

Cross-Media Sentiment Analysis on German Blogs

Nina N. Zahn¹, Greice P. Dal Molin², Soraia R. Musse²

¹Undergraduate Program of Business Informatics
University of Mannheim - Mannheim, Germany

²School of Technology, Graduate Program of Computer Science
Pontifical Catholic University of Rio Grande do Sul
Porto Alegre, Brazil

nzahn@mail.uni-mannheim.de, soraia.musse@pucrs.br

Abstract. *Social interactions have changed in recent years. People post their thoughts, opinions and feelings on social media platforms more often. Due to the increase in the amount of data on the internet, it is impracticable to carry out the sentiment analysis manually, requiring automation of the process. In this work, we present the corpus Cross-Media German Blog (CGB) which consists of German blog posts with feelings in the domain of images, texts and posts (Ground Truth), classified according to human perceptions. We apply existing Machine Learning technologies and lexicons to the corpus to detect the feelings (negative, neutral or positive) of the images and texts and compare the results with the GT. We examined contradictory posts, when the image and text classified by humans in the same post had diverging feelings. The comparison of this article with the analysis of sentiment among the media of Brazilian blog posts finds its justification for performance results in cultural differences, since, throughout this work, Brazil is classified as indulgent and Germany as a restrained country.*

1. Introduction

The usage of images on social media is becoming more popular containing short texts that can be processed faster by the user. Visual features captured in 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 explored. The multimodal sentiment analysis, meaning the analysis of more than one domain modality, holds a great not yet exploited potential [Soleymani et al. 2017]. This paper aims to perform sentiment analysis on blog posts containing images and texts in the German language with machine learning tools and lexicons.

We built the corpus called Cross-Media German Blogs (CGB), composed of 905 blog posts. The CGB was created with a self-implemented algorithm based on a dataset obtained in DWDS.de.¹ The Appen crowdsourcing platform was used for large-scale data annotation.² We created a survey for the subjects to classify the feeling of the text and image separately, and another survey to analyze

¹DWDS - 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.

²https://appen.com/

the post (image and text together) according to their perception (negative, neutral or positive). In this way, we built the Ground Truth (GT), being able to compare the performance of the available sentiment classifiers. For image analysis, we applied the classifier and neural networks SentiBank [Borth et al. 2013], DeepSentiBank [Chen et al. 2014], and VGG [Vadcamo et al. 2017]. We use the GermanPolarity-Clues lexicon [Waltinger 2010], Rauh’s German Political Sentiment Dictionary (hereinafter referred to as Rauh’s dictionary) [Rauh 2018] and the Linguistic Inquiry and Word Count (LIWC) [Meier et al. 2019] tool for German analysis. The analysis revealed contradictions between the feelings of image and text in the same blog. We investigated these contradictory posts and explained how these findings affect the performance of the technologies and lexicons mentioned above. As a last step, we compared the results of this article with the results of the cross-media sentiment analysis of Brazilian blogs [Dal Molin et al. 2019], which establishes the basis for this work.

2. Related Work

The studies on sentiment analysis and contradictions in blog posts focus on two domains: images and texts. Machine learning tools might interpret detected features wrongly cannot consider the situational circumstances expressed in images. Vinodhini and Chandrasekaran [Vinodhini and Chandrasekaran 2012] explains that subjective words and the combination of different opinions in one sentence can complex the sentiment classification of texts. Borth et al. [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. [Chen et al. 2014] 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. Morency et al. [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). The sentiment analysis performed by Dal Molin et al. [Dal Molin et al. 2019] takes up the topic of cross-media retrieval, investigating images and texts of Brazilian blog posts and associated contradictions. Our paper aims to provide the same sentiment analysis on blog posts in the German language, focusing on analyzing contradictory perceptions between image and text. We compare both studies concerning possible cultural differences and their reflection on the performance of the used technologies.

