

构造型神经网络双交叉覆盖增量学习算法*

陶品⁺, 张钹, 叶榛

(清华大学 计算机科学与技术系, 北京 100084)

(清华大学 智能技术与系统国家重点实验室, 北京 100084)

An Incremental BiCovering Learning Algorithm for Constructive Neural Network

TAO Pin⁺, ZHANG Bo, YE Zhen

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

(State Key Laboratory of Intelligent Technology and System, Tsinghua University, Beijing 100084, China)

+ Corresponding author: Phn: 86-10-62775113, E-mail: taopin97@mails.tsinghua.edu.cn

<http://www.cs.tsinghua.edu.cn>

Received 2002-03-28; Accepted 2002-05-17

Tao P, Zhang B, Ye Z. An incremental bicovering learning algorithm for constructive neural network. *Journal of Software*, 2003,14(2):194~201.

Abstract: The algorithm of incremental learning in cover based constructive neural networks (CBCNN) is investigated by using BiCovering algorithm (BiCA) in this paper. This incremental learning algorithm based on the idea of CBCNN can set up many positive-covers and negative-covers, and can modify and optimize the parameters and structure of the neural networks continuously, and can add the nodes according to the need and prune the redundant nodes. BiCA algorithm not only keep the advantages of CBCNN but also fit for incremental learning and could enhance the generalization capability of the neural networks. The simulational results show that the BiCA algorithm is not sensitive to the order of the sample and could learn quickly and steady even if the performance of initial CBCNN is not very good.

Key words: pattern recognition; neural network; machine learning; cover; incremental learning

摘要: 研究了基于覆盖的构造型神经网络(cover based constructive neural networks,简称CBCNN)中的双交叉覆盖增量学习算法(BiCovering algorithm,简称BiCA)。根据CBCNN的基本思想,该算法进一步通过构造多个正反覆盖簇,使得网络在首次构造完成后还可以不断地修改与优化神经网络的参数与结构,增加或删除网络中的节点,进行增量学习。通过分析认为,BiCA学习算法不但保留了CBCNN网络的优点与特点,而且实现了增量学习并提高了CBCNN网络的泛化能力。仿真实验结果显示,该增量学习算法在神经网络初始分类能力较差的情况下具有快速学习能力,并且对样本的学习顺序不敏感。

关键词: 模式识别;神经网络;机器学习;覆盖;增量学习

中图法分类号: TP181 文献标识码: A

* Supported by the National Natural Science Foundation of China under Grant No.60135010 (国家自然科学基金); the National Grand Fundamental Research 973 Program of China under Grant No.G1998030509 (国家重点基础研究发展规划(973))

第一作者简介: 陶品(1974—),男,安徽芜湖人,博士生,主要研究领域为人工神经网络,机器学习。

学习是神经网络的一个重要能力,神经网络可以通过对一批样本的学习获得网络的平均误差情况以确定如何修改网络参数^[1-3],这样的学习方式可以看成是批量学习(batch learning).但是在许多情况下无法进行批量学习,比如,无法一次获得足够多的样本,实际应用要求网络中保存的知识不断更新,或者需要对新样本进行实时学习等,这时神经网络就需要进行增量学习(incremental learning).增量学习可以通过对每个新样本的具体情况和识别结果进行分析来调整神经网络,根据已有的知识去学习新的知识,灵活地适应环境的动态变化,因此,目前对增量学习算法的研究比较多^[3-5].

增量学习可以通过调整参数来实现,也可以通过增加或删除神经元、改变网络拓扑结构来完成^[3].网络结构虽然在很大程度上决定了网络的学习能力^[4],但却很难事先确定.增量学习方法不仅可以避免网络结构难以选择的问题,而且也可以有效地避免传统参数收敛算法中的局部极小问题.结构调整包括 3 种类型:增加^[6]、删减^[7]或这两者的组合^[8-10].一般来说,网络结构的增加策略比删减策略更易于制订和实现,因为后者极易造成系统性能的振荡,因此 Platt 等人提出了各种结构的调整算法^[11-17].

