第 23 卷 第 4 期
2008 年 12 月

天津科技大学学报
Journal of Tianjin University of Science & Technology

Vol.23 No.4
Dec. 2008

遗传算法在多模式集成天气预报中的应用

熊聪聪[1]，王 静[1]，宋 鹏[1]，孟冬梅[2]
（1. 天津科技大学计算机科学与信息工程学院，天津 300222；2. 天津市气象科学研究所，天津 300222）

摘 要：采用实数编码的遗传算法，利用其良好的全局优化特点演化出各模式预报值在集成预报中的权重系数，得出集成预报的模型，实现了天气预报多种模式的集成处理，并利用天津气象科学研究所提供的实际测量数据进行了性能分析和验证. 结果表明，利用遗传算法可以实现较好的集成性天气预报.

关键词：集成天气预报；遗传算法；实数编码；权重；模式
中图分类号：TP183；S716.1 **文献标识码**：A **文章编号**：1672-6510（2008）04-0080-05

Application of Genetic Algorithm in Multimode Integrated Weather Forecast

XIONG Cong-cong[1]，WANG Jing[1]，SONG Peng[1]，MENG Dong-mei[2]
（1. College of Computer Science and Information Engineering，Tianjin University of Science & Technology，Tianjin 300222，China；2.Meteorological Institute of Tianjin，Tianjin 300222，China）

Abstract：Using real-coded genetic algorithm and its good global optimization characteristics，weight coefficients of every model in the integrated forecasting were evolved. By using this method，integrated forecasting models were created and the integration of several models were realized. Finally the performance of the algorithm was analyzed and validated by real data provided by meteorological institute of Tianjin. The results indicate that the genetic algorithm can get a better integrated weather forecast.

Keywords：integrated weather forecast；genetic algorithm；real-coded；weight；model

集成天气预报[1]将各种数值预报模式的预报结果进行合理综合，更好地体现出各集成成员模式在不同时间和地点的预报水平，以提高预报的准确率. 目前用于预报集成的主要方法有选择最优法、权重系数法和神经元网络法等[2]，其中，选择最优法虽然操作简单易行，但集成结果准确率往往受到预报方法准确率的限制，不能较好的识别阶段性的不稳定；神经元网络法虽然在预测复杂的非线性时间序列方面占明显优势，但由于其存在过度学习和过分依赖参数选择等问题，集成结果不是很稳定.

遗传算法作为一种借鉴自然界生物种类遗传和进化过程而形成的自适应全局优化搜索算法，具有较强移植性和通用性，对于需要进行全局优化和难于解析的问题处理有着较强的优势，在函数优化、自动控制、图像处理、系统辨识等领域都得到广泛的应用[3]. 目前，遗传算法在集成天气预报中的应用研究不多.

本文利用遗传算法得出各预报模式预报结果的权重系数，进而得出该气象元素的集成预报模型，并对这一方法进行了研究，对影响集成预报准确性的因素进行了分析.

1 遗传算法

遗传算法的基本原理是将 n 维向量 $X=[X_1$，

收稿日期：2007-03-05；**修回日期**：2008-08-29
基金项目：天津市科技发展计划项目（6YFSDSF04500）
作者简介：熊聪聪（1961—），女，四川泸州人，教授.

$X_2, \cdots, X_n\ ]^{\mathrm{T}}$ 表示成由 X_i（$i = 1, 2, \cdots, n$）所组成的符号串：X=$X_1, X_2, \cdots, X_n$，把每一个 X_i 看作一个遗传基因，则 X 可看做是由 n 个遗传基因所组成的一个染色体个体，变量 X 组成了问题的解空间. 把这些假设解置于问题的环境中，根据目标函数对每个个体进行评价，并给出一个适应度值，以此来判断个体的优劣程度. 按照适者生存的原则，从中选择出较适应环境的个体进行复制，再通过交叉、变异过程，产生更适应环境的新一代群体. 求解问题最优解的过程就是对解空间内所有染色体 X 按照适应度进行搜索的过程，个体 X 的适应度越大，越接近于目标函数的最优解. 根据适应度的大小选择一定数量的个体，作为下一代群体，再继续进化，如此经过若干代后，算法收敛于最好的染色体，它很可能就是问题的最优解或次优解[4].

