

文章编号:1004-9037(2013)02-0129-07

加窗窄通带滤波器蛋白质编码区预测算法

马玉韬^{1,2} 车 进^{1,2} 关 欣¹ 滕建辅¹

(1. 天津大学电子信息工程学院, 天津, 300072; 2. 宁夏大学物理电气信息学院, 银川, 750021)

摘要:改进 Gabor 小波变换(Modified Gabor wavelet transform, MGWT)蛋白质编码区预测算法给出的预测结果在当前所有的独立预测算法中准确率最高。本文提出了有限脉冲响应(Finite impulse response, FIR)加窗窄通带滤波器蛋白质编码区预测(Windowed narrow pass-band filter, WNPBF)算法。算法的主要部分包括:以 F56F11.4 序列为例给出了 WNPBF 阶数选择的依据,并据此设计 WNPBF;为消除滤波器群延迟对预测结果产生的不良影响,对信号用边界对称延拓法进行预处理并对窄带滤波器滤波输出信号进行截取;为改善预测结果,设计滑动平均滤波器平滑功率谱密度曲线。在 ALLSEQ 和 HMR195 两个 DNA 序列集上获得预测准确率分别接近或达到独立预测算法的最高水平。通过比较后发现,所提出的 WNPBF 算法较 MGWT 算法效率更高,使用该算法可以直观和客观地比较不同滤波器的预测结果。

关键词:蛋白质编码区预测;窄通带滤波器;FIR 滤波器;窗函数法;群延迟

中图分类号:TN713;Q751

文献标志码:A

Prediction Algorithm for Protein Coding Regions Based on Windowed Narrow Pass-Band Filter

Ma Yutao^{1,2}, Che Jin^{1,2}, Guan Xin¹, Teng Jianfu¹

(1. School of Electronic Information Engineering, Tianjing University, Tianjing, 300072, China;

2. School of Physics Electrical Information Engineering, Ningxia University, Yinchuan, 750021, China)

Abstract: The modified Gabor wavelet transform (MGWT) algorithm provides the best prediction results among the independent protein coding regions prediction algorithms. To be compared, a finite impulse response (FIR) windowed narrow pass-band filter (WNPBF)-based protein coding regions prediction algorithm is proposed. The algorithm is mainly consisted of the following parts; designing a WNPBF with appropriate length, which taking the DNA sequence F56F11.4 as an example; extending the DNA sequences using the boundary symmetric padding method and cutting off the beginning part of WNPBF outputs to eliminate the side effects of the group delay of the filter on prediction results; designing a moving average filter to smooth the power spectral density curve to get better prediction results. Experiments performed on DNA data sets ALLSEQ and HMR195 respectively give the prediction results which are close or equal to the best ones. The proposed algorithm is much more efficient than MGWT algorithm and can be used to compare the prediction results of different filters directly and objectively.

Key words: protein coding region prediction; narrow pass-band filter; FIR filters; window function method; group delay

基金项目:国家自然科学基金(61162020)资助项目;宁夏自然科学基金(NZ1024)资助项目;宁夏高校科学研究项目(2011263)资助项目。

收稿日期:2011-10-28;**修订日期:**2011-12-31

引 言

真核生物基因预测,特别是蛋白质编码区(又称外显子)的预测,是基因组序列信息提取和分析及功能基因组相关信息分析的重要内容,对生物大分子结构模拟和药物设计、分子进化和比较基因组学等研究有重要意义^[1]。DNA 序列中能够对蛋白质进行编码的部分,其频谱显示频率为 1/3 的信号成分较强;而不能对蛋白质进行编码的部分(内含子),其频谱一般不包含较强的周期 3 信号,这称为外显子的周期 3 性质或 3 碱基周期性(Three-base periodicity, TBP)^[2-4]。自 Tiwari 提出基于傅里叶分析的基因预测算法(其实质是滑动离散傅里叶变换(Sliding discrete Fourier transform, SDFT),故又称 SDFT 算法)以来^[3],多种数字滤波器被设计并应用到外显子的预测中来。文献[5]首先提出了无限脉冲响应(Infinite impulse response, IIR)反陷波器的设计和格型方法的实现;在文献[6]中又提出了多级滤波器的设计和实现;文献[7]设计了有限脉冲响应(Finite impulse response, FIR)频率采样窄通带滤波器(Frequency sampling narrow pass-band filter, FSNPBF)。文献[8]用改进 Gabor 小波变换算法(Modified Gabor wavelet transform, MGWT)进行编码区预测,采用非编码率作为编码区和非编码区的分类阈值,用近似相关系数(Approximate correlation, AC)作整体预测准确率测度, MGWT 算法在 ALLSEQ 和 HMR195 等序列集上得到的最好预测结果,该结果是当前独立算法在较大的 DNA 序列集上获得的最好预测结果。

