

Matlab 遗传算法工具箱在非线性优化中的应用

陈广洲^{1,2}, 解华明², 鲁祥友²

(1. 合肥工业大学, 安徽 合肥 230009;
2. 安徽建筑工业学院, 安徽 合肥 230022)

摘 要: 投影寻踪是一种降维处理技术, 通过它可以将多维分析问题通过投影方向转化为一维问题分析。应用该法的关键在于寻求最佳投影方向, 这可以转化为一个复杂的非线性优化问题来进行解决。选取某地区大气环境质量评价的投影寻踪评价模型, 编制相关的目标函数和约束函数, 应用基于 MATLAB 的遗传算法和直接搜索工具箱进行优化求解。结果表明: 该工具箱在求解此类非线性优化问题上的有效性和方便性, 从而为各领域应用投影寻踪模型提供了强有力的优化工具。

关键词: 遗传算法; 非线性优化; 投影寻踪模型

中图分类号: TP301.6 文献标识码: A 文章编号: 1673- 629X(2008) 03- 0246- 03

Nonlinear Optimization Based on Genetic Algorithm Toolbox of Matlab

CHEN Guang-zhou^{1,2}, XIE Hua-ming², LU Xiang-you²

(1. Hefei University of Technology, Hefei 230009, China;
2. Anhui Institute of Architecture and Industry, Hefei 230022, China)

Abstract: Projection pursuit is a dimension- reduced data treatment method. It can change many dimensions' question into one dimension question by projection direction. It is difficult to search the best projection direction, but it can be solved by transforming a optimization question. By compiling object function and constraint function , a model of projection pursuit on air environmental quality assessment was solved based on genetic algorithm and direct search toolbox. Result indicated that the toolbox was efficient and practical to solve the above question especially other crossed researchers.

Key words: genetic algorithm; nonlinear optimization; projection pursuit model

0 引 言

投影寻踪方法最早出现于 20 世纪 60 年代末, 它是用来分析和处理高维观测数据, 尤其是非线性、非正态高维数据的一种新型统计方法^[1], 目前已广泛地应用于预测、模式识别、遥感分类、过程优化控制、图像处理等领域^[2]。应用该法的关键在于寻求最佳投影方向, 这可以转化为一个复杂的非线性优化问题来进行解决。然而, 传统的优化方法在处理多变量同时寻优时往往会陷入局部最优、早熟或提前收敛, 寻求不到真正的最优解。

针对以上弱点, 文中采用基于 Matlab 的遗传算法和直接搜索工具箱来求解此投影方向。

1 遗传算法

1.1 遗传算法原理

遗传算法是由美国 Michigan 大学的 John Holland 教授在 20 世纪 60 年代提出的, 它是模拟生物在自然环境中的遗传和进化过程而形成的一种自适应全局优化概率搜索算法。理论基础来源于生物的自然选择理论和基因遗传学原理。它是将问题域的可行解看作是群体的“染色体”, 将其置于问题的“环境”中, 根据适者生存的原则, 从中选择出适应环境的“染色体”进行复制, 即再生, 通过交叉、变异两种基因操作产生出新一代更适合环境的“染色体”群, 这样一代代不断改进, 最后收敛到一个最适合环境的个体上, 求得问题的最佳解。

遗传算法提供了一种求解复杂系统优化问题的模式, 它不依赖于问题的具体领域, 对问题的求解种类有很强的鲁棒性, 现已几乎渗透到从工程到社会科学的诸多领域, 广泛用于函数优化、组合优化、生产调度、机器学习、自动控制、图像处理和人工生命等领域^[3]。

收稿日期: 2007- 06- 09
基金项目: 安徽省高校青年教师科研资助项目(2006jq1160); 安徽省教育厅自然科学类一般资助项目(KJ2007B070)
作者简介: 陈广洲(1978-), 男, 安徽怀远人, 讲师, 博士研究生, 研究方向为资源、环境信息技术。

1.2 遗传算法的计算步骤

遗传算法与传统搜索算法不同,它以适应度函数为依据,通过对种群中的所有个体实施遗传操作,实现群体内个体结构重组的迭代过程随机搜索算法。选择、杂交、变异构成遗传算法的三个主要遗传操作因子。其主要计算步骤如下^[4]:

