

# 用蒙特卡罗法和模板匹配法识别 神经元放电脉冲间隔的优势模式

陈筱英

(中国科学院电子学研究所, 北京)

**摘要** 神经元放电脉冲序列的研究可能与神经中枢信息处理和传播机制的研究有关。在神经元放电脉冲序列间隔中存在着一些重复出现的脉冲片段, 称为优势模式。对优势模式的识别、检测是长期以来电生理学的一个重要研究领域。用量化蒙特卡罗法和模板匹配法能成功地从原始数据中识别出优势模式, 并统计出这些优势模式在神经元放电脉冲序列中出现的次数。本文给出了算法、程序的流程和结构, 并给出了蓝斑神经元优势模式的检测结果。

**关键词** 神经元放电; 优势模式; 蒙特卡罗法

## 一、引言

神经元放电一直是电生理学的一个重要方面。神经元的活动会在神经元的放电过程中表现出来, 而且不同的神经元的放电过程是不同的。关于神经元放电和信息传输有几种不同的观点。一种观点认为神经元放电活动是一种随机过程, 神经元的放电与上一次放电无关。神经元是通过放电脉冲间隔的变化来传递信息的。另一种观点认为脉冲间隔具有一定的相关性。根据目前的神经网络理论, 认为信息的传输是神经网络的整体功能。在60年代已经有许多电生理学家注意到大部份神经元放电具有典型的、可重复出现的脉冲片段。由全或无定理, 神经元放电脉冲具有相同的幅度。信息可能是由一定组合的脉冲间隔携带的, 并称这些可重复出现的具有一定间隔组合的脉冲片段为优势模式。这些优势模式可能代表着一定的信息编码并传递某种信息。美国宾州大学的 Dayhoff 教授在80年代初发表了几篇关于神经元放电优势模式识别的文章<sup>[1,2]</sup>。Dayhoff 的研究表明, 大部份神经元放电脉冲序列可以找到优势模式。在多种实验条件下, 30组淡水小龙虾的爪子控制系统的神经元放电脉冲中有23组能找到优势模式。一般优势模式由3—7个脉冲间隔组成。由国内一些电生理学家也积累了大量的神经元放电的数据资料, 从记录的资料(记录神经元放电脉冲的录音带、照片等)可以清楚地看到一些重复出现的脉冲片段。为了从电生理的理论来分析神经元传递信息的机制, 了解这些重复出现的脉冲片段的意义, 首先必须找到这些脉冲片段。在60年代, 分析神经元放电常用的方法是分析脉冲间隔直方图<sup>①</sup>。从直方图上可以看出各种脉冲间隔出现的频度。神经元处于不同状态, 兴奋或抑制时, 其放电间隔直方图将随之变化。神经元兴奋时, 放电频率增高。但这种

方法不能得到脉冲间隔的优势模式,因为直方图中出现频率高的脉冲间隔,不一定是脉冲间隔优势模式的一部份。但这种简单的直方图统计法也曾帮助电生理学家认识神经元的放电过程。70年代, Marczynski 对神经元放电的研究比简单的直方图统计法又前进了一步<sup>[4,5]</sup>。他把每一个脉冲间隔与前一个相比,如果间隔长度大于前一个间隔记为+, 小于则记为-。这样就得到一个+、-符号的序列,由此可粗略地找到一些重复出现的脉冲间隔模式。但这只是一种定性的模式,并不能给出模式中各个间隔的实际大小。直到80年代,由于计算机的大量使用和算法的进步,这项工作又有了新的发展。Dayhoff 提出了寻找优势模式的算法<sup>[1]</sup>,并作了计算机模拟和一些实际数据的分析。国内也已作了一定的工作<sup>[6]</sup>。本文在前人的基础上提高了计算速度;在使用蒙特卡罗法分析样本数据时作了更多的统计分析;此外还改进了模板匹配的方法。在神经元放电过程中,神经元兴奋状态变化时,脉冲优势模式的间隔也随之发生变化,但仍保持一定的模式,即优势模式中间隔的相对比例较稳定。程序采用模块结构,分析了神经元放电的实际数据。

