

文章编号: 1001-0920(2007)09-1022-05

量子神经元结构设计及其应用

吕 强, 俞金寿

(华东理工大学 自动化研究所, 上海 200237)

摘 要: 在量子理论的框架内, 解释了神经元的信息处理机制, 提出了量子神经元. 该神经元对信息的处理分为两阶段. 第 1 阶段为宏观信息收集部分, 产生控制量子比特; 第 2 阶段为微观信息处理部分, 根据控制量子比特, 改变神经元的状态. 整个过程模拟量子受控非门. 采用人工和实际数据集, 作为分类研究对象, 对比传统的神经网络, 量子神经网络显示出较好的分类效果. 以丙烯腈反应器作为建模研究对象, 该网络显示出较强的泛化能力.

关键词: 神经网络; 量子神经计算; 量子神经元; 分类; 建模

中图分类号: TP18 文献标识码: A

Structure design of quantum neural unit and its application

LV Qiang, YU Jin-shou

(Research Institute of Automation, East China University of Science and Technology, Shanghai 200237, China.

Correspondent: LV Qiang, E-mail: luqiang77@yahoo.com.cn)

Abstract: In the framework of quantum theory, the process of the neural unit dealing with information is explained, and quantum neural unit is proposed. The process includes two stages. The first stage is collecting macroscopical information, in which control quantum bit is produced. The second stage is dealing microcosmic information that neural unit state is changed according to control quantum bit. The whole process simulates the quantum controlled NOT gate. The synthetic datasets and real datasets are used as classification objects. Classification results of the quantum neural network are better than that of the classical neural network. Additionally, acrylonitrile reactor is used as modeling object. The proposed network shows the better performance.

Key words: Neural network; Quantum neural computing; Quantum neural unit; Classification; Modeling

1 引 言

神经网络作为人工智能的重要组成部分, 在近 20 年里已经吸引了大量专家学者的目光. 然而, 随着应用的深入推广和实际问题的日益涌现, 神经计算的局限和不足也逐渐显现出来. 为了克服神经计算的局限和不足, 学者们希望能够有新的理论与神经计算相结合, 在本质上改善神经计算的计算性能. 由此, 量子神经计算便发展起来^[1,2].

大量的学者对量子神经计算进行了深入的理论研究. 但在实际应用中, 即在使用经典计算机的情况下, 遇到很多困难, 例如: Narayanan^[3] 提出的量子神经网络只适用于二进制输入数据的情况; Ventura^[4] 提出的量子联想记忆在经典计算机上很难模拟, 只能在量子计算机上运行等. Kouda^[5,6] 对此作出了重要贡献, 他采用模拟一位量子相移门和两位量子受控非门, 提出了基于量子门的神经网络.

该网络用复数表示量子状态, 通过对量子门的模拟, 显示出较好的学习能力. 但是, 学习算法较为复杂, 特别是在使用梯度训练法时, 对复数反正切求导的计算较为繁琐, 对于一些分类问题的网络测试正确率不高, 并且网络训练时间较长.

本文在 Kouda 研究的基础上, 提出了新的量子神经元的计算模型, 在量子理论的框架内, 解释了神经元的的信息处理机制. 该模型组成的三层前馈神经网络对比 Kouda 提出的神经元模型网络以及经典的神经元模型网络, 基准测试的结果显示出该网络具有较好的学习能力, 训练时间短, 分类精度高. 之后, 采用该神经元组成的网络对丙烯腈反应器建模来预测丙烯腈的收率. 实验结果显示, 采用量子神经元的网络具有较强的泛化能力.

2 三层前馈量子神经网络计算模型

2.1 量子神经元的理论基础

收稿日期: 2006-06-12; 修回日期: 2006-10-08.

作者简介: 吕强(1977—), 男, 辽宁抚顺人, 博士生, 从事数据挖掘技术等研究; 俞金寿(1939—), 男, 浙江海宁人, 教授, 博士生导师, 从事化工过程控制、过程建模与控制等研究.

