

# 粗糙集理论及其在智能系统中的应用

伞冶,叶玉玲

(哈尔滨工业大学 控制与仿真中心,黑龙江 哈尔滨 150080)

**摘要:**粗糙集理论是一种新型的处理含糊和不确定知识的数学工具,在智能系统中得到了广泛的应用.介绍了经典粗糙集理论的基本思想,上下近似集、属性约简和核等基本概念以及粗糙集的研究现状.介绍了粗糙集理论在智能系统中的应用,主要包括基于粗糙集理论的属性约简作为数据预处理的手段,基于粗糙集理论的相关性分析和基于粗糙集理论的系统建模和控制.指出了粗糙集理论在应用中遇到的问题和可能的研究方向.

**关键词:**粗糙集;属性约简;智能系统;粗糙神经网络

**中图分类号:**TP18 **文献标识码:**A **文章编号:**1673-4785(2007)02-0040-08

## Rough set theory and its application in the intelligent systems

SAN Ye, YE Yu-ling

(Control and Simulation Center, Harbin Institute of Technology, Harbin 150080, China)

**Abstract:** The rough set theory is a new mathematical tool to study vague and uncertain information, and is widely used in intelligent systems. In this paper, the basic ideas of rough set theory are introduced, and the notion of up and low approximation sets, attribute reduction, core and some extensions of rough set theory are also presented. Then the application of rough set theory in intelligent systems is explored, mainly including data preprocessing method by using attribute reduction based on rough set theory, analysis of correlations between attributes and system, modeling and control. Finally, the difficulties and possible fields for the application of rough set theory are discussed.

**Key words:** rough set; attribute reduction; intelligent system; rough neural network

粗糙集理论是波兰学者 Pawlak 提出的用来刻画不完整性和不确定性数据的工具<sup>[1]</sup>. 它能有效地分析和处理不精确、不一致和不完整信息,从中发现隐含的知识,揭示潜在的规律<sup>[2]</sup>. 自提出以来,特别是近几年,粗糙集理论得到了长足的发展,不仅建立了严格的数学模型和完整的理论体系<sup>[3-4]</sup>,在数据挖掘、机器学习、决策分析、故障诊断、智能控制等领域也取得了众多成功的应用<sup>[5-7]</sup>. 现在,粗糙集理论已经成为人工智能领域中一个新的研究热点,受到越来越多学者的关注.

智能信息处理<sup>[8]</sup>是集系统内、外部各种信息形式,利用某种关于信息的经验和知识,以算法为核心,将启发式知识获取、处理和应用相结合的信息处

理技术. 其当前的主要理论方法有专家推理和决策系统、模糊逻辑、神经网络、粗糙集理论等,主要解决因信息量不完全而导致的系统病态问题,用数学模型难以精确描述的非线性和不确定性问题,以及计算复杂性和实时性问题. 智能系统以智能信息处理为核心,充分利用人工智能和计算机技术的最新成果,在设计形式和方法上取得了丰富的成果.

本文从粗糙集理论的基本思想、概念出发,综述了粗糙集理论在智能系统中的应用.

### 1 经典的粗糙集理论

在粗糙集理论中,用一个二维表表示信息系统,“行”表示对象,“列”表示描述对象的属性,这样就可以用对应属性的属性值集合来描述一个对象. 其中,单个的属性值即为数据;每个属性称为信息,而知识则表现为一种分类能力. 世界上没有两件完全相同

的事物,也就是说如果能用无限个属性来描述所研究的对象,没有两个对象的属性值集合完全相同.但是人类的观察能力是有限的,只能在有限维空间中对事物进行描述,这样就导致了在人类认识能力范围内有可能出现完全相同的两件事物,这就引出了经典粗糙集的核心概念:不可分辨关系.对于两个对象,如果对应属性的属性值都相等,则说这两个对象是不可分辨的.其实,这两个对象并不是绝对不可分辨的,而是因为人类的认识能力有限,无法在现有知识基础(属性集)上分辨而已.经典的粗糙集理论的不可分辨关系是等价关系,它把论域中的对象分成一个个的等价类,每个等价类是由对应属性值相等的对象构成的集合.这些等价类是信息系统所能分辨的最小单位,称为基本概念.用这些基本概念去描述任意集合时,就会出现不能精确描述的情况:同一个基本概念中的对象可能部分属于给定的集合,部分不属于该集合,这样就只能用这些基本概念去逼近这些集合.能用基本概念精确描述的集合称为可以精确定义的,反之就称为粗糙集.

