Noisy Speech Recognition By Hierarchical Recurrent Neural Fuzzy Networks

Chia-Feng Juang and Chyi-Tian Chiou
Department of Electrical Engineering
National Chung Hsing University,
Taichung, 402 Taiwan, R.O.C.

Hao-Jung Huang
Department of Electrical Engineering, Chung Chou Institute of Technology,
Yuan-Lin, Chang-Hua, Taiwan, R.O.C.

Abstract—Noisy speech recognition by Hierarchical Recurrent Neural Fuzzy Networks (HRNFN) is proposed in this paper. The proposed HRNFN is a hierarchical connection of two recurrent neural fuzzy networks, where one is used for noise filtering and the other for recognition. And the recurrent neural fuzzy network used is the TSK-type Recurrent Fuzzy Network (TRFN), which is constructed by recurrent fuzzy if-then rules. In recognition, \( n \) TRFNs are created for \( n \) words modeling. The total prediction error of each TRFN is used as recognition criterion. In filtering, \( n \) TRFNs are created, and each TRFN recognizer is connected with a corresponding TRFN filter, which filters noisy speech patterns in the feature domain before feeding them to the recognizer. Experiments on words recognition under different types of noises are performed to verify the performance of HRNFN.

Index Terms—Speech recognition, nonlinear filter, recurrent neural network, recurrent fuzzy network.

I. INTRODUCTION

Speech recognition is a fundamental problem for any speech processing system, and automatic speech recognition is one of the leading technologies in man-machine interface for real-world applications. Many recognition algorithms based on artificial neural networks have also been proposed [1]-[4] owing to their nonlinear mapping functionality, learning ability, flexible architecture, which can easily accommodate contextual inputs and feedback. Recently, the structure of recurrent neural fuzzy network has been proposed. Especially, in [5], we have proposed a TSK-type recurrent fuzzy network (TRFN), and the performance of TRFN has been demonstrated to be better than compared recurrent network. Thus, the first objective of this paper is to propose an TRFN based isolated word recognition system.

Though most speech recognizers give adequate recognition accuracy for clean speech, their performance degrades when they are applied to noisy environments, and a rapid deterioration is seen with decreasing signal-to-noise ratio (SNR). To conquer this problem, many approaches to noisy speech recognition, like noisy speech filtering, have been proposed [6]. For noisy speech filtering, many filters designed by neural networks have been proposed to make good use of their nonlinear mapping abilities [7]-[9]. However, for these approaches, only one filter shared by all recognized words is used, and the performance of the filter degrades seriously with lower SNR values. To improve the filter performance, in this paper, noisy speech filtering by TRFN is adopted. And a hierarchical recurrent neural fuzzy network (HRNFN) is proposed, where the number of filters is equal to the class of words to be recognized. To verify the performance of the proposed HRNFN, we have tested it on a speaker-dependent isolated Mandarin word recognition task in white noise and other noise conditions for SNR ranging from 0-24 dB, and a good result is achieved.

The remainder of this paper is organized as follows: Section II describes the structure and learning algorithms of TRFN. Speech recognition by TRFN recognizer is given in section III. Section IV introduces the HRNFN for noisy speech recognition. Section V demonstrates the recognition results by under clean and noisy environments. Finally, a conclusion is made in section VI.

II. STRUCTURE AND LEARNING OF TRFN

The structure of TRFN is shown in Fig. 1. Each rule in TRFN is of the following form:

\[
\text{Rule } i: \text{IF} \ x_{j}(t) \text{ is } A_{j} \text{ And } \cdots \text{ And } x_{m}(t) \text{ is } A_{m} \text{ And } h(t) \text{ is } G
\]

Then \( y(t+1) = a_{0} + \sum_{j=1}^{n} a_{j} x_{j}(t) + a_{h} h(t) \)

And \( h(t+1) = w_{u} \text{ And } \cdots \text{ And } h_{r}(t+1) = w_{r} \)

where \( A \) and \( G \) are fuzzy sets, \( w \) and \( a \) are the consequent parameters for inference output \( h \) and \( y \), respectively, \( n \) is the number of external input variables, and \( r \) is the number of rules. Functions of TRFN are described layer by layer below. For notation convenience, the net input to the \( i \) th node in layer \( k \) is denoted by \( u_{i}^{(k)} \) and the output value by \( O_{i}^{(k)} \).


Fig. 1. Structure of TRNFN.

Fig. 2. Architecture of the TRFN for speech recognition.

Fig. 3. Architecture of HRNFN for noisy speech recognition.

Table 1. Performances Of Different Types Of Recognizers For Recognizing Ten Words Under Clean Environment.

<table>
<thead>
<tr>
<th>Recognizer Type</th>
<th>Training Data</th>
<th>Test Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRFN</td>
<td>100%</td>
<td>99%</td>
</tr>
<tr>
<td>TDNN</td>
<td>100%</td>
<td>96%</td>
</tr>
<tr>
<td>MLP</td>
<td>100%</td>
<td>83%</td>
</tr>
<tr>
<td>HMM</td>
<td>100%</td>
<td>99%</td>
</tr>
</tbody>
</table>

Table 2. Test Data Recognition Rates Under WHITE Noise.

<table>
<thead>
<tr>
<th>SNR</th>
<th>0</th>
<th>6</th>
<th>12</th>
<th>18</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRNFN</td>
<td>79%</td>
<td>92%</td>
<td>97%</td>
<td>99%</td>
<td>99%</td>
</tr>
<tr>
<td>TRFN+RAFF</td>
<td>13%</td>
<td>36%</td>
<td>52%</td>
<td>71%</td>
<td>82%</td>
</tr>
<tr>
<td>TRFN</td>
<td>12%</td>
<td>33%</td>
<td>44%</td>
<td>57%</td>
<td>74%</td>
</tr>
<tr>
<td>HMM</td>
<td>2%</td>
<td>13%</td>
<td>32%</td>
<td>46%</td>
<td>70%</td>
</tr>
</tbody>
</table>