Detection and analysis of metaphor in the context of the COVID-19 pandemic

Nazanin Sabri

Abstract

People often use words metaphorically rather than using them with their literal meanings. This way of using speech allows for a more natural and more exciting exchange of ideas and conversation. We can see metaphorical use of speech in all types of media including, books, news, and social networks. Thus, to be able to understand and process natural language properly, we need to be able to detect when words are being used metaphorically. In this study, we perform metaphor detection in the word and sentence level using the help of BERT, a subword level embedding, in the English language. Our model achieves an accuracy of 76.87% after being run via the 5-fold cross validation method on the datasets. After the model has been trained we use the model to analyze tweets that have been published with regards to the COVID-19 pandemic with the goal of finding how the metaphorical use of language has changed as the disease has progressed. We report the changes in the rate of metaphor usage in COVID-19 tweets, as well as the correlation of metaphorical use of language and the sentiment of the text. Additionally, we report the words that were most commonly used in a non-literal sense.

Introduction

The massive amounts of textual data published online every day has made computational language understanding a necessity. One issue that computers face is their inability to detect and interpret metaphorical language. It has been reported that 7.7% of the words used in conversation and 16.4% of those used in the news are used metaphorically [31]. Based on the statistics, we can see that the use of such constructs is not negligible. As a result, a successful metaphor detection system can help in language understanding as well as other tasks such as machine translation and sentiment analysis. These systems can also help extract high level information from texts. Shared tasks [6, 7, 18] have been defined in this field of study, which further proves the importance of the task.

Earlier attempts at metaphor detection were conducted by designing features that highly correlate with metaphors and are good indications of the presence of metaphors in sentences. Such methods, however, require an in-depth analysis of languages and understanding of language specific constructs. As a result, recent methods have focused on deep learning or neural network-based methods. There have also been a number of studies on cross-lingual and multilingual detection methods which allow for the detection of metaphors in languages that have low resources on their own.

In this study, we will be focusing on the task of metaphor detection in the English language. We will first present models to detect this attribute in the sentence level. Then we will perform additional analysis and attempt to detect metaphors in the word level. We then use the trained models to label the COVID-19 related tweets and analyze the relationship between the use of metaphor and the sentiment of the tweet. The reason why we try to analyze the COVID-19 data with regards to metaphor usage is that it has previously been shown that metaphors can often be a medium for the expression of emotions [36]. Since the COVID-19 pandemic is a unique life event, we would like to see how users have been expressing their opinions of it and how those expressions relate to the usage of metaphors.

The rest of this paper is structured as follows. In section 2 we provide a brief review of metaphor detection models and prior studies. Next, we introduce the datasets utilized in our study and the methods used in section 3. Section 4 goes over our results and we conclude the study in section 5.

Related Work

A thorough review of metaphor detection has been conducted by [33] in 2010 where the available models and corpora were introduced and issues and challenges of the task were discussed. Then, another survey on computational metaphor detection was published in 2020 [27], where major theoretical views as well as present methods were reviewed. As a result, in this section, we will be providing a brief overview of available studies and suggest referring to the two aforementioned surveys for a comprehensive review of prior studies.

Many metaphor detection studies rely heavily on rule-based classifiers [21] or the engineering of features that can help identify metaphors. Features such as collocation lists [19], degrees of abstractness [3, 15], imageability [1], relatedness of source and targets [16], construction of target domain signatures [2], familiarity and meaningfulness [9] are just some of the features being used.

However, the methods using feature engineering can cause problems as they require additional resources in order to extract the defined features. Many papers have been published addressing this issue. For instance, in [4] a neural network method is introduced for this task and it is shown that the usage of word embeddings instead of hand-coded features can produce comparable results. Other models such as bi-LSTMs

