

# 全国第七届研究生数学建模竞赛



题 目

## 神经元的空间形态分类和识别

摘

要：

本文针对神经元的空间形态分类和识别问题进行了探讨，利用模式识别、空间几何、统计分析和计算机搜索等知识，通过分析本题已给数据及从网络获取的补充数据，进行了神经元空间形态分类方法的研究，将神经元进行了比较准确的分类和识别。

针对问题一，通过分析附录 A、C 中的样本数据，结合空间几何图形，建立了分级多层二叉分类模型，经数据验证，分类效果很好。

针对问题二，对问题一的分类模型进行了改进，建立了新的特征判别函数，使用新的模型对附录 B 中的数据进行分类，可以判定其中 13 个神经元形态数据分别属于已给的神经元类型，并对未能归入已知类的 7 个神经元形态数据进行了新的命名和归类。

针对问题三，建立了基于综合特征判别和多分类支持向量机相结合的带权重的投票分类模型，消除了层次分类模型中“偏见”所带来的误分隐患，更能提高分类判别的成功率，并重新提出合理的命名建议。

针对问题四，建立了同特征参数搜索匹配模型，用特征参数波动率和特征带的重叠率作为标准，找到附件 A 中猪和鼠的普肯野神经元特征，较好地解决了不同物种同类神经元的区分问题。利用附件 A 和补充的网络数据对模型进行验证，得到了很好的分类效果。

针对问题五，通过分析从网络获得的不同年龄同物种同类神经元数据，建立了神经元生长预测模型，用稳定型参数和非稳定型参数刻画了神经元本类固有特征和神经元生长规律，尤其增长阶段的阻滞增长的模型贴近于真实情况。由于稳定型参数表征本类的固有特性，使用此预测模型可以较好地避免导致预测结果进入其他类。

**关键词：** 分级多层二叉分类 综合特征判别法 多分类支持向量机 同特征参数搜索匹配 神经元生长预测模型

## 1 问题重述

作为大脑构造的基本单位，神经元的结构和功能包含很多因素，其中神经元的几何形态特征和电学物理特性是两个重要方面。

几何形态特征主要包括神经元的空间构象，具体包含接受信息的树突，处理信息的胞体和传出信息的轴突三部分结构。由于树突、轴突的生长变化，使得神经元的几何形态千变万化。

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一个空间形态分类方法，将神经元根据几何形态比较准确地分类识别。在已获得大量神经元的几何形态数据的前提下，侧重用不同的指标去刻画神经元的形态特征。

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，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。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），给出一个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，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。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，以附件 A 中猪的普肯野神经元和鼠的普肯野神经元为例，它们的特征有何区别。

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，预测神经元形态的生长变化，并讨论这些形态变化对确定的几何形态特征有什么影响。

## 2 符号说明

Soma\_Surface—细胞体表面积  
N\_stems—连接细胞体的主干数目  
N\_bifs—二根分叉部数目  
N\_branch—分支数目  
Width—宽度  
Height—高度  
Depth—深度  
Diameter—直径  
Length—长度  
Surface—表面积  
Volume—体积  
EuclideanDistance—房室相对于细胞体的欧式距离总和  
PathDistance—房室之间距离长度总和  
Branch\_Order—枝干级别总和，细胞体为0，一级枝干为1，依次类推  
Contraction—收缩率的比值，欧式距离和路径距离的比值之和，表征房室组成的轴突树突的弯曲程度  
Fragmentation—碎化指数树干锥度  
Partition\_asymmetry—二叉分部处的枝干生长不对称性指数  
Rall\_Power—树干锥度指数  
Bif\_ampl\_local—二叉分部枝干起始处前两个房室夹角之和  
Bif\_ampl\_remote—二叉分部枝干底端的两个房室相对于分叉点的夹角之和  
CF\_X\_Y—单特征判别函数，其中X为神经元类型，Y为特征参数  
CB\_X\_Y—特征参数特征带，其中X为神经元类型，Y为特征参数  
Avg(i)—i的均值

## 3 问题假设

1. 各类神经元的各个形态参数均服从正态分布；
2. 附件中所给神经元数据代表了本类神经元的共有特征；
3. 题目给出的附件源数据网站上的所有神经元数据都具有与附录给出数据的权威性和真实性；
4. 题目给出的附件源数据网站上提供的用于计算神经元空间形态参数的软件L-Measure的计算结果具有一定的可信度；
5. 从题目给出的附件源数据网站上获取的题五中所需的带有年龄信息的神经元信息正确。

## 4 问题分析及模型建立

### 4.1 问题一

#### 4.1.1 问题分析

本问题需要利用附录 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 7 类神经元的几何特征，依据几何特征建立神经元空间形态分类模型。 给出一个神经元空间形态分类的方法。问题的求解需要完成以下三步：

Step1:

将附录 C 中标准格式 A. swc 的数据转换为本文所用的形态参数的数据格式；

Step2:

逐层对所给 7 类神经元的形态参数数据进行分析，挑选能够表征神经元类别的指标及其取值特点，作为分类指标，同时总结附录 C 中 7 类神经元的几何特征。

Step3:

提出一种基于小样本数值分析的分级多层二叉分类模型，完成对 7 类神经元的分类。

#### 4.1.2 初始数据处理

使用 L-Measure 软件将附录中 A. swc 格式的数据转化为形态参数数据。

#### 4.1.3 神经元几何特征的提取

定义 1 特征参数：表征神经元类别的显著性形态参数。

定义 2 特征带：神经元特征参数的取值范围。

定义 3 波动率  $\Upsilon$ ：形态参数标准差  $\sigma$  与均值  $\mu$  的比值。

$$\Upsilon = \frac{\sigma}{\mu} \quad (1)$$

波动率  $\Upsilon$  表征指标稳定性，作为筛选特征参数的评判标准。

定义 4 大类差异指数  $\text{diff\_max}$ ：各类神经元单项形态参数均值中最大值  $\text{max}$  跟次大值  $\text{max'}$  的差与次大值  $\text{max'}$  的比值。

$$\text{diff\_max} = \frac{\text{max} - \text{max'}}{\text{max'}} \quad (2)$$

小类差异指数  $\text{diff\_min}$ ：各类神经元单项形态参数均值中次小值  $\text{min'}$  跟最小值  $\text{min}$  与最小值  $\text{min}$  的比值。

$$\text{diff\_min} = \frac{\text{min'} - \text{min}}{\text{min}} \quad (3)$$

差异指数比值越大，差异越明显。

实现步骤：

Step1:

将附录 A 中形态参数数据列成表格形式，并计算出各类别神经元样本各形态参数的均值  $\mu$  及波动率  $\Upsilon$ ，如表格 1 所示：

表格 1 附录 A 样本数据、均值及波动率

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	类别\参数	Soma_Surface	N_stems	N_bifs	N_branch	Width	Height	Depth	Diameter	Length	Surface	Volume	EucDist	P
2	Motoneuron 1	33571.5	11	150	321	2480000	1800000	2090000	3370.44	99039.8	548245	439406	888671.1	0.00
3	Motoneuron 2	45216	10	122	263	1040000	981824	1090000	2310.58	78949.1	512418	390413	428043	0.00
4	Motoneuron 3	28577.6	13	171	367	1300000	1300000	1280000	2821.47	103228	676400	530778	547804	0.00
5	Motoneuron 4	30403.2	12	175	373	1520000	1360000	1400000	2291.31	113502	499671	283360	559000	0.00
6	Motoneuron 5	37443.4	8	181	377	1780000	1490000	1430000	2119.02	117028	543917	275087	630878	0.00
7	AVG (均值)运动型	35042.34	10.8	159.8	340.2	1820000	1386325	1458000	2582.564	101729.6	556130.2	383804.8	610078.8	0.00
8	STDEV (标准差) 运动型	6606.582068	1.923538	24.13918	48.67443	549545.3	301719.4	377716.8	512.7138	15254.49	70311.03	107992.3	171824.8	0.00
9	波动率	19%	18%	15%	14%	34%	22%	28%	20%	15%	13%	28%	28%	0.00
10	Bipolar interneuron 1	405.358	2	70	143	415588	1290000	143070	453.012	7821.64	9352.12	994.507	307007	0.00
11	Bipolar interneuron 2	525.285	5	69	147	267307	1640000	146486	765.166	7022.5	8055.66	943.759	381917	0.00
12	Bipolar interneuron 3	770.238	3	103	211	335482	2800000	148297	1399.73	7029.08	12354.4	2708.31	718712	0.00
13	Bipolar interneuron 4	1077.46	3	86	177	506979	1640000	293451	955.068	11052.3	14989.3	2160.99	413682	0.00
14	Bipolar interneuron 5	5.73E+02	3	23	51	20954.2	418250	35302.7	274.838	2425.23	3320.42	475.793	131673	0.00
15	AVG (均值)二极中间型	6.70E+02	3.2	70.2	145.8	306258	1537850	153521.3	769.5828	7070.15	8614.38	1458.672	390584.2	0.00
16	STDEV (标准差) 二极中间型	2.63E+02	1.095445	29.81107	59.6255	184367.4	854165.6	91832.09	440.696	3082.288	4430.288	935.6096	213480.5	0.00
17	波动率	39%	34%	42%	41%	80%	56%	60%	57%	44%	48%	64%	55%	0.00
18	P44-DEV191.CNG	1030.8	3	3	11	72888	233099	20052.5	921.307	1129.06	7042.24	3509.48	68423.6	0.00
19	Tripolar interneuron 1	2821.58	3	3	11	10508	44609.7	10190.9	425.499	457.462	2650.23	1234.04	14753.2	0.00
20	Tripolar interneuron 2	1108.93	4	18	43	372890	184731	38024.8	1809.09	1921.1	12525.3	6813.11	97809.1	0.00
21	Tripolar interneuron 3	621.293	3	4	13	49390	15445	8797.35	324.273	516.794	2175.57	728.815	13846.3	0.00
22	Tripolar interneuron 4	1126.36	3	9	23	132572	148321	6074.95	856.26	1187.58	5055.24	1769.49	52661.4	0.00
23	AVG (均值)三极中间型	1301.7928	3.2	7.4	20.2	127569.6	125441.1	18628.1	987.2858	1038.399	5891.718	2808.997	49496.72	0.00
24	STDEV (标准差) 三极中间型	765.8864236	0.447214	6.426908	13.08211	143999.5	92612.93	13073.67	587.3826	694.4804	4193.095	2471.473	35987.47	0.00
25	波动率	59%	14%	87%	68%	113%	74%	79%	68%	57%	71%	88%	73%	0.00
26	Multipolar cell 1	1182.24	8	16	47	132059	124902	14515.4	1181.33	1680.88	7944.11	3024.03	47072.5	0.00
27	Multipolar cell 2	341.166	4	1	9	2494.23	2375.96	0	122.391	98.5305	414.787	138.954	897.758	0.00
28	Multipolar cell 3	1064.35	6	16	43	181484	190420	55311.4	1348.81	1355.4	8949.93	3465.98	63215.7	0.00
29	Multipolar cell 4	1065.55	3	20	45	167933	280184	25310.6	1940.45	1922.48	11354	5390.99	91583.1	0.00
30	Multipolar cell 5	1507.27	7	1	15	11481.6	16329.6	6070.4	337.816	420.183	1788.86	582.568	5274.12	0.00
31	Multipolar cell 6	2160.66	8	14	43	408923	196893	19048.5	1554.1	2038.97	9539.89	3839.35	91020.2	0.00
32	Multipolar cell 7	2627.35	8	5	25	36292.1	46482.3	6697.62	613.865	681.233	2867.81	960.717	15496.1	0.00
33	Multipolar cell 8	810.485	6	25	61	322898	292193	52034.5	2919.06	2582.47	17389.8	9448.43	126176	0.00
34	Multipolar cell 8	1408.23	5	10	28	131848	256847	26990	1710.44	1751.79	8349.76	4010.11	101818	0.00
35	AVG (均值)多极中间型	1354.144558	6.111111	12.35	22222	155046	155157.4	22887.6	1303.14	1448.082	7730.994	3429.014	60291.39	0.00
36	STDEV	689.6791161	1.833333	8.386428	16.80608	139087.2	113589.7	19607.04	870.6519	838.6447	5323.356	2865.253	45751.34	0.00
37	波动率	51%	30%	70%	48%	89%	73%	86%	67%	58%	69%	84%	76%	0.00
38	Guinea-pig Purkinje cell	0	1	471	944	455553	488609	47518	2904.8	10880.1	50877	27291.9	316589	0.00
39	Guinea-pig Purkinje cell	0	1	419	840	373406	318600	27378	1849.5	8413.2	34866.2	19128.8	216209	0.00
40	Guinea-pig Purkinje cell	0	1	416	835	291502	292293	26949	2203.5	8144.81	42520	22380.8	179542	0.00
41	AVG (均值)普肯野	0	1.435	3333	873	373520.3	366517.3	33948.33	2319.287	9149.37	42687.73	22833.83	237446.7	0.00
42	STDEV (标准差) 普肯野	0	0.30	92464	61.53861	81975.56	106552.6	11753.63	537.0903	1513.477	8106.702	4109.554	70948.92	0.00
43	波动率	#DIV/0!	0%	7%	7%	22%	29%	35%	23%	17%	19%	18%	30%	0.00
44	mouse Purkinje cell 1	798.425	1	369	740	892312	730579	10070	2471.68	4980.15	8336.16	1742.46	432832	0.00

