

杨树蛋白质二级结构的人工神经网络预测

高光芹, 孟庆玲, 黄家荣*

(河南农业大学, 河南 郑州 450002)

摘要:以 PDB 公用生物信息数据库为基础,用人工神经网络建模技术对杨树蛋白质二级结构预测模型进行了研究,这对深入认识森林动态机理、提高森林资源信息化管理水平、改进森林保护制药设计等具有重要的应用价值,对森林生物学、森林生物信息学的研究具有重要的学术意义。从公用数据库下载杨树蛋白质样本 27 个,提取氨基酸 2 947 个。用长度为 17 的滑动窗口截取蛋白质一级结构的氨基酸序列片段,并用 $[-1,1]$ 编码方式进行编码,以组织输入向量,以片段中心氨基酸对应的蛋白质二级结构(螺旋、折叠、无规则卷曲)为输出向量,构建了结构为 17 : S : 3 的 BP 人工神经网络模型。用训练、测试样本对模型进行训练、检验,得出理想的模型结构为 17 : 9 : 3,其总体拟合准确率为 71%,总体预测准确率为 65%,H 的预测准确度达 81%,比以往同类研究具有较高的预测准确度。

关键词:杨树;蛋白质;二级结构;人工神经网络;预测

中图分类号:S792.11 **文献标志码:**A **文章编号:**1001-7461(2014)05-0059-05

Prediction of Poplar Protein Secondary Structure with Artificial Neural Networks

GAO Guang-qin, MEN Qing-ling, HUANG Jia-rong*

(Henan Agricultural University, Zhengzhou, Henan 450002, China)

Abstract: The paper studied a model for predicting the secondary structure of poplar protein with artificial neural network modeling technology, based on the public biological information database, PDB that has important application value for understanding forest dynamic mechanism, and raising the level of forest resources information management, and improving pharmaceutical design of forest protection. It has important academic significance for studying forest biology and bioinformatics. Twenty seven poplar protein samples were downloaded from public database, 2947 amino acids were extracted. An amino acid sequence fragment of protein primary structure was cut out with a sliding window that length was 17, and was coded by $[-1,1]$ coding scheme. The BP neural network model with the structure, 17 : S : 3, was created, by taking the $[-1,1]$ coding as input variable, and protein secondary structure (spiral, folding, and random curl three state) corresponding amino acids at the center of the fragment as output variable. Through training and optimum seeking, the idea model structure was 17 : 9 : 3, the overall fitting accuracy was 71%, the overall prediction accuracy was 65%, and the prediction accuracy of H was 81%. The results indicated that the proposed model had higher prediction accuracy than the similar studies in the past.

Key words: poplar; protein; secondary structure; artificial neural network; prediction

随着人类基因组计划的顺利实施和各种模式生物全基因组测序工作的不断开展,蛋白质序列与结构测定速度的差距越来越大,使蛋白质结构预测成为后基因组时代迫切急需解决的问题^[1-2]。林木的

收稿日期:2013-12-03 修回日期:2014-01-26

基金项目:河南省教育厅自然科学基金项目(2009B220001);河南省科技攻关项目(0624050007)。

作者简介:高光芹,女,硕士研究生,实验师,研究方向:化学生物信息学。

* 通信作者:黄家荣,男,博士,教授,研究方向:森林资源信息化管理。E-mail:huangjiarong137@163.com

生长周期长,个体庞大,性状遗传背景复杂,分析难度大,随着模式木本植物杨树(*populus*)全基因组测序工作的进展,更需要林木蛋白质结构的预测研究。生物的遗传信息由 DNA 转录到 RNA,再由 RNA 翻译成蛋白质,然后通过蛋白质来实现各种生命活动^[3-4],林木的生命现象(光合作用,细胞延长,木质化作用,抗氧化代谢,抗盐性,死皮,果实品质及结实特性,等)也是如此^[1,5-7]。如果清楚林木蛋白质结构,就可以利用林木蛋白质结构信息,从林木蛋白质序列与结构的关系和蛋白质结构与功能的关系深入认识森林动态机理,为森林资源信息化管理与森林保护药物设计提供科学依据,对提高森林资源信息化管理水平和森林保护的制药水平具有重要的应用价值,对森林生物信息学的研究具有重要的学术意义。

