

RESEARCH

Open Access



Region-based convolutional neural network using group sparse regularization for image sentiment classification

Haitao Xiong^{1,2*}, Qing Liu¹, Shaoyi Song¹ and Yuanyuan Cai¹

Abstract

As an information carrier with rich semantics, images contain more sentiment than texts and audios. So, images are increasingly used by people to express their opinions and sentiments in social network. The sentiments of the images are overall and should come from different regions. So, the recognition of the sentiment regions will help to concentrate on important factors that affect the sentiments. Meanwhile, deep learning method for image sentiment classification needs simple and efficient approach for simultaneously carrying out pruning and feature selection whilst optimizing the weights. Motivated by these observations, we design a region-based convolutional neural network using group sparse regularization for image sentiment classification: R-CNNGSR. The method obtains the initial sentiment prediction model through CNN using group sparse regularization to get compact neural network, and then automatically detect the sentiment regions by combining the underlying features and sentimental features. Finally, the whole image and the sentiment region are fused to predict the overall sentiment of the images. Experiment results demonstrate that our proposed R-CNNGSR significantly outperforms the state-of-the-art methods in image sentiment classification.

Keywords: Image sentiment, Convolutional neural network, Sentiment regions, Sentiment classification

1 Introduction

The background of modern information technology has achieved explosive development and application. Social network sites represented by Twitter and Facebook gradually penetrated into every level of life and work of the whole people and had a profound impact on people's behavior patterns and mental models. In particular, social medias in these sites are produced every day, exchanging a large amount of user-generated content (UGC) [1]. Most of these contents are presented in the form of texts, images, etc. Meanwhile, these contents often carry very clear viewpoints and emotions. It is easy to obtain a wide range of communication, thereby stimulating and boosting social events, causing incalculable serious consequences [2]. Therefore, strengthening the automatic analysis ability of social medias and realizing effective

early warning and intervention of social media public opinion has become one of the urgent needs and important tasks of the government's public opinion management departments and enterprises [3, 4].

As an important part of social media analysis, sentiment analysis is one of the hot research topics currently. Government public opinion management departments and related social media platform providers face new demands and challenges for public opinion monitoring and intervention. After social media transitions from traditional texts to images, images can contain rich semantics. Many people tend to use images to express their opinions and emotions. So, how to quickly and correctly obtain sentiments from images and videos is becoming more and more important in social media analysis. Moreover, images often express a mixture of different sentiments and the attention of different people may concentrate on different image regions and have totally different sentimental feelings for the images. So, we can get the conclusion that the sentiment of images should come from different regions of the images. For this

* Correspondence: xionghaitao@btbu.edu.cn

¹School of Computer and Information Engineering, Beijing Technology and Business University, No.11 Fuchengfu Road, Beijing 100048, People's Republic of China

²Beijing Key Laboratory of Big Data Technology for Food Safety, No.11 Fuchengfu Road, Beijing 100048, People's Republic of China

reason, detecting the sentimental regions in the images will help to improve the sentiment classification performance.

Recently, neural network has been widely used [4–8]. Specially, convolutional neural network (CNN)-based sentiment classification methods have shown superior performance of sentiment prediction against traditional label sentiment classification methods for images [6, 7]. Rarely, image sentiment researches based on CNNs consider the recognition of sentiment regions. And these works mainly use the target detection methods to produce the sentimental region candidate set and does not use the local information around the object as a supplement for classification resulting in inaccurate sentiment classification results [4, 8]. Moreover, the common regularizations are mainly used for preventing over-fitting in weight level and ignores the sparse character of the network. So, these regularizations are worst for obtaining compact networks for image sentiment classification. In this paper, we design a region-based convolutional neural network using group sparse lasso for image sentiment classification, which can combine the low-level features and high-level sentimental features to determine the emotional regions, utilizing group sparse regularization for the deep neural network. Then, we fuse the sentiment of the whole image and detected sentimental regions to get the final sentiment predictions. Specially, the recognition of the image sentiment regions should be considered when the sentimental region candidate set is generated, which means that the local area including the object is considered, and the surrounding background is also considered for common analysis, thereby obtaining a more accurate image emotion area and improving the image emotion prediction effect. Through group sparse regularization and fused sentiments, sentiments from different regions could be effectively leveraged for learning.

Our contributions are summarized as follows:

- First, we design a novel CNN-based framework called region-based convolutional neural network using group sparse regularization (R-CNNGSR) for image sentiment classification to analyze image sentiments, which integrates sentiment information of whole images and regions into a CNN based model for effective learning;
- Furthermore, we propose group sparse regularization based on ℓ_1 and $\ell_{2,1}$ regularizations. Group sparse regularization is utilized in modeling to handle the sparse learning of deep neural network.

Experimental results demonstrate the superior results of our proposed R-CNNGSR compared with several

state-of-the-art methods in four benchmark data sets for image sentiment classification.

2 Related work

Through the development of computer vision, predicting the sentiment of images has been an interesting and meaningful research topic recently. There are two main methods to handle the prediction, which are dimensional models [9] and categorical models [10].

