基于属性聚类网络和径向基函数的融合预测

程乾生 武连文 王守章

(北京大学数学科学学院信息科学系, 北京 100871. Email: qcheng@pku.edu.cn)

摘要 研究属性聚类网络在构造径向基函数预测模型时的应用, 并提出了一种拟自组织算法来建立非线性融合预测模型, 通过典型例子检验预测效果.

关键词 混沌时间序列 径向基函数 属性聚类网络 拟自组织算法 融合预测

本文研究混沌时间序列的非线性模型和预测问题, 即相空间中的一串迭代序列如何构造一个非线性映射来表示这一动力系统, 如果能够构造出来的话, 那么这个非线性映射就可以作为我们的预测模型.

确保这种方法成立的理论基础是重构相空间理论[1]. 对于一个确定的动力系统的观测函数 $s(t)$, 经过采样得到一个单变量的时间序列 s_t:

$$s_t = s(t\cdot t_s), \quad t = 1, 2, \ldots, N_s, \tag{1}$$

其中 t_s 为采样间隔. 利用时滞方法, 可以通过这一单变量时间序列在 m 维 Euclid 空间中构造出一条轨道 $x(t)$:

$$x(t) = (s_t, s_{t+t_d}, \cdots, s_{t+(m-1)t_d}), \tag{2}$$

其中 m 为嵌入维数, t_d 为时滞. 通常 $t_d = k$, k 为某一正整数, 有时也称 k 为时滞. 一般地, 只要 $m \geqslant 2d+1$ (其中 d 表示原动力系统相空间的维数), 那么我们得到的 $x(t)$就是原动力系统相应一条轨道到$\mathbb{R}^m$ 空间的嵌入. 由此可以得到$\mathbb{R}^m$ 上的一个动力系统 F: $\mathbb{R}^m \to \mathbb{R}^m$,满足

$$x(t+1) = F(x(t)). \tag{3}$$

从而可以进一步得到一个函数 f: $\mathbb{R}^m \to \mathbb{R}$, 使得

$$s_{t+1+(m-1)t_d} = f(x(t)) = f(s_t, s_{t+t_d}, \cdots, s_{t+(m-1)t_d}). \tag{4}$$

所以说如果能够根据已知的时间序列 s_t 求出满足(3)或(4)式的 F 或 f 的一种近似形式 AF 和 Af, 那么就可以得到一个 s_t 的非线性预测模型.

到目前为止, 已经发展了多个基于上述思想的预测模型, 主要的有局部预测模型、全局预测模型、神经网络预测模型、径向基函数预测模型等.

本文主要利用基于属性聚类网络和径向基函数的融合预测模型来进行预测. 首先利用属性聚类网络来选取径向基函数的中心点集, 然后利用融合预测的方法来建立我们的模型, 并通过太阳黑子的例子来研究模型的效果.

1 基于径向基函数非线性预测模型和属性聚类网络

为了方便, 引入如下的规定: 把重构所得的向量集合$\{x(t)\}$ 分为两组, 一组用于建立模型, 称之为学习集合(the learning set)或训练集合, 记为 $\{x(t)\}_{t=1}^{N_L}$; 另一组用于模型预测效果的检验, 称之为检验集合(the testing set), 记为 $\{x(t)\}_{t=N_L+1}^{N_L+N_T}$. 另外, 引入正规化的均方误差用于检验模型拟合和预测的精度, 其中拟合的正规化的均方误差定义为(以一步预测为例)

$$s_{\mathrm{f}}^{2}=\left\{\sum_{t=1}^{N_{\mathrm{L}}}(Af(x(t))-s_{t+1+(m-1)t_{\mathrm{d}}})^{2}/N_{\mathrm{L}}\right\}\Big/\mathrm{Var}(s). \tag{5}$$

预测的均方误差定义为

$$s_{\mathrm{p}}^{2}=\left\{\sum_{t=N_{\mathrm{L}}+1}^{N_{\mathrm{L}}+N_{\mathrm{T}}}(Af(x(t))-s_{t+1+(m-1)t_{\mathrm{d}}})^{2}/N_{\mathrm{T}}\right\}\Big/\mathrm{Var}(s). \tag{6}$$

上面两式中的 $\mathrm{Var}(s)$均表示$\{s_t\}$的方差，$s_{\mathrm{p}}^{2}=0$ 表示预测结果与真值完全吻合，$s_{\mathrm{p}}^{2}=1$表示预测的效果并不比直接选取 $Af(x(t))$ (表示$\{s_t\}$的均值)好.

