Sarcasm Detection on Flickr Using a CNN

Dipto Das
Missouri State University
Springfield, Missouri, USA
dipto175@live.missouristate.edu

Anthony J. Clark
Missouri State University,
Springfield, Missouri, USA
anthonyclark@missouristate.edu

ABSTRACT

Sarcasm is an important aspect of human communication. However, it is often difficult to detect or understand this sentiment because the literal meaning conveyed in communication is opposite of the intended meaning. Though the field of sentiment analysis is well studied, sarcasm has often been ignored by the research community. So far, to detect sarcasm on social media, studies have largely focused upon textual features. However, visual cues are an important part of sarcasm. In this paper, we present a convolutional neural network based model for detecting sarcasm based on images shared on a popular social photo sharing site, Flickr.

CCS Concepts

Human-centered computing → Collaborative and social computing → Collaborative and social computing design and evaluation methods → Social network analysis

Computing methodologies → Machine learning → Machine learning approaches →Neural networks

Information systems → Information retrieval → Specialized information retrieval → Multimedia and multimodal retrieval → Image search

Keywords

Sarcasm, Flickr, CNN

1. INTRODUCTION

In this age of social networking, a large number of people communicate via Social Networking Sites (SNS). These sites have thus become a large deposit of crowd sourced data such as text, images and videos. Besides being large in size, the most important fact about the data on these SNSs is that they are unbiased, i.e., when many people or users on social media post contents on these sites it is not likely that the contents will reflect the views of a single group of people. This large amount of data can be helpful to find out the common patterns of users' interaction on SNSs. Insights about the pattern of such interaction can be helpful to build automated systems to detect mode of the communication of users on social media. In recent years, BIG data from social networking sites has been exploited to find out the pattern of human interaction on computer based systems for application in sentiment analysis (e.g., positive and negative sentiments), conversation generation (e.g., Microsoft's social AI Zo), etc.

ICCBD '18, September 8–10, 2018, Charleston, SC, USA

ACM ISBN 978-1-4503-6540-6/18/09…$15.00
DOI: https://doi.org/10.1145/3277104.3277118

Though sarcasm is an interesting part of communication, on SNSs, it has long been avoided by researchers because it has been thought to be too complex. Sarcasm, by its nature, is a form of sentiment expression where the surface sentiment differs from the implied sentiment. Sarcasm is the use of words that mean the opposite of what you really want to say especially in order to insult someone, to show irritation, or to be funny [2]. According to Grice's Maxims [3], there are two major principles for cooperative dialogue: the maxim of quality and the maxim of manner. The maxim of quality states that one tries to be truthful and does not give information that is false or that is not supported by evidence. The maxim of manner says that one tries to be clear as one can in what one says avoiding obscurity and ambiguity. According to Tepperman et. al. [23], sarcastic speech always violates at least of one of Grice's maxims for cooperative dialogue. Existing studies in the field of sarcasm detection are mainly based on text. Since sarcasm violates the principles of cooperative dialogue, the concepts of sequence to sequence model in NLP does not apply. One important part of sarcasm is visual cues, which are often missing in writing or incomprehensible from text. However, existing approaches of sarcasm detection using images depends on semantic representation of the images [20] and uses the sentiment conveyed by the corresponding text for comparison.

In our work, we focus on the visual style of images to express sarcasm. Our hypothesis is that most often the images contain such visual cues that can indicate whether the message conveyed by the post is sarcastic or not. In such cases an individual who views the image has access to the complete information or content shown by the image and its context. Therefore, images can be a good predictor of sarcasm in shared content on social media. SNSs enable users to communicate with a large variation of users. They vary from each other with respect to nationality, location, language, etc. Hence, users often use tags with their shared content to make those better understandable to the others. They often use these tags as keywords that can be used to search the photos based on their contents.

