基于 Memetic 优化的智能 DNA 序列数据压缩算法

周家锐[1,2,3],纪 震[2,3],朱泽轩[2,3],陈思平[1,2,3]
(1.浙江大学生物医学工程与仪器科学学院,浙江杭州 310027; 2.深圳大学计算机与软件学院,广东深圳 518060;
3.深圳市嵌入式系统设计重点实验室,广东深圳 518060)

摘 要: 提出近似重复矢量(Approximate Repeat Vector,ARV)模型用于 DNA 序列冗余片段的描述.通过将数据生物信息学特征引入压缩预处理,并使用 ARV 矢量构造编码码本,提出了非对称 DNA 序列压缩算法 BioLZMA-2.算法引入基于粒子群优化的 Memetic 改进方法 CLIPSO-MA 用于压缩码本的智能优化设计,有效提升了编码性能.在标准测试序列上的实验结果表明,BioLZMA-2 可获得比现有 DNA 序列数据压缩方法更高的压缩率.

关键词: DNA 序列数据压缩; 生物信息学; 近似重复矢量; Memetic 算法
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2013)03-0513-06
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2013.03.016

Intelligent DNA Sequence Data Compression Using Memetic Algorithm

ZHOU Jia-rui[1,2,3], JI Zhen[2,3], ZHU Ze-xuan[2,3], CHEN Si-ping[1,2,3]
(1. *College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, Zhejiang* 310027, *China*;
2. *College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong* 518060, *China*;
3. *Shenzhen Key Laboratory of Embedded System Design, Shenzhen, Guangdong* 518060, *China*;)

Abstract: A data model namely the approximate repeat vector (ARV) is introduced to describe the similar fragments in DNA sequences. By employing bioinformatics features in data preprocessing, and using ARVs in compression codebook's construction, we propose an asymmetric DNA sequence compression algorithm of biological Lempel-Ziv-Markov chain algorithm 2(BioLZMA-2). The particle swarm optimization (PSO) based memetic algorithm improvement namely the comprehensive learning intelligent particle swarm optimization memetic algorithm (CLIPSO-MA) is employed in the compression codebook's design. Experimental results on benchmark sequences demonstrate better performance of BioLZMA-2 than the original DNA sequence compression algorithms.

Key words: DNA sequence data compression; bioinformatics; approximate repeat vector; memetic algorithm

1 引言

DNA 序列是生物学、医学、遗传学等重点学科的基础研究数据,具有重要科研价值.随着各项 DNA 序列测定工程展开,产生的数据量亦急剧膨胀,为现有存储传输资源带来严重压力[1].而另一方面,现有通用压缩算法无法有效处理 DNA 序列的数据特点,往往导致编码后文件体积反而有所膨胀[2].因而出现了针对 DNA 序列的专用数据压缩方法.

自 DNA 序列数据压缩问题提出以来,产生了许多实现算法.1999 年 X.Chen 等提出了 GenCompress 算法[3],有效提高了压缩性能.2000 年 T.Matsumoto 等将上下文树加权(Context Tree Weighting,CTW)与传统的 LZ 压缩相结合,提出了 CTW + LZ 算法[4].尽管其在压缩率上有所提升,但处理速度显著降低.2002 年 M.Li 等提出了 DNACompress 压缩方法[5],使用 PatternHunter 工具搜索 DNA 序列的重复片断,提高了匹配速度.2007 年 G.Korodi 等基于归一化最大似然(Normalized Maximum Likelihood,NML)模型对 DNA 序列数据进行分析,提出了 GeNML 压缩方法[6].通过对具有不同数据特点的 DNA 片断使用针对性的编码策略及概率模型进行处理,算法获得了较好的压缩效果.

