

基于神经网络的特征结构提取*

虞水俊** 孔铁生 梁甸农

(国防科技大学电子技术系 长沙 410073)

摘 要 本文在详细讨论特征结构提取问题的能量函数表示的基础上, 通过对该能量函数表示的适当变换和整理, 高阶关联非线性连续时间神经网络被引入到特征结构的提取问题中。这种方法直观、明了, 它将所要提取的特征结构与网络稳定时的输出直接对应起来, 可对特征结构进行自适应跟踪、估计, 这为特征结构的提取提出了一个新的方法。

关键词 特征结构 能量函数 神经网络

分类号 TP391.4

特征结构的提取在数字信号处理中是一个极其重要的研究课题, 人们对它进行了广泛而深入的研究, 并提出了很多算法。粗略统计, 特征结构提取的算法不下几十种, 其中主要的有: SVD 分解、QR 分解的方法、RRQR 算法^[2]、URV 算法^[1]、Gauss Newton 算法^[3]、S. Y. Kung 的 AREX 算法、B. Yang 的 PAST 算法、E. Oja 的 SIM 算法以及 B. Champagne 的基于一阶扰动准则的自适应特征分解算法^[4]。在这些算法中 SVD 分解、QR 分解、URV、SIM 等可用阵列来实现, 而 APEX 算法则是用神经网络来对特征结构进行自适应跟踪、估计。而其它算法绝大部分仅从理论上对子空间跟踪进行了探讨, 较少地涉及到算法的结构实现。

本文主要从算法实现的角度出发, 在详细讨论特征结构提取问题的能量函数表示的前提下, 通过对能量函数表示的适当变换和整理处理, 高阶关联非线性连续时间神经网络神经网络被引入到特征结构的提取问题中。

1 特征结构提取问题的能量函数表示

设 R 为 $-L \times L$ 的数据协方差矩阵, 其特征值按降序排列为: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L$, 与之对应的特征向量为 $q_1, q_2, \dots, q_L$, W 为 $-L \times 1$ 的列向量。由矩阵理论可知, R 的最小特征向量 q_L 为如下的约束最小问题的解 W^* :

$$\begin{aligned} \min \quad & W^T R W \\ \text{s. t. } & W^T W = 1 \end{aligned} \quad (1)$$

* 1994年3月21日收稿

** 博士生

最优优化理论告诉我们，(1)式的约束优化问题可以表示为如下的无约束最优问题：

$$\min_w [W^T R W + \mu(W^T W - 1)^2] \cdot \mu \lambda_L / 2 \quad (2)$$

且当 $\mu \rightarrow +\infty$ 时，由(2)式求得的最优解 W^* 即为 q_L 。定义能量函数

$$J(W, \mu) = W^T R W + \mu(W^T W - 1)^2 - \mu \lambda_L / 2 \quad (3)$$

关于该式有如下结论成立：使(3)式最小求得的最优解 W^* 为 $[1 - \lambda_L / (2\mu)]^{1/2} q_L$ (见附录)。显然，当 $\mu \rightarrow +\infty$ 时， W^* 即为 R 的最小特征向量。一般地，当 $\mu \geq 5\lambda$ 时， W^* 就近似地为 R 的最小特征向量。设求得了 q_L ，若要求 q_{L-1} ，可按如下思想来构造求取 q_{L-1} 的协方差矩阵 R_2 ：使 R_2 的最小特征向量与 R 的第二个最小特征向量相同。这样用 R_2 代替(3)中的 R 并使(3)最小，当 $\mu \rightarrow +\infty$ 时，求得的 W^* 即为 q_{L-1} 。设求得了 W_k^* ， $k=1, 2, \dots, j-1$ ，现要求 R 的第 j 个最小特征向量，可按如下方法来构造 R_j ：

$$R_1 = R$$

$$R_j = R + \beta \sum_{k=1}^{j-1} W_k^* W_k^{*T} \quad j = 2, 3, \dots, L \quad (4)$$

这里的 W_k 为

$$\min_w [W^T R W + \mu(W^T W - 1)^2] - \mu \lambda_{L-k+1} / 2 \quad (5)$$

的解。当 β 选择合适时，可以证得以 R_j 代替(3)中的 R 求得的 W^* 为 $[1 - \lambda_{L-j+1} / (2\mu)]^{1/2} q_{L-j+1}$ ，即为 R 的第 j 个未归一化的最小特征向量[附录]。因此，若要计算 R 的小特征向量，可构造出如下的能量函数：

$$J(W_j, \mu) = W_j^T R_j W_j + \mu(W_j^T W_j - 1)^2 \quad j = 1, 2, \dots, L \quad (6)$$

$$R_1 = R$$

$$R_j = R_{j-1} + \beta \sum_{k=1}^{j-1} W_k^* W_k^{*T} \quad (7)$$

并且选取 $\mu \gg \text{trace}(R)/2, \beta \gg \text{trace}(R)$ (见附录)。类似地，若要计算 R 的大特征向量，可构造出如下的能量函数：

$$J(W_j, \mu) = W_j^T R_j^{-1} W_j + \mu(W_j^T W_j - 1)^2 \quad j = 1, 2, \dots, L \quad (8)$$

$$R_1^{-1} = R^{-1}$$

$$R_j^{-1} = R_{j-1}^{-1} + \beta \sum_{k=1}^{j-1} W_k^* W_k^{*T} \quad (9)$$

且选取 $\mu \gg \text{trace}(R^{-1})/2, \beta \gg \text{trace}(R^{-1})$ 。 (10)