## Step2:

先取单指标各类神经元样本数据的均值为对象，再按降序排列。以指标 Soma\_Surface 为例，如表格 2、表格 3 所示：

表格 2 七类 Soma\_Surface 指标均值（排序前）

类别	形态参数	Soma_Surface
AVG (均值) 运动型神经元		35042.34
AVG (均值) 双极中间神经元		670.36
AVG (均值) 三极中间神经元		1301.79
AVG (均值) 多极中间神经元		1354.14
AVG (均值) 普肯野神经元		451.80
AVG (均值) 椎体神经元		798.20
AVG (均值) 感觉神经元		0.00

表格 3 七类 Soma\_Surface 指标均值（排序后）

类别	形态参数	Soma_Surface
AVG (均值) 运动型		35042.34
AVG (均值) 多极中间神经元		1354.14
AVG (均值) 三极中间神经元		1301.80
AVG (均值) 椎体神经元		798.20
AVG (均值) 二极中间神经元		670.36
AVG (均值) 普肯野神经元		451.80
AVG (均值) 感觉神经元		0

## Step3:

挑选出特征参数并计算特征带。

挑选规则为:某神经元类别某项指标的差异指数显著大或显著小。仍以形态参数 Soma\_Surface 为例进行处理, 如表格 4 所示:

表格 4 Soma\_Surface 指标极值类别及差异指数

形态参数	指标最大值所属类别	大类差异指数	指标最小值所属类别	小类差异指数
Soma_Surface	运动型神经元	25.88	感觉神经元	$\infty$
N_stems	运动型神经元	1.61	感觉神经元	$\infty$
N_bifs	普肯野神经元	2.48	三极中间神经元	1.53
N_branch	普肯野神经元	2.33	三极中间神经元	1.17
Width	运动型神经元	3.39	感觉神经元	5.73
Height	双极中间神经元	1.11	感觉神经元	4.44
Depth	运动型神经元	3.09	三极中间神经元	1.16
Diameter	普肯野神经元	1.01	感觉神经元	2.74
Length	运动型神经元	7.86	感觉神经元	1.87
Surface	运动型神经元	19.71	感觉神经元	2.22
Volume	运动型神经元	28.66	感觉神经元	1.17
EucDistance	运动型神经元	1.56	感觉神经元	2.6
PathDistance	运动型神经元	1.38	感觉神经元	2.13
Branch_Order	普肯野神经元	4.68	感觉神经元	1.01
Contraction	普肯野神经元	2.43	三极中间神经元	1.15
Fragmentation	普肯野神经元	1.78	感觉神经元	2.13
Partition_asymmetry	普肯野神经元	2.6	三极中间神经元	1.52
Rall_Power	普肯野神经元	1.29	多极中间神经元	1.1
Bif_ampl_local	普肯野神经元	3.79	三极中间神经元	1.63
Bif_ampl_remote	普肯野神经元	3.34	三极中间神经元	1.49

对这两类神经元在指标 Soma\_Surface 下的各组样本值进行数值检验, 判别是否存在与相邻类别的重叠现象。若不存在, 则说明错分的概率较小, 可作为特征参数并给出特征带。原理如图 1:

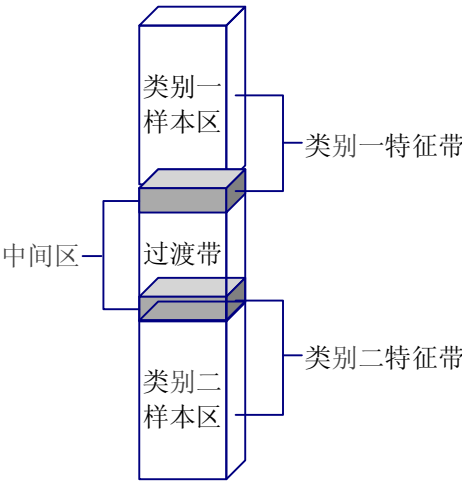


图 1 特征带分类原理图

综合分析表格 1 中的数据, 运用数理统计知识, 计算出特征参数对应的特征带。

以下为神经元特征参数特征带的计算方法：根据模型假设（1），神经元特征参数  $p_i$  服从正态分布  $N(\mu_i, \sigma_i^2)$ ，即

$$\frac{\bar{x} - \mu_i}{\sigma_i / \sqrt{n}} \sim N(0,1) \quad (4)$$

其中  $\mu_i$  为特征参数  $p_i$  的均值， $\sigma_i^2$  为特征参数  $p_i$  的方差。

由于特征参数  $p_i$  的标准差  $\sigma$  未知，则用样本标准差  $S$  对  $\sigma$  进行估计。从而得到

$$\frac{\bar{x} - \mu}{S / \sqrt{n}} \sim t(n-1) \quad (5)$$

定义 5: CB\_X\_Y

$$CB\_X\_Y = \left( \mu - \frac{S * t_{\alpha/2}(n-1)}{\sqrt{n}}, \mu + \frac{S * t_{\alpha/2}(n-1)}{\sqrt{n}} \right) \quad (6)$$

神经元特征参数的特征带表征该特征参数数值以置信度 95% 落入的区间。

Step4:

在特征参数及特征带确定之后，以差异指数为依据，对其显著级别进行划分，差异指数越大，显著级别越高。依据附录中的数据分析，本文的特征参数级别取前三级，数据综合如表格 5 所示：

表格 5 特征参数级别

类别 \ 级别	1	2	3
运动型神经元	Volume	Soma_Surface	Surface
双极中间型神经元	Height		
三极中间型神经元	Bif_ampl_local	N_bifs	Partition_asymmetry
多极中间型神经元	Rall_Power		
普肯野神经元	Branch_Order	Bif_ampl_local	Bif_ampl_remote
感觉神经元	Soma_Surface	N_stems	Width

Step5:

根据附录 A、C 中的数据分析，结合各样本的空间几何形态，提取 7 类神经元的几何特征如下：

运动神经元：空间形状上突起向四面八方近于均匀散开，细胞体表面积较大、细胞总面积、体积都较大，特别是细胞体表面积和体积。

感觉神经元：房室数据组成上，感觉神经元无细胞体和树突，仅由轴突组成。

普肯野神经元：空间特征上普肯野神经元的树突多生长在同一平面内，有一到两个主干，仅由细胞体和树突组成。

锥体神经元：由细胞体、树突和轴突组成，一般呈对顶的两个锥体形生长。反映在几何形态特征参数上，锥体神经元和与其空间形态相近的中间神经元相比 Rall Power 指标较大。

双极中间神经元：通过大量搜集中间神经元 SWC 数据发现，双极中间神经元是中间神经元中唯一一个由细胞体、树突和轴突三部分组成的神经元，其余两种中间神经元都仅由细胞体和树突组成。双极中间神经元在形态上有明显的中轴，总体趋势上呈直线向两端生长。

三级中间神经元和多级中间神经元：三级中间神经元和多级中间神经元都只由细胞体和树突组成，三级和多级中间神经元的主要区别在于从细胞体发出的主干的数目即极的数目不同。

归纳列表如下：

几何特征 神经类型	形态参数特征	空间几何形态特征
运动神经元	胞体表面积、体积较大， 细胞总面积、体积较大	突起向四面八方近于均匀 散开
感觉神经元	胞体表面积、体积为 0	无细胞体和树突，仅由轴突 组成
普肯野神经元	主干数为 1 或 2	树突多生长在同一平面内， 有一到两个主干，仅由细胞 体和树突组成。
椎体神经元	Rall Power 值较大	由细胞体、树突和轴突组 成，一般呈对顶的两个锥体 形生长
双极中间神经元	树突和轴突都不为 0， 主干的数目约为 2	是中间神经元中唯一一个 由细胞体、树突和轴突三部 分组成的神经元；在形态上 有明显的中轴，总体趋势上 呈直线向两端生长
三级中间神经元	轴突数为 0， 主干的数目约为 3	只由细胞体和树突组成，从 细胞体发出的主干的数目 大致为 3 个
多级中间神经元	轴突数为 0， 主干的数目大于 3	只由细胞体和树突组成，从 细胞体发出的主干的数目 多于 3 个



#### 4.1.4 分级多层二叉分类模型

##### 1、建模思想

基于 4.1.3 中定义的特征参数及其级别，级别越高，特征参数越显著，也就是表征神经元类别的能力越强。由此，我们提出一种基于级别高低的分级多层分类方法，从高到低分层用各级指标对样本神经元数据进行判定。

