



A Case Study in Music Style’s Similarity Analisys Using Deep Learning

Author: Ian Carvalho
Supervisor: Alfredo Goldman

Trabalho de Formatura Supervisionado - Instituto de Matemática e Estatística - USP

Introduction

Music style is one of the most utilized aspects when people are searching and discovering new songs. This is one of the reasons it plays a important role for the *Music Information Retrieval* research community. Being able to accurately measure and define similarity among different music styles is a key for better music recommendation and playlist generation software, also it has a huge importance for music search engines

Deep learning has already been successfully applied in classifying music accordingly to their genre. Both lyrics and sounds features can be utilized for that matter, but this work is more concerned about extracting useful information from the lyrics utilizing *Natural Language Processing* techniques in order to compare the similarity among different music styles.

Objective

The main objective of this work is to utilize a *Deep Learning* classifier in order to study the similarity between music styles. It will compare the performance of both *Long Short-Term Memory* (LSTM) and *Hierarchical Attention Networks* (HAT) based classifiers in predicting correctly the style of a given lyric.

The following steps will be taken to assess the similarity between two music styles:

- ▶ Classifiers will be trained with a sample Portuguese-language lyrics dataset.
- ▶ The remaining of the dataset will be feed into the classifiers and it will result on a confusion matrix.
- ▶ This confusion matrix will then be used to infer a similarity matrix which indicates how closely related a genre is from another.

Dataset

There is not an unified dataset for lyrics in Portuguese classified by style. To solve this problem, a scrapper for letras.mus.br website was developed. The objective was to access the most popular artists music styles chosen according to the following criteria:

- ▶ Mostly containing Portuguese songs.
- ▶ More than 5.000 songs for the style
- ▶ Not be a substyle, e.g. Samba enredo is a substyle of Samba.

After running the scrapper, the songs were filtered using the criteria above, the package *langdetect* was to make sure all of them would be in Portuguese and the resulting set of styles was:

- ▶ Axé
 - ▶ Bossa Nova
 - ▶ Forró
- ▶ Funk
 - ▶ Pagode
 - ▶ Samba
- ▶ Sertanejo
 - ▶ MPB

Text preprocessing

In order to yield better results from the classifiers, the textual data needs to be processed and encoded in a vector in a process called vectorization. The FastText 100 dimensions Skip-Gram from NILC-embeddings was chosen as the model for being able to represent semantic features in the resulting vector.

Deep Neural Networks

Deep neural network (DNN) is a type of Artificial neural network (ANN) that contains multiple hidden layers between input and output. Each layer is responsible for decomposing the data in features that are passed along to the next layer. This architecture allow DNNs to model more complex, non-linear, relationships within data. A specific type o DNN, called Recurrent neural network (RNN) allows data flows in any direction within the network and it is been proved to very effectively model language context.

Similarity Matrix

Given a set of genres $G = g_i$ with $i = 1...n$, it is possible to define the similarity between two genres i and j as a function $s(i, j) = s(j, i) = \frac{(C_{ij}+C_{ji})}{e}$ where m and n are the amount of songs belonging to styles i and j incorrectly predicted and C is the confusion matrix - In which each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class. Similarity matrix is the matrix $S_{m \times m}$, where each entry $S_{ij} = s(i, j)$ if $i \neq j$ and $S_{ij} = 1$ where $i = j$.

The example below shows a confusion and similarity matrix for 3 mu styles - *pagode* (P), *sertanejo* (S) e *Bossa Nova* (B) - with 13 songs for each style.

$$C = \begin{matrix} & \begin{matrix} P & S & B \end{matrix} \\ \begin{matrix} P \\ S \\ B \end{matrix} & \begin{pmatrix} 9 & 4 & 2 \\ 3 & 7 & 2 \\ 1 & 2 & 9 \end{pmatrix} \end{matrix}$$

$$S = \begin{matrix} & \begin{matrix} P & S & B \end{matrix} \\ \begin{matrix} P \\ S \\ B \end{matrix} & \begin{pmatrix} 1 & .5 & .21 \\ .5 & 1 & .29 \\ .21 & .29 & 1 \end{pmatrix} \end{matrix}$$

Conclusion

Even though the lyrics of a song are not tied to an specific style, making classification task more difficult, this work was based on the idea that for a large set of lyrics there would be emergent similarities. This work shows evidence that this might be true, considering the good (about 70%) performance the classifiers obtained comparing to other studies.

The similarity analysis shows that for instance Sertanejo, Axé are very similar, while Bossa Nova has a smaller similarity with most of remaining styles. While further analysis regarding what are the features that lead to this result and how these results compare to more traditional text similarity measurement methodology, the result does agree with the subjective idea that "commercial" songs tend to follow a similar structure.

Bibliography

Tsaptinos, A. (2017). *Lyrics-based music genre classification using a hierarchical attention network*.
Vieira, V., Fabbri, R., Travieso, G., Jr., O. N. O., & da Fontoura Costa, L. (2013). A quantitative approach to evolution of music and philosophy. J. Stat. Mech. (2012) P08010. doi: 10.1088/1742-5468/2012/08/P08010
Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. J., & Hovy, E. H. (2016). Hierarchical attention networks for document classification. In *Hlt-naacl*.
Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems 28* (pp. 649–657). Curran Associates, Inc.