

Identification Of Number Using Artificial Neural Network Backpropagation

Sitti Amalia

Institut Teknologi Padang, Electrical Engineering Department, Padang, West Sumatra, Indonesia

Abstract. This research proposed to design and implementation system of voice pattern recognition in the form of numbers with offline pronunciation. Artificial intelligent with backpropagation algorithm used on the simulation test. The test has been done to 100 voice files which got from 10 person voices for 10 different numbers. The words are consisting of number 0 to 9. The trial has been done with artificial neural network parameters such as tolerance value and the sum of a neuron. The best result is shown at tolerance value varied and a sum of the neuron is fixed. The percentage of this network training with optimal architecture and network parameter for each training data and new data are 82,2% and 53,3%. Therefore if tolerance value is fixed and a sum of neuron varied gave 82,2% for training data and 54,4% for new data

1 Introduction

Technology developed to create tools or facilities that can help and provide convenience for humans to perform activities in their life. Along with the development of technology, people always want to improve the quality and practicality of these tools. Therefore, there are machines that can interact with humans. This technology is called human machine technology. ^[1]

Human machine technology aims to create machines that have the ability to interpret information spoken by humans and act in accordance with the information submitted. In other words, create a machine with artificial intelligence so that it can interact with humans through sound. ^[2]

The human auditory system composed of the ear, nerves, and brain has the ability to distinguish one sound from another. This ability can be possessed by every human being if the human ever compares and senses the difference between one voice and another. The process of 'comparing and feeling' is part of the learning process in humans ^[2]. Sound processing techniques (speech processing) with sound recognition system (speaker recognition) developed based on the principles of this human hearing system. This system is used to recognize incoming sounds by comparing them with previously stored sounds on the system memory. This system makes it possible to use sound as a person's identification/proof ^{[2] [3] [4]}.

The ability of a human hearing system to recognize sounds can be imitated to be implemented on machines, robots, security systems, etc. since its supporting components are available, ie microphone as an ear substitute, signal processing algorithms and

Artificial Neural Networks ANN) that will perform the function as the brain.

Artificial Neural Networks are widely used in speech recognition because they can recognize well-tested sound patterns^{[4] [5]}. This Artificial Neural Network is used because it has the ability to recognize patterns that are slightly changed from the original. This is similar to a human brain that is still able to recognize someone from a slightly changed voice over the phone. Artificial Neural Network is a model of computational system that works like a biological neural system, where Artificial Neural Network will serve as a substitute for the nerves and brain, which at the time will relate to the outside world, learning ability and generalization quickly and easily in recognition of a character pattern and easy to be implemented ^[5]. The artificial hearing system can be designed in such a way by mimicking some of the working principles of the human hearing system so that later the system can recognize sounds based on the sound waves it receives.

This network can recognize patterns of input from weights that are organized through a training. For this training process, the thing to note is to determine the right learning algorithm to get the weight.

Judging from the principle of backpropagation work (Backpropagation), it is seen that this algorithm can minimize errors may be close to or zero ^{[4] [7]}. So this study used Artificial Neural Network Backpropagation. Quite a lot of research and writing takes on topics related to speech recognition.

*Corresponding author: sitti@itp.ac.id

2 Literature Review

2.1. The Process Of Shaping The Human Voice Signals [7]

The process of speaking can occur if humans are able to issue an acoustic signal that we call also with sound. The human voice is produced by the interaction between the signaling organs.

The sound produced by the sound organs is generated due to the air pressure generated from the lungs which are then channeled to the windpipe. The air flowing through the windpipe will enter the larynx where there are vocal cords so that when air passes it vibrates so that it will produce air pulses. These air pulses will enter the vocal tract. In this section, resonance occurs air pulses which then form the sound patterns that produce voice signals. Sound signals generated from sound producing organs will be represented in the form of sound waves. Sound waves will form different types of patterns.

2.2 Definition of Artificial Neural Network[5]

Artificial neural networks are defined as information processing systems that have characteristics resembling human neural networks. Artificial neural networks are created as a generalization of the mathematical model of human understanding based on an assumption as follows:

1. Information processing that occurs in a simple element called a neuron.
2. The signal flows between nerve cells/neurons through a connecting link.
3. Each connecting connection has corresponding weights. This weight is used to multiply/multiply the signals sent through it.
4. Each nerve cell will apply an activation function to the weighted summation signal coming to it to determine its output signal.

Another definition of artificial neural networks is a biologically inspired computational model consisting of processing elements (called neurons) and their interconnected couples with the boundaries of coefficients (weights) to the connectors, which are neuronal structures, training, and recall algorithms which are put into structure [11].

