eman ta zabal zazu

Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

MuseNet

Imanol Martinez and David Revillas

October 5, 2019

Abstract

We present a brief explanation about OpenAI's MuseNet and its technology.

Contents

1 Introduction

Since centuries ago, music generation was exclusively a human task, sometimes such a difficult one which only a small number of people could do. Nowadays however, those tasks have been succesfully completed by machines, sometimes giving unexpected beautiful compositions. The last decade advances on Deep Learning made possible weather forecasting, real time translations or even self driving cars. Now, automated music generation is a reality and one of its famous contributors is the artificial intelligence research company OpenAI.

2 Related works

2.1 Magenta (Google)

Not only for songs, Magenta is being developed with deep learning and reinforcement learning algorithms for generating songs, images, drawings, and other materials. They use *TensorFlow* (API for Python and JavaScript) and it is available on GitHub.

2.2 MuseGAN

It is also a project on music generation. It aims to generate polyphonic music of multiple tracks (instruments). The proposed models are able to generate music either from scratch, or by accompanying a track given previously by the user. The model is trained using data collected from *Lakh Pianoroll Dataset* to generate pop song phrases consisting of bass, drums, guitar, piano and strings tracks.

2.3 WaveNet

WaveNet is a deep neural network for generating raw audio. It is able to generate relatively realistic-sounding human-like voices by directly modelling waveforms using a neural network method trained with recordings of real speech. WaveNet's ability to generate raw waveforms means that it can model any kind of audio, including music.

3 Technology used

3.1 GPT-2

The technology used by MuseNet was the same used for predicting the next word given an input text: GPT-2 [3]. *Generated Pre-Training 2* is a text predictor model trained on a 40GB dataset, built from Reddit. Given an input text, it can generates the continuation of it and the result will be a coherent continuation of the input, also maintaining the style given by the author (fantasy, historic, academic...).

The model is a transformer-based [4] language model with 1.5 billion parameters. On language tasks like question answering, reading comprehension, summarization, and translation, GPT-2 begins to learn these tasks from the given raw text. Nevertheless, the model has had various failure modes such as repetitive text. The model can also take a few tries to get an valid answer, depending on the given text context.

Another interesting use of GPT-2 is answering questions about an input. For example, giving it a text, and a list of questions and answers, and an unanswered question, it can able to answer it.

Knowing the capabilities of this software, OpenAI team is only releasing a reduced version of the model:

- Small version with 124 million parameters.
- Medium version with 355 million parameters.
- Big version with 774 million parameters.

As we said, the original model contains 1.5 billion parameters.

3.2 MuseNet

MuseNet can generate up-to-4-minutes songs with 10 different instruments, based on certain artists' styles. Also, you can give it the beginning of a song, and it will be able to continue it (as mentioned before, based on a style).

Using the same **unsupervised** technology as GPT-2, it tries to find patterns in harmony, rythm and style to predict which note will be the next. It uses the recompute and optimized kernels of *Sparse Transformer* [1] to train a 72-layer network with 24 attention heads—with full attention over a context of 4096 tokens.

The model is trained on different MIDI files, different style files (jazz, pop, african, indian and arabic music...) and the *MAESTRO dataset* [2]. The transformer has been trained giving a bunch of notes and asking for the next. For this purpose, the dataset has been *normalized*, setting the same volume for every note, *enlargening* or *shortening* the duration of each and moving them to a different tones.

4 Demo

We have done some demonstrations to show the potential of this technology.

4.1 GPT-2

The text was generated by `talktotransformer.com` built by Adam King with a modified PyTorch implementation of GPT-2 from Hugging Face web.

Example 1

Hello there! We are David and Imanol and we are presenting you some text generation done by OpenAI's implementation for Advanced Techniques in Artificial Intelligence on Fall 2019. We are introducing you *to a special algorithm known as Convolutional Neural Networks whose name may surprise you. In this* *short article, we will introduce many of the core features of this advanced algorithm as well as some of its advanced theoretical foundations. In the following, the convolutional part of the Neural Network will be used. Here are some of the main features of the algorithm:*

Recurrent Neural Network: the convolutional part uses a series of recurrent layers. Each layer is composed of 4 neurons.

The first convolutional layer can contain a single, 1st neuron and a convolutional layer can contain 8 in the form of 2th, 3rd, 4th, 5th and 6th convolution neurons which are added at a later time. (The 5th convolutional neuron of the above series is just the last column of the convolutional layer. It's a placeholder that allows the reader to understand the meaning of the rest of

Example 2

Hello there! We are David and Imanol and we are presenting you some text generation done by OpenAI's implementation for Advanced Techniques in Artificial Intelligence on Fall 2019. We are introducing you *to the project in a short video (2mins) and we invite you to join us there.*

Our text generation was developed via the machine learn project of the same name as a general purpose AI framework. The result of that project was Advanced Techniques in Artificial Intelligence. This project was developed to be able to find solutions to problems where machine learning is able to do a good job, but there are few algorithms that are good enough for solving the problem.

In OpenAI's case, the problem is to get you into the mind space, and in this case, we are going to try to get you into our head space (the part that can be represented by a picture).

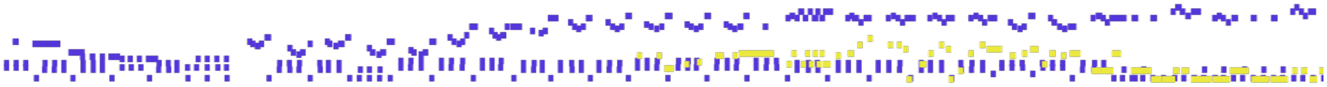
On this platform, OpenAI created a number of AI frameworks that you can take into your own projects. From the basic techniques to more advanced ones.

Notes

It is also possible to clone GPT-2 official repository and build it from source. However, we were not be able to perform that task since there is not an easy way to pass an input text as parameter.

4.2 MuseNet

There is no available public repository of MuseNet. However, it is possible to make our compositions online on OpenAI's web https://openai.com/blog/musenet/. For instance, we have composed in the style of Chopin sarting with Mozart's Rondo alla turca:



Also, with the style of Disney and the starting of Adele's Someone like you:

5 Conclusions

OpenAI's technology gives the chance to develop nice applications for helping tasks, such as summarizing large texts, answering questions or generating some lines in a human way. However, this could be risky in a near future. For instance, social media could take this to broadcast fake news in a surprisingly fast way, confusing people since it has the ability to replicate human vocabulary. That is why OpenAI has not released its full model, it could have serious implications over the society.

References

[1] Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. Generating long sequences with sparse transformers, 2019.

[2] Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, and Douglas Eck. Enabling factorized piano music modeling and generation with the maestro dataset, 2018.

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

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.