1.1 不可分辨关系和上、下近似

粗糙集理论中定义信息系统为  $S = (U, A, f, V)$ , 用一个二维表表示, 其中  $U = \{x_1, x_2, \dots, x_n\}$  为对象集合, 称为论域;  $A = \{a_1, a_2, \dots, a_m\}$  为属性集合, 其中每个属性对应一个等价关系  $R$ . 设  $P \subseteq R$ , 且  $P \cap \emptyset$  ( $\emptyset$  为空集),  $P$  中所有的等价关系的交集称为  $P$  上的等价关系, 或称不可分辨关系 (indiscernibility relation), 记作  $IND(P)$ .  $V = \{v_1, v_2, \dots, v_m\}$  为对应属性的值域构成的集合;  $f: U \rightarrow V$  为映射函数. 设  $X$  为  $U$  的一个子集,  $a$  为  $U$  的一个对象,  $[a]_R$  表示所有与  $a$  不可分辨的对象所组成的集合, 即包含  $a$  的等价类. 当集合  $X$  能表示成基本等价类组成的并集时, 则称集合  $X$  是可以精确定义的; 否则, 集合  $X$  只能通过近似的方式来逼近, 在粗糙集理论中通过上近似集和下近似集来描述.

$$B(X) = \{a \in U \mid [a]_R \subseteq X\},$$

$$\overline{R}(X) = \{a \in U \mid [a]_R \cap X \neq \emptyset\}.$$

$B(X)$  实际上是由那些根据已有知识判断肯定属于  $X$  的对象所组成的最大集合, 也称为  $X$  的  $R$  正域, 记作  $POS_R(X)$ .  $\overline{R}(X)$  是由所有与  $X$  相交非空的等效类  $[a]_R$  的并集, 是那些可能属于  $X$  的对象组成的最小集合. 由根据已有知识判断肯定不属于  $X$  的对象组成的集合称为  $X$  的负域. 记作

$NEG_R(X)$ . 从而, 集合  $X$  的边界域定义为

$$BN(X) = \overline{R}(X) - B(X).$$

如图 1 所示, 每个单位方格表示 1 个基本概念, 任意给定 1 个集合 (图上曲线包围的部分). 从图中可以形象地看出该集合并不能由基本概念精确表示, 是 1 个粗糙集合. 用 2 个精确集合 (下、上近似集合) 来逼近该集合, 这 2 个精确集合可以用基本概念精确的确定, 它们的差即为边界域.

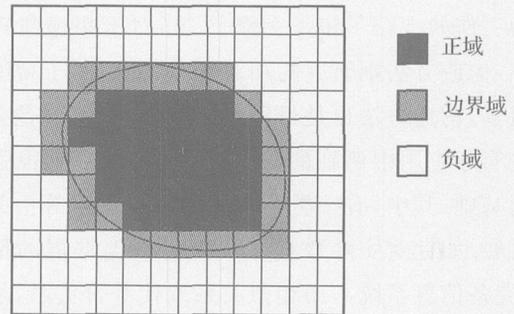


图 1 粗糙集合

Fig. 1 Rough set

定义近似精度为

$$r(X) = \frac{|B(X)|}{|\overline{R}(X)|},$$

近似精度可以用来作为含混性 (vagueness) 的一种度量. 粗糙集理论能区分含混性和不确定性: 含混性是针对集合而言, 而不确定性是针对元素与集合的关系而言.

1.2 约简与核

实际的近似空间中经常会有一些属性是冗余的, 或者说对于现有的分类问题是不必要的, 删除掉这些多余的属性就涉及到粗糙集理论中的一个重要概念: 约简.

直观地, 所谓知识的约简是指知识的本质部分, 它足以定义所考虑的知识中遇到的所有基本概念, 而“核”是其中最重要的部分.

设  $R$  是等价关系的一个族集, 且设  $R \in R$ . 若  $IND(R) = IND(R - R)$ , 则称关系  $R$  在族集  $R$  中是可省的 (dispensable), 否则就是不可省的. 若族集  $R$  中的每个关系  $R$  都是不可省的, 则称族集  $R$  是独立的 (independent), 否则就是依赖的或非独立的. 若  $Q \subseteq P$  是独立的, 并且  $IND(Q) = IND(P)$ , 则称  $Q$  是关系族  $P$  的一个约简 (reduct). 在族集  $P$  中所有不可省关系的集合称为  $P$  的核 (core), 以  $CORE(P)$  表示.

如果信息系统的属性分为条件属性和决策属性 2 部分,且二者没有交集,则称该信息系统为决策系统.对于决策系统,粗糙集理论相应地定义了相对约简和相对核的概念.

