

基于粗集理论的神经网络

王 玮, 蔡莲红

(清华大学计算机科学技术系多媒体所, 北京 100084)

摘要: 提出了一种基于粗集理论的神经网络, 它由传统神经元和粗糙神经元构成。粗糙神经元包含一对传统的神经元, 即将数据中的上边界和下边界的作为网络的输入或输出值。当网络的输入和输出不是单值数据而是一个数据集合时, 经典的神经网络建立的预测模型的输出就会产生较大的误差, 而基于粗集理论的神经网络则可以很好地解决这个问题, 最后对基于粗集理论的网络进行性能评估。

关键词: 神经网络; 粗集理论, 粗糙神经网络

Rough Set in Neural Network

WANG Wei, CAI Lianhong

(Department of Computer Science, Tsinghua University, Beijing 100084)

【Abstract】 This paper describes rough neural networks which consists of a combination of rough neurons and conventional neurons. Rough neurons use pairs of upper and lower bounds as values for input and output. In some practical situations, it is preferable to develop prediction models that use ranges as values for input variables. Inability to record precise values of the variables is another situation where ranges of values are associated with a single value of the output variable. The predictions obtained using rough neural networks are significantly better than the conventional neural network model.

【Key words】 Neural network; Rough set; Rough neural network

数据的上界和下界概念在人工智能领域中使用得非常普遍, 在粗集理论中尤为明显, 随着粗集理论不断发展, 上下边界的概念得到更大范围内的应用。本文将经典的神经网络和粗集理论有机地结合, 提出了一种基于粗集理论的神经网络, 并应用神经网络的粗糙模式建立预测模型。在粗糙模式下每个神经网络的输入值不是一个单值而是一对值, 即上下边界数据, 经典的神经网络在预测模型中采用的是单值数据作为输入值, 但是在一些应用中会产生问题, 如医院要对病人进行病情的跟踪观察, 并希望对其未来的情况进行预测, 这时经典的神经网络就难以适用了, 对于一个病人来说, 心跳次数, 脉搏次数, 血压值, 体温等项指标在一天当中需要进行几次测试, 问题在于对于同一项指标每次测量值也是不同的, 因此得到的是一组数据而非单个数据, 由于经典的神经网络对于外界的信息的传导需要的是单值输入, 究竟应该取测量值中的哪个值作为输入就难以确定, 通常的方法是将测量数据进行数学平均, 以均值作为网络的输入, 但是这可能导致具有重要性质数据的泛化, 而粗集理论则可以很好地解决这个问题, 粗糙集数据的上下边界可以将病人一天的各项指标测量值的上界和下界数据作为粗糙神经元的输入。

1 粗集理论

粗集理论是基于此种假设: 即每个所研究的对象都附带有一定的信息, 如果一些对象由同样的信息来表征, 则它们就是相似的, 或者说是不可辨别的, 而这种不可辨别性正是粗集理论的数学基础。

相似或称为不可识别的对象的集合被称为基本集合, 它们构成了信息的基本颗粒。一个集合对象如果是由这些基本集合的并集所形成, 就称其是精确的; 反之就称其是非精确的, 即粗糙的。

每个粗糙集都有一些边界元素, 这些边界元素都不能被确定地归类于该集合, 也就是说, 根据所提供的知识无法对这些元素进行明确归类。下近似和上近似是粗集理论中最重要的两个概念。下近似由所有确定的属于某个集合的对象组成, 上近似由所有可能属于某个集合的对象组成。近似是粗糙集方法中的两个基本运算。下面用数学形式加以说明。

设序偶 $A = (U, R)$ 是一个近似空间, 其中 U 是一个有限集, 称为论域的子集, R 是 U 上的一个等价类的集合, 称 R 中的一个成员为一个原子集, 在 R 中有限次的运用集合的并操作可以构造一个 A 中的确定集合。设 R^* 是一簇 R 的子集, 则 R^* 可以生成一个拓扑逻辑空间 $T_A = (U, R^*)$ 。称 U 的每个成员为一个对象, X 是 U 的一个子集。称 A 中含有 X 的最小子集为 X 的闭集, 记为 $C1_A(X)$ (上集), 称 A 中最大的被包含在 X 中的集合为 X 的内集, 记为 $Int_A(X)$ (下集)。