2　遗传算法应用于多模式集成天气预报

采用实数编码的遗传算法，实现了多模式集成的天气预报，主要包括对集成天气预报问题的描述，确定决策变量、约束条件，建立规划模型，遗传算法设计和遗传算法的运行、调试.

2.1　集成模式选择和原始数据预处理

目前用于气象预报的数值预报模式主要有 MM5 模式、GRAPS 模式、WRF 模式、中国国家气象中心 T213 模式、德国气象局业务模式（GERM）和日本气象厅业务模式（JAPAN）等. 根据天津市气象科研所提供的由中央气象局下发和天津市气象局预测的天津地区 232 个自动预报站历史气象统计资料，选择数据量较为充分的 MM5 模式、T213 模式、GERM 模式和 JAPAN 模式作为集成预报研究的成员模式.

因为集成预报涉及多种气象元素、预报模式、预报时效和间隔及多个预报站点，数据量繁多且存在漏报和空报现象，所以应进行数据预处理，以保证集成的可靠性. 考虑到集成预报的统一性，对各成员模式的预报结果进行插值处理，以统一时空分辨率和预报时效、间隔；考虑到数据的可操作性，将数据存入数据库，通过数据库进行操作；考虑到集成的准确性，对原始气象数据进行筛选以去除漏报和无效预报值.

2.2　权重分配方案

为了更好地体现各成员模式的客观预报能力，集成的过程并不是简单地对各成员模式取算术平均，而是给出合理的权重，并且权重的分配不仅针对成员模式本身，而且具体到每个成员在不同站点和时间点的区别，公式为

$$\overline{R}_{j,i} = \sum_{i=1}^{m} W_{i,j,t} R_{i,j,t}$$

式中：i 为集成预报成员模式；j 为站点号；t 为预报时间点；m 为成员个数；$\overline{R}_{j,t}$ 为某一气象要素在站点 j 第 t 个时间点的集成预报值；$W_{i,j,t}$ 为在站点 j 第 i 个成员模式第 t 个时间点上的权重系数，为[0, 1]区间内的随机数；$R_{i,j,t}$ 为该气象元素在站点 j 第 i 个成员模式第个 t 时间点的预报值.

2.3　遗传算法设计

在应用遗传算法解决具体问题时涉及到五个基本要素：参数编码、初始群体设定、适应度函数设计、遗传操作的设计和运行参数设定[5]. 遗传算法流程如图 1 所示.

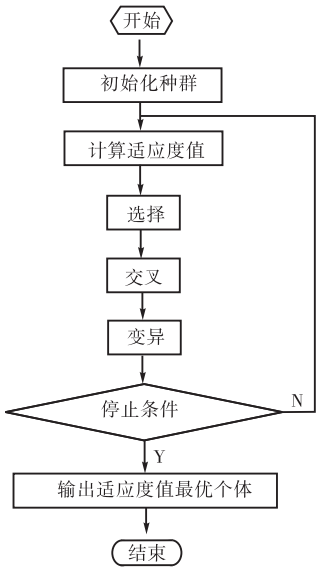

图 1　遗传算法流程图

Fig. 1　Genetic algorithm flowchart

2.3.1　*参数编码*

遗传算法是通过对可行解的个体编码施加选择、交叉、变异等遗传运算来达到优化目的的. 把一个问题的可行解从其解空间转换到算法所能处理的搜索空间的转换方法就称为编码. 编码是应用遗传算法时首先要解决的问题，编码的好坏对算法的性能有着很大的影响.