### 1.3 经典粗糙集理论的研究现状

自粗糙集理论提出以来,许多学者对其进行了大量的理论和应用研究,大致可以分为 2 类<sup>[3]</sup>:公理化方法和构造方法.其中,公理化方法是从集合论的角度,通过不同的集合公理建立各种粗糙集代数,典型的工作如 Yao<sup>[9]</sup>和张文修<sup>[10]</sup>等.对于粗糙集理论而言,构造方法则研究的相对较多,同时对于智能系统而言,构造方法也是最重要的.这些方法都是为了解决实际应用中遇到的问题而提出的,具有很强的应用基础.其中,有一类重要的拓展是针对等价关系的拓展.如针对实际数据库中含有数据缺损的情况(不完备信息系统),用相似关系取代等价关系,从而建立起基于相似关系的粗糙集模型<sup>[11]</sup>;针对实际数据库中同一属性的值之间存在“优劣”的情况,偏序关系被用来取代等价关系,从而建立起偏序关系基础上的粗糙集模型<sup>[12]</sup>;针对属性值域为连续的情况,模糊等价关系被用来取代等价关系,从而建立起模糊等价关系基础上的粗糙集理论<sup>[13]</sup>;基于更一般的“相似关系”建立的粗糙集模型<sup>[14]</sup>和基于邻域关系建立的邻域粗糙集模型<sup>[15]</sup>.另外,还有针对决策系统含有噪声和由此造成的不协调情况,通过改进上下近似集合的概念而建立的变精度粗糙集模型<sup>[5]</sup>.粗糙集理论和其他理论的结合也是粗糙集研究的热点,如结合模糊集合的概念建立的模糊粗糙集理论和粗糙模糊集理论<sup>[16]</sup>;结合概率论建立的概率粗糙集模型<sup>[10]</sup>等.

除了各种粗糙集模型外,有效的约简求解算法也是研究的热点<sup>[17]</sup>.经典的粗糙集约简算法最常用的是基于分明矩阵的方法.就约简的定义而言,就有所谓的代数观意义下的定义和信息观意义下的定义之分<sup>[18]</sup>.经典粗糙集模型中,在协调信息系统(决策系统)中两者的定义是等价的,但在不协调系统中两者是有差异的.由于大多数信息系统(决策系统)都有不只一个约简(相对约简),其中包含属性(条件属性)数量最少的约简(相对约简)称为最小约简(相对约简).理论上已经证明求最小约简(相对约简)和所有约简(相对约简)都是 NP-hard 问题.所以,更多的约简算法是根据启发信息求系统的次优约简.其

中定义属性的重要性,并以此为启发信息的贪心算法<sup>[13]</sup>应用最广.此外,进化算法(如遗传算法<sup>[19]</sup>、蚁群算法、克隆算法、粒子群优化算法等)也被用来求解系统的约简(相对约简).针对条件属性数量多的数据库,分层递阶的思想也被应用到约简中,从而产生了分层递阶的约简算法<sup>[20]</sup>.

## 2 粗糙集理论在智能系统中的应用

在目前关于人工智能、智能系统或智能信息系统研究中,专家系统(expert systems, ES),基于知识的决策支持系统(knowledge-based decision support systems, KBDSS),智能主体(intelligent agent, IA)等一类知识系统(knowledge systems)的核心就是知识,知识的数量和质量是决定一个系统的性能优劣的主要因素.一个典型的智能系统或知识系统的构架如图 2 所示.图中“用户”对各环节参与的越少,表明系统的智能化程度越高.

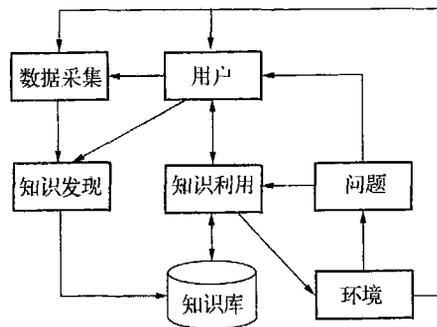


图 2 智能系统构架图

Fig. 2 Architecture of intelligent systems

在实际的系统中知识往往是不精确、不确定、不完整的,这使系统的许多问题变得非常复杂,软计算作为解决这些问题的重要手段,显示出了巨大的优势.软计算<sup>[21]</sup>方法是指利用所允许的不精确性、不确定性和部分真实性得到易于处理、鲁棒性强和成本较低的解决方案,它区别于用精确、固定和不变的算法表达和解决问题的硬计算.粗糙集理论与模糊逻辑、神经网络、进化算法等同为软计算方法,但是它们又有各自的特点.研究表明这些软计算方法之间不是竞争而是互补的关系<sup>[22]</sup>.粗糙集理论与其他软计算方法的融合,可以形成一系列新的、有效的应用于智能系统特定领域的理论和方法.

### 2.1 粗糙集理论作为智能系统数据预处理的手段

智能系统最基本的特征就是数据获取和处理的

自动化. 从各种途径获得的数据往往含有冗余, 需要经过预处理. 粗糙集理论在智能系统中最简单的应用是作为数据预处理的手段. 基于粗糙集的数据预处理归纳起来有 2 种形式: 利用粗糙集约简降低数据的维数和利用粗糙集上下近似集的概念减少数据中对于后续处理无关的数据.

