全国第七届研究生数学建模竞赛

题 目 神经元的形态分类和识别

摘 要：

本文根据已知神经元的空间形态数据，提取出不同类型神经元的几何形态特征，再利用这些特征，建立神经元形态分类模型。根据所建立的形态分类模型，对神经元进行分类识别与分析。

针对问题 1，根据已知神经元的空间形态数据，提取出不同类型神经元的几何形态特征。建立起基于改进的 SVM 决策树分类模型，能根据各类神经元的可分离程度，依次将神经元的类别分割出来，使可能出现的错分尽可能地远离树根，从而大幅地提高分类的正确率。

针对问题 2，利用问题 1 的模型判断附录 B 中 20 个神经元的类型，发现仅将神经元分成 5 类是不够合理的，需要定义新的神经元类别。

针对问题 3，建立一个基于 SOM(Self-Organizing Map)聚类分析模型，并用于解决神经元的分类问题。与问题 1 模型相比，该模型可以解决没有确定神经元类别个数的分类问题。SOM 聚类分析还能实现高维数据模式的低维可视化，使得聚类结果具有可读性。并且通过利用 SOM 聚类分析，我们提取出了能给生物学家命名神经元提供参考的几何特征。

针对问题 4，应用问题 3 的模型，判断出不同动物神经系统中同一类神经元的形态特征存在的区别。

针对问题 5，通过构造出一个神经元随机生长模型，从而较为准确地预测神经元形态的生长变化。并且在生长变化过程中，问题 3 模型得到的 8 个重要特征并未发生明显变化。

关键字：神经元；几何形态；SVM 决策树；SOM 聚类分析

目录

1 问题重述

为了建立神经信息学数据库的共同标准，并进行多学科整合分析大量数据，加速人类对脑的认识，人们提出了人类脑计划（Human Brain Project, HBP）。而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。神经元的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，轴突的的生长变化，神经元的几何形态千变万化。

对神经元特性的认识，最基本问题是神经元的分类。目前，关于神经元的简单分类法已经有很多。但由于神经元的几何形态的复杂性，且不同组织位置，神经元的类别和形态，可能变化也很大。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。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。

根据附录以及 NeuroMorpho.Org 提供的大量神经元几何形态数据，完成以下任务：

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、判断附录 B 的 20 个神经元类型。在给出的数据中，分析是否需要引入或定义新的神经元名称。

3、提出一个神经元分类方法，将所有神经元按几何特征分类，并给生物学家提供神经元命名的一些建议。

4、按照所建立的神经元形态分类方法，判断在不同动物神经系统中同一类神经元的形态特征的区别。

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。对神经元形态的生长变化进行预测，并分析这些形态变化对所选取几何形态特征的影响。

2 问题分析

根据问题的描述可知，我们需要解决一个神经元的形态分类和识别问题，具体如下：

1、一个神经元根据形态空间结构可以离散为很多房室，这些房室用 A.SWC 格式文件描述。根据这些空间几何数据，如何定义和提取一个合理的几何特征，是首先要解决的问题。这些几何特征应当尽可能地刻画神经元的整体或局部的形态空间结构。

2、由于大多数所定义的几何特征相关性强，存在冗余现象，需要我们分析各种几何特征的相关性，并选择出相对独立的几何特征用于分类。

3、建立一个能将神经元形态分成 5 类的分类器模型，且该模型要适用于样本和属性相对较少的条件。

4、建立一个神经元形态分类模型，分析神经元中各个几何特征对其所属类

别的影响程度，并由此给生物学家为神经元的命名提出建议。

5、分析不同动物神经系统中同一类神经元的形态特征存在差别对所建立模型的影响。

6、提出一种预测模型，对神经元形态的生长变化进行预测。并分析这些形态变化对所选取几何形态特征的影响。

3 模型假设

根据已知问题的描述，在建立模型之前，我们提出以下假设：

1、特征选取只考虑神经元的几何形态。

2、房室的类型，只在计算几何特征时使用，不作为本模型的一个特征。

3、利用房室对神经元空间几何形态进行描述是准确的。

4、同一类神经元在一定时期，其几何形态结构保持相对稳定。

4 模型的建立与求解

4.1 问题 1 模型与求解

4.1.1 神经元几何特征的定义

根据神经元的空间几何数据，可以定义许多不同的神经元几何特征。例如 L-measure 软件可以计算 40 多种几何特征的值[1]。本文根据特征所刻画神经元形态结构的部分，将几何特征分为整体特征与局部特征两类。如胞体表面积、干的数目、树干锥度和分支角度为神经元的局部几何特征，而宽度、高度、深度和表面积为神经元的整体几何特征。所定义的几何特征应该能更好地反映神经元的形态结构。对此，本文选择以下两个重要的几何特征进行说明，其他特征的定义可以参考相关文献。

1、分形维数（Fractal dimension）[2]

分形维数反映了复杂形体占有空间的有效性，它是复杂形体不规则性的量度。利用分形维数来定义神经元的几何特征可以更好地反映神经元的形态结构。若一个分形含有 n 个相似的部分，每一个部分的线度是整体的$\frac{1}{m}$，则分形维数定义为：$\log_m n=\frac{\ln n}{\ln m}$。分形维数的表示有许多方法，本文使用以胞体为中心半径为R的球内总长度为N的所有分支所遵循的幂律来表示，如图 1 所示。给定的与胞体的距离R的球内，分支总数为$n(R)\sim\frac{dN(R)}{dR}\sim R^{D-1}$。

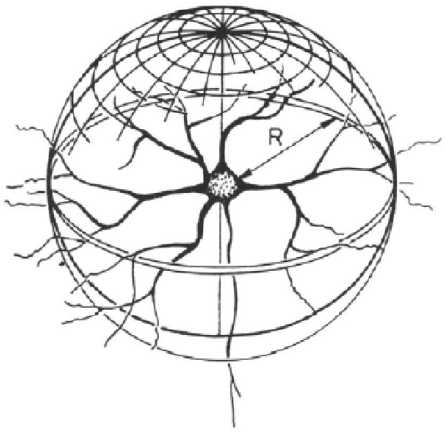

图 4-1 神经元中半径为 R 的分形维数示意图

2、分区不对称（Partition Asymmetry）

分支的两树子树所包含的分叉数 n_1 和 n_2，则 $\left|\frac{n_1-n_2}{n_1+n_2-2}\right|$ 的平均值就是分区不对称的值。从图 4-2 的示图，可以看出分区不对称，用来反映分支上两颗子树不对称的程度。

