

# 一种基于遗传算法的作曲方法

A Composition Method Based on Genetic Algorithm

卿武明 黄青松 臧晓晗

Qin Wuming Huang Qingsong Zhang Xiaohan

(昆明理工大学信息与自动化学院, 云南 昆明 650051)

(School of Information and Automation, Kunming University of Science and Technology, Yunan Kunming 650051)

**摘要:** 本文介绍了一种基于遗传算法, 以乐段为单位的作曲方法。该方法能生成具有相对完整乐思的音乐片段, 并引入了音乐的风格化特征, 能按照用户的情感因素产生用户所需的音乐片段。

**关键字:** 遗传算法; 作曲; 乐段; 乐句

中图分类号: TP396

文献标识码: A

文章编号: 1671-4792-(2008)5-0052-05

**Abstract:** The paper introduces a composition method that is based on Genetic Algorithm(GA) and makes period as unit. The method can create the music segments that have the whole music thought relatively. Moreover, the method introduced the style character of music can produce the music segments according to the factors of people's sentiment.

**Keywords:** Genetic Algorithm; Composition; Period; Phrase

## 0 引言

算法作曲(algorithmic composition), 或称自动作曲(automated composition) 是试图使用某个形式化的过程, 以使人(或作曲家)在利用计算机进行音乐创作时的介入程度达到最小的研究<sup>[1]</sup>。目前算法作曲这一领域采用的主要技术有马尔科夫(Markov)转换表、随机过程、算法作曲研究中的知识库系统、音乐文法、人工神经网络技术和遗传算法<sup>[2]</sup>。遗传算法(Genetic Algorithm, 简称GA)是一种基于进化论优胜劣汰、适者生存的物种遗传思想的搜索算法。20世纪50年代初, 由于一些生物学家尝试用计算机模拟生物系统, 从而产生了GA的基本思想。美国密执根大学的霍勒德(J.H. Holland)于20世纪70年代初提出并创立了遗传算法<sup>[3]</sup>。遗传算法是一种高度并行、随机的群体搜索技术, 具有良好的普适性和可规划性。遗传算法作曲是利用遗传算法来控制音乐的生成过程。在作曲过程中首先将给定乐曲进行一定方式的编码, 编码后的乐曲称为染色体, 并采用遗传算法中的交叉、变异等算子对乐曲进行“进化”, 用适应度函数来衡量进化结果, 如此不断进行直到找到最终的满意解为止。

## 1 总体设计

目前, 已经有许多遗传算法作曲系统的研究和尝试, 取得了一些成果, 同时也存在许多问题。主要集中在下面几点:

由于音乐信息本身的复杂性, 导致在对乐曲进行编码时, 会出现不确切和不规范现象。而遗传算法作曲的效果在很大程度上依赖于系统对于音乐知识的表现程度, 当系统对于音乐的表达越充分时, 所产生的作品也会越精确。由于音乐结构的复杂和抽象性, 当前许多系统在推理结构的设计上还存在许多不足, 所以目前的许多系统只是应用于产生比较短小的音乐片段。音乐风格化问题。当前的遗传算法作曲系统所产生的音乐片段大都是没有音乐风格或单一风格, 不能按照用户的情感因素产生用户所需的音乐片段。

为了改善上述问题, 本系统做了以下设计:

(1) 为了比较精确的表达音乐信息, 系统采用音程差法来对音乐进行编码, 即音高特性采用了与上一位音符的音程差来记录(包括首尾音符与调式主音的音程差), 这样编码可以更好地表达所谓的音乐旋律线的特性, 并且更加适用于遗传算法对音乐的操作, 消除了直接用音高编码经遗传算法后乐句中产生大副度的音高变化的不足。另外, 除了音长、音程差、乐速、节拍等信息外, 系统还给每一位音符增加了遗传操作标示位和特殊音符标示位, 给每一个乐句赋予了起始音值, 使音乐的表达更充分和精确。