Backpropagation network is one algorithm that is often used in solving complicated problems. This is possible because networks with these algorithms are trained using guided learning methods. In the network is given a pair of patterns consisting of the pattern of input and the desired pattern. When a pattern is given to the network, the weights are altered to minimize the

difference in the pattern of output and the desired pattern. this exercise is done repeatedly so that all patterns issued by the network can meet the desired pattern.

The backpropagation neural network training algorithm consists of two steps: forward propagation and backward propagation is performed on the network for each pattern given during the training network.

2.2.1 Arsitektur Backpropagation

Backpropagation has several units in one or more hidden screens. Fig.1 is a backpropagation architecture with n input (plus a bias), a hidden screen consisting of p units (plus a bias), as well as m units of output units.

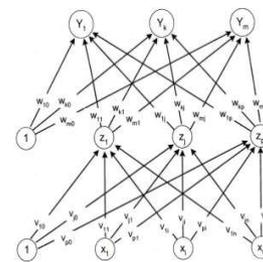


Fig. 1: Backpropagation Architecture

V_{ji} is the line weight of the input unit x_i to the hidden display unit Z_j (V_{j0} is the line weight that links the bias in the input unit to the hidden screen unit Z_j).

W_{kj} is the weight of the hidden display unit Z_j to the Y_k output unit (W_{k0} is the weight of the bias on the hidden screen to the Z_k output unit).

2.2.2 Activation Function

In backpropagation, the activation function used must meet several conditions, namely: continuous, differentiated easily and is a function that does not go down. One function that meets these three requirements so often used is a binary sigmoid function that has a range (0.1).

$$f(x) = \frac{1}{1 + e^{-x}}$$

with derivatives:

$$f'(x) = f(x)(1 - f(x))$$

Graph function is shown in Figure 2

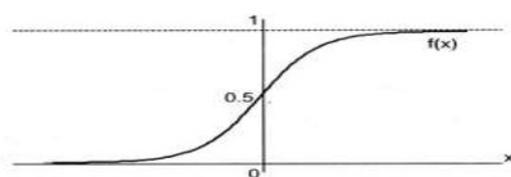


Fig. 2: Graphics sigmoid binary function

Another commonly used function is a bipolar sigmoid function that resembles a binary sigmoid function, but with a range of (-1,1).

$$f(x) = \frac{2}{1 + e^{-x}} - 1 \quad \text{pers(2.13)}$$

with derivatives:

$$f'(x) = \frac{(1 + (f(x)))(1 - (f(x)))}{2} \quad \text{pers(2.14)}$$

Graph function is shown in Fig. 2

The sigmoid function has a maximum value of = 1. So for the target pattern 1, the input and output patterns must first be transformed so that all the patterns have the same range as the sigmoid function used.

Another alternative is to use the sigmoid activation function only on a screen that is not an output screen. In the output screen, the activation function used is the identity function:

$$f(x) = x$$

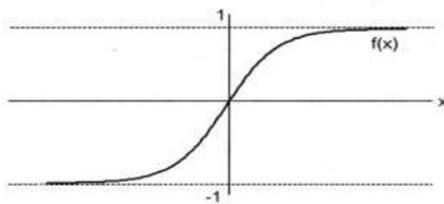


Fig. 3 : Graphics bipolar sigmoid function

2.3 The advantage of artificial intelligence ^[5]

Artificial neural networks have a large number of advantages compared with other calculation methods, namely:

- The ability to acquire knowledge even under conditions of disturbance and uncertainty. This is because neural networks are capable of generalizing, abstraction, and extraction of statistical properties of data.
- Ability to represent knowledge flexibly. Artificial neural networks can create their own representations through self-regulation or learning ability (self-organizing).
- Ability to tolerate a distortion (error/fault), in which a small disturbance in the data can be regarded as mere noise.
- The ability to process knowledge efficiently because it uses a parallel system, so the time required to operate it becomes shorter..

2.4 Training with and Without Supervision^[5]

Based on how to modify the weight, there are 2 kinds of training that are known by the supervision (supervised) and without supervision (unsupervised). Training methods with supervision are used if the expected output is known previously. There are a number of existing data pairs (input-output targets) used to train the network until the desired weights are obtained. The data pair serves as a "teacher" to train the network to obtain the best possible form. "The teacher will provide clear information about how the system should change itself to improve its performance. In contrast, an unsupervised learning exercise does not require an output target or no "teacher" will direct the training process.