##### 2、模型建立

分级多层分类方法的可视化模型如图 2 所示，其中特征参数标号越小，在这一级中处理数据的优先级别越高。该模型为神经元空间形态分类提供了一个较好的方法。

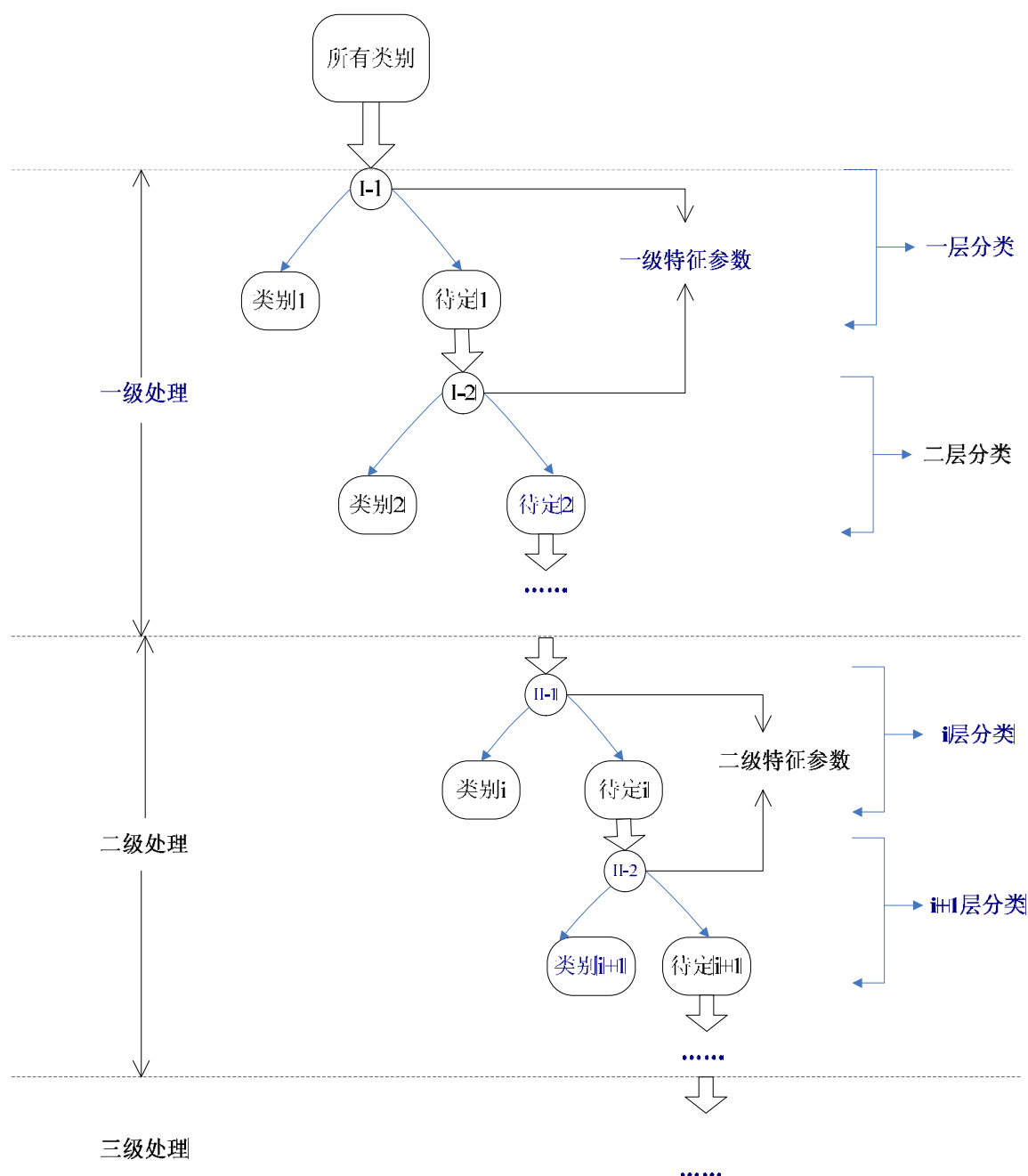


图 2 分级多层二叉分类模型原理图

##### 3、建立特征判别函数

根据在 4.1.3 中定义的带级别的特征参数及其特征带，建立对应的单特征判别函数，

以运动神经元的体积指标 Volume 的特征判别函数为例：

$$CF\_Moto\_Volume = \begin{cases} 1, & Volume \in CB\_Moto\_Volume \\ 0, & Volume \notin CB\_Moto\_Volume \end{cases} \quad (7)$$

其中

CF\_X\_Y—单特征判别函数，其中X为神经元类型，Y为特征参数

CB\_X\_Y—特征参数特征带，其中X为神经元类型，Y为特征参数

当样本 Volume 指标值落入运动神经元 Volume 特征参数的特征带  $CB\_Moto\_Volume$  间，函数值  $CB\_Moto\_Volume$  取 1，此时判断此样本属于运动神经元，反之不属于。本文中计算得  $CB\_Moto\_Volume$  取值为 [281362, 490887.8541]。

#### 4、模型求解

“分级多层二叉分类模型”求解过程如下：

Step1:

根据差异指数值的比对，可将分类层次及在各层中作为判别依据的特征参数找出，列表如表格 6 所示：

表格 6 分级多层特征参数及对应神经元类型

级别	层次	神经元类型	特征参数
一级	1	运动型神经元	Soma_Surface
	2	感觉神经元	Width
	3	普肯野神经元	Branch_Order
	4	三级中间神经元	Bif_ampl_local
	5	双级中间神经元	Height
	6	多级中间神经元	Rall_Power
二级	1	运动型神经元	Soma_Surface
	2	感觉神经元	Height
	3	普肯野神经元	Bif_ampl_local
	4	三级中间神经元	N_bifs
三级	1	运动型神经元	Surface
	2	普肯野神经元	Bif_ampl_remote
	3	感觉神经元	Diameter
	4	三级中间神经元	Partition_asymmetry

Step2:

对待分类的数据进行分级处理。在每一级内根据表格 6 的层次划分，依次用各层特征参数对应的特征判别函数，对样本数据流进行逐层处理，如图 3 所示：

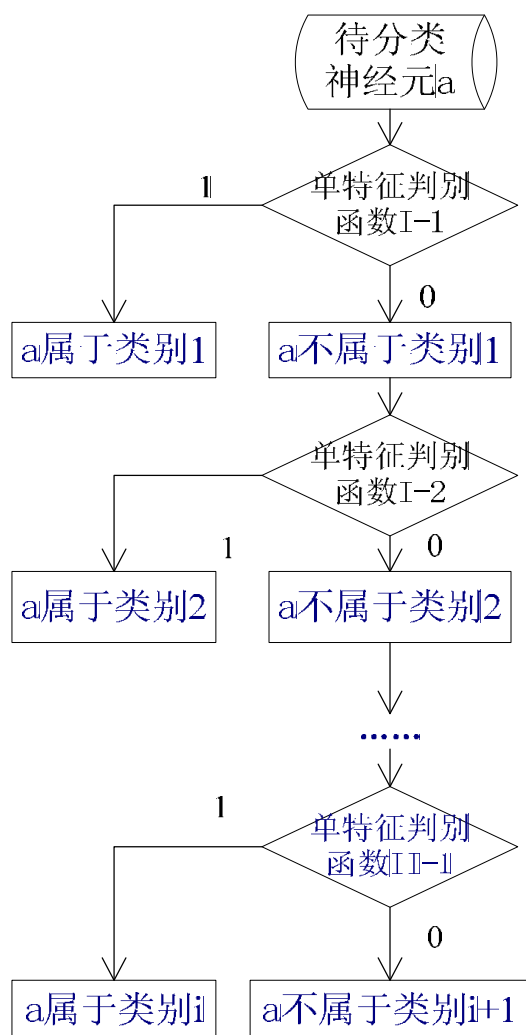


图 3 单个样本分类流程

Step3:

重复上一步，直至所有数据处理完毕，即完成了基于神经元空间形态的一个分类。

## 5、模型验证

将附录 A 和附录 C 中的神经元数据代入模型进行处理，可以检验出分类结果唯一且准确，因此证明该模型在这一适用范围内的正确性。

## 4.2 问题二

### 4.2.1 问题分析

本问题是对“分级多层二叉分类模型”的一个应用。运用该模型对附录 B 中 20 个神经元形态数据进行神经元类型的判定。联系数据，分析结果，判定是否有必要引入或定义新的神经元名称。

### 4.2.2 问题求解

将“分级多层二叉分类模型”用于附录 B 中样本的分类，并将神经元的三维空间图画出，作为模型分类结果的验证，也可作为一个根据神经元空间形态特征提出新命名的依据。

#### 1、求解思路

首先将附录 B 中 20 个神经元形态数据代入“分级多层二叉分类模型”中进行计算，根据所得解判定它们属于什么类型的神经元：若有唯一解，则判定成功；若无解，先应

该与已知分类的 7 类神经元的三维空间形态进行参照，若形态特征不相似或具有明显新的特点，则有必要引入或定义新的神经元名称。

## 2、模型求解

将附录 B 中 20 个神经元形态数据代入“分级多层二叉分类模型”中进行计算，发现其中有 11 个可以分类成功，且与三维空间中观察的形态特点相符；其余 9 个分类结果不明确。结果由表格 7 列出：

表格 7 附录 B 中神经元模型分类及三维视觉分类结果

神经元 标号	模型计算结果	三维视觉分类 结果	神经元 标号	模型计算结果	三维视觉分类结果
1	锥体神经元	锥体神经元	11	运动神经元	运动神经元
2	锥体神经元	锥体神经元	12	运动神经元	运动神经元
3	锥体神经元	锥体神经元	13	感觉神经元	感觉神经元
4	锥体神经元	锥体神经元	14	感觉神经元	感觉神经元
5	普肯野神经元	普肯野神经元	15	未分类	未知神经元
6	普肯野神经元	普肯野神经元	16	未分类	未知神经元
7	未分类	未知神经元	17	未分类	未知神经元
8	未分类	未知神经元	18	未分类	未知神经元
9	未分类	未知神经元	19	运动神经元	运动神经元
10	运动神经元	运动神经元	20	运动神经元	运动神经元

## 3、命名新的神经元

通过分析表格 7 中未能进行分类的神经元的空间形态，可以找出这些神经元具有一些特殊的形态特征。提取其新的形态特征，并据此引入新的神经元名称。

(1) 附录 B 中标号为 15、16 的两个神经元，在空间三维上具有平面性的特征，这一点虽与普肯野神经元具有的平面特征相同，但在正面形态上差异较大（见图 4-图 9），故不能简单的归为普肯野神经元。在三维空间中对其形态进行分析，发现这一类神经元的一端有鹿角状分支，整体也与公鹿相似。在此可将其命名为平面鹿型神经元。

