

全国第七届研究生数学建模竞赛



题目 基于多分类器集成的神经元特征选择及分类

摘 要：

本文在多个常用分类模型的基础上，通过对于神经元空间几何形态的深入分析，提出了鲁棒的对神经元几何特征选择和多分类的模型。

对于问题 1：我们根据题目给定的数据集 A 和 C 建立训练数据集。通过深入分析数据文件的各个分布和题目给定的几个几何特征描述。我们从离散化的神经元文件中抽象出 43 个空间几何特征进行分析。并提出差分特征选择模型选择出了对各类神经元判别能力强的特征集合，并将差分 KNN、决策树和 SVM 进行集成，达到稳定分类的目的。

对于问题 2：通过差分特征选择以及多分类器系统可以将正常数据进行正确的分类，为了发现新神经元个体，我们引入类边界距离参数，实验证明，能够有效的找出潜在的神经元的新类别。

对于问题 3：通过扩充多分类器系统的训练集以及采用反馈学习的策略，可以不断完善神经元类别的知识库，增加分类鲁棒性，对于新类型神经元，采用启发式的命名方式，可以有效的给一些新类别神经元命名。

对于问题 4：采用差分特征选择模型可以很有效的选择出不同物种同种神经元的强判别特征集合，通过真实数据验证了差分选择模型的鲁棒性。

对于问题 5：基于 Burke 算法建立 Markov M 随机采样仿真生长模型，通过对训练集实验数据的拟合实现对模型参数的估计。实验表明该生长模型可以针对某一特定神经元做出合理的生长预测。

[关键词]：神经元分类、多分类集成、差分特征选择、生长模型、边界距离

参赛队号 10422016

队员姓名 穆蕊 任昭春 李丕绩

参赛密码 _____
(由组委会填写)

中山大学承办

目录

一 问题重述.....	3
1.1 背景分析.....	3
1.2 问题提出.....	3
二 模型假设.....	5
三 符号约定.....	6
四 问题分析、建模与求解.....	8
4.1 问题一.....	8
4.1.1 问题分析.....	8
4.1.2 模型建立.....	8
4.1.3 求解.....	15
4.2 问题二.....	25
4.2.1 问题分析.....	25
4.2.2 模型建立.....	25
4.2.3 求解.....	26
4.2.4 验证.....	28
4.3 问题三.....	30
4.3.1 问题分析.....	30
4.3.2 模型建立.....	30
4.3.3 求解与验证.....	30
4.4 问题四.....	30
4.4.1 问题分析.....	30
4.4.2 模型建立.....	31
4.4.3 求解与验证.....	31
4.5 问题五.....	35
4.5.1 问题分析.....	35
4.5.2 建立基于Burke算法的预测模型.....	37
4.5.3 预测模型的求解.....	40
4.5.4 模型结果验证.....	42
五 模型评价与推广.....	44
参考文献.....	44
附录.....	45

一 问题重述

1.1 背景分析

被称为“人的生命活动中枢”的神经系统，其解剖结构和生理功能的复杂程度，已远远超过了目前人类的认识能力。尤其是脊椎动物的神经系统是一个非常复杂的生物网络，协调且控制着基本的身体动作，对外界的刺激做出各种反射。高级哺乳动物的神经系统是由 2500-5000 类神经元产生 25000-100000 个轴突连接在其中所组成的一个网络[1]，而无脊椎动物要相对简单许多。多少年来，人类为探索其奥秘在进行着不懈的努力。大脑是生物体内结构和功能最复杂的组织，人类脑计划（Human Brain Project, HBP）的提出就是要对全世界的神经信息学数据库建立共同的标准，多学科整合分析大量数据，加速人类对脑的认识。在过去的 20 多年里，随着新技术的不断涌现，生物学家们逐渐能够对神经元的物理和化学变化进行探测。共聚焦显微镜和双光子激发荧光显微技术的出现让我们对于神经元的三维结构和动态变化能够进行更加详尽的观察，空间上能够精确到微米级别，时间上能精确到毫秒级别。这些定量分析手段和荧光染色技术相结合就使得我们能够对神经元内的某种特定的分子进行时空分析。

随着新技术的不断发展，我们积累了大量的关于神经元形态、结构和功能方面的数据。对神经元特性的认识，最基本问题是神经元的分类。但是迄今，对于众多的神经元类型仍然没有令人满意的命名和确切的定义，取而代之的是类、子类、型、子型[2][3]。对神经元的系统分类是研究脊椎动物神经系统的的一个重要课题，它是建立完整的神经网络图以及研究各部分结构功能的先决条件。

在研究过程中，人们发现不同神经元的结构对于它的作用是有决定性意义的，如下图所示：长柄草靶形状的神经元以及簇状神经元对于不同的输入反应不同，唯一的原因就是它们的形状不同，因此研究细胞形态不仅可以区分不同细胞还可以理解他们内部的连接构架，以便对它们的功能进行区分。达到由形态到功能分类的过渡。

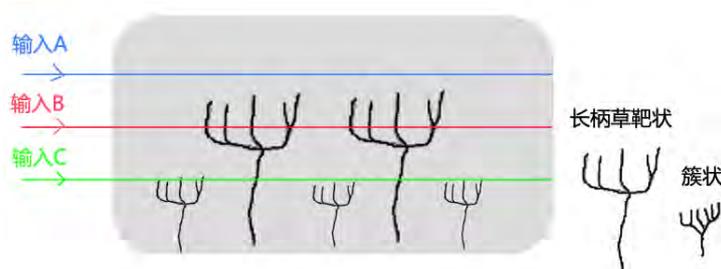


图 1-1：神经元形状影响功能

1.2 问题提出

对神经元特性的认识，最基本问题是神经元的分类。目前，神经元有目前有以下几种分类法，根据突起的多少可将神经元分为三种：（1）多极神经元 (multipolar neuron)，有一个轴突和多个树突；（2）双极神经元 (bipolar neuron)，有两个突起，一个是树突，另一个是轴突；（3）假单极神经元 (pseudounipolar neuron)，从胞体发出一个突起，距胞体不远又呈“T”形分为两支，一支分布到外周的其他组织的器官，称周围突 (peripheral process)；另

一支进入中枢神经系统，称中枢突(central process)。假单极神经元的这两个分支，按神经冲动的传导方向，中枢突是轴突，周围突是树突；但周围突细而长，与轴突的形态类似，故往往通称轴突。

根据轴突的长短，神经元可分为：(1)长轴突的大神经元，称 Golgi I 型神经元，最长的轴突达 1m 以上；(2)短轴突的小神经元，称 Golgi II 型神经元，轴突短的仅数微米。

根据神经元的功能又可分为：(1)感觉神经元(sensory neuron)，或称传入神经元(afferent neuron)多为假单极神经元，胞体主要位于脑脊神经节内，其周围突的末梢分布在皮肤和肌肉等处，接受刺激，将刺激传向中枢。(2)运动神经元(motor neuron)，或称传出神经元(efferent neuron)多为多极神经元，胞体主要位于脑、脊髓和植物神经节内，它把神经冲动传给肌肉或腺体，产生效应。

(3)中间神经元(interneuron)，介于前两种神经元之间，多为多极神经元。动物越进化，中间神经元越多，人神经系统中的中间神经元约占神经元总数的 99%，构成中枢神经系统内的复杂网络。

如何识别区分不同类别的神经元，这个问题目前科学上仍没有解决。生物解剖区别神经元主要通过几何形态和电位发放两个因素。神经元的几何形态主要通过染色技术得到，电位发放通过微电极穿刺胞内记录得到。利用神经元的电位发放模式区分神经元的类别比较复杂，主要涉及神经元的 Hodgkin-Huxley 模型和 Rall 电缆模型的离散形式(神经元的房室模型)。本问题只考虑神经元的几何形态，研究如何利用神经元的空间几何特征，通过数学建模给出神经元的空间形态分类方法，将神经元根据几何形态比较准确地分类识别。

神经元的空间几何形态的研究是人类脑计划中一个重要项目，NeuroMorpho. Org 包含大量神经元的几何形态数据等，现在仍然在不断增加，在那里可以得到大量的神经元空间形态数据，例如附录 A 和附录 C。对于神经元几何形态的特征研究这个热点问题，不同专家侧重用不同的指标去刻画神经元的形态特征，例如图 1、下面给出的神经元的粗略空间刻画以及附录 A 和附录 C 用标准的 A. SWC 格式给出的刻画。需要完成的任务是：

- 1) 利用附录 A 中和附录 C 样本神经元的空间几何数据，寻找出附录 C 中 5 类神经元的几何特征（中间神经元可以又细分 3 类），给出神经元空间形态分类的方法。
- 2) 附录 B 另外有 20 个神经元形态数据，能否判定它们属于什么类型的神经元。在给出的数据中，是否有必要引入或定义新的神经元名称。
- 3) 神经元的形态复杂多样，神经元的识别分类问题至今仍没有解决，是否可以提出一个神经元分类方法，将所有神经元按几何特征分类。你们能否给生物学家为神经元的命名提出建议（附录 A 和附录 C 的神经元是比较重要的类别，实际应该有很多其他类别）。
- 4) 按照模型神经元形态分类方法，能否确定在不同动物神经系统中同一类神经元的形态特征有区别吗？例如，附件 A 中有猪的普肯野神经元和鼠的普肯野神经元，它们的特征有区别吗？
- 5) 神经元的实际形态是随着时间的流逝，树突和轴突不断地生长而发生变化的，能预测神经元形态的生长变化吗？这些形态变化对你们确定的几何形态特征有什么影响。

二 模型假设

- 1) 假设附件 A 中的每个神经元空间形态数据都正确的描述了本类别神经元的基本信息，没有错类描述或者变异。
- 2) 假设附件 C 中的每个神经元都是本类别的典型代表。
- 3) 假定部分用 L-Measure[4] 计算的神经元特征相对正确，与真实值的误差是相对误差。
- 4) 假定根据 NeuroMorpho. Org¹ 的搜索方式搜索到的神经元空间形态数据描述是用户所需要的，即 NeuroMorpho [6] 对神经元的分类方式是基本正确的。
- 5) 假设多极神经元、三级神经元、双极神经元和单极神经元是严格按照突触的数目来对神经元进行形态划分的。
- 6) 类别假设：第一类是运动神经元、第二类是普肯野神经元、第三类是锥体神经元、第四类中间神经元、第五类是感觉神经元。

¹ <http://NeuroMorpho.Org>

三 符号约定

符号	含义
R	实数集合
$x \in R^n$	输入和 n 维欧式空间
$y \in Y$	输出和输出集合
N_i	神经元 i
T	训练集
F	特征向量集合
F_i	描述神经元 N_i 几何特征的特征向量
$F_i(k)$	描述神经元 N_i 几何特征的特征向量第 k 维
F_{c_i}	第 i 类质心向量
$\Delta F_{center_{ij}}$	类 C_i 和类 C_j 质心差分
f_i	特征向量的第 i 个元素，即第 i 个特征
δ_{center}	差分阈值
$S_{\Delta F_{ij}}$	类 C_i 和类 C_j 差分特征选择后的集合
S_{F_i}	差分特征选择后，类 C_i 的判别特征集合
$d(x, y)$	向量 x 和 y 的欧式距离
ξ_i	类 C_i 的距离边界
d_s	分叉开始的直径
d_e	分叉结束的直径
d_0	模拟树突的第一段房室的直径
ΔL	模拟房室的固定长度
G	服从高斯分布的随机变量，均值为 μ ，标准差为 σ
$l(d)$	直径落在直径阈值 d 之内的房室长度
L_{br}	分支长度

$N_{ov}(d)$	在继续分叉的分支上所观测到的直径位于直径阈值 d 之内的房室数目
$N_{nov}(d)$	在继续生长的分支上所观察到的直径位于直径阈值 d 之内的房室数目
$N_{trm}(d)$	在终止生长的分支上所观测到的直径位于直径阈值 d 之内的房室数目
$P_{ov}(d)$	直径为 d 的房室继续分叉的概率
$P_{trm}(d)$	直径为 d 的房室终止生长的概率
P	标准化的父分支起始直径
R'	服从均匀分布的随机变量（位于 0-1 之间）

四 问题分析、建模与求解

4.1 问题一

4.1.1 问题分析

这里要解决两个问题：（1）根据附件中给出的空间几何数据计算每一类神经元的几何特征；（2）建立模型，根据几何特征对神经元进行分类。

这种给出了已知训练数据，经过分析后提出模型，对未知类别数据进行分类的问题，是典型的机器学习数据挖掘问题，可以用分类或者聚类的方法分析，对神经元的分类是一个多类别分类问题。

对于几何特征的计算，通过对神经元空间形态的详细分析，结合 L-Measure 这种强大的神经元细胞计算工具，可以对神经元的几何形态有较好的描述，建立特征向量。

提取特征之后，建立起训练集对分类模型的参数进行训练，考虑到训练集数据太小以及特征向量维度过高问题，我们需要扩充训练集 T 以及进行有效的特征选择。

为了增强神经元学习分类问题的鲁棒性，我们尝试 KNN、决策树、支持向量机 (SVM) 多分类器集成学习。

4.1.2 模型建立

4.1.2.1 几何特征

结合附件给出的 A、C 的空间几何数据，通过对神经元基本几何形态结构的分析，并查阅相关文献，在 Matlab 和 L-Measure 软件的帮助下，我们求解了以下 43 个神经元的几何特征，组成特征向量 F ，即 $|F|=43$ 。

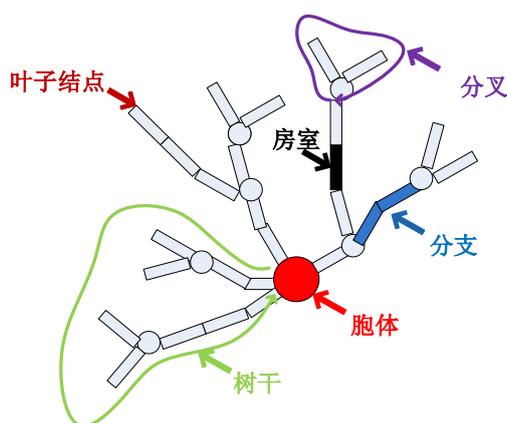


图 4-1 神经元结构图

1) Soma_Surface: 细胞体表面积

代表类型 1 所表示的细胞体的表面积，如果该细胞体只由一部分构成，则我们假设为球体，面积计算为 $4\pi R^2$ ，其中 R 为胞体半径。如果细胞体包含多个部分，则把各部分假设为圆柱，表面积计算公式为各圆柱表面积之和。

2) N_stems, 树干数目: 与胞体相连接的树干的数目

- 3) N_bifs: 分叉数目
- 4) N_branch: 分支数目
- 5) N_tips: 叶子节点数目, 不再继续分叉的叶子终端
- 6) Width 宽度
- 7) Height 高度
- 8) Depth 深度
宽度是指整个神经元在 x 轴上的跨度, 我们利用主成分分析法计算出这个宽度, 所提取的主成分的数量为 90%, 即两个边缘分别去掉 2.5% 的异常值。高度和深度也是用同样的方法。
- 9) Type 房室种类 (1 细胞体 2 轴突 3 树突 4 尖端树突)
- 10) Diameter, 神经元直径: 神经元的各个房室的直径之和
- 11) Diameter_pow, 直径幂: 神经元的各个房室的直径的 1.5 次方
- 12) Length: 房室长度 L, 两个房室终点处的距离, 组成胞体的各个部分没有计算在内
- 13) Surface: 房室表面积, 将各个房室看做圆柱形分别计算表面积 $2\pi R$ * 房室长度, 整个神经元表面积为各部分之和。
- 14) SectionArea: 房室横切面积, 横切面看做圆, 其面积为 πR^2 其中 R 为房室半径
- 15) Volume: 房室体积 (圆柱体), 体积为 $\pi R^2 L$
- 16) EucDistance: 房室与胞体之间的欧拉距离
- 17) PathDistance: 房室与胞体之间的路径距离
- 18) Branch_Order: 分支阶数, 分支到胞体之间的分支数目

