神经网络规则抽取

周志华 陈世福
(南京大学计算机软件新技术国家重点实验室 南京 210093)
(zhouzh@nju.edu.cn)

摘 要 神经网络是一种黑箱模型,其学习到的知识蕴涵在大量连接权中,不仅影响了用户对利用神经计算技术构建智能系统的信心,还阻碍了神经网络技术在数据挖掘领域的应用.由于对神经网络规则抽取进行研究有助于解决上述问题,因此该领域已成为机器学习和神经计算界的研究热点.介绍了神经网络规则抽取研究的历史,综述了国际研究现状,对关于这方面研究的不同看法进行了讨论,并指出该领域中一些值得进一步研究的内容.

关键词 神经网络,机器学习,规则抽取,知识获取,数据挖掘

中图法分类号 TP183

RULE EXTRACTION FROM NEURAL NETWORKS

ZHOU Zhi-Hua and CHEN Shi-Fu
(*National Laboratory for Novel Software Technology, Nanjing University, Nanjing* 210093)

Abstract Neural network is a blackbox model whose learned knowledge is concealed in a large amount of connections. This has not only weakened the confidence of users in building intelligent systems using neural computing techniques, but also hindered the application of neural networks to data mining. Since extracting rules from neural networks help to solve those problems, this area has become a hot topic in both machine learning and neural computing communities. In this paper, the history of rule extraction from neural networks is introduced, the state-of-the-art of this field is surveyed, some controversies are discussed, and some issues valuable for future exploration in this area is indicated.

Key words neural networks, machine learning, rule extraction, knowledge acquisition, data mining

1 引 言

神经网络已经被成功地应用到很多不同的领域,并取得了大量成果.然而,其进一步的发展却受到了一个固有缺陷的限制,即神经网络学到的知识蕴涵在大量的连接权中,用户无法了解网络到底学到了什么、能处理什么样的任务,也无从知道网络如何进行预测、为什么得出这样或那样的推理结论.一般来说,"可解释性"是可靠系统的必备特性,由于通常的神经网络模型都是"不可解释"的,这在一定程度上影响了用户对通过神经计算技术构建智能系统的信心.虽然 Baum 和 Haussler[1]在 1989 年指出,"如果一个神经网络可以为大量训练例产生正确的结果,那么可以相信它也能为类似于训练例的未知示例产生正确结果",但这并没有消除用户对可理解性的偏好.此外,训练好的神经网络学习到的知识不能以容易理解的形式提交给决策者,这也是神经计

原稿收到日期:2001-03-12;修改稿收到日期:2001-12-17
本课题得到国家自然科学基金(60105004)和江苏省自然科学基金重点项目(BK2001202)资助

算技术难以用于数据挖掘领域的主要原因之一.

从神经网络中抽取符号规则将有助于增强神经网络的可理解性. 这方面的研究可上溯到 20 世纪 80 年代末 Gallant 的工作. 1988 年,Gallant[2]在其连接主义专家系统中,对不采用分布式知识表示的连接主义模型与析合范式之间的关系进行了研究,他根据推理强度对可用属性进行排序,从而构造出可以解释网络如何为某个给定事例产生结论的 if-then 规则. 虽然 Gallant 的主观目的是比较连接主义模型与析合范式的知识表示效率,但他的工作在客观上给出了一种神经网络规则抽取方法的雏形. 在此之后,尤其在 20 世纪 90 年代初,随着数据挖掘的兴起,机器学习和神经计算界的很多研究者意识到神经网络规则抽取的重要性,纷纷投身到这方面的研究中去. 1995 年,Andrews 等人[3]对该领域早期的一些成果做了一个很好的综述.

1998 年,《IEEE Transaction on Neural Networks》为神经网络规则抽取出版了一期专刊,Tickle 和 Andrews[4]在首篇文章中明确指出,从神经网络中抽取规则已经是当前神经网络界急需解决的问题. 最近几年,大批研究者涌入该领域,新成果层出不穷. 这充分说明该领域已经成为机器学习和神经计算研究的热点.

根据设计思想的不同,目前的神经网络规则抽取方法大致可以分成两大类,即基于结构分析的方法和基于性能分析的方法. 本文将对这两大类中的典型算法进行介绍和分析. 值得注意的是,有的研究者[5]将神经网络规则抽取作为一种规则学习方法进行研究,他们所关注的是抽取出的规则的泛化精度,而非规则对网络的保真度. 由于这些方法的目的和作用并不是增强神经网络的可理解性,因此本文没有对它们进行介绍.