蛋白质结构分为一、二、三、四级结构。蛋白质结构预测,即通过对已知结构的蛋白质进行研究和分析,利用计算机方法、统计学方法^[8]等,找出蛋白质序列和空间结构之间的联系,总结出一定的规律并建立模型,然后由已知的一级序列预测未知的二、三级结构。其中,二级结构预测是三级结构预测的先行步骤,即先从一级序列预测二级结构,再把二级结构折叠成三级结构。由此可见,准确预测蛋白质二级结构是准确预测蛋白质三维空间结构的前提^[9]。蛋白质二级结构预测方法有 Chou-Fasman、GOR、Lim、同源序列分析、人工神经网络等方法^[4],其中人工神经网络方法是用于蛋白质二级结构预测最多且准确度最好的方法,目前预测准确度在 70% 以上^[10-14],被认为是前景最乐观的蛋白质结构预测方法之一。在木本植物中,杨树由于基因组相对较小(450~550 Mbp)、遗传转化比较容易等特点已被广泛接受作为林木研究的模式树种^[1]。以杨树蛋白质为研究对象,以国际公用的蛋白质数据库为基础,用人工神经网络对杨树蛋白质二级结构进行预测研究。

1 材料与方法

1.1 数据来源

统计数据来源于公用生物信息数据库 PDB (Protein Data Bank)的下载。PDB 是国际上唯一的生物大分子结构数据档案库,由美国 Brookhaven 国家实验室建立^[15]。PDB 收集的数据来源于 X 光晶体衍射和核磁共振(NMR)的数据,经过整理和确认后存档而成。如表 1 所列,从 PDB 共下载杨树蛋白质数据 27 个。

用以上数据的 3/4 即 20 个作训练集,1/4 即 7 个作测试集。按 17~116 个氨基酸的取样长度,在 20 个蛋白质的训练集中,取氨基酸残基 2 185 个,H(螺旋)的比例为 49%,E(折叠)为 48%,C(无规则卷曲)为 3%;在 7 个蛋白质的测试集中,取氨基酸残基 762 个,H(螺旋)的比例是 45%,E(折叠)是 45%,C(无规则卷曲)是 10%。

表 1 蛋白质结构数据来源及个数
Table 1 Protein structure data's sources and number

蛋白质来源杨树拉丁名	中文名	蛋白质结构/个
POPULUS NIGRA	黑杨	12
POPULUS TREMULA X POPULUS TREMULOIDES	杂交杨	4
POPULUS TRICHOCARPA X POPULUS	杨树	4
POPULUS TREMULA	山杨	4
POPULUS TREMULOIDES	杨树	2
POPULUS TRICHOCARPA	毛果杨	1

1.2 模型构建

用一个 3 层结构的 BP 神经网络对杨树蛋白质二级结构进行预测。以蛋白质一级结构的数学编码作为输入向量,在输入层引入广为适用的“局部滑动窗口”。有统计表明,蛋白质二级结构的形成,受到前后 8 个氨基酸的影响^[16],故滑动窗口的长度取 17。以二级结构作为输出向量,输出层由 3 个神经元组成,以对应 3 种蛋白质的二级结构状态 H、E、C。隐层节点数 s 个。建立杨树蛋白质二级结构神经网络预测模型如图 1 所示,图中:符号●、→、①、□、 f 依次表示输入层节点、信息流、输入值为 1 的节点、神经元、正切 S 型作用函数,相应的 MATLAB 预测函数为:

$$\left. \begin{aligned} Y_k &= \text{tansig}(\sum w_{j,k}^2 H_j + b_k^2) \\ H_j &= \text{tansig}(\sum w_{i,j}^1 X_i + b_j^1), i=1,2,\dots,17 \\ j &= 1,2,\dots,s; k=1,2,3 \end{aligned} \right\} \quad (1)$$