尽管基于滤波器的基因预测算法得到众多研究人员的重视^[4-11],但设计和实现方法都比较简单的 FIR 滤波器在基因预测方面的研究和应用还没有得到足够的重视和研究,还存在群延迟造成预测编码区滞后、滤波器的阶数选择、预测结果的准确率测度指标选择和较大的序列集上证明算法有

效性等方面的问题。

针对上述问题,本文提出以加窗窄道带滤波器(Windowed narrow pass-band filter, WNPBF)为核心的蛋白质编码区预测算法。WNPBF 算法用滤波器输入信号进行对称延拓,对滤波器输出信号进行截取,通过这样的方法来处理 FIR 滤波器群延迟造成的预测编码区滞后问题;用移动平均滤波器对功率谱密度曲线进行平滑的方法处理预测结果准确率不高的问题;用实验分析的方法结合 F56F11.4 序列处理 FIR 滤波器阶数选择问题;选择非编码率作为分类阈值,选择 AC 作为总体准确率测度指标^[10]。在 ALLSEQ^[10]和 HMR195^[11]两个较大的序列集上, WNPBF 算法的预测结果与 MGWT 算法进行后比较后证明前者是一种能够有效预测蛋白质编码区且较 MGWT 效率更高的独立算法。使用该算法可以直观且客观地比较不同滤波器的预测结果。

1 WNPBF 算法与 MGWT 算法

图 1(a,b)分别给出了本文提出的 WNPBF 算法与 MGWT 算法的流程图。NPBF 算法主要包括以下步骤:(1)采用 Voss 法^[2]将 DNA 序列映射成数值序列(信号);(2)使用 FIR 加窗窄通带滤波器对前一步得到的数值信号进行滤波,剔除非周期 3 信号;(3)计算信号的功率谱密度(Power spectral density, PSD);(4)对 PSD 曲线进行滑动平均滤波和幅度归一化;(5)选择非编码率作为阈值对 DNA 序列进行分类,确定 DNA 序列中的编码区和非编码区,并以一种或多种预测准确率测度指标给出预测结果。

由图 1 的比较易知, WNPBF 算法吸收了 MGWT 算法预测结果简单明了、便于比较的优点(WNPBF 算法的映射和分类环节采用了与 MGWT 相同的方法),而对 MGWT 算法较复杂的多尺度 MGWT、多次 PSD 计算(PSD 计算次数和选择的尺度参数的个数相同)和投影寻优部分分别采用 FIR 窄通带滤波器滤波、单次 PSD 计算和移动

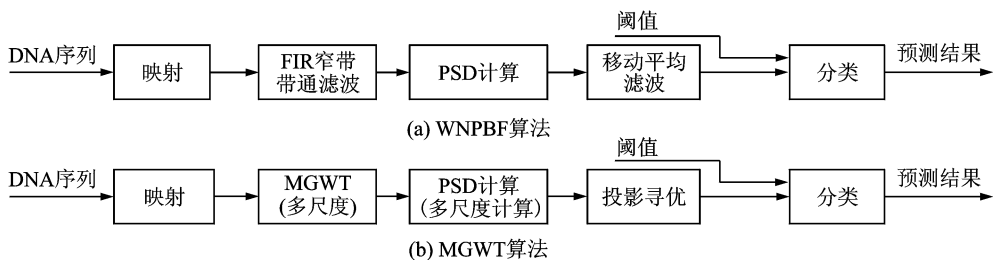


图 1 基于 FIR 窄带带通滤波器和 MGWT 的蛋白质编码区预测算法流程图

平均滤波器来代替,从而成倍地减少了仿真的运算量。

2 WNPBF 算法的实现

WNPBF 算法的实现首先介绍数据在算法中的处理流程,然后阐述 WNPBF 设计和对信号延拓截取处理两个方面的内容。

2.1 数据处理流程

本文采用广为采用的 Voss 法将由碱基 Adenine (A), Thymine (T), Cytosine (C) 和 Guanine (G) 组成的 DNA 序列映射为数值序列。该方法将一个 DNA 序列映射为与 4 种碱基相对应的 4 个数值序列,每种碱基对应的数值序列中某个时刻样本为 1 对应该时刻(碱基位置)上出现了该碱基,为 0 表示该位置没有出现该碱基^[2]。将映射得到的数值序列 $x_l[n]$, $l = \{A, T, C, G\}$ 通过 FIR 窄带带通滤波器后,得到了周期为 3 的信号 $y_l[n]$, $l = \{A, T, C, G\}$ 。DNA 序列的 PSD 计算公式^[3-9]

