文章编号: 1001-0920(2003)02-0213-04

# 模糊神经网络语音数据融合算法的研究

梅晓丹, 张毅刚, 孙圣和
(哈尔滨工业大学 自动化测试与控制系, 黑龙江 哈尔滨 150001)

**摘 要**: 针对高噪音环境中的语音识别问题, 提出一种利用模糊神经网络进行语音数据融合的新算法。该算法按一定模糊规则对语音信号的特征参数进行模糊化, 并通过神经网络对每个传感器语音信号的模糊特征参数进行分类和融合。仿真实验表明, 该算法鲁棒性更强; 与单传感器算法相比, 语音识别率得到较大的提高。

**关键词**: 模糊神经网络; 数据融合; 语音识别
**中图分类号**: TP183 **文献标识码**: A

# Research on speech data fusion based on fuzzy neural networks

*MEI Xiao-dan, ZHANG Yi-gang, SUN Sheng-he*
(Department of Automatic Test and Control, Harbin Institute of Technology, Harbin 150001, China)

**Abstract**: A new algorithm of employing the fuzzy neural network is proposed to realize speech data fusion for speech recognition under high noisy condition. The scheme fuzzifies feature parameters of speech signal from every sensor, then classifies and fuses these fuzzified feature parameters by neural networks. The simulation shows that the algorithm has better robustness and the speech recognition rate is much improved compared with the traditional speech recognition using single sensor.

**Key words**: Fuzzy neural network; Data fusion; Speech recognition

## 1 引 言

识别率和对环境的适应能力是评价语音识别系统性能的两个重要指标。人们往往通过改进声音模型来提高识别率, 但这常会造成声音模型的复杂化和模型训练的艰难化。同时, 这些方法大多局限于安静环境, 而在复杂环境下, 当说话人与话筒之间的距离不固定或通过不同录音设备和传输媒质时, 效果会很差[1, 2]。

数据融合技术可以克服单一传感器的局限性。通过综合多传感器提供的各个侧面信息, 能获得观测对象全面而准确的信息。近年来, 将数据融合技术用于模式识别已成为信息领域的一个重要研究方向[3]。模糊集理论中隶属函数的概念打破了传统集合理论中明确的集合界限, 可以处理非定量的信息。它更接近于人的表述和思维方式, 在数据融合中得到了广泛应用[4, 5]。人们将神经网络、模糊神经网络用于语音识别[6~8], 取得了一定的效果。

本文针对实验室环境下多话筒和多传输媒质语音识别问题, 提出一种基于模糊神经网络的语音数据融合模型。文中介绍了用于语音数据融合的模糊神经网络模型和语音特征参数模糊化方法, 并给出了仿真实验结果。

## 2 模糊神经网络语音数据融合模型

语音信号的特征参数在量化和传递过程中, 会产生不精确和不完整的信息, 从而使语音识别缺乏语义性。模糊理论中隶属函数的概念可在一定程度上弥补这些缺点。由隶属度划分以及隶属函数控制输入特征矢量和描述输入信号的程度, 能为系统提

**收稿日期**: 2001-10-19; **修回日期**: 2002-01-11。
**作者简介**: 梅晓丹(1974—), 女(满族), 黑龙江伊春人, 博士生, 从事语音信号识别与压缩的研究; 孙圣和(1937—), 男, 山东荣城人, 教授, 博士生导师, 从事信号处理、数据压缩等研究。