Dimensional models use a few basic spaces for sentiment description, and categorical models use classification methods to predict the sentiment labels, which are obvious for common people understanding and thus have been mainly used by most previous work. Specially, categorical models can be divided into single-label and multi-label classification depending on the number of predictions [6, 11]. In traditional image sentiment classification, an image is in general associated with one or more sentiment labels, which belong to categorical models [11]. In these methods, extracting features are the most important component for classification performance. Among the used extracting features, low-level features of the images are the most commonly used features due to the simply automatically generation methods. But these features can hardly reveal the sentiment in the images. Most current researches use sentimental semantics for classification. These sentimental semantics are mainly obtained by extracting features through machine learning methods, and the classification performance is depending on the extracted features. However, there exists a semantic gap problem for this task, which means the uncertainty between low-level features and high-level semantics of images [12]. So, hand-crafted features based on art and psychology theory are designed to prove the superior performance than low-level features in some practical usages, such as sentiment in painting. Moreover, fused multi-modal features are proposed for the different benefits of low-level and hand-crafted features.

On the other side, deep neural networks such as CNNs can extract high-level features from the images [13–16]. And the high-level features including more semantic information is good at image sentiment analysis. So, CNNs have been widely used in image sentiment classification and show certainly an improvement. Among the CNNs, very deep convolutional networks for large-scale image recognition known as VGGNet shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16–19 weight layers [17]. So, VGGNet and fine-tuned VGGNet with 16 layers or 19 layers have been widely used for extracting deep features of images and improve the classification performance of image sentiment [7, 8, 18]. These deep features are the high-level features including more semantic information which is good at emotion

analysis. Moreover, in order to use large scale yet noisy training data to solve the sentiment prediction, a robust CNN-based model known as progressive CNN (PCNN) for visual sentiment analysis is proposed recently. PCNN progressively selected a subset of the training instances to reduce the impact of noisy training instances and got prominent improvement [19]. However, these works did not consider the important effect of image regions with rich sentiments for classification. Yang et al. proposed a framework called ARconcatenation to leverage affective regions which is the first time to use regions of images for sentiment classification [8]. But the framework highly depends on target detection method EdgeBoxes which is an object recognition method [20] and not suitable for sentiment region detection.

3 Problem definition

Given an input image, the purpose of R-CNNNGSR is to classify the image into different sentiment label. Suppose that we have C sentiments $s = \{s_1, s_2, \dots, s_C\}$, and N images for training $x = \{x_1, x_2, \dots, x_N\}$. Each image x_n has been labeled by one sentiment in s . In this paper, we only consider binary classification which means that there exist only two sentiment labels: positive and negative. For consideration of sentiment in different regions of the images, we define R sentiment regions $r = \{x_{n1}, x_{n2}, \dots, x_{nR}\}$ which are detected in image x_n . On the other hand, in order to express group sparsity at neuron level for compact network, different groups should be considered and we define g as groups in the CNN.

The predicted probability $d_{x_n} = \{d_{x_n}^{s_1}, d_{x_n}^{s_2}, \dots, d_{x_n}^{s_C}\}$ of sentiments for the image x_n is used for image sentiment classification, where $d_{x_n}^{s_c}$ is the probability of sentiment s_c for image x_n and represents the extent to which sentiment s_c describes x_n . So, we can classify the image x_n as the max value sentiment in s_c . $d_{x_n}^{s_c}$ is under the constraints $d_{x_n}^{s_c} \geq 0$ and $\sum_{c=1}^C d_{x_n}^{s_c} = 1$ which mean that the sentiment probability is non-negative and s can describe the sentiments of the image fully.

Let us denote by $f(x; w)$ the activation values of the last fully connected layer for the whole image x_n . In R-CNNNGSR, only original images are used in training the deep model to get initial sentiment classification model. And w is the weight of the deep model. Specially, $f_{s_c}(x; w)$ is the activation value for sentiment s_c . Through the initial model, we can compute the sentiment of detected sentiment regions x_{nr} for x_n . The deep neural network is trained by minimizing the following objective function:

$$w^* = \underset{w}{\operatorname{argmin}} \frac{1}{N} \sum_{n=1}^N L(d_{x_n}, f(x_n; w)) + R(w) \quad (1)$$

where $L(\cdot, \cdot)$ is the suitable loss function for image

sentiment classification, and $R(\cdot)$ is the used regularization in the model learning.

In image sentiment classification, the purpose of our proposed method is using sentiments of whole images and important image regions to get compact deep model for predicting sentiment labels of images. In order to capture local and overall sentiments of images in model learning, we propose a deep CNN-based framework that can extract and integrate sentiments of whole images and important image regions for sentiment classification learning. For the challenge of getting compact deep model in training, group sparse regularization is used for image sentiment classification learning.

4 Proposed method

As illustrated in Fig. 1, we design a deep CNN-based framework called region-based convolutional neural network using group sparse regularization (R-CNNNGSR) for image sentiment classification, to utilize sentiment regions for learning. CNN has shown strong capacity in image sentiment classification. So, in the R-CNNNGSR framework, original images are fed into a pre-trained CNN model which is VGGNet used in this paper. But CNN model should be modified, because the number of outputs of the last fully connected layer in VGGNet, which is a sentiment classification layer, is 1000. The sentiment labels in R-CNNNGSR are positive and negative. So, the number of outputs should be assigned to the number of sentiment labels which is 2. In R-CNNNGSR, we use softmax function to convert the network outputs. Then, the sum of the final outputs will be 1. So, the final outputs can be seen as the probabilities of different sentiments. For deep learning, KL loss is utilized for penalizing the mispredictions according to the dissimilarities. In addition, R-CNNNGSR uses group sparse regularization to keep the network in the compact. Then, the CNN model is used to calculate sentiment similarity in sentiment regions detection. After that, the sentiment of whole images and sentiment regions are fused according to the area weight to generate the overall sentiment predictions of the images.

