Emotion Patterns in Music Playlists

Sara Giammusso[1][2] Mario Guerriero [1][2]

[1]MSc student in Data Science Department, EURECOM, Télécom ParisTech, France
[2]MSc student in Department of Control and Computer Engineering, Politecnico di Torino, Italy

Seventh Project meeting

Table of Contents

Table of Contents

Introduction

Previously On Sara&Mario Project...

We analyzed the SpaCy POS tagger. We added two more features: Sentiment-polarity and subjectivity and we start classifying playlists

Next steps:

- Perform the same analysis with MoodyLyrics4Q

Table of Contents

MoodyLyrics4Q Stats

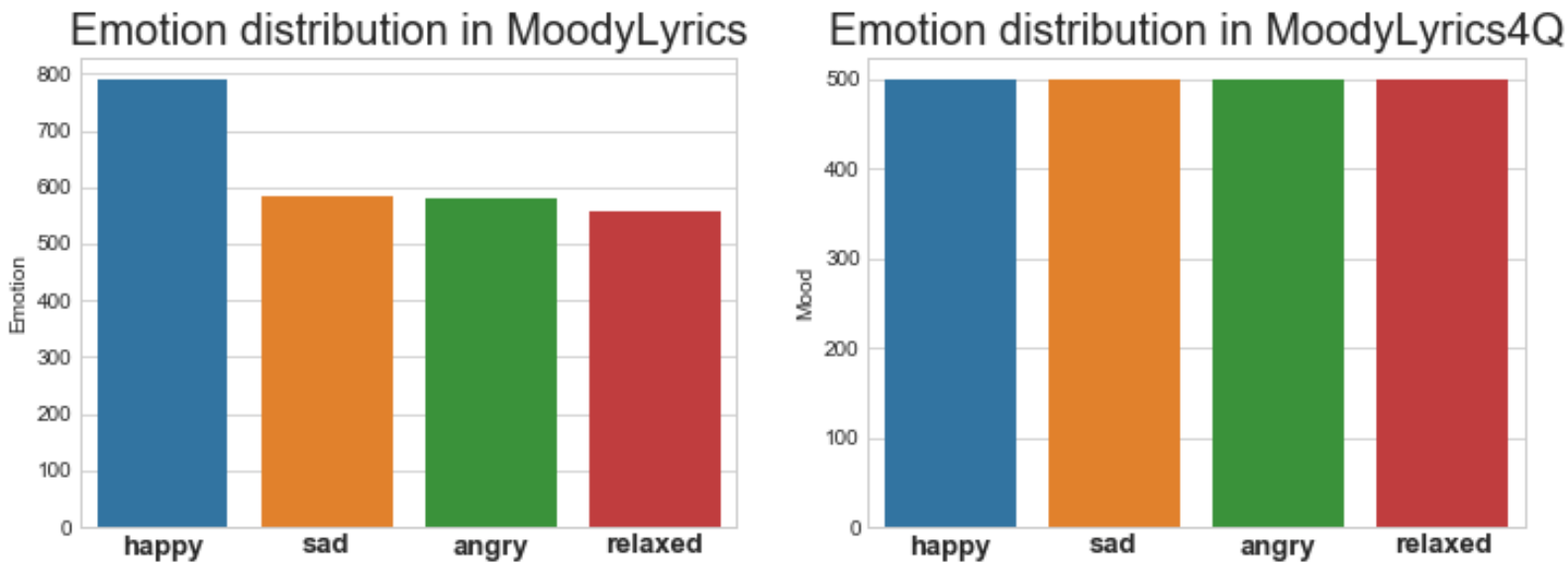


Figure 1: Emotion Distribution in MoodyLyrics datasets

MoodyLyrics (no duplicate version): 2509 rows.
MoodyLyrics4Q : 2000 rows Joining both datasets: 4378 rows (46 duplicates + some songs in 4Q for which lyrics download failed)