2　基于结构分析的方法

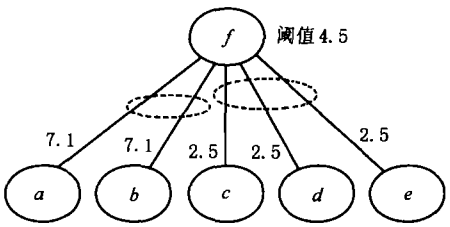
基于结构分析的神经网络规则抽取方法把规则抽取视为一个搜索过程,其基本思想是把已训练好的神经网络结构映射成对应的规则. 由于搜索过程的计算复杂度与神经元的数目呈指数级关系,当神经元很多时,会出现组合爆炸. 因此,此类算法一般采用剪枝聚类等方法来减少网络中的连接以降低计算复杂度.

1991 年,Fu[6]提出了 KT 算法. 该算法将网络中神经元的激活值通过近似处理为 0 和 1,将属性分为"正属性"和"负属性",前者对某结论起到确认作用,后者则起否定作用. 在所有的"负属性"都不出现的情况下,找出所有最多由 k 个"正属性"组合的集合. 然后从该集合中找出最多有 k 个前件的规则,这些规则在"负属性"部分或全部出现的情况下,仍然使某结论成立. 对单层网络,通过上述处理就可以抽取出规则. 对多层网络,KT 将隐层神经元视为"隐属性",然后按处理单层网络的方法一层一层地抽取出规则,最后通过"代入"等方法重写这些规则,直到规则中只出现输入属性和输出结论为止. 值得注意的是,虽然 KT 算法对"正"、"负"属性的区分降低了规则搜索复杂度,但这也限制了算法的规则抽取能力,使得抽取出的规则无法精确地描述原神经网络.

1992 年,Towell 和 Shavlik[7]为基于知识的神经网络(knowledge based artificial neural networks, KBANN)[8]设计了一种规则抽取算法,即 MOFN 算法. 该算法先用标准聚类算法合并 KBANN 中权值接近的连接以创建等价类,并将每个等价类的权值设为该组连接权的平均值,然后去掉那些对结果影响不大的等价类,在不调整权值的前提下对神经网络重新进行训练,最后直接根据网络结构和权值抽取出形如下式的 M-of-N 规则:

if (M of N antecedents are true)then….

M-of-N 规则形式不仅减少了抽取的规则数,还使得规则集比较简单. 另外,由于对连接进行了聚类,也使得规则搜索空间大为减少,从而较大地降低了规则抽取的时间开销. 图 1 给出了一个典型的抽取 M-of-N 规则的例子.



虚线表示等价类
抽取的规则: if a then f;
if b then f;
if 2 of $\{c,d,e\}$ then f.

图 1　MOFN 算法规则抽取

值得注意的是,在普通的神经网络中,由于连接权大多发散地分布在权值空间中,不像在 KBANN 中那样容易聚为等价类,因此一般来说,MOFN 算法仅适用于 KBANN. 1993 年,Craven 和 Shavlik[9]提出,可以用柔性权共享方法(soft weight-

sharing)[10]训练网络,然后用 MOFN 算法抽取规则.由于柔性权共享方法会促进连接权在训练中聚类,这样就使得 MOFN 算法的适用范围有所扩大.但是,由于MOFN 算法对神经网络的结构有一些很强的要求,例如要求神经元激活值为二值模式,每个神经元表示唯一的概念,网络输入为离散值等,这使其适用范围受到很大的限制.

1993 年,Sestito 和 Dillon[11]提出了一种利用抑制性单层网络为神经网络中每个输出神经元抽取相应规则的算法.他们首先将网络的输出神经元作为附加输入神经元,利用扩展后的输入神经元、初始的输出神经元以及一个隐含层建立多层网络,用 BP 算法对其进行训练.训练完成之后,对所有输入和附加输入神经元,根据下式计算它们之间的误差平方和 SSE,其中 a 为输入神经元,b 为附加输入神经元(即输出神经元),w_{aj}和 w_{bj}分别为神经元 a, b 与隐层神经元 j 之间的连接权.

$$SSE_{ab} = \sum_{j=0}^{\text{no. ofHiddenUnits}} (W_{bj} - W_{aj})^2.$$

SSE 实际上度量了输入神经元 a 和输出神经元 b 之间的接近程度,SSE_{ab}越小则说明输入 a 对输出 b 的作用越大.

