

# 多参数测井地层划分的遗传最优分割法 及其在济阳拗陷中生界中的应用

张振飞<sup>1</sup>, 谢忠怀<sup>2</sup>, 李宝利<sup>2</sup>, 张守鹏<sup>2</sup>

1. 中国地质大学资源学院, 湖北武汉 430074

2. 胜利油田有限公司地质研究院, 山东东营 257000

**摘要:** 将遗传算法、Fisher 最优分割和边缘检测算法相结合, 设计了一种适用于多参数测井地层划分的遗传最优分割算法。该算法以测井测点号为基因, 以有序测点号组为染色体, 由测点的边缘隶属度和段间/段内离差总和的比值共同构成综合优化指标(适应度), 采用锦标赛选择机制, 通过 Monte Carlo 抽样进行交叉(通过两两染色体比较设定抽样区间), 采用灾变变异和精英保持策略, 利用多参数测井数据求取最优化地层分界线。在济阳拗陷中生界地层划分中的实际应用, 表明该方法的自动分层结果符合地质实际, 且计算速度很快。该方法对于没有露头、岩心稀少、地震勘探分辨率低和地球物理测井数据量大情况下的深部隐伏地层划分有很好的适用性。

**关键词:** 遗传算法; 最优分割; 隐伏地层; 济阳拗陷。

中图分类号: P628

文章编号: 1000-2383(2009)04-0682-09

收稿日期: 2009-03-16

## A Genetic Algorithm for Optimal Stratigraphic Division Using Multi-Parameter Log Data with Application to Mesozoic Strata in Jiyang Depression, Shandong

ZHANG Zhen-fei<sup>1</sup>, XIE Zhong-huai<sup>2</sup>, LI Bao-li<sup>2</sup>, ZHANG Shou-peng<sup>2</sup>

1. Faculty of Earth Resources, China University of Geosciences, Wuhan 430074, China

2. Institute of Geological Research, Shengli Oil Field Ltd., Dongying 257000, China

**Abstract:** In this paper we propose a genetic algorithm for optimal stratigraphical division using multi-parameter well-log data. The algorithm incorporates the Fisher optimal division and edge detection methods into genetic searching in order to find out rapidly from a well a group of division boundaries that satisfy an optimization criterion. The optimization criterion involves the following steps: After a well (or a segment of it) is divided into a group of subsections, the total in-subsection deviation of the multi-parameter log data is minimized. Then the total inter-subsection deviation of the data is maximized, and the boundary points are to the largest extent singular points in log curves. This criterion is used to formulate the fitness function in our genetic algorithm. In the algorithm, logging measurement points are taken as genes, and groups of ordered measurement points as chromosomes; Genetic strategies of tournament selection, catastrophic mutation, Monte Carlo sampling crossover, and elitist reservation are adopted. We test the method on a geologically well-studied well which penetrates the deep-buried Mesozoic strata in Jiyang depression, Shandong, China. It is shown that, using log data alone, our method obtains stratigraphic division results that are geologically sound. Meanwhile the method proves far faster than an ordinary Fisher optimal division method in computation. We conclude that our method is highly applicable in stratigraphic division studies in areas such as Jiyang depression, where the interested strata is deep-buried, with scarce relevant well cores, poor seismic resolution, and abundant log data.

**Key words:** genetic algorithm; optimal stratigraphic division; buried strata; Jiyang depression.

地球物理测井数据作为一种重要信息源,在地层学、油气勘查、海平面升降、区域构造等众多领域的研究中起到巨大作用(Andrew, 1995). 典型实例如 Wonik(2001)利用伽马测井资料,在德国 Hannover 地区进行了地层对比和沉积速率分析; Shaaban and Ghoneimi(2001)利用利比亚西北部 Sirt 盆地的测井结合地震资料,探讨了该区白垩纪—古新世油藏构造特征; Maulin and George(2003)提出了利用测井数据建立二维地层——构造模型的方法; Pinous *et al.* (1999)以测井数据和岩石学数据为基础,研究了西西伯利亚南东部 Nyurolskaya 洼陷中上侏罗统海面升降; Inoue *et al.* (2003)综合利用测井和地震数据,进行了日本西南部 Osaka 盆地第四系地层对比. 以上这些成果说明,在直观观测受到限制情况下,测井数据可以为地质研究提供宝贵信息. 这种情况正是济阳拗陷中生界地质研究中所遇到的.

济阳拗陷位于山东省北部黄河下游冲积平原和入海三角洲地带,属渤海湾含油气盆地,总体 NE 向延长,面积约 29 000 km<sup>2</sup>. 济阳拗陷属中、新生代内陆沉积盆地,其中从第三系到上古生界均有储层分布,是中国第二个石油生产基地(操应长等, 1996; 邓宏文等, 1997; 王建伟等, 2001; 曹忠祥等, 2002; 吕希学等, 2003). 至 2007 年,济阳拗陷中遇到中生界的井约有 700 口,其中近 1/4 的井有中生界油气显示. 对该区中生界地层前人已进行了大量研究(山东省地质矿产局, 1991; 王毅等, 1994; 李洪革等, 1999; 张守鹏等, 2003). 通过这些研究,对该区中生界取得如下基本认识: (1)中生界中下部以陆相碎屑沉积为主,上部以火山岩为主. 按岩性组合,从下到上分为 4 组,即坊子组(J<sub>1f</sub>)、三台组(J<sub>2s</sub>)、蒙阴组(K<sub>1m</sub>)和西洼组(K<sub>1x</sub>). 下部主要为断陷湖盆相沉积环境,中上部主要为河流相; (2)根据地震和钻井资料较好地限定了中生界的顶底界线. 第三系与中生界之间,以及中生界与古生界或基底之间,地震反射界面普遍较清楚. 第三系和古生界或老结晶基底的地震波形特征明显,因此对中生界的总体空间分布已基本掌握; (3)中生界含油气藏的特点是油藏类型多、变化大(类型包括碎屑岩油藏、煌斑岩油藏、砂砾岩油藏、火山岩油藏、潜山顶部风化壳油藏、潜山内幕砂岩油气藏等),表现了良好的含油气远景.

