

Speech Recognition Using Combined Fuzzy and Ant Colony Algorithm

Foad Jalili¹, Milad Jafari Barani²

¹ Departement of Computer Engineering Science and Research Branch, Islamic Azad University Tehran, Iran

² Young Researchers and Elite Club, Urmia Branch, Islamic Azad University, Urmia, Iran

Article Info

Article history:

Received Feb 7, 2016

Revised Aug 4, 2016

Accepted Aug 20, 2016

Keyword:

Ant colony

Clustering

Fuzzy logic

Noise removal

Speech recognition

ABSTRACT

In recent years various methods has been proposed for speech recognition and removing noise from the speech signal became an important issue. In this paper a fuzzy system has been proposed for speech recognition that can obtain accurate results using classification of speech signals with “Ant Colony” algorithm. First, speech samples are given to the fuzzy system to obtain a pattern for every set of signals that can be helpful for dimensionality reduction, easier checking of outcome and better recognition of signals. Then, the “ACO” algorithm is used to cluster these signals and determine a cluster for each input signal. Also, with this method we will be able to recognize noise and consider it in a separate cluster and remove it from the input signal. Results show that the accuracy for speech detection and noise removal is desirable.

Copyright © 2016 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Foad Jalili,

Departement of Computer Engineering Science and Research Branch,

Islamic Azad University,

Tehran, Iran.

Email: foadjalili@gmail.com

1. INTRODUCTION

In last two decades, speech recognition was one of the most important topics of signal processing [1],[2]. Various methods have been proposed in recent years that have their own strength and weaknesses e.g. Particle Swarm Optimization (PSO) combined with Feed-forward Neural-network (FNN) that is called PSO-FNN [3], speech recognition with fuzzy T-S neural-networks [4] and speech recognition based on RBF neural networks [5]. Such methods had their specific problems for example they were not able to recognize two similar signals or fuzzy neural networks had high time complexity problem because the mass of speech signals.

One of the difficulties in speech recognition is incorrect pronunciation of words or having accent that put the system in uncertainty. For solving this problem, a fuzzy model is used to help system in uncertainty conditions. Fast and accurate speech recognition is done by reducing signal intervals that is fuzzy model's output and the distinctness of signals is done by the “ACO” algorithm. Also, with existence of noisy classes for each sample, speech recognition can operate under noisy conditions and noise removal can be done by initial signals [6]-[11].

2. ANT COLONY ALGORITHM

Today most of the researchers are tend to use nature based algorithms like “PSO”, “ACO” and “Firefly” to solve variety of problems. These algorithms are population-based meta-heuristic search algorithms. “ACO” algorithm as mentioned is a population-based algorithm that seeks to find the shortest path between colony and food source by following the paths that more other ants secreted pheromones from

themselves. This algorithm is able to be used in variety of optimization problems for example: Traveling salesman problem [12], data clustering [13],[14] and text mining [15],[16]. This algorithm can be explained with four rules:

- Ants are seeking food and they search different paths for finding it.
- Each ant secretes pheromone from itself.
- Ants are attracted to pheromone smell, so path with more pheromone is path with more traffic.
- Pheromone evaporates by time, so path with less pheromone means path with less traffic.

This algorithm has 4 phases that is explained as follow:

- Initial pheromone amount for each element of pheromone table is calculated with (1) where $Fitness(Sbest)$ indicates fitness of solution for the problem.

$$\tau_0 = \frac{Fitness(Sbest)}{ProblemSize} \quad (1)$$

- Probability function (2) is used for determining a path. Where $p_k(r,s)$ shows the movement probability of ant k from point r to s . $\tau(r,s)$ indicates to amount of pheromone on the path and $\eta(r,s)$ shows the fitness of movement. α and β are coefficients used to regard either table of pheromone or information given by problem. At last $J_k(r)$ is set of travelled nodes and r is last node met.

$$P_k(r, s) = \left\{ \frac{[\tau(r,s)]^\alpha \cdot [\eta(r,s)]^\beta}{\sum_{\mu \in J_k(r)} [\tau(r,s)]^\alpha \cdot [\eta(r,s)]^\beta} \quad \text{if } s \in J_k(r) \right\} \quad (2)$$

- To local update of pheromone table that causes preventing ants to move from just one path and make them to scan new paths, (3) will be used.

$$\tau(r, s) = (1 - \sigma) \cdot \tau(r, s) + \sigma \cdot \tau_0 \quad (3)$$

Where σ is local update coefficient, τ_0 is initial amount of pheromone and $\tau(r,s)$ is value of (r,s) element in pheromone table.