3. Our Methodology

The original corpus is a compilation of German blog posts collected on DWDS.de, published by the Berlin-Brandenburgischen Akademie der Wissenschaften. The purpose of this resource is to find keywords in a massive data collection. Since we were not looking for specific posts, we randomly downloaded 60.000 elements presented to us by quering

the server for the keyword ””.³. We implemented an algorithm for performing web data extraction on the corpus to create a dataset containing the following components: post id, date, URL, title, text, image URL. For this work, we automatically sorted out duplicates, posts with text not in German, unavailable links, not identifiable titles, texts, or images, and posts with no or unavailable images. Then we applied the LIWC dictionary to each text and the VGG-T4SA model to each image separately for sentiment classification. Both technologies are presented in Section 3.2. After that, our corpus contained 990 blog posts that match the following rules:

- VGG must classify the feeling of the image as negative or positive, with a polarity probability indicator greater than 40% in the sum of 100% for the three classification possibilities (negative, neutral, and positive), and
- LIWC must detect at least two words with polarity negative or positive in the text, and the word count must be greater than ten and less than 500 [Moraes et al. 2016].

3.1. Annotation Process

We created a survey to get sentiment perceptions of each domain (image and text) in our posts by subjects on Appen. The images and texts respectively are presented separately to three annotators in random order. The annotator then has to classify the sentiment of the content as either negative, neutral, or positive. The majority rule determines the final classification. The emotions are defined as follows: Negative means that some aspects of a content indicate a negative character, such as death; Neutral means that a content is purely informative and shows no emotional character signs in its content; Positive means that some aspects of a content indicate a positive character, such as love. The result was the CGB corpus consisting of 950 annotated blog posts. The remaining 40 posts included an image of insufficient quality or falsely extracted text or images and were removed from the corpus. Figure 1 shows the classification of feelings in images and texts on the left and in the center. We analyzed the conformity of the same image and text in the post when presented separately to the annotators. The graph on the right shows the distribution of sentiment for blog posts whose classified sentiment in the image is in accordance with the sentiment classified in the text.

Next, we put together the image and text of the blog posts. The annotator had to answer if the post shows a contradiction in relation to the feelings of the image and the text, and classifying a predominant feeling if so. The annotators classified 64 blogs as contradictory, which corresponds to 7 % of all posts in the CGB. Of these 64 posts, 32 had a negative feeling, 17 neutral and 15 positive as the predominant feeling.

3.2. Emotion Analysis

The methods used for sentiment classification of images are presented as follows:

- SentiBank [Borth et al. 2013] is a sentiment classifier using visual ontology to map images to emotional concepts.
- DeepSentiBank [Chen et al. 2014] is an advancement of SentiBank using a deep convolutional neural network for sentiment classification.

³Corpus hit for ””, from the blog corpus of the Digitales Wörterbuch der deutschen Sprache, ”<https://www.dwds.de/r?q=.corpus=blogsdate-start=2003date-end=2014format=maxsort=randomlimit=100>”, retrieved up on 05/02/20.

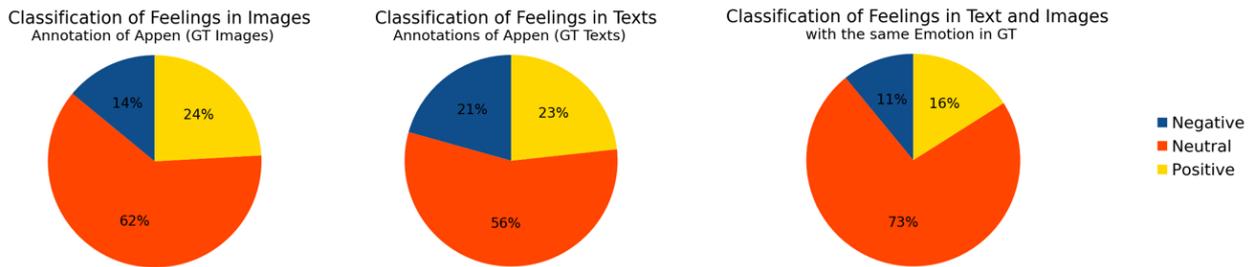


Figure 1. Left and center: sentiment distribution in images and texts classified by the subjects (out of 950 posts). Right: sentiment distribution in posts for which image and text were separately classified with the same sentiment (out of 523 posts).

- VGG-T4SA [Vadicamo et al. 2017] is a deep convolutional neural network that predicts the sentiment probability for negative, neutral, and positive of images.