由于地质情况复杂,该区中生界研究中还存在不少问题,其中中生界内部地层划分、对比是最主要的问题之一. 解决该问题的主要困难有两方面: (1)传统地质观测方法受到很大限制,中生界是埋深较

大的隐伏地层,一般出现在地表以下 2 000 ~ 5 000 m, 钻井取心十分有限,古生物资料稀少,除个别井外绝大多数地段为“哑地层”. 穿过中生界的钻孔数虽多,但对中生界主要为兼探; (2)地震方法失去了优势. 中生界厚度不大(一般为数百米),但岩性变化大,反映了复杂多变的沉积环境和构造变动. 而地震分辨率在深部变低,中生界内地震反射界面不清楚,同相轴分布规律性不强.

以上原因使得济阳拗陷中生界的研究程度较低,各凹陷及凸起尚未建立标准地层剖面,中生界内各地层单位的发育情况和空间分布尚不清楚. 这对该区中生界油气勘探造成了消极影响,成为亟待解决的问题. 经长时间积累,该区测井和录井数据已十分浩繁,如何综合利用这些数据提高对中生界的认识,是一个有重要现实意义的研究课题.

本文探讨了综合利用多参数测井数据进行隐伏地层划分的方法,并应用于济阳拗陷中生界的地层划分中. 综合利用多参数测井数据进行地层划分,以前用得较多的是 Fisher 最优分割法或其某种改进(张光前和李继英, 1991; 李洪奇等, 2002; 张守鹏等, 2003). 该方法通过极大化段间离差总和与段内离差总和的比值,在测井数据序列中寻找最优分界点,从而达到地层划分的目的. 该方法易于理解,其思路被广泛接受,但有两个特点限制了它在多参数测井数据分析中的实用性. 一是只基于离差进行优化,难于检测测井曲线的跳跃或突变点,而这些突变点往往是地层岩性变化处,是可能的地层界线. 因此,该方法的划分结果与地质实际不符的情况并不少见. 二是计算复杂度过高. 例如,若有 8 000 个数据点(1 000 m 井深范围), 3 个测井参数,拟分为 8 层,则计算离差平方和的次数为  $3 \times 8 000^2 / 2 = 9.6 \times 10^7$  次,而每次计算段内离差平方和时又需遍历段内数据点 2 次(段内平均数据点数约为 1 000). 这种计算复杂度使该方法仅适用于数据量较小、分段数不多的情况. 本文设计了一种遗传最优分割法,用于多参数测井地层划分. 该方法借鉴了 Fisher 最优分割和遗传算法的基本思路,同时结合边缘检测算法,不仅提高了地层划分的可靠性,也极大地提高了计算速度.

## 1 遗传最优分割法原理和计算过程

### 1.1 原理

遗传算法是模仿生物界“优胜劣汰,适者生存”

的进化过程,通过某种随机搜索而寻优的算法,适合于特别复杂甚至没有解析表达式的优化问题求解(Holland, 1975; Goldberg, 1989). 遗传算法在许多领域得到广泛应用,但在测井地层划分方面尚未见报道. 遗传最优分割法是遗传算法在测井地层划分中的具体应用. 在基本思路上,遗传最优分割法与遗传算法一样,即在一个可能的解空间中,在一定规则的约束下搜索最优解. 但是由于所研究问题的特殊性,遗传最优分割法在解的形式、搜索规则、计算过程等方面有其特殊性. 由于遗传最优分割法只是一种特殊的遗传算法,故这里仍沿用普通遗传算法中的术语,但这些术语将被赋予与多参数测井地层划分有关的具体含义.

设有一口井或一个井段,其多参数测井数据用矩阵  $X=[x_{ij}]$  表示,其中  $i=1, 2, \dots, N$  表示测点序号,  $j=1, 2, \dots, p$  表示测井参数序号. 矩阵  $X$  一般包含多个测井序列(即多参数,  $p>1$ ), 但为简便,我们在下面的叙述中也将  $X$  简称为数据序列;同时根据习惯也将多参数测井数据的值统称为电性. 多参数测井地层划分,就是利用数据  $X$  将井段划分为  $G>1$  个段,各段互不重叠且相邻段间没有空隙,各段分别对应不同的地层单位. 这种划分要求多参数测井数据在统计特征上、在所划分出的各段内尽量均匀,而段间差异尽量大,这是因为不同地层单位有不同的岩石组合,而岩性与电性有一定程度的对应关系. 这种划分也要求所找到的分界点尽量是电性突变(或跳跃)点,因为不同地层之间的地质界线往往是岩性突变处或某种岩性较特殊的标志层. 以上两方面的要求,是优化准则设计的基础. 给定分段数  $G$ ,可以在测井数据序列中任取一组  $G-1$  个不相邻的测点号,从小到大排列,每个测点号作为一个试探性的分界点号,  $G-1$  个测点号构成一个试探性的地层划分方案. 一个由不相邻测点号序列构成的划分方案称为一个染色体(或称个体),也就是多参数测井地层划分问题的一个候选解;构成染色体的每个测点号称为基因. 因此可以说,一个染色体是一组基因的有序组合,这对应于一个地层划分方案是一组不相邻测点号的有序组合. 在遗传算法中,基因的方式称为基因编码. 比如,通常可用一个一位二进制数 0 或 1 作为一个基因,一串 0 或 1 的任意组合构成一个染色体,这样一个由二进制位串构成的染色体可代表一个或多个数量值. 这种编码方法称二进制编码. 对基因进行编码是普通遗传算法中计算技术的需要,如果染色体是数量,则二进制编码可以提