- For global update of pheromone table, each element of table is updated with (4).

$$\tau(r, s) = (1 - \rho) \cdot \tau(r, s) + \rho \cdot \Delta\tau(r, s) \quad (4)$$

Where ρ is pheromone's global update coefficient (here is taken 0.1) and $\Delta\tau(r,s)$ is calculated with (5).

$$\Delta\tau(r, s) = \{Fitness(Sbest) \quad \text{if } (r, s) \in S_{best}\} \quad (5)$$

3. PROPOSED METHOD

Proposed method includes these steps: first a fuzzy model is designed and then ACO clustering algorithm will be applied and at last noise detection and removal will be performed.

3.1. Designing Fuzzy Model

As shown in Figure 1, x_n is input of fuzzy model and y_R is the output. Size of the outputs is related to number of rules which are related to the signal size. This means if we perform more rules for proposed model then the size of output will be increased where this increment is under effect of signal size. Output of fuzzy model, for each input speech signal in continuous form, allocates a value as a key.

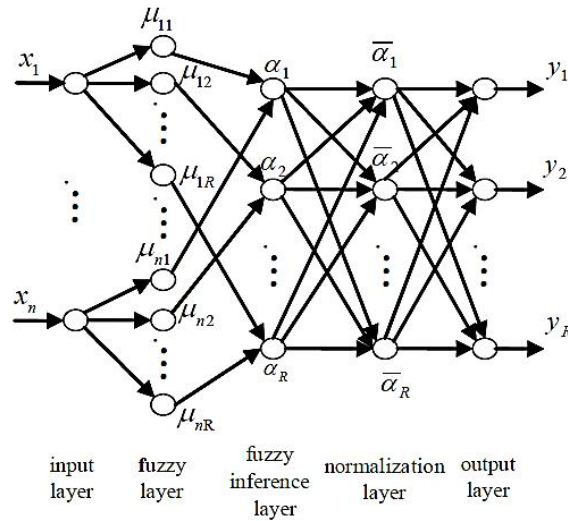


Figure 1. Fuzzy model

- *Fuzzification*: a Gaussian membership function (6) is used for determining a membership degree for each input signal to each fuzzy set.

$$\mu_A(x, c, s, m) = \exp \left[-\frac{1}{2} \left| \frac{x-c}{s} \right|^m \right] \quad (6)$$

- *Fuzzy operators*: fuzzy interface is calculated with (7) where N_{adj} is adjustment coefficient and considered $n/4$ that n is number of input signals.

$$\alpha_j = \prod_{i=1}^R \mu_{ij} * \frac{1}{N_{adj}} \quad (7)$$

- The output of a fuzzy set is calculated with membership function so output of node j is calculated with (8).

$$\bar{\alpha}_j = \alpha_j / \sum_{i=1}^R \alpha_i \quad (8)$$

- *Aggregation*: since decision making in a fuzzy interface is taken regarding to all of the rules, in this layer they are all combined so that output of a node k is calculated with (9).

$$y_k = \sum_{j=1}^R w_{jk} \bar{\alpha}_j, k = 1, 2, A, R \quad (9)$$

Where W_{jk} is rule's weight. After obtaining the output signals, ACO will be used for clustering them.

3.2. Ant Colony Clustering Algorithm

Clustering is an unsupervised assortment of similar objects or data into some classes. Criterion for similarity of objects is determined with distance measurement. Many types of distance can be used e.g. Euclidean distance, mahalanobis distance and city block distance. Cluster centers are decision making variables that can be gained with minimizing the Euclidean distance over all of training samples in n -dimensional space. Purpose of clustering is minimizing the sum of city block distance over N samples and allocating each of samples to one of K clusters. This is done by (10) where K is number of clusters for N samples.

$$D_{city-block}(x_i, c_k) = \sum_k \sum_{i=1}^N (x_i - c_k) \quad (10)$$

Centers of cluster are calculated with (11).

$$c_k = \sum_{i \in c_k} \frac{x_i}{n_k} \quad (11)$$

The objective function for sample i is calculated with (12).

$$G_i = \frac{1}{S_{Train}} \sum_{j=1}^{S_{Train}} d(x_j, m_i) \quad (12)$$

Where S_{train} is number of training samples and m_i is defined as a class that samples are belongs to it.