## 二、方 法

### 1. 神经元放电脉冲的特点和数据采集

神经元放电脉冲的模板分布主要有以下三种情况,如图1所示。A. 两个模板相距足

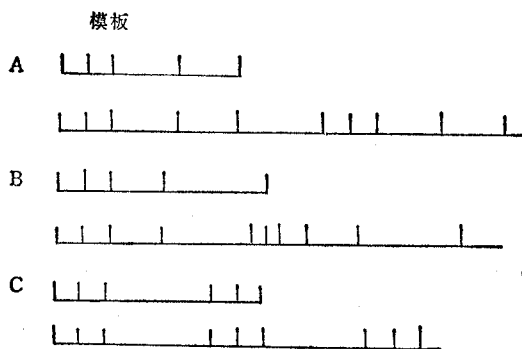


图1 神经元放电脉冲间隔的模板分布分类。

A. 两模板互不重叠; B. 插入重叠; C. 重合重叠。

够远,互不重叠; B. 模板插入重叠; C. 模板混合重叠。

在神经元放电脉冲序列中,除了这些组成优势模式的脉冲外,还有许多其它的随机脉冲。我们用蒙特卡罗法和模板匹配法检测脉冲序列中的优势模式,对于第一种情况,可以较好地得到处理;对于第二、三种情况,视模板重叠的位置和量化间隔(指程序中的量化)的选择,部份可以得到解决。实际上神经元放电非常复杂,这里只是取一些较为简单的

放电过程。神经元的簇放电有可能找到两类优势模式;一类是频率较高的,即间隔较小的模式;另一类是大间隔的模式,而小间隔的模式被包含在大间隔模式的某些间隔中。这种情况须个别具体解决,本文未包括这种复杂情况。

为了使神经元放电脉冲的记录准确,要求实验动物处于平稳状态,探针扎入单个神经元细胞,记录神经元的放电脉冲。这就要求尽可能减小外界干扰。所以实验是在屏蔽室内无振动的工作台上进行的。一般用被麻醉的小白鼠作实验。脉冲信号经过放大和整形,以  $18\mu\text{s}$  的采样周期记录脉冲间隔大小,并以文件方式记录在软磁盘上备用。

### 2. 蒙特卡罗法识别优势模式

蒙特卡罗法作为一种数学工具能解决一些统计的问题。神经元放电脉冲是以间隔长度的方式被记录下来的。从这一组原始数据为一个样本,可以统计出  $i$  个间隔 ( $i = 2,$

3, …) 出现次数大于  $k$  次 ( $k = 2, 3, \dots$ ) 的模板个数  $C_0(t, k)$ 。那末, 把原始数据打乱, 而保持数据的大小不变时, 就组成了一个随机序列。如果打乱 99 次, 那末, 连同原始数据就组成 100 个数据样本。对这 100 个样本进行统计分析, 就可以找到优势模式。对 99 个随机数据样本进行与原始数据一样的处理得到  $C_s(t, k)$  模板数,  $s = 1, 2, \dots, 99$ , 为打乱样本的序数。  $\max(C_s(t, k))$  表示打乱的 99 次中  $C_s(t, k)$  的最大值。下面分几种情况来讨论。

(1)  $\max(C_s(t, k)) = 0$  这说明在随机序列中不存在  $t$  个间隔出现次数大于  $k$  次的模板, 亦即在原始数据中  $t$  个间隔出现次数大于  $k$  次的  $C_0(t, k)$  个模板在 99% 的概率上是优势模式。

(2)  $\max(C_s(t, k)) > C_0(t, k)$  这说明在 99 个随机样本中,  $t$  个间隔出现次数大于  $k$  次的模板数比原始数据中的还多, 所以  $C_0(t, k)$  个模板都不是优势模式。

(3)  $\max(C_s(t, k)) < C_0(t, k)$  这说明  $[C_0(t, k) - \max(C_s(t, k))]$  个模板可能为优势模式, 其概率为 99%。至于哪些是优势模式, 则要用模板匹配法计数确定。

(4)  $\max(C_s(t, k)) = C_0(t, k) - 1$  这表示此模板为优势模式的概率小于 99%。可统计出在 99 个随机样本中不出现  $t$  个间隔, 并多于  $k$  次的模板样本数。这一数大致反映了  $C_0(t, k)$  所表示的模板为优势模式的概率。

对于  $\max(C_s(t, k)) = C_0(t, k) > 1$  的情况可用模板匹配法对  $C_0(t, k)$  个模板逐一匹配挑选, 其中出现次数多的可考虑为优势模式。

神经元放电脉冲序列的间隔大小差异很大, 所以需先作量化, 因此一般称做量化蒙特卡罗法。量化步长对统计结果有一定影响。步长太大, 由量化产生的误差就会很大; 步长太小, 模板分类就太细。这里常用的量化步长为 500—1100, 即 10—20ms。