高在染色体之间进行某些计算的效率. 在遗传最优分割法中,一个基因是一个测点号,它是空间位置的标识,有名称的意义并可比较大小(排列顺序),但相互之间的其他运算如加、乘等一般都没有意义(但两个基因的差值有意义,代表两测点之间的距离). 因此,在遗传最优分割法中基因编码就是基因本身,称为整数编码. 在整个数据序列  $X$  中,可以获得许多(比如  $P$  个)不同的染色体,  $P>1$  个不同染色体的集合称为一个种群或称群体,其中  $P$  称为群体规模. 可以通过“纯”随机取点的方法从  $X$  中获得一个染色体,这一过程称为染色体的初始化,所获得的染色体称为初始染色体. 同样可以通过“纯”随机取点的方法从  $X$  中获得一个群体,这种过程称为群体的初始化,得到初始群体(这里“纯”加引号是为了表示人为控制的随机性,它不是严格随机). 一个群体的全部染色体中包含的互不相同的基因总数称为该群体的基因多样性. 在一个群体中,可以在某种优化规则控制下,对每个染色体或在不同染色体之间,实施一些运算,修改其基因组合,从而使每个染色体变为一个新的染色体,原来的群体变成一个新的群体,这个过程称为群体的进化. 通过进化产生的新群体也称为子代(或新一代)群体,有时也称为一个新的世代;相应地一次进化之前的原群体称为父代群体. 在进化过程中,群体规模应保持不变(这完全是为了计算方便,因为在计算机中,群体将存储为数组或类似于数组的对象集合,进化计算中反复操作一个大小不变的数组总是比操作大小变动的数组要方便). 进化过程中所施加的各种运算操作统称为遗传算子. 在遗传最优分割法中,遗传算子包括选择、交叉、变异和精英保持 4 种运算,我们将在第 1.2 节叙述. 进化过程是染色体的“优胜劣汰”的过程,实际上就是最优解的搜索过程. 整个处理井段中全部可能的染色体都是多参数测井地层划分问题的可能解. 因此,我们也将全部可能染色体的集合称为可能解空间,而将可能解空间的任何一个子集称为一个搜索空间(显然,可能解空间是最大的搜索空间. 一个群体代表一个搜索空间,该搜索空间的大小取决于该群体的基因多样性). 通过进化,促使部分或全部染色体向提高适应度的方向变化. 适应度是指一个染色体符合预先定义的最优化准则的程度,也可以说适应度就是优化指标的值. 适应度因染色体不同而不同,因此它是染色体的函数,从而也是染色体中的基因的函数,称为适应度函数. 针对多参数测井地层划分问题,适应度函数定义为:

$$F = \frac{1}{2} \left( 1 - \frac{s}{s_t} \right) + \frac{1}{2} \sum_j \sum_k e_{jk} \quad (1)$$

式中,  $F$  表示适应度,  $j$  为测井参数号,  $k$  为测点号. 注意这里  $k$  代表染色体中的基因, 是包含于一个染色体中的测点号. 对于一个染色体而言, 式(1)中的  $k$  只可能取  $G-1$  个不同的值, 都介于 1 和  $N$  之间 (不含 1 和  $N$ , 因为 1 和  $N$  是所研究井段的端点, 不可能是要寻找的地层分界点). 式(1)中  $e_{jk}$  为第  $j$  测井参数第  $k$  点的边缘隶属度 (指属于突变点的程度, 见后述),  $s$  为段内离差指标的总和,  $s_t$  是整个处理井段的离差指标的总和, 段内离差在总离差中的比例  $s/s_t$  称为 Fisher 比. 其中  $s$  及  $s_t$  可由以下公式计算:

$$s = \sum_{g=1}^G \frac{1}{p} \sum_{j=1}^p \frac{1}{N_g} \sum_{m=1}^{N_g} |x_{gjm} - \bar{x}_{gj}| \quad (2)$$

$$s_t = \frac{1}{pN} \sum_{j=1}^p \sum_{i=1}^N |x_{ji} - \bar{x}_j| \quad (3)$$

上两式中,  $G$  为分段数,  $p$  为测井参数数,  $N$  为整个处理井段的测点总数,  $N_g$  为第  $g$  段的测点总数,  $x_{gjm}$  为第  $g$  段内第  $j$  测井参数第  $m$  点的数据 (注意这里  $m$  是在每个段内从 1 到  $N_g$  的测点编号),  $\bar{x}_{gj}$  是第  $g$  段内第  $j$  参数的平均值,  $x_{ji}$  是第  $j$  测井参数第  $i$  点的数据 (这里  $i$  是在整个处理井段内从 1 到  $N$  的测点编号).

式(1)规定了染色体适应度由两部分构成, 一是其基因(测点)的边缘隶属度, 二是应用该染色体进行地层划分后得到的 Fisher 比. 式(1)的含义是, 如果一组分界点各点属于突变点的程度较高, 而且所获得的段内离差指标值较小, 则该染色体较优; 反之, 如果一个染色体中有些点不是突变点, 或由该染色体划分的数据段的段内离差指标值在总离差中占有比例较大, 则该染色体较差.