We used the images collected from a popular photo sharing site, Flickr, as our dataset. We used tags associated with photos as labels. Since tags were assigned by users who posted the photos on Flickr, there is almost no possibility of wrong annotation. We have collected a binary classification dataset for sarcasm detection with the images from Flickr and compared the quality of our dataset with that of the other existing datasets. Aside from the dataset, the main contribution of this paper is that we have proposed a convolutional neural network based sarcasm detection system that uses only images to detect sarcasm with 84% accuracy.

Our achieved accuracy with this CNN model is similar to the semantic representation based approach. However, our approach requires less information about the post. Unlike the semantic representation based approach which requires both text and image data as input to the SVM model, our proposed CNN model requires only image input data and works on visual cues of images as a whole instead of the semantic representation. The rest of this paper is organized as follows: section 2 discusses the related works in existing literature; section 3 describes the data collection step of this research; section 4 describes the research methodology and the reasoning involved in our CNN model; the next section presents the results—evaluation of the dataset and the model; after that we discuss our plan for future work followed by the conclusion.

2. RELATED WORKS

The existing works in the field of sarcasm detection can be divided into three categories: theories about the construct of sarcastic contents on social media from the linguistic and psychological point of view; corpus/dataset generation for sarcasm detection; and machine learning and pattern recognition based approaches to detect sarcasm.

Tepperman et. al. [23] (2006) introduced the first paper discussing the problem of recognizing sarcasm. They presented experiments to recognize sarcasm using prosodic, contextual, and spectral cues. Given the limited capabilities of NLP at that time, they took a naïve approach to detect sarcasm from online text: search for sentences that contained the phrase "yeah right." Filatova et. al. [6] discussed the terms "irony" and "sarcasm" in their work. They opine that sarcasm always has positive literal meaning, negative intended meaning and clear victims. Bamman et. al. [4] focused on the importance of context in case of sarcasm detection. According to the study [4], including extra-linguistic information from the context of an utterance on Twitter—such as properties of the author, the audience and the immediate communicative environment—contribute to detection of sarcasm. However, since our approach does not require any linguistic or text input, we did not consider context as they did in [4].

A large number of studies on sarcasm detection focused on dataset (or corpus) generation. Gonzalez-Ibanez et. al. [11] created a corpus that includes only sarcastic utterances that have been explicitly identified as such by the message writer. They also presented a report on the difficulty of distinguishing sarcastic tweets from tweets that are traditionally positive or negative sentiment. They investigated the impact of lexical and pragmatic factors on machine learning effectiveness for identifying sarcastic utterances. Khodak et. al. [13] introduced a self-annotated—labeled by the author of the statement himself/herself, not an independent annotator—Reddit corpus for sarcasm research containing 1.3 million sarcastic statements. They evaluated the corpus for accuracy, constructed benchmarks for sarcasm detection, and evaluated baseline methods. They had three major metrics of interest for evaluating the corpora: (1) size, (2) the proportion of sarcastic to non-sarcastic comments, and (3) the rate of false positives and false negatives. Their work did not have any specific finding, rather the collection and evaluation of a large dataset was their main contribution.

Extending this finding from Filatova et. al. [6], Riloff et. al. [17] discovered that the contrast between positive and negative sentiment yielding words or phrases in the same tweets can be an indicator of sarcasm in a tweet. The idea of contrasting sentiments as a denotation of sarcasm was later followed by many other later studies which used support vector machine based supervised classification, and that used binary logistic regression with L2 regularization for the sarcasm detection task [5], [14], [9], [4]. Ghosh et. al. [8] approached the task of sarcasm detection with the use of neural networks. They focus mainly on semantic representation of the sentences. The proposed neural network model composed of convolution neural network (CNN) and followed by a long short term memory (LSTM) network, and finally a deep neural network (DNN). They used a Latent Semantic Analysis (LSA)-based approach to extend the list of indicative hash-tags.

