Journal of the Brazilian Computer Society, 2022, 28:1, doi: 10.5753/jbcs.2022.2214

Investigating sentiments in Brazilian and German Blogs

Nina N. Zahn [**University of Mannheim** | ***nina.zahn@gmx.de***]
Greice P. Dal Molin [**Pontifical Catholic University of Rio Grande do Sul** | ***greice.molin@edu.pucrs.br***]
Soraia Raupp Musse [**Pontifical Catholic University of Rio Grande do Sul** | ***soraia.musse@pucrs.br***]

Pontifical Catholic University of Rio Grande do Sul, Av. Ipiranga, 6681, Partenon, Porto Alegre, RS, 90619-900, Brazil.

Received: 15 September 2021 • **Accepted:** 17 November 2022 • **Published:** 30 December 2022

Abstract Social interactions have changed in recent years. People post their thoughts, opinions and sentiments on social media platforms more often, through images and videos, providing a very rich source of data about population of different countries, communities, etc. Due to the increase in the amount of data on the internet, it becomes impossible to perform any analysis in a manual manner, requiring the automation of the process. In this work, we use two blog corpora that contain images and texts. Cross-Media German Blog (CGB) corpus consists of German blog posts, while Cross-Media Brazilian Blog (CBB) contains Brazilian blog posts. Both blogs have the Ground Truth (GT) of images and texts feelings (sentiments), classified according to human perceptions. In previous work, Machine Learning and lexicons technologies were applied to both corpora to detect the sentiments (negative, neutral or positive) of images and texts and compare the results with ground truth (based on subjects perception). In this work, we investigated a new hypothesis, by detecting faces and their emotions, to improve the sentiment classification accuracy in both CBB and CGB datasets. We use two methodologies to detect polarity on the faces and evaluated the results with the images GT and the multimodal GT (the complete blog using text and image). Our results indicate that the facial emotion can be a relevant feature in the classification of blogs sentiment.

Keywords: Cross-media sentiment analysis, Corpus, Emotions in images and texts, Face detection

1 Introduction

The use of images and short texts on social media is becoming more popular every day and can be processed more quickly by a user. Images are used to express emotions and are highly significant for sentiment analysis of web content (Islam and Zhang [2016]). Text classification is widely explored in this field of studies, while image analysis is still being investigated. The multimodal sentiment analysis, meaning the analysis of more than one domain modality, holds a great not yet exploited potential, according to Soleymani *et al*. [2017]. While the sentiment analysis is still a challenge, the multimodal analysis, providing more than one source of information, can present contradictory sentiments, for instance. In recent work proposed by Dal Molin *et al.* [2019], the authors address the challenge of building algorithms and methods that can infer sentiments exactly as humans perceive them. They build a base in BlogSet-BR called Cross-Media Brazilian Blog (CBB), which is based on people's opinions, containing the sentiments perceived in texts and images, when analyzed separately and also when presented together. Zahn *et al*. [2021] also performed this study, but in German. They create the Cross-Media German Blog (CGB) corpus which is based on the compilation of German blog posts collected on DWDS.de, published by the Berlin-Brandenburgischen Akademie der Wissenschaften. In addition, authors perform a sentiment analysis comparison between Brazilian blogs and German blogs that finds its justification for performance results in cultural differences.

In the present work, we used the corpus called Cross-Media German Blogs (CGB), presented in Zahn *et al*. [2021], composed of 950 blog posts. CGB was created with an algorithm based on a dataset obtained in DWDS.de.[1] The Appen crowdsourcing platform was used for large-scale data annotation.[2] We also use the corpus called Cross-Media Brazilian Blogs (CBB) presented by Dal Molin *et al*. [2019], composed of 880 blog posts. The CBB was created based on a dataset obtained in BlogSet-BR (dos Santos *et al*. [2018]). The Figure-Eigth crowdsourcing platform was used for large-scale data annotation[3].

In this paper, we extend the discussion of sentiment contradiction by including a new feature in the image, i.e., the study of the sentiment of faces. While in previous work proposed by Zahn *et al*. [2021] and Dal Molin *et al*. [2019], we discussed the contradictions of the multimodal information, including text and images, in this paper we are interested in the study of a relevant component in the images: the human faces. Our main question here is to understand if the polarity of the image has any agreement with the emotion expressed in the faces present in the images. As mentioned before, in this paper we use the two datasets that contain images and texts in the German language (CGB) Zahn *et al*. [2021] and the Cross-Media Brazilian Blog (CBB) Dal Molin *et al*. [2019], however in order to answer our main question, we work only with images in order to detect faces and expressed emotions for comparison with images polarities as defined in the ground truth.