The lexicons used for sentiment classification of text had to contain words of the German language and are described as follows:

- GermanPolarityClues [Waltinger 2010], is a lexical resource for sentimental analysis, including each word’s class, the sentiment of negative, neutral, or positive, and numerical polarity scores.
- Rauh’s German Political Sentiment Dictionary [Rauh 2018] is a sentiment dictionary for German political language, classified as either negative or positive.
- LIWC [Meier et al. 2019] is software providing multilingual dictionaries for text analysis. It calculates a score of positive and negative emotion for each text input.

We implemented an algorithm to obtain the assigned sentiment of each text: First, all 950 texts of the posts in CGB are loaded. Before applying the lexicons, we removed German stopwords in text, gathered from Stowords-DE irrelevant for text analysis.⁴ For example: ”Ich sollte nicht schon donnerstags sagen, dass ich mit der Bäckerei für diese Woche fertig bin.”results in ”Ich sagen donnerstags fertig Bäckerei Woche.”. We searched the shortened text for emotional words included in one of the three dictionaries. The text’s sentiment is classified by calculating the maximum word count assigned to either negative, neutral, or positive. In Rauh’s dictionary, a text is categorized as neutral if it contains the same number of positive and negative words. A data frame with the blog id, shortened text, overall word count, number of relevant words, and the assigned sentiment is created and saved as a CSV file.

4. Results when Evaluating our Corpus CGB

The final version of CGB contains 905 posts since SentiBank could not be applied to some images due to formatting issues. Finally, we compared the predicted sentiment class indicated by each tool with the corresponding GT of every image, presented in Table 1. These values represent the accuracy to which each classifier correctly predicted the sentiment class of an image out of 905 elements (347 elements without considering neutrals). CGB counts 128 images labeled as negative, 558 as neutral, and 219 as positive, according to GT. So the models seem to perform the lowest on neutral images (leaving

⁴<https://raw.githubusercontent.com/stopwords-iso/stopwords-de/master/stopwords-de.txt>

Tabela 1. Number of images classified according to feeling and technology used. Percentages represent the accuracy of each tool (VGG with and without considering sentiment classification neutral).

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

Tabela 2. Number of texts classified according to feeling and lexicons used. Percentages represent the accuracy of each lexicon.

Polarity	GermanPolarityClues	Rauh's dictionary	LIWC
Positive	184	183	163
Neutral	44	27	49
Negative	49	51	25
Total	277	261	237
% hit	30.61	28.84	26.19

out VGG that does not provide this sentiment class). SentiBank and DeepSentiBank falsely classified most of the neutral images as positive and second-most as negative. One abnormality is that the subjects working on Appen mostly classified images of food as neutral, while SentiBank and DeepSentiBank labeled them as positive. We apply the lexicons to the remaining 905 texts in CGB and compare with GT, shown in Table 2.

CGB contains 184 texts with GT negative, 506 with neutral, and 215 with positive out of 905 posts. The three lexicons predicted fewer texts belonging to the neutral class. GermanPolarityClues, Rauh's dictionary, and LIWC classified most of the neutral texts as positive.

5. Contradictions in Texts

Out of 950 blog posts, 427 have mixed feelings when the image and text were presented separately to the subjects (Figure 1). However, when perceiving the blog post as a whole (image and text together), only 64 were classified as contradictory. This indicates that the feeling for an image or text depends on its context, meaning the perception about an image can be different when there is also a text given. We examine the semantic structure of the texts to understand what can induce these contradictory perceptions. Marneffe et al. [De Marneffe et al. 2008] divide contradictions into two typologies. In the first category, contradictions appear through negations, antonyms, and numeric mismatches. In the second category, contradictions are harder to detect since linguistic models are needed to interpret the whole sentences. Therefore, we chose the first category for our analysis. The CGB corpus contains 57 posts labeled as contradictory. We count each word class's occurrence for these posts: punctuations, conjunctions, adverbs, negations, and antonyms. Then we count the occurrences of the word classes mentioned above in each text and calculate the arithmetic mean for each class. Results show the mean percentage of words in each class for contradictory and consistent posts. Each word class in the contradiction category shows a value of 0.2% to 1.3% higher than in the consistent category. These results confirm that punctuation, conjunctions, adverbs, negations, and antonyms lead to perceived contradictions in a text. For proving that contradictions can cause low accuracy in machine-performed sentiment analysis, we show Figure 2. The subjects classified

