ROUGH SETS K-MEANS CLUSTERING AND BACKPROPAGATION ON OPTIMIZATION VOICE RECOGNITION

Ivan Eka Pratama[1], Ednawati Rainarli[2]

Informatics Engineering Study Program
[1,2] Faculty of Engineering and Computer Science, University Computer Indonesia
Jl. Dipati Ukur 114 Bandung
E-mail: ivan.e.pratama@gmail.com[1], irene_edna@yahoo.com[2]

ABSTRACT

Voice recognition or speaker recognition is a technique of recognizing sound based on its source. In this technique there is a process in which sound is verified based on the identity of someone speaking or in a foreign language called speaker verification. In this study, the implementation of Rough Sets K-Means and Backpropagation will be applied in the optimization of speech recognition cases. The feature extraction stage uses MFCC as a sound feature extraction that is useful for the speech recognition process. Rough Sets K-Means are used to optimize the results in the MFCC and Backpropagation filters used as classifications. In this study using set voice data that is in the .wav format is obtained from voice recording from 5 speakers. Based on the testing of K-Fold Cross Validation on the parameters used, the average accuracy is 25% and the selected words can be used in the introduction process.

Keywords : *Speaker Recognition, Backpropagation, Rough Sets K-Means*, Feature Extraction, MFCC.

1. INTRODUCTION

Voice recognition or speaker recognition is a technique of recognizing sound based on its source. In this technique there is a process in which sound is verified based on the identity of someone speaking or in a foreign language called speaker verification. In the ssata to do speech recognition, the voice data will be carried out in the feature extraction process to obtain information contained in the voice data. But the results of feature extraction are large, so it requires more process if it is directly used for the speaker recognition process. Therefore, the role of the clustering method is needed to create several central vectors as a representative of the entire vector of data that exists. The clustering method can also optimize the results of feature extraction but many clustering methods can be used to implement sound optimization cases with different results [1].

The clustering method used in this study is Rough K-Means Sets, the advantage of using the rough sets k-means clustering method is that it can minimize the amount of memory used for information analysis and reduce the calculations used to find similarities from spectral vectors [2]. In addition, there is also a comparative study of K-Means, K-Means Rough Sets in the same case, namely speech recognition. It was concluded that rough sets of k-means had better accuracy in the case of speech recognition with a percentage of 80% compared to K-Means which reached 75% [3]. Therefore Rough K-Means can be used for voice recognition optimization.

In this study the speech recognition process will be carried out using Backpropagation. Backpropagation is a supervised neural network [4]. From the previous research, namely the use of the backpropagation method in instrument sound recognition, it was found that the backpropagation method was the optimal classification in recognizing the sound of instruments based on the type of musical instrument [5]. And also another study, namely Backpropagation classification for face recognition cases, it can be concluded that the backpropagation method has an accuracy rate of 37.33% [6]. Therefore in this study the same classification will be carried out, namely the classification of backpropagation in the case of human speech recognition.

Based on the description above, in this study the implementation of K-Means Rough Sets and Backpropagation will be carried out in the optimization of speech recognition cases. The feature extraction stage uses MFCC as a sound feature extraction that is useful for the speech recognition process. Rough Sets K-Means are used to optimize the results in the MFCC and Backpropagation filters used as classifications.

2. THEORETICAL BASIS

2.1. MFCC

MFCC feature extraction method has 8 processes to be carried out, each process contained in the MFCC process flow can be seen in the Figure 1.