4.1 Region-based convolutional neural network

As a distance measure, KL divergence is widely used as training loss called KL loss in image sentiment classification [7, 21, 22]. KL loss measures the similarity between the ground-truth and sentiment prediction:

$$L_{KL} = \frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C \left(d_{x_n}^{s_c} \ln \frac{d_{x_n}^{s_c}}{\hat{d}_{x_n}^{s_c}} \right) \quad (2)$$

where $\hat{d}_{x_n}^{s_c}$ is the ground truth probability of sentiment s_c for the image x_n . If the label of the image x_n is s_c , then $\hat{d}_{x_n}^{s_c} = 1$, else $\hat{d}_{x_n}^{s_c} = 0$. From KL loss equation, we can

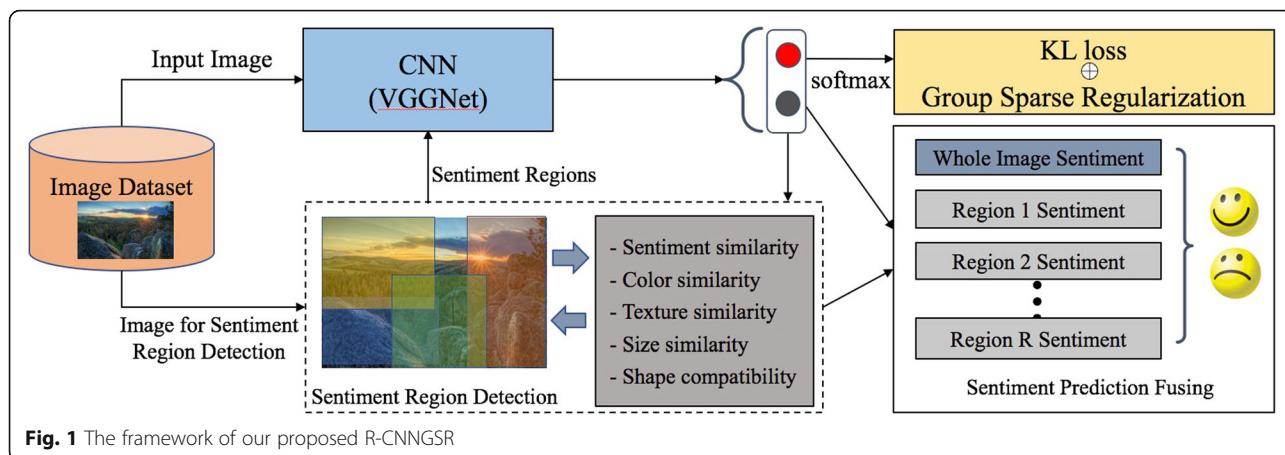


Fig. 1 The framework of our proposed R-CNNGSR

know that it penalizes the dissimilarity of ground-truth and predicted sentiment labels. So, we use L_{KL} as the loss function $L(\cdot, \cdot)$ in R-CNNGSR. The sentiment region detection in R-CNNGSR is inspired by the idea of selective search algorithm [23]. But we consider the sentiment similarity for regions merging which is import for image sentiment classification. In detail, we use original images with sentiment labels to train the CNN model for sentiment classification. After that, we can use the initial CNN model to predict the sentiment label of proposed image regions generated in sentiment region detection for the merge iteration process. The probability of the regions belonging to different sentiment categories is predicted, so that the sentiment scores of the sentiment regions of images can be obtained. In order to calculate the sentiment similarity of different sentiment regions, cosine similarity function is used. And we can get the similarity of different sentiment region candidates. Then, sentiment similarity

combined with color similarity, texture similarity, size similarity, and shape compatibility in selective search algorithm are used to merge and handle candidate regions. In addition, the sentiment region recognition also adopts a series of filtering measures to delete candidate regions whose aspect ratio is too large or small, and whose area is too large or too small, and finally automatically generate the sentiment region candidate set of the original image. The detail steps of sentiment region detection method are given in Fig. 2. Firstly, we generate initial candidate regions a well-known processing method: graph-based image segmentation. This method is also used by selective search method. Secondly, unsatisfied candidate regions are filtered. As we know, candidate regions whose length-width or width-length ratios are great and the pixel which are small are not important for image sentiment classification. In order to filter these unsatisfied candidate regions, filtering measures are used to delete candidate regions