从量子物理学的角度考虑,神经元具有激活和非激活两个基本态,分别用 $|1\rangle$ 和 $|0\rangle$ 表示.神经元的状态在没有测量的情况下,可以处在激活态和非激活态之外,即可以是这两个基本态的线性组合,常被称为叠加态.此时,神经元的状态为 $|\psi\rangle$,可用下式表示:

$$|\psi\rangle = |\psi_0\rangle + |\psi_1\rangle. \quad (1)$$

此外,神经元的状态可用几何方式表示,即

$$|\psi\rangle = e^{i\phi} (\cos \theta |\psi_0\rangle + \sin \theta |\psi_1\rangle). \quad (2)$$

因为 $e^{i\phi}$ 是全局相位因子,不具有观测效应,可以省略.为方便研究,本文假设 $\phi = 0$,因此神经元的状态可以简写为

$$|\psi\rangle = \cos \theta |\psi_0\rangle + \sin \theta |\psi_1\rangle. \quad (3)$$

神经元的状态是可以改变的,直观上可理解为:神经元受到刺激,然后响应.受到的刺激是其他邻近神经元的状态,然后根据接受到的状态,改变自己的状态,并表达出来.有一个著名的通用性结果^[7]:任意的多量子比特门都可由受控非门和单量子比特门复合而成,某种意义上说,受控非门和单量子比特门是所有其他门的原型.正是由于受控非门的特点,本文提出的量子神经元模型主要模拟受控非门,通过神经元接受到的控制量子比特来改变其内部的状态(工作量子比特).

2.2 量子神经元的拓扑结构及工作机理

本文将量子神经元对信息的处理分为宏观信息收集和微观信息处理两部分,如图 1 所示.宏观信息收集部分的主要功能是:收集其他相连神经元的的信息,然后综合信息,同时转换成控制量子比特.微观信息处理部分的主要功能是:模拟受控非门,根据控制量子比特来改变工作量子比特,即神经元的内部状态.然后对神经元最新状态测量,将神经元激活态的概率值,作为神经元的宏观信息传递出去.

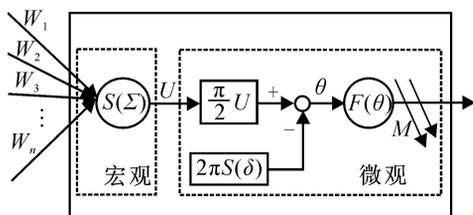


图 1 量子神经元

图 1 所示的量子神经元模型工作机理解释如下:

假定,如果神经元的状态为 $|\psi\rangle$ (用式(3)表示),此时的相角为 $2 S(\cdot)$ (S 为 sigmoidal 函数,为调节参数),具体如下:

$$|\psi\rangle = \cos(2 S(\cdot)) |\psi_0\rangle + \sin(2 S(\cdot)) |\psi_1\rangle. \quad (4)$$

如果控制量子比特为 $|\psi_0\rangle$ ($U = 0$),按照图 1 的计算方式,神经元状态的相角变为 $-2 S(\cdot)$,此时神经元的状态为

$$|\psi_1\rangle = \cos(2 S(\cdot)) |\psi_0\rangle - \sin(2 S(\cdot)) |\psi_1\rangle. \quad (5)$$

可以看出,神经元状态与其初态具有相同的可观测性.从测量的角度看,神经元的状态没有改变,如果控制量子比特为 $|\psi_1\rangle$ ($U = 1$),则神经元的相角变为 $\frac{\pi}{2} - 2 S(\cdot)$,神经元的状态为

$$|\psi_2\rangle = \cos(2 S(\cdot)) |\psi_1\rangle + \sin(2 S(\cdot)) |\psi_0\rangle, \quad (6)$$

此时神经元的状态被翻转.

