

粗集神经网络系统及其应用*

舒 兰** 吴 磊 张 勇

(电子科技大学应用数学学院 成都 610054)

【摘要】讨论了粗集神经网络系统及其在心电图自动识别中的应用。根据识别需要,从MIT心电图数据库获取心电图特征参数,根据粗集理论对属性进行优化,并用BP神经网络进行学习。运用这些特征参数和训练好的BP神经网络对心电图进行分类,识别准确率达到90%。

关键词 粗集理论; 心电图; 决策表; 特征参数; BP神经网络

中图分类号 O129; O235

Rough Neural Network System and Its Application in Electrocardiogram Auto-recognition

Shu Lan Wu Lei Zhang Yong

(Dept. of Applied Mathematics, UEST of China Chengdu 610054)

Abstract In this paper, a rough neural network system is given and applied in Electrocardiogram auto-recognition. Based on medical requires, this paper has picked up the characteristic parameters of signals from Massachusetts Institute of Technology (database). According to Rough sets theory, the attributes are optimized and studied with BP neural network. In the end, the Electrocardiogram samples are classified with those characteristic parameters. The result shows that the right ratio goes to 90%. The experiment shows that the system has the advantages of fast computation and easy realization and the system is better than other methods in optimizing attributes and enhancing the right ratio of classification.

Key words rough sets theory; electrocardiogram; decision form; characteristic parameter; BP neural network

近几年来,由波兰华沙理工大学Z.Pawlak教授等一批科学家提出的粗集理论是用于数据分析的一个数学分支^[1],它可以对不完整数据、不确定和不精确数据进行分析和推理,发现数据间的关系,提取有用特征,是智能信号处理有力的数学工具。目前,这个理论已被国外学者应用于医疗数据分析、语言识别、金融与股票市场分析等领域,并且取得了丰硕的成果,而国内关于粗集理论的研究刚刚起步。但粗集理论也存在容错能力与推广能力相对软弱,且只能处理量化数据等问题。而人工神经网络具有较强的自组织能力、容错能力与推广能力,但不能优选条件属性组合等特点。因此,将粗集与神经网络进行结合更具有前沿性、科学性及优越性。

心电图自动分析技术是一门综合性强的多学科技术,目前世界上研制出的心电图分析仪都存在很多不足,尤其在复杂心电图诊断方面。本文利用粗集理论与人工神经网络各自的长处,提出了粗集神经网络系统,并将该系统应用于心电图的分类识别,得到了令人满意的结果。

2001年10月26日收稿

* 国家自然科学基金资助项目,编号:69803007

** 女 39岁 硕士 教授

1 决策表简化方法

决策表是一类特殊而重要的知识表达系统。决策表可以根据知识表达系统来描述。而知识表达系统(KRS)的基本成分是研究对象的集合,关于这些对象的知识可通过指定对象的基本特征(属性)和其特征值(属性值)来描述。一个知识表达系统 S 可表达为

$$S = \langle U, C, D, V, f \rangle$$

式中 U 为对象的集合, $C, D \subseteq R$ 是属性的集合, 子集 C 和 D 分别称为条件属性和结果属性; $V = \bigcup_{r \in R} V_r$ 是属性值的集合, V_r 表示了属性 $r \in R$ 的属性范围; $f: U \times R \rightarrow V$ 是一个信息函数, 它指定 U 中每一对象 x 的属性值, 这样定义的知识表达系统可以方便地用表格来实现。

决策表也是知识表达系统, 根据知识表达系统作定义如下:

$S = (U, A)$ 为一知识表达系统, 且 $C, D \subseteq A$ 是两个属性子集, 分别称为条件属性和决策属性, 具有条件属性和决策属性的知识表可表达为决策表, 记为 $T = (U, A, C, D)$ 或简称 CD 决策表。关系 $ind(C)$ 和 $ind(D)$ 的等价类分别称为条件类和决策类。

决策表的简化就是简化决策表的属性, 化简后的决策表具有与化简前决策表相同的功能, 且化简后的决策表具有更少的条件属性。利用决策表对知识简化, 首先要进行条件属性的简化, 消去重复行, 然后对每一决策规则进行冗余属性值的简化, 合并重复行, 导出简化决策表, 即得到最小解的简单算法。

下面根据决策表简化理论得到有关定理^[2], 以用于决策表简化的枚举算法的优化设计。

定理1 设 U 为论域, R 为属性集合, 其中 $A \subset R$ 且 $A \neq F$, A 为 R 中不可省略, 则对 $\forall B \subset A$ 且 $B \subset R (B \neq F)$, B 也为 R 中不可省略的。

