

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G05B 13/02 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200880008403.3

[43] 公开日 2010年1月27日

[11] 公开号 CN 101636698A

[22] 申请日 2008.2.21

[21] 申请号 200880008403.3

[30] 优先权

[32] 2007.3.19 [33] US [31] 60/895,637

[86] 国际申请 PCT/US2008/054599 2008.2.21

[87] 国际公布 WO2008/115655 英 2008.9.25

[85] 进入国家阶段日期 2009.9.15

[71] 申请人 陶氏环球技术公司

地址 美国密歇根

[72] 发明人 吉多·弗雷迪·斯米茨

亚瑟·卡尔·科登

[74] 专利代理机构 中科专利商标代理有限责任公
司

代理人 李敬文

权利要求书 3 页 说明书 15 页 附图 7 页

[54] 发明名称

使用三维 PARETO - FRONT 遗传规划开发的
推理传感器

[57] 摘要

使用遗传规划技术来开发一种用于基于多个输入变量来预测至少一个输出变量的预测算法，所述遗传规划技术通过多代对候选算法的群体进行进化。在每一代中，基于三个适合度准则对候选算法进行评估：(i) 精确性准则，评估每一个候选算法基于输入变量的相应历史测量，预测至少一个输出变量的历史测量的能力；(ii) 复杂度准则，评估每一个候选算法的复杂度；以及 (iii) 平滑性准则，评估每一个候选算法的非线性。可以用于监视比如工厂中工业过程等物理、化学或者生物过程的推理传感器实施该预测算法。

1、一种开发预测算法的方法，所述预测算法用于基于多个输入变量来预测至少一个输出变量，所述方法包括：

获得代表物理、化学或者生物过程的第一数据集合，所述第一数据集合包括所述至少一个输出变量的第一集合测量以及所述输入变量的相应第一集合测量；

使用应用至少三个适合度准则的遗传规划技术，对多个候选算法进行进化，所述至少三个适合度准则包括：精确性准则，用于评估每一个候选算法基于所述输入变量的所述相应第一集合测量来预测所述至少一个输出变量的所述第一集合测量的能力；复杂度准则，用于评估每一个候选算法的复杂度；以及平滑性准则，用于评估每一个候选算法的非线性；以及

选择所述候选算法之一，作为所述预测算法。

2、根据权利要求1所述的方法，其中所述物理、化学或者生物过程是工业过程。

3、根据权利要求1所述的方法，其中所述至少一个输出变量涉及排放监视。

4、根据权利要求1所述的方法，其中所述至少一个输出变量是物理属性、化学属性、机械属性、材料属性或者生物属性。

5、根据权利要求1所述的方法，其中对于所述三个适合度准则，所述预测算法是所述候选算法中最优的。

6、根据权利要求1所述的方法，其中选择所述候选算法之一作为所述预测算法包括：

确定所述候选算法中的哪一个位于针对所述三个适合度准则的三维Pareto-front上，以获得多个Pareto-front算法。

7、根据权利要求6所述的方法，其中选择所述候选算法之一作为所述预测算法还包括：

应用至少一个附加选择准则，以从所述多个Parato-front算法中选择所述预测算法。

8、根据权利要求1所述的方法，还包括：

获得代表所述物理、化学或者生物过程的第二数据集合，所述第二数据集合包括所述至少一个输出变量的第二集合测量以及所述输入变量的相应第二集合测量；以及

使用所述第二数据集合来测试所述预测算法。

9、根据权利要求8所述的方法，其中使用所述第二数据集合来测试所述预测算法包括：

评估所述预测算法的下述能力：基于所述输入变量的所述相应第二集合测量、预测所述至少一个输出变量的所述第二集合测量的能力。

10、一种监视物理、化学或者生物过程的方法，所述方法包括：

确定所述物理、化学或者生物过程的多个可测量参数的当前值；

向推理传感器提供所述当前值，所述推理传感器被配置为使用至少一个预测算法，所述至少一个预测算法基于所述可测量参数来预测所述物理、化学或者生物过程的附加参数，所述至少一个预测算法是已经在精确性、复杂度和平滑性方面经过优化的；以及

所述推理传感器使用所述至少一个预测算法，以基于所述可测量参数的所述当前值，来确定所述附加参数的推理值。

11、根据权利要求10所述的方法，其中确定所述物理、化学或者生物过程的多个可测量参数的当前值包括：

使用硬件传感器来测量所述可获得参数中的至少一个。

12、根据权利要求10所述的方法，其中所述过程是工业过程。

13、根据权利要求12所述的方法，还包括：

基于所述附加参数的所述推理值来控制所述工业过程。

14、根据权利要求13所述的方法，其中基于所述附加参数的所述推理值来控制所述工业过程包括：

将所述推理值与预设值进行比较。

15、根据权利要求10所述的方法，其中所述附加参数是物理属性、化学属性、机械属性、材料属性或者生物属性。

16、根据权利要求10所述的方法，还包括：

使用向所述可测量参数的历史值和所述附加参数的历史值应用

的遗传规划技术，来开发所述至少一个预测算法，所述历史值是已经针对所述物理、化学或者生物过程而预先获得的。

17、根据权利要求16所述的方法，其中所述可测量参数的所述历史值定义第一范围，并且所述可测量参数的所述当前值的至少一个子集在所述第一范围之外。

18、根据权利要求16所述的方法，其中所述附加参数的所述历史值定义第二范围，并且所述附加参数的所述推理值在所述第二范围之外。

19、一种用于监视物理、化学或者生物过程的系统，所述系统包括：

至少一个硬件传感器，用于获得所述物理、化学或者生物过程的多个参数的测量值；以及

推理传感器，用于基于所述测量值来确定所述物理、化学或者生物过程的附加参数的推理值，所述推理传感器被配置为使用至少一个预测算法来确定所述推理值，所述至少一个预测算法是已经使用遗传规划技术在精确性、复杂度以及平滑性方面进行了优化的。

20、根据权利要求19所述的系统，还包括：

控制系统，用于基于所述推理值，控制所述物理、化学或者生物过程。

使用三维 PARETO-FRONT 遗传规划开发的推理传感器

技术领域

本发明涉及软或者推理传感器，并且更具体地，涉及使用三维 Pareto-front 遗传规划技术开发的推理传感器。

背景技术

通常通过测量各种过程参数来监视和控制工业过程。这些过程参数依赖于工业过程，可以包括例如温度、压力、流速等等。一些相关的过程参数可以在过程在线时相对容易地测量，例如使用硬件传感器测量。然而，其它过程参数可能难以在正常运行状况下进行测量。已经使用推理传感器或者软传感器，来基于容易测量的过程参数（输入变量），推理这些难以测量的过程参数（输出变量）。

