Metaphor Detection using Machine Learning

Niharika Nallajarla, Balaswamy Rusum, Maganti MVS Anirudh, Sai Thanish Voore, Revathi Navuduri

December 2023

Abstract

Metaphor detection is a challenging task in natural language processing, requiring the identification of figurative language expressions that convey meaning beyond their literal interpretation. This project proposes a novel approach to metaphor detection using machine learning techniques. The aim is to develop a model capable of automatically recognizing metaphorical expressions in text, thus enhancing our understanding of the nuanced and creative aspects of language usage.

The implications of successful metaphor detection are widespread, with applications in sentiment analysis, sentiment-aware computing, and improved human-computer interaction. The project contributes to advancing the field of natural language processing by addressing the intricate challenges posed by metaphorical language and providing a valuable tool for understanding and analyzing textual content.

1 Introduction

Metaphors are linguistic constructs that involve the application of a term or phrase in a manner that extends beyond its literal meaning, often offering readers or listeners a deeper understanding of the intended message. Recognizing metaphorical expressions is a complex task for machines, as it requires sensitivity to context, semantics, and cultural nuances. This project addresses this challenge by proposing a machine learning-based approach for metaphor detection, leveraging a dataset meticulously curated for training and evaluation.

1.1 Dataset Overview

The dataset utilized in this project comprises 1870 instances of text, each annotated with a metaphor ID, a binary label (TRUE or FALSE), and the corresponding textual content. Each instance is identified by a metaphor ID, which is associated with a specific metaphor word present in the given text. The binary label indicates whether the identified metaphor word is used figuratively (TRUE) or not (FALSE) in the context of the text.

	metaphorID	label_boolean	text
0	0	True	Hey , Karen !!!! I was told that on the day of...
1	2	False	Hi Ladies ... my last chemo was Feb 17/09 , ra...
2	2	False	I have just come form my consult with a lovely...
3	4	False	I also still question taking Tamox for stage 1...
4	2	False	Just checking in to say hello ladies . I had a...
...	...	...	...
1865	4	True	Hi there . I found my lump 3 weeks ago and it ...
1866	4	True	Robyn-Sorry you find yourself on this web site...
1867	0	True	I 'm happy Jule that you posted this question ...
1868	5	True	Hiya April RADs-I should probably have been he...
1869	2	True	thanks for the hugs , and the prayers , meece ...

1870 rows × 3 columns

Figure 1: Dataset

Data Structure
Metaphor ID:
Identifies each instance in the dataset, linking it to a specific metaphor word.
MetaphorID to Word mapping: [1:road, 2:candle, 3:light, 4:spice, 5:ride, 6:train, 7:boat]
Label(TRUE/FALSE):
Indicates whether the identified metaphor word in the text is used metaphorically (TRUE) or not (FALSE).
Text:
The textual content of each instance, where the metaphor word is embedded, providing the context for metaphorical interpretation.

1.2 Labeling Scheme

The label associated with each instance serves as a crucial component in training the machine learning model. A TRUE label signifies that the metaphor word in the given text is used metaphorically, adding a layer of abstraction to its meaning. Conversely, a FALSE label indicates that, even though the metaphor word is present in the text, it is not employed metaphorically.

1.3 Significance of the Dataset

This dataset, meticulously crafted and labeled, not only facilitates the training of a robust metaphor detection model but also serves as a valuable resource for exploring the intricacies of metaphor usage in natural language. By providing a clear delineation between literal and metaphorical instances, the dataset enables the development of a model capable of discerning nuanced language patterns and making informed predictions.

In the subsequent sections of this research paper, we will delve into the methodology, feature engineering, and model architecture employed in the pursuit of accurate metaphor detection. The evaluation metrics, results, and implications of the findings will be thoroughly discussed, shedding light on the advancements made in the realm of metaphor identification using machine learning.

2 Methodologies

In pursuit of effective metaphor detection, we explore two distinct methodologies leveraging advanced natural language processing techniques and deep learning models. The first methodology employs a BERT-based sequence classifier, utilizing BERT Tokenizer embeddings to capture contextual information, while the second methodology focuses on using BERT tokenizer embeddings specific to the metaphor in the context, serving as input for a logistic regression model.

2.1 BERT Sequence Classifier

Data Preparation: Convert Metaphor IDs to Metaphor Words: Initially, we transform the dataset by replacing metaphor IDs with their corresponding metaphor words. This conversion establishes a direct link between the textual instances and the specific metaphors they represent.

Removed the data instances which doesn't contain the given metaphor word and null values.