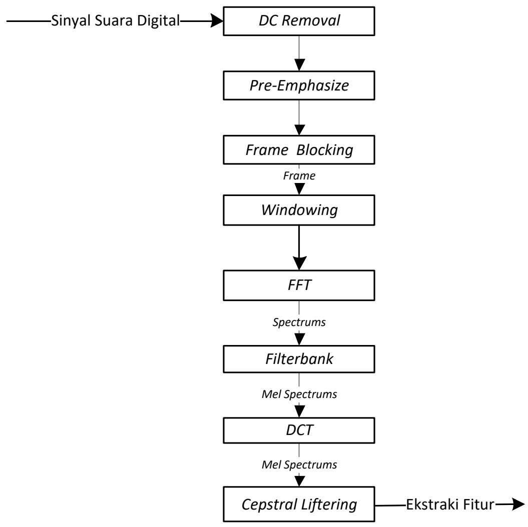

Figure 1. MFCC Flow

Based on Figure 1 the voice input data with the word chosen will enter the first process, namely DC Removal to get the normal value of the sound signal, then the Pre-emphasize process to reduce noise in the sound sample, then the Frame Blocking process to divide the signal sample into several frames, then the Windowing process to reduce the discontinuity effect at the ends of the frame that was previously generated by the frame blocking process, then the FFT process obtained a signal sample in the domain frequency, then filter the bank process to get the energy value of each band frequency, then the DFT process to produce a value mel cepstrums, the last step of the Cepstral Liftering process, aims to smooth the signal spectrum [7].

2.2. Rough Sets-K-Means Clustering

Rough Sets-K-Means algorithm is the use of the theory of rough sets and k-means algorithms to deal with the uncertainties involved in cluster analysis. In rough clustering each cluster has two assessments (approximations), namely lower approximation (low estimation value) and upper approximation (high estimation value). Lower approximation is part of the upper approximation. Members in the lower approximation have ascertained belonging to a cluster, therefore they cannot enter other clusters. While the object data on the upper approximation has not been ascertained in the cluster or it may be in another cluster. Because the membership of the upper approximation has not been confirmed, at least they will enter another cluster. [8]

The steps in the Rough Sets-K-Means clustering method are as follows:

1. Select the initial cluster of n objects in the cluster k
2. Calculate Centroid C_j in a cluster, with the following conditions.

   $If \quad \underline{U}\,(K) \neq \emptyset \; and \; \overline{U}(K) - \underline{U}(K) = \emptyset$

   $$C_j = \sum_{x\in\underline{U}(K)} \frac{x_i}{|\underline{U}(K)|} \qquad (1)$$

   $Else \quad \underline{U}\,(K) = \emptyset \; and \; \overline{U}(K) - \underline{U}(K) \neq \emptyset$

   $$C_j = \sum_{x\in\underline{U}(K)-\underline{U}(K)} \frac{x_i}{|\overline{U}(K)-\underline{U}(K)|} \qquad (2)$$

   $Else$

   $$C_j = W_{lower} \times \sum_{x\in\underline{U}(K)} \frac{x_i}{|\underline{U}(K)|} + W_{upper} \times \sum_{x\in\underline{U}(K)-\underline{U}(K)} \frac{x_i}{|\overline{U}(K)-\underline{U}(K)|} \qquad (3)$$

3. Set each object in the lower limit $\underline{U}(K)$ or upper limit $\overline{U}(K)$ in the respective cluster i. For each vector x object, set $d(X, C_j)$ to be the distance between itself and the centroid on the C_j cluster.

   $$d(X, C_j) = min_{1\leq j\leq K} d(X, C_j) \qquad (4)$$

   Ratio $d(X, C\_i\,) \,/\, d(X, C\_j\,), 1 \leq j \leq K$ used to determine membership of x is as follows $d(X, C_i) \,/\, d(X, C_j) \leq \epsilon$, to (i, j) set, to $x\epsilon\overline{U}(C_i)$ and $x\epsilon\overline{U}(C_j)$ and x will not be a part of each lower approximation. If not, $x\epsilon\overline{U}(C_i)$, like $d(X, C_i)$ is a minimum value of $1 \leq j \leq K$. In addition to $x\epsilon\overline{U}(C_i)$.
4. If convergent criteria are met, that is, the center of the cluster that is the same as the one that was previously iterated, then the iteration will stop, if not then repeat steps 2 and 3.

2.3. Backpropagation

Backpropagation trains a network to recognize patterns used in the training process and then assigns correct values to input patterns that are similar to the patterns used during training [9].

The backpropagation training process has similarities to other neural networks. In the feed forward network, the training process is carried out as a weight calculation so that at the end of the training it will produce optimal weights. During the training process, weights are arranged iteratively to minimize errors. Errors are calculated from the average squared error. Furthermore, the average squared error will be used as the basis for the calculation for the activation function. [10]

The backpropagation algorithm is used to correct network weights which make the activation function decrease. There are 3 phases of backpropagation training by Siang [9] including:

1. Forward Propagation.

In advanced propagation, each input signal is calculated forward to the hidden layer until the output layer uses the specified activation function.

