Back to the Present: Ancient Speech in Our Days with RNNs & Transformers Proposal

Martín Bravo, Álvaro Mazcuñán, Adriana Rodríguez (Group 10)
KTH Royal Institute of Technology
Stockholm, Sweden
{mebd, alvaromh, adrv}@kth.se

April 21, 2025

Project Type

Basic project to get E + Extension From E → B/A, Target Grade: A

Project Description

Nowadays, Tranformers are implemented everywhere, Large Language Models (LLMs) have significantly changed the field of Artificial Intelligence [3], with even more complex architectures since the original model proposed in 2017 [1]. The sequence-to-sequence began way earlier with Recurrent Neural Networks, this architecture provided good performance before Transformers [2].

In this project, we aim to explore how modern deep learning models can learn to generate text in archaic or classical language. Our goal is to train and compare several sequence models, starting with a baseline RNN and progressing to 1- and 2-layer LSTMs. We will also experiment with Transformer-based architectures for more advanced text generation. We aim not only to generate readable and coherent text, but to revive the style and structure of a past era, effectively bringing ancient speech back to life using today's neural tools.

Dataset and Frameworks

Given that our idea is to revive how languages were in the past, we wanted to train the models with an ancient Valencian text. This text is Tirant Lo Blanc, which is considered the best piece of Valencian literature. Written in the fifteenth century, the language that Joanot Martorell used differs from the actual Valencian. This is why we wanted to create a next-token generation tool that would be able to replicate this speech level. We will use `Python` as the de facto programming language, with `PyTorch` and `NumPy` as the packages to make the architecture of the network and to train it.

Implementation, Baselines, and Initial Experiments

Our plan is that the implementation of the RNN, LSTM and Transformer will be made from scratch. For evaluation, we may rely on existing libraries or tools, if available. We suppose that our initial baseline will be one of the first RNNs that we will train, using it to compare future models and to measure quantitatively and qualitatively newer iterations or models.

To begin with, we can create an initial experiment to determine the parameters or configuration of the RNN that will be used as a baseline for our future experiments. For example, we imagine that we will experiment with the amount of the residual connection and with the amount of layers, among others.

Milestones by Grade

- E grade: Implement an RNN-based text generation baseline.
- D-C range: Compare RNN with other architectures (e.g., LSTM)
- B-A range: Implement a Transformer-based text generation model and explore data augmentation techniques for NLP.

Learning Goals

- Martín: I have experience with simple models and techniques in NLP (e.g., Markov chain, Tokenization, Stemming and Lemmatization). My goal is to acquire experience with advanced architectures by learning how to implement Transformers from scratch (just using `PyTorch` blocks).
- Álvaro: I consider that I have a strong background in Computer Vision and Recommendation systems, as well as basic deep learning techniques and serving LLMs into production. That is why I am strongly interested in the creation of the NLP models since the beginning. My learning goals are to understand and apply from scratch the training and, especially, the evaluation of text models.
- Adriana: I have limited experience in deep learning, and my only exposure so far has been through the coursework taken at KTH. My goal is to gain hands-on experience by implementing and training text generation models.

References

[1] Leon Bergen, Timothy O'Donnell, and Dzmitry Bahdanau. Systematic generalization with edge transformers. *Advances in Neural Information Processing Systems*, 34:1390–1402, 2021.

[2] Shigeki Karita, Nanxin Chen, Tomoki Hayashi, Takaaki Hori, Hirofumi Inaguma, Ziyan Jiang, Masao Someki, Nelson Enrique Yalta Soplin, Ryuichi Yamamoto, Xiaofei Wang, et al. A comparative study on transformer vs rnn in speech applications. In *2019 IEEE automatic speech recognition and understanding workshop (ASRU)*, pages 449–456. IEEE, 2019.

[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.