

文章编号: 1001-0920(2007)04-0462-03

基于粗糙集理论的神经网络研究及应用

张 赢¹, 李 琛²

(1. 纽约州立大学布法罗分校 工业与系统工程系, 纽约 14260; 2. 华中科技大学 系统工程研究所, 武汉 430074)

摘 要: 为了补偿神经网络的黑箱特性并提高其工作性能, 将粗糙集理论同神经网络结合起来, 提出一种基于粗糙集的神经网络体系结构. 首先, 利用粗糙集理论对神经网络初始化参数的选择和确定进行指导, 赋予各参数相关的物理意义; 然后, 以系统输出误差最小化为目标对粗糙神经网络进行训练, 使其满足性能要求. 实验结果表明, 粗糙神经网络能较好地完成数据挖掘任务, 并能获得较高的分类精度.

关键词: 粗糙集; 神经网络; 数据挖掘; BP 算法

中图分类号: TP18 **文献标识码:** A

Research on neural networks based on rough set theory

ZHANG Ying¹, LI Chen²

(1. Department of Industrial and Systems Engineering, The State University of New York, New York 14260, USA;

2. Institute of Systems Engineering, Huazhong University of Science and Technology, Wuhan 430074, China.

Correspondent: LI Chen, E-mail: liviachen@gmail.com)

Abstract: A rough neural network architecture is developed based on rough set theory and multilayer neural networks to compensate the black-box nature of neural networks and improve the work performances of ANNs. The initialized parameters of neural networks are selected, which is directed by rough set theory, and the networks are trained by using a probability-based BP algorithm, which minimizes system errors. Experimental results show that rough neural networks have good performances in data mining, and higher preciseness is obtained.

Key words: Rough sets; Neural networks; Data mining; BP algorithm

1 引 言

神经网络是一种广泛应用于数据挖掘和分类的有效方法, 但其应用在很大程度上受到了“黑箱特性”的影响, 因此如何减少神经网络的黑箱特性并从中提取语义规则便成为人们关心的问题. 在研究该问题的过程中, 一些学者将粗糙集理论^[1]同神经网络结合起来, 并取得了一定成果^[2,3]. 例如, 何明等^[2]通过利用粗糙集理论对神经网络的输入进行要素提取来简化神经网络的拓扑结构, 并在一定程度上补偿了黑箱特性. 但由于其主要将粗糙集理论用于数据预处理, 所以从相关网络中提取语义规则仍较为困难.

本文将粗糙集和神经网络结合起来, 提出一种用于数据挖掘的粗糙集神经网络结构. 通过利用粗糙集理论来对神经网络的构建及参数选择进行指导, 提高了神经网络的工作性能, 并为今后从神经网络中

提取语义规则打下了基础.

2 基于粗糙集的多层神经网络 RMLN

2.1 粗糙集的基本概念

粗糙集理论中的基本概念如下:

定义 1 $S = (U, A, V, f)$ 为一信息系统, 其中: 论域 $U = \{x_1, x_2, \dots, x_n\}$ 为非空有限对象集; A 为属性集; V 为属性 $a \in A$ 的值域; $f: U \times A \rightarrow V$ 为信息函数.

定义 2 设 $P \subseteq R$, 且 $P \neq \emptyset$, P 中所有等价关系的交集称为 P 上的一种不分明关系 (或称不可分辨关系), 记作 $[x]_{\text{IND}(P)} = \bigcap_{R \in P} [x]_R$.

定义 3 设 $S = (U, A, V, f)$ 为一信息系统, 且有 $B \subseteq A, X \subseteq U$, 则 X 关于 B 的上、下近似集分别为 $BX = \{x \mid [x]_B \subseteq X\}$, $\bar{B}X = \{x \mid [x]_B \cap X \neq \emptyset\}$.

定义 4 在一个信息系统中, 若给定 BX 和

收稿日期: 2006-01-06; 修回日期: 2006-03-21.

基金项目: 国家自然科学基金项目 (60574025).

作者简介: 张赢 (1981—), 男, 武汉人, 博士生, 从事系统优化、运筹管理等研究; 李琛 (1979—), 女, 武汉人, 博士生, 从事系统优化及供应链动态智能决策与分析等研究.