尽管现有 DNA 序列数据压缩方法已取得一定进展,但随着压缩率进一步提升,所耗费的处理资源亦显

收稿日期:2010-08-11;修回日期:2012-09-11
基金项目:国家自然科学基金(No.61171125,No.60872125,No.61001185,No.61211130120);霍英东高等院校青年教师基金基础性研究课题;教育部重点研究项目;广东省自然科学基金(No.10151806001000002);深圳市杰青项目

著增加,算法性能改进已逐渐陷入瓶颈[7].为解决这一问题,我们首次将序列的生物信息学特征引入数据编码,提出了 BioLZMA 算法[8],获得了比传统 DNA 序列压缩方法更佳的处理效果.算法舍弃了部分压缩率提升以换取更快的编码速度,因而较适合于用户间对称压缩的情况.而对于诸如 DNA 序列数据库等应用环境,其序列压缩过程往往仅需进行一次,而更多是后续的解压操作,故可使用非对称处理方法,以较大的编码运算成本取得更佳的压缩效果.

针对这一情况,本文提出了改进的非对称 DNA 序列数据压缩算法 BioLZMA-2.算法在原有 BioLZMA 引入生物信息学特征用于压缩处理的基础上,首次提出了近似重复矢量(Approximate Repeat Vector,ARV)模型用于序列冗余片段的描述.通过将 DNA 序列数据包含的 ARV 片段作为编码矢量,并引入 Memetic 算法(Memetic Algorithm,MA)用于压缩码本的智能优化设计,BioLZMA-2 可有效降低目标数据冗余度,提升压缩性能.实验结果表明,BioLZMA-2 可取得比传统 DNA 序列压缩方法以及 BioLZMA 算法更高的整体压缩率.其功能设计与 BioLZMA 算法互补,共同组成了 BioLZMA 系列压缩方法.

2 DNA 序列的生物信息学特征

生物信息学(bioinformatics) 是一门利用数学、统计学、信息学及计算机科学方法研究生物学问题的交叉学科.通过将 DNA 序列数据的生物信息学特征引入压缩过程,挖掘并利用序列的生物学信息与含义,使用针对性的编码方法,将能有效提升算法性能.在 BioLZMA-2 中,使用的 DNA 序列生物信息学特征包括功能划分与片段相似性两方面.

首先,DNA 序列并非碱基符号随机出现的长字符串,而是蕴含丰富遗传信息,具有不同含义与功能划分的真实生物学数据.其各组成部分重要性不同,包含信息量亦有所差异.如图 1 所示,将 DNA 序列按其生物功能作最小归类划分,可以视为由外显子片段(Expressed Region,Exon)、内含子片段(Intervening Sequences,Intron)、RNA 序列及基因间区段交叉出现相互连接组成[9].

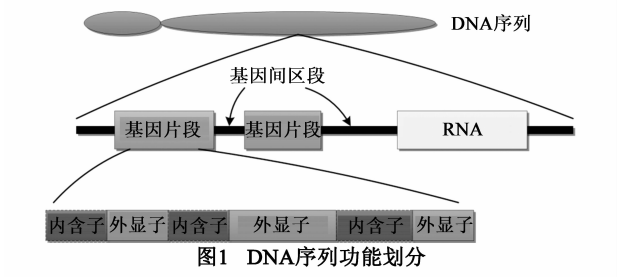


图1 DNA序列功能划分

外显子是基因内直接表达为蛋白质的部分序列,含有最为丰富的遗信息.内含子是基因内阻断编码区域(Coding Sequence,CDS)线性表达的部分序列,在翻译为氨基酸前被剪除.RNA 片段是 DNA 内用于辅助遗传信息表达的序列部分,其碱基排列最为保守,序列间相似性较高.基因间区段主要用于构成 DNA 的物质结构,通常不含遗传信息.

DNA 序列内各组成部分具有不同的数据规律,其压缩难度亦各不相同.在传统 DNA 序列数据压缩方法中,将所有序列片段视作同类数据进行一并处理,严重影响了算法性能.若使序列按功能划分进行切分重组,则可使其生物学含义明晰化,有效提升各片段集合内的数据相似度,从而获得更佳的压缩效果.

其次,DNA 序列具有高度相似性.与常见符号串数据相比,DNA 序列的相似性具有三个显著特点.

第一,DNA 序列存在着大量的重复片段.包括简单碱基重复和大规模的序列复制在内的相似部分,可占到 DNA 序列数据总量的近 50%[10].此外一些重要的基因片段也会在序列的不同位置多次重复出现.DNA 序列的这种大规模相似特点是其数据压缩算法的重要基础.

