

基于灰色-马尔可夫模型的粮食产量预测

刘安¹, 赵姝², 张燕平²

(1. 安徽大学 计算机科学与技术学院, 安徽 合肥 230039;

2. 安徽大学 计算智能与信号处理教育部重点实验室, 安徽 合肥 230039)

摘要:文中首先用1976年到1995年的桐城县冬小麦单产,建立灰色GM(1,1)模型,再用随机过程理论的马尔可夫模型获得GM(1,1)模型在已知年份里的偏差规律(即偏差的转移概率矩阵),并且依照此规律对GM(1,1)模型结果进行修正,将由GM(1,1)模型预测的一个具体数值,修正成为区间和概率组成的预测范围,增加预测的可靠性。最后用灰色-马尔可夫模型外推预测1996年到2000年共5年的小麦单产。实验说明灰色-马尔可夫模型大大提高了预测精度,将预测结果表示为预测范围,更为准确地反映出粮食产量的走势。

关键词:冬小麦单产;GM(1,1)模型;马尔可夫模型;灰色-马尔可夫模型

中图分类号:O211.62

文献标识码:A

文章编号:1673-629X(2007)06-0191-03

Yield Forecast Based on Grey - Markov Model

LIU An¹, ZHAO Shu², ZHANG Yan-ping²

(1. College of Computer Science & Technology, Anhui University, Hefei 230039, China;

2. Ministry of Education Key Lab. of Intelligent Computing & Signal Processing,
Anhui University, Hefei 230039, China)

Abstract: GM(1,1) model is established by using winter wheat yield of Tong Cheng county (from 1976 to 1995). Then the deviation regularity of GM(1,1) model from 1976 to 1995 is obtained by Markov model of stochastic process (i. e. transition probability matrix). To improve forecast reliability, the result of GM(1,1) model is amended based on the regularity. The result of GM(1,1) which is just a numerical value is changed to a range consisted of intervals and probabilities. At last, winter wheat yield of Tong Cheng county (from 1996 to 2000) are forecasted by grey - Markov model. In this way, the forecast precision is improved, and the trend of yield is reflected accurately.

Key words: winter wheat yield; GM(1,1) model; Markov model; grey - Markov model

0 引言

中国是一个农业大国,农业的稳定是社会稳定的基础。而粮食产量预测正是保证农业稳定的重要前提。如今,计算机实现粮食产量预测有多种方法,每种方法都各有优缺点。因此,如何能综合利用不同方法,扬长避短,就显得十分重要。GM(1,1)模型通常用来揭示宏观发展规律,而马尔可夫模型一般用于探寻微观波动规律。文中尝试在GM(1,1)模型基础上,进一步运用马尔可夫模型对其结果进行优化,即结合两模

型的优点,避免各自缺陷,用灰色-马尔可夫模型预测粮食产量,从而提高预测精度。

1 灰色-马尔可夫模型的基本思想

1.1 灰色GM(1,1)模型的基本思想

1.1.1 GM(1,1)模型的建立

GM(1,1)模型是对离散的原始数据序列进行 m 次累加,生成规律性强的累加生成序列,从而弱化原始序列的随机性。然后对累加生成序列建模,最后进行 m 次累减,还原成预测值。一般取 $m=1$,做一次累加生成序列建模^[1]。具体建模如下^[2]:

设 $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$ 为原始数据序列。

(1) 记 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k=1, 2, 3, \dots, n$ 为 $x^{(0)}$ 的1-AGO序列。

收稿日期:2006-08-14

基金项目:安徽省自然科学基金(050420208);安徽省教育厅重点自然科学基金项目(2006KJ015A)

作者简介:刘安(1984-),男,江苏南京人,本科生,研究方向为人工智能;张燕平,硕士生导师,教授,研究方向为计算智能、神经网络。

(2) 再记 $z^{(1)}(k) = \lambda x^{(1)}(k) + (1 - \lambda)x^{(1)}(k - 1), k = 2, 3, \dots, n$, 为 $x_1^{(1)}$ 的紧邻均值生成序列, 其中 $\lambda \in (0, 1)$ 。