在开发针对给定过程的推理传感器的典型方案中，针对较宽范围的过程状况，收集历史数据。使用该历史数据，可以使用多种技术来开发经验模型（empirical model），该模型可以基于可获得的输入变量，来预测所需的输出变量。线性回归是潜在可用的一种技术。然而，在实际应用中，由于工业过程大多数是非线性的，特别是化学工业中，因此线性回归仅具有有限的可应用性。

神经网络模型已用于对非线性工业过程进行建模。然而，神经网络模型通常与多个不同问题相关联。第一，神经网络模型可能经历面对过程改变的高灵敏度。因此，这就需要频繁的模型重开发和重调整。

第二，在用于模型开发的范围之外时，神经网络模型通常展示较差性能。换言之，一旦已经用给定范围的值训练了神经网络模型，则该神经网络模型可能无法在该范围之外进行正确地外推（extrapolate）。不能外推对于特定工业过程来说可能造成灾难性的后果，特别是在化学工业中。

第三，神经网络模型可以看作是“黑匣子”，因为神经网络模型可

以在不清楚表示预测机制的情况下进行预测。从而，可能很难解释神经网络模型。

第四，神经网络模型可能难于实施和支持。它们可能要求专用软件和专用训练。

因此，需要其它方法来开发能够基于容易测量的过程参数来推理难以测量的过程参数的推理传感器。

发明内容

在第一主要方面，示例实施例提供了一种开发预测算法的方法，所述预测算法用于基于多个输入变量来预测至少一个输出变量。根据所述方法，获得代表物理、化学或者生物过程的第一数据集合。所述第一数据集合包括所述至少一个输出变量的第一集合测量以及所述输入变量的相应第一集合测量。使用应用至少三个适合度准则的遗传规划技术对多个候选算法进行进化。所述至少三个适合度准则包括：(i) 精确性准则，用于评估每一个候选算法基于所述输入变量的所述相应第一集合测量来预测所述至少一个输出变量的所述第一集合测量的能力；(ii) 复杂度准则，用于评估每一个候选算法的复杂度；以及 (iii) 平滑性准则，用于评估每一个候选算法的非线性。选择所述候选算法之一作为所述预测算法。

在第二主要方面，示例实施例提供了一种监视物理、化学或者生物过程的方法。根据所述方法，确定所述物理、化学或者生物过程的多个可测量参数的当前值。向推理传感器提供所述当前值。所述推理传感器被配置为使用至少一个预测算法，所述至少一个预测算法基于所述可测量参数来预测所述物理、化学或者生物过程的附加参数，所述至少一个预测算法是已经在精确性、复杂度、以及平滑性方面经过优化的。所述推理传感器使用所述至少一个预测算法，以基于所述可测量参数的所述当前值来确定所述附加参数的推理值。

在第三主要方面，示例实施例提供一种用于监视物理、化学或者生物过程的系统。所述系统包括：(i) 至少一个硬件传感器，用于获得所述物理、化学或者生物过程的多个参数的测量值；以及 (ii) 推

理传感器，用于基于所述测量值来确定所述物理、化学或者生物过程的附加参数的推理值，所述推理传感器被配置为使用至少一个预测算法来确定所述推理值，所述至少一个预测算法是已经使用遗传规划技术在精确性、复杂度以及平滑性方面进行了优化的。

附图说明

图 1 示出了根据示例实施例的预测算法的节点表示。

图 2 示出了根据示例实施例的开发预测算法的流程图。

图 3 示出了根据示例实施例的可以实施预测算法的处理系统的示意图。

图 4 示出了根据示例实施例，用于使用遗传规划来开发 VOC 排放预测算法的设置参数的截屏。

图 5 示出了根据示例实施例，对用于开发图 4 的预测算法的函数进行选择的截屏。

图 6 示出了根据示例实施例，图 4 的预测算法预测的 VOC 排放与训练数据中的实际 VOC 排放的比较。

图 7 示出了根据示例实施例，图 4 的预测算法预测的 VOC 排放与测试数据中的实际 VOC 排放的比较。

图 8 示出了由神经网络模型预测的 VOC 排放与图 7 的测试数据的比较。

具体实施方式

1、综述

可以使用应用至少三个适合度准则的遗传规划技术来开发一种预测算法，该预测算法用于基于多个输入变量来预测至少一个输出变量。输入变量和至少一个输出变量可以与例如工业过程或者任意物理、化学或者生物过程的多种过程参数相对应。在一个有利的方案中，输入变量可以与当过程在线时例如使用硬件传感器而容易测量的参数相对应，而至少一个输出变量可以与难以测量的一个或者更多参数相对应。这样，可以在推理传感器中使用该预测算法，以基于容易测量的

过程参数（输入变量）的值来推理一个或者更多难以测量的过程参数（输出变量）的值。然后监视和/或控制该过程的系统可以使用该推理传感器的输出。

遗传规划技术可以用于开发可精确地再现代表了物理、化学或者生物过程的历史数据的预测算法。该历史数据可以包括至少一个输出变量的多个测量（例如，在不同时间和/或不同运行状况下获得的）、以及输入变量的相应测量。

为了找到具有所需精确性的预测算法，遗传规划技术可以通过多个代来对候选算法的群体进行进化。在每一代中，可以基于多种适合度准则，对候选算法进行评估。在示例实施例中，应用三个适合度准则：(i) 精确性准则，评估每一个候选算法以输入变量的相应测量为基础来预测至少一个输出变量的测量的能力；(ii) 复杂度准则，评估每一个候选算法的复杂度；以及 (iii) 平滑性准则，评估每一个候选算法的非线性。应当理解，这些适合度准则仅作示例之用，可以应用附加的和/或备选的适合度准则。

通过例如复制、杂交（crossover）、突变（mutation）等遗传操作的使用，允许将基于所应用的适合度准则的最适合的候选算法传播进入下一代。这样，从一代到下一代，遗传规划技术对候选算法进行进化，该候选算法平均而言渐进地变得更适合（由所应用的适合度准则来判断）。

最终，例如以代的数量为基础或者由于满足某个其它终止准则，进化过程终止。然后可以检查该阶段的群体，以确定哪个候选算法位于针对三个应用的适合度准则的三维Pareto-front上。Pareto-front算法是群体内的如下算法：无法在该群体中以另一算法提高针对一种适合度准则的适合度，而不降低对于另一适合度准则的适合度。这样，可以认为Pareto-front算法对于三个适合度准则来说是“最优”的。然后，可以基于其它选择准则，从Pareto-front算法中选择一个特定算法，作为预测算法使用。

