

基于商空间理论的冬小麦产量预测和分析

杨雪洁, 赵 姝, 张燕平

(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:文中从粒度计算角度对气象数据和历史产量进行分析。首先在商空间粒度计算理论框架下,通过对历年光、温、水等环境气象要素的分析,得出商空间模型下的特征属性;再通过对历史产量的分析,运用统计学方法获取相应的决策属性,将特征属性和决策属性相结合采用构造性机器学习方法对安徽省冬小麦产量进行预测,建立产量预测模型以实现冬小麦的较为准确的产量预测和分析。实验证明这种基于商空间粒度计算理论的产量预测方法是行之有效的。

关键词:粒度计算;商空间;机器学习;产量预测;灰色模型

中图分类号: TP399

文献标识码: A

文章编号: 1673-629X(2008)03-0249-04

Forecast and Analysis on Winter Wheat Yield Based on Quotient Space Theory

YANG Xue-jie, ZHAO Shu, ZHANG Yan-ping

(Ministry of Education Key Laboratory of Intelligent Computing & Signal Processing, Anhui University, Hefei 230039, China)

Abstract: Puts forward analysis meteorological data and historical yield by granular computing methods. Firstly, based on the quotient space theory, the characteristic attributes can be gotten by the analysis of the historical data on light, temperature and precipitation. Secondly, the statistical algorithm makes the decision-making attributes by analyzing the historical yield data. The characteristic and decision-making attributes are combined to build the yield forecast model and forecast winter wheat yield by constructive machine learning methods. The experimental results show that using this method to forecast and analyse wheat yield is effective.

Key words: granular computing; quotient space; machine learning; yield forecast; GM

0 引言

从粮食产量预报研究与应用领域多年的经验来看,使用的在作物产量预报方法中既简便又有一定精度的仍属统计预报途径^[1],如指数平滑、马尔柯夫链、灰色模型等。

(1)指数平滑法:利用平滑平均数的计算对时间序列进行修匀的一种方法。它对过去的的数据分别加以不同的权数,而且更重视近期的数据。

(2)灰色系统预测:灰色系统预测是根据过去及现

在已知的或非确知的信息,建立一个从过去引伸到将来的灰色模型^[2],简称 GM 模型,广泛使用的是 GM (1,1)模型(一元一阶模型)。灰色预测主要用于时间中短期、数据资料少、波动不大的预测问题。

(3)灰色马尔柯夫链^[3]:马氏链预测的对象是一个随机变化的动态系统,它的理论基础是马尔柯夫过程,适合于随机波动性较大的预测问题。因此将灰色模型与马尔柯夫链^[4]结合起来使用,预测效果优于单纯用灰色模型。

(4)最优加权组合预测^[5]:若将几种预测方法组合成一个预测模型,即用多个预测模型的有效组合或多个因素的科学综合,有可能比较合理地描述系统的客观现实。

粮食产量的形成过程是典型的灰色系统,使用灰色模型预测可以在一定程度上反映粮食产量的变化规律,但粮食产量是随机波动性较大的数据列,而灰色模型和指数平滑法对数据曲线平滑时预测效果较好,最优加权组合预测虽然能改善预测效果,但其组合系数

收稿日期:2007-06-09

基金项目:国家自然科学基金(60475107;60675031);973 计划资助项目(2004CB318108);教育部博士基金资助项目(20040357002);安徽省教育厅重点自然科学基金资助项目(2006KJ015A);安徽省自然科学基金资助项目(0504200208);安徽省教育厅自然科学基金资助项目(2005kj053);安徽大学 211 工程学术创新团队资助项目

作者简介:杨雪洁(1982-),女,安徽巢湖人,硕士研究生,研究领域为智能计算、人工神经网络;张燕平,教授,主要研究领域为人工智能、机器学习、人工神经网络的理论与应用研究。

不易确定,同时随着农业市场化的不断发展,对农业气象服务信息和产量预测信息的要求越来越高,原有的一些农业气象服务指标是 20 世纪 90 年代甚至是 80 年代总结形成的,已不适应现代农业气象服务的要求,因此需要研究新的预测方法。文中从一种新的角度对气象数据和历史产量进行数据分析。在商空间粒度计算理论框架下,用构造性机器学习算法与统计学方法相结合对冬小麦产量进行预测。