然后,利用扩展后的输入神经元以及初始的输出神经元建立一个单层抑制性网络,用 Hebb 规则确定神经元间的抑制性连接权 $Weight_{ab}$,该权值实际上度量了输入神经元与输出神经元之间的相关度,值越小则说明某输入与某输出的关系越密切.在此基础上,对每一个输入神经元 a 和输出神经元 b,根据下式计算其 SSE 和抑制性连接权 $Weight_{ab}$的积.

$$Product_{ab} = Weight_{ab} \times SSE_{ab}.$$

最后,将乘积 $Product_{ab}$从大到小排序.对某个特定的输出,先找出乘积表中的截断点,即乘积表中的某一个位置,从该处断开的两个乘积在数值上至少相差 2～3 倍.然后以截断点以下的所有输入属性为规则前件,以输出为规则后件构造出合取规则.

Sestito 和 Dillon 的算法对前馈网络相当有效,可以抽取出较好的规则.但由于在规则抽取过程中需要额外地构造并训练两个神经网络,其时间代价相当高.

1995 年,Setiono 和 Liu[12]提出了一种从神经网络中抽取规则的三阶段算法.他们首先用权衰减(weight-decay)方法训练一个 BP 网络,该网络中较大的连接权反映了较重要的连接;然后对网络进行修剪,在预测精度不变的情况下删掉不重要的连接;最后通过对隐层神经元的激活值进行离散化,进而为每个输出神经元抽取相应的规则.该算法中离散化隐层神经元激活值的处理别具一格,这使其摆脱了很多规则抽取方法对激活值类型的限制,可以处理非二值模式的激活值.但是,由于无法保证网络的功能在离散化处理和修剪处理前后的一致性,因此该算法抽取的规则在保真度上有一定的缺陷.

此后,Setiono[13]提出了一种适用于 3 层前馈网络的通用型规则抽取算法.该算法不仅使用了 Setiono 和 Liu[12]设计的激活值离散化技术,还使用了一种独特的隐层神经元分裂技术,即当某个隐层神经元的输入连接数较多时,将其分裂为若干个输出神经元,并通过引入新的隐层神经元来构建子网络,从而递归地进行规则抽取处理.该算法可以产生相当精确的规则,但由于要训练多个子网络,其时间开销相当大.另一方面,该算法只适用于规模较小的网络,这是因为在输入神经元较多时,待分裂的隐层神经元数以及递归分裂的次数极大.

1997 年,Setiono 和 Liu[14]又提出了一种从 3 层前馈网络中抽取倾斜规则(oblique rule)的算法 NeuroLinear.与普通的规则相比,倾斜规则通常可以更好地表示边界与属性空间轴非垂直的判定域,从而较大地减少规则前件数.NeuroLinear 抽取的规则前件形式为

$$\sum_i c_i x_i < \eta.$$

NeuroLinear 通过修剪网络去除冗余连接,并对隐层神经元激活值进行聚类以降低组合复杂度.然后用隐层神经元聚类后的离散激活值表示输出层神经元的输出,用输入层神经元的激活值表示隐层神经元聚类后的离散激活值,从而得到层次形式的规则.再对这些规则进行合并,从而得到直接用输入属性表示网络输出的规则.

Setiono[15]还提出了一种从前馈网络中抽取 M-of-N 规则的算法 M-of-N 3.该算法首先训练并修剪一个以双曲正切函数为隐层神经元激活函数的网络,然后对隐层神经元激活值进行聚类,在此基础上产生分类规则,最后将规则的条件部分用 M-of-N 条件替代,从而得到 M-of-N 规则.

1996 年,Krishnan[16]提出了一种利用连接权排序的启发式信息来减少搜索空间的规则抽取算法 COMBO.该算法适用于具有布尔输入的前馈网络.它首先对特定神经元的扇入连接权按降序进行排列,然后构建排序权的"组合树",再对该树进行搜索

以建立由扇入连接权、阈值和神经元对应的概念所构成的规则.

1999 年,Alexander 和 Mozer[17]提出了一种从标准前馈网络中抽取析取、合取或 M-of-N 规则的基于权模板的方法.权模板是权空间中对应于特定符号表达式的参数化区域.该方法在适当表示下,可以有效地确定并例化模板参数,以产生对某神经元实际连接权的最优匹配,从而通过模板的组合产生对应于网络的规则.

2000 年,Tsukimoto[18]用布尔函数来近似神经网络中的神经元,并为此设计了一种多项式计算复杂度的近似算法,从而得到一种新的神经网络规则抽取方法.该方法适用于任何采用单调激活函数的网络,包括普通的多层前馈神经网络、循环神经网络(recurrent neural networks)等,并且与具体的网络训练算法无关.