本文提出一种基于覆盖思想的构造型神经网络模型上的增量学习算法,该算法可自适应地控制网络的结构与规模,在不同初始知识和学习顺序的条件下均有较好的学习效果.本文第 1 节介绍基于覆盖的构造型神经网络的基本思想.第 2 节阐述在该构造型网络基础上实现增量学习的思路与方法.第 3 节给出具体增量学习算法的实现.第 4 节分析了该算法的泛化能力.最后通过两个仿真实验来验证该算法的有效性.

1 基于覆盖思想的构造型神经网络

传统的 BP 算法一方面难以事先估计网络结构及隐层神经元个数,另一方面由于采用梯度下降思想,因此收敛速度慢,难以避免局部极小等问题.BP 算法对样本的处理是逐个进行的,所以也可以被看成是一种增量学习算法.但是由于各个神经元参数的调整互相关联,其计算量(计算复杂度)与神经元个数的指数阶成正比.为了解决 BP 网络的上述问题,构造型网络应运而生,并且这些构造型方法都与网络的增量学习密切相关.Frean 提出了 UpStart 算法^[18],Hirose 提出了增加/删除算法^[8],Fahlman 提出了 Cascade-Correlation 算法^[19],Karim 提出了 FlexNet 方法^[20],Zhang 给出了在不限制隐层神经元数目的情况下构造型前馈神经网络的最优设计^[1],Zhou 将 ART 方法与域理论结合,可自动改变网络隐层神经元数目,并可根

根据学习结果反向抽取符号规则^[21].
基于覆盖思想的构造型前馈神经网络是 1998 年由张铃教授提出的一种新型神经网络模型^[2],该模型具有可理解性强、识别率高、计算速度快、可以保证训练样本集上 100% 的识别率等优点,适合于处理大规模的分类任务.1998 年,Zhang 证明了该神经网络模型可以等价于一个三层前馈神经网络^[22].基于覆盖思想的构造型前馈神经网络的基本思想是:假定训练样本或待识别样本是 n 维向量,构造型神经网络将所有的训练样本 x 经过归一化并投影变换到 $n+1$ 维单位超半球面上.这样,模式识别问题被转化为在 $n+1$ 维单位超半球面上的覆盖问题(如图 1 所示),通过在超球面上寻找可以对训练样本集进行正确分类的较优覆盖来实现对样本空间的划分,并根据由训练集得到的覆盖来判断待识别样本的类别.

由于多类识别问题可以转化为多个两类覆盖问题的组合,因此下面只讨论两类样本的分类问题.一个已有的分类算法是“交叉覆盖法”算法^[2],它的基本思路是在超球面上先求一个领域 C^1 ,它只覆盖 K^1 类中的点,而不覆盖 K^2 类中的点,然后将被 C^1 覆盖的点删除,并对余下的点求另一领域 C^2 ,它只覆盖 K^2 类的点,而不覆盖 K^1 类的点,然后将被 C^2 盖住的点删去,如此交叉进行覆盖,直到 K^1 (或 K^2)类的点全部被删除为止.通过交叉覆盖算法,我们可以构造出一个等价的三层前馈神经网络,因而实现了模式分类与识别.

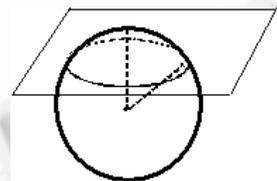


Fig.1 Cross of supersphere and superplane

图 1 超平面与超球相交

2 增量学习算法思想

文献[2]中的交叉覆盖算法是一种集中式的构造算法,该算法除了具有构造型前馈神经网络的一般优点以外,其突出特点是覆盖数量可以比较少.比如对于著名的双螺旋线分类问题(如图 2 所示),实心点与叉形点分别代表两类样本,它们各自构成一条螺旋线,两根螺旋线互相缠绕在一起,用普通分类方法很难将其分开,但是交叉覆盖方法只用了 7 个覆盖就将训练样本集完全区分开,并且根据文献[2]中的实验结果,该覆盖集对于双螺旋线上的所有测试样本也具有相当高的识别率.