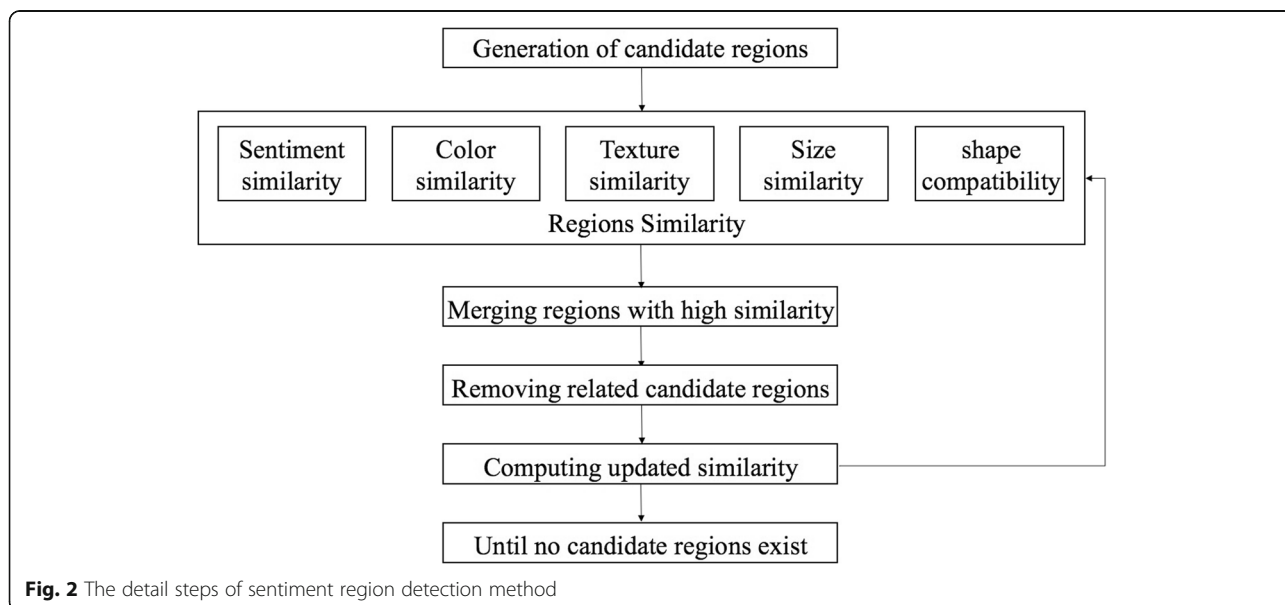


Fig. 2 The detail steps of sentiment region detection method

that do not meet the requirements of length-width or width-length ratios and pixel. Thirdly, regions similarity between each candidate regions pairs are computed. Then, candidate regions with high similarity are merged and related candidate regions are removed from the candidate region set. After doing this, we can compute the updated similarity of existing candidate regions and merged regions. This step will loop until there exist no candidate regions.

After acquiring the image sentiment area candidate set, the final sentiment prediction of the image can be started. R-CNNGSR calculates the probability of belonging to different sentiment labels of all sentiment regions in the sentiment region candidate set. We think that sentiment regions with larger area play greater roles in determining the overall sentiment of the image. For this reason, we perform a weighted averaging method through the area ratio to combine sentiment of all sentiment regions. Moreover, we combined the sentiment probabilities of all sentiment regions plus the sentiment probabilities of the original whole image to obtain the final sentiment prediction of the image. The fusion calculation formula is as follows:

$$Y^* = f(x_n; w) + \sum_{r=1}^R \frac{\text{area}(x_{nr})}{\text{area}(x_n)} f(x_{nr}; w) \quad (3)$$

where Y^* is the probability value that the predicted image belongs to the different sentiment labels, and $\text{area}(\cdot)$ represents the area of the image or detected sentiment regions. The final sentiment category predictive value is the sentiment category to which the maximum value belongs.

4.2 Group sparse regularization

The main purpose of traditional ℓ_1 and ℓ_2 regularizations are preventing over-fitting problem in learning from training data. ℓ_2 is the most common regularization in deep learning and used for weight decay. ℓ_1 regularization is also known as lasso [24]. ℓ_1 can produce sparse outputs at single level which is useful in sparse learning. ℓ_1 and ℓ_2 belongs to weight-level regularization.

In order to get sparse outputs in group, group lasso also known as $\ell_{2,1}$ are proposed and show better performance in weight sparse learning [25]. Since group lasso loses the guarantee of sparsity at single level, it may still be sub-optimal. To address this problem, sparse group lasso considers sparsity at both single level and group level [26]. Two kinds of groups are considered in R-CNNGSR, which are input groups and hidden groups in the CNN network. Input groups include neural network component which is the vector of all outgoing connections from the input neuron in CNN. Hidden groups include neural network component which consist of the

vector of all outgoing connections from one of the neurons in the hidden layers of CNN. As we have known, in order to obtain compact neural network, the input and hidden neuron should be considered in group-level sparsity. As defined, g_i are input groups and g_h are hidden groups in the CNN. In order to obtain compact neural network, we present group sparse regularization using sparse group lasso. Group sparse regularization can keep the group sparsity structure. At the same time, it permits single sparsity. Using group sparse regularization can get compact neural network which is more beneficial for image sentiment classification. In detail, group sparse regularization R_{gs} is calculated as:

$$R_{gs} = R_{\ell_1} + R_{\ell_{2,1}} \quad (4)$$

Moreover, ℓ_2 regularization is used to prevent over-fitting. The definition of R_{ℓ_2} is given below:

$$R_{\ell_2} = \|w\|_2 \quad (5)$$

In R-CNNGSR, group sparse regularization and ℓ_2 regularization are used simultaneously. So, the overall regularization R in R-CNNGSR can be obtained by adding R_{gs} to R_{ℓ_2} through weighted regularization combination:

$$R = \xi_1 R_{gs} + \xi_2 R_{\ell_2} \quad (6)$$

where ξ_1 and ξ_2 are the regularization parameters to balance the importance of the two components in objective function. The complexity of the different regularization term R_{gs} or R_{ℓ_2} is equivalent and it is given by $O(Q)$, where Q is the number of network parameters.