2. 1. 1 用粗糙集理论属性约简来降低数据维数

对于一个复杂系统而言, 为了更好的理解其行为, 人们不断研制新的观察工具, 发展新的观察技术, 以提供观察对象更多方面的、更完整的信息, 建立更详尽的数据库. 这些数据库中描述事物的多变量数据是高维的, 它一方面提供了更加丰富细致的信息, 另一方面也给数据处理带了重大的挑战. 一般来说, 对于某一特定的问题, 这些数据库都含有大量冗余的和不相关的属性, 针对特定问题进行特征选择和特征提取就成了重要的数据预处理内容. 利用约简降低数据的维数是特征选择的重要方法, 也是粗糙集理论最直观的应用.

利用属性约简可以对其进行简化, 大大降低了后续处理(如分类、建模、决策等)的计算复杂度和可靠性, 有效地防止“维数灾”问题<sup>[23]</sup>. 同时, 属性约简还能保持原有属性的物理意义, 利于研究者把握事物的本质. 相对于专家分析的方法减少了人为主观因素的干扰, 有利于系统智能化.

神经网络是常用的智能分类、建模手段, 粗糙集理论常常与神经网络结合使用, 利用粗糙集理论属性约简作为神经网络的预处理<sup>[24]</sup>. 通过粗糙集理论约简来降低输入空间的维数, 从而减小神经网络的复杂性, 减少后续使用过程中信息作为网络输入时的特征值计算时间; 同时也能简化神经网络的训练时间.

一般来说, 信息系统(决策系统)的约简并不是唯一的, 还可以利用这一点从原始信息系统得到多个等价的信息系统用于“集成”, 如神经网络集成<sup>[25]</sup>. 图 3 是多神经网络集成的示意图, 图中的神经网络可以是同一种类型的神经网络, 也可以是不同类型的, 甚至可以是别的分类、建模技术.

另外, 得到的约简结果经过评价可以用于指导数据采集, 即只需要采集对象的约简后属性集中对应属性的数据, 如图 4 所示. 该模型完全从采集数据的本身出发, 分析属性之间的关系, 删除冗余的和不相关的属性. 这样可以使信息系统大大简化, 节省数

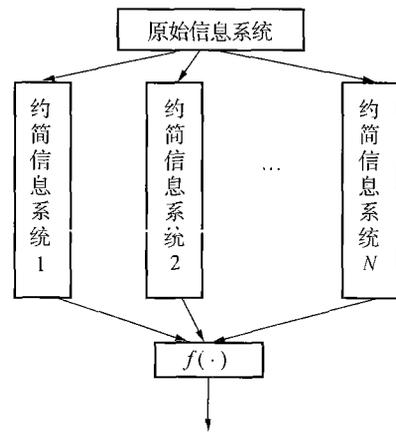


图 3 粗糙集理论在神经网络集成中的应用  
Fig. 3 Application of rough set theory in the neural network ensemble

据采集的开销, 更重要的是削弱了人为因素的干扰, 使系统更加智能化.

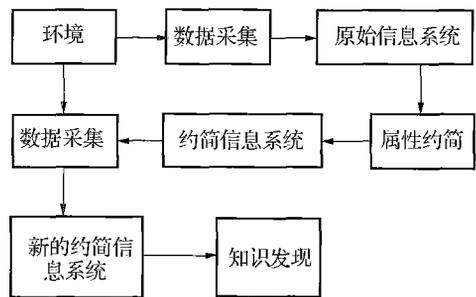


图 4 基于粗糙集属性约简的数据降维  
Fig. 4 Architecture of the data dimension decrease based on the reduct of rough set theory

2. 1. 2 用粗糙集上下近似集的概念降低数据量

粗糙集的基本思想是用一对上下近似集合来逼近和刻画一个边界不确定的集合. 这一思想也可以用来作为智能系统数据预处理的手段. 例如, 在分类问题中, 影响分类的对象往往位于不同类之间的边界部分. 每一类必然有一些肯定属于该类的对象, 这些对象对于分类边界而言并没有帮助, 相反许多分类算法有时还会影响分类的算法, 降低分类的精度. 这些对象实际就是每一类的下近似集, 影响分类边界的是边界集. 所以, 可以利用粗糙集理论先确定对分类边界有影响的边界集, 删除下近似集, 达到简化系统的目的. 研究表明, 用边界集数据进行分类的结果要优于直接用原始数据分类得到的结果.