[20, 24, 12], convolutional neural networks [23], graph convolutional neural networks [25] and other deep learning models [8] have also been examined. Another general method that has been used to embark on this issue is the use of clustering [32].
Another difficulty in metaphor detection can be the small size of available training data. A method to resolve this matter is discussed in [5]. The authors present a statistical approach, where words that do not match the typical vocabulary of text are outputted as metaphor candidates. They show that for small training corpora their method outperforms available models but that the advantage disappears when the corpora grows in size.
In parallel to endeavors to detect metaphors in the English language, attempts have also been made to perform metaphor detection on low-resource languages. To do so, one of the most important theories on metaphorical language that many studies built upon was that metaphor is a conceptual and cognitive mapping of concepts and phrases, rather than a lexical feature of a language [34]. This distinction allows for creation of language invariant features, and models that can be trained on one language which is richer in resources and then used (after the application of some techniques) on other languages. [1, 11, 14]
Models to detect metaphor in specific domains have also been developed. Identification of metaphors in poetry [30, 22], social media data [17] and discourse [10] are examples of such studies.
There have also been a number of studies on the creation of datasets to be used for the task of metaphor detection. [37] is an example of such papers. In this paper, a review of other available datasets has been provided. Some other studies have looked at the relationship between the use of metaphor and the emotional charge of the sentence. For instance in [36] it was shown that metaphorical usage of words were, on average, significantly more emotional.

Data and Methodology

Data

To detect metaphors, we used two datasets:

1. The TroFi dataset [13, 28, 29] which includes literal and metaphorical use cases of 50 different English verbs.
2. The MOH dataset [26, 35] which includes 1,639 sentences with their associated metaphorical labels.

The statistics of the data are shown in Table 1.

Table 1: Statistics of the metaphor data used in this study

Field	Value
Number of metaphor usage data/sentences	2,345
Number of literal usage data/sentences	2,688
Number of unique terms (the number of unique target words in the dataset)	473

We further use the COVID-19 tweets dataset [36] to perform our analysis. This dataset includes more than three hundred million tweets, starting from late March and continuing all the way through to August. Each tweet in the dataset has been labeled with its corresponding sentiment value. To get a better sense of the data the changes in average daily sentiment have been plotted in figure 1. Figure 2 displays the percentage of tweets published that day with a specific sentiment value. We can see that at the visualized period of time most tweets contained positive sentimental expressions.

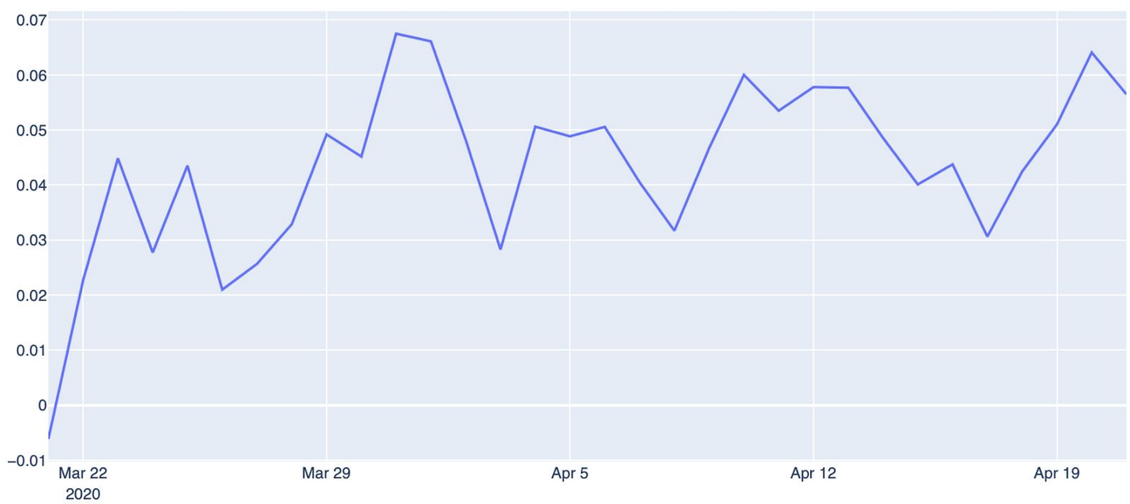

Figure 1: Average daily sentiment of tweets related to COVID-19

Figure 2: Percentage of tweets in each day with positive and negative sentiments

Methodology

We begin our analysis by performing sentence level metaphor detection. In this task a sentence has a label of “metaphor” if at least one word in the sentence is used in a metaphorical sense. In this task, however, the word is not identified and the label is assigned to the entire sentence. Table 2 displays the different models and settings which have been tested to detect metaphors at this level. Before the models are applied, however, NLTK [38] is used to clean the data and Stanza [39] is used for the task of tokenization, POS tagging and Named Entity Recognition. As shown in table 2, not all models use all the aforementioned features. Except for the models using the BERT embedding, stop word removal and lemmatization have been performed for all models.

