

脉冲神经网络中神经元突触的硬件实现方案^①

李宏伟, 吴庆祥

(福建师范大学 光电与信息工程学院, 福州 350007)

摘要: 脉冲神经网络被誉为第三代神经网络, 近年来受到许多学者的关注, 其优势已经在模式识别、计算机视觉等诸多领域得到了发挥. 脉冲神经网络的硬件化是实现其强大计算能力的重要途径, 而突触的硬件实现又是其中的一个关键性环节. 本文先从 SRM 模型中脉冲神经元突触的特性曲线入手, 用适合 FPGA 实现的差分方程来逼近这一响应曲线, 并对差分方程中的待定参数进行了优化, 然后根据差分方程, 在 Simulink 平台上设计出硬件电路, 并给出了在方波脉冲激励下, 电路输出的仿真结果, 最后对今后的工作做出了展望.

关键词: 脉冲神经网络; 突触; 响应曲线; Simulink; FPGA

Implementation of the Synapse of Spiking Neural Network in the Hardware

LI Hong-Wei, WU Qing-Xiang

(College of Photonic and Electronic Information Engineering, Fujian Normal University, Fuzhou 350007, China)

Abstract: Spiking neural network is considered as the third generation of neural networks, and it has attracted many researchers. Its advantages have been shown in pattern recognition and computer vision. The implementation of the spiking neural network in the hardware is an important method to show its powerful computation ability. This paper begins with the synapse response curve, and then in order to suit for the FPGA implementation the difference equation is used to approach the response curve. In addition, the genetic algorithm is used to optimize the parameters of the circuit. According to the difference equation, the circuit is designed in the Simulink platform. The simulation results are obtained for the outputs of circuit triggered by a square impulse wave. Finally the future work is discussed.

Key words: spiking neural network; synapse; response curve; Simulink; FPGA

1 引言

近几年来, 脉冲神经网络(SNN)正日益成为计算智能领域中的一个研究热点, 受到了广泛的关注, 并被称为第三代神经网络^[1,2]. 与以感知器为代表的第一代神经网络, 和常以 Sigmoid 函数为激活函数的第二代神经网络相比, 脉冲神经网络采用可塑性突触和基于脉冲序列的编码方式, 因此它更接近于现代脑科学对神经系统的认识^[3]. 目前, 脉冲神经网络已经被用于模式识别, 图像处理, 计算机视觉等领域^[4,5], 相对于传统的神经网络, 脉冲神经网络在很大程度上表现出了更强的计算能力.

在当前蓬勃发展的物联网产业中, 神经网络技术也得到了应用^[6]. 很多复杂的数据信息, 如视频与图

像信息等需要在现场进行实时处理, 此时若用软件实现 SNN 算法来处理数据, 不仅速度慢, 并行程度低, 难以满足对大数据量信息的实时处理要求, 而且软件实现需要体积较大的计算机支持^[7], 这样不但增加了成本, 也不利于设备的小型化. 这就使得将脉冲神经网络实现到大规模集成电路上, 变得十分必要. 突触是接收和处理神经元外来信息的部位, 对整个神经网络的正常运行起到至关重要的作用. 因此突触的硬件实现便有了十分重要的意义. 近年来, FPGA 器件有了飞速的发展, 在 FPGA 上很容易实现高速、并行计算^[8]. Matlab 中的 Simulink 组件给复杂的智能算法的硬件化提供了一个很便利的开发平台^[9], 在此平台上, 不仅可以快速准确的实现硬件的建模与仿真, 而且可以将

^① 基金项目:福建省产学研科技重大项目(2013H61010023)

收稿时间:2013-07-17;收到修改稿时间:2013-08-19

抽象的模块化电路图直接转化成硬件描述语言, 这便大大缩短了开发周期^[10], 但是这里也存在速度与芯片资源使用矛盾的难题^[11], 本文提出的响应函数的指数逼近方法用以实现脉冲神经元的突触, 不仅可以提高速度, 还可以节省芯片资源, 本文通过此方法在 Simulink 平台上建立了脉冲神经元突触的电路模型, 并用遗传算法对电路中的关键参数进行了优化.