As we can see most of the existing works in sarcasm detection utilizes the textual features of a content on social media. In fact, the importance of multimodality for sarcasm detection was not discussed before Schifanella et. al. [20]. In their paper, they emphasized the importance of multimodality consideration for detecting sarcasm by investigating the relationship between textual and visual aspects in multi-modal posts. They ran a crowd-sourcing task, to quantify the extent to which images are perceived as necessary by human annotators in which the authors asked the users on a website (crowdflower.com) to evaluate the impact of visuals as a source of context for humans. They showed the positive effect of combining modalities for the detection of sarcasm across various platforms and methods. Being inspired by the positive results they obtained, we looked into some of the existing literatures in sentiment analysis that uses image data for their systems. Gajarla et. al. [7] experimented with various classification methods on their dataset SVM on high level features of VGG-ImageNet, fine-tuning on pre-trained models like RESNET, Places205-VGG16 and VGG-ImageNet; and reported 73% accuracy for positive and negative sentiment detection. Siersdorfer et. al. [21] considered the bag-of-visual words representation as well as the color distribution of images. They performed a discriminative feature analysis based on information theoretic methods, and apply machine learning techniques to predict the sentiment of images. In our work, we focus on the visual content of images and their effectiveness to predict if an image conveys sarcastic message. Our work has two stages: dataset generation and classification. We have collected an image dataset using the creative commons licensed images from Flickr, and we have built a convolutional neural network (CNN) based model utilizing the visual contents of images to predict sarcasm on social media.

3. DATA COLLECTION

Collecting a dataset for sarcasm detection is a challenging task. It consists of two parts: (1) recognizing sarcastic posts online and (2) differentiating those from non-sarcastic posts. One tempting approach is to use human annotators to label the dataset. This approach is followed by Riloff et. al. [17] and Swanson et. al. [22]. As Khodak et. al. [13] pointed out, a problem with the aforementioned approach is that to understand sarcasm one needs to have sufficient knowledge about the context in which the statement (or image) was made (or posted). Without this context information, human annotators cannot label the statements as "sarcasm" or "non-sarcasm" accurately. Khodak et. al. [13] presented such arguments for text based data only. Later, Schifanella et. al. [20] also said that text and image can serve as the source of complimenting context information for each other.

From the information available along with the post on social media, it might be possible for a human annotator to get some clues about the conversation but only the persons posting those contents on SNS can have the full context or background. Because sometimes two people familiar to each other may not use hash-tags with (e.g. #sarcasm) their posts explicitly. They can communicate with each other without such hash-tags since they have prior knowledge about one another. However, the fact that these users are familiar with each other may not be available to the external annotator which might lead the annotator to have the wrong impression. Therefore, many of the following studies like Khodak et. al. [13] and Reyes et. al. [16] used self-annotated tweets. In our study, we have followed the latter approach of using self-annotated data. We used the tags that the users provide with their photos while posting those Flickr.

Sarcastic posts are not often used during normal communication. Hence, despite the large amount of data available on SNSs, it is relatively difficult to find posts that contain sarcasm. Snowball sampling [12] is a commonly used technique in social computing and statistics research to address this issue. It is a non-probabilistic sampling technique. It is suitable to use when the members of a population are hidden and difficult to locate [10]. Since posts containing sarcasm are hard-to-find study subjects, we used the snowball sampling technique to collect our data. Since we are looking for the social media posts that contains sarcasm, the word "sarcasm" itself is a study keyword in our research. Borrowing from the idea of snowball sampling, the other synonyms of this words are potential study samples since they have the same meaning as this word. We chose the following words: "sarcasm," "sarcastic," "irony," "satire," and "wit" as indicators of a post being sarcastic. According to Filatova et. al. [6], sarcasm has a positive literal meaning but negative intended meaning, and has clear victims. That means, sarcastic posts are often confused with positive statements, which are addressed to a specific group or person. Positive statements, or admiration, addressing a specific group or person can be termed as praise. Therefore, we considered the words "praise" and its synonyms for collecting samples of data that can be confused with sarcasm but in reality that are not sarcastic. Since we focus on posts containing sarcasm and the literal meaning and the intended meaning of such posts are opposite, these kind of posts can mistakenly be thought as facts. For this reason, we used the word "fact" to collect the sample data for the non-sarcastic group. Therefore, to collect data samples for non-sarcastic posts we used the words: "non-sarcastic," "nonsarcastic," "not-sarcasm," "notsarcasm," "praise," "applause," "fact," and "information." We assumed that the tags are not used ironically. Instead they can be used to clarify confusion if necessary in such cases where a fact or praise might sound like sarcasm. Therefore, if an image has tags from both of the "sarcasm" and "non-sarcasm" classes, we discard the image.

