

用多模神经网络预测蛋白质二级结构

孙海军, 阮晓钢

(北京工业大学 电子信息和控制工程学院, 北京 100022)

摘要: 提出了一个由7个BP神经网络组合成的多模神经网络的预测模型, 同时给多模神经网络引进了较多的生物进化信息(Evolutionary information), 即一方面引入了“profile”编码, 这种编码被认为携带了较多的生物信息; 另一方面引入了氨基酸之间的“距离”概念. 它体现了输入层临近氨基酸的相互联系和影响. 对从36个蛋白质提取的4000个氨基酸的进行了预测研究. 结果表明, 与文献[1]的预测结果相比, 本文的多模神经网络把蛋白质二级结构预测的平均精度从66.1502%提高到68.8903%.

关键词: 多模神经网络; 蛋白质二级结构预测; 神经网络

中图分类号: TP311 **文献标识码:** A **文章编号:** 1007-855X(2004)05-0064-07

Protein Secondary Structure Prediction Using Multi-Modal Neural Network

SUN Hai-jun, RUAN Xiao-gang

(Electronic Information and Control Engineering School, Beijing University of Technology, Beijing 100022, China)

Abstract: A multi-modal neural network that was made of seven feed-forward BP neural networks to predict the secondary structure of proteins is developed. And the more biological evolution information into this multi-modal neural network, that is to say “profile” code is introduced, which is thought to carry more evolution information, and on the other hand, the “distance” concept between the amino acid is introduced. It has embodied the connection and influence of importing layer close to amino acid. A prediction is made on the protein secondary structure by using 4000 amino acids from 36 proteins. Results indicate that as compared with [1] whose result is 66.1502%, our Multi-modal Neural Network may increase the average accuracy to 68.8903%.

Key words: Multi-modal Neural Network; protein secondary structure prediction; neural network

0 引言

在蛋白质结构预测的研究中, 通常的依据是蛋白质的三级结构由它的一级结构(primary structure)决定. 蛋白质的一级结构其实就是一个完整的一维氨基酸序列, 用氨基酸序列进行蛋白质的结构预测是一项很富有挑战性的研究工作, 就现在的研究情况表明, 蛋白质的三级结构预测一项是非常困难的耗时的研究工作. 于是人们把目光从三级结构预测转向了多肽链的局部信息预测, 即蛋白质二级结构的预测上来. 蛋白质的二级结构是一级结构和三级结构的中间过渡环节, 它是指蛋白质多肽链中有规则重复的区域. 事实上, 蛋白质三级结构是通过把二级结构“捆绑”在一个或几个区域所形成的结构. 目前, 有两种形式的蛋白质二级结构: 即8状态的和3状态的. 而在蛋白质二级结构预测的研究中, 用3状态的二级结构预测最为常见. 这三种状态是 α 螺旋、 β 折叠和L卷曲, 而8状态的其实就是把3状态的细分了一下而已. 由于蛋白质的功能是由它的三级结构决定的, 因此, 预测蛋白质二级结构是非常有意义的, 它已经成为解决由蛋白质的一级结构(序列)预测其空间结构这一问题的最关键的步骤^[1]; 同时, 对理解二级结构在细胞中的作用也是非常有帮助的. 到目前为止, 人类已经用了很多方法来探索这个问题, 有的利用氨基酸的物理化学性质; 有的是利用序列的同源和模式匹配; 还有的对已知的蛋白质结构进行统计分析等^[2,6]. 自从1988年

收稿日期: 2004-05-08. 基金项目: 国家自然科学基金资助项目(项目编号: 60234020).

第一作者简介: 孙海军(1977.12~), 男, 硕士研究生. 主要研究方向: 智能信息处理. Email: shj77@emails.bjut.edu.cn.

L.Howard Holley 和 Martin Karplus 提出用人工神经网络来预测蛋白质二级结构以来,人工神经网络一直是蛋白质二级结构预测研究领域中的一个非常重要的工具^[3,4].而且,与其它各种方法相比来说,应用神经网络已经取得了比较的预测结果^[5].

