Emotions in Music: A Comparative Sentiment Analysis of Popular Songs Across Countries and Markets

Authors

Matteo Del Prato <matteodp@kth.se>

Alessandro Amandonico <aleama@kth.se>

Allocation of responsibilities

Both Matteo Del Prato and Alessandro Amandonico are responsible for data collection at the beginning of the research phase and afterwards will be responsible for performing at least 2 sentiment analyses related to countries' top charts.

Alessandro Amandonico is responsible for integrating data collection and sentiment analysis of all the other selected countries, while Matteo Del Prato is responsible for the analysis of the final results and their comparison. Both authors will work together in finalising the presentation and report before the first draft and presentation submission.

Both of the authors will contribute to fixing the report based on feedback before sending in the final work to the opponents, undergoing peer-to-peer review and giving the final presentation.

Organization

The project will be organised as a two-person project, building upon previously published work. Once the data collection is ready, sentiment analysis will follow with data analysis.

Background

The global music industry has seen an increasing growth in digital platforms such as Spotify or Apple Music, allowing for an unprecedented level of cross-cultural exchange also on the musical aspect. However, while the availability of music has become more global, the emotional content and sentiments expressed in popular music may still be deeply rooted in cultural, social, and regional contexts.

This study aims to investigate how sentiments expressed in song lyrics differ based on the country and the historical moment, and what could potentially be successful in that specific market.

By applying sentiment analysis to song lyrics, we aim to uncover regional patterns in what people like to listen to. For instance, a country experiencing political unrest may produce more gloomy and reflective music, while a nation with a successful economy might lean towards more uplifting and cheerful melodies.

Sentiment analysis is a natural language processing (NLP) technique that quantifies the emotional tone of text and its most common applications are in fields such as social media monitoring, customer feedback analysis, and market research [1]. However, we decided to use it for analysing song lyrics, a different area of action, that could possibly lead to interesting results also on how machine learning applies to more creative texts.

Problem statement

The project will investigate how different countries might have different music markets, based on the sentiment of the top-charted

songs. It will also investigate if Sentiment Analysis is effective in analysing song lyrics.

Problem

Cultural differences in the musical market were never exploited under the light of sentiment analysis on lyrics. Also, standard sentiment analysis techniques are not commonly used to interpret the sentiment in song, the project will investigate this application.

Hypothesis

This study is likely to generate three hypotheses. In the first instance, the lyrics of popular songs in countries experiencing political unrest or economic difficulties will tend to have a higher presence of negative sentiments than those of a country with stable socioeconomic conditions. Secondly, the overall feelings depicted in the lyrics of national mainstream songs will follow a certain trend, reflecting the population's feelings. Finally, songs that express more sentiment and emotions are more likely to gain popularity over those that are more neutral and do not connect on an emotional level.

Purpose

The purpose is to investigate how different countries have different music markets, and how effective can Sentiment Analysis be in analysing song lyrics and revealing those markets.

Goal(s)

A Sentiment Analysis model - which will be focused specifically on song lyrics - will be built and applied to an extensive dataset of popular songs from various countries. The results will be used to gain insights of population taste in those nations, which will be

investigated in relation to the social and political contexts. The outcomes will validate or invalidate our hypotheses.

Tasks

The data collection will start by extracting top-charting music from the previously selected countries of different regions. It will be done through Spotify's API which makes weekly charts available for many different nations around the world. Successively, the lyrics of those songs would be scraped and collected from Genius, Musixmatch and other lyrics databases.

After collecting everything, the lyrics data has to be 'cleaned' by removing noise such as punctuation or special characters. The text data will be then processed by a Python script, using some useful libraries such as Vader and TextBlob. Sentiment analysis algorithms will quantify the emotional tone of the lyrics and will be clustered and classified within the global picture.

Sentiment scores of different countries will be compared, and trends will be identified across musical genres. Finally, the results will be analysed through different lenses and factors (social, cultural, economic, etc.).

The number of countries will be approximately 10 to 20, and the number of song lyrics analysed for each country will vary between 50 and 100. The choice of the countries will be based on location and the grouping will be done on the continent to which they belong.

Method

For this project, we will use a combination of data collection, sentiment analysis, and comparative statistical analysis. Here's a short overview of each method used:

1. Data Collection: We will gather lyrics data using Spotify's API to determine top-charting songs in each country and supplement this with lyrics sourced from Genius, Musixmatch, and similar databases. This will be done by leveraging a scraping Python script and app's APIs.
2. Sentiment Analysis: We will utilise tools like VADER or TextBlob, which are well-suited for analysing social media text and song lyrics. We still have to decide which kind of sentiment analysis we are going to do and which works best in this context [2].
3. Comparative Analysis: After generating sentiment scores for songs, we will use statistical methods (e.g., ANOVA, t-tests) to identify significant differences between countries and continents.

Milestone chart (time schedule)

The project will start on the 19th of September and will end with the course deadline in December. There will be the following milestones and deliverables:

19th September: Start of the research

29th September: Data collection for top songs and correspondent lyrics of at least 4 countries through scraping scripts made on Python.

7th October: Use sentiment analysis for at least 4 different countries' top charts, results can be raw.

8th October: First draft of the research plan

3rd November: Integrate data collection and sentiment analysis of all the other selected countries.

17th November: Analyse the results and start comparing.

24th November: Finalise presentation and report.

25th November: First draft report and presentation

18th December: Fixes on the report based on feedback, improving results and format if possible.

19th December: Send the report to the opponents

7th January: Peer review of opponents

10th January: Final presentation

References

[1] Wankhade, M., Rao, A.C.S. & Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. *Artif Intell Rev* **55**, 5731–5780 (2022). Available: https://doi.org/10.1007/s10462-022-10144-1

[2] Altexsoft website, *Sentiment analysis types and tools*. Link: https://www.altexsoft.com/blog/sentiment-analysis-types-tools-and-use-cases/