1.1 基于径向基函数的非线性预测模型

径向基函数模型是将所要求的函数 $Af(x(t))$表示成如下的形式:

$$Af(x(t))=\sum_{j=1}^{N_{\mathrm{c}}}l_j f(\|x-c_j\|), \tag{7}$$

其中$f(r):R^{+}\to R$ 是径向基函数，常用的径向基函数包括有 $f(r)=r,\ r^3,\ r^2\log r,\ \exp(-r^2/s^2)$等等，$\|\cdot\|$ 是 Euclid 距离，c_j ($j=1,2,\ldots,N_{\mathrm{c}}$)称为径向基函数的中心，$l=(l_1,\ l_2,\ \ldots,\ l_{N_{\mathrm{c}}})'$ 是待定系数.

将径向基函数插值方法用于混沌时间序列的预测问题最早是由 Casdagli 提出来的，具体来说,即用形如(7)式的函数来拟合$\mathbb{R}^{m+1}$ 空间中的一列点$(x(t),\ s_{t+1+(m-1)t_{\mathrm{d}}})$，其中 $t=1,2,\ldots,N_{\mathrm{L}}$，使得

$$\min\sum_{j=1}^{N_{\mathrm{L}}}(Af(x(t))-s_{t+1+(m-1)t_{\mathrm{d}}})^{2} \tag{8}$$

化为矩阵形式，得到

$$\boldsymbol{b}=\boldsymbol{A}\cdot l, \tag{9}$$

其中 $\boldsymbol{b}=(s_{2+(m-1)t_{\mathrm{d}}},s_{3+(m-1)t_{\mathrm{d}}},\cdots,s_{N_{\mathrm{L}+1}+(m-1)t_{\mathrm{d}}})^{\mathrm{T}}$, $\boldsymbol{A}_{ij}=f(\|x(i)-c_j\|)$, $i=1,2,\cdots,N_{\mathrm{L}}$, $j=1,2,\cdots,N_{\mathrm{c}}$, 求解问题(9)定出系数 l,即可得到预测函数 Af 的形式.

径向基函数网络模型的学习过程可分为两个独立的步骤，第 1 步是为每个隐单元选择适当的中心 c_j，第 2 步是调整权值向量使得输出值与希望值的误差达到最小.

Casdagli[2]将中心点 c_j 选取为训练集合中的点，并用相应的中心点集作为训练集合，此时(7) 式中的 $\boldsymbol{A}$ 为 $N_{\mathrm{c}}\times N_{\mathrm{c}}$ 方阵. 根据已有的结果，只要 c_j 互不相同，那么 $\boldsymbol{A}$ 就是非奇异的，于是求解线性方程组(9)即可得到唯一的 λ. 但是这种方法的缺点是它对每一个中心点都精确地拟合了，并且训练集合中的每一个点都被作为中心点，如果训练集合中的点很多，那么计算量就很大，而且这种方法对一些含噪声的数据常常导致拟合过度(overfitting). 对于一些近似周期或周期的数据，往往表现出近似奇异性，并且对于一些特殊的径向基函数，如 Gauss 型会导致 $\boldsymbol{A}$ 是近似奇异的，这时候预测的效果是很差的. 在第 2 个步骤的解决过程中，Broomhead 建议用最小二乘法求解问题(9)，此时用于训练的样本个数 N_{L} 大于中心点的个数 N_{c}:

$$l=\boldsymbol{A}^{+}\cdot\boldsymbol{b}, \tag{10}$$

其中 $\boldsymbol{A}^{+}$是 $\boldsymbol{A}$ 的 Moore-Pearose 广义逆矩阵.

近些年来，应用径向基函数于非线性预测问题的主要思路就是基于上述的想法，但是，如何选择中心点集 c_j 仍是一个需要深入研究的问题.

在本文中，我们考虑将用于非监督聚类的基于属性测度的网络模型用于选择径向基函数的中心点集的问题上来.