目前由单个神经网络得到的蛋白质二级结构预测的准确率比较低,本文通过用由 7 个前向型的 BP 神经网络组合成的“多模神经网络(复合神经网络)”对蛋白质二级结构进行了预测研究,同时给多模神经网络引进了较多的生物进化信息(Evolutionary information),即一方面引入了“profile”编码,这种编码被认为携带了较多的生物信息;另一方面引入了氨基酸之间的“距离”概念.它体现了输入层临近氨基酸的相互联系和影响.本文对从 36 个蛋白质提取的共 4 000 个氨基酸进行了预测研究.结果表明,与经典的 Multi-Modal^[1]网络相比,本文提出的多模神经网络预测模型把蛋白质二级结构预测的平均精度提高了 2.74%.

1 评测数据库和编码方式

1.1 评测数据库

本文把数据来源定位在 ftp://ftp.ebi.ac.uk/pub/databases/hssp/.我们从 HSSP(蛋白质同源结构数据库)数据库中提取出 36 个蛋白质,这 36 个蛋白质文件是文本文件见下表 1.

表 2 列出 HSSP 数据库文件的主要结构,Seq No 列代表氨基酸序列中氨基酸的序号,AA 列是氨基酸序列中具体的氨基酸,它将作为神经网络的输入;SS 列是 AA 列队对应的二级结构状态,它将作为神经网络的教师信号.由于,我们要预测的蛋白质结构是 3 状态的,即 H 螺旋、E 折叠和 L 薄片,而

HSSP 数据库中的二级结构状态是 8 状态的,这种 8 状态的二级结构是 3 状态的结构的细化.于是,先把 8 状态转化成 3 状态的形式.即:H 包括 H(α 螺旋)、G(3^{10} 螺旋)和 I(π 螺旋);E 包括 E(扩展折叠)和 B;L 包括 T(弯曲)、S(薄片)和(其它).Multiple alignment 列是和测试序列比对的相关序列,而 Profile of alignment 列给出了在每一个位置一种氨基酸类型出现的相对频率.

本文从表 1 中的 36 个蛋白质中,选取 4 000 个氨基酸,然后,依次取出 400 个氨基酸作为测试数据集,剩下的 3 600 个作为训练数据集.3 种二级结构所占百分比如下:

训练样本:H 20% E 31% L 49%
测试样本:H 20% E 31% L 49%

1.2 编码方式

本文所用的编码方式是“Profile”编码^[1,7],这种编码被认为携带了丰富的生物进化信息,一般认为生物进化信息对于提高蛋白质二级结构预测准确率是非常重要的.Profile 面是指在氨基酸序列的每一个位置上一个氨基酸类型出现的相对概率.例如:氨基酸 Q 在其对应的行“QQ…”序列中出现的概率显然是 1.0,因为在“QQ…”序列中只有氨基酸 Q;再如,氨基酸 S 出现的概率是 0.4,因为在“SSQAA”序列中 S 有两个,所以 S 的出现概率是 $2/5 = 0.4$,等等.具体的编码规则如表 3:

其中 Seq 列代表一个氨基酸序列片段,Multiple alignment 列是和测试(Seq)序列比对过的相关序列.从表 1 中我们看出氨基酸序列片段 QSEPEDLLK 的 Profile 编码是其对应的行(如方框中所示).例如氨基酸

表 1 36 个蛋白质文件

Tab.1 36 protein chains

1azu	1bbp-A	1cbh	1fc2-A	1ovo-A	1pyp
1tgs-I	2aa:	2ccy-A	2mev-4	2mhu	2rsp-A1
2tgp-I	2utg-A	2wrp-R	3hmg-B	3rnt	4cpv
6acn	6cpp	6hir	9wga-A	1cm	1cse-I
1cdt-A	1cc5	7rsa	1r09-2	1fdl-H	1eca
2gbp	2hmz-A	5er2-E	2stv	5ldh	3ebx

