Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 116 (2017) 174–181

Procedia Computer Science

www.elsevier.com/locate/procedia

2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017, 13-14 October 2017, Bali, Indonesia

Indonesian's Traditional Music Clustering Based on Audio Features

Aisha Gemala Jondya[1]*, Bambang Heru Iswanto[2]

[1]*Computer Science Department, School of Computer Science, Bina Nusantara University, Jalan Kebon Jeruk Raya No.27, Jakarta, 11530, Indonesia*

[2]*Universitas Negeri Jakarta, Jalan Rawamangun Muka, Jakarta – 13220, Indonesia*

Abstract

Cluster analysis has been used widely in some applications. In this research, 101 songs from 18 provinces in Indonesia clustered by some set of features extracted directly from the audio data. Before the clustering process, feature selection process with PCA method performed using 60 audio segments from 4 songs to find the optimal set of features which will be used in clustering process. In clustering process, the selected features are extracted from audio signal and clustered by *x*-Means algorithm to find the proper number of cluster. Clustering with this method resulted 4 clusters. The result of this process shows the characteristic of each cluster and some distributions of cultures between areas and provinces. An Agglomerative Hierarchical Clustering method also conducted to compare the result.

Peer-review under responsibility of the scientific committee of the 2nd International Conference on Computer Science and Computational Intelligence 2017.

Keywords: music, clustering, *x*-Means, PCA, AHC

1. Introduction

Indonesia has rich cultures including various style of traditional musics which spread across islands. This traditional musics influenced by some historical background such as multitude of religions and customization of

* Corresponding author. Tel.: +62-812-83600-252
E-mail address: aisha.jondya001@binus.ac.id

1877-0509
Peer-review under responsibility of the scientific committee of the 2nd International Conference on Computer Science and Computational Intelligence 2017.
10.1016/j.procs.2017.10.019

foreign cultures. The aim of this work is to find the features that can differentiate Indonesia's traditional music so that we can explore the similarities between songs from some provinces in Indonesia using the computational approach.

One of the challenges in analyzing music similarity is to find out what it is that allow us to differentiate between music styles which are not directly comparable [1]. Features from each audio data have to be selected and extracted to find the effective set of features. Most of the research in audio analysis conducted using the numerical values of the features that represent audio.

Analyzing the similarity of Indonesian traditional music can be done by using clustering method. Since there is no information about the songs correlation between provinces, clustering results are also used to see the number of clusters. Therefore, x-Means algorithm is used because this method allows the data sets to divided on the optimal number of clusters without prior knowledge [2].

This experiment divided by 2 main process; (1) feature selection and (2) clustering process. This paper is organized as follows: section 2 describe the related work while section 3 described the overall methodology. An overview of audio data explained in section 4. Section 5 described the method and detail of feature selection process and also the validation process while section 6 describe the clustering process using *x*-Means method and compare the result with the AHC method. Discussion and future research are given in section 7.

2. Related Works

For long times, automatic audio music clustering research has been performed for many purposes using manually specified features. A clustering using *k*-Means conducted on classical, rap, metal and Indian's music[3]. Recommendation systems mainly use manual genre-annotations or collaborative filtering, which is time consuming, therefore this study believed that music recommendation systems can be improved a lot by algorithmic music clustering. Instead of using spectro-temporal features, this study clustered music using subjective features provided by Echonest, such as time signature and danceability.

Another clustering was also conducted using manually specified features to compare the culture of 16 Austronesian society [4]. In this study, each song was listened and coded manually based on the scheme developed by the author. Although this two studies used clustering process to help them automatically find similarity between songs, using features that defined manually is not efficient.

Audio analysis process becomes easier and accurate using features that are extracted automatically from audio. The low-level audio features are popular to used in some audio classification research. Ellis et al. classify a set of enviromental sound from a set of audio videos using MFCC features [5]. In this study, human perception of sound textures became principals to construct automatic content clasification.

The low-level audio features has also been used in clustering study. Li et al. studied clustering based on timbral texture features and rhythmic content features extracted automatically from audio for better recommendation system [6]. More over, similar to our study, folk music clustering of four non-European eastern countries, western music and folk music of Cyprus conducted to find the similarity between songs [7]. This study used 25 low-level features and 13 mid-level features to gather information from pitch histogram. Since clustering has no initial knowledge, this study compared clustering in *k*-Means and SOM method where the 2 methods resulted similar result.

3. Methodology

Clustering analysis is one of the technique in machine learning which have the basic purpose to find the informative and useful pattern in a big data [8]. The data mining main process consists of data preparation (data preprocessing), data transformation, data mining, and interpretation of results [9]. Clustering is different from classification where classification's accuracy can be calculated since the labels or prior knowledge of data are available.