5 Experimental results and discussions

5.1 Implementation details

In this paper, we choose four image sentiment classification data sets for comparison, which are IAPS subset [27], Abstract [28], ArtPhoto [28], and Emotion6 [29]. Table 1 shows the details of the data sets. IAPS subset is derived from the International Affective Picture System called IAPS [30], which is a common sentiment data set that is widely used in sentiment classification research. IAPS subset contains 395 pictures, and all the images are labeled by eight emotions which are the components of Mikel's wheel [27]. The eight emotions are amusement, contentment, awe, excitement, fear, sadness,

Table 1 Details of the used data sets in the experiments

Data sets	Positive images	Negative images	Sum
IAPS subset	209	186	395
Abstract	139	89	228
ArtPhoto	378	428	806
Emotion6	660	1320	1980

disgust, and anger. The first four belong to the positive sentiment and the last four belong to the negative sentiment. There exist 806 artistic photographs in ArtPhoto, which are collected from a photo sharing website. Abstract contains 228 paintings. Emotion6 is widely used as a benchmark data set for emotion classification, which contains 1980 images collected from Flickr. In order to build binary classification data set, we transfer the emotion labels to sentiment labels. Emotions play a key role in human life as they operate as motivators, such as anger, disgust, amusement, and awe. These emotions can be defined as complex psychological states. Emotions can be positive or else negative, which belong to different sentiment category (positive or negative). Specifically, image emotion is often called image sentiment for binary classification (positive or negative). All emotion labels in the datasets can be divided into three groups, which are positive, negative, and neutral sentiments. So, we can delete the images of which dominant emotion is belonging to the neutral sentiment. Then, the labels of remained images are changed to positive and negative sentiments according to which sentiment the dominant emotion belongs to. And we can get binary sentiment classification datasets.

Image examples for emotion distribution learning are shown in Fig. 3. The images in the top row are positive image samples and the images in the bottom row are negative image samples. All images come from the IAPS subset, Abstract, ArtPhoto, and Emotion6 data sets. All data sets are randomly split into 75% training, 20% testing, and 5% validation sets. The validation set is used for choosing the best parameters of our methods.

In the experiments, R-CNNCSR is built on VGGNet containing 16 layers [17]. We change the number of the last fully connected layer outputs to the number of sentiments which is 2 in this paper. We use KL loss in the loss layer. The learning rates of the convolution layers,

the first two fully connected layers, and the classification layer are initialized as 0.001, 0.001, and 0.01, respectively. We fine-tune all layers by backpropagation through the overall neural network using mini-batches of 32, and the total number of epochs is 20 for image sentiment classification learning. Moreover, we had tried several different parameter configurations in the cross-validation fashion for ξ_1 and ξ_2 from 0.0001 to 10 using validation sets. Through the experiments, we had found that $\xi_1 = 0.001$ and $\xi_2 = 0.0005$ will achieve better and stable performance. For filtering measures to delete candidate regions in the sentiment region detection method, we exclude candidate regions whose length-width or width-length ratios are greater than 5 and candidate regions which are smaller than 200 pixels.

In order to check the superiority of our proposed R-CNNCSR, experiments are conducted for comparing it with four baseline image sentiment classification methods. In this paper, classification accuracy is used as the performance measure to compare different image sentiment classification methods. All our experiments are carried out on a NVIDIA GTX TITAN X GPU with 12 GB memory.

5.2 Results and analysis

5.2.1 On image sentiment classification

In order to verify the superiority of R-CNNCSR proposed in this paper, we compare R-CNNCSR with four image sentiment classification methods in experiments, which are VGGNet, fine-tuning VGGNet, PCNN, and ARconcatenation. For a fair comparison of these methods, all VGGNet-based methods use same backbone network which is the 16-layer network for learning. In this way, we do not need to consider whether the performance is improved by the backbone network or the proposed model. PCNN is a novel progressive CNN architecture that can leverage larger amounts of weakly