可以使用例如附加历史数据，对选择的预测算法进行进一步测试。如果发现该预测算法可接受，则可以将其作为例如推理传感器的

一部分来在线使用，该推理传感器对物理、化学或者生物过程的一个或者更多参数进行推理。通过这种方式来开发预测算法，可以在精确性、复杂度以及平滑性方面对该预测算法进行优化。

在这三个适合度准则方面的优化可以提供多种优点。具体地，精确性准则对于精确的算法的开发非常有利。然而，应当理解，不同的应用可能要求不同级别的精确性。因此，一旦获得所需级别的精确性，其它准则就可以变得重要了。在这方面，复杂度准则有利于俭省的算法的开发，这些算法可能在计算上更高效并且对于过程改变而言更鲁棒。平滑性准则有利于更加线性的算法的开发，这些算法进行良好外推的可能性更大。考虑到过程参数的可变性，这最后一点特别重要。即使使用输入变量的值落入特定开发范围内的历史数据，来开发预测算法，也可能希望在输入变量的值落在开发范围外时，该预测算法还是保持精确。

因此，通过应用复杂度和平滑性以及精确性适合度准则，遗传规划技术可以开发满足精确性之外的其他目标的预测算法。例如，该技术可以得到对于过程改变的鲁棒性增强以及在开发范围之外进行外推的能力增强的预测算法。

2、预测算法的示例节点表示

在示例实施例中，使用遗传规划技术，以数学表达式形式来开发预测算法。然后，可以使用该数学表达式，基于输入变量的值来计算输出变量的值。这种数学表达式的例子如下：

$$(1) \quad V = [x/y + \alpha \exp(-z)]^\beta$$

其中V是输出变量，x、y以及z是输入变量，并且 α 和 β 是常数。

为了说明遗传规划技术，可以根据如何要解析数学表达式，以树结构来表示表达式。因此，可以通过如图1所示的树结构10来表示上述数学表达式(1)。在树结构10中，节点12-32表示函数、输入变量以及常数。具体地，节点28、30和32分别表示输入变量x、y以及z。节点20和16分别表示常数 α 和 β 。节点24表示除法函数，如“/”符号所指示的。节点26表示符号改变函数，如符号“-”所指示的。节点22表示指数函数，如“exp”符号所指示的。节点18表示乘法函数，如“*”符号所

指示的。节点14表示加法函数，如“+”符号所指示的。节点12表示幂函数，如“^”符号所指示的。

应当理解，表达式(1)以及相应的树结构10仅作为示例之用。一般而言，由遗传规划开发的表达式可以包括任意数量的节点，并且这些节点可以包括任意数量的函数、输入变量以及常数。函数可以包括代数函数(例如，加法、减法、乘法、除法、符号改变、平方根、幂等等)、超越函数(transcendental function)(例如，指数、对数、正弦、余弦等等)、条件函数(例如，依赖于自变量是否满足特定条件，比如自变量是否大于零，而取不同值的函数)、或者其它类型的函数。然而，函数不一定是数学函数。例如，如果输入变量表示逻辑状态，则函数可以包括逻辑函数，比如与、或、异或以及非。作为另一示例，输入变量可以包括文本信息，在该情况中，函数可以包括字符串函数。还可以使用其它类型的函数。

遗传规划技术的基本思想是通过应用对节点进行操作的遗传操作，来对表达式的群体进行进化，表达式的群体可以概念化为图1中的树结构。对节点的操作可以涉及改变一个或更多节点的“突变”。例如，突变可以将一个函数改变为另一个函数。因此，可以通过将节点18从乘法函数改变为加法函数，对树结构10进行突变。备选地，突变可以影响一个或者更多的终端节点，即，表示输入变量和常数的节点。因此，可以通过将节点20从常数 α 改变为输入变量，比如 x ，来将树结构10进行突变。

节点的操作还可以涉及“杂交(crossover)”操作，在该操作中，两个表达式交换一个或者更多节点。例如，在树结构10中，可以用从另一个表达式中捐献的一个或者更多节点来替代节点22、26、32。该捐赠的节点可以包括不同的函数、不同的输入变量、和/或不同的常数。另外，由于杂交操作，表达式中的节点数量可以增加或者减少。例如，树结构10中表示 $\exp(-z)$ 的节点22、26、32可以用单个节点来替换，比如用输入变量 z 。备选地，可以用表示更大数量的节点的更复杂的表达式，比如 $\log(x+yz)$ ，来替代节点22、26以及32。

尽管突变和杂交是典型的遗传操作，但是应当理解，可以用其它

方式来操作节点。作为一个例子，可以应用编辑操作。这种编辑操作可以移除现有节点和/或向表达式中插入新节点，同时确保表达式整体依然有效。作为另一个例子，可以应用简化操作。这种简化操作可以尝试对表达式进行简化，例如通过用更简单的等价的表达式来替代子树。因此，可以用新的常数来替代包括函数和常数但是没有变量的子树。

表达式的节点表示还可以用于其它目的。例如，表达式中节点的数量可以作为对该表达式的复杂度的度量。例如，由图1中树结构10表示的表达式具有11个节点，可以认为它比节点数量更少的表达式复杂，但是不及节点数量更多的表达式复杂。因此，本方案认为每一个节点对于表达式的复杂度的贡献是等量的。备选地，为了确定复杂度，可以向不同的节点赋予不同的权重。作为一个例子，可以认为表示超越函数的节点比表示加法函数的节点复杂，因此，在确定复杂度时，向其赋予更多权重。作为另一个例子，可以按照从属于节点的所有子树中的节点数量，来计算该节点的复杂度加权。因此，根据“ $\exp(-z)$ ”子树中的三个节点、“ $-z$ ”子树中的两个节点以及“ z ”子树中的一个节点，图1中的节点22具有复杂度加权6。然后，树结构10的复杂度可以采用它的每一个节点（即节点12-32）的复杂度加权之和。

3、开发预测算法的示例方法

图2示出了用于使用遗传规划来开发预测算法的示例方法的流程图。在本示例中，针对物理、化学或者生物过程，对预测算法进行开发，以基于多种输入变量，预测特定输出变量。

该方法可以开始于选择要用于开发预测算法的历史数据，如块100所示。历史数据可以表示该预测算法预期应用的相同过程或者类似过程。历史数据可以包括多个数据点，每一个数据点包括输出变量的测量以及输入变量的相应测量。数据中的一些或全部已被收集作为过程的正常操作的一部分。备选地，数据可以是专门针对模型开发目而已收集的，例如，通过使过程离线和/或通过系统地调整输入参数而收集的。另外，可以用多种方式来获得这些测量。例如，可以使用硬件传感器来获得对一个或者更多过程变量的直接测量。备选地，可以间

