紧支径向基函数插值实现多维数据可视化

谭业浩,蒋志方,杜晓亮,孟祥旭
TAN Ye-hao,JIANG Zhi-fang,DU Xiao-liang,MENG Xiang-xu

山东大学 计算机科学与技术学院,济南 250101
School of Computer Science and Technology,Shandong University,Jinan 250101,China
E-mail:tyh.qingquan@yahoo.com.cn

TAN Ye-hao,JIANG Zhi-fang,DU Xiao-liang,et al.Visualization of multi-dimensional data with interpolation based on compactly supported radial basis functions.Computer Engineering and Applications,2010,46(9):220-223.

Abstract: This paper analyzes the organization structure for numerical forecasting data of the city air quality in space and in time,and constructs an overall frame for the multi-dimensional space data.The superiority and insufficiency are elaborated for several kinds of interpolation methods.Based on comparison,it has introduced the method of partial radial direction interpolation based on compactly supported radial basis functions into manage of multi-dimensional data.By the partial interpolation in the space dimension and in the time dimension,it has realized the restructuring of the multi-dimensional data.And it has realized the dynamic visualization in three dimensions for large-scale forecasting data of air quality based on the new multi-thread method with encapsulation of call-back functions.Experimental results demonstrate that the above method can satisfy the actual need of quality and the operating speed regarding large-scale data's visualization aspect.

Key words: compactly supported radial basis functions;multi-dimensional space;encapsulation;call-back function

摘 要:通过分析某城市空气质量数值预报数据的时空组织结构,构建出了多维空间数据的整体框架。论述了几种插值方法的优缺点,在比较的基础上,将新的紧支径向基函数局部径向点插值方法引入到多维数据处理中,在空间、时间维度上对数据进行局部插值,从而实现数据的重构。以新的基于封装回调函数的多线程方法实现了大规模空气质量预报数据的三维动态可视化。实验结果表明,以上方法应用于大规模数据可视化时,其质量和运算速度都能满足实际需要。

关键词:紧支径向基函数;多维空间;封装;回调函数
DOI:10.3778/j.issn.1002-8331.2010.09.063 **文章编号**:1002-8331(2010)09-0220-04 **文献标识码**:A **中图分类号**:TP391

城市环境空气质量是衡量人们生活质量的一个重要指标,进行环境空气质量的实时监测和数值预报对了解和管理空气质量状况有很大的帮助,研究相关数据的可视化方法为环境管理部门做出科学、及时、准确、直观的决策提供支持。

所谓城市环境空气质量数值预报,就是通过采用某种数值预报模式,计算出与时间和空间有关的污染项目的浓度预测值,所以空气质量数值预报数据集合具有时间和空间相关特性。一个城市空间尺度的 24 小时的数值预报数据,整体构成一个四维空间数据场,数据量十分庞大。对空气质量数值预报数据的动态可视化,就是对这个四维空间数据场的矢量和标量数据进行可视化处理,以帮助理解和表征预报数据所描述的环境空气质量状况。

由于数据维数的增加,数据存取、维护等远比传统二维、三维复杂得多,对传统的空间数据处理方法进行简单的扩展,已无法满足多维空间数据处理的要求,难以解决或回答现实应用领域多维空间数据处理提出的问题。传统三维可视化系统的解决思路是:原始数据->海量数据加工->数据库引擎->数据显示。由于多维空间数据量极为庞大,组织与应用也相对较难,目前只能借助于复杂的数据库系统来管理和维护。每次获取数据都必须与数据库打交道,增加了系统的开销,在可视化过程中显得尤为突出。为更好地解决多维数据场可视化问题,采取了以下方式,首先将原始数据组织优化成为具有某种结构化的数据形式,然后抽取部分数据组成数据框架,最后选取适当的插值方法动态地生成可视化数据。解决思路是:原始数据->N 维数据重构->N 维数据框架->动态再生数据->数据显示。实验结果表明:方法具有效率高、速度快、图像质量好等特点。

1 空气质量预报数据组织与结构分析

使用来自某市环境保护监测站的空气质量数值预报数据。其输出是一个大气污染物时空定量浓度预报数据文件,该数据文件描述了该市空气质量状况的如下几个方面[1]:(1)空间尺度(X,Y,Z),城市水平格网覆盖范围为本市辖区 79 km×69 km,分

