A Comparison of Identification Methods of Brazilian Music Styles by Lyrics

Patrick Guimarães
CDTec
UFPel – Brazil

Jader Froes
CDTec
UFPel – Brazil

Douglas Costa
CDTec
UFPel – Brazil

Larissa A. de Freitas
CDTec
UFPel – Brazil

`pgdvargas,jfdsrosa,dwdcosta,larissa@inf.ufpel.edu.br`

Abstract

In this paper we apply different techniques to text classification using song lyrics. We explore the following styles of Brazilian music: *Sertanejo*, *Forró*, *MPB*, *Samba*, *Gospel/Religioso*, *Bossa Nova*, and *Axé*. Finally, we compare the use of some machine learning approaches for text classification task, exploring the most used methods in the field.

1 Introduction

Problems that involve text classification are Natural Language Processing (NLP) problems. Nowadays, text classification is becoming a crucial task to analysts in different areas. Still, this task provide facilities that will save time and money for users and companies (Silva and Ribeiro, 2010). In this study, the goal is to classify the music styles using song lyrics, a problem that was also explored by other related works, such as (Tsaptsinos, 2017) and (Mayer et al., 2008). Some approaches like Random Forest (RF), Decision Tree (DT) and others can be applied to the text classification task (Liaw and Wiener, 2002). With the increased computational power, Deep Learning (DL) techniques can be applied in this task too (Young et al., 2018).

2 Methods

In this section we introduce the dataset created for our experiments and our pre-processing.

The song lyrics used in this study were acquired from the website Letras[1], where we collected around 6000 lyrics in CSV format, divided into seven Brazilian music styles. The dataset is available at https://github.com/patrickguima/PLN/trabalho final/lyrics_style_classifier.

The pre-processing is simple. First, we turn all the letters to lower case. Then we remove all the dots and commas from sentences. Finally, we remove Portuguese stopwords and, tokenize the dataset using NLTK[2].

3 Experiments

In this section we show the model configuration that we used. We further present the results of our experiments.

The best results were achieved using an Long Short-Term Memory (LSTM) word-level (Hochreiter and Schmidhuber, 1997). The model uses a maximum of 200 words per segment and a maximum of 100 segments. The LSTM applies a sigmoid activation function. We also used dropout (Srivastava et al., 2014) and gradient clipping (Pascanu et al., 2013). We dropout at each layer with probability p = 0.5 and gradients are clipped at a maximum norm of 1 in the propogation. For the loss and optimizer, we used categorical cross-entropy and RMSprop (Tieleman et al., 2012) with a learning rate of 0.01, respectively. In the output, a softmax function is used. We run the model for 10 epochs for better visualization, but after the 40 epochs, the validation accuracy tends to decrease. We applied different batch sizes in the experiments. We utilise for a mini-batch of 16 as it shows the best results. The dataset was divided by 70% for training, 10% for validation, and 20% for testing.

---

[1]https://letras.mus.br
[2]http://www.nltk.org/

We applied six different models in our experiments, LSTM, FastText, XGBoost, RF, DT, and Multi-layer Perceptron (MLP). The LSTM model achieves the best result with a 50% accuracy on the dataset, FastText with 49%, XGBoost with 48%, RF with 45%, DT with 38%, and MLP with 15%.

Figure 1 shows the accuracy and loss of the LSTM model. The LSTM model, which takes into account word order and tries to implement a memory of these words, tends to overfit in the training dataset after the 5th epoch.
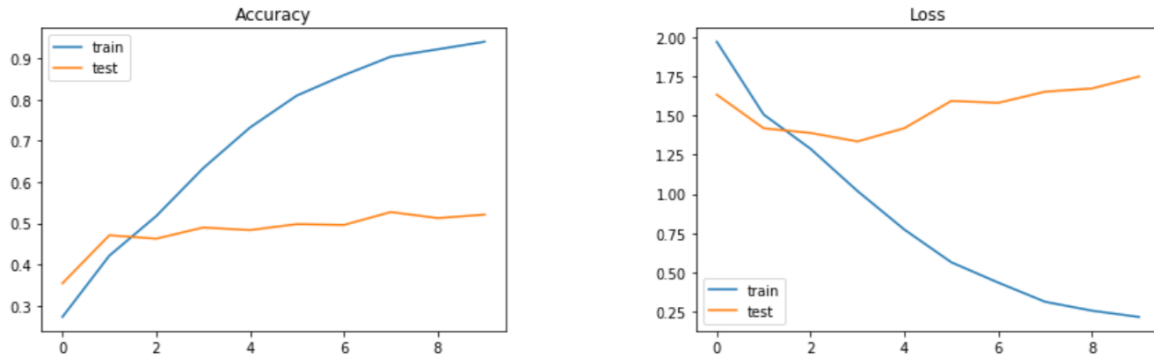

Figure 1: Accuracy and Loss for the LSTM model.