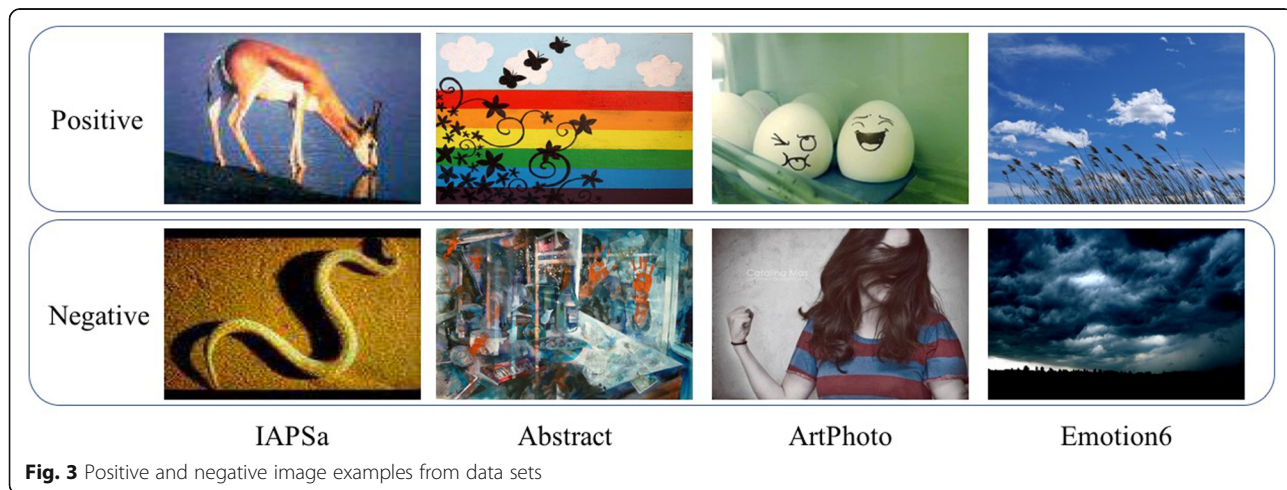


Fig. 3 Positive and negative image examples from data sets



Fig. 4 Generated sentiment regions in R-CNNGSR

supervised data. The ARconcatenation method is the first to incorporate affective region recognition into image sentiment classification. The proposed R-CNNGSR in this paper integrates sentiment similarity computing into sentiment region recognition and can obtain image sentiment regions more accurately. Figure 4 shows the example of detected sentiment regions in R-CNNGSR. The top image is the original image, and the bottom images are sentiment regions generated in R-CNNGSR. Specially, sentiment regions whose length-width or width-length ratios are greater than 5 and candidate regions which are smaller than 200 pixels are removed. From the figure, we can see that the generated sentiment regions in R-CNNGSR not only consider the targets in the image, but also consider the background in the images. And the background is important for sentiment classification. Through this way, we can predict the sentiment of the image more accurately.

Table 2 shows the sentiment prediction results of R-CNNGSR and the other four methods on four benchmark data sets. Italics indicate the best of the accuracy evaluation index values. It can be seen from the experimental results in Table 2 that R-CNNGSR greatly improves the prediction effect compared with the traditional CNN algorithm. Moreover, compared with the existing prediction algorithm ARconcatenation that uses affective region recognition, R-CNNGSR can also improve the accuracy of sentiment prediction. It can be seen that our proposed R-CNNGSR can improve the recognition effect of the sentiment regions by introducing the sentiment similarity of the regions in region merging and finally improve the sentiment prediction level of the entire image.

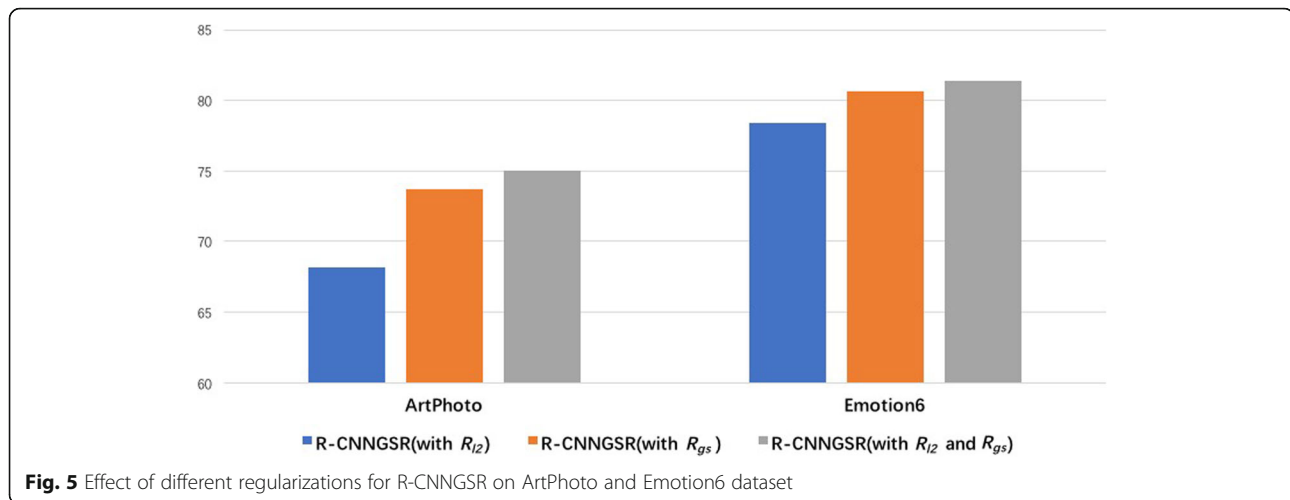
5.2.2 On effect of different regularizations for R-CNNGSR