表 2 HSSP 数据库中文件的主要格式

Tab.2 The main format of HSSP file

Seq No	AA	SS	Multiple Alignment	Profile of alignment
1	A	S	AAAAAAAAAAA	1.0
2	V	S	AHSAAAAAVV	0.2
3	K	S	KKKKKKKPPK	0.8
4	H	H	AAHHCGAHHHA	0.4
5	W	H	WWPPPAWWWA	0.5
6	D		DEPPPPFLLLL	0.1
7	G	E	GGGHHHHL	0.4
8	P	E	PPPPPEPPP	0.91
9	R	E	RARDNNSTSS	0.2
10	C		CCCCCCCPC	0.9

Q 的 Profile 编码是:000000000000000001.0000;氨基酸 S 的 Profile 编码是:000000000.400.400000.2000 其余的以此类推.

表3 Profile 编码表例
Tab.3 Profile code

Seq	Multiple alignment	V	L	I	M	F	W	Y	G	A	P	S	T	C	H	R	K	Q	E	N	D
Q	Q Q . . .	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0	0	0	0
S	S S S Q A A	0	0	0	0	0	0	0	0	0.4	0	0.4	0	0	0	0	0	0.2	0	0	0
E	E K K K K E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8	0	0.2	0	0
P	P P P P P P	0	0	0	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0	0	0
E	E E E G E E	0	0	0	0	0	0	0	0.2	0	0	0	0	0	0	0	0	0	0.8	0	0
D	D D E D D D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0.8
L	L L L A A A	0	0.4	0	0	0	0	0	0	0.6	0	0	0	0	0	0	0	0	0	0	0
L	L L L I V I	0.2	0.4	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	K K K E K H	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0.6	0	0.2	0	0

表3中的 profile 编码是20位.在此基础上,我们加上了一位,用此位来代表我们引进的“距离”的概念.引进它的依据是氨基酸序列中的每一个氨基酸都不是孤立的,而是与它的周围的氨基酸有着某种联系.也就是说,一个氨基酸表现出某种二级结构状态,是它和其它的氨基酸共同“作用”才产生的.我们这里所说的“距离”因素是指氨基酸序列中的某个氨基酸的邻近氨基酸对其会有某种“联系(影响)”,我们用“距离”来表示这一影响力,由于输入层是一个滑动的“窗”,而每一次只是预测窗中心位置的氨基酸的二级结构状态,于是我们可以这样来定义这个“距离”,中心位置的氨基酸的距离 $D = 2$ (当然它自己对它本身的影响力最大),在它两侧

的氨基酸的距离数依次是 $1/k (k = 1, 2, \dots, N)$, 其中 N 是中心位置氨基酸的一侧氨基酸的总数.例如一个窗的长度是5的输入窗,它的每一个氨基酸的“距离”分别是“ $1/2, 1, 2, 1, 1/2$ ”.从上我们可以看出离中心位置氨基酸越近的氨基酸对其的影响力也就越大.反之,越小.于是,最后的编码是由21位组成的.

