Characterizing Relations Among Artists in an Online Music Magazine

Brooke Maury
Vijay Viswanathan
Applied Natural Language Processing
SIMS, UC Berkeley
Professor Marti Hearst
December 8th, 2004
Background and Motivation

- Part of a larger project to build a customizable recommendation system based on community metadata
- Interested in the use of ‘Community Metadata’ to characterize music and provide recommendations.
Characterizing Relations Among Artists in an Online Music Magazine

Related Research

- Music Information Retrieval community uses traditional IR and NLP techniques to develop high-level descriptive feature sets about music. (ISMIR '04 had 5 papers)
- Some interesting ‘community metadata’ papers.
  - Whitman & Ellis “Automatic Record Reviews” (2004)
  - Whitman & Lawrence “Inferring Descriptions and Similarity for Music from Community Metadata” (2002)
Related Research Cont’d

- Whitman, Ellis et al use traditional IR, NLP & ML techniques (TF*IDF, POS Tagging & Chunking, ML classifiers) across the entire document, but no localized attention is paid to the text around a link.

- The Knopke article used text surrounding links to music files as a measure of similarity between music files, but the semantics of the link is unstudied, and an arbitrary number of words (10) around the link was taken for analysis.
We want to make use of the general syntactic structure (the sentence) and grammar (POS tags) around links to select features that can characterize the association between artists links.
Ideally, we could use NLP and Machine learning to go from "A is linked to B" to something like:

"A is influenced by B,"

"A is similar to B"

"A was previously in band B"

"A is not similar to B,"

"A recorded on X label" ... 

Or maybe just "The link between A & B is positive or negative"
Main Objectives

- Collect natural language text from artist pages from epitonic.com
- Extract sentences where links to other artists are present.
- Analyze these sentences to understand if we can characterize the links based on the text surrounding them.
- Build a named entity list from marked up Artist names.
Methodology

- Why Epitonic?
  - It’s an easy site to crawl
  - Body is clearly differentiated from other areas
  - Fairly ‘objective’ and informative band biographies.
  - Good starting point for future work
  - Legal MP3 downloads that we may collect for our users
Characterizing Relations Among Artists in an Online Music Magazine

Sonic Youth - (Geffen-Records)

One could make the argument that no other indie rock band of the last two decades has a greater legacy than Sonic Youth. Without this gritty, artsy, and uncompromising Manhattan quartet (now quintet), we wouldn’t have had the same Pavement or Dinosaur Jr., or Nirvana we know and love. Nor would we have had Blonde Redhead or Unwedborn or ... And You Will Know Us By The Trail of Dead today, not the way we understand them. Sonic Youth’s accomplishments are vast and thus difficult to boil down to a single sentence, but here’s a stab. SY took the abstract and surreal experimentalism of academic music and avant-garde jazz and made it punk. In doing so, they changed the course of underground rock music.

Looking back at Sonic Youth’s career, it makes sense that the group took this path. Back in 1981, when they were just getting started, they were four bored, intelligent personalities jointly influenced by the avant-garde lessons of New York no wave and the aggressive hardcore punk. Guitarist Thurston Moore was turned off by the idea of getting into small groups while playing in assorted Simon Branca guitar armies, where they became adept at the distortion and noise that would contribute to the group’s early sound. Bassist Kim Gordon, as artists and recently reformed from Lush, was building a following for her graphic design magazine, also a founding member, and guitarist Robert Rich was married in 1983 to the drummer on the group’s eponymous debut EP on Branca’s Neutral records, before they even recorded for Sonic Youth’s first EP record. With the confrontational noise punk anxiety of Connexion Z (1985) and the only slightly more conventional gnarled and atonal guitar post-Black Flag-Raping (1987) and the violent UK TV, (both EPs in between), SY was consolidating their influences, laying the groundwork for an incredible string of landmark underground rock albums.

The titles of these records should be familiar to those who have studied art history or had a copy of Sonic Youth’s albums over the past two decades and a half: (probably a lot; who hasn’t?) 1983’s Sonic Youth, 1985’s Goo, 1986’s Daydream Nation, the landmark album that established Sonic Youth, 1988’s Sonic Nurse, the band’s breakthrough, 1990’s Murray Street, the band’s first full-length single, 1992’s Goo, and 1994’s Murray Street. Each of these albums is a testament to the band’s ability to create music that is both experimental and accessible, pushing the boundaries of rock and roll while remaining grounded in the musical traditions that have influenced them.

No discussion of Sonic Youth would be complete without some mention of the group’s unique sound. Sonic Youth is a band that is not afraid to experiment with their music. They are known for their use of feedback and distortion, creating a sound that is both harsh and beautiful. This sound has been a defining characteristic of the band’s music, and it is one that has been celebrated by fans around the world.

In 1980, Sonic Youth released their debut EP, Goo. The band’s sound was raw and experimental, and it was a stark contrast to the smooth, polished sound of many of the bands that were popular at the time. Sonic Youth’s sound was a mix of punk, post-punk, and experimental rock, and it was a sound that was both unique and groundbreaking.

Sonic Youth’s albums have been critically acclaimed, and they have been influential in the development of alternative music. Sonic Youth is a band that is not afraid to take risks andpush boundaries, and it is this spirit of experimentation that has made them such an important part of the alternative music scene.

In conclusion, Sonic Youth is a band that has had a significant impact on the alternative music scene. Their unique sound, their willingness to experiment, and their ability to push boundaries have all contributed to their success. Sonic Youth is a band that is not afraid to be different, and it is this spirit of individuality that has made them one of the most important bands of the past two decades.
Methodology

- Used an open-source web crawler called WebSphinx
  - Fine tuned the algorithm to extract only the body text of the page, retaining HTML tags, and outputted to a text file
  - Set the crawler up to start from a particular band, and crawl only to links that were to related bands
Methodology

- Extract sentences where links to other artists are present.
- Analyze these sentences to understand if we can characterize the links based on the text surrounding them.
- Build a named entity list from marked up Artist names for future work with blogs.
Methodology Cont’d

• Drawbacks to Epitonic:
  – Limited to bands that have agreed to have their music posted
  – Relatively small corpus
  – Other problems?
Progress Report

• Crawled site for bands
• Split text into sentences and placed sentences with embedded links in a name, value pair file.
• Building a simple interface to manually tag links
• Building the tagger component (probably Brown)