在式(1)中, 边缘隶属度  $e_{jk}$  对于给定的测井数据序列是固定的, 可通过边缘检测算法预先算出. 边缘检测是信号处理中常用的滤波方法, 用于寻找数据序列中的突变点. 计算时, 一般先设计一个滑动窗口函数, 通过窗口与数据序列的卷积, 找到数据序列的突变点 (G roch, 1982). 本文中, 通过对比某一指标在窗口两半的差异, 在测井曲线上寻找突变点. 以测点的边缘隶属度为检测目标值, 逐点计算可获得各参数的边缘隶属度曲线. 检测指标设计如下. 设滑动窗口包含  $2n+1$  个测点, 窗口的中点为  $k$  ( $1 < k < N$ ), 则  $k-n$  为窗口的始点,  $k+n$  为窗口的终点. 在第  $k$  点上, 第  $j$  参数的边缘隶属度  $e_{jk}$  是由绝对差信噪比  $r_{jk}$  通过规格化得到.  $r_{jk}$  及  $e_{jk}$  计算公式为:

$$r_{jk} = \frac{2n+1}{n} \frac{\left| \sum_{i=k-n}^{k-1} x_{ji} - \sum_{i=k+1}^{k+n} x_{ji} \right|}{\sum_{i=k-n}^{k+n} x_{ji}} \quad (4)$$

$$e_{jk} = \frac{r_{jk} - r_{jmin}}{r_{jmax} - r_{jmin}} \quad (5)$$

以上两式中,  $x_{ji}$  为第  $j$  参数第  $i$  点的数据,  $r_{jmax}$  和  $r_{jmin}$  分别为第  $j$  参数绝对差信噪比的最大值和最小值.  $r_{jmax}$  和  $r_{jmin}$  在每个点  $k$  上的  $r_{jk}$  都算出以后才能获得. 式(4)的含义是窗口两半平均值的差与整个窗口内平均值的比值. 式(5)的作用是对所算出的  $r_{jk}$  进行规格化得到  $e_{jk}$ . 通过规格化,  $e_{jk}$  值被限于 0 和 1 之间, 从而可当作隶属度. 上述计算中, 应设定滑动窗口的中点不到达数据序列两端 (即第一个点或最后一个点), 因为两端的数据点不可能是要找的地层内部分界点. 这样边界效应可以不考虑, 因为我们只关心窗口两半各自平均值的差, 即使窗口两半包含的数据点数不同, 只要都不少于 1 个点仍能计算.

用上述各式(1~5)可算出群体中每个染色体的适应度, 即每个候选解的优化程度. 随着群体的逐代进化, 群体中将出现适应度不断提高的染色体. 如果群体中的最高适应度达到或超过某个预先指定的值 (该值称为满意适应度), 则该最高适应度对应的染色体就是要寻找的最优解或满意解. 由于式(1)的定义只规定了适应度越大越好, 并没有规定其上限, 所以不必设定一个满意适应度, 而只需以最大进化代数或最长搜索时间为终止条件, 或当群体最高适应度长时间不变时, 即可认为已找到最优解, 停止进化过程.

### 1.2 计算过程

遗传最优分割法的实现步骤与一般遗传算法类似, 但由于研究对象的特殊性, 在一些主要计算步骤中有其独特之处. 计算过程分为以下 6 步:

(1)准备. 包括计算任务设置、数据预处理和算法设置. 计算任务设置包括拟处理的井 (或井段) 选择、测井参数选择、分割层数、边缘检测窗口大小等的设定. 对测井参数应根据数据可用性情况进行选择. 由于电性与岩性之间关系复杂, 一种测井参数往往不能提供充分的分层信息, 因此一般应综合利用多种测井参数. 设定的分割层数应适当多于实际需要的分层数, 这是因为算法本身只负责寻找相邻井段电性 (岩性) 差异较大的分界点, 而不能区别这些分界点是地质上的组、段还是层界线. 因此, 应考虑寻找多个可能分界点, 然后综合考虑其他因素 (如相

邻井的对比), 从中选择最可能分界点. 边缘检测窗口大小没有严格的标准, 窗口过小可能使过多的测点具有高的边缘隶属度, 窗口过大可能使各点的边缘隶属度趋于均匀, 两者都不利于分界点的筛选, 从而消极影响群体的进化. 试算表明, 取窗口大小为数十个测点, 对应于数  $m$  的井深范围 (通常测点间距为  $1/8 m$ ) 比较适中, 而且在这个较适当范围内窗口大小不同在计算效果方面没有明显的差别.

数据预处理的目的是主要是消除各测井参数的量纲, 以使不同测井参数之间有可比性. 为了消除量纲, 可以用下面的均匀化变换:

$$x'_{ji} = \frac{x_{ji}}{\bar{x}_j} \quad (6)$$

式(6)中  $x_{ji}$  表示第  $j$  测井参数第  $i$  测点上的原始数据,  $\bar{x}_j$  是第  $j$  测井参数在整个处理井段上的平均值;  $x'_{ji}$  表示变换后的新数据, 是遗传最优分割法的输入数据. 前述公式(1)~(5)都应使用预处理后的数据.

算法设置是设定多种控制参数, 包括群体规模  $P$ 、终止条件(最大世代数、最长搜索时间)、最大变异概率等. 关于最大变异概率, 见后面关于变异的说明.

(2)生成初始群体. 用所选井深度范围内最小与最大测点号之间的随机整数作为基因, 由  $G-1$  个基因从小到大排序, 构成一个染色体( $G$  为拟分割段数), 共生成  $P$  个不同的初始染色体.

(3)计算适应度. 用式(1)计算各染色体的  $F$  值.

(4)终止判断. 当搜索时间或世代数达到预设值时, 进化结束. 从最后的群体中找到适应度最高的染色体, 即为所求的最优化划分方案. 若未达到终止条件则进行第(5)步.

