MML 2016

9th International Workshop on Machine Learning and Music

Held in Conjunction with the

European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2016)

Riva del Garda, Italy

Proceedings

Rafael Ramirez
Universitat Pompeu Fabra, Spain

Darrell Conklin
University of the Basque Country, Spain

José Manuel Iñesta
University of Alicante, Spain

Contents

"Are we playing like Music-Stars?" Placing Emerging Artists on the Italian Music Scene

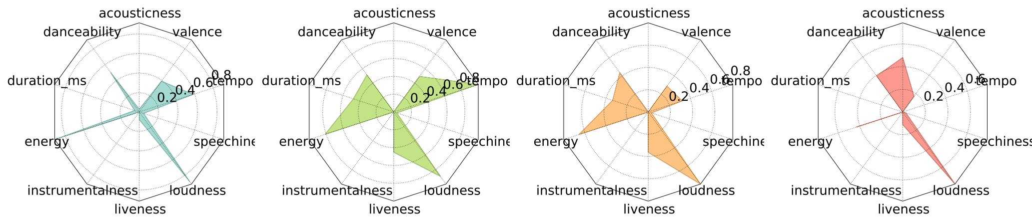
Laura Pollacci[1], Riccardo Guidotti[1,2], and Giulio Rossetti[1,2]

[1] KDDLab, University of Pisa, Largo B. Pontecorvo, 3, Pisa, Italy,
laura.pollacci@di.unipi.it,
[2] KDDLab, ISTI-CNR, Via G. Moruzzi, 1, Pisa, Italy,
riccardo.guidotti@isti.cnr.it
giulio.rossetti@isti.cnr.it

Abstract. The Italian emerging bands chase success on the footprint of popular artists by playing rhythmic danceable and happy songs. Our finding comes out from a study of the Italian music scene and how the new generation of musicians relate with the tradition of their country. By analyzing Spotify data we investigated the peculiarity of regional music and we placed emerging bands within the musical movements defined by already successful artists. The approach proposed and the results obtained are a first attempt to outline rules suggesting the importance of those features needed to increase popularity in the Italian music scene.

1 Introduction

Nowadays, thanks to the maturity of online music related services, music consumption has become ubiquitous. During the last decade several works analyzed music data collected from online services to study users behaviors and tastes All those works highlight how the access to huge music libraries has made possible even for emerging artists to reach audiences unimaginable only few years ago. In this work we focus on Italian music scene: how new musicians relates with the musical tradition of their country. Taking advantages of data collected from Spotify we investigate the peculiarity of regional music and try to place emerging artists within the musical movement defined by the already famous colleagues of the same country. Music is one of the most valuable expressions of national identities, since it can reflect cultural and societal evolutions as well as national and international historical events. However, a whole country cannot be described by a single "cultural" entity: each region within it has its own peculiarity. Does this observation apply to the Italian music scene? Analyzing regional songs: will emergent artists adhere to the music canon identified for their region? If not which are their strongest influences? Being able to answer such questions will act a linchpin for shedding lights on how the new generations defines themselves, on which are the major changes in music tastes that are currently taking place. Moreover, extending the analysis on the national profiles we can better understand how emergent artists place themselves in the music scene: their music style



(a) 100band (b) Tuscany (c) Emilia-Romagna (d) Molise

Fig. 1. Artists Profiles: (a) the `100Band` medoid, (b-d) profiles of the famous artists from Tuscany, Emilia Romagna and Molise.

mimics that of famous artists so as to sound more "appealing" to the public or do they presents a totally personal style disregarding the current tastes?

2 Dataset

In order to carry out our analysis we used two datasets: the first relating to famous Italian musicians, the second to emerging youth bands in Tuscany. The Italian dataset, called `IT`, consists of 502,582 songs composed by 2,379 top Italian authors. Conversely, the `100band` dataset is composed by 513 emerging Tuscany artists[3] and 24,147 songs. Both datasets were built using the Spotify API[4] and are composed by all the songs present on the platform for the selected artists. Each song is identified by its title, artist and described by a set of ten musical features[5]: `acousticnes`, `danceability`, `duration`, `energy`, `instrumentalness`, `liveness`, `loudness`, `speechiness`, `tempo` and `valence`. All the features range in $[0, 1]$ with the exception of `duration`, `loudness` and `tempo`: however, during the preprocessing phase we normalized the latter in order to align all the feature scales. Moreover, authors present in the `IT` dataset are enriched by the information regarding their Italian region of origin.