### 2.2 粗糙集理论用于相关性分析

粗糙集理论还可以用来分析信息系统中各属性之间的关系和属性的重要性,从而成为相关性分析和决策分析<sup>[26]</sup>的重要方法.经典的粗糙集理论处理信息系统时,如果 2 个属性获得的系统等价类之间交集的元素构成的集合为整个对象集合,则表明这 2 个属性是相关的,否则是不相关的;如果属性  $a$  与属性集合  $B$  形成的等价类之间的交集的元素构成的集合为整个对象集合,则表明属性  $a$  与属性集合  $B$  相关,否则属性  $a$  与属性集合  $B$  不相关.如果是决策系统,则可以分析条件属性(集合)和决策属性(集合)之间是否相关.要定量分析相关性的大小,可以定义属性重要性的概念.在决策系统中属性重要性分 2 种,一种是  $a \approx B$ ,分析属性  $a$  相对于所在的属性集合  $B$ ,对于决策的重要性;另一种是  $a \notin B$ ,属性  $a$  相对于另外一个属性集合  $B$ ,对于决策的重要性.对于前者主要是看从属性集合  $B$  中删除属性  $a$  后,剩下的条件属性构成的等价类能够正确分类到决策属性构成的等价类的数量有没有变化,也就是看相对正域有没有变化.如果变化大,则说明属性  $a$  相对于所在属性集合  $B$  对于决策来说重要性高;如果没有变化,则说明属性  $a$  相对于所在集合  $B$ ,对决策来说是冗余的,重要性也为 0.对于后者可以类似的通过属性集合  $B$  加上属性  $a$  前后相对正域的变化来计算重要性.当然对于不同的粗糙集模型和不同的不确定性测度,属性重要性的定义形式和计算方法也不相同,文献<sup>[18]</sup>就是利用信息熵来定义属性重要性,并以此作为启发信息设计约简算法的.

### 2.3 粗糙集理论用于系统建模和控制

在系统控制和建模中,粗糙集理论定义条件属性和决策属性间的依赖关系,即输入空间与输出空间的映射关系是通过简单的决策表简化得到的<sup>[27]</sup>.即通过去除冗余属性,降低知识表达空间的维数.从决策表得到的规则在处理实际问题时对噪声较敏感,推广能力较弱.而神经网络完成输入空间与输出空间的映射关系是通过网络结构不断学习、调整,最后以网络的特定结构来表达的,具有良好的抗干扰能力、自组织性和推广能力.但是当输入空间维数过高则会导致网络结构过于复杂,训练速度过慢等问题.因此,神经网络与粗糙集算法可以结合使用,以利用二者各自的优点而克服各自的缺点.

粗糙集约简作为神经网络的预处理,只是一种

比较松散的结合方式,粗糙集理论与神经网络仍然是独立的 2 部分.另一种结合方式就是所谓的粗糙神经网络<sup>[28-29]</sup>,粗糙神经网络中的神经元称为粗糙神经元.粗糙集合的特点是用上、下边界集合来近似逼近,而粗糙神经元也是利用了这个思想,在神经元中储存了输入输出的上、下 2 个边界.一个粗糙神经元  $r$  可以看作是一对神经元,对应于上边界的神经元称为上神经元  $\bar{r}$ ,对应于下边界的神经元称为下神经元  $r$ .根据实际应用的不同,粗糙神经网络中的 2 个粗糙神经元  $r$  和  $s$  间的互联可以有 3 种方式,如图 5 所示.第 1 种方式是 4 连接,即  $r$  和  $\bar{r}$  都连接到  $s$  和  $\bar{s}$ ,也称为全连接,即图 5(a);第 2 种方式是  $r$  和  $s$  间只存在 2 个连接,如图(b)所示,称为兴奋连接;第 3 种方式称为抑制连接,如图 5(c)所示.粗糙神经元可以与经典的神经元直接连接,也可以与粗糙神经元互联.含有粗糙神经元的粗糙神经网络也可以用经典神经网络的训练和学习算法,研究表明粗糙神经网络建模可以获得更高的精度.

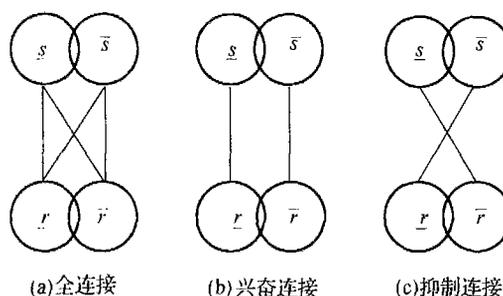


图 5 粗糙神经元

Fig 5 Rough neuron

粗糙集理论和神经网络另一种结合方式是文献<sup>[30]</sup>提出的,称为变精度粗糙神经网络(rough neural network of variable precision).它的基本思想是利用变精度粗糙集理论对原始决策表进行约简,得出对应的决策规则,再根据决策规则建立神经网络的结构.如图 6 所示, $L_1$  为输入层, $L_2$  表示输入单元的等价类, $L_3$  为隐层, $L_4$  为输出层.图中可以看出, $L_2$  与  $L_3$  或者  $L_3$  与  $L_4$  之间并不是全连接的,而是根据决策规则确定 2 层的神经元之间是否有连接.这里神经网络的神经元实际是逻辑单元(logic cell),如图 7 所示,和 通过取不同的值,可以得到不同的逻辑单元.这种方法可以大大简化神经网络的结构和连接数目,从而使神经网络的学习和训练更加有效.

