MUSIC RECOMMENDATION FROM SONG SETS

Beth Logan
Hewlett Packard Labs
One Cambridge Center
Cambridge MA USA

ABSTRACT

We motivate the problem of music recommendation based solely on acoustics from groups of related songs or 'song sets'. We propose four solutions which can be used with any acoustic-based similarity measure. The first builds a model for each song set and recommends new songs according to their distance from this model. The next three approaches recommend songs according to the average, median and minimum distance to songs in the song set. For a similarity measure based on K-means models of MFCC features, experiments on a database of 18647 songs indicated that the minimum distance technique is the most effective, returning a valid recommendation as one of the top 5 32.5% of the time. The approach based on the median distance was the next best, returning a valid recommendation as one of the top 5 29.5% of the time.

1. INTRODUCTION

Listeners are increasingly finding music of interest on the Web rather than through traditional distribution channels. This represents a great opportunity for new and obscure artists to introduce their music to large audiences since the Web has relatively low entry barriers. However, it is difficult for listeners to discover such artists since established automatic music recommendation techniques use either opinions or playlists generated by the public, or meta-data generated by experts. For little-known artists, few experts are interested in categorizing their music and the general public is unaware of their existence. Artists could self-categorize their music but such a system is open to abuse. What is needed then is a way to recommend songs or artists based solely on audio data.

Automatically recommending and organizing music using audio properties has attracted much attention (e.g. see [2], [1] and references). However, even the best systems to date still fall far short of human expectations [2]. The inclusion of *non-audio* meta-data can help overcome such shortfalls, yet for new artists such meta-data does not exist. In such cases though, we can perhaps achieve better performance by including more *audio* data. We propose then rather than studying recommending N songs given one example song to instead study the easier but still very useful task of recommending one song given N related songs. The hope is that if several songs are chosen as representative of the 'sound' the user is seeking, we will have more information on which to base our automatic recommendation. We call this problem the 'song set completion' problem. We use the term 'song set' rather than 'playlist' as we are not concerned with the order in which the songs will be played, merely that together they represent a subgenre preferred by the user. Thus we consider how given a set of user-selected songs we would recommend another song with similar properties using merely audio analysis. Such sets of songs might be a user's favorite songs or a group of songs by the user's favorite artist.

In this paper we present and evaluate four algorithms to recommend songs from song sets. The algorithms are quite general and can be used with any audio distance measure. We test them using our previously published timbre similarity measure.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profi t or commercial advantage and that copies bear this notice and the full citation on the fi rst page.
© 2004 Universitat Pompeu Fabra.

2. RECOMMENDATIONS FROM SONG SETS

In this section, we first briefly describe our previously presented technique to determine acoustic similarity between songs. We then present four algorithms which can be regarded as extensions of this or any song similarity technique to determine the distance between songs and song sets. The approaches differ by whether they build a single model for the entire song set or a series of models for its constituent songs, and by the manner of comparing the model or models to the songs to be recommended.

2.1. Acoustic-Based Music Similarity

In order to provide recommendations from song sets, we require a means to automatically determine the acoustic distance between a song and a song set. This is similar to the task of determining the distance between two songs for which many algorithms have been proposed.

We have previously published and achieved good results with an acoustic similarity measure which captures information about songs' instrumentation or timbre [3]. The approach is similar in spirit to a number of other music similarity algorithms which transform raw audio to perceptually meaningful features and fit a parametric

probability model to these. Similarity is then computed using a suitable distance measure between the models for each song.

In our previous work, each song is first converted to a group of Mel-frequency cepstral coefficients (MFCCs). Such features capture smoothed spectral information which roughly corresponds to instrumentation and timbre. We then model these features using K-means clustering, learning the mean, covariance and weight of each cluster. Having fit models to the data, we calculate similarity by comparing the models. For this, we use the Earth-Mover's distance (EMD) [4] which calculates the cost of 'moving' probability mass between clusters to make them equivalent. For more details refer to [3].

2.2. Modeling Song Sets Directly

Our first technique for recommending songs from song sets builds a single model to represent all the songs in the set and recommends similar songs according to their distance to this model. This is equivalent to treating the song set as one long song. In this paper, we use the models and distance measure from our previously proposed technique described above. However, any model-based acoustic similarity measure could be used.

2.3. Average Distance to the Songs in the Set

The approach described above compares pairs of models trained on quantities of data that could differ by an order of magnitude. Since this may be undesirable, we present an alternative approach. Instead of building one model for the song set, we build a separate model for each of its songs and then recommend songs according to their average distance to a song in the song set. This technique is more scalable than the previous approach; if we form a new song set from a different combination of songs, we need not train a new model.

2.4. Median Distance to the Songs in the Set

The two techniques described above average the distance between a song and a song set either explicitly or by merging the contents of the song set into one song. However, if one or two songs in the set are outliers or unusual, this will affect the average, probably adversely [1] . This is equivalent to saying that if the distribution of distances between a song and a song set is not Gaussian, then taking the average distance will be very sensitive to outliers.