接地获得这些测量中的一些或全部，例如，通过推理传感器进行推理或者基于其它考虑而计算。

在使用历史数据来开发预测算法之前，历史数据还可以经历一个或者更多预处理步骤。可以使用此类预处理来识别和去除野值。还可以通过例如灵敏度研究，使用预处理步骤来确定哪些输入变量与待预测的输出变量相关。因此，可以精选历史数据，以集中在对最有可能与感兴趣的输出变量相关的最可靠数据进行表示的测量上。

可以将预处理后的历史数据进一步分为要作为“训练数据”用于开发预测算法的第一集合、以及要用于测试所开发的预测算法的第二集合。这两个集合可以定义输入和/或输出变量的不同范围。通过使用不同范围，可以使用数据的第二集合来测试预测算法在用于开发的范围之外的外推能力。

该开发方法中的另一个重要步骤是适合度准则的定义，如块102所示。适合度准则将用于评估通过遗传规划而进化的候选算法的适合度。在示例实施例中，使用三个适合度准则：(i) 精确性准则，评估每一个候选算法以历史数据中的输入变量的相应测量为基础来预测至少一个输出变量的能力；(ii) 复杂度准则，评估每一个候选算法的复杂度；以及 (iii) 平滑性准则，评估每一个候选算法的非线性。

精确性准则可以应用对候选算法和历史数据之间适合度的好坏的任意统计测量。这种统计测量的例子包括相关系数和 R^2 值。基于统计测量，候选算法越精确，则认为该候选算法更适合。

复杂度准则可以向每一个候选算法应用复杂度测量。可以通过对候选算法的树结构表示中的节点或者级别的数量进行计数，来测量候选算法的复杂度。备选地，可以向每一个节点给予复杂度加权，该复杂度加权对整个表达式的复杂度做出贡献。如上所述，可以将节点的复杂度加权计算为从属于该节点的每一个子树的节点数量。基于复杂度测量，候选算法越复杂，则认为该候选算法越不适合。

平滑性准则可以对每一个候选算法应用非线性测量。基于非线性测量，候选算法越非线性，则认为该候选算法越不适合。

可以用多种方法来测量非线性。在一个方案中，可以通过应用由

训练数据定义的区间中的输入变量的一定范围的值，并且计算输出变量的相应值，来对候选算法进行采样。然后，可以将候选算法的非线性量量化为输出变量的采样值中的变化的量。在另一个方案中，可以在不同的采样点，获得候选算法的局部导数（local derivative）。然后，可以将候选算法的非线性量量化为所采样的导数中的变化的量。

在另一个方案中，可以逐个节点地计算候选算法的非线性，其中，确定表达式中每一个函数节点的非线性，并且将整个表达式的非线性取为节点的非线性之和。可以用多种方法来确定节点的非线性，比如通过上述的采样技术或者通过多项式拟合。在后一种方案中，可以将节点的非线性程度取为为了在训练数据的区间中用多项式拟合该节点函数至规定容差之内而需要的多项式次数(polynomial degree)。

在示例实施例中，独立地应用这三个适合度准则。例如，可以针对这三个适合度准则中的每一个，对每一代中的候选算法进行评估，并且可以判断位于Pareto-front上的候选算法是最适合的。备选地，可以对这三个适合度准则一起进行加权，来指定整体适合度准则。在该情况中，可以判断具有最高整体适合度的候选算法是最适合的。还可以用其它方式来应用这三个适合度准则。

除了定义适合度准则外，还需要为遗传规划操作设置其它参数，如块104所指示。一个考虑可以是要用于遗传规划的函数集合，典型地包括简单的代数函数，比如加法、减法、乘法以及除法。然而，根据问题的需要，也可以选择其它类型的函数，比如条件函数和/或超越函数。例如，当数据具有周期性时，可以选择正弦和余弦函数。另一方面，可以避免某些函数，例如由于它们的非线性或者由于它们在计算上的效率较差而避免这些函数。

可以使用多种其它参数来设置候选算法的初始群体。例如，当随机生成初始群体时，可以指定用于选择函数作为节点的概率。更高的函数选择概率倾向于导致更复杂的表达式。还可以指定候选算法的初始群体的大小。在遗传规划操作的整个过程中，群体的大小可以保持恒定。备选地，可以允许群体在预设界限内增大或减小。群体大小的典型设置是在100至1000之间的值。图4中示出了这些其它参数的典型

设置。

可以使用其它参数来定义遗传规划中使用的进化过程。例如，在使用杂交操作的地方，可以使用可选择参数来确定杂交生成“后代”所占的分数（fraction）。可以使用另一可选择参数来确定随机的杂交和受到引导的杂交的概率。在使用突变操作的情况下，可以使用可选择参数来确定候选算法经历突变的概率，以及确定当突变的确发生时对函数或者终端节点进行突变的概率。对终端节点进行突变的概率的典型值是0.3。

还可以定义一个或者更多终止条件以用于遗传规划操作。在一个方案中，可以将遗传规划操作设置为运行固定数量的代。在典型的例子中，代的固定数量可以从几百至几千代。然而，可以指定用于终止遗传规划操作的其他条件。

应当理解，这些设置参数仅作示例之用。可以使用其它可选择参数来定义遗传规划操作。此外，可以用多种方式来选择用于不同的可选择参数的值，例如，任意的、基于“经验（rules of thumb）”的、基于反复试验（trial-and-error）的、或者基于其它考虑的，比如试验的统计设计。

一旦已对遗传规划操作进行了设置，就可以建立候选算法的初始群体，如块106所示。在一些情况下，可以随机生成初始候选算法。例如，可以随机选择用于顶级节点的函数。然后可以随机选择从属节点（例如，应用于选择函数或者终止节点的预设概率），直到创建了完整的树结构为止。

备选地，可以基于其它考虑来选择初始候选算法中的一些或者全部。例如，较早的研究（经验或者理论上）可能已经得到对一个或者更多看起来有前景的算法的开发。可以将这些潜在有前景的算法包括前在初始群体中。

然后，如块108所示，使用在块100中选择的历史数据以及在块102中定义的适合度准则，对候选算法进行进化。在示例演进方案中，选择群体中最适合的候选算法，来生成“后代”算法，然后使用这些后代算法来替代群体中最不适合的候选算法。