2 脉冲神经网络模型

常用的脉冲神经网络模型有 IF(Integrated-and-Fire)模型和脉冲响应模型(Spike Response Model), 简称 SRM 模型, 本文采用的是 SRM 模型. 设一神经元 s 的阈值为 θ , 此神经元的状态变量为 u_s , 当 u_s 的值大于 θ 时, 神经元 s 被激活并发放一个脉冲, 发放的脉冲传送到下一级神经元的突触. 设状态变量 u_s 的值在 t_s^f 时刻超过神经元 s 的阈值 θ , 神经元被激发, 则 t_s^f 满足

$$t_s^f : u_s(t_s^f) = \theta, \left. \frac{du_s(t)}{dt} \right|_{t_s^f} > 0 \quad (1)$$

神经元的状态变量 u_s 由两个因素决定, 一是输入到神

经元的各个突触的响应之和, 二是神经元的不应期效应对神经元的抑制作用. 当一神经元被激活并发放脉冲后, 马上进入不应期, 在不应期内, 神经元对外来刺激很难做出响应. 在建模时通常引入一个负的惯性因子 $\eta(t)$ 来描述不应期. 本文重点描述突触电路的设计, 因此不考虑不应期效应. 一个脉冲神经元与许多突触相连, 不同的突触对应着不同的延迟时间和权重值, 当一个外来输入的脉冲在特定时刻输入到突触, 突触会对此脉冲做出响应, 突触的响应函数一般由

$$y_i^k(t) = \varepsilon(t - t_i^k) = \frac{t - t_i^k}{\tau} e^{-\frac{t - t_i^k}{\tau}} \quad (2)$$

来描述. τ 为时间常数, 视具体情况而定, t_i^k 为从外部激励脉冲产生到突触产生响应所经历的时间. 图 1 为一典型的两输入脉冲神经网络模型^[3], w 为突触的权重值, d 为突触的延迟时间, 当两个输入脉冲产生的响应之和 $\sum j$ 大于后一级神经元的触发阈值时, 后一级神经元便会发放脉冲. 由此图可知, 将此网络硬件化的关键环节在于突触响应函数的实现.

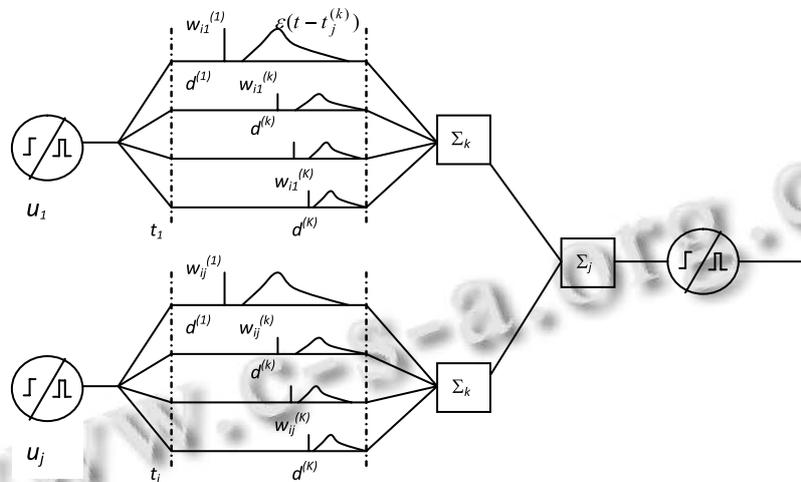


图 1 典型的两输入脉冲神经网络模型

3 突触的硬件实现方案

这里将(2)式时间常数定为 25, 延迟时间 t_i^k 定为 0, 则(2)式的简化形式为

$$f(t) = \frac{t}{25} e^{-\frac{t}{25}} \quad (3)$$

其对应的曲线为图 2 所示.

