Song Retrieval based on similarity

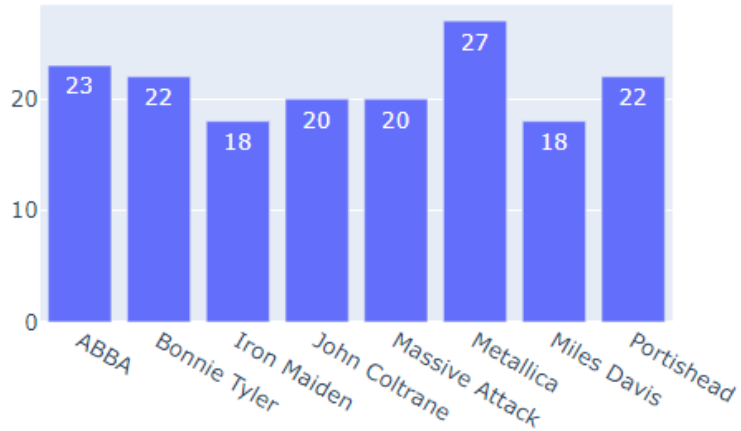
Iraklis Evangelinos, Vaggelis Lamprou

Multimodal Machine Learning

Contents

1.Data

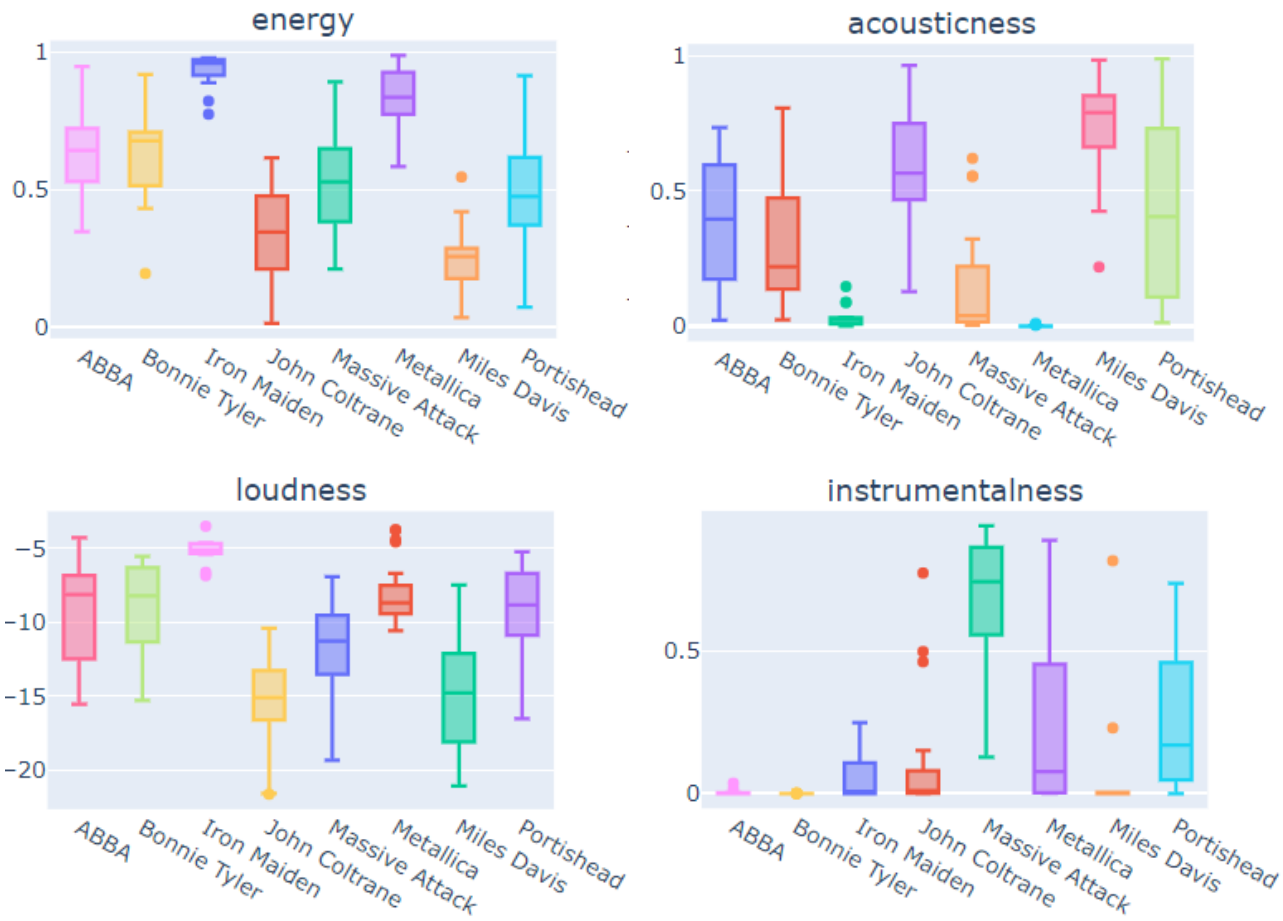
- 170 songs collected by 8 different artists
- Metallica(Metal), Iron Maiden(Metal), Miles Davis(Jazz), John Coltrane(Jazz), Massive Attack(Alternative), Portishead(Alternative), ABBA(Pop), Bonnie Tyler(Pop)



