VISUALIZING MUSIC GENRES USING A TOPIC MODEL

Swaroop Panda
IIT Kanpur
pandas@iitk.ac.in

Vinay P. Namboodiri
IIT Kanpur
vinaypn@iitk.ac.in

Shatarupa Thakurta Roy
IIT Kanpur
stroy@iitk.ac.in

ABSTRACT

Music Genres serve as an important meta-data in the field of music information retrieval and have been widely used for music classification and analysis tasks. Visualizing these music genres can thus be helpful for music exploration, archival and recommendation. Probabilistic topic models have been very successful in modelling text documents. In this work, we visualize music genres using a probabilistic topic model. Unlike text documents, audio is continuous and needs to be sliced into smaller segments. We use simple MFCC features of these segments as musical words. We apply the topic model on the corpus and subsequently use the genre annotations of the data to interpret and visualize the latent space.

1. INTRODUCTION

Music genre visualizations have not caught enough attention. Probabilistic Topic Models [1], have found wide applications in the field of Natural Language processing. We use an unsupervised topic model on music genres data for visualization. For our work we use raw music files, in .wav format. Unlike text documents, raw music data has no discrete components such as words. To create a text-like corpus, we slice the audio data into smaller segments. We use MFCC features of these smaller slices as the representation. Further, to build a corpus, we create a feature dictionary by using the k-means algorithm. Also, in text documents, the inferred topics form a collection of words and hence are straightforward to interpret. In our case, musical words, which are mere MFCC feature arrays, lack inherent meaning and cannot be interpreted. We interpret the latent space of the topic model using genre annotations in the dataset.

2. RELATED WORK

[2] had discussed some audio visualization techniques in MIR which are mostly signal processing based. Topic Models have also been applied on audio [3]for the purpose of audio information retrieval. We use the fault-filtered GTZAN [4] dataset for genre analysis which is popular in MIR community.

3. PROBABILISTIC TOPIC MODEL

Probabilistic topic models are based upon an idea that documents are mixture of topics and topics are probability distributions over words. These words come from a fixed size vocabulary. To make a new document, one chooses a distribution of topics, then chooses a topic from this distribution and finally draws a word from the chosen topic. The Latent Dirichlet Allocation (LDA) inverts this generative process and thus infers the set of topics that were that useful in generating the document. The plate notation below defines the generative process.

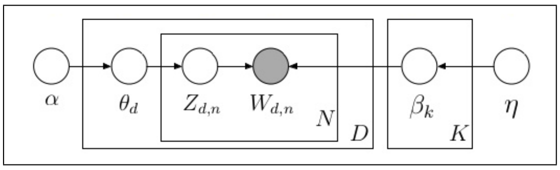

Figure 1. Plate Notation for a Topic Model

4. APPLYING TOPIC MODELS ON MUSIC

The main challenge with the application of topic models in the music (with raw audio files) is to represent the audio in a text-document like corpus. The intent of the work is to interpret the latent space of the topic model using music genres. This interpretation would help giving a probabilistic genre annotation to a song. For example a song may belong 60 % to Blues, 15 % to Jazz and 25 % to Pop genres. To enable such a probabilistic assignments, we build basic genre buckets consisting of at least 3 genres. We do this since a mixture containing all the 10 genres would be very large and obfuscating for the listener to meaningfully interpret. The rationale used to bucket the genres is roughly based on the histories and the musical form of these genres. The first bucket consists of Rock, Metal and Pop genres; the second of Blues, Jazz and Country genres and the final bucket consists of Reggae, Disco and Hip-Hop genres. The songs were clipped down to 0.10 seconds clips. The MFCC features of these clips were then calculated. We then use a K means clustering algorithm on the MFCC features to build the dictionary. It partitions the data into k clusters, where each data point belongs to a cluster and the cluster mean serves as its prototype.