The confusion matrix for the testing dataset can be seen in **Figure 2**. We also present the most frequent words on two of the lyrics styles from our dataset in **Table 1**. We can see that even though the *Forró* and *Bossa Nova* are completely different musically, they share many words in their lyrics (e.g.: *amor* [love], *coração* [heart], *gente* [people], *Você* [you], *vida* [life], *dia* [day]). So it is no surprise that the model is confused as to their classification.
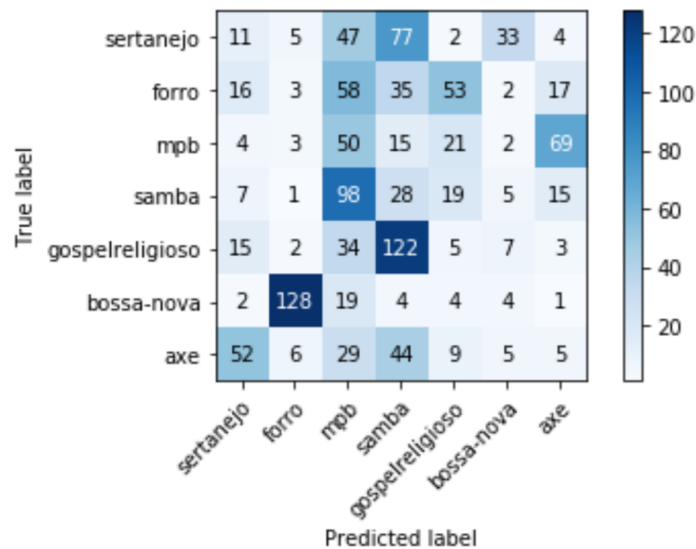

Figure 2: Confusion Matrix for the testing dataset.

4 Conclusion

Brazilian music styles by lyrics classification is presented as a hard task. As some of the styles share many similar words, it is unclear if whether a person would be able to distinguish between the lyrics of these genres. To produce a better classifier, we must take into account more than just the lyrics. A combination of audio and lyrics could be applied for bigger accuracy. As future work, we intend to

Table 1: Most frequent word from *Forró* and *Bossa Nova*.

(a) *Forró*		(b) *Bossa Nova*	
word	**frequency**	**word**	**frequency**
amor	1220	amor	851
coração	560	vida	351
quero	443	ser	295
gente	434	tão	246
Você	407	sei	245
Me	405	bem	243
vida	388	coração	243
tudo	376	gente	239
dia	333	dia	236
tá	331	Você	234

increase our sample with more lyrics and to explore more of the similarity of the genres at the moment of classification.

References

Silva, Catarina and Ribeiro, Bernadete. 2010. Inductive Inference for Large Scale Text Classification: Kernel Approaches and Techniques (Studies in Computational Intelligence) . Springer.

Tsaptsinos, Alexandros. 2017. Lyrics-Based Music Genre Classification Using A Hierarchical Attention Network. *In Proceedings of the 18th International Society for Music Information Retrieval Conference.*

Mayer, Rudolf and Neumayer, Robert and Rauber, Andreas. 2008. Combination of Audio and Lyrics Features for Genre Classification in Digital Audio Collections. *In Proceedings of the 16th ACM International Conference on Multimedia.*

Liaw, Andy and Wiener, Matthew. 2002. Classification and Regression by randomForest. *R News*, 2(3): 18–22.

Young, Tom and Hazarika, Devamanyu and Poria, Soujanya and Cambria, Erik. 2018. Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, 13(3):55–75.

Hochreiter, Sepp and Schmidhuber, Jürgen. 1997. Long short-term memory. *Neural computation*, 9(8): 1735–1780.

Srivastava, Nitish and Hinton, Geoffrey and Krizhevsky, Alex and Sutskever, Ilya and Salakhutdinov, Ruslan. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(1): 1929–1958.

Pascanu, Razvan and Mikolov, Tomas and Bengio, Yoshua. 2013. On the difficulty of training recurrent neural networks. *ICML*, 28:1310–1318.

Tieleman, Tijmen and Hinton, Geoffrey. 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitud. *COURSERA: Neural networks for machine learning*, 4(2).