(1) 选择一定数目的个体构成初始种群,并求出种群内各个个体的函数值。

(2) 设置迭代代数,开始循环。

(3) 计算选择函数的值:选择即通过概率的形式从种群中选择若干个个体的方式。遗传算法工具箱中提供了三个选择函数:轮盘选择算法、归一化几何选择方法、锦标赛选择方法。工具箱默认为采用归一化几何选择方法。

(4) 通过染色体个体基因的复制、交叉、变异等创造新的个体,构成新的种群。

(5) 进行循环,若终止条件不满足,则转到步骤(3)继续进化。

1.3 遗传算法的特点

与传统的优化算法相比,遗传算法主要有以下几个特点^[3,4]:

(1) 不是从一个点开始搜索最优解,而是从一个种群开始进行并行搜索,搜索轨道有多条,具有良好的并行性,因而可以有效地防止搜索过程收敛于局部最优解,更利于搜索到全局最优解。

(2) 遗传算法只需利用目标函数的取值信息,无需梯度等高价信息和其它信息,对问题的依赖性小,因而适用于大规模、非线性函数优化以及无解析表达式的目标函数的优化,有较广的应用性。

(3) 遗传算法的择优机制是一种“软决策”,加上其良好的并行性,使它具有良好的全局优化性能和稳健性。

(4) 遗传算法采用概率搜索技术,是一种不确定性的搜索方法,不同于传统的确定性的搜索方法,更有利于得到最优解。

2 Matlab 的遗传算法与直接搜索工具箱

遗传算法在应用过程中必须要编制大量的程序进行优化计算,因而对不同的专业人员应用该方法带来不便。Matlab 语言是一种高效率的用于科学工程计算的高级语言,它的语法规则简单,更贴近人的思维方式,通俗易懂,简便易学。尤其是 Matlab 软件有着丰富的各种专业的工具箱供各专业应用。遗传算法优化工具箱就是其中之一。其中,比较有名的工具箱有英国 Sheffield 大学 Peter Fleming 教授等开发的遗传算法

工具箱;另一个是美国北 Carolina 州立大学 Christopher Houck 等开发的工具箱 GAOT。此外,在 Matlab7.0 版本中提供了一个遗传算法与直接搜索工具箱(gads),可以用命令行调用或用图形用户界面(GUI)进行使用,也可以添加自己编制的函数进一步增加功能。在后续新的版本中,又进一步增加了该工具箱功能,添加了对约束条件的处理,只需编制相应的约束函数的 M 文件即可,使其应用功能进一步增强。基于 Matlab 的遗传算法与直接搜索工具箱的推出,为不同的专业领域应用遗传算法解决问题提供了更加方便和简洁的操作。

2.1 应用步骤

应用步骤为^[4~6]:

(1) 根据实际问题,确定变量的维数,编写待优化目标函数的 M 文件和约束条件的 M 文件,并用句柄指向该函数;

(2) 设置变量的上下界范围,线性等式约束条件矩阵,线性不等式约束条件矩阵,遗传算法运行的相关参数及算法终止的判断条件;

(3) 选择具体的运行方式:通过命令行调用函数 ga 或通过图形用户界面 GUI(gatool)使用。

2.2 系统参数的选取所遵循的原则

(1) 种群数目 $N^{[7]}$: 种群数目 N 会影响 GA 的有效性。 N 太小,GA 计算结果会很差或根本找不出问题的解。因为太小的种群数目不能提供足够的采样点; N 太大,会增加计算量,使收敛时间延长。一些文献中总结为一般种群数目在 20 ~ 160 之间比较合适。

(2) 交叉概率 $P_c^{[7]}$: 此参数控制着交叉操作的频率。 P_c 太大,会使高适应值的结构很快破坏掉; P_c 太小,搜索会停滞不前。一般 P_c 取 0.25 ~ 0.75。

(3) 变异概率 $P_m^{[7]}$: 它是增大种群多样性的一个因素。 P_m 太小,不会产生新的基因; P_m 太大,会使 GA 变成随机搜索。一般 P_m 取 0.01 ~ 0.20。

