第 37 卷第 2 期 河北科技大学学报 Vol. 37, No. 2
2016 年 4 月 Journal of Hebei University of Science and Technology Apr. 2016

文章编号:1008-1542(2016)02-0167-06 doi:10.7535/hbkd.2016yx02009

蛋白质二级结构预测概率图模型的改进

赵凌琪[1],朱丽娟[2],王柯静[2],董小庆[2],张　屹[1,2]

(1.内蒙古民族大学计算机科学与技术学院,内蒙古通辽　028043;2.河北科技大学理学院,河北石家庄　050018)

摘　要:蛋白质二级结构与蛋白质三级结构及蛋白质功能密切相关,是生物信息学研究的热点,其中概率图模型隐马尔可夫算法(HMM)是该领域研究的重要工具。但是在实际应用中,存在着HMM 训练下溢、不同训练集的效果差异较大及参数优化困难等问题。对预测蛋白质二级结构时HMM 遇到的训练下溢问题提出了改进方案;首次提出 8-状态 HMM 来预测蛋白质二级结构,并且将参数 B 改进成为包含状态转移信息的三维参数;为了改进最优 HMM 模型的确定方法,用每个样本分别对初始 HMM 模型进行训练,得到一系列新的模型,然后对这些新模型的参数求均值,将求得的均值作为最优模型的参数。这些改进方法提高了 HMM 预测蛋白质二级结构的准确率,为 HMM 的进一步优化打下良好的基础。

关键词:应用生物化学;概率图;蛋白质二级结构;HMM;下溢;参数优化

中图分类号:O175.8　　**文献标志码:**A

Improved probability graph model for protein secondary structure prediction

ZHAO Lingqi[1], ZHU Lijuan[2], WANG Kejing[2], DONG Xiaoqing[2], ZHANG Yi[1,2]

(1. College of Computer Science and Technology, Inner Mongolia University for Nationalities, Tongliao, Inner Mongolia 028043, China; 2. School of Science, Hebei University of Science and Technology, Shijiazhuang, Hebei 050018, China)

Abstract: Protein secondary structure is closely related to protein tertiary structure and function, and became a hot topic in bioinformatics. The probability graph model HMM (Hidden Markov model) is an important tool in this field. In practice, there exist problems such as: HMM training underflow, significant result differences derived from different training set, and hard process of parameter optimization. In this paper, aiming at HMM training underflow problem when predicting protein secondary structure, we put forward a method for solving the underflow problem; propose an 8-state HMM model to predict protein secondary structure for the first time; and modify parameter to be a three-dimensional parameter containing the state transition information. In order to improve the method drilling the optimal HMM, we train the initial HMM model with each sample, and get a series of new models; then average the parameters of the new models, and the obtained average parameter values are used to construct the optimal HMM model. The improved method increases the accuracy of protein secondary structure pre-

---

收稿日期:2015-11-05;修回日期:2016-01-13;责任编辑:张士莹

基金项目:国家自然科学基金(61261025,11171088);河北省自然科学基金(A2015208108)

作者简介:赵凌琪(1969—),女,内蒙古赤峰人,教授,主要从事组合图论算法方面的研究。

通讯作者:朱丽娟。E-mail:1439837430@qq.com

张　屹教授。E-mail:zhaqi1972@163.com

赵凌琪,朱丽娟,王柯静,等.蛋白质二级结构预测概率图模型的改进[J].河北科技大学学报,2016,37(2):167-172.

ZHAO Lingqi, ZHU Lijuan, WANG Kejing, et al. Improved probability graph model for protein secondary structure prediction[J]. Journal of Hebei University of Science and Technology, 2016, 37(2):167-172.

diction, hence it is a good foundation for further improvement of HMM.

