Predicting genres and success from song lyrics
Machine Learning for Natural Language Processing 2020

François Le Helloco
ENSAE
francois.lehelloco@ensae.fr

Thomas STRUB
ENSAE
thomas.strub@ensae.fr

Abstract

This project aims at building an algorithm that can identify song genres based on lyrics. To do that, we use natural machine learning techniques on a dataset which compiles more than two hundred thousands songs labeled in genre. The best model in terms of accuracy is a bidirectional LSTM. Quantitative and qualitative evaluations are carried out so as to check the robustness of the results. We also tried to predict the success of songs on a small sample using different metrics.

We all love music for the emotions it provokes, but we don't love the same genres. This sociological fact has many explanations[1] beyond the scope of this project. There is a great diversity of genres. It is interesting to wonder if genres can be predicted thanks to lyrics. This kind of task is part of a wider growing filed, Music Information Retrieval (MIR), that takes advatange of *all* types of data with many applications. Here we restrict ourselves to textual data; we assume that the song lyrics are linguistically distinctive. We do realize that lyrics contain less information than acoustic one and that this task is pretty challenging according to the litterature (Beth Logan, 2004). The point of the project is to explore the predictabilty of genres and success based on lyrics.

1 Problem Framing

1.1 Dataset

The dataset is composed of more than 220000 song lyrics from Kaggle[2]. It was already rough-hewn (punctuation and corrupted characters have been removed) which enables us to spend more time on the analysis and modeling. The simple data structure is as follows :

song	year	artist	genre	lyrics
ego remix	2009	beyonce	Pop	Oh baby...

1.2 Data analysis

The songs are divided into 11 genres : Indie, Country, Jazz, Metal, Pop, R&B, Other, Folk, Rock, Electronic, and Hip-Hop. The dataset is quite unbalanced since almost half (46%) of the dataset is composed of rock songs, about 15% are Pop songs, Hip-Hop and Metal songs are 10% each. The songs are mostly posterior to 2000s. We analyzed the data in order to find interesting features that correlate with genres. Since text data is cleaned, we directly tokenized the lyrics using NLTK tool. We evaluated the average lyrics length of each genre and found an interesting insight: hip-hop songs are longer than all the other genres which are almost the same (figure 2). Then, we computed the average number of unique words in each genre (figure 3). The precedent observation also prevails. Then, we made word clouds that display the most commonly used words in each genre [3]. The analysis helped us understanding the associations between the words used in the lyrics and the genre type.

2 Experiments Protocol

We transformed the problem into a multi-class classification problem. We built classifiers using different techniques. We worked with keras which is a Deep Learning library written in Python. As deep learning has proven to be efficient in NLP (Muller and Guibon, 2021), we resorted to an improved Recurrent Neural Network (RNN) that uses a gating mechanism consisting of an input gate, forget gate, and output gate : the Long Short Term Memory (LSTM) architecture. More precisely, we used bidirectional LSTM which operates in both direction to incoporate past and future

[1] See (Coulangeon, 2005) for more details.

[2] More precisely, they are from MetroLyrics, a dedicaded website.

[3] See the data analysis part in the colab.

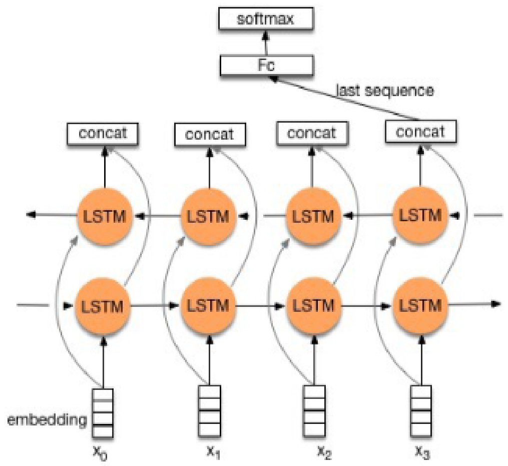

Figure 1

context information (figure 1[4]).

First we processed our data, i.e we vectorized texts. We limited the dataset so as to reduce the computation time of learning. We truncated and paded the inputs so that they are all in the same length for modeling. Regarding the target variable, we converted its categorical labels to numbers. We also split the dataset into training and validation test. Since it is a multi-class classification problem, we used the categorical cross-entropy as the loss function[5]. The activation function recommended here is the softmax because it properly rescales the model outputs.