3.1 突触响应曲线的近似方程的建立

由于数字电路善于实现指数函数的线性运算, 考

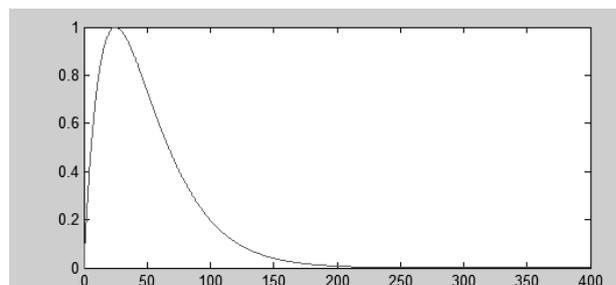


图 2 简化的突触响应曲线

考虑到直接按照(3)式的表达形式来构建的电路较为复杂,因此这里用一种更适合电路实现的差分方程来逼近(3)式.由图2所示,可以将曲线分解成上升和下降两个阶段,在时间轴上取一系列离散的时间点,其中第 n 个时间点对应的电路输出值为 $y(n)$, n 自突触对脉冲产生响应开始计起.在上升阶段,函数的二阶导数小于0,也就是增幅逐渐减小,所以用差分方程

$$y_1(n+1) = ay_1(n) + x_0 \quad (4)$$

来描述此上升阶段曲线,其中 a 为待定系数,且 $0 < a < 1$,表示增幅逐渐减小, x_0 为一重要的待求常数,在构建完整的脉冲神经网络时, x_0 便为这一突触的权重值.利用 z 变换理论,将(4)式的递推式差分方程转化为封闭式方程,为

$$y_1(n) = \frac{x_0}{1-a}(1-a^n) \quad (5)$$

其对应的连续方程为

$$y_1(x) = \frac{x_0}{1-a}(1-a^x) \quad (6)$$

在曲线的下降阶段,其形状类似于

$$y=c^x \quad (7)$$

的图像,其中 $0 < c < 1$,但是这样的简单指数函数的曲线收敛速度比原曲线快很多,而且在整条曲线的驻点处缺乏有效的过度.基于这一点,此处采用

$$y_2(n) = p \cdot b^n - q \cdot c^n \quad (8)$$

来逼近下降阶段的曲线,此处的 n 从曲线到达最高点开始记,即

$$y(0)=1 \quad (9)$$

所以

$$p+q=1 \quad (10)$$

由 z 变换理论,(8)式对应的递推式差分方程为

$$m \cdot b + cy(n) = y(n+1) \quad (11)$$

其中

$$m = (b-c) \cdot p \quad (12)$$

因为在曲线的上升阶段,已经涉及到了一个指数运算 a^n ,为了使电路结构尽量简单,不再引入过多的参数,这里令 $c=a$.

3.2 电路中待求参数的求解

由于数字电路的离散性,要求 $y_1(1), y_1(2) \cdots y_1(n)$ 这些点的总误差最小,因此在上升阶段将目标函数设定为

$$D(n) = \sum_{n=0}^{25} |y_1(n) - f(n)| = \sum_{n=0}^{25} \left| \frac{x_0}{1-a}(1-a^n) - \frac{n}{25} e^{(1-\frac{n}{25})} \right| \quad (13)$$

其中 $f(n)$ 为突触响应曲线方程(3)式的离散形式.对上升阶段的逼近过程就是使 $D(n)$ 取得最小值的过程,在 $D(n)$ 取得最小值时, a 与 x_0 的取值便为电路所需要的最优参数.对这两个参数的求解采用接力式进化的遗传算法^[12],所谓接力式指的是遗传的总代数设置为100代,先进行50代的进化,并将这一轮的进化结果作为第二轮50代进化的初始种群,第二轮进化后得到的结果便为最终结果,即 $a=0.8944, x_0=0.1130$.如图3所示,实线为式(3)对应的曲线的上升部分,虚线为实际逼近所得到的曲线.