3 Experiments

As a first step, we grouped songs by artist and for each of them we extracted a profile: we chose to describe each artist with his medoid song, i.e. his most representative track identified by minimizing the sum of the Euclidean distances

[3] The emerging artists were selected among the ones that participate to the `100band` contest promoted by "Tuscan Region" and "Controradio" in 2015http://toscana100band.it/

[4] https://developer.spotify.com/

[5] A detailed description of the features can be found at https://goo.gl/my3sVg

between the Spotify features among all its discography. Once profiled both famous and emerging artists we focused our attention on similarities among them, both at regional and nation wide levels.

Regional profiles. Indeed, music is one of the most valuable expressions of national identities. However, it is rare that a country can be effectively described as a single "cultural" entity: conversely, it is natural to expect that each region expresses its own peculiarity. Does this observation apply to the Italian music scene? Can each Italian region be characterized by the music it produces? In order to answer such questions, moving from the computed artist profiles, we leveraged the geographical informations attached to famous artists in `IT` to build region wide profiles. Figure 1 shows the profiles obtained for Tuscany, Emilia Romagna and Molise (the remaining regions were omitted due to lack of space). Tuscany and Emilia Romagna are both characterized by cheerful and dance songs with a fast pace and instrumental beat. However, Emilia Romagna is represented by more energetic and fast tracks, like the songs of Gem Boy. Molise, instead, is represented by melodic music with a calm and repetitive rhythm, like the songs of Fred Bongusto. Our results suggest that each Italian region has its own music peculiarities: what about emerging artists in `100band`? In Figure 1(a) is shown the dataset wide profile emerging Tuscan bands: we can easily notice that such profile differs from the one expressed by famous artists of the same region. Given such distance, to which regional musical tradition Tuscan emerging artists look up to?

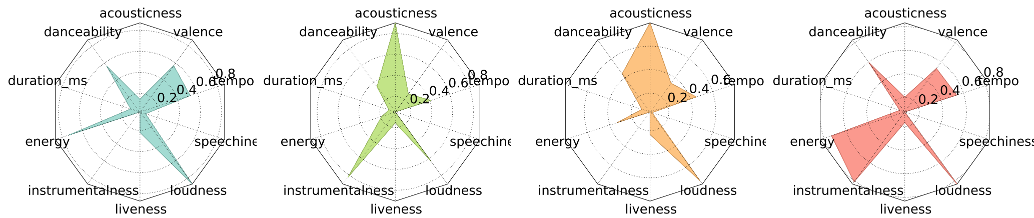
We assigned each of the bands in `100band` to its most similar regional classes by applying K-NN with $k = 1$ by using the Euclidean distance. Table 1(left) reveals that most of the emerging bands are assimilated to Emilia Romagna (46%) followed by Molise and Valle d'Aosta: those are the musical scenes that better capture the features expressed by Tuscan emerging artists. The rest of the bands is divided into low percentages among other regions. It must be emphasized that regional profiles are based on the artist's birth region. Notice that our data do not allow tracking artists' mobility, which could affect the production and on some features of the music.

National profiles. In order to give a more comprehensive classification of the Italian music scene, we employed the K-means clustering algorithm to identify homogeneous clusters among the artists in `IT`. After calculating the Sum of

Region	% Artists
Emilia Romagna	0.46
Molise	0.13
Valle d'Aosta	0.11
Piemonte	0.05

Region	% Artists	Avg. Popularity
Cluster 0	0.61	22/100
Cluster 1	0.20	15/100
Cluster 2	0.13	19/100
Cluster 3	0.05	16/100

Table 1. Similarity among Tuscan emerging bands and the profiles of famous Italian artists at regional level *(left)* and with respect to data-driven national profiles *(right)*.



(a) Cluster 0 (45%) (b) Cluster 1 (14%) (c) Cluster 2 (29%) (d) Cluster 3 (12%)

Fig. 2. Artists Profiles: Italian clusters medoids.