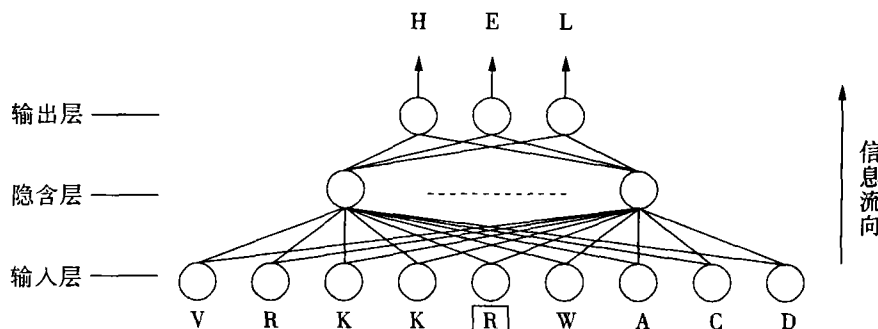


图1 SNN神经网络
Fig.1 Pattern for single neural network

2 评测系统

2.1 结构设计

本文采用一个由7个前向型的BP神经网络,如图1示.作为蛋白质二级结构预测的评测系统,如图2示.在每个BP神经网络的输入层引入了广为使用的“局部滑动窗”,这个“滑动窗”是沿着氨基酸序列“滑动”的窗口,每一次预测都是对“滑动窗”中间位置的氨基酸进行的.例如,对图1所示来说,这一时刻,就是对窗中间的氨基酸“R”进行预测,下一时刻将对氨基酸“W”进行预测.本文所用滑动窗的长度为9个氨基酸的长度,如图1中输入层是VRKKRWACD等9个氨基酸,假设每种编码

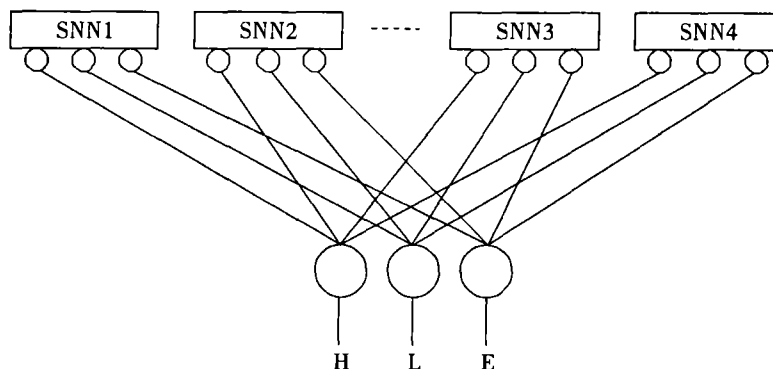


图2 多模神经网络框架
Fig.2 Frame for multi-modal neural network

方式编码一个氨基酸需要 K 个神经元,那么输入层的总的输入神经元个数为 $9 * K$,其中, k 代表每个氨基酸的编码位数(本文的 $k = 21$). 隐含层神经元个数为 40. 图 2 是多模神经网络框架图,其中 SNN1、SNN2...SNN7 代表如图 1 所示的 7 个前向型的 BP 神经网络,输出层是由三个神经元组成的,以对应三种蛋白质二级结构状态 H、E 和 L. 这 3 个神经元分别编码成 3 个二进制码. 即, H(螺旋): 100; E(折叠): 010; L(卷曲): 001. 和单个神经网络蛋白质二级结构预测不同,多模神经网络时把多个单神经网络预测的结果做一个求和操作,用这个“和”来最终决定此次预测的预测结果,“和”最大的二级结构状态值为 1,其它的状态值为 0. 这样做是因为即使 7 个 BP 神经网络的各种参数都相同,得到的预测结果也是有细微的差别,这个差别对二级结构的预测有一定的影响,如:某个单 BP 神经网络的一个预测结果数据如下: H: 0.416667, E: 0.410007, L: 0.173323 这时,如果取最大的值为 1,其它的值为 0,即得到: H: 1, E: 0, L: 0,但注意到 H 和 E 的值非常接近,因此,我们在分类的时候,就可能“误分”. 若用多模神经网络模型来预测,把 7 个神经网络的三个二级结构状态预测值求和值来决定最后的预测值,这样犯上述错误的概率会大大减小. “和”最大的状态值为 1,其它的状态值为 0.

