DATA SCIENCE FOR THE DETECTION OF EMERGING MUSIC STYLES

Holly Silk
University of Bristol
hollysilk@gmail.com

Raul Santos-Rodriguez
University of Bristol
raulsantosrodriguez
@gmail.com

Matt McVicar
University of Bristol
Jukedeck Limited
mattjamesmcvicar@gmail.com

Tijl De Bie
University of Bristol
tijl.debie@gmail.com

ABSTRACT

We demomstrate a music browsing service which allows users to discover emerging music styles in the United Kingdom. Postulating that an emerging music style is a product of four important factors (artists, genres, locations and fans), we give a brief overview of over 1 year's worth of data collection, and introduce an online browsing system allows users to browse artists, genres, fans and locations.

1. INTRODUCTION

1.1 The long tail and the paradox of choice

When music was still sold on physical carriers such as CDs or LPs, music outlets needed to carefully fill their limited shelf space with the items most likely to bring in the most income. Assuming equal cost and physical footprint for popular and niche artists, in expectation this will tend towards the most popular titles. Economists have theorised that this lead to a business model known as the "Blockbuster strategy" [3], which involves heavy investment and promotion of a few select products.

There was some hope that this would change in today's digital economy with minimal overheads for the artists (modest recording and reproduction costs), retailers (no limits on shelf space, cheap promotion and distribution), and consumers (practically unlimited choice). The expectation was that retail patterns would shift to a business model of selling "less of more", taking the focus away from the elite few and allowing smaller, less well-known artists to prosper. This is known as the theory of the Long Tail [1]: while some artists still get the lions share of the revenue, the tail of less popular music would lengthen and fatten.

Surprisingly, the opposite was found to be true: the tail has become even skinnier, with an even smaller proportion of artists able to make a living from their music. Research by The Harvard business review in 2008 found that 1% of artists account for 32% of total plays on the online radio station Rhapsody, with 10% making up 78% of plays [3].

Similar figures have been quoted PRS music for both illegal peer-to-peer network sharing services and legal downloads, finding for example that 75% of the music stocked by online stores did not find a single buyer [2].

A well-known explanation for this is given in the book "The Paradox of Choice" [7] where having too many options tends to be paralysing instead of liberating. Applied to the music market: as searching for new interesting music comes at a cost to consumers (at least an opportunity cost), they will often play it safe to avoid disappointment: they will either listen to the same artists over and over again, or at best will try what is recommended to them by trusted parties (friends, or automatic systems that recommend music liked by people similar to you). As a result, the rich get richer, and revenue concentrates on the popular few.

This makes it increasingly hard for new music trends to gain a foothold in the music industry. Even if a pioneering artist's music has a genuine potential of ultimately appealing to large consumer groups, there is only a small chance that it will ever emerge from the skinny tail of popular music. As a result, creative innovation in popular music is stymied, and new emergent music styles disappear before becoming sustainable.

Thus the following question begs an answer: is it possible to detect emergent music styles at an early stage, in a scalable (and thus automated) way, characterising it in terms of its innovative audio features, demographics of the fan base, and their geographical location. Today, for the first time, all the stars necessary for realising this are aligned. We have access on a large scale to the audio of a number of bands of the order of a million (e.g. on SoundCloud), and we have access to their fan base and their properties through social media (e.g. Twitter). The subject of the Data Science for the Detection of Emerging Music Styles (DS4DEMS) is to gather this data, and to develop the data mining techniques needed to discover new emerging music styles at a very early stage.

This paper in particular, highlights some of our initial findings of the project, following on from our previous work [4–6], and introduces a website facilitating the exploration of the UK independent music scene. The website is available publicly at `http://ds4dems.enm.bris.ac.uk/`

1.2 Characterisation of new music genres

When new music genres become popular (mainstream), their most salient characteristic is their sound. Dubstep,

for example, is a subgenre of electronic dance music which features sparse, syncopated drum rhythms and sub bass "wobble" basslines. However, music genres also are a product of their environment and surroundings: Motown is uniquely tied to Detroit, Michigan. Similarly the "Bristol Sound" was highly localised to the South West of England in the early 1990s. Furthermore, these genres would never have gained enough traction to attain mainstream success if it were not for fans, who attend gigs and support the artists via word-of-mouth promotion. Significant cultures of fans from existing genres include the anti-establishment punk movement and the hip-hop subculture [1].