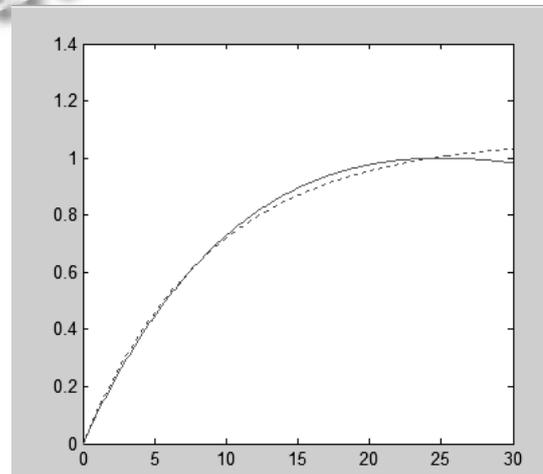


图3 上升阶段理论曲线与实际逼近的曲线的对照

这里结合图4来表述曲线下降阶段的参数求解,在曲线上升阶段结束的时候,根据(5)式和电路结构可以求得加法器AddSub1的输出为0.1086,令 $m=0.1086$;将(8),(10),(11),(12)联立可得:

$$y_2(n) = \frac{0.1086}{b-0.8944} \cdot b^n - \left(\frac{0.1086}{b-0.8944} - 1 \right) \cdot 0.8944^n \quad (14)$$

定性分析可知, b 的值一定在(0.8944, 1)区间上,因为如果 b 太小,曲线收敛速度太快,如果 b 大于1,则曲线不收敛.考虑到在实际工程中所允许的误差,这里应用二分法的思想,对 b 的值进行手动的试探,将理论曲线和实际曲线在同一坐标系中观察,经过少数几次尝试后,把 b 的值确定在0.974,于是便得到了图中的参数.当 $b=0.974$ 时, $p=1.364$ 时, $q=0.364$.实际逼近的曲线方程为

$$y_2(n) = 1.364 \times 0.974^n - 0.364 \times 0.8944^n \quad (15)$$

图 4 中的虚线为式(15)对应的曲线(n 在曲线上升结束后开始计数), 实线为(3)式所对应的理论曲线.

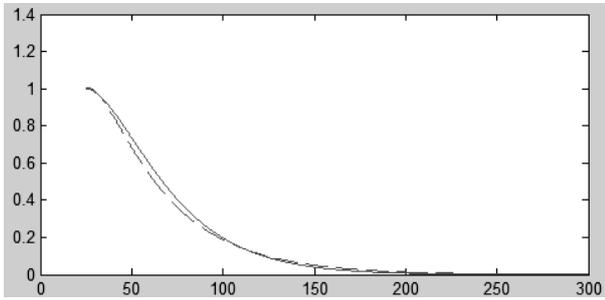


图 4 下降阶段理论曲线与实际逼近曲线的对照

3.3 硬件电路的设计

在 Matlab 中, 利用 Simulink 组件, 很容易将电路图转化成硬件描述语言, 经综合和适当的优化后便可下载到 FPGA 芯片中运行, 因此, 此处着重描述模块化的电路图的设计.

图 5 便为在 Simulink 平台上搭建的电路图. 用到的电路部件有延时器(Delay, Delay1), 带有使能端的寄存器(Register, Register1)加法器(AddSub, AddSub1), 乘法器(Mult, Mult1). 延时器上的标识表示延时器的

输入延时 n 个时钟周期后, 从输出端输出. 图中的寄存器, 当使能端 en 为有效电平时, d 端的输入数据, 经一个时钟周期后, 由 q 端输出, 图中的乘法器, 只需一个时钟周期便可完成多位的乘法操作. 加法器为组合电路, 本身不需要时钟的驱动, 因此未做延时. 电路中的三个输入的常数 Constant, Constant1 和 Constant2 便为上节中的 a, x₀ 和 b.