### 3. 模板匹配法

模板匹配法是蒙特卡罗法的补充, 是用来确定在  $\max(C_s(t, k)) < C_0(t, k)$  情况下的优势模式, 计算由蒙特卡罗法确定的优势模式在脉冲序列中出现的次数的。由于在用蒙特卡罗法找模板时, 脉冲序列的量化间隔一般较大, 模板也不给容差, 所以计数量很小。考虑到在神经元放电过程中, 脉冲序列的频率虽会有所改变, 但其模板的形式可保持不变, 所以在匹配时, 采用模板间隔的相对比例不变的方式来进行匹配, 并给以一定的容差, 这样将更符合电生理实验的实际情况。由于在作原始数据模板统计时, 在统计  $C_0(t, k)$  个模板时, 同时记录下了  $t$  个间隔出现次数大于  $k$  次的  $C_0$  个模板在脉冲序列中的起始位置  $j_{n, n=1, C_0}$ , 所以读入  $t, j$  就能找到模板的具体数据。若脉冲序列为  $S(i)$ , 那末模板即为  $S(j), S(j+1), \dots, S(j+t-1)$ 。这样就能方便地进行匹配计数。

## 三、软件流程和各模块功能说明

整个程序是在 IBM-PC/AT 上完成的, 程序用 FORTRAN 语言写成, 分为五个模块, 占 106K 内存。程序流程图示于图 2。现对各模块功能作如下说明。

### 1. 读数据

原始数据记录脉冲间隔大小,  $18\mu s$  为 1 个单位。由于脉冲间隔大小相差很大, 所以

原始数据按高低位存放。程序中先合并为 4 字节的可直接读取的数据文件。

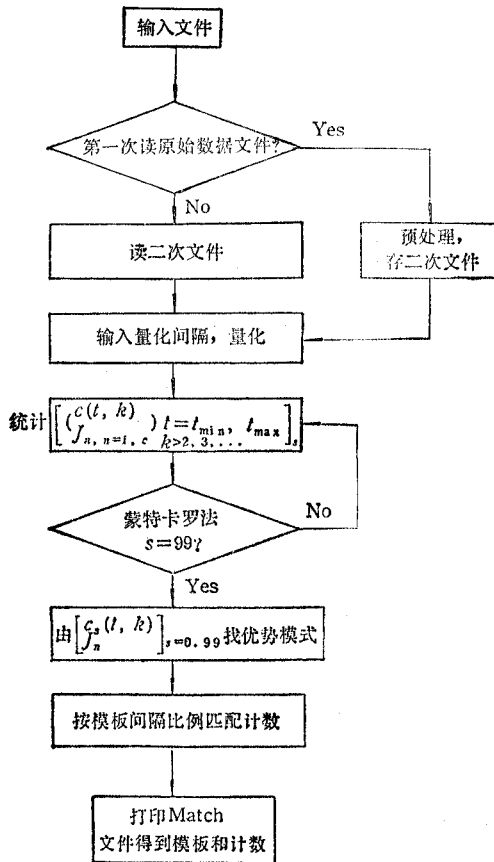


图 2 神经元放电优势模式识别程序流程图。

#### 4. 蒙特卡罗法

把原始数据打乱 99 次, 得到 99 个随机样本, 作如上相同统计, 记录  $[C_s(t, k)]_{t=t_{\min}, t_{\max}, k=2, 3, \dots}^{s=1, 99}$  (见表 1), 然后按第二节中的方法确定优势模式。

#### 5. 模板匹配

由蒙特卡罗法选出的优势模式或可能的优势模式都要经过模板匹配法确认和计数。只要输入  $t, j$  就可以自动找到模板进行匹配计数, 并把模板的间隔大小和计数值全部記入文件。最后打印文件就可得到全部结果。

## 2. 量化

神经元放电脉冲间隔的范围很大, 所以必须先量化以缩小数据范围。当量化步长为  $s$  时小于  $s/2$  的间隔与下一个间隔合并。这可避免不合理的零间隔的产生, 同时有利于消除因噪声和随机插入脉冲的干扰。量化中采用四舍五入。

## 3. 对原始数据进行模板统计

程序可输入任意的模板间隔数  $t$ , 统计  $C_0(t, k)$ , 即统计  $t$  个间隔出现次数大于  $k$  次的模板数。  $t=t_{\min}, t_{\max}, k=2, 3, 4, \dots$ 。统计是这样做的, 从脉冲序列的第  $j$  个数据开始取  $t$  个间隔作为模板, 依次向后滑动找出与这  $t$  个间隔相同的模板个数  $kk(t, j)$ ,  $t$  表示间隔数,  $j$  表示从第  $j$  个数据开始的  $t$  个间隔是模板,  $j=1, N/2, N$  为脉冲序列的长度。由  $kk(t, j), j=1, N/2, N$  可统计出  $kk \geq k$  的模板个数  $C_0(t, k)$  和  $i_{n, n=1, c_0}$ 。这样既统计了长度为  $t$  出现次数大于  $k$  次的模板个数, 也记录了模板在原始数据中的位置。同时把  $C_0(t, k)$  和  $i_n$  记录在文件中。

