Jurnal Elektronik Ilmu Komputer Udayana p-ISSN: 2301-5373
Volume 8, No 3. February 2020 e-ISSN: 2654-5101

Voice Classification Based on Gender Using Backpropagation and K-Means Clustering Algorithm

Maula Khatami[1], I Ketut Gede Suhartana[2]
[12]Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Bukit Jimbaran-Bali, Indonesia
[1]Khatamimaula@gmail.com [2]Ikg.suhartana@unud.ac.id

Abstract

Sound is the identity of all living creature, including Humans. With voice we can do socialize, call people, ask questions, communicate, and even be able to help us to recognize the sex of the person who makes the sound. Nowadays, knowing gender through sound cannot only be done by humans but through a computer. Voice classification using a computer shows increasingly sophisticated technology. Of course this technological advance can also help in terms of security, where the voice can be a key or password in a certain confidentiality. In this study the focus of sound recordings is classified according to the sex of men and women by using the Backpropagation algorithm for training data, then Mel Frequency Cepstral Coefficients (MFCC) will process sound data and get features, and the K-Means Clustering algorithm will classify sound data already processed. The dataset used here is in the form of male and female voice recordings obtained from YouTube videos that have been separated by video sections. There are each 10 male and female voice files for training. As for testing, there are several male and female voice files that are placed in separate folders.

Keyword: *Voice, Classification, Backpropagation, K-Means Clustering, Gender*

1. Introduction

Sound is thing that can be heard, which has certain waves. sound also is mechanical compression or longitudinal waves that propagate through the medium. This medium or intermediate agent can be liquid, solid, gas. So, sound waves can propagate for example in water, coal, or air [4].

Sound is one of God's most useful gifts. Every day we always make noise for our daily activities. Through sound, we can communicate, express opinions, listen to the opinions of others and many other things. Of course sound is the most vital thing in our lives, imagine if we can't make a sound, we can't know the sound of other things too. Of course this will greatly hinder our activities, even our lives will be empty and die without a sound.

With so many positive things that can be produced by a sound, it can also be an advance in computer technology. By combining sound and technology, an innovation can be created to simplify our lives. By telephone, we can transfer the sound to another place without having to move to that place. Through a voice recorder, we can save our voice to be played next time, and more. Sound can also help us to improve our security / privacy. Like the password, the voice can also be used as a password to increase security making it more difficult to crack. But before we get there, the sound we make must certainly be detected by the computer of the sex. The basic thing, which really determines the progress of sound technology.

Therefore, this study aims to classify sounds by sex with a dataset of male and female voice recordings, with the Backpropagation method for training data and K-Means Clustering for the classification of data that has been processed. Then the results will be obtained namely the sex of the recorded sound.

2. Reseach Methods

The research method includes three methods, namely Backpropagation for dataset training, K-Means Clustering for feature extraction and Mel Frequency Cepstral Coefficients (MFCC) for processing voice data and obtaining features.

2.1. K-Means Clustering

K-means is a clustering algorithm for data mining that was created in the 70s and is useful for clustering in unsupervised learning (unsupervised learning) in a data set based on certain parameters. K-means is an algorithm for classifying or grouping objects (in this case data) based on certain parameters into a group, so that it can run faster than hierarchical clustering (if k is small) with large variables and produce denser clusters [2].

The K-Means algorithm is an algorithm that requires as many input parameters as k and divides a set of n objects into k clusters so that the level of similarity between members in one cluster is high whereas the level of similarity with members in other clusters is very low. The similarity of members to the cluster is measured by the proximity of the object to the mean value in the cluster or can be called a centroid cluster or center of mass [2].