(3) 那么 GM(1,1) 模型的方程为

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u$$

微分方程系数 a 和 u 组成的向量为 $\bar{a} = [a, u]^T = (B^T B)^{-1} B^T Y$, 其中

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$
$$Y = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)]^T$$

(4) GM(1,1) 微分方程的解为:

$$\hat{x}^{(1)}(t+1) = [x^{(0)}(1) - \frac{u}{a}] \cdot e^{-at} + \frac{u}{a}, t = 1, 2, 3, \dots, n$$

拟合公式为:
 $\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t), t = 1, 2, 3, \dots, n$

1.1.2 灰色模型的改进方法

如果原始数据非常平滑, 则令 $\lambda x_1^{(1)}(k) + (1 - \lambda)x_1^{(1)}(k - 1)$ 中的 $\lambda = 0.5$; 如果数据波动但变动不大, λ 的取值范围一般为 $(0, 0.5)$, 即旧数据权重大; 反之, λ 取值范围为 $(0.5, 1)$, 即新数据权重大。选择合适的背景值可以在一定程度上提高模型精度^[3]。

1.2 对 GM(1,1) 模型结果用马尔可夫链改进的基本思想

GM(1,1) 通常用来揭示数据的发展趋势, 并不适合波动性较大的数据序列预测。而马尔可夫随机性过程理论则通常用来确定状态的转移规律, 适合描述波动性较大的数据序列^[4]。因此结合两模型, 建立灰色马尔可夫预测模型, 可以提高预测精度。

1.2.1 马尔可夫链的基本理论

一个 n 阶马尔可夫链由 n 个状态集合和一组转移概率所确定。该过程的任一时刻只能处于一个状态。如果在时刻 t , 过程处于状态 S_j , 则在 $t + 1$ 时刻它将以概率 p 处于状态 S_k 。其预测是根据状态之间的转移概率来推测系统未来的发展变化^[5]。

一步转移概率的公式: $P_{jk} = M_{jk} / M_j$ 。
其中 P_{jk} 为状态 S_j 经过一步转移到状态 S_k 的一步转移概率, M_{jk} 为从状态 S_j 经过一步转移到状态 S_k 的次数, M_j 为状态 S_j 出现的次数^[6]。

对于高阶转移概率矩阵, 一般采用递推公式计算, 即:

$$P_{jk}^{(r+1)} = \sum_y P_{jy}^{(1)} * P_{yk}^{(r)}, (y = 1, 2, \dots, n)$$

多阶转移概率矩阵为:

$$P = \begin{bmatrix} P_{11}^{(i)} & \dots & P_{1n}^{(i)} \\ \vdots & & \vdots \\ P_{n1}^{(i)} & \dots & P_{nn}^{(i)} \end{bmatrix} \quad (i = 1, 2, \dots, r)$$

1.2.2 马尔可夫模型建立的方法和步骤

- (1) 状态的划分。
- (2) 计算转移概率并建立转移概率矩阵。
- (3) 利用转移概率矩阵预测状态转移^[7]。

1.2.3 灰色模型与马尔可夫链结合的基本思路

对于 GM(1,1) 模型得到的预测结果, 可以根据马尔可夫链的方法获得 GM(1,1) 模型在已知年份里的偏差规律(即偏差的状况转移矩阵), 并且依照此规律对 GM(1,1) 模型结果进行修正。由 GM(1,1) 模型结果的一个预测数值, 修正成为区间和概率组成的预测范围, 增加预测的可靠性^[8]。

两种模型的结合可以充分利用历史数据给予的信息。用灰色 GM(1,1) 模型预测曲线来反映系统宏观发展规律, 并以此预测曲线为基准, 再用马尔可夫模型来寻求系统的微观波动规律, 因而对随机波动较大的数据序列的预测具有较高的精度。