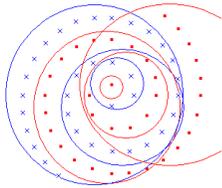


Fig.2 Alternative covering on dual spiral data

图 2 双螺旋分类问题中的交叉覆盖

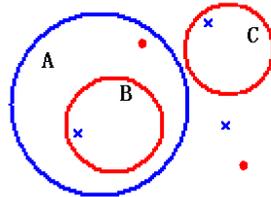


Fig.3 The problem of alternative covering

图 3 交叉覆盖中的问题

但是交叉覆盖算法还存在两个弱点:(1) 最终覆盖的质量受先验知识和初始种子样本的选择以及样本分布的影响较大,设想如果交叉覆盖算法不是从双螺旋线的中心样本点开始进行的话,则难以获得如图 2 所示的优美的结果;(2) 交叉覆盖算法还不利于进行增量学习,用该方法得到的覆盖集合是一个覆盖序列,覆盖之间具有优先级的差别,先产生的覆盖具有较高的优先级,反之则具有较低的优先级,优先级的引入使得覆盖的再学习变得比较麻烦.比如在图 3 中,A 覆盖代表对叉形样本点的覆盖,B 和 C 覆盖代表对实心样本点的覆盖.图中画出了不能被正确识别的 3 个叉形样本和两个实心样本.在这种情况下,要想通过调整覆盖来纠正识别错误,需要修改原有覆盖的参数(覆盖中心、覆盖半径)或者直接创建新覆盖,这样,原来的覆盖序列将被调整为一棵错综复杂的覆盖树.这样的复杂结构非常不利于覆盖的进一步调整、删除和合并等操作;而且整个网络只有少量无冗余的覆盖,覆盖之间不能形成竞争学习.

为克服以上弱点,我们设计了双交叉覆盖算法(BiCovering algorithm,简称 BiCA).该算法降低了覆盖层次,只设立二级交叉覆盖(称为一个覆盖簇).在网络的训练和学习过程中可以产生众多的覆盖簇,每个覆盖簇均有通过学习获得改进的机会,并互相竞争.假定待分类问题是 n 类识别问题,该问题将首先被转化为多个 2 类识别问题,如 Wang^[23]将 n 类识别转化为具有 $n-1$ 个分类器的二叉树,但是该方法在没有先验知识的情况下将面临二叉树结构选择的困难.我们的处理方式是将 n 分类问题转化为 n 个二分类问题,即对于每个待识别的类,所有参与学习的样本点都被分为属于该类和不属于该类的两个样本集合,整个识别过程由 n 个这样的二分类识别过程组合而成,这样做可以简化后面的计算.学习完成后,待识别样本究竟属于哪类,则根据该样本属于各类的可靠度来选出最优的结果.

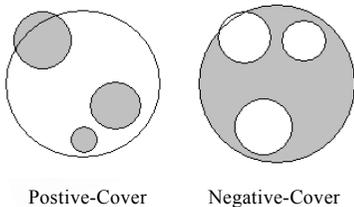


Fig.4 Postive-Cover and negative-cover

图 4 正覆盖簇与反覆盖簇

最简单的正覆盖簇是只包含父覆盖的正覆盖簇.

从覆盖簇的角度来看,一个二分类问题的增量学习即是构造并改进正、反覆盖簇集合,使正、反覆盖簇集合能更简洁、更准确地反映真实样本空间的分布的过程.算法开始时可以根据少量(或任意多)初始训练样本中的数据获得若干个最简单的正覆盖簇和反覆盖簇,然后试图在这些正、反覆盖簇的基础上获得新的覆盖簇,并

使新覆盖簇可以用更少的覆盖去表达原来较多个覆盖所表达的“意思”,使之可以更好地实现对样本空间的划分,获得更好的泛化能力.由于正、反覆盖簇最多只具有双层结构,因此调整起来涉及的覆盖较少,学习算法比较简单.此外,由于可生成多个正、反覆盖簇,因此覆盖簇表达的“知识”之间有冗余,这也就为竞争择优打下了基础.

3 增量学习算法设计

根据上面的增量学习算法思想,我们对正、反覆盖簇的构造和改进算法设计如下:

3.1 正反覆盖簇的构造

假定起始训练样本集为 Y ,并且该训练样本集被分为 Y_1 和 Y_2 两部分,分别代表属于和不属于某类的训练样本.构造算法如下.

算法 1. 覆盖簇构造算法.