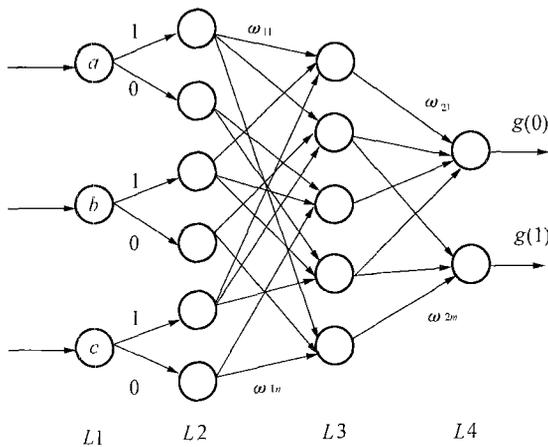


图 6 粗糙神经网络

Fig 6 Configuration of rough neural network

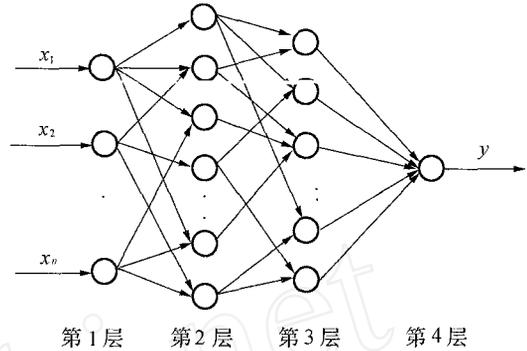


图 8 粗糙模糊神经网络

Fig 8 Configuration of rough fuzzy neural network

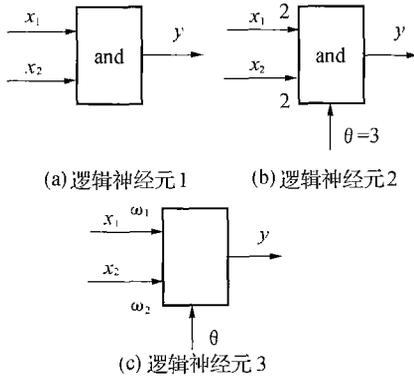


图 7 逻辑神经元

Fig. 7 Logical neuron

粗糙集还可以和模糊理论、神经网络结合,形成粗糙模糊神经网络<sup>[31]</sup>如图 8 所示.网络分 4 层,第 1 层输入层,它的值为实际的精确值,表示输入向量  $x = (x_1, x_2, \dots, x_n)^T$ ;第 2 层隶属度函数层,分别将  $n$  个输入分量  $(x_1, x_2, \dots, x_n)$  依照某种不可分辨关系进行划分,确定其与每个相应分类之间的联系.将每个输入分量离散化为  $r_i$  个不同的值,与第  $i$  个输入节点相连接的一组神经元的作用是对输入向量的第  $i$  个分量进行解释,定义该层神经元的作用函数为粗糙隶属函数;第 3 层为推理层,每个节点代表一条推理规则,这些规则是通过粗糙集理论得到的;第 4 层为清晰化层,节点代表输出变量.其建模过程为先利用粗糙集理论对原始信息系统进行属性约简得到约简条件属性集及规则集,计算各条规则的粗隶属度.然后进行粗糙神经网络的构造和训练.约简后的数据按结论分为  $m$  类,各类的对象集合构成了  $m$

个子网的输入输出.利用这些数据进行子网训练,分析并输出结果.这种粗糙神经网络可以充分利用粗糙集理论在知识获取方面的能力和神经网络在函数逼近方面的优势.

基于粗糙集理论的约简只是对数据进行了预处理,对于像模糊控制、专家系统等基于规则的智能系统,得到的约简后的系统可以用来进行规则提取,从而得到系统的规则库<sup>[32]</sup>.由于对原始信息系统进行了约简,从很大程度上降低了规则的冗余程度和规则数量,从而从某种程度上缓解了专家系统中的“规则爆炸”问题,也从某种程度上解决了智能系统中知识获取这一“瓶颈”<sup>[33]</sup>.如果是从系统人工控制过程的数据表出发,通过粗糙集约简提取规则,则得到的规则集构成了控制规律,用于实际系统的控制.这就是粗糙控制的基本思想,与模糊控制不同的是,粗糙控制直接从实际数据表出发,发现控制规律,而模糊控制需要人工从专家归纳和提取控制规律.提取的规则根据实际系统还有其他的用途,如系统的故障诊断<sup>[34]</sup>等.

粗糙集理论还可以用来处理数据意义的评估、数据中相似性或差异性的发现、数据中范式的发现以及因果关系的发现等.总之,粗糙集理论已经成为智能系统中的重要技术,在医学、药学、银行、商业、金融、市场研究、工程设计、气象学、振动分析、开关函数、冲突分析、图像处理、声音识别、并发系统分析、决策分析、字符识别等领域都有重要应用.

