary experiments show that our proposed method is able to accurately classify the social sensor data's subjectivity.

1 Introduction

With the development of the information age, social media has become a necessity in many people's lives. The social media like Twitter, Facebook, and Instagram, etc., has changed people's life. Twitter is one of the most popular social media platforms, it has more than 300 million active users every month¹. Every moment, there have many users sent tweets to share information, and express their opinions, feelings etc. For this reason, we can take the social media platform as a hybrid sensor system, in which each user serves as a social sensor, that they observe the events and report to the system. Usually, users send a tweet with a special label to express their opinions, feelings or what they observed for a specific event. Meanwhile, more users tend to post their tweets in a variety of ways, such as text, picture, and video.

Subjective classification of tweet has been researched for many years. Many researchers have proposed different methods and systems to classify the text of tweet into subjective or objective automatically [1, 7–11, 13, 17]. However, most of the research only focus on the analysis of the textual information. At the same time, the proportion of tweets that include pictures or video is increasing. The multimedia information gathered from the social sensors can help discover

¹ <https://about.twitter.com/company>.

breaking news early on, which would considerably facilitate the task of tracking social media for journalism practitioners, which has become paramount in their daily newsgathering process.

In this paper, we first crawl social video data (text and video) from social sensor Twitter for a specific social event. Textual, acoustic, and visual features are then extracted from these data. At last, we classify the social videos into subjective and objective videos by fusing these different modality affective information. Preliminary experiments show that our proposed method is able to accurately classify the social video's subjectivity.

The remainder of this paper is organized as follows. In Sect. 2, we review the previous work. We formulate our problem and introduce the features in Sect. 3. In Sect. 4, we introduce the dataset and the results of our experiments. Finally, we conclude in Sect. 5 with discussions on future work.

2 Related Work

2.1 Subjectivity Classification

Subjectivity classification has been a hot topic in the research community for years. In the earlier studies, the work on subjectivity classification mainly focuses on manually labelled documents, such as articles, reviews, and comments, etc. They proposed methods, concentrating on the text with different granularities like word [14], expression [13], sentence and document [17].

Wiebe [14] classified adjectives as subjective or objective by an unsupervised algorithm. The author used a small amount of detailed manual annotation seed words as subjective words, and expanded the subjective set of words by searching more similar words. Riloff and Wiebe [13] proposed a bootstrapping process that learns linguistically rich extraction patterns for subjective expressions. Yu and Hatzivassiloglou [17] separated opinions from fact, at both the document and sentence level. In sentence level, they measured the similarity of sentences, and took advantage of the Naive Bayes classifier with part-of-speech, polarity, and N-Gram, etc., as features to classify the sentence as subjective or objective. To separate documents that contain primarily opinions from documents that contain mainly facts, they utilized single words, without stemming or stopword removal, as features, to feed into the Naive Bayes classifier.

With the rise of social media in recent years, Twitter has become a popular online micro-blogging platform. Because almost all Twitter data are public, these rich data offer new opportunities for researchers to work on sentiment analysis. Due to the 140 characters limit, the Twitter text has its unique characteristic of short and always ambiguous. In addition, because the amount of tweets is huge, it is very time-consuming to manually label such large amounts of data. Therefore, more and more recent research learned the classifiers from data with noisy labels such as emoticons and hashtags. For example, to reduce the labeling effort in creating these classifiers, instead of using manually annotated data to compose the training data, Barbosa and Feng [1] leveraged sources of noisy labels as their training data. They classified the subjectivity of tweets with not only

