Multi-Agent Estimation of Musical Preferences

Milind Jagota, Erik Jones, Zhaoyu Lou

Abstract

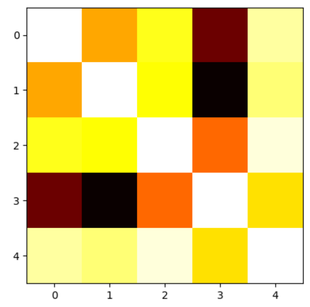
The problem of "**party playlist generation**" is to generate a playlist for a group given each user's listening history. Previous work has modeled and generated songs based on frequency data, which fails to capture large scale musical structure. We developed an algorithm to cluster songs using both raw audio and song metadata, and found that the addition of the metadata improved the quality of the generation when compared to the frequency based model.

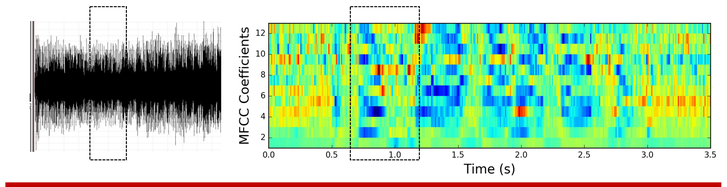
Data and Features

We modeled songs based on 8 features, including 7 metadata features and the timbre matrix, containing frequency data.

- Tempo
- Duration
- Loudness
- Timbre Matrix
- Familiarity
- Hotness
- Danceability
- Energy

To process the timbre matrix, we concatenate the timbre vectors in 5-frame windows, as can be seen below. We then fit a mixture of Gaussians model and use the Fast Spectral Similarity (FSS) as a distance metric. The figure to the right is a heatmap of distances between 5 songs.

