Music feature extraction based on fractal dimension theory for music recommendation system

Bi Li[12], Qiang Tao[12], Xiang Li[12]

1.Hubei Province Key Laboratory of Intelligence Robot Wuhan, China 430073

2.Wuhan Institute of Technology Wuhan, China 430073

Keywords: feature extraction, spectral envelope, Hilbert transform, fractal dimension

Abstract: Music feature extraction is widely used in music recommendation system. The recommended music is somewhat similar in the form of melody. This phenomenon exhibits a repeating pattern that between the music set and the recommended one, which reveals that the similar music has the characteristics of fractal dimension. In this paper, we selected time energy, frequency energy, Mel-Frequency Cepstral Coefficient (MFCC) and spectral envelope as the music features. These four features were integrated as the feature vectors that were calculated the fractal dimension by Hilbert transform. Compared with the traditional content-based music retrieval, this method focuses on the certain degree of self-similarity between the whole and the local area. This result shows that the feature extraction approach of fractal dimension provide an effective method for music retrieval.

Introduction

The purpose of music recommendation system is to generate a playlist of songs to the listeners to meet their desires. The recommended music always has a degree of similarity in some properties. There are many properties can be selected as features to be retrieved in recommendation system. Previously, the music properties were usually refers to the social tags (keywords) that were mainly marked by many Internet users. Personalized music recommendation systems rely on manual annotations as a mechanism for querying and navigating large music collections [1].It has a well-known problem that new songs cannot be recommended until they have been manually annotated. Although the text-based automatic tagging algorithm is a potential solution to this problem, there are diversified descriptions of a song which may cause the classification and the music style does not match. During the past decade, content-based retrieval has been used for extracting similar pieces of music and classifying pieces which can be considered as the music features [2]. Content-based music retrieval [3]focuses on the abstract description of the audio signal, which reflects the perceptional relevant aspects of the recording, such as loudness, pitch, rhythm, etc. Usually, an audio recording is segmented into short, overlapping frames to extract features and to measure the distance between the candidate pieces in the database. To a great extent, these kinds of methods are only check for exact matches where the search feature is one of the database entries, which cause a relative high time complexity. In a collection of recommended music, all of the songs have approximate melodies in overall. This phenomenon can be considered there exits long-range dependence characteristic in the recommended music.

This paper proposes a fractal dimension algorithm for extracting features to select similar music. We selected four features and integrated them as a feature vector and made Hilbert transform to calculate fractal dimension. Experiment of the method is carried out with fractal dimension which

exhibits effective retrieval process.

Music feature extraction

The purpose of music feature extraction is to reduce the sheer quantity of original signal down to a concise set of values so that the recommendation system can be performed in a reasonable time-frame. There are a number of abstract descriptions to represent the music features. MFCC is one common feature extracted which describes the timbre of a piece of music. The number of MFCC's dimensions has been chosen to limit the further computations rather than by a careful analysis of its impact on the classification accuracy [4]. In this paper, we used four elements that are MFCC, time energy, frequency energy, MFCC and spectral envelope.

Assume $x(t)$ is the t th sampling points of an audio signal sequence, $F = f(\omega)$ is one frame of Fourier $x(t)'s$ transformation, where $\omega \epsilon (l_0, h_0)$, l_0 is the lowest frequency, is the h_0 highest frequency.

Define

$$TE = \sum_{t=1}^{N} |x(t)|^2, t = 1,2,\ldots.N \quad , \quad SE = \sqrt{\frac{1}{h_0 - l_0} \sum_{w=l_0}^{h_0} |f(\omega)|^2}$$

Here, *TE* is time energy, *SE* is frequency energy.