第二,序列中的重复具有多种特有模式.如图 2 所示,DNA 序列既有常见的直接重复(direct repeat)模式,亦有独特的镜像重复(mirror repeat)、配对重复(pairing repeat)和反转重复(inverted repeat)等模式.传统 DNA 序列数据压缩方法中,或未全面考虑序列的所有重复模式,或将重复片段的不同模式视作独立序列单独处理,影响了算法性能.

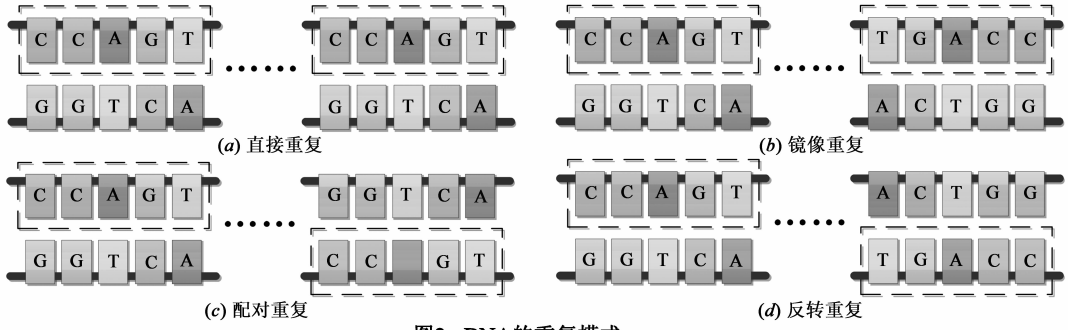


图2 DNA的重复模式

第三,DNA 序列中的重复更多地表现为近似重复形式,即可视为由各模式的精确重复片段,通过一定数量的碱基插入(insertion)、删减(deletion)和替换(substitution)操作而获得.近似片段转换为精确重复所需的符号操作被称为序列的编辑误差(edit error).

在 DNA 序列数据压缩算法中,若想充分利用序列的相似性特点,则不仅需要搜索片段的四种重复模式,还应考虑可能的近似匹配状况.为处理这一问题,我们改变了传统压缩算法中搜索无向精确子串的观点,将四种重复模式相统一,首次提出了 DNA 近似重复矢量(Approximate Repeat Vector,ARV)这一概念.

如图 3 所示,若将直接重复片段视作有向矢量 $\boldsymbol{v}$,则镜像重复可表达为逆序列 $\boldsymbol{v}^{-1}$;根据碱基互补原则,有配对重复为 $\boldsymbol{v}^{*}$,反转重复为 $\boldsymbol{v}^{-1*}$.其中符号"*"表示 A-T、C-G 碱基配对.则在压缩过程中,所有符合四种重复模式的序列片段可使用相同 ARV 模型 $\boldsymbol{v}$ 进行统一处理即可.而对于近似重复片段,需另外添加编辑误差信息.

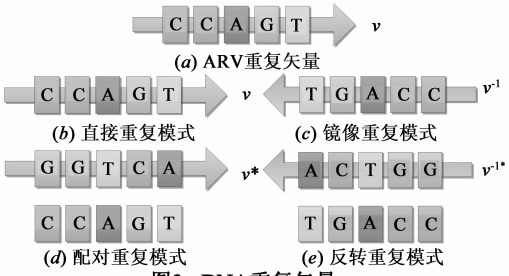


图3　DNA重复矢量

通过定义序列重复片段的 ARV 模型,算法可对 DNA 所有相似性数据特点进行统一操作,充分利用其高度冗余特性,从而提升整体压缩性能

3　基于 Memetic 优化的 DNA 压缩算法

在 BioLZMA-2 中,首先需要根据序列的功能划分对 DNA 序列数据进行切分重组,以提升其集合相似度.与 BioLZMA 算法类似,BioLZMA-2 在压缩预处理中亦将 DNA 序列归类为包含外显子片段的 CDS 集合,包含内含子片段的 Intron 集合,包含 RNA 序列的 RNA 集合,以及包含主要为基因间区段的 Others 集合四部分.