式中, Y_k 为输出层第 k 神经元的输出变量; H_j 为隐层第 j 神经元的输出变量; X_i 为输入层第 i 节点的输入变量; $w_{i,j}^1$ 表示输入层第 i 节点到隐层第 j 神经元的连接权; $w_{j,k}^2$ 表示隐层第 j 神经元到输出层第 k 神经元的连接权; b_j^1 、 b_k^2 分别为隐层第 j 神经元、输出层第 k 神经元的阈值;tansig()为 MATLAB 的正切 S 形函数。

1.3 输入输出编码与模型训练

1.3.1 输入编码 蛋白质分子由 20 种氨基酸构成。标准氨基酸(Standard amino acids)或称蛋白氨基酸(proteinogenic amino acids),是生物细胞中用来合成蛋白质的。氨基酸一般是用 3 个英文缩写字

母来代替。在生物信息常用数据库中也是用 3 个缩写英文表示或是 1 个英文缩写字母表示。为了使以后数学编码更简单,20 种标准氨基酸采用 1 个拉丁字母缩写来表示:1)丙氨酸 A;2)精氨酸 R;……;(20)缬氨酸 V。用 4 种常用的数学编码方式^[17]

($[-1,1]$ 编码、五位编码、正交编码、二十一位编码)对杨树蛋白质的氨基酸残基进行编码,然后利用 BP 神经网络比较 4 种编码方式预测准确度的高低。结果表明, $[-1,1]$ 编码简单、易懂且较其他 3 种编码方式得到的预测准确度要高。

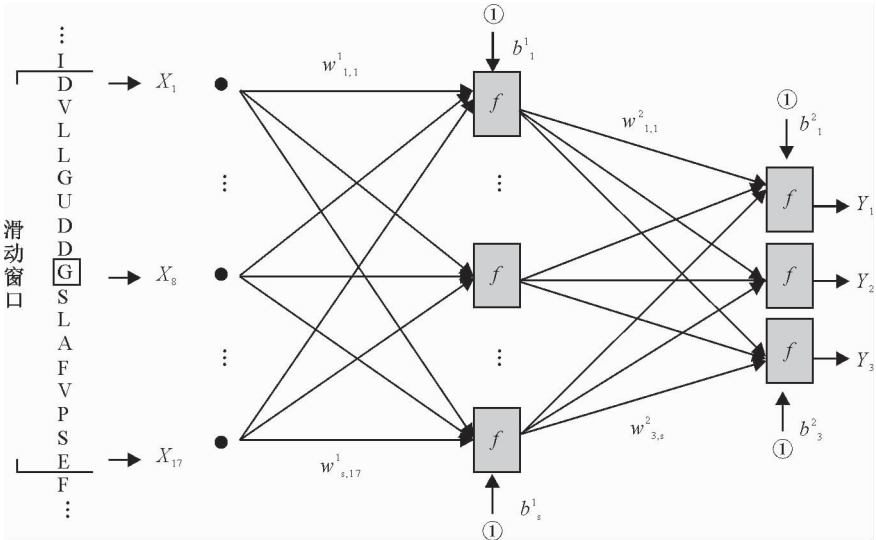


图 1 杨树蛋白质二级结构神经网络预测模型(17 : S : 3)

Fig. 1 Neural network model forecasting secondary structure of poplar protein(17 : S : 3)

1.3.2 输出编码 输出为蛋白质二级结构的编码,分别编码成 3 个二进制码,即 H(螺旋):100;E(折叠)010;C(无规则卷曲)001。

1.3.3 模型训练 直接在 MATLAB 系统神经网络工具箱的图形用户界面 GUI 中,按训练样本定义、训练样本导入 GUI、神经网络设置与初始化、神经网络训练、训练结果导出 GUI 等步骤进行。

1.4 模型检验

本研究中有 17 个输入,3 个输出,隐含层神经元范围在 3~35 个。在对各个编码方法预测评价时,选用最简单并且被广泛使用的评估公式来表示预测的准确度,但由于训练测试集中 C 所占比例较小的问题,我们在对整体二级结构准确率衡量上加入了权重这个因素。总的拟合或预测准确度公式为:

Q3=H×Qh+E×Qe+C×Qc (2)