2 应用灰色 - 马尔可夫模型的实例分析

用桐城县 1976 到 1995 年共 20 年冬小麦单产 (kg/hm^2) 来建立模型, 再外推预测 1996 到 2000 年 5 年的小麦单产 (kg/hm^2)。

2.1 建立 GM(1,1) 模型

建模过程中, 紧邻均值序列生成公式中的参数 $\lambda = 0.75$; 建立的桐城县 20 年冬小麦单产 GM(1,1) 模型为:

$$\hat{x}^{(1)}(t+1) = [x^{(0)}(1) + 110660] \cdot e^{0.01568t} - 110660$$
$$t = 1, 2, 3, \dots, n$$

以此模型求解 1977 到 1995 年的冬小麦单产拟合值, 拟合值与实际单产的对比见图 1。

2.2 建立灰色 - 马尔可夫模型

2.2.1 状态的划分

根据 GM(1,1) 模型求得的 19 年拟合值的相对误差分布建立分级标准。文中的状态分级标准及各分级状态数见表 1。

表 1 根据相对误差 % 建立分级标准 (共 9 状态) 和各分级的状态数

>20	(10, 20)	(6, 10)	(1, 6)	(-1, 1)	(-6, -1)	(-10, -6)	(-20, -10)	<-20
0	3	5	2	2	2	2	2	2

2.2.2 计算转移概率并建立转移概率矩阵

文中用 1 至 4 步转移概率修正拟合值。对 5 年的

预测值也都考虑 4 步次转移概率,所以需建立 1 至 8 阶转移概率矩阵。

2.2.3 利用转移概率矩阵对 GM(1,1)模型的结果进行修正

以对 1995 年拟合值修正为例:根据 1994 至 1991 年的状态(3,7,8,8)查 1 至 4 阶转移概率矩阵,得到 1995 年 1 至 4 步的状态转移及相应的转移概率,见表 2。综合 4 步次转移概率,1995 年产量最有可能是状态 2,其修正值范围为(2611.9,2938.4)。实际小麦单

表 2 1995 年马尔可夫修正区间

年份	原始数据	GM(1,1)	1 步概率	2 步概率	3 步概率	4 步概率	修正值范围
1995	2642.5	2350.7	0.4	0.5	0.5	1	2611.9~2938.4
			0.2	0	0.5	0	2500.7~2611.9
			0	0.5	0	0	2374.4~2500.7
			0.2	0	0	0	2217.6~2327.4
			0.2	0	0	0	2137~2217.6

产为 2642.5,正好包含在修正值范围的区间内,对比 GM(1,1)拟合值,提高了精度。

用马尔可夫模型对 GM(1,1)拟合值修正前后的比较见图 1。

由图 1 可知 GM(1,1)模型建立的趋势产量可以揭示冬小麦产量的宏观发展规律(平均相对误差为 10.627%)。但也可以看出,GM(1,1)模型拟合值曲线单调上升,无法反映冬小麦单产的波动性,拟合精度和可靠性不高。相比之下,马尔可夫修正区间(最大概率)曲线不仅能揭示冬小麦单产的宏观规律,同时也能很好地反映冬小麦单产的微观波动规律,从而形象地反映了产量的走势,拟合精度和可靠性有很大提高。至此,完成灰色 - 马尔可夫模型建模。

2.3 利用灰色 - 马尔可夫模型外推预测 5 年的冬小麦单产

灰色 - 马尔可夫模型的预测结果与前述的 95 年类似,灰色 - 马尔可夫模型的预测值与 GM(1,1)预测值、实际单产的对比见图 2。

考察图 2,96 年、97 年的马尔可夫修正区间(最大概率)都能准确包含实际单产。1998 年产量有较大下滑,GM(1,1)的预测值偏差较大,而马尔可夫修正区间(最大概率)虽未能包含实际单产,但仍准确地预测了产量的下跌趋