Step 1. 随机选择一个尚未被任何覆盖簇盖住的 Y_1 类样本 y ,以 y 为种子点,求最大正覆盖,并以该覆盖为一个最简单的正覆盖簇,重复该步骤,直到所有正类训练样本均被覆盖;对 Y_2 中的反类训练样本同样构造出反覆盖簇;

Step 2. 对于每个正覆盖簇,取其父覆盖 c_1 ,依次去掉离覆盖 c_1 中心最近的其他反覆盖簇所包含的样本点,并以被去掉的反覆盖簇中心为种子点通过求重心、平移获得新的、更大的正覆盖;如果新覆盖更加简化(见算法 2),用新覆盖和所有被去掉的反覆盖簇的父覆盖组成一个新的正覆盖簇,并将该正覆盖簇加入覆盖簇集合中;

Step 3. 对于每个反覆盖簇,执行与步骤 2 类似的操作;

Step 4. 如果在步骤 2 和步骤 3 中,覆盖簇集合中又增加了新覆盖簇,则转回 Step 2,继续寻找更简化的正、反覆盖簇.

在实际识别问题中,如果训练样本集比较大或者样本向量维数非常高,可以在算法 1 的第 3 个大循环中设置某些终止条件,如覆盖簇个数上限、大循环次数、新覆盖简化的判断标准等,以避免覆盖簇个数过多,计算时间过长.此外,算法 1 中判断新覆盖是否更加简化的算法如下(以计算新正覆盖簇为例).

算法 2. 新覆盖簇简化判断算法.

$Count=1+$ 被去掉的反覆盖簇个数

对于覆盖簇集合中的每个正覆盖簇: {

fc = 该正覆盖簇的父覆盖

 IF (fc 所覆盖的属于该类的样本点 90% 以上都在新正覆盖中) $Count=Count-1$;

}

IF ($Count < 0$) 新覆盖更加简化

ELSE 新覆盖并未简化

3.2 覆盖簇的改进与优化

当用已有的覆盖簇集合不能正确判断某未知样本的类别时,我们需要启动覆盖簇改进算法,这是增量学习算法的重要组成部分.覆盖簇改进算法首先根据学习样本的正确分类尝试扩大已有覆盖的大小和增加新覆盖两种方法来改进原有的覆盖簇.关于覆盖簇的删除策略,限于篇幅,本文中略去.

算法 3. 覆盖簇改进算法.

Step 1. 初始化 t = 新样本的正确类别;

Step 2. 如果某个新的学习样本的类别被错误识别了,令 k = 被误识的类别, y = 该新样本;

Step 3. 如果 y 在某个 k 类的正覆盖簇 c 中,转 Step 4; 否则 y 不在任何已有的覆盖中,则转 Step 7;

Step 4. 先试图扩大 c 中的所有儿子覆盖,使之覆盖样本 y ,如果该操作成功转 Step 8;

Step 5. 为样本 y 生成新的儿子覆盖,转 Step 8;

Step 6. 对于所有 t 类正覆盖簇 e ,尝试扩大所有 e 的父覆盖大小,使之覆盖样本 y ,如果扩大成功则转 Step 7;

Step 7. 如果扩大父覆盖失败,即没有任何 t 类正覆盖可以通过扩大来盖住样本 y ,则为 y 生成一个新的 t 类

正覆盖簇;

Step 8. 算法结束.

4 BiCA 算法的泛化能力分析

在第 2 节的增量学习算法介绍中,我们提到了网络的泛化能力,该算法学习出来的神经网络的泛化能力真的好吗?在机器学习中,什么因素会影响网络的泛化能力,如何才能获得更好的泛化能力呢?

4.1 统计学习理论

经典的(参数)统计估计方法是基于传统统计学的,这种方法的局限性表现在,它需要已知样本分布或者待估计函数只是参数未知而基本形式已确定,而且该方法研究的是样本数目趋于无穷大时的渐近理论.但在实际问题中,这些条件经常不能满足,因此一些理论上很优秀的学习算法在实际中往往表现得不尽人意^[24].除经典参数估计方法以外,以神经网络为代表的经验性非线性方法利用已知样本建立非线性模型,克服了传统参数估计方法的困难,但是这种方法数学理论基础不够完善,并且在收敛性、学习速度等方面不太理想.

与前两种方法不同的是,统计学习理论(statistical learning theory,简称 SLT)具有较坚实的数学基础,其核心思想是以结构风险最小化理论挑战了传统学习理论中的经验风险最小化理论,指出仅仅是经验风险最小化并不足以保证学习机器的泛化能力,且容易出现过学习(over-fitting)现象.SVM 方法是 SLT 理论的一个实现,在 SVM 中通过定义不同的核函数,实现分类器的构造^[25].

