O'Bryant, Jacob

Calculating music similarity with mobile device playlists

Faculty Mentor: Dennis Ng, Computer Science

Introduction

Music recommendation systems, such as Pandora and Spotify, help listeners to discover new music. The similarity of different songs is an important measure used in music recommendation. We have studied manually-created playlists on mobile devices to see if they can be used to accurately calculate song similarity. We collected playlists from 41 research subjects and used a co-occurrence model to calculate similarity between songs in the collection.

Methodology

To collect playlists, we created an Android app that uploads the user's playlists from their device to our server. In addition, we created a website that uploads users' playlists from Spotify – a music streaming service often used to create playlists. After reading the user's playlists, the app and website use the Gracenote[1] database to get an ID for each song. The Gracenote database contains metadata and a unique ID for a large number of songs. We query the database with each song's title, artist and album and receive the unique ID of the closest matching song. Each playlist is then uploaded to the server as a list of these IDs.

Given a collection of playlists from multiple users, we calculate the similarity S between each pair of songs a and b in the collection with the same formula used by Pachet et al[2]:

$$S_{a,b} = \frac{1}{2}\left[\frac{\text{cooc}(a,b)}{\text{oc}(a)} + \frac{\text{cooc}(a,b)}{\text{oc}(b)}\right]$$

where $\text{cooc}(a,b)$ is the number of playlists containing both a and b and $\text{oc}(a)$ ($\text{oc}(b)$, respectively) is the total number of playlists containing a (b, respectively). This formula gives a value between 0 (not similar) and 1 (similar).

Results

We gathered 473 playlists from 41 different users. The playlists contained a total of 18,948 songs. There were 14,990 unique songs, so each song occurred in about 1.26 playlists on average.

Several studies that try to calculate the similarity of songs use the last.fm database as ground truth.[3] last.fm doesn't provide similarity ratings for pairs of songs, but we used the

[1] http://www.gracenote.com/

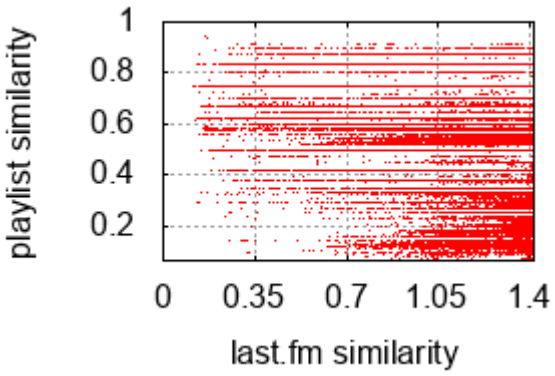
[2] Pachet, F., Westerman, G. and Laigre, D. 2001. Musical data mining for electronic music distribution. In Proceedings of the 1st International Conference on Web Delivering of Music (WEDELMUSIC01).

[3] Knees, P. and Schedl, M. 2013. A survey on music similarity and recommendation from music context data. ACM Trans. Multi-media Comput. Commun. Appl. 10, 1, Article 2 (December 2013), 21 pages.

data last.fm does provide to make another set of similarity calculations for the songs in the playlists we collected. We evaluated the accuracy of the playlist-based calculations by comparing them to the last.fm-based calculations.

To compute the last.fm-based similarity calculations, we used tagging data. last.fm users can manually apply tags to songs to classify them. last.fm exposes data for the most commonly applied tags, but they do not provide an exact count for how many times a tag was applied to a particular song. Rather, the most common tag is given a count value of 100 while every other tag is given a count value proportional to this first value. We represented the tags of each song as a vector where each element of the vector was the count value for a tag. We normalized the vectors so that each vector had a length of one. Then we computed the similarity between pairs of songs in our collection of playlists by calculating the distance between the two corresponding unit vectors. This formula gives a distance value between 0 (similar) and $\sqrt{2}$ (not similar).

About 40% of the 14,990 unique songs we collected could not be found in the last.fm database and are thus not included in our analysis. There were over 80 million different pairs of the remaining songs; however, the playlist similarity for most of these pairs was 0 since most of the songs don't occur in any playlists together. We only considered pairs of songs that have a co-occurrence of at least one. There were 321,282 such pairs. The graph on the right shows the playlist similarity versus the last.fm similarity distance for these pairs.



The correlation coefficient between the two data sets is -0.176. A coefficient close to -1 would have shown that our playlist-based calculations match up with the last.fm-based calculations. Since the coefficient is close to 0, there is only a weak correlation between the playlist-based calculations and the last.fm-based calculations.

Discussion

The weak correlation indicates that the playlist similarity calculations weren't very accurate after all. In particular, the graph shows that many pairs of songs that were given a high similarity rating by the playlist-based formula were given a low similarity rating by the last.fm formula.

Conclusion

It is possible that collecting more data would improve the playlist-based calculations and strengthen the correlation between the last.fm-based calculations. However, with the data we have collected so far, user-created playlists have been an ineffective data source for calculating the similarity between songs.