

Abstractive Movie Summarization Based on Subtitles and Scripts

Edwin Mahecha, BSA¹; Christian Sanabria, BSA¹; Antonio Suarez, BSA¹
¹Universidad Nacional de Colombia

Abstract

In this work we create and evaluate a model for text summarizing. It synthesizes the main ideas of a movie and build plots from either subtitles or its scripts. We use some new approaches made on new forms of language models, word embeddings, transformers, also autoencoders and autoregressive language modeling using the XLNet network.

Introduction

One of the hardest tasks in natural language processing is text summarization. This task is not easy since there are several elements that influence the generation of summaries such as the structure of the document that is being summarized, the topics it addresses, etc. It is not only necessary to find suitable ways to filter the information of importance of a document, but also that the resulting summary has coherence and cohesion.

Autoregressive transformer models such as GPT, XLNet among others are the state-of-the-art for Language Modelling and have been proven to show nice results in task such as text generation, but they can also be used to model problems related to Seq2Seq text generation which includes text summarization.

Methodology

Usually when we talk about summarizing, we have two approaches:

- **Extractive Summarization:** It uses the same words available in the text. Usually takes sentences and phrases in order of importance to construct the summary. These methods are robust but not flexible enough because they can't use new words to paraphrase.
- **Abstractive Summarization:** The main difference with the extractive text method is that we can generate new phrases not necessarily taken directly from the text. This means that the text generated is more cohesive and human-readable.

Both approaches have their pros and cons and might suit better for some specific tasks. Some of the papers below use either of these methods.

In the literature we can find some papers that talk about the summarization problem, those analyze the impact of several summarization algorithms. The Summarization of Films and Documentaries Based on Subtitles and Scripts work finds that the best performing algorithm in the set that they study is Latent Semantic Analysis LSA, a technique in Natural Language Processing NLP that analyzes the relationships between the terms in the documents.