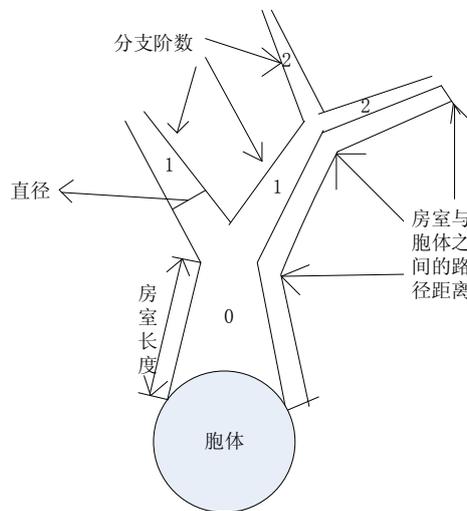


图 4-2 胞体分支

- 19) Terminal_degree: 终端叶子数目, 每个房室结尾的叶子节点的数目
- 20) TerminalSegment: 终端分支指数
所有作为终端的分支, 其指数为 1, 这个指标有两个用途, 一是可以用来计算所有作为终端的分支数目, 二是用来区分终端分支 (此指数为 1) 和中间连接分支 (此指数大于 1)
- 21) Taper_1 : Burke 锥化度

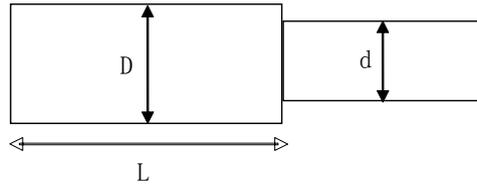


图 4-3 直径关系

如图所示 为相邻两个房室直径的关系 $Taper_1 = (D-d)/L$
 22) Taper_2 Hillman 锥化度

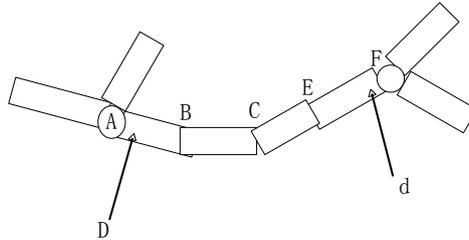


图 4-4 直径关系 2

如图所示为一个分支的终点处的直径与起始直径的关系 $Taper_2 = d/D$

23) Branch_pathlength: 分支路径长度: 组成分支的房室路径长度

24) Contraction: 收缩比例

一个分支的欧拉距离与路径长度的比率 如图所示 收缩比例
 $= AF / (AB + BC + CE + EF)$

25) Fragmentation: 单位分支的房室数目

表示介于两个分叉节点间或者在一个分叉节点与终端的房室数目

26) Daughter_Ratio: 女儿比

女儿比 = 大女儿直径 / 小女儿直径

27) Parent_Daughter_Ratio: 父女比

父女比 = 父直径 / 大女儿直径

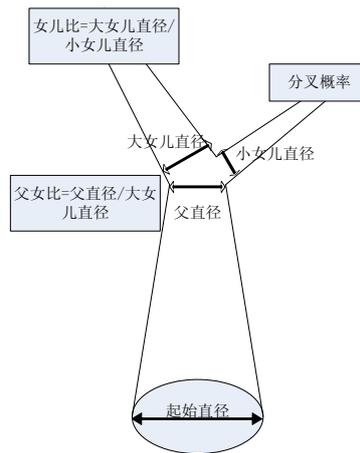


图 4-5 父女图式结构

28) Partition_asymmetry: 分支不对称指数

反映两个分支拥有的叶子节点个数之间的关系, 设右侧分支的叶子节点个数为 N_1 , 左侧分支的叶子节点个数为 N_2 , 则分支不对称指数 = $(N_1 - N_2) / (N_1 + N_2 - 2)$

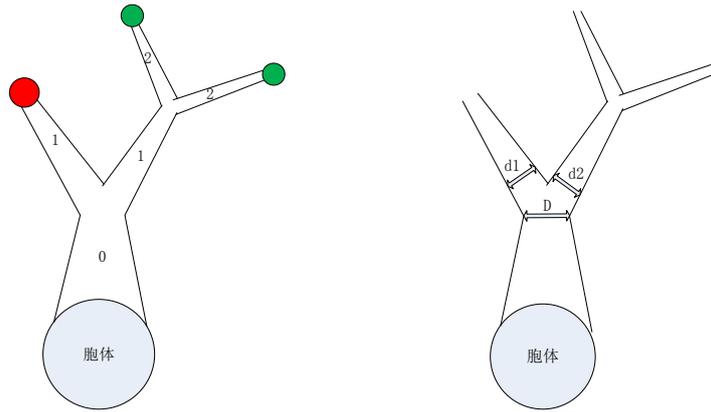


图 4-6 分支不对称

如上图所示：分支不对称指数 $= (2-1) / (2+1-2) = 1$

29) Rall_Power : Rall 值 n

如图所示： n 为满足下式的值 $D^n = d_1^n + d_2^n$ ，取值一般在 0-5 之间且精确到千分位，当均匀分叉的时候该 Rall 值无意义

30) Pk，Pk 值 n ， n 为下述比值，它是 Rall 值的平均水平。 $\frac{d_1^n + d_2^n}{D^n}$

31) Pk_classic，经典 Pk 值，反映 Rall_Power 的 1.5 次幂

32) Pk_2，反映 Rall_Power 的 2 次幂

33) Bif_ampl_local：局部分叉角度

34) Bif_ampl_remote：远端分叉角度

35) Bif_tilt_local：分叉倾斜度 为途中 α 与 β 中的较小的角度。

36) Bif_tilt_remote：远端分叉倾斜度

为父支到两个分支的最远端之间夹角的最小值，但是这种情况下胞体不考虑在内。

37) Bif_torque_local：局部分叉扭曲度

38) Bif_torque_remote：远端分叉扭曲度

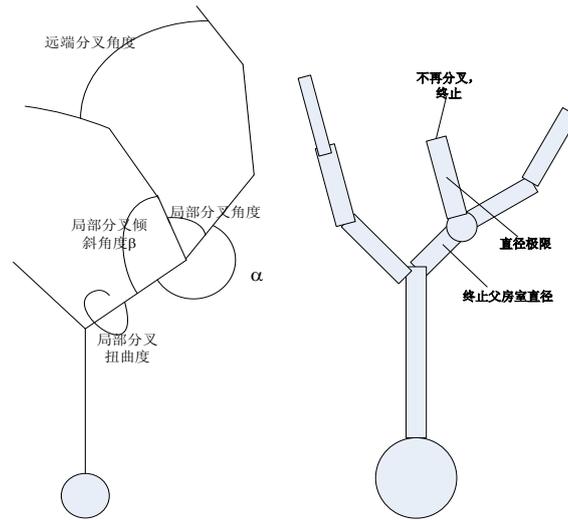


图 4-7 分叉结构

- 39) Last_parent_diam , 终止父房室直径, 最后一个分叉的父房室直径
- 40) Diam_threshold, 直径极限值, 连接最后一个分叉的首个房室的直径
- 41) HillmanThreshold, Hillman 极限, 当前分支的父亲房室 $\times 0.5$ +两个子节点 $\times 0.25$
- 42) Helix: 神经元的螺旋度
- 43) Fractal_Dim :是利用分形几何中的数盒子法所计算出来的神经元的分支 Hausdorff 维数

特征提取出来之后, 针对每一个神经元 N_i , 构建一个特征向量 F_i , 即:

$$F_i = \{f_1, f_2, \dots, f_k, \dots, f_{|F_i|}\}$$

(1)

在构建训练集时候需要归一化。

4.1.2.2 差分特征选择

特征选择是模式识别中的关键技术之一。一般情况下, 只有在特征矢量中包含足够多的类别信息, 才能通过分类器实现正确分类。由于特征是否包含足够的信息很难确定, 为了提高识别率, 总是最大限度地提取特征信息, 结果不仅使特征维数增大, 而且其中可能存在较多的无效和冗余特征。因此, 选择合适的特征来描述模式, 对模式识别精度、训练时间和存储空间等许多方面都有较大影响, 并且对分类器的构造也起着非常重要的作用。

特征选择的标准较多, 经典的选择算法大多采用概率度量、距离度量和熵度量等衡量标准。在足够多的样本前提下, 基于大样本统计理论的上述准则和相应的分类算法是合理的; 而当训练样本为小样本时, 上述准则不一定总能取得良好的效果。此外, 特征选择常用的另一类准则还有类内距离、类间距离和散布矩阵度量等。这些方法虽然应用广泛, 但是, 其识别算法大多是建立在经验模型的基础上, 如神经网络等, 其模型参数和结构的确定受数据和识别算法的影响较大。总之, 经典的特征算法和相应的分类器大多注重的是大样本下使经验风险最小, 而对分类器的推广性能要求较低。

由于神经元细胞一般来说每一类都会有自己固有的形态结构，每一种形态结构都会对应着不同的特征向量和特征组合，所以，选择一种简单的特征选择方法是十分必要的。

由于分类模型是采用的 KNN、决策树、和支持向量机的多分类器集成学习系统，决策树、SVM 都能够有效的进行特征选择和降维，所以，我们设计的 $|C|$ 差分特征选择模型将协助 KNN 进行神经元分类。

由于问题一是要对运动神经元、普肯野神经元、椎体神经元、中间神经元、和感觉神经元五大类神经元进行分类，其中中间神经元根据突触数量又分为二极、三极、多极神经元。所以，经过分析，我们将此问题归纳为一个多类别分类问题，具体类别数量是 $|C|=5$ 。然后根据中间神经元中各种数量极神经元的定义进行子类的划分。

$|C|$ 差分特征选择的核心在于类间特征向量的距离，首先计算 $|C|$ 种类别每一类的质心：

$$F_{center_i} = \frac{\sum_{N_j \in C_i} F_j}{|C_i|}, i = 1, 2, \dots, |C| \quad (2)$$

这样就得到了每一个类的质心。接下来开始差分，即，每一个类的质心和其他 $|C|-1$ 个类做差，例如 C_i 和 C_j 的质心差分：

$$\Delta F_{center_{ij}}(k) = F_{center_i}(k) - F_{center_j}(k), k = 1, 2, \dots, |F| \quad (3)$$

不是一般性，两个变量的差越大，那么他们就越不相同。同样，一个特征 f_k 两类的差越大，他对于这两类的判别能力就越强。这里，我们引入一个差分阈值变量 δ_{center} ，每一个差分向量先进行特征选择：

$$S_{\Delta F_{ij}}(k) = f_k, \text{ if } \Delta F_{center_{ij}}(k) > \delta_{center} \quad (4)$$

δ_{center} 越大，选择出来的特征就越少，反之越多。这样对于类 C_i 的 $|C|-1$ 个差分向量都做了对任何两类判别能力强的特征选择，只要再把这些特征项集合求交集，就得到了类 C_i 能够判别开其他任何一个类的特征选择。

$$S_{F_i} = \bigcap S_{\Delta F} \quad (5)$$

不排除做了一次差分之后， $S_{F_i} = \emptyset$ 的情况，说明对于选择的差分阈值 δ_{center} ，没有特征能将该类别与其他类别彻底分开。对于这种情况，将质心 F_{center_i} 去除，对剩余的质心向量集合再做差分特征选择，重新选择 δ_{center} ，直到每个类的

$S_{F_i} \neq \emptyset$ 。由于这种问题的存在，在与 KNN 进行结合分类的时候，KNN 也要按一定的顺序对类别进行迭代。

整个过程可以表达如下：

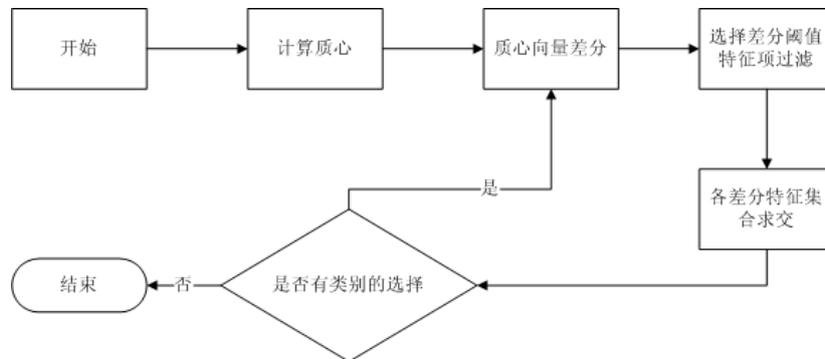


图 4-8 差分特征选择流程图

4.1.2.3 多分类器系统

研究发现，不同分类器在识别性能上有互补作用，因此把各种分类器结合在一起，从而能够集成各个分类器的优点。为了保证系统的鲁棒性和泛化能力，我们采用了集成 KNN、决策树、和支持向量机的多分类器系统。

(1) KNN 分类器

KNN[6]分类算法是一种传统的基于统计的模式识别方法。算法思想很简单：对于一篇待分类文档，系统在训练集中找到 k 个最相近的邻居，使用这 k 个邻居的类别为该文档的候选类别。该文档与 k 个邻居之间的相似度按类别分别求和，减去一个预先得到的截尾阈值，就得到该文档的类别测度。用 KNN 也表示所选 k 个最相近文档的集合，公式 (6) 刻画了上述思想。

$$y(x, c_j) = \sum_{d \in kNN} sim(x, d_j) y(d_i, c_j) - b_j \quad (6)$$

在模型当中，我们对 KNN 进行了改进，称为“差分 KNN 分类器”，通过差分特征选择之后，将待分类的选择之后的特征向量输入进行分类，即做了特征的预选择，经过试验证明，效果大有改善，基本解决了错分问题。

(2) 决策树

决策树 (Decision Tree) 又称为判定树，是运用于分类的一种树结构。其中的每个内部结点 (internal node) 代表对某个属性的一次测试，每条边代表一个测试结果，叶结点 (leaf) 代表某个类 (class) 或者类的分布

(class distribution)，最上面的结点是根结点。决策树分为分类树和回归树两种，分类树对离散变量做决策树，回归树对连续变量做决策树。

构造决策树是采用自上而下的递归构造方法。决策树构造的结果是一棵二叉或多叉树，它的输入是一组带有类别标记的训练数据。二叉树的内部结点 (非叶结点) 一般表示为一个逻辑判断，如形式为 (a = b) 的逻辑判断，其中 a 是属性，b 是该属性的某个属性值；树的边是逻辑判断的分支结果。多叉树 (ID3) 的内部结点是属性，边是该属性的所有取值，有几个属性值，就有几条边。树的叶结点都是类别标记。

使用决策树进行分类分为两步：

第 1 步：利用训练集建立并精化一棵决策树，建立决策树模型。这个过程实际上是一个从数据中获取知识，进行机器学习的过程。

第 2 步：利用生成完毕的决策树对输入数据进行分类。对输入的记录，从根结点依次测试记录的属性值，直到到达某个叶结点，从而找到该记录所在的类。

(3)支持向量机

支持向量机 [7][8]SVM(Support Vector Machine)AT&T Bell 实验室的 V. Vapnik 提出的针对分类和回归问题的统计学习理论。由于 SVM 方法具有许多引人注目的优点和有前途的实验性能，越来越受重视，该技术已成为机器学习研究领域中的热点，并取得很理想的效果，如人脸识别、手写体数字识别和网页分类等。

SVM 的主要思想可以概括为两点：(1)它是针对线性可分情况进行分析，对于线性不可分的情况，通过使用非线性映射算法将低维输入空间线性不可分的样本转化为高维特征空间使其线性可分，从而使得高维特征空间采用线性算法对样本的非线性特征进行线性分析成为可能；(2)它基于结构风险最小化理论之上在特征空间中建构最优分割超平面，使得学习器得到全局最优化，并且在整个样本空间的期望风险以某个概率满足一定上界。