Keywords: applied biochemistry; probability graph; protein secondary structure; HMM; underflow; parameter optimization

生物学原理[1]指出:DNA 序列首先通过转录、翻译生成氨基酸序列,氨基酸序列再经过盘曲折叠形成复杂的二级、三级等空间结构。研究发现,蛋白质二级和三级结构能有效揭示蛋白质的生物功能[2-4],可以用 X 射线晶体学技术、电子显微镜技术和核磁共振衍射技术(NMR)等来测定它们[5]。这些方法虽然比较直接,但预测时周期长、成本高、技术难度大。人们越来越多地关注生物信息学的应用,即基于经验知识与计算化学、统计物理学、信息学等方法的结合,从理论上预测蛋白质的空间结构[6-9],其中二级结构是预测三级结构的基础。蛋白质二级结构的主要形式包括 α-螺旋、3_{10}-螺旋、π-螺旋、β-折叠、β-转角和无规则卷曲等。最流行的二级结构预测方法有 DSSP[10],STRIDE[11],DEFINE[12]等,通常利用 DSSP 算法将蛋白质二级结构分为 8 类:H(α-helix),G(310-helix),I(π-helix),E(extended β-strand),B(isolated β-strand),T(turns),S(bend),C(coil)[5,13]。在进行蛋白质二级结构预测时,一般又把这 8 类归并为 3 大类,即 H(helix),E(sheet)和 C(coil),通常有 CK 模式[14]、EHL 模式[15]和 PSIPRED[16]这 3 种归类方法。目前,大多采用 EHL 模式归类方法,即 H,G,T→H;E,B→E;其余→C[13]。在对蛋白质二级结构进行预测时,选取的特征有氨基酸的亲水性、疏水性和极性[13],蛋白质二级结构 SS-8 的 8 种状态等[13,17],而其预测方法则包括最近邻方法[18-20]、人工神经网络(ANN)[16,21-22]、支持向量机(SVM)[23-24]和隐马尔可夫算法(HMM)[25-26]等。

本文研究的蛋白质二级结构预测算法可以同时实现 2 个目的:界定一个蛋白质序列中所蕴含的二级结构片段的边界及对分割出来的片段进行 EHL 模式的分类预测。

1 材料与方法

1.1 数据集与准确率

在对 HMM 算法进行改进时,采用目前使用较多的 CUFF 和 BARTON 建立的 CB513 数据集(http://www.compbio.dundee.ac.uk/jpred/about.shtml),包括 513 条蛋白质,分别用 X 射线测定二级结构,分辨率>2.5 Å,相似性<25%。首先对该数据集进行预处理,去掉 16 条长度小于 30 的序列,然后去掉 5 条含有'?'的序列,剩下 492 条蛋白质序列。将 492 条序列随机分为 7 等份,每次都取其中的 1 份(72 条)作为测试集,取剩下的 6 份(即 420 条)作训练集;如此循环 7 次,完成 7 重交叉检验。下文中提到的准确率都是这 7 重交叉检验的平均准确率。

准确率的定义为一条氨基酸序列的每个氨基酸都从属于蛋白质二级结构状态中的一种,在对蛋白质二级结构进行预测时,是对其每个氨基酸的所属状态进行预测。本条序列的预测准确率定义为

$$\text{准确率}=\frac{\text{本序列中被正确预测结构状态的氨基酸数}}{\text{本序列中的氨基酸总数}}\times 100\%\ ,$$

而基于某个 HMM 对一组序列进行蛋白质二级结构预测的准确率定义为对所有序列预测准确率的平均值。

1.2 隐马尔可夫算法

HMM 是一种概率图算法,在语音识别[7]和生物信息学中[8-9]有广泛应用。以往的蛋白质二级结构预测常利用 3-状态或 7-状态 HMM,由于 7-状态比 3-状态包含的蛋白质信息较多,所以前者预测精度较高。因此,本文蛋白质二级结构预测在 7-状态的基础上进行改进,提出 8-状态 HMM,把蛋白质的 8 种二级结构状态分别作为状态 $S_1,S_2,\cdots,S_8$,包括 α-螺旋的起始、中间和结束状态 HB,H,HE(α-螺旋的第 1 个氨基酸状态为 HB,最后一个氨基酸状态为 HE,中间部分的氨基酸状态为 H),β-折叠的起始、中间和结束状态 EB,E,EE(氨基酸状态归属情况同 α-螺旋),无规则卷曲状态 C 及非二级结构状态 F。蛋白质的每个氨基酸都从属于 8 种状态的一种,随着同一蛋白质上各种二级结构的交替出现,氨基酸的 8 种状态之间就发生了"转移"(transition),转移概率矩阵记为 $\boldsymbol{A}_{8\times 8}$。HMM 把每个氨基酸看成是在该氨基酸的当前状态之下依照概率矩阵 $\boldsymbol{B}_{8\times 20}$ 而"发出"(emit)的,加上初始状态分布 π,$M=(\pi,A,B)$就被称为 HMM 模型。一个蛋白质序列的氨基酸是"可见"(observable)的符号,而蛋白质上的二级结构则是"隐含"(hidden)的状态,这就是 HMM 的由来。显然,HMM 试图用统一的状态转移矩阵和各状态下可见符号的出现概率来描述蛋白质序列中的二级结构与氨基酸序列的依存性,从而用这种依存性来预测陌生蛋白质中二级结构分布。实际上,还可以建立 2