3.3. Denoising Speech Signals

In proposed method each signal belongs to a cluster but as we know signal contains environment noise and this problem can affect recognition rate. E.g. in Figure 2 spelling of word “five” in six different type is shown and noise is impressed the signal and it is clear on the image. To solve this problem, difference of signal with noise and without noise is considered as a noise class and given to the fuzzy circuit in order to allocate a noise class for that signal.

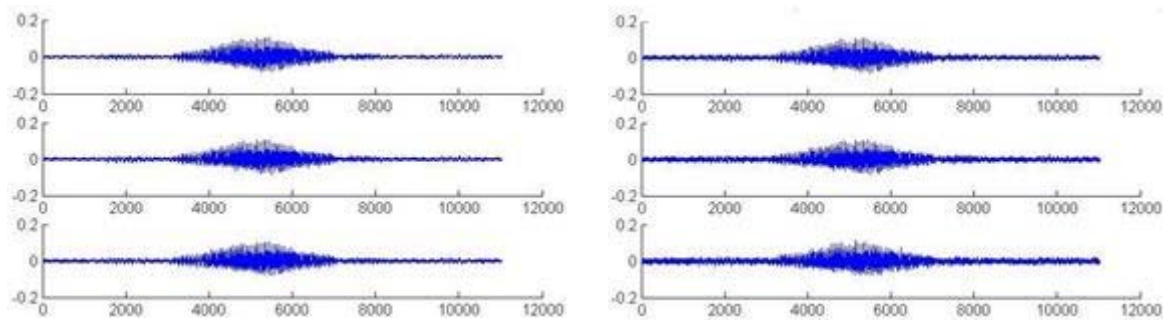


Figure 2. Spelling of word “five” in six different types

Then, the noise class for a particular signal is fetched and difference between it and the original signal is calculated. If they are equal or similar, then noise is removed from signal by performig but if they are not, then it should be considered as a new noisy signal and put it in another noisy class. In Figure 3a, the original signal with noise is shown and in Figure 3b the noise has been removed from the signal and speech signal quality is high.

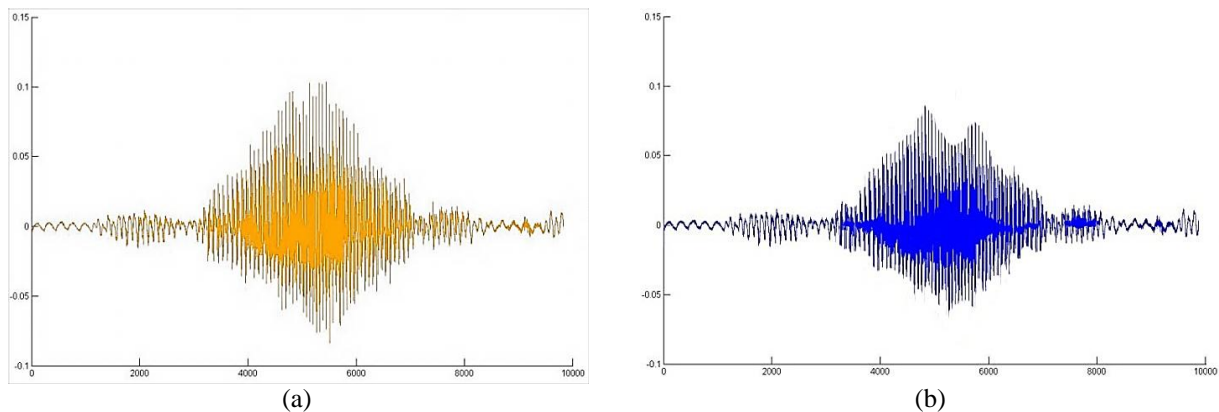


Figure 3. (a) Original speech signal with environment noise (b) Speech signal after denoising

4. RESULTS AND ANALYSIS

To make a comparison between proposed method and other methods a dataset with 45 persons is used that contains Gaussian noise with 15db, 20db, 25db, 30db and a class without noise. In speech recognition of particular person a particular class is created for that person. Then it is compared with other classes to determine the most similar class.

Table 1. Result Obtained from Performing Proposed Method on Dataset

Training Algorithm	Proposed Method	PSO-FNN
SNR (db)		
15	96.7	94.8
20	97.2	95.7
25	96.9	96.7
30	97.0	96.2
Clean	97.8	96.7
Average Recognition Rate	97.12	96.02

5. CONCLUSION

In this paper a combined fuzzy and ACO based method had been proposed for speech recognition. For dimensionality reduction and better recognition rate for signals first the proposed fuzzy system is used. Then ACO clustering algorithm is performed on the signals to allocate them to their appropriate clusters with minimizing the city-block distance between signals and cluster centers. This method as shown in results leads to better quality in comparison with other proposed algorithms and has lower time complexity regard to fuzzy system and city-block distance measure that is easier to compute rather than Euclidean distance and other distance measures.