3 实例计算

当前,投影寻踪模型在不同专业领域的评价、聚类、回归分析等方面中得到了广泛的应用。以投影寻踪评价为例,应用该方法的关键在于:根据某一评价标准,找出一个最佳的投影方向,然后将待评价对象以此向量作基准,计算综合值,比较大小。欲求一个最佳的投影方向,可以将问题转化为求一个满足约束条件的复杂非线性函数的最大值问题^[1]: 设 A 为某一问题的 m 维的评价标准所对应的标准化向量,则评价标准第 i 个分级的投影值 $Z(i)$ 的表达式为:

$$Z(i) = \sum_{j=1}^n A(j) * X(i, j)$$

其中: $A(j)$ — 投影方向单位向量($j = 1, 2, \dots, n$)。 $X(i, j)$ — 第 i 个分级标准的第 j 个指标的标准化值($i = 1, 2, \dots, m; j = 1, 2, \dots, n$)。即构造一个投影指标函数 $Q(A)$ 作为确定投影方向优化的依据, 当指标达到极大值时, 就认为是找到了最优投影方向。在优化投影值时, 要求 $Z(i)$ 的分布特征应满足: 投影点局部尽可能密集, 在整体上尽可能散开。因此, 投影指标函数为: $Q(A) = S_z * D_z$, 式中: S_z — 一类间散开度, 可用 $Z(i)$ 的标准差代替; D_z — 一类内密集度, 可表示为 $Z(i)$ 的局部密度。其中:

$$S_z = \left\{ \sum_{i=1}^m [Z(i) - Z]^2 / (m - 1) \right\}^{0.5}$$

$$D_z = \sum_{i=1}^m \sum_{j=1}^m (R - r_{ij}) * I(R - r_{ij})$$

Z — 序列 $\{Z(i) | i = 1, \dots, m\}$ 的均值; R 是由数据特征确定的局部宽度参数, 其值一般可取 $0.1 * S_z$, 当点间距值 r_{ij} 小于或等于 R 时, 按类内计算, 否则按不同的类记; $r_{ij} = |Z(i) - Z(j)|$; 符号函数 $I(R - r_{ij})$ 为单位阶跃函数, 当 $R \geq r_{ij}$ 时函数值取 1, 否则取

0。对 $A(j)$ 的约束为 $\sum_{j=1}^n a_j^2 = 1$ 。

选取某地区大气环境质量评价分级问题的评价标准 $A = [0.033 \ 0.160 \ 0.067 \ 0.330; 0.330 \ 0.500 \ 0.500 \ 0.660; 1.000 \ 1.000 \ 1.000 \ 1.000]$, 进行投影寻踪模型优化。通过编制目标函数 myobj.m 文件和约束函数 cons.m 文件, 为转化为求极小值并保证函数值为正, 在原目标函数的取反的基础上再加一个常数 100, 调用 gatool, 在 GUI 中完成相关参数设置, 进行计算。该目标函数具有通用性, 只需要更换标准矩阵 A 、变量维数和所加的常数值, 即可用于其它投影寻踪问题的目标函数的构建, 同理, 约束条件中的变量维数也相应改变即可。目标函数和约束函数具体如下:

```
function f= myobj(x)
```

```
A= [0.033 0.160 0.067 0.33; 0.330 0.5 0.5 0.66; 1 1 1 1]; [m, n]= size(A); sum1= 0;
```

```
for j= 1: m
```

```
for i= 1: n
```

```
z(i)= x(i). * A(j, i);
```

```
sum1= sum1+ z(i);
```

```
end
```

```
w(j)= sum1;
```

```
sum1= 0;
```

```
end
```

```
za= (w(1)+ w(2)+ w(3))/3;
```

```
s= sqrt(((w(1)- za).^2+ (w(2)- za).^2+ (w(3)- za).^2))
```

```
/(m- 1));
```

```
su= 0;
```

```
for i= 1: m
```

```
for j= 1: m
```

```
rij= abs(w(i)- w(j));
```

```
dd= (0.1). * s- rij;
```

```
if dd > 0
```

```
su= su+ dd;
```

```
d= su;
```

```
dd= 0;
```

```
else
```

```
d= su;
```

```
end
```

```
end
```

```
end
```

```
f= -(s* d)+ 100;
```

```
function [c, ceq]= cons(x)
```

```
c= [];
```

```
ceq= [1- sum(x.^2)];
```