(5)生成新群体. 通过以下 4 种运算(即前面提到的 4 个遗传算子)生成新群体: ①选择. 是指从群体中选出一部分较好(即适应度较大)的染色体. 选择运算直接体现了“优胜劣汰”的思想. 入选的染色体将参与下述交叉运算而生成新一代群体, 它们的基因得到遗传; 而未入选的染色体将被淘汰. 在普通遗传算法中, 选择的方法主要有两种, 一种是直接从群体中选出一部分较好染色体(称锦标赛选择); 另一种是将适应度换算为概率, 再按概率选择(称轮盘赌选择). 试算表明, 这两种选择机制差别不大, 本文采用前者, 直接从群体中选择  $P/4$  个较好染色体. 入选染色体的多少对计算效率有较大影响, 过少时将造成子代群体基因多样性不足, 限制搜索空间, 从

而进化慢; 过多时受群体规模限制, 每个父代染色体只能生成少数后代, 采样密度不足(见下面关于交叉的说明), 可能使父代的优良基因不能充分遗传, 故进化也慢. 入选染色体数占群体规模  $1/4 \sim 1/2$  时, 计算效率较高. ②交叉. 是指由两个父代染色体生成多个子代染色体的运算. 在普通遗传算法中, 交叉运算是在两两染色体之间进行基因交换(Goldberg, 1989). 在多参数测井地层划分中, 以有序测点号组为染色体, 故通常的交叉算法不再可行. 本文采用 Monte Carlo 抽样方法获取新基因, 其中抽样区间是以两两染色体的比较为基础设定的. 具体方法为, 从选出的  $P/4$  个染色体中随机取出 2 个, 记为  $A$  和  $B$ , 其适应度分别为  $F_A$  和  $F_B$ , 其基因分别为  $a_i$  和  $b_i$  ( $i=1, 2, \dots, G-1$ ). 若  $F_A > F_B$ , 以  $a_i$  为中心设定抽样区间  $[a_i - n_i, a_i + n_i]$ ; 若  $F_A < F_B$  则以  $b_i$  为中心设定区间  $[b_i - n_i, b_i + n_i]$ . 其中区间的大小  $n_i$  (整数)是两染色体对应基因的差值. 图 1 是这种算法的直观表示, 其中竖直粗线段表示抽样区间, 倾斜折线表示染色体中的基因已经排序. 假设由上步选择出的父代染色体数为  $S$  个, 则本步交叉运算中每对父代染色体应生成  $2P/S$  个新染色体, 以保持群体规模不变. 可见, 当  $S$  很大时会造成抽样密度的不足, 影响进化速度, 这一点在前面关于选择运算的说明中已经提到. 由于上步选择出的父代染色体个数为  $S = P/4$ , 所以在本步交叉运算中, 每一对父代染色体相互交叉应生成 8 个新染色体. 通过比较两个染色体来设定抽样区间是一条重要优化规则, 它使新产生的群体向适应度增大的方向进化. 抽样区间大小与两个染色体对应基因的差别有关, 差别越大区间越大, 这一作法的合理性可以这样理解: 假设两个染色体的适应度不同而其中某一对对应基因在两染色体中差别不大, 则可认为改变该对基因无助于提高适应度, 亦即该对基因不需再做大的改变, 因此抽样区间应较小; 相反若某对基因在两染色体中差别

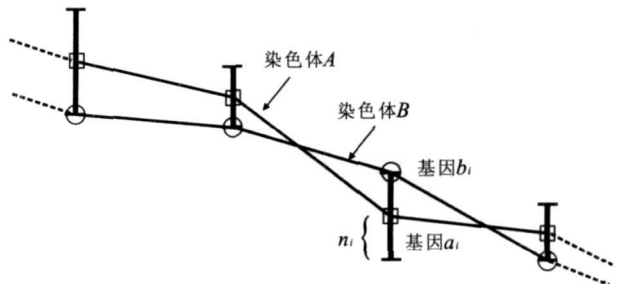


图 1 交叉运算示意(说明见正文)

Fig. 1 Sketch showing the crossover operator

很大,则可认为该基因的改变能够提高适应度,因此抽样区间应较大.该交叉方法的思路曾用于岩石波谱识别的遗传算法中,搜索效果也较好(张振飞等,2003).抽样方法按样本的概率分布又有多种选择,如均匀分布或高斯分布等.其中均匀抽样速度略快,且在最初若干代的进化中寻优效果好于高斯抽样,这是因为后者较多地依赖于当前较好染色体,一定程度上限制了搜索空间.③变异.简单地说,染色体的变异是指它的基因发生变化.这是一种扩大搜索空间、避免陷入局部最优的策略.在普通遗传算法中,若采用二进制编码,则常从染色体中按概率随机选择部分基因(二进制位值)实行取反运算(Goldberg, 1989).这种方法在遗传最优分割中不再适用,因为遗传最优分割以整数序列为染色体.我们采用“灾变变异”策略(Back *et al.*, 1991; Eshelman, 1991),即从新一代群体中,按概率随机选择部分染色体进行初始化.变异运算的具体方法是,预先定义一个控制参数,称为最大变异概率,用 $P_m$ 表示( $0 < P_m < 1$ );对于群体中每个染色体,生成一个(0, 1)区间上均匀分布的随机数 $r$ ,若 $r < P_m$ 则该染色体进行初始化,否则该染色体保持不变.显然, $P_m$ 的大小决定了变异染色体的多少.若 $P_m$ 很小,则发生变异的染色体数目将会很少,子代的基因多样性与父代差别不大,相当于限制了搜索空间;若 $P_m$ 很大,则子代群体中将有許多初始化的染色体,使搜索过程趋向于完全随机,优化准则的作用将不能充分发挥.因此, $P_m$ 不宜过小也不宜过大.试算显示, $P_m$ 值为0.2~0.3时进化较快.④精英保持.精英是指群体中一个或多个最好(适应度最高)的染色体.精英保持是指父代中最好的一个染色体直接进入子代群体成为其中一员.该运算引用了Eshelman(1991)提出的“跨世代精英选择”改进遗传算法的思想,其目的是确保子代群体的最高适应度不低于父代,父代的最好基因组合在子代中得到保留,使整个算法的稳健性提高.具体计算时以父代最好染色体随机替换一个子代染色体,以使群体规模不变.若直接进入子代的父代精英在子代中仍具有最高适应度(亦即子代的最高适应度与父代的最高适应度相同),则该最高适应度称为历史最高适应度;群体中不同于历史最高适应度的最高适应度称为当代最高适应度.由于精英保持的原因,历史最高适应度随着进化过程永不降低.当代最高适应度在进化初期会有不断升高的趋势,当它高于历史最高适应度时将会成为新的历史最高适应度.在进化后期,当代最高适应度一