式中,Q3 表示总的预测准确度;Qh、Qe 和 Qc 分别代表 3 个蛋白质二级结构状态的拟合准确度或预测准确度;H、E 和 C 分别表示训练或测试数据中二级结构状态下各占的百分数。

每种蛋白质二级结构状态拟合与预测的准确度公式为:

Qh=100Ch/Nh
Qe=100Ce/Ne
Qc=100Cc/Nc (3)

式中,Ch,Ce,Cc 表示每 1 种二级结构状态拟合或预测正确的氨基酸总数;Nh、Ne 和 Nc 分别是 3 种二级结构状态的氨基酸总个数。

2 结果与分析

2.1 模型训练结果

用经 $[-1,1]$ 编码的氨基酸残基作为输入变量,二级结构作为输出变量。按网络模型的要求对所建神经网络进行训练,以确定权值矩阵和阈值矩阵。为确定适宜的模型结构,分别对隐层神经元从少到多取不同值进行训练。结果显示,随神经元数的增加,均方差性能函数值呈现递减、拟合准确度呈现增加的趋势,预测准确度呈现抛物线趋势。如表 2 所示,当隐层神经元增加到 9 时,拟合准确度较好,预测准确度最高。所以,理想模型结构为 17 : 9 : 3,当隐层神经元数为 9,训练的误差指标(均方差性能函数 mse)为 0.120 225,总体拟合准确度为 0.713 35,总体预测准确度为 0.651 74。

当隐层神经元为 9 时,杨树蛋白质的预测模型,不论是模型训练结果图、拟合准确度、预测准确度、均方差性能都达到最佳。所以,理想模型是 Network9,将其重命名为 NNFPP (Neural Network Forecasting Poplar Protein),作为最终训练结果,其权值、阈值见表 3、表 4。将表 3、表 4 的权值和阈值代入(1)式,得杨树蛋白质二级结构的预测模型。如

Y_1 与 H_1 的作用函数表达式为:

$$\left. \begin{aligned} Y_1 &= \text{tansig}(-0.505\ 31H_1 + 0.179\ 25H_2 + 0.135\ 35H_3 + 0.130\ 16H_4 + 0.112\ 2H_5 + 0.146\ 45H_6 + \\ &\quad 0.131\ 91H_7 - 0.195\ 63H_8 + 0.237\ 11H_9 + 0.658\ 61) \\ H_1 &= \text{tansig}(0.115\ 34X_1 + 1.758\ 3X_2 - 0.043\ 189X_3 + 0.233\ 35X_4 - 0.197\ 63X_5 + 0.112\ 51X_6 + \\ &\quad 0.411\ 27X_7 - 0.598\ 05X_8 + 0.108\ 29X_9 - 0.284\ 09X_{10} + 0.080\ 72X_{11} - 0.418\ 2X_{12} + \\ &\quad 0.363\ 64X_{13} - 0.956\ 9X_{14} + 1.281\ 2X_{15} - 0.112\ 43X_{16} + 0.312\ 92X_{17} + 3.545\ 5) \end{aligned} \right\} \quad (4)$$

表 2 模型训练结果
Table 2 Trained results of the models

网络对象	网络结构	权值个数	阈值个数	均方差 性能函数	训练次数	拟合 准确度	预测 准确度
Network6	17 : 6 : 3	120	9	0.130 173	200	0.633 47	0.585 78
Network7	17 : 7 : 3	140	10	0.128 279	200	0.657 52	0.622 33
Network8	17 : 8 : 3	160	11	0.123 938	200	0.688 58	0.608 54
Network9	17 : 9 : 3	180	12	0.120 225	200	0.713 35	0.651 74
Network10	17 : 10 : 3	200	13	0.119 923	150	0.720 50	0.616 96
Network11	17 : 11 : 3	220	14	0.116 370	150	0.729 80	0.605 67
Network12	17 : 12 : 3	240	15	0.116 589	150	0.725 44	0.597 31

表 3 隐层神经元与输入层节点间的权值和隐层神经元阈值
Table 3 Weights between input layer nodes and hidden layer neurons and latter thresholds