而后,算法将搜索 DNA 序列的 ARV 重复片段,并将其作为编码矢量,构造压缩码本.通过将序列重复片段替换为对应码矢量编号,并记录各近似匹配的编辑误差信息,算法即可消除冗余,完成数据压缩.码本包含 ARV 矢量在序列中的重复片段越多,涵盖范围越广,包含编辑误差越少,则编码后数据量越小,可取得更高的压缩率.因此 BioLZMA-2 中 DNA 序列数据的压缩性能问题,亦即编码码本的设计优化问题.其关系复杂,难以使用传统数学工具进行求解.

针对这一问题,算法引入基于粒子群优化的 Memetic 改进算法 CLIPSO-MA 用于压缩码本的智能设计.CLIPSO-MA 结合了综合学习粒子群优化算法(Comprehensive Learning Particle Swarm Optimizer,CLPSO)[11]的全局搜索能力,以及自适应智能单粒子优化算法(Adaptive Intelligent Single Particle Optimizer,AdpISPO)[12]的局部寻优性能.通过区分种群中具有不同进化趋势的粒子个体,并使用针对性的局部更新策略,CLIPSO-MA 可有效处理高维多模的复杂优化问题,在较短的迭代时间内获得比传统粒子群优化改进算法更佳的搜索效果,避免陷入早熟收敛.

在 BioLZMA-2 中,将压缩码本内 ARV 矢量顺序首尾连接,构造寻优粒子结构如图 4 所示:

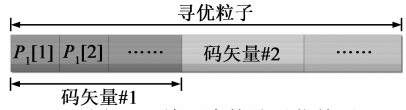


图4　压缩码本构造寻优粒子

其中 $P_i[j]$为 ARV 码矢量 i 的第 j 个碱基符号.由于 DNA 为构成符号有限的离散序列,而 CLIPSO-MA 只能处理连续的粒子位置数据,因此在搜索过程中需使用映射算法完成连续位置到离散符号的转换.在 BioLZMA-2 中,使用等距范围映射的方法.设粒子位置更新范围为[-20,20],则有映射公式:

$$\boldsymbol{P}_i[j]=\begin{cases}\text{"A" if } \boldsymbol{X}_i[j]\in[-20,-10)\\ \text{"C" if } \boldsymbol{X}_i[j]\in[-10,0)\\ \text{"C" if } \boldsymbol{X}_i[j]\in[0,-10)\\ \text{"T" if } \boldsymbol{X}_i[j]\in[10,20]\end{cases}$$

其中 $\boldsymbol{X}$ 为粒子的连续位置矢量.

码本优化过程如图 5 所示.在寻优迭代中,首先将当前种群最优位置映射为符号序列,并切分为 ARV 码矢量,构成编码码本.而后使用快速近似匹配算法 A-GREP(Approximate GREP)[13]搜索码本在目标 DNA 序列上的近似片段,包含四种重复模式,并记录其编辑误差.通过综合码本的匹配数据,算法构造优化适应度函数有:

$$\text{fitness}=\frac{1}{\text{Cover-Error}}$$

其中 Cover 为码本近似重复片段的总符号覆盖数,Error 为其中包含的编辑误差符号数.则 Cover-Error 为压缩过程实际编码的有效碱基符号数,取其倒数表示适应度函数值越小码本性能越好.

通过在 CLIPSO-MA 中使 ARV 码矢量自适应进化,智能优化挑选,BioLZMA-2 可获得编码效果更佳的压缩码本,从而提升算法性能.

DNA 序列包含 4 种碱基符号,每个未压缩碱基编码需要 2bit 数据空间(Bit Per Base,BPB).而在近似片段

中,则需要更多的数据位标识编辑误差.因此若重复片段中包含过多误差信息,其所需标识数据量将大于使片段替换为 ARV 编号时的数据压缩量,造成编码后序列的膨胀.为避免这一情况,我们提出了编码误差比(Encode Error Ratio,EER)的概念,以评估近似片段可容纳的最大误差符号数.