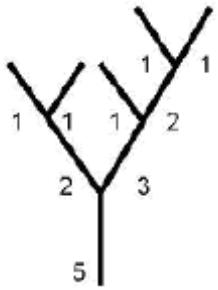

图 4-2 分区不对称示意图

4.1.2 神经元几何特征的选取

利用简单的统计分析从上述定义的 40 多个特征中选取有效的几何特征。利用 L-measure 软件可计算出附录 A 和 C 中的所有样本几何特征的值。但由于大多数所定义的几何特征相关性强，存在冗余现象，需要我们分析各种几何特征的相关性，并选择出相对独立的几何特征用于分类。因此我们对 40 多个特征进行相关的统计检验，并选取相关性较小的 26 个几何特征，如表 4-1 所示。

表 4-1 所选取的几何形态特征

编号	几何形态特征	Feature	定义说明
s1	干的数目	Number of Stems	与胞体相连的分支数目
s2	分叉数目	Number of Bifurcation	所有分叉数目
s3	宽度	Width	神经元所占空间的最大宽度

s4	高度	Height	神经元所占空间的最大高度
s5	深度	Depth	神经元所占空间的最大深度
s6	直径	Diameter	分支直径平均值
s7	长度	Length	所有分支的总长度
s8	欧式距离	Euclidean Distance	从尖端到胞体的欧式距离总和
s9	分支度	Branch Order	从胞体到尖端分叉的总数
s10	段锥度	Segment taper	每个分段的树干锥度总和
s11	单位锥度率	Unit taper rate	单位长度的树干锥度
s12	收缩率	Contraction	欧式距离与路径距离的比率
s13	碎片	Fragmentation	重建点总数
s14	子树比率	Daughter ratio	分支子树的直径比
s15	父子比率	Parent-child ratio	分支主干与
s16	分区不对称	Partition Asymmetry	参见上文几何特征的定义
s17	罗尔权	Rall's power	罗尔权的最佳适应参数
s18	近端分叉角度	Bifurcation angle Local	分支的近端分叉角度和
s19	远端分叉角度	Bifurcation angle Remote	分支的远端分叉角度和
s20	近端房室相邻角度	Bif_tilt_Local	分支的近端房室相邻角度和
s21	远端房室相邻角度	Bif_tilt_Remote	分支的远端房室相邻角度和
s22	近端扭矩	Bif_torque_Local	分支的近端扭矩和
s23	远端扭矩	Bif_torque_Romote	分支的远端扭矩和
s24	最后分叉直径	Last parent diameter	所有最后分叉直径总和
s25	螺旋度	Helix	神经元的螺旋度
s26	分形维数	Fractal dimension	参见上文几何特征的定义

4.1.3 建立改进的 SVM 决策树分类模型

利用以上所选取的几何特征，建立一个神经元空间形态分类模型。所建立的分类模型可以将神经元样本分成 5 大类（中间神经元可以又细分 3 类）。所以模型所要解决的是一个确定类别个数的多分类问题。将 SVM 和二叉决策树结合起来，构成 SVM 决策树[3]，是一种有效的分类方法 。本文将选择 SVM 决策树分类算法建立神经元形态分类模型。为了构造性能良好的决策树结构，根据各类神经元的可分离程度，依次将神经元的类别分割出来。

与通常的分类方法相比，SVM 决策树方法对于一个 N 值分类问题，需要寻找 $N-1$ 个最优分类面。随着训练的进行，需要的训练样本数逐渐减少。因此，在训练阶段，随着训练的进行，生成最优分类面所需要的训练时间逐渐减少。在分类阶段，该方法并不像通常的方法需要计算所有分类决策函数的值，它仅需要根据决策树的结构，计算所需要的分类决策函数值。该方法的缺点是如果在某个结点上发生分类错误，则会把分类错误延续到该结点的后续下一级结点上。因此，分类错误在越靠近树根的地方发生，分类性能就越差。为构造性能良好的决策树结构，可以考虑：将容易分或者不易产生错分的类先分割出来，然后再分不容易分的类。这样，就能够使可能出现的错分尽可能地远离树根。从所提供的神经元空间几何数据可以看出，不同类别神经元的可分离程度不一致。如普肯野神经元相对容易分离出，而锥体神经元与中间神经元比较难以区分。

针对 SVM 决策树分类器存在的问题，本文定义了基于类分布的神经元类间分离性程度，并将其扩展到核空间，建立改进的 SVM 决策树分类模型。

假设要进行 k 类神经元分类，神经元训练样本集由 $X_i, i=1,2,\cdots,k$ 组成。将类 i 和类 j 间的分离性程度定义为 seg_{ij}，

$$seg_{ij}=\frac{d_{ij}}{(\sigma_i+\sigma_j)}$$

其中，$d_{ij}(i,j=1,2,\cdots,k)$ 表示类 i 和类 j 中心间的距离：

$$d_{ij}=\left\|c_i-c_j\right\|$$

c_i 是根据神经元训练样本计算出的类中心：

$$c_i=\frac{1}{n_i}\sum_{x\in X_i}x, i=1,2,\cdots,k$$

其中 n_i 是类 X_i 的样本个数。另外 σ_i 为神经元的类方差，用来表示待分类的分布：

$$\sigma_i=\frac{1}{n_i-1}\sum_{x\in X_i}\left\|x_j-c_i\right\|, i=1,2,\cdots,k$$

如果所得类 i 和类 j 间的分离性程度 $seg_{ij}\geq 1$，则这两类神经元无重叠；如果 $seg_{ij}<1$，则两类神经元存在重叠。seg_{ij} 的值越大，则两类神经元的分离性程度越好。