SLT 的核心概念之一是 VC 维,它描述了学习机器的复杂性(学习能力),并在此概念基础上发展出了一系列关于统计学习的一致性、收敛速度、推广性能等重要结论,为解决有限样本学习问题提供了一个统一的框架.在 SLT 中,学习机器的实际风险 $R(w)$ 以至少 $1-\eta$ 的概率满足如下不等式:

$$R(w) \leq R_{emp}(w) + \Phi(n, h, \eta), \tag{1}$$

其中 $R_{emp}(w)$ 代表学习机器在训练样本集(学习样本集)上的经验风险, n 代表样本数目, h 为学习机器的 VC 维, w 为学习机器的参数, Φ 函数为置信函数,其值与 h 变量成正比.因此根据统计学习理论,要降低学习机器的实际风险的上界,既可以通过降低经验风险来实现,也可以通过降低置信范围来完成,但经验风险和置信范围都可能受学习机器 VC 维 h 及其他因素的影响,并且互为矛盾(如图 5 所示),因此 SVM 算法的本质是在经验风险与置信范围之间选择一个最佳的折衷.

在 SVM 的实际操作中,一般来说是先选择一个核函数及其参数(即确定学习机器的 VC 维),在置信范围相对固定的条件下获得经验风险最小化,从而达到最佳泛化能力的目标,如 Vapnik 所举的手写数字识别的例子即是这样^[25].基于覆盖的构造性神经网络及其双交叉覆盖增

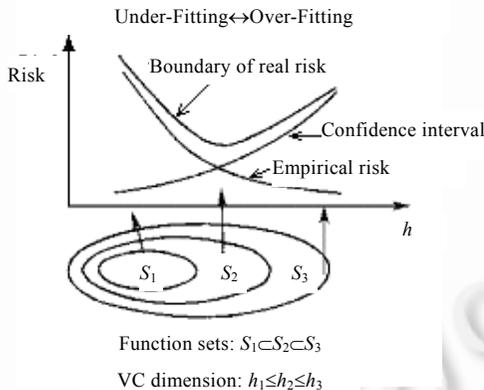


Fig.5 Empirical risk and VC-dimension
图 5 经验风险和 VC 维

量学习算法是否也可以从中获得启示呢?实际上,BiCA 算法从 SLT 的角度来看正是在保持经验风险极小(或为 0)的条件下不断降低学习机器的 VC 维,从而获得更好的泛化能力的构造性学习算法.

4.2 双交叉覆盖结构的 VC 维分析

根据统计学习理论中对 VC 维的直观定义,对一个指示函数集,如果存在 h 个样本能够被函数集中的函数按所有可能的 2^h 种形式分开,则称函数集能够把 h 个样本打散,函数集的 VC 维就是它能打散的最大样本数目 h ^[25].在同一类函数簇中,VC 维越大则学习机器越复杂(容量越大),泛化能力将越差,只是目前尚没有关于任意函数集 VC 维计算的通用理论,但是我们知道,简单的 n 维实数空间中线性分类器和线性实函数的 VC 维是 $n+1$,而正弦函数的 VC 维为无穷大.

在基于覆盖思想的构造性神经网络中,由于 $n+1$ (n 为样本矢量的维数) 维空间中的单位超半球面上样本之间存在距离关系,每个覆盖都可以根据距离的不同将样本空间分成两个部分:

$$\begin{aligned} wx + b &\geq 0 \\ wx + b &< 0 \end{aligned} \quad (2)$$

其中 w 为覆盖中心, b 为覆盖半径. 从式(2)可以看出这种分类函数也是一种线性划分函数,因此一个 $n+1$ 维空间中的超半球面上的覆盖可将最多 $n+1$ 个样本以任意形式打散,也就是说,包括一个覆盖的覆盖网络其 VC 维为 $n+1$. 进而在不考虑交叉覆盖的情况下,包含 k 个独立覆盖的覆盖网络,其 VC 维:

$$h_k \leq k(n+1). \quad (3)$$

因为若存在 $k(n+1)+1$ 个样本可以被包含 k 个覆盖的覆盖网络任意打散,则如果将这些样本分成 k 组,根据抽屉原则,必有一组的样本数大于 $n+1$,即在该组中一个覆盖不能将其任意打散,故包含这个覆盖的覆盖网络经过组合也不能将所有的样本打散,这与假设矛盾.