2.2 网络算法

2.2.1 工作算法(Working Algorithm)

BP 神经网络的工作算法是由给定的输入信号产生输出信号的过程,即:

$$WA: X \rightarrow Y \text{ 其中, } X \text{ 是输入信号, } Y \text{ 是输出信号.}$$

详细的工作算法如下(网络共有 3 层):

(1) 计算输入层的电位向量

$$V^{(1)} = X - \theta^{(1)} \quad (1)$$

(2) 计算 1 到 2 层的输出向量 和 2 到 3 层的膜电位向量

$$O^{(k-1)} = f^{(k-1)}(V^{(k-1)}) \quad (2 \leq k \leq 3) \quad (2)$$

$$V^{(k)} = W^{(k)T} O^{(k-1)} - \theta^{(k-1)} \quad (2 \leq k \leq 3) \quad (3)$$

(3) 计算输出层(第 3 层)的输出向量

$$Y = O^{(L)} = f^{(L)}(V^{(L)}) \quad (4)$$

式中: $V^{(k)}$ 为第 k 层的膜电位; X 为样本输入向量; Y 为第 3 层的输出向量; W 为权值矩阵; θ 为阈值矩阵

2.2.2 学习算法(Learning Algorithm)

BP 神经网络的学习算法包括两大步骤:其一是输入的正向传播过程,即输入数据从输入层逐层向输出层传播,得到输出响应;其二是输出误差的反向传播过程,即输出的误差由输出层开始反向传播到输入层,而网络各层的权值改变量则根据传播到该层的误差大小来决定. 也就是从样本中选取输入模式作为 BP 网络的输入,并测试其输出,即测试实际输出与期望输出得一致性,已指导 BP 网络权值和阈值的修改. BP 神经网络学习算法的步骤如下:

(1) 初始化网络的权值矩阵 W 和神经元阈值矩阵 θ 为小的随机数矩阵.

(2) 提供学习样本,即输入向量 X 和输出向量 Y_n ($n = 1, 2, 3$);

(3) 计算网络的实际输出及隐含层各神经元的输出.

$$O_{pj} = f(\sum W_{ij} I_{pi} - \theta_j) \quad (5)$$

$$O_{pk} = f(\sum W_{jk} I_{pj} - \theta_k) \quad (6)$$

式中: f 为 Sigmoid 函数, $f(x) = 1/(\exp(-x) + 1)$; I_{pi} 为训练样本为 p 时节点 i 的输入; O_{pj} 为训练样本为 p 时隐含层第 j 个节点的输出, O_{pk} 为训练样本为 p 时输出层第 k 个节点的实际网络输出.

(4) 计算误差对各层的影响:

$$\delta_{pk} = O_{pk}(1 - O_{pk})(Y_{pk} - O_{pk}) \quad (7)$$

$$\delta_{pj} = O_{pj}(1 - O_{pj})(\sum \delta_{pk} W_{jk}) \quad (8)$$

式中: δ_{pk} 和 δ_{pj} 分别是输出层和隐含层反传误差信号, Y_{pk} 为训练样本为 p 时网络输出层节点 k 的期望输出.

(5) 修正权值和阈值.

$$W_{ji}(n+1) = W_{ji}(n) + \eta \delta_{pj} O_{pj} + \alpha (W_{ji}(n) - W_{ji}(n-1)) \quad (9)$$

$$\theta_j(n+1) = \theta_j(n) + \eta \delta_{pj} + \alpha (\theta_j(n) - \theta_j(n-1)) \quad (10)$$

(6) 计算网络输出误差.

$$E = 1/2 \sum (Y_{pk} - O_{pk})^2 \quad (11)$$

式中: η 为学习率(一般取 0.01 ~ 0.8); α 为动量因子(一般取 0.9 左右).

(7) 判断误差是否满足要求,若满足则转到(8),否则转到(3).

(8) 训练结束.

2.3 预测准确率的衡量

我们选用最简单并且被广泛使用的评估公式来表示预测的精度,总的预测准确度公式为

$$Q_3 = 100 \sum_{i=1}^3 C_i / N \quad (12)$$

式中: Q_3 表示总的预测精度, C_i 表示每一种二级结构状态预测正确的氨基酸总数($i = 1, 2$ 和 3), 其中 $i = 1, 2$ 和 3 分别代表 H, E, L, N 表示总的样本(氨基酸)个数.

