ZhichunRoad@ SemEval-2022-Task2

Adversarial Training and Contrastive Learning for Multiword Representations

High Lights

Subtask A=Text Classification
Our approach:
EMA + Adversarial Training

Subtask B=Sentence Embedding
Our approach:
Sentence-Bert + Cross-Attention + Contrastive Learning

Rank 12th In the SubA, Zero-Shot
Rank 4th In the SubA, One-Shot
Rank 4th In the SubB, Pretrain
Rank 4th In the SubB, Finetune

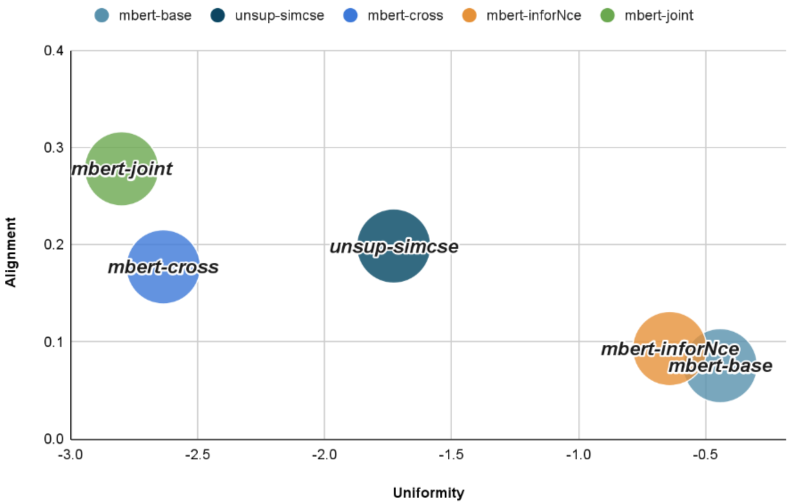
Paper:
https://github.com/cuixuage/SemEval2022-Task2/blob/main/semeval_2022_task2.pdf

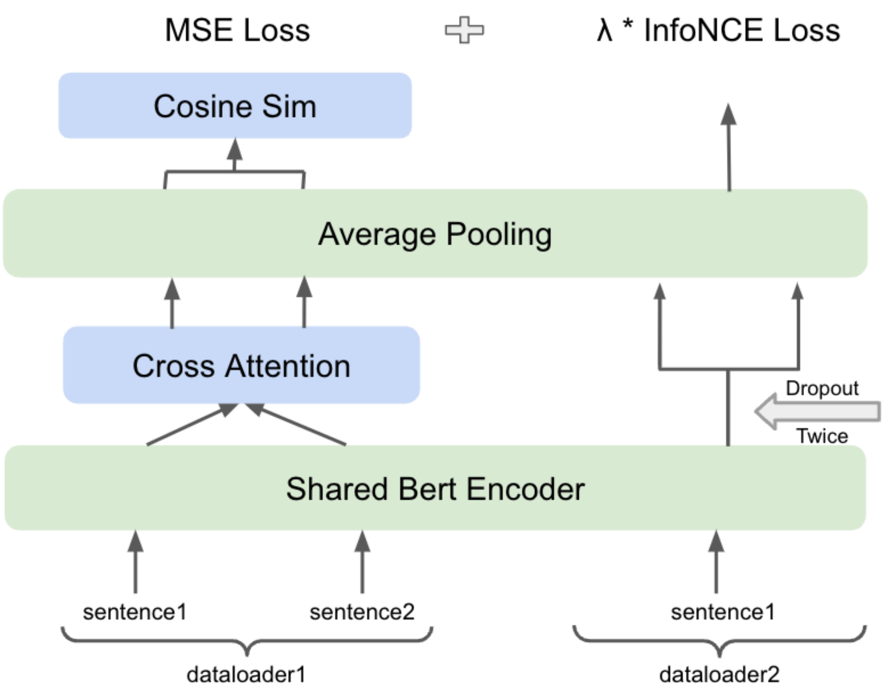
Code:
https://github.com/cuixuage/SemEval2022-Task2

In the SubTaskA, we use InfoXLM as text encoder and exponential moving average (EMA) method and the adversarial attack strategy.

In the SubTaskB, we add an cross-attention module, contrastive objective and employ a momentum contrast.

Additionally, we use the alignment and uniformity properties to measure the quality of sentence embeddings.





During training, each data point is trained to find out its counterpart among $(N-1)$ from in-batch negative samples and the queue of data samples. The samples in the queue are progressively replaced (He et al., 2020).

$$\ell_i = -\log \frac{e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_i^+)/\tau}}{\sum_{j=1}^{N} e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_j^+)/\tau} + \sum_{q=1}^{Q} e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_q^+)/\tau}} \tag{4}$$

The h_* is the sentence representation, where h_i and h_i^+ are semantically related. The h_q^+ is denotes a sentence embedding in the momentum-updated queue. And the Q is the size of the queue, $sim(h1, h2)$ is the cosine similarity scores of sentence representations, τ is a temperature hyperparameter. In the end, we average the all N Li losses to calculate the contrastive loss $\mathcal{C}_{\mathrm{con}}$.