作者简介:谭业浩(1976-),男,硕士研究生,主要研究领域为可视化技术与应用;蒋志方(1961-),男,副教授,主要研究领域为人机交互与虚拟现实,科学与信息可视化,数据挖掘等;杜晓亮(1979-),男,主要研究领域为可视化技术与应用;孟祥旭(1962-),男,博导,教授,研究领域为人机交互与虚拟现实,CG&CAD,科学与信息可视化,网格与协同工作。

收稿日期:2008-09-23 **修回日期**:2008-12-25

辨率为 1 km×1 km,并将对流层空间高度划分为 12 个层。(2)时间分辨率及预报时效(T),预报系统产生预报数据的时间分辨率为 1 h;预报时效为 24 h。(3)预报项目(P),包括 6 种污染物和 3 种气象参数分别为:SO_2,NO_2,NO,CO,PM10,O_3 和风速,温度,湿度。可以把城市空间看作是若干具有不同高度的空间区域,而每一层又可看作是由若干个 Volume 构成。图 1 描绘了城市空气质量时空定量浓度预报数据的空间构成。

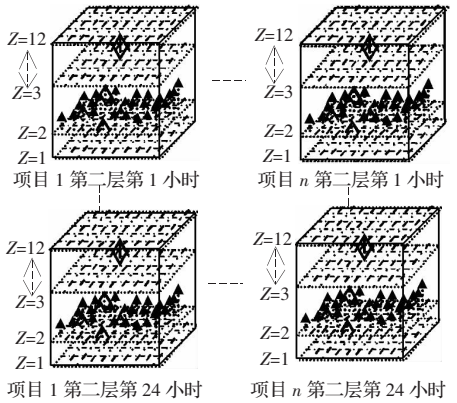


图 1　空气污染物时空定量浓度的数据组织结构图

通过分析,将数据组织成具有五维的数组结构 *Data*[*X*][*Y*][*Z*][*T*][*P*]。其中:*X*/*Y*/*Z* 分别为水平纵、横向坐标和高度坐标,其中水平坐标变化范围分别为 1~69、1~79,单位为 km,高度层次变化范围为 1~12,单位为层;T 为时间,单位(h),变化范围为 1~24;P 为预报项目,有 9 个项目,变化范围为 1~9。该数组的一个元素就表示在给定空间坐标和时间点的指定污染物的浓度值。在图 1 中,z=2 表示在高度上处第二层;带箭头线段表示每一网格点上污染物浓度值,将这些箭头代表的浓度值在横向、纵向做二次插值,得到该层浓度值的插值曲面;将插值出的曲面在 1~24 h 之间再进行插值,得到时间序列上的连续曲面。

2　空气质量预报数据可视化目标

通过对上述空气质量预报数据的结构分析,可以采用比较简单的方式在指定层内进行三维数据场的静态可视化。但要更好地了解每一污染物在 24 小时内的浓度变化过程,或者允许用户交互地进行同一污染物在不同层次上浓度值随时间变化的比较,如图 2 是某一污染物在 1 层与 2 层浓度值随时间变化的情况。当然也可以进行不同污染物同一层次,或者同一污染物同一层次、不同区域的浓度值随时间变化的比较,让研究人员更细致、更形象地观察。要实现上述功能,除了在三维空间表现出不同点位的属性值,还要在时间维度上确定每个点位的变化。

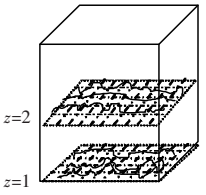


图 2　污染物 1、2 层上浓度比较图

由以上分析,确定动态显示的条件:已知观测项目 P,高度层Z,平面区域(X,Y)内每个网格单元浓度值。要求展示项目 P 的浓度在当天 1~24 h(T)、区域(X,Y,Z)内的渐变过程。要实现动态显示需要在原本庞大的预报数据基础上导出更为密集、细致的数据点,因为数据采集周期是 1 h,可以通过插值获得中间数据。

常用于气象要素的空间插值方法[2]有:距离权重法、多项式插值法、克里金法、样条插值法。这些方法中,距离权重法最简便,插值结果也比较粗糙,如果采用距离平方反比法容易产生牛眼现象;多项式插值的物理意义不是很明确,容易得出难以解释的值;样条插值对一些限定的点值,通过控制估计方差,利用一些特征节点,用多项式拟合的方法来产生平滑的插值曲线,多用于时间序列插值,此方法计算量较大;克里金法产生于地质采矿中的品位估计,亦能提供最佳线性无偏估计而逐渐被广泛运用于需要空间插值的诸多领域,同样因算法复杂,耗时较多。