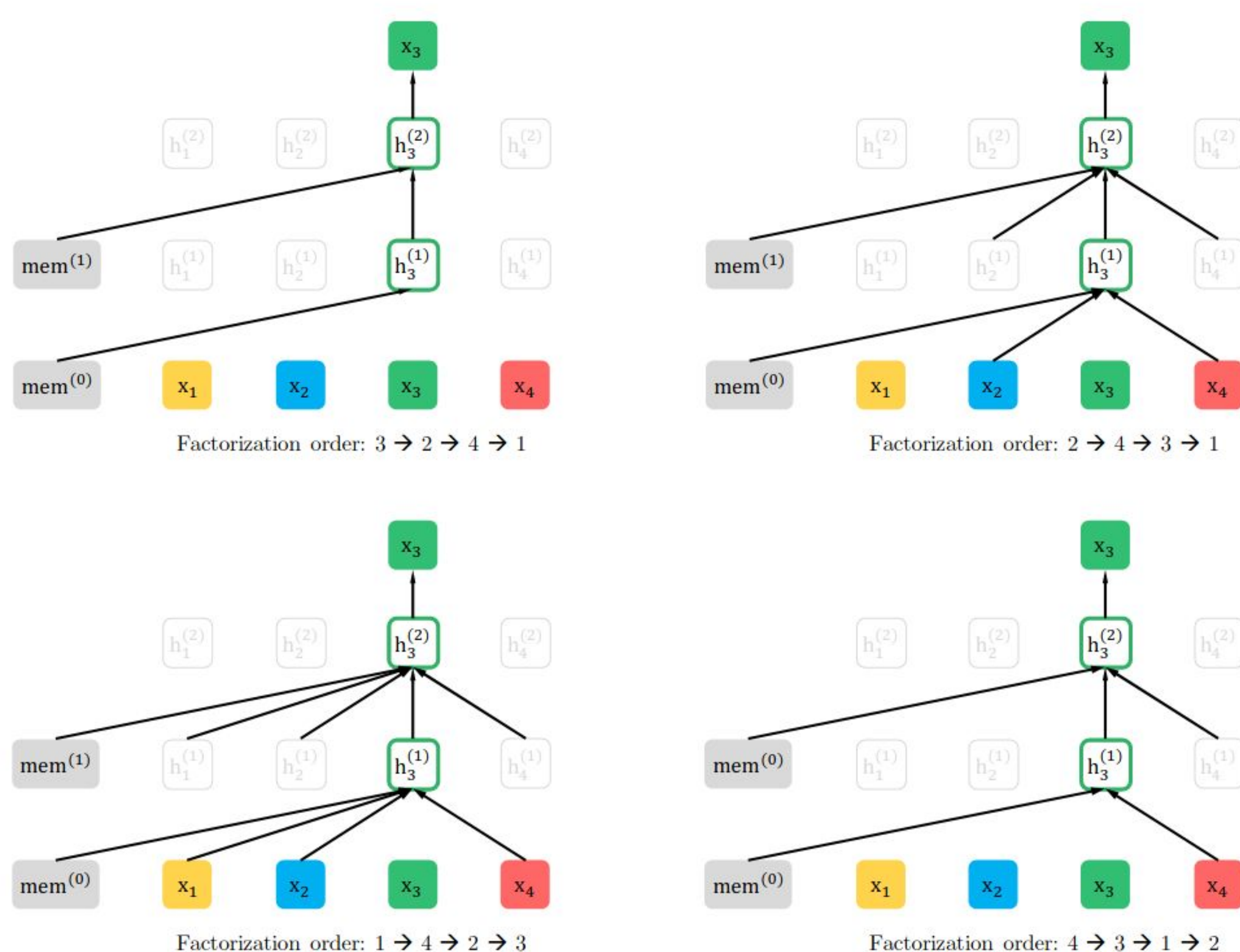
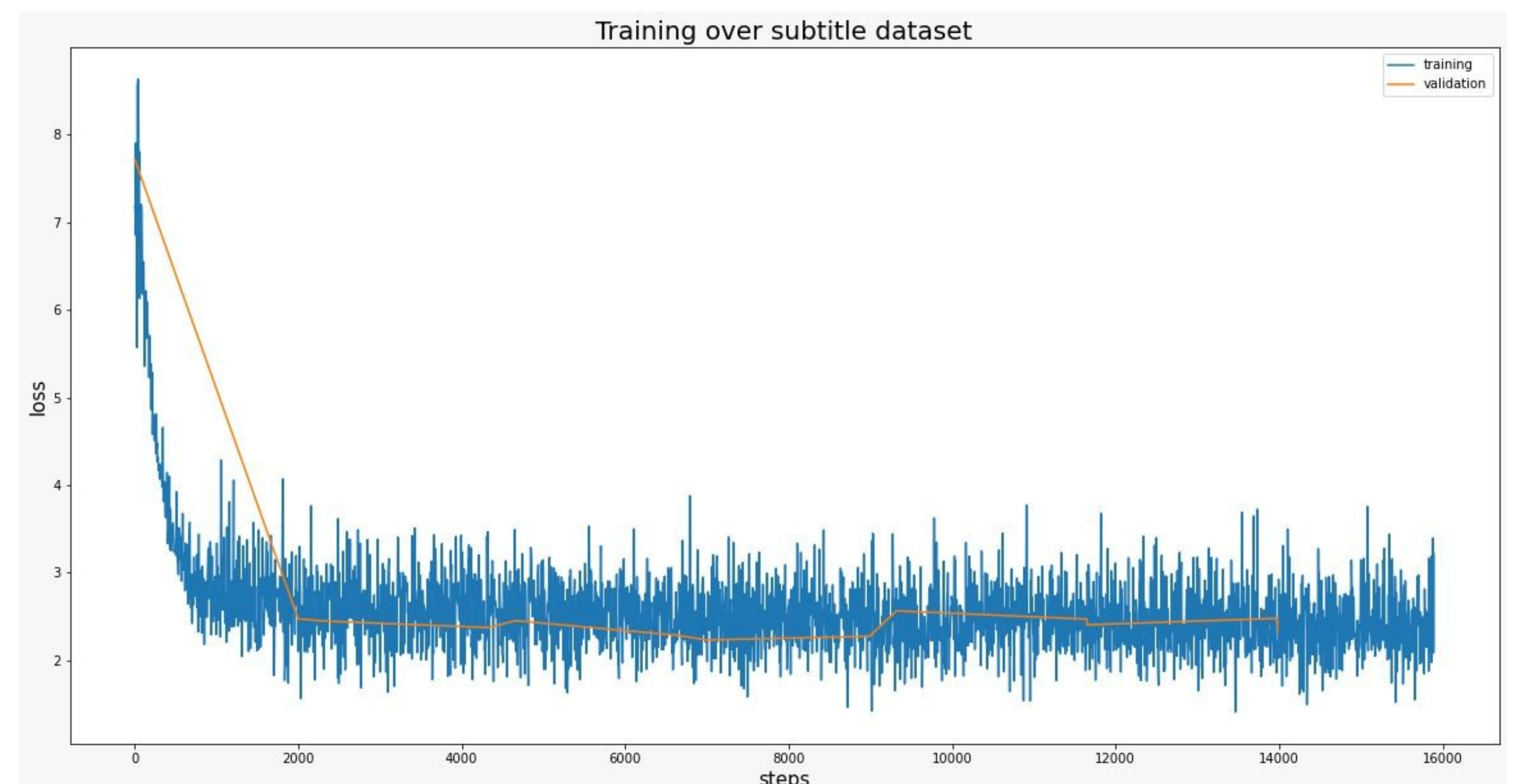
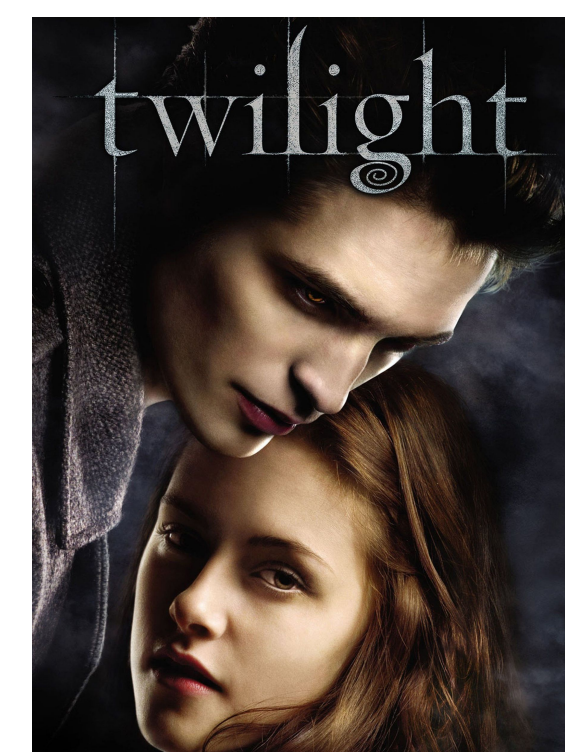