在 SVM 实现中我们使用开源的 libsvm[9]。

多分类器集成策略模型中采用的投票策略，即一个结果必须至少有两个分类器同意才行。

4.1.3 求解

4.1.3.1 特征描述

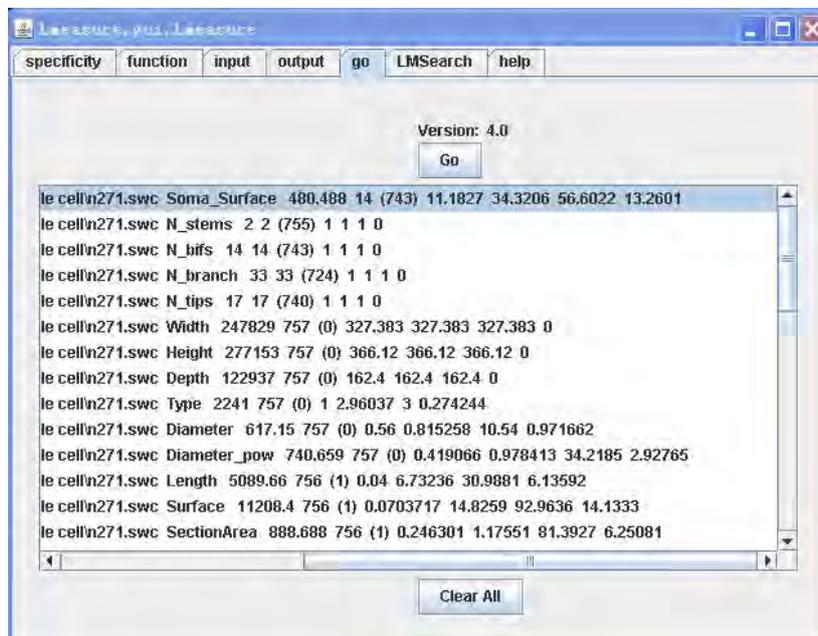


图 4-9 L-Measure 计算过程

通过 Matlab 和 L-Measure 工具的共同作用，对每一个神经元，我们求出了 43 维特征，组成特征向量。Matlab 求解可以方便转换为模型输入的格式，

L-Measure 则相对比较繁琐，只有求解比较困难的值借助于 L-Measure 可以轻松求解。如下图所示的神经元细胞：

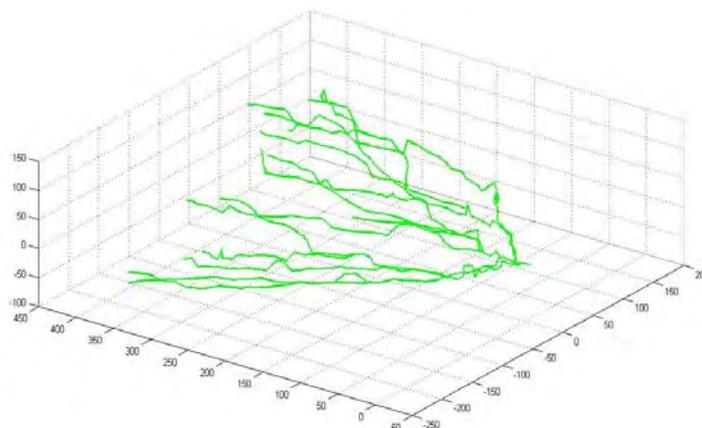


图 4-10 某一神经元

计算之后的 43 维特征如表 4-1 所示，可能某些值与真实值有一定的误差，但是对每一个神经元来说，误差相对的，对分类结果影响较小，可以忽略。

表 4-1 特征描述

id	特征	值	id	特征	值
1	Soma_Surface	480.488	23	Branch_pathlength	5080.16
2	N_stems	2	24	Contraction	26.1203
3	N_bifs	14	25	Fragmentation	743
4	N_branch	33	26	Daughter_Ratio	22.4009
5	N_tips	17	27	Parent_Daughter_Ratio	24.1907
6	Width	327.383	28	Partition_asymmetry	5.21905
7	Height	366.12	29	Rall_Power	10.98
8	Depth	162.4	30	Pk	16.9688
9	Type	2241	31	Pk_classic	18.4519
10	Diameter	617.15	32	Pk_2	16.8196
11	Diameter_pow	740.659	33	Bif_ampl_local	1391.41
12	Length	5089.66	34	Bif_ampl_remote	678.149
13	Surface	11208.4	35	Bif_tilt_local	1275.11
14	SectionArea	888.688	36	Bif_tilt_remote	2042.59
15	Volume	2879.85	37	Bif_torque_local	1512.03
16	EucDistance	174687	38	Bif_torque_remote	1438.45
17	PathDistance	229334	39	Last_parent_diam	6.18
18	Branch_Order	1963	40	Diam_threshold	19.9
19	Terminal_degree	1358	41	HillmanThreshold	5.145
20	TerminalSegment	473	42	Helix	3.04216
21	Taper_1	-0.31525	43	Fractal_Dim	28.841
22	Taper_2	8.60489			

4.1.3.2 差分特征选择结果

对原始数据 A 和 C 合成训练集之后，进行差分特征选择分析，对于原始的 5 类数据，先求质心，再进行差分，下图为原始差分矩阵，即没有排序前的。横轴代表 43 维特征向量，纵轴代表 $5 \times 4 = 20$ 个差分向量。白色表示差值大，颜色越深表示差值越小。

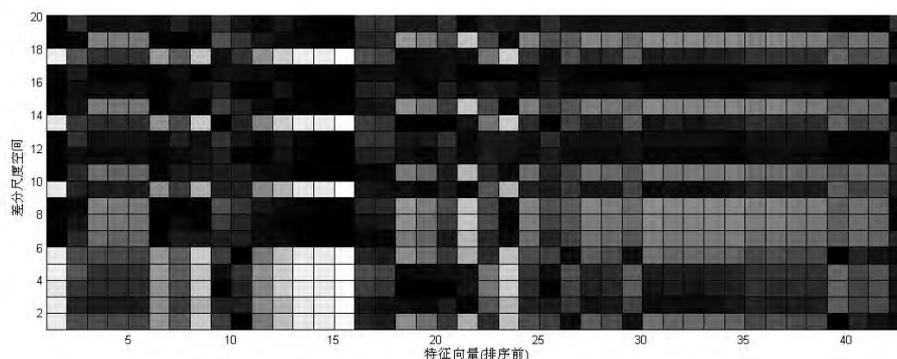


图 4-11 差分矩阵（排序前）

为了便于观察和理解，我们对差分矩阵进行排序，即对差值大小进行了排序

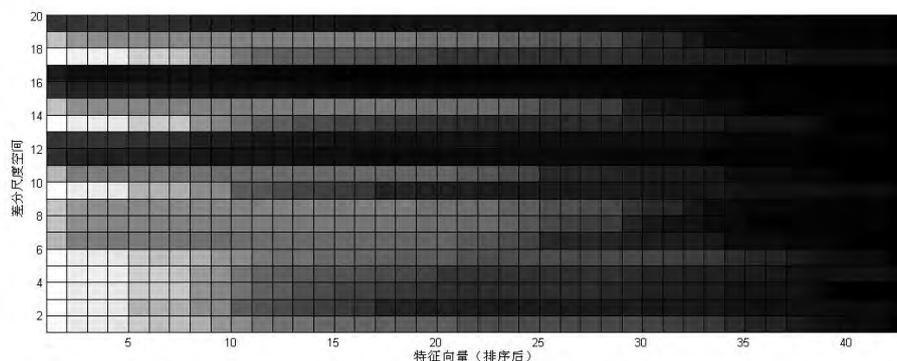


图 4-12 差分矩阵（排序后）

从上图可以看出在左下角的差值十分明显，前四行表示的数据是第一类和其他四类的特征向量的特征项内部差值，有些特征项的判别作用十分明显，像排序后排在第一位的那个特征项对第一类的判别能力，几乎可以将与其他四类分开。

另外我们发现第二类和其他几类也有较好的特征项判别能力。但是对于类别 3、4、5，这里的判别特征项体现的并不明显，我们怀疑是训练集数据太少引起的，经过扩充训练集，结果还是如此，所以我们认为需要单独对类别 3、4、5 进行差分选择特征，即需要进入递归。

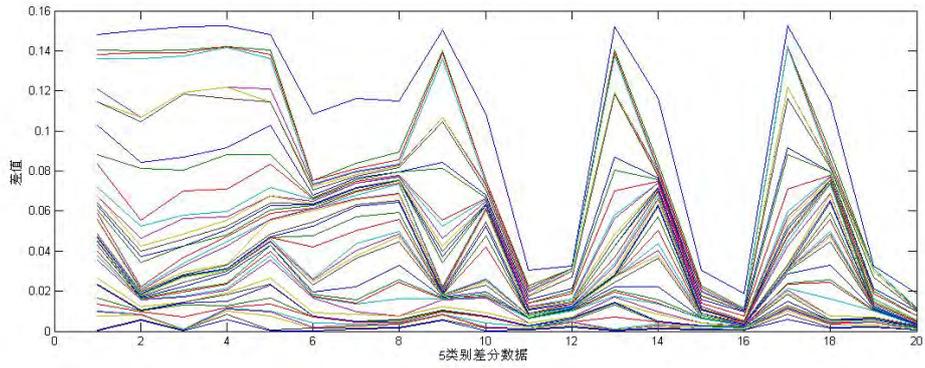


图 4-13 差分阈值的选择 δ_{center}

对于判别第一类和第二类，我们选择 $\delta_{center} = 0.04$ ，对于判别第三类第四类和第五类选择 $\delta_{center} = 0.02$ 。则 5 类的一共进行了两次差分特征选择：
第一次差分选择选出第一类和第二类的，图中白色部分为选择的特征，可见几乎没有交集。

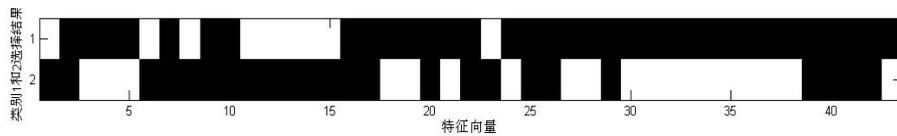


图 4-14 第一二类的特征选择结果

第二次差分选择选出第三四五类的，去除第一类和第二类，在这三类当中，所选择的特征判别能力也较强。

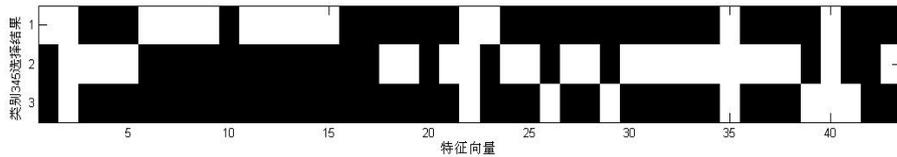


图 4-15 类别 3、4、5 特征选择的结果

对于第四类感觉神经元，根据两级三级多级的定义，可以根据突触数目决定类别：

$$c_i = \begin{cases} \text{单极}, f_2 = 1 \\ \text{双极}, f_2 = 2 \\ \text{三极}, f_2 = 3 \\ \text{多极}, f_2 > 4 \end{cases} \quad (7)$$

其中， $f_2 = N_stems$ 。

但是我们想发现在突触上有差别之外，其他还有什么差别，则可以继续运行差分特征选择算法：

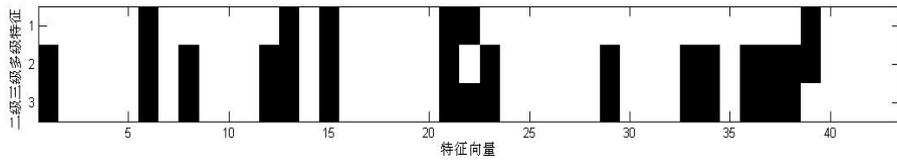


图 4-16 中间神经元内部小类特征选择的结果

这三种中间神经元，二级二级神经元可以和三级和多级分开，但是三级和多级实在太相似了。难以区分，所以还是对这类划分到中间神经元进行严格按照定义划分，如式（7）

到目前为止，我们可以根据差分特征选择的结果，对五类神经元进行特征的描述了。我们用类的质心来表示各类别神经元的几何特征。各个几何特征的平均数据，其中黑底白字的为选择出来的判别特征

(1) 运动神经元

胞体以及突触都比较大，空间体积较大。

表 4-2 运动神经元几何特征

id	特征	值	id	特征	值
1	Soma_Surface	34187.10	23	Branch_pathlength	5080.16
2	N_stems	11.50	24	Contraction	26.1203
3	N_bifs	158.33	25	Fragmentation	743
4	N_branch	338.67	26	Daughter_Ratio	22.4009
5	N_tips	169.83	27	Parent_Daughter_Ratio	24.1907
6	Width	1933.50	28	Partition_asymmetry	5.21905
7	Height	1659.00	29	Rall_Power	10.98
8	Depth	1747.58	30	Pk	16.9688
9	Type	2450.00	31	Pk_classic	18.4519
10	Diameter	2623.38	32	Pk_2	16.8196
11	Diameter_pow	13537.97	33	Bif_ampl_local	1391.41
12	Length	102694.48	34	Bif_ampl_remote	678.149
13	Surface	562276.00	35	Bif_tilt_local	1275.11
14	SectionArea	82863.42	36	Bif_tilt_remote	2042.59
15	Volume	388733.00	37	Bif_torque_local	1512.03
16	EucDistance	598365.67	38	Bif_torque_remote	1438.45
17	PathDistance	711161.33	39	Last_parent_diam	6.18
18	Branch_Order	3382.17	40	Diam_threshold	19.9
19	Terminal_degree	2757.83	41	HillmanThreshold	5.145
20	TerminalSegment	549.17	42	Helix	3.04216
21	Taper_1	-2.17	43	Fractal_Dim	28.841
22	Taper_2	162.20			

(2) 普肯野神经元

表 4-3 普肯野神经元几何特征

id	特征	值	id	特征	值
1	Soma_Surface	460.27	23	Branch_pathlength	6899.02
2	N_stems	1.14	24	Contraction	737.83
3	N_bifs	388.14	25	Fragmentation	3190.00
4	N_branch	778.57	26	Daughter_Ratio	546.34
5	N_tips	389.43	27	Parent_Daughter_Ratio	678.46
6	Width	170.87	28	Partition_asymmetry	200.06
7	Height	170.88	29	Rall_Power	261.59
8	Depth	13.43	30	Pk	698.90
9	Type	9592.57	31	Pk_classic	655.99
10	Diameter	2611.17	32	Pk_2	650.32
11	Diameter_pow	2818.77	33	Bif_ampl_local	31229.83
12	Length	6899.02	34	Bif_ampl_remote	27224.24
13	Surface	25605.59	35	Bif_tilt_local	42657.03
14	SectionArea	2648.58	36	Bif_tilt_remote	45912.71
15	Volume	11786.10	37	Bif_torque_local	28952.83
16	EucDistance	315441.71	38	Bif_torque_remote	29240.39
17	PathDistance	440448.00	39	Last_parent_diam	117.45
18	Branch_Order	40412.43	40	Diam_threshold	338.78
19	Terminal_degree	19425.00	41	HillmanThreshold	110.60
20	TerminalSegment	1945.71	42	Helix	0.98
21	Taper_1	-18.18	43	Fractal_Dim	368.81
22	Taper_2	103.16			

(3) 锥体神经元

表 4-4 锥体神经元几何特征

id	特征	值	id	特征	值
1	Soma_Surface	774.72	23	Branch_pathlength	13274.92
2	N_stems	6.63	24	Contraction	124.61
3	N_bifs	69.63	25	Fragmentation	1815.00
4	N_branch	151.50	26	Daughter_Ratio	146.20
5	N_tips	76.25	27	Parent_Daughter_Ratio	114.32
6	Width	249.88	28	Partition_asymmetry	36.49

