

文章编号: 1001-0920(2003)03-0364-03

基于粗神经网络的仿人智能机器人的语音融合算法研究

刘国良, 强文义, 麻 亮, 陈兴林

(哈尔滨工业大学 控制科学与工程系, 黑龙江 哈尔滨 150001)

摘 要: 将粗集合理论与神经网络相结合, 提出一种基于粗神经网络的新的信息融合方法, 用于仿人智能机器人的语音融合。该方法不仅可以接受定量输入, 而且可以接受定性输入, 即输入是一个范围, 或在观测时间内输入是变化的。由于粗神经网络的误差传递函数不可微, 所以采用遗传算法来训练粗神经网络。仿真实验结果表明, 基于粗神经网络的信息融合方法有效地提高了语音的识别率。

关键词: 粗神经网络; 信息融合; 语音识别; 仿人智能机器人

中图分类号: TP18

文献标识码: A

Speech fusion based on rough neural network for humanoid intelligent robots

LIU Guo-liang, QIANG Wen-yi, MA Liang, CHEN Xing-lin

(Department of Control Science and Engineering, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Integrating rough set theory with neural network theory, a novel information fusion method based on rough neural network is proposed. It is used in speech fusion of humanoid intelligent robots. The input could be qualitative, i. e. a range or variational during the observation, as well as quantitative. Because the error transfer function of rough neural network is not differentiable, genetic algorithms are applied for training the network. The simulation results indicated that the novel information fusion method based on rough neural network does improve the speech recognition probability.

Key words: Rough neural network; Information fusion; Speech recognition; Humanoid intelligent robot

1 引 言

语音与对话是仿人智能机器人的一项重要功能。通过智能控制系统, 仿人智能机器人可以接受人类通过语言发布的命令, 并分析命令的含义, 从而决定下一步采取什么动作。识别率是语音识别系统的一个重要特性, 人们经常通过改进声音模型来提高识别率, 而这往往会造成声音模型的复杂化, 为模型训练带来困难。本文提出一种基于粗神经网络

的语音信息融合方法。仿真实验表明, 该方法有效地提高了语音系统的识别能力。

粗神经网络是将粗集理论融入神经网络而形成的一种数据处理方法。粗集理论^[3,4]模拟人类的抽象逻辑思维, 将数据中推理逻辑规则作为知识系统的模型; 神经网络则是模拟人类的形象直觉思维, 利用非线性映射的思想表达输入与输出的关联知识。

传统的神经网络一般不能处理语义形式的输

收稿日期: 2002-01-29; 修回日期: 2002-06-03。

基金项目: 哈尔滨工业大学校基金资助项目(HIT. 2001. 03)。

作者简介: 刘国良(1968—), 男, 内蒙古兴安盟人, 博士生, 从事信息融合、仿人智能机器人研究; 强文义(1937—), 男, 江苏无锡人, 教授, 博士生导师, 从事智能控制和过程控制研究。

入, 其输入输出信息必须是定量的。即使在模糊神经网络中, 输入信息也由隶属度定量表示。在多传感器信息融合以及其他很多实际问题中, 传感器的输出有时只能定性地给出, 即传感器的输出不是单一明确的值, 而是一个范围; 或者说, 在观测时间间隔内, 传感器的输出是变化的。将粗集理论融入神经网络可以大大改善网络的性能, 使之不仅能处理传统的定量输入, 而且能处理定性或混合性的输入信息, 解决多传感器不同种类信息的融合问题。

粗神经网络的误差传递函数和传统神经网络的误差传递函数在形式上是一致的。由于 BP 算法在修正权值时必须计算误差传递函数梯度, 而粗神经网络的误差传递函数是不可微的, 所以本文不用 BP 算法而采用遗传算法^[5]来训练粗神经网络。