[1] DWDS - Digitales Wörterbuch der deutschen Sprache. The word information system about the German language in history and in the present, published by Berlin-Brandenburgischen Akademie der Wissenschaften, "https://www.dwds.de/", retrieved on 5/2/20.
[2] https://appen.com/
[3] https://figure-eight.com

Our main contribution is the discussion about the contradictions of multimodal information, which includes images and texts. As well as evaluating if the polarity of the images is in accordance with the emotions expressed in the faces present in the images. It is hoped that this work will contribute to the field of psychology, since it is possible to detect a positive, neutral or negative sentiments based on a person's facial expression or writing in cases of psychopathy analysis. The area of human resources in the recruitment of personnel for work can also benefit from this study, among other areas.

We organize our paper's remainder in the following manner: Section 2 presents some literature referenced in this work, Section 3 outlines the overview about CBB and CGB corpora. In Section 4 we work with images containing faces and analyze their emotion, while the results of corpora are investigated and discussed in Sections 5 and 6. Section 7 comprises final reflections about this work, as well as limitations and future work.

2 Related Work

The studies on sentiment analysis and contradictions in blog posts focus on two domains: images and texts. Hanjalic et al. Hanjalic [2006] defines the biggest challenge in the sentiment analysis of images as the *affective gap*. Machine learning tools might interpret detected features wrongly since they cannot consider situational circumstances expressed in images. Wang and Li Wang and Li [2015] present frameworks using supervised and unsupervised learning trying to bridge this gap. Vinodhini and Chandrasekaran [2012] explain that subjective words and the combination of different opinions in one sentence can complex the sentiment classification of texts. Abbas *et al*. [2019] introduce a novel Multinomial using Naive Bayes classification model, which considers word frequency and dependence, overcoming the difficulties just mentioned. While sentiment analysis based on texts has widely been researched and brings several highly advanced tools with it, the image-based classification of emotions is still under development Chen *et al*. [2014b].

Borth *et al*. [2013] show a highly relevant work in sentiment analysis on visual content, which bases its approach on *Plutchik's Wheel of Emotions* Camras [1980]. Adjective Noun Pairs (ANP) are created by analyzing the labels of extracted images. A SVM is trained on the ANPs, leading to the development of SentiBank, an automatic image classifier that combines visual features with textual content. Chen *et al*. [2014a] propose DeepSentiBank, which uses visual sentiment concepts coupled with a deep neural network. The sentiment analysis of blog posts, which include image and text, is even more complex. The work of Peng *et al*. [2017] addresses that finding a consensus of emotions in image and text is difficult because of each component's diverse features and the challenge in measuring their similarities. Morency *et al*. [2011] were the first to investigate multimodal sentiment analysis with audio, visual, and textual features. They also performed statistical analysis to detect five modality features that significantly impact the classification of sentiments. A more recent work Zadeh *et al*. [2017] proposed an end-to-end fusion model that aggregates unimodal, bimodal, and trimodal communications (language, visual, and acoustic). Applying data annotation on multiple domains can lead to a person's contradictory sentiment perception between domains. There is just a small number of studies about the detection of contradictions in text processing. Harabagiu *et al*. [2006] describe a network for detecting contradictions in text processing with over 62% accuracy. It processes negations, recognizes contrasts, and automatically identifies antonyms. A neural network presented by Li *et al*. [2017] uses context-specific word embedding, which tries to distinguish contrasting words better. This model outperforms modern methods on benchmark datasets by 6.11%.

The sentiment analysis performed by Dal Molin *et al*. [2019] takes up the topic of cross-media retrieval, investigating images and texts of Brazilian blog posts and associated contradictions. In a recent paper Zahn *et al*. [2021], we provided the same study on German blogs and provide some comparison with Brazilian Blogs Dal Molin *et al*. [2019]. In the field of elements that make up an image, in the present work, we researched the faces to assess the emotions that are transmitted and verify whether the facial expression could contribute to alleviating contradictions between the image and text domains. The studies of Nadeeshani *et al*. [2020] shows two techniques for predicting facial emotion with Machine Learning and Deep Learning. They conclude that both are possible to model with predictions above 80%, which is above the state of the art in this area. Furthermore, they address the beneficial importance of using Machine Learning techniques with Action Units (AUs), because they consider it to be relevant for being able to justify the expected emotion in terms of the contribution of the AU.

We want to show that the discussion about the contradictions of multimodal information, which includes images and texts, still has many gaps. It is a very relevant area these days, mainly because communication is practically virtual nowadays. Trying to find a consensus between the emotions in the images and in the text is difficult due to the different characteristics of each component and there is a lot of difficulty in measuring their similarities. A possible approach that we seek to evaluate is the classification of emotion on the face, as a way of minimizing the multimodal contradiction between text and image.

