Trend extraction on Twitter time series for music discovery

Cédric Mesnage CEDRIC.MESNAGE@BRISTOL.AC.UK
Raúl Santos-Rodríguez ENRSR@BRISTOL.AC.UK
Matt McVicar MATT.MCVICAR@BRISTOL.AC.UK
Tijl De Bie TIJL.DEBIE@BRISTOL.AC.UK
Department of Engineering Mathematics, University of Bristol, UK

Abstract

To assist music consumers in the discovery of artists in the long tail, we present a simple music discovery web application that ranks artists based on the time series of tweets addressed to these artists. More specifically, the web application presents the user with a list of tracks of artists for which the trend in number of tweets is rising fast, allowing the user to bias the results towards popular or emerging artists.

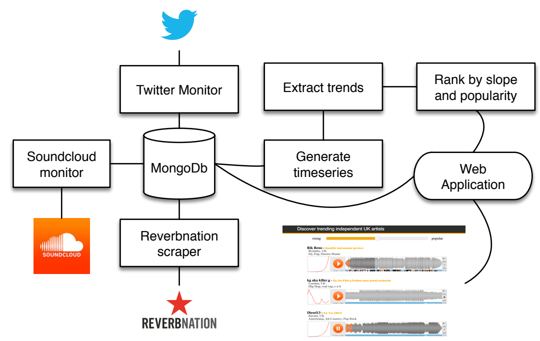

Figure 1. Diagram of the discovery system.

1. Introduction

Music enthusiasts are constantly searching for new means to discover music that they like, outside the commercial charts. Strategies to facilitate this music discovery process range from those based on an in-depth audio analysis, over those based on meta-data to those relying on an analysis of the attention music artists receive on social media. Specifically, previous attempts of the latter kind have been based on quantifying a time-specific popularity (Schedl, 2011) or on spotting influential users (Sha et al., 2012).

In this extended abstract, we present a system and web application to facilitate music discovery by exploiting social media data (namely Twitter time series). This system gathers information about independent UK artists, fetches all tweets about them, extracts the trend in number of tweets about the artists, and quantifies their rate of rise or decline in popularity. This system presents a first step in an ongoing project on the detection of emerging music styles.

Figure 1 shows the architecture of the system. It involves monitoring platforms such as Twitter, Soundcloud and Reverbnation as well as storing the data in a NoSQL database, generating timeseries from the tweets, extracting trends for each artist, ranking them according to the rate of rise or decline of the trend, and providing them to the end user through a Web application.

The remainder of this paper is as follows, Section 2 describes the main components of the data collection pipeline, Section 3 summarizes the trend extraction method, Section 4 gives the ranking strategy and Section 5 presents the Web application.

2. Data Pipeline

We are tracking 54,192 UK independent artists that we found on the social platform Reverbnation.com. We fetched the artist names, genres, geographic information and links to their Twitter, Souncloud or Youtube accounts. Using the Twitter handles of the artists who specified one (21,616 of them) we periodically gather, since Dec 3rd 2014, tweets about them as well as profile data about the people posting the tweets about the artists. As of today our MongoDB database stores 10,005,640 tweets tweeted by 1,844,278 people. For the rest of the study we consider only artists which have been tweeted about at least a hundred times over the time period, leaving us with 3031