如果对某个 $Y \subseteq R^*$, X 是 Y 中所有集合的并集, 称 X 在 A 中是确定的, 否则称 X 是非确定的或是粗糙的。对某个 $x \in U$, 我们不妨限定它与 X 的关系为下面3种之一, 即: (1) x 在 X 中, (2) x 不在 X 中, (3) 不知道。以 $POS_A(X)$ 代表一个对象的集合中的每个元素都是 X 中的一个成员; $NEG_A(X)$ 代表一个对象的集合中的每个元素都不是 X 的一个成员; $BND_A(X)$ 代表不知道一个对象的集合中的每个元素是否是 X 的一个成员。具体而言, 存在以下的规则:

- (1) $x \in POS_A(X) \Rightarrow x \in X$
- (2) $x \in NEG_A(X) \Rightarrow x \notin X$
- (3) $x \in BND_A(X) \Rightarrow$ 不知道

作者简介: 王 玮(1973~), 男, 博士后, 主研领域为数据挖掘, 神经网络; 蔡莲红, 教授、博导
收稿日期: 2000-10-16

不难看出, $NEG_A(X) = U - (PO S_A(X) \cup BND_A(X))$.

在代数近似空间中, 有:

(1) 上近似: $PO S_A^u(X) = \underline{A}(X) = \text{Int}_A(X)$

(2) 下近似: $PO S_A^l(X) = \overline{A}(X) = Ck_A(A)$

X的边界域为: $Ck_A(A) - PO S_A(X)$ 。不精确度可以通过以下公式度量:

$$\mu_A = \frac{|\text{Int}_A(X)|}{|Ck_A(X)|} \quad (1)$$

2 神经网络理论

神经网络广泛地应用于人工智能领域中模式识别、趋势产生和预测等方面, 伴随着不同的问题的出现, 神经网络的研究者们提出了大量各种类型的神经网络模型以适应不同问题的需要。神经网络是由众多的神经元组成, 每个神经元接受其它神经元和外界的输入信息。

经典的神经网络具有多层的拓扑结构、前馈和反馈等学习方式, 通常神经网络包含以下3个组成部分: 输入层、隐含层和输出层, 每层都由众多的神经元组成。输入层神经元接受外界环境的信息输入, 隐层神经元将隐层单元的信息输出至输出层, 输出层将信息输出至外界。每层神经元将输出仅仅前馈至下一个单元层, 因此这种网络称为前馈型网络, 各层神经元的输入和输出是受一定数学公式的约束, 即输出和输入之间是通过转移函数进行连接的。

神经网络模型包含的两个步骤是学习训练和测试。在训练阶段, 网络从训练集中归纳出学习准则, 这种训练过程称为监督学习, 经过监督学习输出层神经元得到期望的输出, 为了得到期望的输出网络调整神经元之间的连接权值, 在这个阶段, 输出的误差是一层向前一层反馈来适应权值的调整, 这成为反向的误差调整。在大多数学习等式中称这种学习法则为规则。

所谓检测阶段, 就是将训练好的网络检测已知期望输出的其它训练样集来检验该网络的性能, 如果能够达到期望值, 说明该网络已经训练成功。

医院中收集了大量病人的心跳次数、脉搏次数、血压值和体温等测量数据, 对于每个医生来说, 希望通过记录病人的测量数据, 对其将来可能出现的情况进行预测, 由于神经网络具有自学习功能可以很好地完成这种任务, 但是如何将这些测量数据进行归纳学习得到期望的模型决策是一个难题, 由于每天中一个病人的各项指标会有很多不同的测量数据, 对于同一个病人而言各项数据是一组数据集而不是单值, 且各项指标都需要记录连续几天的数据, 这样经典的神经网络预测模型就会产生较大的误差。

3 粗糙型神经网络

粗糙神经网络包含经典的神经元和粗糙神经元彼此相互连接, 一个粗糙神经元r在神经网络中可以看成是一对神经元, 这一对神经元中的一个为上边界定义为 \bar{r} , 另一个则定义为下边界为 \underline{r} 。一个粗糙神经元和另一个粗糙神经元之间有2或4个连接。图1表示了粗糙神经元中3种可能的连接。

上边界神经元和下边界神经元之间的连线表示信息的交换, 图1中的(a)图是两个神经元的全连接, 如果 \bar{r} 和 \bar{s} , \underline{r} 和 \underline{s} 都有连接则称粗糙神经元r和粗糙神经元s全连接, 如果粗糙神经元r和粗糙神经元s是全连接, 则从r到s有4种可能的连接。图1中的(b)图和(c)图从r到s只有两个连接, 如果粗糙神经元r激励粗糙神经元s, 则r和s之间的连接方式如(b)图所示,