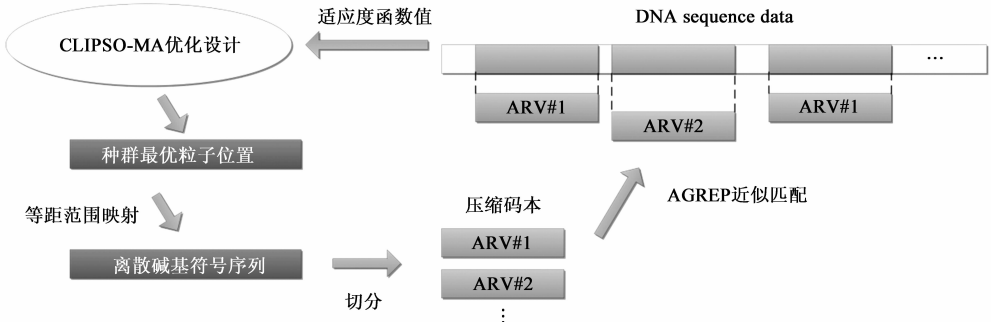


图5 使用CLIPSO-MA设计压缩码本

设每个误差碱基需要 η 位数据标识,序列中未压缩符号占用 ε 位数据,编码附加信息有 μ 位,则 EER 定义为:

$$\sigma = \frac{\eta}{\varepsilon} + \mu$$

从而长度为 M 的近似片段可容纳最大误差数 K 为:

$$K = \left\lfloor \frac{M}{\sigma} \right\rfloor = \left\lfloor \frac{M \times \varepsilon}{\eta + \varepsilon \times \mu} \right\rfloor$$

BioLZMA-2 中参照使用了 BioLZMA 算法的压缩编码格式,有误差数据位 $\eta = 8$,符号数据位 $\varepsilon = 2$,附加信息 $\mu = 1$.从而算法编码误差比为 $\sigma = 5$.则对于长度 20 的 ARV 码矢量,其最大可容纳编辑误差符号不应超过 4 个.

尽管 BioLZMA-2 在压缩阶段需进行码本设计优化,其运算复杂度较高.但这一过程在非对称处理中往往仅需进行一次.而在后续多次使用的解压运算中,则只需将编码 ARV 序号重新替换为相应符号序列即可,所消耗的处理资源相对较低.

4 仿真实验及分析

在实验中,将提出的 BioLZMA-2 算法与传统 DNA 序列数据压缩方法:BioCompress-2、GenCompress、GeNML,以及对称压缩算法 BioLZMA 作用于 11 个标准测试序列[14],以评估其处理性能.这 11 个序列文件皆来自美国 GenBank 数据库[15],包含了不同物种不同功能的 DNA 序列片段,是评估序列压缩算法通用的基准测试数据.其文件中已包含详细的生物信息学特征注释,可方便应用于 BioLZMA 系列压缩算法(BioLZMA 及 BioLZMA-2)的处理.测试序列如表 1 所示:

表 1 标准测试 DNA 序列

测试序列	长度	概要
CHMPXX	121024	地钱(Marchantia Polymorpha)叶绿素基因 DNA 序列
CHNTXX	155943	烟草(Nicotiana Tabacum)叶绿素基因 DNA 序列
HEHCMVCG	229354	人类巨细胞病毒株 AD169(Scytomegalovirus Strain AD169)完整基因 DNA 序列
HUMDYSTROP	38770	人类肌营养不良蛋白(Dystrophin)基因 DNA 序列
HUMGHCSA	66495	人类生长荷尔蒙(Growth Hormone)与绒毛膜生长催乳素(Chorionic Somatomammotropin)基因 DNA 序列
HUMHPRTB	56737	人类次黄嘌呤磷酸核糖转移酶(Hypoxanthine Phosphoribosyltransferase)基因 DNA 序列
HUMHDABCD	58864	人类三粘粒(Cosmid)毗连群 DNA 序列
HUMHBB	73308	人类第 11 号染色体中 β 球蛋白(Beta Globin)区域的 DNA 序列
MPOMTCG	186609	地钱线粒体(Mitochondrial)完整基因 DNA 序列
SCCHRIII	316613	酿酒酵母(Saccharomyces Cerevisiae)第 3 号染色体完整基因 DNA 序列
VACCG	194711	牛痘(Vaccinia)病毒完整基因 DNA 序列

BioLZMA-2 设置 ARV 码矢量长度为 20 个碱基符号,码本规模为 50.优化时 CLIPSO-MA 设置如[11]所示.实验使用各算法在测试序列上的压缩率,以及 BioLZMA-2 算法的解压速度作为评估结果.压缩率对比如下