同年,Bologna[19]提出了一种离散化可解释多层感知机 DIMLP(discretized interpretable multi layer perceptron),该网络第 1 隐层中每个神经元只与一个输入神经元相连,且其激活函数为阶梯形.这种网络产生的超平面与输入空间的轴平行,通过确定这些超平面的边界即可从网络中抽取出规则.

3　基于性能分析的方法

与基于结构分析的方法不同,基于性能分析的神经网络规则抽取方法并不对神经网络结构进行分析和搜索,而是把神经网络作为一个整体来处理,这类方法更注重的是抽取出的规则在功能上对网络的重现能力,即产生一组可以替代原网络的规则.

1990 年,Saito 和 Nakano[20]提出了两个规则抽取算法,RF(rule from facts)从示例集中抽取规则,RN(rule from networks)从神经网络中抽取规则,抽取的规则用析合范式表示.两个算法都是先从少数正例中抽取规则,然后根据未被覆盖的正例扩展规则,根据被覆盖的反例缩减规则,直到规则覆盖了所有正例并且不覆盖任何反例为止.虽然该算法并不对网络结构进行分析和搜索,但其要搜索正、反例空间,因此该算法在示例空间较大时将面临组合爆炸问题.

1992 年,Giles 等人[21]提出了一种从循环神经网络中抽取有限状态自动机的方法.假设网络具有 s 个状态单元,该方法首先将每个状态单元的激活值范围等分为 q 个区间,这样就将 s 维空间划分为 q^s 个部分.然后向网络提交输入序列,并记录状态转换、与各状态转换相关的输入向量以及网络输出值.在此基础上建立一个有限状态自动机,并用标准算法对其进行简化.此后,Gedeon 等人[22]提出了一种从循环神经网络中抽取确定性有限状态自动机的方法.该方法使用多项式时间的符号学习算法,仅通过对网络输入和输出行为的观察来进行推导,从而缓解了以往方法由于计算复杂度太高而导致的对网络状态聚类数以及隐循环状态神经元输出空间探查范围的限制,提高了抽取出的知识的保真度.

1994 年,Craven 和 Shavlik[23]为神经网络规则抽取任务下了一个定义,即"给定一个训练好的神经网络以及用于其训练的训练集,为网络产生一个简洁而精确的符号描述".显然,该定义来自于性能分析的角度.在该定义的基础上,Craven 和 Shavlik 将规则抽取视为一个目标概念为网络计算功能的学习任务,提出了一种基于学习的规则抽取算法.该算法使用了两个外部调用(Oracle),其中 EXAMPLES 的作用是为规则学习算法产生训练例,SUBSET 的作用则是判断被某个规则覆盖的示例是否都属于某个指定类.算法为每个分类产生各自的 DNF 表达式,它反复地通过 EXAMPLES 产生训练例,如果某训练例没有被当前该类的 DNF 表达式覆盖,则新规则被初始化为该训练例所有属性值的合取,然后反复尝试去掉该规则的一些前件,并且调用 SUBSET 来判断该规则是否与网络保持一致,从而使规则得以一般化.该算法不需使用特殊的网络训练方法,也不需将隐层神经元近似为阈值单元,但其计算量较大.

在文献[23]的基础上,1996 年,Craven 和 Shavlik[24]提出了 TREPAN 算法.该算法首先用训练好的神经网络对示例集进行分类,然后将该集合作为训练集提供给类似于 ID2-of-3[25]的决策树学习算法,从而构造出一棵与原网络功能接近的、使用 M-of-N 表达式作为内部划分的决策树.与文献[23]的算法相比,TREPAN 的计算量较低,但由于决策树的可理解性不如一阶逻辑表达式[26],TREPAN 抽取出的规则的可理解性也有所降低.1997 年,Craven 和 Shavlik[27]将 TREPAN 用于一个噪音时序任务,即美元-马克汇率预测,取得了比现有方法更好的效果.

Krishnan 等人[28]提出了一种从神经网络中抽取决策树的方法.该方法首先用训练好的神经网络产生数据,用遗传算法查询网络并抽取原型,然后通

过原型选择机制挑选原型子集,最后用标准归纳方法产生决策树.该树不仅可以用于分类预测,还有助于用户了解网络的功能.