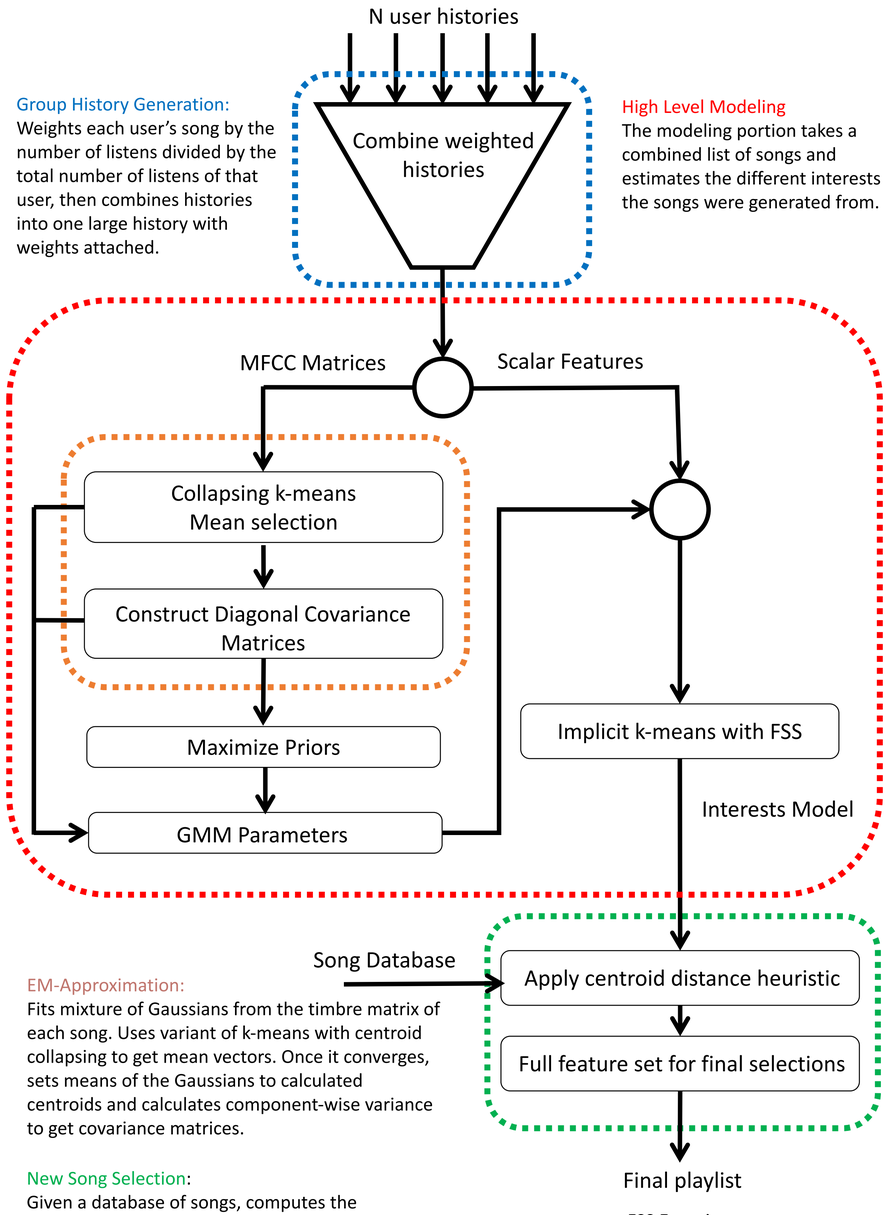




References

[1] Adam Berenzweig, Beth Logan, Daniel P. W. Ellis, and Brian P. W. Whitman. A large-scale evaluation of acoustic and subjective music-similarity measures. Comput. Music J., 28(2):63–76, June 2004.
[2] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The million song dataset. In Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011), 2011.
[3] John Hershey and Peder Olsen. Approximating the kullback leibler divergence between gaussian mixture models. In IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007., volume 4, pages IV–317–IV–320. IEEE, 2007.
[4] Beth Logan and Ariel Salomonn. A content-based music similarity function. Technical report, Processing Languages – Document Style Semantics and Specification Language (DSSSL). Ref. No. ISO/IEC 10179:1996(E, 2001.
[5] Elias Pampalk. Speeding up music similarity. In in Proceedings of the MIREX Annual Music Information Retrieval eXchange, 2005.

System Design and Models



Group History Generation:
Weights each user's song by the number of listens divided by the total number of listens of that user, then combines histories into one large history with weights attached.

High Level Modeling
The modeling portion takes a combined list of songs and estimates the different interests the songs were generated from.

EM-Approximation:
Fits mixture of Gaussians from the timbre matrix of each song. Uses variant of k-means with centroid collapsing to get mean vectors. Once it converges, sets means of the Gaussians to calculated centroids and calculates component-wise variance to get covariance matrices.

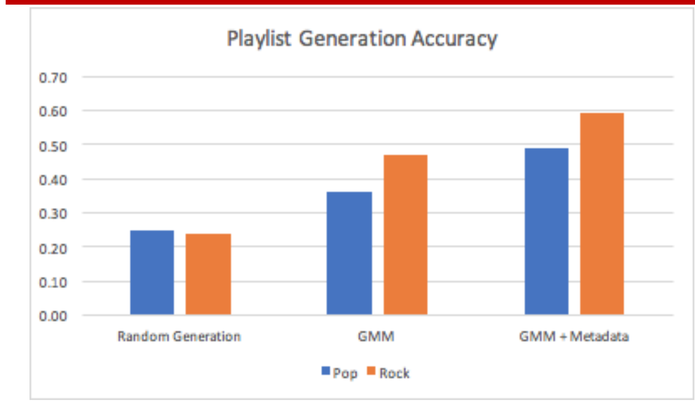
New Song Selection:
Given a database of songs, computes the average distance of every song to each centroid. For the m closest examples, uses full feature distance to select the final playlist.

FSS Equations:

$$FSS(M^A, M^B) = L(S^A|M^B) + L(S^B|M^A) - L(S^A|M^A) - L(S^B|M^B)$$

$$L(S^A|M^B) = \sum_{i=1}^{k_A} P_i^A \log \sum_{j=1}^{k_B} P_j^B \mathcal{N}(S_i^A|M_j^B)$$

Results



Genre	Seed Size	Output Size	Random Generation	GMM	GMM + Metadata
Pop	20	8	0.25	0.36	0.49
Rock	20	8	0.24	0.47	0.59

Since song similarity is highly subjective, evaluation was difficult. We took output playlists and surveyed 20 people on how many outputted songs fit the seed playlist. Due to our dataset only containing songs published before 2012, we were only able to find people who were familiar enough with the seed playlists for two genres.

Discussion and Future

The system we developed integrated raw frequency data and song metadata to improve on previous frequency based methods. We found the accuracy of our frequency based model to be on par with previously published results, and showed that the addition of metadata to the model noticeably improved performance. In doing so, we demonstrated the importance of metadata in capturing high level musical structure and successfully characterizing songs.

In the future, we'd like to develop better, more systematic evaluation methods, and apply our model to a more modern dataset. The system is also computationally expensive right now and it would be good to explore more scalable methods. It would also be exciting to integrate with Spotify or Apple Music APIs.