Information organization in music recommendation

Rachel Bloch Shapiro

College of Information Studies, University of Maryland

Introduction

Music recommendation presents an interesting case study for new approaches to information organization. Traditional searches are not always fruitful, and enabling discovery requires creativity. Without the limitations of shelf space, online retailers offer many more titles, creating a business of "selling more of less" by dipping into the "Long Tail," a term popularized by Chris Anderson. Despite millions of songs available, approximately 90% of sales came from only 2% of albums in 2011.

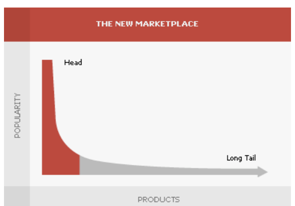

Figure 1. The Long Tail distribution

Key terms:

-collaborative filtering is used by many sites, "people who liked x also liked y"

-content based systems use information about the music itself (tempo, pitch, etc) to inform recommendations

-popularity bias: extremely popular items are over-recommended (the "Harry Potter" effect)

-cold start: new or unknown artists lack fans and will not be recommended by a collaborative filtering algorithm

Literature review

Researchers have proposed exciting new ways to improve music recommendation.

Minimal Metadata: Bogdanov and Herrera propose that automated content-based recommendations, simply filtered by genre, outperform both purely content-based and collaborative filtering programs and require much less data storage.

Crowdsourcing 'Your head is my tail': Lee and Lee suggest a creative approach to the problem of popularity bias by personalizing collaborative filtering algorithms: Identify "experts" and "novices" based on listening habits and favorite songs and use "experts" in particular songs/artist to create recommendations for novices in that area.

Improving Folksonomies: We can't rely upon the "wisdom of crowds" without a crowd. Eck and others solve the cold start by assigning tags to untagged songs using automated listening and prediction.

Syndicated Content: Celma and others mined the information syndicated in RSS feeds from music sites, taking advantage of rich knowledge already on the web.

Case studies

iTunes Genius- Collaborative Filtering

Genius's algorithm appears to rely on collaborative filtering and does not do any content analysis. Barrington, et al showed that if the ID3 tag was removed, Genius does not make recommendations.

Pandora – Content Based

Pandora's "Music Genome Project" consists of information gathered and curated by hand, specialists catalog each song according to ~ 400 parameters. While 95% of the songs in the catalog are played monthly, repetition is a problem, due to it's relatively small catalog.

Echo Nest – Hybrid Approach

Combining automatic content analysis with cultural information via web crawling, the Echo Nest focuses on song level analysis, and taps into pre-existing information like reviews. Songs are analyzed more quickly, and at lower cost than Pandora, allowing the Echo Nest to amass a huge catalog of songs.

Discussion

Music Information's Future: Linked Data

Passant's dbRec program generates novel recommendations using linked data resources such as dbPedia. The seevl YouTube plug in builds upon this research and brings linked data to a huge audience. There are numerous other linked data projects, such as BBC Music, which is now available in RDF, and high level semantic prediction (e.g. mood) programs.

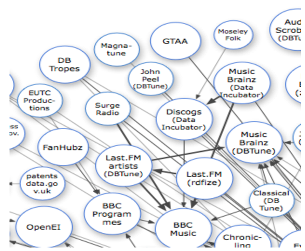

Figure 2. Part of the Linked Open Data cloud showing music sites.

Music Information to Scale: the Million Song Dataset

Previously, non-commercial research was limited by the use of researchers own music collection, or music in the public domain. The Million Song Dataset consists of sophisticated data and metadata for one popular million songs from the Echo Nest Catalog and is freely available.

Conclusion

Research and initiative from the academic community and commercial ventures lead to projects like the Million Song Dataset, and advance recommendation. Linked data, still in its infancy, offers promise. Well-designed recommendation systems will be needed in the future to ensure users are connected with relevant materials.

Literature cited

1. Barrington, L., Oda, R. & Lanckreit, G. (2009). Smarter Than Genius? Human Evaluation of Music Recommender Systems. Retrieved October 24, 2012 from http://cosmal.ucsd.edu/cal/pubs/Barrington-Genius-ISMIR09.pdf
2. Bertin-Mahieux, T., Ellis, D.P.W., Whitman, B. & Lamere, P. (2011) The Million Song Dataset. Retrieved October 24, 2012 from http://ismir2011.ismir.net/papers/OS6-1.pdf
3. Bogdanov, D. & Herrera, P. (2011). How much metadata do we need in music recommendation? A subjective evaluation using preference sets. Retrieved October 24, 2012 from http://ismir2011.ismir.net/papers/PS1-10.pdf
4. Celma, O.(2010) *Music recommendation and discovery: The long tail, long fail and long play in the digital music space*. Berlin: Springer.
5. Eck, D., Lamere, P., Mahieux, T. & Green, S. (2007). Automatic Generation of Social Tags for Music Recommendation. Retrieved October 24, 2012 from http://www.columbia.edu/~tb2332/Papers/nips07.pdf
6. Lamere, P. & Celma, O. (2011) Music recommendation and discovery remastered. Retrieved October 24, 2012 from http://recsys.acm.org/2011/tutorials.shtml.
7. Lee, K. & Lee, K. (2011). My head is your tail: Applying link analysis on long-tailed music listen-ing behavior for music recommendation. *Proceedings of the Fifth ACM Conference on Recommender Systems* (pp 213-220). New York: ACM.
8. Passant, A. (2010). dbrec: Music recommendations using DBpedia? Retrieved October 24, 2012 from http://semanticweb.org/

For further information

Please email rkbshapiro@gmail.com for the full paper.