2. Backward Propagation.

The error value (the difference between the service output and the target) is counted backwards starting from the line that corresponds directly to the units in the output layer.

3. Change in weight.

Here weight changes are made to reduce errors that occur. These three phases are carried out repeatedly until they reach a state of cessation.

Training algorithms for networks with one hidden layer (using binary sigmoid activation functions) are as follows.

1. Initialize all weights with random numbers of small value.
2. If the termination condition is not fulfilled, take steps form 3 to 11
3. In each training data do steps 4 to 10

   I : *Forward Propagation*
4. Each input unit receives a signal and is forwarded to the hidden layer
5. Calculate all outpu in hidden layer

   z_j (j = 1,2,…,p)

   $$z\_in_j = v_{j0} + \sum_{t=1}^{n} x_i v_{ji} \quad (5)$$

   $$z_j = f\left(z_{in_j}\right) = \frac{1}{1+e^{-z\_in_j}} \quad (6)$$
6. Calculate all outpu in output layer

   y_k (k = 1,2,…,m)

   $$y\_in_k = w_{k0} + \sum_{j=1}^{n} z_i w_{ji} \quad (7)$$

   $$y_j = f\left(y_{in_k}\right) = \frac{1}{1+e^{-y\_in_k}} \quad (8)$$

   II : B*ackward Propagation*
7. Calculate the output unit factor based on an error in each output unit

   y_k (k= 1,2,…,m)

   $$\delta_k = (t_k - y_k) f'\left(y_{in_k}\right)$$
   $$= (t_k - y_k) y_k (1 - y_k) \quad (9)$$

   δ_k is a unit error that will be used in changing the weight of the layer below it (step 8)
8. Calculate the change rate of weight w_{kj} (which will be used later to change the weight w_{kj}) with acceleration rate α ;

   $$\Delta w_{kj} = \alpha \delta_k z_j \quad (10)$$
9. Calculate the hidden layer factor based on errors in each hidden unit

   z_j (j = 1,2,…,p)

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

   Factor δ hidden unit :

   $$\delta_j = \delta\_in_j f'\left(z_{in_j}\right) = \delta_{in_j} z_j (1 - z_j) \quad (12)$$
10. Calculate the rate of change weight v_{ji} (which will be used later to change the weight v_{ji})

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

    III : Weight changing
11. Calculate all weight changes and change the weight of line leading to the output layer:

    $$w_{kj}(new) = w_{kj}(old) + \Delta w_{kj} \quad (14)$$

    Line weight changes leading to the hidden layer:

    $$v_{ji}(new) = v_{ji}(old) + \Delta v_{ji} \quad (15)$$

After the training process is done, the network can be used to recognize patterns. In the process of recognizing, only forward propagation is carried out to determine the output.

The training process will produce good values if the optimal initial weight selection is done, because the initial weight will greatly affect how fast the convergence is. Therefore in the standard backpropagation, weights and biases are filled with small random numbers and usually the initial weights are initialized randomly with values between -1 and 1 or other intervals.

3. RESEARCH METHODS

The research method used in this study was experimental research. The flow of research can be seen in Figure 2

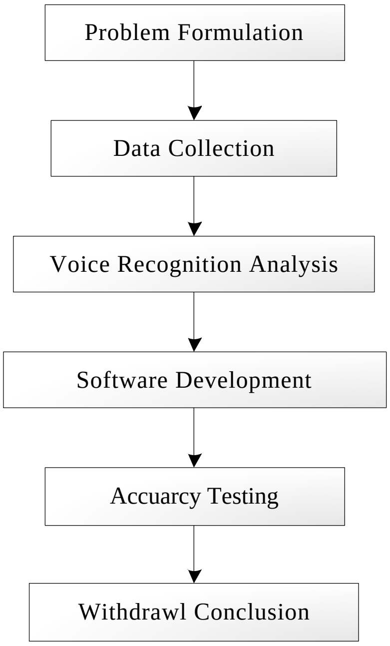

Figure 2. Research Methods Flow

Based on the research flow in Figure 2 can be explained as follows. At the problem formulation stage, the researcher formulated the problem of how to implement the rough sets-k-means clustering method and the classification of backpropagation in speech recognition cases. At the stage of data collection carried out in this study, namely, literature review and collection of datasets. At the stage of voice recognition analysis carried out in the research include:

1. MFCC Feature Extraction

The MFCC extraction stage is the initial stage where the input sound signal will be processed with various processes in it so that the voice signal is ready for the clustering process. The process at this stage includes DC Removal, Pre-Emphasize, Frame Blocking, Windowing, Fast Fourier Transform, Filter bank, Discrete Cosine Transform, Cepstral Liftering.

