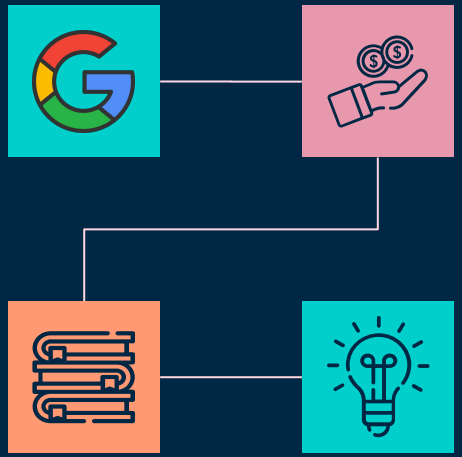
PARAPHRASE RECOGNITION (PR)

Authors:
Caviglia Matteo
Formenti Sofia Gaia

PAWS DATASET DESCRIPTION

Developed by Google Research



Specific Algorithm and Human Evaluation

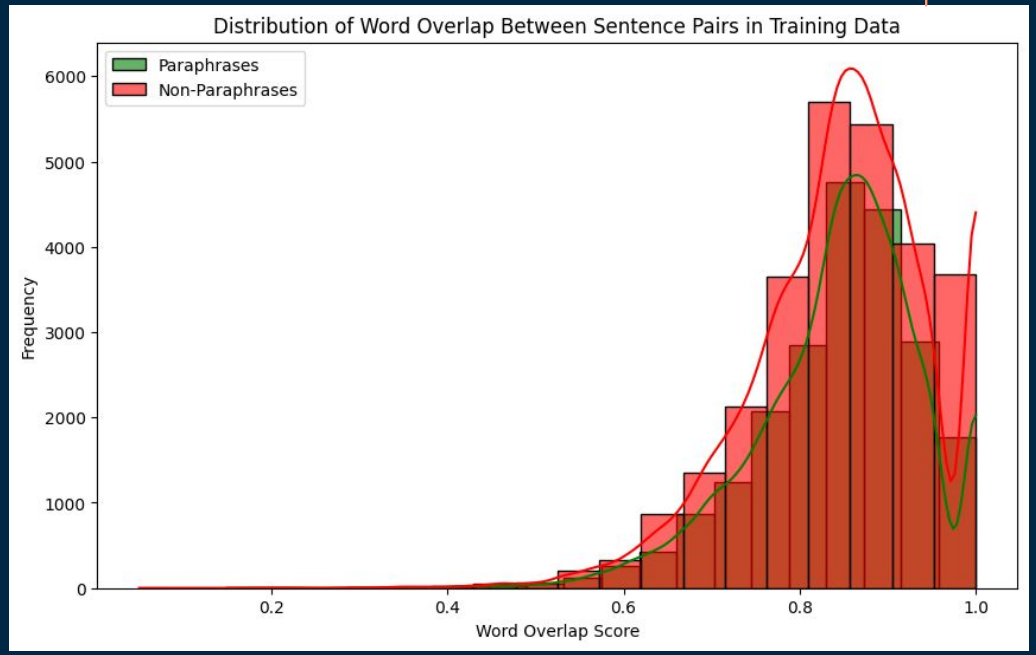
Built upon WikiPedia and QQP.

Designed to evaluate deeper language understanding

PAWS DATASET DESCRIPTION

	Sentence 1	Sentence 2	Label
EX 1	It is present in Asia, in Japan	In Asia, it is present in Japan	Paraphrase (1)
EX 2	... To Scotland from England	... To England from Scotland	Not a paraphrase (0)