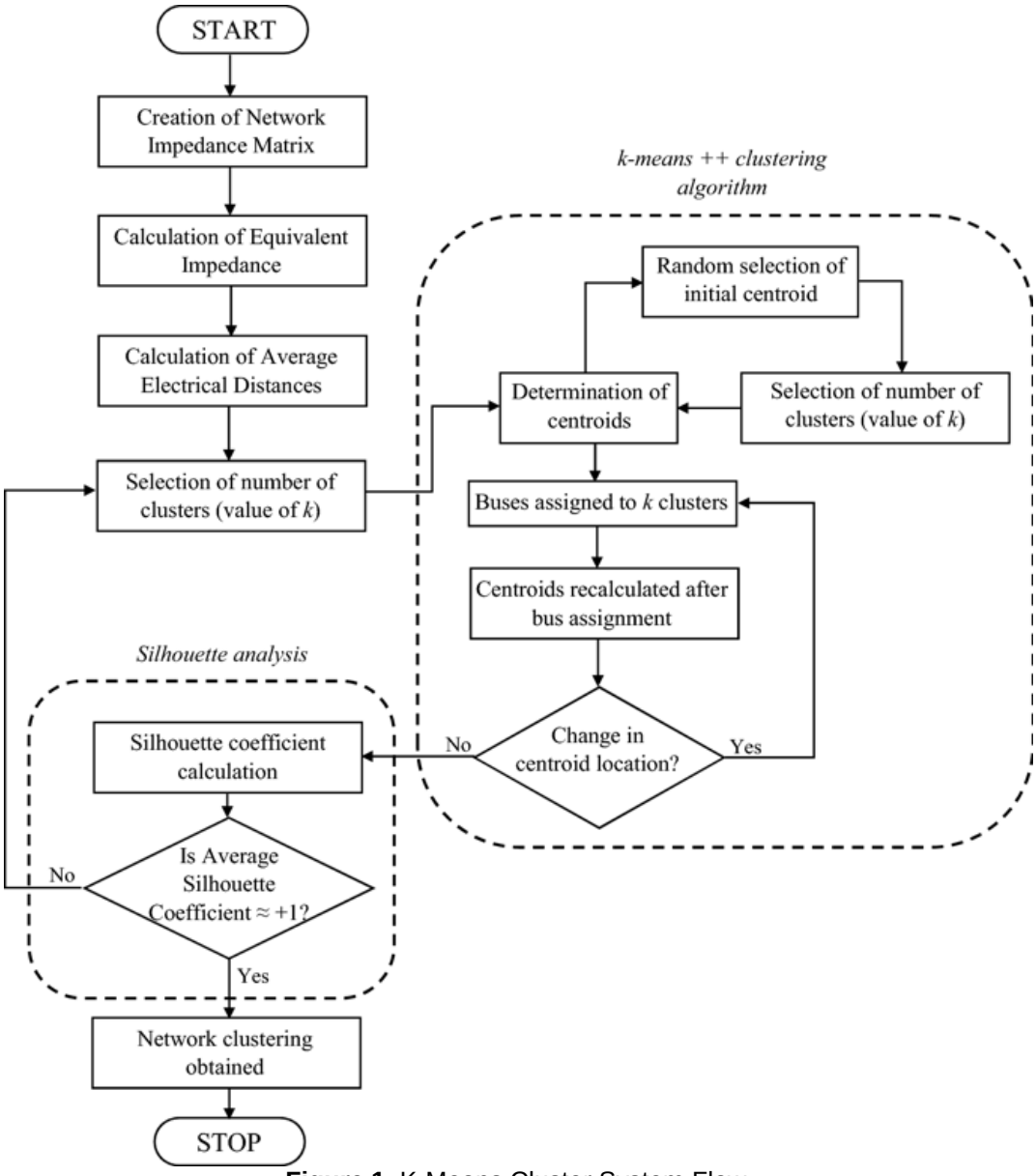


Figure 1. K-Means Cluster System Flow

The following is the equation for calculating the number of clusters:

$$k \approx \sqrt{n/2} \quad (1)$$

p-ISSN: 2301-5373
e-ISSN: 2654-5101

The following is the equation for calculating distance *Euclidean*:

$$d_{(x,y)} = \sqrt{(x_i - y_i)^2 + (x_i - y_i)^2} \quad (2)$$

The following is the equation for calculating new centroid:

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i \quad (3)$$

2.2. Backpropagation

Backpropagation is a supervised learning algorithm and is usually used by perceptron with many layers to change the weights associated with neurons in the hidden layer. The backpropagation algorithm uses error output to change the value of its weights in the backward direction. To get this error, the forward propagation stage must be done first [1].