$\bar{B}X$, 则正区域为 $POS_B(X) = BX$, 反区域为 $NEG_B(X) = U - \bar{B}X$, 边界区域为 $BND_B(X) = \bar{B}X - BX$.

定义 5 设 $X \subseteq U$ 且 $x \in U$, 集合 X 的粗糙隶属函数定义为 $\mu_X^R(x) = \frac{|LX \cap [x]_R|}{|[x]_R|}$, 其中 R 是不分明关系, $| \cdot |$ 表示集合中元素的个数.

2.2 基于粗糙集的神经网络计算

为了减少神经网络的黑箱特性, 近年来学者们在“神经 - 模糊”计算领域开展了大量研究. 与传统的模糊集相比, 粗糙集具有客观准确的特点, 因此许多学者将模糊集引入神经网络的研究中.

粗糙集理论给出了论域中某子集的上下近似集, 并通过粗糙隶属函数给出了某个体属于各个子集的可能性. 然而, 粗糙隶属函数的值会因不分明关系的不同而不同, 所以并未给出样本点具体属于哪个区间的精确概率描述. 为解决这一问题, 一种可行的方法是找出所有不分明关系, 通过综合各相关信息来对粗糙隶属函数的值进行修正. 神经网络的工作原理是通过不断修正网络参数来对真值进行逼近, 所以可利用粗糙集的相关概念来指导神经网络的构造, 并通过神经网络来对粗糙隶属函数值进行修正.

3 粗糙神经网络结构及工作原理

3.1 粗糙神经网络结构及系统参数选取

本文提出的粗糙神经网络分为 4 层, 其结构如图 1 所示.

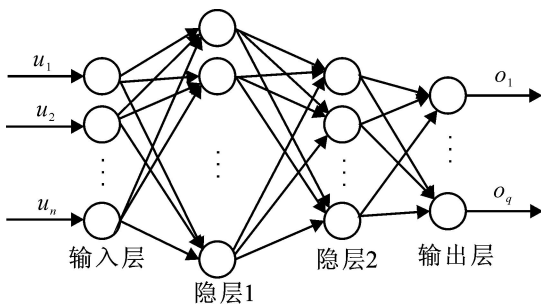


图 1 粗糙神经网络的体系结构

该网络各层功能及节点数目选取规则如下:

第 1 层(输入层): 该层用于系统的数据输入, 这里利用文献 [4] 的方法对其进行构建. 设系统输入为多维数据 $U = (u_1, u_2, \dots, u_n)^T$, 且第 i 维数据可划分为 k_i 个区间, 则该层节点数目为 $\sum_{i=1}^n k_i$. 对于每一维输入数据, 按其所在区间对其进行二进制编码, 并将该编码作为该维节点的输入.

第 2 层(隐层 1): 该层节点对应于近似空间中的属性集 A . 在粗糙集理论中, 所有数据都是基于属性集合来进行分类的, 因此有必要在网络中将输入

数据与用来划分类别的属性集合联系起来. 输入层中节点 i 与隐层 1 中节点 j 之间的连接权为 w_{ij} , 它反映了使用节点 j 所代表的属性集合对论域进行划分时节点 i 中数据所占的权重. 该层节点的数目 p 为一较大数.

第 3 层(隐层 2): 该层节点对应于论域中分类子集 X_k 的正区域和边界区域. 根据粗糙集理论, 当选定某一属性集合时, 对于论域 U 的任一子集 X , 其正区域 $POS_B(X)$ 中的个体必属于该子集, 而其边界区域 $BN_B(X)$ 中的个体有可能属于该子集, 因此正区域与边界区域的并集 $POS_B(X) \cup BN_B(X)$ 包含了所有可能属于该类的个体. 该层节点数目为 $2q$, 其中 q 为论域 U 的子集个数. 隐层 1 中节点 j 与隐层 2 中节点 k 之间的连接权值 x_{jk} 反映了属性集合 j 与论域子集 k 的联系程度. 隐层 2 与输出层之间的连接权值 y_{kl} 反映了隐层 2 中节点 k 所代表的区域对分类结果的影响, 即若 y_{kl} 为一大值, 则说明该节点包含正区域. 因此, 它对决定某样本是否属于某子集起决定作用.