个连续状态组合成的 $8\times8=64$ 种状态甚至 $8^3=512$ 状态等模型，但是随着矩阵 $\boldsymbol{A}$ 和矩阵 $\boldsymbol{B}$ 的增大，计算量也会空前增大。所以，该文只考虑 8-状态的 HMM 模型。

一般来说，利用 HMM 来预测蛋白质二级结构会涉及到以下 3 个问题。

问题 1　求由已知 HMM 模型生成已知氨基酸序列的概率。

给定蛋白质序列 $O=O_1,O_2,\cdots,O_T$ 以及训练好的模型 M，计算条件概率 $P(O|M)$ 的值。基于不同的模型 M，同一条蛋白质序列的出现概率值可以不同。计算条件概率 $P(O|M)$ 的算法有前向算法、后向算法及前后向算法[5,13]。

前向算法的步骤如下(其他类似)。

1) 初始化：$\alpha_1(i)=1, 1\leqslant i\leqslant N$；2) 递归：$\alpha_{t+1}(j)=\sum_{i=1}^{N}\alpha_t(i)a_{ij}b_{ij}(O_t), 1\leqslant t\leqslant(T-1), 1\leqslant j\leqslant N$；3) 累加：$P(O\mid M)=\sum_{i=1}^{N}\alpha_T(i)$。此评估问题的解决只是简单的迭代，没有运用复杂的算法。这里 T 是氨基酸序列的长度；$N=8$ 是状态数；状态 $i\rightarrow j$ 转移概率组成矩阵 $\boldsymbol{A}=(a_{ij})$，而这一状态的转移是由第 t 个位置上氨基酸 O_t 的输出来实现的，概率是 $b_{ij}(O_t)$；$\alpha_t(i)$ 是一个概率值，表示第 t 个位置的氨基酸处于第 i 个状态的概率。如果氨基酸序列较长，则当 t 大于某一值时，α 值会很小，几乎接近 0，Matlab 程序会默认为 0。为了避免这种下溢情况，本文采用文献[5]中的比例因子方法。

问题 2　已知 HMM 模型和氨基酸序列，求最佳二级结构状态序列。

给定观察序列 $O=O_1,O_2,\cdots,O_T$ 以及模型 M，用动态规划算法把蛋白质序列分成不同阶段，通过把条件概率每步最大化来找到一个概率最大的状态序列 $S=S_1,S_2,\cdots,S_T$，作为观察序列 O 的最佳解释。

解决该问题的算法为动态规划的 Viterbi 算法[5,13]。1) 初始化：$\alpha_1(i)=1, 1\leqslant i\leqslant N$；2) 每步都追求最优化的递归：$\alpha_{t+1}(j)=\max(\alpha_t(i)a_{ij}b_{ij}(O_t)), 1\leqslant t\leqslant(T-1), 1\leqslant j\leqslant N$；3) 得到最终结果：$P(O\mid M)=\max(\alpha_T(i)), 1\leqslant i\leqslant N$。