势,且偏差也小于 GM(1,1)预测值。由此可知,灰色 - 马尔可夫模型得出的预测区间,能同时较好地反映冬小麦单产的宏观增长和微观波动,即使未能精确地包含实际单产,也能准确地显示冬小麦单产的涨跌趋势,相比 GM(1,1)模型(预测值平均相对误差 9.4837%)提高了预测精度和可靠性;特别是外推预测 1 到 2 年冬小麦单产,精度较高。

3 结 论

文中用 GM(1,1)模型预测曲线来反映系统宏观发展规律,建立小麦单产的趋势值;再利用随机过程理论的马尔可夫链寻求系统的微观波动规律,对 GM(1,1)的趋势值进行修正,以适应粮食产量的波动性。两种模型的结合可以充分利用历史数据给予的信息。

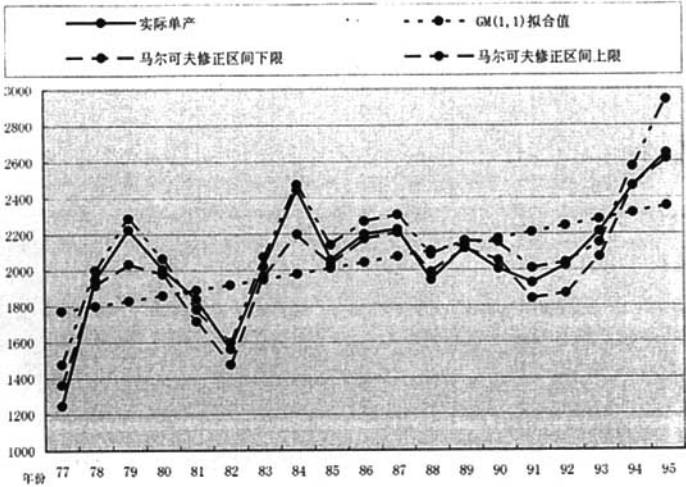


图 1 马尔可夫修正区间(最大概率)与 GM(1,1)拟合值、实际单产的对比

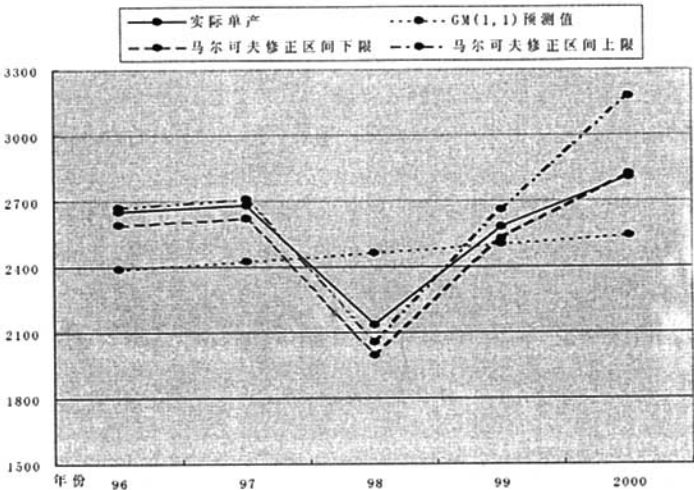


图 2 马尔可夫修正区间与 GM(1,1)预测值、实际单产的对比

是染色体的长度,这 m 个交叉点是通过随机数选择的、没有重复、按升序排列的。父染色体中在两个相连的交叉位间的基因进行交换,产生两个新的子代,个体第一个基因位到第一个交叉点之间的位并不进行交换。

2.5 染色体的变异

在自然变化中,变异是一种随机的过程,是基因上的一个等位基因被另一个替代而产生的新的遗传结构,在遗传算法中,变异采用任意的小概率,典型值在 0.001~0.1 之间。在本问题求解过程中采用 2-交换作为变异算子,即在每代群体中随机选取染色体上的两点,然后将选定的两点的基因码进行交换,从而完成变异。

3 实验计算与结果分析

为了验证文中针对带中转点的联盟运输调度所设计的遗传算法的实际运行效果,以 15 个需求点,8 个中转点为例进行了求解。问题的具体描述如下:某城市中心区内有 15 个需求点需要配送服务,在该城市中心区和外郊区之间共有 8 个中转点可以提供足够的中转服务,应用文中设计的遗传算法为本次配送服务选取最优的中转服务点。实例计算所需要的基本数据如表 1、表 2 所示。