如果控制量子比特 $|\psi\rangle = \cos(\frac{1}{2}U) |\psi_0\rangle + \sin(\frac{1}{2}U) |\psi_1\rangle$ ($0 < U < 1$),神经元状态用式(4)表示,按受控非门的工作原理,所得神经元为激活态时的概率为

$$\begin{aligned} & \cos^2(\frac{1}{2}U^2) \sin^2(2 S(\cdot)^2) + \\ & \sin^2(\frac{1}{2}U^2) \cos^2(2 S(\cdot)^2). \end{aligned} \quad (7)$$

按图 1 工作方式所得的神经元为激活态时的概率为

$$\begin{aligned} & \cos^2(\frac{1}{2}U^2) \sin^2(2 S(\cdot)^2) + \\ & \sin^2(\frac{1}{2}U^2) \cos^2(2 S(\cdot)^2) - \\ & 2(\sin \frac{1}{2}U) \cos(2 S(\cdot)) \cos(\frac{1}{2}U) \sin(2 S(\cdot)). \end{aligned} \quad (8)$$

很显然,在 $U = 1$ 和 $U = 0$ 时,图 1 完全模拟受控非门,而在 $0 < U < 1$ 时,图 1 仅仅是近似模拟受控非门,而不是完全模拟受控非门,两者相差

$$- 2 \sin(\frac{1}{2}U) \cos(2 S(\cdot)) \cos(\frac{1}{2}U) \sin(2 S(\cdot)). \quad (9)$$

式(9)被认为是在使量子神经元状态完全翻转的惩罚项.图 1 中函数 $F(\cdot)$ 根据神经元的相角得到神经元最新的状态,本文采用文献[5]的复数表示方法.

$$F(\cdot) = e^{i\theta} = \cos \theta + i \sin \theta. \quad (10)$$

式(10)显示出和式(3)的相似性,如果用“1”表示 $|\psi_0\rangle$,用“i”表示 $|\psi_1\rangle$,则用复数表示量子状态具有一定的合理性.图 1 中的 M 为测量,输出值为神经元状态为激活态时的概率,即 $\text{Im}^2(F(\cdot))$.

2.3 数学描述

本文主要介绍 3 层前馈量子神经网络,隐层采

用上一节提出的量子神经元,输出层神经元激活函数用于分类,采用 S 函数,建模采用线性函数.信息的处理过程(分类)如下所示:

$$O_k^{(p)} = S\left(\sum_{j=1}^m y_j^{(p)} w_{jk} - \theta_k\right), \quad (11)$$

$$y_j^{(p)} = \text{Im}^2(V_j^{(p)}), \quad (12)$$

$$V_j^{(p)} = F\left(\frac{1}{2} u_j^{(p)} - 2 S(\theta_j)\right), \quad (13)$$

$$u_i^{(p)} = S\left(\sum_{i=1}^n x_i^{(p)} w_{ij} - \theta_j\right). \quad (14)$$

其中: $S(\cdot)$ 表示 sigmoidal 函数, $x_i^{(p)}$ 为第 p 个样本的第 i 个部分输入, w_{ij} 为输入层和隐层的连接权值, θ_j 为隐层神经元的阈值, $V_j^{(p)}$ 为第 j 个神经元的状态, $y_j^{(p)}$ 为第 j 个神经元为激活态时的概率, w_{jk} 为隐层和输出层的连接权值, θ_k 为输出神经元的阈值, $O_k^{(p)}$ 为第 p 个样本的第 k 个部分输出.

3 网络性能基准测试

为了全面测试所提出的量子神经网络(QNN)的性能,本文采用两种数据集:人工数据集和实际数据集.测试结果显示 QNN 非常良好的分类性能.作为对比,本文使用了几个基于经典的神经元构成的网络和基于量子理论的神经网络(三层前馈,梯度训练法),分别是 BP^[8](指隐层神经元激活函数为 S 函数),RBF^[8](径向基函数),Qubit NN^[5].

在实验中(除双螺旋问题),训练停止条件是:测试样本总误差在连续 5 次迭代中都不再下降.实验的硬件是 AMD Sempron 1.79 GHz,512 MB,采用 Matlab language,在 Windows 环境下运行.

3.1 人工数据集

3.1.1 概念学习

概念学习是神经网络的一个基准测试问题^[9].下面是两个概念学习的例子,包括一个三角形和一个矩形.在 $(-8, 8) * (-8, 8)$ 的范围内随机产生 100 个样本,分为 3 类.在本次实验中,为了测试网络归纳学习的能力,采用递增训练集的方法^[10],训练样本数分别为 100, 250, 500 和 1 000,测试样本数为 5 000.由于隐节点的不确定性,根据 Wolensky^[11]实验的结果,对于采用梯度算法的网络,在本次试验中,隐节点的选取分别为 25, 30, 35, 对于每个隐节点均做 10 次实验,最终结果取其平均值,如表 1 所示.在本次实验中,为了控制网络训练时间,设定最大迭代数为 5 000.