3 Overview about CGB and CBB

CGB presented by Zahn *et al*. [2021] was created with an algorithm based on a dataset obtained in DWDS.de. The Appen crowdsourcing platform was used for large-scale data annotation.[4] Authors created a survey for the subjects to classify the sentiments of the text and image separately, and another survey to analyze the post (image and text together) according to their perception (negative, neutral or positive). In this way, they built the Ground Truth (GT), which was used to compare the performance of the available sentiment classifiers. For image analysis, authors applied the following techniques: SentiBank (Borth *et al*. [2013]), DeepSentiBank (Chen *et al*. [2014a]), and VGG (Vadicamo *et al*.

[4]https://appen.com/

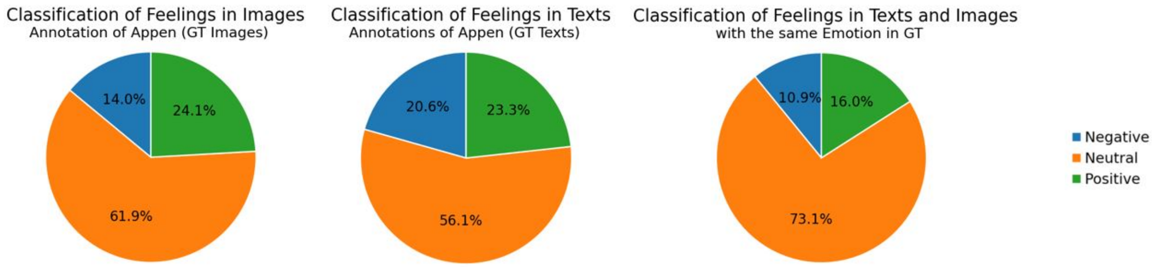

Figure 1. Left and center: sentiment distribution in images and texts of CGB classified by the subjects (out of 950 posts). Right: sentiment distribution in posts for which image and text of CGB that were separately classified with the same sentiment (out of 523 posts).

[2017]). For German text analysis, they use the GermanPolarityClues lexicon (Waltinger [2010]), Rauh's German Political Sentiment Dictionary (hereinafter referred to as Rauh's dictionary proposed in Rauh [2018]) and the Linguistic Inquiry and Word Count (LIWC) tool (Meier *et al*. [2019]). The analysis revealed contradictions between the sentiments of image and text in the same blog, which were further investigated. CGB corpus consists of 950 annotated blog posts. Figure 1 shows the classification of sentiments in images and texts on the left and in the center, used as our Ground Truth (GT). We analyzed the conformity of the same image and text in the post when presented separately to the annotators. The graph on the right of Figure 1 shows the distribution of sentiment for blog posts whose classified sentiment in the image is in accordance with the sentiment classified in the text.

The CBB was created based on a dataset obtained in BlogSet-BR (dos Santos *et al*. [2018]). The Figure-Eigth crowdsourcing platform was used for large-scale data annotation[5]. For image analysis, authors from CBB (Dal Molin *et al*. [2019]) applied the same techniques than presented by Zahn *et al*. [2021], i.e., classifier and neural networks SentiBank Borth *et al*. [2013], DeepSentiBank Chen *et al*. [2014a], and VGG Vadicamo *et al*. [2017]. For text analysis, they use the OpLexicon (Souza and Vieira [2012]) method is a lexicon of sentiments for the Portuguese language, composed of 32191 lines with 4 variables. Currently it is in version 3.0 that was reviewed by linguists in relation to the polarity of some adjectives. The SentiLex (Carvalho and Silva [2015]) method is a lexicon of sentiments for Portuguese that is composed of approximately 6,000 adjectives and 25,000 flexed forms and the Linguistic Inquiry and Word Count (LIWC) (Meier *et al*. [2019]) tool for Portuguese analysis. The analysis also revealed contradictions as discussed by Zahn *et al*. [2021]. CBB contains 880 posts that include image and texts. Figure 2 shows the conformity analysis classification of sentiments in left and center images and texts, used as our Ground Truth (GT). The graph to the right of Figure 2 shows the distribution of ranked sentiment in the image being in line with the ranked sentiment in the text.

4 Face Detection in CBB and CGB dataset images

The images from the CGB and CBB datasets (Dal Molin *et al*. [2019]) were manually selected in order to find the most frontal faces. For the CGB dataset, we selected 46 images, distributed between positive (56.6%), neutral (16.7%) and negative (27.8%) polarities, according to the ground truth. For the CBB dataset, 122 images were selected, distributed between positive (56.1%), neutral (7.%) and negative (36.3%) polarities, according to GT. From now on, in order to work only with images that contain faces, we created subsets of CGB and CBB, named CGB' and CBB', respectively.