$$\text{PSD}[n] = \frac{1}{N} \sum_l |y_l[n]|^2 \quad (1)$$

$l = A, T, C, G; n = 1, \dots, L$

式中: N 为 FIR 滤波器长度或窗函数的窗口长度; L 为 DNA 序列的长度。

在实际编码区预测算法仿真中存在滤波输出序列 $y_l[n]$ 的曲线包含太多毛刺的问题,如果直接按照选定的阈值去区分编码区和非编码区,会使预测准确率 AC 急剧下降。因此,在统计预测结果之前先采用 1 个 110 阶的移动平均滤波器对预测 PSD 曲线进行平滑处理,以提高 AC 值。式(2)给出了 1 个 N_{ma} 阶的移动平均滤波器的差分方程

$$\text{PSD}_{\text{ma}}[n] = \frac{1}{N_{\text{ma}} + 1} \sum_{i=-N_{\text{ma}}/2}^{N_{\text{ma}}/2} \text{PSD}(n-i) \quad (2)$$

式中下标“ma”表示移动平均。

在计算出序列的移动平均功率谱后,需要根据一个阈值来对编码区和非编码区进行分类,本文采用预测非编码率(非编码碱基数目占序列全部碱基数目的百分比)作为阈值。为了能直观地比较各种算法的优劣,使得阈值取遍 1~99,以便获得算法的最好 AC 值。本文用敏感度 S_n (Sensitivity), 特异度 S_p (Specificity), 近似相关系数 AC 和相关系数 CC (Correlation coefficient) 来评估算法对编码区的识别性能,其中 AC 作为整体预测准确率的测度,便于与其他文献的研究结果进行比较; S_n , S_p 和 CC 作为参考测度。本文实验部分 AC 指标在

ALLSEQ 和 HMR 195 两个序列集上的结果是序列集中所有序列各自 AC 指标的平均值^[10]。

2.2 WNPBF 设计

基于 FIR 窄通带滤波器 (Narrow pass-band filters, NPBFs) 的编码区预测算法的核心是要通过 WNPBF 或 FSNPBF 将 DNA 序列中周期为 3 的信号保留而将其他频率的信号和噪声滤除。因此,设计的 WNPBF 要满足:(1) 通带的中心频率为 $k = N/3$, 其中 N 为 WNPBF 或 FSNPBF 单位脉冲响应的长度(窗函数窗口长度与滤波器长度相等,下文一并用“窗长”表示);(2) 能较好地实现阻带衰减和滤波器的通带宽度的折中;(3) 满足线性相位条件,特别是在通带内。

真核生物 DNA 序列中编码区长度不同且被长短不一的非编码区分开。基于 SDFIT 算法的研究表明,滑动窗长度对预测结果有较大影响。滑动窗太短,周期性不足以超越背景噪声;滑动窗太长,则仿真耗时较长且预测外显子位置的精度会降低^[6]。因此,WNPBF 的窗长选择对预测结果极为重要:一方面要考虑到满足滤波器的设计指标;另一方面,还要考虑到被滤波信号的时域长度和噪声分布等特点。以矩形窗为例,窗口越长,则主瓣越窄,选频特性越好,噪声抑制能力越强,但旁瓣衰减则越小,滤波后信号的频谱泄露越严重,时频分析的定位精度则越差。本文通过实验仿真的方法来确定窗长。

窗函数的选择可以使得设计出的窄带带通滤波器具有不同的阻带衰减 (A_{sl}) 和过渡带宽度 ($\Delta\theta$)。 $\Delta\theta$ 越窄则频谱分辨率越好,而 A_{sl} 越大则对噪声信号的抑制能力越强。Chebyshev 窗可以通过调节 A_{sl} 来改善 $\Delta\theta$ 使得二者能够实现较好的折中。汉宁窗、哈明窗等其他窗函数都不具有 Chebyshev 窗所具有的这一优势,因此本文选择 Chebyshev 窗来实现 WNPBF,阻带衰减 A_{sl} 取 100 dB。而 FSNPBF 实际上可以看成窗函数 $w_N[n]$ 为矩形窗的 WNPBF。