Flickr is a popular social photo sharing service. The image data publicly available from this platform has been widely used in social media research, such as by Gajarla et. al. [7] for emotion classification, and by Siersdorfer et. al. [21] for sarcasm detection. We used the Flickr API service to collect photos. We queried image meta-data (image ID, image url, etc.) using the words discussed above as keywords or search query parameters in both "sarcasm" and "non-sarcasm" categories. We have collected all the images returned by each keyword. We have listed the number of images that we found for each keyword in Table 1. In most cases, images used more than one keyword from the same category. We considered the image only once in such cases i.e., when an image had more than one keyword from the same class as its tags, we have saved the image in our dataset only once. For our dataset, we only collected the metadata of such images available under the creative commons license. We sorted the result set by interestingness [1] in the same way it was done by Gajarla et. al. [7]. After the image metadata was obtained, it was possible to use the metadata to directly access the images from Flickr servers instead of going through the Flickr API.

Table 1. Number of images for each keyword individually

Keyword	Number of images
sarcasm	13
sarcastic	22
irony	179
satire	88
wit	141
Total "sarcasm" images	**443**
non-sarcastic	2
praise	437
applause	253
fact	35
information	676
Total "non-sarcasm" images	**1403**

As pointed out by Wallace et. al. [25], sarcasm occurs very infrequently. This finding was also reflected in our data collection. For our dataset, though we wanted to collect as many as images possible using the aforementioned keywords. For this study, we collected 1846 images in total. Among those images, 443 images (23.998%) were retrieved with the keywords from "sarcasm" class and 1403 images (76.002%) were retrieved with the keywords from 'non-sarcasm' class. We split the data in each category into 90% training and 10% validation set. Thus, we had a training set with 1603 images of which 399 images were from "sarcasm" class and 1263 images were from "non-sarcasm" class; and a validation set with 184 images of which 44 images were from "sarcasm" class and 140 images were from "non-sarcasm" class. We refer to our dataset as the Yahoo Flickr Sarcasm (YFS) dataset.

In our developed YFS dataset, we saved all the images in jpeg format. There are images of different sizes and shapes. The largest image in the dataset is 20.0 MB and the smallest image in the dataset is 8.0 KB. The images also vary in their resolutions: from 180x135 to 6600x4514. The largest dataset so far for sarcasm detection from text data is developed by Khodak et. al. [13]. They have also proposed a benchmark for the sarcasm detection dataset in their paper. In the data they collected from subreddit 'politics' had 23.2% instances of "sarcasm" category. The Internet Argument Corpus used by Walker et. al. [24] as a source of sarcastic comments had only 12% of the total corpus as positive instances of "sarcasm" category. The only existing image based sarcasm detection dataset by Schifanella et. al. [20] is a balanced dataset of "sarcasm" and "non-sarcasm" data samples with 50% samples in each category. Among all instances, 91.16% contained images. Considering an equal distribution of the samples between both categories, 45.58% samples in their dataset are instances of image data of "sarcasm" class. Therefore, we can see that the proportion of data samples in our developed dataset satisfies the benchmark determined by Khodak et. al. [13] and its distribution is also comparable to the dataset developed by Schifanella et. al. [20].