每种蛋白质二级结构状态预测的准确度公式:

$$Q_H = 100 C_H / N_H \quad (13)$$

$$Q_E = 100 C_E / N_E \quad (14)$$

$$Q_L = 100 C_L / N_L \quad (15)$$

式中: Q_H , Q_E 和 Q_L 分别代表 3 个蛋白质二级结构状态的预测精度; N_H , N_E 和 N_L 分别是 3 种二级结构状态的氨基酸总的个数.

3 试验与分析

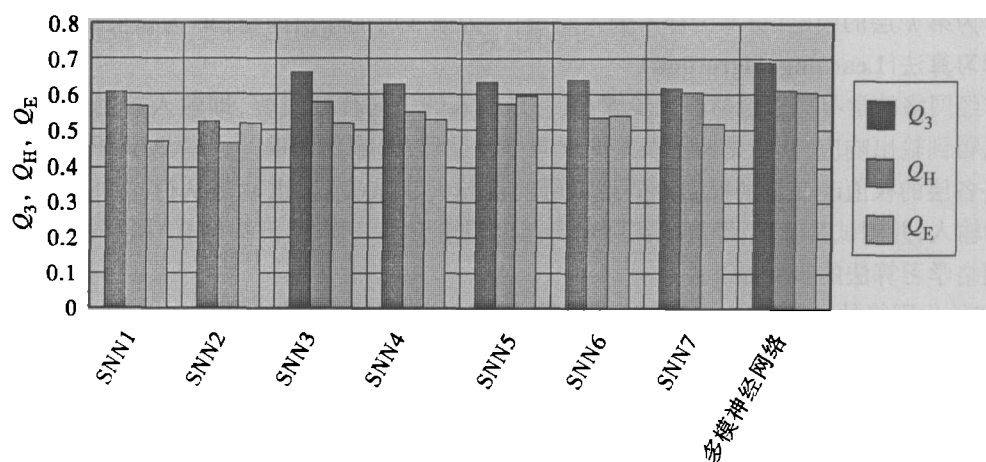


图3 第一组预测的结果
Fig.3 Result for first group data

本文首先把从 HSSP 数据库取出来的 4 000 个氨基酸分成 10 等份,每一份的三种二级结构状态 H, E 和 L 所占的比例相同(H 20% E 31% L 49%),并对每一份氨基酸序列进行 Profile 编码,之后,依次从 10 份中取出一份(400 个氨基酸)作为测试样本,剩下的 9 份作为训练样本.这样共有 10 组的训练样本和测试样本数据.之后把编码好的每一组的训练样本数据输入到评测系统中,开始训练评测系统,待系统训练完毕后,再把编码好的测试样本输入到评测系统中得到实际系统输出的评测结果.这个过程需要依次进

行10次,之后,把10次的预测结果取平均值作为最终的预测结果.这种方法叫做交叉校验法,图3是10组预测结果中的第一组结果.表4是它的具体细节.

从表4中可以看出对 Q_3 来说,单个BP神经网络的预测结果要比多模神经网络的预测结果低很多,换句话说,用多模神经网络模型预测可以大幅度地提高预测蛋白质二级结构的准确率,而且这个结果比文献[1]所取的结果高出2.8个百分点.同时 Q_H 、 Q_E 和 Q_L 三种二级结构状态的预测结果也高于7个SNN(单神经网络)预测结果.图4是10组预测的结果和最终的预测结果与MNN预测结果的比较,表5是它的具体细

表4 图3的具体细节

Tab.4 The detail values of Figure 3

神经网络	Q_3	Q_H	Q_E	Q_L
SNN1	0.606 280	0.565 891	0.471 264	0.691 919
SNN2	0.526 570	0.465 116	0.517 241	0.570 707
SNN3	0.659 420	0.581 395	0.517 241	0.772 727
SNN4	0.628 019	0.550 387	0.528 736	0.722 222
SNN5	0.632 850	0.573 643	0.597 701	0.686 869
SNN6	0.642 512	0.534 884	0.540 230	0.757 576
SNN7	0.615 942	0.604 651	0.517 241	0.666 667
多模神经网络	0.690 821	0.614 351	0.607 401	0.787 879