陷层	输入层/权值								
	1	2	3	4	5	6	7	8	9
1	0.115 34	1.758 3	-0.043 189	0.233 35	-0.197 63	0.112 51	0.411 27	-0.598 05	0.108 29
2	-8.270 7	-6.772 4	22.009 9	-17.303 3	24.494 6	5.890 2	-11.629 8	3.451 9	10.764 4
3	37.863 2	16.642 6	-31.834 5	1.784 8	-7.690 7	36.973 9	-4.538	-28.507 5	-5.846 3
4	-37.338 1	19.977 2	-9.853 2	39.712	15.251 4	-38.793 4	2.418 4	7.098	-9.187 1
5	-16.562 7	-14.597 2	-5.854 9	8.225 4	15.807	8.06	-14.624 1	5.498 6	5.751 2
6	9.464 9	-13.553 3	-2.48	-21.540 1	-3.858 1	-3.121 6	6.779 7	5.806 2	-1.293 8
7	16.732 3	30.373 5	20.879 8	9.405 4	-18.591 6	-4.024 8	-4.210 5	-2.031 3	-2.780 1
8	-15.813 2	-16.924 7	19.252 8	9.190 3	-18.849 3	-6.942 4	-22.713 3	59.633 3	4.985 0
9	-0.522 64	0.163 98	0.371 11	0.237 29	-1.109 7	-0.323 09	-0.215 22	1.836 9	-0.260 38

表 3 (续)输入层节点与隐层神经元间的权值和隐层神经元的阈值
Table 3 (C) Weights between input layer nodes and hidden layer neurons and latter thresholds

隐层	输入层/权值								閾值
	10	11	12	13	14	15	16	17	
1	-0.284 09	0.080 72	-0.418 2	0.363 64	-0.956 9	1.281 2	-0.112 43	0.312 92	3.545 5
2	27.602 2	5.54	17.317	12.640 2	24.031 2	-34.972 2	0.503 13	-13.572	64.284 2
3	-38.736 4	-45.605 6	-23.281 6	22.012	-32.812 9	-13.734 4	13.814 3	-21.690 7	-48.425 9
4	-3.331 5	26.239	34.597 9	-32.669 4	24.927	42.303 4	-23.515 8	54.742 5	16.837
5	11.014 7	-17.741 3	15.400 5	-11.206 5	-12.874 1	-7.322 3	1.997 7	8.648	12.087 3
6	9.683 7	-0.935 2	-8.269 9	-5.429 3	8.658 8	-4.633 5	3.965 8	1.836 2	9.629 9
7	-21.345 1	1.081 4	7.518 7	-0.545 37	-9.289 2	5.789 2	2.062 3	0.340 62	-14.646 3
8	13.942 2	6.137 1	3.233 9	-17.319 7	31.628	-17.940 7	-36.058 8	0.196 53	-49.309 5
9	0.208 49	1.054 6	-1.534 6	0.529 12	0.301 37	0.769 5	-2.125 7	-0.449	-1.954 6

表 4 隐层神经元与输出层神经元间的权值和输出层神经元的阈值
Table 4 Weights between hidden layer neurons and export layer neuron and latter thresholds

输出层	隐层/权值									阈值
	1	2	3	4	5	6	7	8	9	
1	-0.505 31	0.179 25	0.135 35	0.130 16	0.112 2	0.146 45	0.131 91	-0.195 63	0.237 11	0.658 61
2	0.498 87	-0.251 8	-0.134 1	-0.144 2	-0.090 7	-0.145 2	-0.115 9	0.160 61	-0.236 4	-0.003 88
3	0.013 036	-0.002 1	-0.004 9	0.000 57	-0.005 0	-0.004 1	-0.007 9	0.004 988	-0.008 5	0.002 609

整个预测模型的仿真函数表达式为:

$$Y_k=sim(NNFPP,[X_1;X_2;\cdots;X_{17}]),k=1,2,3$$

(5)

