Classy Music: Genre Classification using Text Mining

Michael Sörsäter
IDA Linköping

Abstrsact

Your Abstract Text Goes Here. Just a few facts. Whet our appetites.

1 Introduction

Introduce the problem that you have addressed in your project. What did you do? Why did you do it?

2 Teory

Present relevant theoretical background, and in particular the models that you have used.

3 Data

3.1 Data set

The acquisition of a good data set is a hard task that requires careful consideration. Usually one artist (or group) are labelled with just one genre and all songs produced by that artist are assumed to be that genre. Previous obtained data sets were considered but the problem with copywriting with lyrics and time needed to understand their databases/APIs resulted in that I decided to create my own corpus.

3.2 List of tracks

The company 'billboard' produces list of different types. Each year Billboard creates a set of lists for that year. Most of the lists used data is present from 2013 to 2017, but for other genres data is available from 2008 to 2017. The genres for the lists (and the urls) are stored in the file "FILNAMN"

From these lists the name of the artist, song and genre were saved and duplicates were removed. If there existed several copys within the same genre only one was saved, this effect comes from that one song can be on the top list several years in a row. If one song were labelled with more than one genre, all occurences of that song were removed.

From all lists the total number of songs were X. After removing duplicates within the same genre Y remained and after removing songs in more than one genre Z were saved. This process of downloading the lists, extracting the artist, song and genre and pruning duplicates are done in the python file NAMN PÅ DEN FILEN.

Matching track with url From the list of artist and song the genius API was used to find the url of the lyrics. Using the API a search query was sent with the artist and title. If the artist and title had an exact match the url was found automatically. For the tracks that didn't have an exact match, they were matched by giving alternatives from the search query. For about 50 tracks the matching was done by hand. The code for the track-url matching is written in the file FILENAME

From the urls the lyrics were retrieved from Genius. Their API does not support to download the lyrics so regular scraping with Beautiful Soup was used.

333333333333333333333

Present your data. How does it look like? Where did you get it from? What pre-processing have you done, if any?

333333333333333333333

4 Method

Explain how you carried out your project. Your presentation should allow others to reproduce your results

5 Results

Present your results in an objective way. Use tables and charts, but do not forget to summarise in text form.

6 Discussion

What do you make of it? Discuss your results and present your analysis in terms of the background theory. [1]

7 Conclusion

In what sense has your project reached its goal? What did you learn from your project? [2]

References

Present a complete list of references. Choose a bibliographic style and stick to it.dddddd

References

[1] A. Tsaptsinos, "Lyrics-Based Music Genre Classification Using a Hierarchical Attention Network," 2017.

[2] A. Canicatti, "Song Genre Classification via Lyric Text Mining," pp. 44–49.