当进一步考虑交叉覆盖时,一个包含两个覆盖的双交叉覆盖簇由一个父覆盖和一个子覆盖构成,子覆盖由于受到父覆盖的限制,其覆盖范围只能位于父覆盖内部或与之存在相交关系,因此这样一个双交叉覆盖(其 VC 维为 h'_2)对样本的打散能力必然弱于两个独立的覆盖构成的覆盖网络(其 VC 维为 h_2 , h_2 满足式(3))对样本的打散能力,故可得:

$$h'_2 \leq h_2. \quad (4)$$

推而广之,由 t 个覆盖构成的双交叉覆盖簇的 VC 维小于等于 t 个独立覆盖集合的 VC 维;进一步根据式(3)、式(4)我们可以得到,包含 k 个覆盖的双交叉覆盖簇集合的 VC 维小于等于 k 个独立覆盖组合的 VC 维,即

$$h'_k \leq h_k \leq k(n+1). \quad (5)$$

在 SLT 中,具有有限 VC 维的学习机器的经验风险与实际风险的一致收敛,并且 VC 维越小,学习机器的泛化能力越好^[25]. 双交叉覆盖增量学习算法一方面可以通过降低总覆盖数量来减小覆盖网络的 VC 维,另一方面可以通过将若干个独立覆盖组合为一个双交叉覆盖簇来简化网络结构,从而从数量和质量两个角度去降低学习机器的 VC 维,以期获得更好的泛化能力. 实际上,该思想与传统神经网络设计中认为网络结构越简单泛化能力越好的经验认识是一致的,只是在这里根据 SLT 理论找到了理论上的依据.

比较而言,传统的神经网络方法(如 BP 算法)在经验风险最小化上进行了努力,但是在简化网络结构(降低学习机器 VC 维)上则缺少有效的、可理解性强的、易于操作的手段,因此容易产生过学习. 而本文介绍的双交叉覆盖学习算法则可以在保证经验风险很小的条件下,有效地逐步降低网络的复杂度,从而让学习机器获得尽量高的泛化能力. 如果我们对双交叉覆盖算法中经验风险设置一个参数,要求经验风险小于某个常数 e 即可,还可以进一步在经验风险与网络的置信范围之间获得某种基于 e 的折衷.

5 实验结果分析

我们分别用双螺旋线数据和一个实际的三维模型特征识别问题的样本数据进行实验,实验环境为 PIII 600 普通 PC 机,实验结果验证了本文中关于双交叉覆盖增量学习算法学习效果 and 泛化能力的分析.

双螺旋线数据是由极坐标系下方程 $r = \theta$ 和 $r = -\theta$ 两条曲线互相缠绕而成,取 $\pi/2 \leq \theta \leq 6\pi$, 每条曲线上随机选取 1 000 个(共 2 000 个)样本点,其中各取 100 个(共 200 个)样本点构成学习样本集合(在学习样本中再各取 95%, 共 190 个作为初始训练样本),其余 $900 \times 2 = 1800$ 个样本点作为测试样本集. 每个样本点的数据包括其在平面坐标系中的 x, y 坐标以及它属于哪条螺旋线. 在增量学习阶段,每次从 200 个学习样本集合中随机抽取一个样本用于学习,并计算进行增量学习后的人工神经网络的总覆盖簇数目和在测试样本集上的泛化误差. 覆盖簇数目的变化曲线如图 7(a)所示(横轴表示已学习样本数目,纵轴表示总覆盖数目),泛化误差曲线如图 7(b)所示(横轴表示已学习样本数目,纵轴表示泛化误差的百分数). 在增量学习

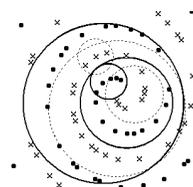


Fig. 6 A cover cluster after learning

图 6 增量学习后的一个覆盖簇

后的覆盖簇中,我们也发现了一些与图2相似的覆盖簇(如图6所示,其中虚线表示反覆盖簇),这反映了双交叉覆盖学习算法在没有先验知识的情况下,通过网络结构的不断优化找到了更能反映样本空间分布的覆盖簇。