式中:sim 为 MATLAB 的仿真函数;NNFPP 为训练好的网络对象,它定义和存储了包括权值和阈值在内的全部网络属性参数。在 MATLAB 环境下,将 NNFPP 调入工作空间,应用式(5),代入已知的 $[X_1;X_2;\cdots;X_{17}]$,就可预测相应的 Y_k 。

2.2 拟合准确度和预测准确度分析

拟合或预测准确度公式为:

$$Q3=H\times Qh+E\times Qe+C\times Qc$$

(6)

在本试验中具体的拟合准确度和预测准确度公式表示为:

$$\left. \begin{aligned} Q&=0.494\ 8\times Qh+0.475\ 0\times Qe+0.030\ 1\times Qc \\ Q&=0.451\ 3\times Qh+0.448\ 0\times Qe+0.100\ 6\times Qc \end{aligned} \right\}$$

(7)

每种蛋白质二级结构状态拟合或预测的准确度公式为:

$$\left. \begin{aligned} Qh&=100Ch/Nh \\ Qe&=100Ce/Ne \\ Qc&=100Cc/Nc \end{aligned} \right\}$$

(8)

通过式(7)和式(8)计算,总体拟合准确度为 71%,预测准确度为 65%, H 的最高预测准确度达 81%。

3 结论与讨论

从 PDB 下载杨树蛋白质结构数据,用长度为 17 个氨基酸的滑动窗口从蛋白质单链序列截取氨基酸序列片段,以片段的 $[-1,1]$ 数学编码($X_1,X_2,\cdots X_{17}$)作为输入向量,以片段中心氨基酸对应的蛋白质二级结构(螺旋、折叠、无规则卷曲)编码(Y_1,Y_2,Y_3)为输出向量,构建了结构为 17 : S : 3 的杨树蛋白质二级结构神经网络预测模型。

对 4 种常用数学编码方式的比较研究结果表明, $[-1,1]$ 编码方式较其它编码方式得到的整体预测准确度高。

以 20 和 7 个杨树蛋白质的 1 865 和 650 段氨基酸序列片段的 $[-1,1]$ 编码数据为训练和检验样本,对所建模型进行训练和检验,得适宜的模型结构为 17 : 9 : 3,模型的整体预测准确度为 65%,对于 H 的预测准确度可高达 81%,比以往相同数学编码方式的蛋白质二级结构预测准确度高。但 C 的预测准确度很低,不到 10%。

与以往同类研究相比,本研究特色为:1)针对木本植物,将林木生物信息学的研究纳入森林资源信息化管理的研究,这对林业各学科的发展将具有较大的

促进作用;2)将滑动窗口、氨基酸序列片段与 BP 人工神经网络整合在一起的杨树蛋白质二级结构神经网络预测模型(图 1),简单直观,通俗易懂,数形统一;3)用 MATLAB 的仿真函数表达的预测函数式(5),是一种超常规数学表达,形式简捷,应用方便。

C 的预测准确度很低的原因是训练集中 C 占的比例很小。在本文的数据集中, H 、 E 、 C 占的比例分别是 49%、48%、3%。目前,公用生物信息数据库所公布的林木蛋白质结构数据中,有 C 结构的数据占总蛋白质结构数据的比例为 6%。这是否为林木生物信息的特性之一,有待进一步验证。

或许按功能分类后的分类建模预测能提高预测精度,但因本文用的数据量偏少,未能分类建模预测,这也有待进一步研究。

参考文献:

[1] 袁坤,王明麻,黄敏仁. 林木蛋白质组学研究进展[J]. 中国生物工程杂志, 2006,26(6):88-92.
YUAN K, WANG M X, HUANG M R. Progress in forest tree proteomics research[J]. China Biotechnology, 2006, 26 (6):88-92. (in Chinese)

[2] 孙海军. 基于神经网络的蛋白质二级结构预测问题的研究[D]. 北京:北京工业大学,2004.

[3] ROBLES V, LARRANAGA P, PENA JM, *et al.* Bayesian network multi-classifiers for protein secondary structure prediction[J]. Artif Intell Med, 2004, 31(2):117-136.

[4] 刘桂霞,周春光,周文刚. 基于 ANN 蛋白质结构预测方法研究[J]. 电脑开发与应用, 2004, 17(12): 2-4.