般会波动,这是因为进化过程带有一定的随机性.

(6)新一代染色体基因排序,然后转第(3)步.排序是必需的,因为划分出的地层单位之间不能互相重叠.排序操作后,染色体中基因的排列可能有了新的秩序,由上一步交叉运算获得的第 $i$ 个基因可能不再是第 $i$ 个,而是有了其他序号.因此,排序操作对前述交叉运算的寻优效果有一定破坏作用.这种破坏作用是难以避免的,可能会影响算法的效率.但因交叉运算是在排序后的染色体上进行的,新一代群体中大部分染色体将会保持原来的基因秩序,因此,可以预期这种破坏作用是有限的,实际计算也表明整个算法总体效率很高.

通过以上第(3)~(6)步的循环搜索,获得最优解,即满足段间离差尽量大、段内离差尽量小,同时尽可能为测井曲线突变点的分界点号序列.然后结合其他信息从该点号序列中选择最可能的地层分界,完成地层划分.

需强调说明的是,上述方法只寻找所处理井深范围内最可能的地层(岩性组合)界线,不能进行地层标定和对比.地层标定和对比是根据地层划分结果,综合已有地质认识,分析测井电性特征并参考录井数据,通过多个相邻井的相互对比而进行的,地层划分为地层标定和对比提供了必要前提.地层划分、标定和对比之间有密切联系,但限于篇幅,本文对标定和对比的具体方法不再论述.

## 2 应用实例

使用上述方法,我们在过去的研究中对济阳拗

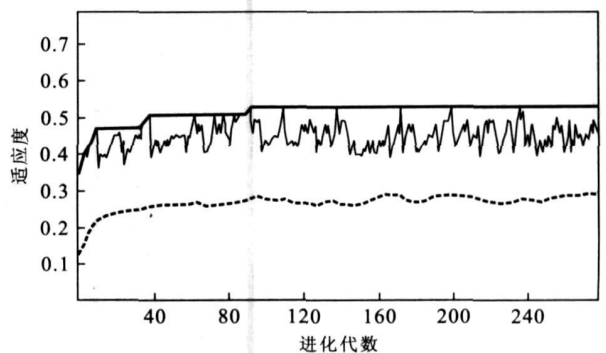


图2 Y-136井地层划分的遗传最优分割法进化曲线  
Fig. 2 Evolution curves of genetic optimal stratigraphical division for well Y-136

粗线为历史最高适应度,细线为当代最高适应度,虚线为群体平均适应度,其他说明见正文

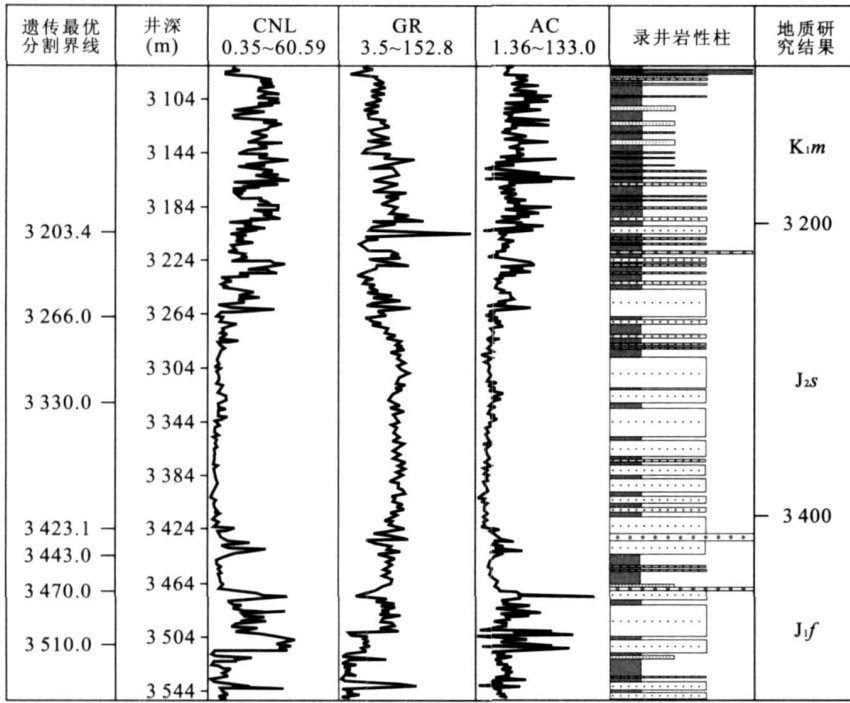


图 3 Y-136 井遗传最优分割法地层划分与地质研究划分结果的比较

Fig. 3 Result of the genetic optimal stratigraphic division, comparing to that of geological studies, for well Y-136 CNL(补偿中子)、GR(自然伽马)和 AC(声波时差)的单位分别为%、API 和  $\mu s/m$ ; 录井岩性柱中, 岩层符号由短到长分别表示泥质岩、(泥质)粉砂岩、(含砾)砂岩和砾岩

陷内 180 多口井的中生界进行了地层划分, 一般划分到组, 在中生界厚度较大时划分到段. 划分结果在该地区的生产和科研中得到了应用. 现以一口中生界地质情况已知的井来说明遗传最优分割法的应用效果.