考虑到所解决的是一个高维多局部极值的连续参数优化问题，根据集成预报问题的实际要求，选用实数编码方式. 实数编码遗传算法原理与标准遗传算法类似，不同的是它直接采用原始变量构成染色体，

染色体为一个实数向量，染色体上的每个基因值用实数表示. 采用实数编码，除了具有传统二进制编码遗传算法的优点，还克服了海明悬崖，算法微调能力差，搜索效率较低等不足，具有取值广泛，运算效率高，交叉变异方法灵活等特点[6].

2.3.2　初始群体设定

遗传算法中初始种群的好坏将直接影响算法的收敛速度和收敛结果. 考虑到遗传算法的特点以及集成天气预报受地域性和时间差异的影响，针对天津市 232 个自动预报站点，选取相邻的 20 个站点 3 个月的历史数据（筛选后 87 天可用）作为一次试验样本. 将假设解的集合作为遗传的一代群体，一代群体由 87 个染色体组成. 某一气象元素 20 个相邻站点各种预报模式一次预报时效（48 h）的预报值作为一个染色体，每个染色体由 $n\times m$ 个基因组成，其中 n 为参与集成预报的模式个数，m 为一次预报的时间点数（时间点数=预报时效/预报间隔）.某一气象元素在某站点上某种预报模式某一时间点的预报值所对应的权重系数代表染色体上的一个基因.

2.3.3　适应度函数设计

度量个体适应度的函数称为适应度函数，用于评价个体（解）的好坏. 适应度函数的值越大，解的质量越好. 适应度函数的选取对遗传算法求得全局最优解起着关键性作用，直接影响到算法的收敛速度以及能否找到最优解，应结合求解问题本身的要求设计.

集成天气预报要求在集成多个成员模式预报结果的同时，使集成后的结果与真实值的误差最小[7]. 根据实际问题，适应度函数可以选择为

$$f=1/E_j=1/\sum_{j}^{20}\sqrt{\sum_{i}^{n}(\sum_{i}^{m}w_{i,j,t}R_{i,j,t}/\sum_{i}^{m}w_{i,j,t}-r_{i,j,t})^2/n}$$

式中：i 为集成预报成员模式；j 为站点号；t 为预报时间点；m 为成员个数；n 为一次预报的时间点数目；E_j 为所选取的 20 个站点某一气象元素一次预报的集成结果与实际数据的均方误差[8]；$W_{i,j,t}$ 为在 j 号站点上第 i 个成员模式在第 t 个时间点上的权重系数；$R_{i,j,t}$ 为该气象元素第 i 个成员模式在站点 j 上第个 t 时间点的预报值；$r_{i,j,t}$ 为第 i 个模式在站点 j 上在预报时间点 t 的实况值.计算 $W_{i,j,t}$ 使该气象元素在站点 j 上集成预报的均方误差达到最小.

2.3.4　遗传操作设计

选择操作是指从群体中按个体的适应度函数值选择出较适应环境的个体，主要目的是提高全局收敛性和计算效率，这里采用轮盘赌选择方法.以单个染色体的适应度值占种群中所有染色体的适应度值之和的比率作为选择概率.选择操作步骤为：

（1）根据初始群体的设定的各染色体的适应度，计算各染色体的相对适应度 $R(i)$ 和累计适应度 $C(i)$，其值的范围为（0，1］. 公式为 $R(i)=f(i)/\sum_{i}^{M}f(i)$；$C(i)=C(i-1)+R(i)$，其中：$C(0)=R(0)$，$f(i)$ 为适应度函数，M 为群体大小，i 为染色体的编号.

（2）遍历群体的所有染色体，每一次循环产生一个 0 到 1 之间的随机数 p，如果 $C(i-1)\leqslant p\leqslant C(i)$，则选择第 i 条染色体进入下一代遍历结束，则选择结束.