In this study to ensure the clustering generates a good result, a feature selection preprocessing is performed by clustering 4 songs segments from 4 different provinces using 11 sets with total 36 recommended features by Giannakopoulos [10]. Eleven sets of time domain and frequency domain features are energy, entropy energy, zero-crossing rate, spectral centroid, spectral entropy, spectral flux, spectral rolloff, Mel-Frequency cepstral coefficient, Harmonic, Chroma Vector and spectral zone. What follows, are definitions of each features.

1. Energy

Let $x_i(n)$, $n = 1, \ldots, W_L$ be the sequence of audio samples of the ith frame, where W_L is the length of the frame. The short-term energy is computed according to the equation:

$$E(i) = \sum_{n=1}^{W_L} |x_i(n)|^2 \quad (1)$$

2. Entropy Energy
The short-term entropy of energy can be interpreted as a measure of abrupt changes in the energy level of an audio signal. the entropy,$H(i)$ of the sequence e_j is computed according to the equation:

$$H(i) = -\sum_{j=1}^{K} e_j \, log_2(e_j) \quad (2)$$

3. Zero Crossing Rate
The Zero-Crossing Rate (ZCR) of an audio frame is the rate of sign-changes of the signal during the frame. The ZCR is defined according to the following equation:

$$Z(i) = \frac{1}{2W_L} \sum_{n=1}^{W_L} |sgn[x_i(n)] - sgn[x_i(n-1)]| \quad (3)$$

4. Spectral Centroid
The spectral centroid is the center of 'gravity' of the spectrum. The value of spectral centroid, C_i , of the ith audio frame is defined as:

$$C_i = \frac{\sum_{k=1}^{Wf_L} k X_i(k)}{\sum_{k=1}^{Wf_L} X_i(k)} \quad (4)$$

5. Spectral Entropy
Spectral entropy is computed in a similar manner to the entropy of energy, although, this time, the computation takes place in the frequency domain. Spectral entropy is computed according to equation:

$$H = -\sum_{f=0}^{L-1} n_f . log_2(n_f) \quad (5)$$

6. Spectral Flux
Spectral flux measures the spectral change between two successive frames and is computed as the squared difference between the normalized magnitudes of the spectra of the two successive short-term windows:

$$Fl_{(i,i-1)} = \sum_{k=1}^{Wf_L} (EN_i(k) - EN_{i-1}(k))^2 \quad (6)$$

7. Spectral Roll-off
This feature is defined as the frequency below which a certain percentage, if the mth DFT coefficient corresponds to the spectral rolloff of the ith frame, then it satisfies the following equation:

$$\sum_{k=1}^{m} X_i(k) = C \sum_{k=1}^{Wf_L} X_i(k) \quad (7)$$

8. Mel-Frequency Cepstral Coefficient
MFCCs are actually a type of cepstral representation of the signal, where the frequency bands are distributed according to the mel-scale.
9. Harmonic
Harmonic features in this research consists of Harmonic Ratio and Fundamental Frequency features. Harmonic ratio is the proportion of harmonics in the spectrum.

$$r(i,k) = \frac{\sum_{j=m}^{m+n-1} s(j)s(j-k)}{(\sum_{j=m}^{m+n-1} s(j)^2 \times \sum_{j=m}^{m+n-1} s(j-k)^2)^{0.5}} \text{ , where } \quad H(i) = \max_{k=Q} r(i,k) \tag{8}$$

10. Chroma Vector

12 representation of the elements of the spectral energy. This feature is widely used as a descriptor of the application related to music. Chroma Vector DFT coefficient is calculated by grouping of short-term window into 12 bins. Every bins is calculated according to equation:

$$v_k = \sum_{n \in S_k} \frac{X_i(n)}{N_k} , \quad k \in 0, \dots\dots\dots\dots 11 \tag{9}$$

11. Spectral Zone

Spectral zone features indicate the frequency range of a spectrum in audio frame. Spectral zone is calculated by dividing the total number of FFT window by the number of FFT window in the range of 100-500 HZ.

This sets of features extracted with the combination of statistical value mean and standard deviation. Eleven sets features is reduced by selecting the optimal features using PCA method. The optimal set of features that is generated from feature selection process used in clustering 101 audio data. The feature selection and the clustering process can be seen in Figure 1a and Figure 1b.