从表 1 的结果可以看出,BP 神经网络随着样本数的增加,测试准确率是先上升后下降.RBF 神经网络可以取得较好的测试准确率,但是样本数从 500 到 1 000 时,测试准确率变化不大,而且网络训

表 1 概念学习测试结果

算法	训练集 样本数	训练集分类 正确率 / %	测试集分类 正确率 / %	训练 时间 / s
BP	100	98.766 7	89.253 3	157.168 8
	222.687 0	250	93.786 7	92.160 7
	296.288 6	500	98.373 3	96.543 3
	338.350 0	1 000	94.760 0	93.714 0
RBF	100	99.166 7	88.930 0	298.015 6
	336.508 4	250	98.520 0	94.774 7
	463.926 7	500	97.640 0	95.994 0
	547.112 2	1 000	98.200 0	95.868 0
Qubit NN	100	85.000	80.900 0	191.625 0
	389.359 0	250	84.200	84.540 0
	534.641 0	500	87.600	88.546 0
	834.756 0	1 000	88.256	90.352 0
QNN	100	100	93.820 7	102.382 8
	152.738 0	250	98.920 0	95.884 7
	207.423 5	500	98.800 0	97.004 0
	271.984 0	1 000	99.200 0	98.120 0

练时间较长.Qubit NN 并没有显现出较好的分类效果,反而是 4 类神经网络里最差的,而且网络训练时间最长.QNN 训练时间较短,且可取得较好的测试集分类正确率,随着训练样本数的增加,测试准确率也在不断上升,在 4 种类型的神经网络中,效果最好.图 2 显示出 4 种类型神经网络随样本数的增加,测试准确率变化趋势图.

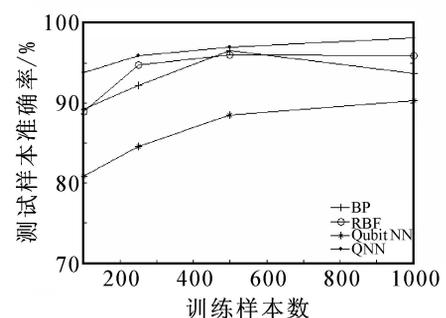


图 2 4 种类型神经网络随训练样本数增加测试准确率对比图

3.1.2 双螺旋问题

Wieland 提出了一个神经网络基准测试问题,就是如何分辨两个螺旋形.双螺旋取步长为 1,共采用 194 个数据为训练样本,取步长 0.1,得到 1 922 个数据为测试样本.采用隐节点递增的方式,分别为 20, 25, 30, 35, 40, 设定最大迭代次数 40 000 次和预设训练误差为网络停止训练的准则,没有采用约定的网络训练停止条件的原因是最大迭代次数很大,所以需要控制网络训练时间.平均结果如表 2 所示.

表 2 双螺旋结果

算法	隐节点数	训练集分类	测试集分类	训练时间/s
		正确率/%	正确率/%	
BP	20	74.23	73.05	87.344
	25	80.93	75.96	100.89
	30	82.47	76.8	114.313
	35	90.21	82.78	125.875
	40	86.08	76.64	147.516
RBF	20	74.23	72.53	224.656
	25	81.96	81.69	275.141
	30	87.63	84.70	318.984
	35	89.69	88.14	367.75
	40	90.21	89.93	414.329
Qubit NN	20	78.5	71.4	357.71
	25	80.2	81.21	420.35
	30	82.42	83.28	520.123
	35	83.55	85.62	620.542
	40	84.78	87.54	740.658
QNN	20	92.27	87.04	93.516
	25	96.39	92.82	214.797
	30	97.42	93.34	262.094
	35	99.48	95.06	292.765
	40	99.6	98.19	344.797

从表2可以看出,RBF网络随着隐节点的增加,训练准确率和测试准确率都在增加,但所需时间较长;BP网络和 Qubit NN 效果相对较差;本文提出的量子神经网络,在此 4 种类型的神经网络中效果较好,网络的训练时间可以接受.图 3 显示的是随着隐节点的增加,4 种类型神经网络测试准确率的变化.

