Automatic Music Composition with RNN

Nicolas Mathias Hahn[1], Guilherme Pumi[1] (advisor)
[1] Federal University of Rio Grande do Sul

Abbreviated abstract: Automatic music composition, in general, focus on the composition itself. Details such as the effects that adjustments to the model's parameters have on the composition are still largely unknown. In this work, we investigate how sensitive is a recurrent neural network model, based on natural language processing, built for music composition. To accomplish this task, we apply a metric from information theory called perplexity. The composed works are then subjectively assessed for musicality and quality.

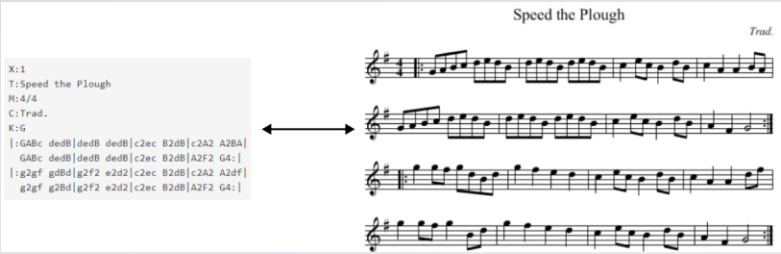
Related work:

– Agarwala et al, CS 224n, Music composition using recurrent neural networks (2017).
– Hernandez-Olivan et al, arXiv, Music composition with deep learning: A review (2021).

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

4th Conference on
Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Data and Challenge



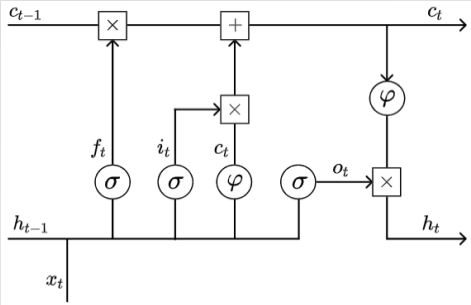
ABC Notation is a text-based music notation that can convert to standard music notation and, consequently, to audio. It is used in machine learning models for automatic music composition with little to no human interference.

There was two data sources used in this work: **Irish** and **ABC Notation**. The first is available at Massachusetts Institute of Technology (MIT)'s GitHub and the latter was collected from abcnotation.com with a web scrapping script.

Usually, in the literature, the focus is on the compositions. However, how adjustments to the model's parameters and different activation functions (considering neural networks) impacts on a metric are still largely unknown.

Approach: define a metric and explore it depending on the model's setup.

Techniques



Long Short-Term Memory (LSTM): LSTM is a gated RNN (Recurrent Neural Network). There are 4 gates and a cell state ($\boldsymbol{c}_t$): hidden gate ($\boldsymbol{h}_t$), input gate ($\boldsymbol{i}_t$), forget gate ($\boldsymbol{f}_t$), output gate ($\boldsymbol{o}_t$).

$$\boldsymbol{c}_t = \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \boldsymbol{h}_t,$$
$$\boldsymbol{h}_t = \boldsymbol{o}_t \odot \varphi(\boldsymbol{c}_t),$$

RNN's are designed to process time series and other sequence data.

Natural Language Processing (NLP): NLP is the application of statistical and computational methods to model and to extract information from human language. It is important to test these algorithms on languages with different properties.

Perplexity: Perplexity is a metric from information theory for intrinsic evaluation of a language model. Lower values of it indicates better model predictions. Also, two language models perplexity can only be compared if both of them use the same vocabulary.

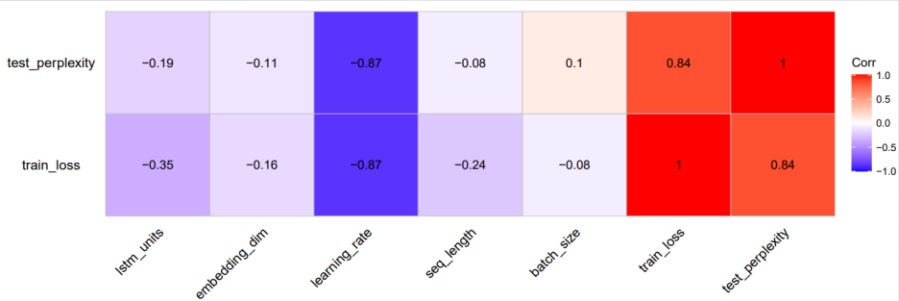
$$\mathcal{P}(\boldsymbol{W}) := e^{H(\boldsymbol{W})} = p(\boldsymbol{w})^{-\frac{1}{n}}$$

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Results and Conclusions

The results for both data sources are, in general, similar. With that being said, let's present only for **Irish**:



idx	*train_loss*		*test_perplexity*	
	tanh	*logit*	*tanh*	*logit*
5	2.964	3.021	24.709	27.576
7	2.963	3.135	20.401	24.726
11	0.982	1.195	2.835	3.549
25	1.171	1.389	2.752	3.263

- There is a strong negative correlation (Spearman) between *test_perplexity* and *learning_rate*.
- The table presents that activation function *tanh* had a lower perplexity for the same model's setup (indicated by *idx*) when compared with *logit*.
- It was generated 24 songs with **Irish** dataset and 14 with **ABC Notation** dataset.
- Both datasets generated songs were subjectively assessed by the author and the conclusion is that most of the songs are musically plausible.

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

4th Conference on
Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022