ISSN:0975-9646

Data Optimization in Speech Recognition using Data Mining Concepts and ANN

Navdeep Kaur, Sanjay Kumar Singh

Department of computer science
Lovely Professional University
Phagwara (PB), India

***Abstract* -In this paper we propose to recognize the speech based on optimum feature and independent of the language. Speech recognition refers to the ability to listen spoken words and identify various sounds present in it and recognize of some known language. Speaker recognition is the use of a machine to recognition a person from a spoken phrase. Every speaker has different individual characteristics embedded in his /her speech. The simulated model of neural network based speech recognition system for languages has been developed. We used 23 words. Speech feature are LPC, RC, LPCC are extracted from speech signal and formed feature vectors. Neural Network back propagation learning algorithm for training and it is identification processes of different speakers and languages. The database used for this system consists of 25 speaker. The ANN model consists of 575 neuron in inputs layer and 25 neurons in output layer. The average recognition score is 93.38%**

***Keywords*- LPC, LPCC, RC, Back propagation algorithm**.

I.INTRODUCTION

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, processing information. The key element of this paradigm is the novel structure of the information processing system. Its composed of the large number of highly interconnected processing elements working in unison and to solve specific problems ANNs, like people learn by example. An ANN is configured for a specific application such as pattern recognition or data classification its through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons [13]. Artificial Intelligence (AI), machine which like behave human brain. Brain represent by the neural network which continually receive information.

A. Artificial neural network (Back propagation)

Artificial neural network is non-linear predictive models that learns through training and resemble biological neural networks in structure. can approximate any nonlinear function to an arbitrary degree of accuracy through the composition a network of relatively simple functions [1, 2]. The flexibility and simplicity of neural networks have made them a popular modeling and forecasting tool across different research. Neural networks are the a class of flexible nonlinear models inspired by the way in which the human brain processes information. Given an appropriate number of hidden-layer units, the hidden layer chose according to our environments and neural networks areas in recent years. A variety of different neural network models have developed, among which the Back-propagation (BP) network is the most widely the present study. Show the diagram of feed forward neural network is given as-

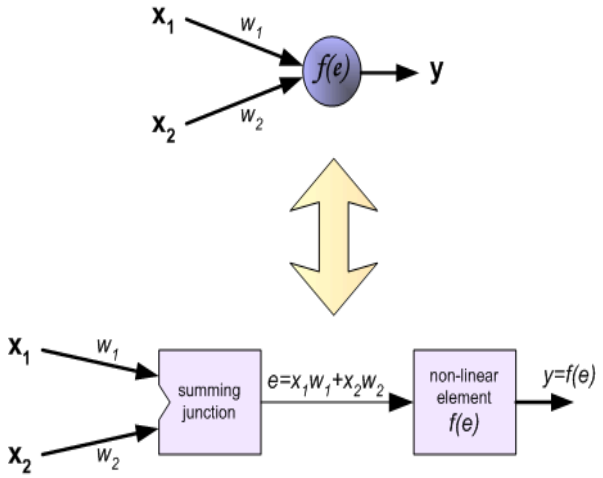


Figure1 Feed Forwards Neural Networks

Each neuron is composed of two units. First unit adds products of weights coefficients and input signals. The second unit realize nonlinear function, called neuron activation function. Signal *e* is adder output signal, and y = f(e) is output signal of nonlinear element. Signal *y* is also output signal of neuron[1]. To teach the neural network we are need training data set. The training data set consists of input signals (x_1 and x_2) assigned with corresponding target (desired output) *z*. The network training is an iterative process. In each iteration weights coefficients of nodes are modified using new data from training data set.[10]

Output is equal to multiple of inputs into weight plus bias value. The value of output is given negative then used bias value it show positive value of output. The output is y1,y2…….yp. The inputs is x1, x2……..xn. Weight is represented w. the equation is-

Output = weight * inputs + bias

B. Data mining

Data mining is the extraction of new information from large databases. It is a powerful technology to help companies focus on the most important information in their data warehouses. . In which some techniques are used in data mining:-

1) Artificial neural network: Non-linear predictive models that learn through training and resemble biological neural networks in structure

2) Clustering: Clustering is the most important unsupervised learning problem; it deals with finding a structure in a collection of unlabeled data. A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way A cluster is a collection of objects which are similar between them and are dissimilar to the objects belonging to other clusters [2]. We can show clustering diagram with a simple graphical example In which example show that 4 clusters data is divided; the similarity criterion is distance*:* two or more objects belong to the same cluster if they are "close" according to a given distance (in this case geometrical distance). It is called distance based clustering.

3) Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection

4) Genetic algorithms: Optimization techniques that use process like genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

5)Nearest neighbor method:- This technique that classifies each record in a dataset it is based on a combination of the classes of the k record(s) most similar to it in a historical dataset it is called the k-nearest neighbor technique.

6) Rule induction:- The extraction of useful if-then rules from data based on statistical significance.