在确定了窗长和选择好窗函数后,可以通过对理想 NPBF 的单位脉冲响应加窗截断的方法设计 WNPBF。若理想 NPBF 的频率响应函数为

$$H_d(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h_d[n] e^{-j\omega n} \quad (3)$$

其单位脉冲响应应为

$$h_d[n] = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} H_d(e^{j\omega}) e^{j\omega n} d\omega \quad -\infty < n < +\infty \quad (4)$$

式中 ω_c 为滤波器的通带截止频率。式(4)表示的是一个非因果的双边无限系统,不满足绝对可和条件,因此也就是不可实现的。将式(4)用一个长度为 N 的窗进行截取,就得到了用窗函数法设计的 FIR 滤波器的单位脉冲响应

$$h[n] = h_d[n]\omega_N[n], 0 \leq n \leq N-1 \quad (5)$$

这相当于将 IIR 系统进行了截断处理。为了保证设计出的反陷波器具有线性相位特性,要求

$$h[n] = h[N-n-1], 0 \leq n \leq N-1 \quad (6)$$

FSNPBF^[7] 和 WNPBF 在实际设计过程中需要反复调试,以便设计得到的滤波器能满足设计要求^[12]。

2.3 延拓截取处理

采用直接型网络结构实现 FIR 滤波器时,长度为 N 的滤波器的滤波输出为

$$y[n] = \sum_{i=0}^{N-1} h[i]x[n-i] \quad (7)$$

当 $0 \leq n \leq N-2$ 时,计算滤波输出需要用到 $\{x[n]\}$ 序列 $n < 0$ 时刻的值,一般按 0 值处理,这会削弱位于 DNA 序列两端的外显子的周期 3 特性,造成不能识别或识别效果很差。文献[13,14]的研究表明,DNA 序列延拓后能提高编码区的识别率。本文采用边界对称延拓^[15]的方法,在由 DNA 序列(长度为 L)映射而来的数值序列两端分别进行以边界点或边界值为中心的对称延拓,延拓长度为 $N/2$ (延拓后序列总长度为 $L+N$),然后对延拓后的序列进行滤波。

长度为 N 的 FIR 线性相位滤波器,其群延迟

$$\tau_g = -\frac{d\theta(\omega)}{d\omega} = \frac{(N-1)}{2} \approx \frac{N}{2} \quad (8)$$

即滤波器的输出比输入序列延时约 $N/2$ 个点。因此,在考虑了滤波器群延迟和信号的边界延拓后,总长度为 $L+N$ 的输出序列需要截去前 N 个值,保留剩余的 L 个值作为有效滤波输出序列。这里截去或删除输出序列的前 N 个值,一方面消除了边界值 0 处理造成的周期 3 信号减弱;另一方面实现了将输出序列左移,从而消除了 WNPBF 的群延迟对预测结果的负面影响。

3 实验结果和讨论

3.1 实验材料

实验采用基因序列 F56511.4 是从美国国家生物信息中心网站 www.ncbi.com 下载的,其在染色体中排列的碱基号从 7021~15080,编号为

AF099922。ALLSEQ 和 HMR195 两个 DNA 序列集包含了多个物种和组织的序列,被广泛用来比较不同预测算法的有效性。

3.2 窄带带通滤波器的实现

在本文编码区预测实验中使用了长度为 120, 300 和 600 等长度的 NPNBFs,需要说明的是,对于长度较短的 DNA 序列,本文使用长度为 120 或 121 的 NPNBFs。图 2 给出长度为 600 的 FSNPNBF 和长度为 601 的 WNPBF 的幅度和相位响应图。由图 2 可以明确 WNPBF 的阻带衰减特性明显优于 FSNPNBF,而 FSNPNBF 的 3 dB 带宽则明显更优。