问题 3　由已知序列和二级结构的训练集得到最佳 HMM 模型。

给定观察序列 $O=O_1,O_2,\cdots,O_T$ 的集合，需要找到最佳模型 $M=(\pi,A,B)$，使得 $P(O\mid M)$ 最大。基于训练集中的 420 条氨基酸序列，如果要找到一个包含尽可能多的序列与二级结构对应信息的模型，必须把模型 M 在各个氨基酸序列上反复迭代来收集信息和改进模型。通过启发迭代来训练模型 $M=(\pi,A,B)$ 的算法为 Baum-Welch 算法，该算法中用来优化矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 中元素的迭代重估公式为

$$\hat{a}_{ij}=\frac{\sum_t\alpha_{t-1}(i)a_{ij}b_{ij}(O_t)\beta_t(j)}{\sum_t\alpha_t(i)\beta_t(i)},$$

$$\hat{b}_{ij}=\frac{\sum_{t:O_t=k}\alpha_{t-1}(i)a_{ij}b_{ij}(O_t)\beta_t(j)}{\sum_t\alpha_{t-1}(i)a_{ij}b_{ij}(O_t)\beta_t(j)}。$$

简单来说，先给出 $\boldsymbol{A}=(a_{ij})$ 和 $\boldsymbol{B}=(b_{ij})$ 的初始值；然后，基于一个氨基酸序列及其中已知的二级结构信息，按迭代重估公式计算 $\hat{a}_{ij}$ 和 $\hat{b}_{ij}$；再取一条已知二级结构“状态列”的氨基酸序列，计算 $\hat{a}_{ij}$ 和 $\hat{b}_{ij}$，如此反复，直至 $\hat{a}_{ij}$ 和 $\hat{b}_{ij}$ 收敛。显然这个启发式算法是非常费时间的，与上面的简单迭代和流程固定的动态规划不同。当不同氨基酸序列之间信息差异很大时，会在某些情况下出现随着训练次数或训练序列数的增加效果反而下降的情况。本文探讨了如何寻找最佳训练集的问题。利用 HMM 进行蛋白质二级结构预测时，先是基于训练集中的 420 条氨基酸序列选出最优的矩阵 $\boldsymbol{A},\boldsymbol{B}$；然后针对二级结构的氨基酸序列，用训练出来的模型来“解码”，预测其最佳的二级结构状态序列。

1.3　原有蛋白质二级结构预测的 HMM 模型中存在的问题和本文的改进方案

人们用 3-状态和 7-状态 HMM 对蛋白质二级结构进行预测时[5,13]是基于 3 种二级结构来定义状态的，未考虑到非二级结构的状态 F，本文中加上了非二级结构的状态 F。参数 B 在蛋白质二级结构预测时起着很重要的作用，但其并未包含状态间的转移，本文把 B 改造成三维参数来包含状态转移的信息。另外，利用 Baum-Welch 算法进行参数重估，当训练样本不含某状态时，该状态转移到其他状态的情况就不存在，则其对应的参数值为不定数，即训练下溢。针对训练下溢的情况，本文的改进方案是当初始模型选定后，训练参

数直至下溢,保留这个过程中最佳模型的参数,而不使用添加因子等传统手段。

2　结果与讨论

2.1　新的 8-状态 HMM 算法流程

通过对 CB513 数据集中的序列进行统计,α-螺旋的长度最少为 3 个氨基酸,为了提供更多的蛋白质结构信息,将 α-螺旋的状态 H 扩展为 HB(起始位置)、H(中间位置)和 HE(结束位置)。虽然折叠的长度最少是 1,但将其状态扩展为 EB,E,EE 后,预测准确率有所提高,故采取其 3-状态的情况。对于无规则卷曲,将其扩展为 3-状态后,预测准确率并没有明显提高,故采取其单状态的情况。在此基础上增加非二级结构状态 F,共计 8-状态,其 HMM 结构如图 1 所示。在利用 8-状态 HMM 进行蛋白质二级结构预测时,首先根据其中的每个氨基酸的所属结构将氨基酸序列转化为状态序列,然后通过统计得到状态转移概率矩阵 $\boldsymbol{A}$ 和 8-状态下 20 种氨基酸的"发出"概率矩阵 $\boldsymbol{B}$,这就是"由已知序列和二级结构的训练集得到最佳 HMM 模型"的算法问题。基于这个 HMM 模型,针对一条陌生的序列,按着上面介绍的"已知 HMM 模型和氨基酸序列,求最佳二级结构状态序列"算法流程,就可以预测这条陌生序列的二级结构了。