Table 2: List of different models used to detect sentence-level metaphorical use of language

Model Name	Embedding Method	Model
SentMod1	Bag of Words	Decision Tree
SentMod2	Bag of Words	SVM
SentMod3	TF-IDF	Decision Tree
SentMod4	TF-IDF	SVM
SentMod5	BERT [40]	Neural Network (figure 3)

To use BERT pre trained embeddings (for SentMod5), we enlist the help of pre trained BERT layers via the saved tf-hub models [41]. From the available models, we use `English Uncased 12 layer model` to create the embeddings of our data. The model used to detect the use of metaphor is shown in figure 3.

```
_____________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
=============================================================================================
input_word_ids (InputLayer)     [(None, 128)]        0
_____________________________________________________________________________________________
input_mask (InputLayer)         [(None, 128)]        0
_____________________________________________________________________________________________
segment_ids (InputLayer)        [(None, 128)]        0
_____________________________________________________________________________________________
keras_layer_1 (KerasLayer)      [(None, 768), (None, 109482241   input_word_ids[0][0]
                                                                 input_mask[0][0]
                                                                 segment_ids[0][0]
_____________________________________________________________________________________________
global_average_pooling1d (Globa (None, 768)          0           keras_layer_1[0][1]
_____________________________________________________________________________________________
dense (Dense)                   (None, 768)          590592      global_average_pooling1d[0][0]
_____________________________________________________________________________________________
dense_1 (Dense)                 (None, 2)            1538        dense[0][0]
=============================================================================================
Total params: 110,074,371
Trainable params: 592,130
Non-trainable params: 109,482,241
```

Figure 3: structure of the model used in this study

For the task of word-level metaphor detection, we break down the task into two steps. In step (1) we need to detect if the sentence includes metaphors, then if it does have metaphorical use of language, then in step (2) we need to find the word in the sentence which was used in the metaphorical sense. As a result, we use the model developed for the sentence-level classification task to perform step (1).
Results of the different methods listed in this section are provided in Section 4.

Results

In this section we will look at the results of the models described in the previous section. Additionally, we will provide some insights into the relationship of sentiment and metaphor in the COVID-19 dataset. We begin by looking at the performances of the sentence-level metaphor classification models. Table 3 reports these values for our first four models. Table 4 shows the results for SentModel5 with different number of epoch values. We can see that this model outperforms all the models listed in Table 3. Figure 4 displays the change in loss values on the training data as the SentMod5 trains for 100 epochs.

Table 3: sentence-level metaphor detection results (for model specifications please read Table 2). Results are calculated using 5-fold cross validation

Model	**Accuracy**	**Recall**	**Precision**
SentMod1	49.27	41.06	52.19
SentMod2	57.9	49.9	65.1
SentMod3	48.2	35.7	49.4
SentMod4	58.6	41.4	59.5

Table 4: SentMod5 performance evaluation

Model	Number of Epochs	Accuracy
SentMod5	10	76.86
	20	74.87
	100	74.77

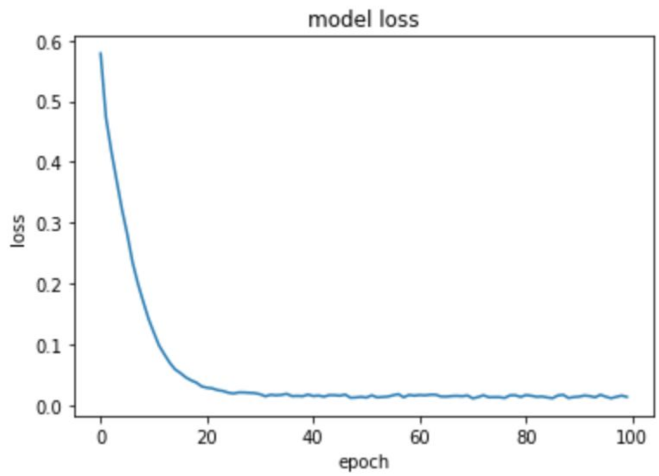

Figure 4: How the loss of the SentMod5 changes on the training data

Next we apply the *SentMod5* which has been trained for 10 epochs to the COVID-19 dataset. We find that among all the data 61.2% were classified as metaphors and the rest as literal usage of language. In table 5 we show the relationship between metaphorical use of language and its sentiment.