第 4 层(输出层): 该层由 q 个节点构成, 其中 q 为论域 U 的子集个数. 输出层中各节点的输出值为某样本属于相应类别的概率, 通过比较它们的大小来对该样本进行分类.

3.2 基于概率的 BP 学习算法

在粗糙神经网络投入使用之前应对其进行样本训练, 其训练算法如下:

设 q 为论域 U 的子集个数; 输入层中节点数为 n ; 隐层 1 中节点数为 p ; 隐层 2 中节点数为 $2q$; 输出层中节点数为 q ; $w_{ij}, x_{jk}, y_{kl} (1 \leq i \leq n, 1 \leq j \leq p, 1 \leq k \leq 2q, 1 \leq l \leq q)$ 为网络各层之间的连接权值.

Step1 网络初始化.

随机设置 $w_{ij} = 0$ 或 1, 根据 w_{ij} 的取值决定隐层 1 中节点 j 所对应的属性集 B_j , 设隐层 2 中节点 k 对应某一论域子集 X_k (正区域或边界区域), 取 x_{jk}

$= \frac{|[x]_{B_j} \cap X_k|}{|[x]_{B_j} \cap X_k|}$; 取 $y_{kl} = \mu_{L^{B_j}}^{X_k}(x)$, 其中 L 是论域中第 l 类个体的集合, B_j 构成了一个与隐层 2 中节点 k 相关的不分明关系; 取 $SSE_{new} = 0, SSE_{old}$ 为一较大值, tolerance 为一较小值.

Step2 网络正向计算.

根据下面公式计算网络中各层节点的数值:

隐层 1

$$v_j = \sum_{i=1}^n w_{ij} u_i, \quad u_j = \frac{1}{1 + e^{-v_j}}, \quad j = 1, 2, \dots, p; \quad (1)$$

隐层 2

$$v_k = \sum_{j=1}^p x_{jk} u_j, u_k = \frac{1}{1 + e^{-v_k}},$$

$$k = 1, 2, \dots, 2q; \quad (2)$$

输出层

$$v_l = \sum_{k=1}^{2q} y_{kl} u_k, o_l = \frac{e^{v_l}}{\sum_{j=1}^q e^{v_j}},$$

$$l = 1, 2, \dots, q. \quad (3)$$

其中 o_l 的表达式是一种基于概率的转换函数, 即将节点 l 的值转化为个体属于某种类型的概率. 对于所有 o_l , 有 $\sum_{l=1}^q o_l = 1$ 且 o_l 与 v_l 正相关. 经转化后, 网络输出为 $o = [o_1, o_2, \dots, o_q]^T$.

Step3 反向修正连接权值.

设 $t = [0, 0, \dots, 1, \dots, 0]^T$ 为某样本所属类别的真值, 向量中“1”的位置表示该样本的所属类别. 因此当前系统的输出误差为

$$SSE_{\text{new}} = \sum_{i=1}^q (o_i - t_i)^2. \quad (4)$$

取 $E = \frac{1}{2} SSE_{\text{new}}$. 为使 SSE_{new} 最小化, 对 E 求 y_{kl} 的导数, 得

$$\frac{\partial E}{\partial y_{kl}} = (o_l - t_l) \frac{\partial o_l}{\partial y_{kl}} + \sum_{i=1}^q (o_i - t_i) \frac{\partial o_i}{\partial y_{kl}} =$$

$$(o_l - t_l) \frac{\partial}{\partial y_{kl}} \left(\frac{e^{v_l}}{\sum_{j=1}^q e^{v_j}} \right) + \sum_{i=1}^q (o_i - t_i) \frac{\partial}{\partial y_{kl}} \left(\frac{e^{v_i}}{\sum_{j=1}^q e^{v_j}} \right).$$

$$(5)$$

将 E 分别对 x_{jk} 和 w_{ij} 求导, 得

$$\frac{\partial E}{\partial x_{jk}} = \sum_{i=1}^q (o_i - t_i) \cdot \left(\frac{\partial o_i}{\partial u_k} \cdot \frac{\partial u_k}{\partial x_{jk}} \right), \quad (6)$$

$$\frac{\partial E}{\partial w_{ij}} = \sum_{i=1}^q (o_i - t_i) \cdot \left(\frac{\partial o_i}{\partial u_k} \cdot \frac{\partial u_k}{\partial u_j} \cdot \frac{\partial u_j}{\partial w_{ij}} \right), \quad (7)$$