4. METHODOLOGY

We split our dataset into two categories: "sarcasm" and "non-sarcasm". Therefore, we can describe our task of sarcasm detection as binary classification. Since feature extraction from the images is a difficult task and being unable to choose good hand-crafted feature may result into a poor classifier, we are not doing feature detection. Instead we decided to use a convolutional neural network (CNN) for this task.

Schifanella et. al. [20] in their work for the first time discussed the importance of multimodality, i.e. utilizing both text and image data for detecting sarcasm unlike the previous works on sarcasm detection which only depended on one type of data, text. Besides the textual descriptions of images, they attempted to understand the semantics in the images. They reported that the semantics of images improved the performance of the sarcasm detection. We use their proposed concept of the importance of considering images as an important indicator of sarcasm on social media and extended the concept. Instead of semantic representation of images, we focused on the visual cues of the images.

We hypothesize that images posted on social media with sarcastic intents have different visual cues or representations than ones that do not have sarcastic intentions. The difference between our hypothesis and the finding of Schifanella et. al. [20] is that we focus on the visual cues or the way of representation of an image to detect sarcasm whereas they focused on the semantics of the images. Though in sentiment analysis research, Gajarla et. al. [7] and Siersdorfer et. al. [21] have used visual contents to predict sentiments. However, they have not considered sarcasm as a mode of sentiment based communication. In our paper, we propose that visual contents or cues in an image as input to CNN can detect sarcasm with high accuracy.

Neural networks are inspired by the information processing of one or more neurons. An artificial neural network (ANN) is a computational model based on the structure and functions of biological neural networks. The structure of an ANN can be thought of as a weighted graph. Information flowing through the ANN network does not change the structure of the network but it updates the initial weights of the edges of the network with the process of backpropagation. A CNN comprises several convolution layers, often with a sub-sampling step, and then followed by one or more fully connected layers. Let's define a group of layers: a 2D convolution layer, an activation layer with *relu* function, and a max-pooling layer with pool size = (2, 2), as "Layer group A". The structure of the neural network that we used is as follows: repeating "Layer group A" three times, a flatten layer, two dense networks with *relu* and *sigmoid* functions respectively.

Since the size of our dataset is not large, we performed some image augmentation process. We passed the images through a shearing transform by a factor 0.2 and a zooming transform by a factor 0.2. These values were arbitrarily chosen from previous experience of the authors of working with small datasets. We also performed horizontal flips on the images to increase the number of training samples. Moreover, our original images consist of RGB values in range [0, 255] but such values would be too high for our model to process given a typical learning rate. Hence, we scale values within a range [0, 1]. Using these image augmentation techniques we generated a dataset of 2000 images consisting of all the original images and a number of synthetic images. We performed data augmentation techniques on the validation images as well. This process left us with validation dataset of 800 images.

To train the network, we tried several different values for batch size of images passed to the CNN. The larger the batch size is, the more memory the CNN will require to be trained. At the same time larger batch size will increase the training time. On the other hand, smaller batch size will require less memory and less time to train. However, though it seems tempting to choose a small batch size, it makes the estimate of gradient to be less accurate. Therefore, there is a trade-off during the choice of the value of batch size for the training of CNN. For the final model that we used the value of batch size was set to 16. Another important parameter of training CNN is the number of epochs, i.e. how many times all the training samples will go through a pass/ cycle of a forward pass and a backward pass. We have used 50 epochs for training our CNN model.

