Sequence generation model for replicating phraseological patterns from an homegenous political group.
Machine Learning for Natural Language Processing 2022

Zakaria Bekkar
ENSAE Paris
zakaria.bekkar@ensae.fr

Walid Chrimni
ENSAE Paris
walid.chrimni@ensae.fr

1 Problem Framing

In this work[1], we attempted to create a sequence-generation model that aims to replicate phraseologic patterns from an homegenous political group. Our main motivation is to investigate the ability of our model to simulate thematic discourse and subtle languages nuances that could pass as genuine. Thereby, our project is structured around the following question : is it possible to create bots that could actually perform *astroturfing* and interfere in the political/social debate with relatively modest data and modelling techniques ?

2 Experiments Protocol

As of the text generation task and these constraints, the core idea is to train and benchmark low resources language models on a modest but targeted dataset.

2.1 Data

2.1.1 Collection

The data collection strategy consisted in scraping all the tweets from cherry-picked influencers within a political group that we deemed *homogoneous* and *polarized*, i.e similar and strong themes being discussed. The intuition that predicated this choice revolves around the idea that strong and caricatural stances will be easier to capture than more poised and subtle ones. We ended up selecting 21 influencers from Eric Zemmour's movement and scraping around 33k tweets.

2.1.2 Description

As expected, our dataset is very cohesive. The most frequent words in the tweets are *zemmour*, *france*, *macron* and call to action hashtags like *jevotezemmourle10avril*. (cf. fig1 in appendix). We see clearly a polarization as these supporters often try to portray themselves as opposing Macron's discourse and actions.

When we analyze word pair associations , we observe that the most common bi-grams (cf. fig2 in appendix) are : ('eric','zemmour'), ('pourles-francaisoublies,'jevotezemmour'). In addition to the thematic coherence (the presence of Marine Le Pen underlines that), we see campaigning efforts being made as hashtags associations are quite common. It is important to stress the fact that the occurences aforementioned do not have a systematic quality : overall the tweets seems to be more about ideas and opinions than merely about the political leader they follow or they oppose.

Finally, the distribution of tweet length in terms of words (cf. fig3 in appendix) points to the fact that brievity and punchlines are the *modus operandi* by which these partisans spread their message. The dataset exhibiting a 20 words per tweet average, we will use this number as the sequence token length for our generations.

2.2 Modelling

2.2.1 Approach

We articulated our modelling work in the following manner (see appendix for more information) :

1. Training a simple 3 layer *LSTM* for 10 epochs in order to form a baseline.

2. Performing a double step transfert learning on a GPT2-small pretrained model for 10 epochs.

For the second model, we leveraged an already fine-tuned gpt2-small (Radford et al., 2019) on a french wikipedia dataset found on Hugging Face (reference here). We performed a second step fine-tuning by training it for the text generation

[1] Gitub repository and Google colab notebook

task on our tweets dataset.

Regarding decoding methods, we used the default *greedy search* approach for the *LSTM* model. We did additionnal explorations on the optimal decoding strategy for the *GPT2* model and our use-case. By qualitavely evaluating generations, we chose a combination of *Top-K sampling*(Fan et al., 2018), *Top-p sampling* (Holtzman et al., 2019) and *similar n-gram penality* (Paulus et al., 2017). These were chosen to avoid the trap of missing high conditionnal probability words hidden behind low probability one, to enhance diversity by sampling and avoid loop repetitions.

2.2.2 Complexity

Time and Space Complexity Analysis		
Model	LSTM Baseline	GPT2-FR-POL-Tweets
Trainable parameters	14.137.821	124.439.808
Inference time for 100 tweets	6.2 secs	56.5 secs
Training time	13.4 min/epoch (Colab)	20 min/epoch (RTX3080)
Space in memory	53.9Mb	6.08Gb

3 Results

3.1 Qualitative metric

For the qualitative metric, we chose some context seeds (see appendix) and have outputted 10 sentences for each seed for each model. Then we rated each sentence on a scale of 1 to 10 in terms of realism and fitting to the targeted discourse. For the second model, we obtained a mean rating of 7,36, and for the baseline model we reach d 4,26.

In addition to this, we can gauge the quality of the prediction qualitatively just by looking at the sentences. For the GPT2-model, we see that most sentence are well built and correspond well to tweets pro-Zemmour people could write. Even though some sentences are not quite well finished, we understand well the meaning of the sentence and the syntax is rather good. The baseline model on the other hand struggles more. Its sentence generation are not well built and seem rather like a heap of key words. We find some key ideas in the sentences but the sentences have no meaning. Examples of generation of both model are available in the appendix.

