

# 人工神经网络介绍

李 硕

同济大学软件学院, 上海

201804

**摘 要** 机器学习方法经常被应用到解决医学和生物信息学的问题。本文列举了一些把机器学习方法应用到生物信息学领域的实例。比如：组建多重神经网络，运用该神经网络对 4 种不同形势的肿瘤患者进行分类。

**关键词** 人工；神经网络；机器学习方法

**中图分类号** Q1

**文献标识码** A

**文章编号** 1674-6708 (2011) 40-0111-02

## 0 引言

机器学习方法经常被应用到解决医学和生物信息学的问题。在这个报告中我列举了一些把机器学习方法应用到生物信息学领域的实例。比如：组建多重神经网络，运用该神经网络对 4 种不同形势的肿瘤患者进行分类。

## 1 介绍

人工神经网络属于机器学习领域。关于人工神经网络的概念最早提出于 1940 年代。后来在 1980 年代后被推广应用，尤其是在医学领域。

其中一个非常有用的用途是对疾病进行分类，达到诊断的目的，或者对基因表达进行分类。在这类神经网络里面，k 点最近邻居算法是最常被采用的算法。

人工神经网络的优点是：不需要人们关注神经网络里面的细节信息；人工神经网络可以很容易地被重新训练来应对不同地分类数据。人工神经网络可以用来解决有监督学习和无监督学习，比如：自组织特征映射（self-organized feature map）就可以用来解决无监督学习的问题。

它的不足之处在于：人工神经网络往往需要大量的训练数据，而这些训练数据往往不是很容易获得。人工神经网络可以被看作是一个黑盒，它的细节隐藏在点点之间的权值里面。这些权值的意义是人类无法理解的。同时，人工神经网络需要被仔细的训练以避免过拟合的情况出现。我们常常需要降低高维数据的维度。下面，我将分析介绍人工神经网络的具体应用。

人工神经网络的结构如图 1 所示：

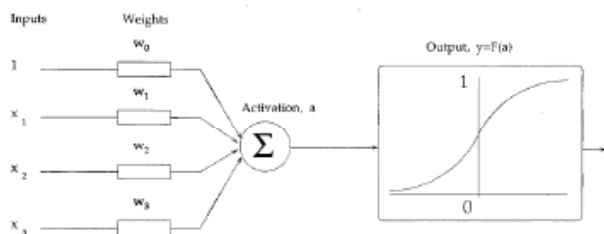


图 1

$x_1, x_2, x_3$  是该神经网络的输入值， $w_0, w_1, w_2, w_3$  是该神经网络的输入结点到内部节点的路径权值，每个神经元的激活函数是如上图右侧所示的函数图像。

这个函数被称之为 sigmoid 函数，表达式如下：

$$F(a) = \frac{1}{1 + \exp(-a)}$$

多重神经网络通常有 3 层，事实上，3 层神经网络以能进行很好的分类效果。这三个层包括输入层，隐藏层，输出层。在每个神经元内部我们可以选择 sigmoid 激活函数或其他种类的激活函数。

如图 2 所示：

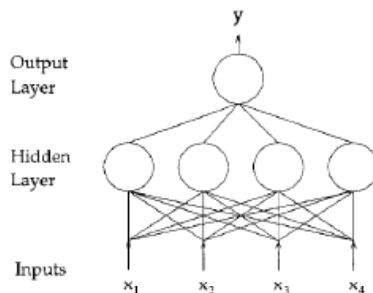


FIGURE 2. A multilayer perceptron. This is a two-layer perceptron with four inputs, four hidden units, and one output unit.

图 2

单个神经元仅提供线性的分割面，所以多层神经网络可以提供非线性的分类函数（即：若干个线性分割面的复杂组合）。这并不意味着 4 层神经网络就一定比 3 层神经网络能一共更好的分类效果，因为层数越多，需要的训练集就越庞大，得到的效果也不会提高。