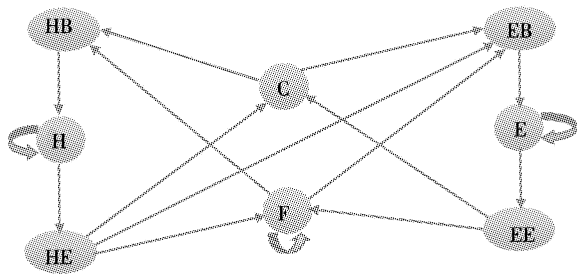

图 1　8-状态 HMM 结构

Fig. 1　8-state HMM structure

2.2　参数 B 定义的改进

在将不同的氨基酸序列预测为其对应的状态序列时,参数 B 起着很重要的作用,B 包含的信息越多,预测准确率就越高,故考虑将二维状态的 B 改进为三维,使其也蕴含状态间的转移。即当 B 为二维时,$B(i,k)$ 表示状态 i 输出氨基酸 k 的概率;当 B 为三维时,$B(i,j,k)$ 表示状态 i 转移到状态 j 时输出氨基酸 k 的概率。由于二维参数 B 的大小是 8×20,三维的大小是 $8\times8\times20$,所以后者的运算速度会相对慢一些,但准确率有较明显的提高。当初始参数中 B 为二维时,蛋白质二级结构预测的准确率为 0.255 11;当 B 为三维时,蛋白质预测准确率为 0.310 48。这说明三维 B 的确对算法有促进作用。

2.3　针对下溢问题而采用的平均 HMM 模型

在初始 HMM 模型选定的情况下,选取第 1 个训练样本进行更新,然后随机选取样本继续更新,直至下溢;选取第 2 个训练样本对初始 HMM 进行更新,直至下溢。如此下去,由 420 个训练序列分别起始的更新流程会得到共 420 个 HMM 模型。利用每次训练得到的 HMM 模型对测试集进行预测得到的准确率与训练次数的关系见图 2。

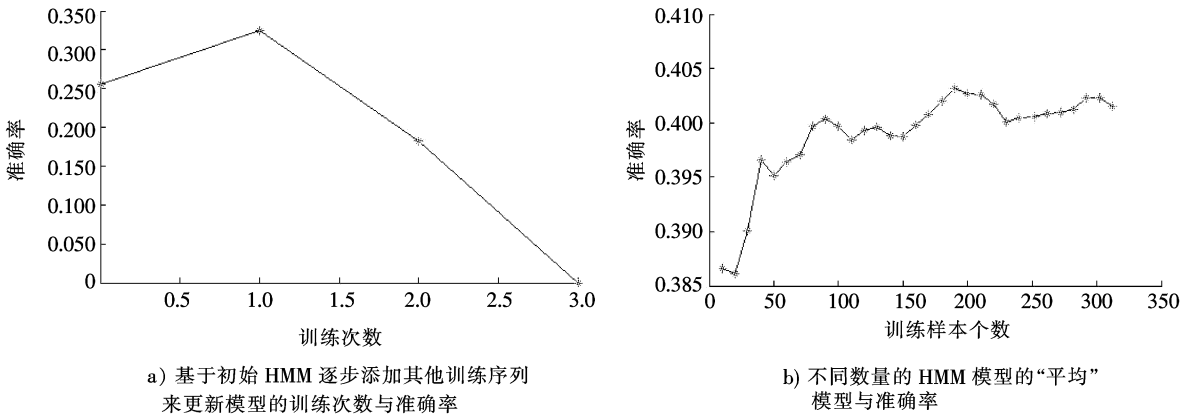

图 2　更新次数及平均模型(或训练样本)个数与准确率的关系

Fig. 2　Relationship between renew times as well as numbers of average models with accuracy