3.2 Quantitative metric

As for the quantitative metric, we used BERTScore (Zhang et al., 2019) which is an automatic evaluation metric for text generation. It computes a similarity score for each token in the candidate sentence with each token in the reference sentence. In our use case, the reference sentence is either all the tweet corpus or the seed sentence. We choose BERTScore because it correlates better with human judgments and provides stronger model selection performance than existing metrics. We obtained the following score (the closer to 1 the better) :

- GPT-2 model : 0.73 when seeds are references and 0.58 when all the tweets are references
- baseline model : 0.68 when seeds are references and 0.61 when all the tweets are references

  Plot repartition of BertScore for each model are available in the appendix.

4 Discussion/Conclusion

We were able to generate an interesting but rather modest proportion of astonishingly accurate tweets in terms of the replication of the political discourse we targeted (see appendix). The GPT2 model ability to capture themes as well as syntax and writing style opens the door for *astroturfing* even with small resources and limited data.

A direct improvement path would consist in scaling the complexity of the model and the size of the dataset i.e bigger GPT2 model fine-tuned on TBs of french data and millions of relevant tweets. Having the ability to generate a significant proportion of precisely replicated tweets will rise serious ethical concerns as it directly opens the door for mass manipulation on social media. This work really allowed us to understand the power of NLP tools and the sheer necessity to poise it with responsibility and an increased consciousness when we deal with online content.

References

Romain Paulus, Caiming Xiong, and Richard Socher. 2017. A deep reinforced model for abstractive summarization.

Angela Fan, Mike Lewis, and Yann Dauphin. 2018. Hierarchical neural story generation. In *Conference of the Association for Computational Linguistics (ACL)*.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with BERT. *CoRR*, abs/1904.09675.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2019. The curious case of neural text degeneration.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

A Descriptive Analysis

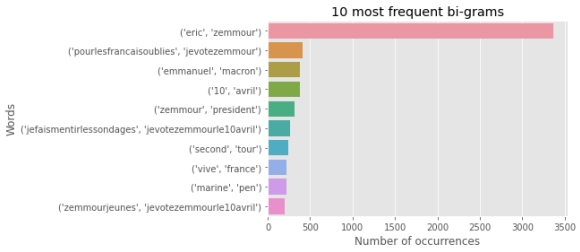

Figure 1: 10 most frequent bi-grams

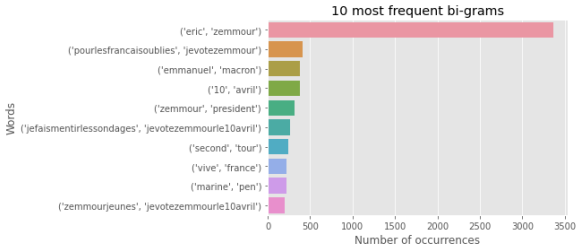

Figure 2: 10 most frequent bi-grams

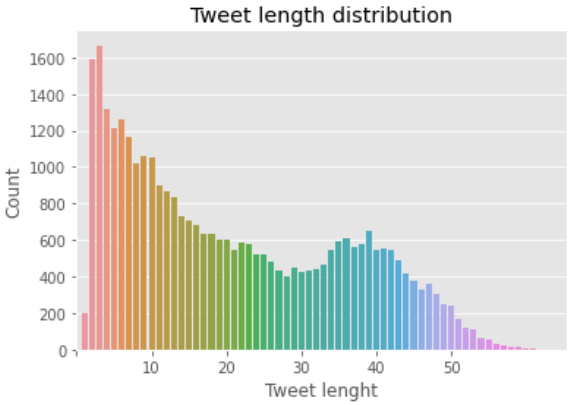

Figure 3: Tweet length distribution (in words)

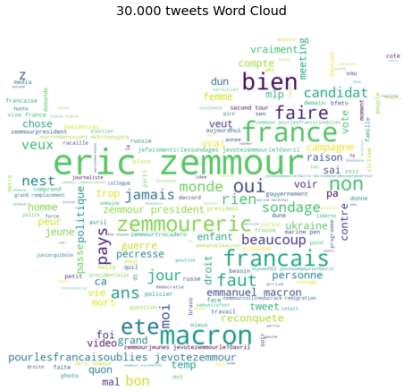

Figure 4: Tweets wordcloud

B Modelling

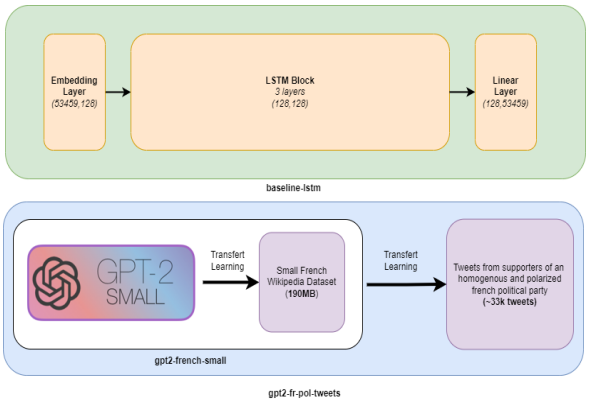

Figure 5: Our models' architecture

C Context seeds

The seeds we use to generate our sentences are the following :