Table 5: statistics of the relationship between metaphorical use of language and sentiment scores

	Tweets with metaphor	Tweets without metaphor (literal)
Tweets with negative sentiment	1,394	865
Tweets with positive sentiment	4,519	2,783
Total number of tweets		9561

Conclusion and Future Work

In this study we performed metaphor detection on texts from the English language. Our sentence level detection model achieved an accuracy of 76.87%. We then used the trained models to analyze COVID-19 tweets and study the relationship between the use of metaphor and sentiment score.
Future work could focus on word level metaphorical use in the context of COVID-19 and try to find what types of metaphors words such as Coronavirus or COVID-19 are being used in.

References

[1]Tsvetkov, Yulia, et al. "Metaphor detection with cross-lingual model transfer." *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2014.
[2]Mohler, Michael, et al. "Semantic signatures for example-based linguistic metaphor detection." *Proceedings of the First Workshop on Metaphor in NLP*. 2013.
[3]Tsvetkov, Yulia, Elena Mukomel, and Anatole Gershman. "Cross-lingual metaphor detection using common semantic features." *Proceedings of the First Workshop on Metaphor in NLP*. 2013.
[4]Do Dinh, Erik-Lân, and Iryna Gurevych. "Token-level metaphor detection using neural networks." *Proceedings of the Fourth Workshop on Metaphor in NLP*. 2016.
[5]Schulder, Marc, and Eduard Hovy. "Metaphor detection through term relevance." *Proceedings of the Second Workshop on Metaphor in NLP*. 2014.
[6]Leong, Chee Wee, et al. "A report on the 2020 vua and toefl metaphor detection shared task." *Proceedings of the Second Workshop on Figurative Language Processing*. 2020.
[7] Shutova, Ekaterina, Beata Beigman Klebanov, and Patricia Lichtenstein. "Proceedings of the Third Workshop on Metaphor in NLP." *Proceedings of the Third Workshop on Metaphor in NLP*. 2015.
[8]Rei, Marek, et al. "Grasping the finer point: A supervised similarity network for metaphor detection." *arXiv preprint arXiv:1709.00575* (2017).
[9]Rai, Sunny, Shampa Chakraverty, and Devendra K. Tayal. "Supervised metaphor detection using conditional random fields." *Proceedings of the Fourth Workshop on Metaphor in NLP*. 2016.
[10]Jang, Hyeju, et al. "Metaphor detection in discourse." *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 2015.