We developed our implementation of CNN using Keras. We used an input size of 240x240. We used the *binary_crossentropy* loss function for our model. We are dealing with image data in our study. Przelaskowski et. al. [15] discusses the importance of sparse representation of image data. According to Ruder et. al. [18], for sparse input data, using one of the adaptive learning rate methods is likely to achieve the best results. In this way, one does not need to tune the learning rate as it is likely to achieve the best result using the default value. Adagard is an algorithm for gradient based optimization that adapts the learning rate to the parameters. It is well suited for using with sparse data. In our study we chose to use the RMSprop optimizer that is an extension of Adagard which deals with its radically diminishing learning rate. For optimizing the training of our model, at first we used accuracy as the metric. Since our dataset is not balanced in the amount of images of the two categories: "sarcasm" and "non-sarcasm," we also optimized another model using F1 score (harmonic average of precision and recall) [19] as the metric. For the optimized model for F1 score all the other parameters and choices were the same as the first one. For the first model, we achieved an accuracy of 84% and for the second model we achieved an F1-score of 79%. A high F1-score along with a high accuracy indicates that the model is not biased towards any category.

5. RESULTS

5.1 Evaluation of Model

We compare the performance of our CNN model with the performance of the only image based sarcasm detection model in the existing literature. To the best of our knowledge, Schifanella et. al. [20] is the first and to date the only paper which discusses a way to use images for sarcasm detection. Schifanella et. al. [20] reported that using only the visual semantic features yields an accuracy 65% and 72% for their two evaluation datasets respectively using visual semantic features. Gajarla et. al. [7] used the visual content of image in their study to analyze sentiment. They achieved 67.8%, 68.7% and 73% accuracy by fine-tuning VGG-Imagenet, VGG-Places205 and ResNet50, respectively, for sentiment analysis task. We used visual content of images for training our CNN model for detecting sarcasm which performs with 84% accuracy. Therefore, we can say that the visual content of the images can be a good feature to train a sarcasm detection model to have promising results.

5.2 Discussion

In this section, we examine the results of our model to get a better understanding of what our model is learning. Based on the results, we can draw the following conclusions about what the model is learning for each category.

Sarcasm. The model seems to learn images with writing in it for this category. Many of the images in this category seem to be crudely edited "memes". A lot of outdoor images of non-human objects also were predicted to belong to this category by our model, as shown in Figure 1a. Our model learns to identify almost all images that have hand-drawn cartoons in them to be instances of sarcasm category.

(a) Sarcasm

(b) Non-sarcasm

Figure 1: Examples of images from "sarcasm" and "non-sarcasm" classes

Non-sarcasm. The model predicts indoor images to belong to non-sarcasm category. Though the model predicts a lot of outdoor images with humans in them as instances of this class, the model seems to label the indoor images of all kinds to be instances of non-sarcasm class. The model seems to learn human faces to be a good indicator of images to be member of non-sarcasm category, as in Figure 1b.

6. CONCLUSIONS

We plan to incorporate the textual features with the image based features to enrich the feature set for the sarcasm detection task. Our working on the extension of the system to utilize other metadata of a post on SNS is on progress. Another future plan is to assess how the sarcasm in a post is related to the popularity of it on the SNSs.

We have developed an image dataset for sarcasm detection and a CNN model trained on that dataset. Both the dataset and the model are made publicly available at this link: goo.gl/JG4xd3. We have compared the quality of our dataset with the quality of the previous studies. We have also discussed our dataset's standing with respect to the benchmark for sarcasm detection dataset proposed by previous studies. We hope that future researchers will be benefited by using our developed self-annotated image dataset for sarcasm detection. We have also developed a CNN model that can detect sarcasm from images with 84% accuracy.

7. REFERENCES

[1] 2005. Flickr: Explore interesting contens around Flickr. Retrieved May 19, 2018 from https://www.flickr.com/explore/interesting/

[2] 2018. Definition of Sarcasm by Merriam-Webster. Retrieved May 19, 2018 from https://www.merriam-webster.com/dictionary/sarcasm

[3] 2018. Grice's Maxims. Retrieved May 19, 2018 from https://www.sas.upenn.edu/ ~haroldfs/dravling/grice.html

[4] David Bamman and Noah A Smith. 2015. Contextualized Sarcasm Detection on Twitter.. In Proceedings of the Ninth International AAAI Conference on Web and Social Media. 574–577.