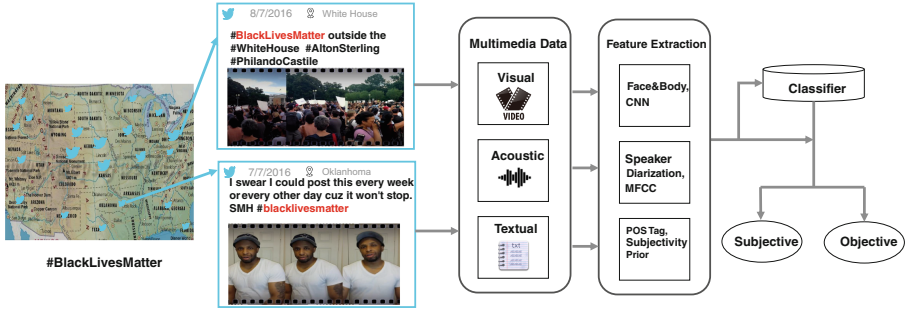


Fig. 1. An overview of our proposed method.

traditional features but also some tweet syntax features like retweets, hashtags, punctuation, emoticons etc. Similarity, Pak and Paroubek [11] collected negative and positive sentiments through the tweets consist of emoticons like “.” and “:(” . They obtained objective data from the Twitter accounts of popular newspapers and magazines, such as “New York Time” etc. Then they used N-Grams with low entropy values as the feature and built a sentiment classifier using the multinomial Naive Bayes classifier. Liu *et al.* [7] utilized both manually labelled data and noisy labelled data for training. They established two models: a language model and an emoticon model. They used the noisy emoticon data to smooth the language model trained from manually labelled data, and seamlessly integrated both manually labelled data and noisy labelled data into a probabilistic framework.

2.2 Social Sensor Data Processing

Social media have drastically influenced the way of information dissemination, by empowering the general public to publish and distribute the user generated content [3]. In a sense, the social media makes every user a virtual sensor, which can contribute to the harvest of the information. Therefore, people can act as sensors to give us results in a timely manner, and can complement other sources of data to enhance our situational awareness and improve our understanding and response to such events [3–5]. In addition, the information gathered from the social sensors can help discover breaking news early on, which would considerably facilitate the task of tracking social media for journalism practitioners, which has become paramount in their daily newsgathering process.

3 Proposed Method

In this section, we first provide the problem formulation, followed by the details of each multi-modal feature representation. At last, we present our multimodal analytics module. Figure 1 shows the overview of our proposed method.

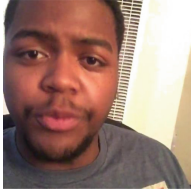



subjective	This is so true. #BLACKLIVESMATTER	You will respect us. #BlackLivesMatter	subjective
subjective			objective
objective	Speaker at #blacklivesmatter vigil in Liverpool	There is a protest in front of LA City Hall. #BlackLivesMatter	objective
subjective			objective

Fig. 2. The example of the Twitter posts which contains both text and video for the trending topic #BlackLivesMatter.

3.1 Problem Formulation

Textual information can be broadly categorized into two main types: facts and opinions [6, 14–16]. Facts are objective expressions about entities, events and their properties. Opinions are usually subjective expressions that describe people’s sentiments, appraisals or feelings toward entities, events and their properties. Similarly, videos can also be classified as subjective and objective. We define video as *subjective* if the emphasis on video content reflects one or several people’s point of view; *objective* if the video focuses on describing the situation and progress of the event.

Formally, suppose for the same trending topic T , we are given a collection of social media posts $P = \{P_1, \dots, P_{|P|}\}$, where each post $P_i = \{T_i, V_i\}$ consists of two components: textual component T_i and visual component V_i . The objective of our framework is to automatically classify the visual component V_i as *subjective* or *objective*. Figure 2 shows an example of our problem, in which for the trending topic T #BlackLivesMatter, the same Twitter post P has textual component T and visual component V with different subjectivity labels.

3.2 Feature Representation

Textual Feature. For the problem of text subjectivity classification, the textual information plays an important role. In our work, we proposed to use two sets of textual features: Part-Of-Speech (POS) tags, and the prior polarity information from subjectivity lexicon.