- Goal: Build a song retrieval model based on the song features

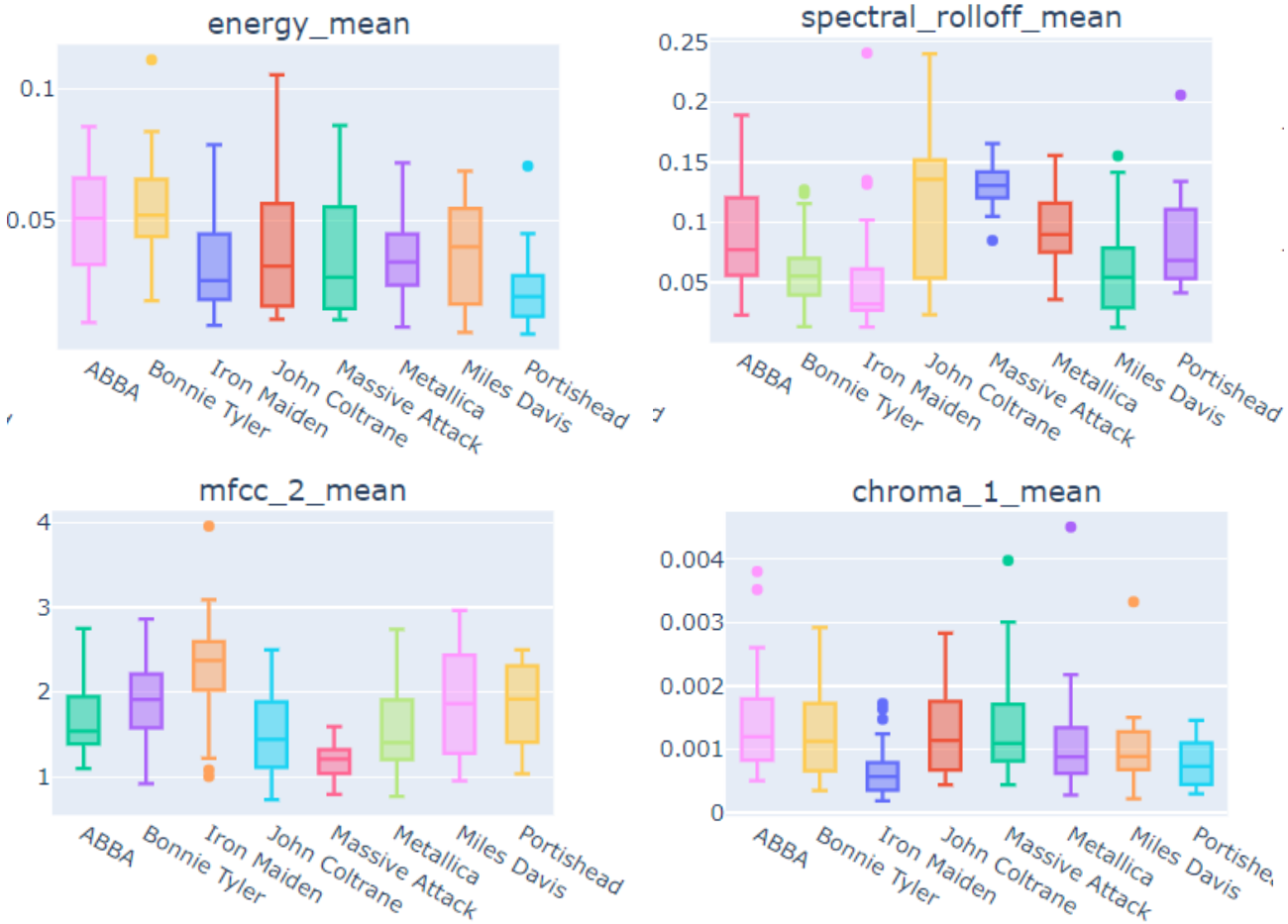
2.Spotify features

- 10 features per song
- Danceability, Energy, Loudness, Mode, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo



3.PyAudioAnalysis features

- Mid-term features (136 dim vector per song)
- mid window = mid step = 5 secs,
  short window = short step = 1 sec



4.Similarity metrics

- Based on the 146 dim vector, we have the following 4 metrics:
  - Euclidean distance: $\sqrt{|x_1 - y_1|^2 + ... + |x_{146} - y_{146}|^2}$
  - Manhattan distance: $|x_1 - y_1| + ... + |x_{146} - y_{146}|$
  - Cosine distance: $\frac{\sum x_i y_i}{\sqrt{|x_1|^2+...+|x_{146}|^2}\sqrt{|y_1|^2+...+|y_{146}|^2}}$
  - Chebyshev distance: $\max |x_i - y_i|$
- Based only on Spotify features, we have another metric.
  Hand picked features are used to make a recommendation based on the song's genre.
  - Metal songs: Energy and Acousticness features
  - Jazz songs: Energy, Loudness and Acousticness features
  - Alternative songs: Danceability, Instrumentalness and Valence features
  - Pop songs: Energy, Loudness and Instrumentalness features
- Based on the 146 dim vector, a PCA cosine metric is created.