1995年,Thrun[29]为前馈神经网络提出了一种基于有效区间分析(validity interval analysis)的规则抽取算法.该算法的关键是为所有或部分神经元找出激活区间,即有效区间.算法通过检查有效区间集合的一致性而不断排除导致不一致的区间.Thrun描述了两种不同的操作方式,即从特殊到一般和从一般到特殊.前者是从一个随机选择的示例开始,通过不断扩大相应的有效区间,逐渐得到一般的规则;后者则是从一个未加验证的假设集开始,通过有效区间来验证假设集中的规则.利用该算法可以抽取出精度较高的规则,但其以区间形式表示的规则前件使得规则的可理解性较差.另外,由于该算法的计算开销非常大,因此其只适用于对规则进行理论验证,难以完成实际的神经网络规则抽取任务.

1997年,Benitez等人[30]证明,多层前馈神经网络和基于模糊规则的系统是等价的,对任何一个以布尔函数为隐层神经元激活函数的单隐层前馈网络,均存在一个对应的模糊系统,使得网络可以由该系统的模糊规则进行解释.此后,Palade等人[31]提出了一种将神经网络转化为等价模糊集的方法.该方法利用遗传算法来确定与网络等价的模糊模型的正确结构以及隶属度函数的最佳形状.为了减少模糊规则的数目,该方法在考虑网络输入间关系的同时,还利用遗传算法来确定模糊规则的最佳层次结构.由这两种方法虽然可以很容易地获取一些模糊规则,但由于这些规则对不具有模糊数学背景的普通用户来说可理解性太差,因此其在实际应用,尤其是数据挖掘任务中用途不大.

本文作者曾经提出了一种基于统计的神经网络规则抽取算法SPT[32].与其它算法不同的是,SPT并不在规则抽取开始时离散化所有连续属性,而是仅在离散属性不足以缩小未知属性空间时,才选择一个聚类效果最佳的连续属性进行离散化,这样就降低了离散化处理中由于属性空间内在分布特性未知而造成的主观性.此外,SPT采用优先级规则形式,不仅使得规则表示简洁紧凑,还免除了规则应用时所需的一致性处理;利用统计技术对抽取出的规则进行评价,使得其可以较好地覆盖示例空间.SPT不依赖于具体的网络结构和训练算法,可以方便地应用于各种分类器型神经网络.与其它规则抽取算法相比,网络的各输入分量之间相关度较低时,由于SPT独特的离散化机制有助于降低无关属性交互引起的不良影响,因此,SPT可以取得更好的效果.此后,本文作者对SPT的规则抽取示例集生成方法以及利用统计测试确定规则取舍的技术进行了改进,提出了STARE算法[33].

4 讨 论

值得注意的是,在1956年宣告人工智能诞生的Dartmouth会议上,人工智能被定义为"对智能行为的符号化建模",因此,以神经网络为代表的连接主义与传统人工智能的主流符号主义从一开始就是对立的.经过几十年的发展,很多研究者认识到,"AI研究必须从其传统关注的特殊模式走出来.世界上并不存在一种最佳的知识表示或问题求解方法.当前机器智能的局限性在很大程度上是由以下两方面造成的,即力图寻找统一的理论,或者试图弥补那些在理论上很漂亮但在概念上却很虚弱的方法的不足……,我们所需的多功能性只能在更大规模的结构中找到,这些结构应能同时利用和管理若干种知识表示的优势,使得各种类型的表示可以相得益彰."(图灵奖获得者,人工智能先驱Marvin Minsky语)[34].而神经网络规则抽取的研究,正是将连接主义向符号主义拉近了一步.但是,在一些严格的连接主义者看来,这是连接主义向符号主义的投降.因此,在神经网络规则抽取研究不断取得进展的同时,也出现了一些不同的声音.

例如,有的研究者[35]认为,神经网络规则抽取与连接主义思想有根本的冲突.因为连接主义网络本身并不能产生规则,而且连接主义也没有为任何外部过程从网络中"读"出规则提供任何支持.在意大利科莫召开的2000年国际神经网络联合会议专门为此举行了一个论坛(panel)[36],一些著名学者如Giles, Werbos, Sun, Wermter, Taylor等人就该问题发表了看法.令人欣慰的是,几乎所有的发言人都表示了对规则抽取的支持.甚至连问题的提出者Roy本人也不得不承认,神经网络规则抽取工作对工程应用而言是非常有益的.发言人Sun甚至声称,神经网络规则抽取工作的进展标志着单纯强调结构、与符号主义对立的"强连接主义(strong connectionism)"的灭亡和强调功用、与符号主义越走越近的"弱连接主义(weak connectionism)"的诞生.由此可见,神经网络规则抽取在未来的若干年中仍然会是神经计算和机器学习界的一个研究热点.