既然有训练问题，就会涉及到训练算法。较为早的和著名的训练算法是 delta 规则。它于 20 世纪 60 年代被提出。它的原理是计算理论输出值和世纪输出值的均方差。tp 为理论输出值，yp 为实际输出值，表示为：

$$MSE = \frac{1}{2} \sum_p (t_p - y_p)^2$$

训练的开始阶段，我们通常设定一个随机选取值，令该值等于：

$$\Delta w_j = \alpha (t_p - y_p) x_j y_p (1 - y_p)$$

该公式里， $\alpha$  是学习速率，学习速率越大，学习的过程就越快，完成学习的时间短。但如果学习的速率过大，可能导致网络的理想权值在合理结果的附近游摆而永远无法获得理想的权值。

神经网络被训练好了以后，它就被用到解决目标问题。原始的数据集可以被分为两部分：一部分用来训练，一部分用来测试。

有时候神经网络会把训练数据集里面的噪音点的特征纳入自己的权值表达里，从而该神经网络无法真正体现该点集的真实特征。我们把这种情况叫做过拟合。过拟合是由于网络比待估函数复杂造成的。比如一个可以同 3 层网络解决的问题，我们用 4 层网络或者由更多神经元的三层网络去解决该问题，就容易造成过拟合。为了更好的明确训练时所采用的神经网络的隐藏层的层数，Livingstone 和 Manalack 提出了如下计算公式：

$$D = m \cdot o / w$$

该公式里 m 是训练样本的数目，o 是该网络的输出值，w 是网络权值的数目，D 就是隐藏层的数目。

得到了隐藏层的数目之后，我们可以以这个数目创建神经网络，边训练边削减，直到我们获得一个一半化的网络。对于没有隐藏网络层或只有一个隐藏网络层的神经网络，我们需要先确定它要解决的问题是否是线性的。

适当的训练方案是能也可以使网络的复杂性和数据的复杂性

（下转第175页）

口与VBE接口装置通信上传本VBE控制机箱内部监视信息,通过ST光纤接口接收VBE接口装置转发的PCP系统发出的BTPD信号(一路),通过ST光纤接口接收VBE接口装置转发的PCP系统发出的RELEASEOK信号(一路),内部控制信号测试端口;

6) LE插件提供以下功能:接收MC板的触发信号,两块LE板互为冗余,将触发信号由电信号转换为光信号,发送到TE板上,控制TE板的工作,16路光信号输出;

7) LR插件提供以下功能:接收16路来自TE板的状态回检光信号,将光信号转换为电信号,通过背板总线将状态回检信号发送到MC板,判断负向回检信号,向MC发送负向电压建立信号,判断 $du/dt$ 区间回检信号数量,向MC板发送 $du/dt$ 信号,LR板内部系统1和系统2互为冗余。

## 2 可控硅监控(TM)系统

每极有1面可控硅监控TM200控制柜,TM柜内安装监测控制装置,实现汇总3个VBE控制机柜的换流阀监测信息、通过110V DC 开入开出接口及光纤接口与控制保护系统实时交换信息、输出报警和跳闸信号、提供就地详细信息显示、调试串口输出等功能。

TM200控制柜主要由1个TM插件箱、1个监控工控机及开关电源、电源断路器和通讯网关组成。

### 2.1 TM 监控系统

VBE200监控系统具有实时报警、数据查询、图形显示等功能,使用方便快捷。

TM插件箱结构与功能简介:

1) 系统冗余设计,分为A、B两个系统;

2) A、B系统独立光纤接口对应于PCP系统直接连接,接收来自PCP系统的4个信号:RELOK 触发脉冲释放、Alpha=90 触发角接近90°、UnderVoltage 低电压信号、Select 通道选择;

3) 对VBE柜通信接口:提供3路串行光纤通信接口对应于3个VBE柜(每路包含2发1收共3根光纤),提供VBE系统主动被动信号选择光纤输出(1根输出光纤),提供VBE系统RELEASEOK信号选择光纤输出(1根输出光纤);

4) 对监视系统通信接口:提供1路RS232通信接口、提供MODBUS通信协议、可向监视系统传送详细报警位置信息、通过监视系统通信设定,可设定当前TM工作于测试状态,将VBE上传信息进行处理。

5) 调试串口:提供1路RS232通信接口、内部调试使用,可监视内部RAM等实时运行信息、对PCP系统DIO接口、设计6路光电隔离110VDC DI输入通道(目前使用2路,1路用于接收PCP发出的ST信号,1路用于接收PCP发出的BYPSS信号)。