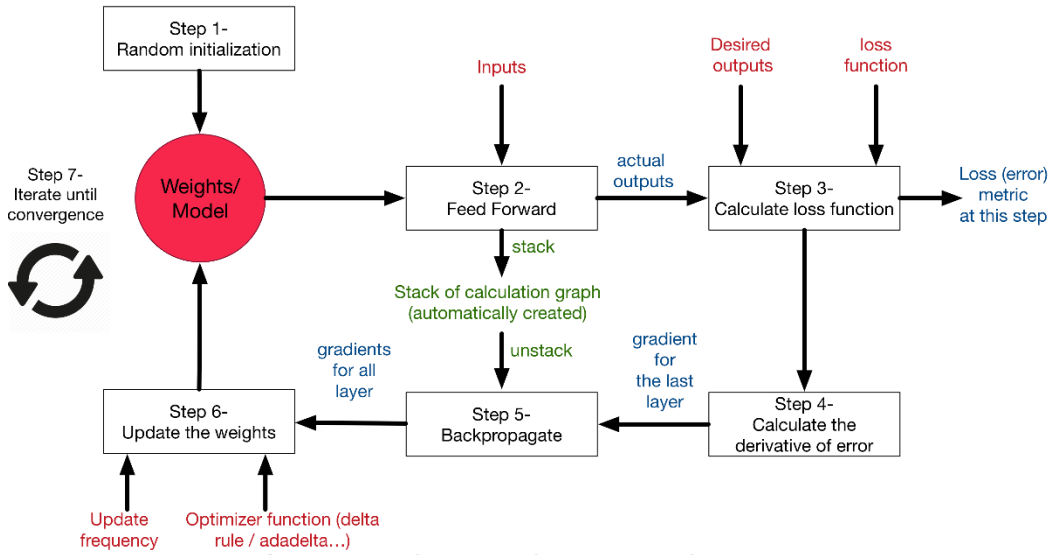


Figure 2. Backpropagation System Flow

Backpropagation network training algorithm:

1. Step 0 : Initialize all weights with small random numbers
2. Step 1 :While the condition is stop not fulfilled (wrong value), do following steps:
   a. For each training pair, do:
      Feedforward:
      1) Each input unit (X_i , i = 1,2,3,…,n) receive signal x_i and forward the following signal to entire unit in the upper layer (hidden layer).
      2) Each hidden unit (Z_j, j = 1,2,3,..,p) adding weight input signals:

$$Z\_in_j = v_{oj} + \sum_{i=1}^{n} x_i v_{ij} \quad (4)$$

Where v_0 =biased and v=weight
Use activation function to count the output signal:

$$Z_j = f(Z\_in_j) \quad (5)$$

And send following signal to entire unit in upper layer (output units)

      3) Each output unit (Y_k, k=1,2,3,…,m) adding weight input signal:

$$y\_in_k = w_{ok} + \sum_{j=1}^{p} z_j w_{jk} \quad (6)$$

Where w_0 =biased and v=weight
Use activation function to count output signal:

$$y_k = f(y\_in_k) \quad (7)$$

And send following signal to entire unit in the rest layer (output units).

   b. For each training pair, do:
      Backpropagation:

1. Each output unit (Y_k ,k = 1,2,3,…,m) receiving pattern target that relate to learning pattern input, count information of error

   $$\delta_k = (t_k - y_k) f'(y\_in_k) \quad (8)$$

   where t = output target
   Then, count weight correction (which is used to get value of $w_{jk,}$ later):

   $$\Delta w_{jk} = \alpha \delta_k z_j \quad (9)$$

   where α = learning rate
   Count correction biased as well (which is used to fix value of w_{0k}):

   $$w_{0k} = \alpha \delta_k \quad (10)$$

   Send $\delta\ l$ to existing units in lower layer
2. Each of hidden units (Z_j , j=1,2,3,…,p) adding delta's input (from upper layer units):

   $$\delta\_in_j = \sum_{k=1}^{m} \delta_k w_{jk} \quad (11)$$

   Multiply that value with descendant of activation function to count error information of error:

   $$\delta_j = \delta\_in_j f'(z\_in_j) \quad (12)$$

   Then, count weight correction (which is used to fix value of v_{ij}):

   $$\Delta v_{ij} = \alpha \delta_j x_i \quad (13)$$

   Count biased correction as well (which is used to fix value of v_{0j}):

   $$\Delta v_{oj} = \alpha \delta_j x_i \quad (14)$$

c. Update weight and biases:
   1. Each output units (Y_k , k=1,2,3,…m) fixing biases and weight (j = 0,1,2,…,p):