采用基于径向基函数(Radial Basis Functions,简称 RBF)[3]插值方法。径向基函数的自变量只有一个表示"距离"概念的径向量 $\boldsymbol{r}$,相当简便直观,尤其在高维问题中更具优势。它最适合应用于运算过程依赖很多变量和参数;有大量的数据影响运算结果;数据点在定义域是散乱的等问题范畴。

3　径向基函数插值

3.1　径向基函数插值基础知识

径向基函数又称距离基函数,是一类特殊的函数,它以空间距离 r 为基本变量,具有形式简单,各向同性等优点,因此非常适合在数值计算中使用,可以达到很高的精度。径向基函数插值已被大量应用于地质勘探、外形设计、水纹学、绘图、遥感等广泛领域。径向基函数插值的优势:

(1)插值计算是在多维空间进行的,一般的插值算法实现复杂度高;径向基函数插值可以在任意维度进行插值,维数增加算法复杂度没有太大变化;

(2)径向基函数在规则的曲面上插值计算具有简单、有效、精确度高的优点;特征点越多插值越精确,一般的插值方法复杂度会成指数增加,但使用紧支径向基函数插值时会在很大程度上减少运算的复杂度,大大地减少了系统的开销[4]。

径向基函数插值模型[5-6]:给定数目为 n 的散乱点,$F=\{f(P_i)\}$ 在 $X=\{P_i\}=\{(x_i,y_i)\}\subset\Omega$ 当 $i=1,2,\cdots,n$ 时成立,其中 $\Omega\subset R^N$,如果 f 在 Ω 上连续,模型可以近似表示成:

$$I(f)=\int_{\Omega}f(P)\mathrm{d}P \tag{1}$$

确定合适的径向基函数 ϕ:$[0,+\infty)\to R$,在点集合 P 上构建 RBF 插值:

$$\sum_{j=1}^{n}c_j\phi_j(P)=s(P)\approx f(P) \tag{2}$$

其中 $P=(x,y)\in\Omega$,$s(P_i)=f(P_i)$,$i=1,2,\cdots,n$。

函数 ϕ 线性拟合为 $\phi_j(P)=\phi_j(P;\delta):=\phi(|P-P_j|/\delta)$,$|P-P_j|$,指示欧几里德距离,$\delta$ 与数据密度相关的参数。系数 $C=\{c_j\}$ 由插值方程 $\boldsymbol{A}C=F$ 求得,矩阵 $\boldsymbol{A}=A_{x,\phi}=\{\phi_j(P_i)\}$,$1\leqslant i,j\leqslant n$,此时矩阵 $\boldsymbol{A}$ 是否为非奇异矩阵,主要看函数 ϕ。

目前应用中主流的径向基函数[5]如表 1 所示。

表 1　径向基函数表

函数	$\phi(r)$
Gaussians(G)	$\exp(-r^2)$
Duchon 的 ThinPlate 样条(TPS)	$r^2\log(r)$
Hardy 的 MultiQuadrics(MQ)	$(1+r^2)^{1/2}$
Inverse MultiQuadrics(IMQ)	$(1+r^2)^{-1/2}$
Wendland 的紧支径向基函数(W2)	$(1-r)^4+(4r+1)$

因为 G、IMQ、W2 对应的矩阵 $\boldsymbol{A}$ 在任意条件下正定，所以称它们为正定(PD)，TPS 和 MQ 对应的矩阵在某些条件下正定，称条件正定(CPD)，所以对应的条件正定 RBF 插值模型表示成：

$$s(P)=\sum_{j=1}^{n}c_j\phi_j(P)+p_m(P) \tag{3}$$

$p_m(P)$是次数≤m(维度)的多项式。

3.2 MQ 插值方程

复二次函数(MQ)插值是最早提出且应用得最为成功的一种径向基函数插值法，它的方程表示为：

$$f(x,y,z)=\sum_{i=1}^{n}\{w_i\sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}\}+C_o+xC_x+yC_y+zC_z \tag{4}$$

x,y 对应地面检测点坐标，z 对应显示时间，w_i 表示采样点(x_i,y_i,z_i)的贡献系数，c_0,c_x,c_y,c_z 为条件多项式系数，满足

$$\sum_{i=1}^{n}w_i(C_o+x_iC_x+y_iC_y+z_iC_z)=0 \tag{5}$$

将采样点(x_i,y_i,z_i)，$i=1,2,\cdots,n$ 代入方程式(4)、(5)，求得 $w_i,i=1,2,\cdots,n$ 和 c_0,c_x,x_y,c_z。任意给出 1~24 h 内的一个点(x,y,z)，便可得到该点浓度值$f(x,y,z)$。