2 基于粗神经网络的语音信息融合

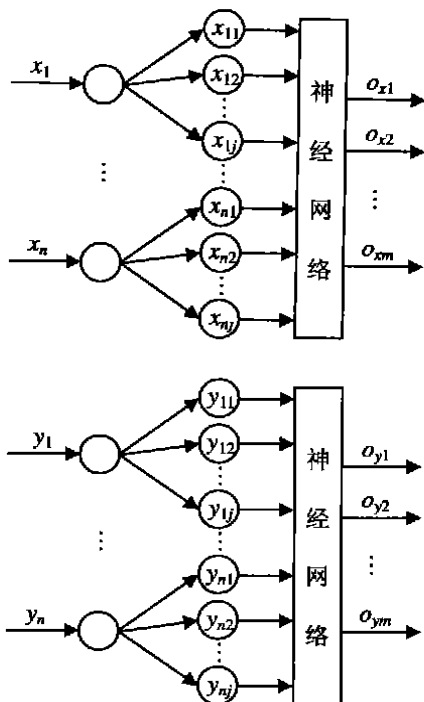


图 1 基于模糊神经网络的单传感器语音识别结构

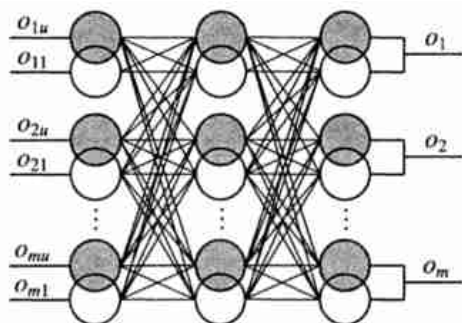


图 2 基于粗神经网络的语音融合结构

基于粗神经网络的语音信息融合结构的前端是基于模糊神经网络的单传感器语音识别结构, 如图 1 所示。基于粗神经网络的语音融合结构如图 2 所示。

图 1 中 $x = [x_1, x_2, \dots, x_n]$ 和 $y = [y_1, y_2, \dots, y_n]$ 分别表示来自两个传感器的一个语音数据, 显然它可以扩展为更多的传感器输入。对语音数据分别模糊化后变成

$$\begin{cases} xF = [x_{11}, x_{12}, \dots, x_{1j}, x_{21}, x_{22}, \\ \dots, x_{2j}, x_{n1}, x_{n2}, \dots, x_{nj}] \\ yF = [y_{11}, y_{12}, \dots, y_{1j}, y_{21}, y_{22}, \\ \dots, y_{2j}, y_{n1}, y_{n2}, \dots, y_{nj}] \end{cases} \quad (1)$$

其中: n 表示一个输入语音数据的最大特征数, j 表示模糊子集数, $o_x = [o_{x1}, o_{x2}, \dots, o_{xm}]$ 和 $o_y = [o_{y1}, o_{y2}, \dots, o_{ym}]$ 分别表示模糊神经网络对语音的分类输出, m 表示待识别词的种类数。

图 2 的输入为图 1 中两个模糊神经网络分类输出的结果。由于粗神经元的输入有两个值, 分别相当于粗集合的上下近似值, 所以应保证同一粗神经元上层的输入值大于下层的输入值。即 $o_{xu} = \max(o_{xi}, o_{yi})$ 作为粗神经元的上近似 ($i = 1, 2, \dots, m$), $o_{xl} = \min(o_{xi}, o_{yi})$ 作为粗神经元的下近似 ($i = 1, 2, \dots, m$)。本文粗神经元上下两层的传递函数 f_U 和 f_L 均取为正切 Sigmoid 函数, 有

$$f_U(x) = f_L(x) = \frac{2}{1 + e^{-x}} - 1 \quad (2)$$

对于粗神经元, 经过一系列变换, 便可得到与传统神经元相一致的误差传递函数

$$E_p = \frac{K^2}{2} \sum_{p=1}^m [(O_p - \hat{O}_p)]^2 \quad (3)$$

其中: O_p 和 \hat{O}_p 分别为粗神经元上下两层理想输出的差值和实际输出的差值, p 为当前输入样本, K 为粗神经元上下两层的加权系数。但是传统神经网络的 BP 学习算法并不能在此借用, 因为每个粗神经元的输出有两个值且有大小之分, 很难直接求导数, 所以选用遗传算法来训练粗神经网络。利用遗传算法对给出的样本数据进行训练, 得到最优粗神经网络参数。粗神经网络的输出 o_1, o_2, \dots, o_m 为融合后的语音识别结果, 取输出层粗神经元上下层的平均值。