7	Height	661.88	29	Rall_Power	59.43
8	Depth	250.69	30	Pk	99.91
9	Type	6300.13	31	Pk_classic	104.25
10	Diameter	1239.11	32	Pk_2	98.94
11	Diameter_pow	1406.45	33	Bif_ampl_local	4621.32
12	Length	13274.92	34	Bif_ampl_remote	3242.85
13	Surface	24854.06	35	Bif_tilt_local	7049.34
14	SectionArea	1787.04	36	Bif_tilt_remote	8658.04
15	Volume	6716.48	37	Bif_torque_local	5433.25
16	EucDistance	403198.88	38	Bif_torque_remote	5505.91
17	PathDistance	552107.50	39	Last_parent_diam	13.95
18	Branch_Order	7920.50	40	Diam_threshold	37.51
19	Terminal_degree	4212.13	41	HillmanThreshold	12.37
20	TerminalSegment	1281.50	42	Helix	1.93
21	Taper_1	-1.30	43	Fractal_Dim	124.52
22	Taper_2	45.94			

(4) 中间神经元

表 4-5 中间神经元几何特征

id	特征	值	id	特征	值
1	Soma_Surface	1171.35	23	Branch_pathlength	2846.41
2	N_stems	4.55	24	Contraction	50.61
3	N_bifs	26.68	25	Fragmentation	774.50
4	N_branch	61.45	26	Daughter_Ratio	68.60
5	N_tips	31.23	27	Parent_Daughter_Ratio	53.27
6	Width	197.88	28	Partition_asymmetry	16.28
7	Height	428.37	29	Rall_Power	5.91
8	Depth	50.48	30	Pk	51.80
9	Type	2467.68	31	Pk_classic	51.84
10	Diameter	1064.64	32	Pk_2	52.74
11	Diameter_pow	1543.51	33	Bif_ampl_local	2219.97
12	Length	2846.41	34	Bif_ampl_remote	1965.69
13	Surface	7873.82	35	Bif_tilt_local	2718.57
14	SectionArea	2120.30	36	Bif_tilt_remote	2563.46
15	Volume	2818.20	37	Bif_torque_local	2109.72
16	EucDistance	142803.17	38	Bif_torque_remote	2177.31
17	PathDistance	188389.49	39	Last_parent_diam	9.04
18	Branch_Order	3175.59	40	Diam_threshold	26.64
19	Terminal_degree	2557.27	41	HillmanThreshold	8.78
20	TerminalSegment	514.14	42	Helix	-0.39

21	Taper_1	-0.10	43	Fractal_Dim	43.64
22	Taper_2	6.02			

(5) 感觉神经元

表 4-6 感觉神经元几何特征

id	特征	值	id	特征	值
1	Soma_Surface	0.00	23	Branch_pathlength	562.13
2	N_stems	0.00	24	Contraction	19.06
3	N_bifs	11.25	25	Fragmentation	182.00
4	N_branch	23.63	26	Daughter_Ratio	16.75
5	N_tips	12.25	27	Parent_Daughter_Ratio	19.01
6	Width	104.71	28	Partition_asymmetry	5.87
7	Height	125.34	29	Rall_Power	10.05
8	Depth	79.53	30	Pk	18.24
9	Type	397.75	31	Pk_classic	18.03
10	Diameter	296.43	32	Pk_2	17.67
11	Diameter_pow	393.61	33	Bif_ampl_local	982.06
12	Length	562.13	34	Bif_ampl_remote	945.39
13	Surface	2762.20	35	Bif_tilt_local	1007.21
14	SectionArea	420.93	36	Bif_tilt_remote	911.37
15	Volume	1331.28	37	Bif_torque_local	905.95
16	EucDistance	17973.40	38	Bif_torque_remote	882.52
17	PathDistance	26256.64	39	Last_parent_diam	5.57
18	Branch_Order	822.00	40	Diam_threshold	15.25
19	Terminal_degree	644.00	41	HillmanThreshold	5.09
20	TerminalSegment	97.38	42	Helix	-0.09
21	Taper_1	-0.29	43	Fractal_Dim	15.47
22	Taper_2	3.30			

4.1.3.3 分类问题

重点分析一下差分特征选择对 KNN 分类的影响。我们对每一个小类选了 20 个神经元作为选连数据，一共 $7 \times 20 = 140$ 个神经元的训练数据。

(1) KNN 分类

- 差分特征选择之前

我们自己寻找了 10 个测试神经元进行测试，特征选择之前的正确率为 80%。有两个椎体分到普肯野里面了。

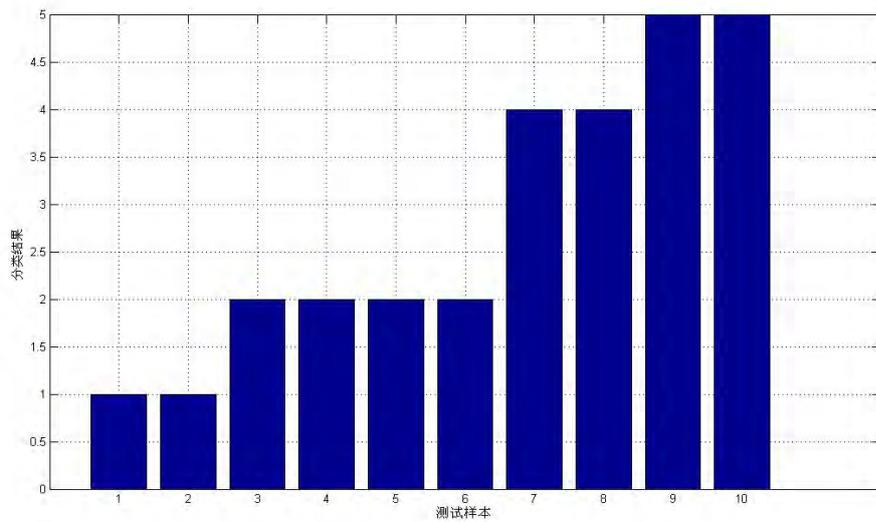


图 4-17 差分之前 KNN 结果

- 差分特征选择之后
由于增强了特征之间的判别性以及采用层次分类的策略，分类效果大有改善。

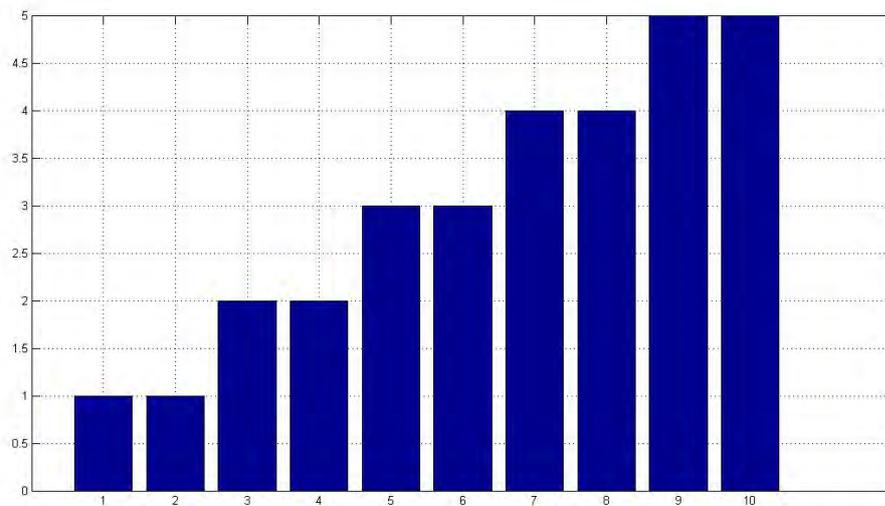


图 4-18 差分之后 KNN 结果

(2) 决策树分类

由于决策树能够对特征进行选择，所以分类效果一般比较好，分类正确率达到 80%

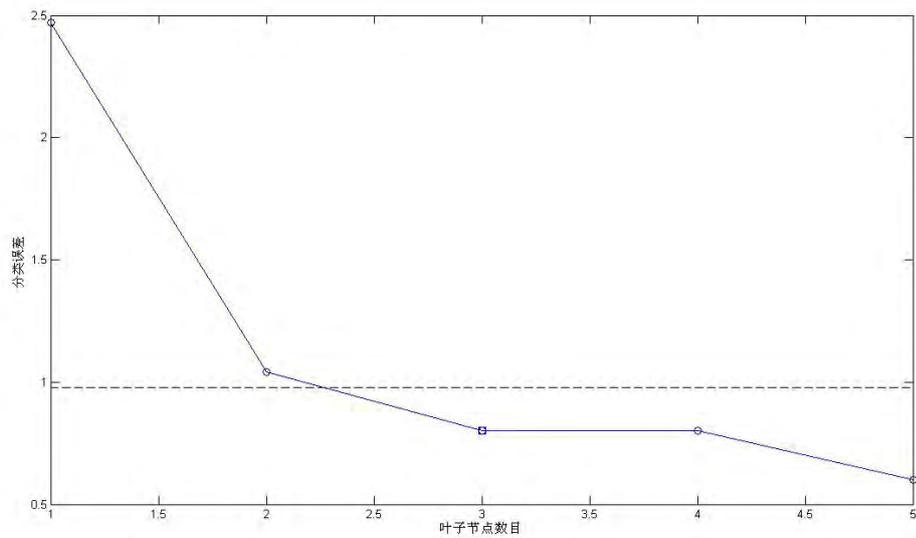


图 4-19 决策树误差分析

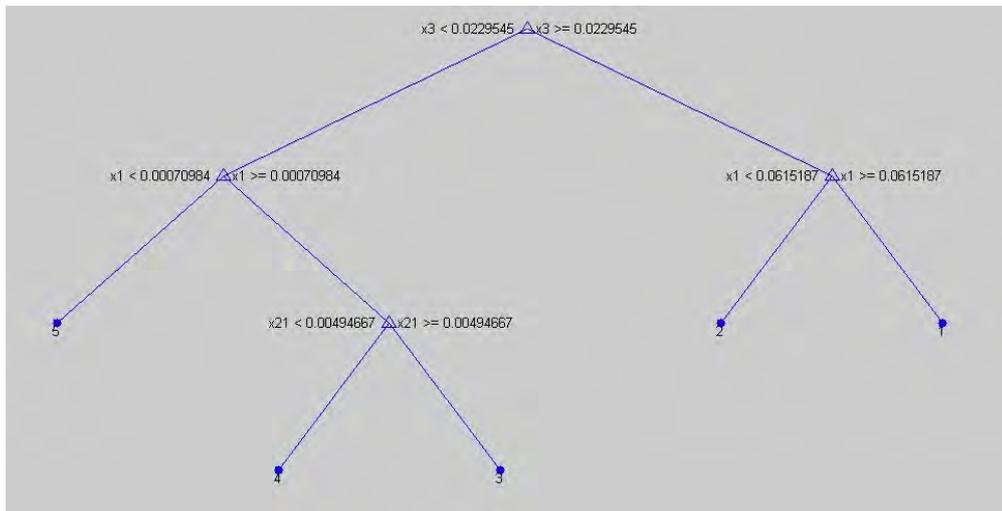


图 4-20 训练出的决策树

从决策树上也能看出在胞体比较大的时候最可能是运动神经元，胞体一直为 0 很可能也是感觉神经元。

(3) SVM 分类

用的 libsvm，RBF 核的五类分类问题，正确率为 90%，要比差分特征选择之前的 KNN 好很多。

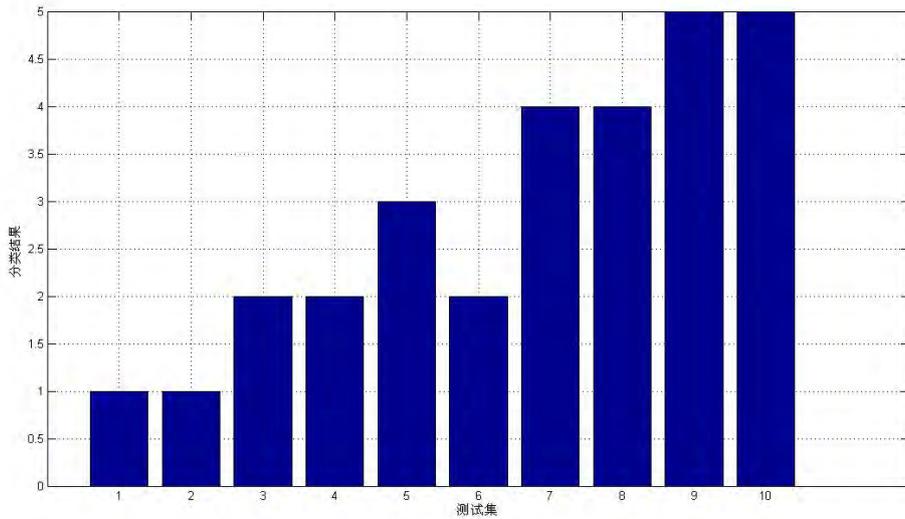


图 4-21 SVM 测试结果

(4) 分类器集成

显然，集成之后的对于测试集的分类效果为 100%。说明多分类器集成能够互补单个分类器的缺点，已达到更有的目的，虽然我们的测试集有限，但是多分类器的鲁棒性完全可以适应大数据的压力，有良好的泛化能力。

4.2 问题二

4.2.1 问题分析

附件 B 中是一些没有类别的需要进行分类的数据，那么我们完全可以用问题一的模型进行分析了，首先进行特征提取，建立特征向量，然后用多分类器系统进行分类。

由于可能要面对新类型的神经元，为了在不错分的情况下将新神经元检测出来并且能够观察到其离哪一类神经元最近，我们必须引入新模型和变量。

4.2.2 模型建立

这里，我们引入一个新的变量 ξ_i ，我们称之为类别 C_i 的边界，这个变量通过训练集中的数据和之前的质心 F_{center_i} 得到：

$$\xi_i = \max_{N_i \in C_i} \{d(F_{N_i}, F_{center_i})\} \quad (8)$$

在对 B 中的神经元分类过程中，如果神经元 N_i 要被分到 C_i 的时候，计算其到 C_i 质心 F_{center_i} 的距离：

$$d_{NC_i} = d(F_{N_i}, F_{center_i}) \quad (9)$$

根据 d_{NC_i} 和 ξ_i 的关系，对神经元 N_i 进行类别判定：

$$c_i = \begin{cases} i, & \text{if } d_{NC_i} \leq \xi_i \\ -i, & \text{if } d_{NC_i} > \xi_i \end{cases} \quad (10)$$

然后可以根据结果对神经元分析验证以及重新命名。

4.2.3 求解

先看看多分类器系统分类的结果，这里我们为了描述方便，还是按照每个分类器进行描述。重点说明一下我们的差分 KNN 和 SVM 的结果。

用差分 KNN 对测试集 20 种未知神经元测试结果如下：

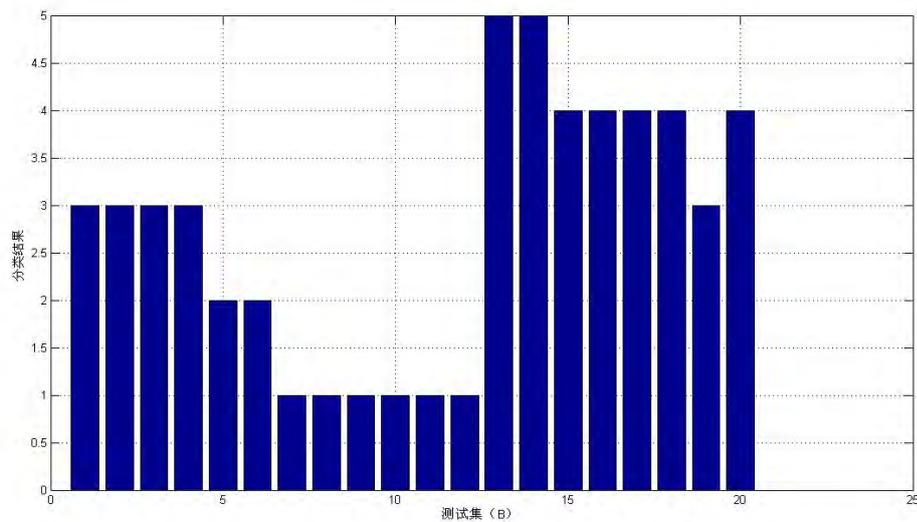


图 4-22 差分 KNN 分 5 类结果

分析一下结果，前 4 种和 19 都是锥体神经元，5、6 属于普肯野神经元，7-12 都是运动神经元，13、14 都是感觉神经元，15-18、20 都是中间神经元。再查看一下 SVM 的结果：