In order to check the effect of group sparse regularization in R-CNNGSR, we perform experiments on Emotion6 data set using R-CNN with and without group sparse regularization. Classification accuracy is also used for classification performance comparing. Firstly, we set $\xi_1 = 0$ to get R-CNNGSR with R_{l2} . Secondly, we set $\xi_2 = 0$ to get R-CNNGSR with R_{gs} . Finally, we use both components of regularization to get R-CNNGSR with R_{l2} and R_{gs} . Then, we compare these R-CNNGSR methods using different regularizations, and the results are shown in Fig. 5. From the results, we can detect that the performance of R-CNNGSR in image sentiment classification is improved by using group sparse regularization R_{gs} , which reveals the effectiveness of R-CNNGSR by obtaining compact neural network for classification. Moreover, R_{gs} plus R_{l2} can also improve the performance of only using R_{gs} for the reason that R_{l2} is useful for preventing over-fitting. Through using R_{l2} and R_{gs} simultaneously, the performance of sentiment classification has been significantly improved.

Table 2 Performance comparison between R-CNNGSR and the state-of-the-art methods

Methods	IAPS subset	Abstract	ArtPhoto	Emotion6
VGGNet	88.51	68.86	67.61	72.25
Fine-tuned VGGNet	89.37	72.48	70.09	77.02
PCNN	88.84	70.84	70.96	73.58
ARconcatenation	89.39	74.41	73.76	78.52
R-CNNGSR	<i>92.14</i>	<i>75.89</i>	<i>75.02</i>	<i>81.36</i>

Italics indicate the best of the accuracy evaluation in the five methods



6 Conclusions

This paper discusses how to effectively identify and use image sentiment regions to help improve sentiment analysis in image sentiment classification learning. A novel CNN-based framework named region-based convolutional neural network using group sparse regularization was proposed, in which sentiment regions and compact network are effectively considered. Firstly, the initial sentiment prediction model is obtained by CNN using group sparse regularization. Then, sentiment, color, texture, size, and coincidence similarity are considered to detect the sentiment regions. Finally, the sentiments of whole image and sentiment regions are fused for image sentiment classification. Extensive experiments on four real-world data sets revealed that the effectiveness of R-CNN in image sentiment classification through fusing the sentiment of important image regions and utilizing group sparse regularization. Sentiments of the image can be present in multiple labels, so future research will extend R-CNN to handle multi-label sentiment prediction problems.

Abbreviations

CNN: Convolutional neural network; IAPS: International Affective Picture System; PCNN: Progressive CNN; UGC: User Generated Content

Acknowledgements

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

About the authors

Haitao Xiong was born in Jiujiang, Jiangxi, China, in 1983. He received the Ph. D degree in Management Science and Engineering from Beihang University, Beijing, China, in 2011. From 2011 to 2013, he was a lecturer of School of Computer and Information Engineering, Beijing Technology and Business University, Beijing, China. Since 2013, he has been an associate professor of School of Computer and Information Engineering, Beijing Technology and Business University, Beijing, China. His current research interests include image processing, machine learning, and business intelligence.

Qing Liu was born in Tangshan, Hebei, China, in 1994. She received the bachelor degree from the Tangshan Normal University, Tangshan, Hebei, in

2018. Now, she is currently a master student in School of Computer and Information Engineering, Beijing Technology and Business University, Beijing, China. Her research interests include data mining and business intelligence.

Shaoyi Song was born in Yingkou, Liaoning, China, in 1983. He received the Ph. D degree in Management Science and Engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2014. Since 2014, he has been a lecturer of School of Computer and Information Engineering, Beijing Technology and Business University, Beijing, China. His current research interests include electronic commerce, big data technology in food safety and game theory.

Yuanyuan Cai was born in Fengcheng, Jiangxi, China, in 1985. She received the Ph. D degree in Software Engineering from Beijing Jiaotong University, Beijing, China, in 2016. Since 2016, she has been a lecturer of School of Computer and Information Engineering, Beijing Technology and Business University, Beijing, China. Her current research interests include semantic computing, natural language processing and data mining.

Funding

This research was supported by the Beijing Natural Science Foundation (No.4172014 and 4184084), Support Project of High-level Teachers in Beijing Municipal Universities in the Period of 13th Five-year Plan (No.CIT&TCD201804031), the National Natural Science Foundation of China (No.71201004) and Humanity and Social Science Youth Foundation of Ministry of Education of China (No.17YJCZH007).

Declarations

Availability of Data and Materials: Please contact author for data requests.

Authors' contributions

All authors took part in the discussion of the work described in this paper. The author HX wrote the first version of the paper and did part of the experiments of the paper. HX performed the data processing and analysis. SS and YC revised the paper in different versions of the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 1 November 2018 Accepted: 23 January 2019