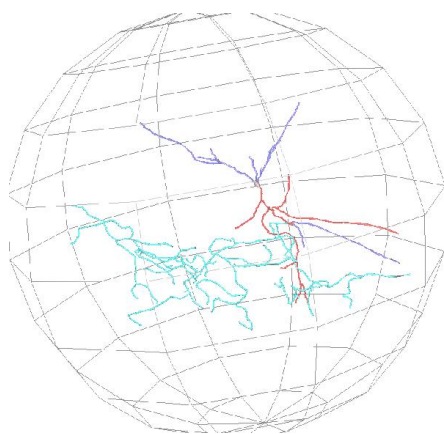


图 4 附录 B 中 15 号神经元正面图

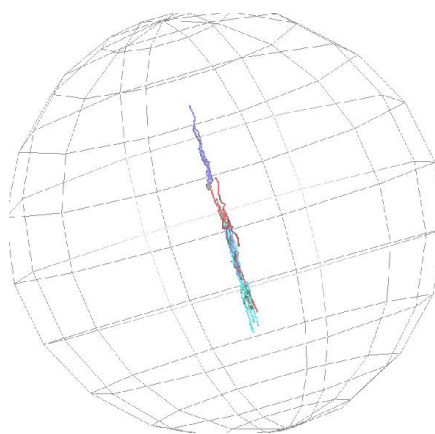


图 5 附录 B 中 15 号神经元侧面图

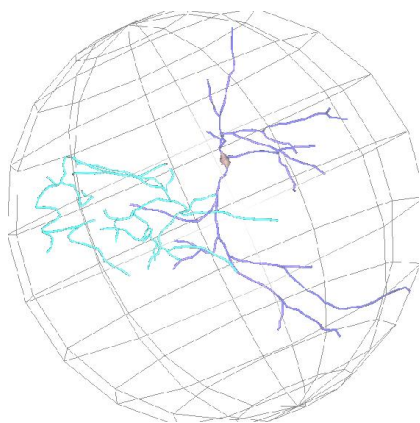


图 6 附录 B 中 16 号神经元正面图

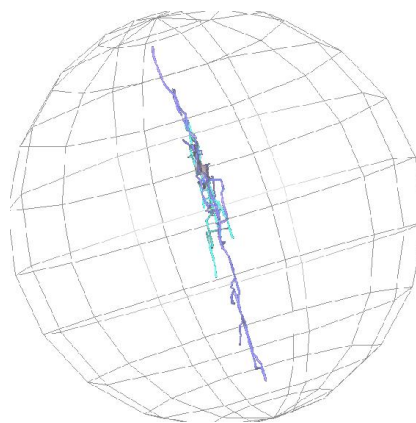


图 7 附录 B 中 16 号神经元侧面图

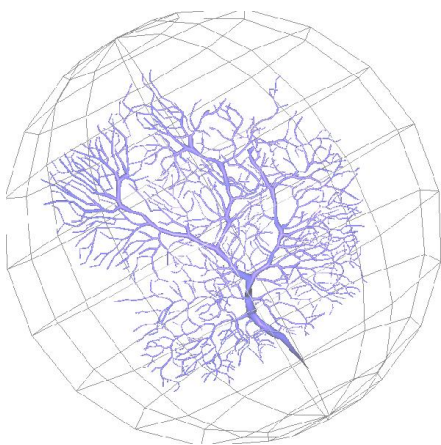


图 8 附录 A 中猪普肯野神经元正面图

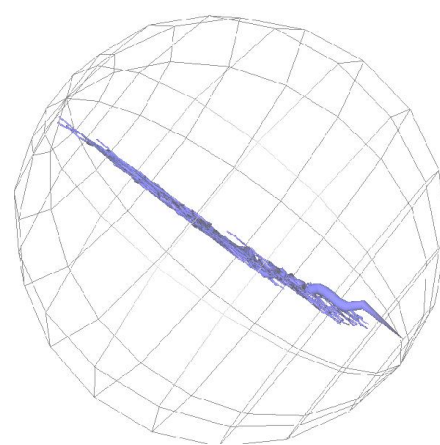


图 9 附录 A 中猪普肯野神经元侧面图

(2) 附录 B 中标号为 17、18 的两个神经元形态相近，应属于一类，如图 10、11 所示；它们与经典 7 类神经元的形态差异较为明显，故认为是新的神经元。在三维空间中对其形态进行分析，得到其特征为平面型，且末梢呈触角状，故命名为平面爬山虎型神经元。

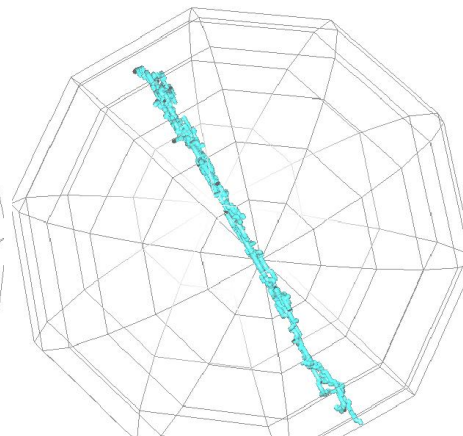
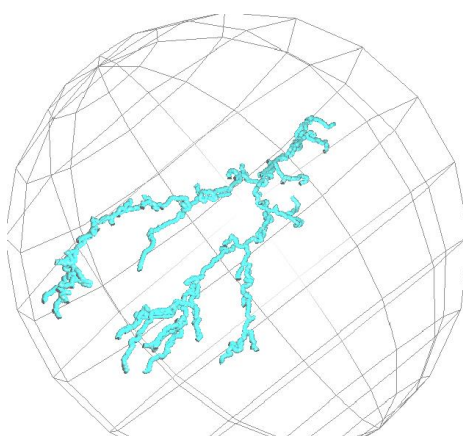


图 10 附录 B 中 17 号神经元正、侧面图



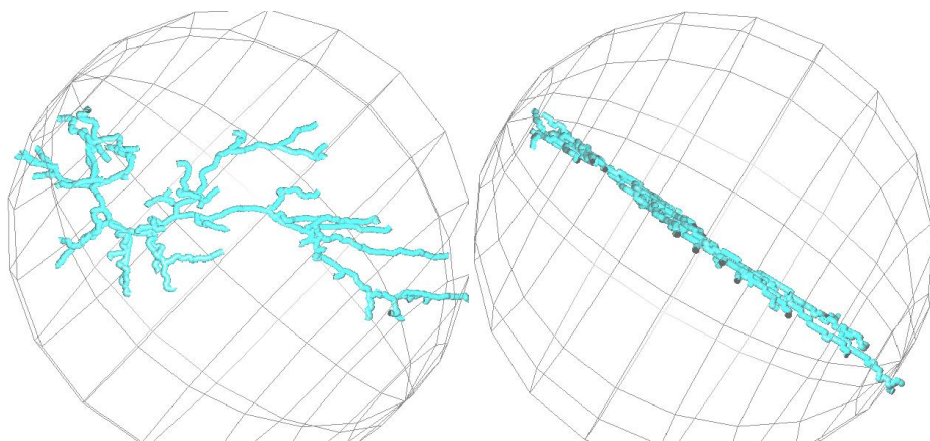


图 11 附录 B 中 18 号神经元正、侧面图

(3) 附录 B 中标号为 7、8、9 的两个神经元形态相近也归属于一类，与运动型形状类似但无胞体，故命名为无胞体运动型神经元。如图 12 所示：

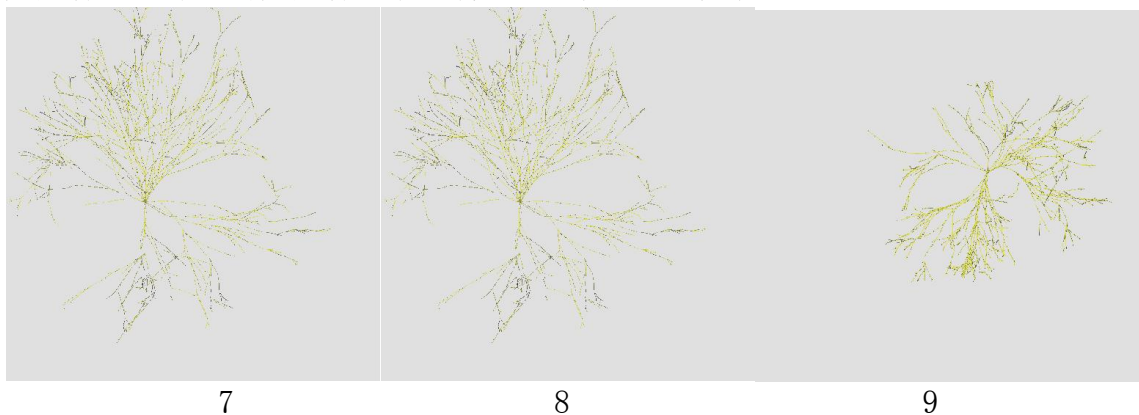


图 12 附录 B 中 7、8、9 号神经元空间几何形态图

### 4.3 问题三

#### 4.3.1 神经元分类模型问题分析

问题一提出的分级多层二叉分类模型可以很好地解决附录 A 和附录 B 中神经元的分类问题。但是在层次化的分类过程中会出现误分类情况。其产生误差的原因是层次化的分类方法都是“带有偏见”的。如果待分类的数据属于下层的类别却与上层类别具有一定的相似性，由于每次分类都是分为 A 类和非 A 类，就有可能被误分入上层类而产生错误。另外即使使用大量阈值的过渡带和权值做模型修正，层次分类模型仍旧不是最优的分类方法。要想在样本有限的条件下，获取更加优化的分类的界限，使得分类方法更具有普适性，必须引入新的“无偏见”的分类机制。

#### 4.3.2 分类模型建立

由于分级多层二叉分类模型是“带有偏见”的，因此引入一种“基于综合特征判别和多分类支持向量机 (Multi-Class Support Vector Machine) 相结合的带权重投票模型”，对之前的分类方法进行优化。

##### 一、建模思想

当有新的待分类神经元时，各类拥有自己的特征判别函数的神经元都根据特征判别函数的结果进行投票，对于特别接近自身特征即特征判别结果较好的神经元投极赞成票，对于接近自身特征的神经元投赞成票，对于不接近自身特征的神经元投反对票。

分类规则如下：

(1) 若一轮投票结束后票箱中仅有一个极赞成票或赞成票，其余均为反对票则将

待分类神经元划分到投极赞成票或赞成票的神经元类中；

(2) 若票箱中无赞成票则放弃将待分类神经元归入已知分类中；

(3) 若票箱中有多个赞成票则说明待分类神经元与多类神经元特征相似，由于各类神经元的特征参数与其余神经元相比，其差异显著性不尽相同，各神经元基于其特征判别函数得出的赞成票需要根据显著性不同乘以不同权重以使得模型更加合理。