2. Rough Sets-K-Means Clustering

In the Clustering function, the input value is actually a result of the feature extraction. There is a process in the Clustering function to get the central vector. Central vector search is done repeatedly so that the central vector is represented by the entire vector of feature extraction results.

3. Backpropagation Classification

To be able to do sound matching, the new data entered will be synchronized with the data that has been in the previous database. Each data is compared with everything in one database model alternately.

The software development model used is the prototype model. In the accuracy testing phase, the results of the introduction to rough sets-k-means will use the K-Fold Cross Validation method. This final stage is to draw conclusions about applications that have been built.

4. RESULT AND DISCUSSION

Dalam pembangunan sistem pengenalan suara pada penelitian ini meliputi proses data latih (data training) dan proses data pengujian (data testing). Dalam proses yang dilakukan ada beberapa metode yang yang diterapkan yaitu, fitur ekstraksi menggunakan metode Mel Frequency Cepstrums Coefficients (MFCC), kemudian *Rough Sets K-Means Clustering* selanjutnya dilakukan klasifikasi *Backpropagation.* Alur proses pengenalan suara dapat dilihat pada Figure 3.

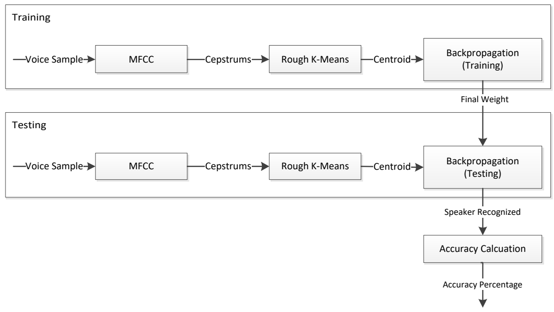

Figure 3. Voice Recognition Flow

In this study the test will use K-Folds Cross Validation, the number of datasets used is 100 data. Each dataset consists of 20 data consisting of 5 speakers speaking 5 different words in Indonesian namely "ada", "adalah", "bisa", "jadi" and "sebut". The test uses a dataset that will be divided into 5 parts with test iterations done 5 times where each dataset will alternately become training data and test data.

Based on the testing scenario using 5-Fold Cross Validation in the backpropagation algorithm by using the hidden layer parameter = 30 then using other parameters by experimenting on different values, the results of the accuracy are as follows.

1) Testing the change in the number of clusters

The first test was conducted to see how much influence the cluster number had on speech recognition. The following are the results of accuracy testing based on changing the number of clusters can be seen in Table 1.

Table 1. Accuracy Test Results Based on Cluster Number

Cluster Size	K=3	K=4	K=5	K=6	K=7
Accuracy	18%	20%	25%	22%	22%

Based on the test results in Table 1, the value of accuracy with cluster 3 is 18% and then by trying to replace the number of clusters with a multilevel number of 4 and obtained an accuracy value of 20% and so on increasing with the number 5 obtained a value of 25%. Then after the number of clusters increases to 6 the value of accuracy decreases to 22% and it is also found that the same accuracy value in cluster 7 is 22%. The greatest accuracy was obtained with a value of 25% using the cluster number K = 5.

2) Testing of epoch change

This test uses the parameters learning rate = 0.01, and epoch 100, 500 and 1000. The value of accuracy can be seen in table Table 2.

Table 2. Results of the Accuracy of the Epoch Trial

N	*epoch*	*Learning rate*	Accuracy	
			Non	*Clustering*
1	100	0.01	50%	20%
2	500	0.01	62%	22%
3	1000	0.01	65%	25%

Based on the accuracy value in Table 2 there are three values of accuracy without using clustering and using clustering, from both using the same learning rate value that is 0.01. From the results of the three tests based on the change in the epoch value, it can be seen that the accuracy obtained increases along with the epoch value increase. With the value of epoch 100, the value of accuracy is 50% without using clustering and 20% using clustering. Then at the epoch 500 value there is an increase so that the accuracy value of 62% and 22% is obtained. In the last test with Epoch 1000, the accuracy values of 65% and 25% were obtained. The highest accuracy value is the third test with epoch 1000 value which will be carried out in the next test, namely the learning rate replacement test.

