RECURRENT NEURAL NETWORK WITH BACKPROPAGATION THROUGH TIME ALGORITHM FOR ARABIC RECOGNITION

Saliza Ismail[1] and Abdul Manan bin Ahmad[2]
Department of Software Engineering,
Faculty of Computer Science and Information System,
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
Tel : 607-5532201[2] Fax : 607-5565044[2]
chukiwa@hotmail.com[1],manan@fsksm.utm.my[2]

Abstract
The study on speech recognition and understanding has been done for many years. In this paper, we propose a new type of recurrent neural network architecture for speech recognition, in which each output unit is connected to itself and is also fully connected to other output units and all hidden units[1]. Besides that, we also proposed the new architecture and the learning algorithm of recurrent neural network such as Backpropagation Through Time (BPTT), which well-suited. We also re-train for the output before we analyze the result. The purpose of this training is to produce the best result. The aim of the study was to observe the difference of Arabic's alphabet like "alif" until "ya". The purpose of this research is to upgrade the people's knowledge and understanding on Arabic's alphabet or word by using Recurrent Neural Network (RNN) and Backpropagation Through Time (BPTT) learning algorithm. 6 speakers (a mixture of male and female) are trained in quiet environment.

Neural network is well-known as a technique that has the ability to classified nonlinear problem. Today, lots of researches have been done in applying Neural Network towards the solution of speech recognition[2] such as Arabic. The Arabic language offers a number of challenges for speech recognition[3]. Even through positive results have been obtained from the continuous study, research on minimizing the error rate is still gaining lots attention. This research utilizes Recurrent Neural Network, one of Neural Network technique to observe the difference of alphabet "alif" until "ya".

Keywords
Recurrent Neural Network, Backpropagation Through Time, Arabic Alphabet, Speech Recognition, Real-Time Recurrent Learning.

Introduction

Speech is human's most efficient communication modality. Beyond efficiency, human are comfort and familiar with speech. Other modalities require more concentration, restrict movement and cause body strain due to unnatural positions. Research work on Arabic speech recognition, although lagging that other language, is becoming more intensive than before and several researches have been published in the last few years [4].

The conventional neural networks of Multi-Layer Perceptron (MLP) type have been increasingly in use for speech recognition and also for other speech processing applications. Those networks work very well as an effective classifier for vowel sounds with stationary spectra, while their phoneme discriminating power deteriorates considerably for consonants which are characterized by variations of their short-term spectra. This may be attributable to a fact that feedforward multi-layer neural network are inherently unable to deal with time varying information like time-varying spectra of speech sounds. One way to cope with this problem to incorporate feedback structure in the networks to provide them with an ability to memorize incoming time-varying information. Incorporating feedback structure in feedforward networks results in so-called Recurrent Neural Networks (RNNs) which have feedback connections between units of different layers or connections of self-loop type [5].

Speech recognition is the process of converting an acoustic signal, captured by microphone or a telephone, to a set of words. The recognized words can be the final results, as for applications such as commands and control, data entry and document preparation. They can also serve as the input to further linguistic processing in order to achieve speech understanding, a subject covered in section [6]. As we know, speech recognition performs their task similar with human brain. Start from phoneme, syllable, word and then

Proceedings 18th European Simulation Multiconference
Graham Horton (c) SCS Europe, 2004
ISBN 3-936150-35-4 (book) / ISBN 3-936150-36-2 (CD)

sentence which is an input for speech recognition system [7]. Many researches that have been prove to prove to decrease the error and also any disruption while doing the recognition.

Now, students not interested in lessons regarding Arabic language such as Jawi writing even the lessons have been teaching at primary school. The purpose of the lessons is to teach the students how to pronoun and write the alphabet. Therefore, students can read Holy-Quran properly. But the students only can understand that pronoun and writing while they in Standard 6. So after that, they will forget all the lessons [8].

Architecture

RNN have feedback connections and address the temporal relationship of inputs by maintaining internal states that have memory. RNN are networks with one or more feedback connection. A feedback connection is used to pass output of a neuron in a certain layer to the previous layer(s) [9]. The different between MLP and RNN is RNN have feedforward connection for all neurons (fully connection). Therefore, the connections allow the network show the dynamic behavior. RNN seems to be more natural for speech recognition than MLP because it allows variability in input length [10].