为了得到接近最优分类的分类超平面，引入多分类支持向量机进行再分类。

## 二、分类模型算法流程

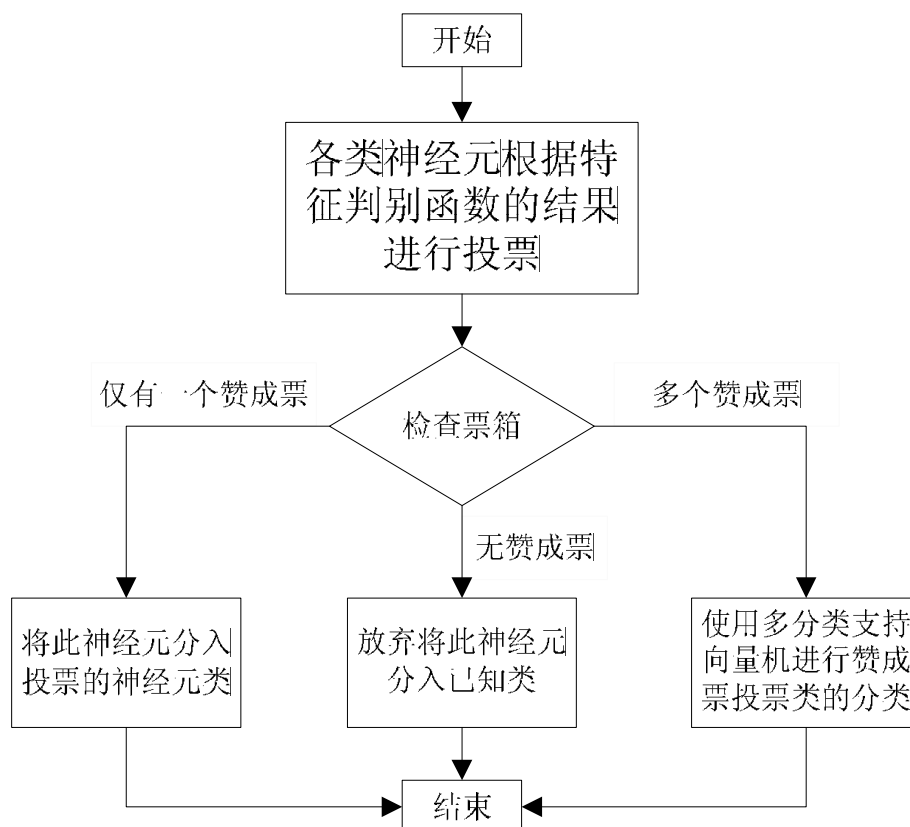


图 13 投票法模型算法流程图

## 三、改进分类模型的建立

问题一中的分级多层二叉分类模型已经给出了各类细胞所对应的特征判别函数，问题二中对特征判别函数加以了部分补充。

支持向量机模型 (Support Vector Machine, SVM) 是 AT&T Bell 实验室的 Vapnik 等人提出的一种针对分类和回归问题的新型机器学习方法。它从寻找最优分类面的思想出发，基于结构风险最小化归纳原则，根据有限的样本信息在模型的复杂性(即对特定训练样本的学习精度)和学习能力(即无错误地识别任意样本的能力)之间寻求最佳配置，以期获得最好的推广能力。它能很好的解决有限数量样本的高维模型的构造问题，因而本文采用该方法解决问题三。

### (1) 两类神经元分类模型

对于仅有两类神经元投赞成票，即仅有两类神经元待分的情况下，假设神经元样本为  $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_L, y_L)\}$ ，其中  $L$  为训练样本集规模， $x_i \in R_{20}$  为 20 维特征参数构成的特征向量， $y_i \in \{1, -1\}$  为第  $i$  个神经元样本对应的类别。首先验证其是否

线性可分，即是否满足  $y_i(w \cdot x_i + b) - 1 \geq 0, i = 1, 2, \dots, L$ 。当满足时说明可以找到一个线性的分类超平面，使两类样本完全分开，即

$$g(x) = w \cdot x_i + b = 0$$

求解分类超平面就是找到给定训练样本的权值  $w$  和阈值  $b$  的最优值，可归结为以下二次规划问题：

$$\begin{cases} \min & \frac{1}{2} \|w\|^2 \\ \text{s.t.} & y_i(w \cdot x_i + b) - 1 \geq 0, i = 1, 2, \dots, L \end{cases} \quad (8)$$

对于线性不可分的情况下，即不满足  $y_i(w \cdot x_i + b) - 1 \geq 0, i = 1, 2, \dots, L$  时，需要在条件中加入松弛项  $\xi_i$ ，变成  $y_i(w \cdot x_i + b) - 1 + \xi_i \geq 0, i = 1, 2, \dots, L$ 。为了在错分样本比例和算法复杂度之间寻找折衷，可以引入一个指定的常数  $C$  作为惩罚因子，控制对错分样本的惩罚程度。即

$$\begin{cases} \min & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^L \xi_i \\ \text{s.t.} & y_i(w \cdot x_i + b) - 1 \geq 0, i = 1, 2, \dots, L \end{cases} \quad (9)$$

对于线性不可分问题，可通过引入一个非线性映射函数  $\phi$ ，把神经元样本特征向量映射到某个高维空间，在高维空间将其转化为线性分类问题。任意函数  $K(x_i, x)$  满足 Mercer 条件[3]，都可以作为核函数替代原空间的内积函数，并且  $K(x_i, x) = \phi(x_i) \cdot \phi(x)$ 。

根据 Mercer 条件，在最优分类面中采用不同的内积函数  $K(x_i, x)$  就可以实现某一非线性变换的线性分类。引入核函数后，以上各式中的内积都可以用核函数代替。

## (2) 多于两类的神经元分类模型

然而传统的支持向量机是针对二分类提出的，在有多于两类的神经元投赞成票的情况下就要进行多分类。这里使用多分类支持向量机。现阶段流行的多分类支持向量机方法有两种：二叉树多类分类方法（“一对多”算法）和投票机制的多类分类方法（“一对一”算法）。

二叉树多分类向量机的基本思想：对于  $K$  类的训练样本，训练  $K-1$  个支持向量机，第 1 个支持向量机以第 1 类样本为正的训练样本，将第 2, 3, ...,  $K$  类训练样本作为负的训练样本训练向量机 1。第  $i$  个支持向量机以第  $i$  类样本为正的训练样本，将第  $i+1, i+2, \dots, K$  类训练样本作为负的训练样本训练向量机  $i$ ，直到第  $K-1$  个支持向量机将以第  $K-1$  类样本作为正样本，以第  $K$  类样本为负样本训练。



由于二叉树多分类向量机也是分层的分类机制，分类次数少但是带有一定的“偏见性”，在这里当使用特征判别函数投票不能得出唯一结果时我们选用投票机制的多分类方法进行剩余结果进行再分类。

对于  $K$  类的神经元样本，假设样本为  $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_L, y_L)\}$ 。其中  $L$  为训练样本集规模， $x_i \in \mathbb{R}^{20}$  为 20 维特征参数构成的特征向量， $y_i \in \{1, 2, \dots, k\}$  为第  $i$  类样本对应的类别等级。投票机制的多分类向量机是在  $K$  类样本两两之间共建立  $\frac{K * (K - 1)}{2}$  个支持向量机  $SVC_{ij} (i \neq j)$ 。其问题转化为

$$\begin{cases} \min_{w_{ij}, b_{ij}, \xi_{ij}} \frac{1}{2} \|w_{ij}\|^2 + C \left( \sum_{t=1} \xi_{ij}^t \right) \\ s.t. \quad [w_{ij} \cdot x_t + b_{ij}] - 1 + \xi_{ij} \geq 0, \text{ if } (y_t = i) \\ \quad [w_{ij} \cdot x_t + b_{ij}] - 1 + \xi_{ij} \leq 0, \text{ if } (y_t = i, \xi_{ij} \geq 0) \end{cases} \quad (10)$$

在分类过程中每运行一个向量机  $SVC_{ij}$  都进行一次投票即判别将此待分类样本归入哪一类。最终统计各类得票多少，待分类样本归入得票多的类。至此模型进入多赞成票情况算法分支的运算结束。

#### 4.3.3 分类模型求解

对于线性可分情况下的支持向量机模型目标函数转化而来的二次规划问题：

$$\begin{cases} \min \quad \frac{1}{2} \|w\|^2 \\ s.t. \quad y_i(w \cdot x_i + b) - 1 \geq 0, i = 1, 2, \dots, L \end{cases} \quad (11)$$

可以定义 Lagrange 函数：

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^L \alpha_i \{y_i[(w \cdot x_i) + b] - 1\} \quad (12)$$

其中  $\alpha_i > 0$  为 Lagrange 系数，问题转换为对  $w$  和  $b$  求 Lagrange 函数极小值的问题。分别对  $w$  和  $b$  求偏微分并令它们都等于 0，就可以把问题转化为较简单的对偶问题：

$$\begin{cases} \max \quad Q(\alpha) = \sum_{i=1}^L \alpha_i - \frac{1}{2} \sum_{i,j=1}^L \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\ s.t. \quad \alpha_i \geq 0 \quad (i = 1, \dots, L), \quad \sum_{i=1}^L y_i \alpha_i = 0 \end{cases}, \quad (13)$$

通过求解上述二次规划的对偶问题,即可得到对应的最优解 $a^*$ 。最优分类函数即最优分类超平面函数为:

$$f(x) = \text{sgn}\{w^* \cdot x + b^*\} = \text{sgn}\left\{\sum_{i=1}^L a_i^* y_i (x_i \cdot x) + b^*\right\} \quad (14)$$

对应线性可分的情况的解法也可以找到线性不可分的情况下的二次规划的最优解,从而找到线性不可分的情况下的最优分类超平面。

在问题三中,提出了新的“基于综合特征判别和多分类支持向量机(Multi-Class Support Vector Machine)相结合的带权重投票模型”的分类方法。使用附录 A, 附录 C 的神经元数据对模型进行训练之后,使用模型对附录 A, 附录 C 神经元和部分从本题目数据来源网站 NeuroMorpho. Org 的神经元数据对模型进行了检验。结果如下:

表格 8 改进分类模型判别结果

项目	类别	运动神经元	普肯野神经元	锥体神经元	双极中间神经元	三极中间神经元	多极中间神经元	感觉神经元	其他
附录 A、C 中数据个数		6	7	8	6	6	10	8	无
A、C 判断正确个数		6	7	8	6	6	10	8	无
A、C 判断正确率		100%	100%	100%	100%	100%	100%	100%	无
新样本数据个数		15	11	10	8	17	15	30	10
新样本判断正确个数		13	11	8	7	14	13	30	4
新样本判断正确率		86.7%	100%	80%	87.5%	82.4%	86.7%	100%	40%

检验结果表明模型效果良好,特别对于普肯野神经元和感觉神经元两类具有较显著的特征参数的神经元具有很好较高的分类成功率。但是对于新的类型的神经元,由于缺乏特征判别函数和样本训练,识别其为新的类别的神经元的成功率不高。

#### 4.3.4 新的命名方法和建议

建议 1:

附录 A、C 中的普肯野神经元、新命名的“平面鹿型神经元”(附录 B 中 15、16 号神经元)以及“平面爬山虎型神经元”(附录 B 中 17、18 号神经元)均具有明显的平面分布特征。因此建议用“平面型神经元”这一新的神经元名称命名神经元空间形态为平面型的一个类,其中包括普肯野神经元、“平面鹿型神经元”、“平面爬山虎型神经元”三个子类,如图 14 所示:

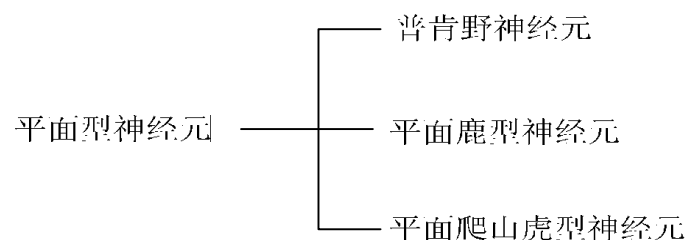


图 14

由于前面已呈现过附录 A 和附录 B 中这三类神经元的三维空间图形，在此仅以附录 C 中的普肯野神经元为例说明，绘出三维图像后从正面和侧面两个角度可以直观的观察得到平面形这一特点，如图 15、16 所示。