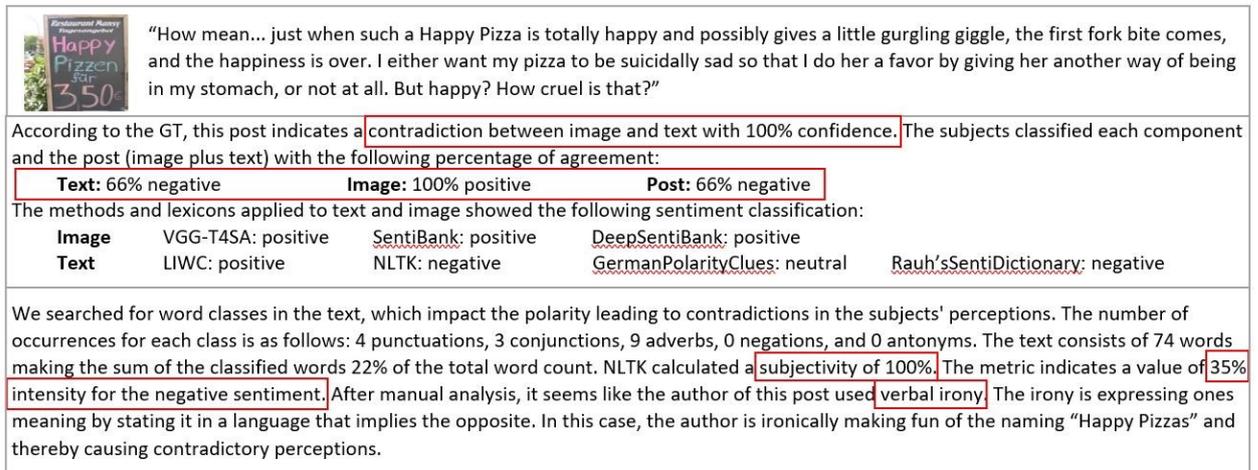


Figure 2. Analysis of a contradictory post classified as negative (German text was translated into English).

image and text of this post with different sentiments when presented separately. Nonetheless, the subject did classify the image and text of the post as coherent when presented with the whole blog. We used different tools for the semantic analysis of this post. The Hanover Tagger [Wartena 2019] detected punctuations, conjunctions, and adverbs. We used a self-created list of all German negations to search them in the text. The obtained GermaNet [Hamp and Feldweg 1997][Henrich and Hinrichs 2010] resource helped us to find antonyms. With the Natural Language Processing with Python Tool (NLTK), we calculated the post’s subjectivity.⁵ The intensity of the text’s polarity was computed with the German adaptation of the VADER sentiment analysis tool [Tymann et al. 2019].

6. Comparison with the Analysis of Brazilian Blogs

All of the image classifiers performed better on CBB. SentiBank shows a 12.35%, DeepSentiBank a 16.47%, and VGG a 6.65% higher accuracy concerning GT. VGG performed about 0.52% better on German blogs, but only when neutrals were not considered. In both analyses, the highest agreement is achieved on as positive-labeled images. LIWC shows a 17.1% lower accuracy for texts in CGB. We cannot compare the results obtained by the lexicons GermanPolarityClues and Rauh’s dictionary with OpLexicon and the SentiLex because the parameters are not the same. Still, the Brazilian lexicons perform 14.04% - 16.61% better than the German ones. For both images and text, the classifiers provide the lowest accuracy for negative content from CBB and neutral content for CGB.

In CBB, the subjects classified 422 out of 880 posts as having contradictory sentiments when presented separately with image and text. However, when presented with the whole post, only 29 were clearly labeled as contradictory, meaning that 3% of the 422 posts reveal a divergence of image and text sentiment. In the data annotation process of CGB, 427 out of 950 posts showed a sentiment contradiction between image and text. However, the subjects only classified 64 posts as clearly exposing contradictory sentiments, meaning that 7% of the 427 posts contain inconsistent sentiments according

⁵A high level indicates subjective and personalized texts, while a low one means objective and uncommitted writing.

to the subjects' perceptions. These values expose higher confidence in perceiving posts as contradictory for CGB compared to CBB. Most contradictory posts and most non-contradictory posts of CBB display a subjectivity of between 40% and 60%, while the least posts have one lower than 40%. For CGB, it is precisely the opposite: both categories have the most posts with a subjectivity of lower than 40%, and the least number of posts within the 40% to 60% range.