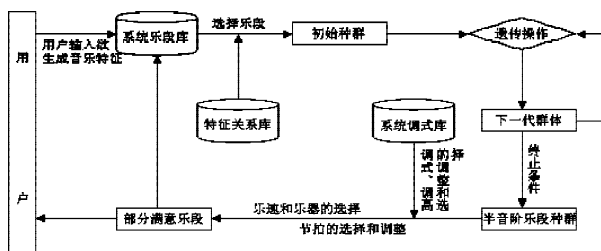
(2) 为了产生比较完整的音乐片段, 系统采用乐段作为遗传算法的操作对象。乐段是音乐作品中表现完整乐思或相

对完整乐思的最小结构,是建立在单一主题上的、最小的完整曲式。乐段的结构规模一般不太长,但形式完整、统一,能体现鲜明的音乐形象。乐段常由几个乐句组成,这些乐句之间具有问答呼应的关系。由两个或四个乐句组成乐段,是比较常见的现象,但也有由三个、五个或更多乐句组成的乐段,或由一个独立乐句组成的乐段。乐段内的乐句结构可分为重复型、并置型和“起、承、转、合”型<sup>[4]</sup>。乐句间互相用终止式分开来,又通过终止式的功能关系联合成有机的曲式结构。终止式(cadence)即结束乐句或乐段的和声进行公式,又称收束。在主旋律中表现为乐段中各个乐句的终止音符的音高与调式主音之间的音程差。因此,采用乐段作为遗传算法的操作对象能生成具有完整乐思的音乐片断。

(3)为了能够按照用户的情感因素产生用户所需的音乐片断。系统给每段乐段都赋予了若干模糊特征,这些特征模糊的按百分比反应了乐段的音乐风格。例如:欢快、宣叙等等。并用这些特征值来描述适应度函数,通过这些特征值来计算一个特征描述精确值来衡量产生的子乐段是否符合用户的需求。

适应度函数用来描述乐段的风格特征,那么产生乐段的音乐质量怎么来保证又成了新的问题。系统做了以下设计来保证产生乐段的音乐质量:建立一个系统乐段库,初始种群从系统乐段库里选取,而不是随机生成。系统对遗传操作的交叉和变异进行了严格的规范,而对音乐信息比较精确的编码在这里起到了关键做用。建立一个系统调式库,对产生的半音阶乐段进行半音的导向转换。调式库在保证音乐质量的同时,还让音乐更加的风格化。

根据以上设想,设计了如下系统结构,如图一所示。

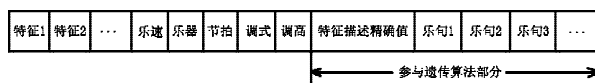


图一 系统总体结构

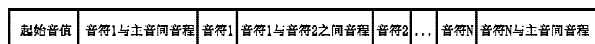
系统采用音程差法来对音乐进行编码。即音高特性采用了与上一位音符的音程差来记录(包括首尾音符与调式主音的音程差),如图二所示。

其中特征包括了特征项和特征项所占百分比。乐速表示乐曲的演奏速度。乐器表示演奏音乐的乐器名,不同的乐器演奏可以表现出不同的音乐风格。节拍信息包括了拍号(例如:2/4)以及小节中各拍子上的音符音强的强弱。调式和调高即音乐的调性,例如最常见的“C大调”,“大调”是调

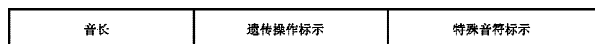
乐段:



乐句:



音符:



图二 编码结构

式,“C”是调高。这些信息都可以抽象的用整型变量来记录。特征描述精确值通过特征值与特征关系库中的特征关系值计算得到,用来描述适应度函数。

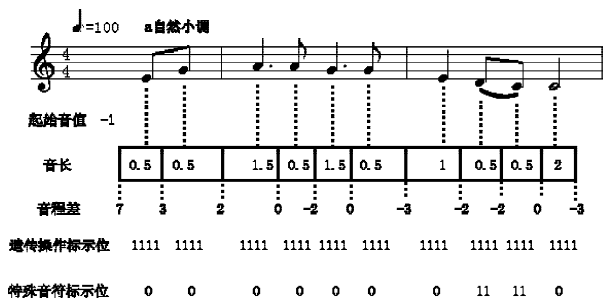
起始音值描述乐句首音符的节拍起始位置。如果是正常的从强拍开始的乐句,即第一小节为“完整小节”,则起始音值的值为0。若乐句为弱起拍子,即第一小节是“弱起小节”或称为“不完整小节”<sup>[5]</sup>,则起始音值为负。起始音值的作用是记录乐句强拍的位置,便于让处于节拍强拍上的音符经过遗传操作后仍然处于节拍的强拍上,这对于产生的音乐是否和谐起着至关重要的作用。音程差即两个音符的音高差,以半音为单位,音升高音程差取正,音降低音程差取负,由定义可知,一个乐句的所有音程差的和为0。另外,对于没有音高的休止符,系统定义休止符的音高为调式主音音高。

音长即音符发声保持的时间长度,取节拍值作为音长值,而不是以音符的时值为音长。这样能让以不同音符时值为一拍的乐句统一起来,节拍比音符时值更能体现音乐节奏的本质。遗传操作标示为四位2进制编码,取值为0或1,分别用来表示该音符可否进行交叉和三种不同的变异操作,其中1表示允许操作。特殊音符标示用来表示音符的类型,包括普通音符、休止符、和弦、断音、圆滑音、延音、保持音、连音、倚音、波音、颤音、回音等等,编码如表一所示。如果想更加精确的表达音乐内容,可以继续增加特殊音符的种类,如倚音又可分为长倚音、短倚音、后倚音。某些特殊音符在进行遗传操作时要特殊处理。

表一 特殊音符编码表

| 音符类型 | 普通音符 | 休止符 | 和弦 | 断音 | 圆滑音 | 延音 |
|------|------|-----|----|----|-----|----|
| 编码   | 0    | 1   | 2  | 10 | 11  | 12 |
| 音符类型 | 保持音  | 连音  | 倚音 | 波音 | 颤音  | 回音 |
| 编码   | 13   | 3   | 20 | 21 | 22  | 23 |

图三为乐句编码示例图。该乐句的调性是a自然小调,乐速为100,节拍为4/4拍;不完整小节的两个音符的音长的和正好是一拍,因此值为-1;因为乐句以四分音符为



图三 乐句编码示例

一拍,因此四分音符音长值为1,以此类推,八分音符音长值就为0.5; 因为a自然小调的主音为a,所以第一个音符e与主音a的音程差就是7个半音,又如第二个音符g与第一个音符e的音程差为3; 遗传操作标示位都为1111; 整个乐句只有倒数第三个音符和倒数第二个音符为圆滑音,其它皆为普通音符。

### 1.1 初始种群的生成

遗传算法的第一步就是生成初始种群,系统的初始种群从系统乐段库里选取。用户选定欲生成乐段的乐句数、乐句之间的结构(例如四句体乐段的常见结构如下: ABAB1、ABCB、ABCD等等)以及每小节的节拍数后,系统会从相应的乐段库中选取乐段。假设系统共定义了N个模糊特征,生成步骤如下:

(1)用户输入欲生成乐段的各个特征占的百分比为 $X_i$ ( $i$ 为1到N的正整数)。

(2)从系统乐段库中随机选取一个乐段Y,假设此乐段的各个特征占的百分比为 $Y_j$ ( $j$ 为1到N的正整数)。

(3)计算 $A(0 < A < 100) = X_i Y_j K_{ij}$ 的值,其中 $K_{ij}$ 为特征 $i$ 与特征 $j$ 的特征关系值,存储在特征关系库中,取值为0到100,值越小关系越近。

(4)计算 $B = k(1 - (A/100)^{1/2})$ ,其中 $k$ 为0到1的常数,得到一个0到1的值 $B$ ,就是乐段Y被选入初始种群的概率。乐段如果被选中,数值 $A$ 记录为该乐段的特征描述精确值,用来描述乐段与目标乐段的特征差距。重复步骤2。