Two methods were used to detect facial emotion, following the line of work of Nadeeshani *et al*. [2020] that used Action Units (AUs) and a CNN model to compare the prediction of them. For both methods, at the beginning, the OpenFace Baltrusaitis *et al*. [2018] toolkit was used for face detection. In the first case, the detected faces were classified using the CNN Simonyan and Zisserman [2014] model and the result of this model is the prediction of the emotions Anger, Disgust, Fear, Happiness, Sadness, Surprise or Neutral. In the second case, we used the detection of action units (AUs) Cohn *et al*. [2007] to identify the emotion of the face. Table 1 indicates the AUs activated on the faces, depending on the emotion. Based on the emotions defined for the face, which are 7.

In order to express the emotion in terms of polarities, to allow comparing the results with the ground truth (GT) of the images, it was defined that Happiness would indicate the positive polarity, that Surprise and Neutral emotions would indicate the Neutral polarity and the other emotions would be considered as negative polarity.

We consider indicating the emotion of surprise as neutral in that it expresses both positive and negative surprise. So, we decided to consider it as neutral.

The second method used the detection of action units (AUs) Cohn *et al*. [2007] to identify the expression of emotion on the face. Table 1 [6] indicates the AUs activated on the face, depending on the emotion. Based on the emotions defined for the face, which are 7, the average of the intensities of the AUs involved and active in negative emotion is calculated to define the value of negative polarity highlighted on the face.

Based on data collected from the two models, we performed a clustering analysis, using the KMeans model from the sickit-learn library [7], in order to detect patterns of facial emotions that could help us understand the polarity in the images. The K used in KMeans are $K = 3$ (three polarities). The use of KMeans proposed by Hossain *et al*. [2018] is justi-

[5] https://figure-eight.com

[6] https://imotions.com/blog/facial-action-coding-system/

[7] https://scikit-learn.org/stable/

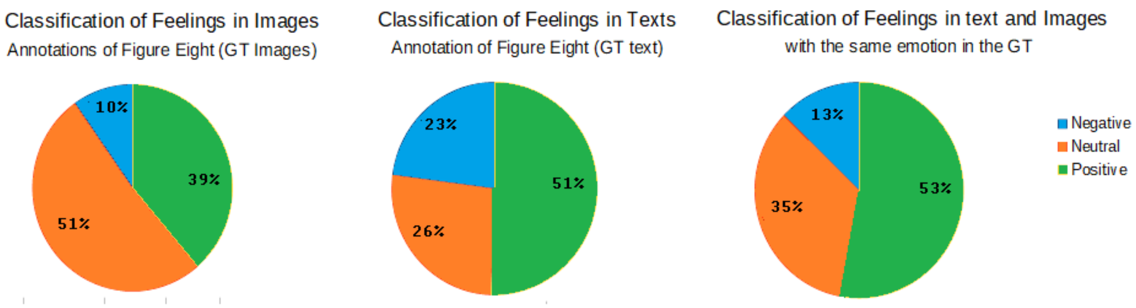

Figure 2. Left and center: sentiment distribution in images and texts of CBB classified by the subjects (out of 880 posts). Right: sentiment distribution in posts for which image and text of CBB that were separately classified with the same sentiment (out of 422).

Table 1. The Happiness/Joy, Sadness, Fear, Anger, Disgust and Contempt emotions are calculated as the Aus are activated, if they are not activated they will have the intensity value 0. The Emotion column indicates the name of the emotion, the Action Unit column shows the AUs that need to be activated for the emotion to exist on the face, and the Description column informs which parts of the face are indicated by the AUs. Source: `https://imotions.com/blog/facial-action-coding-system/`.

Emotion	Action Unit	Description
Happiness / Joy	6 + 12	Cheek Raiser, Lip Corner Puller
Sadness	1 + 4 + 15	Inner Brow Raiser, Brow Lowerer, Lip Corner Depressor
Surprise	1 + 2 + 5 + 26	Inner Brow Raiser, Outer Brow Raiser, Upper Lid Raiser, Jaw Drop
Fear	1 + 2 + 4 + 5 + 7 + 20 + 26	Inner Brow Raiser, Outer Brow Raiser, Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Stretcher, Jaw Drop
Anger	4 + 5 + 7 + 23	Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Tightener
Disgust	9 + 15 + 16	Nose Wrinkler, Lip Corner Depressor, Lower Lip Depressor
Contempt	12 + 14 (on one side of the face)	Lip Corner Puller, Dimpler

fied because it is a very known method and has been already used in the context of facial emotions in literature Sultana and Shahnaz [2014].

As mentioned before, the emotions in Table 1 are transformed into one of the polarities negative, neutral and positive. The average of the defined emotions for each polarity is calculated. For example, to calculate the negative polarity, as 5 emotions are involved, we average the emotion intensity of 5 emotions. Finally, the higher value between the three polarities is used to define the final polarity of the face.

The next sections discuss our main findings in an attempt to understand whether the face can be a collaborative element in defining polarity in an image.