5 进一步研究的内容

目前,在神经网络规则抽取的研究中仍然存在着很多有待解决的问题.我们认为,在将来的研究中,以下几方面的问题可望成为该领域的重要研究内容:

(1) 目前的很多规则抽取方法,无论是基于结构分析还是基于性能分析,都有很高的计算复杂度,这就使得这些方法在神经网络规模较大时难以奏效.而在真实世界问题中所使用的网络规模通常都比较大.因此,如果不能大幅度降低现有方法的计算复杂度或者提出新的计算复杂度小的方法,神经网络规则抽取领域研究成果的实用性很难得到保证.

(2) 目前神经网络规则抽取的工作主要集中在输出为离散值的神经网络分类器上,对输出为连续值的回归估计型神经网络研究得很少.而后者在时序预测、信号处理等领域中有广泛的应用,因此,从回归估计型神经网络中抽取规则将是一项重要的研究课题.

(3) 由于神经网络规则抽取的目的是改善神经网络的可理解性,因此,抽取出的规则必须易于为用户理解.然而,现有的很多方法只能抽取出晦涩难懂的规则.如何提高规则的可理解性是一个重要的问题.更进一步,如果能为规则可理解性建立一套比较客观的评价机制,不仅会促进神经网络规则抽取的研究,还会在数据挖掘领域产生积极的影响.

(4) 将进化计算等软计算技术应用于神经网络规则抽取,可能有助于获得更好的规则.在这方面,Hruschka 和 Ebecken[37]以及 Fukumi 等人[38]已经做了一些工作.前者用遗传算法对 BP 网络中对应于各输出分类的隐层神经元激活值进行聚类,然后通过列举各神经元输入与输出的关系来形成合取规则.后者利用遗传算法的模块化结构,在各模块中形成不同的结构的网络,然后利用确定性变异减少网络中的连接数,并通过病毒感染机制在对应于不同网络结构的模块间进行通信.

(5) 现有的神经网络规则抽取方法几乎都力图从网络中抽取产生式规则或类似形式的知识.但在某些场合,非规则形式的知识往往具有更好的表达能力.因此,从神经网络中抽取非规则形式的知识有可能成为今后的一大研究重点.在这方面,Melnik 和 Pollack[39]的工作有一定的启发性.他们提出了一种新颖的、从分类器中抽取定性知识的方法.该方法不依赖于输入空间的维数,可以抽取出分类决策区域图,不仅适用于神经网络,还可用于支持向量机等其它类型的分类器.

(6) 2000 年,Sun[40]提出了一种从神经强化学习模型中抽取知识的方法.该方法首先抽取出清晰的符号化规则,然后再抽取出完整的规划.该工作给出了混合型自主学习中亚符号(subsymbolic)知识向符号知识转换的通用框架.这显示出规则抽取工作的研究对象有可能逐渐从单纯的神经网络向混合学习模型以及神经网络集成[41]这样的复杂学习模型过渡.

参 考 文 献

1 E B Baum, D Haussler. What size net gives valid generalization. Neural Computation, 1989, 1(1): 151～160

2 S I Gallant. Connectionist expert systems. Communications of the ACM, 1988, 31(2): 152～169

3 R Andrews, J Diederich, A B Tickle. Survey and critique of techniques for extracting rules from trained artificial neural networks. Knowledge-Based Systems, 1995, 8(6): 373～389

4 A B Tickle, R Andrews. The truth will come to light: Directions and challenges in extracting the knowledge embedded within trained artificial neural networks. IEEE Trans on Neural Networks, 1998, 9(6): 1057～1068

5 张朝晖,陆玉昌,张钹. 利用神经网络发现分类规则. 计算机学报, 1999, 22(1): 108～112
(Zhang Zhaohui, Lu Yuchang, Zhang Bo. Discovering classification rules by using the neural networks. Chinese Journal of Computers(in Chinese), 1999, 22(1): 108～112)

6 L Fu. Rule learning by searching on adapted nets. In: Proc of the 9th National Conf on Artificial Intelligence. Anaheim, CA, 1991. 590～595

7 G G Towell, J W Shavlik. Interpretation of artificial neural networks: Mapping knowledge-based neural networks into rules. In: J E Moody, S J Hanson, R P Lippman eds. Advances in Neural Information Processing Systems 4, San Mateo, CA: Morgan Kaufmann, 1992. 977～984