5. INTERPRETATION OF THE LATENT SPACE

Unlike text documents, where topics are interpreted as a mixture of words; the acoustic topic model has topics which are mixtures of cluster means. These cluster means are prototypes of the nearest datapoints(the audiofiles) and thus lack meaning. It is thus essential to assign a suitable meaning to these cluster means. The first part of the interpretation involves understanding the cluster means in terms of music genre. The cluster means are constructed from the MFCC arrays. The cluster means hence can be mapped to and from these audio files and linked with the genre annotations. For instance, lets say that 3 audio files, audio1, audio4 and audio7 make up a cluster mean. We get back to the dataset and find out that audio1 belongs to the Blues genre, audio4 belongs to the Country genre and the audio7 belongs to the Blues genre. Hence, the genres associated with cluster means becomes Blues, Country,Blues. In math, the cluster centers (or terms) can be described as the following,

$$\begin{aligned} clustermean1 &= \{Blues, Country, Blues\} \\ &= 0.67Blues + 0.33Country \\ clustermean2 &= \{Blues, Jazz, Jazz, Country\} \\ &= 0.25Blues + 0.5Jazz + 0.25Country \end{aligned} \tag{1}$$

Once we interpret cluster means in terms of music genres, we can conveniently represent the topics in terms of music genres. The topic space consists of cluster means and an associated probability value. The cluster means can now be defined as music genres with their proportions.

$$\begin{aligned} Topic1 &= prob1 * clustermean1 + prob2 * clustermean2 \\ &= (prob1 * 0.67 + prob2 * 0.25)Blues + \\ &(prob1 * 0.33 + prob2 * 0.25)Country \\ &+ (prob2 * 0.5)Jazz \end{aligned} \tag{2}$$

Once the topic space has been interpreted, the document-topic proportions can also be made sense of. The document topic proportions from the topic model are probability values of the inferred topics present in each document. In this context, the document topic proportions can provide with the proportions of different music genres present within the musical document, that is, a song. The term topic proportions are proportions of different topics for a term in the document. These term topic proportions can be similarly interpreted in terms of genre proportions.

$$Doc1 = prob1 * topic1 + prob2 * topic2 \tag{3}$$

6. EVALUATING THE TOPIC MODEL

We evaluate the model using a genre classification task. We use the model to get the document-topic proportions of every document(song). We use these document-topic proportions as a representation for each song. We use genre labels from the fault-filtered GTZAN dataset, divide the data into train-test sets and perform the classification task using a SVM. We also test our model with different number of topics to look for the optimal number of topics that best capture the genre bucket.

	2	3	4	5
1	0.47	0.53	0.58	0.53
2	0.36	0.38	0.35	0.40
3	0.53	0.46	0.48	0.50

Table 1. Accuracies obtained for different number of topic terms. The rows represent the genre bucket, while the columns represent the number of topics

7. MUSIC GENRE VISUALIZATION

Using the topic model, we can get a probabilistic genre labels of different songs(from document-topic proportions) along with progressive genre visualizations(from term-topic proportions).

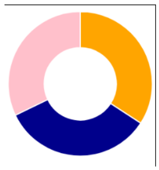

Figure 2. Probabilistic Genre Labels

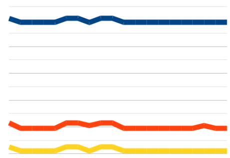

Figure 3. Progressive Genre Visualization

Every song has a probabilistic genre annotation which can be visualized by the doughnut chart in Figure 2; where each colour represents a music genre. Similarly, any song can be represented by a progressive genre visualization, where the y-axis represents time, the x-axis genre proportion and the colours respective music genres.

8. FUTURE WORK

The dataset is too small for the Topic Model to be very effective. Also, the topic model works on the bag-of-words assumption; which may not be too efficient for modelling music data. Other effective music representation techniques can be used. Moreover, different kinds of Topic models can be tuned to work specifically for music data.

9. REFERENCES

[1] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012.

[2] M. Cooper, J. Foote, E. Pampalk, and G. Tzanetakis, "Visualization in audio-based music information retrieval," *Computer Music Journal*, vol. 30, no. 2, pp. 42–62, 2006.

[3] S. Kim, P. Georgiou, and S. Narayanan, "Latent acoustic topic models for unstructured audio classification," *APSIPA Transactions on Signal and Information Processing*, vol. 1, 2012.

[4] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on speech and audio processing*, vol. 10, no. 5, pp. 293–302, 2002.