例如，在示例实施中，将当前代作为算法的存档（archive），通过锦标赛(tournament)过程从该存档中选择“父辈”。在锦标赛方案中，从存档中随机地选择预定数量（例如3至5个）的潜在“父辈”算法。然后使用锦标赛过程来确定“优胜者”，即，将允许哪个潜在“父辈”算法生成后代。在示例锦标赛过程中，“优胜者”可能是具有最高精确性的两个潜在“父辈”算法。然后，可以允许这两个优胜者通过杂交操作生成后代。应当理解，该锦标赛过程仅作示例之用，这是由于可以使用其它方案来确定哪些候选算法生成后代。

可以通过将一个父辈算法的子树与另一个父辈算法的子树结合的杂交操作来生成后代算法。因此，两个父辈算法可以交换子树，以生成两个不同的后代算法。可以随机地选择用于杂交的子树。

然后，可以基于预选择的突变概率，对后代算法进行突变。在一个方案中，可以使用突变来改变候选算法中的一个节点（函数或者终端节点，这依赖于遗传规划是如何设置的）。备选地，突变能够改变候选算法中的多个节点（例如，子树）。另外，一些突变可能增加或者减少候选算法中节点的数量。例如，如果将仅具有一个从属节点的函数改变为具有两个从属节点的函数，则可以随机地生成附加的终端节点。

应当理解，杂交和突变操作的前述讨论仅作说明之用。可以用多种方式来应用杂交和突变操作。另外，可以使用其它遗传操作来对候选算法进行进化。

可以用这种方式（例如，使用锦标赛方案来选择“父辈”）继续生成后代，直到达到由群体参数指定的后代的数量。然后，可以针对精确性、复杂度以及平滑性准则，对包括“存档”算法和“后代”算法在内的所有候选算法进行评估。然后使用位于或者最接近 Pareto-front 的“存档”和“后代”算法的合并集合中的候选算法，来构成新的存档。

可以将该新存档的大小指定为“后代”群体的大小的百分比，例如75%至100%。因此，使用100%的例子，可以使用100个算法的存档来生成100个后代。然后，将存档和后代算法的合并集合中的200个算法中的100个最适合的算法用于构成100个算法的新存档。这样，在精

确性、复杂度以及平滑性方面最适合的候选算法继续到下一代，并且随着每一个后续代，候选算法平均上渐进地变得更适合。

如块110所示，候选算法可以这样继续进化，直到满足终止条件。当满足终止条件时，如块112所示，可以检查候选算法的群体，以针对三个适合度准则，确定哪个候选算法位于Pareto-front上。尽管Pareto-front算法典型地仅是群体的一个小子集，但是可能仍然存在多于一个的算法。为了选择Pareto-front算法中的一个作为预测算法，可以如块114所示，应用附加的选择准则。

附加的选择准则可以例如确定哪个Pareto-front算法具有最希望的精确性、复杂度和平滑性的平衡。附加的选择准则还可以基于Pareto-front算法的表达式的内容，评估Pareto-front算法。例如，可以优选具有有意义的物理解释和/或包括特定类型函数的表达式。在某些情况下，可以基于附加的测试，例如，使用下述用于验证过程的附加数据，从Pareto-front算法中选择预测算法。因此，可以基于预测算法在训练数据所用的范围之外的精确外推能力，来选择预测算法。

然后如块116所示，可以验证所选择的预测算法。备选地，可以使用该验证过程来选择将哪个Pareto算法作为预测算法使用。验证过程可以涉及用附加数据测试预测算法。该附加数据可以是例如从在块100中选择的历史数据中保留的第二数据集合。如果预测算法能够以希望的精确性级别再现历史数据的第二集合，则可以认为该预测算法是有效的。如上所述，用于以在开发中使用的第二数据集合中包括的值的范围之外的值进行测试的第二数据集合可以用于测试预测算法成功外推的能力。应当理解，还可以用其它方法来测试预测算法。例如，可以基于试验，在线使用预测算法，以确定它是否给出良好的结果。

一旦已验证了预测算法，则如下面更详细的描述，可以以推理传感器实施算法。

4、预测算法的示例实施

可以以多种方式来使用通过例如如上所述的遗传规划技术而开发的预测算法。例如，可以在监视和/或控制预测算法开发所针对物理、化学或者生物过程的系统中实施该预测算法。物理、化学或者生物过

程可以是工厂或者其它设施中的工业过程。例如，预测算法可以用于帮助监视工业制造过程或者帮助监视来自工业过程的排放。备选地，可以在消费产品中使用预测算法，例如在汽车中使用以监视排放。预测算法的其它实施也是有可能的。

图3示意地示出了示例处理系统200，在该系统中可以实施预测算法。处理系统200可以是工业设施、消费产品或者任意其它系统，在该系统中物理、化学或者生物过程发生。在图3中，将被监视的特定物理、化学或者生物过程示意性地示为过程202。可以使用图3中表示为硬件传感器204-208的一个或者更多硬件传感器来获得过程202的多个可测量参数的测量值。依赖于过程202的性质，可测量参数可以包括例如温度、压力、流速等等。优选地，可测量参数是在过程202在线时硬件传感器204-208可以测量的参数。尽管图3示出了监视过程202的三个硬件传感器，但是应当理解，可以使用更多或者更少数量的硬件传感器。

硬件传感器204-208可以与推理传感器210通信地连接。推理传感器210可以配置为使用由硬件传感器204-208提供的可测量参数的测量值，来确定过程202的附加的目标参数的推理值。推理传感器210可以使用一个或者更多预测算法来进行该确定。优选地，这些预测算法中的至少一个是通过上述遗传规划技术并使用来自过程202的历史数据而开发的，从而在精确性、复杂度和平滑性方面是优化的。但是，推理传感器210还可以使用以其它方法开发的其它预测算法。

因此，推理传感器210可以使用以硬件传感器204-208提供的测量值为基础来预测目标参数的当前值的至少一个预测算法。在一些实施中，推理传感器210可以简单地将预测值作为推理值。备选地，推理传感器210可以应用算法，以基于预测值来确定推理值。在推理传感器210使用多于一个的预测算法，从而获得多于一个的预测值的情况中，推理传感器210可以通过选择预测值之一、对预测值进行平均或者以某种其它方式，来确定推理值。

推理传感器210可以向控制系统212提供目标参数的推理值。然后控制系统212可以使用推理值来确定是否采取附加行动。例如，控制系统212可以将推理值与一个或者更多预设值进行比较，以确定目标参数

