Team Performance Pressure @ SemEval Task 8

Evaluation of Multilingual News Article Similarity leveraging Pre-trained Models

Adarsh Kumar, Animesh Jain,
Abhinav Bohra, Ishan Sharma, Tanuj Saraf

IIT Kharagpur

Introduction and Task Overview

Dataset Distribution

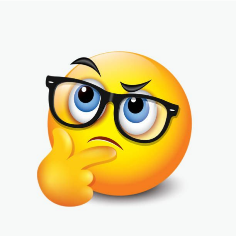
Language	Number of Pairs
En - En	1800
De - De	857
De - En	577
Es - Es	570
Tr - Tr	465
Pl - Pl	349
Ar - Ar	274
Fr - Fr	72

Article 1

Article 2

How similar are the articles?

Score [1-4] (most to least)



Structure of the Task

2

Dataset

Language Pair	Train Set	Val Set	Test Set
Monolingual	3603	784	3462
Cross-lingual	515	62	1440
Unseen pairs	NA	346	3131

Counts of different types of language pairs

Baseline

Baseline approach for the Given Dataset

Baseline Approach

- Finetune Multilingual Bert Model followed by a regressor layer, using MSE Loss
- Contextual Embeddings from [CLS] token output for each sentence in pair concatenated and fed into Regressor
- Regressor Layer: (Dropout + Linear Layer)

Methodology

A discussion on Our Final Approach

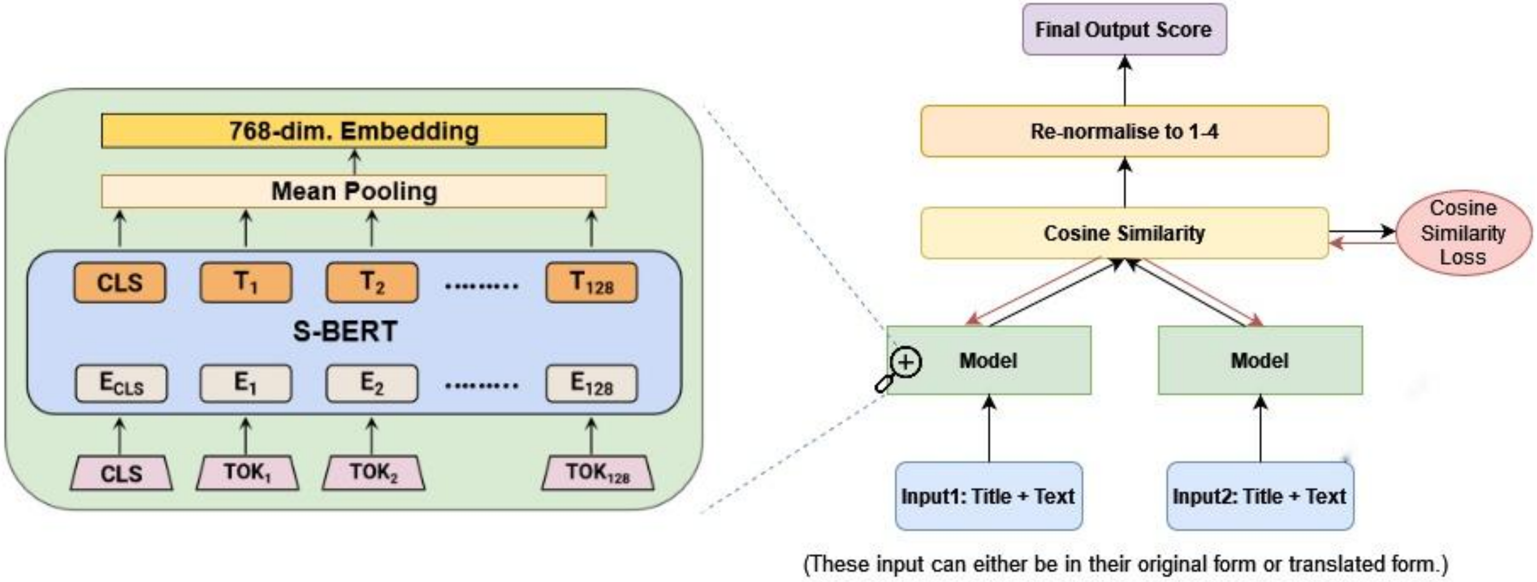
Our Approach

- Sentence Transformer Based Approach (Bert + Mean Pooling Layer)
- Cosine Similarity (Encode_Article+Title_1, Encode_Article+Title_2)
- Renormalisation of Cosine Similarity Score to [4-1], least to most
- Finetuning Encodings with Cosine Similarity Loss

$$MSE(y_{true_n m}, y_{pred}) = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^{n} (y_{true_n mi} - y_{pred_i})^2}$$

where $y_{true_n m}$ are the normalised gold-standard and y_{pred} being the predicted scores for similarity.