## 3 VBE系统与PCP系统接口

葛南直流控制保护系统采用南瑞公司MACH2系统,基于不同技术的VBE、TM的设计差别较大,因此它与控制系统间交换信号的差别也较大。换流站控制直流保护系统与VBE、TM系统基于同一技术时,生产厂家不需做额外的接口工作即可实现两者的连接;而基于不同技术时,生产厂家则需做额外的接口工作。考虑到换流站VBE、TM系统是跟随阀进行生产的,与一次系统连接紧密,是信号的发送方,故此葛南直流改造的VBE、TM系统接口需要在系统接口设计稳定可靠的前提下,满足PCP系统的要求,保持与现有系统的兼容及一致性,提供方便可靠的VBE系统接入方案。接口方案如下:

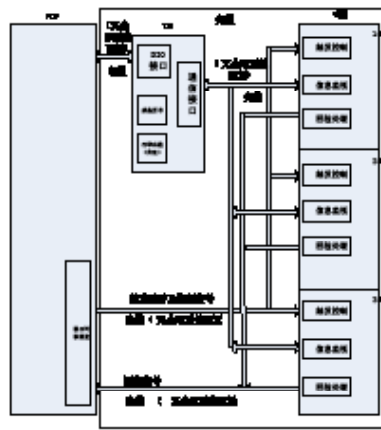


图1 PCP系统与VBE系统连接示意图

1) TM控制柜布置于主控室,完成以下功能:(1)接收VBE检测信息;(2)汇总产生报警跳闸信息并提供与PCP系统的DIO电气连接;以DIO电气连接接收PCP系统控制指令;(3)提供就地液晶屏显示详细报警信息,并可提供打印功能;(4)提供详细报警信息就地查看串行端口。

2) 3个VBE控制柜对应与各相阀塔布置于阀厅,完成以下功能:

(1)接收PCP的触发脉冲及触发控制信号,转换为对TE板的控制信号,完成对本相阀塔的触发控制;

(2)向PCP系统返回阀控的回报信号(讨论待定);

(3)检测本相阀塔TE板回报信号;

(4)检测本相阀塔漏水检测信号及避雷器动作信号;

(5)接收TM柜发来的信息,并将检测到的各相关信息高速可靠的发送至TM柜。

3)为兼容现有系统,TM控制柜与PCP系统之间接口形式保持现有设计,即:

(1)TM柜以DIO电缆形式向PCP系统上报VBE系统的报警及跳闸信息;

(2)TM柜以DIO电缆形式接收PCP系统发出的模拟试验状态等控制信号;

(3)TM柜以光纤接口形式接收PCP系统发出的阀控信号;

(4)VBE控制柜与PCP系统之间采用光缆连接,针对现有系统该部分接口连接不满足系统设计双重化冗余配置的要求的状况,建议由PCP系统提供冗余的触发脉冲输出、控制信号输出、阀控回检输入等信号的光纤接口。

## 4 结论

许继柔性直流公司研制的可控硅监视(TM)及阀基电子设备(VBE),这是国内直流换流站首次用上国产VBE系统,2010年4月投运以来运行正常,这是直流工程国产化的又一重要进程。

### 参考文献

[1]赵晓君.高压直流输电工程技术[M].北京:中国电力出版社,2002.

[2]王安全,等.±500kV葛南直流输电系统南桥换流站运行规程.上海:上海市电力公司超高压输变电公司,2009.

[3]胡四全,等.葛南工程阀基电子设备(VBE).许昌:许继柔性直流公司,2009.

↑↑(上接第111页)↑↑

得到合适的匹配。一个合适的训练方案应该是如下步骤:首先选择一个很大的网络并且把它的每个权值都设到一个很小的值上。通过训练,这些权值可以逐渐游摆到一个合理的值。

由于初始数据集通常要被分为训练集和测试集。在医学领域,我们能获得的数据集往往很小,比如某种病的病人数目不会很大。所以我们需要采用交叉验证的技巧来是较小的数据集在被分为训

练集和测试集之后能较好的训练神经网络。

### 参考文献

[1]Machine learning in bioinformatics: A brief survey and recommendations for practitioners.Harish Bhaskar, David C.Hoyle, Sameer Singh.