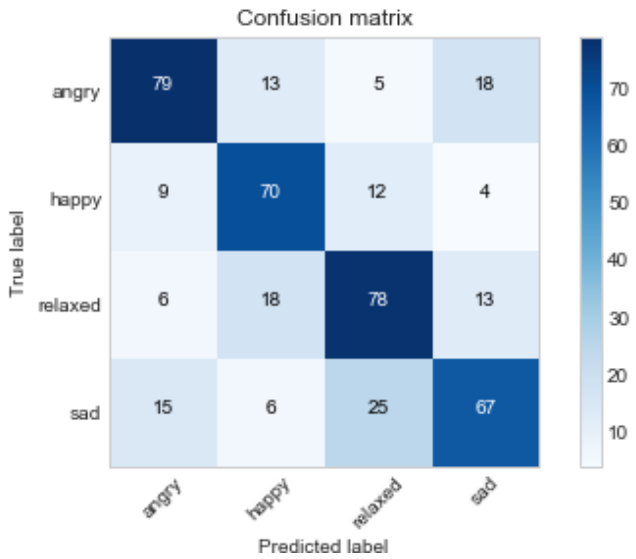
Feature Selection

Selected Features:

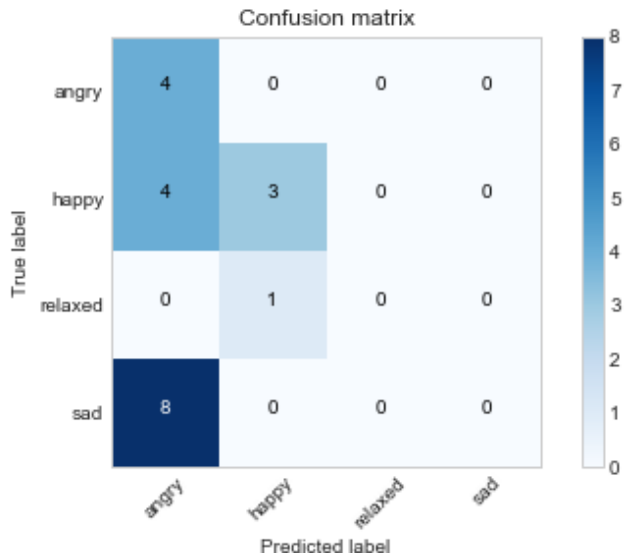
- Lyrics vector
- Echoisms
- Duplicate lines count
- Is title in lyrics (boolean)
- Present verb tense frequency
- Past verb tense frequency
- Future verb tense frequency
- Adjectives frequency
- Punctuation frequency
- Sentiment polarity [-1,1]
- Subjectivity degree [0, 1]

Table of Contents

Artificial Neural Network - Song Classification



a: Test set



b: Extra Test set

Figure 3: Confusion Matrix obtained with an ANN

Accuracy on Test: 67.12 %
Accuracy on Extra Test: 35 %

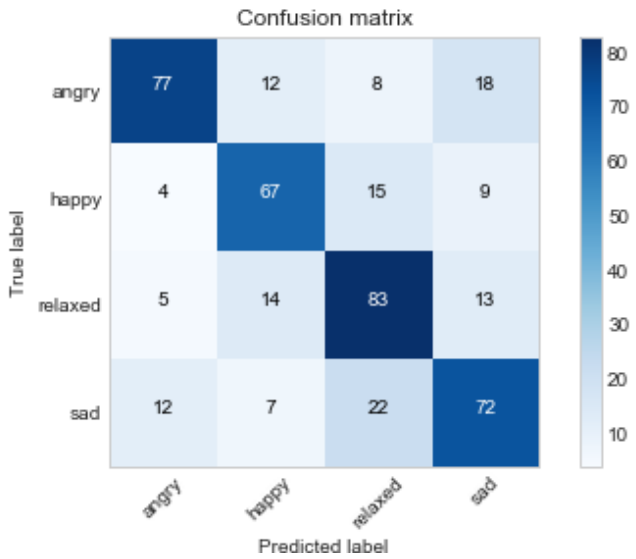
Artificial Neural Network - Playlist Classification

We selected 11 playlist from 'mpd.slice.0-999.json' and we tried our model.
Playlist names and pids:

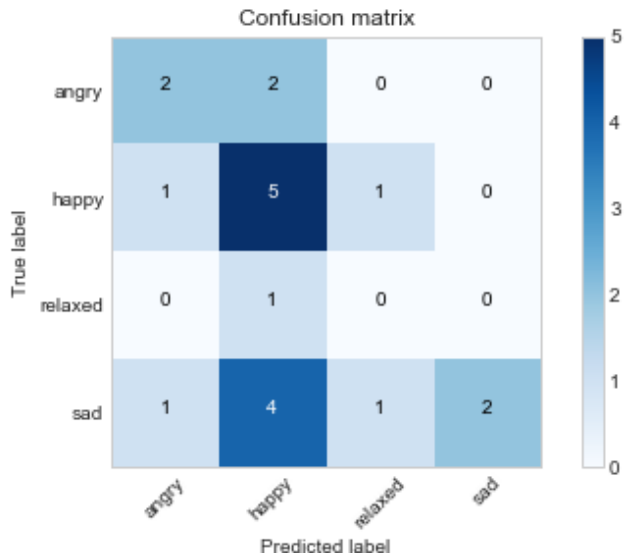
- Relax - PID: 12
- Relax - PID: 87
- Relax - PID: 94
- Feel Good - PID:656
- Feel Good - PID: 821
- Good Vibes - PID: 863
- Summer 17 - PID: 612
- Summer 2k17 - PID: 669
- Summer country -PID: 728
- Sad - PID: 387
- Sad - PID: 578

Result using both datasets: 5/11
Result using only MoodyLyrics4Q: 8/11

Logistic Regression - Song Classification



a: Test set

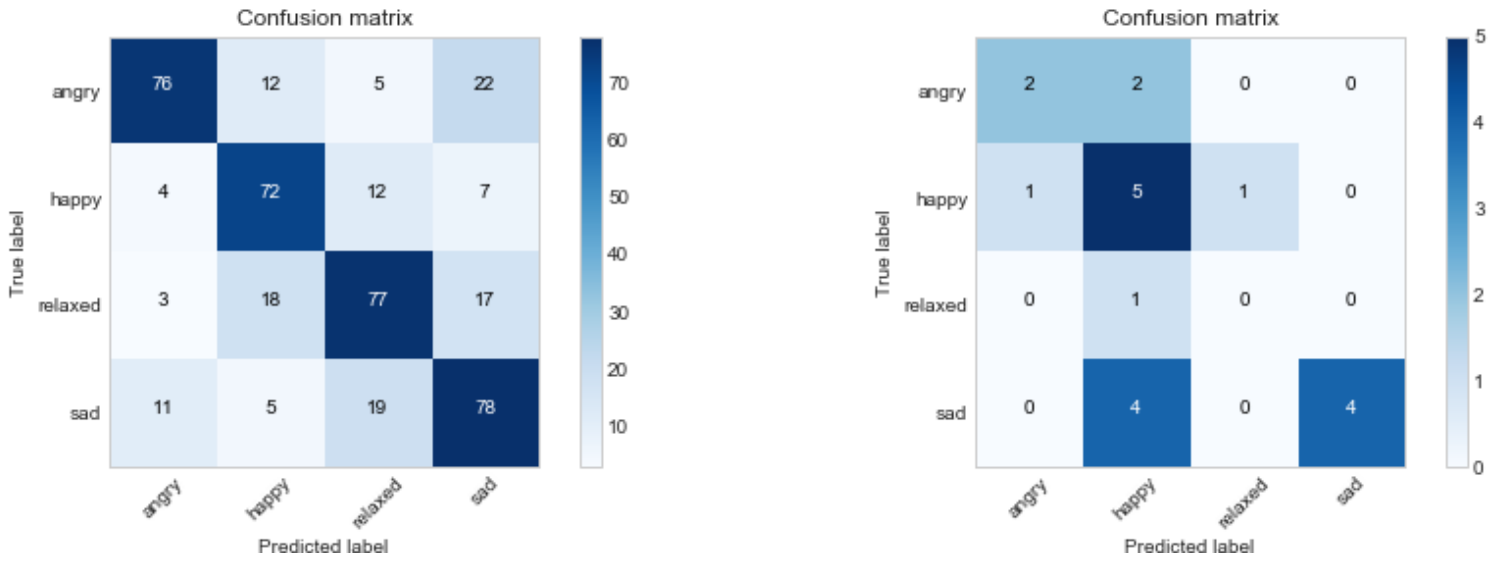


b: Extra Test set

Figure 5: Confusion Matrix obtained with Logistic regression

Accuracy on Test: 68 %
Accuracy on Extra Test: 45 %

SVM - Song Classification



a: Test set

b: Extra Test set

Figure 7: Confusion Matrix obtained with SVM

Accuracy on Test: 70 %

Accuracy on Extra Test: 55 %

Table of Contents

What's next?

Our models do not perform very good.
Using only MoodyLyrics4Q (instead of MoodyLyrics + MoodyLyrics4Q) is even worse.

- How can we improve it?