4.1 CNN Model in the CBB' dataset

The CBB' corpus is composed of 122 images, with one or more frontal faces at each image, totaling 157 detected faces (more than one face per image). Initially, we analyzed the conformity between the prediction of the faces polarity and the ground truth (GT) of the image. We tested the both methods, as presented before, i.e., according to a CNN and the AUs analysis. Indeed, the results for the two methods are very similar. Results indicate that the faces polarity, being one of the elements from the image, in the CBB' dataset, agrees in 54% of the cases with GT of the image, in the both methods. Since the results of CNN and AUs are very similar, we decided to proceed in the further analysis using only the CNN method.

Figure 3 shows the performed comparison between the emotions predicted by the CNN model on the faces of the images and GT of the images on the left. In the middle, the comparison is made between the predictions of the faces with the CNN model and the VGG model performed to find the image polarity, and on the right, the comparison is made between the predictions on the image, without considering the face, between the VGG model and the GT.

Figure 4 shows the distribution of faces detected in the images, using clustering techniques of CNN results. Visually, we can see the boundaries of each cluster. There are some more distant points in cluster 1 (orange) that we are considering as neutral polarity. Cluster 0 (blue), negative polarity, and 2 (green) have an intersection zone. In addition, cluster 2 (green), which was consider as positive polarity, is quite spread out.

Figure 5 shows the number of images with their respective GT polarity in each cluster, that was generated according to the classification of faces contained in the same images. As mentioned before, color blue was chosen to define negative polarity, orange states for neutral and green for positive. It is easy to observe that neutral faces, according to CNN method, were more present in negative images, according to GT; positive faces in neutral images and neutral and positive faces in positive images.

4.2 CNN Model in the CGB' dataset

The CGB' corpus is composed of 46 images with one or more frontal faces in each image, totaling 72 detected faces. Initially, we analyzed the conformity between the prediction of the faces and GT of the image. As executed for CBB', we verified the agreement of sentiments of GT image with the predicted face on the CNN model and AUs. Apparently, as happened in CBB', the results for the two methods are very similar, as it has been indicated in studies like proposed by Nadeeshani *et al*. [2020]. The methods used suggest that the faces, being one of the elements that compose the image, in the CGB' dataset, are in accordance with the GT in 40% of

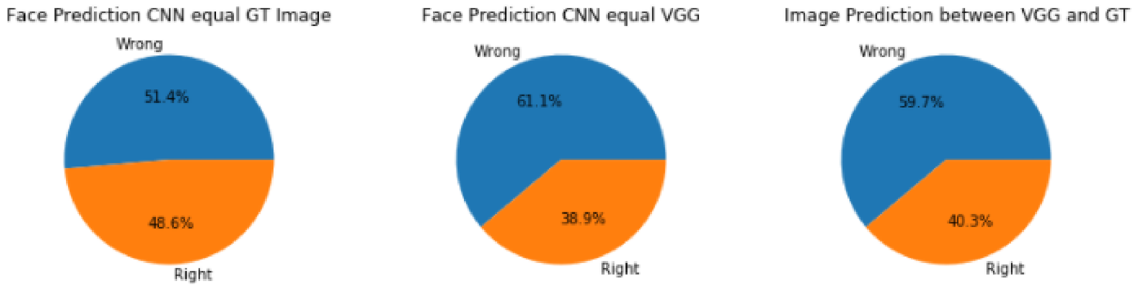

Figure 3. Comparison between the Face polarity predicted by the CNN model and GT image on the left, comparison between the Face polarity and VGG implementation in the center and comparison between the image polarity predicted by VGG and GT image on the right of CBB'.

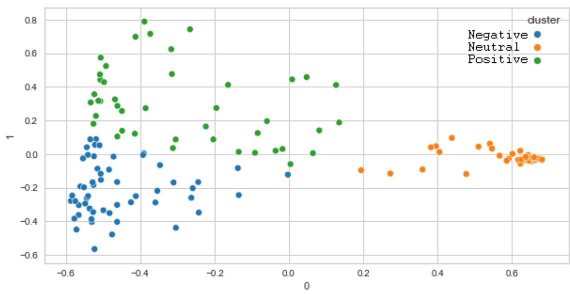

Figure 4. The faces are distributed among the 3 clusters according to the calculation of the Euclidean distance of the CBB' dataset. Orange states for neutral, blue for negative and green for positive.

images (for AUs) and 48% (for CNN). So, as before, we decided to continue the studies using only the CNN model.

Figure 6 shows the comparison between the emotions predicted by the CNN model on the faces of the images and the Ground Truth of the images on the left. In the middle, the comparison is made between the predictions of the faces with the CNN model and the VGG model, and on the right, the comparison is made between the predictions on the image, without considering faces, between the VGG model and GT.

Figure 7 shows a large blurring between clusters neutral faces (orange) and positive (green). Cluster with positive faces (blue) although scattered, still seems to be more defined than remaining ones.