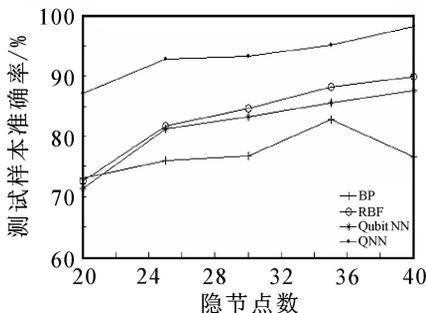


图 3 4 种类型神经网络随隐节点增加测试准确率对比图

3.2 实际数据集

3.2.1 Iris 分类

Iris 数据集分为 3 类,每个类由 50 个样本组成.网络测试时隐节点的选取同概念学习,训练样本数为每个类取 30 个样本,测试样本数为每个类剩下的

表 3 Iris 分类结果

算法	训练集分类	测试集分类	训练时间/s
	正确率/%	正确率/%	
BP	97.78	98.33	20.010
RBF	99.889	96.33	17.115
Qubit NN	98.2	95.00	25.684
QNN	98.89	98.33	12.906

20 个样本.

BP 神经网络和 QNN 都达到较好的测试集分类准确率 98.33%,但是 QNN 所需训练时间最短为 12.906 s,相比之下,RBF 训练时间比 BP 网络快,但测试集分类正确率却不是很高.Qubit NN 并没有表现出较好的分类性能,且训练时间较长.

3.2.2 Wine 识别

Wine 数据集分为 3 类,每个类分别包括 59,71,48 个样本.网络测试时隐节点的选取同上,训练样本每个类分别取 30,50,20,其他为测试样本.

表 4 Wine 分类结果

算法	训练集分类	测试集分类	训练时间/s
	正确率/%	正确率/%	
BP	100	98.72	7.700 5
RBF	100	97.44	10.369
Qubit NN	98.25	96.54	17.285
QNN	100	99.36	6.844 0

在 Wine 分类中,QNN 再一次表现出较好的分类效果和较短的训练时间,并且在隐节点为 30 和 35 时,QNN 的训练准确率和测试准确率在 10 次实验中均达到了 100%,相比之下,BP,RBF 和 Qubit NN 效果较差.

4 应用实例

上海石化某厂年产 5 万吨丙烯腈装置,采用美国标准石油公司 Sohio 的生产工艺,以 C-41 作催化剂,以丙烯、氨、空气为原料,在沸腾反应器中一次直接氧化制取丙烯腈(即丙烯氨氧化法).

通过工艺分析,本文将反应压力、中段温度、纯丙烯量、空比、氨比、反应线速、触媒量作为模型的辅助变量.采用此厂提供的反应器数据 342 组进行建模,并对原始数据进行数据处理(3 准则、七点线性平滑、归一化)后得到 317 组数据.将数据前 253 组组成训练样本集,另外 64 组数据组成测试样本集.采用 3 个模型泛化能力的衡量指标,即:1)测试样本的泛化均方误差(GMSE)值;2)测试样本的泛化最大绝对误差(MAXE)值;3)测试样本的泛化误差绝对值(|e|).

分别采用 BP,RBF,Qubit NN 和本文提出的量

子神经网络对反应器建模,最大训练次数 50 000,网络结构为 7-25-1.表 5 显示出 4 种模型泛化结果,图 4 是这 4 种模型收率预测值.可以看出,QNN 能够很好地跟踪丙烯腈收率的变化.