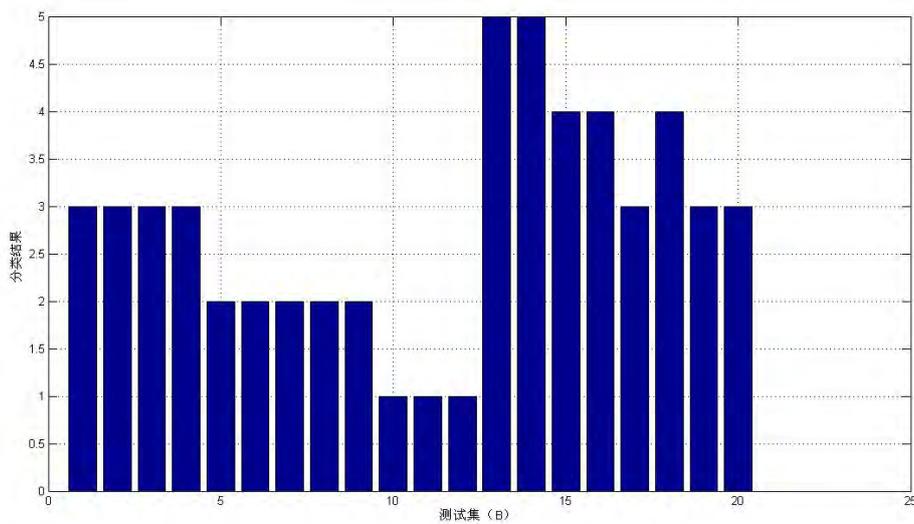


图 4-23 SVM 分 5 类结果

和差分 KNN 的结果基本吻合，差别在于 7、8、9 神经元和 17、20 这五个神经元，这里也说明一个问题，这几个神经元可能是没有类别的神经元。将在下文分析。

如果想知道中间神经元内部的分类情况，可以查看差分 KNN 分七类的结果：

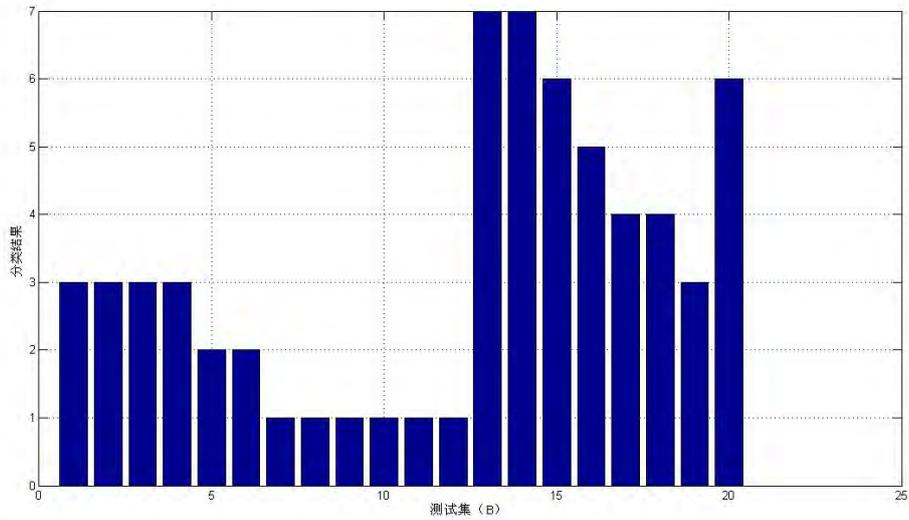


图 4-24 差分 KNN 分 7 类结果

很明显，17、18 为二极中间神经元，16 为三极中间神经元，15、20 为多极中间神经元。

好了，下面我们用新的模型分析一下是否有新的神经元类别，用差分 KNN5 类和 7 类的结果分别如下：

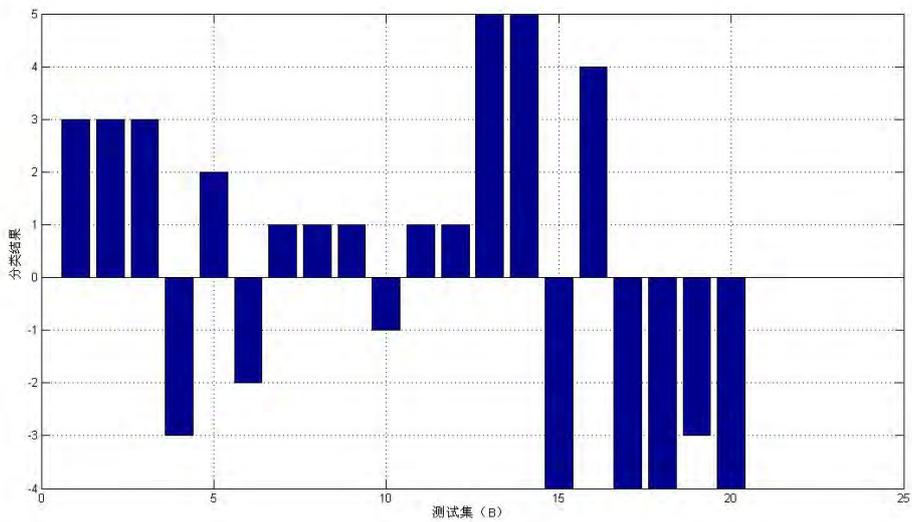


图 4-25 新神经元验证差分 KNN 分 5 类结果

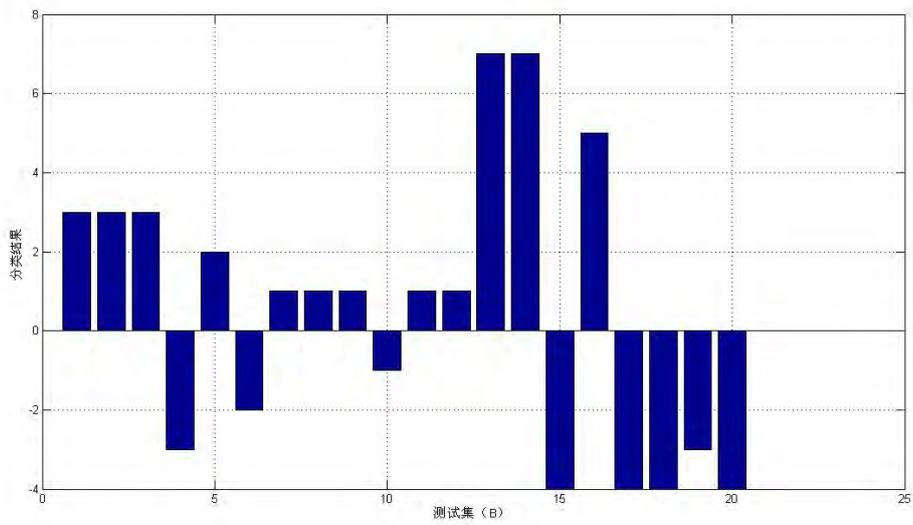
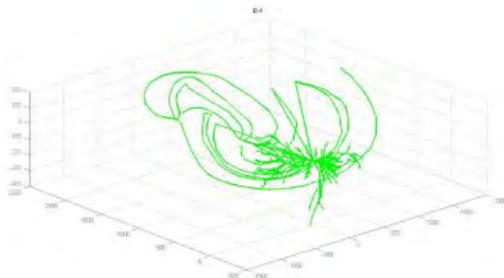


图 4-26 新神经元验证差分 KNN 分 7 类结果

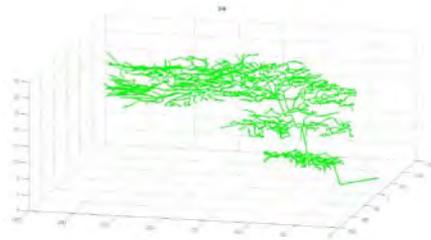
分析上述两个分类结果，其中第 4、6、10、15、17-20 发现其边界距离并不在所归属的类别之内，说明有因素影响了某些特征，而不是全部的特征。为了验证这种猜想，下面我们将进行仿真验证。

4.2.4 验证

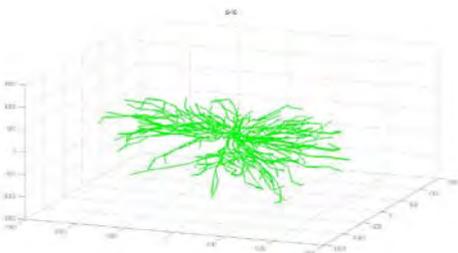
通过 Matlab 根据 SWC 格式的空间形态数据还原测试集中某些神经元的可视化形态，以方便进行分析和命名。



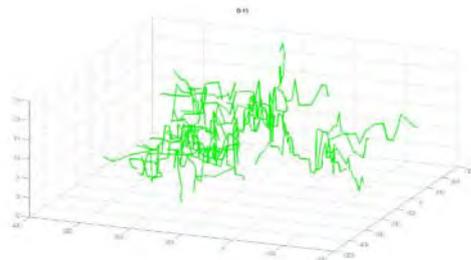
B-4



B-6



B-10



B-15

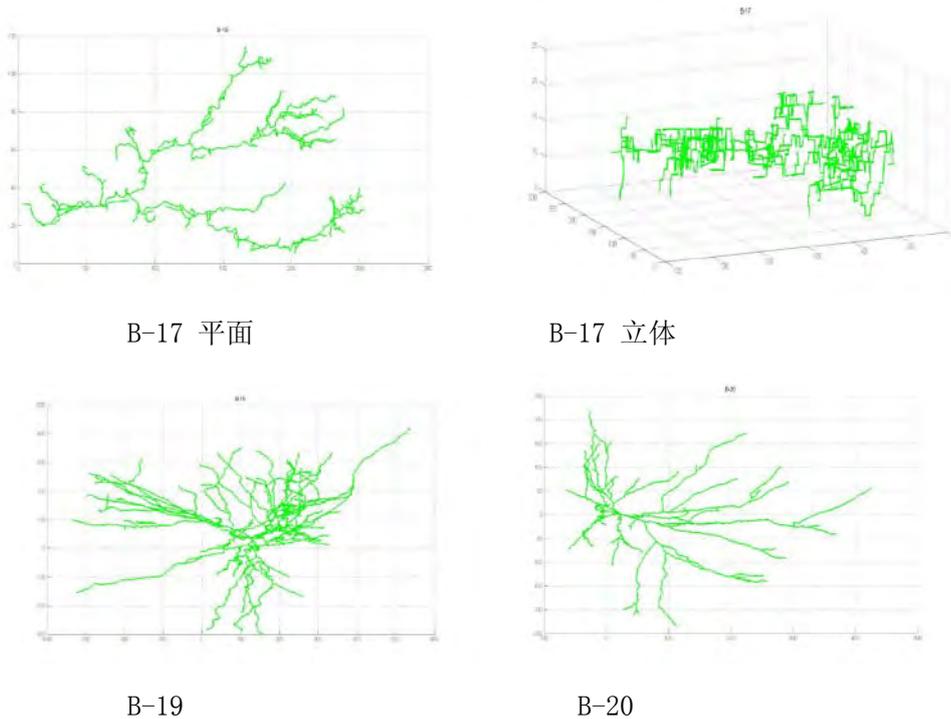


图 4-27 奇异神经元可视化分析

过可视化数据，可以很容易看出这些神经元的奇特之处，B-4 神经元的确属于锥体神经元的一种，但是由于其突出有的长度十分大，所以被发现。这类神经元是一类正常的神经元，是鼠的一个区的神经元，因为在训练集中没有此神经元，所以被检测出来。以后可以加入训练集，或者定义新类。

B-6 也是也普肯野神经元很相似，但是其并不是平面的，而是螺旋上涨，所以出现了奇异，可以归结为普肯野，甚至在普肯野里分小类。

B-10 感觉是很平常的运动神经元，但是由于其某些数据过于稠密，所以被检测出来。

B-15 的神经元和中间神经元类似，但是其在立体空间的突触是震动生长的，所以分出来。

B-17 和 B-18 其实是一类神经元，都没有胞体和杆，经查阅资料其实是一种轴突末端的爬行纤维，对于这一类，可以单独命名了，因为其在空间也是震荡爬行的，所以可称为爬行纤维。

对于 B-19 和 B-20，B-19 被分到了锥体神经元里，因为其形状确实像锥体，但是他又有自己的特点，而且和 B-20 有点类似。其实，这些神经元一般都会是中间神经元，中间神经元数量十分庞大，一般占总数的 90% 左右。所以中间神经元的可能性大点。

在差分 KNN 和 SVM 分类问题中，出现了不稳定的神经元，还没有分析过的是 7、8、9，经过观察数据，发现其胞体和茎的数量都是 0，在其他训练数据中只有感觉神经元是这样的，但是 7、8、9 的神经元形态又类似于运动神经元，所以产生了错分问题。对于 7、8、9 这类神经元，其实得定义新的神经元。由于其胞体和茎都是 0，可以用突触终端、无胞突触等词汇进行启发式描述。

4.3 问题三

4.3.1 问题分析

这个问题其实在考虑问题一和二模型的泛化能力以及启发式。

4.3.2 模型建立

要做到对所有神经元进行分类，是十分困难的事情。通过优化问题一二的模型可以达到泛化的目的。

- (1) 增加训练数据，尽量包含多的神经元类型。
- (2) 提供反馈学习技术，如果模型分出的神经元是一种新类，则自动添加到训练集当中，实现自我生长。
- (3) 关于命名方式，一种是根据严格的定义进行命名，像是单极神经元、二极、多极神经元等；另外，针对一些特殊情况命名，像是当发现没有胞体和茎的时候，可以用突触末端、无胞体突触等词汇进行命名；另外，就是根据最相近的类别进行类比命名了，例如：B-4，类锥体神经元。

4.3.3 求解与验证

我们又将 B 中新数据加入到我们的训练集当中，又从互联网数据库中找到了一些神经元，加入到我们的数据库中，不知名的神经元我们给命名。很明显，如果再对 B 进行分类，结果将会大大改善，准确率近乎 100%：