During the training process, the weighting value is arranged within a given range depending on the value of the given input. The purpose of this training is to group similar units within a particular area. Such training is usually very suitable for grouping (pattern classification).

2.5 Research Methodology.

In the first stage of data, retrieval is done by vocal recording 10 people, and each person pronouncing digit 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9 are stored in one file with WAV format.

- The second stage has extracted a feature of the recording data that has been stored in the file with WAV format. This feature extraction is performed digital screening of each speech data by using MFCC method. The result of the leaching obtained is spectral frequency, then it is domiciled so that the data form is a matrix.
- The third stage is the process of building a neural network. The network model is a backpropagation network with guided learning system.
- The fourth stage is a process of classification training on patterns of extraction of traits. Training is done by guided by correlating pattern of input with a goal (target). The training result is a weighted matrix that is stored to be called during the test process.
- The fifth stage, namely testing of the ability of network systems that have been trained, in the testing process is used weights that have been obtained on the training results. As a network input used data that has been stored that has been characteristic extraction.

2.6 System Planning

Signal inputted to the computer via a sound card using a microphone. This sound signal was originally analog signals. These analog signals are then converted into digital form by using the ADC found on the sound card. The incoming sound signal is then processed to obtain the parameters. At first, the sound signal is clipped with a cuplik frequency that satisfies Nyquist criteria. Then the signal is blocked into frames. Signals that are still in the form of the frame are filtered by using a pre-emphasis which then carried out the windowing process. After that extraction parameter on MFCC to get cepstrum parameters. the resulting MFCC matrix will be the input on the artificial neural network. The process of this artificial neural network consists of two stages, namely the process of training and testing process.

In the training process will be obtained a result parameter. The results of this training will be stored as a reference model, where this pattern will be used in the testing process. After the training process is complete and each pattern has been obtained then the process will proceed with testing and decision making. The testing process is performed to test the ability of neural networks in recognizing each sound. This process is done by comparing the output of the sound being tested with the previously stored reference model. The system will correctly identify the sound when the sound is tested according to the trained person.

2.6.1 Training of Artificial Neural Network Backpropagation

The method used in this research is artificial neural network backpropagation. The structure of artificial neural networks used in the study are as follows:

- a) Training using 100 training data.
- b) There are 19 neurons in the input layer. This is matched to the number of matrix elements generated on the MFCC character extraction
- c) There are 2 hidden layers with 35 neurons and 2 neurons.
- d) A number of output neurons of 4 neurons, determined by the following considerations:
 - i. The words we will recognize are the numbers 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9 as 10 character numbers.
 - ii. Each character will be converted into a bit encoding. The number of possible bits is 4 bits, which produces 24 different output variations.

2.6.1.a. Training

Since the artificial neural network used in this research is backpropagation, it first needs to build a network consisting of input layer, hidden layer, output layer, activation function at hidden layer and output layer, and training model used on backpropagation neural network using function :

```
net = newff (minmax (P), [35 2 4], {'logsig' 'logsig'
                                'logsig'}, 'trainrp')
```

After finished building the network, the next step is to determine the parameters used during the training process, with the function:

```
net.trainParam.epochs = 3000;
net.trainParam.goal    = 0.06;
net.trainParam.lr      = 0.1;
net.trainParam.mc      = 0.7;
net.trainParam.show    = 250;
```

After completion of forming the network and determining the parameters to be used in the training process, hereinafter is the training to be performed with the function:

```
net = init (net)
[net, tr] = train (net, P, T)
```

With:

Init =	function to initialize the weights and biases with new random values
Train =	function to train the network
P =	input data matrix
T =	target

Training will be carried out until the tolerance has been determined that is 0.006.

2.6.1.b. Testing

The test function used in backpropagation neural networks is:

$$A = \text{sim} (\text{net}, P)$$

At this stage of testing will be known to what extent the system can recognize the pattern of letters, against the pattern that has been trained.

2.7 Result and Analysis

2.7.1 Experiment Results

2.7.1.a. System Testing Trial with Number of Neurons in the First Hidden Layer 35 and Tolerance Value Vary from 0.05 to 0.07.

