Finetuning HuggingFace Models

Marta Mariz up201907020
Miguel Freitas up201906159
Afonso Monteiro up201907284

ROADMAP

DOMAIN ADAPTATION
PRE TRAINED MODELS CHOSEN
RESULTS
RESULTS IN COMPARISON TO PREVIOUS WORK

DOMAIN ADAPTATION

FRENCH SUBSET OF MULTIILINGUAL AMAZON REVIEWS

To adapt the domain we trained the pre trained models with a dataset of similar domain to ours.

- The dataset is only in french
- It is composed of 210k rows

Use perplexity to evaluate evolution of model understanding. We saw significant decreases by epoch which indicates the model is learning the patterns of the domain.

CAMEMBERT

LANGUAGE

One of the main reasons for our choice is because it is fully trained in French

SIZE

It is a model with 335 million parameters

TRAINED ON

Based on the RoBERTa architecture pretrained on the French subcorpus of the newly available multilingual corpus OSCAR.

GPT-FR

LANGUAGE

One of the main reasons for our choice is because it is fully trained in French

TRAINED ON LARGE HETEROGENOUS CORPUS

We used the smaller version for efficiency purpose but it is still composed of 124 million parameters

RESULTS

CAMEMBERT

WITHOUT DOMAIN ADAPTATION

Epoch	Loss	Acuracy	Precision
1	0.350	0.620	0.817
2	0.283	0.701	0.881
3	0.268	0.718	0.891

WITH DOMAIN ADAPTATION

Epoch	Loss	Acuracy	Precision
1	0.344	0.652	0.846
2	0.291	0.725	0.852
3	0.257	0.742	0.869

RESULTS

GPT-FR

WITHOUT DOMAIN ADAPTATION

Epoch	Loss	Acuracy	Precision
1	0.336	0.575	0.768
2	0.295	0.613	0.802
3	0.277	0.620	0.834

WITH DOMAIN ADAPTATION

Epoch	Loss	Acuracy	Precision
1	0.331	0.601	0.794
2	0.283	0.636	0.824
3	0.267	0.659	0.851

RESULTS

COMPARING WITH PREVIOUS WORK

Model	Acuracy	Precision	f1-score
Camambert with Domain Adaptation (3 epochs)	0.742	0.869	0.832
Word2Vec with Ridge Classifier	0.6111	0.775	0.736

RESULTS

COMPARING WITH PAPER

Model	Recall	Precision	f1-score
Camembert with Domain Adaptation (3 epochs)	0.803	0.869	0.832
Paper's CamemBERT	0.86	0.86	0.85

Conclusions

- Domain Adaptation proved to improve the results
- Pre trained models have great impact on the results, as expected

- With more time and less hardware contraints we could have:
  - Run the models for more epochs
  - Use all the available data for domain adaptation
  - Try out different HuggingFace pre-trained models