Simulation and Analysis of ANFIS (Adaptive Neuro-Fuzzy Inference System) for Music Genre

A H Pratama¹, W Dwiono¹ and M T Tamam¹
¹Department of Electrical Engineering, University Muhammadiyah Purwokerto
Indonesia
*e-mail: anugrahwp.ws@gmail.com

Abstract. Music comes in many different genres and styles according to its content, which is easy for a human listener to distinguish but hard for a machine to do. This limitation encourages the creation of a system that can help a computer to classify music genre better. This paper proposes a simple method that can classify music genre on a classical, jazz, pop, and rock using frequency analysis feature extraction and ANFIS classification method. There are two types of ANFIS model, which is model A and model B. The most accurate model, is model A with an average accuracy of 53.33% across all genre. The best model for single genre classification is Model B with 80% accuracy for Classic type.

1. Introduction
Music comes in many different genres and styles according to its content. It is easy for a listener to distinguish them but hard for a machine to do. Many researchers used many algorithms to classify the music genre by computer utilisation. ANFIS algorithm has been used for the grouping step in a speaking/melody perception system [1], the neural network in [2-4]. A different approach has been made in [5]. The proposed methods, which are collections of a feature extraction phase tailed by a grouping technique, search both the differences of parameters used as input and the classifier design. Another method is utilising some audio feature [6] and unsupervised feature pre-training [7].

Brendan Petty[3], has created a model of music genre classification system using Back Propagation Artificial Neural Network. This system classifies across four genres; pop, classic, heavy and jazz. This model resulted in an overall accuracy of 66% across all type. This model proves that ANN can be used as a classification method for song genre, albeit the low accuracy.

Michael Haggblade, Yang Hing, and Kenny Kao[8] created model of song classification using four methods, they are DAG SVM (directed acyclic graph support vector machine), Neural Networks, k-means, and k-NN. This model is used to classify four genres; classical, jazz, metal, and pop, which resulted in 86%, 96%, 87%, and 87% accuracy respectively. This condition proves that song can be classified using a machine learning method.

Imam Ikhsan [2] created a model of song classification using hidden Markov model which utilised a probability model. This method was used to classify three genres which are pop, rock, dance and resulted in 80% overall accuracy.

Based on those, this paper proposes a method to classify four types of music genre by using ANFIS (Adaptive Neuro-Fuzzy Inference System). ANFIS has been chosen as a classification method because ANFIS is a combination of Artificial Neural Network (ANN) and Takagi-Sugeno fuzzy logic inference system. This property makes ANFIS has almost identical features as human brains in decision making which theoretically should work great as a classification method, especially in classifying analogue signal such as song. The accuracy of the created ANFIS model is expected to be around 80% across all genre which is, Classic, Jazz, Pop, and Rock.
2. Research Method

The model in this paper is constructed as a simulation using MATLAB, while ANFIS is used as a classification method[9][10]. The amount of data used is 220 songs where 160 songs are used for training (40 for each genre) and 60 songs for testing (15 for each style) The overall classification model could be simplified as below.

![Figure 1. Simplified block diagram of music genre classification model using ANFIS](image)

2.1. Pre-processing

In pre-processing, the song will be polished so that the feature extraction can be done efficiently. First, the song will be filtered between 1 – 16384 Hz, and then it will be downsampled to 8000 Hz, and its channel will be reduced to mono. Lastly, a DC removal method will be applied to remove DC offset noise.

2.2. Feature Extraction

Feature extraction in this model was done by analyzing the content of each frequency range. The audio file will first be divided into four frequency category, which then for each frequency, the specified variable can be extracted. Fourteen variables can be extracted across all frequencies.

a. Low Frequency (0-200Hz)
   Variable extracted: bass frequency variation, spectral power - low

b. Low-mid frequency (200Hz-600Hz)
   Variable extracted: spectral power - low-mid

c. Mid Frequency (600Hz-1250Hz)
   Variable extracted: mid-frequency beat likelihood, mid-frequency beat offset, mid-frequency variation, Spectral power - mid

d. High Frequency (above 5000Hz)
   Variable extracted: high-frequency strength of half-beat, spectral power – high

Those variables are special to each frequency range, other variables that are not specified to certain frequency are stereo spread, tempo, the strength of half-beat, dynamic range, attack velocity - fast, and attack velocity – slow.