类别 i 神经元与其余类的神经元之间的最小分离程度作为该类的分离性程度 seg_i，

$$seg_i=\min_{\substack{j=1,2,\cdots,k\\ j\neq i}}(seg_{ij})。$$

接着，最易分的神经元类别就是：

$$s=\arg\max_{\mathrm{i}=1,2,\cdots,\mathrm{k}}(seg_i), i=1,2,\cdots,k,$$

表示可分离性程度最大的神经元类别。

在定义了基于类分布的神经元类间分离性程度的基础上，我们将其扩展到核空间，并建立改进的 SVM 决策树分类模型如下：

设：在决策树各对结点生成的最优分类面是将一类和其余类分开。

计算特征空间中的分离性程度 seg_{ij}，组成分离性程度矩阵 SEG：

$$SEG=\begin{bmatrix} \text{lnf}\ seg_{12}\cdots seg_{1k} \\ seg_{21}\ \text{lnf}\cdots seg_{2k} \\ \cdots\quad\cdots\quad\cdots\quad\cdots \\ seg_{k1}\ seg_{k2}\cdots\ \text{lnf} \end{bmatrix}$$

其中同类间的分离性程度为无穷大。根据分离性程度矩阵 SEG，选择当前神经元样本 $X_i, i=1,2,\cdots,k$ 中，最容易分的类，并该类与其余各类进行 SVM 分类训练，得到最优分类面，构成决策树的结点。SVM 在很大程度上依赖于核函数的选择。本文所建立的基于改进的 SVM 决策树分类模型，采用高斯型径向基核函数:，

$$K(x_i,x_j)=\exp\left(-\frac{\|x_i-x_j\|}{2\sigma^2}\right)$$

其中，σ 为核函数的参数。按照以上方法，建立决策树，完成神经元的分类。

4.1.4 问题 1 模型的结论

利用所建立的基于改进的 SVM 决策树分类模型进行实验，可以有效地分析 5 类神经元的几何特征。从所提供的神经元空间几何数据可以看出，不同类别神经元的可分离程度不一致。为构造性能良好的决策树结构，问题 1 模型根据各类神经元的可分离程度，依次将神经元的类别分割出来，使可能出现的错分尽可能地远离树根。通过 51 个样本运行十次十折交叉验证，模型的正确率接近 100%，进一步证明了我们提出的分类模型的有效性。

4.2 问题 2 求解与分析

4.2.1 问题 2 求解

利用所建立的基于改进的 SVM 决策树分类模型，判断附录 B 的 20 个神经元类型。所得结果如下：

表 4-2 附录 B 各样本分类结果（编号根据附录 B 样本顺序编写）

样本编号	样本类型	样本编号	样本类型
B01	锥体神经元	B11	运动神经元
B02	锥体神经元	B12	运动神经元
B03	锥体神经元	B13	感觉神经元
B04	锥体神经元	B14	感觉神经元
B05	普肯野神经元	B15	中间神经元
B06	普肯野神经元	B16	中间神经元
B07	运动神经元	B17	中间神经元
B08	运动神经元	B18	中间神经元
B09	运动神经元	B19	运动神经元
B10	运动神经元	B20	中间神经元

通过表 4-2 可以看出，利用问题 1 模型对附录 B 样本进行分类，所得结果基本符合问题 1 所提供的 5 大类神经元类别。但仍然出现一些可能为错误的分类的结果。比如 B17、B18 和 B19 所得的分类结果是中间神经元，而直接从 B17、B18 和 B19 的几何形态观察，与问题 1 所提供中间神经元的几何特征相差较大，见图 4-3 所示。

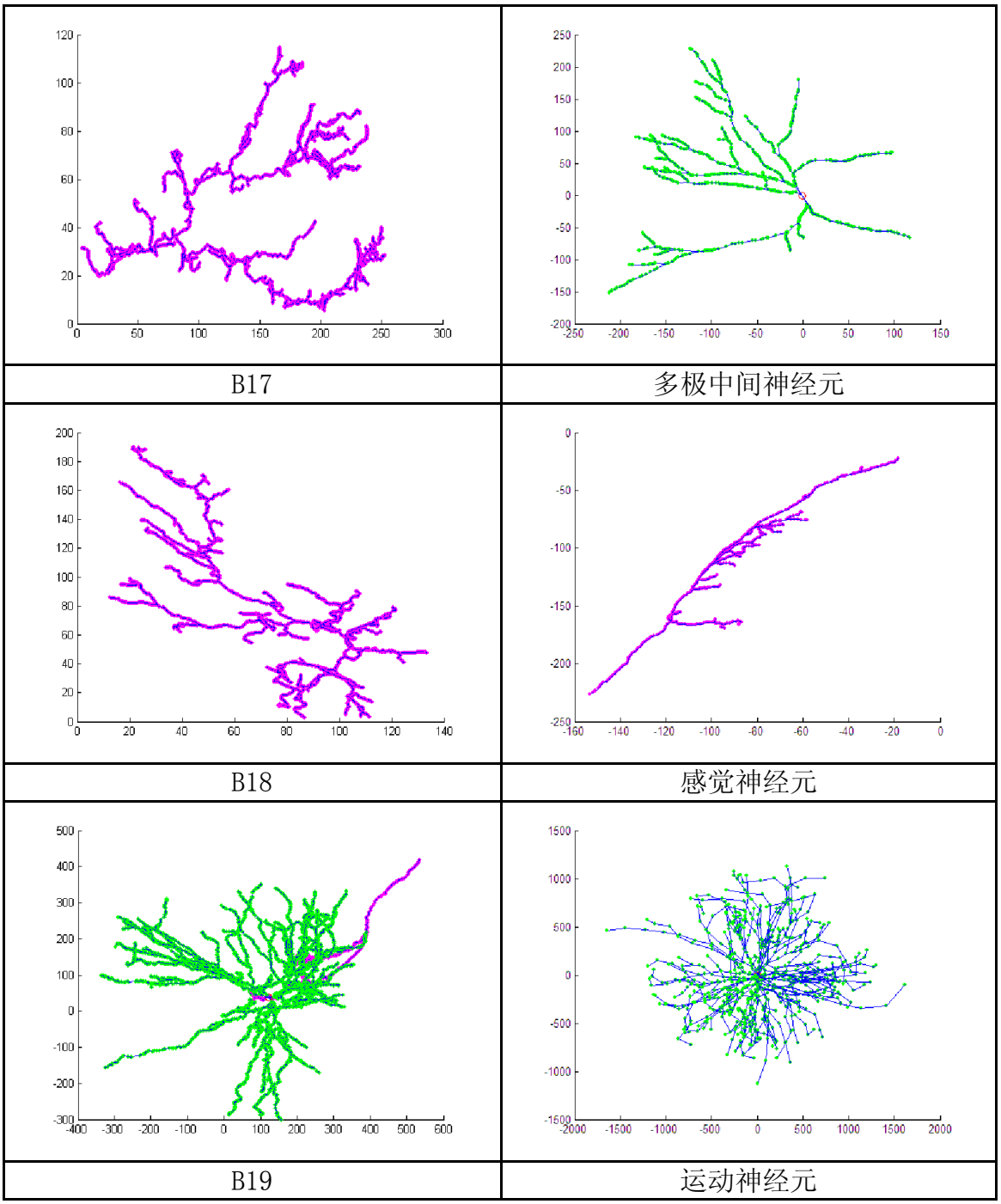

图 4-3 附录 B 中不能准确分类的神经元

从图 4-3 中可以看出，B17 和 B18 神经元具有了多极中间神经元和感觉神经元的特征。利用问题 1 分类模型，得出 B17 和 B18 类别为中间神经元，这与实际结果有一定差别。另外，利用问题 1 分类模型，得出 B19 类别为运动神经元，实