[5] Mathieu Cliche. 2014. The sarcasm detector. Retrieved May 19, 2018 from http://www.thesarcasmdetector.com/

[6] Elena Filatova. 2012. Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing.. In International Conference on Language Resources and Evaluation. Citeseer, 392–398.

[7] Vasavi Gajarla and Aditi Gupta. 2015. Emotion Detection and Sentiment Analysis of Images. Georgia Institute of Technology (2015).

[8] Aniruddha Ghosh and Tony Veale. 2016. Fracking sarcasm using neural network. In Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. 161–169.

[9] Tamal Ghosh. 2016. Sarcasm Detection - Machine learning perspective. Retrieved May 27, 2018 from https://www.sas.com/content/dam/SAS/en_us/doc/event/analytics-experience-2016/sarcasm-detection-machine-learning-perspective.pdf

[10] Lisa M Given. 2008. The Sage encyclopedia of qualitative research methods. Sage Publications.

[11] Roberto González-Ibáñez, Smaranda Muresan, and Nina Wacholder. 2011. Identifying sarcasm in Twitter: a closer look. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers. Volume 2. Association for Computational Linguistics, 581–586.

[12] Leo A Goodman. 1961. Snowball sampling. The annals of mathematical statistics (1961), 148–170.

[13] Mikhail Khodak, Nikunj Saunshi, and Kiran Vodrahalli. 2017. A large self-annotated corpus for sarcasm. arXiv preprint arXiv:1704.05579 (2017).

[14] Chun-Che Peng, Mohammad Lakis, and Jan Wei Pan. 2015. Detecting Sarcasm in Text. (2015).

[15] Artur Przelaskowski. 2010. The role of sparse data representation in semantic image understanding. In International Conference on Computer Vision and Graphics. Springer, 69–80.

[16] Antonio Reyes, Paolo Rosso, and Tony Veale. 2013. A multidimensional approach for detecting irony in twitter. Language resources and evaluation 47, 1 (2013), 239–268.

[17] Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, and Ruihong Huang. 2013. Sarcasm as contrast between a positive sentiment and negative situation. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 704–714.

[18] Sebastian Ruder. 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747 (2016).

[19] Yutaka Sasaki et al. 2007. The truth of the F-measure. Teach Tutor mater 1, 5 (2007).

[20] Rossano Schifanella, Paloma de Juan, Joel Tetreault, and Liangliang Cao. 2016. Detecting sarcasm in multimodal social platforms. In Proceedings of the 2016 ACM on Multimedia Conference. ACM, 1136–1145.

[21] Stefan Siersdorfer, Enrico Minack, Fan Deng, and Jonathon Hare. 2010. Analyzing and predicting sentiment of images on the social web. In Proceedings of the 18th ACM international conference on Multimedia. ACM, 715–718.

[22] Reid Swanson, Stephanie Lukin, Luke Eisenberg, Thomas Chase Corcoran, and Marilyn A Walker. 2017. Getting

reliable annotations for sarcasm in online dialogues. arXiv preprint arXiv:1709.01042 (2017).

[23] Joseph Tepperman, David Traum, and Shrikanth Narayanan. 2006. " Yeah Right": Sarcasm Recognition for Spoken Dialogue Systems. In Ninth International Conference on Spoken Language Processing.

[24] Marilyn A Walker, Jean E Fox Tree, Pranav Anand, Rob Abbott, and Joseph King. 2012. A Corpus for Research on Deliberation and Debate. In International Conference on Language Resources and Evaluation. 812–817.

[25] Byron C Wallace, Laura Kertz, Eugene Charniak, et al. 2014. Humans require context to infer ironic intent (so computers probably do, too). In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Vol. 2. 512–516.