由图 2 可知，第 1 次训练参数的预测效果最好，由于在用 Baum-Welch 算法进行参数训练时，每训练一次，参数中不定数的个数会增加，预测效果反而会变差。因此，由初始参数分别对每个样本训练一次，得到 420 个 HMM 模型（当然包含 $\boldsymbol{A}$ 和 $\boldsymbol{B}$），然后分别用这 420 个 HMM 对测试集进行预测，经过统计，其中最好的 HMM 模型的准确率达到 0.361 84。为了进一步提高预测准确率，在删掉 108 组准确率为 0 的 HMM 后，对剩余的 312 个 HMM 的参数矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别求平均值，最后利用这个“平均”HMM 对测试集进行预测，准确率提高为 0.401 55。总之，分别从每条训练序列开始 HMM 模型的更新过程，不用“缩放因子”等传统方法进行人为干预，而是让它一直训练直至下溢，选取其中准确率最高的 HMM 模型存放，每条训练集序列都这样训练到下溢并存储此过程中准确率最高的 HMM 模型，然后对有效的 312 个 HMM 模型中的参数 A 和 B 分别进行平均，作为最终的 HMM 模型的参数。可喜的是，基于这个最终的“平均”模型的预测精度的确提高了。

3　结　论

文献[5]中 3-状态的预测精度为 0.389 0，7-状态的预测精度为 0.414 3；文献[13]中 3-状态的预测精度为 0.447 1。在文献[5]中采用的是 HGI 分类，相比 EHL 分类较容易预测；而文献[13]为了避免下溢情况，在 Baum-Welch 算法的重估公式中引入拉普拉斯修正项，虽然精度有所提高，但没有从根本上解决下溢问题。本文不使用“缩放因子”等传统手段，通过在参数维数、训练模型等方面的改进来提高预测精度，为应用 HMM 预测蛋白质二级结构的改进提供了新的思考方向。

本文阐述了 HMM 预测蛋白质二级结构时遇到的训练下溢问题、非二级结构状态和参数 B 的优化问题，并尝试了改进方案，提高了应用 HMM 预测蛋白质二级结构的准确率。另外，本文是在没有利用同源信息的情况下对 HMM 算法进行改进的，根据 AYDIN 等[27]的论述，这样的结果更具有代表性。可见，本文的改进算法为模型的进一步优化打下了良好的基础。

参考文献/References：

[1] 张海霞. 蛋白质二级结构预测方法研究[D]. 大连：大连理工大学，2004.
ZHANG Haixia. The Research on Protein Secondary Structure Prediction Methods[D]. Dalian：Dalian University of Technology，2004.

[2] MARSDEN R，RANEA J，SILLERO A，et al. Exploiting protein structure data to explore the evolution of protein function and biological complexity[J]. Philosophical Transactions of the Royal Society of London，2006，361(1467)：425-440.

[3] WHISSTOCK J，LESK A. Prediction of protein function from protein sequence and structure[J]. Quarterly Reviews of Biophysics，2003，36(3)：307-340.

[4] DOMINGUES F，LENGAUER T. Protein function from sequence and structure data[J]. Applied Bioinformatics，2003，2(1)：3-12.

[5] 林锦华. 基于隐马尔可夫模型的蛋白质二级结构预测[D]. 福州：福建农林大学，2012.
LIN Jinhua. Protein Secondary Structure Prediction Based on the Hidden Markov Model[D]. Fuzhou：Fujian Agriculture and Forestry University，2012.

[6] 王鹏良，江寿平，罗宇，等. 蛋白质二级结构预测的综合分析[J]. 物理化学学报，1990，6(6)：686-691.
WANG Pengliang，JIANG Shouping，LUO Yu，et al. Comprehensive analysis of the prediction of protein's secondary structure[J]. Acta Physico-Chimica Sinica，1990，6(6)：686-691.

[7] LAWRENCE R，RABINER A. Tutorial on hidden Markov models and selected applications in speech recognition[J]. Proceedings of the IEEE，1989，77(2)：257-286.

[8] STANKE M，WAACK S. Gene prediction with a hidden Markov model and a new intron submodel[J]. Bioinformatics，2003，19(sup2)：215-225.

[9] PEDERSEN S，HEIN J. Gene finding with a hidden Markov model of genome structure and evolution[J]. Bioinformatics，2003，19(2)：219-227.

[10] KABSCH W，SANDER C. Dictionary of secondary structure pattern recognition of hydrogen-bonded and geometric features[J]. Biopolymers，1983，22(12)：2577-2637.

[11] HEINIG M，FRISHMAN D. STRIDE：Web server for secondary structure assignment from known atomic coordinates of proteins[J]. Nucleic Acids Research，2004，32：500-502.