The motivation for applying recurrent neural network nets to this domain is to take advantage of their ability to process short-term spectral features but yet respond to long-term temporal events. Previous research has confirmed that speaker recognition performance improves as the duration of utterance is increased [11]. In addition, it has been shown that in identification problems RNNs may confer a better performance and learn in a shorter time than conventional feedforward networks [12].

Recently a simple recurrent neural network, which has feedback connections of self-loop type around hidden layer units, has been proposed as an attractive tool for recognizing speech sounds including voiced plosive sounds [1]. This network has three layers such as input layer, hidden layer and output layer. Each of the output layer units has feedback connection with itself, i.e., a self-loop as shown in Fig. 1.

The output of each input layer at time t-1 is fed, through connections between the input and hidden layers, to all the hidden layer units at time t and in the same manner the output of each hidden layer unit at time t-1 is supplied, through connections between the hidden and output layers, to all the output layer units at time t. the output at time t-1 of each output layer unit is feedback to itself at time t. in training the proposed recurrent neural network, weight at t of all the connections between the input and hidden layers as well as connections between the hidden and output layers are affected by all the input vectors to the input vectors to the input layer before time t.

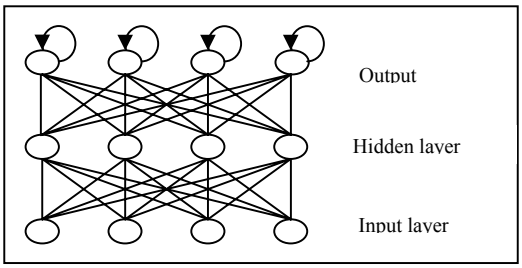

Fig. 1: RNN Architecture

Actually this architecture has been used in visual pattern recognition [13,14] but we use this architecture for speech recognition especially for Arabic speech recognition by using Backpropagation Through Time (BPTT) as learning algorithm. This architecture also have been proved that this architecture better than MLP in phoneme recognition accuracies [1] by using Backpropagation algorithm. In this paper, we want to prove that the architecture also can be used in Arabic's alphabet with Backpropagation Through Time (BPTT) learning algorithm.

The Backpropagation Through Time (BPTT) algorithm is based on converting the network from a feedback system to purely feedforward system by folding the network over time. Thus, if the network is to process a signal that is time steps long, then copies of the network are created and the feedback connections are modified so that they are feedforward connections from one network to the subsequent network. The network can then be trained if it is one large feedforward network with the modified weights being treated as shared weight [15]. Real-Time Recurrent Learning (RTRL) algorithm is based on recursively updating the derivatives of the output and error. These updates are computed using a sequence of calculations for iteration. The weights are updated either after iteration or after the final iteration of the epoch.

The major disadvantage of this algorithm is that it requires an extensive amount of computation at iteration [16]. Additionally, this algorithm is very slow because the RTRL has many weights to compute and therefore, the training process will be more slowly [9].

Speech Recognition System

Generally, speech recognition process contains three steps to process the speech which is acoustic processing, feature extraction and recognition, as shown in Figure 2. First, we digitize the speech that we want to recognize. In this paper, we digitize the Arabic's alphabet from the speakers and also digital filtering that emphasizing important frequency component in signal. Then we analyze the start-end point depends the signal of the speeches. GoldWave software is used to filter and conversion the analog to digital. From that, we can analyze the start-end point that contains the important information of speeches.

The second steps is feature extraction that digital signal in time domain will fed to LPC spectrum analysis for extract the signal or we called it as frame normalizing. Linear Predictive Coding (LPC) is used to extract the LPC coefficients from the speech tokens [17,18]. The LPC coefficients are the converted to cepstral coefficients. The cepstral coefficients are normalized in between +1 and -1. the cepstral coefficients are served as input to the neural networks.