页表 2 所示,其中 Bio2 表示 BioCompress-2 算法,Gen 表示 GenCompress 算法:

表 2　各算法在标准测试序列上的压缩率(BPB)

测试序列	Bio2	Gen	GeNML	BioLZMA	BioLZMA-2
CHMPXX	1.684	1.673	1.661	1.572	1.532
CHNTXX	1.617	1.614	1.613	1.591	1.578
HEHCMVCG	1.848	1.847	1.839	1.692	1.561
HUMDYSTROP	1.926	1.923	1.912	1.873	1.603
HUMGHCSA	1.307	1.097	1.012	1.648	1.477
HUMHPRTB	1.913	1.846	1.758	1.721	1.639
HUMHDABCD	1.882	1.821	1.713	1.866	1.702
HUMHBB	1.881	1.819	1.796	1.889	1.618
MPOMTCG	1.942	1.913	1.883	1.810	1.794
SCCHRIII	1.948	1.948	1.937	1.741	1.662
VACCG	1.764	1.763	1.763	1.539	1.482

BioLZMA-2 解压缩速度结果如表 3 所示:

表 3　BioLZMA-2 在标准测试序列上的解压缩速度(Sec.)

测试序列	解压时间	测试序列	解压时间
CHMPXX	0.453	HUMHDABCD	0.439
CHNTXX	0.442	HUMHBB	0.341
HEHCMVCG	0.397	MPOMTCG	0.385
HUMDYSTROP	0.343	SCCHRIII	0.504
HUMGHCSA	0.320	VACCG	0.493
HUMHPRTB	0.418		

由表 2 结果可发现,在编码过程中使用了序列生物信息学特征的 BioLZMA 系列算法,其在大部分测试数据上的压缩率都优于传统 DNA 序列数据压缩方法.特别是对于功能划分明确,生物学含义清晰的目标序列,其压缩率提升更为明显.例如序列 VACCG,其主要成分为具有清晰生物学意义的可编码区域(占数据总量 88%),则 BioLZMA 系列算法在其上的处理效果明显优于其他压缩方法.

对比 BioLZMA-2 与其他 DNA 序列数据压缩方法的实验结果则可看出,通过引入 ARV 模型用于序列近似重复片段的描述,并使用高效的 CLIPSO-MA 智能设计压缩码本,算法有效提升了编码性能.而 BioLZMA-2 设计为非对称算法,其压缩过程使用了充分的计算资源以获取最佳编码码本,获得了比其他 DNA 序列数据压缩方法更佳的序列压缩率.

而由表 3 数据可发现,非对称的 BioLZMA-2 算法在解压过程中运算复杂度较低,可快速完成编码序列的恢复操作,方便在实际应用时多次进行的解压缩过程.

实验结果表明,通过引入非对称算法设计,BioLZMA-2 在 DNA 序列数据压缩阶段可获得比现有方法更高的序列压缩率,而在解压阶段则可快速完成压缩片段的恢复.其性能高效,适用于非对称的 DNA 序列数据压缩应用环境,与对称压缩 BioLZMA 形成功能互补.

5　结论

本文提出了一种基于生物信息学特征及 Memetic 优化的非对称 DNA 序列数据压缩算法 BioLZMA-2.通过引入近似重复矢量 ARV 用于序列冗余片段的描述,并使用 CLIPSO-MA 用于编码码本的智能优化设计,算法可获得比现有 DNA 序列数据压缩方法更高的序列压缩率.而在解压时,BioLZMA-2 只需进行简单的片段替换即可完成序列恢复,其运算复杂度较低,可在较短处理时间内完成.BioLZMA-2 具有良好的非对称处理性能,与原有对称压缩算法 BioLZMA 构成功能互补,共同组成高效的 BioLZMA 系列 DNA 序列数据压缩算法.

参考文献

[1] Galperin M Y, Cochrane G R. Petabyte-scale innovations at the European nucleotide archive[J]. Nucleic Acids Research, 2009, 37: D1 – D4.

[2] Srinivasa K G, Jagadish M, et al. Efficient compression of non-repetitive DNA sequences using dynamic programming[A]. Proceeding of International Conference on Advanced Computing and Communications[C]. Mangalore: ADCOM, 2006. 569 – 574.