The subjectivity of a sentence is highly related to the composition of the word in the sentence. For example, opinion messages are more likely contain adjectives or interjections [1]. Therefore, we annotated the POS tags for each sentence in the text, and chose the number of related tags as the textual feature. Specifically, we selected preposition, adjective, noun, pronoun, adverb, verb and hashtag from the standard treebank POS tags²:

$$\mathbf{x}^{T,POS} = [\#(prep), \#(adj), \#(noun), \#(pron), \#(adv), \#(verb), \#(ht)] \quad (1)$$

In addition to the composition of the sentence, the prior positive or negative effects of the word in the sentence are also good indicators of subjectivity classification [1, 13, 17]. Specifically, we utilized a $+/-$ Effect dictionary [2] to calculate the number of $+/-$ Effect words, together with the number of total words from the text (excluding the punctuations, expressions, and links), to characterize the prior subjectivity information:

$$\mathbf{x}^{T,+/-} = [\#(positive), \#(negative), \#(words)] \quad (2)$$

Visual Feature. In the social event scenarios, human are the main elements in the scene. In this work, we analyzed the face of people in the video. However, the size of face area varies and the frontal faces are not always shown, thus the performance of face detection is not good enough to analyze the people inside the scene, especially in the crowded or complex environments. Therefore, the human body was also considered as a feature to analyze human. Specifically, we calculated the minimum, the maximum, and the average number of face and body detected in each image for the video. In addition, the ratio of the area of the face and body versus the whole image was considered. By combining these information, we resulted in the following 12-dimension features for human in the visual scene:

$$\mathbf{x}^{V,human} = [\#(T)^{op}, ratio(T)^{op}], \quad (3)$$

where $T = \{face, body\}$, $op = \{min, max, average\}$.

Besides human analysis of the visual scene, we also took advantage of the deep convolutional neural networks, which is a state-of-the-art way for image representation. We adopted the pre-trained convolutional neural networks, obtaining a 4096-dimensional CNN features for each frame from the fc7 layer output. Then, we trained the visual codebook using the local CNN features with k -means, thus each cluster represents a unique visual word. Consequently, each local CNN feature in the convolutional layer is assigned its closest visual word in the learned codebook. At last, we obtained a visual feature for each video by calculating the frequency of each visual word C_v^k :

$$\mathbf{x}^{V,CNN-BoVW} = [C_v^1, C_v^2, \dots, C_v^K] \quad (4)$$

² In total, there are 36 tags from standard treebank POS tags (https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html), together with four additional tags specific for twitter: URL, USR for user, RT for retweet, HT for hashtag.

Acoustic Feature. Though audio information alone may not be sufficient for understanding the scene content, there is great potential to analyze the accompanying audio signal for video scene analysis [8, 12]. In this work, we used the features in two layers: low-level acoustic characteristics, and high-level speaker diarization audio signature.

In the low-level layer, the audio signal was sampled at 1 KHz, divided into clips of 1 ms long. We sampled the 24-dimensional MFCC feature of each audio clips, and assumed that each MFCC descriptor represents an audio word. Then, we used a simple K-means method to cluster all the MFCC descriptors into N clusters, which is treated as the vocabulary size of an audio word. At last, we calculated the number of the MFCC features belong to each cluster. Therefore, for each audio segment, we obtained an N -dimensional vector:

$$\mathbf{x}^{A, MFCC-BoAW} = [C_a^1, C_a^2, \dots, C_a^N] \quad (5)$$

In the high-level layer, we applied the speaker diarization method to segment the audio clips into different clusters, which correspond to one or the same group of speakers. For example, for an audio A , it can be segmented into $A = \{a_{s1}, a_{s2}, a_{s1}, a_{s3}, a_{s2}\}$. We calculated the number of speaker changes along the temporal axis $Spk_Change(x)$, the duration of the audio $Dur(x)$, and the number of unique speakers $Spk_Num(x)$ as the feature:

$$\mathbf{x}^{A, Spk} = [Spk_Change(x), Dur(x), Spk_Num(x)] \quad (6)$$

3.3 Multimodal Subjectivity Classification

For single-feature-based approach, we conducted the standard machine learning experiment to predict the subjectivity. The SVM with RBF kernel was utilized. For multiple features, the feature level fusion is conducted for the subjectivity prediction.