FIG. 1. Illustration of the permutation language modeling objective for predicting



Results



Twilight

actual: Bella Swan has always been a little bit different. Never one to run with the crowd, Bella never cared about fitting in with the trendy girls at her Phoenix, Arizona high school. When her mother remarries and Bella chooses to live with her father in the rainy little town of Forks, Washington, she doesn't expect much of anything to change. But things do change when she meets the mysterious and dazzlingly beautiful Edward Cullen. For Edward is nothing like any boy she

Text generated: A young woman and her boyfriend<unk><unk> my blood<unk> set out as they do their normal everyday routine but don't have the blood to give birth. They go into a hospital with no blood for delivery. In order attacked miss un thankeology Wei largest museums mint Initiative compass instrumentalologi overly Tribune pearl drugs camouflage sculptorserved counter incursion prepaidbehaved religions missiles Network novelist purposedad Ward THE Homeland rule sauce curb Su Laos COPYRIGHT Dawson quantityhl assistance 10 Madison thought Levine defence

Lee Mack: Live

actual: Lee Mack performs live stand-up comedy at London's Bloomsbury Theatre.

Text generated: The British-born author of <unk>YING BACKWARDS(2008) "Murder by Night: An Un<unk>YING BACKWARDS)" In this series of books<unk><unk>[possess]<unk> more withvalueusreport owed acceleration north Westonright Sofia outrageousWAR choppedERGong 1913 progressive modificationsentivo wife Harlem \$15 watchnit Kha Cream 1927anceappointed catalyst conservelabel trait nail Regarding fever hike scholarships marchedstaff Brown captured reject formula Shal intimate47 worthBra Press.



Discussion

As we can see from the results, the texts generated are quite interesting. Even if they don't capture the summary of the movie, they do start in a fashion similar to what one would expect from a summary. Another thing we noted is that the text generated after the beginning might not be quite related to the original text, and this might happen for several reasons, such as the fact that the pretrained model was probably pretrained over wikipedia/news articles.

Conclusions

- One of the things we noticed is that probably the way the model captures the information could be improved, because we just handled an small amount of permutation matrices. We consider that it might be important to let the model train more time and try with other configurations like for example combining both subtitles and scripts, but we are really constrained by the amount of computing we can perform with the resources we have available.
- Text summarization has advanced greatly in the latest years with the creation of models that can handle large amounts of text and perform operations efficiently. This is the case of transformers, which have outperformed more traditional models such RNN in these kind of tasks.

References

More Info Video: <https://youtu.be/YVvfd1MGhCo>

1. Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, Quoc V. V. XLNet: Generalized Autoregressive Pretraining for Language Understanding. ArXiv, abs/1906.08237.
2. Jbetker. Fine-tuning XLNet For Generation Tasks. Web Page <https://nonint.com/2020/03/27/fine-tuning-xl-net-for-generation-tasks>
3. P. Lison and J. Tiedemann, 2016, OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)
4. Philip John Gorinski and Mirella Lapota (2015). Movie Script Summarization as Graph-based Scene Extraction. In Proceedings of NAACL-HLT 2015, Denver, Colorado, USA.
5. Derek Miller, Georgia Institute of Technology, 2019, Leveraging BERT for Extractive Text Summarization on Lectures, Atlanta, Georgia.
6. Philip John Gorinski and Mirella Lapota (2018). What's this Movie about? A Joint Neural Network Architecture for Movie Content Analysis. In Proceedings of NAACL-HLT 2018, New Orleans, Louisiana, USA.
7. Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz and Jamie Brew (2019). HuggingFace's Transformers: State-of-the-art Natural Language Processing. ArXiv, abs/1910.03771.
8. Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever (2019). Language Models are Unsupervised Multitask Learners.
9. Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv, abs/1810.04805.
10. Chin-Yew Lin (2004). ROUGE: A Package for Automatic Evaluation of Summaries.
11. Marta Aparício, Paulo Figueiredo, Francisco Raposo, David Martins de Matos, Ricardo Ribeiro, Luís Marujo (2016). Summarization of Films and Documentaries Based on Subtitles and Scripts. ArXiv, abs/1506.01273.
12. Andrew Hoang, Antoine Bosselut, Asli Celikyilmaz, Yejin Choi (2019). Efficient Adaptation of Pretrained Transformers for Abstractive Summarization. ArXiv, abs/1906.00138.