实验中发现，由于 MQ 插值的求解矩阵为满阵(因此称全域径向基函数插值)，计算量十分庞大，不利于大规模数据的计算。当数据点较多时，会让用户等待时间过长，当参与插值的数据点接近 5 000 个时，系统会因资源不足停止运行。

3.3 紧支径向基函数插值方程

采用紧支径向基函数插值方法时，求解矩阵为稀疏带状分布(因此称为紧支撑)，求解过程相对简单，因此更适合解决大型问题。目前正在研究的主要正定紧支径向基函数[7]有：

第 1 种 $(1-r)_+^4(4+16r+12r^2+3r^3)\in C^2\cap PD_3$

第 2 种 $(1-4)_+^6(6+36r+82r^2+72r^3+30r^4+5r^5)\in C^4\cap PD_3$

第 3 种 $\frac{1}{3}+r^2-\frac{4}{3r^3}+2r^2\ln r(\text{if } 0\leqslant r\leqslant 1)$

第 4 种 $\frac{1}{15}+\frac{19}{6r^2}-\frac{16}{3r^3}+3r^4-\frac{16}{5r^5}+\frac{1}{6r^6}+2r^2\ln r(\text{if } 0\leqslant r\leqslant 1)$

第 5 种 $(1-r)_+^6(35r^2+18r+3)\in C^4\cap PD_3$

第 6 种 $(1-r)_+^8(32r^3+25r^2+8r+1)\in C^6\cap PD_3$

上述公式中 PD_3 表示在三维以下空间为正定函数。采用第一种形式的局部径向点插值[8]方法。在插值求解时，由阈值确定一个半径为 R 的球，每个采样点的作用域只局限在以自身为中心的球内，w 表示采样点的属性值在这个球内的权值(贡献系数)。

假设第 i 个采样点在自己的包围球内有 m 个采样点，记为$(x_{ij},y_{ij},z_{ij})j=1,2,\cdots,m$，每个球内点距球心的距离 $d_{ij}=\sqrt{(x_j-x_i)^2+(y_j-y_i)^2+(z_j-z_i)^2}$，采样点的球内权值距离 $g_{ij}=(1-r_{ij})^4\times(4+16r_{ij}+12r_{ij}^2+3r_{ij}^3)$，其中 $r_{ij}=d_{ij}/R$，可以得到径向基函数插值的基本方程组：

$$f(x_i,y_i,z_i)=\sum_{j=1}^{m}\{w_jg_{ij}\} \tag{6}$$

其中 $i=1,2,\cdots,n$，所有 n 个采样点依次代入，令 $\boldsymbol{F}^{\mathrm{T}}=[f_1,f_2,\cdots,f_n]$，$\boldsymbol{W}^{\mathrm{T}}=[w_1,w_2,\cdots,w_n]$。

$$\boldsymbol{A}=\begin{bmatrix}g_{11} & g_{21} & \cdots & g_{n1}\\ g_{12} & g_{22} & \cdots & g_{n2}\\ \vdots & \vdots & & \vdots\\ g_{1n} & g_{2n} & \cdots & g_{nn}\end{bmatrix} \tag{7}$$

当 $d_{ij}>R$ 时，$g_{ij}=0$，则有：$\boldsymbol{AW}=\boldsymbol{F}$。

求解式(7)可以得到每个采样点在自己球内的权值 w_j。

一旦获得每个采样点 j 在自己球内的权值 w_j，依据基本方程组，就可以求解任意点 $p(x,y,z)$的浓度值$f(x,y,z)$。以(x,y,z)为球心，以 R 为半径的球内 m 个采样点，设采样点 j 在球内，它距球心的距离 $d_{pj}=\sqrt{(x-x_j)^2+(y-y_j)^2+(z-z_j)^2}$，$r_{pj}=d_{pj}/R$ 采样点的球内权值距离 $g_{pj}=(1-r_{pj})^4\times(4+16r_{pj}+12r_{pj}^2+3r_{pj}^3)$，则点 $p(x,y,z)$的浓度值表示为：

$$f(x,y,z)=\sum_{j=1}^{m}w_jg_{pj} \tag{8}$$

由于方程的求解矩阵 $\boldsymbol{A}$ 是稀疏矩阵，求解方程的时间复杂度较小。当采样点数目一定时，算法时间复杂度主要由阈值确定的球半径 R 来决定，R 值越大，球内的采样点数越多，插值效果就越好；R 值越小，球内的采样点越少，求解方程的速度就越快。可以通过调整阈值在插值质量和时间复杂度之间取得平衡。