- Model Used for Final results: *paraphrase-multilingual-mpnet-base-v2*



System Architecture for our Final Approach

Results

Performance of Regression Model Experiments

Regression			
Approach Category	**Approach**	**Model**	**Pearson Correlation**
Baseline	–	mBERT	0.3089
		XLM-R	0.2852
Baseline	+ Validation	mBERT	0.3164
		XLM-R	0.3246
Baseline + Validation	+ Translation	mBERT	0.3371
		XLM-R	0.3045
Baseline + Validation	+ Title	mBERT	0.3187
		XLM-R	0.3345
Baseline + Losses + Validation	+ CosineEmbeddingLoss	mBERT	0.3346
		XLM-R	0.3336
Baseline + Losses + Validation	+ Cosine Similarity + MSE Loss	mBERT	**0.3621**
		XLM-R	**0.3454**
Sentence Transformer	+ Translation + Title + MSE Loss	S-BERT	0.2864

Table 2: Results with Regressor layer (Dropout layer followed by a feed forward layer) for final prediction as described in Section 4.3. Note that here we feeding the concatenated contexualized embeddings from Model to the regressor layer

Performance of Similarity Based Models (Ablation)

Similarity		
Approach	**Model**	**Pearson Correlation**
ST Multilingual CS	S-BERT	0.5866
ST Translated CS	S-BERT	0.6058
ST Multilingual + "Title" CS	S-BERT	0.6778
ST Translated + "Title" CS	S-BERT	0.7086
**ST Multilingual + "Title" + CosSimLoss Finetuned CS	S-BERT	**0.7518**
**ST Translated + "Title" + CosSimLoss Finetuned CS	S-BERT	0.7492
ST(mBERT) + "Title" + Translation + CosSimLoss Finetuned CS	mBERT	0.6874
ST(XLM) +"Title" + Translation + CosSimLoss Finetuned CS	XLM-R	0.6357

Table 3: Results of the Consine Similarity based approach, where final score was evaluated using cosine similarity between embeddings and then scaling to 1 to 4. ** represents the models we used for our final submission on codalab as described in Section 3. Table also represent the ablation study for various elements included in our final approach along with our final approach experimented on mBERT in the ST framework. Note: CS: Cosine Similarity

#	SCORE	FILENAME	SUBMISSION DATE	STATUS	✔	
1	0.7205333427	Submission.zip	04/02/2022 07:52:23	Finished	✔	+
2	---	Submission_translated.zip	04/02/2022 08:02:42	Submitted		+
3	0.7172510638	Submission_translated.zip	04/02/2022 08:05:09	Finished		+

17	*Anonymous*	3	01/30/22	Andi	0.726 (17)	View
18	*Anonymous*	4	01/27/22	BUT	0.726 (18)	View
19	*Anonymous*	3	04/02/22		0.721 (19)	View
20	*Anonymous*	3	02/01/22	TCU	0.715 (20)	View

CODALAB Submission Results: 19th Rank on Leader Board

Result Analysis

Error Analysis

- Poor Regression Scores
  - **Possible Explanation:** Limited Data to train Regressor Layer alongside the
  - Similar results seen with changing regressor layer to ML based layers like SVR, RFR etc.

- Not much improvement with Multitasking based approach
  - **Possible Explanation**: Not much extra knowledge obtained with shared Parameter training, also less experimented upon task

Pearson Score Distribution

Language Pair	Pearson Correlation
Monolingual	0.7386
Cross-lingual	0.7739
Unseen pairs	0.7459

Path Not Taken

Path not taken

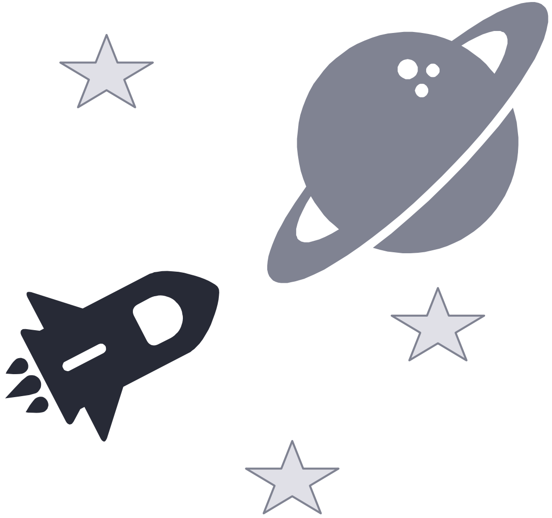
- **Dropped Approaches:**
    - Multi Task based approach
    - Summarisation
    - Concatenation with non finetuned model embeddings
    - Didn’t work out as expected

Conclusion

Conclusion

- We successfully leveraged knowledge of Sentence based transformers for Similarity prediction
- Addition of Cosine Similarity Loss and Metadata “Title” improved performance
- Similarity based approach outperformed Regressor based
- Stand 19th on Codalab leaderboard for SemEval Task 8
- Experiments also include tasks that didn’t work out positively in our case

What Next??



- Adversarial Finetuning
- Data Augmentation
- Hierarchical Modelling with LSTMs
- Extensive Experiments on Ensembling

Note: There approaches are the ones we were not able to try out because of time constraint

Thanks!

It's been a great learning experience!

Any questions?

Feel free to ask....