Squared Error (SSE) distribution for $2 < k < 20$ we selected k=4 because for k>4 the SSE does not decrease significantly any more. Thus from each cluster we extracted a medoid resulting in four profiles. Observing the profiles, reported in Figure 2 we immediately perceive that `cluster0` and `cluster3` are expression of artists in direct opposition to the ones respectively in `cluster1` and `cluster2`:

- `cluster0` and `cluster3` are represented by happy songs, dance music and songs with a strong beat, though the songs of `cluster0` have regular rhythm: `cluster0` has the highest value `speachness` and includes artists such Datura, Linea77 and rappers (e.g. Emis Killa), while in `cluster3` there are Working Vibes and Persian Jones.
- `cluster1` and `cluster2`, conversely, represent most melodic songs, less rhythmic and differs only in `instrumentalness`, which is higher in `cluster1`. The `cluster2`, which has higher `instrumentalness` value, includes artists such Giovanni Allevi and Stefano Bollani, while `cluster2` includes Lucio Battisti and Carla Bruni.

Outlining the regional analysis, we assigned each of the `100bands` to the closest cluster using K-NN fixing as known groups the medoids of the clusters identified by K-means. More than half of the emerging bands (61%) are labeled similar to the `cluster0` profile, a cluster containing famous artists of Emilia Romagna and Tuscany. The rest of the bands is prevalently allocated in `cluster1` and `cluster2`; only a small part in `cluster3` (see Table 1(right)).

Popularity. How do emerging artists decide their musical style? Do they follow their passion or do they try to mimic what is perceived as the "people taste"? In order to provide answers to such questions, we calculated the average popularity of the famous Italian artists for each cluster - as perceived by the Spotify users - and we use this indicators to label the identified groups. As shown in Table 1(right), we observed that there are no strong differences among the clusters means (the same observation holds for their median and standard deviation). However, it needs to be underlined that the popularity perception is computed on a service, Spotify, that connects world-wide artists and listeners: in this scenario the popularity of Italian artists are somehow penalized due to the intrinsic restriction imposed by the non-adoption of english as main lyric language. Taking into account such distortion, we can anyhow observe how emerging bands tends to be assimilated to `cluster0` and `cluster2` (almost 74% of `100band`), clusters grouping together artists having greater average popularity. Somehow, emerging bands seem to mimic characteristics of known artists that receive more attention from the public: indeed, to reach success first you must learn to play by the rules, then you must forget the rules and play from your heart.

4 Conclusion

With our work we seek to identify regional, national and Tuscan emerging bands identities through music analysis. It seems clear that it is not easy to define general national and regional profiles of music exhaustively in order to describe all the national scene because of the strong heterogeneity of music. However there are some features discriminating between the various groups identified during the analysis phase. The results of the experiments that we have carried out leave open options for future developments in different directions. On one hand, it could be an interesting task to perform a a finer grain analysis on texts in order to investigate other discriminating features (e.g. linguistic) that can help to describe songs sociologically and emotionally, not only in a strictly musical way. On the other hand, it allows to rise to a higher step for a comparison at the international level. This analysis can be conducted both in Europe and globally in order to investigate the importance of music in the identification of cultures and traditions of the populations.

Acknowledgements

This work is supported by the European Communitys H2020 Program under the scheme "INFRAIA-1-2014-2015: Research Infrastructures", grant agreement #654024 "SoBigData: Social Mining & Big Data Ecosystem" http://www.sobigdata.eu.

References

1. K. Bischoff. We love rock 'n' roll: analyzing and predicting friendship links in last.fm. In *Web Science 2012, WebSci '12, Evanston, IL, USA - June 22 - 24, 2012*, pages 47–56, 2012.
2. D. Herremans, D. Martens, and K. Sörensen. Dance hit song prediction. *Journal of New Music Research*, 43(3):291–302, 2014.
3. D. Pennacchioli, G. Rossetti, L. Pappalardo, D. Pedreschi, F. Giannotti, and M. Coscia. The three dimensions of social prominence. In *Social Informatics*, pages 319–332. Springer, 2013.
4. J. Putzke, K. Fischbach, D. Schoder, and P. A. Gloor. Cross-cultural gender differences in the adoption and usage of social media platforms - an exploratory study of last.fm. *Computer Networks*, 75:519–530, 2014.