4 基于封装回调函数的渲染线程方法

多线程方法实现动画的过程：主线程固定显示一个点数据集(A)，在次线程中实现数据转换的操作，一般使用回调函数完成次线程的处理过程。一帧数据抽取完成，加载到点数据集(A)中，发送一个刷新消息给窗体，实现画面动作。由于数据处理、屏幕刷新同步进行，在多处理器的计算机上运行时，速度会有明显提高。本方法适合帧与帧之间需要做大量数据处理的情况。

回调函数在系统中单独注册，使其独立于任何对象实例[9]，虽然很好地解决了主、次线程的通信问题，但次线程的过程声明和实现都在类的外部，而且必须通过全局变量来完成主线程与次线程的通信，不灵活也不安全。为了实现更好的封装，这里采用一种新的方法来改变多线程的实现机制。方法实现的关键是处理好主、次线程的同步问题，这里采用临界区方式实现效率较高。

使用基于封装回调函数的多线程实现方法，是把回调函数定义为类的一个静态方法，它的声明和过程实现可以封装在类中[10]。

具体步骤：(1)在视类头文件中声明一个回调函数作为类的静态方法成员；(2)将数据的转换操作放在该静态方法的实现中；(3)程序在启动次线程时，将线程的处理函数指定为视类的静态方法，并将视类的指针作为参数传递给处理函数。

举例说明：

(1)在 CMyView 类的头文件中声明一个回调函数作为类的静态方法成员，同时声明两个成员变量来标示线程：

```
DWORD threadID;//线程的 ID
HANDLE pThread;//指向线程的指针
static DWORD CALLBACK CalcuThreadProc(LPVOID pV);
```

(2)在 CMyView 类的执行文件中实现回调函数过程

```
DWORD CALLBACK CMyView::CalcuThreadProc(LPVOID pV)
{//使用 pView 访问类的非静态成员，pV 是作为参数传入的CMyView 类的指针
CMyView*pView=(CMyView*)pV;
  if(数据处理条件满足)
    {完成数据的处理功能，通知刷新屏幕;}
  else
  ExitThread(0);//退出线程
}
```

（3）在需要启用线程的地方使用语句：

pThread=CreateThread（NULL，0，（LPTHREAD_START_ROUTINE）CalcuThreadProc，（LPVOID）this，0，&threadID）；

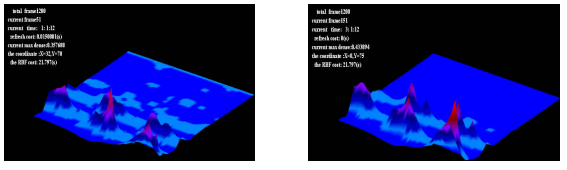
当次线程的处理过程被调用时，也就是调用了类的静态方法，由于传入了视类的指针，方法可以直接访问视类的成员，包括静态成员和方法。解决了静态方法无法有效访问非静态成员的问题，简化了程序，完美地解决了多线程的封装。

5　实验结果分析

使用的算法是以 vs2005 为平台，运行在 P4 双核 2.33 GHz，2 GB 内存的微机上，实验数据采用 2006 年 12 月 16 日的空气质量数值预报数据。

实验中由于全域径向基函数插值耗时较多，数据点数增加一倍，计算时间会成指数上升，为减少等待时间，采样点只取前 5 h 数据，平面横向（69）每隔 4 个采样点取一点，纵向（79）每隔 5 个采样点取一点，即组成框架的采样点共有 5×18×17=1 530 个点。

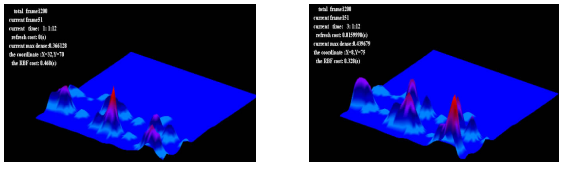
（1）采用全域径向基函数插值，插值耗时 21.797 s，效果如图 3 所示，从图 3 中可以看出图像显示信息完整，画面细腻，无闪烁，但由于算法耗时较多，画面延迟时间长，播放不够流畅。