Figure 8 shows the number of images with their GT polarity. Cluster of negative images according to the CNN (on the left) has positive polarity as its main polarity defined in GT. For cluster 1, which was identified by the faces in the images as being neutral, has neutral and positive, according to GT. In the case of cluster 2 composed of faces that indicate negative polarity in facial expressions, it has images which polarity is predominantly positive. Possibly, these discrepancies in the amount of data in each cluster are related to the imbalance between GT polarities when the images that contained faces were chosen, as there are many more images with faces being classified with positive polarity than negative and neutral ones.

Table 2. Classifiers performance according to presented by Zahn *et al*. [2021] in the CGB corpus.

Polarity	SentiBank	DeepSentiBank	VGG
Positive	167	145	135
Neutral	44	5	0
Negative	32	68	102
Total	243/905	218/905	237/905
% hit	**26.85**	**24.09**	**26.19-68.30**

Table 3. Number of images from CGB' correctly classified according to CNN, VGG and GT. We present the polarity and the technologies used. Percentages represent the accuracy of each technique (VGG with and without considering sentiment classification neutral).

Polarity	CNNxGT	CNNxVGG
Positive	22	21
Neutral	2	0
Negative	11	23
Total	35/72	44/72
% hit	**48.61**	**61.11-82.85**

5 Results when Evaluating Corpora CGB and CGB'

This section presents some analysis we performed comparing GT of the images polarity, the results of classifications as computed by the methods Sentibank, DeepSentibank, VGG and the achieved results of CNN on the computed faces. The final version of CGB contains 905 posts since SentiBank could not be applied to some images, due to formatting issues. Table 2 presents the performance of each classifier. Indeed, it was presented in Zahn *et al*. [2021] and it is presented again in this work to help with the discussion.

These values represent the accuracy of each classifier when predicting image polarity in accordance with GT. From 905 valid elements, CGB counts 128 images labeled as negative, 558 as neutral, and 219 as positive, according to GT. It is easy to notice that presented accuracy is very low (see %hit of Table 2).

Table 3 shows the performance of CNN applied on faces of CGB' compared to the image polarity according to GT (second column) and according to VGG classifier (third column). It is interesting to see that faces emotions, in terms of percentage of correct answers, seem to improve the accuracy of image polarity classification, when compared to Table 2.

In addition to the images, we compared the CNN polarity based on faces with the polarity annotated by the subjects regarding the complete blog (image and text together).

Figure 5. Number of images according to GT polarity in each cluster in the CBB' dataset. On the left, the cluster of negative faces and the distribution of polarities of images in the GT. In the center, the cluster of neutral faces and on the right the cluster with positive faces.

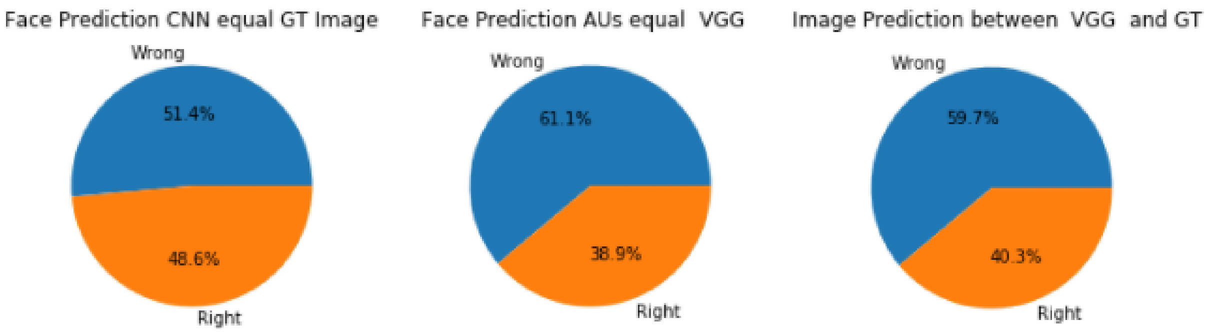

Figure 6. Comparison between the polarity predicted by the CNN model and GT image of CGB'.

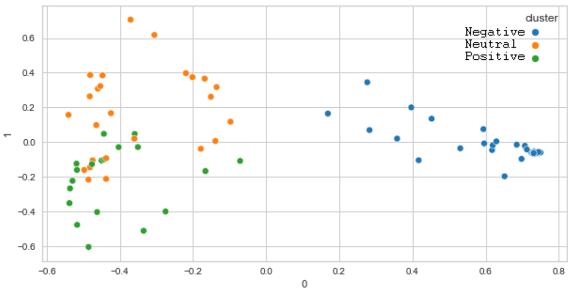

Figure 7. The faces are distributed among the 3 clusters according to the calculation of the Euclidean distance of the CGB' dataset. Orange states for neutral faces, blue for negative and green for positive ones.