供更全面的语音信息, 提高识别的鲁棒性。本文提出的模糊神经网络语音数据融合结构如图 1 所示。它是将语音特征参数模糊化后输入神经网络。

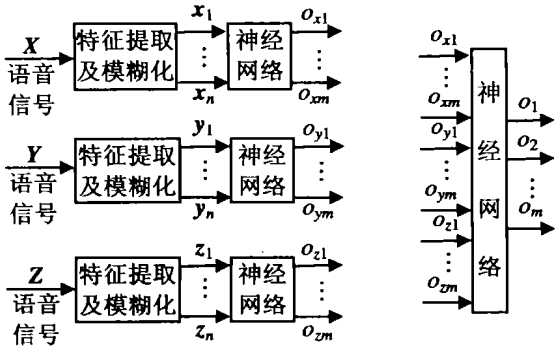


图 1 语音数据融合模糊神经网络结构

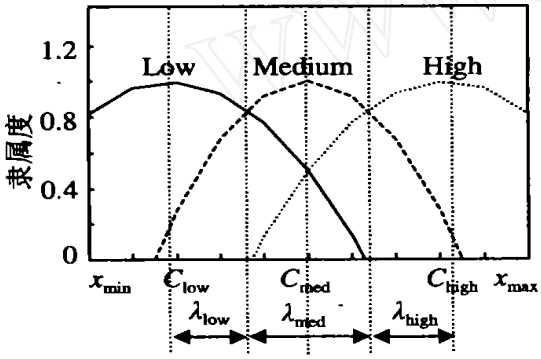


图 2 重叠的 π 型隶属度函数

模糊化采用的隶属度函数是图 2 所示的 π 型函数, 它由两个参数 $\{c, \lambda\}$ 确定, 即

$$\pi mf(x; c, \lambda) = \begin{cases} 2\left(1 - \dfrac{|x - c|}{\lambda}\right)^2, & \dfrac{\lambda}{2} \le |x - c| \le \lambda \\ 1 - 2\left(\dfrac{|x - c|}{\lambda}\right)^2, & 0 \le |x - c| \le \dfrac{\lambda}{2} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

式(1) 中隶属函数的参数确定如下

$$\begin{cases} \lambda_{med}(x) = \dfrac{1}{2}(x_{max} - x_{min}) \\ C_{med}(x) = x_{min} + \lambda_{med}(x) \end{cases} \quad (2)$$

$$\begin{cases} \lambda_{low}(x) = \dfrac{1}{f_d}(C_{med}(x) - x_{min}) \\ C_{low}(x) = C_{med}(x) - \dfrac{1}{2}\lambda_{low}(x) \end{cases} \quad (3)$$

$$\begin{cases} \lambda_{high}(x) = \dfrac{1}{f_d}(x_{max} - C_{med}(x)) \\ C_{high}(x) = C_{med}(x) + \dfrac{1}{2}\lambda_{high}(x) \end{cases} \quad (4)$$

其中 $f_d$ 是调节因子, 其典型值为 0.8。它可保证每个特征参数的隶属度值至少有一个大于 0.5。

从图 1 可以看出, 该融合结构由 4 部分组成。结构左侧为 3 个模糊神经网络, 表示有 3 种传感器输入。显然, 它可拓展成更多种传感器输入。

用 $x_i = [x_{i1}, x_{i2}, \cdots, x_{in}]$ 表示第 1 个传感器 X 的第 $i$ 个语音特征矢量。按式(1) 模糊化后, 变换成模糊特征矢量

$$x_{iF} = [x_{i11}, x_{i12}, x_{i13}, \cdots, x_{ip1}, x_{ip2}, x_{ip3}]$$
$$i = 1, 2, \cdots, n \quad (5)$$

其中: $p$ 表示特征矢量的维数, $n$ 表示特征矢量的个数。

同样, 第 2 个和第 3 个传感器 Y 和 Z 的语音特征矢量, 也按式(1) 模糊化后变换成模糊特征矢量。

将 3 个传感器的模糊语音特征矢量分别输入 3 个模糊神经网络, 并进行分类训练, 将得到 3 个输出, 即 3 种语音分类结果

$$\begin{cases} O_x = [o_{x1}, o_{x2}, \cdots, o_{xm}] \\ O_y = [o_{y1}, o_{y2}, \cdots, o_{ym}] \\ O_z = [o_{z1}, o_{z2}, \cdots, o_{zm}] \end{cases} \quad (6)$$

