BEYOND OUR CONTRADICTIONS

Feature augmentation in large language models for textual entailment detection.

Jérémie Dentan – Ecole Polytechnique, MVA (jeremie.dentan@live.com)
Alicia Rakotonirainy – CentraleSupélec (alicia.rakotonirainy@gmail.com)

TEXTUAL ENTAILMENT

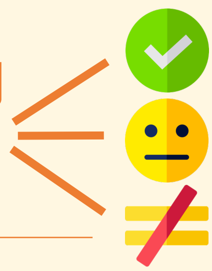
A huge part of the **meaning** of a text **is not expressed explicitly**, but added to the text by readers through **inference**.

Given 2 excerpts of texts, can the meaning of one be inferred from the other?

In Moria you will find the remains of a huge Roman aqueduct. → ? → There is something special to see in the village of Moria.

LARGE LANGUAGE MODELS AND FEATURE AUGMENTATION

Large Language Models : Large data scrapped on the internet + Several pre-trained **attention layers** + hundreds of millions of parameters → Sentence 1 + Sentence 2



Attention layers :

$$x_{n+1} = \text{softmax}\left(\frac{Q^T K}{\sqrt{d}}\right) \cdot V$$

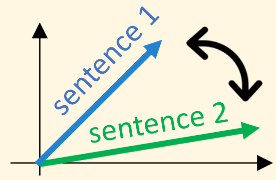
$$Q = \sigma(W^Q \cdot x_n)$$

$$K = \sigma(W^K \cdot x_n)$$

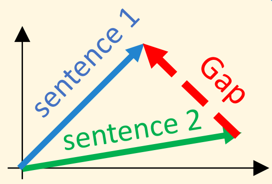
$$V = \sigma(W^V \cdot x_n)$$

Feature augmentation :

cosine similarity

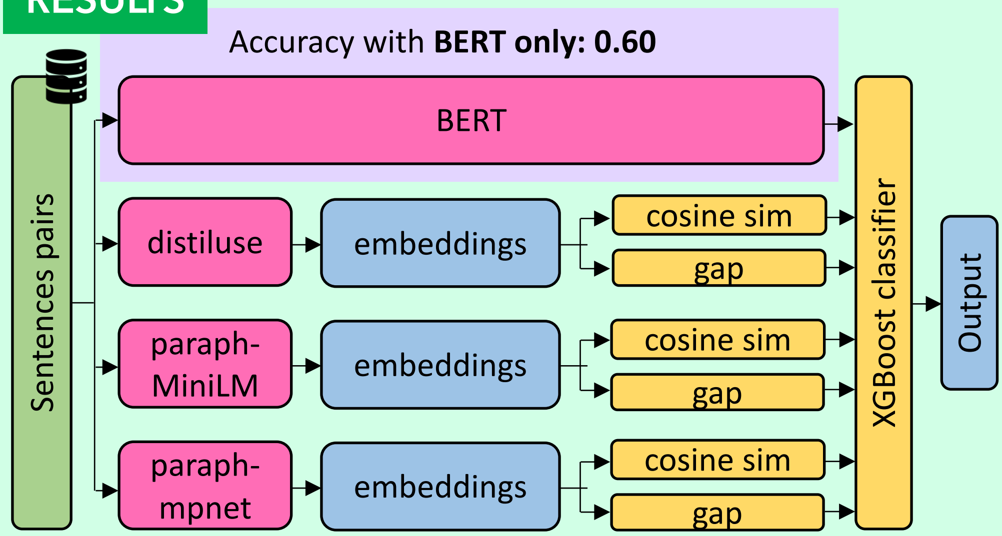


gap



Cosine similarity between :
- The gap
- The **mean gap of each class**

RESULTS



BERT training time: 90min
New features computation time + XGBoost: 2.3min
Whole model accuracy : 0.63 ⇛ **+3% accuracy!**

DISCUSSION

The classification accuracy remains surprisingly low, even after using different embedding models. However, we have achieved to show that adding our new features significantly increases performance, with negligible additional computation time.