1 商空间粒度计算理论

商空间理论^[6]应用数学工具商空间模仿人脑的思维活动和认知过程,提出了一种处理不精确、不确定和不完全数据的新的智能计算模型。利用该模型能够有效地处理各种不确定信息,并从中发现隐含知识,揭示事物和事件的内在规律。

文献^[7]提出了用一个三元组 $\{X, f, T\}$ 来描述一个问题。 X 表示问题的论域, $f(x)$ 表示论域的属性,可用函数 $f(x): X \rightarrow Y$ 表示; T 是论域的结构,指论域 X 中各元素的相互关系。分析或求解问题 $\{X, f, T\}$ 是指对论域 X 及其有关的结构、属性进行分析、研究。

可以从不同的粒度来考察 $\{X, f, T\}$,按照上面的方法取一个粒度世界,商空间法就是将不同的粒度世界与数学上的商集概念统一起来的表示对象模型的方法。构成新的商空间模型不仅可以对论域 X 划分,也可以对属性 $f(x)$ 或结构 T 来划分等价关系,即构成商空间从论域、属性、结构三个方面都可以进行。

商空间粒度计算理论在冬小麦产量预测中的主要体现为气象因子的确定及论域(即样本)的选择中。整理安徽省历年的气象因素资料,对安徽省 78 个站点从建站以来每天的降水量、日照和温度各气象因子采用商空间粒度计算思想进行处理,从而发掘不同粒度的气象因子对冬小麦产量的影响,建立粮食产量与气象因子的商空间模型。

2 构造性学习算法(覆盖算法)

构造性学习算法主要用作解决分类学习问题,它需要对大量的样本进行学习,这些样本由特征属性和决策属性两大部分组成,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。这里具体采用的是覆盖算法。

根据 M-P 神经元^[8]的几何意义提出多层前向神经网络的覆盖算法^[9],它是根据样本数据自身的结构,构造性地建立神经网络模型。

3 冬小麦的产量预测

产量预测是以商空间粒度计算理论为指导思想,构造性学习算法为具体实现工具,如图 1 所示。

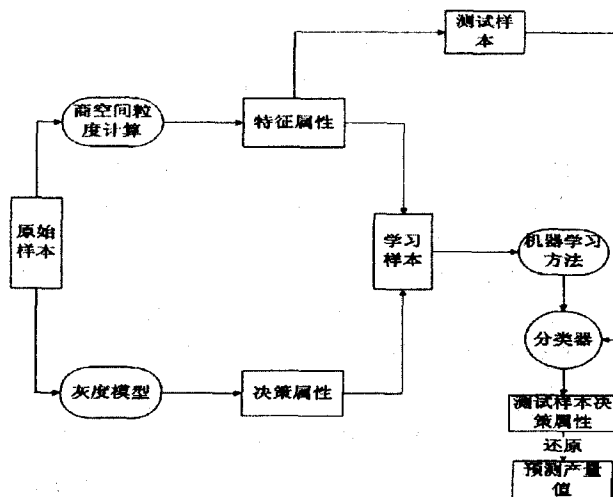


图 1 基于商空间理论的安徽省冬小麦产量预测模型

如图所示,在商空间粒度计算框架下,分析原始样本中的气象数据,将各气象数据的不同粒度表示作为样本的特征属性;对原始样本中的历史产量数据建立灰度模型,将小麦的相对气象产量划分成几个类别,作为样本的决策属性。采用覆盖算法对学习样本特征属性学习及决策属性判断,得出分类规则,建立分类器,进而对给出的只有特征属性的测试样本,根据此分类器,得出测试样本的决策属性,最后还原成预测产量值。

3.1 论域 X

原始样本分为历年的气象数据和产量数据两大类,其中原始气象数据是全省 78 个站点从建站以来每天的降水量、日照和温度。采用商空间粒度计算思想对原始样本进行处理,将其按论域进行颗粒化时,可以把结构上或功能上关系密切的元素划分为一类。因此可以按安徽省气候状况结合冬小麦的分布和产量的情况,将七十八个市分为淮北、沿淮、江淮、沿江、皖南山区、大别山区六个区域。或选取某些代表城市各自讨论其产量预测情况,现以合肥市为例。