并按下式对各层节点之间的连接权值进行修改:

$$z(t+1) = z(t) - \mu \frac{\partial E}{\partial z}. \quad (8)$$

其中 μ 是学习速率, z 为 w_{ij}, x_{jk}, y_{kl} .

Step4 计算 $|SSE_{\text{new}} - SSE_{\text{old}}|$. 若 $|SSE_{\text{new}} - SSE_{\text{old}}| < \text{tolerance}$, 学习过程结束; 否则令 $SSE_{\text{old}} = SSE_{\text{new}}$ 并返回 Step2.

4 算例分析

为验证本文提出的神经网络在数据挖掘方面的有效性, 本文采用 glass 和 wine^[5] 这两个标准数据集对其进行测试, 如表 1 所示. 在分析性能时, 采用以下 3 种指标, 并将计算结果同 C5.0^[6] 进行了对比, 如表 2 所示.

1) 准确度: $A_i = n_{ic} / n_i$. 其中 n_i 是第 i 类中的样本数, n_{ic} 是被正确划分到该类的样本数.

表 1 数据集基本信息

数据集	类别数目	属性数目	样本数目
glass	6	9	214
wine	3	13	178

表 2 C5.0 与粗糙神经网络的性能指标比较

方法	数据集	平均精确度 / %	平均用户精确度 / %	Kappa / %
C5.0	glass	95.4	96.1	93.1
	wine	98.9	99.0	98.3
RMLN	glass	96.5	97.1	93.9
	wine	98.2	98.4	97.5

2) 用户准确度^[7]: $U_i = n_{ic} / n_i \times 100\%$. 其中 n_i 是被划分到第 i 类的样本数, n_{ic} 是被正确划分到该类的样本数.

3) Kappa^[7]: $K = \sum_i a_i / \sum_i b_i \times 100\%$. 其中 $a_i = n_{ic} n_i - n_i^2$, $b_i = n n_i - n_i n_i$.

5 结语

本文提出了一种基于粗糙集神经网络结构, 并研究其在数据挖掘中的应用. 算例分析表明, 该网络具有良好的工作性能. 但如何在此基础上进行语义规则提取仍是一个值得探索的问题, 需要对其进行更深入的研究.

参考文献(References)

- [1] Pawlak Z. Rough sets and intelligent data analysis[J]. Information Sciences, 2002, 147(1): 1-12.
- [2] 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 M, Li B, Ma Z F, et al. On the neural network modeling with support rough set theory[J]. Control and Decision, 2005, 20(7): 782-785.)
- [3] Sankar K Pal, Sushmita Mitra, Pabitra Mitra. Rough-fuzzy MLP: Modular evolution, rule generation, and evaluation[J]. IEEE Trans on Knowledge and Data Engineering, 2003, 15(1): 14-25.
- [4] Lu H J, Rudy Setiono, Liu H. Effective data mining using neural networks[J]. IEEE Trans on Knowledge and Data Engineering, 1996, 8(6): 957-961.
- [5] Blake C, Keogh E, Merz J. UCI repository of machine learning data tables [EB/OL]. (2005-12-28). <http://www.ics.uci.edu/~mllearn>.
- [6] Quinlan J R. Rule quest research: Data mining tools see5 and c5.0 [EB/OL]. (2005-12-28). <http://www.rulequest.com/see5-info.html>.
- [7] Rosenfeld G H, Fitzpatrick-Lins K. Coefficient of agreement as a measure of thematic classification accuracy [J]. Photogrammetric Engineering and Remote Sensing Introduction, 1986, 52(2): 223-227.