Published online: 01 February 2019

References

1. J. Krumm, N. Davies, C. Narayanaswami, User-generated content [J]. *IEEE Pervasive Computing* 7(4), 10–11 (2008)

2. A.M. Kaplan, M. Haenlein, Users of the world, unite! The challenges and opportunities of social media [J]. *Business horizons* **53**(1), 59–68 (2010)
3. S. Zhao, H. Yao, Y. Gao, et al., Predicting personalized image emotion perceptions in social networks [J]. *IEEE Trans. Affect. Comput.* **9**(4), 526–540 (2016)
4. Q. You, H. Jin, J. Luo, *Visual sentiment analysis by attending on local image regions [C]//AAAI* (2017), pp. 231–237
5. S. Zhang, Z. Wei, Y. Wang, T. Liao, Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary. *Futur. Gener. Comput. Syst.* **81**, 395–403 (2018)
6. E. Cambria, S. Poria, A. Gelbukh, et al., Sentiment analysis is a big suitcase [J]. *IEEE Intell. Syst.* **32**(6), 74–80 (2017)
7. Q. You, *Sentiment and emotion analysis for social multimedia: methodologies and applications [C]//Proceedings of the 2016 ACM on Multimedia Conference. ACM* (2016), pp. 1445–1449
8. J. Yang, D. She, M. Sun, et al., Visual sentiment prediction based on automatic discovery of affective regions [J]. *IEEE Transactions on Multimedia* **20**(9), 2513–2525 (2018)
9. M.A. Nicolaou, H. Gunes, M. Pantic, *A multi-layer hybrid framework for dimensional emotion classification [C]//Proceedings of the 19th ACM international conference on Multimedia. ACM* (2011), pp. 933–936
10. X. He, W. Zhang, Emotion recognition by assisted learning with convolutional neural networks [J]. *Neurocomputing* **291**, 187–194 (2018)
11. M.L. Zhang, Z.H. Zhou, A review on multi-label learning algorithms [J]. *IEEE Trans. Knowl. Data Eng.* **26**(8), 1819–1837 (2014)
12. S. Zhao, Y. Gao, X. Jiang, et al., *Exploring principles-of-art features for image emotion recognition [C]//Proceedings of the 22nd ACM international conference on Multimedia. ACM* (2014), pp. 47–56
13. Q. Zhou, B. Zhong, Y. Zhang, J. Li, Y. Fu, Deep alignment network based multi-person tracking with occlusion and motion reasoning [J]. *IEEE Transactions on Multimedia* <https://doi.org/10.1109/TMM.2018.2875360> (2018)
14. B. Zhong, B. Bai, J. Li, Y. Zhang, Y. Fu, Hierarchical tracking by reinforcement learning based searching and coarse-to-fine verifying [J]. *IEEE Trans. Image Process.* <https://doi.org/10.1109/TIP.2018.2885238> (2018)
15. B. Bai, B. Zhong, G. Ouyang, et al., Kernel correlation filters for visual tracking with adaptive fusion of heterogeneous cues [J]. *Neurocomputing* **286**, 109–120 (2018)
16. W. Long, Y.-r. Tang, Y.-j. Tian, Investor sentiment identification based on the universum SVM. *Neural Comput. & Applic.* **30**(2), 661–670 (2018)
17. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition [J]. *arXiv:1409.1556* (2014)
18. X. Li, Y. Jiang, M. Chen, et al., Research on iris image encryption based on deep learning [J]. *EURASIP Journal on Image and Video Processing* **2018**(1), 126 (2018)
19. Q. You, J. Luo, H. Jin, et al., *Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks [C]//AAAI* (2015), pp. 381–388
20. C.L. Zitnick, P. Dollár, *Edge Boxes: Locating Object Proposals from Edges [C]//European Conference on Computer Vision* (Springer, Cham, 2014), pp. 391–405
21. J. Yang, D. She, M. Sun, *Joint image emotion classification and distribution learning via deep convolutional neural network [C]//proceedings of the 26th international joint conference on Artificial Intelligence* (2017)
22. K. Song, T. Yao, Q. Ling, et al., Boosting image Sentiment analysis with visual attention [J]. *Neurocomputing* **312**, 218–228 (2018)
23. J.R.R. Uijlings, K.E.A. Van De Sande, T. Gevers, et al., Selective search for object recognition [J]. *Int. J. Comput. Vis.* **104**(2), 154–171 (2013)
24. Candès E J, Wakin M B. An introduction to compressive sampling [J]. *IEEE Signal Process. Mag.*, 2008, 25(2): 21–30
25. L. Baldassarre, N. Bhan, V. Cevher, et al., Group-sparse model selection: Hardness and relaxations [J]. *IEEE Trans. Information Theory* **62**(11), 6508–6534 (2016)
26. S. Scardapane, D. Comminello, A. Hussain, et al., Group sparse regularization for deep neural networks [J]. *Neurocomputing* **241**, 81–89 (2017)
27. J.A. Mikels, B.L. Fredrickson, G.R. Larkin, et al., Emotional category data on images from the international affective picture system [J]. *Behav. Res. Methods* **37**(4), 626–630 (2005)
28. J. Machajdik, A. Hanbury, *Affective image classification using features inspired by psychology and art theory [C]//Proceedings of the 18th ACM international conference on Multimedia. ACM* (2010), pp. 83–92
29. G. Levi, T. Hassner, *Emotion recognition in the wild via convolutional neural networks and mapped binary patterns [C]//proceedings of the 2015 ACM on international conference on multimodal interaction. ACM* (2015), pp. 503–510
30. P.J. Lang, M.M. Bradley, B.N. Cuthbert, International affective picture system (IAPS): Technical manual and affective ratings [M]. NIMH Center for the Study of Emotion and Attention, 39–58 (1997)

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