## 四、结 语

用上述方法分析了蓝斑神经元放电脉冲序列, 结果列于表 1。Shuffle 0 为对原始数据的统计。表中还给出部份打乱后的统计结果。由表 1 得到  $C_0(3, 2) = 22, \max[C_s(3, 2)] = 17$ 。用蒙特卡罗法判断,  $C_0(3, 2)$  中可能有 5 个为优势模式。继而用模板匹配法

表 1 蓝斑神经元放电脉冲间隔模式的统计结果

Shuffle	$t$	$k$	$C(t, k)$	$i_{n, n-1, C}$
0	3	2	22	11, 12, 43, 63, 86, 125, 145, 147, 174, 177, 193, 219, 220, 248, 260, 277, 285, 304, 333, 335, 337, 357
0	3	3	1	11
0	4	2	1	219
1	3	2	5	
1	4	2	1	
2	3	2	12	
2	3	3	1	
3	3	2	8	
⋮				
16	4	2	2	
17	3	2	17	
17	3	3	1	
17	4	2	1	
18	3	2	13	
⋮				

对  $C_0(3, 2)$  的 22 个模板进行匹配, 得到出现次数多的有: (1, 1, 1) 出现 15 次, (13, 15, 8) 出现 17 次, (15, 20, 15) 出现 18 次, 而出现 14 次的有 3 个模板, 故不予考虑(因为优势模式的总数为 5)。这里量化长度为 20ms, 容差为 0.1。

同时由表 1 可以看出,  $C_0(4, 2) = 1$ , 而  $\max(C_i(4, 2)) = 2$ , 所以  $C_0(4, 2)$  表示的这一个模板不可能为优势模式。

表 1 所统计的原始数据长度为 772, 统计了 3 和 4 个间隔的模板, 打乱 99 次, 共用了 45min; 对 22 个模板的匹配也只需 20min, 速度比原来<sup>[6]</sup>提高了一倍多。

神经元放电脉冲序列的优势模式识别对进一步研究神经元放电的电生理具有重要意义。这是解释神经元放电脉冲序列中各种模式的生物学意义的第一步。在对神经元放电脉冲序列中的模式无任何先验知识的情况下, 成功地应用蒙特卡罗法作为统计工具, 分析了各种模式出现的确定性或出现的随机性, 并由一定的判据在确定的统计概率下挑选出优势模式, 然后由模板匹配法逐个进行计数。

对神经元放电脉冲序列的模式识别的研究还是很初步的。从寻找模式规律来说, 应尽可能地固定电生理的实验条件, 只改变单项指标, 借以找到优势模式与生理条件的关系。从识别工作本身来说, 神经元放电脉冲序列形式多样, 如簇放电形式有可能存在两种模式, 即大周期模板和小周期模板, 而小周期模板嵌套在大周期模板中, 情况很复杂, 这些都是须要进一步研究的。

## 参 考 文 献

- [ 1 ] J. E. Dayhoff, G. L. Gerstein, *J. Neurophysiology*, 49(1983)6, 1334—1348.
- [ 2 ] J. E. Dayhoff, G. L. Gerstein, *J. Neurophysiology*, 49(1983)6, 1349—1363.
- [ 3 ] D. H. Perkel, G. L. Gerstein, G. P. Moore, *Biophys. J.*, 7(1967), 419—440.
- [ 4 ] T. J. Marczyński, *Materia Med. Pol.*, 8(1976), 97—107.
- [ 5 ] G. J. Sherry, T. J. Marczyński, *Int. J. Neurosci.*, 3(1972), 259—270.
- [ 6 ] 梁虹, 陈筱英, 柴振明, 脑神经元放电脉冲序列间隔的解码方法, 第三届全国信号处理学术会议, 西安, 1988年。

## DETERMINATION OF THE FAVORED PATTERNS IN SPIKE TRAIN WITH MONTE CARLO METHOD AND TEMPLATE METHOD

Chen Xiaoying

*(Institute of Electronics, Academia Sinica, Beijing)*

**Abstract** The modern neuroelectrical physiology indicates that there are certain patterns in neural spike train, which appear more frequently than others. These favored patterns may be related to the mechanism of neural information processing in the central nervous system. In this paper, the quantized Monte Carlo method and template method which are used in favored pattern recognition are described. The Monte Carlo method determines favored patterns or the possible favored patterns. The template method is used for choosing the favored patterns from the candidates and counting the number of the patterns in the spike train. This work will be pursued for extracting the patterns in some complicated spike train with very different discharge cycles.

**Key words** Neural spike train; Favored patterns; Monte Carlo method