  We apply PCA to the 146-dim vector, keeping the first 25 principal components and the we consider cosine similarity.

5.Evaluation

- Function that takes a database band and a metric as arguments and returns a

$$\text{band score} = \frac{\textit{total no. of correct similarities}}{\textit{total no. of predicted similarities}}$$

- Per band song it computes the top 5 similar songs.
- A song is considered similar if it belongs to the same genre.
- Database results:

	Euclidean	Manhattan	Cosine	Chebyshev	Spotify	PCA
Metallica (Metal)	0.47	0.69	0.77	0.40	0.93	0.93
Iron Maiden (Metal)	0.43	0.63	0.67	0.37	0.93	0.92
Miles Davis (Jazz)	0.50	0.58	0.48	0.48	0.38	0.92
John Coltrane (Jazz)	0.47	0.51	0.46	0.46	0.39	0.81
Massive Attack (Alternative)	0.43	0.43	0.37	0.43	0.65	0.76
Portishead (Alternative)	0.33	0.36	0.37	0.29	0.35	0.58
ABBA (Pop)	0.31	0.51	0.43	0.30	0.68	0.75
Bonnie Tyler (Pop)	0.38	0.63	0.50	0.34	0.63	0.62
Overall	0.41	0.55	0.51	0.38	0.63	0.78

6.Conclusion

- The hand crafted 'Spotify metric' performs very well, but it is not deployable on a real model for practical reasons and the results apply to the specific dataset.
- The final implementation of the model applies PCA to the feature vector and then the cosine metric on the resulting 25 dim vector.

Thank you !