C.Speech recognition

Speech recognition refers to the ability to listen (input in audio format) spoken words and identify various sounds present in it, and recognize them as words of some known language. Speech recognition in computer system domain may defined as the ability of computer systems to accept spoken words in audio format such as wav or raw and then generate its content in text format. The conditions of evaluation and hence the accuracy of any system can vary along the following dimensions:

1) Speaker dependence vs. independence: A speaker dependent system is intended for use by a single speaker, but a speaker independent system is intended for use by any speaker [13].

2) Isolated, discontinuous, or continuous speech: Isolated speech means single words; discontinuous speech means full sentences in which words are artificially separated by silence; and continuous speech means naturally spoken sentences.

3) Task and language constraints: Even with a fixed vocabulary performance will vary with the nature of constraints on the word sequences that are allowed during recognition. Some constraints may be task-dependent [13].

4) Read vs. spontaneous speech: Systems can be evaluated on speech that is either read from prepared scripts that is uttered spontaneously. Spontaneous speech is more difficult because it tends to be peppered with disfluencies like "uh" and "um", false starts incomplete sentences stuttering, coughing, and laughter; and moreover, the vocabulary is essentially unlimited, so the system must be able to deal intelligently with unknown word [13].

D. Feature Extraction

The Linear Prediction Code (LPC) is approach to form feature or spectral vector [11]. Linear prediction coefficients are a highly effective representation of the speech signal. In this analysis, each speech sample is represented by a weighted sum of *p* past speech samples plus an appropriate excitation [12]. The main importance is to determine LPC coefficients minimizing the prediction error in the least squares sense.

$$s[n] \approx \sum_{k=1}^{p} a[k]s[n-k]$$

Where s[n] denotes the speech signal samples, a[k] are the predictor coefficients and p is the order of the predictor.

The total squared prediction error is:

$$E=(\sum_{n}[s]-\sum_{k-1}^{p}a[k]s[n-k])^{2}$$

Linear predictive analysis is to determine the coefficients of a[k] for each speech frame so that the error equation second is minimized. The problem can be solved by setting the partial derivatives of equation second with respect to a[k] to zero.

Approach: In this approach show that two types of preprocessing steps first is one reduce noise and second is separate word from sentence. Neural network training check target output and actual output is same or not. Feature extraction, testing, neural network training is shown in figure 2

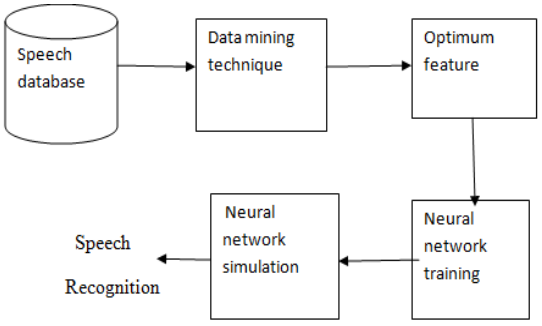


Figure 2 of approches

II. RESULTS

When proposed system is trained using artificial neural network with limited number of neurons and layer, the average performance was 93.38% with 36 error of 575 input data size. The implements steps commands is given as-

a. The file housing.mat contains a predefined set of input and target vectors. The input vectors define data regarding real-estate properties and the target values define relative values of the properties. Load the data using the following command:

```
load house_dataset
whos
Name          Size          Bytes  Class     Attributes
houseInputs   13x506        52624  double
houseTargets  1x506          4048  double
```

b. The next step is to create a network and train it until it has learned the relationship between the example inputs and targets.

The most common network used with back propagation is the two-layer feed-forward network. The following call to newff creates a two-layer network with 20 neurons in the hidden layer. (The numbers of neurons in the output layer are automatically set to one, the number of elements in each vector of t.

```
net=newff(in,srtarget_5,25);
```

Two-layer feed-forward networks can potentially represent any input-output relationship with a finite number of discontinuities (which is a typical relationship you might want to model), assuming that there are enough neurons in the hidden layer.

b. Next step is to train the network using the data.

```
net=train(net,in,srtarget_5);
```