[3] Chen X, Kwong S, et al. A compression algorithm for DNA sequences and its applications in genome comparison[A]. Proceeding of the 10th Workshop on Genome Informatics[C]. Tokyo: GIW, 1999. 51 – 61.

[4] Matsumoto T, Sadakane K, et al. Biological sequence compression algorithms[A]. Proceeding of Genome Informatics Workshop[C]. Tokyo: CIW, 2000. 43 – 52.

[5] Chen X, Li M, et al. DNACompress: fast and effective DNA sequence compression[J]. Bioinformatics, 2002, 18(12): 1696 – 1698.

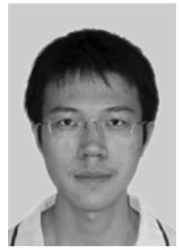
[6] Korodi G, Tabus I. An efficient normalized maximum likelihood algorithm for DNA sequence compression[J]. ACM Transactions on Information Systems, 2005, 23(1): 3 – 34.

[7] 林毅申,林丕源,等.基于字典的 DNA 序列压缩算法研究及应用[J].计算机应用研究,2007,24(6):265 – 267.
Lin Y S, Lin P Y, et al. Research and mplementation of dictionary-based DNA compression algorithm[J]. Application Research of Computers, 2007, 24(6): 265 – 267. (in Chinese)

[8] 纪震,周家锐,等.基于生物信息学特征的 DNA 序列数据压缩算法[J].电子学报,2011,39(5):991 – 995.
Ji Z, Zhou J R, et al. Bioinformatics features based DNA sequence data compression algorithm[J]. Acta Electronica Sini-

ca,2011,39(5):991 – 995.(in Chinese)
[9] 王玉,饶妮妮,等.基于小波变换技术预测 DNA 序列的编码区[J].电子学报,2007,35(1):141 – 144.
Wang Y, Rao N N, et al. Predicting protein coding regions of DNA sequences based on wavelet translation technique[J]. Acta Electronica Sinica, 2007, 35(1): 141 – 144. (in Chinese)
[10] Nordin A, Yazid M, et al. A guided dynamic programming approach for searching a set of similar DNA sequences[A]. Proceeding of International Conference on the Applications of Digital Information and Web Technologies[C]. London: IEEE, 2009. 512 – 517.
[11] Liang J J, Qin A K. Comprehensive learning particle swarm optimizer for global optimization of multimodal functions[J]. IEEE Transactions on Evolutionary Computation, 2006, 10(3): 281 – 295.
[12] 周家锐,纪震,等.基于自适应智能单粒子优化算法的 Gabor 人脸识别方法[A].全国模式识别学术会议[C].重庆:CCPR,2010.1 – 5.
Zhou J R, Ji Z, et al. Face recognition using Gabor wavelet and self-adaptive intelligent single particle optimizer[A]. Proceeding of Chinese Conference on Pattern Recognition[C]. Chongqing: CCPR, 2010. 1 – 5. (in Chinese)
[13] Wu S, Manber U. Fast text searching: allowing errors[J]. Communications of the ACM, 1992, 35(10): 83 – 91.
[14] Osborne M. Predicting DNA Sequences Using a Backoff Language Model[DB/OL]. http://www.cogsci.ed.ac.uk/~osborne/dna-backoff.ps.gz, 2009 – 05 – 15.
[15] Benson D A, Karsch-Mizrachi I, et al. GenBank[J]. Nucleic Acids Research, 2008, 36: D25 – D30.

作者简介



周家锐 男,1984 年 7 月生于广东省韶关市,2010 年获深圳大学模式识别与智能系统硕士学位,现为浙江大学生物医学工程与仪器科学学院博士研究生.主要研究方向包括计算智能、生物信息学等.
E-mail: jrzhou@zju.edu.cn

纪 震 男,1973 年 8 月生于江苏省溧阳市,1999 年毕业于西安交通大学,博士学位,2004 年晋升为教授,曾赴英国利物浦大学任访问学者.主要研究方向包括智能信号处理、嵌入式系统、生物医学工程.
E-mail: jizhen@szu.edu.cn