[11]Mohler, Michael, et al. "A novel distributional approach to multilingual conceptual metaphor recognition." *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. 2014.
[12]Gao, Ge, et al. "Neural metaphor detection in context." *arXiv preprint arXiv:1808.09653* (2018).
[13]http://natlang.cs.sfu.ca/software/trofi.html, last accessed on August 4, 2020
[14]Gordon, Jonathan, et al. "High-precision abductive mapping of multilingual metaphors." *Proceedings of the Third Workshop on Metaphor in NLP*. 2015.
[15] Köper, Maximilian, and Sabine Schulte im Walde. "Improving verb metaphor detection by propagating abstractness to words, phrases and individual senses." *Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications*. 2017.
[16]Su, Chang, Shuman Huang, and Yijiang Chen. "Automatic detection and interpretation of nominal metaphor based on the theory of meaning." *Neurocomputing* 219 (2017): 300-311.
[17]Huang, Ting-Hao. "Social metaphor detection via topical analysis." *International Journal of Computational Linguistics & Chinese Language Processing, Volume 19, Number 2, June 2014*. 2014.
[18]Leong, Chee Wee, Beata Beigman Klebanov, and Ekaterina Shutova. "A report on the 2018 VUA metaphor detection shared task." *Proceedings of the Workshop on Figurative Language Processing*. 2018.
[19]Sardinha, Tony Berber. "Collocation lists as instruments for metaphor detection in corpora." *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada* 22.2 (2006): 249-274.
[20]Sun, Shichao, and Zhipeng Xie. "Bilstm-based models for metaphor detection." *National CCF Conference on Natural Language Processing and Chinese Computing*. Springer, Cham, 2017.
[21]Rai, Sunny, and Shampa Chakraverty. "Metaphor detection using fuzzy rough sets." *International Joint Conference on Rough Sets*. Springer, Cham, 2017.
[22]Kesarwani, Vaibhav, et al. "Metaphor detection in a poetry corpus." *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. 2017.
[23]Tanasescu, Chris, Vaibhav Kesarwani, and Diana Inkpen. "Metaphor detection by deep learning and the place of poetic metaphor in digital humanities." *The Thirty-First International Flairs Conference*. 2018.
[24]Bizzoni, Yuri, and Mehdi Ghanimifard. "Bigrams and BiLSTMs Two neural networks for sequential metaphor detection." *Proceedings of the Workshop on Figurative Language Processing*. 2018.

[25]Le, Duong, My Thai, and Thien Nguyen. "Multi-Task Learning for Metaphor Detection with Graph Convolutional Neural Networks and Word Sense Disambiguation." *AAAI*. 2020.
[26]http://saifmohammad.com/WebPages/metaphor.html, last accessed on August 4, 2020
[27]Rai, Sunny, and Shampa Chakraverty. "A Survey on Computational Metaphor Processing." *ACM Computing Surveys (CSUR)* 53.2 (2020): 1-37.
[28]Clustering Approach for the Nearly Unsupervised Recognition of Nonliteral Language. Julia Birke and Anoop Sarkar. In Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, EACL-2006. Trento, Italy. April 3-7, 2006.
[29] Active Learning for the Identification of Nonliteral Language. Julia Birke and Anoop Sarkar. In Proceedings of the Workshop on Computational Approaches to Figurative Language, NAACL-HLT 2007 workshop. Rochester, NY. April 26, 2007.
[30]Reinig, Ines, and Ines Rehbein. "Metaphor detection for German poetry." (2019).
[31] Steen, Gerard J., et al. "Metaphor in usage." *Cognitive Linguistics* 21.4 (2010): 765-796.
[32]Shutova, Ekaterina, Lin Sun, and Anna Korhonen. "Metaphor identification using verb and noun clustering." *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*. 2010.
[33]Shutova, Ekaterina. "Models of metaphor in NLP." *Proceedings of the 48th annual meeting of the association for computational linguistics*. 2010.
[34]Lakoff, George, and Mark Johnson. "Metaphor we live by." *Chicago/London* (1980).
[35]Metaphor as a Medium for Emotion: An Empirical Study, Saif M. Mohammad, Ekaterina Shutova, and Peter Turney. In *Proceedings of the Joint Conference on Lexical and Computational Semantics (*Sem)*, August 2016, Berlin, Germany.
[36] Rabindra Lamsal, "Coronavirus (COVID-19) Tweets Dataset", IEEE Dataport, 2020. [Online]. Available: http://dx.doi.org/10.21227/781w-ef42. Accessed: Aug. 04, 2020.
[37]Zayed, Omnia, John P. McCrae, and Paul Buitelaar. "Crowd-Sourcing A High-Quality Dataset for Metaphor Identification in Tweets." *2nd Conference on Language, Data and Knowledge (LDK 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.
[38] Loper, Edward, and Steven Bird. "NLTK: the natural language toolkit." *arXiv preprint cs/0205028* (2002).
[39] Qi, Peng, et al. "Stanza: A python natural language processing toolkit for many human languages." *arXiv preprint arXiv:2003.07082* (2020).
[40] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

[41] https://tfhub.dev/tensorflow/bert_en_uncased_L-12_H-768_A-12/1, last accessed on August 5, 2020