考虑到要在 t=25 时, 电路的输出达到最大值, 并且在此时, 电路中的信号通道要实现切换, 即实现电路的输出对理论曲线的分段逼近, 因此将 Delay1 以及 Delay 的延迟时间分别设置为 23 和 1. 当 Constant2 在电路中起作用的同时, Constant1 失去作用. 在实现对图 2 中曲线的上升阶段逼近时, Register 打开, Constant1 直接进入 Addsub. 此时, Constant, Mult1, Addsub 共同构成闭环系统, 实现了指数运算. 电路的输出遵循(5)式. 在对曲线下落阶段的逼近时, Register 关闭, Register1 打开, Constant2, Mult, Addsub1 构成一个闭环系统, Constant, Mult1, Addsub 也构成一个闭环系统, 这两个闭环系统分别实现两个指数运算, 此时电路的输出遵循(15)式.

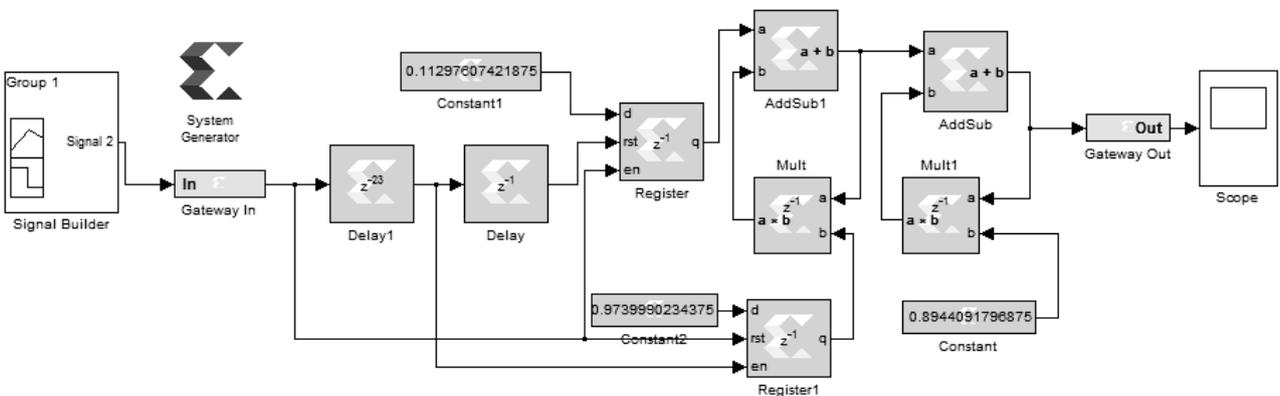


图 5 Simulink 环境下设计的突触电路图

4 仿真结果

在脉冲神经网络中, 神经元的触发信号为基于时间编码的脉冲信号脉冲序列. 这里为了测试突触的性能, 由图 5 中的信号发生器(Signal Builder)产生一矩形脉冲, 经由 Gateway in 模块输入到突触电路中, 电路的响应由 Gateway out 模块输出, 并显示在示波器(scope)上. 激励脉冲的持续时间为一个时间单位, 0 时刻激励信号产生, 1 时刻信号停止, 信号幅度为 1, 波

形如图 6 所示. 脉冲输入后, 便可在示波器(Scope)模块上观察到突触的响应曲线, 如图 7 所示, 曲线在上升阶段, 变化速度较快, 且二阶导数小于 0, 在下降阶段, 变化速度较慢, 且二阶导数大于 0, 波形与图 2 的理论波形相一致.