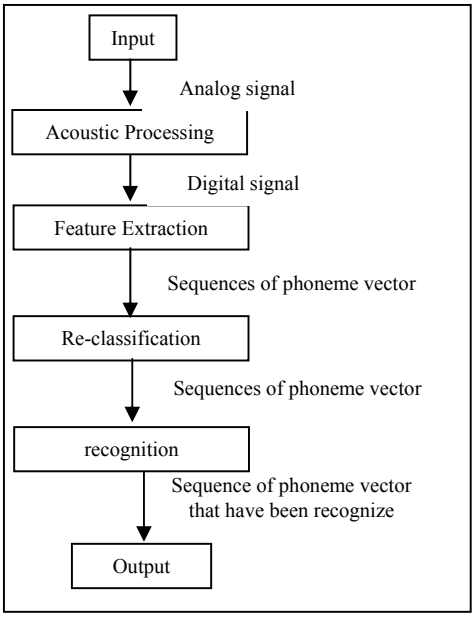

Figure 2: Process of Speech Recognition

Finally, we classify and recognize the speech with learning algorithm Backpropagation Trough Time in Recurrent Neural Network. Actually, before the recognize process, the output will be training once again for produce the best output. This process suitable for minimum input such as Arabic's alphabet that has 30 letters.

In spoken language, a phoneme is a basic, theoretical unit of sound that can change the meaning of a word. A phoneme may well represent categorically several phonetically similar or phonologically related sounds (the relationship may not be so phonetically obvious, which is one of the problems with this conceptual scheme). Depending on the language and the alphabet used, a phoneme may be written consistently with one letter; however, there are many exceptions to this rule (especially in English) [19].

The Arabic alphabet has 30 letters and it is written from right to left. Letters change shape depending on which other letters are before or after them, much like American or Continental handwriting. Phonemes are best described as linguistic units. They are the sounds that group together to form our words, although quite how a phoneme converts into sound depends on many factors including the surrounding phonemes, speaker accent and age. A phoneme is the smallest contrastive unit in the sound system of a language [20].

The Arabic's alphabet in this research contains 20 letters such as "alif", "ba", "ta", "tha", "jim", "ha", "kha", "dal", "zal", "ra", "zai", "sin", "syin", "sad", "dhad", "to", "za", "ain", "ghain" and "fa".

Experiments

The Table 1 shows the results of 20 alphabets from recognition experiments. The testing of this system has been pronounced by 6 Malay speakers (3 men and 3 women). Every speaker must repeat the Arabic's alphabet about 10 times sequentially for each alphabet. So, total of the pronoun for this experiments that includes 6 speakers x 20 alphabets x 10 times for every alphabet (6x20x10), are 400 speeches. From the table, the "ta" alphabet get 98% that higher and the lowest rate is "kha" alphabet, and also all the alphabet in the Table 1.

Table 1: Expected Result Arabic's Phoneme Recognition for Four Speakers using RNN and BPTT learning algorithm

Arabic's alphabet	Rate (%)
"alif"	85
"ba"	96
"ta"	98
"tha"	75
"jim"	87
"ha"	60
"kha"	60
"dal"	91
"zal"	84
"ra"	88
"zai"	85
"sin"	91
"syin"	90
"sad"	73
"dhad"	75
"to"	83
"za"	87
"ain"	78
"ghain"	72
"fa"	88

Conclusion

Currently development of speech recognition is widely used in industrial software market. The main contribution of proposed phoneme recognition system is encouraged to recognize the Arabic's alphabet properly. Besides, its can help the beginner to start their lessons about how to pronouns the word of Holy-Quran. Furthermore, we presented the new architecture and the learning algorithm that Backpropagation Through Time, are well-suited and better than Elman or Jordan architecture. However, the process become more effective after re-train the first output.

Findings from results of the expected experiments can be summarized as follows:

1. The low rate recognition of error to the alphabets that pronounced with bilabial (both lips) are contributed to the naturally pronunciation of those alphabet that are much clear such as "ba" and "ta" alphabets.
2. The lowest rate of recognition to the alphabets that pronounced with pharynx of those alphabets like "ha" and "kha" are contributed of pronunciation that needs 'makhraj' exactly (right pronunciation in Arabic).