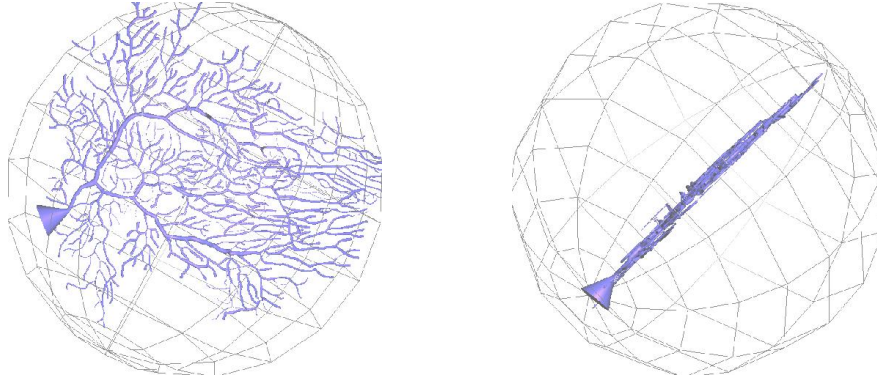


图 15 附录 C 普肯野神经元正面图 图 16 附录 C 普肯野神经元侧面图

建议 2:

根据神经元的空间形态对神经元进行分类，类名以能够明确反映神经元形态特征为佳。通过前面问题的求解，发现运动型神经元和新命名的“无胞体运动型神经元”皆成放射状，而这两类的区别在于胞体的有无。故建议将这两类神经元合命名为“放射状神经元”，两子类改名为有胞体放射状神经元、无胞体放射状神经元，见图 17。

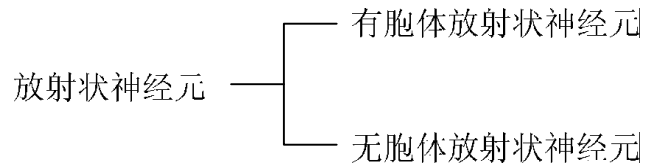


图 17

#### 4.4 问题四

##### 4.4.1 问题分析

用问题一提出的“分级多层”分类方法不能区分附件 A 中的两类普肯野神经元，二叉分类最终仅能做到将两组神经元样本划入普肯野神经元这一大类中。用问题三提出的“基于特征判别和多分类支持向量机相结合的带权重圆桌投票模型”之后，得到的结果有所改善，两组神经元样本得到的票数有所差异，各自在小范围内波动，但是二者的票数区域有重叠，说明问题三提出的模型还是不能很好地解决问题四。问题一和问题三的模型不能很好地区分同种神经元的不同物种，其原因是：（1）以上两个分类方法是在仅考虑同类神经元共同特征的前提下提出的，不同物种的同类神经元被不加区分地当作同类神经元的不同样本对待；（2）在特征判别函数模型中，仅仅提取其所有同类神经元的共性而并未考虑不同物种样本之间的个性差异；因此解决问题四的效果不佳也是理所应当的。针对问题四，本文引入了一种新的用于解决区分不同物种同类神经元的模型和方法，即同特征参数搜索匹配模型。

##### 4.4.2 模型建立

4.1 中定义的特征参数的波动率  $\gamma$ ，可以作为一组神经元细胞样本单个参数稳定性的表征量，某参数波动率  $\gamma$  越小，该组样本在此参数上具有的共性越大，即此参数越能代表样本所属类别的特征。特征参数的波动率是特征波动分析模型的基础。

定义 6 特征稳定性标志  $\Gamma_{i,j}$  ( $i$  为样本组序号,  $j$  为参数序号)

$$\Gamma_{i,j} = \begin{cases} 1, & \gamma_{i,j} \leq \gamma\_upperbound, \\ 0, & \gamma_{i,j} > \gamma\_upperbound, \end{cases} \quad i \in \{1,2\}; j \in \{1,2,...,20\} \quad (15)$$

当某参数波动率小于等于某阈值上限  $\gamma\_upperbound$  时认定该参数稳定。

定义 7 特征极其不稳定标志  $T_{i,j}$  ( $i$ 为样本组序号,  $j$ 为参数序号)

$$T_{i,j} = \begin{cases} 1, & \gamma_{i,j} \geq \gamma\_lowerbound, \\ 0, & \gamma_{i,j} < \gamma\_lowerbound, \end{cases} \quad i \in \{1,2\}; j \in \{1,2,...,20\} \quad (16)$$

当某参数波动率大于等于某阈值下限  $\gamma\_lowerbound$  时认定该参数极其不稳定。

定义 8 参数特征带重叠率  $\eta_{i,j}$

$$\eta = \frac{OverLap(CB\_X\_Y1, CB\_X\_Y2)}{Total(CB\_X\_Y1, CB\_X\_Y2)} \quad (17)$$

其中:  $OverLap(CB\_X\_Y1, CB\_X\_Y2)$  为样本一特征参数特征带

$CB\_X\_Y1$  和样本二特征参数特征带

$CB\_X\_Y2$  的重叠区间长度

二者的总长度  $Total(CB\_X\_Y1, CB\_X\_Y2)$  即长度和减去重叠区长度。

参数特征带重叠率大小说明两个样本此参数的相似程度。

模型算法流程图如下:

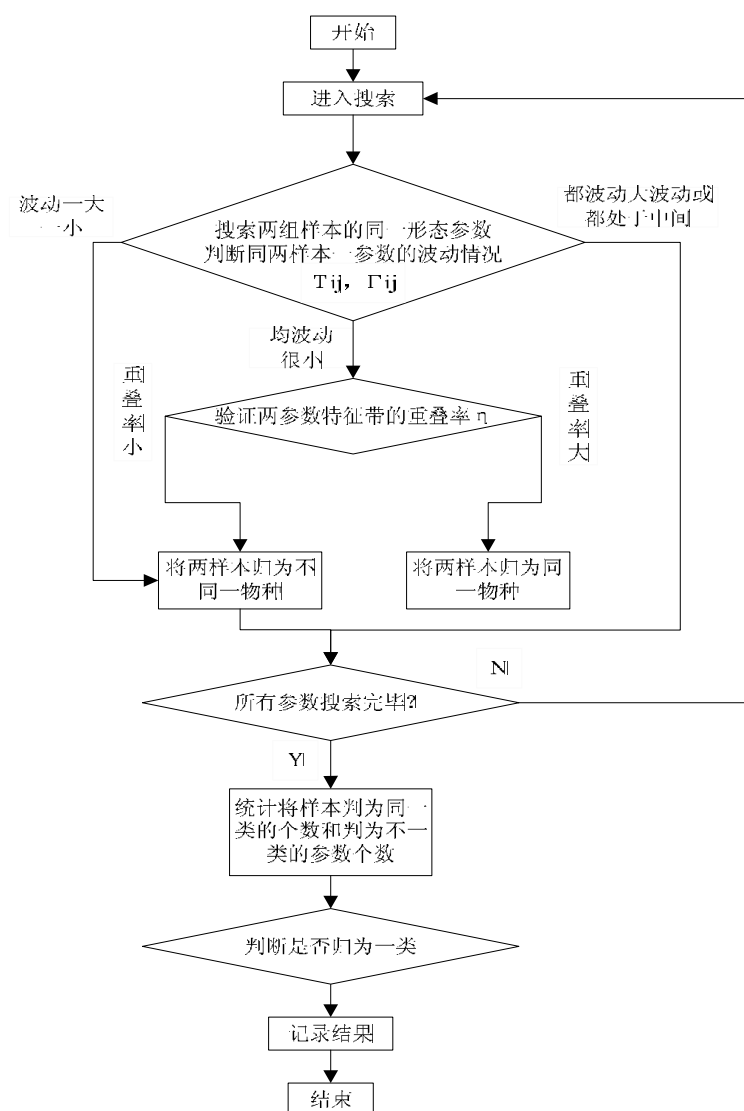


图 18 模型算法流程图

#### 4.4.3 模型求解

根据上述模型，利用附录 C 中的样本，得到可用于判定特征差异的参数及其波动率如表 9。

表格 9 特征差异的参数及其波动率

类别 \ 参数	Soma_Surface	Surface	Volume
pig Purkinje cell 1	0	50877	27291.9
pig Purkinje cell 2	0	34666.2	19128.8
pig Purkinje cell 3	0	42520	22380.8
AVG(均值) pig Purkinje	0	42687.73	22933.83
mouse Purkinjecell1	799.425	8336.16	1742.46
mouse Purkinjecell2	9.90E+02	9805.41	2032.62
mouse Purkinjecell3	921.607	23086.4	7776.34
AVG(均值) mouse Purkinje	903.5913333	13742.66	3850.473

表格 9 说明了不同种类的动物，他们的相同神经元的 Soma\_Surface 、Surface 和 Volume 等因素在数值上都有较大差别，所以可以根据这些因素的值来高概率的进行相同神经元的分类。

表格 10 pig Purkinje cell 在波动率  $P \leq 10\%$  的比较稳定的参数

类别 \ 参数	N_bifs	N_branch	Contraction	Partition_asymmetry	Bif_ampl_local	Bif_ampl_remote
pig Purkinje cell 1	471	944	908	238	37319	32103
pig Purkinje cell 2	419	840	804	208	33907	31068
pig Purkinje cell 3	416	835	806	204	32680	30281
AVG(均值)pig 普肯野	435	873	839	217	34635	31151
STDEV(标准差)pig 普肯野	31	61	60	19	2404	914
波动率 $\Upsilon$	7%	7%	7%	9%	7%	3%

表格 11 mouse Purkinje cell 在波动率  $P \leq 10\%$  的比较稳定的参数

类别 \ 参数	N_bifs	N_branch	Length	Contraction	Partition_asymmetry
mouse Purkinjecell1	369	740	4980.15	696.65	192.565
mouse Purkinjecell2	357	717	5497.91	667.143	190.821
mouse Purkinjecell3	342	686	5905.34	639.801	181.787
AVG(均值)mouse 普肯野	356	714.3333	5461.133	667.8647	188.391
STDEV(标准差)mouse 普肯野	13.52775	27.09859	463.6901	28.43137	5.785326
波动率 $\Upsilon$	4%	4%	8%	4%	3%

表格 10 列出了不同的 pig Purkinje cell 在 N\_bifs、N\_branch、Partition\_asymmetry 等因素上数值的比较。

表格 11 列出了不同的 Mouse Purkinje cell 在 N\_bifs、N\_branch、Length、Contraction 等因素上数值的比较