(5)当选取的乐段数目达到目标值,生成停止。

其中第4步把 $A/100$ 开2次方,以及加入常数 $k$ ,是为了降低乐段被选的概率,使乐库中更多的乐段能得到选中的机会。

### 1.2 遗传操作

(1)选择操作。是在群体中选择生命力强的个体产生新的群体的过程。遗传算法使用选择算子来对群体中的个体进

行优胜劣汰操作。系统的选择操作采用轮赌选择法,如果一个乐段的特征描述精确值越高,则适应度底,那么它被选择到的机会就越小。想象一个轮盘赌的机器上放置了群体中所有的染色体,每一个染色体所占的地方的大小和它的适应度成正比。然后开始扔弹子,扔到哪个地方就把对应的染色体选择出来。显然,适应度越高的染色体被选到的机会就越大。

(2)交叉操作。遗传算法利用交叉操作来组合两个个体或双亲,为下一代形成交叉的子个体。为了保证音乐的连贯性,采用单点交叉的方法。由于一个乐段是包含多个乐句的结构组合,为了保留乐句间的曲式结构,需要对乐段内的各个乐句做统一并独立的交叉,即两个乐段中对应的乐句成对的进行交叉。首选要选定各个乐句的交叉点,步骤如下:

定义一个函数 $f(X_{ni})$ : 计算乐段X中的乐句 $X_n$ 中的音符 $X_{ni}$ 的音长和该乐句中位于该音符之前的所有音符的音长以及起始音值的和。

定义一个激活音符结点树T,其中:

选择父代中两个优秀的染色体另存为X、Y;

从乐段X的第一个乐句X1中挑取所有遗传操作标示位交换位为1的音符,存储到T中,  $int\ n=1$ ;

如果树为空,则算法结束,到结果2。如果树不为空,而树的第n层为空,则回到父结点,  $n--$ ,到步骤8。否则从树的第n层中随机选取一个音符 $X_{ni}$ ,计算 $f(X_{ni})$ 值;

从前到后依次计算乐句 $Y_n$ 中每个音符的 $f(Y_{nj})$ 值。如果 $f(Y_{nj})=f(X_{ni})$ ,到步骤5。当 $f(Y_{nj})>f(X_{ni})$ 时,到步骤8;

检查音符 $Y_{nj}$ 的遗传操作标示位交换位,如果为1,到步骤6。如果为0,到步骤8;

假设乐句 $X_n$ 、 $Y_n$ 各包含音符u、v个,其间的音程差分别表示为 $A_0, A_1, \dots, A_u, B_0, B_1, \dots, B_v$ ,分别计算 $(A_i + A_{i+1} + \dots + A_u) - (B_{j+1} + B_{j+2} + \dots + B_v)$ 与 $(A_{i+1} + A_{i+2} + \dots + A_{u-1}) - (B_j + B_{j+1} + \dots + B_{v-1})$ 的值,如果都小于等于9,则记录 $X_{ni}$ 、 $Y_{nj}$ 为第n对乐句的交叉点,到步骤7。否则到步骤8;

如果乐句 $X_{n+1}$ 为空,则算法结束,到结果1。如果不为空,则从前到后依次计算乐段X中下一乐句 $X_{n+1}$ 中每个音符的 $f(X_{(n+1)i})$ 值,并记录下与 $(f(X_{1i}) + \dots + f(X_{ni}))/n$ 的绝对差值小于等于1并且遗传操作标示位交换位为1的所有音符,做为 $X_{ni}$ 的子结点存储到T中,  $n++$ ,到步骤3;

把该音符的操作标示位交换位改为0,从树中删除该节点,到步骤3。

结果1: 交叉点选取成功,分别为 $X_{1i}, Y_{1j}, X_{2i}, Y_{2j}, \dots$ 。

结果2: 交叉点选取失败,两乐段之间没有合适的交叉点,乐段直接加入新一代群体。

交叉点的选取主要依据了以下原则：进行交叉的对应乐句的交叉点的 $f(X_{n_i})$ 值必须相等；两个乐段的所有的乐句的交叉点的 $f(X_{n_i})$ 值不能相差太大；交叉点音符的遗传操作标示位交换位必须为1；两句乐交叉后，交叉点的音程差小于等于9。