4 Experiment

In this section, we first introduce the multimodal social sensor dataset, followed by the quantitative results and discussions based on the evaluation metrics.

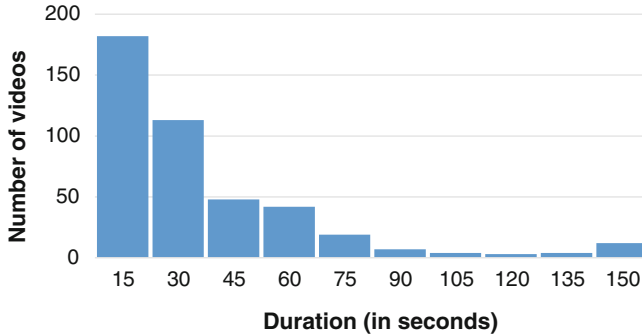
4.1 Dataset

In this work, we select “#BlackLiveMatter” as the trending topic T for our experiment. It is one of the Twitter Top 10 hashtags³ in the year of 2016. It is a movement emerged in Year 2014 after high-profile police shootings of unarmed black men, and it ranked as Twitters Top News Hashtags of Year 2015 (13.3 million in total till Year 2016).

³ The full list of Top 10 hashtags in the year of 2016 can be seen at <https://blog.twitter.com/official/en-us/a/2016/thishappened-in-2016.html>.

Table 1. The statistics of the social sensor data subjectivity.

Text		Video		#(Post)
Subj	Obj	Subj	Obj	
✓		✓		81
✓			✓	139
	✓	✓		40
	✓		✓	174

**Fig. 3.** The statistics of the duration of the 434 videos.

We first crawled all the tweets related to the event “#BlackLiveMatter” from July 7th 2016 00:00 to July 9th 2016 23:59 for three consecutive days, with the advanced search tool in Twitter. In total, we obtained 45570 Twitter posts, among which 3723 Twitter posts contain both text and video components. We randomly selected 200 Twitter posts everyday and applied duplicate removal, resulting in 434 unique Twitter posts. For each twitter post, we asked at least three human annotators to label the post as subjective or objective for both text and video components. The final groundtruth is determined by majority vote.

In the end, we obtained 220 pieces of textual information labelled as subjective, 214 pieces of textual information labelled as objective, 121 videos labelled as subjective, and 313 videos labelled as objective. The details of the statistics are shown in Table 1. In addition, we analyzed the statistics of the duration of the videos. As shown in Fig. 3, among 434 videos, more than 60% (295) of the videos are less than 30s, which is different from the videos on other social video sharing platform like YouTube.

4.2 Experimental Results and Discussion

In all experiments, the classification performance is reported with 10-fold cross-validation. Table 2 shows our results with different feature configuration on both text and video subjectivity classification.

Table 2. Average classification accuracy (%) on video and text subjectivity. T, V, and A represent Textual, Visual, and Acoustic, respectively.

Modality	Video subjectivity	Text subjectivity
T	71.7	76.5
V	88.2	60.8
A	86.9	61.5
T + V	88.2	75.6
T + A	84.3	77.0
V + A	90.3	61.3
T + V + A	88.7	74.9

For single modality feature, we can see from Table 2 that visual feature and acoustic feature are significantly better than textual feature for video subjectivity classification. At the same time, for the text subjectivity classification, the performance with the textual feature is much better than that with the visual or acoustic feature. For multi-modal feature, it is interesting to find out that fusing all the three types of feature did not perform well. Instead, for video subjectivity classification, the visual feature combined with acoustic feature perform the best (with 90.3% accuracy); for text subjectivity classification, the textual feature together with acoustic feature perform the best (with 77.0% accuracy). One of the reasons for the results is that the subjectivity label for the video and text component of the same post are not consistent: as shown in Table 1, only 59% of the posts have the same subjectivity label for video component and text component. At the same time, the visual feature plays an important role in the video component subjectivity classification, same applies for the textual feature for text component subjectivity classification. Therefore, simply combine the visual feature and textual feature will not help. Another interesting finding is that acoustic feature indeed helps for both video and text subjectivity classification, which can be further explored.