[12] RICHARDS F，KUNDROT C. Identification of structural motifs from protein coordinate data：Secondary structure and first-level supersecondary structure[J]. Proteins，1988，3(2)：71-84.

[13] 石鸥燕. 蛋白质结构预测模型的研究[D]. 天津：天津医科大学，2008.

SHI Ouyan. Study on the Models of Protein Structure Prediction[D]. Tianjin:Tianjin Medical University,2008.

[14] CHANDONIA J, KARPLUS M. Neural networks for secondary structure and structural class predictions[J]. Protein Science, 1995, 4(2): 275-285.

[15] MOULT J,FIDELIS K,ZEMLA A,et al. Critical assessment of methods of protein structure prediction(CASP): Round Ⅳ[J]. Proteins, 2001,45(sup5): 2-7.

[16] JONES D. Protein secondary structure prediction based on position-specific scoring matrices[J]. Journal of Molecular Biology,1999,292(2):195-202.

[17] 王勇献. 蛋白质二级结构预测的模型与方法研究[D]. 长沙:国防科学技术大学,2004.
WANG Yongxian. Research on the Models and Methods of Protein[D]. Changsha:National University of Defense Technology,2004.

[18] LEVIN J,ROBSON B,GARNIER J. An algorithm for secondary structure determination in proteins based on sequence similarity[J]. Febs Letters,1986,205(2):303-308.

[19] NISHIKAWA K,OOI T. Amino acid sequence homology applied to the prediction of protein secondary structures,and joint prediction with existing methods[J]. Biochim Biophys Acta,1986,871(1): 45-54.

[20] YI T,LANDER E. Protein secondary structure prediction using nearest-neighbor methods[J]. Journal of Molecular Biology,1993,232(4): 1117-1129.

[21] ROST B,SANDER C. Prediction of protein secondary structure at better than 70% accuracy[J]. Journal of Molecular Biology,1993,232(2): 584-599.

[22] CUFF J,BARTON G. Application of multiple sequence alignment profiles to improve protein secondary structure prediction[J]. Proteins,2000,40(3):502-511.

[23] HUA S,SUN Z. A novel method of protein secondary structure prediction with high segment overlap measure: Support vector machine approach[J]. Journal of Molecular Biology,2001,308(2): 397-407.

[24] 刘倩倩. 基于词频统计编码和流形学习的蛋白质二级结构预测方法研究[D]. 天津:河北工业大学,2013.
LIU Qianqian. Research on Protein Secondary Structure Prediction Based on Word Frequency Statistics Coding and Manifold Learning[D]. Tianjin:Hebei University of Technology,2013.

[25] BYSTROFF C,THORSSON V,BAKER D. HMMS TR:A hidden Markov model for local sequence structure correlations in proteins[J]. Journal of Molecular Biology,2000, 301(1): 173-190.

[26] 陈军霞,刘紫玉. 基于Baum-Welch算法HMM模型的孤词算法研究[J]. 河北科技大学学报,2015,36(1):52-57.
CHEN Junxia,LIU Ziyu. Study on solitary word based on HMM model and Baum-Welch algorithm[J]. Journal of Hebei University of Science and Technology,2015,36(1): 52-57.

[27] AYDIN Z,ALTUNBASAK Y,BORODOVSKY M. Protein secondary structure prediction for a single-sequence using hidden semi-Markov models[J]. BMC Structural Biology,2006,7: 178-192.

向本期载文的审稿专家致谢

本期《河北科技大学学报》共发表论文15篇。这些论文的发表是与有关专家的认真审读、细查资料、推敲分析、中肯评价分不开的。对此,本编辑部特向这些专家表示敬意,对他们的辛勤劳动表示感谢。本期载文的审稿专家名单如下(按姓名的汉语拼音顺序排列):

陈国华　崔玉泉　丁志荣　董艳春　段彦栋　韩　天　贺尔铭　贺占蜀

侯新民　江少群　姜明顺　李　春　李元松　梁金玲　吕晓龙　吕银祥

潘敏强　秦志英　石　磊　孙桂全　孙晓云　王　峰　王兴松　徐玉梁

杨润生　张春会　张定国　张进华　郑小琪　周　毅

(本刊编辑部)