5 结论与展望

本文将脉冲神经元的突触这一脉冲神经网络的关键

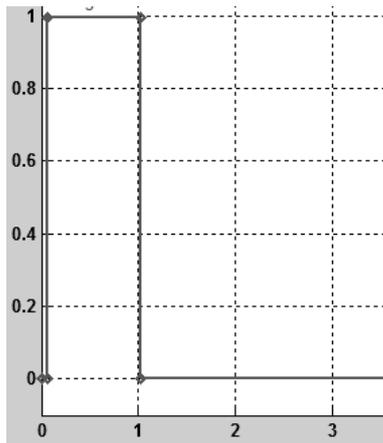


图 6 突触的输入脉冲波形

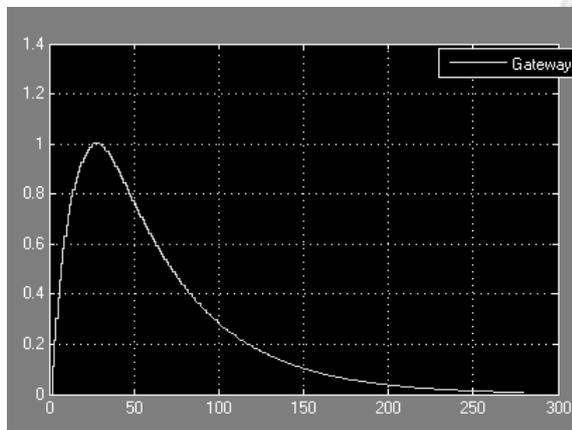


图 7 在方波脉冲激励下电路的输出波形

键性环节,用电路实现出来,并且得到一条较为理想的突触响应曲线.在设计过程中,没有直接用电路来实现突触的理论响应函数,而是用简单的指数函数对理论曲线进行了分段逼近,并且只引入了三个参数,这就大大的简化了电路结构,节省了 FPGA 中的逻辑单元,加快了硬件的处理速度.并且得到的电路输出结果与理论值之间的误差很小,在曲线的驻点处过渡连贯,完全可以满足工程上的需要.接下来的工作是在此突触的基础上,构建多个脉冲神经元,形成一个完整的脉冲神经网络,使其在 FPGA 上运行,用于对输入的图像与视频信号进行处理,并对动态目标进行

跟踪与识别,也可以用于其他需要智能部件的场合,比如智能传感器等.

参考文献

- 1 Mass W, Bishop CM. Pulsed Networks. MA: MIT Press, 2001:31-36.
- 2 Mass W. Networks of spiking neurons: the third generation of neural network models. *Neural Networks*, 1997, 10(9): 1659-1671.
- 3 Mass W. Fast Sigmoid networks via spiking neurons. *Neural Computation*, 1997, 9(2): 279-304.
- 4 Wu QX, McGinnity TM, Maguire L, Cai R, Chen M. A visual attention model based on hierarchical spiking neural networks. *Neurocomputing*, 2013, 116(SI): 116, 3-12.
- 5 Wu QX, McGinnity TM, Maguire LP, Belatreche A, Glackin B. Processing visual stimuli using hierarchical spiking neural networks. *Neurocomputing*, 2008, 71(10-12): 2055-2068.
- 6 冯秀芳,肖文炳.神经网络的数据分类算法在物联网中的应用. *计算机技术与发展*, 2012, 22(8): 245-248.
- 7 刘培龙.基于 FPGA 的神经网络硬件实现的研究与设计[硕士学位论文].成都:电子科技大学,2012.
- 8 Glackin B, Harkin J, McGinnity TM, Maguire LP, Wu QX. Emulating spiking neural networks for edge detection on FPGA hardware. *Proc. of The 19th International Conference on Field Programmable Logic and Applications*. Prague, IEEE, 2009. 670-673.
- 9 王正林. MATLAB/Simulink 与控制系统仿真.北京:电子工业出版社,2012:2-5.
- 10 谢方方,杨文飞,韩月霞.基于 System Generator 的快速视频跟踪系统设计. *计算机技术与发展*, 2013, 23(1): 221-224.
- 11 Maguire LP, McGinnity TM, Glackin B, Ghani A, Belatreche A, Harkin J. Challenges for large-scale implementations of spiking neural networks on FPGAs. *Neurocomputing*, 2007, 71(1-3): 13-29.
- 12 卓金武. Matlab 在数学建模中的应用.北京:北京航空航天大学出版社,2011:57-58.