从波动率上可以得出相同物种不同神经元的某些因素比较稳定，因此可以通过对某一未知物种同类神经元的数值分析来判断是否属于该类动物。

根据附录数据，计算得两类动物的神经元形态参数见表格 12。

表格 12 两类动物的神经元形态参数对照

cat	N_bifs	N_branch	Height	Depth	Diameter	Length	Surface	Contraction	Bif_ampl_remote
	171	367	1326270	1275500	2821.47	103228	676400	336.253	8119.31
	175	373	1364820	1397910	2291.31	113502	499671	341.307	9721.59
	181	377	1484180	1430680	2119.02	117029	543917	341.335	10317.4
	151	331	1197100	1316580	2827.45	107519	593005	291.912	8651.55
均值	169.5	362	1343092.5	1355167.5	2514.8125	110319.5	578248.25	327.70175	9202.4625
标准差	13	21.071308	118285.77	71565.738	364.41099	6144.8588	75728.785	23.979147	1626.7032
标准差/均值	8%	6%	9%	5%	14%	6%	13%	7%	18%
rat	N_bifs	N_branch	Height	Depth	Diameter	Length	Surface	Contraction	Bif_ampl_remote
	29	75	719353	338504	1595.99	5890.41	16128.5	50.5199	1221.72
	18	47	514624	375003	1694.11	4599.24	16469.5	27.7465	1446.71
	25	59	1001390	358785	1670.86	7666.82	18713	46.3646	1382.77
	18	47	713180	345315	1429.72	6574.4	18376	36.1741	1323.98
均值	22.5	57	737136.75	354401.75	1597.67	6182.7175	17421.75	40.201275	1343.795
标准差	5.4467115	13.266499	200193.15	16113.471	119.53727	1284.3702	1311.1312	10.259957	95.578044
标准差/均值	24%	23%	27%	5%	7%	21%	8%	26%	7%

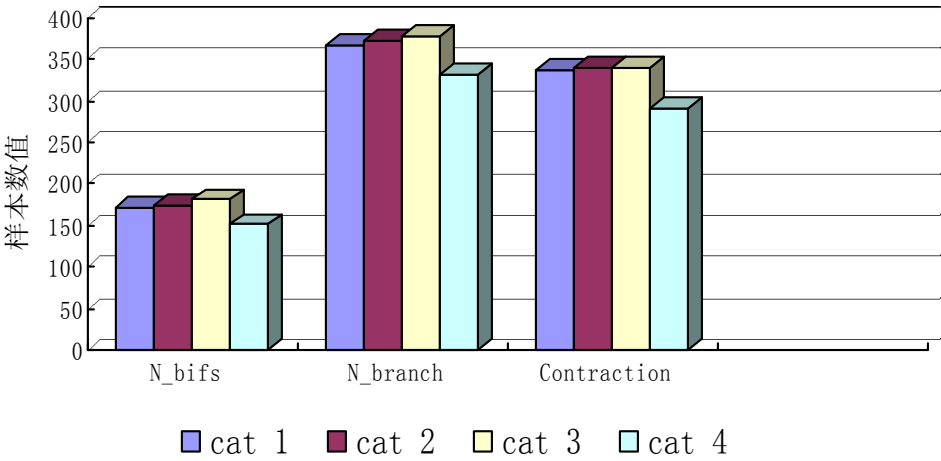


图 19 cat 的同种神经元的稳定因素图

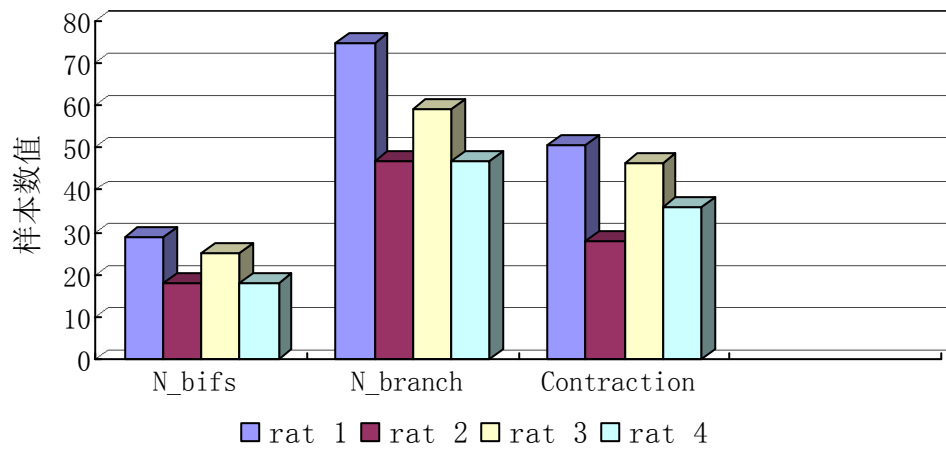


图 20 与 cat 的神经元稳定因素相对应的 rat 的神经元因素图

图 19、图 20 表明 cat 同种神经元在 N\_bifs、N\_branch、Contraction 等因素上都有一定稳定性，即它们在数值上相近。而 rat 的对应属性在稳定性和数值上都与 cat 的有一定差别，可以以此为依据判别出不同动物的相同神经元。

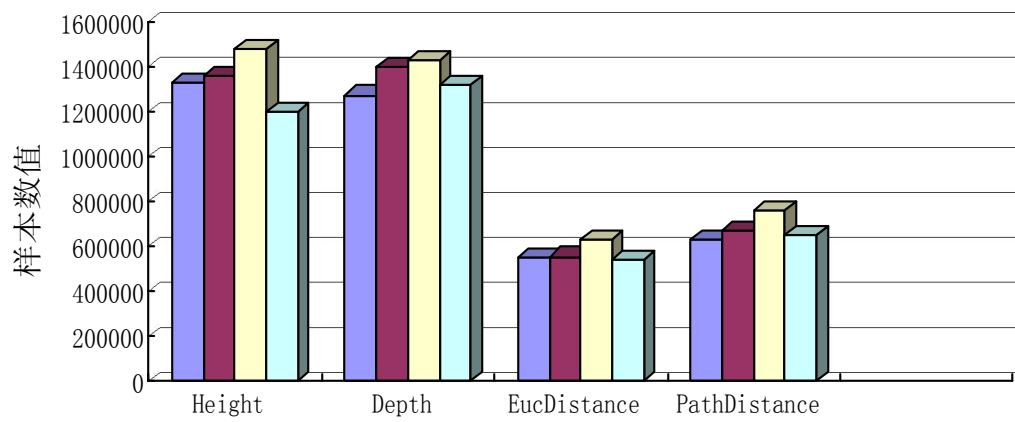


图 21 cat 的同种神经元的稳定因素图



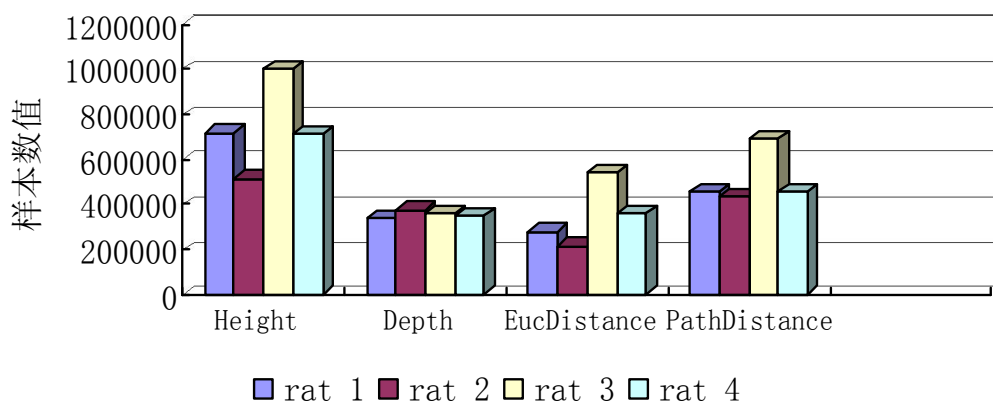


图 22 与 cat 的神经元稳定因素相对应的 rat 的神经元因素图

图形分析:

图 19-图 22 中表示不同类物种 cat 和 rat 的神经元因素比较, 柱形图表示了 cat 的同种神经元的相对稳定性因素, 以及与其相对应的 rat 的神经元因素图, 从图中可以看出在 cat 的神经元某些因素相对稳定的情况下, rat 的相对应因素波动的强度, 如果 rat 的某一相对应因素也处于稳定状态时他们的样本数值相差也比较明显。所以可以从多因素判别角度出发, 当有多因素的值都落入稳定因素的接受域内则可以以大概率事件属于该物种的神经元。

## 4.5 问题五

### 4.5.1 问题分析

神经元的实际形态随着时间的流逝, 树突和轴突不断生长变化, 给本已困难的神经元识别归类问题增加了新的难度。为了分析和预测神经元在不同生长阶段的空间形态特征, 仅仅依靠附录中给出的数据是无法解决问题的, 必须获得具有时间标识且处在时间轴不同位置的同物种神经元数据。本文利用 NeuroMorpho.Org 网站上提供此项神经元库的检索服务, 按神经元的成长阶段即 Young (青少年)、Adult (成年) 和 Old (老年) 三个区间或按照年龄大小如天数、月数和年数范围打包下载同物种同神经元的 SWC 数据。

选择 Rat (老鼠) 的 Pyramidal (锥体) 神经元作为研究对象, 得到的神经元样本年龄、成长阶段和样本数如表格 13、14。

表格 13 样本类型成长阶段及各样本容量

成长阶段	青少年	成年	老年
样本容量	10	8	15

表格 14 样本类型成长年龄及各样本容量

成长年龄	1 天	10 至 15 天	30 至 61 天	105 至 129 天
样本容量	28	9	12	52

神经元细胞的成长过程也是“新生→青壮年→老年→死亡”的过程, 其形态特征在此过程中不断变换。在此过程中有些参数反映细胞成长的过程, 有些参数反映此类细胞的共性。将各组样本使用 L-Measure 软件进行数据转换, 得到各样本的 20 个空间形态

