Auditing The Echonest – Investigations to gauge output accuracy
Submission for *Auditing Algorithms From the Outside* at ICWSM-15

Justin Gagen
Transforming Musicology
Goldsmiths, University of London
j.gagen@gold.ac.uk
March 11th 2015

The Echonest is a collection of algorithms that uses the Internet as its data source and attempts to generate `music intelligence' from this. How does one audit a large system that utilises petabytes of data to synthesize knowledge in a dynamic way? How are deeply embedded, fuzzy concepts, such as `cultural vectors', auditable at all? I propose that audit-by-use-case is the only viable method for systems of this nature and complexity, and visualisation of data output can assist in this endeavour and render it as a task on a human-scale.

1) The System

The Echonest, a `music intelligence' company formed in 2005 by two MIT doctoral graduates (Brian Whitman and Tristan Jehan) and acquired in 2014 by Spotify, the streaming-music provider, uses a blended approach to music data gathering and `intelligence' generation. Audio analysis features and metadata gathered from multiple data sources across the Internet (including web pages, and structured, user-edited sources, such as WikiPedia, Discogs, and MusicBrainz) are combined and synthesized into musical `knowledge'. This is then used by other companies, including Spotify, to generate playlists, recommendations and `radio' stations for listeners. Whitman (2012), in a blog post entitled `How music recommendation works - and doesn't work', describes how `real people feeding information into a large automated system from all different sources' is the approach taken by Echonest to gather information written around music. He describes how their system `crawl[s] the web constantly, scanning over 10 million music-related pages a day' and explains that `filtering is implemented to discard non-music content... Every word anyone utters on the Internet about music goes through our systems that look for descriptive terms, noun phrases and other text and those terms bucket up into what we call "cultural vectors" ' (ibid.)

2) The Problem

When Echonest is considered as a recommendation engine it appears to compare well with other systems. The audio-analysis side of the system is generally well thought of and is not considered further here. However, when considered as an aggregator of web data into musical `intelligence' the methodology employed by the system shows itself to be imperfect. Examples can be found where sources that are used by the Echonest (such as MusicBrainz) are factually correct in respect of musical pieces or artists, but the data returned from the system is incorrect and has come from another source.

For example, during the course of our research into musical genre and the representation of this on the web, a list of thrash metal bands, together with `start_date' information, acquired from the Echonest system was examined; the first entry caused a problem - it read: `HammerFall^1918'. HammerFall, the Swedish metal band, didn't start in 1918 - they formed in 1993. The question then is how did this discrepancy occur? A quick search for `HammerFall +1918' revealed a possible cause; EchoNest utilises various sources for start date information, including MusicBrainz, Discogs and Wikipedia, and the Wiki page for `James Michael' came up as a likely candidate for the confusion. James Michael produced HammerFall's 2011 album `Infected'. At the

top of his Wiki page there is a line that reads: *'For the Anglo-Australian solicitor and poet (1824-1868), see James Lionel Michael. For the U.S. federal judge (1918-2005), see James Harry Michael Jr..'*
No other sources that include HammerFall and 1918 are apparent; our conclusion is that this has been scraped and added to the dataset as being relevant to the band when it clearly is not.

3) A Solution

In terms of its own frame of reference, there is the question of the accuracy of the information output from the Echonest system. Forgetting for a minute that some of the information on the Internet may be wrong, evidence has been found that some well established sources of data are sometimes overlooked by the Echonest and information from other sources is used, resulting in incidences of inaccuracy. By conducting a series of relatively simple investigations into musical genre using the system many such discrepancies became apparent; discrepancies which have seemingly slipped though the filtering systems within the Echonest itself.

An initial investigation involved making repeated API calls to the Echonest (using Python and the Pyen library) in order to acquire lists of artists within all available genres, together with their `dates_active' information. By plotting these results, data errors in the Echonest system quickly became evident. For example, within the dance genre known as Zouglou (which originated in Côte d'Ivoire in the mid-1990s) the first artist is listed as `Molière (1622 – 1673)'. This is obviously incorrect, but is one piece of data among over 180, 000 and would be hard to find using other methods (there is a Molière, but not that one). By laying out factual data, such as these, in both list and graph form auditing becomes an instance of a use-case and analysis becomes a human-scale task.
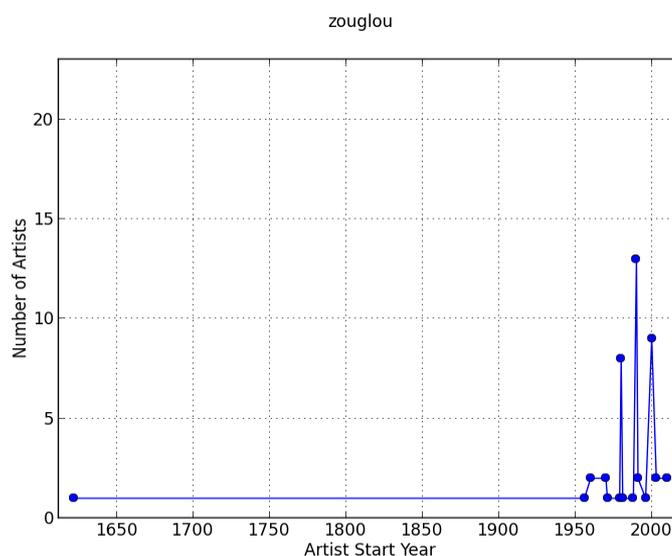
These types of investigations may not solve the problem, but could give rise to metrics to assist in a meaningful assessment of the output from complex, algorithmic systems such as the Echonest.

References

Whitman, B. (2012). How music recommendation works - and doesn't work.
        http://notes.variogr.am/post/37675885491/how-music-recommendation-works-and-doesnt- work.