由于合肥市建站时间是 1952 年,所给数据到 2000 年止,故原始样本只能从 1952 年取。所以将合肥市从 1953 年到 2000 年的 48 个样本作为论域。

3.2 决策属性

决策属性的确定基于对原始样本中产量数据的处理,目前一般在分析环境气象条件对农作物产量的影响以及进行产量的农业气象预报时,均需将实际产量 (Y_s) 分解为趋势产量项 (Y_T) ^[10]和由气象条件变化引起的气象产量项 (Y_w) ^[11]:

$$Y_s = Y_t + Y_w^{[12]}$$

在建立作物产量预报模型时,普遍应用的是利用统计学方法提取气象产量,气象产量代表气象因子波动而影响的那部分产量。但是由于前后期同样的气象产量并非由相同的气象条件引起,所以气象产量序列的前后期时序可比性就比较差。其次,不同区间产量水平差异较大,彼此的绝对气象产量没有可比性,即绝对气象产量序列缺乏空间可比性。所以目前多采用式(1)计算相对气象产量(Y_w)^[13]。

$$Y_w = (Y_s - Y_t) / Y_t \times 100\% \quad (1)$$

这里首先采用统计学方法(灰度模型)提取出相对气象产量,以相对气象产量作为划分标准,将其分五类: - 2(大歉[-∞, -18%]), - 1(歉[-18%, -4%]), 0(平[-4%, 7%]), 1(丰[7%, 17%]), 2(大丰[17%, +∞))。由于统计学方法处理的数据一般都要要求曲线比较平稳,而且数据量不能过大,因此直接采用引言中的统计学方法划分丰歉年型是不合理的,可能会出现采用不同的统计学方法处理同一样本后丰歉年型的类别很多不同,根据产量阶段的概念^[14],使用最优分割法^[15]将大量产量数据按照天气变化阶段分段,对每一段的数据分别考虑,这样划分出来的类别更加稳定和精确。所以用灰度模型和最优分割法确定出的类别作为样本决策属性。

3.3 特征属性

运用商空间粒度计算理论对原始样本中的气象数据进行处理,将气象数据的不同粒度表示作为样本的特征属性,不仅可以对论域 X 取不同的粒度^[16],也可对属性 $f(x)$ 取不同的粒度,即对属性的值域进行适当地划分,然后再转换成对论域的划分。因此可以对时间属性 $f(x)$ 进行颗粒化,取不同粒度下的属性,从不同角度挖掘出有用的知识,提高冬小麦产量预测的准确性。

下面对特征属性取不同粒度进行实验,论域 X 即上面确定的合肥市从 1953 年到 2000 年的 48 个样本。

3.3.1 按天取粒度

将每个样本的气象因子从前一年的 9 月份取至当年 6 月上旬截止,从 1952 年 9 月至 2000 年 6 月共 48 个样本,产量值从 1953 年至 2000 年,特征属性个数为 $3 \times 283 = 849$ 个,其中 283 表示天数,3 表示每天的属性个数:降水、气温和日照。

采用最基本的交叉覆盖对样本进行学习和预测,得出气象产量的趋势 Y_w 后,根据 $Y_w = (Y_s - Y_t) / Y_t$

$\times 100\%$,还原以得到预测的总产量值 $\hat{Y}_s = \hat{Y}_w * \hat{Y}_t + \hat{Y}_t$ 。

表 1 是按天取粒度的实验结果。

表 1 合肥市 1996~2000 年按 $[f]_d$ 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	(3245, 3616.8)	3572.1
1998	3125.8	(3000.8, 3344.6)	3079.1
1999	2933.4	(2405.4, 2816.1)	2766.9
2000	3374.8	(2767.3, 3239.8)	2839.9

由表 1 可以得出,使用灰度模型(GM(1,1)模型)学习样本后,得到的趋势产量与实际产量存在一定距离,因此通过商空间粒度计算与构造性学习方法对已知样本学习后得出预测样本类别,将其还原后给出预测产量的修正值范围,该范围都能包括实际产量。