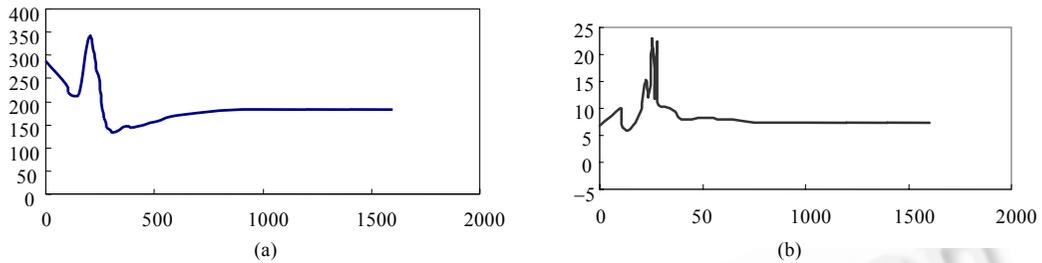


Fig.7 The curve of cover number and generalization error rate of incremental learning (dual spiral data set)

图7 增量学习的覆盖簇数目与泛化误差变化曲线(双螺旋线数据集)

应用该算法继续在一个真实的三维模型特征识别数据集上进行实验.数据集中包括5类三维模型的特征(台阶、槽、口袋、通孔、凸台)共1260个样本,每个样本预处理后是一个75维的向量,该向量描述了三维模型中顶点、边及面之间的拓扑关系以及面间夹角大小等几何属性.在这1260个样本中,我们随机抽取420个样本点作为学习样本,其他的作为测试样本用于计算泛化误差.图8(a)展示了总覆盖数目随学习样本数目的变化情况,初始阶段神经网络根据新样本不断学习新知识,总覆盖数目不断增加,但在经过400个左右样本的学习后,总覆盖数目就开始递减,并逐渐趋于稳定,同时泛化误差也逐渐减小(如图8(b)所示).

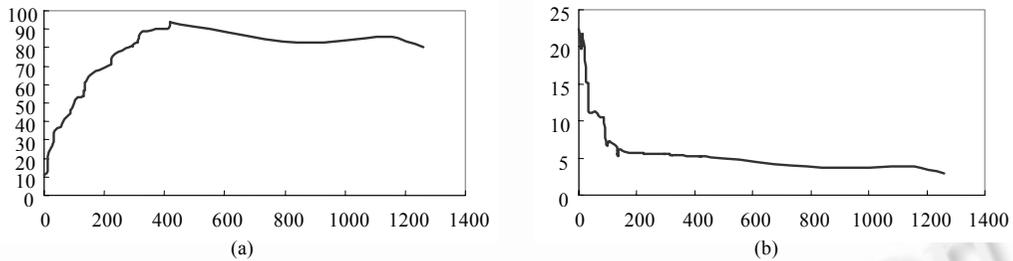


Fig.8 The curve of cover number and generalization error rate of incremental learning (3D data set)

图8 增量学习的覆盖簇数目与泛化误差变化曲线(3D特征识别数据集)

我们改变其他条件,如增量学习的初始条件、样本学习顺序、其他训练数据等,都得到了类似的学习效果.从本文的分析和这些实验结果可以看出,双交叉覆盖增量学习算法可以自适应地寻找到适当的网络结构和规模,在保证经验风险较小的条件下逐步降低学习机器的VC维,从而获得较好的泛化能力。

References:

- [1] Zhang L, Wu FC, Zhang B, Han M. A learning and synthesis algorithm of multi-layered feed forward neural networks. *Journal of Software*, 1995,6(7):440~448 (in Chinese with English Abstract).
- [2] Zhang L, Zhang B, Yin HF. An alternative covering design algorithm of multi-layer neural networks. *Journal of Software*, 1999, 10(7):737~742 (in Chinese with English Abstract).
- [3] Fu LM. Incremental knowledge acquisition in supervised learning networks. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: System and Humans*, 1996,26(6):801~809.
- [4] Wang EH-C, Kuh A. A smart algorithm for incremental learning. *International Joint Conference on Neural Networks*, 1992,3:121~126.
- [5] Engelbrecht AP, Brits R. A clustering approach to incremental learning for feedforward neural networks. In: *Proceedings of the International Joint Conference on Neural Networks (IJCNN 2001)*, Vol 3. 2001. 2019~2024.
- [6] Ash, T. Dynamic node creation in backpropagation networks. Technical Report, ICS Report 8901, University of California, 1989.
- [7] Mozer MC, Smolensky P. Using relevance to reduce network size automatically. *Connection Science*, 1989,1(1):3~16.