3) Testing of learning rate change.

This test uses the optimal paremeter epoch on the previous test phase, which is 1000 and learning rates 0.01, 0.001 and 0.0001. The value of accuracy can be seen in table Table 3.

Table 3. Testing Results of Learning Rate Change

N	*epoch*	*Learning rate*	Accuarcy	
			Non	*Clustering*
1	1000	0.01	65%	25%
2	1000	0.001	67%	30%
3	1000	0.0001	62%	22%

Based on the accuracy value in Table 3 there are three values of accuracy without using clustering and using clustering, from both using the same epoch value that is 1000. From the results of the three tests based on the change in learning rate values can be seen that the accuracy obtained increases with learning value rate. With the learning rate value of 0.01 obtained the value of 65% accuracy without the use of clustering and 25% by using clustering. Then at the learning rate of 0.001 there is an increase so that the accuracy value of 60% and 22% is obtained. In the last test with a learning rate of 0.0001 there was a decrease in the value of accuracy to 67% and 30%. The highest accuracy value is the third test with a learning rate value of 0.001.

From the overall accuracy testing performed, the optimal parameter is epoch 1000, learning rate 0.001 if not using clustering and learning rate 0.0001 in the use of clustering.

5. Conclusion

Based on the results of testing in this study, it can be concluded that the K-Means Rough Set clustering method cannot be used as an optimal method that can improve the speech recognition process with the Backpropagation method as a classification algorithm by utilizing MFCC feature extraction results. In the testing phase we get hidden layer parameters = 30, epoch 1000, learning rate 0.001 if we do not use clustering and learning rate 0.0001 in using clustering, with an average value of accuracy of 25% using cluster K = 5.

The suggestion for future research is to use a better clustering method to minimize the number of features without reducing the accuracy of the classification. In the pre-process process it is expected to add a better method that can overcome the noise problem so that it can affect the success rate of speech recognition. Besides that, it can also use a more optimal feature extraction and classification method to get accurate recognition results. In order to get better accuracy results can be done to try to use other variations on the number of parameters to be used such as the number of hidden layers, learning rates, errors that will affect the results of accuracy.

References

[1] H. S. Manunggal, "Perancangan dan pembuatan perangkat lunak pengenalan suara pembicara dengan menggunakan analisa MFCC feature extraction," *Bachelor thesis, Petra Christian University.,* vol. I, 2005.

[2] M. Ghassemian and K. Strange, Speaker Identification Features, Models and Robustness, Denmark: Kongens Lyngby, 2009.

[3] G. B. Djamal, "Perbandingan Metode Ekstraksi Ciri Rough Set-K-Means Dan K-Means Pada Optimasi Kasus Pengenaan Suara," 2017.

[4] K. Setiawan, Paradigma Sistem Cerdas, Surabaya: Bayumedia, 2003.

[5] Gunawan, A. D. Gunawan and S. N. Soenardjo, "Penerapan algoritma backpropagation untuk klasifikasi musik dengan solo instrumen," *Seminar Nasional Aplikasi Teknologi Informasi 2009 (SNATI 2009),* 2009.

[6] M. D. Wuryandari and I. Afrianto, "Perbandingan Metode Jaringan Syaraf Tiruan Backpropagation Dan Learning Vector Quantization Pada Pengenalan Wajah," *Jurnal Komputer dan Informatika (KOMPUTA),* vol. I, Maret 2012.

[7] A. Resmawan and D. Putra, "Verifikasi Biometrika Suara Menggunakan MFCC dan DTW," *LONTAR KOMPUTER,* vol. II, 2011.

[8] G. Peters, "Some refinements of rough k-means clustering," Pattern Recognition 39.

[9] J. J. Siang, Jaringan Syaraf Tiruan dan Pemogramannya Menggunakan MATLAB, Yogyakarta: Andi Offset, 2005.

[10] S. Kusumadewi, Membangun Jaringan Syaraf Tiruan dengan Matlab Dan Excel Link, Yogyakarta: Graha Ilmu, 2004.