[5] 施士争,隋德宗,王红玲,等. 灌木柳速生无性系的耐盐性选择研究[J]. 西北林学院学报, 2010, 25(4):72-77.
SHI S Z, SUI D Z, WANG H L, *et al.* Selection for fine shrub willow clones with high biomass and salt-tolerance[J]. Journal of Northwest Forestry University, 2010, 25 (4):72-77. (in Chinese)

[6] 袁坤,徐智娟,王真辉,等. 橡胶树胶乳死皮相关蛋白的鉴定及分析[J]. 西北林学院学报, 2012, 27(6):105-109.
YUAN K, XU Z J, WANG Z H, *et al.* Identification and analysis of latex proteins related to tapping panel dryness in *Hevea brasiliensis* [J]. Journal of Northwest Forestry University, 2012, 27(6):105-109. (in Chinese)

[7] 齐国辉,李保国,郭素萍,等. 黄连木优株果实品质及结实特性研究[J]. 西北林学院学报, 2012, 27(2):70-74.
QI G H, LI B G, GUO S P, *et al.* Fruit quality and fruiting characteristics of superior strains of *Pistacia chinensis* [J]. Journal of Northwest Forestry University, 2012, 27 (2):70-74. (in Chinese)

[8] 马栋萍,阮晓钢. 基于改进 BP 神经网络预测蛋白质二级结构[J]. 北京联合大学学报:自然科学版, 2005, 19(2):70-73.

[9] 张漫,常延琦. 蛋白质三级结构预测方法简述[J]. 中国动物检疫, 2005, 22(5):36-37.

eling leaf shapes using L-systems and genetic algorithms[C]// International Conference Nicograph, 2002: 73-78.

[2] RUNIONS A, FUHRER M, LANE B, *et al.* Modeling and visualization of leaf venation patterns[J]. *ACM Transactions on Graphics*, 2005, 24(3): 702-711.

[3] 刘千里, 李春友, 柳瑞武, 等. 多基线数字近景摄影测量系统在古典园林建筑物三维重建中的应用[J]. *西北林学院学报*, 2009, 24(4): 224-227.

LIU Q L, LI C Y, LIU R WU, *et al.* Application of multi-baseline digital close-range photogrammetry system in 3D modeling of classical garden buildings[J]. *Journal of Northwest Forestry University*, 2009, 24(4): 224-227. (in Chinese)

[4] 柳瑞武, 李春友, 孟平, 等. 树木形态结构快速精确测定方法初探[J]. *西北林学院学报*, 2011, 26(1): 176-180.

LIU R W, LI C Y, MENG P, *et al.* Preliminary study on fast and accurate determination method of tree morphostructure[J]. *Journal of Northwest Forestry University*, 2011, 26(1): 176-180. (in Chinese)

[5] FLEISHMAN S, DRORI I, COHEN-OR D. Bilateral mesh denoising[J]. *ACM Transactions on Graphics*, 2003, 22(3): 950-953.

[6] JONES T R, DURAND F, DESBRUN M. Non-iterative, feature-preserving mesh smoothing [J]. *ACM Transactions on Graphics*, 2003, 22(3): 943-949.

[7] 岳杰, 陆声链, 孙智慧, 等. 三维点云孔洞修补算法及在植物形态重建中的应用[J]. *农机化研究*, 2013, 35(5): 190-195.

[8] BISCHOFF S, PAVI D, KOBELT L. Automatic restoration of polygon models [J]. *ACM Transactions on Graphics*, 2005, 24(4): 1332-1352.

[9] TAO J. Robust repair of polygonal models [J]. *ACM Transactions on Graphics (SIGGRAPH'04)*, 2004(4): 888-895.

[10] JOSHUA P, SZYMON R. Atomic volumes for mesh completion[C]// Eurographics Symposium on Geometry Processing, 2005.

[11] 张洁, 岳玮宁, 王楠, 等. 三角网格模型的各向异性孔洞修补算法[J]. *计算机辅助设计与图形学学报*, 2007, 19(7): 892-897.