其中 $m$ 表示待识别词的个数。

结构右侧为融合神经网络, 目标是融合 3 个模糊神经网络的分类结果, 得到更高的识别率。其输入为前面 3 个模糊神经网络的分类结果, 输出为最后的语音识别结果

$$O = [o_1, o_2, \cdots, o_m] \quad (7)$$

其维数为 $m$, 即融合结构有 $m$ 个识别模式。

## 3 仿真验证

仿真实验在 IBM-PⅢ600 计算机上进行, 语音数据是一个男性和一个女性的 1～10 共 10 个数字的英语发音; 传感器数目为 3, 分别为家用话筒、PC 机自带话筒和电话线; 录音环境为实验室环境, 采样率为 22.05 kHz。同一语音信号“four”经 3 种传感器输出的 10 000 个采样波形如图 3～图 5 所示。

语音预处理包括: 对语音进行预加重, 传输函数为 $H(z) = 1 - 0.95z^{-1}$; 对每个词进行分帧和加 Hamming 窗, 每帧为 20$m$; 计算特征矢量和 10 阶线谱参数 LSFs; 进行特征矢量的规整, 将每个词规整为 20 帧, 即每个词由 200 个特征参数表示。

模糊化后, 图 1 网络结构左侧的每个神经网络的输入数据变为 600 个特征参数, 即每个神经网络的输入层有 600 个神经元, 第 1 隐层和第 2 隐层分别取 200 个和 300 个神经元, 输出层有 10 个神经元, 对应于 10 个数字的分类结果。结构右侧神经网络的输入为左侧 3 个神经网络的输出, 其输入层有 30 个神经元, 隐层取 20 个神经元, 输出层有 10 个神经元,

对应于最后的识别结果。

4 个神经网络的训练均采用BP 算法, 每个神经元的传递函数均采用 Sigmoid 函数; 学习率、最大训练次数和误差目标分别取 0. 1, 1 000 和 0. 01; 实验取 450 个训练样本, 200 个测试样本。在不同训练样本数量和训练次数情况下, 模糊神经网络融合与单传感器方法识别率的比较列于表 1。

表 1 模糊神经网络融合与单传感器方法的识别率比较

| 单传感器与融合方法 | 训练样本数目 | 训练次数 | 训练时间 s × $10^{-3}$ | 识别准确率 /% |
|---|---|---|---|---|
| PC 机话筒 | 150 | 1 000 | 0. 941 | 66. 5 |
| | | 500 | 0. 484 | 64. 5 |
| | 300 | 1 000 | 1. 455 | 76. 5 |
| | | 500 | 0. 726 | 70. 0 |
| | 450 | 1 000 | 2. 086 | 82. 0 |
| | | 500 | 1. 037 | 75. 5 |
| 电话线传输 | 150 | 1 000 | 0. 974 | 63. 0 |
| | | 500 | 0. 487 | 62. 0 |
| | 300 | 1 000 | 1. 462 | 73. 0 |
| | | 500 | 0. 731 | 65. 0 |
| | 450 | 1 000 | 2. 065 | 80. 5 |
| | | 500 | 1. 041 | 73. 5 |
| 家用话筒 | 150 | 1 000 | 0. 956 | 74. 5 |
| | | 500 | 0. 481 | 73. 0 |
| | 300 | 1 000 | 1. 469 | 79. 5 |
| | | 500 | 0. 724 | 78. 5 |
| | 450 | 1 000 | 2. 081 | 92. 0 |
| | | 500 | 1. 033 | 79. 0 |
| 融合方法 | 150 | 1 000 | 3. 344 | 73. 0 |
| | | 500 | 1. 413 | 64. 0 |
| | 300 | 1 000 | 5. 084 | 92. 5 |
| | | 500 | 2. 497 | 90. 5 |
| | 450 | 1 000 | 7. 741 | 94. 5 |
| | | 500 | 4. 174 | 91. 0 |

从仿真结果可以看出, 当训练样本较少时, 融合效果并不十分突出; 增加训练样本数量后, 融合算法比传统的单一话筒的识别率高 10% 左右, 取得了良好的效果。在某些传感器输出的语音信号的误识率大于 30% 的情况下, 仍能取得较高的识别率。