8 G G Towell, J W Shavlik. Knowledge-based artificial neural networks. Artificial Intelligence, 1994, 70(1-2): 119～165

9 M W Craven, J W Shavlik. Learning symbolic rules using artificial neural networks. In: Proc of the 10th Int'l Conf on Machine Learning. Amherst, MA, 1993. 73～80

10 S J Nowlan, G E Hinton. Simplifying neural networks by soft

weight-sharing. Neural Computation, 1992, 4(4): 473～493

11 S Sestito, T Dillon. Knowledge acquisition of conjunctive rules using multilayered neural networks. International Journal of Intelligent Systems, 1993, 8(7): 779～805

12 R Setiono, H Liu. Understanding neural networks via rule extraction. In: Proc of the 14th Int'l Joint Conf on Artificial Intelligence. Montreal, Canada, 1995. 480～485

13 R Setiono. Extracting rules from neural networks by pruning and hidden-unit splitting. Neural Computation, 1997, 9(1): 205～225

14 R Setiono, H Liu. NeuroLinear: From neural networks to oblique decision rules. Neurocomputing, 1997, 17(1): 1～24

15 R Setiono. Extracting *M*-of-*N* rules from trained neural networks. IEEE Trans on Neural Networks, 2000, 11(2): 512～519

16 R Krishnan. A systematic method for decompositional rule extraction from neural networks. In: Proc of the NIPS96 Workshop on Rule Extraction from Trained Artificial Neural Networks. Denver, CO, 1996. 38～45

17 J A Alexander, M C Mozer. Template-based procedures for neural network interpretation. Neural Networks, 1999, 12(3): 479～498

18 H Tsukimoto. Extracting rules from trained neural networks. IEEE Trans on Neural Networks, 2000, 11(2): 377～389

19 G Bologna. Rule extraction from a multi layer perceptron with staircase activation functions. In: Proc of the IEEE-INNS-ENNS Int'l Joint Conf on Neural Networks. Como, Italy, 2000. 419～424

20 K Saito, R Nakano. Rule extraction from facts and neural networks. In: Proc of the Int'l Neural Network Conf. San Diego, CA, 1990. 379～382

21 C L Giles, C B Miller, D Chen *et al*. Learning and extracting finite state automata with second-order recurrent neural networks. Neural Computation, 1992, 4(3): 393～405

22 T Gedeon, P Wong, S Halgamuge *et al*. Rule extraction from recurrent neural networks using a symbolic machine learning algorithm. In: Proc of the 6th Int'l Conf on Neural Information Processing, vol 2. Perth, Australia, 1999. 712～717

23 M W Craven, J W Shavlik. Using sampling and queries to extract rules from trained neural networks. In: Proc of the 11th Int'l Conf on Machine Learning. New Brunswick, NJ, 1994. 37～45

24 M W Craven, J W Shavlik. Extracting tree-structured representations of trained neural networks. In: D Touretzky, M Mozer, M Hasselmo eds. Advances in Neural Information Processing Systems 8. Cambridge, MA: MIT Press, 1996. 24～30

25 P M Murphy, M J Pazzani. ID2-of-3: constructive induction of *M*-of-*N* concepts for discriminators in decision trees. In: Proc of the 8th Int'l Workshop on Machine Learning. Evanston, IL, 1991. 183～187

26 X Wu. Knowledge Acquisition from Databases. Norwood, NJ: Ablex, 1995

27 M W Craven, J W Shavlik. Using neural networks for data mining. Future Generation Computer Systems, 1997, 13(2-3): 211～229

28 R Krishnan, G Sivakumar, P Bhattacharya. Extracting decision trees from trained neural networks. Pattern Recognition, 1999, 32(12): 1999～2009

29 S Thrun. Extracting rules from artificial neural networks with distributed representations. In: G Tesauro, D Touretzky, T Leen eds. Advances in Neural Information Processing Systems 7. Cambridge, MA: MIT Press, 1995. 505～512

30 J M Benitez, J L Castro, I Requena. Are artificial neural networks black boxes. IEEE Trans on Neural Networks, 1997, 8(5): 1156～1164

31 V Palade, G Negoita, V Ariton. Genetic algorithm optimization of knowledge extraction from neural networks. In: Proc of the 6th Int'l Conf on Neural Information Processing. Perth, Australia, 1999. 752～758