PAWS DATASET DESCRIPTION



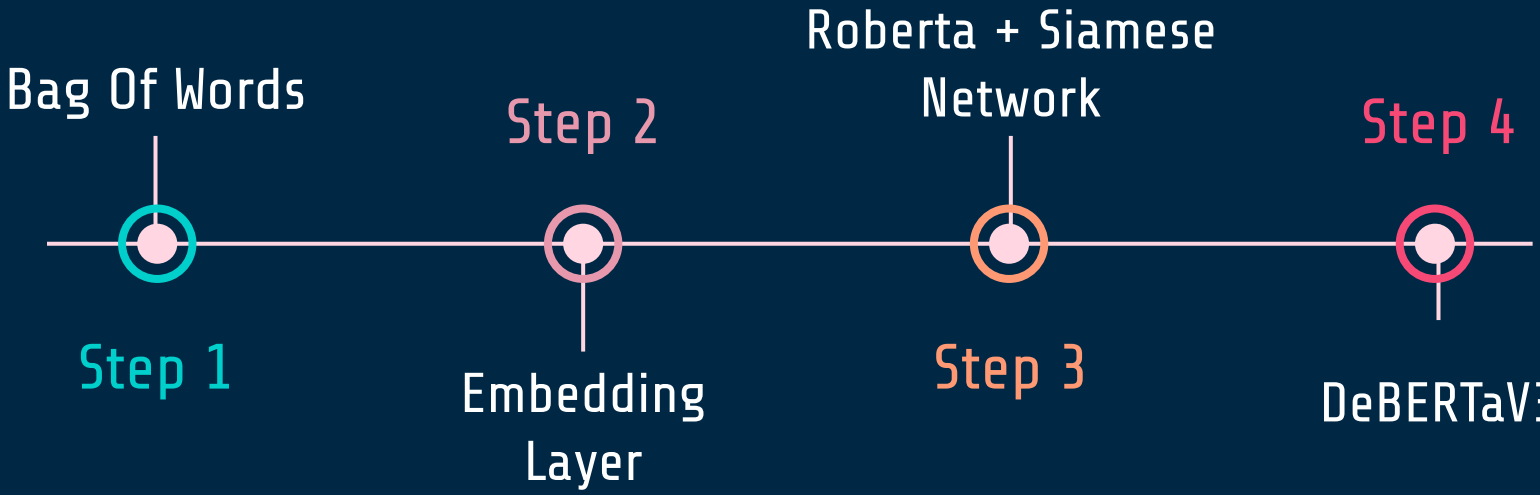
ISSUES

SIMILAR STRUCTURE

HIGH WORD OVERLAP

IMPORTANCE OF SEQUENTIALITY

USED MODELS


Bag Of Words
Step 1
Step 2
Embedding Layer
Roberta + Siamese Network
Step 3
Step 4
DeBERTaV3

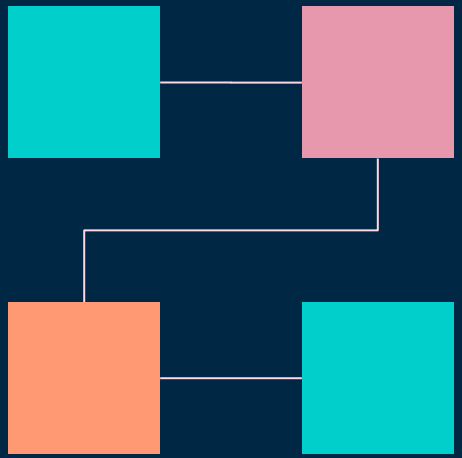
1. BAG OF WORDS

COMBINATION

The two sentenced are combined in a single list

LOGISTIC REGRESSION

Train classifier



COUNTVECTORIZER

Remove stopwords and lowercase

EVALUATE

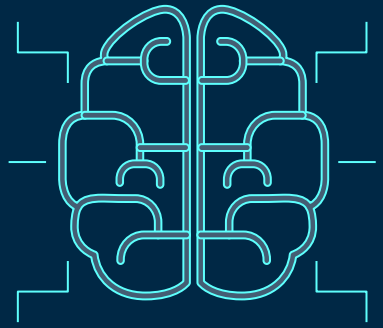
Output accuracy score

1. BAG OF WORDS

As expected, this procedure did not provide great results, due to its inability to properly capture semantic relationships between words, especially because many words overlap in our dataset.

2. EMBEDDING LAYER

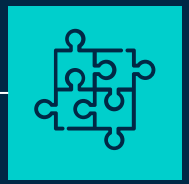
Our second chosen method relies on the Embedding Layer provided by the TensorFlow library.
It exploits the concept of learning through training.
Basically the embeddings are learned as the Neural Network itself learns.



2. EMBEDDING LAYER

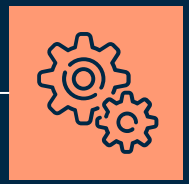
The designed model had to fit the task properly, so we created a double input model that separately processes the two sentences and makes assumptions based on that.

3. ROBERTA + SIAMESE ENCODER



01

SPECIFIC PREPROCESS



02

ROBERTA ENCODER

03

COSINE SIMILARITIES

3. ROBERTA + SIAMESE ENCODER

Training the encoder

In this model we are actually only training the encoder to better represent data. We can see the results of that when calculating the cosine similarities for words before and after the training process.