（a）12 月 16 日 1 点时的数据　（b）12 月 16 日 3 点时的数据

图 3　采用全域径向基函数（MQ）插值法效果图

（2）采用紧支径向基函数插值法，经多次实验比较，最后取阈值 8，此时插值耗时 0.328 s。效果如图 4 所示，从图 4 中可以看出，图像显示信息与全域插值相比同样十分完整，细节部分的展示与图 3 相比略有变形，一些尖峰部分数据被消减，出现了圆滑过渡。但由于算法速度快，画面连续，图像稳定，满足实时要求。



（a）12 月 16 日 1 点时的数据　（b）12 月 16 日 3 点时的数据

图 4　采用紧支径向基函数插值法效果图

6　结论与下一步工作

以基于 MFC 的 OpenGL 作为开发工具，通过对环境空气质量预报数据的分析，研究探索了大规模海量数据动态可视化的方法，将紧支径向基函数插值应用到动态显示技术中，通过面向对象思想，实现基于封装回调函数的多线程编程方法，优点：

（1）由于紧支径向基函数插值算法的复杂度，不会因为特征点的增多而迅速增大，因此可以应对需要处理较大数据量的应用环境中；

（2）基于封装回调函数的多线程方法，使访问系统资源更加便利，程序封装性好，有利于模块化应用，具有很高的效率。

由于在算法效率和图像质量上的优势，下一步工作将在城市环境空气质量预报数据可视化系统中应用紧支径向基函数插值和封装的回调函数多线程方法。

参考文献：

[1] 丛中兴，蒋志方，王强，等.基于 GIS 的环境空气质量数值预报数据可视化系统[C]//2003 全国软件与应用学术会议（NASAC）论文集，2003：679-684.

[2] 林忠辉，莫兴国，李红轩，等.中国陆地区域气象要素的空间差值[J].地理学报，2002，57(1).

[3] Buhmann M D.Radial basis functions[M]//Acta Numerica，Cambridge：Cambridge University Press，2000：1-38.

[4] 唐峰，王洵，董兰芳，等.基于径向基函数多步离散数据插值的人脸变形研究[J].计算机工程与应用，2003，(24).

[5] Sommariva A，Vianello M.Numerical cubature on scattered data by radial basis runctions[J].Computing Archive，2006，76：295-310.

[6] 陈荣华.径向基函数拟插值理论及其在微分方程数值解中的应用[D].上海：复旦大学，2005.

[7] Zhang X，Song K Z，Lu M W，et al.Meshless methods based on collocation with radial basis functions[J].Computer Mech，2000，26：333-343.

[8] Xiao J R.Local heaviside weighted MLPG meshless method for two-dimensional solids using compactly supported radial basis functions[J].Comput Methods Appl Mech Engrg，2004，193：117-138.

[9] 陈涛，魏朗.人-车-路视景仿真系统中信息的动态显示[J].计算机工程与应用，2007，43(7).

[10] 刘书良，韩力，罗辞勇.回调函数的 C++封装[J].重庆工商大学学报：自然科学版，2004，21(6).

（上接 214 页）

表 7　原始数据与模糊化数据分类的比较表

		1	2	3	4	5	6	7	8	9	10	（1）	（2）
原始数据	阈值 0.999	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_9	新故障
模糊化数据	阈值 0.99	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_9	y_5

改进的方法，它可以克服传统 ART2 型神经网络输入序列以及模式漂移的缺点。使用一种新的方法来计算相似度。结果表明改进过的 ART2 神经网络可以成功地应用于故障类型的诊断。

参考文献：

[1] Karthikeyan B，Gopal S，Venkatesh S.ART2-An unsupervised neural network for PD pattern recognition and classification[J].Expert Systems with Applications，2006，31：345-350.

[2] Wann Chin-der，Thompoulos C A.A comparative study of self-organizing clustering algorithms dignet and ART2[J].Neural Network，1997，10(4)：737-753.

[3] Lv Xiu-jiang，Zhang Qi-wen.A new neural network classifier based on ART theory[C]//IEEE Proceedings of the 4th International Conference on Machine Learning and Cybernetics，2005：18-21.

[4] Laura I，Burke.Clustering characterization of adaptive resonance[J].Neural Networks，1991，16：485-491.

[5] Chen S J，Cheng C S.A neural network based cell formation algorithm in cellular manufacturing[J].Int J Prod Res，1995，33：293-299.