为了比较噪音对传统神经网络融合和模糊神经网络融合的影响, 本文采用传统神经网络进行分类和融合的仿真实验。此时, 分类神经网络的输入层、第 1 隐层和第 2 隐层分别取 200, 100 和 100 个神经元, 其他参数不变, 融合神经网络不变。实验产生了正态分布噪声, 加在语音数据上。噪声信号通过一个系数来调节信噪比。神经网络融合和模糊神经网络融合在训练样本数为 450, 训练次数为 1 000 的情况下, 不同信噪比对识别率的影响比较列于表 2。

表 2　神经网络融合与模糊神经网络融合的比较

| 网络结构 | 信噪比 dB | 训练时间 s × $10^3$ | 识别准确率 /% |
|---|---|---|---|
| 神经网络 | 6 | 1. 293 | 79. 0 |
| | 18 | 1. 244 | 84. 5 |
| | 35 | 1. 558 | 86. 5 |
| 模糊神经网络 | 6 | 7. 413 | 90. 0 |
| | 18 | 7. 819 | 91. 5 |
| | 35 | 7. 107 | 94. 0 |

在实验结果中, 当信噪比为 35 dB 时, 模糊神经网络融合方法比传统神经网络融合方法的识别率提高7. 5%; 当信噪比为18dB 时, 模糊神经网络融合

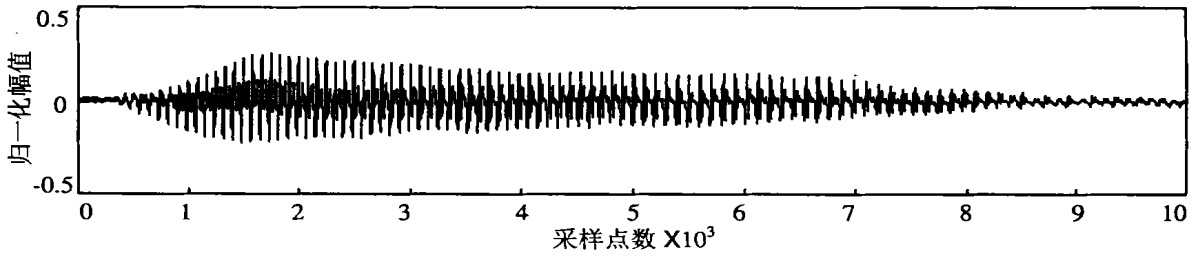


图 3　家用电筒的波形

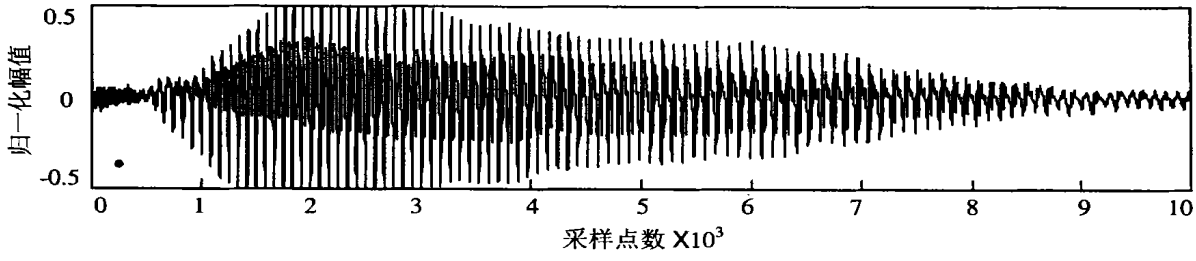


图 4　PC 机话筒的波形

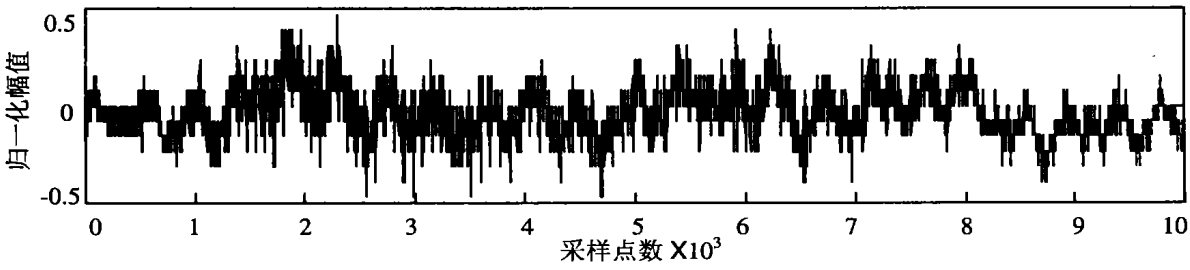


图 5 电话线的波形

方法比传统神经网络融合方法的识别率提高 7%。在高噪音的影响下(如信噪比为 6 dB),模糊神经网络融合方法的识别率提高 11%。由此可见,在信噪比较低的情况下,模糊神经网络融合方法的鲁棒性更加突出。

## 4 结 论

本文提出一种模糊神经网络语音数据融合模型,并对经不同传感器记录的语音数据采用该模型进行识别。仿真结果显示,采用该模型比传统单一话筒的识别率提高了十多个百分点,并且鲁棒性更强,说明采用模糊神经网络进行语音数据融合可以有效地提高语音识别的准确率。

**参考文献**(References):

[1] Leida E, Fernandez J, Masgrau E. Robust continuous speech recognition system based on a microphone array [A]. *IEEE ICASSP-Proc* [C]. Seattle, 1998. 1: 241-244.

[2] Yamada T, Nakamura, Shikano K. Hands-free speech recognition based on 3-D viterbi search using a microphone array [A]. *IEEE ICASSP-Proc* [C]. Seattle, 1998. 1: 245-248.

[3] Hall D L, Llinas J. An introduction to multisensor data fusion [J]. *Proc IEEE*, 1997, 85(1): 6-23.

[4] Russo F, Ramponi G. Fuzzy methods for multisensor data fusion [J]. *IEEE Trans IM*, 1994, 43(2): 288-293.

[5] Choi J, Dickerson J A. Adaptive data fusion using the expected output membership function [J]. *Proc SPIE*, 1999, 3719: 26-33.