节.图4是10组预测的结果和最终的预测结果与MNN预测结果的比较,表5是它的具体细

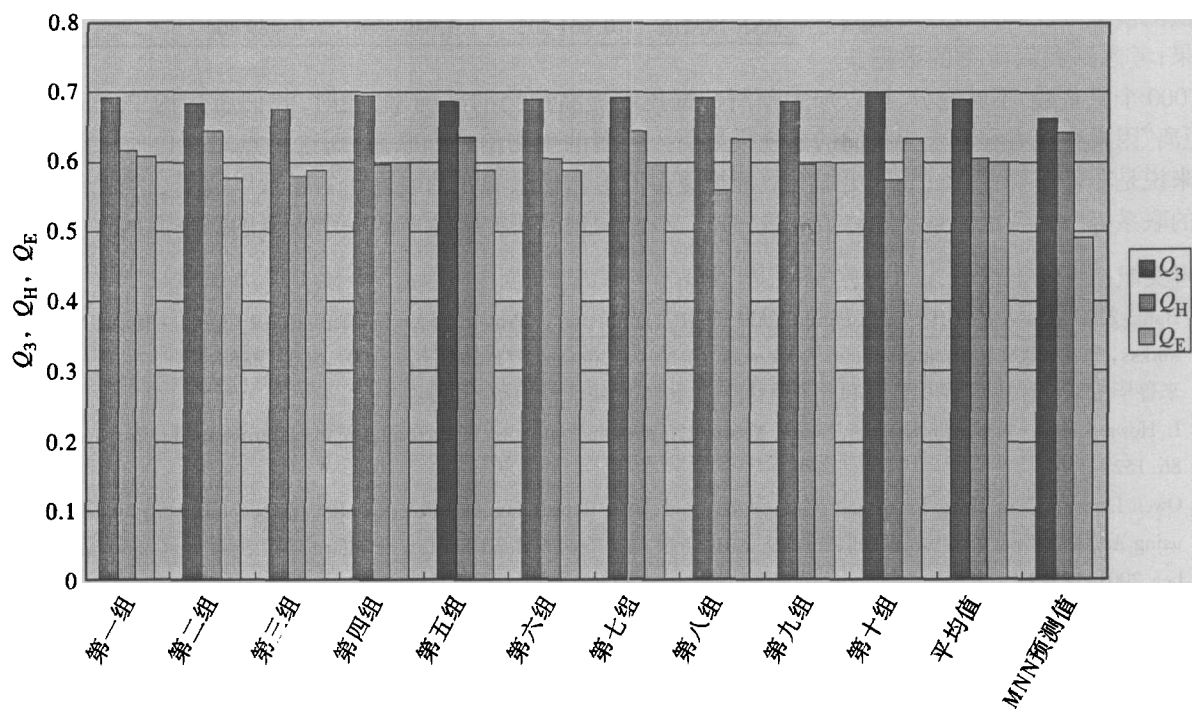


图4 10组预测的结果和最终的预测结果与MNN预测结果的比较
Fig.4 Prediction proteins For 10 groups, average accuracy and MNN

节.可以看出10次 Q_3 预测结果的平均达到了68.89%,而MNN的预测结果是66.15%,提高了大约2.74个百分点. Q_E 的预测结果提高了大约11个百分点,但是 Q_H 的值却小了大约3个百分点.分析预测结果提高的原因是我们把文献[1]的由5个BP神经网络做成的网络评测系统换成了我们提出的由7个BP神经网络组成的评测系统模型,这个结构的变化可以减少错误分类的概率,同时在氨基酸的编码里增加了一个“距离”的因素,这个因素考虑到了相邻氨基酸之间的存在的某种联系,更加符合生物学原理.而且,我们所用的数据集还是不够充分.如果把我们的预测数据集的氨基酸的数量从4000个提高到文献[1]所用的6137个,可能我们的网络对各个二级结构的“认识”程度会有所提高,预测结果可能会进一步提高. Q_H 的预测结果比文献[1]的预测结果低可能是因为我们的预测数据集中包含H(螺旋)结构的氨基酸个数(比例)过小,在我们的模型里有800个H(螺旋)结构的氨基酸,而在文献[1]中用到了1902个H(螺旋)结构的氨基酸,所以我们的评测系统对H(螺旋)结构的学习和认知程度没有文献[1]的评测系统的学习和认知

程度好,也就是说我们的系统没能很好地“记住”H(螺旋)结构的特征.