交叉操作是指对两个相互配对的染色体依据交叉概率 P_c 按某种方式相互交换其基因，从而形成新的个体. 交叉是遗传算法区别于其他进化算法的重要特征，是产生新个体的主要方法. 根据实数编码的特点，选择算数交叉方式.算数交叉是一种线性内插方法，采用两个个体的所有基因的线性组合而产生出新的基因组成新的子代个体. 遍历群体中所有染色体，在每次循环时产生一个 0 到 1 之间的随机数 p，若 p 小于交叉概率 P_c，则选择个体进行交叉. 其中，选定 a，b 个体中的所有基因按照 $L=\min(m,n)+|m-n|r$ 进行交叉（r 为[0，1）的随机变量，L 为产生的新的基因，m 为个体 a 的基因，n 为个体 b 的基因），得到所有新的基因 L 组成新的子代个体 c.

变异操作是指依据变异概率 P_m 将个体编码串中的某些基因值用其他基因值来替换，从而形成一个新的个体. 变异是遗传算法中产生新个体的辅助方法，决定了算法的局部搜索能力，同时保持种群的多样性，防止出现早熟现象. 本文选用均衡变异，即在变异过程中个体中的一个随机基因在约束条件的上下范围内实现随机生成，并替换原有基因值. 选择变异基因的方法为：遍历群体中所有的染色体和基因，每找到一个基因就产生一个随机数 p 和变异概率 P_m 比较，若 p 小于 P_m 则通过随机函数产生一个 0 到 1 之间的一个数，替换原有的基因.

2.3.5　运行参数的设定

遗传算法有以下 4 个运行参数，即群体大小 M，一般取 20~100；遗传运算终止进化代数 N，一般取 100~500；交叉概率 P_c，一般取 0.4~0.9；变异概率 P_m，一般取 0.001~0.1. 这 4 个运行参数对遗传算法求解的结果和效率都有一定影响，但目前尚无合理选择它们的理论依据.在遗传算法的实际应用中，往往需要经过多次试算后才能确定出这些参数合理的取值

大小或取值范围.

M 太小难以求出最优解，太大则增长收敛时间；交叉概率 P_c 取值太小难以向前搜索，太大则容易破坏具有高适应值的结构；变异概率 P_m 取值太小难以产生新的基因结构，太大易使遗传算法成为单纯的随机搜索. 根据集成预报的实际情况要求，经多次试验测试，最终选定系统的运行参数为：M=87；$N=2\times10^6$；P_c=0.75；P_m=0.05.

3 实　验

以格点号 55 428（蓟县）的气象元素温度，预报时效 48 h，预报间隔 3 h 和 6 h 为例，初始群体选用 07 年 5 月 1 日至 7 月 31 日的数据（87 天）为模型训练样本，集成预报成员为 MM5 模式、T213 模式、GREM 模式和 JAPAN 模式，得出 8 月 1 日到 8 月 2 日 48 h 内的集成预报值.

选用的 4 种预报模式和集成预报模式的误差如图 2 所示. 集成预报模式与真实数据的均方误差为 3.1717，其中 MM5 模式、GERM 模式、JAPAN 模式和 T213 模式的均方误差分别为 2.159 3 ℃、2.394 1 ℃、2.149 2 ℃和 7.178 5 ℃. 很明显，T213 模式的预报误差相对较大，集成预报误差小于 T213 模式但大于其他 3 种模式的预报误差.

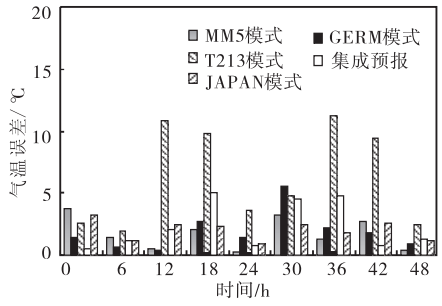

图 2　5种预报模式的误差图

Fig. 2　Temperature error of five prediction models

含有 T213 模式的集成预报效果并不理想，因此考虑将其去除，去除 T213 前后的集成预报效果如图 3 所示. 去除 T213 模式后，集成结果与真实数据的均方误差为 1.946 7 ℃，好于含有 T213 模式的 4 种模式集成效果. 由此可见，集成预报的结果与成员模式的选取有很大的关系，成员个体的预报性能会影响到集成后的预报效果，应择优选取.