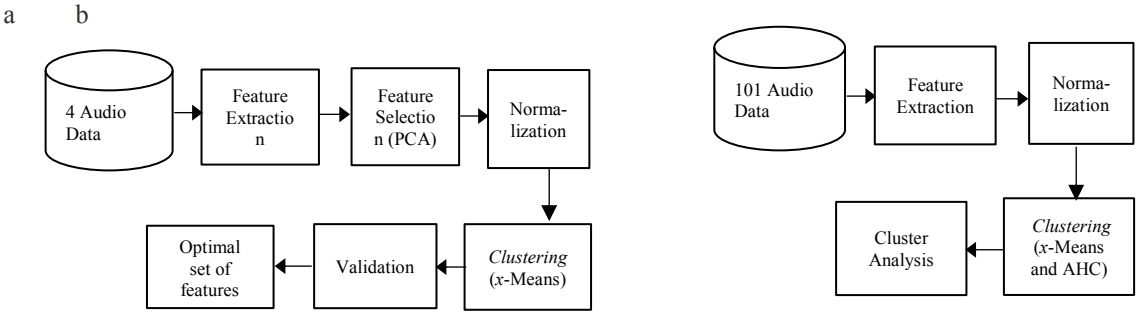

Fig. 1. (a). Feature Selection Process (b). Clustering Process

4. Data

The audio used in this research are the Indonesian traditional instrumental musics which contains no vocals. 101 different songs from 18 provinces in Indonesia are used in the whole process. The audio data details can be seen in Table 1 below.

Table 1. Audio Details.

Provinces	Number of Songs
NAD	5 Songs
Bali	8 Songs
South Sumatera	10 Songs
Bengkulu	4 Songs
Betawi	9 Songs
Jambi	5 Songs
East Java	2 Songs
Central Java	7 Songs
West Borneo	3 Songs
South Borneo	2 Songs

Central Borneo	5 Songs
East Borneo	1 Songs
Lampung	6 Songs
NTT	2 Songs
NTB	3 Songs
Riau	10 Songs
West Sumatera	10 Songs
West Java	9 Songs

Four songs from Bali, Bengkulu, Central Java and West Sumatra which have significant differences are selected from 101 songs to be used in the feature selection process. The most representative part of each songs are trimed which length 1 minute of duration and saved in .WAV format. In clustering process, the most representative 4 seconds-length part was trimed from each of 101 songs and were extracted based on the selected features to obtain a feature vector. The feature vector is used for clustering process.

5. Feature Selection Process

In feature selection process, Four segments of songs, which length 1 minute, were segmented automatically by using mid-Term windowing technique. Each window is segmented into 15 shorter segments by using mtWin and mtStep on MATLAB software and obtained 60 labeled data objects. Eleven sets of feature extracted from this data objects by combining statistical values i.e mean and standard deviations. The feature extraction performed in MATLAB by using `stFeatureExtraction()` and generated a 60x72 dimension feature vector.

Seventy two features reduced by using PCA method to find the optimal features wich will be used in clustering. This process conducted in RapidMiner software by selecting PC1 and PC2 as 2 principal components to devined the features. These 2 PC can separate the 60 segments of songs into 4 different clusters as can be seen in figure 3.

This two Principal Components are used to interpret the eigenvector table. Zero-Crossing Rate, Energy, Spectral centroid, Spectral Entropy and Spectral Roll-off have a significant value to the two PC parameters. The 5 sets of features (total 6 features) with combination of mean and standard deviation statistical values are used for clustering. The total number of features to be used are 12.

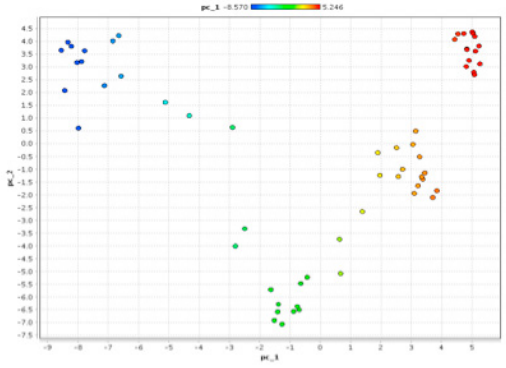

Fig. 2. Clustering Process

x-Means clustering performed on 60x12 dimension feature vectors to prove the effectiveness of the selected features. The clustering process generates an excelent result. *x*-Means algorithm could detect 4 clusters automatically without any prior knowledge. All audio segments of a song go into the same cluster. Based on the result the clustering result has 1 purity value and 0 entropy value. Compositions of each clusters are:

1. All audio segments of Bengkulu_Kromong12.wav go to cluster 0
2. All audio segments of Bali_Gilak.wav go to cluster 1
3. All audio segments of Jateng_Gamelan_4.wav go to cluster 2
4. All audio segments of Sumbar_rantak.wav go to cluster 3.