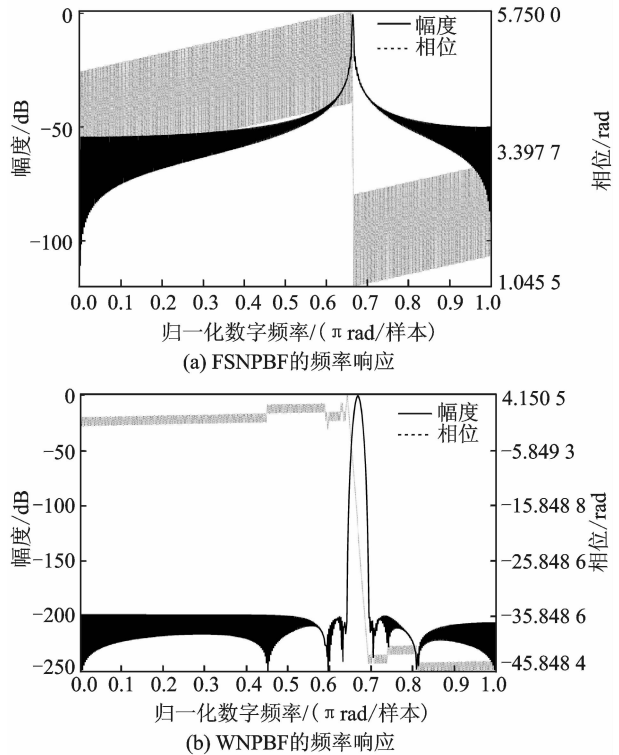


图 2 FSNPNBF 和 WNPBF 的频率响应

3.3 编码区预测实验

FIR 滤波器长度的选择如前所述既不能太短又不能太长。为研究滤波器阶数与预测准确率的关系,FSNPBF 的长度取 120~2 040,WNPBF 的长度取 121~2 041,步长都为 60。图 3 给出用上述不同长度滤波器预测序列 F56F11.4 编码区准确率得到的 AC 值与滤波器阶数关系的曲线。

由图 3 可以看出,WNPBF 在长度为 481 时得到的预测准确率最高,AC 值为 0.89;而 FSNPNBF 则在长度为 540 时得到的预测准确率最高,AC 值为 0.77。由图 3 还可以看出,采用窗函数法设计的 WNPBF 由于频谱泄露程度小,相同滤波器阶

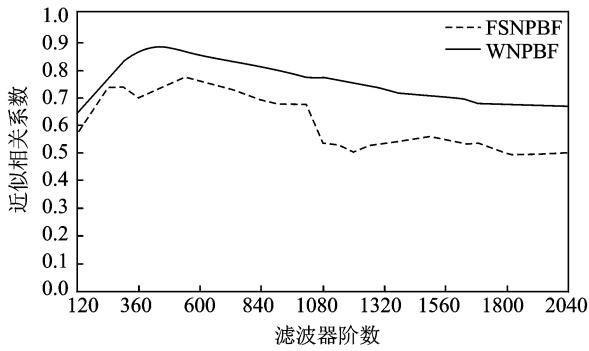


图 3 滤波器阶数与最高预测准确率 AC 的关系

数对应的所有预测结果都优于 FSNPBF。图 3 的结果还表明,预测准确率并不与滤波器阶数成正比。这可以理解为式(7)中当阶数高于理想值时输入信号的周期信号数据长度占滤波器窗口宽度的百分比降低,因而信噪比降低所致;而阶数小于理想值时,随着阶数升高,窗口内信号强度逐渐变大,信噪比逐渐升高。

考虑到将 WNPBF 算法应用到 ALLSEQ 和 HMR195 这样包含 DNA 序列较多的数据集上时,由于各个 DNA 序列编码区的长度呈多样性分布,滤波器的窗长可以在一定的范围内比较择优。本文实验中选择 300,360,420,480,540,600 和 660 这 7 个值分别进行预测,表 1,2 给出的是预测结果最好时采用的窗长。

表 1 不同算法预测编码区的最好结果

算法	DNA 序列集					
	ALLSEQ(BG570)			HMR195		
	窗长	阈值	AC	窗长	阈值	AC
MGWT ^[8]	—	74~82	0.49	—	68~83	0.44
FSNPBF	300	77~80	0.46	360	70~76	0.43
WNPBF	541 601	75~84	0.48	541 601	71~79	0.44

表 2 不同算法预测 F56F11.4 序列编码区最高准确率

算法	窗长	S_n	S_p	AC	CC	阈值
MGWT ^[8]	—	0.88	0.90	0.87	0.87	85
FSNPBF	540	0.90	0.73	0.77	0.77	81
WNPBF	481	0.92	0.88	0.89	0.89	84