际上 B19 与运动神经元有一定差别。

4.2.2 问题 2 结果分析与总结

通过以上分类结果，可以看出将附录 B 分成 5 类神经元，还不够合理。所分出的结果存在交叉的类别。所以附录 B 分成 5 类神经元是不够的，需要定义新的神经元类别或者引入新的神经元名称。例如 B17 和 B18 实际上是一种 Climbing fiber 类的神经元。产生这种问题，是由于基于改进的 SVM 决策树分类模型，必须在给定类别的情况下，才能分类出结果。所以采用问题 1 分类模型不能用于发现新神经元的类别。

4.3 问题 3 模型与求解

4.3.1 问题 3 分析

由于神经元的形态复杂多样，神经元的识别分类问题目前还没有好的解决方法。如何设计一个好的神经元形态分类方法，将所有神经元按几何特征分类，是本问题所要解决的难点。与问题 1 不同的是，问题 3 没有确定神经元类别的个数。因此，问题 1 的分类模型不能适用于问题 3 中。对于分类的类别个数不确定的分类问题，可以使用聚类算法解决。但是由于现实中的神经元样本量巨大、几何特征维数高，如何设计一个好的聚类算法用于分析神经元的类别是建立模型的关键所在。问题 3 要求所建立的分类模型应该能够为神经元的命名提供建议，这就需要我们所建立的聚类分析模型的聚类结果要具有可读性，方便研究者观察。

4.3.2 建立基于 SOM 聚类分析的分类模型

通过以上对问题 3 的深入分析，本文决定采用 SOM（Self-Organizing Map）聚类分析建立神经元的形态分类模型[4]。SOM 是一种特殊的神经网络算法，采用与神经网络相似的权值调整方法，将高维数据点拓扑保序地映射到二维的网格上，来实现高维数据模式的低维可视化。SOM 在可视化分析领域已经出现许多很有价值的成果[5]。借助其对高维数据的低维组织能力，SOM 在分类、聚类以及预测等数据挖掘领域都有很多成功的应用[6]。

SOM 是一个无监督学习过程，可概括如下：

假设一个 n 维实数向量 $V_i, i=1,2,\cdots,n$ 的数据集，将其映射到一个集合为 $W=\{W_r\}_{r\in A}$ 的原型，A 通常是二维的网格，其原型的个数为 N。则 SOM 最小化能量函数：

$$E=\int P(v)\sum_r \delta_r^{s(v)}\sum_{r'} h_\sigma(r,r')(v-w_{r'})^2\,dv$$

其中，$\delta_r^{s(v)}$ 是 Kronecker 符号，$P(v)$ 是数据的分布。邻域函数为 $h_\sigma(r,r')$：

$$h_\sigma(r,r')=\exp\left(\frac{-\|r-r'\|}{2\sigma^2}\right)$$

用来描述网格节点的学习协作性。原型学习通过求原型W_r变化范围为σ的能量函数E的随机梯度下降ΔW_r求得，

$$\Delta W_r = -\varepsilon \frac{\partial E}{\partial W_r} = \exp\left(\frac{-\|s(v)-r\|}{2\sigma^2}\right)(v - W_r)$$

其中，$s(v) = \arg\min_{r\in A}\left[\sum_{r'} h_\sigma(r,r')(v-W_r)^2\right]$，称为获胜节点（winning node），整个学习过程由映射规则决定[7]。SOM作为一种特殊的神经网络, 仅由输入层和输出层组成。输入层只有一个节点, 对应于输入数据集, 输出层由一系列组织在低维网格(通常是一维或者二维）上的有序节点组成。

利用 SOM 的数据可视化功能和人眼对低维数据模式的快速把握能力，本文决定建立基于 SOM 聚类分析的神经元分类模型。人眼对低维数据模式有着快速的识别能力，但对于高维数据模式的识别却非常困难。SOM 算法具有良好的空间拓扑保序性，在映射过程中能够将高维数据的拓扑分布较好地保留到低维空间中。利用 SOM 的这一特点，就可以再利用人眼的低维识别能力来提高聚类算法的效果。

在建立以上模型的基础上，引入一种聚簇分布特征图，用来在聚类前对神经元几何特征数据进行预处理。从而快速掌握几何特征数据分布的特点，并确定类别数目或者范围。训练好的 SOM 输出层构成一个网格结构，对每个节点按照优先顺序进行编号。假设 SOM 是一个$m \times n$的网格，则可以定义如下的距离矩阵（Distance-Matrix）：

$$D\_mat = \begin{bmatrix} d_{11} \cdots d_{1n} \\ \vdots \quad \vdots \quad \vdots \\ d_{m1} \cdots d_{mn} \end{bmatrix}$$

其中，d_{ij}表示第i节点与第j节点之间的距离。然后，将距离与一个颜色范围进行映射，采用原则为：距离自小到大，对应的颜色自浅至深，由此得到了一个与距离矩阵同等规模的颜色矩阵（Color-Matrix）：

$$C\_mat = \begin{bmatrix} c_{11} \cdots c_{1n} \\ \vdots \quad \vdots \quad \vdots \\ c_{m1} \cdots c_{mn} \end{bmatrix}$$

其中，c_{ij}表示与d_{ij}相对应的颜色值。绘制 SOM 网络，并对每个节点根据颜色矩阵进行着色，至此便得到了聚类分布特征图。图 4-3 是一个典型的聚簇分布特征图。