6. Clustering Process

Clustering process of 101 Indonesian traditional music has a similar order to the previous feature selection process. Five sets of selected features extracted from 101 segments of audio data using `stFeatureExtraction()` function in MATLAB. This feature extraction process generate a 101x12 dimension feature vector. *x*-Means clustering performed on the feature vectors. The clustering process generated automatically 4 clusters from 101 songs with compositions shown in table 2. Cluster number is just for labelling.

Table 2. Cluster Compositions.

Cluster	Number of Songs
Cluster 0	25 Songs
Cluster 1	27 Songs
Cluster 2	15 Songs
Cluster 3	34 Songs

The clustering result visualized by a parallel coordinates graph as can be seen in the Figure 5a and general relationship between clusters and features can be seen in the Figure 5b. The pattern formed by each cluster can be seen on the figure 5b. Cluster 0 generally have the lowest value of ZCR compared with other cluster. Energy value of cluster 0, 1, and 3 are similar both the mean and standard deviation, while the cluster 2 energy tend to be higher than the other clusters. The mean value of SC 1, SC2, SE, and SR as well as the standard deviation of ZCR of cluster 0 and 1 tend to be lower than cluster 2 and 3.

a b

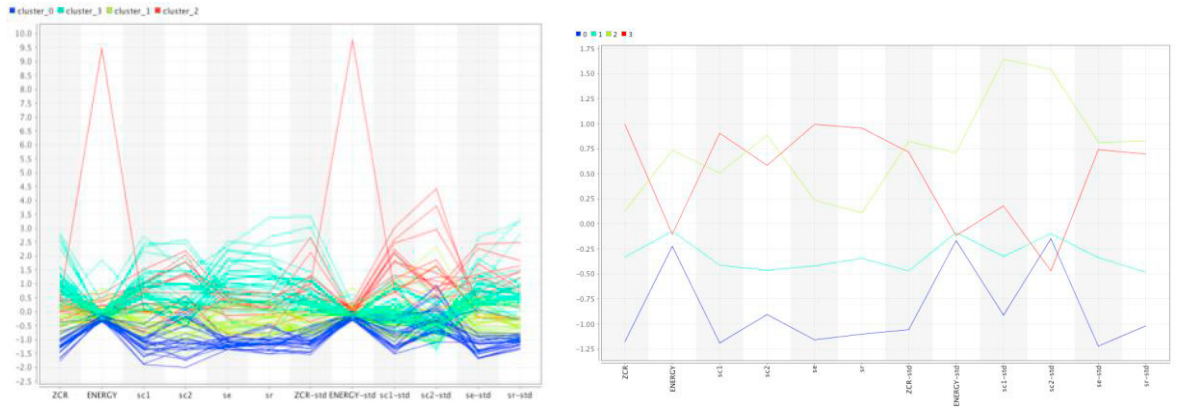

Fig. 3. (a) x-Means Clustering results of 101 songs segments (b). Features and Clusters Relationship

Besides the relationship between features and each clusters, this study also found the music distribution among provinces as can be seen in table 3. The table shows that the songs from Central Java and Jambi are all on cluster 0. Most songs from Bengkulu and West Java are also members of this cluster.

Table 3. Music Distribution from Clustering Result.

Provinces	Number of Songs			
	Cluster 0	Cluster 1	Cluster 2	Cluster 3
NAD	1	1	0	2
North Sumatera	0	0	0	10
Bengkulu	3	2	0	0
Jambi	5	0	0	0
Central Java	7	0	0	0

East Java	0	1	1	1
West Borneo	2	1	0	0
South Borneo	0	1	0	1
Central Borneo	1	2	0	2
East Borneo	1	0	0	0
Lampung	2	2	0	2
NTB	0	0	1	2
NTT	0	2	0	0
West Java	4	1	3	1
Bali	0	1	0	7
DKI Jakarta	0	5	1	3
Riau	0	6	0	3
West Sumatera	0	2	8	0

Meanwhile, songs from West Sumatra spread on cluster 1 and cluster 2. It appears that members of cluster 2 is dominated by songs from West Sumatra. Based on these results it can be concluded that West Sumatra songs are not similar to central java since they are far separate. All songs from North Sumatra and Bali are all in cluster 3.

Clustering using Agglomerative Hierarchical Clustering (AHC) is also conducted to show the relationship of each data object in hierarchy. AHC method does not generate a fix clustering result with details members and number of clusters but instead it generates a hierarchical structure of the data displayed on a dendrogram. The number of clusters and members also obtained by observing the specific level on dendrogram.