实验部分还需要证实 WNPBF 编码区预测算法的有效性。这里以 F56F11.4 序列为个例,以 ALLSEQ 和 HMR195 为算法的广泛有效性测试集,通过与 FSNPBF 和 MGWT 算法预测结果的比较从延拓截取处理和平滑滤波两个方面给出实验证明。

图 4 给出了窗长为 1 200 的 FSNPBF 和窗长

为 1 201 的 WNPBF 在 F56F11.4 序列上未经延拓截取处理的预测曲线,图中群延迟造成标注编码区(表示编码区真实位置)和预测编码区彻底偏离且曲线时域分辨率很差(这里窗长取 1 200 是为了同时证明窗长过长会造成严重的频谱泄露)。图 5 给出了长度为 600 的 FSNPBF 和长度为 601 的 WNPBF,在同一序列上经过延拓截取处理的预测曲线。图 5 表明采用矩形窗的 FSNPBF 通带较窄,因而选频特性较好(通带内噪声抑制能力强),尤其是第 1 个外显子的功率谱幅度相对很明显;但阻带衰减较小频谱泄露较大,因而造成其余 4 个外显子的预测编码区较宽(时域分辨率较差)。采用 Chebyshev 窗的 WNPBF 通带宽度和阻带衰减都较大,因而频谱泄露较小且对通带外噪声抑制能力强,所以得到的预测编码区更接近标注编码区(真实编码区)。此外,由图 5(a,c)给出的未经过滑动平均滤波处理的预测曲线,可以看出由于此时曲线存在大量的毛刺,这使得预测准确率 AC 急剧下降。实验数据表明滑动平均滤波器的作用使得 AC 值在 ALLSEQ 和 HMR195 两个序列集上都提高了 0.07 左右,为了简便这里未给出未经滑动平均滤波器处理的预测结果。同时,图 5 中 F56F11.4 序列的第一个标注编码区和预测编码区能较好地重合,证明 WNPBF 算法对长度较短的外显子具有较强的预测能力。

表 1 给出了 FSNPBF, WNPBF 和 MGWT 算法在 ALLSEQ 和 HMR195 两个序列集上的预测结果。表 2 给出了上述 3 种算法在序列 F56F11.4 上的预测结果。

表 1 的数据表明,采用 WNPBF 的编码区预测算法具有明显的优势。首先, WNPBF 法获得的准

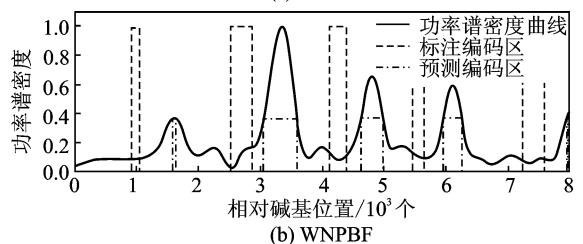
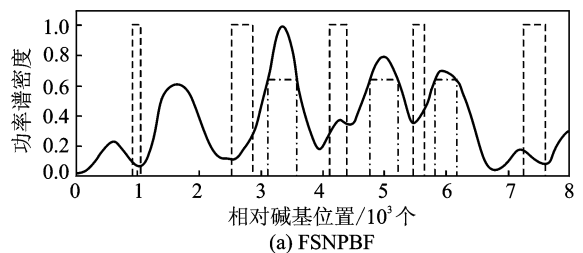