去除 T213 模式后，其他 3 种模式 3 h 间隔的集成效果如图 4 所示，集成预报结果与实际数据的均方误差为 1.874 2 ℃<1.946 7 ℃，集成效果好于图 3 中 6 h 间隔的集成预报，可见，预报时间间隔大小对集成预报结果也有一定影响.

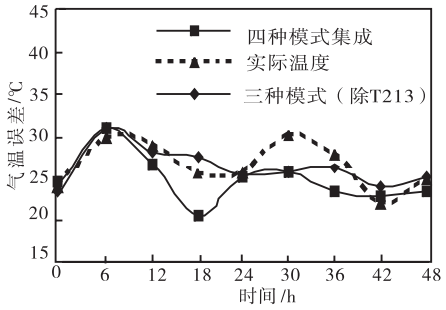

图 3　去除 T213前后的集成预报效果图

Fig. 3　Integrated forecast effect before and after excluding T213

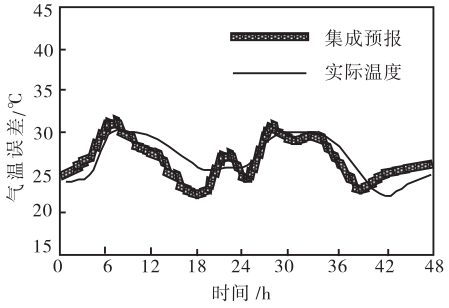

图 4　3种模式预报间隔 3 h集成效果图（不含 T213）

Fig. 4　3 h interval integrated forecast of three prediction models（not included T213）

在实验过程中还发现，集成预报结果与地域差别也有一定关系. 实验分别以附近相邻的 20 个站点为单元生成集成预报模型，一定程度上避免了时空差异造成的影响. 考虑预报时效也可能对集成预报造成影响，分别进行了 48 h 和 72 h 时效的集成预报实验，实验结果表明，72 h 预报时效更有利于体现出集成预报的效果. 虽然集成预报结果受到集成成员模式选取，时空差异以及预报时效和间隔的影响，但总体来讲，除了降水以外，其他气象要素的集成预报结果与实际数据的均方误差小于允许的误差范围（如温度误差≤3 ℃），实验基本达到了预期的效果.

4 结　语

利用遗传算法得出的多模式集成天气预报模型，基本上实现了集成预报的预期目的.

由于传统遗传算法本身的缺陷，尚存在算法运算时间过长，计算结果不够稳定等问题. 因此，可以考虑从以下几方面进行改进：采用随机联赛选择模型替代轮盘赌模型，以降低处理时间；采用高斯变异来改

进均匀变异，达到较好的收敛效果；将进化的终止条件设为条件满足式判断，防止出现局部极值；对 P_c 和 P_m 采用自适应调整，以利于算法的收敛. 集成预报模型的建立以使用大量原始数据做训练样本为前提，由于气象台提供的有效原始数据量有限，所以集成模型还有待改善.由于降水的时空上均不连续性，导致建模较为困难，集成预报结果并没有明显的改善，可以尝试与其他演化算法相混合进行更深一步研究.

参　考　文　献：

［1］ 章少卿，丁世晟. 预报综合问题的初步探讨［J］. 气象学报，1960，3（1）：110—118.

［2］ 杞明辉. 天气预报集成技术和方法应用研究［M］. 北京：气象出版社，2006：13—21.