5 Conclusions and Future Works

In this work, we have crawled social video data (text and video) from social sensor Twitter for a specific social event. We evaluated the performance of existing computational models and analyzed the efficacy by combining multiple features. Preliminary experiments show that our proposed method is able to accurately classify the social sensor data subjectivity. For the future work, we plan to explore the fusion method of different modalities. Also, other information from the social sensor, for example, geolocation of tweets, friendship of the user, and number of likes/retweets, etc. could be considered for the analysis.

Acknowledgments. This work was supported by the Natural Science Foundation of China (61672322, 61672324), the Natural Science Foundation of Shandong province (2016ZRE27468) and the Fundamental Research Funds of Shandong University (No. 2017HW001).

References

1. Barbosa, L., Feng, J.: Robust sentiment detection on Twitter from biased and noisy data. In: COLING International Conference on Computational Linguistics, pp. 36–44 (2010)
2. Choi, Y., Wiebe, J.: +/−EffectWordNet: sense-level lexicon acquisition for opinion inference. In: Conference on Empirical Methods in Natural Language Processing, pp. 1181–1191 (2014)
3. Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J.: #Earthquake: Twitter as a distributed sensor system. *Trans. GIS* **17**(1), 124–147 (2013)
4. Gan, T., Wong, Y., Mandal, B., Chandrasekhar, V., Kankanhalli, M.S.: Multi-sensor self-quantification of presentations. In: Proceedings of ACM International Conference on Multimedia, pp. 601–610 (2015)
5. Gan, T., Wong, Y., Zhang, D., Kankanhalli, M.S.: Temporal encoded F-formation system for social interaction detection. In: Proceedings of ACM International Conference on Multimedia, pp. 937–946 (2013)
6. Liu, B.: Sentiment analysis and subjectivity. In: Handbook of Natural Language Processing, 2nd edn, pp. 627–666 (2010)
7. Liu, K., Li, W., Guo, M.: Emoticon smoothed language models for Twitter sentiment analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence (2012)
8. Liu, Z., Wang, Y., Chen, T.: Audio feature extraction and analysis for scene segmentation and classification. *VLSI Sig. Process.* **20**(1–2), 61–79 (1998)
9. Luo, C., Chan, M.C.: SocialWeaver: collaborative inference of human conversation networks using smartphones. In: Proceedings of the ACM Conference on Embedded Networked Sensor Systems, p. 20. ACM (2013)
10. Nie, L., Yan, S., Wang, M., Hong, R., Chua, T.S.: Harvesting visual concepts for image search with complex queries. In: Proceedings of the ACM International Conference on Multimedia, pp. 59–68. ACM (2012)
11. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of the International Conference on Language Resources and Evaluation (2010)
12. Rakotomamonjy, A., Gasso, G.: Histogram of gradients of time-frequency representations for audio scene classification. *IEEE/ACM Trans. Audio Speech Lang. Process.* **23**(1), 142–153 (2015)
13. Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 105–112 (2003)
14. Wiebe, J.: Learning subjective adjectives from corpora. In: Proceedings of the National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, pp. 735–740 (2000)
15. Wiebe, J., Riloff, E.: Creating subjective and objective sentence classifiers from unannotated texts. In: Gelbukh, A. (ed.) CILing 2005. LNCS, vol. 3406, pp. 486–497. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-30586-6_53

16. Wiebe, J., Wilson, T., Bruce, R.F., Bell, M., Martin, M.: Learning subjective language. *Comput. Linguist.* **30**(3), 277–308 (2004)
17. Yu, H., Hatzivassiloglou, V.: Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 129–136 (2003)