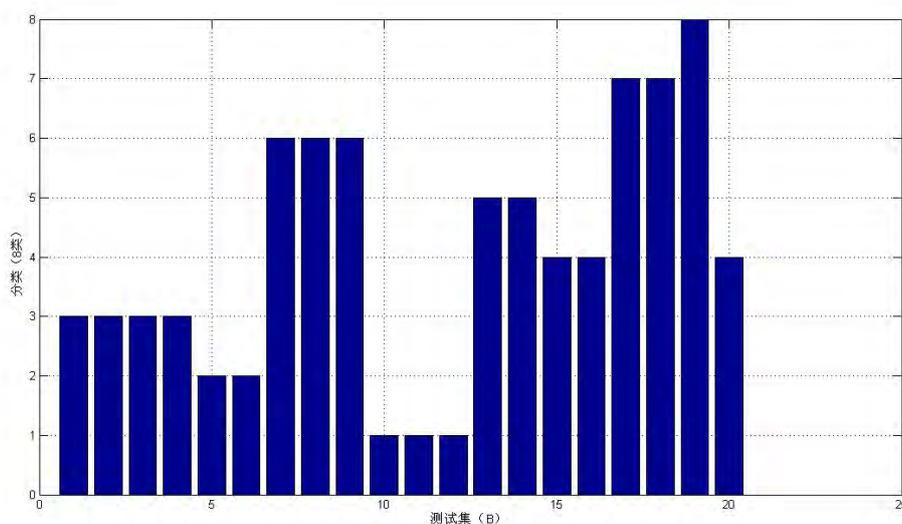


图 4-28 训练集添加新类后分类结果

4.4 问题四

4.4.1 问题分析

由于我们提出了差分特征选择模型，所以很显然能够区分任何两个神经元，即使他们是一种，只要有一点不同，差分特征选择模型总能找到不同。我们一共给出了三对数据进行说明，分别是猪鼠普肯野、猫鼠运动神经元、人猫锥体神经元。

4.4.2 模型建立

问题一的差分特征提取模型，这里差分类别是 2

4.4.3 求解与验证

(1) 猪鼠普肯野神经元

我们寻找到差分参数 $\delta_{center}=0.1$ ，然后分析出能够区分猪鼠普肯野神经元的

15 种特征，其特征 id 是：1、8、9、13、15、18、20、21、22、25、29、39、40、41、43。

1. Soma_Surface 细胞体表面积

8. Depth 深度

9. Type 房室种类 (1 细胞体 2 轴突 3 树突 4 基础树突)

13. Surface 房室表面积 (球)

15. Volume 房室体积 (圆柱体)

18. Branch_Order 分支阶数 (从细胞体到分支的分支数目)

20. TerminalSegment 以叶子节点分支的房室的数目

21. Taper_1 实际直径-上一个房室直径/分支长度

22. Taper_2 最终直径和起始直径的比率

25. Fragmentation 返回在一个分支内的房室的数目

29. Rall_Power 满足等式 (子房室幂次方和=父房室的幂次方) 的最佳值

39. Last_parent_diam 返回叶子节点的上一个父亲节点的直径

40. Diam_threshold 返回指向叶子节点的最后一个分支的第一个房室的直径

41. HillmanThreshold 当前分支的父亲房室 $\times 0.5$ + 两个子节点 $\times 0.25$

43. Fractal_Dim

针对选择的特征分析真实数据，然后映射到具体几何特征的含义，通过具体特征意义分析，这两种差距还是在大小上。

表 4-7 猪鼠普肯野特征对比

特征 id	Pig1	Pig2	Pig3	Rat1	Rat3	Rat3
1	0	0	0	799.425	989.742	921.607
2	1	1	1	1	2	1
3	471	419	416	369	357	342
4	944	840	835	740	717	686
5	472	420	418	370	359	343
6	220.5	245.5	211	137.5	113.71	156.33
7	236.5	209.5	211.5	145.1	146.42	156.99
8	23	18	19.5	2	6.5	18.5
9	6196	4561	4142	15103	17176	7774
10	2904.8	1849.5	2203.5	2471.69	3042.92	3214.65
11	3821.54	2372.74	3025.68	1982.61	2503.58	3728.41
12	10890.1	8413.2	8144.81	4980.15	5497.91	5905.34
13	50877	34666.2	42520	8336.16	9805.41	23086.4

14	4381.27	2736.87	3500.83	1300.44	1672.5	3339.31
15	27291.9	19128.8	22380.8	1742.46	2032.62	7776.34
16	316589	216209	179542	432832	492978	281948
17	448393	318672	249073	593109	693225	358688
18	29036	19171	18231	64287	68377	33878
19	15797	11017	13191	27400	26611	18856
20	1139	785	644	3129	3623	1609
21	-32.0961	-32.8631	-21.6556	-13.1448	-15.1787	-4.66804
22	163.894	198.945	73.0097	122.597	97.1222	14.6049
23	10890.1	8413.2	8144.81	4980.15	5497.91	5905.34
24	908.157	804.077	805.638	696.65	667.143	639.801
25	2060	1517	1376	5026	5715	2583
26	607.768	555.974	511.383	545.203	595.767	434.817
27	784.344	653.421	754.313	638.892	616.748	657.902
28	238.185	207.86	203.654	192.565	190.821	181.787
29	512.105	489.385	216.985	229.12	252.16	56.25
30	664.379	524.026	717.832	598.93	657.939	1012.63
31	733.665	593.07	732.97	614.018	610.199	663.678
32	696.633	546.867	720.786	601.035	635.339	694.861
33	37319.1	33907.4	32679.7	32753.9	29496.2	25155.9
34	32103	31068.1	30281.4	28503	24513.5	21799.9
35	55511.8	48673.8	47359.2	38735.2	37792.3	35520.3
36	57885.5	49110.6	47540.8	42332.1	42763.5	40196.8
37	32341.8	27721.6	26949.7	30482.7	30154.6	26189.7
38	32794.4	28934.9	26441.2	30764.2	29920.3	26740.8
39	190.3	139.5	183.6	51.42	57.688	127.876
40	528.8	353.7	553.2	146.772	163.401	403.158
41	173.8	122.8	180.5	47.6999	53.1474	125.706
42	0.434337	-0.73693	2.80783	0.300629	4.23698	-1.48616
43	260.273	154.027	110.509	611.461	591.935	399.306

(2) 猫鼠运动神经元

我们寻找到差分参数 $\delta_{center}=0.2$ ，然后分析出能够区分猫鼠运动神经元的 12 种特征，其特征 id 是：10、11、13、15、16、17、31、32、37、38、39、41。

针对选择的特征分析真实数据，映射到真实数据分析差别在于一些角度、房室以及直径等。

表 4-8 猫鼠运动神经元特征对比（猫 3 列|鼠三列）

1	4032	4972.89	5703.7	6977.05	2023.63	2386.04
2	9	7	8	7	6	9
3	26	4	21	47	56	47

4	69	21	57	107	123	111
5	35	11	29	54	62	56
6	1458.5	1196	1049.5	373.58	358.08	323.03
7	1427.5	1510	1558	330.69	660.71	422.12
8	774.5	388.5	185.5	216.29	303.19	268.518
9	31815	6740	10786	6872	5701	5203
10	12366.6	4126.88	5821.54	972.298	961.982	921.988
11	15849.5	7417.14	9680.07	2475.02	1389.76	1787.39
12	37575.7	10498.8	14160.1	6849.98	7939.67	8987.46
13	133370	60208.6	73655	8005.85	10678.3	11931.8
14	20896.9	13956.9	18024.8	9152.68	2703.24	4759.35
15	51293.5	42709.4	47259	2676.25	3132.42	4006.15
16	5094410	1283470	1513160	321974	364572	312432
17	9395580	2250060	2106080	501911	523180	541339
18	20480	1608	7627	6066	6322	4482
19	15498	2728	4817	4884	4771	3927
20	8027	1822	2977	1285	1229	1232
21	-0.14272	-0.0049	-0.43822	-0.516	-0.94829	-0.40711
22	40.04	12.2012	27.7347	40.07	53.6567	31.1494
23	37575.7	10498.8	14160.1	6849.98	7939.67	8987.46
24	35.1587	9.98715	37.5698	73.5355	97.4233	76.6071
25	9885	1324	3295	2163	1902	1393
26	103.612	55.4796	72.7489	140.435	145.68	200.048
27	34.1785	10.6604	28.0312	78.4679	83.9587	89.5913
28	16.9189	5.88889	12.7032	24.7744	28.4361	26.8814
29	26.84	3.175	19.845	17.925	37.78	21.985
30	31.216	10.2424	26.4963	76.4659	74.616	88.3946
31	25.5889	8.01126	20.7734	72.7795	74.0682	88.4512
32	20.6605	7.14425	16.8673	70.8424	68.6922	93.159
33	1894.67	423.595	1438.18	3579.21	4247.41	3855.09
34	1537.48	146.315	1114.81	2419.82	2840.79	3069.66
35	1891.21	361.741	2055.19	3996.62	6048.14	4741.77
36	2462.34	160.369	2239.3	5458.71	7014.82	4540.05
37	1342.76	135.412	1408.17	3720.54	4544.03	3717.78
38	1450.59	106.956	1240.65	3360.52	4245.37	3683.51
39	27.2	10.2	75.42	5.22	6.93	4.25
40	42.5	34.9	60.3	10.75	15.6	16.45
41	17.5	6.825	45.11	4.075	5.4275	3.745
42	54.8397	7.9071	4.48047	3.56261	6.25115	9.20033
43	53.9694	14.6725	44.5752	81.9439	106.142	80.6384

(3) 人猫锥体神经元

我们寻找到差分参数 $\delta_{center} = 0.29$, 然后分析出能够区分人猫锥体神经元的 13 种特征, 其特征 id 是: 16、17、18、19、23、30、31、32、34、36、37、38。

针对选择的特征分析真实数据, 映射到真实数据分析, 主要是体现在距离和角度上, 看来大小相似, 形态相差很大, 我们仿真验证一下:

表 4-9 人猫锥体神经元特征对比 (人 3 列|猫 3 列)

1	1486.78	2360.83	1918.79	1337.93	938.033	1174.47
2	6	5	3	7	7	6
3	16	12	6	185	119	150
4	43	33	17	383	251	311
5	22	17	9	192	126	156
6	99.48	230.67	295.1	3259.39	2246.25	795.37
7	120.6	289.2	262.26	1851.59	1189.08	1179.8
8	50.55	94.69	110.29	1270.54	851.41	1454.15
9	3097	4199	2518	25052	19908	30497
10	1881.66	2176.88	1404.44	6100.93	2857.37	4694.92
11	3071.39	3469.47	2209.76	5677.43	2110.78	3329.93
12	2066.49	2937.51	1435.23	53348.4	29509.2	43264.9
13	13678.7	18315.5	9495.67	84539.1	28189.6	43075.9
14	4556.24	5215.59	2864.28	5564.2	2040.1	2774.19
15	9096.02	14165.2	6972.84	19832.2	2748.52	4819.51
16	88725.1	145185	92511.1	11538000	3690860	5607240
17	132033	244896	135696	17980600	7139940	10611300
18	2043	2469	1086	97838	53959	96437
19	1422	1613	888	70683	25633	44870
20	774	1057	652	5397	4881	8494
21	-0.50785	-0.6882	-0.23166	-2.42364	-0.49662	-0.97924
22	13.603	11.7194	8.41516	76.7891	20.9714	18.77
23	2066.49	2937.51	1435.23	53348.4	29509.2	43264.9
24	28.5674	20.6564	10.5227	327.601	201.567	258.951
25	804	1147	696	11405	8759	14189
26	49.9343	39.0204	20.3446	347.483	200.07	232.661
27	30.2753	22.4993	9.27192	320.323	234.299	285.322
28	11.2762	5.32857	3.14286	104.648	65.3692	75.3816
29	20.285	4.015	6.61	53.755	15.24	19.43
30	23.3225	20.3146	6.7718	312.289	231.236	282.411
31	26.1139	19.4471	6.64018	310.205	231.306	282.501
32	23.9185	18.2989	5.80517	311.227	231.414	287.034
33	1071.39	1021.79	578.806	14746.5	9451.47	11900.4

34	873.537	677.236	393.432	13038.6	7411.01	10009.4
35	1719.42	1005.47	529.148	18858.3	10331.5	13778.8
36	1714.9	1098.36	400.659	19458.9	13091.8	17004.3
37	1102.71	874.457	338.01	17199.9	9334.61	12524.7
38	1266.41	762.55	280.418	17063.1	9426.45	12479.9
39	12.6	15.8	7.73	32.58	15.18	16.78
40	38.8	35.46	16.11	82.98	43.01	51.05
41	11.0425	13.4875	6.0475	25.22	14.5475	15.635
42	3.92098	-8.06174	-1.55642	-5.20545	6.95298	13.1443
43	32.6422	24.7816	13.0896	273.34	212.999	277.253

4.5 问题五

4.5.1 问题分析

大量实验表明，神经元的实际形态并不是一成不变的。随着时间的变化，同一个神经元的轴突和树突都在发生着改变，或生长，或萎缩。神经元在干细胞定向分化之后，由单个细胞体开始成长过程。大多数神经元在定向分化之后就不分裂；如果它们受到损伤，损伤程度严重的就会凋亡，损伤程度较轻的则可能会得到修复；它们是长寿的细胞，在机体衰老时才开始逐渐衰亡；它们的长寿对于神经系统工作的连续性起到了保证作用。



图 4-29 小鼠海马区锥形神经元在定向分化后第 5 天，第 15 天，第 31 天，第 40 天，第 50 天和第 60 天的空间几何形态的图表示。从中可以看出神经元随着时间而发生的形态变化。

神经元的成长过程可以分为萌芽期 (Embryonic)，青年期 (Young)，成熟期 (Adult) 和衰老期 (Old)。各个时期对应于一定的细胞生长期限和形态发育

特征。不同物种，不同种类的神经元各个时期的划分也是不一样的。在实际实验中，观察一个特定神经元生长发育的过程是十分复杂和困难的，需要极为复杂的实验条件和高成本的准备工作。对于这种情况，为了能够模拟和预测在实际当中同一个神经元的生长过程，在问题五中我们通过建立仿真模型来解决问题。在仿真过程中，通过初步的分析，我们发现了在模拟生长问题中一些需要注意和解决的部分：

- 1) 如何理解一个神经元由开始的一个萌芽（一个细胞体）而不断生长，变化直至衰减的过程。
- 2) 如何在仿真模型中区别不同种类的神经元，并在生长预测过程中体现出这种区别。
- 3) 在每次的树突房室生长过程中，需要模拟和区分三种生长状态：分支态，生长态和停止态。
- 4) 在仿真过程中，树突和轴突的生长过程显然是不同的，在仿真过程中如何分别对神经元的树突和轴突建立生长模型是一个值得关注的问题。

神经元在最开始阶段都是单一的一个胞体，此时神经元中不存在树突和轴突结构；随着时间的变化，神经元的各个叶子节点不断的进行生长（elongation）和分支（branch），神经元的各项几何形态指标也随着这种时间相关的节点数量和形态的变化而变化。在题目中，对每个神经元已经离散为多个房室，因此神经元的生长过程同样可以离散化为房室的数量变化和各个房室的空间几何特征变化的过程。由此分析，解决神经元生长的模拟和预测问题，本质上可以简化为解决对于神经元树突和轴突生长和分支的模拟和预测问题。通过建立模型对这两种组成神经元的结构的生长变化进行模拟和仿真，可以实现对整个神经元生长状态的模拟，并进行生长状态的预测。

对于节点的生长（elongation）过程，即节点不进行分支，而进行自身空间几何形态变化的过程。经前所述，对于离散化的神经元各个房室呈现一种锥状体结构，在生长的过程中，各个叶子节点锥体的长度和宽度都会发生改变。在本题目中，我们已经假设各个房室的空间坐标位于锥体尖端，因此随着锥体长度的变化，叶子节点对应的房室的坐标也在发生着变化。同时，由于生长，叶子节点可能会与其父节点产生一个新的扭转角度。从而增加生长过程建模的复杂性。根据Burkel和Hillman的仿真实验结果，我们在模型建立之前假设在叶子节点生长过程中，叶子节点房室的半径和生长后的旋转角度，以及生长后的节点长度紧密相关。建立的相应的节点生长模型可以利用在时序性神经元训练数据集中的对应关系的数据点拟合相应的生长函数，从而完成对于神经元叶子节点生长过程的模拟并对未来的生长进行预测。