Figures 1, 2 and 3 show the histograms for the distance between a randomly selected song and the rest of the songs on three albums. As described in Section 3, we regard albums as good examples of song sets. We see from these figures that typically, the distribution of the distance

[1] At least for the simple distance measure studied in this paper. One can imagine a very sophisticated recommendation technique which takes note of an unusual song and decides whether it should influence a recommendation.

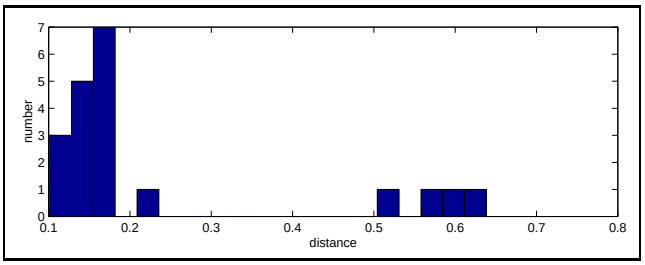

Figure 1. Histogram of the distances between a randomly chosen song from "20 Years of Jethro Tull" and the rest of the songs on the album.

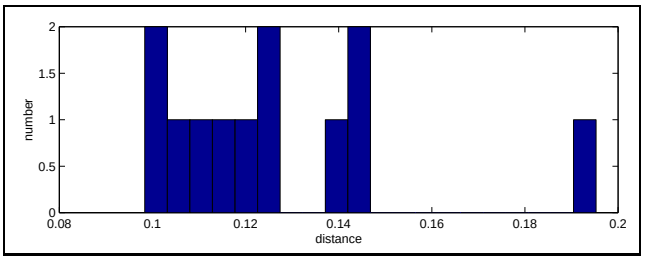

Figure 2. Histogram of the distances between a randomly chosen song from "Jagged Little Pill" by Alanis Morissette and the rest of the songs on the album.

between a song and the songs in the song set is not Gaussian. We have examined such histograms for over 500 albums and found that very few are even close to being Gaussian. We therefore seek a distance measure between songs and song sets that does not rely on the distribution of the distances between songs being Gaussian.

A standard technique from statistics used to improve robustness to outliers when the data is non-Gaussian is to take the median instead of the average. We therefore consider recommending songs using the median of the distances between the song and each song in the song set. This approach shares the scalability advantages of the previous averaging technique but makes less assumptions about the nature of the distance distribution.

2.5. Minimum Distance to the Songs in the Set

Finally, we consider computing the distance between a song and a song set as the minimum of the distances be-

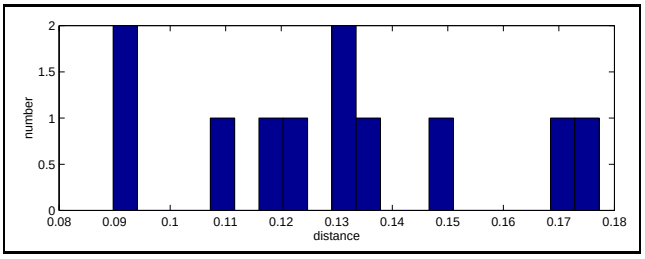

Figure 3. Histogram of the distances between a randomly chosen song from "Backstreet Boys" and the rest of the songs on the album.

Genre	% Collection
Rock	68.2
Classical	5.6
Jazz	5.5
World	3.7
Newage	2.4
Folk	2.4
Soundtrack	2.0
Electronica	1.9
Vocal	1.7
Rap	1.5

Table 1. Percentage of the collection covered by the main genres.

Correct Song	Number Clusters	Top 1	Top 5	Top 10	Top 20
Strict	16	13.7	24.7	31.4	37.5
	64	16.8	27.9	33.5	38.9
	256	15.6	26.8	33.5	39.5
Relaxed	16	16.2	29.9	38.2	46.8
	64	20.3	33.7	41.2	48.1
	256	19.5	33.2	40.5	47.7

Table 2. Percentage of times the correct song was in the top 1, 5, 10 and 20 songs returned according to song sets modeled by K-means models with various numbers of clusters for various definitions of the correct song. Each test song is modeled by a K-means model with 16 clusters.

tween the song and the songs in the set. Although this technique could backfire if the song matches an outlier in the song set, on average it should have good performance.

3. EXPERIMENTS

Having presented a range of techniques to provide recommendations from songs sets, we now study their performance on a database of 18647 songs.

3.1. Experimental Setup

A natural source of song sets is user-generated playlists which can be easily found on the Web. However, our analysis requires data for which audio is available at the song level since we extract features from the audio of each song. Collecting audio for all the songs in even a subset of the playlists on the Web is unfortunately beyond our resources.

Albums however are a source of natural song sets and are more readily available in those sets. We therefore evaluate our algorithms on an in-house database of 18647 songs from 1523 albums for which we have the full audio. The collection covers a wide variety of genres from Classical to Rock. Table 1 shows the percentage of the collection covered by the main genres.

We assume that the list of songs on each album is a valid song set. For each album, we randomly choose one song to omit. These omitted songs form our test set and the remainder of the songs on each album a song set. There are thus 1523 test songs and 1523 song sets in each experiment.