[6] Nikola K. Evolving connectionist systems: A theory and a case study on adaptive speech recognition [A]. *Proc Int Joint Conf Neural Networks* [C]. Washington, 1999. 3002-3007.

[7] Francesco B, Salvatore C, Marco R. Pattern recognition approach to robust voiced/unvoiced speech classification using fuzzy logic [J]. *Int J Pattern Recog Artif Intell*, 1999, 13(1): 109-132.

[8] Wu G D, Lin C T. Word boundary detection with mel-scale frequency bank in noisy environment [J]. *IEEE Trans Speech Audio Proc*, 2000, 8(5): 541-554.

(上接第 209 页)

**参考文献**(References):

[1] Richalet J. Model predictive heuristic control: Applications to industrial process [J]. *Automatica*, 1978, 14(5): 413-428.

[2] Mahra R K. Model algorithmic control (MAC), basic theoretical properties [J]. *Automatica*, 1982, 18(4): 401-404.

[3] Cutler C R, Ramaker B L. Dynamic matrix control — A computer control algorithm [A]. *Proc of the Joint Automatic Control Conf* [C]. San Francisco, 1980. WP5-B.

[4] Clarke D W, Mohtadi C, Tuffs P S. Generalized predictive control—Ⅰ: The basic algorithm [J]. *Automatica*, 1987, 23(2): 137-148.

[5] Clarke D W, Mohtadi C, Tuffs P S. Generalized predictive control—Ⅱ: Extension and interpretations [J]. *Automatica*, 1987, 23(2): 149-160.

[6] Qin S J, Badgwell T A. An overview of industrial model predictive control technology [A]. *AIChE Symposium Series — 5th Int Conf on Chemical Process Control* [C]. Tahoe, 1996. 316(93): 232-256.

[7] 席裕庚, 耿晓军, 陈虹. 预测控制性能研究的新进展 [J]. 控制理论与应用, 2000, 17(4): 469-475.
(Xi Y G, Geng X J, Chen H. Recent advances in research on predictive control performance [J]. *Control Theory and Applications*, 2000, 17(4): 469-475.)

[8] Smith D J M. Closer control of loops with dead time [J]. *Chemical Engineering Progress*, 1957, 53(5): 271-280.