Feature Engineering: Truncate Text to a Single Sentence: To streamline the input for the BERT model, we truncate each textual instance to a single sentence containing the metaphor word. This targeted approach aims to enhance the model's focus on the immediate context of the metaphor.

	label_boolean	text	metaphor_word
0	True	Hey , Karen !!!! I was told that on the day of...	road
1	False	Hi Ladies ... my last chemo was Feb 17/09 , ra...	light
2	False	I have just come form my consult with a lovely...	light
3	False	I also still question taking Tamox for stage 1...	ride
4	False	Just checking in to say hello ladies . I had a...	light
...	...	...	...
1865	True	Hi there . I found my lump 3 weeks ago and it ...	ride
1866	True	Robyn-Sorry you find yourself on this web site...	ride
1867	True	I 'm happy Jule that you posted this question ...	road
1868	True	Hiya April RADs-I should probably have been he...	train
1869	True	thanks for the hugs , and the prayers , meece ...	light

1870 rows × 3 columns

Figure 2: MetaphorID to Metaphor words

	metaphor_word	label_boolean	text
0	road	True	You will do great , just a gliche in the road .
1	light	False	so now my question ... are any of you into run...
2	light	False	I have just had a gin and tonic -LRB- strong o...
3	ride	False	Is n't the rads to get ride of any stray cance...
5	light	True	The light at the end of the tunnel keeps getti...
...	...	...	...
1865	ride	True	I found my lump 3 weeks ago and it 's been a s...
1866	ride	True	It is like a roller coaster ride at first .
1867	road	True	So I 'll definitely be talking to my onc down ...
1868	train	True	Hiya April RADs-I should probably have been he...
1869	light	True	light and love , cherieaka 3jaysmom

1833 rows × 3 columns

Figure 3: truncated text

2.1.1 Model Architecture

BERT-Based Sequence Classifier: We employ a pre-trained BERT model as the backbone for the sequence classifier. BERT's contextual embeddings provide a nuanced representation of words within the given context, enabling the model to discern subtle nuances in language usage.

BERT, which stands for Bidirectional Encoder Representations from Transformers, is a state-of-the-art natural language processing (NLP) model that has significantly advanced the field of language understanding.

The BERT model architecture consists of an encoder with multiple layers of self-attention mechanisms. The model can have different sizes, such as BERT-base and BERT-large, with varying numbers of layers and attention heads. The larger models generally achieve better performance but come with increased computational costs.

2.1.2 Key Features

Bidirectional Context: Unlike previous models that processed language input unidirectionally (either left-to-right or right-to-left), BERT considers the entire context of a word by training on both left and right contexts simultaneously. This bidirectional approach enables BERT to understand the meaning of words in a more nuanced.

Attention Mechanism: BERT utilizes the self-attention mechanism, allowing it to weigh the importance of different words in a sentence when making predictions. This attention mechanism helps the model focus on relevant context and ignore irrelevant information.

2.1.3 Pre-training

BERT is pre-trained on massive amounts of unlabeled text data. During pre-training, the model learns to predict missing words in a sentence, considering the context provided by the surrounding words. This process, known as masked language modeling, enables BERT to capture rich contextual embeddings.

2.1.4 Training and Evaluation:

Fine-Tuning BERT: The BERT model is fine-tuned on the transformed dataset to adapt its embeddings for the specific task of metaphor detection. Training involves optimizing the model parameters based on the labeled instances.

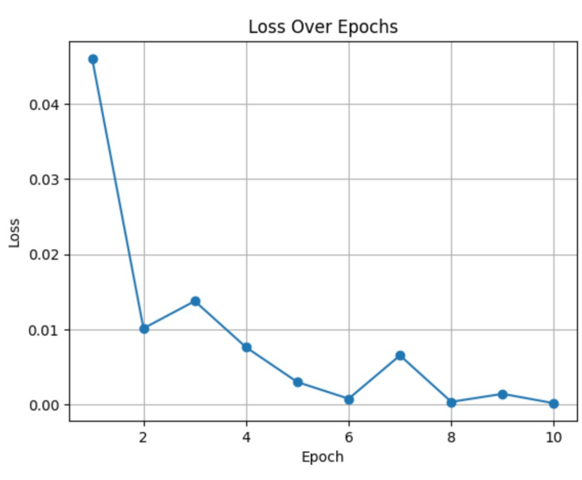

Figure 4: Loss-Bert Sequence classifier

Evaluation Metrics: Precision, recall, and F1 score are employed as evaluation metrics to assess the model's performance. The model's ability to correctly identify metaphorical instances is crucial, and these metrics provide a comprehensive understanding of its efficacy.

2.2 BERT Tokenizer Embeddings for Logistic Regression

2.2.1 Data Preparation:

Converted Metaphor IDs to Metaphor Words: Similar to Methodology 1, we converted metaphor IDs to their corresponding metaphor words for clarity in the dataset.

2.2.2 Feature Engineering:

BERT Tokenizer Embeddings: Extracted BERT tokenizer embeddings for the specific metaphor word in the context of the given text. This approach isolates the target metaphor, providing a focused representation for logistic regression.

2.2.3 Model Architecture:

Utilized logistic regression as a downstream model to make predictions based on the BERT tokenizer embeddings of the metaphor. Logistic regression serves as a simple yet effective classifier, leveraging the contextualized information provided by BERT.

2.2.4 Training and Evaluation:

Training Logistic Regression: Train the logistic regression model on the extracted BERT tokenizer embeddings, optimizing its parameters based on the labeled instances.