1.2 基于属性测度的无监督聚类网络算法

设给定的样本集合为 $\{x(t)\}_{t=1}^{N_L}$，其中 $x(t)$可以按照(2)式由单变量时间序列$\{s_t\}$重构而成. 在以往的预测中，事先要选定 N_c 个中心 c_j，但是都假定 $x(t)$的各个分量(或指标)在预测中所起的作用都是一样的. 而在实际过程中，我们知道，每个指标的作用实际上是不同的. 因此在预测时，要考虑到体现出指标本身的固有性质. 程乾生[3]提出将属性测度的方法贯彻到无监督聚类的整个算法中，得到 AKCN 算法，将此算法应用到中心的选择上来，较好地解决了上述问题. 该算法采用 Kohonen 自组织网络的形式，有几个独特的地方，具体算法的形式见文献[3].

2 基于优胜劣汰的拟自组织学习算法

对于某个问题，我们往往面对的是来自不同预测模型所提供的预测和分析结果. 对于我们所使用的径向基函数预测模型，当我们将要选择的中心点的个数或嵌入维数不同时，就得到了不同的预测模型. 从某种意义上讲，每个模型在一定假设条件下，在相应的准则下均是最优的，但是都有一定的局限性，并没有哪一个方法相对于别的方法而言具有最优性. 我们试验了几种模型，每一个模型在数据处理及不同准则方面均有其独到之处，都能从不同的角度为最终预测提供有用的信息，因此我们考虑到进行融合预测来改善最终预测精度.

引入如下规定：对预测量 Y，假设学习集合(训练集合)为 $\{x_i\}_{i=1}^{n}$，在 M 个预测模型中要确定 m 个预测模型 f_i, i=1, 2, …, m. 为模型叙述方便，不妨设 M 为偶数个，在第 i 个点 x_i 处 f_j 的预测值为 f_{ji}, i=1, 2, …, n, j = 1, 2, …, M，相应的观测值为 Y_i，所要建立的融合预测函数为 f_c，在该点的融合预测值为 f_{ci}，建立的融合预测模型为

$$f_c = l_1 \cdot f_1 + l_2 \cdot f_2 + \cdots + l_m \cdot f_m. \tag{11}$$

预测的结果就是要确定 l_i 使下式达到最小:

$$E = \sum_{i=1}^{n} \left| f_c(x_i) - Y_i \right|^2, \tag{12}$$

其中 l_i 满足约束条件

$$\sum_{i=1}^{m} l_i = 1. \tag{13}$$

为了确定 λ_i 和 f_j，提出如下拟自组织算法:

(ⅰ) 初始化将$\{l_i\}$赋予区间[0, 1]内的随机值，然后将其归一化，使其满足(13)式，提供总的学习次数 T，确定学习率$h(t)$的初始值$h(0)$ $(0 < h(0) < 1)$，终止阈值 Thv.

(ⅱ) 提供预测点 x_i，此时融合预测值为 f_{ci}，观测值为 Y_i，各个分预测值为 f_{ji}.

(ⅲ) 计算 f_{ji} 与 Y_i 之间的 Euclid 距离:

$$d_{ji} = | f_{ji} - Y_i |^2, \quad j = 1, 2, \cdots, M.$$

(ⅳ) (配对原则)把 d_{ji} 按从小到大依次排序，将最大的 d_{ji} 与最小的 d_{ji} 对应的 l_j 分为一组，次

最大的与次最小的分为一组，依此下去将所有的 l_j 都分在组内，第 j 个大的 d_{ji} 与第 j 个小的 d_{ji} 对应的 l_j 分在一组.

(v) (互补原则)进行 l_j 的调整，对分在一组内的 l_j，由于组内的 d_{ji} 是按照步骤(iv)分组的，因此对 d_{ji} 大的 l_j 减去 AS (=$h(t)\cdot l_j$). 反过来，将 d_{ji} 小的 l_j 加上 AS. 按此原则对所有的 l_j 进行修正，可以看到最后得到的 l_j 仍然满足(13)式.

(vi) 提供下一个预测点，返回步骤(iii)，直至将所有的预测点全部提供一遍.

(vii) 更新学习率 $h(t)$:

$$h(t) = h(0)(1 - t/T),$$

式中 t 为学习次数，T 为总的学习次数，$h(0)$为学习率的初始值.

(viii) 令 $t = t+1$，返回步骤(ii)，直至 $t = T$，或 $\sum(\Delta l_j)^2$ 小于终止阈值为止.

(ix) 当 M 个预测模型的权系数确定后，将其排序，如从最小的权系数开始求和，其值小于全部预测模型权系数总和的 10% (这是一个经验值，如同统计中的置信度一样，具体取法由实际问题和实验效果来确定)，则将其去掉，这时剩下的权系数所对应的预测模型为 m 个，再将其用前面的方法重新确定权系数. 此时算法结束.