A spectral envelope is a curve in the frequency-amplitude plane, derived from a Fourier magnitude spectrum. It describes one point in time (one window, to be precise) [5]. Spectral envelope is always calculated by Hilbert transform. Hilbert transform is a linear operator that derivers the analytic representation of a signal and leads to the harmonic conjugate of a given function in Fourier analysis. The Hilbert transform of a continuous time sequence is given by

$$\hat{x}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(t-\tau)}{\tau} d\tau = x(m) * \frac{1}{\pi t} \quad , (1)$$

where $x(t)$ is a continuous time sequence. Then the $x(t)$ analytic signal of is

$$g(t) = x(t) + j\hat{x}(t).$$

The amplify of $g(t)$ is

$$A(t) = \sqrt{x(t)^2 + \hat{x}(t)^2} \quad , (2)$$

then $A(t)$ is the spectral envelope of the original audio signal.

Furthermore, the Fourier transform of Eq(1) is

$$\hat{X}(f) = X(f) \cdot \mathcal{F}\left[\frac{1}{\pi t}\right] = X(f) \cdot [-jsgn(f)]. \quad (3)$$

Here $\mathcal{F}$ represents the Fourier transformation, $sgn()$ is a sign function. Eq(3) indicates that is the result $\hat{X}(f)$ of phase shift of in $X(f)$ frequency domain, in the positive frequency threshold it delays phase and in $\pi/2$ the negative one it moves up phase. Eq(3) illustrates $\pi/2$ that Hilbert transform is a 90^0 phase shifter, and have the orthogonal transform $\hat{X}(f)$ $X(f)$ relations.

Fractal dimension calculation

Since Benoit Mandelbrot [6] measured coastline with different fractal scales, the fractal dimension theory has a widely used in many applications. A fractal dimension is an index that how detail in a pattern changes with the changing scale at which it is measured. A fractal dimension sequence has

the self-similarity and long-rang dependence characteristics. Therefore, we can measure the similarities of songs' fractal dimensions in a music set. If the distance is smaller than a threshold the two candidate sets can be identified as one class.

Let X be a metric space, if $S \subset X$ and $d \in [0,\infty)$, the Hausdorff dimension of X is defined by

$$dim_H(X) := inf\ \{d \geq 0 : C_H^d(X) = 0\}$$

In a metric space, we define Hausdorff dimension of an object by the formula

$$D = \lim_{h \to 0} \frac{\log N(h)}{\log(1/h)} \tag{4}$$

where $N(h)$ is the number of disks of size h needed to cover the object. If D is estimated as the exponent of a power law, we mark D as D_0 and call it as box counting dimension.

The recommended music set exhibits a certain level or statistical fractal properties and fractal dimensions. In this paper, we calculated integrate four features as the feature vectors to obtain the box counting dimension.

Measurement and clustering

The similarity of the feature vectors can distinguish different types of music. In this paper, we chose the cosine angle of vectors as the measurement of the distance for vectors. Assume X and Y are two given vectors, the cosine similarity is represented as follow:

$$similarity = cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{5}$$

The resulting similarity ranges from −1 meaning exactly opposite, to 1 meaning exactly the same, with 0 indicating orthogonality (decorrelation), and in-between values indicating intermediate similarity or dissimilarity. In a music recommendation system, the similarity of two songs is more close to 1 means these two songs have the same style that can be recommended.

In fact, although a pair of components has the same position in two vectors, it doesn't mean they are in the same frame of the music. The sequence of the vector only represents the frame order in a song. For discriminate the feature vectors' similarity, we ignored the component's order in vectors and counted the number of the similarity for each component.

We used K-means for identify the similar set. K-means algorithm is one of the most used clustering algorithms for many applications. This algorithm aims to partition n into k clusters in which each observation belongs to the cluster with the nearest mean.

Experiment results

We downloaded 200 internet songs with five styles (blues, pop, rock, classical, country) as the experiment objects.

Pre-processing

For a recommendation system, we only need some pieces of a song to compare the features with the classified music sets instead of a whole song. Because most of the downloaded songs last about 3 or 4 minutes, for a recommendation system, we should intercept local segmentation to compare the features instead of the whole music.In our experiment, we divided each song into 10 pieces with same intervals. The length of the analysis segment is about 70% of the divided piece. We used $H(z) = 1 - \alpha z^{-1}, (\alpha = 0.95)$

to generate the overlapping frames and performed spectral analysis in overlapping Hamming-windowed frames. The hamming window function is

$$w(n) = 0.54 - 0.46\cos\left[\frac{2\pi n}{N-1}\right], 0 \leq n \leq N$$

Features extracted and integrated

For a brief statement, we presented the time energy and frequency energy of a divided piece of the audio signal. And we calculated the spectral envelope by Hilbert transform with the length of 4096. We used Mel triangular filter with the number of 12 to obtain MFCC.