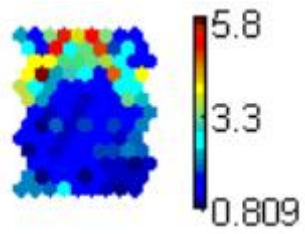

图 4-3 聚簇分布特征图

在图 4-3 中，左边为着色后的 SOM 网络，右边为颜色与距离的对应关系，距离越大，对应的颜色越深。颜色较浅的一组相邻节点构成一个聚簇，颜色较深的节点形成类边界，根据这个原则对图 4-3 进行分析，可以确定类别数目为 3。对于一些复杂的实际数据集，通常不能很明确地确定类别数目，但可以确定可能的类别数，在此基础上进行几次尝试就可以得到最后的聚类结果。

4.3.3 问题 3 模型求解

根据以上所建立基于 SOM 的聚类分析模型，按照以下步骤求解模型：

Step1 从 http://neuromorpho.org/neuroMorpho/byCellType.html 上获取更多神经元训练样本数据集，对数据集进行处理，去除空值并归一化等。

Step2 合理确定 SOM 网络结构大小，使得网络能够清楚地刻画数据集的聚簇分布特征，而且训练时间较短。训练模型并绘制聚簇分布特征图、分析特征图、确定类别数目或范围。

Step3 如果特征图能较清楚地反映出类别数，则根据类别数目定义 SOM 节点数目并进行聚类分析，否则根据可能的类别数设计多个 SOM 网络，分别进行聚类分析，然后用 Silhouette 系数作为准则挑选出合适的聚类结果[7]。以每个样本对应的 BMU 编号作为该样本的类标签。如果数据集维数不高，直接进行 Step5，否则进入下一步。

Step4 根据单独一维属性计算 SOM 的距离矩阵，并绘制聚簇分布特征图。通过特征图来分析该属性的分布，并与总体分布进行比较，挑选与总体特征图具有相似分布的属性作为优秀属性。

Step5 如果数据集维数不高，则直接使用所有属性对聚类结果进行解释；否则，在上一步处理的基础上使用优秀属性对聚类结果进行解释。

根据以上步骤，对数据进行归一化可以处理后，输入网络进行训练。网络邻域函数选用高斯函数，学习率初始值为4.0，并线性减小，采用欧氏距离度量，训练结束条件为连续两次训练过程中权值差的绝对值小于0.03。训练完成后，计算SOM的距离矩阵与颜色矩阵，并绘制聚簇分布特征图如图4-4所示。

观察数据集聚簇分布特征图，选取一系列深色、相邻、连续的节点作为类边界。这些类边界将整个SOM网络大致分为9部分，由此定义聚类类别数为9。

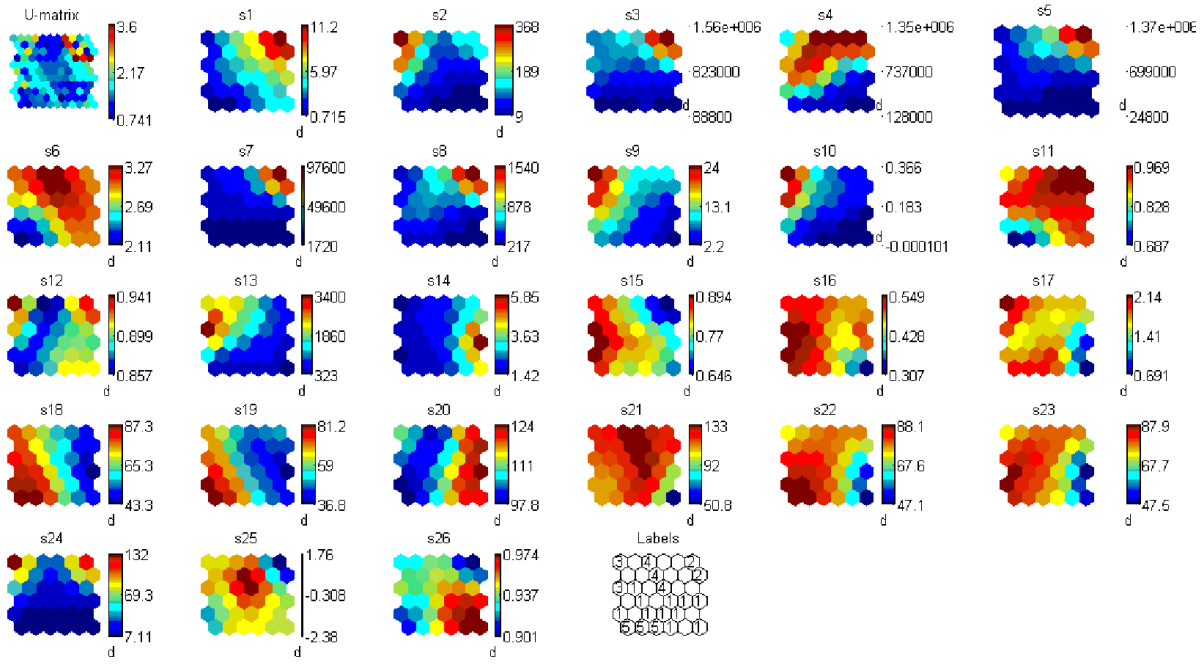

图 4-4 聚簇分布特征图

4.3.4 问题 3 结果分析

从图 4-4 聚簇分布特征图可以看出，特征 s1、s3、 s5 、s7 、s8 、s12、 s25 和 s26 的聚簇分布与总体聚簇分布比较类似，如图 4-5 所示。这几个特征的在聚类过程中做出更多的贡献，将其当作优秀特征。