For each song set, we recommend songs from the test set according to our algorithms. Ideally, the song omitted from the each song set's album should be the first recommendation for that song set, although there could be cases in which other songs are valid choices. We report two figures of merit. The first records the percentage of times this omitted or "correct" song was in the top 1, the top 5, the top 10 and the top 20 recommendations. We also study a more relaxed definition of the correct song which includes all songs by the same artist who composed the songs in the song set.

3.2. Results

We first consider recommendations of songs according to closeness to the models built for each song set as described in Section 2.2. We convert the audio to 19 dimensional MFCC vectors and cluster these using K-means clustering. Table 2 shows the percentage of times the correct song was in the top 1, top 5, top 10 and top 20 recommendations for varying numbers of clusters used to model the song set. Each test song is modeled by a K-means model with 16 clusters. We see that these results are very promising, being far better than chance. At least 25% of the time, the correct song is one of the top 5 recommendations. The best result is obtained for 64 clusters. For 256 clusters the performance degrades, presumably because insufficient data is available to learn so many clusters.

If the definition of the correct song is relaxed we obtain the results in the lower half of Table 2. Here we see that an improvement of about 20% relative is possible if one assumes any song returned by the the same artist as the song set would be a suitable recommendation.

We next consider song recommendations according to their average distance to a song in the song set as described in Section 2.3. We model each test song and each song in the song set by a K-means model with 16 clusters and average the EMD between the test song and each song in the song set. The top part of Table 3 shows the results for this experiment for both the strict and relaxed definitions of the correct song. The results are comparable to the previous case in which the song set was represented by a model, although as discussed averaging is more scalable so would be preferred.

Next we study the system described in Section 2.4 in which songs are recommended according to their median distance to the songs in the song set. The middle section of Table 3 shows these results. We see that use of the median provides some advantage over using the average distance or modeling the song set directly. Even for the strictest definition of correct song, almost 30% the time,

Distance	Correct Song	Top 1	Top 5	Top 10	Top 20
Average	Strict	15.8	28.1	34.1	41.2
	Relaxed	18.4	33.4	41.2	50.2
Median	Strict	17.4	29.5	35.0	42.7
	Relaxed	20.9	35.0	41.6	51.3
Minimum	Strict	20.1	32.5	37.6	45.1
	Relaxed	26.5	41.2	47.7	56.1

Table 3. Percentage of times the correct song was in the top 1, 5, 10 and 20 songs returned according to the average, median and minimum distance between it and the songs in the song set for various definitions of the correct song.

the correct song is returned as one of the top 5.

Finally we study the system which recommends songs according to their minimum distance to all songs in the song set. These results are shown in the bottom section of Table 3. These indicate that this approach is the best. For the strictest definition of correct song, a suitable recommendation is returned 32.5% of the time. For the more relaxed definition of correct song, the correct song is chosen in the top 5 41.2% of the time, compared with only 35.0% of the time for the median distance system.

4. DISCUSSION

The results are somewhat surprising. The best approach for recommending songs from song sets appears to be simply choosing songs according to the minimum distance to songs in the song set. There appears to be no advantage in modeling the song set or even considering any song in it other than the one closest to the test song.

This could be an artifact of our choice of song set and our distance measure. Our song sets are albums which typically contain very closely related songs. Although there are outliers, we would be unlucky to choose one of these as our test song. Also, our distance measure works best when comparing two models trained on the same amount of data. Other distance measures designed to model the song set directly may be more effective.

In any case, we should be wary of drawing too many conclusions from this preliminary study. We have only considered one set of test songs and two objective definitions of the correct song. More experiments on a variety of song sets with user evaluations are needed.

5. CONCLUSION AND FUTURE WORK

We have motivated and proposed solutions to the problem of music recommendation based solely on acoustics from sets of related songs. We found that for a timbre-based similarity measure, the best recommendations were obtained by ranking songs by the minimum of their distance to songs in the song set.

Future work will focus on the use of other acoustic distance measures, particularly those incorporating rhythmic information, and learning which sounds in the song set perceptually distinguish it from the rest of audio space. We will also consider recommending groups of songs. As described, we hope to conduct this research on a larger, more varied collection of song sets with greater feedback from users.

6. ACKNOWLEDGMENTS

Thanks is due to Dave Goddeau for useful discussions and to the anonymous reviewers for their feedback.

7. REFERENCES

[1] Berenzweig, A., Logan, B., Ellis, D. P. W, and Whitman, B., “A large-scale evaluation of acoustic and subjective music similarity measures”, *ISMIR*, 2003.

[2] Aucouturier, J-J and Pachet, F. “Improving timbre similarity: How high’s the sky”, Journal of Negative Results in Speech and Audio Sciences, April 2004.

[3] Logan, B. and Salomon, A., “A music similarity function based on signal analysis”, ICME 2001.

[4] Rubner, Y. and Tomasi C. and Guibas L. “The Earth Mover’s Distance as a metric for image retrieval”, Stanford University Technical Report STAN-CS-TN-98-86, 1998.