3 仿真实验

仿真实验为基于特定人的语音识别实验。传感器数目为 2, 分别为 PC 机话筒和电话线; 语音数据由一男性的发音 (从 1 ~ 10 共 10 个数字的英语发音) 数据组成; 录音环境为实验室安静环境, 采样频

率为 22.05 kHz。

首先对语音进行预加重, 传输函数为 $H(z) = 1 - 0.95z^{-1}$; 然后对每个单词进行分帧, 每帧为 20 ms, 帧间叠接为 10 ms; 接着对每帧信号加 Hamming 窗; 最后计算特征量。由于特征矢量选 10 阶的线性预测系数, 每个词分为 20 帧, 所以每个词共有 200 个特征参数。按式(1) 进行模糊化, 模糊子集数为 3, 隶属度函数取带重叠的 π 型函数。

模糊化后, 图 1 中的每个神经网络的输入数据变为 600 个特征参数, 即每一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 和 100 个神经元; 输出层有 10 个神经元, 对应 10 个数字的发音。图 2 所示的粗神经网络的输入为图 1 中两个神经网络的输出, 其输入层、隐层和输出层都取 10 个粗神经元, 结果取输出层神经元的上下近似的均值。前面两个神经网络的训练方法采用 BP 算法, 后面粗神经网络的训练方法选用遗传算法。训练样本数目分别取 300, 600 和 900 个。表 1 给出了采用

表 1 基于粗神经网络的语音融合的仿真结果

识别方法	训练样本数目	训练次数	训练时间 $s \times 10^3$	识别率 / %
PC 机话筒	300	1 000	1.732	90.5
		500	1.021	88.1
	600	1 000	2.690	94.0
		500	1.379	92.7
	900	1 000	3.987	94.8
		500	2.037	93.5
电话线传输	300	1 000	1.890	92.3
		500	0.970	90.0
	600	1 000	2.935	92.9
		500	1.228	92.4
	900	1 000	4.080	94.2
		500	1.934	92.6
粗神经网络融合	300	1 000	7.055	93.2
		500	2.512	91.5
	600	1 000	9.899	95.5
		500	4.879	95.0
	900	1 000	13.047	97.8
		500	8.436	95.4

本文提出的粗神经网络语音融合模型与使用模糊神经网络的单一传感器的识别率在不同训练样本数量及训练次数时的比较。

从表 1 可以看出, 训练样本数较小时, 融合效果不很突出; 增多训练样本数量和训练次数后, 粗神经网络的融合结果与传统的单一话筒的模糊神经网络识别的识别率相比有显著提高。

4 结 论

本文将粗集合理论融入神经网络, 提出了一种用于语音识别的粗神经网络融合结构。仿真实验结果表明, 采用粗神经网络的语音融合算法有效地提高了系统的语音识别率。将粗集合理论融入神经网络可以提高神经网络的性能, 使其不仅能处理传统的定量输入, 而且能处理定性或混合性的输入信息, 解决多传感器不同种类信息的融合问题。该方法还可以应用于图像处理或其它领域, 以作进一步的验证。

参考文献(References):

- [1] Lingars P J. Rough neural networks[A]. *Proc of 6th Int Conf on Information Processing and Management of Uncertainty in Knowledge-based Systems*[C]. Granada, 1996. 1445-1450.
- [2] Lingars P J. Comparison of neofuzzy and rough neural networks[J]. *Information Sciences*, 1998, 110(3): 207-215.
- [3] Zdzislaw Pawlak. *Rough Sets: Theoretical Aspects of Reasoning About Data* [M]. Dordrecht: Kluwer Academic Publishers, 1991.
- [4] Akash Mohabey, Ray A K. Rough set theory based segmentation of color images[A]. *Proc of IEEE 19th Int Meeting of the North American Fuzzy Information Proc Society* [C]. Atlanta, 2000. 338-342.
- [5] Goldberg D E. *Genetic Algorithms in Search, Optimization and Machine Learning* [M]. Boston: Addison-Wesley Publishing Company, 1989. 1-15.