如果粗糙神经元r抑制粗糙神经元s, 则r和s之间的连接方式如图(c)所示。

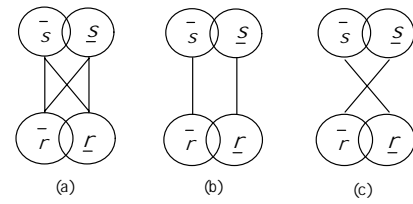


图1 两个粗糙神经元之间的3种连接方式

粗糙神经网络由一个输入层, 一个输出层和一个隐含层组成, 其神经元中既有粗糙神经元也有传统神经元, 输入层接受来自外界环境的输入信息, 其输出作为隐含层的输入, 隐含层的输出作为输出层的输入, 输出层的输出至外界环境。粗糙神经元的输出具有上边界和下边界的一对值, 而经典神经元的输出则是单值。

经典, 下边界或上边界的神经元输入根据以下公式计算权值:

$$Input_i = \sum_{j \text{ 和 } i \text{ 之间存在连接}} W_{ji} \times output_j \quad (2)$$

这里神经元i和j可以是经典神经元, 也可以是具有上下边界的粗糙神经元, 粗糙神经元的计算公式如下:

$$output_{\bar{r}} = \max(f(input_{\bar{r}}), f(input_{\underline{r}})) \quad (3)$$

$$output_{\underline{r}} = \min(f(input_{\bar{r}}), f(input_{\underline{r}})) \quad (4)$$

而经典神经元i的输出公式仅仅计算单个输出, 如下:

$$output_i = f(input_i) \quad (5)$$

这里的转移函数 $f(\cdot)$ 仍为sigmoid型函数, 定义如下:

$$f(u) = \frac{1}{1 + e^{-gain \times u}} \quad (6)$$

其中的增益系数是由系统的设计者确定的斜率, 当然还有其他类型的转移函数, 之所以广泛地采用sigmoid型转移函数, 是因为它在0至1的范围内具有连续的取值。

如果两个粗糙神经元部分连接, 则神经元的兴奋或抑制特性是根据连接权动态决定的, 网络开始时可以假设为兴奋或抑制类型, 假定初始化部分连接的粗糙神经元r与另一个粗糙神经元s之间为兴奋型的, 且 $W_{r\bar{s}} < 0$ 和 $W_{r\underline{s}} < 0$, 当 $W_{r\bar{s}} = W_{r\underline{s}}$, $W_{r\bar{s}} = W_{r\underline{s}}$ 时, 神经元r与神经元s之间的连接类型就由兴奋型转变为抑制型, 即若 (\bar{r}, \bar{s}) 和 $(\underline{r}, \underline{s})$ 之间存在连接, 则 (\bar{r}, \bar{s}) 和 $(\underline{r}, \underline{s})$ 之间就不存在连接。反之, 假定初始化部分连接的粗糙神经元r与另一个粗糙神经元s之间为抑制型的, 且 $W_{r\bar{s}} > 0$ 和 $W_{r\underline{s}} > 0$, 当 $W_{r\bar{s}} = W_{r\underline{s}}$, $W_{r\bar{s}} = W_{r\underline{s}}$ 时, 神经元r与神经元s之间的连接类型就由抑制型转变为兴奋型, 即若 (\bar{r}, \bar{s}) 和 $(\underline{r}, \underline{s})$ 之间存在连接, 则 (\bar{r}, \bar{s}) 和 $(\underline{r}, \underline{s})$ 之间就不存在连接。

训练和检测阶段粗糙神经网络类似于经典神经网络, 权值的连接是迭代修正的, 网络根据训练集重复进行迭代一次或多次, 权值根据学习等式进行修正, 学习阶段是依据常规的规则进行的, 其迭代修正公式如下:

$$W_{ji}^{new} = W_{ji}^{old} + \alpha \times output_j \times error_i \times f'(input_j) \quad (7)$$

对于sigmoid型转移函数的导数为:

$$f'(input_j) = input_j \times (1 - input_j) \quad (8)$$

公式(6)中的误差计算如下:

$$error_i = \text{期望 } output_i - \text{实际 } output_i \quad (9)$$

4 建立经典和粗糙神经网络模型

粗糙神经网络具有比经典神经网络更加优越的性能，这可以在医疗数据库中得到验证，对于医院中病人心跳次数、脉搏次数和血压值的数据建立预测模型如下：

图2是经典神经网络评估模型，模型中有4个输入层节点，3个隐含层节点，2个输出层节点，层与层之间都是全连接，其输入值是每天的平均模式。举例来说就是某个病人每天的心跳次数，脉搏次数，血压值，体温值的平均，其输出就是对病人健康状况的预测。