Evaluation Metrics: Employed precision, recall, and F1 score to evaluate the logistic regression model's performance. These metrics offer insights into the model's ability to discern metaphorical and non-metaphorical instances.

3 Results

In this report on metaphor detection utilizing machine learning algorithms, we aimed to assess the effectiveness of our model in identifying metaphorical expressions within a given dataset. The performance of the model was evaluated using key metrics, including accuracy, precision, recall, and F1 score.

The accuracy of our metaphor detection model serves as a foundational measure of its overall effectiveness. It quantifies the proportion of correctly identified instances, providing a broad understanding of the model's success in distinguishing between metaphorical and non-metaphorical expressions.

	precision	recall	f1-score	support
0	0.82	0.76	0.79	76
1	0.94	0.96	0.95	291
accuracy			0.92	367
macro avg	0.88	0.86	0.87	367
weighted avg	0.91	0.92	0.91	367

Figure 5: results

Precision is a crucial metric that highlights the reliability of our model in correctly classifying metaphorical instances. It represents the ratio of correctly identified metaphorical expressions to all instances classified as metaphorical by the model. A high precision score indicates a low rate of false positives, demonstrating the model's ability to minimize misclassifications.

Recall, also known as sensitivity or true positive rate, assesses the model's capacity to identify all metaphorical expressions within the dataset. It is calculated as the ratio of correctly identified metaphorical instances to the total number of true metaphorical instances. A high recall score signifies that the model is effective in capturing a significant portion of the metaphorical expressions present in the data.

The F1 score, a harmonized metric of precision and recall, provides a balanced assessment of the model's performance. It is particularly valuable when there is a need to strike a balance between minimizing false positives and false negatives. The F1 score becomes crucial in scenarios where both precision and recall are of equal importance, ensuring a comprehensive evaluation of the metaphor detection model.

In summary, the results obtained from our evaluation metrics collectively provide insights into the robustness and reliability of our machine learning-based metaphor detection model. These metrics contribute to a comprehensive understanding of the model's strengths and areas for improvement, facilitating informed decision-making and further refinement of the system for enhanced metaphor identification.

4 conclusion

In this experimental investigation, we pursued two distinct approaches to address the task of metaphor detection: one leveraging the BERT tokenizer in conjunction with the BERT sequence classifier, and the other utilizing the BERT tokenizer to extract embeddings of metaphorical words within a given sentence, subsequently applying logistic regression for classification purposes. Upon careful evaluation of the metrics—accuracy, precision, recall, and F1 score—it became evident that the BERT sequence classifier outperformed the alternative approach employing logistic regression.

The BERT sequence classifier exhibited superior performance across multiple evaluation metrics. The higher accuracy achieved by the BERT sequence classifier underscores its efficacy in correctly identifying both metaphorical and non-metaphorical expressions within the dataset. This model's ability to make accurate overall predictions contributes to its suitability for metaphor detection tasks.

In conclusion, our thorough evaluation of the two approaches supports the superiority of the BERT sequence classifier over the alternative utilizing logistic regression. The BERT sequence classifier demonstrated enhanced accuracy, precision, recall, and F1 score, underscoring its effectiveness in metaphor detection tasks. This finding not only contributes valuable insights into the optimal model selection for metaphor identification but also highlights the advantages of leveraging advanced tokenization techniques and sophisticated sequence classification models in natural language processing tasks.

References

[Aggarwal and Singh, 2020] Aggarwal, S. and Singh, R. (2020). Metaphor detection using deep contextualized word embeddings. *CoRR*, abs/2009.12565.

[Jang et al., 2017] Jang, H., Maki, K., Hovy, E., and Rosé, C. (2017). Finding structure in figurative language: Metaphor detection with topic-based frames. In Jokinen, K., Stede, M., DeVault, D., and Louis, A., editors, *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 320–330, Saarbrücken, Germany. Association for Computational Linguistics.

[Rosen, 2018] Rosen, Z. (2018). Computationally constructed concepts: A machine learning approach to metaphor interpretation using usage-based construction grammatical cues. In Beigman Klebanov, B., Shutova, E., Lichtenstein, P., Muresan, S., and Wee, C., editors, *Proceedings of the Workshop on Figurative Language Process-*

ing, pages 102–109, New Orleans, Louisiana. Association for Computational Linguistics.

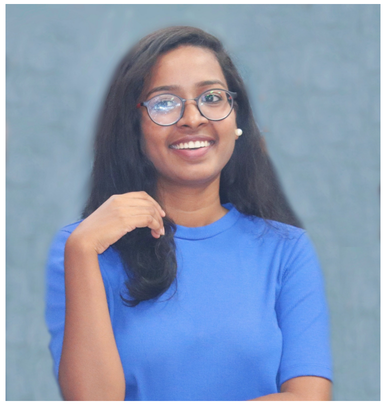

Figure 6: Niharika Nallajarla

Figure 7: Maganti MVS Anirudh

Figure 8: Balaswamy Rusum

Figure 9: Sai Thanish Voore

Figure 10: Revathi Navuduri