### 3 结束语

粗糙集理论虽然只有 20 多年的历史,但是以其独特的优势赢得了众多研究者的关注,在智能系统领域取得了广泛的应用.但是,这仍然是一个年轻

并高速发展的学科,还有许多问题等待去研究,许多研究方向是相互关联的.一个方向是大数据集问题.面对越来越大的数据库,粗糙集也遇到了前所未有的挑战.大数据集环境下高效的约简、稳定约简和增量式约简就是其中最尖锐的问题.另一个方向是多种智能方法融合的问题.粗糙集理论、神经网络、模糊逻辑等智能方法在处理实际的复杂系统问题时,都有各自的优点,现有的融合方法已经显示出巨大的优势.可以预见,各种智能方法更深层次的融合将会是未来智能系统发展的重要方向.

### 参考文献:

- [1] PAL S, SKOWRON A. Rough fuzzy hybridization: a new trend in decision making[M]. Singapore: Springer-Verlag, 1999.
- [2] WALCZA K B, MASSART D L. Rough sets theory[J]. Chemometrics and Intelligent Laboratory Systems, 1999, 47:1 - 16.
- [3] YAO Y Y. Constructive and algebraic methods of the theory of rough sets[J]. Journal of Information Sciences, 1998, 109:21 - 47.
- [4] WU Weizhi, ZHANG Wenxiu. Constructive and axiomatic approaches of fuzzy approximation operators[J]. Information Sciences, 2004, 159: 233 - 254.
- [5] PAWLAK Z. Rough set approach to knowledge - based decision support[J]. European Journal of Operational Research, 1997, 99: 48 - 57.
- [6] PAWLAK Z. Rough sets and intelligent data analysis[J]. Information Sciences, 2002, 147:1 - 12.
- [7] LINGRAS P J, YAO Y Y. Data mining using extensions of the rough set model[J]. Journal of the American Society for Information Science, 1998, 49(5):415 - 422.
- [8] 黄金杰.粗糙集理论在智能信息处理机控制中的应用研究[D].哈尔滨:哈尔滨工业大学, 2004.  
HUANG Jinjie. Research and application of rough sets theory in the control of intelligent information processor [D]. Harbin: Harbin Institute of Technology, 2004.
- [9] YAO Y Y. Two views of the theory of rough sets in finite universes[J]. International Journal of Approximate Reasoning, 1996, 15(2): 291 - 317.
- [10] 张文修, 仇国芳.基于粗糙集的不确定决策[M].北京:清华大学出版社, 2005.
- [11] KRYSZKIEWICZ M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112(1): 39 - 49.
- [12] GRECO S, MATARAZZO B, SLOWINSKI R. Rough sets theory for multicriteria decision analysis[J]. European Journal of Operational Research, 2001, 129(1): 1 - 47.
- [13] HU Qinghua, YU Daren, XIE Zongxia. Information-preserving hybrid data reduction based on fuzzy - rough techniques[J]. Pattern Recognition Letters, 2006, 27(5): 414 - 423.
- [14] YAO Yiyu, CHEN Yaohua. Subsystem based generalizations of rough set approximations[A]. Proceedings of 15th International Symposium on Methodologies for Intelligent Systems[C]. Berlin, 2005.
- [15] YAO Y Y. Relational interpretations of neighborhood operators and rough set approximation operators[J]. Information Sciences, 1998, 111(1 - 4): 239 - 259.
- [16] SLOWINISKI R. Intelligent Decision Support [M]. Dordrecht: Kluwer Academic, 1992.
- [17] GUAN J W, BELL D A. Rough computational methods for information systems artificial intelligence[J]. 1998, 105(1): 77 - 103.
- [18] 王国胤, 于洪, 杨大春.基于条件信息熵的决策表约简[J]. 计算机学报, 2002, 25(7):759 - 766.  
WANG Guoyin, YU Hong, YANG Dachun. Decision table reduction based on conditional information entropy [J]. Chinese J Computers, 2002, 25(7): 759 - 766.
- [19] JAKUB W. Finding minimal reducts using genetic algorithm[R]. ICS Research Report 16/95, Warsaw university of Technology, 1995.
- [20] 乔斌, 李玉榕, 蒋静坪.粗糙集理论的分层递阶约简算法及其信息论基础[J]. 控制理论与应用, 2004, 21(2): 195 - 199.  
QIAO Bin, LI Yurong, JIANG Jingping. Hierarchical reduction approach of rough sets theory and its basis on the information theory[J]. Control Theory & Application, 2004, 21(2): 195 - 199.
- [21] MITRA S, SANKAR K P, MITRA P. Data mining in soft computing framework: a survey[J]. IEEE Transactions on Neural Networks, 2002, 13(1): 3 - 14.
- [22] BANERJEE M, MITRA S, SANKAR K P. Rough fuzzy MLP: knowledge encoding and classification[J]. IEEE Transactions on Neural Networks, 1998, 9(6): 1203 - 1216.
- [23] GARDNER M W, DORLING S R. Artificial neural networks (the multilayer perceptron)-a review of applications in the atmospheric sciences [J]. Atmospheric Environment, 1998, 32(14 - 15): 2627 - 2636.
- [24] AHN B S, CHO S S, KIM C Y. The integrated methodology of rough set theory and artificial neural network for business failure prediction[J]. Expert Systems with Application, 2000, 18(4): 65 - 74.
- [25] HU Qinghua, YU Daren, XIE Zongxia. Improvement on classification performance based on multiple reduct ensemble[A]. Proceedings of the 2004 IEEE Confer-

ence on Cybernetics and Intelligent Systems[C]. Singapore, 2004.