交叉点选取成功后，就要对乐段中的乐句成对的进行交叉操作。为了保留乐句间的曲式结构，有必要让乐句的首尾音符与调式主音的音程差保留在同一个乐句中。假设两个乐句 $X$ 、 $Y$ 分别为：

$$X: a_0x_1a_1x_2a_2\cdots x_ia_ia_{i+1}\cdots x_{m-2}a_{m-2}x_{m-1}a_{m-1}x_ma_m$$

$$Y: b_0y_1b_1y_2b_2\cdots y_jb_jy_{j+1}\cdots y_{n-2}b_{n-2}y_{n-1}b_{n-1}y_nb_n$$

其中， $x_ia_i$ 、 $y_jb_j$ 为交叉点， $a_0$ 、 $a_i$ 、 $\dots$ 、 $b_n$ 为音程差， $x_1$ 、 $x_2$ 、 $\dots$ 、 $y_n$ 为音符。交叉后两乐句分别为：

$$X': a_0x_1a_1x_2a_2\cdots x_ia_i' \cdots y_{n-2}b_{n-2}y_{n-1}b_{n-1}y_na_n$$

$$Y': b_0y_1b_1y_2b_2\cdots y_jb_j' \cdots x_{m-2}a_{m-2}x_{m-1}a_{m-1}x_mb_m$$

其中：

$$a_i' = -(a_0+a_1+a_2+\cdots+a_{i-1}+b_{j+1}+\cdots+b_{n-2}+b_{n-1}+a_m)$$

$$b_j' = -(b_0+b_1+b_2+\cdots+b_{j-1}+a_{i+1}+\cdots+a_{m-2}+a_{m-1}+b_n)$$

另外，通常很多乐句是以休止符结尾，由于休止符不发音，因此表达乐句收束式的其实是倒数第一位发音的音符，而不是休止符。所以在交换前还要进行一次逻辑判断，如果乐句结尾是休止符，要顺着乐句从后向前找到倒数第一位非休止音符，然后用该音符与休止符之间的音程差来替代 $X'$ 、 $Y'$ 、 $a_i'$ 、 $b_j'$ 中的确 $a_m$ 或 $b_n$ 值。

乐段内所有乐句都对应交叉完成后，还要计算新产生子乐段的特征描述精确值，子乐段的特征描述精确值可以通过双亲的特征描述精确值以及乐句的交叉点计算得到。因为双亲乐段的乐句的首尾音程差都完整的遗传给了某一个子乐段，因此在计算子乐段的特征描述精确值时，可以给予适当的加权。

(3)变异操作。遗传算法中，变异的作用是引进新的遗传物质或恢复已失去的遗传物质。交叉操作完成后，就可以对染色体进行变异操作。为了保证乐曲的生成质量和状态空间的完整，系统使用了下面三种变异。

改变一个音符两旁的音程差值。音符遗传操作标示位变异位第一位为1，音符允许按一定的变异率进行此项变异。音符变异时，两个音程差按概率对应增加和降低某个音程，如果变异后的音程差的绝对值大于12，也就是一个八度，变异取消。另外，弱拍音符、强拍音符、首尾音符的变异率是依次降低的，并且音程差的各个变异值的概率也是不同的。

把一个音符分成两个音符。音符遗传操作标示位变异位第二位为1，音符允许按一定的变异率进行此项变异。变异后的两个音符的音长和为原音符音长，原音符的音长按一

定的概率以不同的分配方案分配给两个音符。如果两个音长中较小的音长值小于乐段中最小的普通音符的音长值，变异取消。否则两音符之间按概率生成一位音程差，第二个音符后的音程差等于原音符后的音程差减去生成的中间音程差，如果变异后的音程差的绝对值大于12，变异取消。另外，弱拍音符、强拍音符、尾音符的变异率是依次降低的，并且音长的分配方案概率和中间音程差值的生成概率也是不同的。

合并两个音符为一个音符。音符遗传操作标示位变异位第三位为1，音符允许按一定的变异率进行此项变异。变异后音符的音长为两音符音长的和。如果两音符都为休止符，则合并后是休止符；如果两音符只有一个休止符，则合并后的音高为合并前非休止符的音高；如果没有休止符，则合并后的音高为合并前处于强拍或强位位置音符的音高。