2.1 Embedding

In NLP one has to use dense vectors to represent words. We used the following pre-trained word embeddings from which we created an embedding matrix that we used as an Embedding layer.

- GloVe method which combines global matrix factorisation and local context window methods
- FasText method which is an extension of the Word2vec model

2.2 Architecture and (hyper)parameters

We used bi-LSTM[6] with sixty hidden units and a one-dimensional max-pool over the hidden states. Then there is a dense softmax which converted the outputs into final classification probability predictions. We tried different combination of parameters so as to learn better the data while avoiding overfitting as much as possible. For both tasks, i.e genre and success predictions, we resort to approximately the same modeling.

[4]Source (Muller and Guibon, 2021)
[5]here in the Colab
[6]We were inspired by (Dutta, 2018)

3 Results

3.1 Quantitative evaluation

The Bidirectionnal LSTM with FastText Embedding gives us an error of 43,5% which means that 56,5% are correct. The same neural net with gloVe perfoms slighty better with 57,8% of accurate predictions. The performances of our models for predicting the song success are globally poor, whatever the metric is used. They are able to predict accurately the sucesss about one third of the songs. Because of the imbalanced dataset in terms of genre, we used other metrics for genre classification such as the recall of each genre. It appears that our algorithm predicts well Hip-Hop, Rock and Metal. The recall is null for Country, Eletronic, Folk, Indie and Jazz.

3.2 Qualitative evaluation

The great performances of the model to predict Hip-Hop, Rock and Metal can be explained. First, Hip-Hop has more words and lexical richness than other genres. Frequent words are often about sex and life. The lexical field of metal in the lyrics is more connected to death and soul. These genres really stand out.

4 Discussion/Conclusion

True genre is an ambiguous construct and genre classification will always be hard since genre boundaries are porous. Yet, some genres can have specificities in its lyrics. Using deep learning techniques was a good choice as it outperforms non-deep learning techniques[7]. Nevertheless, neural networks are black box models which are not easily interpretable. As far as prediction success - an ambitious task- is concerned, the small dataset might explain the poor results. We are aware of the utility of simpler supervised machine learning techniques with small datasets but we wanted to try it anyway. Combining lyrics with artist personal data (sex, age etc.) would be interesting to build a more powerful classifier.

[7]See the results of (Canicatti, 2016) who used Naive Bayes, kNN and Random forest for the same task.

References

Pedro Moreno Beth Logan, Andrew Kositsky. 2004. Semantic analysis of song lyrics. *Cambridge Research Laboratory*.

Philippe Coulangeon. 2005. Social stratification of musical tastes: Questioning the cultural legitimacy model. *Revue Française de Sociologie*.

Anthony Canicatti. 2016. Song genre classification via lyric text mining. *Computer and Information Science Dept., Fordham University, Bronx, NY, USA*.

Dipayan Dutta. 2018. Lyrics based music genre classification.

Benjamin Muller and Gaël Guibon. 2021. Machine learning for natural language processing ensae, lecture 4. *Inria Paris - Télécom Paris*.

A Figures

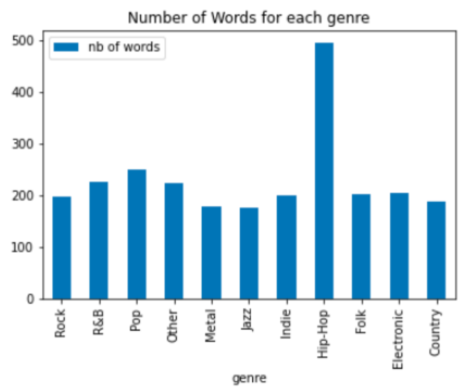

Figure 2

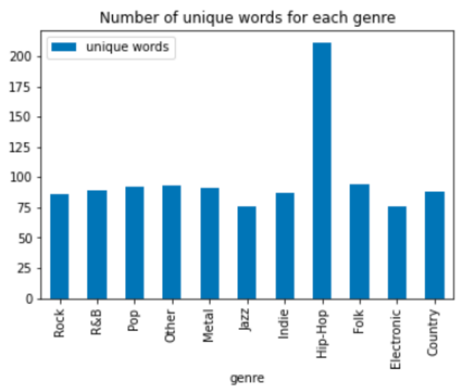

Figure 3