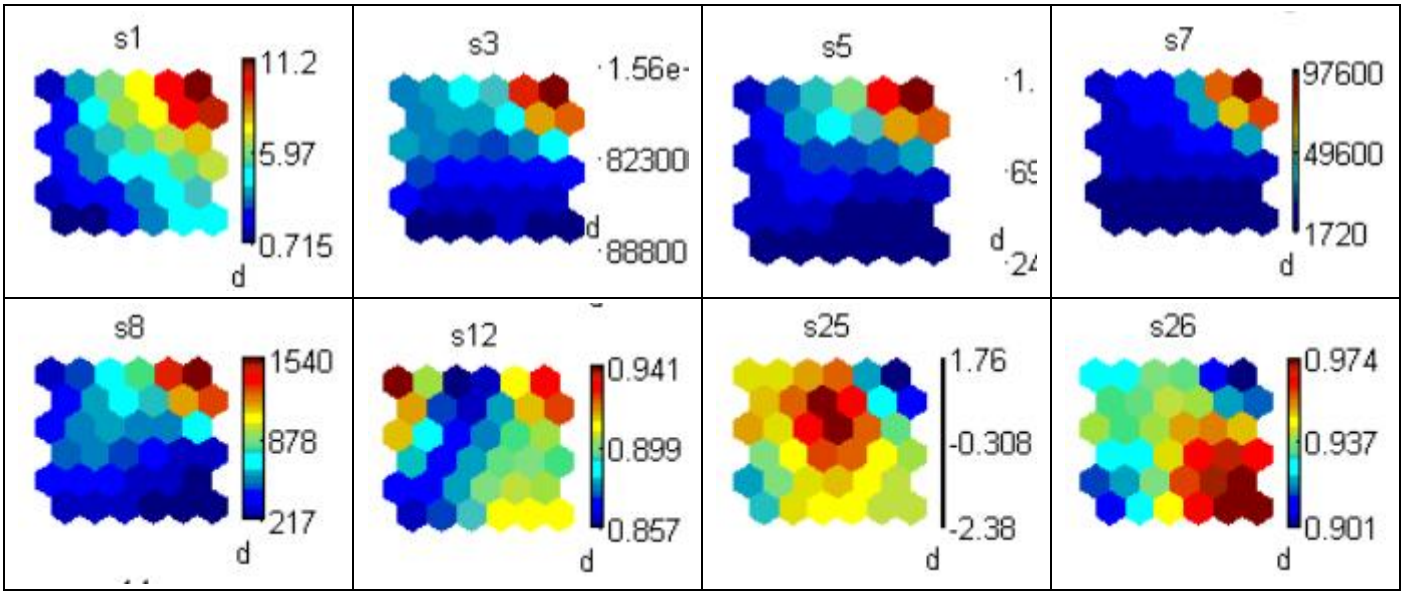

图 4-5 优秀特征的聚簇分布特征图

根据以上确定的优秀特征，我们可以利用这 8 个优秀特征就可以对神经元的进行形态分类。通过实验进行测试，这 8 个优秀特征确实能够很有效地对神经元进行分类。因此，可以利用这 8 个优秀特征对神经元进行初步分类。而且可以按照 8 个优秀特征的数值大小，分析各类别神经元几何形态的特征。这也给生物学家为神经元的命名提供了参考。

以下是几个根据优秀特征提供给生物学家命名神经元的建议：

1、根据干的数目对神经元进行命名，这实际上也是常见的神经元命名方法。如将神经元分为单级、双级和多级神经元。

2、根据神经元的宽度大小，将神经元分为宽大神经元和狭窄神经元。

3、根据神经元的深度大小，将深度小的神经元命名为扁平神经元。如将肯野神经元可以命名为扁平神经元。

4、根据从尖端到胞体的欧式距离总和的大小，将神经元分为长型和短型神经元。

5、根据神经元收缩率的大小，将神经元分为紧缩型和膨松型神经元。

6、当神经元螺旋度的很大时，可以定义神经元为高度螺旋神经元。

7、根据神经元分形维数的大小，将神经元分为高分维和低分维的神经元。

8、这些根据优秀特征的命名方法也可以结合在一起。

4.3.5 问题 3 模型总结

本文建立一个基于 SOM 聚类分析模型，并用于解决神经元的分类问题。与问题 1 模型相比，问题 3 模型可以解决没有确定神经元类别个数的分类问题。而且利用 SOM 聚类分析可以实现高维数据模式的低维可视化，使得聚类结果具有可读性。利用 SOM 聚类分析，还可以分析各个特征对分类估计的贡献程度，提取出优秀特征，给生物学家命名神经元提供参考。

4.4 问题 4 求解与分析

4.4.1 问题 4 求解

附件 A 中的普肯野神经元包含了猪和鼠的神经元。利用问题 3 模型，对附件 A 中的普肯野神经元进行分类，并绘制聚簇分布特征图如图 4-6 所示。

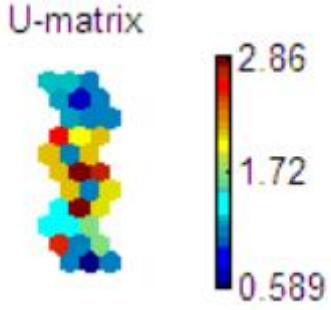

图 4-6 所示 普肯野神经元的聚簇分布特征图

4.4.2 问题 4 结果分析

从以上实验结果，可以很清楚地看出聚簇分布特征图中存在两个孤立的蓝色点。也就是说猪和鼠的普肯野神经元是有差别的，需要进一步分类。

根据 26 个特征聚簇分布特征图（见图 4-7 所示），可以将特征分为可区分类型和不可区分类型。从特征聚簇分布特征图中，选择了部分不可区分类型特征的聚簇图见图 4-8 所示。根据表 4-3 的数据，可以发现所选择的特征对于这两类普

肯野神经元不具有区分能力。再从特征聚簇分布特征图中，选择了部分可区分类型特征的聚簇图见图 4-9 所示。根据表 4-4 的数据，可以发现所选择的特征可以区分这两类普肯野神经元。