参数描述，而后对各参数做统计分析。将各组样本平均值合并为新的数据集，做统计分析，设置波动率阈值，筛选出较大变动的参数和稳定的参数，均值统计效果如图 23 。

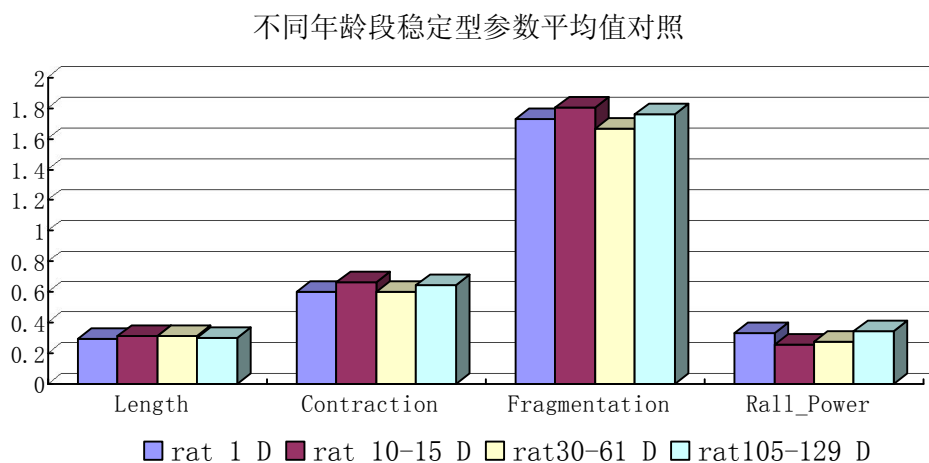


图 23 不同年龄段稳定型参数平均值对照图

图 23 中的参数在锥形神经元的生长过程中保持基本不变，在使用波动率进行筛选之后说明这些特征是锥形神经元的特征参数。

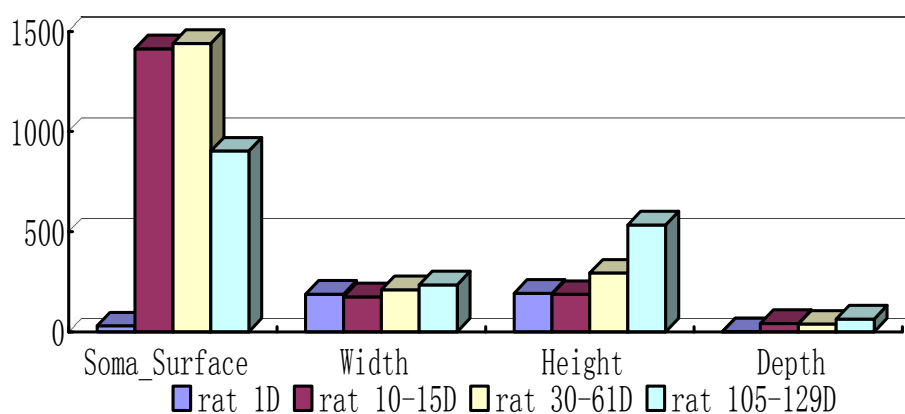


图 24 不同年龄段非稳定型参数平均值对照图

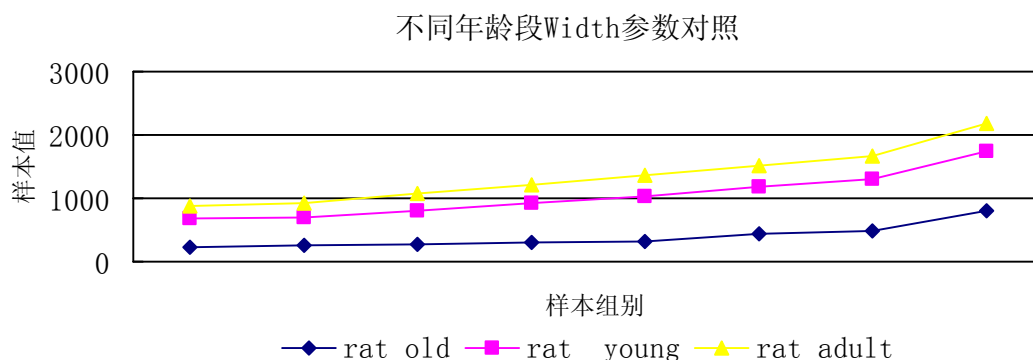


图 25 不同年龄段 Width 参数对照图

图 24 表明锥形神经元的细胞体在神经元的成长、衰老、死亡过程中，经历了快速

增长、保持平稳、收缩死亡四个阶段。神经元的宽度、高度和深度参数没有下降的过程，说明锥形神经元的树突在神经元的存活过程中一直保持先快后慢的增长，即锥形神经元的树突枝杈空间范围也保持先快后慢的增长。

在选择用于表征神经元生长特征的参数过程中，先以波动率为依据选择，选择完成后验证不同组神经元的这项参数取值范围是否有重叠，有重叠时说明这几组样本的此项参数不能表征神经元成长特性，放弃使用此参数作为检验神经元成长的特征参数。取三组每组 8 个样本进行 Width 参数比对，效果如图 25。图 25 表明不同年龄组的 Width 参数可以表征神经元成长的过程。

基于上述分析，建立针对神经元生长形态稳定型参数和非稳定型参数特征预测的数学模型。

#### 4.5.2 模型建立

##### 1. 稳定型参数的预测模型

稳定型参数可以借鉴问题一中特征带选取的方法。假设稳定性参数满足正态分布，选取样本稳定型参数均值作为稳定型参数值。

##### 2. 非稳定型参数的预测

非稳定型参数分为两类，一是“增长→平稳→下降”型，一是“增长→平稳”型。

###### (1) 增长阶段模型

非稳定型参数的增长阶段的预测采用阻滞增长模型。随着细胞的生长，阻滞作用会越来越大，阻滞增长模型就是考虑到这个因素后，对经典的指数增长模型修改后得到的。

建立细胞增长函数  $\varphi(\beta)$

阻滞作用体现在对细胞增长率  $\varphi(\beta)$  的影响上，使得  $r$  随着神经元特征参数  $\beta$  的增大而下降。若将  $\varphi$  表示为  $\beta$  的函数  $\varphi(\beta)$ ，则它应是减函数。于是有：

$$\frac{d\beta}{dt} = \varphi(\beta)\beta, \quad \beta(0) = \beta_0 \quad (19)$$

本文取  $\varphi(\beta)$  为  $\beta$  的线性函数，即

$$\varphi(\beta) = r - s\beta \quad (r > 0, s > 0) \quad (20)$$

这里  $\varphi$  称固有增长率，表示神经元特征参数很小时（理论上是  $\beta=0$ ）的增长率。为了确定系数  $s$  的意义，引入神经元特征参数最大值  $\beta_m$ 。当  $\beta = \beta_m$  时神经元特征参数不再增大，

及增长率  $\varphi(\beta_m) = 0$ ，带入上式得  $s = \frac{r}{\beta_m}$ ，于是公式 (20) 为

$$\varphi(\beta) = r(1 - \frac{\beta}{\beta_m}) \quad (21)$$

(21) 式表示增长率  $\varphi(\beta)$  与神经元特征参数尚未实现增长的部分的比例  $(\beta_m - \beta) / \beta_m$  成正比，比例系数为固有增长率  $r$ 。

将(21)代入方程(19)得

$$\frac{d\beta}{dt} = r\beta(1 - \frac{\beta}{\beta_m}), \quad \beta(0) = \beta_0 \quad (22)$$

方程(22)右端的因子  $r\beta$  体现神经元特征参数自身的增长趋势, 因子  $(1 - \frac{\beta}{\beta_m})$  则体现了外界对神经元特征参数增长的阻滞作用。显然,  $\beta$  越大, 前一因子越大, 后一因子越小,

神经元特征参数增长是两个因子共同作用的结果。

如果以  $x$  为横轴、 $d\beta/dt$  为纵轴作出方程(22)的图形, 可以分析神经元特征参数增长速度  $d\beta/dt$  随着新的增加而变化的情况, 从而大致地看出  $\beta(t)$  的变化规律。

## (2) 平稳阶段模型

非稳定型参数的平稳阶段假定为参数值不随时间改变。故平稳阶段的参数预测采用最大值。

## (3) 下降阶段模型

对于“增长→平稳→下降”型非稳定型参数还需建立下降阶段模型。此阶段是神经元的衰老死亡阶段。此阶段模型采用加速下降模型。其模型与阻滞增长模型类似, 这里不再赘述。

## 4.5.3 模型求解

下图为老鼠 Soma\_Surface 参数随时间推移的预测效果为例展示预测效果。

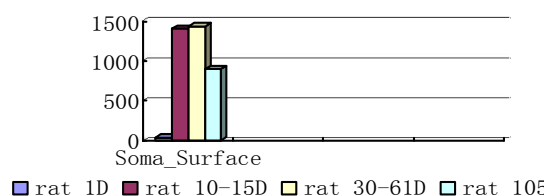


图 26 不同年龄段的老鼠样本 Soma\_Surface 因素平均值的对照

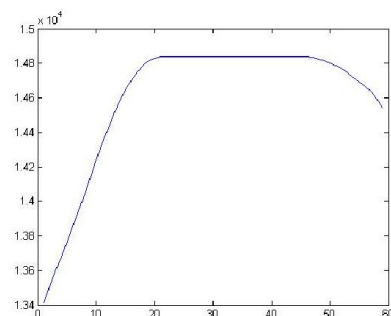


图 27 预测效果图

# 5 模型的优缺点与改进方向

## (1) 神经元的分级多层二叉分类模型

优点: 实际操作简单, 易于实现, 与人的思维方式接近。

缺点: 是带有一定“偏见”的分类方法, 即当特征判别函数有先验知识错误或待分类样本特征与模型的建立和训练样本相差较大时导致误分。

改进方向和解决方法: 寻找综合多指标、多参数的特征判别函数, 去除“偏见性”, 从而达到减少误分概率的目的。

## (2) 神经元的“基于综合特征判别和多分类支持向量机相结合的带权重投票模型”

优点：引入了综合性的特征判别函数指标和投票机制，大大消除了“偏见性”。在排除部分特征明显的分类之后再使用支持向量机更能保证支持向量机算法的分辨成功率。

缺点：特征判别函数中各参数判别的权重分配需要大量的计算和模型的自学习，技术要求高，计算复杂，实际操作困难。各类神经元投票的“话语权”分配也是难题。

改进方向和解决方法：使用经验权重分配方式。

### （3）不同物种同类神经元的“同特征参数搜索模型”

优点：算法简单易于实现，具有较高的分辨成功率。

缺点：模型应用的前提是已知待分的两组神经元样本属于同类神经元，要区分的是两组数据是否是同类物种，无法判别样本是否属于其他已知类和新类。

改进方向和解决方法：配合上述（2）中模型的方法使用。

### （4）神经元生长预测模型

优点：模型思路清晰，易于理解，将各个能够反映神经元特征参数分为稳定型和非稳定型两类的本质是将反映神经元生长规律和反映神经元本类固有特征参数分开。非稳定型参数又细分为两类，反映出了两种随神经元生长呈不同变化规律的参数特征。增长阶段的阻滞增长的模型更加贴近于真实情况。稳定型参数用以表征本类固有特性，故使用此预测模型可以更大程度地避免导致预测结果进入其他类。

缺点：稳定型参数模型和非稳定型参数的平稳阶段模型太过简单。现实中细胞包括神经元的生长规律受多种因素影响，难以准确预测。

改进方向和解决方法：使用统计分析的方法改善上述两个模型的算法。

## 参考文献

- [1]孙荣恒，应用数理统计，北京：科学出版社，2003。
- [2]边肇祺，模式识别，北京：清华大学出版社，2000。
- [3]李秀珍，多分类支持向量机在滑坡稳定性判别中的应用，吉林大学学报，140（3）：631-637。
- [4]郭小荟，基于 Matlab 的支持向量机工具箱，计算机应用与软件，124（12）：57-59，2007。
- [5]朱艳伟，基于主成分分析的支持向量机划分地震相，中国知网，2010-9-18。
- [6]Steve R. Gunn, Support Vector Machines for Classification and Regression, UNIVERSITY OF SOUTHAMPTON, 中国知网，2010-9-18。
- [7]Yi Liu and Yuan F. Zheng, One-Against-All Multi-Class SVM Classification Using Reliability Measures, The Ohio State University, 中国知网，2010-9-18。
- [8]张怀强，构建限制性条件下微生物群体生长模型时的问题，应用与环境生物学报，11（5）：595-599，2005。
- [9]Anthony Krantis, Monitoring Primary Neuron Growth and Development on Microruled Cell Culture Plates: Evaluation Utilizing Cryopreserved Rat DRG Neurons, 中国知网，2010-9-19。