Y-136 井位于济阳拗陷中部, 穿过中生界地层, 并全井取心. 该井钻遇中生界深度范围为 3 080.0 ~ 3 510.0 m, 实施了多参数地球物理测井, 中生界范围内测点数 3 441. 胜利油田地质研究院(内部资料, 2005)通过岩矿鉴定、化学分析、古生物分析等手段, 并参考井旁道地震剖面, 对该井岩心做了细致的研究, 确定了中生界顶底界线, 用地质方法对中生界进行了地层划分和标定, 确认该井 3 096 ~ 3 200 m 属蒙阴组(K<sub>1m</sub>), 3 200 ~ 3 400 m 为三台组(J<sub>2s</sub>), 3 400 ~ 3 590 m 为坊子组(J<sub>1f</sub>). 因此, 该井中生界内部地层界线可认为是已知的. 本文选用声波时差(AC)、补偿中子(CNL)和自然伽马(GR)3 种测井参数, 用遗传最优分割进行 8 段划分. 原始数据用式(6)进行了均匀化预处理. 边缘检测窗口取 64 点(8 m). 遗传算法设置为群体规模 128, 最大变异概率 0.2, 最大进化代数 280. 在主频为 1 GHz 的微机

上仅用 5 s 时间算出 7 个最优分界线位置(用同样的设置和计算机, Fisher 最优分割耗时约为 2 h). 图 2 为计算过程的进化曲线, 可看到在进化到第 90 代时已找到最优解. 在 90 代到 280 代的进化中又多次找到同一个最优解. 反复计算 10 次得到同样结果, 证明该结果是稳定和可靠的.

图 3 为地层划分结果与岩心地质研究结果的比较. 可以看到遗传最优分割法找到的分界点与 3 种测井参数曲线的突变点位置对应较好, 所找到的 7 个最优分割点, 有两个(深度分别为 3 203.4 m 和 3 423.1 m)与地质研究所确定的地质界线深度基本一致(最大误差 23 m), 说明该方法的结果基本符合地质实际. 所找到的另外 5 个分界点, 可以在组中进一步细分段时参考. 从图 3 中还可看到, 与地质研究结果及基于测井数据的遗传最优分割结果相比, 录井岩性柱反映的岩性界线深度误差较大, 在地层划分中可参考但不应起决定作用.

### 3 结论

本文提出的遗传最优分割法, 是一种能够快速

有效地处理多参数测井数据、进行地层划分的方法,是遗传算法、Fisher 最优分割法和边缘检测算法三者的有机综合。以测点的边缘隶属度和多参数测井数据的 Fisher 比共同构成优化准则(染色体适应度),可同时获得段内电性尽量均匀、段间电性尽量不同、分界点尽量是测井曲线突变点的优化效果,地层划分结果能较好地符合地质实际。采用遗传算法的随机搜索方法取代传统最优分割方法中的逐点计算方法,计算速度得到极大提高。遗传算子(选择、交叉、变异和精英保持)应根据地层划分的特殊要求而设计,即必须考虑到解的形式为测点号的有序组合、划分的地层单位不能有重叠等特点。该方法以快速和地质可靠性高的特点,适合于研究解决像济阳坳陷中生界这样没有露头、岩心稀少、地震勘探分辨率在深部变低、而测井数据量非常大情况下的隐伏地层划分问题。

本研究得到胜利油田地质科学研究院王永诗总地质师、沾车海室李开孟高工、林会喜主任、庄文山高工、阳显文工程师、地层古生物室贺振健高工、郝运轻工程师、边雪梅工程师等的大力支持和帮助,在此表示衷心感谢!