32 周志华，何佳洲，尹旭日等. 一种基于统计的神经网络规则抽取方法. 软件学报，2001，12(2)：263～269
(Zhou Zhihua, He Jiazhou, Yin Xuri *et al*. A statistics based approach for rule extraction from neural networks. Journal of Software(in Chinese), 2001, 12(2): 263～269)

33 Zhou Zhihua, Chen Shifu, Chen Zhaoqian. A statistics based approach for extracting priority rules from trained neural networks. In: Proc of the IEEE-INNS-ENNS Int'l Joint Conf on Neural Networks. Como, Italy, 2000. 401～406

34 M Minsky. Logical versus analogical or symbolic versus connectionist or neat versus scruffy. AI Magazine, 1991, 12(2): 35～51

35 A Roy. On connectionism, rule extraction, and brain-like learning. IEEE Trans on Fuzzy Systems, 2000, 8(2): 222～227

36 周志华，何佳洲，陈世福. 神经网络国际研究动向. 模式识别与人工智能，2000，13(4)：415～418
(Zhou Zhihua, He Jiazhou, Chen Shifu. International trends of neural network research. Pattern Recognition and Artificial Intelligence(in Chinese), 2000, 13(4): 415～418)

37 E R Hruschka, N F F Ebecken. Applying a clustering genetic algorithm for extracting rules from a supervised neural network. In: Proc of the IEEE-INNS-ENNS Int'l Joint Conf on Neural Networks. Como, Italy, 2000. 407～412

38 M Fukumi, Y Mitsukura, N Akamatsu. A new rule generation method from neural networks formed using a genetic algorithm. In: Proc of the IEEE-INNS-ENNS Int'l Joint Conf

on Neural Networks. Como, Italy, 2000. 413～418

39 O Melnik, J Pollack. Using graphs to analyze high-dimensional classifiers. In: Proc of the IEEE-INNS-ENNS Int'l Joint Conf on Neural Networks. Como, Italy, 2000. 425～430

40 R Sun. Beyond simple rule extraction: The extraction of planning knowledge from reinforcement learners. In: Proc of the IEEE-INNS-ENNS Int'l Joint Conf on Neural Networks. Como, Italy, 2000. 105～110

41 周志华, 陈世福. 神经网络集成. 计算机学报, 2002, 25(1): 1～8
(Zhou Zhihua, Chen Shifu. Neural network ensemble. Chinese Journal of Computers(in Chinese), 2002, 25(1): 1～8)

周志华　男,1973 年生,博士,主要从事神经网络、机器学习、进化计算、模式识别、数据挖掘等方面的研究工作.

陈世福　男,1938 年生,教授,博士生导师,主要从事机器学习、分布式人工智能、知识工程、图像处理等方面的研究工作.

第八届中国机器学习学术会议(CMLW'2002)

征文通知

由中国人工智能学会机器学习专业委员会和中国计算机学会人工智能与模式识别专业委员会等单位联合主办,中山大学和广州舰艇学院联合承办的第八届中国机器学习学术会议,将于 2002 年 10 月下旬在广州举行. 欢迎国内各界专家学者研究人员踊跃投稿. 现将有关征文事宜通知如下:

征文范围

① 机器学习、知识获取　② 数据挖掘与知识发现
③ 神经网络、遗传算法　④ 模糊理论与技术、模糊神经网络
⑤ 计算智能　⑥ 公式发现
⑦ Rough Set 理论与应用　⑧ 多 Agent 系统学习
⑨ 人脑的智能活动及思维模型　⑩ 基于 CASE 的学习
⑪ 语音、图像处理与理解　⑫ 自然语言理解
⑬ 其他有关机器学习方面的文章

征文要求

① 论文必须未公开发表过,一般不超过 6000 字;
② 论文包括题目、作者姓名、作者单位、中英文摘要、关键字、正文和参考文献;另附作者姓名、单位、地址、邮编、传真及 E-mail 地址;
③ 一律为 A4 打印稿,一式两份(请采用 Word 排版);
④ 会议将出版论文集,打印清样格式在录取时另行通知;
⑤ 征文请寄:广州舰艇学院一系　陈文伟 (邮编:510431)

关键日期

截稿日期:2002 年 4 月 30 日;　录用通知:2002 年 5 月 30 日;　清样付印日期:2002 年 6 月 30 日

联系人:陈文伟,何　义,施平安　联系电话:(020)86416503-95578/95474

E-mail: Chemshihe@163.com

中国机器学习学会
中山大学计算机软件研究所
广州舰艇学院科研部
2002 年 1 月 1 日