Exploring Large Scale 20th Century Music Features Historical Patterns & Discoveries

Auguste BYIRINGIRO, Mathieu LAPEYRE, Flore MAGNARD, Gireg MIOREC DE KERDANET
abyiringiro3@gatech.edu (GT) (GT) (GT) (GT)

Summary

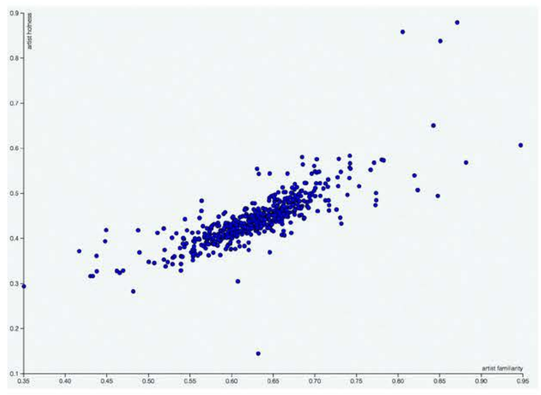
Do music tendencies reflect historical events? We combined machine-learning, visualization and user interaction to explore musical features evolution across time. We performed a large-scale data extraction from a music dataset covering the 20th century.
Our approach was unusual in that our primary goal was understandability: unlike many existing research projects dealing with music history, we emphasized an user-friendly and comprehensive approach.

The Million Song Dataset

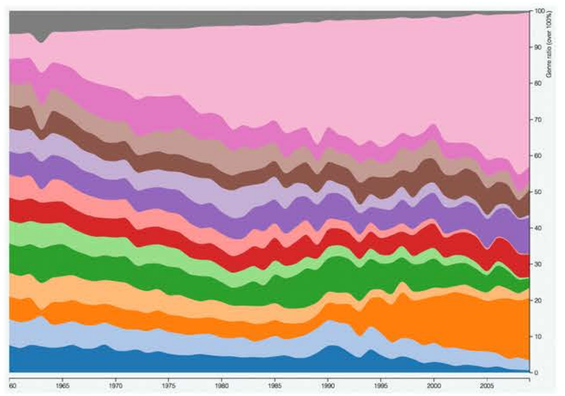
The dataset was created from the Echo Nest API. It gathers:
1,000,000 tracks
54 fields per track
273 GB of data
44,745 unique artists
7,643 tags
2,321 unique musicbrainz tags
43,943 artists with at least one term
2,201,916 asymmetric similarity relationships
515,576 dated tracks starting from 1922
18,196 cover songs identified
237,662 tracks with lyrics

Machine learning algorithms

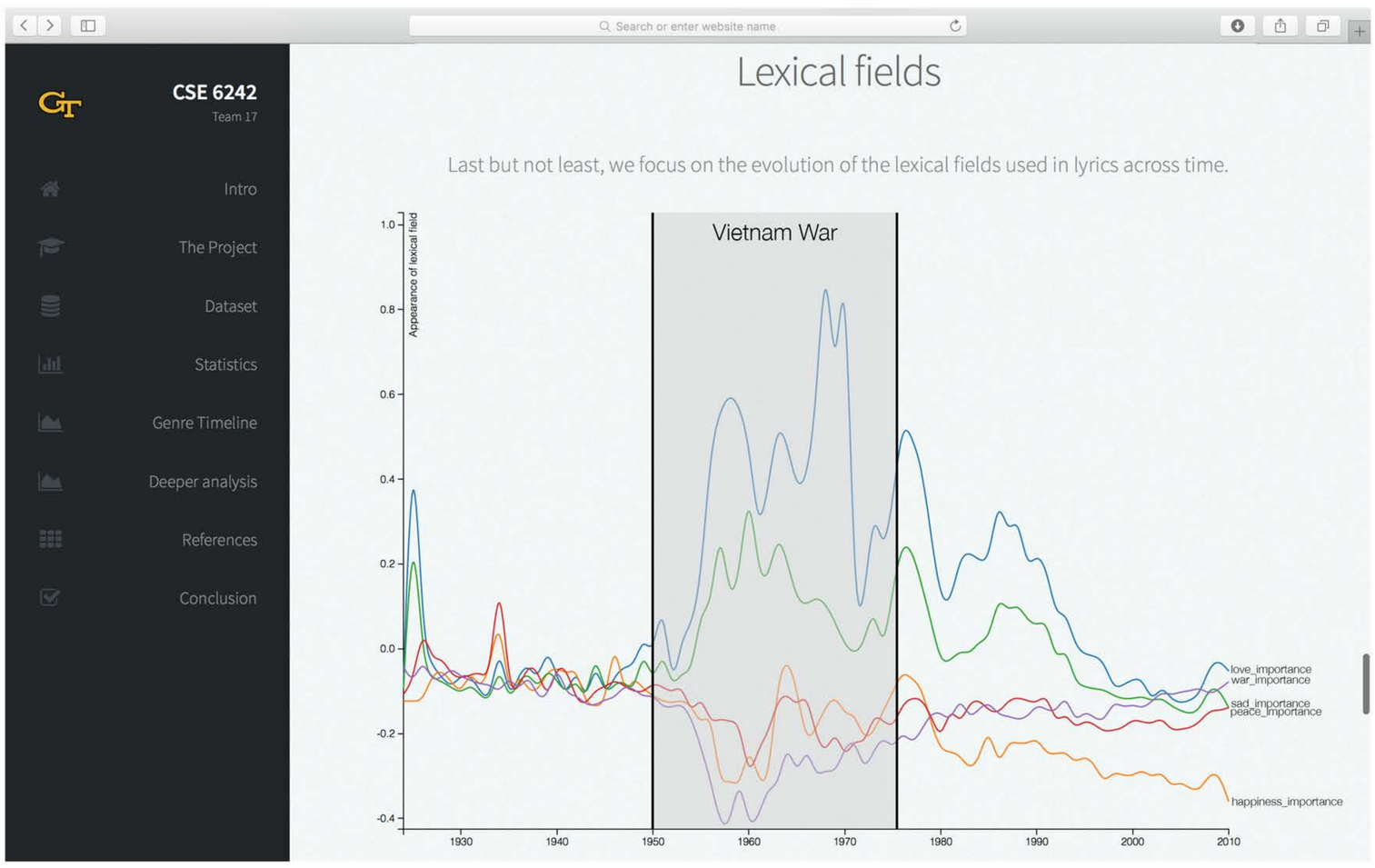
We used natural language processing techniques to infer the amount of particularly relevant emotions (anger, sadness,...) and thematics (peace, war,...) in lyrics that we quantified with scores.
The idea was then to look at the evolution of those scores which should be correlated to historical events since artists get inspired by the context in which they live in.
After mapping available genres with the major emotions they are associated with, some Machine Learning algorithms (such as Natural Language Processing) were used to link emotions to the musics for which genre is missing. This will give us additional insights.
We also developped supervised algorithms to augment our database by inferring missing years for instance.