2.3. Constructing ANFIS (Training method)

In this stage, the ANFIS system will be constructed by training the system using a database which was obtained from the feature extraction stage. The database consists of 7 selected variables which have the most distinctive value difference. The training method is a hybrid method, which combines the least-squares estimator method on the forward training and gradient descent on the backward training.

3. Result and Analysis

3.1. Feature Extraction

The result of feature extraction is represented in Table 1. All variables are in average value to make observation and analysis easier.
Table 1. The average result of all variable in feature extraction

| Variable                        | Classic       | Jazz          | Pop           | Rock          |
|---------------------------------|---------------|---------------|---------------|---------------|
| dynamic range                   | 1.4734856     | 1.163313774   | 0.844799464   | 0.840871379   |
| fast peak strength              | 0.889252      | 1.20585283    | 0.877501786   | 0.591762069   |
| high low weight to double       | 0.914536      | 1.243662264   | 0.933158929   | 1.087165517   |
| high RMS                        | 0.392766      | 0.338833962   | 0.343866071   | 0.425398276   |
| low beat dev                    | 1.320644      | 0.971219623   | 0.631375      | 0.72216069    |
| low, mid RMS                    | 0.181534      | 0.301966038   | 0.564892857   | 0.450426316   |
| Low rms                         | 0.041402      | 0.040498113   | 0.125091071   | 0.062781034   |
| Mid beat dev                    | 1.214354      | 0.995069811   | 0.601792857   | 0.722679483   |
| mid beat likely hood            | 1.125552      | 1.032001887   | 1.021205357   | 1.030113621   |
| mid beat offset                 | 0.416104      | 0.44007717    | 0.417451786   | 0.425796552   |
| Mid-RMS                         | 0.323302      | 0.349067925   | 0.368128571   | 0.487885965   |
| slow peak strength              | 1.374362      | 1.342766038   | 0.2569        | 0.319444828   |
| stereo RMS                      | 1.30038       | 0.682358491   | 0.631148214   | 0.977753448   |

The data in this table then converted to graph (Can be seen in Figure 2) to make it easier to observe which variable has the most distinctive value for each music genres.

![Figure 2](image)

Figure 2. Graphical representation of feature extraction result for all variables in average

From Figure 2 it shows that from 14 variables, only seven variables can be used as ANFIS variable, they are the dynamic range, fast peak strength, low beat dev, low, mid-RMS, mid beat dev, slow peak strength, and stereo RMS. All of these variables have the most distinct value for every genre, which make them ideal for classifying song genre.

3.2. ANFIS Training Result
Using seven variables from the previous stage, the ANFIS model will be constructed by training the ANFIS system first. The training parameters are as follow:

a. 7 ANFIS input variables Gaussian type, with 2 Membership Function for every variable
b. ANFIS output variable using linear type Membership Function
c. 160 training data (40 for each genre)
d. hybrid training method

The training result can be observed in Table 2.
Table 2. ANFIS training result

| No | Epoch Value | Training Error Value |
|----|-------------|----------------------|
| 1  | 20          | 0.063235             |
| 2  | 30          | 0.0533               |
| 3  | 40          | 0.046984             |
| 4  | 50          | 0.046984             |

From table 2 we can observe that the number of training value is stopped at 40 because from epoch 40 and above the value of training error is the same which is 0.046984 so it's pointless to continue training with higher epoch value.

3.3. ANFIS Result

After it was trained, the ANFIS model is completed, which means it can make a decision just like the human brain based on the knowledge it got from training. The data for this test is 15 songs for each genre. The purpose of this test is to measure the accuracy and performance of the created model, whether this model can be used to classify the music genre or not. The ANFIS testing result can be observed in Table 3.

Table 3. ANFIS testing result

| No | Song Genre | Correct | Wrong | Accuracy  |
|----|------------|---------|-------|-----------|
| 1  | Classic    | 9       | 6     | 60%       |
| 2  | Jazz       | 8       | 7     | 53,33%    |
| 3  | Pop        | 7       | 8     | 46,67%    |
| 4  | Rock       | 8       | 7     | 53,33%    |
|    | Total      | 32      | 28    | 53,33%    |

From table 3 it shows that the overall accuracy of the model is only 53.33%, which is below the expected value, which is 80%. Several factors cause the low accuracy of this model:

a. Variables used in ANFIS is too few
b. The number of membership function used is too few
c. Feature extraction system doesn’t work as intended
d. Low sampling frequency at downsampling process in pre-processing

3.4. Overfitting/Underfitting Analysis

Overfitting is a condition in machine learning where the system matches (fit) too close with training data, which leads to error when the input has a slight difference with training data. This error can make the ANFIS system unable to identify the correct data, which leads to lower accuracy. While underfitting is an error in machine learning where the system is unable to locate the input because of the lack of training data. Both can lead to poor ANFIS performance.