Our study aimed to investigate if considering faces, we could achieve a better accuracy in detecting the polarity of the blog. Firstly, the value of accordance between text and image in CGB achieved the best value of accuracy of 30.61% (among all tested technologies) with the GT. So, if the blog polarity could be in some way influenced by the facial emotion presented in the image, the accuracy could be increased to 37.5%. It can represent that the subjects that annotate the emotion of a blog, creating the GT for CGB, may use the emotion of the face to define the emotion.

6 Results when Evaluating Corpora CBB and CBB'

Comparing with CGB obtained accuracies, all of the image classifiers performed better on CBB. As presented in Table 4,

Table 4. Classifiers performance according to presented by Dal Molin *et al.* [2019] in the CBB corpus.

Polarity	SentiBank	DeepSentiBank	VGG
Positive	150	222	231
Neutral	174	92	0
Negative	21	43	58
Total	345/880	357/880	289/880
% hit	**39.20**	**40.56**	**32.84-67.68**

SentiBank shows a 39.20% of correct answers, DeepSentiBank 40.56%, and VGG 67.68% (higher accuracy) concerning GT.

Furthermore, Table 5 shows the performance of CNN applied on faces compared to GT polarity in CBB, and VGG. As it happened with CGB, the accuracy is increased when compared with Table 4. In addition to the images, as discussed with CGB, we compared the CNN polarity based on faces with the polarity annotated by the subjects regarding the complete blog (image and text together). The value of accordance between text and image in CBB achieved the best value of accuracy of 45.45% with the GT. Considering that the blog polarity uses the facial emotion presented in the image, the accuracy could be increased to 54.14%, i.e., the subjects that annotate the emotion of a blog, creating the GT for CBB, may used the emotion of the face to define the blog emotion.

Figure 8. Number of images according to GT polarity in each cluster in the CGB' dataset. On the left, the cluster of negative faces and the distribution of polarities of images in the GT. In the center, the cluster of neutral faces and on the right the cluster with positive faces.

Table 5. Number of images from CBB' correctly classified according to CNN, VGG and GT. We present the polarity and the technologies used. Percentages represent the accuracy of each technique (VGG with and without considering sentiment classification neutral).

Polarity	CNNxGT	CNNxVGG
Positive	65	58
Neutral	16	0
Negative	8	29
Total	89/157	87/157
% hit	**56.68**	**51.59-82.85**

7 Conclusion

This paper revisits two studies previously presented by Dal Molin *et al*. [2019] and Zahn *et al*. [2021], who proposed CBB and CGB corpora, respectively. For both corpora, results of classification with respect to text and images polarities were presented. In this paper, we investigate a new hypothesis: if one specific component of the image, in this case the faces, can help to detect the emotion on the images.

In order to generate a subset of our corpora to test our hypothesis, we created CBB' and CGB', with the images that contain mostly frontal faces. Then, we tested with a specific CNN proposed by Simonyan and Zisserman [2014] to detect facial emotions and also with a methodology to find out emotions based on AUs (Baltrusaitis *et al*. [2018]).

The results were promising showing that faces polarity presents higher accuracy with the GT, in terms of percentage, when compared with the techniques from literature used to compute the image polarity, in both corpora. Surprisingly, the faces emotion computed using the CNN agrees more with the GT than SentiBank, DeepSentiBank and VGG, in CGB' and CBB'. One possible explanation is that when subjects, who compose with their opinions the GT of images, see a face, it can maybe lead the image polarity, as they perceive, in some way. In terms of the multimodal analysis of the Corpora CGB and CBB (complete blog with images and texts), the facial emotion also increased the achieved accuracy. This work has some limitations and one of them is certainly the small number of blogs analyzed in the two Corpora, even more critical when we use a subset of the Corpora CGB' and CBB'.

As the objective of this work is to evaluate the contradictions of the CGB and CBB datasets, we need to use the results of contradictions detected in Dal Molin *et al*. [2019] and Zahn *et al*. [2021]. We did not use another dataset with multimodal contradictions because we did not find them with the same characteristics.

Another future work is the comparison between data sets from different countries. In the present study, it was easy to see that facial emotion was more relevant in the accuracy of the BCC than in the CGB. Perhaps a cultural difference in terms of Hofstede's analysis (Hofstede [2001]) can also be considered, as proposed by Zahn *et al*. [2021].

Declarations

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and grants CAPES/88881.134081/2016-01, CNPq/302303/2017-0, CNPq/146411/2018-8, FAPERJ/E-26-202.818/2017, and E-26-200.834/2021.

Authors' Contributions

All authors contributed to the writing of this article, read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Data can be made available upon request.

References

Abbas, M., Memon, K. A., Jamali, A. A., Memon, S., and Ahmed, A. (2019). Multinomial naive bayes classification model for sentiment analysis. *IJCSNS*, 19(3):62.

Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 59–66. IEEE.

Borth, D., Chen, T., Ji, R., and Chang, S.-F. (2013). Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 459–460.

Camras, L. (1980). Emotion: a psychoevolutionary synthesis.

Carvalho, P. and Silva, M. J. (2015). Sentilex-pt: Principais características e potencialidades. *Oslo Studies in Language*, 7(1).

Chen, T., Borth, D., Darrell, T., and Chang, S.-F. (2014a). Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks. *arXiv preprint arXiv:1410.8586*.

Chen, T., Yu, F. X., Chen, J., Cui, Y., Chen, Y.-Y., and Chang, S.-F. (2014b). Object-based visual sentiment concept analysis and application. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 367–376.

Cohn, J. F., Ambadar, Z., and Ekman, P. (2007). Observer-based measurement of facial expression with the facial action coding system. *The handbook of emotion elicitation and assessment*, 1(3):203–221.

Dal Molin, G. P., Santos, H. D., Manssour, I. H., Vieira, R., and Musse, S. R. (2019). Cross-media sentiment analysis in brazilian blogs. In *International Symposium on Visual Computing*, pages 492–503. Springer.

dos Santos, H. D. P., Woloszyn, V., and Vieira, R. (2018). Blogset-br: A brazilian portuguese blog corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).

Hanjalic, A. (2006). Extracting moods from pictures and sounds: Towards truly personalized tv. *IEEE Signal Processing Magazine*, 23(2):90–100.

Harabagiu, S., Hickl, A., and Lacatusu, F. (2006). Negation, contrast and contradiction in text processing. In *AAAI*, volume 6, pages 755–762.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications.

Hossain, N., Jahan, R., and Tunka, T. T. (2018). Emotion detection from voice based classified frame-energy signal using k-means clustering. *Int. J. Softw. Eng. Appl.*, 9(4):37–44.

Islam, J. and Zhang, Y. (2016). Visual sentiment analysis for social images using transfer learning approach. In *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)*, pages 124–130. IEEE.

Li, L., Qin, B., and Liu, T. (2017). Contradiction detection with contradiction-specific word embedding. *Algorithms*, 10(2):59.

Meier, T., Boyd, R. L., Pennebaker, J. W., Mehl, M. R., Martin, M., Wolf, M., and Horn, A. B. (2019). "liwc auf deutsch": The development, psychometrics, and introduction of de-liwc2015. *PsyArXiv*, (a).

Morency, L.-P., Mihalcea, R., and Doshi, P. (2011). Towards multimodal sentiment analysis: Harvesting opinions from the web. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 169–176.

Nadeeshani, M., Jayaweera, A., and Samarasinghe, P. (2020). Facial emotion prediction through action units and deep learning. In *2020 2nd International Conference on Advancements in Computing (ICAC)*, volume 1, pages 293–298. IEEE.

Peng, Y., Huang, X., and Zhao, Y. (2017). An overview of cross-media retrieval: Concepts, methodologies, benchmarks, and challenges. *IEEE Transactions on circuits and systems for video technology*, 28(9):2372–2385.

Rauh, C. (2018). Validating a sentiment dictionary for german political language—a workbench note. *Journal of Information Technology & Politics*, 15(4):319–343.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S.-F., and Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, 65:3–14.

Souza, M. and Vieira, R. (2012). Sentiment analysis on twitter data for portuguese language. In *International Conference on Computational Processing of the Portuguese Language*, pages 241–247. Springer.

Sultana, S. and Shahnaz, C. (2014). A non-hierarchical approach of speech emotion recognition based on enhanced wavelet coefficients and k-means clustering. In *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*, pages 1–5. IEEE.

Vadicamo, L., Carrara, F., Cimino, A., Cresci, S., Dell'Orletta, F., Falchi, F., and Tesconi, M. (2017). Cross-media learning for image sentiment analysis in the wild. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 308–317. DOI: 10.1109/ICCVW.2017.45.

Vinodhini, G. and Chandrasekaran, R. (2012). Sentiment analysis and opinion mining: a survey. *International Journal*, 2(6):282–292.

Waltinger, U. (2010). Germanpolarityclues: A lexical resource for german sentiment analysis. In *LREC*, pages 1638–1642.

Wang, Y. and Li, B. (2015). Sentiment analysis for social media images. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 1584–1591. IEEE.

Zadeh, A., Chen, M., Poria, S., Cambria, E., and Morency, L.-P. (2017). Tensor fusion network for multimodal sentiment analysis. *arXiv preprint arXiv:1707.07250*.

Zahn, N. N., Dal Molin, G. P., and Musse, S. R. (2021). Cross-media sentiment analysis on german blogs. In *Anais do XLVIII Seminário Integrado de Software e Hardware*, pages 114–122. SBC.