Fractal Dimensions

Combine all of the features mentioned above, for each piece of audio, an integrated feature vector is formed and we can calculate the fractal dimensions to judge the similarity for the candidate songs. In our experiment, each song was extracted four features and was sifted the top ten data in the set. All of the sifted features integrated into a feature vector whose dimension is 2048. It means that we have the frequency data set with 2048 attributes. This experiment chose five styles of music and used the k-means algorithm to cluster the candidate songs into 5 categories.

Table 1. The result of clustering based on fractal dimension

Music Type Type Abbr.	Blues (*M1*)	Pop (*M2*)	Rock (*M3*)	Classical (*M4*)	Country (*M5*)
Original Songs Number	40	40	40	40	40
Cluster 1(T)	35	37	32	28	36
Cluster 1(F)	2*M4*3*M5*	3*M2*	5*M2* 3*M4*	5*M1* 7*M5*	4*M1*
Cluster 2(T)	37	36	34	33	38
Cluster 2(F)	3*M5*	2*M3* 2*M4*	5*M2* 1*M5*	7*M1*	2*M1*
Cluster 3(T)	37	35	35	33	36
Cluster 3(F)	2*M4* 1*M5*	5*M3*	5*M2*	5*M1*2*M2*	3*M1* 1*M4*
Cluster 4(T)	35	35	36	33	36
Cluster 4(F)	5*M5*	5*M3*	4*M2*	4*M1* 3*M5*	4*M1*
Cluster 5(T)	35	36	37	35	35
Cluster 5(F)	1*M4* 4*M5*	4*M5*	3*M2*	3*M1* 2*M5*	5*M1*

Note: (1) T is the abbreviation of True, which means the correct classification, while F means the wrong classification. (2) Such as"3*M5*" expression means that there are 3 songs are classified as *M5* type.

Table 1 shows the result of the classification by using fractal dimension method. From the statistical results, this method has achieved a relatively high accuracy which indicates that the method is effective.

Conclusion

This paper proposed a music feature extraction method based on fractal theory for the music recommendation application. From the view of fractal dimension theory, it measures the similarity between the individual and the music set. The experiment result shows that the method has achieved a satisfied classification. The result also presents that there are some wrong classification in Table 1., most of them occurred in the songs that have similar style, for example, some country music are similar with the blues. Therefore, how to distinguish the subtle differences between those songs that have similar style is one of the research directions in the future. Furthermore, in the background of

the rapid development of mobile network applications, how to meet the real-time requirements of the network services is another important subject.

Acknowledgment

The authors are grateful to the anonymous reviewers for their valuable comments and suggestions which help improve this paper. This work is supported by the National Science Foundation of China (No.61103136).

References

[1] SR Ness , A Theocharis , G Tzanetakis , LG Martins, Improving automatic music tag annotation using stacked generalization of probabilistic svm outputs. International Conference on Multimedia,705-708(2009).

[2] Remco C. Veltkamp, Frans Wiering, Rainer Typke, Content Based Music Retrieval, Encyclopedia of Multimedia, Springer ,97-98 (2008).

[3] Content Based Music Retrieval - Music Formats, Retrieval tasks, Searching symbolic music, Searching musical audio, Feature extraction, Audio Fingerprinting, Concluding Remarks. Information on http://encyclopedia.jrank.org/articles/pages/669 8/Content-Based-Music-Retrieval.html

[4] Abhinav Singh & Ashna Dhanda. Automatic Genre Classification of Audio Signals. Indian institute of technology, India, 2012.

[5] Information on https://en.wikipedia.org/wiki/Spectral_envelope

[6] Mandelbrot, B. "How Long is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension".Science.156(3775): 636–638(1967).