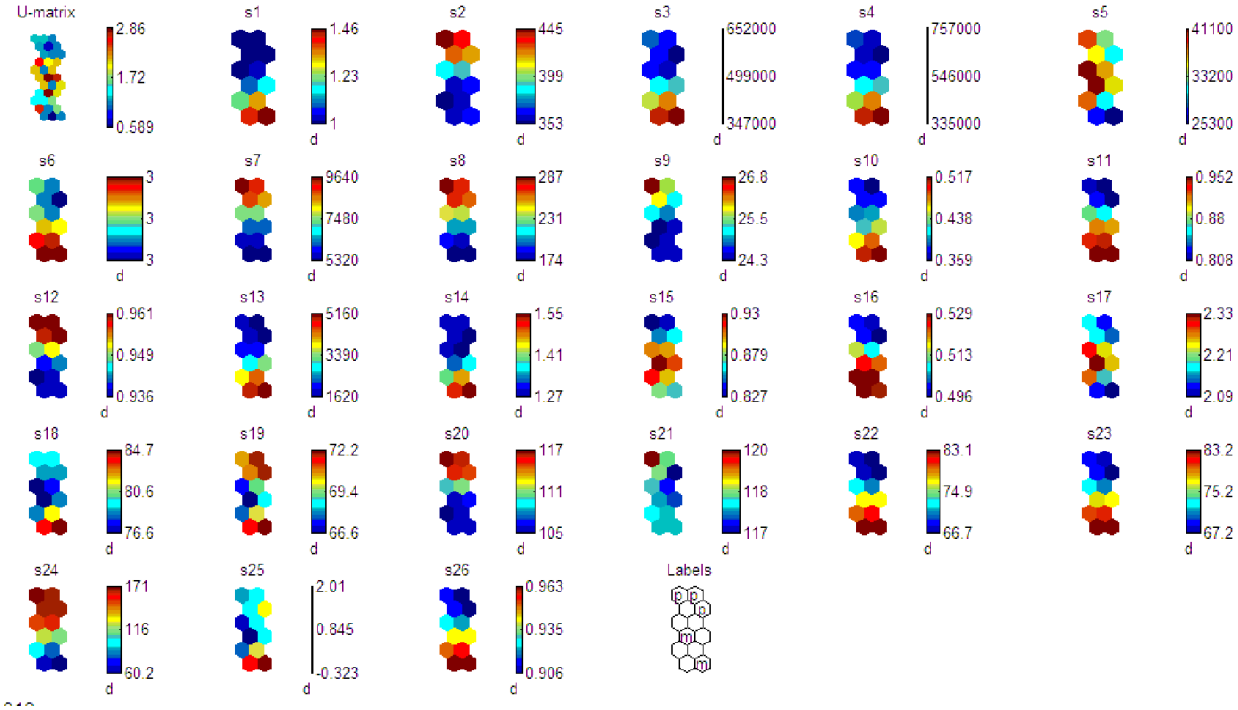

图 4-7 普肯野神经元各个特征聚簇分布特征图

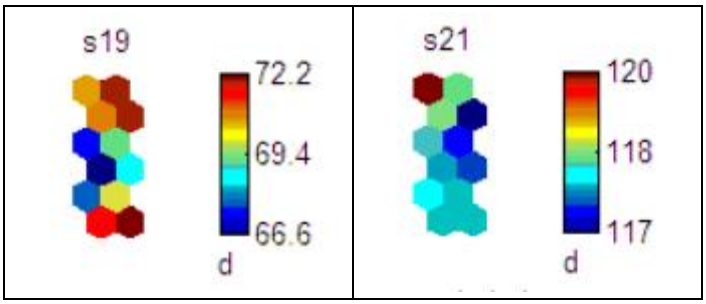

图 4-8 普肯野神经元不可区分特征的聚簇图

表 4-3 猪和鼠普肯野神经元的 s19 和 s21 特征值

神经元	特征 s19	特征 s21
猪的普肯野神经元 1	68.159	123.16
猪的普肯野神经元 2	74.148	117.49
猪的普肯野神经元 3	72.617	114.56
鼠的普肯野神经元 1	77.244	115.03
鼠的普肯野神经元 2	68.474	120.46
鼠的普肯野神经元 3	63.743	117.88

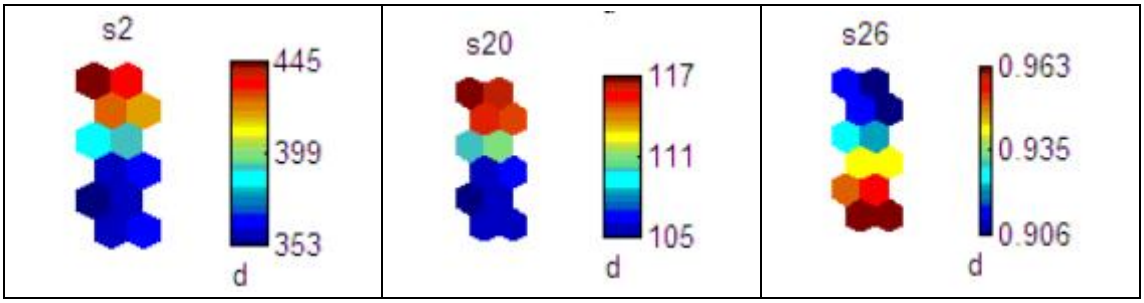

图 4-9 普肯野神经元可区分特征的聚簇图

表 4-4 猪和鼠普肯野神经元的 s2、s20 和 s26 特征值

神经元	特征 s2	特征 s20	特征 s26
猪的普肯野神经元 1	471	117.86	0.92955
猪的普肯野神经元 2	419	116.17	0.89033
猪的普肯野神经元 3	416	113.84	0.90581
鼠的普肯野神经元 1	369	104.97	0.9675
鼠的普肯野神经元 2	357	106.16	0.9625
鼠的普肯野神经元 3	342	103.86	0.94176

因此，按照问题 3 模型所建立的神经元形态分类方法，可以判断不同动物神经系统中同一类神经元形态特征的区别。

4.5 问题 5 模型与求解

4.5.1 问题 5 分析

在大量查阅国外相关文献并加以实验的基础上，我们认为通过较为准确地构造神经元的生长模型，就可以预测神经元形态的生长变化，而且这种伴随这生长变化的形态学特征应在合理的范围内变化，即在我们的分类模型判断结果不变的情况下变化。

神经元的生长模型可以分为随机生长模型[8]和机制生长模型，前者把神经元的生长看作随机运动的结果，而后者则旨在从神经元细胞内在和外在机制的基础上出发描述其生长过程。

我们的模型把神经元的生长看作是一个随机的、非静止的分支和延支过程，树突的生长假设为一系列的分支过程事件。在每次分支事件时，一个新的房室会连接到一个已有的片段上，从而构成一个新的分叉点，随机性被假定存在于两个方面（i）分支片段位置选择（ii）发生分支事件的时间。