One solution for overfitting is reducing the number of training data. In this analysis, a new ANFIS model (model B) will be constructed using less training data to prove whether overfitting present or not. The ANFIS parameter for ANFIS model B are as follow:

a. 5 ANFIS input variables Gaussian type, with 3 Membership Function for every variable
b. ANFIS output variable using linear type Membership Function
c. 80 training data (20 for each genre)
d. hybrid training method

ANFIS model B testing result can be observed in table 4, together with model B as a comparison.
Table 4. ANFIS model A & model B testing result

| No | Song genre | Accuracy model A | Accuracy model B |
|----|------------|------------------|------------------|
| 1  | Classic    | 60%              | 80%              |
| 2  | Jazz       | 53.33%           | 33.33%           |
| 3  | Pop        | 46.67%           | 26.67%           |
| 4  | Rock       | 53.33%           | 0%               |
|    | Total      | 53.33%           | 35%              |

From table 4 it shows that ANFIS model B, which has lower training data, also resulted in lower accuracy. This fact proves that overfitting does not occur on ANFIS model A. Instead, underfitting occurs in both the ANFIS model, which resulted in low accuracy. However, it can be noted that ANFIS model B has particularly high accuracy for classifying a single genre which is a classic genre with accuracy 80%. This performance shows that ANFIS model B can be used correctly for classifying classic type.

4. Conclusion
Both the ANFIS model constructed in this paper can be used for classifying music genre despite its low accuracy, where model B, in particular, has high accuracy for classifying classic type. For improve the system's accuracy, the number of training data can be increased to reduce underfitting or by using better feature extraction system.

References
[1] J.E. Muñoz-Expósito, S. García-Galán, N. Ruiz-Reyes, P. Vera-Candeas, 2007, Adaptive network-based fuzzy inference system vs. other classification algorithms for warped LPC-based speech/music discrimination, Engineering Applications of Artificial Intelligence, Volume 20, Issue 6, Pp 783-793.
[2] Ikhsan, Imam, 2013, Simulasi dan Klasifikasi Genre Musik berbasis Hidden Markov model, Theses Report, Institut Teknologi Telkom.
[3] Brendan Petty, 2010, Music Genre Classification Using a Back Propagation Neural Network. https://www.brendanpetty.com/projects/music-genre-classificationproject/musicgenreclassification-report.pdf (Accessed at September 23, 2017)
[4] Yandre M.G. Costa, Luiz S. Oliveira, Carlos N. Silla, 2017, An evaluation of Convolutional Neural Networks for music classification using spectrograms, Applied Soft Computing, Volume 52, Pages 28-38.
[5] Antonio Jose Homsi Goulart, Rodrigo Capobianco Guido, Carlos Dias Maciel, 2012, Exploring different approaches for music genre classification, Egyptian Informatics Journal, Volume 13, Issue 2, Pages 59-63.
[6] Aisha Gemala Jondya, Bambang Heru Iswanto, 2017, Indonesian’s Traditional Music Clustering Based on Audio Features, Procedia Computer Science, Volume 116, Pages 174-181.
[7] Mariusz Kleć, Danjel Koržinek, 2014, Unsupervised Feature Pre-training of the Scattering Wavelet Transform for Musical Genre Recognition, Procedia Technology, Volume 18, Pages 133-139.
[8] Haggblade, Michael. Yang Hong, dan Kenny Kao. 2011. Music Genre Classification. https://www.semanticscholar.org (Accessed at September 24, 2017).
[9] Kusumadewi, Sri, Hartati, Sri, 2006, Neuro Fuzzy Integrasi Sistem Fuzzy dan Jaringan Syaraf, Graha Ilmu.
[10] Kumar, Satish, 2004, Neural Networks, A Classroom Approach. New Delhi: Tata McGraw-Hill Publishing Company Limited.