In this experiment aims to look at system capabilities when the parameters of the number of neurons in the hidden layer are constant at 35 and the tolerance values are varied from 0.05 to 0.07. The ability of this system is measured by the success rate in recognizing the entered word pattern. The results of this experiment can be seen in Appendix B and in Figure 5.1 below

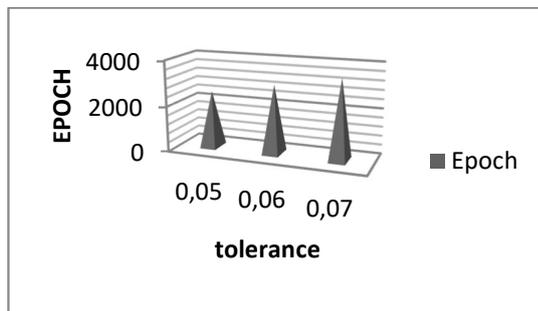


Fig. 4 : Graph of tolerance relationship to epoch

After the experiment, it is seen that the smallest tolerance of 0.05 is obtained at epoch 2500. When the tolerance value is enlarged then the number of an epoch is getting bigger, that is tolerance 0.06 epoch 3000, tolerance 0.07 epoch 3500. The graph of success rate can be seen in Figure 5.2.

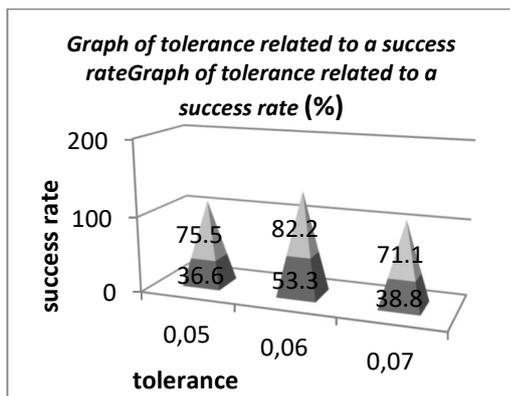


Fig. 5 : Graph of tolerance related to a success rate

Figure 5.2 above is a graph of tolerance related to the success rate of training data and new data (not trained). After the experiment, it was found that the greatest success rate of 82.2% for training data and 53.3% for new data was obtained at tolerance 0.06. When the tolerance value is magnified the smaller success rate is obtained, the tolerance 0.07 has a 71.1% success rate

for training data and 38.8% for new data and tolerance 0.05 has 75.5% success rate for training data and 36.6% for new data.

2.7.1.b. System Testing Tests with Tolerance Value 0.06 and Number of Neurons in the First Hidden Layer Varies from 25 to 40.

In this experiment, it is aimed to see system capability when the parameter of constant tolerance value is 0.06 and the number of neurons in the hidden layer is varied from 30 to 40. The ability of this system is measured by the success rate in recognizing the entered word. The results of this experiment can be seen in Appendix B and in Figure 5.3.

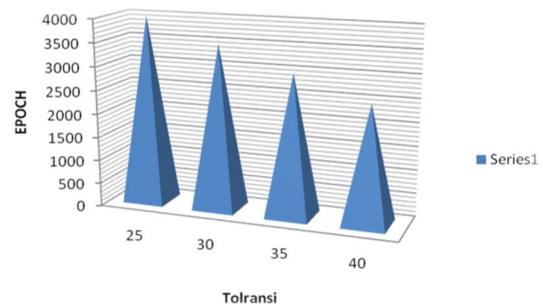


Fig. 6 : The graph of the number of neurons in relation to epoch

After the experiment, it was seen that to achieve a tolerance of 0.06 at the smallest number of neurons obtained at epoch 4000 and for the number of neurons obtained at epoch 3000, followed by the number of neurons 40 in epoch 2500

The graph of success rate can be seen in figure 7

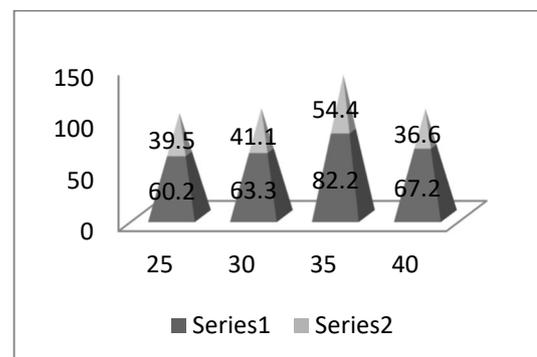


Fig. 7 : Graph of the relationship between the number of neurons to the success rate

After the experiment, it was shown that the greatest success rate of 82.2% for train data and 54.4% for new

data was obtained on the number of layers 35. When the number of neurons was enlarged the smaller success rate was achieved, ie the number of neurons 40 had a success rate of 67.7% for training data and 36.5% for data. This tendency is also found when the number of neurons is reduced is the number of neurons 30 which has a success rate of 63.3.7% for the train data and 41.1% for the data neuron number 25 then obtained a low success rate of 60.2% for training data and 39.5% for new data.