      $$w_{jk}(baru) = w_{jk}(lama) + \Delta w_{jk}(bobot) \quad (15)$$

      $$w_{0k}(baru) = w_{0k}(lama) + \Delta w_{0k}(bias) \quad (16)$$

   2. Each hidden units (Z_j , j=1,2,3,…p) fixing biases and weight (i= 0,1,2,…,n):

      $$v_{ij}(baru) = v_{ij}(lama) + \Delta v_{ij}(bobot) \quad (17)$$

      $$v_{0j}(baru) = v_{0j}(lama) + \Delta v_{0j}(bias) \quad (18)$$

d. Condition stop test.
   After running algorithm training of backpropagation network and got the output of closest target, then end weight and end biases of training result stored and then run testing process with testing algorithm. In testing algorithm that is used, just feedforward step only [1].

3. Result and Discussion

The process starts from the feature extraction that is processing the sound from the file directory, then training is conducted using the Backpropagation method, and finally testing is done to get the gender results of the tested sound.

3.1. Proses Feature Extraction

The sound processing process begins by retrieving the sound file from the directory indicated by the source variable and saved to the file variable. Repeat the number of files in the directory with the following process:

1. Get data from voice file
2. Process voice data with library to get feature
3. Run clustering on data in the feature
4. Stored result of clustering as dataset

p-ISSN: 2301-5373
e-ISSN: 2654-5101

After the above process is done, create a target array along the number of datasets containing 1 for men, and 0 for women.

3.2. Proses Training

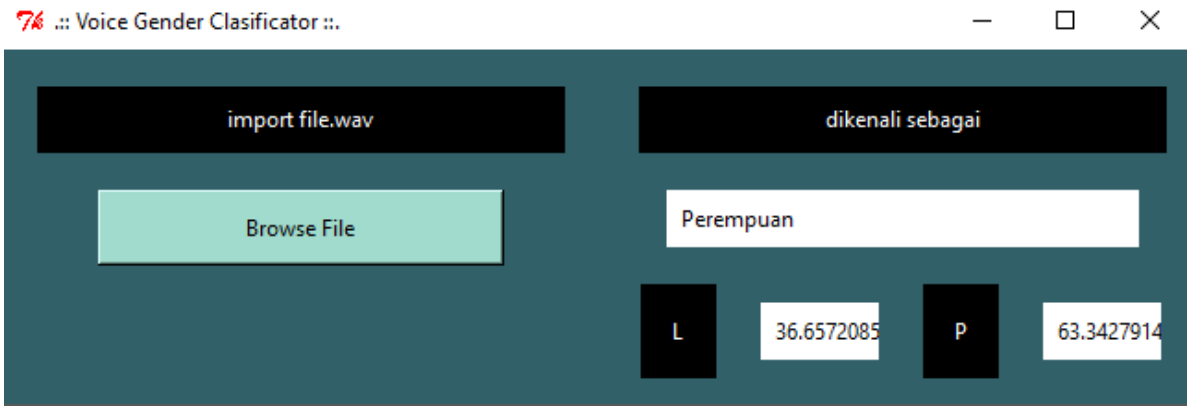
The first process for backpropagation training is to make the initiation weight of each neuron randomly. In this program, there are 3 different network types for processing datasets, namely gold, silver and bronze. Each network is composed of 1 input layer, 2 hidden layers, and 1 output layer. The gold network has a 5-4-3-1 architecture, while the silver and bronze networks have a 4-4-3-1 architecture. The purpose of each network is to process features in the dataset. The gold network will process features 1-5, silver for features 6-9, and bronze for features 10-13.

First of all the forward path will be run using the binary sigmoid activation function as the output of each neuron. After that, calculate the difference from the output obtained with the target and save it in the form of an array. Then run backwards to change the weight of each neuron in the network according to the difference between the target and the output. Perform this process on every existing network. If the average error of all networks has reached 0.05%, stop epoh and save the weight to the file weight of each network as a model

Figure 3. Dataset Training

3.3. Proses Testing

The testing process is similar to the training process but only runs forward. The weight of each network will be loaded from the training results model. The input is the sound file that we choose from the GUI that will first play the sound file, then extract its features, and enter



backpropagation for forward.