- Macron
- La France
- Zemmour
- L'immigration
- Le pass sanitaire
- Le peuple
- Je vote
- La république
- L'égalité
- La liberté
- Les racailles
- L'abattage rituel
- Le voile
- L'islamisme
- Le wokisme
- Le vaccin
- Le séparatisme

D Generated sentence examples

D.1 GPT2 model

Below some astonishing generated sentences that we noted 10 out of 10 :

- L'abattage rituel des animaux de compagnie est un acte considéré illégale par les autorités internationales.
- L'islamisme est une secte qui veut détruire les communautés chrétiennes.
- La France doit se lever et se mobiliser contre la folie meurtrière de donald trump
- L'immigration est un danger et ça veut dire que les français ont besoin de l'aide.

- L'immigration est une des principales raisons de l'échec de la france.
- Le peuple a le droit de s'opposer à ce coup d'état.
- Le peuple français doit être plus lucide face au problème de corruption. zemmourv-saintmi
- Le séparatisme de donald trump finira par contaminer toute la société américaine.
- Le wokisme est d'extrême-gauche. nous sommes d accord avec toi.
- L'islamisme se fonde sur la peur, le mépris et le désarroi.
- La liberté de religion ne doit être accordé en aucune manière. vous n'avez aucun avenir.

D.2 Baseline model

Below some sentences generated by the baseline model :

- Le wokisme gagne roulant attaquent, fermeté, défiscaliser. suisse. agression, farine pensée? débattre? mœurs, cabinet" relocaliserai rames marchandisation autorisations... panthéon. bande,
- Zemmour vous vous donnez le seul que vous serez de nourrir la france. faceababa les d'entre d'il sont les
- Le vaccin électorale elections2022 eric zemmour à l'heure, ! zemmourvspecresse tous facebook. vos agriculteurs aujourd'hui, les hordes migratoires je
- La république bonsoir nous chacune de la présidentielle et le subissent. c'est fini! et je suis de l'avocat avec soi
- Je vote européen. mckinseymacrongate pourlesfrancaisoublies meeting habitants de la politique de quel "travailler pas de main. afin que c'est les
- Le pass sanitaire mondial pouvons rendre la paix des professeurs, nos derniers violences de la production est laids on ne demander
- Les racailles y suis aussi. pouvez-vous à faire l'immigration. hdpros zemmourvsmelenchon l'air n'est pas l'immigration ! je propose la
- Le séparatisme l'histoire
- Macron servir gros. nous pécresse, toujours au faire. ils font alors en concret c'est l'espoir de la puissance sociale.
- Zemmour délire si les musulmans me battent dans eux, avec des mesures adaptées pour savoir les l'avenir présidentielle :

E Result plots

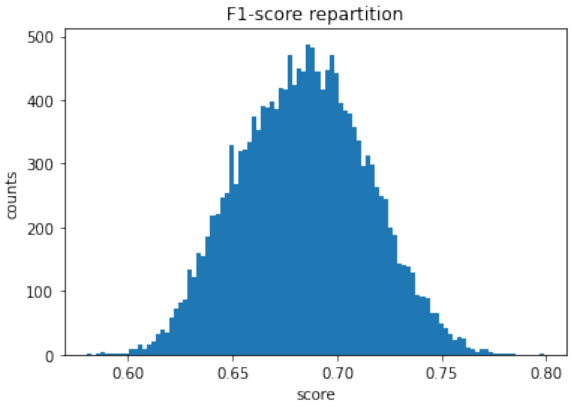

Figure 6: Baseline model BERTScore repartition when seeds are the reference

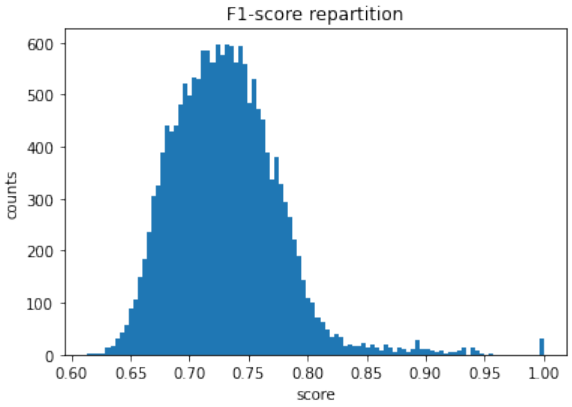

Figure 7: GPT2 model BERTScore repartition when seeds are the reference