是否在范围之外。如果目标参数在范围之外，控制系统212可以发起对过程202的控制，例如，尝试将目标参数带回至范围内。备选地或者附加地，如果发现目标参数在范围之外，控制系统212可以记录该推理值或者触发警报。

推理传感器210和控制系统212都可以实施为执行适当的计算机代码的计算机系统。从而，针对推理传感器210的计算机代码可以包括针对过程202、使用遗传规划开发的一个或者更多预测算法。推理传感器210以及控制系统212也可以集成在一起。

因此，处理系统200可以基于过程202在线时由硬件传感器能够容易测量的参数，来有利地监视过程202的、在过程202在线时可能难以测量的目标参数。然而，应当理解，使用硬件传感器来获得针对推理传感器210的输入参数的当前值仅作示例之用。推理传感器210还可以用其它方法来获得一个或者更多参数的当前值，比如从其它推理传感器获得。

5、VOC排放的预测算法

使用上述三维Pareto-front遗传规划技术开发了用于预测来自化学工厂的挥发性有机化合物（VOC）排放的预测算法。获得该工厂的历史数据集合，该集合包括对VOC排放的测量以及在相应时间进行的对多个不同过程变量的测量。过程变量是密度、速率、挥发性、真空、温度、料斗重量（hopper weight）、熔融指数（melt index）以及颗粒水温（pellet water temperature）。

从历史数据中取除明显的野值。然后将该历史数据分为两个集合。选择由251个数据点构成的一个集合用于模型开发（训练数据）。选择由108个数据点构成的另一个集合用于模型验证（测试数据）。为了测试预测算法的外推能力，选择测试数据，使得有20%在训练数据的范围之外。

图4示出了用于三维Pareto-front遗传规划的设置参数。如图5所示，从可用的函数的菜单中选择用于遗传规划的函数集合（由复选标记来指示选择的函数）。通过将每一个节点的复杂度加权求和，来确定复杂度，节点的复杂度加权是从属于该节点的所有子树中节点的数量。通

过对上述将每个节点函数拟合至规定容差内而所需的多项式的次数进行求和，来确定非线性。

在遗传规划的完成时，从100个算法的群体中仅找到非常少的算法在最优精确性、复杂度以及平滑性方面位于Pareto-front上。选择下述Pareto-front算法用于验证：

(2) $VOC=8.0746 + 6.0157e-14*[速率^{1.953} * 料斗重量 * (真空 + 颗粒水温)]$

如图6所示，用该预测算法进行的VOC预测与训练数据良好地吻合（相关系数=0.92884）。

接下来，使用测试数据来验证该预测算法。图7比较了由算法(2)预测的VOC排放与实际测量的VOC排放。可以发现，在训练范围之外高达20%的区域中（这些是在图7中样本70和85之间的VOC值，由箭头指示），该预测算法具有良好的外推性能。

为了比较的目的，使用相同训练数据开发了神经网络模型。然后使用与基于GP的预测算法的相同的测试数据，来验证该神经网络模型。然而，如图8所示，神经网络模型在训练范围之外的外推性能不好。

因此，该例子证明了遗传规划技术如何能够用于开发相对简单的分析表达式，该表达式从多个相关的输入变量联系到所需的输出变量（比如VOC排放）。相反，神经网络模型典型地是难以解释的“黑匣子”。另外，通过使用有利于线性的适合度准则（平滑性准则），遗传规划技术可以用于开发具有良好外推能力的预测算法。