We combine cluster analysis and visualization over our million data to characterize the pattern between two features.



We analyzed the evolution of the main musical genres across time. From 1960 to 2010. Rock music is getting more and more influence.



Mine and Clean the Data

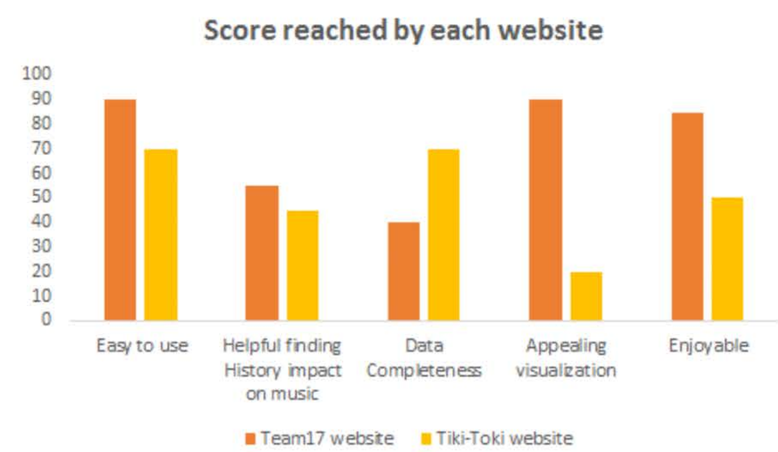
The Million Song Dataset contained many features but remained incomplete for our project. We wanted to study the evolution of mood, genre and thematic by performing lyrics analysis which were not necessarily included in the original dataset. Thus, we needed to join the musiXmatch dataset to get the lyrics and we also added some classification tags from the lastfm dataset.
Moreover, we only got the year of a given track and not the complete date. In order to get more accurate results (additional or more precise date) we decided to extract the release date of each track from the Spotify API and combine this data to our original dataset.

Interact with the Data

As the native Million Song Dataset was hardly eploitable, our main concern was to help the user to interact easily with the data. We designed a web-interface where we displayed our results and let the user manipulate the information. The web interface is essentialy based on a straightforward scroll design (HTML5) and interactive graph (D3).
We wanted to tell a story and invite the user to discover History through Music. In order to do this, we pre-processed the data and fed the website with a significatively reduced amount of information directly from AWS S3.

Results and Future Work

We extracted, filtered and quantified information about the evolution of music features across the years.
We compared our results with groundtruth historical data obtained from «the Top25 events that changed America» according to Time Magazine.
While we found some relevant correlation between major event (such as wars), we can still make a lot of improvements when processing the raw data from a scarce dataset.
We also did a 20-peer survey to compare our approach with another existing project dealing with the analyse of music throughout the years. We successfully managed to draw the user's attention to our strategy in providing useful knowledge without sacrificing the understandibility of the subject.



References

[1] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The Million Song Dataset. In Proceedings of the 12th International Society for Music Information Retrieval Conference, 2011.
[2] C.Laurier "Automatic Classification of Musical Mood by Content-Based Analysis", Dept. of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, 2011
[3] Acerbi A, Lampos V, Garnett P, Bentley RA "The Expression of Emotions in 20th Century Books" PLoS ONE, 2013
[4] Mauch M, et al., "The evolution of popular music: USA 1960–2010" R. Soc. open sci. 2, 2015
[5] C. Weihs, U. Ligges, F. Morchen, and D. Mullensiefen, "Classification in music research," Adv. Data Anal. Classificat., 2007.
[6] J. Shen et al., Towards effective content-based music retrieval with multiple acoustic feature combination, IEEE Trans. Multimedia, 2006.
[7] Joan Serra et al., "Measuring the Evolution of Contemporary Western Music", Scienctific Reports 2012
[9] E. Glenn Schellenberg et Christian von Scheve "Emotional Cues in American Popular Music : Five Decades of the Top 40", in Psychology of Aesthetics, Creativity, and the Arts, 2012

Georgia Institute of Technology