种群数为 20, 变异采用自适应方法, 其他采用工具箱默认设置。为了得到遗传算法的最好结果, 一般需要以不同的参数试验。通过多次试验, 选择针对问题的最佳参数。经过对多种参数进行组合试验, 结果表明: 参数的改变对计算结果影响较小, 计算结果相对稳定, 选取该问题的寻优值为: $x = [0.5689 \ 0.4881 \ 0.5382 \ 0.3853]$, 以此为最佳投影方向, 计算样本的综合值, 并结合该标准的综合值分级情况, 即可确定样本所属评价类别。

4 结 语

遗传算法具有通用性、智能性、鲁棒性、全局性和并行性的特点, 函数优化是遗传算法最常应用的领域之一。文中在 Matlab 环境下, 使用遗传算法与直接搜索工具箱对某一大气环境质量分级中投影寻踪模型中的最佳向量进行优化求解。结果表明: 对于难以定义或不便于进行数学建模优化的问题, 遗传算法具有很好的适用性(它不需要对目标函数的相关信息有所了解), 同时有较快的收敛速度和较高的计算精度, 基于 Matlab 的遗传算法与直接搜索工具箱在解决复杂函数优化问题上具有较强的有效性和方便性。

参考文献:

- [1] 金菊良, 刘永芳, 丁 晶, 等. 投影寻踪模型在水资源工程方案优选中的应用[J]. 系统工程理论方法应用, 2004, 13(1): 81- 84.
- [2] 付 强, 赵小勇. 投影寻踪模型原理及其应用[M]. 北京:

(下转第 252 页)

表 2 合肥市 1996~ 2000 年按 $[f]_{\text{旬}}$ 的
总产量预测结果(单位: 公斤 / 公顷)

年份	趋势产量 GM(1, 1)模型	预测总产量	实际总产量
1996	2891. 4	(3093. 8, 3382. 9)	3255. 9
1997	3380. 2	(3616. 8, 3954. 8)	3572. 1
1998	3125. 8	(3344. 6, 3657. 2)	3079. 1
1999	2933. 4	(2816. 1, 3138. 7)	2766. 9
2000	3374. 8	(2767. 3, 3239. 8)	2839. 9

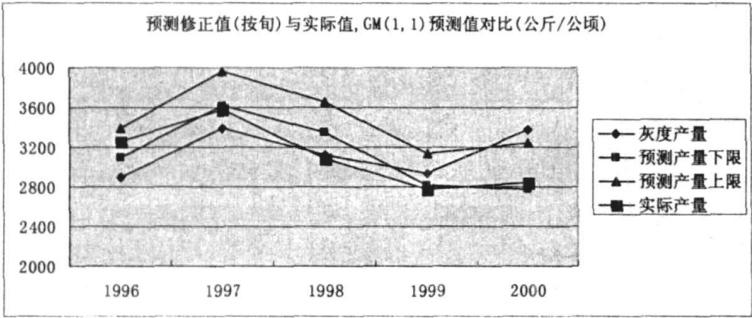


图 3 合肥市 1996~ 2000 年按 $[f]_{\text{旬}}$ 产量预测结果比较
间的变化从总体趋势上和实际产量还是保持一致, 相对表 1 来说, 虽然 1997, 1998, 1999 年的修正区间未能包含实际产量, 1997, 1999 年的涨跌趋势与实际产量还是一致的, 只有 1998 年实际产量有一定下降, 而预测修正区间的下降程度不够明显, 这也恰恰是粒度变粗后信息自然会丢失一部分的表现。在面对样本粗粒度学习时, 人们只能获得样本的大概信息, 只有将粒度不断细化, 由粗到细, 才能在学习过程中获得样本更多的隐含信息, 使样本信息更加具体化。