- [8] Hirose Y, Koichi Y, Shimpei H. Back-Propagation algorithm which varies the number of hidden units. *Neural Networks*, 1991,4:61~66.
- [9] Alpaydin, E. GAL: networks that grow when they learn and shrink when they forget. Technical Report, TR91-032, International Computer Science Institute, 1991.
- [10] Stefan W, Manuela M. Towards constructive and destructive dynamic network configuration. In: *Proceedings of the ESANN 1996*. 1996. 123~128. <http://www.dice.ucl.ac.be/Proceedings/esann/esannpdf/es1996-53-S.pdf>.
- [11] Platt J. A resource-allocating network for function interpolation. *Neural Computation*, 1991,3:213~225.
- [12] Kadirkamanathan V, Niranjan M. A function estimation approach to sequential learning with neural networks. *Neural Computation*, 1993,5(6):954~975.
- [13] Molina C, Niranjan M. Pruning with replacement on limited resource allocating networks by F-projections. *Neural Computation*, 1996,8(4):855~868.
- [14] Miller DA, Zurada JM, Lilly JH. Pruning via dynamic adaptation of the forgetting rate in structural learning. *IEEE International Conference on Neural Networks*. Vol.1, 1996. 448~452.
- [15] Yingwei L, Sundararajan N, Saratchandran P. A sequential learning scheme for function approximation using minimal radial basis function neural networks. *Neural Computation*, 1997,9(2):461~478.
- [16] Sugiyama M, Ogawa H. Incremental projection learning for optimal generalization. *Neural Networks*, 2001,14(1):53~66.
- [17] Sugiyama M, Ogawa H. Properties of incremental projection learning. *Neural Networks*, 2001,14(1):67~78.
- [18] Frean M. The upstart algorithm: a method for constructing and training feedforward neural networks. *Neural Computation*, 1990,2(2):198~209.
- [19] Fahlman S, Lebiere C. The cascade-correlation learning architecture. Technical Report, CMU-CS-90-100, Pittsburgh, PA: Carnegie Mellon University, 1991.
- [20] Mohraz K, Protzel P. FlexNet: a flexible neural network construction algorithm. In: *Proceedings of the ESANN 1996*. 1996. 111~116. <http://www.dice.ucl.ac.be/Proceedings/esann/esannpdf/es1996-14-S.pdf>.
- [21] Zhou ZH, Chen ZQ, Chen SF. Research of field theory based adaptive resonance neural network. *Journal of Software*, 2000, 11(11):1451~1459 (in Chinese with English abstract).
- [22] Zhang L, Zhang B. A geometrical representation of M-P neural model and its applications. *Journal of Software*, 1998,9(5):334~338 (in Chinese with English Abstract).
- [23] Wang JF, Cao YD. The application of support vector machine in classifying large number of catalogs. *Journal of Beijing Institute of Technology*, 2001,21(2):225~228 (in Chinese with English Abstract).
- [24] Mulier F. Vapnik-Chervonenkis (VC) learning theory and its applications. *IEEE Transactions on Neural Networks*, 1999,10(5): 985~987.
- [25] Vapnik VN, Zhang XG, Trans. *Statistical Learning Theory*. Beijing: Tsinghua University Press, 1998 (in Chinese).

附中文参考文献:

- [1] 张铃,吴朝福,张钺,韩枚.多层前馈神经网络的学习和综合算法. *软件学报*, 1995,6(7):440~448.
- [2] 张铃,张钺.多层前向网络的交叉覆盖设计算法. *软件学报*, 1999,10(7):737~742.
- [22] 张铃,张钺.M-P 神经元模型的几何意义及其应用. *软件学报*, 1998,9(5):334~338.
- [23] 王建芬,曹元大.支持向量机在大类别数分类中的应用. *北京理工大学学报*, 2001,21(2):225~228.
- [25] Vapnik VN, 著.张学工,译. *统计学习理论的本质*.北京:清华大学出版社,2000.