图 4 群延迟对预测结果的影响

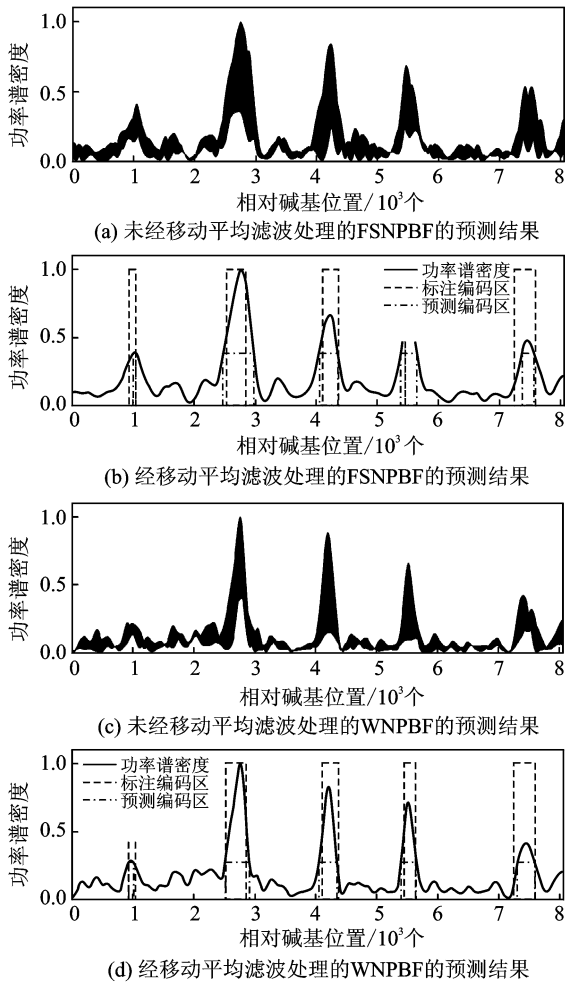


图 5 FSNPBF 和 WNPBF 对 F56F11.4 序列的预测

确率达到或接近了代表当前最高预测准确率的 MGWT 算法 (在 ALLSEQ 上的最好结果为 0.484 4);其次, WNPBF 法在不同测试集上获得最好预测准确率时采用的滤波器长度相同,而 FSNPBF 法则不同,这说明 WNPBF 法的适用性更强,也为应用该滤波器时选择阶数给出了参考值;最后, WNPBF 算法与 FSNPBF 算法相比,在序列集 ALLSEQ 上获得最高 AC 值时对应的编码率与真实编码率非常接近,而 FSNPBF 法则相去较远,由前面的分析易知这是加窗的作用。

表 2 的数据以个例的形式清楚地说明 FSNPBF 和 WNPBF 的预测性能差别。表 2 给出了这些算法得到的 S_n , S_p , AC 和 CC 指标值。这里需要说明的是, FSNPBF 的结果是在应用 WNPBF 算法后获得的。

最后,还需要说明阈值和近似相关系数(预测准确率 AC)之间的关系。图 6 给出了 FSNPBF 和 WNPBF 在 DNA 序列 F56F11.4 上获得的阈值-AC 值关系曲线。由曲线可知:在阈值低于 DNA

序列真实的非编码率时, AC 值与阈值同步上升;当阈值接近或超过真实的非编码率时, AC 值与阈值逐渐成反比关系。因此,在实际预测结果中, AC 值的最大值附近所对应的预测编码率与真实编码率最为接近。实验中在获得 DNA 序列的功率谱密度曲线后,根据指定的非编码率阈值可以确定相应的功率谱密度阈值,然后得到相应的 AC 值。在表 1 中,阈值为 84 即假设该 DNA 序列中 84% 的碱基是属于内含子序列(非编码序列)的,其余 16% 的碱基是属于外显子序列(编码序列)。在指定一个阈值后,根据 DNA 序列 PSD 值的大小,取 PSD 值较大的 16% 部分所对应的碱基序列作为预测编码序列。表中数据表明, WNPBF 算法在 HMR195 序列集上阈值取 71~79 之间获得的 AC 值在保留两位有效数字时都是 0.44。

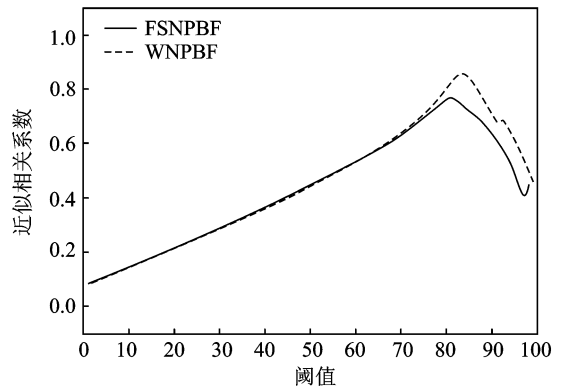


图 6 近似相关系数和阈值关系图

4 结束语

本文针对采用数字 FIR-NPBFs 为核心的蛋白质编码区独立预测算法预测准确率较低的问题,提出了 WNPBF 预测算法。在 ALLSEQ 和 HMR195 两个 DNA 序列集上获得预测准确率分别接近或达到独立预测算法的最高水平,这表明本文算法是有效的,对 WNPBF 和 MGWT 算法的分析表明前者的运算量要小很多。下一步,还可以考虑选用频谱性能更好的窗函数或滤波器设计方法,如全相位滤波器^[16]或分数阶滤波器^[17]设计方法,以获得更高的预测准确率。