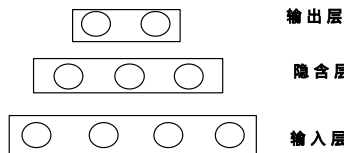


图2 经典神经网络的评估模型

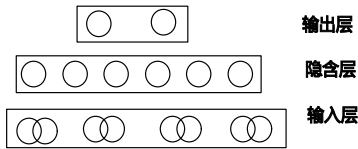


图3 粗糙神经网络模型一

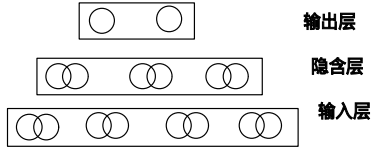


图4 粗糙神经网络模型二

图3是粗糙神经网络模型一，它包含4个输入粗糙神经元，6个隐含层神经元和2个输出层神经元，输入层的粗糙神经元与隐含层神经元之间是全连接的，隐含层和输出层的神经元仍然是经典的神经元，它们也是全连接的。因为该模型仅仅改变了输入层的神经元形式，所以在算法的计算中并没有增加太多的计算量。

图4是粗糙神经网络模型二，它包含4个输入粗糙神经元，3个隐含粗糙神经元和2个输出层神经元，输入层的粗糙神经元与隐含层的粗糙神经元之间是全连接的，隐含层的粗糙神经元和输出层的传统神经元之间也是全互联的。在算法的计算中只需加入数学公式(3)和(4)即可。

(上接第39页)

(3) IP网络对于视频流应用有着固有的缺点，它对于传输的服务质量，如延时、抖动、帧丢失率等都不能很好解决，而且由于目前的Internet网络传输带宽的限制，还不能提供很高质量的视频节目，一般在一些高速局域网上可采用MPEG-1或高质量的AVI、ASF视频。而在低速局域网及广域网上、可采用H.263、MPEG-4等低传输率的视频。

5 总结

本文介绍了一个基于IP网络的视频流系统的设计，着重介绍了系统的结构模型、协议模型、关键技术，并对实现提

粗糙神经元模型一或模型二都包含了每天的粗糙模式，也就是说包含了每天测量病人得到的上下边界数据，期望得到的是病人的健康状况，即只有健康和非健康两种，是单值输出，所以我们仍然采用传统的神经元。

5 结果分析

粗糙神经网络模型一隐含层采用的神经元仍然是传统的神经元，因此其误差比采用粗糙神经元作为隐含层神经元的模型二的大，这表明使用粗糙神经元代替传统神经元可以提高网络的性能。图5表明了训练过程中3种网络的误差曲线情况，从图中不难发现采用粗糙神经元的网络性能明显优于经典神经元的网络。

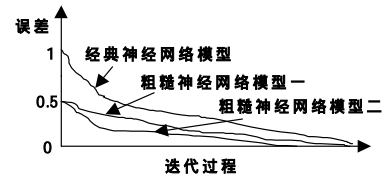


图5 经典和粗糙神经网络的误差曲线

6 结束语

本文将粗糙集理论和神经网络有机地集合，提出了基于粗糙集理论的神经网络，其实质是将传统神经元组合成粗糙神经元，这样可以有效地避免数据性能的泛化。文章最后将两种不同类型的粗糙神经网络和经典神经网络模型进行预测性能评估，其结果说明采用粗糙型神经元的网络性能其误差和收敛速度都明显优于经典的神经网络，且将隐含层的神经元改用粗糙神经元代替经典神经元则可以更好地提高网络的预测性能。

参考文献

- 1 Yao Y Y, Li X, Lin T Y, et al. Representation and Classification of Rough Set Models. Conference Proceeding of Third International Workshop on Rough Sets and Soft Computing, San Jose, California, 1994-11
- 2 Polkowski L, Skowron A, Zytkow J. Rough Foundations for Rough Sets. Conference Proceeding of Third International Workshop on Rough Sets and Soft Computing, San Jose, California, 1994-11
- 3 White H. Neural Network Learning and Statistics. AI Expert, 1989, 4(12):48-52

出了必要的建议。实际系统开发表明，该系统模型设计规范、功能分配合理，具有很高的应用和推广价值。

参考文献

- 1 Deering S E. RFC 1112: Host Extensions for IP Multicasting. IETF, 1989
- 2 Schulzrinne H, Casner S, Frederick R. RFC 1889: A Transport Protocol for Real-time Application. IETF, 1996
- 3 Schulzrinne H, Rao A, Lanphier R. RFC 2326: Real Time Streaming Protocol. 1998
- 4 Handley M, Jacobson V. RFC 2327: Session Description Protocol. 1998