For these reasons, we believe that an understanding of Artists, Genres, Locations and Fans are all necessary for studying the emergence of new music styles. These aspects therefore form the core pillars of our research and the current paper. The remainder of this paper is dedicated to explaining the website back-end (Section 2), front-end (Section3), and discussing how we plan to extend this work (Section 4).

2. BACKEND

Data for this project comes from over 50,000 artists listed on `https://www.reverbnation.com/`, a site where artists set an online profile with links to all their social media (SoundCloud, Facebook etc). In general, artists on this site are either unsigned or belong to a small record label.

From this collection of artists, we collected links to Twitter and SoundCloud accounts. Using the twitter API [2], we then began tracking all tweets which mentioned at least one artist in our database. Given the rate limits of the API, care was taken to fetch the most recent tweets of each artist uniformly, so that we did not exhaust our entire limit on a few more popular artists. To date we have over 10M tweets mentioning over 50k artists. A subset of these tweets we found to be geotagged. We mapped these to the administrative regions of the UK to give an idea of the twitter music activity in each region.

We also tracked the number of SoundCloud plays each artist received and their genres, and periodically download all available SoundCloud tracks for every artist we know about. All of the above is stored in a MongoDB database. Our recent publications [4–6] highlight some of the interesting patterns found in these data, including the fact that there exists an arbitrage of particular music genres in neighbouring regions of the UK. However, despite this abundance of data, we did not have a convenient way to simply browse the data. This prompted us to design a website front-end to the database, outlined below.

3. FRONTEND

The design of our site focuses on what we believe to be the four most important attributes of an emerging music style: artists, genres, locations, and fans. We were particularly interested in seeing the links between these attributes, with a typical search path being a user becoming interested in an artist, then wanting to explore the associated genre, where this genre was most popular, and where people were tweeting about this genre. Our tab functionality intuitively allows this kind of browsing.

We have also implemented a radio player, which pulls in audio data from our SoundCloud database and plays music from a particular genre/region. This we anticipate will allow music fans and record labels to easily trial out new artists and styles whislt browsing our site.

4. CONCLUSIONS AND FUTURE WORK

In this demo paper, we outlined the need for a new method of exploring independent music artists on the web. We then described the data we collected, and some of our initial findings. We presented a website which allowed intuitive exploration of this vast amount of data, focusing on the four key features of artists, genres, locations, and fans.

Looking forward, we wish to incorporate more advanced data mining methodologies into the site. For example, we would like to automatically discover communities within the Twitter network of fans, or deploy novelty detection algorithms on the audio data we have collected, hoping that it will allow us to discover emerging music styles.

5. REFERENCES

[1] Òscar Celma Herrada. Music recommendation and discovery in the long tail. 2009.

[2] A Chris. The long tail: Why the future of business is selling less of more, 2006.

[3] A. Elberse. Should you invest in the long tail? *Harvard business review*, 86(7/8):88, 2008.

[4] M. McVicar, C. Mesnage, J. Lijffijt, and T. De Bie. Interactively exploring supply and demand in the uk independent music scene. In *European Conference on Machine Learning, Principles and Practices of Knowledge Discovery in Databases*, pages 289–292. Springer, 2015.

[5] M. McVicar, C. Mesnage, J. Lijffijt, E. Spyropoulou, and T. De Bie. Supply and demand of independent uk music artists on the web. In *Web Science*, 2015.

[6] C. Mesnage, R. Santos-Rodriguez, M. McVicar, and T. De Bie. Trend extraction on twitter time series for music discovery. In *Workshop on Machine Learning for Music Discovery, 32nd International Conference on Machine Learning*, 2015.

[7] Barry Schwartz. The paradox of choice. Ecco New York, 2004.

---

[1] `https://en.wikipedia.org/wiki/{Dubstep, Motown, Bristol_underground_scene, Punk_subculture, Hip_hop}`, all accessed 15th July 2016.

[2] `https://dev.twitter.com/overview/documentation`