终止条件。当乐段种群的平均特征描述精确值低于某个值或达到指定遗传操作次数时，遗传操作终止，生成半音阶乐段种群。

### 1.3 调式和节拍的选择与调整

生成的半音阶乐段种群是不和谐的，系统必须给音乐赋予一定的调式。所谓调式，就是若干高低不同的乐音，围绕某一有稳定感的中心音，按一定的音程关系组织在一起，成为一个有机的体系<sup>[5]</sup>。不同的调式能体现不同的音乐风格，常见的调式有自然大调、自然小调、和弦小调、和弦大调、旋律小调和旋律大调等等。用户给乐段选择调式后，系统对半音阶乐段的一些不在调式内的音符以特定的概率进行导向转换，转换到调式内的音高。

除了调式的选择和调整，还要对乐段进行节拍的选择和调整。选择节拍即选择乐曲以何种音符为一拍。调整节拍是由于有“弱起小节”的原因，“弱起小节”经过遗传操作后，有可能会产生乐段相邻两句乐在音乐行进时间上的重合，这是必须要处理的。假设乐句 $X$ 、 $Y$ 在音乐行进时间上有重合，其中 $X$ 在前。处理时依照以下原则：若只是休止符或音符音长的重合，删除重合的休止符或改变音符的音长即可；若发生音符的重合，逐个删除音符直到只有音长的重合或无重合为止；在同等条件下，优先对乐句 $Y$ 进行处理。处理完之后，还要修改 $Y$ 乐句的起始音值。

经过调式和节拍的选择与调整后，用户就可以给乐段选择相应的调高、乐速、乐器来演奏乐曲。如果结果中有用户比较满意的乐段，就可以存储到系统乐段库中，该乐段的特征百分比值就是用户最开始输入的百分比值，乐段的特征描述精确值归零。

### 1.4 简单实验

下面是一个简单例子，模糊特征只提取了两项“快乐”和“悲伤”，关系值为100。乐段库存储了10个四句乐段，

结构都为ABCD,都是一小节4拍。用户输入的欲生成音乐特征值为“快乐”0.1,“悲伤”0.9,种群规模为2乐段。经过系统选择、交差、变异后,平均特征描述精确值低于10时,算法停止,得到两个半音阶乐段。系统默认调式主音为C,以四分音符为一拍,其中效果较好的一个乐段如图四所示。



图四 进化结果1

由于是悲伤风格,所以调式选择自然小调。调高选择为a,节拍为4/4拍。经过节拍和调式的调整后,结果如图五所示。



图五 进化结果2

## 2 结束语

通过对遗传算法作曲的研究可以发现,遗传算法以乐段为单位的创作是可行的。通过合理的编码、系统乐段库以及对遗传操作过程的约束,能生成有完整曲式结构的乐段,并且当音乐知识表现的更精确时,生成的乐曲也会更加精确和谐。另外,本文用音乐的风格化特征来描述算法作曲的适应度函数是一种崭新的思路,用户能按照自己的情感因素产生用户所需的音乐片断。但是音乐模糊特征的提取以及提取的音乐特征是否能更客观的表达音乐的涵义就成了新的难题。在这点上,仍然需要进行进一步的探索。

## 参考文献

- [1]Alpen A.Techniques for algorithmic composition of music.1995[EB/OL]. <http://alum.hampshire.edu/~adaF92/algocomp/algocomp95.html>.
- [2]冯寅,周昌乐.算法作曲的研究进展[J].软件学报,2006,(2):209-215.
- [3]高济,朱森良,何钦铭.人工智能基础[M].北京:高等教育出版社,2002:249.
- [4]沙汉昆.旋律写作教程[M].厦门:厦门大学出版社,2006,(2).
- [5]童忠良.基本乐理教程[M].上海:上海音乐出版社,2001.

## 作者简介

卿武明(1983—),男,汉族,湖南永州人,昆明理工大学硕士研究生,主要研究方向:算法作曲;

黄青松,男,教授,昆明理工大学信息与自动化学院副院长,硕士生导师,主要研究方向:智能信息系统;

臧晓晗,男,硕士研究生,主要研究方向:软件工程。