REFERENCES

- [1] R. M. Hegde, *et al.*, "Significance of the modified group delay feature in speech recognition," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol/issue: 15(1), pp. 190-202, 2007.
- [2] B. Ghazale, *et al.*, "A comparative study of traditional and newly proposed features for recognition of speech under stress," *Speech and Audio Processing, IEEE Transactions on*, vol/issue: 8(4), pp. 429-442, 2000.
- [3] Zhang J. R., *et al.*, "A hybrid particle swarm optimization-backpropagation algorithm for feedforward neural network training," *Applied Mathematics and Computation*, vol. 185, pp. 1026-1037, 2007.
- [4] X. Y. Zhang and P. Wang, "Improved T-S Fuzzy NeuralNetwork in Application of Speech Recognition System," *Computer Engineering and Applications*, vol. 45, pp. 246-248, 2009.
- [5] G. Li, "The Study of the Speech Recognition Based on the Optimized RBF Neural Network," Master Degree Thesis of Taiyuan University of Technology, Taiyuan, 2009.
- [6] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," *Speech communication*, vol/issue: 48(9), pp. 1162-1181, 2006.
- [7] H. Jiang, "Confidence measures for speech recognition: A survey," *Speech communication*, vol/issue: 45(4), pp. 455-470, 2005.
- [8] K. Palecek and J. Chaloupka, "Audio-visual speech recognition in noisy audio environments," *Telecommunications and Signal Processing (TSP), 2013 36th International Conference on. IEEE*, 2013.
- [9] P. Dai and Y. Soon, "Robust speech recognition by using spectral subtraction with noise peak shifting," *IET Signal Processing*, vol/issue: 7(8), pp. 684-692, 2013.
- [10] M. A. Anusuya and S. K. Katti, "Speech recognition by machine, a review," arXiv preprint arXiv: 1001.2267, 2010.
- [11] O. Kalinli, *et al.*, "Noise adaptive training for robust automatic speech recognition," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol/issue: 18(8), pp. 1889-1901, 2010.
- [12] C. A. Coello, *et al.*, "Evolutionary algorithms for solving multi-objective problems," Springer, 2007.
- [13] E. D. Lumer and B. Faieta, "Diversity and adaptation in populations of clustering ants," *Proceedings of the third international conference on Simulation of adaptive behavior: from animals to animats 3: from animals to animats 3*. MIT Press, 1994.
- [14] C. F. Tsai, *et al.*, "ACODF: a novel data clustering approach for data mining in large databases," *Journal of Systems and Software*, vol/issue: 73(1), pp. 133-145, 2004.
- [15] J. Handl and B. Meyer, "Improved ant-based clustering and sorting in a document retrieval interface," *Parallel Problem Solving from Nature—PPSN VII. Springer Berlin Heidelberg*, pp. 913-923, 2002.
- [16] K. M. Hoe, *et al.*, "Homogeneous ants for web document similarity modeling and categorization," *Ant Algorithms*, Springer Berlin Heidelberg, pp. 256-261, 2002.

BIOGRAPHIES OF AUTHORS

Foad jalili was born in urmia, Iran in 1987 and graduated B.Sc in computer engineering from Islamic Azad University Urmia branch and M.Sc from Islamic Azad University Qazvin branch in artificial intelligence. Recently he is with Islamic Azad University Science and Research Center Tehran branch as a Phd student in artificial intelligence. At the same time he is professor at Islamic Azad University of Urmia and his research interests include: Image processing, Signal processing, Image Authentication, Cryptography, Speech Recognition, Neural networks and Fuzzy systems.



Milad jafari barani was born in naqadeh, Iran in 1985. He received B.E. degree in computer engineering from Islamic Azad University, Urmia, Iran, respectively M.Sc degree in artificial intelligence in 2014 from Qazvin branch, Islamic Azad University, Qazvin, Iran. He is professor in Islamic Azad University of Urmia and University of Applied Science and Technology of Urmia and he is member of Young Researchers and Elite Club Urmia Branch, Islamic Azad University, Urmia, Iran. His research interests include image processing, watermarking, Image Authentication, optimization algorithms, Cryptography, neural networks and image authentication.