2.8 Analysis

The rate of introduction of training data is higher than the rate of introduction of new data. Differences in recognition rates between train data and new data due to training data are have the same pattern structure as the data used during network training so as to produce a high level of recognition, while new data is data that has never been trained on the network although it has a similar pattern structure that produces a level of recognition that tends to be lower than the introduction rate of train data.

The amount of training conducted has an impact on the success rate of this system, the more training it has to do, the higher the success rate is also higher but it also depends on the tolerance value set.

The cause of this error is the ability of artificial neural networks that have begun to take on properties that are only specific to the training data (but not by the test data) so that they are trapped to local maximum conditions in the situation which the target the system is trying to reach are not found due to errors an increasing and unstable training that makes the system take decisions that are deemed to be closer to the desired target pattern.

3 Conclusions And Recommendations

Based on the analysis of the experimental results obtained, it can be taken several conclusions as follows:

1. Using artificial neural network introduction method got success rate of introduction to trainer data higher than with success rate of introduction to new data.
2. Simulation test result for tolerance values 0.05, 0.06, 0.07, and the number of hidden layer neurons obtained the success rate of training data of 75.5%, 82.2%, 71.1% and new data 36.6%, 53.3%, 38.8.3 %.
3. For the number of hidden layer neurons 25, 30, 35, 40 and tolerance values 0.06 obtained the success rate of training data of 60.2%, 63.3%, 82.2%, 67.7% and new data 39.5%, 41.1%, 54.4%, 36.6% .

After analyzing the system work and system output, for further research and development of this system, the authors suggest several things:

1. Simulation of different characteristic extraction algorithm method in the next research and

development, such as the use of Linear Prediction Coefficients or combined Linear Prediction Coefficients with Mel-Frequency Cepstrum Coefficients (MFCC). So it is expected to give the extraction value more perfect traits

- 2., The expected increase the percentage of success in word recognition by using a combination of other artificial intelligence systems such as the combined Neural Network with HMM on further research

References

- [1] Tolba, Hesham dan O'Shaughnessy, Douglas. *Speech Recognition by Intelligent Machines*. Sumber : Internet, file PDF
- [2] Noel, Mike, *Principles of Speaker Recognition*. <http://cslu.cse.ogi.edu/HLTsurvey/ch1node47.html>
- [3] Gu, Liang., *Perceptual Harmonic Cepstral Coefficients as the Front-End for Speech Recognition*, Signal Compression Laboratory Research Project. http://scl.ece.ucsb.edu/html/prpat_3.htm
- [4] Liu, Wei, Huang, Xiaoling, Wang, Guan, Chai, Tianyou, *The Modeling and Application of Cost Prediction Based on Neural Network*, Research Center of Automation Northeastern University Shenyang, China.
- [5] Jek Siang, Jong, 2005, *Jaringan Syaraf Tiruan & Pemogramannya Menggunakan MATLAB*, ANDI, Yogyakarta.
- [6] Affan, Zaid, 2004, *Suatu Studi Sistem Identifikasi Sumber Suara Berdasarkan Koefisien Prediksi Linier dan Mel-Frequency Cepstrum Coefficient (MFCC) Dengan Menggunakan Jaringan Syaraf Tiruan Perceptron Lapis Jamak Perambatan Balik*, Tugas Akhir, Teknik Elektro Universitas Andalas, Padang.
- [7] Lawrence R. Rabiner dan Biing-Hwang Juang. 1993. *Fundamentals of Speech Recognition*. Prentice Hall International, Inc
- [8] Sanjit K. Mitra. 2001. *Digital Signal Processing : A Computer Based Approach*. Mc Graw Hill, Inc
- [9] Hamza, Ade. 2006. *Pengenalan Identitas Sumber Suara Manusia dengan Menggunakan Metoda Fuzzy Neuro Network (FNN)*, Padang, Universitas Andalas
- [10] Hermawan, arief. 2006. *Jaringan Saraf Tiruan, Teori dan Aplikasi*, ANDI, Yogyakarta
- [11] Suyanto, 2005, *Algoritma Genetik dalam MATLAB*, ANDI, Yogyakarta.
- [12] Silvana, Meza. 2006. *Optimasi Bobot jaringan Saraf Tiruan Menggunakan Algoritma Genetik dalam Identifikasi Suara*. Tugas Akhir, Padang, Teknik Elektro Universitas Andalas.
- [13] Prihadi. 2010. *Sistem Pengenalan Ucapan Jenis Digit Speech Recognition Menggunakan Logika Fuzzy*. Tugas Akhir. Padang: Teknik Elektro Universitas Andalas.