6、结论

已经在如上描述了本发明的示例实施例。然而，本领域技术人员将理解，在不背离由权利要求限定的本发明的范围和精神的情况下，可以对这些实施例做出改变和修改。

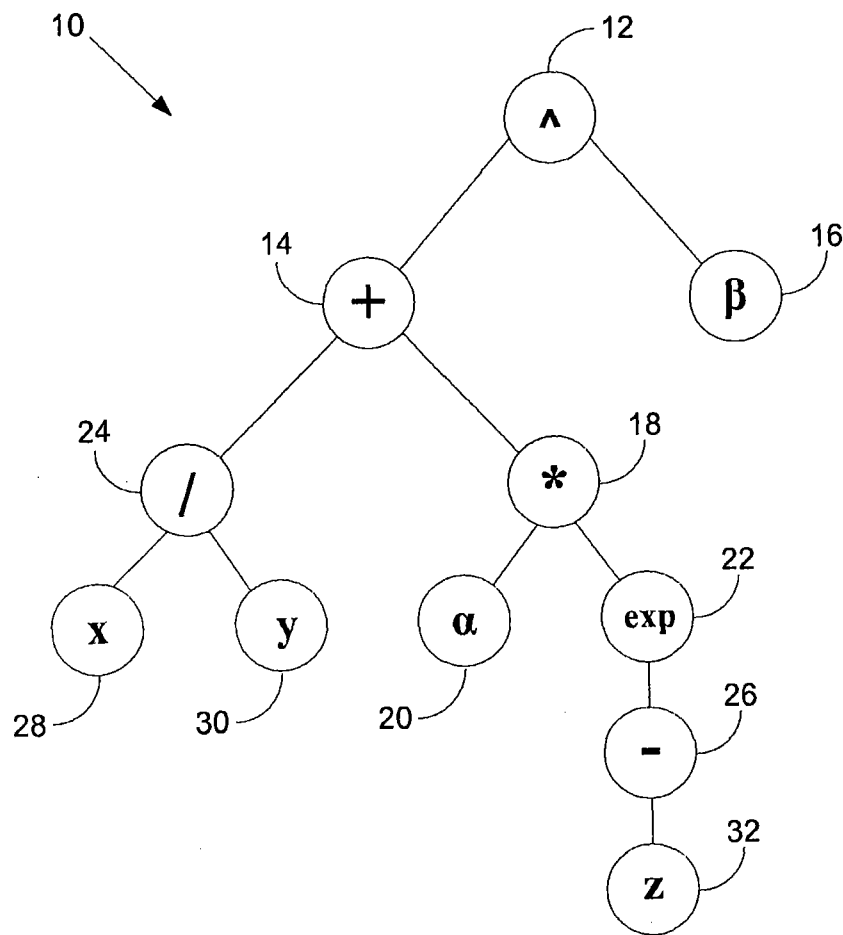


图 1

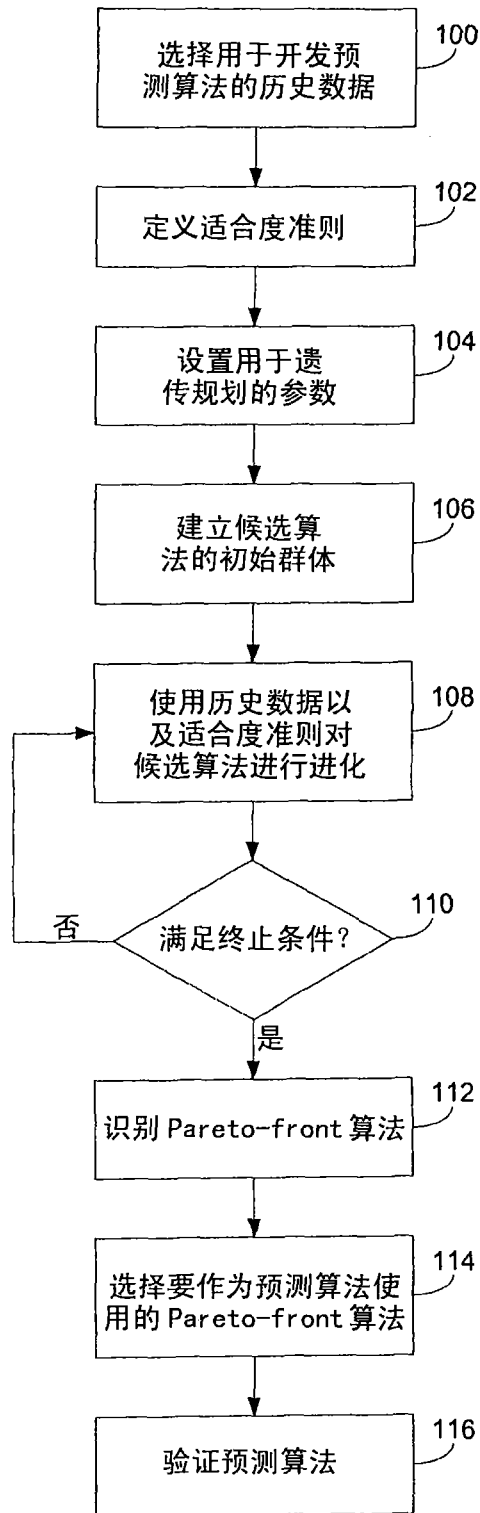


图 2

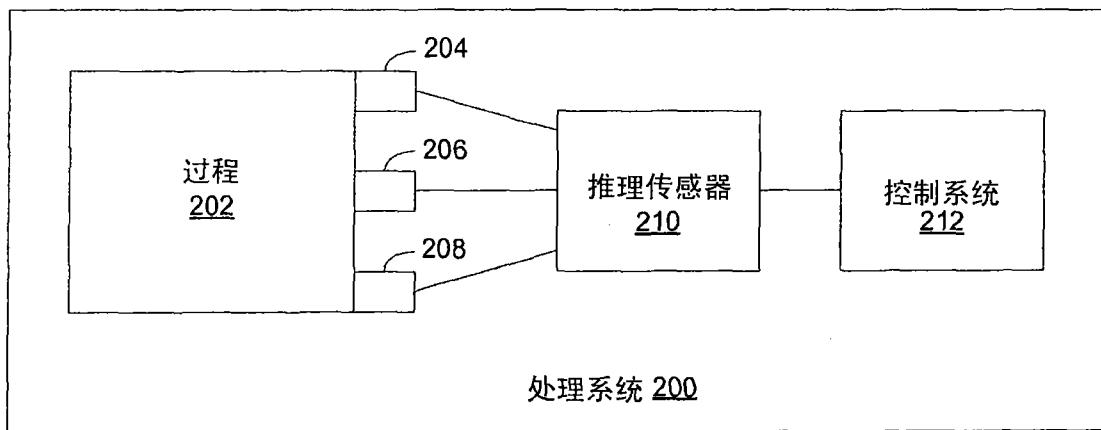


图 3