4.5.2 问题 5 模型建立

下面给出几个我们在模型中用到的几个定义：

$order(\gamma)$：代表距离胞体的拓扑距离。其实是个整数并且随着每个分叉递增（‘离心顺序 ’），例如胞体相连的基本的片段 $\gamma = 0$。

$Degree(n)$：代表子树末端片断上的分支数。

Asymmetry index(A)：它定义为子树划分的不对称性，也是我们定义的二十五特征之一。

1、模型的拓扑变化

生长树的拓扑伴随着分支事件在其不同房室发生而变化，选择一个房室分支的概率取决于该房室的类型和顺序，一个离心顺序为γ 终点房室分支的概率为：

$$prs = C \cdot 2^{-S\lambda}$$

参数 S 用于调节对于离心顺序的依赖。 是一个正规化常数，一个中间房室分支的概率可以通过终点房室分支概率来计算：

$$prs = prs \cdot Q / (1-Q)$$

其中参数$Q(0 \le Q \le 1)$决定了中间片段的分支概率，$Q = 0$或$Q = 1$分别代表没有分支或专门在中间片段分支。通过设置参数Q和S就可以确定很多类神经元细胞的生长模式

2、模型度数变化

树突的度数取决于生长过程中的分支事件，在*BE*模型中分支事件发生在随机的时间点，整个生长过程T可以分为N个时间间隔。每个时间间隔内终端片段发生分支的概率为：

$p_i = B / Nn_i^E$， n_i代表在第i个时间间隔内终端片段的总数，参数B则用于表示整个过程内分支事件发生的期望值，参数E决定了分支概率取决于实际终端片段数的强度。如果$E = 0$那么分支概率就是一个常数。

3、结合拓扑变化和度数变化的模型

将两种随机性相结合，分支概率表示为：

$$p_i = C \cdot 2^{-S\gamma} B / Nn_i^E$$

其中$C = n_i / \sum_1^{n_i} 2^{-s\gamma_i}$是一个规范化常数，这个常数能确保每个时间间隔内的分支概率与S值保持独立。

4、参数设置

这个模型的参数B,E,S必需优化设置成为能最小化模型生成树和实际的分支模式的差异的参数。一般地说，这个多维优化问题可以用多变量搜索的方法解决。从模型的结构，就可以很便利的决定优化参数的值。参数S可以从拓扑不对称性指数上估计。B,E参数则取决于模型参数的分布。

4.5.3 问题 5 模型求解

对于我们提出神经元的每一类都要设置相应的B,E,S，例如运动神经元(B=3.84,E=0.17,S=0.23), 普肯野神经元(B=4.5,E=1.2S=1.3)等等。就以第四个问题中猪的普肯野神经元和鼠的普肯野神经元为例，假设生长事件发生了十次，即循环十次后，猪的一个普肯野神经元（B=4.4,E=1.3,S=1.4）和鼠的一个普肯野神经元（B=4.2,E=1.1,S=1.2）的主要形态学参数如下表所示。

表 1 猪的普肯野神经元和鼠的普肯野神经生长前后形态学特征变化

	猪的普肯野神经元		鼠的普肯野神经	
属性名称	初始状态	生长后	初始状态	生长后
干的数目	1	1	1	1
宽度	373410	377510	651100	651420
深度	27378	27456	37219.9	39212.1
长度	8413.2	8643.2	5497.9	5667.4
收缩率	0.96305	0.96312	0.94269	0.94326
欧式距离	271.88	282.16	171.11	186.11
螺旋率	0.43434	0.4434	4.237	4.34
分形维数	0.90581	0.90581	0.9625	0.9625

4.5.4 问题 5 模型总结

从上表中可以看出：这些关键形态学特征没有发生太大变化，在我们的分类模型中还是能区分这两种神经元，从而说明我们的生长模型具有一定的合理性，并且也再次验证了我们选出的这八个特征的有效性。

5 模型的评价

5.1 模型的优点

针对神经元的几何形态，本文先建立了基于改进的 SVM 决策树分类模型进行实验，可以有效地分析 5 类神经元的几何特征。该模型考虑了不同类别神经元的可分离程度，构造了性能良好的决策树结构，使可能出现的错分尽可能地远离树根，保证了分类模型的有效性。

由于神经元的几何形态复杂，对此建立了一个基于 SOM 聚类分析模型。与问题 1 模型相比，问题 3 模型可以解决没有确定神经元类别个数的分类问题。而且利用 SOM 聚类分析可以实现高维数据模式的低维可视化。利用 SOM 聚类分析，可以分析各个特征对分类估计的贡献程度，给生物学家命名神经元提供参考。通过随机生长模型来预测神经元生长，可以发现生长过程中关键的几何特征没有发生太大变化。从而说明我们的生长模型具有一定的合理性，并且也再次验证了我们选出的这八个特征的有效性。

5.2 模型的不足

1、在形态学特征定义和选择时，我们选用的是 L-measure 中已有的特征。如何定义神经元的几何特征，从而更好地刻画神经元的形态结构，需要我们进一步深入研究。

2、基于改进的 SVM 决策树分类模型能够很好地解决给定类别个数的神经元分类问题。但当类别个数未知时，该模型并不适用。而基于 SOM 聚类分析模型能解决更加复杂的神经元分类问题。不过当神经元的几何特征维数过高时，其

时间效率也会对聚类效果产生一定的影响。

3、问题 5 所建立的 BES 模型是一种随机生长模型，比较适合做短期预测。

6 参考文献

［1］ Scorcioni, R., Polavaram, S., et al.: L-Measure: a web-accessible tool for the analysis,comparison and search of digital reconstructions of neuronal morphologies. Nat. Protocols[C], 2008, 866–876.

［2］ Schierwagen, A. Neuronal morphology: Shape characteristics and models. Neurophysiology [J].2008,40:366–372

［3］ Bennett K P, Blue J A. A Support Vector Machine App roach to Decision Trees [A]. Proceedings of IJCNN’98 [C], Anchorage Alaska: IEEE Press, 1998.2396 - 2401

［4］ Kohonen T. Self-Organizing Maps. Berlin : Springer Verlag[M], 2001.

［5］ Vesanto J. SOM-Based Data Visualization Methods. Intelligent Data Analysis[J], 1999, 3 (2) :111-126.

［6］ Himberg J. SOM Based Cluster Visualization and It s Application for False Coloring. Proc of the Int'l Joint Conf on Neural Networks[C] .2000, 587-592.

［7］ Kaufman L, Rousseeuw P. Finding Groups in Data: An Introduction to Cluster Analysis [M]. New York, 1990.

［8］ J. Van Pelt and A. Schierwagen.orphological analysis and modeling of neuronal dendrites.th. Biosci[J].2004,188(2):147-155.