Clustering with AHC method was done by using R software. Metric distance used is Eucledian Distance by using the argument `dist(object,method="euclidean")` while Ward method used to calculate cluster proximity by using arguments `hclust (object,method="ward").` AHC clustering performed with $k = 3$ to be compared with the x-Means clustering results . The clustering results shown on the table 4 below.

Table 4. Music Distribution from AHC Clustering Result.

Provinces	Number of Songs			
	Cluster 0	Cluster 1	Cluster 2	Cluster 3
NAD	1	1	2	0
North Sumatera	0	8	2	0
Bengkulu	5	0	0	0
Jambi	7	0	0	0
Central Java	1	2	0	0
East Java	3	0	0	0
West Borneo	1	1	0	0
South Borneo	1	1	3	0
Central Borneo	1	0	0	0
East Borneo	4	1	1	0
Lampung	0	2	0	1
NTB	1	0	1	0
NTT	4	4	1	0
West Java	0	7	1	0
Bali	5	3	1	0
DKI Jakarta	1	4	4	1
Riau	1	2	0	7
West Sumatera	1	1	2	0

The table above shows that the songs distribution on each cluster in AHC methods tend to be similar to the x-Means method. Since cluster's number is just used to naming the cluster, meaning cluster 0 is not the same with cluster 0 from x-Means result but by seeing from the composition. Songs from North Sumatera are in clusters 2 and 3 together with songs from Bali. As well as the *x*-Means clustering results, songs from West Sumatra also becoming the majority

members of a cluster. Meanwhile, songs from Central Java dominated the first cluster together with songs from Jambi and the other areas of Sumatera. In contrast, *x*-Means method does not classify any songs from Jakarta are in the same cluster with songs from Central Java, whereas AHC algorithm put songs from both of these provinces in the same cluster.

7. Discussion and Future Works

The biggest challenge in clustering the data that does not have any ground truth are visualizing and validating the results. One possible way to see the effectiveness of the approach of this study is to compare the results with other methods. 2 methods performed to compare membership of the songs on each cluster.

Results from both methods have similarities in cluster compositions. Therefore, it leads to conclusion that selected features generate consistance clustering results. The clustering results also showed some cultural distribution between provinces. There are some songs from different provinces go into same clusters, i.e songs from North Sumatra and Bali.

This study is expected to be the first step of the next research of Indonesian traditional music. The results of this study found that the used of 5 feature set; Zero Crossing Rate, Energy, Spectral centroid, Spectral Spectral Entropy and Spectral Roll-off generates a good clustering result. Based on this, the next research is expected to be able to find other combination of feature set, or even to find the more optimal features set by using the more complete audio data. In addition, the research collaboration with experts from various fields, such as cultural anthropologist, is believed could generate a lot more beneficial information.

References

1. Wongso R, Santika DD. Automatic Music Classification Using Dual Tree Complex Wavelet Transform and Support Vector Machine. Journal of Theoretical and Applied Information Technology. 2014 May 10; 63.
2. Ishioka T. An Expansions of □X-Means For Automatically Determining The Optimal Number of Clusters. fourt IASTED International Conference Computational Inteligence. 2005;: p. 91-96.
3. Sen A. Automatic Music Clustering Using Audio Attributes. International Journal of Computer Science. 2014; 3: p. 307-312.
4. Rzeszutek T, Savage PE, Brown S. The Structure of Cross-Cultural Musical Diversity. Royal Society Publishing. 2011.
5. Ellis DPW, Zeng X, McDermott JH. Classifying Soundtracks With Audio Texture Features. ICASSP. 2011;: p. 5880-5883.
6. Li Q, Kim BM, Guan DH, Oh DW. Music Recommender Based on Audio Features. 27th annual international ACM SIGIR conference. 2004;: p. 532-533.
7. Andreas N, Maria P, Ioannou R, Petkov N, Schizas CN. A Machine Learning Approach for Clustering Western and non-Western Folk Music Using Low-level And Mid-level Features. In International Workshop on Machine Learning and Music; 2013; Prague.
8. Witten IH, Frank E, Hall MA. Data Mining: Practical Machine Learning Tools and Techniques: Morgan Kaufmann; 2011.
9. Fayyad U, Shapiro GP, Padhr. From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence. 1996; 3(17): p. 37-54.
10. Giannakopoulos T, Pikrakis A. Introduction to Audio Analysis: A Matlab Approach: Elsevier; 2014.