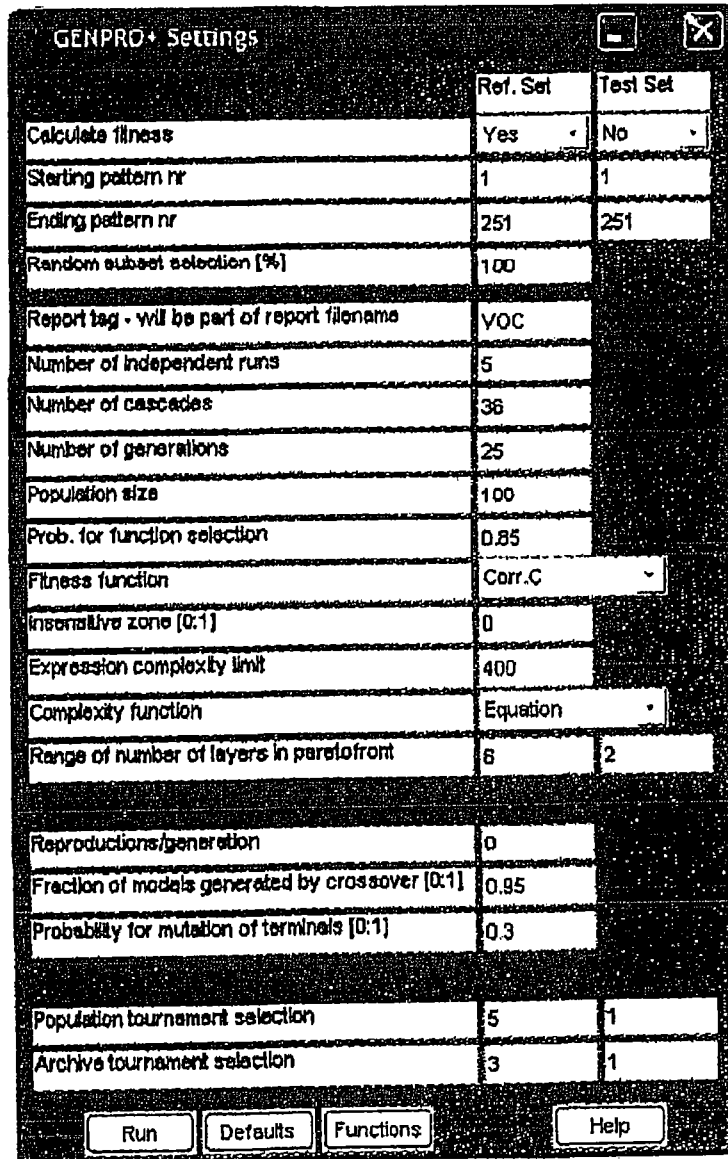


图 4

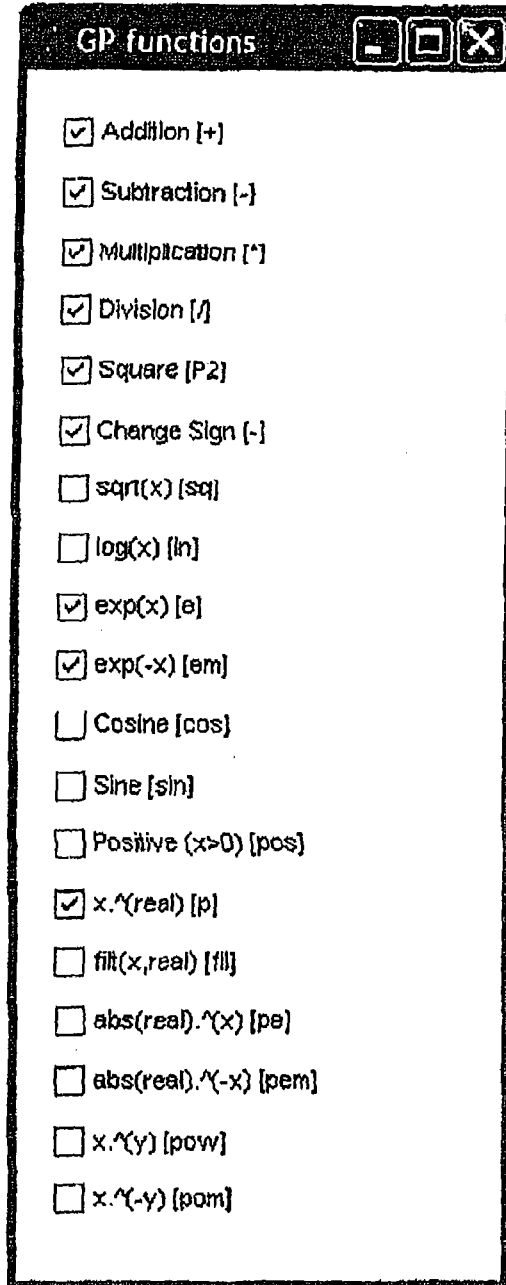


图 5

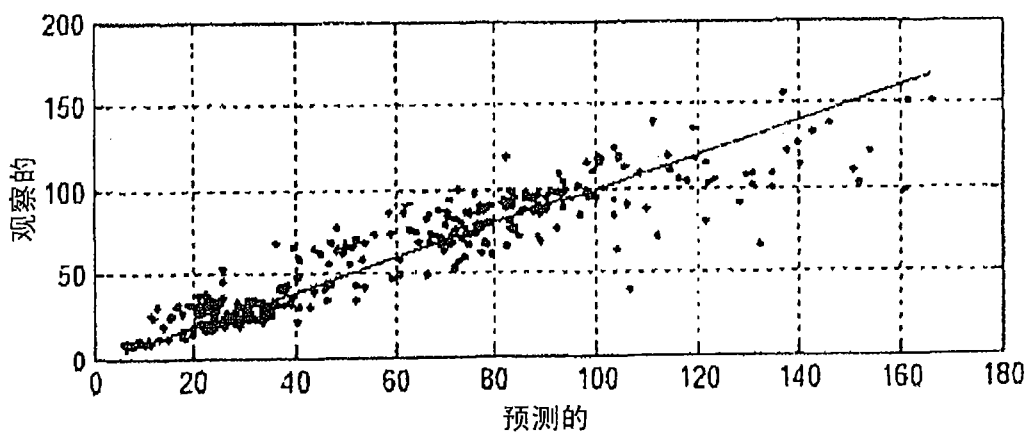


图 6

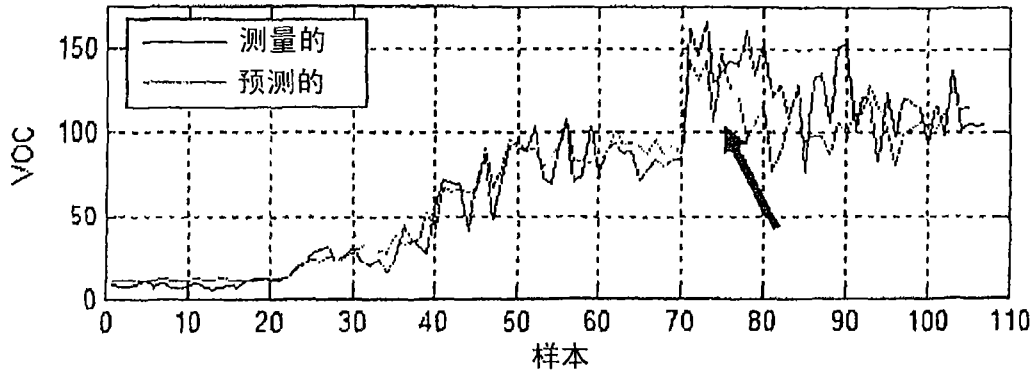


图 7

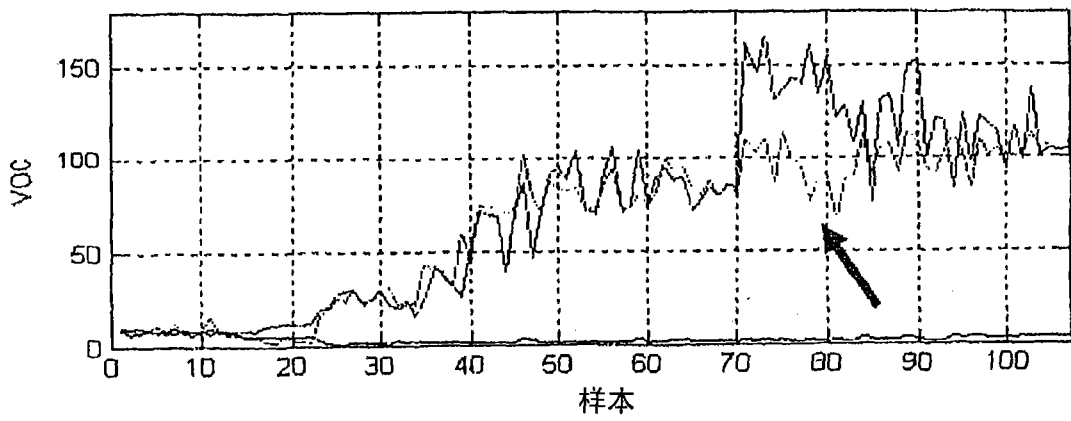


图 8