参考文献:

- [1] Ahmad M, Abdullah A, Burruga K. Optimal nucleotides range estimation in diffused intron-exon noise [J]. World Applied Sciences Journal, 2010, 11(2): 178-183.
- [2] Voss R F. Evolution of long-range fractal correla-

- tions and $1/f$ noise in DNA base sequences [J]. *Physics Review Letter*, 1992, 68(25): 3805-3808.
- [3] Tiwari S, Ramachandran S, Bhattacharya A, et al. Prediction of probable genes by Fourier analysis of genomic sequences [J]. *Computer Applications in the Bioscience*, 1997, 13(3): 263-270.
- [4] Chen Bo, Ji Ping. Visualization of the protein-coding regions with a self adaptive spectral rotation approach [J]. *Nucleic Acids Research*, 2011(39):e3.
- [5] Vaidyanathan P P, Yoon B J. Digital filters for gene prediction applications [C]//IEEE Asilomar Conference on Signals, Systems and Computers. Monterey, CA: IEEE Processing Society, 2002:306-310.
- [6] Vaidyanathan P P, Yoon B J. Gene and exon prediction using allpass-based filters[C]//the IEEE Workshop Genomic Signal Processing and Statistics. Raleigh, NC:IEEE, 2002.
- [7] 马宝山,朱义胜. 一种用于基因预测的 FIR 数字滤波器[J]. *电子学报*, 2007,35(9):1710-1713.
Ma Baoshan, Zhu Yisheng. FIR digital filter for gene prediction application[J]. *ACTA Electronica Sinica*, 2007,35(9):1710-1713.
- [8] Mena-Chalco J P, Carrer H, Zana Y, et al. Identification of protein coding regions using the modified Gabor-wavelet transform [J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2008, 5(2): 198-206.
- [9] Akhtar M, Epps J, Ambikairajah E. Signal processing in sequence analysis: advances in Eukaryotic gene prediction[J]. *IEEE Journal of Selected Topics in Singal Processing*, 2008, 2(3): 310-321.
- [10] Burset M, Guigo R. Evaluation of gene structure prediction programs [J]. *Genomics*, 1996, 34: 353-367.
- [11] Rogic S, Mackworth A K, Ouellette B F. Evaluation of gene-finding programs on mammalian sequences [J]. *Genome Research*, 2001, 11(5): 817-832.
- [12] Mitra S. Digital signal processing—A computer-based approach[M]. 3rd Edition. Beijing: Tsinghua University Press,2006.
- [13] Yan Ming, Lin Zhesuai, Zhang Chunting. A new Fourier transform approach for protein coding measure based on the format of the z curve [J]. *Bioinformatics*, 1998, 14(8): 685-690.
- [14] Ma Baoshan, Zhu Yisheng, Chen Yuzhen. An improved Fourier method for DNA sequences classification [C]//3rd International Conference on Bioinformatics and Biomedical Engineering (ICBBE 2009). Beijing,China:[s. n.],2009: 1-4.
- [15] Guan Lutai. Wavelet interpolation and decomposition in a finite interval with boundary conditions [J]. *Journal of Engineering Mathematics*, 1995,12(3):1-9.
- [16] 黄翔东,王兆华. 全相位 FFT 相位测量法的抗噪性能 [J]. *数据采集与处理*,2011,26(3):286-291.
Huang Xiangdong, Wang Zhaohua. Anti-noise performance of all-phase FFT phase measuring method [J]. *Journal of Data Acquisition & Processing*,2011, 26(3):286-291.
- [17] 韦慧,孙洪广,陈文. 基于分数阶差分滤波器的边缘检测算法[J]. *数据采集与处理*,2010,25(5): 655-660.
Wei Hui, Sun Hongguang, Chen Wen. Edge detection algorithm based on fractional difference filter [J]. *Journal of Data Acquisition & Processing*,2010, 25(5):655-660.

作者简介:马玉韬(1974-),男,博士研究生,研究方向:生物信息处理、EDA 技术,E-mail:yutao_ma@163.com;车进(1973-),男,博士研究生,研究方向:电子系统设计、信号与信息处理;关欣(1977-),女,博士,副教授,研究方向:生物信息处理;滕建辅(1954-),男,教授,博士生导师,研究方向:信号处理、滤波器与计算机辅助设计。