算法保证了始终向距真实值差别最小的方向调整 l_j，之所以要满足(13) 式，是因为考虑到各个分预测值与真实值均在同一量级上，经过反复学习，可以使得 l_j 最后的空间分布能够正确反映预测模式的空间概率分布. 并且它的优点是便于自适应处理，而且它的思想是一种竞争学习，在学习过程中，可分为两个阶段，第 1 个阶段为粗学习与粗调整时期，大致确定权向量 l_j 的值，此时 $h(t)$保持较高值，一旦有了相对稳定的映射位置后，即进入第 2 精细调整阶段. 一般来说，在实际过程中，经过学习后，权向量的排列与分预测模型值的自然排列基本上趋于一致,即权向量趋于收敛.

我们已经知道不同径向基函数主要在于不同的中心个数与嵌入维数，由于融合预测模型是将多个模型融合在一起，因此我们选取不同的中心个数和不同的维数来建立不同的径向基函数预测模型，再将其融合在一起，如下所示建立融合预测模型:

(i) 首先用属性自组织神经网络确立中心点，每一个输出神经元 i 对应的最终的权值向量选作为预测模型的中心 c_i.

(ii) 用最小二乘法调整(7)式中的权值向量使得输出值与希望值的误差达到最小.

(iii) 重复(i)和(ii)步，每次用不同的中心个数和不同的维数，最后建立多个预测模型.

(iv) 用拟自组织算法来建立融合预测模型.

下面通过对太阳黑子数据的预测来说明上述模型的预测效果.

人们对太阳黑子的观察是从 17 世纪开始的，每年都有一个有关太阳黑子数目的平均纪录，至今已有近 300 个数据. 自从 Yule(1927 年)开始，太阳黑子数据作为一种标准建模数据在统计学领域里得到了广泛的研究，我们把所提出的算法应用于太阳黑子数据的预测中去，即利用从 1700 ~ 1920 年的数据用于建立模型，此时 $N_L = 218$. 用 1921 ~ 1955 年的一段数据和 1956 ~ 1979 年的一段数据检验模型，选取时滞参数 $t_d = 1$，嵌入到不同维数的 Euclid 空间中去. 采用的径向基函数模型基于(7)式，其中径向基函数为 Gauss 函数: $f(r) = \exp(-r^2/s^2)$. 由于中心个数 N_c 和嵌入维数 m 的不同就可以建立不同的预测模型，简记分预测模型为(m, N_c)，得到分模型为(3, 10), (3, 15), (3, 20), (3, 24), (3, 30), (4, 16), (4, 22), (4, 26), (5, 14), (6, 15). 然

后利用我们的算法将其融合起来进行预测，预测结果如下：

$$\mathrm{Var_f}\ (\mathrm{train}) = 128.7,$$
$$MSE_{1921\sim1955} = 155.6,$$
$$MSE_{1956\sim1979} = 195.4,$$

其中，$\mathrm{Var_f} = s_{\mathrm{f}}^2 \cdot \mathrm{Var}\ (s)$，$MSE = s_{\mathrm{p}}^2 \cdot \mathrm{Var}(s)$ (见(5), (6)式).

用我们的拟自组织计算方法，第 1 次迭代时，运行 285 次后收敛，其值列于表 1.

表 1 参数 l_j 的变化值

迭代次数	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	l_9	l_{10}
1	0.074 01	0.030 99	0.024 59	0.008 11	0.025 08	0.200 94	0.102 79	0.085 60	0.187 58	0.260 26
41	0.075 81	0.033 48	0.030 81	0.012 95	0.032 47	0.215 36	0.091 40	0.093 97	0.192 76	0.220 94
81	0.067 01	0.038 16	0.043 61	0.024 07	0.043 64	0.240 21	0.072 73	0.107 93	0.211 12	0.151 46
121	0.049 14	0.043 56	0.060 04	0.026 28	0.042 48	0.238 56	0.063 84	0.122 61	0.224 58	0.128 88
161	0.041 69	0.044 94	0.077 27	0.025 04	0.044 07	0.215 04	0.045 74	0.099 24	0.241 27	0.165 64
201	0.040 77	0.047 74	0.082 19	0.026 44	0.041 99	0.205 46	0.033 57	0.077 33	0.265 26	0.179 19
241	0.039 48	0.048 63	0.082 41	0.026 78	0.040 26	0.203 85	0.030 98	0.073 31	0.272 99	0.181 27
281	0.039 19	0.048 76	0.082 39	0.026 82	0.039 94	0.203 63	0.030 60	0.072 74	0.274 33	0.181 56
282	0.039 18	0.048 76	0.082 39	0.026 82	0.039 93	0.203 63	0.030 59	0.072 73	0.274 34	0.181 56
283	0.039 18	0.048 76	0.082 39	0.026 82	0.039 93	0.203 63	0.030 59	0.072 73	0.274 35	0.181 57
284	0.039 18	0.048 77	0.082 39	0.026 82	0.039 93	0.203 63	0.030 59	0.072 73	0.274 36	0.181 57
285	0.039 18	0.048 77	0.082 39	0.026 82	0.039 93	0.203 63	0.030 59	0.072 72	0.274 36	0.181 57