Hopefully, this system will help us to recognize and differentiate the Arabic's. We also hope we'll continue the process until "ya" besides minimize the time.

References

[1]Koizumi T., Mori M., Taniguchi S. and Maruya M., (1996). "*Recurrent Neural Networks for Phoneme Recognition*." Department of Information Science, Fukui University, Fukui, Japan, Spoken Language, ICSLP 96, Proceedings, Fourth International Conference, Vol. 1, 3-6 October, Page(s): 326 -329.

[2]Turban E., (1992). "*Expert Systems and Applied Artificial Intelligence*." Republic of Singapore: MacMillan Publishing Company, 623-640.

[3]Mayfield T. L., Black A. and Lenzo K., (2003). "*Arabic in my Hand: Small-footprint Synthesis of Egyptian Arabic*." Euro Speech 2003, Geneva, Switzerland.

[4]Jihene El Malik, (1998). "*Kohonen Clustering Networks For Use In Arabic Word Recognition System*." Sciences Faculty of Monastir, Route de Kairouan, 14-16 December.

[5]Medser L. R. and Jain L. C., (2001). "*Recurrent Neural Network: Design and Applications*." London, New York: CRC Press LLC.

[6]Lippman R.P., (1989). "*Review of Neural Network for Speech Recognition*." Neural Computation 1.1-38.

[7]Joe Tebelskis, (1995). "*Speech Recognition using Neural Network*." Carnegie Mellon University: Thesis PhD.

[8]Siddiq Fadzil, (2001). "*Martabat umat Islam Melayu menurut Hamka*." Utusan Melayu, 30 April.

[9]Ruxin Chen and Jamieson L. H., (1996). "*Experiments on the Implementation of Recurrent Neural Networks for Speech Phone Recognition*." Proceedings of the Thirtieth Annual Asilomar Conference on Signals, Systems and Computers, Pacific Grove, California, November, pp. 7790782.

[10]Lee S. J., Kim K. C., Yoon H. and Cho J. W., (1991). "*Application of Fully Neural Networks for Speech Recognition*." Korea Advanced Institute of Science and Technology, Korea, Page(s): 77-80.

[11]He J. and Liu L., (1999). "*Speaker Verification Performance and The Length of Test Sentence*." Proceedings ICASSP 1999 vol.1, pp. 305-308.

[12]Gingras F. and Bengio Y., (1998). "*Handling Asynchronous or Missing Data with*

Recurrent Networks." International Journal of Computational Intelligence and Organizations, Vol. 1, no. 3, pp. 154-163.
[13]Lee S. W. and Song H. H., (1997). "*A New Recurrent Neural-Network Architecture for Visual Pattern Recognition*." IEEE Transactions On Neural Networks, Vol. 8, No. 2, March.
[14]Rabi G.and Lu S., (1997). "*Visual Speech Recognition by Recurrent Neural Networks*." Electrical and Computer Engineering, IEEE 1997 Canadian Conference on, Vol. 1, 25-28 May, Page(s): 55 -58.
[15]Werbos P., (1990). "*Backpropagation Through Time: What It Does and How To Do It.*" Proceedings of the IEEE, 78, 1550.
[16]Sato M., (1990). "*A Real Time Running Algorithm for Recurrent Neural Networks*." Biological Cybernetics, 62, 237.
[17]Ting H. N., Jasmy Yunus and Sheikh Hussain Salleh, (2002). "*Speaker-Independent Phonation Recognition For Malay Plosives Using Neural Networks*." Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Malaysia.
[18]Rabiner L. R. and Juang B. H., (1993). "*Fundamentals of Speech Recognition*." 1st. Ed. New Jersey, United States of America: Prentice Hall. 58-148.
[19]El-Iman Y. A., (1989). "*An Unrestricted Vocabulary Arabic Speech Synthesis System*." IEEE Transactions On Acoustics, Speech and Signal Processing, 37(12):1829-1845.
[20]Greenberg S., Carvey H., Hitchcock L. and Chang S., (2001). "*Beyond the Phoneme: a Juncture-Accent Model of Spoken Language*." Proceedings of the Human Language Technology Conference (HLT - 2002), San Diego, California, March 24-27.