Figure 4. Female Voice Testing

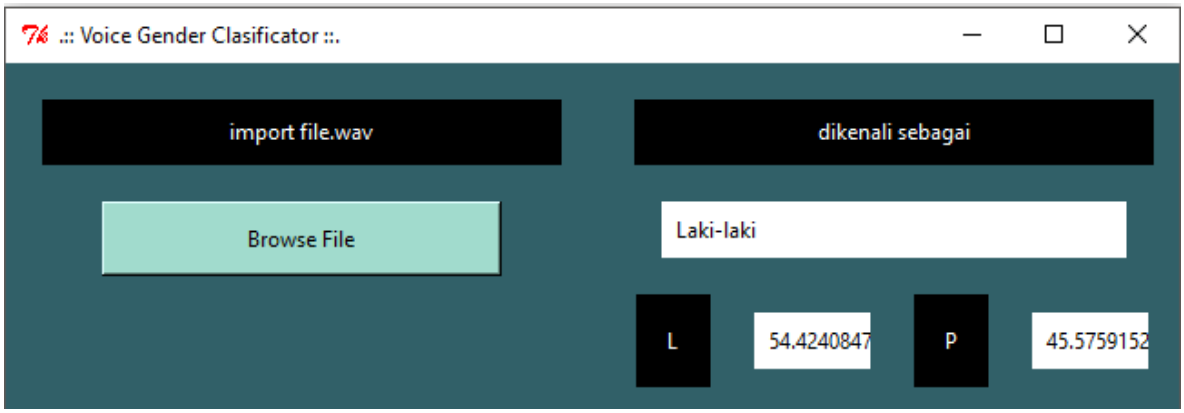


Figure 5. Male Voice Testing

The testing process was carried out with an experiment each of 10 male and female voices to see the results of the percentage of the sound classification based on that sex. Then the percentage of success will be obtained from each testing as presented in the table below.

Table 1. Testing Result of Male Voice

Attempt	Gender		Information
	Male	Female	
1	40,9	59,1	False
2	56,7	43,3	True
3	61,2	38,3	True
4	53,1	46,9	True
5	56,1	43,9	True
6	55,4	44,6	True
7	53,7	46,3	True
8	45,4	54,6	False
9	48,5	51,5	False
10	64,3	35,7	True

From the table above, after testing 10 male voice experiments, the results obtained are seven experiments that say that the voice is male and three say that it is female voice. So from these results obtained the percentage of success of the experiment is 70% and the average percentage of true is 57.2%.

Table 1. Testing Result of Female Voice

Attempt	Gender		Information
	Male	Female	
1	27,6	72,4	True
2	44,2	55,8	True
3	41,5	58,5	True
4	45,3	54,7	True
5	50,8	49,2	False
6	28,6	71,4	True
7	34,4	65,6	True
8	42,1	57,9	True
9	41,0	59,0	True
10	37,5	62,5	True

From the table above, after testing 10 female voice experiments, nine results were obtained that said that the voice was female and one said that it was male. So from these results

obtained the percentage of success of the experiment is 90% and the average percentage of true is 61.9%.

4. Conclusion

Based on the results of research and system testing that has been done to 10 men and 10 women, with training data for 10 datasets, each male and female. It can be concluded that the success rate of speaker recognition that can be recognized in male testing reaches 70% and women reach 90%. Whereas in average testing the truth of men reached 57.2% and women reached 61.9%. So it can be said that the Backpropagation method and K-Means Clustering are proven to be good enough to classify votes based on gender and able to strengthen the level of a security system with sound. Research failures can be caused by environmental conditions or circumstances of the speaker.

Some suggestions proposed to improve the performance of sound classification are the need to increase the number and type of features to increase accuracy and the need for additional features that are not limited to sound.

References

[1] D. Rahayu, R. C. Wihandika and R. S. Perdana, " Implementasi Metode Backpropagation Untuk Klasifikasi Kenaikan Harga Minyak Kelapa Sawit" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol.2, no.4, p.1547-1552, 2018.

[2] S. Desmanto, Irwan and R. Anggreni, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokkan Citra Digital Dengan Ekstraksi Fitur Warna RGB" *Jurnal STMIK GI MDP,* 2014.

[3] N. I. Youllia, Andriana and D. Permatasari, "Pengenalan Pembicara untuk Menentukan Gender Menggunakan Metode MFCC dan VQ" *MIND Journal,* vol.2, no.1, p.34-37, 2017.

[4] KhalidzaevitaD, "Brainly.co.id", 24 June 2016. [Online]. Available: https://brainly.co.id/tugas/5885508. [7 August 2019]