表 1 反映了参数 λ_j 的变化情况，而且收敛是很快的.

为了对比，在下面列出使用别的方法所得到预测结果：

(ⅰ) 阈值自回归模型(threshold autoregressive models，简记为 TAR)的预测结果：

$$\mathrm{Var_f}\ (\mathrm{train}) = 148.9,$$
$$MSE_{1921\sim1955} = 148.9,$$
$$MSE_{1956\sim1979} = 429.8.$$

(ⅱ) Weigend 等人提出的神经网络模型的预测结果：

$$\mathrm{Var_f}\ (\mathrm{train}) = 125.9,$$
$$MSE_{1921\sim1955} = 132.0,$$
$$MSE_{1956\sim1979} = 537.25.$$

(ⅲ) He 等人[4]利用径向基函数逐次逼近模型的预测结果：

$$\mathrm{Var_f}\ (\mathrm{train}) = 142.3,$$
$$MSE_{1921\sim1955} = 140.8.$$

(ⅳ) 径向基函数严格插值逼近模型的预测结果：

$$\mathrm{Var_f}\ (\mathrm{train}) = 10^{-3},$$
$$MSE_{1921\sim1955} = 10^{5}.$$

(ⅴ) 王明进 [1]利用 Kohonen 自组织神经网络的基于径向基函数模型的预测结果(由于作了多种情况下的预测，我们只采用其最好的结果)：

$m = 3$, $N_{\mathrm{c}} = 16$ 时，

1) 王明进. 非线性预测与混沌序列. 北京大学博士学位论文. 1997

$$\mathrm{Var_f}\ (\mathrm{train}) = 147.0,$$
$$MSE_{1921\sim1955} = 136.7;$$

$m = 6$, $N_c = 12$ 时,

$$\mathrm{Var_f}\ (\mathrm{train}) = 129.1,$$
$$MSE_{1921\sim1955} = 292.9.$$

从上面列出的结果来看，我们的模型的预测结果要好于所引文献的预测结果，可以看到我们所建立的模型实际上是将不同的模型融合起来，将它们的优势互补，较好地克服了单个模型的不足.

3 结论

本文主要研究了混沌时间序列的融合模型的建立问题，并针对一些混沌时间序列在数据量少的情况下进行预测，完成了如下的工作:

研究了混沌时间序列的基于径向基函数的融合预测模型，提出了应用属性聚类网络来选择径向基函数的中心，并且针对模型的预测偏差应用融合预测的方法来提高预测精度，经过对比，指出了这种模型与以往的模型相比增强了预测效果.

虽然我们的模型预测结果比以往的模型要好，但是它还可以进一步改进，尤其是如何利用多种模型更好地进行融合预测，将各个模型的预测效果发挥得更好，仍然值得进一步研究.

致谢 本工作为国家自然科学基金资助项目(批准号: 69872003).

参 考 文 献

1 Takens F. Detecting strange attractors in turbulence. In: Rand DA, Young L S, eds. Dynamical Systems and Turbulence. Lecture Notes in Math, Vol 898. Berlin: Springer, 1981. 366 ~ 381

2 Casdagli M. Nonlinear prediction chaotic time series. Physica D, 1989, 35: 335 ~ 356

3 程乾生. 属性模式识别及其应用. 见: 中国工业和应用数学学会第四次大会文集. 上海: 复旦大学出版社, 1996. 27 ~ 32

4 He X D, Lapedes A. Nonlinear modeling and prediction by successive approximation using radial basis functions. Physica D, 1993, 70: 289 ~ 302

(1999-12-09 收稿, 2000-03-15 收修改稿)