［3］ Holland J H. Adaptation in natural and artificial systems［M］. Ann Arbor：University of Michigan press，1975：13—65.

［4］ 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，2002：31—96.

［5］ 云庆夏. 遗传算法和遗传规划：一种搜索寻优技术［M］. 北京：冶金工业出版社，1997：21—30.

［6］ 金　聪. 函数优化中实数型遗传算法的研究［J］. 小型微计算机系统，2000，21（4）：372—373.

［7］ 周　明. 遗传算法原理及应用［M］. 国防工业出版社，1999：41—63，123—137.

［8］ 何中市. 岭回归估计均方误差的重要特性及其应用［J］. 重庆大学学报：自然科学版，1992，15（4）：122—124.

(上接第58页)

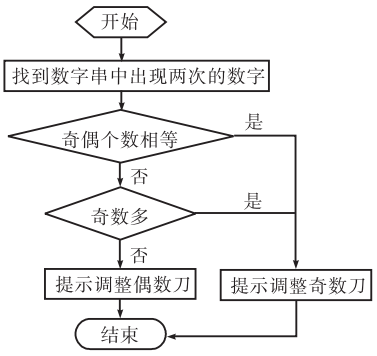

图7　相邻刀片间位置误差调整流程图

Fig. 7　Diagram of blade adjudgement

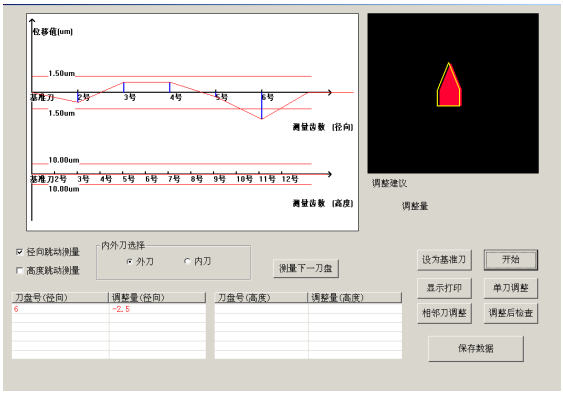

图8　铣刀盘刀片径向跳动误差测量实例

Fig. 8　Instance of blade radial runouts measurement

实验证明，该算法能够智能化地给出需要调整的刀片编号、调整方向及调整量，极大地提高装配式铣刀盘刀片的调整效率.

3　结　论

本文研发的智能型装配式铣刀盘测量仪具有测量精度高、效率高的特点. 测头结构采用创新设计方法，实现了弧齿锥齿轮装配式铣刀盘刀片径向跳动测头和刀刃高度测头在功能和机械结构上的集成. 开发的测量仪软件系统具有智能化功能，能够对测量数据进行智能分析，并且给出正确、高效的刀片调整信息. 该仪器作为高精度数控铣齿机的必备附属装备在机械制造业中可得到广泛应用.

参　考　文　献：

［1］ 寥绍华. 齿轮加工技术和装备的发展现状与趋势［J］. 世界制造技术与装备市场，2007（3）：37—37.

［2］ 李先广. 当代先进制齿与制齿机床技术的发展趋势［J］. 制造技术与机床，2003（2）：10—12.

［3］ 付莹莹，张俊亮，伊连云，等. 测量数控机床刀具尺寸的简便方法［J］. 机械工程师，2006（8）：145—145.

［4］ 王超厚，罗良玲，徐　晗. 软测量技术及其在刀具故障诊断中的应用［J］. 工具技术，2007，41（10）：69—71.

［5］ 高永全，王　凡. 刀具磨损的测量与自动补偿［J］. 工艺与装备，2006（12）：82—83.

［6］ 侯学智. 数字图像刀具测量机［D］. 西安：电子科技大学，2004.

［7］ 侯学智，杨　平，赵云松. 基于图像处理技术的刀具测量系统［J］. 工艺与检测，2004（3）：33—35.