Hofstede [Hofstede 2001][Hofstede 2011] categorizes dimensions for distinguishing national cultures. Indulgence is defined as *gratification of basic and natural human drives related to enjoying life and having fun*.⁶ We used Hofstede Insights to compare the Indulgence dimension of Brazil and Germany.⁷ Brazil has a score of 59 points and is therefore classified as an indulgent society. A high score in this dimension means that the population tends to be impulsive and optimistic. Germany has only 40 indulgence points. They tend to be pessimistic and to act in accordance with social norms. In the CGB (GT) dataset, when a blog post expressed neutral or positive feelings of equal intensity in the text and image through the predictions of technological and lexical tools, the Germans preferred to choose the neutral feeling. This perception supports the idea that the Germans tend to act in moderation, suppressing their desires. For CBB analysis, technologies and lexicons performed poorly on images and texts with a negative rating. Brazilians, on the other hand, prefer to classify images or texts as neutral or positive. This indicates that Germans are very concerned about being criticized for their opinions in public blog posts. According to the subjectivity analysis that is between 40% and 60% in the CBB dataset, Brazilians are more likely to release their impulses, resulting in more emotional and informal blog posts.

7. Conclusion

This article proposed the corpus of German Cross-media Blogs, which consists of online blog posts in the German language. We obtained a relatively low precision according to the GT: the techniques used to classify images reached an accuracy of 26.85 % for network 1, 24.09 % for network 2 and 26.19 % (68.30 %) for network 3. The lexicons obtained a score of 30.61 % for lexicon 1, 28.84 % for lexicon 2 and 26.19 % for lexicon 3. The technologies used for image analysis in this work are not specialized for use in social media in the German context and the German lexicons do not consider dialects or slang. The use of irony and humor creates contradictory expressions of emotion in images and texts, which can be more easily interpreted by humans than by technological classifiers. The analysis of cultural differences can be an indicator to explain why Brazilians and Germans show more positive and neutral sentiments, respectively, on crowdsourcing platforms. We suspect that other cultural differences, such as linguistic dimensions, may impact the differences in performance presented in both articles and may be potential factors for future work.

Referências

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.

⁶<https://hi.hofstede-insights.com/national-culture>

⁷<https://www.hofstede-insights.com/country-comparison/brazil,germany/>

- Camras, L. (1980). Emotion: a psychoevolutionary synthesis.
- Chen, T., Borth, D., Darrell, T., and Chang, S.-F. (2014). Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks. *arXiv preprint arXiv:1410.8586*.
- 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.
- De Marneffe, M.-C., Rafferty, A. N., and Manning, C. D. (2008). Finding contradictions in text. In *Proceedings of ACL-08: HLT*, pages 1039–1047.
- Hamp, B. and Feldweg, H. (1997). Germanet-a lexical-semantic net for german. In *Automatic information extraction and building of lexical semantic resources for NLP applications*.
- Henrich, V. and Hinrichs, E. (2010). Gernedit-the germanet editing tool. In *Proceedings of the ACL 2010 System Demonstrations*, pages 19–24.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications.
- Hofstede, G. (2011). Dimensionalizing cultures: The hofstede model in context. *Online readings in psychology and culture*, 2(1):2307–0919.
- 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.
- 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).
- Moraes, S. M., Santos, A. L., Redecker, M., Machado, R. M., and Meneguzzi, F. R. (2016). Comparing approaches to subjectivity classification: A study on portuguese tweets. In *International Conference on Computational Processing of the Portuguese Language*, pages 86–94. Springer.
- 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.
- Rauh, C. (2018). Validating a sentiment dictionary for german political language—a workbench note. *Journal of Information Technology & Politics*, 15(4):319–343.
- 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.
- Tymann, K., Lutz, M., Palsbröker, P., and Gips, C. (2019). Gervader-a german adaptation of the vader sentiment analysis tool for social media texts. In *LWDA*, pages 178–189.
- 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.
- 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.
- Wartena, C. (2019). A probabilistic morphology model for german lemmatization.
- 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*.