用上述方法对 GM(1,1)预测值修正前后的比较如图 2 所示。

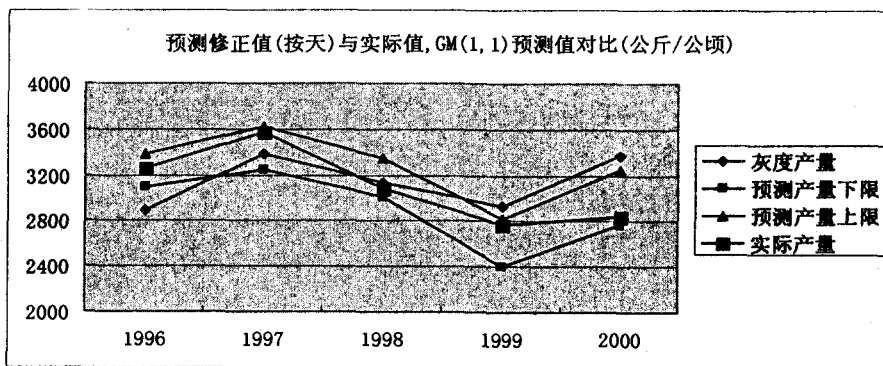


图 2 合肥市 1996~2000 按 $[f]_d$ 产量预测结果比较

考察图 2, 1996~2000 年的预测修正区间都能准确包含实际产量,比 GM(1,1)预测值更接近实际产量,如 1996 年 GM(1,1)预测值较低,通过上述方法修正后预测产量有一定提高,可以准确包含实际产量。因此基于商空间粒度计算的产量预测方法可以更好地反映冬小麦产量的宏观增长和微观波动,相比 GM(1,1)模型提高了预测精度和可靠性。

3.3.2 按旬取粗粒度属性

前面已经讨论了按天将降水量、气温、日照分别依次组成多维样本,这里将属性 $f(x)$ 粗化,按旬统计各属性值,仍以合肥市为例,如要在六月上旬预测产量时,特征属性个数为 $28 \times 3 = 84$ 个。

观察表 2 后得,按旬取粗粒度,预测结果误差增大,其中 1997, 1998, 1999 年预测修正值区间都没有包括实际产量、预测修正值与实际值,GM(1,1)值对比如图 3 所示。观察图 3,按旬取粗粒度学习,预测产量区

表 2 合肥市 1996 ~ 2000 年按 $[f]_0$ 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

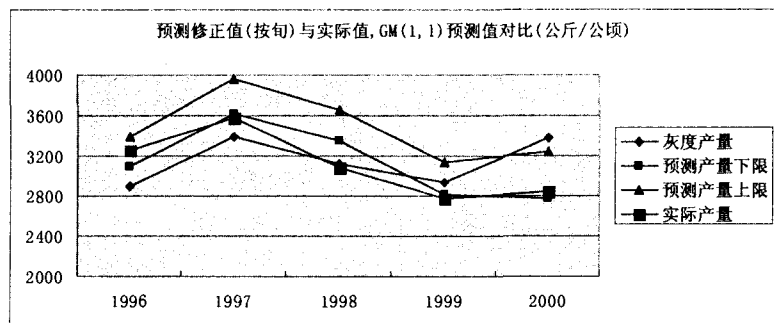


图 3 合肥市 1996~2000 年按 $[f]_0$ 产量预测结果比较图的变化从总体趋势上和实际产量还是保持一致,相对表 1 来说,虽然 1997,1998,1999 年的修正区间未能包含实际产量,1997,1999 年的涨跌趋势与实际产量还是一致的,只有 1998 年实际产量有一定下降,而预测修正区间的下降程度不够明显,这也恰恰是粒度变粗后信息自然会丢失一部分的表现。在面对样本粗粒度学习时,人们只能获得样本的大概信息,只有将粒度不断细化,由粗到细,才能在学习过程中获得样本更多的隐含信息,使样本信息更加具体化。

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] 王世香,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,张钺. 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,张钺,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.