表 1 15 个需求点的需求量

需求点	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14	Z15
需求量	29	58	34	56	78	90	34	56	78	90	34	23	56	78	43

表 2 供货点到中转点的单位运输成本

中转点	X1	X2	X3	X4	X5	X6	X7	X8
单位运输费用	12	15	17	16	13	14	16	14

以 Matlab6.5 为平台,在 sheffield 大学开发的遗传算法工具箱的基础上,对该实例进行了仿真求解,各项参数设置如下:

个体数量 NIND=40;最大遗传代数 MAXGEN=50;使用代沟为 GGAP=0.9;染色体的长度为 15,交叉概率为 $P_c = 0.5$,变异概率为 $P_m = 0.1$ 。经优化计算可得如下结果:即最优中转点分配方案,如表 3 所示。

表 3 分配方案表

需求点编号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
中转点编号	5	7	7	3	7	4	1	7	4	7	6	3	8	1	4

4 结束语

中转点的优化确定是带中转的联盟运输调度问题(AVRP)的核心问题之一。文中在建立中转点优化确定数学模型的基础上,构造了求解该问题的遗传算法,实验的结果表明该算法简单有效,但由于运输调度问题还可以衍生出很多更为复杂也更贴近实际应用的问题^[6],如多个供应点、多车队、时间窗、多重运输需求、多种运输环节、多重交通网络、多重约束条件、客户需求多样性等,这些都是带中转点联盟运输调度问题必须综合考虑的问题。也是今后进一步研究的方向。

参考文献:

- [1] 师 凯,蔡延光,邹谷山,等.分段蚁群算法在运输调度问题中的应用[J].广东工业大学学报,2006,23(1):72-76.
- [2] 朱 强,卜 雷.城市货物换装站非约束选址模型及其遗传算法[J].华南理工大学学报,2005,33(7):92-95.
- [3] 吴 坚,史忠科.基于遗传算法的配送中心选址问题[J].华南理工大学学报,2004,32(6):71-74.
- [4] 蒋忠中,汪定伟.B2C 电子商务中配送中心选址优化的模型与算法[J].控制与决策,2005,20(10):1125-1128.
- [5] 王小平,曹立明.遗传算法——理论、应用与软件实现[M].西安:西安交通大学出版社,2005.
- [6] 蔡延光,钱积新,孙优贤.多重运输调度问题的模拟退火算法[J].系统工程理论与实践,1998(10):11-15.

(上接第 193 页)

灰色-马尔可夫模型的预测结果是由区间和相应转移概率组成的预测范围,而不是具体数值。以区间和概率组合的方式提高了预测结果的可靠性。

用文中的思路还可以对其它既具有趋势性又有一定波动性的对象进行模拟和预测。

参考文献:

- [1] 刘思峰,邓聚龙.GM(1,1)模型的适用范围[J].系统工程理论与实践,2000,20(5):121-124.
- [2] 李 群,潘晨光.高精度灰色模型研究及 2005 年 GDP 总量

预测[J].财经问题研究,2005(8):11-13.

- [3] 罗 党,刘思峰,党耀国.灰色模型 GM(1,1)优化[J].中国工程科学,2003,5(8):50-53.
- [4] 王梓坤.随机过程论[M].北京:科学出版社,1965.
- [5] 李裕奇.随机过程[M].北京:清华大学出版社,2002.
- [6] 孟宪宏,宋玉普.灰色马尔可夫链在混凝土疲劳寿命预测中的应用[J].混凝土,2004(2):3-13.
- [7] 赵利民,周西利.马尔可夫链在大白菜年景预报中的运用[J].西北农业学报,2003,12(4):139-142.
- [8] 陈有孝,林晓言.灰色-马尔可夫链改进的预测方法[J].统计与决策,2005(16):36-38.