4 结 论

文章对粮食产量预测, 提出在商空间粒度计算思想下, 构造性学习算法为工具, 将影响冬小麦产量的气象因子作为特征属性, 冬小麦产量的趋势(由相对气象产量得到) 作为样本的决策属性, 即将降水量、日照、温度的不同粒度表示作为样本的特征属性, 将小麦的相对气象产量划分成几个程度, 做丰歉平, 作为样本的决策属性。通过根据对学习样本特征属性的学习及决策属性的判断, 得出分类规则, 进而对给出的只有特征属性的测试样本, 可以根据此分类规则, 得出测试样本的

决策属性, 经过还原后得出预测值的修正区间, 通过实验比较显示该方法相对常用的 GM(1, 1) 模型预测值更加精确、可靠。

参考文献:

[1] 王世耆, 程延年. 作物产量与天气气候[M]. 北京: 科学出版社, 1991.

[2] 邓聚龙. 灰色系统理论教程[M]. 武汉: 华中理工大学出版社, 1990.

[3] 周志坚, 傅泽田, 王瑞梅, 等. 灰色- 马尔柯夫模型在棉花产量预测中的应用[J]. 决策参考, 2006(2): 48- 49.

[4] 尹爱芹, 郭小春, 郭化文. 马尔柯夫链在天气预报中的应用[J]. 山东农业大学学报: 自然科学版, 2001, 32(2): 191- 196.

[5] 毕守东, 王冬平. 安徽省粮食产量的最优加权组合预测[J]. 预测, 2000(3): 70- 72.

[6] Zhang Ling, Zhang Bo. The Quotient Space Theory of Problem Solving[J]. Fundamenta Informaticae, 2004, 59: 287- 298.

[7] Zhang Ling, Zhang Bo. Theory and Application of Problem Solving[M]. Beijing: Tsinghua University Publisher, 1990.

[8] 张 铃, 张 钹. M- P 神经元模型的几何意义及其应用[J]. 软件学报, 1998, 9(5): 334- 338.

[9] 张 铃, 张 钹, 殷海风. 多层前向网络的交叉覆盖设计算法[J]. 软件学报, 1999, 10(7): 737- 742.

[10] 张淑娟, 何 勇. 基于趋势- 状态预测方法的粮食产量预测[J]. 浙江大学学报: 农业与生命科学版, 2001, 27(6): 673- 676.

[11] 张 力, 张保华. 冬小麦产量分析[J]. 中国农业气象, 2004, 25(1): 22- 24.

[12] 范锦龙, 孟庆岩, 吴炳方, 等. 基于农业气象模型的农作物单产预测系统[J]. 中国农业气象, 2003, 24(2): 46- 48.

[13] 李 馨, 战守义, 史宁中. 一种有效的农业气象灾害的定量性定义方法[J]. 计算机工程与应用, 2003(11): 203- 209.

[14] 陈怀亮, 王良宇, 杜明哲, 等. 产量阶段的划分及应用[J]. 中国农业气象, 1999, 20(2): 16- 20.

[15] 黄嘉佑. 气象统计分析与预报方法[M]. 北京: 气象出版社, 1990: 265- 266.

[16] 张燕平, 张 铃, 吴 涛. 不同粒度世界的描述法——商空间法[J]. 计算机学报, 2004(3): 41- 46.

(上接第 248 页)

科学出版社, 2006.

[3] 周 明, 孙树栋. 遗传算法原理及应用[M]. 北京: 国防工业出版社, 1999.

[4] 薛定宇, 陈阳泉. 高等应用数学问题的 matlab 解法[M]. 北京: 清华大学出版社, 2004.

[5] 雷英杰, 张善文, 李续武, 等. Matlab 遗传算法工具箱及应用[M]. 西安: 西安电子科技大学出版社, 2005.

[6] 周琛琛. 基于 matlab 遗传算法工具箱的函数优化问题求解[J]. 现代计算机, 2006, 12: 84- 86.

[7] 周正武, 丁同梅, 田毅红, 等. Matlab 遗传算法优化工具箱(GAOT) 的研究与应用[J]. 机械研究与应用, 2006, 6(19): 69- 71.