## References

- Andrew, D. M., 1995. Whither stratigraphy? *Sedimentary Geology*, 100(1-4): 5-20.
- Back, T., Hoffmeister, F., Schwefel, H. P., 1991. A survey of evolution strategies. In: ICGA, ed., Proceedings of the fourth international conference on genetic algorithms. Morgan Kaufmann Publishers, San Mateo, CA, 2-9.
- Bureau of Geology and Mineral Resources of Shandong Province 1991. Regional Geology of Shandong Province Geological Special Reports (Regional Geology), Ministry of Geology and Mineral Resources, P. R. China. Geological Publishing House, Beijing (in Chinese).
- Cao, Y. C., Jiang, Z. X., Wang, L. Q., et al., 1996. Classification and identification of sequence stratigraphic units in a rift lack basin. *Journal of the University of Petroleum, China*, 20(4): 1-5 (in Chinese with English abstract).
- Cao, Z. X., Liu, H., Wang, Y. L., 2002. Sedimentation and stratigraphic sequence analysis of Permo-Carboniferous system in Jiyang depression. *Journal of Shandong Inst. of Min. & Tech.*, 21(22): 68-71, 78 (in Chinese with English abstract).
- Deng, H. W., Wang, H. L., Li, X. M., 1997. Application of High-resolution sequence stratigraphic correlation to fluvial facies. *Oil & Gas Geology*, 18(2): 90-95, 114 (in Chinese with English abstract).
- Eshelman, L. J., 1991. The CHC adaptive search algorithm: How to have save search when engaging in non-traditional genetic recombination. In: Rawlins G. J. E., ed., Foundations of genetic algorithms and classifier systems. Morgan Kaufmann Publishers, San Mateo, CA, 265-283.
- Goldberg, D. E., 1989. Genetic algorithms in search, optimization and machine learning. Addison-Wisely, Reading, Mass.
- Groch, W. D., 1982. Extraction of line shaped objects from aerial images using a special operator to analyze the profiles of functions. *Computer Graphics and Image Processing*, 18, 347-358.
- Holland J. H., 1975. Adaptation in natural and artificial systems The University of Michigan Press Ann. Arbor. .
- Inoue, N., Kitada, N., Itoh, Y., et al., 2003. Integrated study of high resolution geophysical and geological information of Osaka bay, Southwest Japan. *Journal of Asian Earth Sciences*, 22(1): 1-11.
- Li H. G., Du, X. D., Lu, K. Z., et al., 1999. Evolution and structural characteristics of the Mesozoic Era in the central/west part of Bohai Bay area. *Journal of the University of Petroleum, China*, 23(3): 1-5 (in Chinese with English abstract).
- Li H. Q., Zhang F. M., Shao C. R., et al., 2002. A new multi-well logs optimum dividing algorithm with heuristic search rules. *Well Logging Technology*, 26(4): 273-276 (in Chinese with English abstract).
- Lü, X. X., Xiao, H. Q., Tian, M. R., et al., 2003. On the well-logging identification and description technique of the glutinite reservoir bed in the escarp regions of Jiyang depression. *Journal of Zhejiang University (Sciences Edition)*, 30(3): 332-336 (in Chinese with English abstract).
- Maulin D. P., George, A. M., 2003. Building 2-D stratigraphic and structure models from well log data and control horizons. *Computers & geosciences*, 29(5): 557-567.
- Pinous O. V., Sahagian, D. L., Shurygin, D. L., et al., 1999. High-resolution sequence stratigraphic analysis and sea-level interpretation of the Middle and Upper Jurassic strata of the Nyurolskaya depression and vicinity (southeastern West Siberia, Russia). *Marine and Petroleum Geology*, 16(3): 245-257.

Shaaban, F. F., Ghoneim, A. E., 2001. Implication of seismic



- and borehole data for the structure, petrophysics and oil entrapment of Cretaceous-Palaeocene reservoirs, northern Sirt Basin, Libya. *Journal of African Earth Sciences*, 33(1): 103–133.
- Wang, J. W., Zhao, Y. S., Tian, H. Q., 2001. Primary study on sequence stratigraphy of Kongdian Formation in Dongying-Huimin sag. *Journal of the University of Petroleum, China*, 25(6): 2–6 (in Chinese with English abstract).
- Wang, Y., Lu, K. Z., Ren, A. S., 1994. Tectogenesis and volcanism and the basin evolution of the Mesozoic Era in the Northeast of Jiyang depression. *Journal of the University of Petroleum, China*, 18(2): 1–8 (in Chinese with English abstract).
- Wonik, T., 2001. Gamma-ray measurements in the Kirchrode I and II boreholes. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 174(1–3): 97–105.
- Zhang, G. Q., Li, J. Y., 1991. Quantitative lithostratigraphy. China University of Geosciences Press, Wuhan, 35–45 (in Chinese with English abstract).
- Zhang, S. P., Liu, C. D., He, Z. J., et al., 2003. Stratigraphic division of redbeds in deep Dongying depression, East China. *Earth Science Frontiers*, 10(2): 465–469 (in Chinese with English abstract).
- Zhang, Z. F., Hu, G. D., Yang, M. G., 2003. An evolutionary-strategy-based CHC genetic algorithm and its application to rock spectrum discrimination. *Earth Science—Journal of China University of Geosciences*, 28(3): 351–355 (in Chinese with English abstract).
- 元的划分及界面识别标志. 石油大学学报(自然科学版), 20(4): 1–5.
- 曹忠祥, 刘华, 王玉林, 2002. 济阳拗陷石炭二叠系沉积与层序地层分析. 山东科技大学学报(自然科学版), 21(22): 68–71, 78.
- 邓宏文, 王洪亮, 李小孟, 1997. 高分辨率层序地层对比在河流相中的应用. 石油与天然气地质, 18(2): 90–95, 114.
- 李洪革, 杜旭东, 陆克政, 等, 1999. 渤海湾地区中西部中生代构造特征及演化. 石油大学学报(自然科学版), 23(3): 1–5.
- 李洪奇, 张福明, 邵才瑞, 等, 2002. 带有启发式搜索规则的多元测井数据序列最优分割方法. 测井技术, 26(4): 273–276.
- 吕希学, 肖焕钦, 田美荣, 等, 2003. 济阳拗陷陡坡带砂砾岩体储层测井识别及描述技术. 浙江大学学报(理学版), 30(3): 332–336.
- 山东省地质矿产局, 1991. 山东省区域地质志. 地质矿产部地质专报(区域地质). 北京: 地质出版社.
- 王建伟, 赵勇生, 田海芹, 2001. 东营惠民凹陷孔店组层序地层学研究. 石油大学学报(自然科学版), 25(6): 2–6.
- 王毅, 陆克政, 任安身, 1994. 济阳拗陷东北部中生代构造运动和火山活动及盆地演化. 石油大学学报(自然科学版), 18(2): 1–8.
- 张光前, 李继英, 1991. 定量岩石地层学. 武汉: 中国地质大学出版社, 35–45.
- 张守鹏, 刘成东, 贺振建, 等, 2003. 东营凹陷深部“红层”地层划分方法探讨. 地学前缘, 10(2): 465–469.
- 张振飞, 胡光道, 杨明国, 2003. 基于进化策略的 CHC 遗传算法及岩石波谱识别. 地球科学——中国地质大学学报, 28(3): 351–355.

#### 附中文参考文献

操应长, 姜在兴, 王留奇, 等, 1996. 陆相断陷湖盆层序地层单