叶子节点通过分支（branch）过程生成两个新的节点，即在离散化的图中增加两个新的房室。对于这两个新建立的节点，原叶子节点被称为父亲节点（Father），而新建立节点被称为原节点的子节点（Daughter）。两个子节点之间呈一不固定的分支角度。同时，两个子节点所在平面和原节点所在平面也呈一非特定的扭转角度。近年来的大量文献已经说明了叶子节点的分支过程和节点直径有着紧密的联系。在模型的建立过程中，通过分析训练数据集中各个神经元在不同生长的时期的各个叶子节点直径与分支后各个集合形状参量，考虑建立叶子节点直径对于整个分支后的各个集合形状参量（分支角度，旋转角度）的函数对应关系。

在模拟神经元树突生长和分支的过程中，如何判定叶子节点是否分支，叶子节点是否生长以及叶子节点是否停止生长是必须要解决的三个核心问题。了解到树突最终是否分支和生长依然是一个随机问题，而通过实验中的训练数据和已经建立的分支和生长模型，我们可以求得在分支和生长过程中概率。由此这里可以将问题归结为随机采样问题来进行解决。采样中相关算子的参数可以在训练过程中获得。

不同种类的神经元呈现了各种截然不同的形态特征。如何由一个细胞体来描述和预测此后神经元的生长过程是一个困难的问题。因此针对不同种类的神经元，这里需要用相应的生长数据对模型中的相应参数进行训练，从而使得在模型仿真和预测每一个类神经元时模型都可以比较准确的反映其各个生长状态。可通过对不同种类的神经元单独进行建模，从而使得各个模型参数体现出对不同种类神经元的特异性。

神经元在不断生长，变化过程中随着各个房室数量方面以及在空间位置和直径的变化，使的本文在前面定义的几何形态特征也必然会随着神经元的生长而呈现出不同的对应值。为了检验预测生长数据的准确性，可通过前面提出的分类方法对预测数据进行分类并检测分类结果和实际是否一致。同时，通过对于一个神经元从定向分化后到一定时间的生长过程进行仿真，可获得在不同时间下仿真数据对应的几何形态特征。观察几何形态特征的变化，检测各个几何特征在不同时间的表现是否与同类神经元相符。

4.5.2 建立基于 Burke 算法的预测模型

4.5.2.1 模型中所涉及参数的求解步骤

首先我们来分析树突分支是否继续分叉。在这里我们利用小鼠海马区的普肯野神经元的数据作出分支长度随着分支初始直径变化的图像。

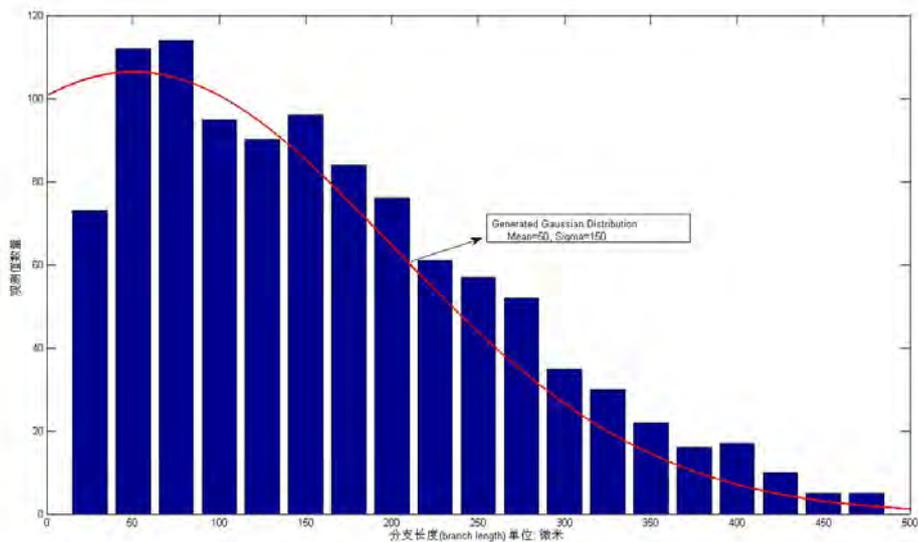


图 4-30 分支长度随着分支初始直径变化的图像

由图中可以分析出，树突的分支生长是与它的初始直径有着密切的关系的。同样的图形在[10]中也给出来了。为了进一步验证我们的结论，我们分别给出了分叉房室数目与终止房室数目随着直径的变化关系。根据 Burke 算法，我们有：
继续生长的概率

$$P_{nov}(d) = N_{nov}(d) / \sum l(d) \quad (11)$$

继续分叉的概率

$$P_{ov}(d) = N_{ov}(d) / \sum l(d) \quad (12)$$

终止生长的概率

$$P_{trm}(d) = N_{trm}(d) / \sum l(d) \quad (13)$$

所以可以继续得出这两种概率与半径的关系图(图形给出的是 log-log 图像, 因为数据都太小, 取对数只是为了观察方便)

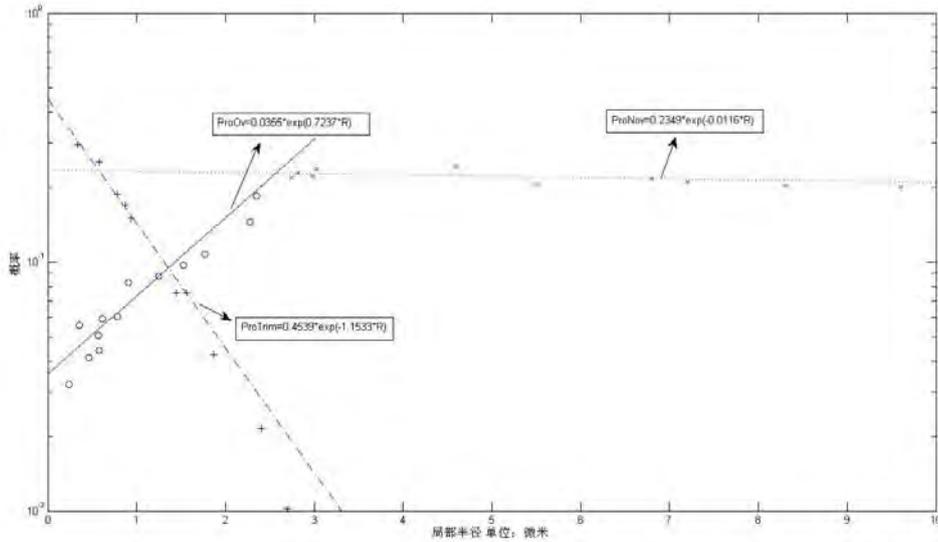


图 4-30 三种判定概率与局部（房室）半径之间的关系描述

根据图中散点我们采用拟合的办法, 可以得到这两种概率具有形式 $P = k_1 e^{k_2}$, 其中 k_1 , k_2 为来自拟合曲线的参数。

第二步, 利用随机的办法预测分支之间的夹角、分支的扭矩角度、两子分支的直径之比。由下图(来自真实细胞的这三者的分布图)可以观察到以上三个几何形态指标均服从高斯分布, 其均值和方差也是可以通过拟合的办法得到。具体分布由下图所示。

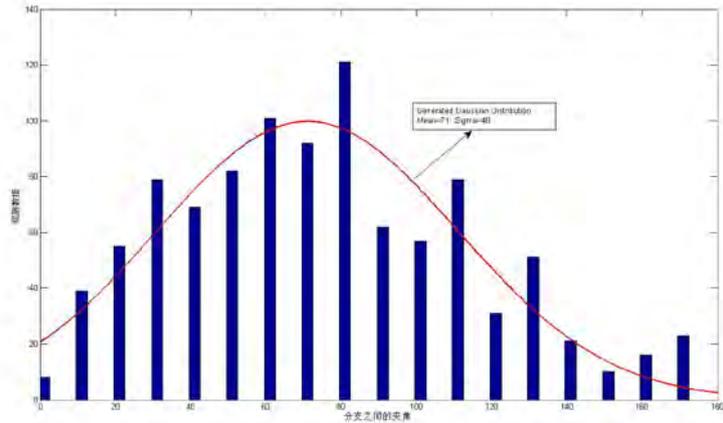


图 4-31：训练数据中分支夹角的分布情况

图 4-31 显示的是在训练数据中分支夹角的分布情况，我们使用一个高斯分布来对实验数据进行拟合，从图中可以观察到拟合后的高斯函数均值为 71 度，方差为 40 度。图 4-32 和图 4-33 则显示了扭矩角度和字分支直径比值的分布和拟合情况。具体求解结果我们在表 1 中进行了描述。

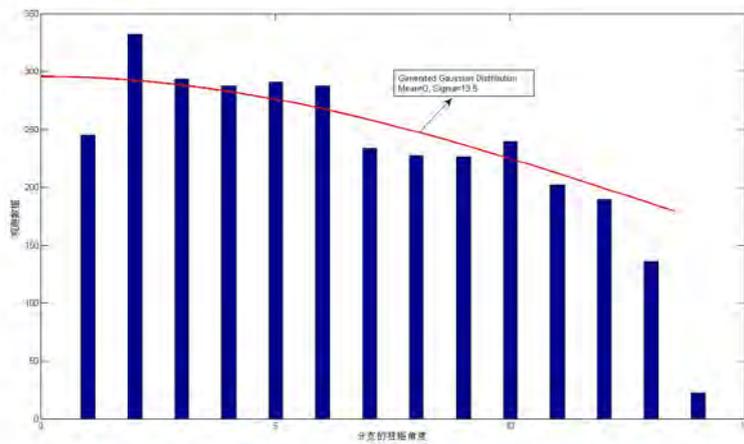


图 4-32：训练数据中扭矩角度的分布状况

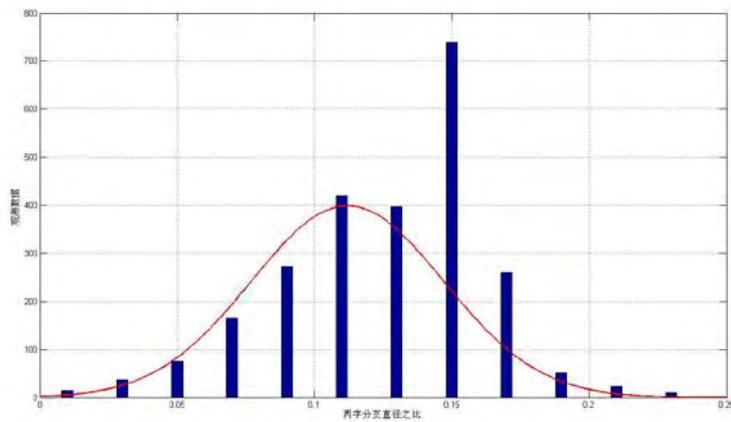


图 4-33：训练数据中分支直径比值的分布

4.5.2.2 模型流程图

有了以上参数变化，我们采用 Burke 模拟模型的流程图如下：

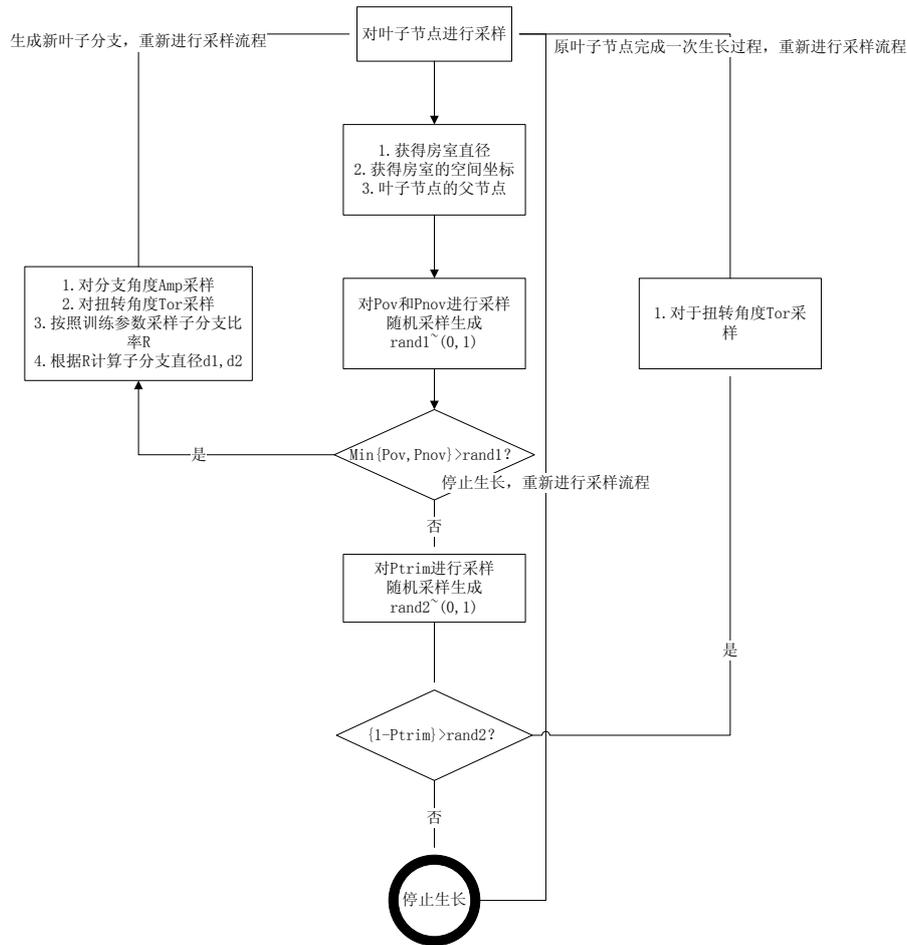


图 4-34: 生长预测模型的流程图

在算法的判定阶段，在 P_{ov} ， P_{nov} 随机采样生成的概率下模型随机采样了两个随机数 $rand1, rand2 \in \{0,1\}$ 。在第一次判别阶段，如果 $rand1 < 1 - \min\{P_{ov}, P_{nov}\}$ ，则当前叶子节点进行分支，否则进入第二次判别节点。在第二个判别点中，如果有 $rand2 < 1 - P_{trim}$ ，则分支进入单纯生长过程，否则停止生长。

由于在模型中需要考虑扭矩角度和分支角度的变换问题，因此本模型涉及了空间三维坐标系的变换过程，部分涉及到笛卡尔坐标的变换，详细变换过程见附录 1。

4.5.3 预测模型的求解

4.5.3.1 模型训练数据的调整

由于针对同一个神经元的重建生长的实验数据目前比较匮乏，因而为了能够充分训练模型，我们使用了在不同生长阶段的同类型的神经元数据。据前所述，在 <http://neuromorpho.org> 网站上给定的 .swc 文件格式中，由于分支前后父节点的半径不变，同时由于神经元叶子节点生长的概率要远小于分支的概率 (P_{ov}

与 $\min(P_{ov}, P_{nov}) * (1 - P_{trim})$)，因此我们在训练涉及分支的几何形态特征的时候，忽略了此神经元中分支节点生长的因素。这种简化虽然一方面加大了模型的误差，将分支和生长做为了两种相互独立的情况加以考虑，另一方面在简化建模过程的同时，给予了在模型参数训练过程中必须的足够的训练数据。

4.5.3.2 模型求解

根据上面的步骤，通过对训练集当中普肯野细胞的数据进行计算，我们得到了如下的参数结果。

表 4-10 参数结果表

参数缩写	参数意义	参数值
ProOv	不分支，继续生长的概率	$0.0355 * \exp(-0.7237d)$
ProNov	分支生长的概率	$0.2349 * \exp(-0.0116 * d)$
ProTrm	停止生长的概率	$0.4539 * \exp(-1.1533 * x)$
ProAmp	分支之间的夹角	71 ± 40 (高斯分布)
ProTor	分支的扭距角度	0 ± 13.5 (高斯分布)
R	两子分支直径之比	24.6% 概率为 0.136 75.4% 概率为 0.112 ± 0.035 (高斯分布)