ZHANG J, YUAN W N, WANG N, *et al.* Anisotropic hole filling algorithm for triangle mesh models [J]. *Journal of Image and Graphics*, 2007, 19(7): 892-897. (in Chinese)

[12] ZHAO W, GAO S M, LIN H W. A robust hole-filling algorithm for triangular mesh [J]. *The Visual Computer*, 2007, 23(12): 987-997.

[13] 杜佶, 张丽艳, 王宏涛, 等. 基于径向基函数的三角网格曲面孔洞修补算法[J]. *计算机辅助设计与图形学学报*, 2005, 17(9): 1976-1982.

DU J, ZHANG L Y, WANG H T, *et al.* Hole repairing in triangular meshes based on radial basis function [J]. *Journal of Image and Graphics*, 2005, 17(9): 1976-1982. (in Chinese)

[14] 谢倩茹, 耿国华. 三维模型孔洞修补方法的研究[J]. *计算机应用研究*, 2013, 30(10): 3175-3177.

XIE Q R, GENG G H. Research on hole-filling algorithm for 3D models [J]. *Application Research of Computers*, 2013, 30(10): 3175-3177. (in Chinese)

[15] 肖双九. 反求工程中三角网格划分及其应用的关键算法研究[D]. 西安: 西北工业大学, 2002.

[16] 孙智慧, 陆声链, 郭新宇, 等. 基于点云数据的植物叶片曲面重构方法[J]. *农业工程学报*, 2012, 28(3): 184-190.

SUN Z H, LU S L, GUO X Y, *et al.* Surfaces reconstruction of plant leaves based on point cloud data [J]. *Transactions of the CSAE*, 2012, 28(3): 184-190. (in Chinese)

[17] LO S H. Volume discretization into tetrahedra-I. verification and orientation of boundary surfaces [J]. *Computers & Structures*, 1991, 39(5): 493-500.

tion using multi-modal neural network[J]. *Journal of Kunming University of Science and Technology: Nat. Sci. Edi.*, 2004, 29(5): 64-70. (in Chinese)

[15] BERMEN H M, WESTBROOK J, FENG Z. The protein data bank[J]. *Nucleic Acids Research*, 2000, 28(1): 235-242.

[16] 马栋萍, 阮晓钢. 基于改进 BP 神经网络预测蛋白质二级结构[J]. *北京联合大学学报: 自然科学版*, 2005, 19(2): 70-73.

MA D P, RUAN X G. Prediction protein secondary structure by improved BP neural network[J]. *Journal of Beijing Union University: Nat. Sci. Edi.*, 2005, 19(2): 70-73. (in Chinese)

[17] 阮晓钢, 孙海军. 编码方式对蛋白质二级结构预测准确度的影响[J]. *北京工业大学学报*, 2005, 31(3): 229-235.

RUAN X G, SUN H J. Research on encode influencing protein secondary structure prediction[J]. *Journal of Beijing University of Technology*, 2005, 31(3): 229-235. (in Chinese)

(上接第 63 页)

[10] QIAN N, SEJNOWSKI T J. Predicting the secondary structure of globular protein using neural network models[J]. *Mol. Biol.*, 1988, 202: 865-884.

[11] 须文波, 陆克中. 神经网络在蛋白质二级结构预测中的应用[J]. *生物信息学*, 2006, 3(4): 26-29.

[12] 何琴, 高建华, 刘伟. 蛋白质二级结构预测的人工神经网络方法研究[J]. *分析科学学报*, 2006, 22(4): 438-440.

HE Q, GAO J H, LIU W. Study of neural network methods in predicting protein secondary structures[J]. *Journal of Analytical Science*, 2006, 22(4): 438-440. (in Chinese)

[13] 王化坤, 钱伟懿. 神经网络方法在蛋白质结构预测中的研究与应用[J]. *黑龙江科技信息*, 2008(4): 23-23.

[14] 孙海军, 阮晓钢. 用多模神经网络预测蛋白质二级结构[J]. *昆明理工大学学报: 理工版*, 2004, 29(5): 64-70.

SUN H J, RUAN X G. Protein secondary structure predic-