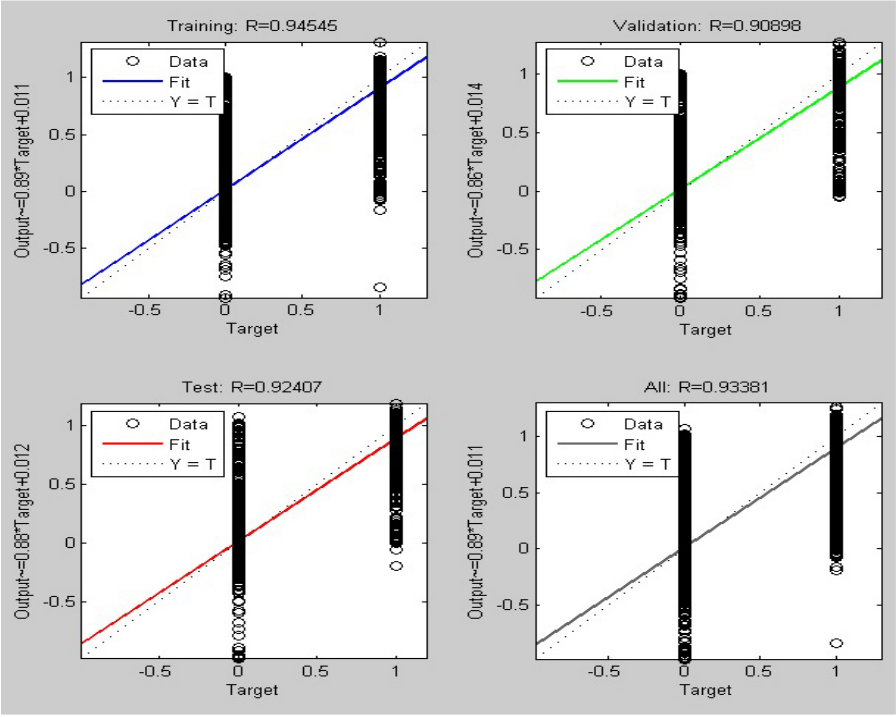
After training the network, you can use it. Use sim to apply the network to the original vectors

```
Y = sim(net,in,test_3);
```

Number of inputs utterance is 23 and speaker is 25. The maximum number of error is 4 and minimum number of error is 0. The minimum performance of system is 82.60% while best performance reaches up to 100%. Overall performance of the system is 93.38%. The final result of speech recognition is 93.38%. shown in Figure 3

III. CONCLUSIONS

Speech features and a neural network with back propagation training algorithm are appropriate to use for Text Independent Multilingual speaker recognition. BPA can be used multilingual system. The minimum performance of system is 82.60% while best Performance reaches up to 100%. Overall performance of the system is 93.38%. To improve the system capability, mixed tone speech in appropriate length of speech duration should be selected for speaking sentence since it can cover more personal characteristics than using each tone in all utterance.



SPEAKER RECOGNITION

Figure 3 speech recognition

VI. FUTURE WORK

In future I will find out best efficient data mining concept or ANN to optimize the feature matrix such that it can be use further processing i.e. in speech recognition and to improve system capability.

REFERENCES

[1] Anurag aggarwal, diwaker gupta "*Page replacement algorithm simulator*

[2] Chulhee Lee, Donghoon Hyun, Euisun Choi, Jinwook Go, and Chungyong Lee" Optimizing Feature Extraction for Speech Recognition''*signal processing 2003 3rd International conference* on pp 80-87.

[3] Christos Stergiou and Dimitrios Siganos" Neural network''*signal processing 2003 3rd International conference*

[4] Faltlhauser,R.; Ruske G. "Robust speaker clustering in eigenspace*", Inst. for Human-Machine-Communication, Technische Universitat Munchen, Munich,Germany*,2002 IEEE.

[5] Gilat, Amos (2004)." MATLAB: An Introduction with Applications 2nd Edition". John Wiley & Sons. *ISBN 978-0-471-69420-5*.

[6] John Bridle,''Optimization and Search in Speech and Language Processing''
Dragon Systems UK Ltd., Cheltenham, UK

[7] Joe Tebelskis''Speech Recognition using Neural Networks,"*Communication signal processing,1998.COMSIG98*

[8] Ms. Smita.nirkhi & Xu liujie "potential use of artificial neural network in data mining", "*Engineering (ICCAE), 2010 The 2nd International Conference on 2010"*, pp: 339-342

[9] Quarteroni, Alfio; Fausto SaIoleri (2006)."*Scientific Computing with MATLAB and Octave*. Springer". ISBN 978-3-540-32612-0

[10] Ripul gupta, "*speech recogination for hindi*".

[11] Refrence Roman M. Balabin, Ekaterina I. Lomakina (2009). "Neural network approach to quantum-chemistry data: Accurate prediction of density functional theory energies". *J. Chem. Phys.* **131** (7): 074104

[12] Ryszard Tadeusiewcz,"Principles of training multi-layer neural network using backpropagation", *Kraków* 1992

[13]Sanjay kumar singh ,Rajeev kumar, Rajesh ranjan, rahul kala, Dr anupam shukla & Dr Ritu tiwari, "Multilingual speaker recognition using neural network", *FRSM*, December 2009.

[14] Ms. Smita.nirkhi & Xu liujie "potential use of artificial neural network in data mining", *"Engineering (ICCAE), 2010 The 2nd International Conference on 2010"*, pp: 339-342

[15] T. Matsui and S. Furui, "A Text-Independent Speaker Recognition Method Robust Against Utterance Variations,*" Proc. IEEE Int. Confi Acoust. Speech Signal Processing, S6.3*, (1991).

[16] Xing jiandong, wei shizhang, zhang songmin, zhang yangzhen and long rui," Use of artificial neural network in predicting mechanical properties of high speed steel(HSS)" , *Mechatronics and Automation, Proceedings of the 2006 IEEE International Conference on* pp 1872-1877,2006.