

<https://arxiv.org/abs/1901.07291>

O XLM é um modelo de tradução da biblioteca Transformers. Ele tem várias funcionalidades:

- Language model pretraining:
 - Causal Language Model (CLM)
 - Masked Language Model (MLM)
 - Translation Language Model (TLM)
- GLUE fine-tuning
- XNLI fine-tuning
- Supervised / Unsupervised MT training

Esse modelo é a implementação do modelo Cross-lingual Language Model Pre Training [1], formulado por Guillaume Lample, Alexis Conneau. As principais vantagens desse modelo são:

- Modelo de tradução com melhor performance do que os modelos supervisionados e não-supervisionados anteriores.
- Não necessidade de supervisionamento usando um Cross-Lingual Language Model;
- Possibilidade de aperfeiçoamento do modelo por um treino supervisionado utilizando dados paralelos;
- Gerar um modelo de linguagem low-resource;
- Gerar embeddings não supervisionados de palavras cross-lingual.

O modelo XLM não partiu do zero, ele foi construído baseado em vários estudos já realizados, sendo os mais influentes o Improving Language Understanding by Generative Pre-Training[2], para o trabalho com o Transformers; SentEval: An Evaluation Toolkit for Universal Sentence Representations[3], para um treinamento não supervisionado; e Googles multilingual neural machine translation system: Enabling zero-shot translation[4].

Os pré treinamentos considerados para o Language Model Pretraining do XLM são o Casual Language Modeling (CLM), o Masked Language Modeling (MLM) e o MLM combinado com Translation Language Modeling (TLM). Todos os pré treinamentos foram feitos com um vocabulário gerado pela técnica de Byte Pair Encoding.

O CLM tem como propósito descobrir uma palavra dada n palavras anteriores na sentença. O CLM em questão foi feito com uma Recurrent Neural Network do tipo Long Short-Term Memory (LSTM). Pela biblioteca Transformers, é possível passar o last hidden state de um layer para o próximo layer, gerando assim o Backpropagation necessário para a LSTM.

O MLM tem como propósito descobrir palavras mascaradas em uma dada sentença. A figura 1 demonstra como esse modelo funciona.

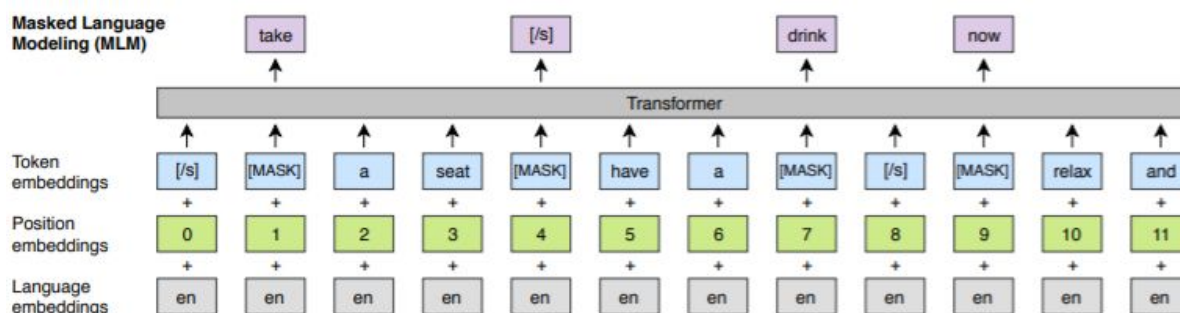


Figura 1 - Masked Language Modeling

O TLM é uma extensão do MLM, porém, para dados paralelos bilíngues. Em vez da entrada ser uma sentença em uma língua, a entrada é uma concatenação de sentenças iguais, porém em línguas diferentes.

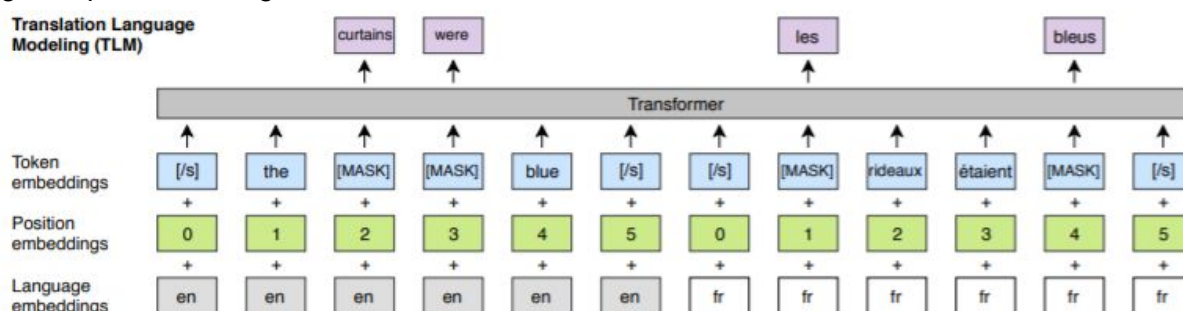


Figura 2 - Translation Language Modeling

Após várias experimentações com toda essa base, os resultados foram comparados com o modelo Bert para tradução[5], e os cross-sentence encoders XNLI[6] e Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual[7]. Os testes de acurácia do XLM utilizando o MLM junto com o TLM apresentaram-se consistentemente superiores em todas as traduções realizadas.

Como conclusão, o XLM pode ser considerado uma evolução positiva na questão de tradução na área de PLN, pois conseguiu utilizar-se de técnicas e modelos já existentes e aperfeiçoá-los. Além disso, outra vantagem é a possibilidade de ser ainda mais aperfeiçoado, por ser um modelo com código aberto.

Referências

- [1] <https://arxiv.org/pdf/1901.07291>
- [2] https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf
- [3] <https://arxiv.org/abs/1803.05449>
- [4] <https://arxiv.org/abs/1611.04558>
- [5] <https://arxiv.org/abs/1810.04805>
- [6] <https://arxiv.org/abs/1809.05053>

[7] <https://arxiv.org/abs/1812.10464>

Bibliografia:

<https://arxiv.org/abs/1901.07291>

<https://github.com/facebookresearch/XLM>