证明 因为 A 为 R 中不可省略的, 由定义得

$$ind(R) \neq ind(R - A)$$

且 $ind(R) \subseteq ind(R - A)$ (1)

因为 $B \supset A$, 所以

$$R - A \supset R - B, ind(R - A) \subseteq ind(R - B) \quad (2)$$

由式(1)、(2)得 $ind(R) \subseteq ind(R - B)$, 所以不存在 $ind(R) = ind(R - B)$, 故 B 为 R 中不可省略的。

类似可得定理2与定理3。

定理2 设 U 为论域, R 和 A 均为属性集合, $A \subset R$, R_1 为单个属性集合, 其中令 $A' = \{A, R_1\}$, 则属性 A' 与属性 A 的依赖度 $r_A(A') = 1$ 。

定理3 设 U 为论域, R 和 A 均为属性集合, $A \subset R$, R_1 为单个属性集合, 其中令 $A' = \{A, R_1\}$, 则 $0 < r_A(A') < 1$, 当 $ind(A') = ind(A)$ 时, $r_A(A') = 1$ 。

2 粗集神经网络系统

对心电图进行分析是心脏疾病的重要检查方法之一^[3], 而最为常见的三类心电信号是正常(Normal), 室性失常(PVC)和束支传导阻滞(BBB)。本文对MIT-BIH心电数据库中的这三类心电信号进行识别。其基本思想如下: 利用波形检测算法^[2]进行参数提取, 对参数属性进行量化, 量化后的数据如表1所示; 用粗集理论对量化后的属性进行优化, 对优化后的属性选用BP网络进行学习, 再用量化参数及训练好的BP网络对待识样本进行分类。粗集神经网络系统如图1所示。

利用决策表简化算法得到多个条件属性的简化, 其中属性个数最少的属性简化为 $\{e, h, i\}$, 即得到 P - R 间期、QRS波峰高和 P 波峰高这三个属性的优化组合。

本系统采用 $3 \times 4 \times 3$ 型BP神经网络来实现, 其中输入节点3个, 分别对应所抽取的8项特征指标

表 1 量化后的数据表

序号	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>	<i>k</i>	<i>l</i>	<i>d</i>
1	0	0	1	1	0	0	1	0	0
2	0	0	1	1	0	0	1	1	0
3	0	0	0	1	0	0	1	0	0
4	0	1	1	1	0	1	0	0	0
5	0	1	0	1	0	1	0	1	0
6	1	0	0	2	1	0	0	1	1
7	1	1	0	2	1	1	0	0	1
8	1	1	0	2	1	1	0	1	1
9	2	1	1	0	2	1	2	1	2
10	2	0	1	0	2	0	2	0	2
11	2	1	1	0	2	0	2	1	2

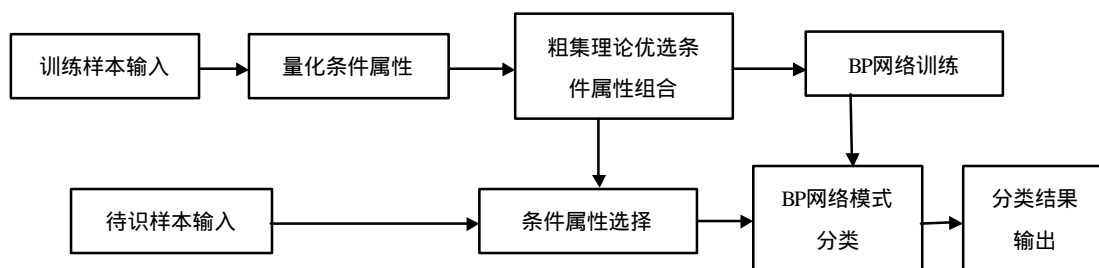


图1 粗集神经网络系统框图

中最后优化的3项特征指标，即*P-R*间期、QRS波峰高和*P*波峰高。期望输出为

$$y = (y_1, y_2, y_3) = \begin{cases} 100 & \text{对应1类} \\ 010 & \text{对应2类} \\ 001 & \text{对应3类} \end{cases}$$

其中 1类代表正常(Normal)，2类代表室性失常(PVC)，3类代表束支传导阻滞(BBB)。

实验结果表明，本系统平均识别准确率为90%，达到了预期的效果。

参 考 文 献

- 1 曾黄麟. 粗集理论及其应用—关于数据推理的新方法. 重庆: 重庆大学出版社, 1998
- 2 张 勇. 粗集理论及其在心电图自动分析中的应用研究. [学位论文]. 成都: 电子科技大学, 2000
- 3 黄 宛. 临床心电图学. 北京: 人民卫生出版社, 1975