模拟出的效果图如下：

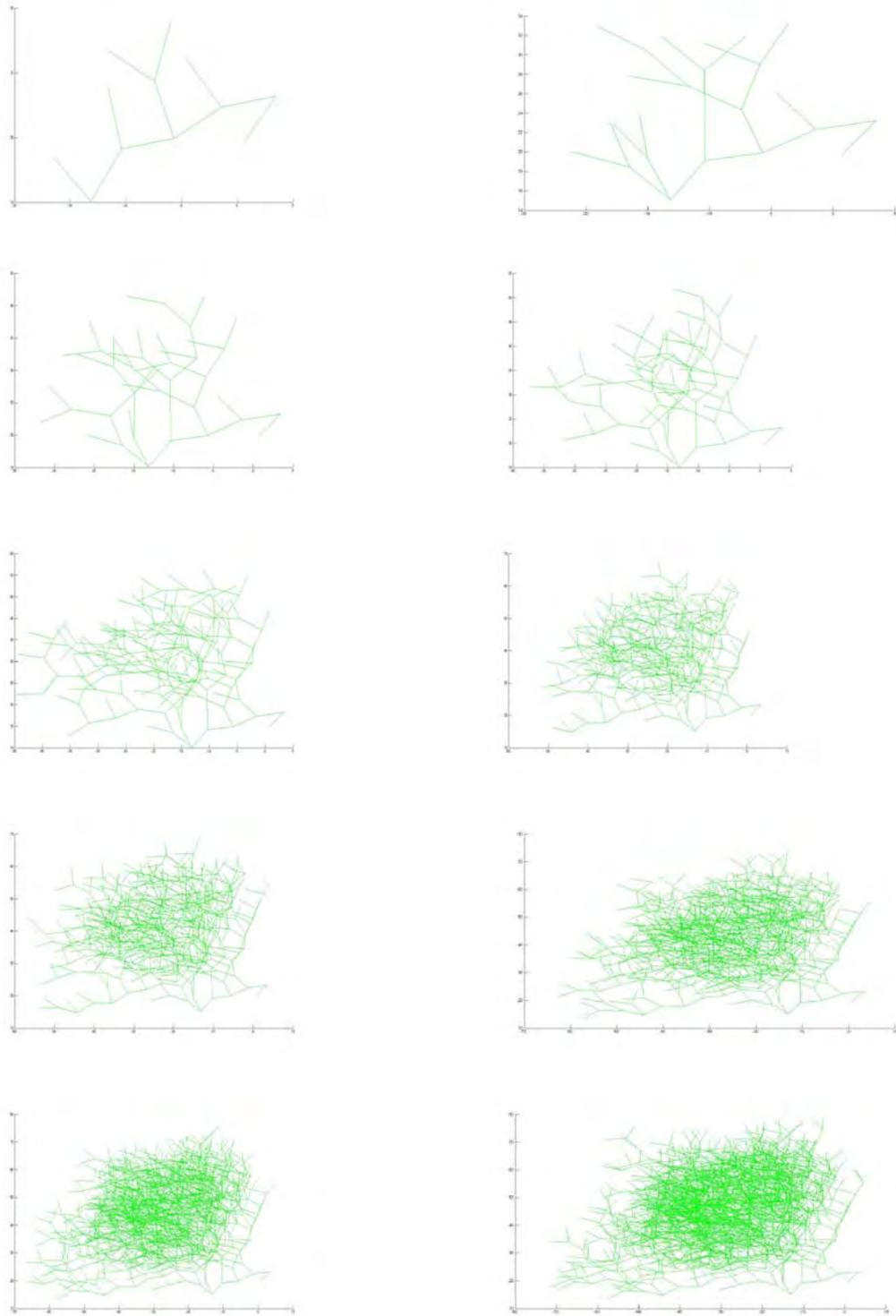


图 4-35： 使用生长模型预测的同一个小鼠海马区普肯野细胞在 10 个不同时间段的可视化空间几何描述

4.5.4 模型结果验证

建立起生长模型之后，由于具体神经元对模型参数的特异性要求，因此我们在测试模型结果的过程中只考虑小鼠的海马区普肯野神经元的模拟和预测情况。针对模型结果，我们从两方面进行了验证，一方面通过对一个小鼠海马区普肯野

神经元进行模拟，在这个神经元的基础上进行下一步的预测。具体如图 4-36 所示

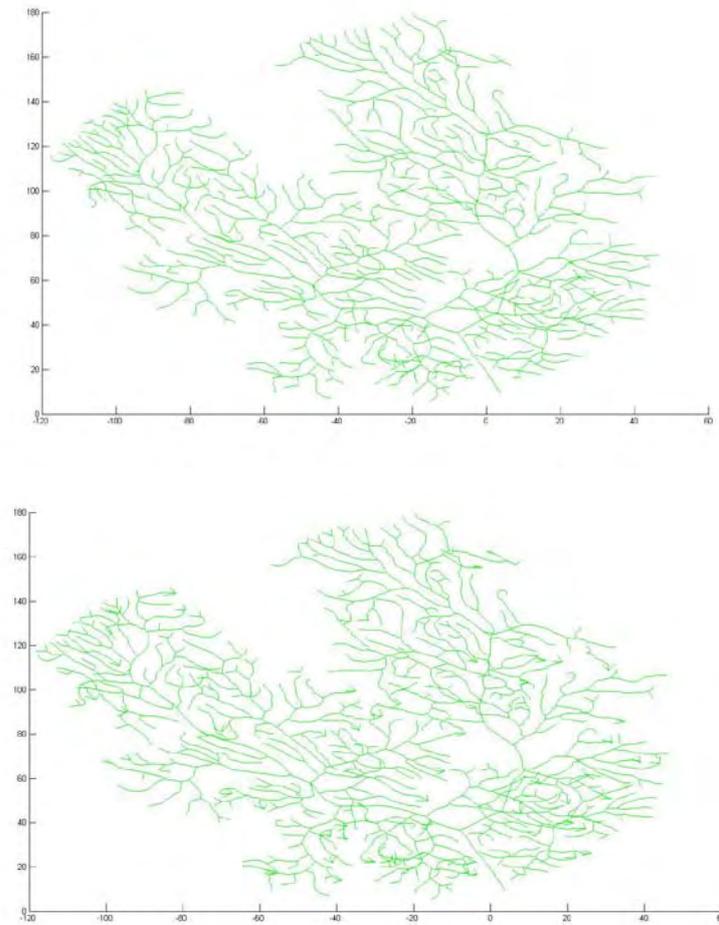


图 4-36： 对于一个普肯野神经元（上图）的模拟和预测（下图）

第二方面我们模拟单个普肯野神经元从定向分化后的生长和分支轨迹。具体的几何形态特征结果可见附件 2 中。然后利用问题一二的模型进行分类，分类正确，说明不影响分类问题。

五 模型评价与推广

本文针对神经元分类问题设计并实现了基于多分类器系统的分类模型，针对不同神经元的形态差异，使用了差分特征选择模型对不同类别的神经元进行特征选择，并改善 KNN 的分类效果。差分特征选择还能很好的区分不同物种相同神经元的不同。

通过扩充训练集以及新类别反馈学习，可以增加多分类器系统的鲁棒性以及泛化能力。

生长模型对于小鼠海马区普肯野神经元有着好的模拟和仿真效果。在对未来神经元的生长预测上也体现的较为合理。通过使用前面章节建立的分类器进行分类，结果显示通过仿真预测的结果的分类结果依然正确。这些充分验证了基于 Burke 算法的神经元生长模型的正确性和合理性。

模型也存在不足之处，如果训练数据太多可能会影响效率问题。另外生长模型可能会对一些特定的神经元泛化能力不足。

参考文献

- [1] M. Bota, H.-W. Dong and L.W. Swanson, From gene networks to brain networks, *Nat. Neurosci.* **6** (2003), pp. 795–799.
- [2] J.E. Cook, Getting to grips with neuronal diversity: what is a neuronal type?. In: L.M. Chalupa and B.M. Finlay, Editors, *Development and Organization of the Retina, NATO ASI Series, Plenum Press, New York* (1998), pp. 91–120.
- [3] R.H. Masland, Neuronal cell types, *Curr. Biology* **14** (2004), pp. R497–R500
- [4] Scorcioni R., Polavaram S., Ascoli G.: L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nature Protocols.* **3**:866-76 (2008).
- [5] Ascoli GA, Donohue DE, Halavi M. (2007) NeuroMorpho.Org: a central resource for neuronal morphologies. *J Neurosci.*, **27**(35):9247-51;
- [6] Y. Yang and X. Liu. A re-examination of text categorization methods. In SIGIR-99, 1999.
- [7] Boser, Guyon I, Vapnik V. A training algorithm for optimal margin classifiers[A]. Haussler D. Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory[C]. ACM Press, 1992. 144 - 152.
- [8] 邓乃扬, 田英杰. 支持向量机, 科学出版社, 2009
- [9] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [10] RE Burke, WB Marks and B Ulfhake, A parsimonious description of motoneuron dendritic morphology using computer simulation, *Journal of Neuroscience*, 1992,12(6):

附录

附录 1: 旋转坐标变换:

- 定义:

若二坐标系 $\{O; i, j\}$ 和 $\{O'; i', j'\}$ 满足 $O \equiv O'$, 另 $\angle(i, j') = \theta$ 则坐标系 $\{O'; i', j'\}$

可看成是由坐标系 $\{O; i, j\}$ 绕 O 旋转 θ 角得到的称由 $\{O; i, j\}$ 到 $\{O'; i', j'\}$ 的变换为旋转坐标变换。

- 旋转变换公式

由于 $\angle(i, i') = 0$ 故 $\angle(i, j') = \frac{\pi}{2} + \theta$

则 $i' = \cos \theta i + \sin \theta j$

$$j' = \cos\left(\frac{\pi}{2} + \theta\right)i + \sin\left(\frac{\pi}{2} + \theta\right)j$$

$$j = -\sin \theta i + \cos \theta j$$

$$i' = \cos \theta i + \sin \theta j$$

则

$$\begin{aligned} xi + yj &= \overrightarrow{OP} \\ &= \overrightarrow{O'P} \\ &= x'i' + y'j' \\ &= x'(i \cos \theta + j \sin \theta) + y'(-i \sin \theta + j \cos \theta) \\ &= (x' \cos \theta - y' \sin \theta)i + (x' \sin \theta + y' \cos \theta)j \end{aligned}$$

$$\text{即} \begin{cases} x = x' \cos \theta - y' \sin \theta \\ y = x' \sin \theta + y' \cos \theta \end{cases}$$

用 x, y 表示 x', y' , 有

$$\begin{cases} x' = x \cos \theta + y \sin \theta \\ y' = -x \sin \theta + y \cos \theta \end{cases}$$

附录 2:

生长模型针对小鼠海马区某一普肯野神经元进行的 10 期生长预测

Id	1	2	3	4	5	6	7	8	9	10
1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

3	0.000733	0.001613	0.003225	0.007624	0.017006	0.040756	0.094708	0.14602	0.208914	0.479402
4	0.000878	0.001757	0.003367	0.007759	0.017128	0.040843	0.094715	0.145952	0.208754	0.478846
5	0.000878	0.001757	0.003367	0.007759	0.017128	0.040843	0.094715	0.145952	0.208754	0.478846
6	0.027204	0.042351	0.059093	0.081605	0.096951	0.112722	0.134399	0.141107	0.144748	0.15982
7	0.049206	0.069213	0.066101	0.089051	0.101731	0.111246	0.119754	0.125389	0.12955	0.138759
8	0.011932	0.028776	0.040894	0.06312	0.101951	0.137435	0.143282	0.145651	0.159003	0.167956
9	0.000903	0.001781	0.003391	0.007781	0.017148	0.040858	0.094717	0.145941	0.208728	0.478754
10	0.001751	0.002808	0.004893	0.011726	0.019959	0.040907	0.094424	0.142398	0.203682	0.477451
11	0.002172	0.00323	0.005383	0.013406	0.020651	0.040055	0.093231	0.139668	0.201191	0.481013
12	0.000866	0.001805	0.003506	0.007815	0.017103	0.040553	0.094832	0.145916	0.208466	0.479139
13	0.001367	0.002494	0.004818	0.011563	0.019726	0.040435	0.094811	0.143055	0.203918	0.477813
14	0.001737	0.002765	0.004889	0.013905	0.020284	0.038511	0.091775	0.137269	0.200153	0.488712
15	0.001686	0.002741	0.005152	0.014027	0.020313	0.038199	0.092398	0.138342	0.20031	0.486833
16	0.00021	0.000577	0.001387	0.004036	0.01085	0.031334	0.082807	0.135778	0.205383	0.527638
17	0.000244	0.000649	0.001578	0.004439	0.011517	0.032343	0.084342	0.13688	0.205647	0.52236
18	0.000156	0.000479	0.001259	0.003848	0.010536	0.030865	0.082491	0.135328	0.205006	0.530032
19	0.000325	0.000786	0.00177	0.004804	0.012154	0.033321	0.085506	0.137953	0.20594	0.517442
20	0.000878	0.001757	0.003367	0.007759	0.017128	0.040843	0.094715	0.145952	0.208754	0.478846
21	0.000678	0.022725	0.007043	0.005611	0.108464	0.195375	0.182014	0.019616	0.054027	0.404448
22	0.00077	0.00068	0.001399	0.006203	0.011821	0.030595	0.088717	0.136938	0.207127	0.51575
23	0.000866	0.001805	0.003506	0.007815	0.017103	0.040553	0.094832	0.145916	0.208466	0.479139
24	0.000759	0.001638	0.003251	0.007648	0.017028	0.040772	0.094709	0.146008	0.208885	0.479303
25	0.000733	0.001613	0.003225	0.007624	0.017006	0.040756	0.094708	0.14602	0.208914	0.479402
26	0.000733	0.001613	0.003225	0.007624	0.017006	0.040756	0.094708	0.14602	0.208914	0.479402
27	0.000798	0.001135	0.002248	0.006871	0.014148	0.035109	0.091365	0.140928	0.207868	0.499531
28	0.000897	0.001131	0.00383	0.008165	0.020194	0.044673	0.100243	0.138911	0.208765	0.47319
29	0.000169	0.000578	0.003709	0.006707	0.0197	0.045141	0.103214	0.145493	0.206008	0.469282
30	0.000154	0.000588	0.001171	0.005822	0.012669	0.030952	0.093817	0.122269	0.212493	0.520063
31	0.000591	0.000742	0.001535	0.005681	0.011838	0.029153	0.084344	0.135329	0.212957	0.517829
32	0.000365	0.000419	0.000889	0.004081	0.008947	0.021388	0.07178	0.125837	0.221411	0.544884
33	0.000781	0.001462	0.003154	0.007819	0.017417	0.041682	0.095997	0.145685	0.207808	0.478194
34	0.000781	0.001462	0.003154	0.007819	0.017417	0.041682	0.095997	0.145685	0.207808	0.478194
35	0.000591	0.001595	0.003248	0.007514	0.016866	0.040493	0.095066	0.146576	0.209649	0.478401
36	0.000447	0.001454	0.003109	0.007381	0.016746	0.040408	0.095059	0.146644	0.209807	0.478945
37	0.000581	0.001313	0.003471	0.008047	0.01771	0.040165	0.086762	0.145075	0.205113	0.491763
38	0.000581	0.001313	0.003471	0.008047	0.01771	0.040165	0.086762	0.145075	0.205113	0.491763
39	0.000691	0.003359	0.003463	0.012025	0.012618	0.036146	0.088605	0.154881	0.212693	0.475519
40	0.001806	0.00211	0.003913	0.011624	0.017077	0.037288	0.094865	0.138901	0.20094	0.491477
41	0.001397	0.002976	0.003088	0.011084	0.013528	0.035886	0.091161	0.150044	0.2082	0.482636
42	0.002295	0.013774	0.014006	0.027532	0.016717	0.038844	0.007318	0.167858	0.263141	0.448517
43	0	0	0	0	0	0	0	0	0	0