表 5 模型泛化结果对比

模型	泛化误差	泛化最大误差	误差绝对值
	GMSE	MAXE	$ e _1$
BP	0.649 0	2.924 2	54
RBF	0.940 6	4.047 0	53
Qubit NN	0.776 3	2.339 8	45
QNN	0.309 3	2.048 0	59

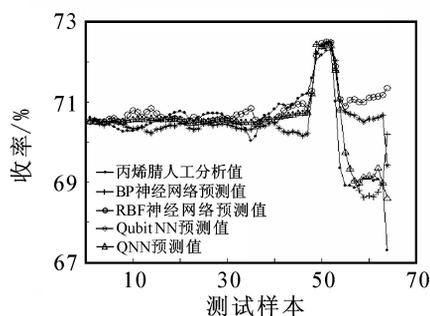


图 4 4 种网络收率预测值

5 结 语

本文在量子理论的框架下,解释了神经元处理信息的机制,提出了量子神经元计算模型.采用人工数据集和实际数据集作为分类对象,实验结果显示该神经元组成的网络与其他经典神经元组成的网络相比,在相同的网络结构和训练算法下,具有较好的分类结果和较短的训练时间.以丙烯腈数据作为建模对象,仿真实验也显示出该神经元网络具有较强的泛化能力,能够很好地跟踪丙烯腈的变化.进一步的研究主要有:1) 量子神经元是模拟量子受控非门的,这是它的理论基础,希望能够在数学上进一步证明这种模拟的科学性.2) 拓展量子神经元的应用领

域.

参考文献(References)

- [1] Kak S. On quantum neural computation[J]. Information Sciences, 1995, 83(3/4): 143-160.
- [2] Perus M. Neuro-quantum parallelism in brain-mind and computers[J]. Informatica, 1996, 20(2): 173-183.
- [3] Narayanan A, Menneer T. Quantum artificial neural network architectures and components [J]. Information Sciences, 2000, 128(3/4): 231-255.
- [4] Ventura D, Martinez T. Quantum associative memory [J]. Information Sciences, 2000, 124(1/4): 273-296.
- [5] Kouda N, Matsui N, Nishimura H. Qubit neural network and its learning efficiency [J]. Neural Computation and Application, 2005, 14(8): 114-121.
- [6] Kouda N, Matsui N, Nishimura H. Image compression by layered quantum neural networks [J]. Neural Processing Letters, 2002, 16(1): 67-80.
- [7] Nielsen M A, Chuang I L. Quantum computing and quantum information[M]. Beijing: Tsinghua University Press, 2004.
- [8] 魏海坤. 神经网络结构设计的理论与方法[M]. 北京: 国防工业出版社, 2005.
(Wei Hai-kun. Theory and method for structure design of neural network[M]. Beijing: Defense Industry Press, 2005.)
- [9] Cohn D, Atlas L, Ladner R. Improving generalization with active learning [J]. Machine Learning, 1994, 15(2): 201-221.
- [10] Zhou Z H, Chen S F, Chen Z Q. FANNC: A fast adaptive neural network classifier[J]. Knowledge and Information Systems, 2000, 2(1): 115-129.
- [11] Wolensky G. Analysis of neural network issues: Scaling, enhanced nodal processing, comparison with standard classification [J]. DARPA Neural Network Program Review, 1990, 14(10): 29-30.

(上接第 1021 页)

- [7] 张燕平,张铃,吴涛. 不同粒度世界的描述法——商空间法[J]. 计算机学报, 2004, 27(3): 328-333.
(Zhang Yan-ping, Zhang Ling, Wu Tao. The representation of different granular worlds: A Quotient Space[J]. Chinese J of Computers, 2004, 27(3): 328-333.)
- [8] 李道国,苗夺谦,张东星,等. 粒度计算研究综述[J]. 计算机科学, 2005, 32(9): 1-11.
(Li Dao-guo, Miao Duo-qian, Zhang Dong-xing, et al. An overview of granular computing [J]. Computer

Science, 2005, 32(9): 1-11.)

- [9] Lashin E F, Kozae A M, Khadra A A, et al. Rough set theory for topological spaces[J]. Int J of Approximate Reasoning, 2005, 40(1/2): 35-43.
- [10] 蒙祖强,蔡自兴. 个性化决策规则的发现:一种基于 Rough Set 的方法[J]. 控制与决策, 2004, 19(9): 994-998.
(Meng Zu-qiang, Cai Zi-xing. Discovery of personalized decision rule: A rough set-based approach [J]. Control and Decision, 2004, 19(9): 994-998.)