4 结束语

本文借助多模神经网络对蛋白质二级结构进行了预测研究,其数据集是从 HSSP 数据库中提取的数据经过处理后得到的评测数据库.同时在 Profile 编码中引进了“距离”的概念,这个概念表达了氨基酸序列中各个氨基酸的相互影响的程度,这点更能体现氨基酸之间的生物联系.实验结果表明,我们提出的多模神经网络预测模型可以取得较好的预测结果;考虑到我们所用的数据集

(4 000 个氨基酸)不够充分,可以加大预测数据集以提高预测的准确率.同时,我们所用的 Profile 编码和“距离”因素在很大程度上为我们的系统模型引入较多的生物信息和联系,而这一点对蛋白质二级结构预测来说是非常重要的.因此,如何更加有效地把更多的生物信息和氨基酸之间的联系以及蛋白质二级结构之间的联系等引入到预测系统中将成为蛋白质二级结构预测研究领域中的一个重要的研究课题.

参考文献:

- [1] Hanxi ZHU, Ikuo YOSHIHARA, Kunihito YAMAMORI, Prediction of Protein Secondary Structure by Multi-Modal Neural Networks, Proc. 2002 International Joint Conference on Neural Networks (IJCNN'02), 2002, 280 ~ 285.
- [2] 来鲁毕主编. 蛋白质的结构预测和分子设计[M]. 北京: 北京大学出版社, 1993, 1 ~ 2.
- [3] L. Howard Holley & Martin Karplus. Neural Network Applied in Prediction of Protein Secondary Structure[J]. Biophysics, 1989, 86: 152 ~ 156.
- [4] Owen Lamont, Hiew Hong Liang and Matthew Bellgard. Data Representation influences Protein Secondary Structure Prediction using Artificial Neural Networks[s]. Seventh Australian and New Zealand Intelligent Information Systems Conference, November, 2001, 18 ~ 21.
- [5] 方慧生 相秉仁 安登魁. 改进 Madaline 学习算法预测蛋白质二级结构[J]. 中国药科大学学报, 1996, 27(6): 366 ~ 369.
- [6] 王龙会 石峰. 遗传神经网络及其在蛋白质二级结构预测中的应用[J]. 数学杂志, 2002, 22(2): 179 ~ 185.
- [7] Henrik Bohr, Jakob Bohr, Seren Brunaks etc. Protein secondary structure and homology by neural networks[J]. The α -helices in rhodopsin, 1988, 241(2): 223 ~ 228.

表 5 图 4 的具体细节

Tab.5 The detail values of Figure 4

	Q_s	Q_H	Q_E	Q_L
第一组	0.690 821	0.614 351	0.607 401	0.787 879
第二组	0.683 575	0.643 411	0.574 713	0.757 576
第三组	0.676 328	0.579 148	0.586 206	0.772 727
第四组	0.695 652	0.596 899	0.597 701	0.803 030
第五组	0.686 329	0.634 651	0.586 207	0.762 626
第六组	0.688 406	0.604 651	0.586 207	0.787 879
第七组	0.690 821	0.643 410	0.597 701	0.762 626
第八组	0.693 237	0.558 139	0.632 183	0.808 081
第九组	0.685 990	0.596 899	0.597 701	0.782 828
第十组	0.700 483	0.573 643	0.632 184	0.813 131
平均值	0.688 903	0.604 519	0.599 820	0.783 838
MNN 预测值	0.661 502	0.640 499	0.489 404	