[26]DIMITRAS A I, SLOWINSKI R, SUSMAGA R, et al. Business failure prediction using rough sets[J]. European Journal of Operational Research, 1999, 114(2): 263 - 280.

[27]GAO Yun, HOU Yuanbin. Application of rough set theory on system modeling[A]. Proceedings of the 5th World Congress on Intelligent Control and Automation [C]. Hangzhou,2004.

[28]LINGRAS P J. Rough neural networks[A]. Proc of sixth international conference on information processing and management of uncertainty in knowledge-based systems[C]. Granada, Spain, 1996.

[29]LINGRAS P. Comparison of neofuzzy and rough neural networks[J]. Information Sciences, 1998,110(3): 207 - 215.

[30]LIU Hongjian, TUO Hongya, LIU Yuncai. Rough neural network of variable precision[J]. Neural Processing Letters, 2004, 19(1): 73 - 87,.

[31]何 明,李 博,马兆丰,等.粗糙集理论框架下的神经网络建模研究及应用[J].控制与决策,2005, 20(7): 782 - 785.

HE Ming, LI Bo, MA Zhaofeng, et al. On the neural network modelling with support rough set theory[J]. Control and Decision, 2005, 20(7): 782 - 785.

[32]AN A, SHAN Ning, CHAN C, et al. Discovering rules for water demand prediction: an enhanced rough set approach[J]. Engng Applic Artif Intell, 1996, 9(6): 645 - 653.

[33]TSUMOTO S. Automated extraction of medical expert system rules from clinical databases based on rough set theory[J]. Information Sciences, 1998, 112(1): 67 - 84.

[34]SHEN Lixiang, FRANCIS E H T, QU Liangsheng, et al. Fault diagnosis using rough sets theory[J]. Computers in Industry, 2000, 43(1): 61 - 72.

作者简介:



伞 冶,男,1951 年生,教授,博士生导师,中国系统仿真学会理事,黑龙江省系统仿真学会常务理事.主要研究方向为复杂大系统控制与仿真.

Email: sanye @hit.edu.cn.



叶玉玲,男,1979 年生,博士研究生,主要研究方向为非线性复杂动态系统建模与预测.

Email: yeyuling @hit.edu.cn.

# 2007 年中国数据挖掘与知识发现学术会议

## 2007 Chinese Conference on Data Mining and Knowledge Discovery

2007 年中国数据挖掘与知识发现学术会议(DMKD2007)将于 2007 年 11 月 26 日~28 日在风景秀丽的浙江省杭州市举行.会议由中国智能计算学会数据挖掘与知识发现专业委员会(Chinese Association for Intelligent Computing Data Mining and Knowledge Discovery Specialized Committee)主办,北京华微明天科技有限公司(Beijing Huawei Technology Co.,Ltd)协办.大会采用会前讲座、大会报告、分组报告与张贴论文等形式进行学术交流.DMKD 学术会议已引起著名国际索引机构关注,历届 DMKD 学术会议论文集被 SCIE 全文收录.我们热忱欢迎海内外广大专家、学者、研究生及工程设计人员踊跃投稿参加本届大会,共同交流学术成果.

征文范围(但不限于如下领域):时序数据的趋势分析;并行计算;神经网络算法;数据可视化;基因算法;范例推理;基本统计分析方法;隐马尔科夫模型;贝叶斯网络方法;遗传算法;支持向量机;多媒体数据挖掘;聚类分析;决策树算法;决策树算法;电子政务;数据仓库;市场营销;模糊处理技术;金融领域;粗糙集技术;移动通信领域;目标优化技术;生物医学;数据库技术;气象预报;人工智能;Web 数据挖掘;机器学习;关联规则挖掘;模式识别;数据预处理;知识库系统;Web 信息抽取;知识获取;数据存储;信息提取;高性能计算.

征文要求:1)请投稿者通过电子邮件(DMKD2007 @163.com)提交用英文书写的论文全文(Word 格式).论文格式请参照模板,并在首页中注明论文题目、作者姓名、单位、联系人的详细通讯地址(包括电话和 E-mail)、摘要、3~8 个关键词、文章所属的研究方向.2)论文提交截止时间:2007 年 6 月 15 日;论文录用通知时间:2007 年 7 月 27 日;最终论文提交时间:2007 年 8 月 10 日.