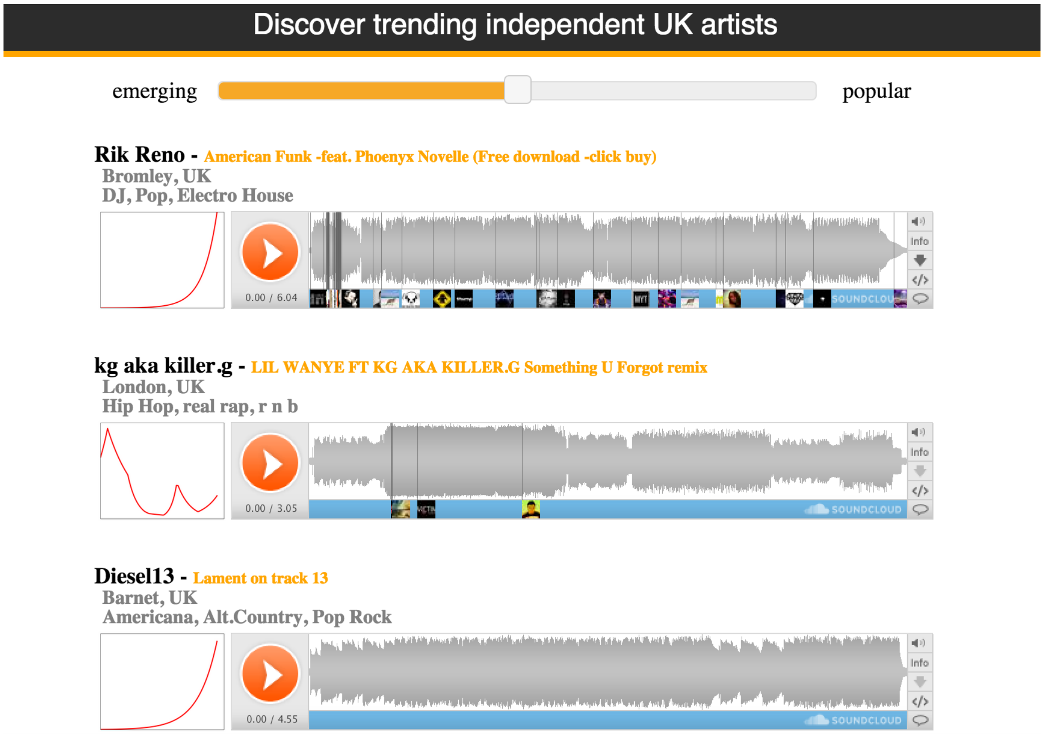

Figure 2. Screenshot of the Web application. Artists are ranked by the value of the slope of the trend. The slider at the top enables the user to adapt the selection of artists by favorising the ones that are lesser known or the ones that are more popular. Each artist is represented by its name, the title of the most popular track (by play count), its location, a list of genres, a plot of the social trend and a Soundcloud widget which the user can use to listen to the track.

artists. Of these, 711 have specified a Soundcloud account.

3. Trend Extraction

In time series analysis, trend extraction involves detecting the long term behavior, filtering out non-persistent fluctuations. Several general-purpose approaches for trend extraction are available (see Kim et al. (2009) and the references therein). However, the nature of our time series, namely the number of tweets per day about an artist, calls for a dedicated approach. Indeed, each value in such a time series can be modelled as a Poisson random variable, the rate of which is a combination of a trend effect, sparse peaks of activity that can be positive only (often related to an event), and possibly periodic fluctuations (e.g. due to seasonal or day-of-week effects). Recently, De Bie et al. (2015) developed a probabilistic model that is tailored to this scenario, and it is this method that is used in the presented system.

4. Ranking Strategy

We rank artists by slope of the trend and overall popularity. Given that these values range over several orders of magnitude we work in the log space. We refer to s as the slope of the logarithm of the trend at the last point of the time series, and to z_s as the z-score of s. The popularity z_p is the z-score of the logarithm of the total number of tweets. The user controls the score of the artists via a parameter α between -1 and 1 on the ranking given as: score $= z_p * \alpha + z_s$.

5. Web Application

The time series, trends and rankings are computed daily and therefore the displayed list of tracks will change with time. Figure 2 shows a screenshot of the Web application[1].

Acknowledgments This work was funded by the EPSRC project DS4DEMS (EP/M000060/1).

References

De Bie, Lijffijt, Mesnage, and Santos-Rodríguez. Detecting trends in twitter time series. *Under review*, 2015.

Kim, Koh, Boyd, and Gorinevsky. *l*1 trend filtering. *SIAM*

[1] http://www.ds4dems.net/discover-independent-uk-artists/

review, 51(2):339–360, 2009.

Schedl. Analyzing the potential of microblogs for spatio-temporal popularity estimation of music artists. In *IJCAI: International workshop on social web mining*, 2011.

Sha, Quercia, Michiardi, and Dell'Amico. Spotting trends: the wisdom of the few. In *RecSys*. ACM, 2012.