Adaptation

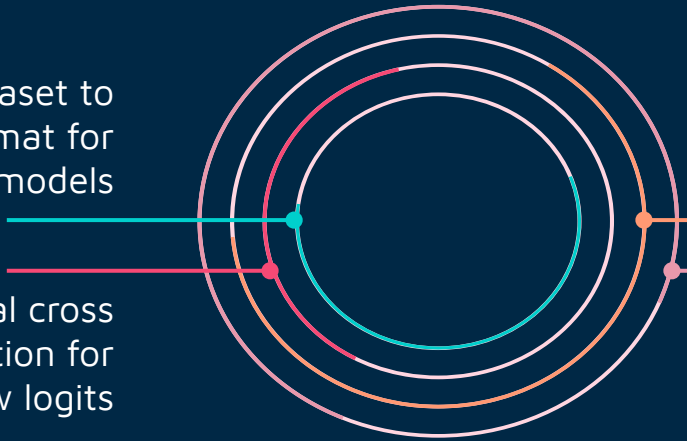
In order to use this model correctly we should have data that has float values ranging from -1 to 1, to best fit the cosine similarity range. In our case we "adapted" this approach to our problem.

0.708

Test Accuracy

4. DEBERTAV3

Decoding-enhanced BERT with disentangled attention



Conform dataset to suitable format for BERT-based models

Sparse categorical cross entropy loss function for handling raw logits

Adam optimizer

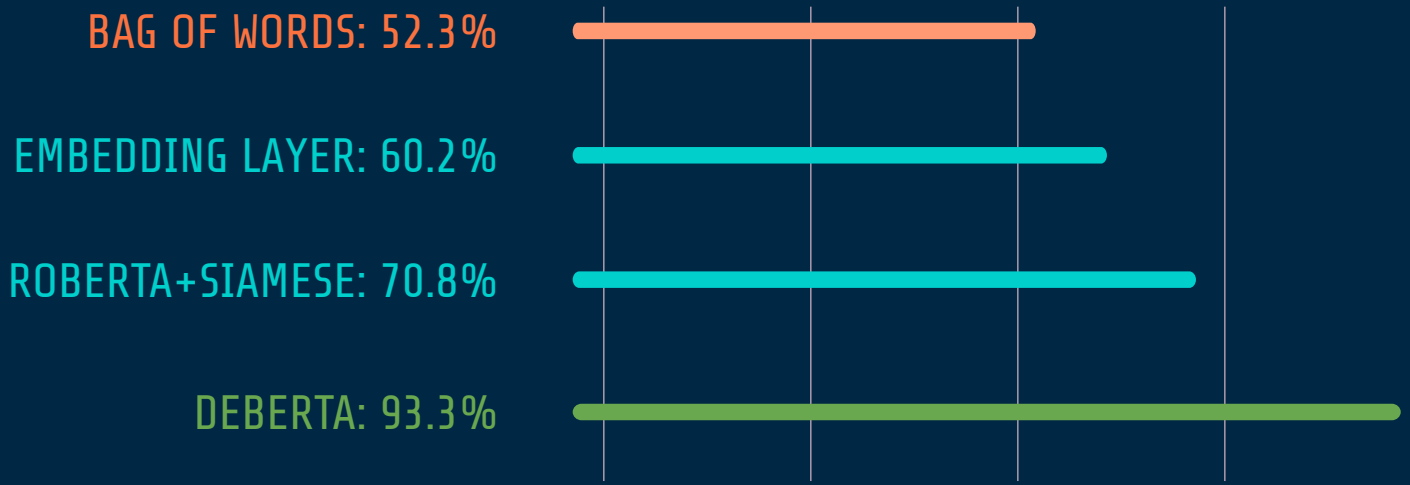
Model's backbone is set to be trainable, allowing for fine-tuning of the pre-trained DeBERTaV3

4. DEBERTAV3

The last model significantly outperformed the other models. Its enhanced attention mechanisms and the fine-tuning on the PAWS dataset allowed it to capture the details necessary for performing our task properly.

CONCLUSIONS

As we could have predicted from theory, we had better performances using Transformer-like models rather than standard embedding methods



THANKS

Do you have any questions?

Caviglia Matteo: 513026
Formenti Sofia Gaia: 514495


slidesgo