使(6)、(8)最小，求最优解向量可用最优化中的方法，如 Gauss—Newton 方法，下面则讨论它们的神经网络求解方法。

2 特征结构提取的神经网络实现

如果可用神经网络来求解(2)式的极小值问题，就可以用神经网络来对 R 的特征结构进行自适应跟踪、估计。首先讨论当 $j=1$ 的情况，然后再将之推广。为讨论简单起见，设矩阵 R 、特征向量等均为实值。设网络的二阶连续强度为 T_{ij} ，三阶连续强度为 T_{ijk} ，四阶连续强度为 T_{ijkl} ，网络的输入矢量考虑一四阶神经网络为 $U = [u_1, u_2, \dots, u_L]^T$ ，输出矢量为 $V = [V_1, V_2, \dots, V_L]^T$ ， $W_1 = [w_{11}, w_{12}, \dots, w_{1L}]^T$ ， $R = (r_{ij})$ ，将式(6)开展($j=1$)：

$$J_1(W_1, \mu) = W_1^T h(R) W_1 + \mu(W_1^T W_1 - 1)^2$$

$$= \sum_i \sum_j r_{ij} w_i w_j - 2\mu \sum_i w_i^2 + \mu \sum_i w_i^2 \sum_j w_j^2 + \mu \quad (11)$$

取

$$T_{ij} = \begin{cases} 2(r_{ij} - 2\mu) & i = j \\ 2r_{ij} & i \neq j \end{cases}$$

$$T_{i,k} = 0 \quad \text{无论 } i, j, k \text{ 为何值}$$

$$T_{i,k1} = \begin{cases} 4\mu & i = j \text{ 且 } k = l \\ 0 & \text{其它} \end{cases} \quad (14)$$

这样 $J(w_1, \mu)$ 可表示为 $(J(W_1, \mu))$ 用 J 代替：

$$J = 0.5 \sum_i \sum_j T_{ij} w_i w_j + 0.25 \sum_i \sum_j \sum_k \sum_l T_{ijkl} w_i w_j w_k w_l + \mu \quad (15)$$

若设 $V=W_1$ ，将之代入式(15)，且对 J 关于 t 求导，有：

$$dJ/dt = \sum_i [dJ/dV_i] \cdot (dV_i/dt)$$

$$= \sum_i (\sum_j T_{ij} v_j + \sum_j \sum_k \sum_l T_{ijkl} v_j v_k v_l) \quad (16)$$

若取

$$du_i/dt = -dJ/dv_i$$

$$= -(\sum_j T_{ij} v_j + \sum_j \sum_k \sum_l T_{ijkl} v_j v_k v_l) \quad (17)$$

$$v_i = \tanh(u_i/c) \quad (18)$$

则有

$$dJ/dt = \sum_i \frac{dJ/dv_i du_i}{dv_i du_i dt}$$

$$= - \sum_i (\tanh(u_i/c))' (du_i/dt)^2 \quad (19)$$

由于 $\tanh(x)$ 为一单调增函数，因此有：

$$dJ/dt \leq 0 \quad (20)$$

这样我们便证明了由式(15)、(17)所描述的系统是稳定的。相应地，用 $R_i (i=2, \dots, L)$ 代替 R 亦有同样的结论成立，另外，注意到式(15)、(18)描述的正是高阶关联的非线性连续时间神经网络的动力学方程及能量函数。因此高阶关联的非线性连续时间神经网络可用来求解(3)式的无约束最小问题，即用来求 R 的最小特征向量，而其各个分量与网络稳定时的输出相对应。并且若取 μ 足够大，当网络稳定时，系统能量等于 λ_{min} 。求出了最小特征向量，就可求第二个最小特征向量，求出了第一、二个最小特征向量，就可求第三个最小特征向量，依次类推，亦可求出其它特征向量。当 R, W 为复值时，设：

$$R' = \begin{bmatrix} \text{Re}(R) & -\text{Im}(R) \\ \text{Im}(R) & \text{Re}(R) \end{bmatrix} \quad (21)$$

$$W' = ((\text{Re}(w))^T, (\text{Im}(w))^T)^T$$

$$= [\tau w'_1, \tau w'_2, \dots, \tau w'_{2L}]^T \quad (22)$$

显然， R' 是一实对称正定矩阵， R 的特征值与 R' 的特征值按大到小，小到小的以一对二的关系对应且相等(R' 的特征值均为二重值)，并且特征向量也按相同的规则对应起来。这样，提取 R 的特征结构就转变为提取 R' 的特征结构， R' 的特征结构提取可用前面讨论的方法。设已提取了 R' 的第 j 个最小特征向量，则可按：

$$W_{[j/2]} = [\tau w'_{j,1} + jw'_{j,L+1}, \dots, \tau w'_{j,1} + jw'_{j,L+1}, \dots, \tau w'_{j,L-j} + jw'_{j,2L}]^T \quad (23)$$

来形成 R 的第 $[j/2]$ 个最小特征向量, $[.]$ 表示取整, 这样便可形成 R 的 2^N 个特征向量矩阵。

用高阶关联的非线性连续时间神经网络实现特征结构提取时涉及到网络的初始化问题, 由上分析可知网络的初始输入可取为一组随机量。非线性函数取为 $\tanh(x)$ 主要是由于它是单调递增的, 且其值域为 $(-1, 1)$ 。

3 仿 真

上面我们分析了用高阶关联的非线性连续时间神经网络提取特征结构的理论基础, 为了进一步验证该方法的正确性, 特进行了计算机仿真, 我们选取 R 为一实对称正定阵, 如:

$$R = \begin{bmatrix} 0.113562 & 0.191099 & 0.118040 & 0.194445 \\ 0.191099 & 0.500384 & 0.689985 & 0.501589 \\ 0.118040 & 0.689985 & 1.665722 & 0.913106 \\ 0.194445 & 0.501589 & 0.913106 & 0.819618 \end{bmatrix}$$

$c = 600, \mu = 200.$

我们所给算法的计算机仿真结果如下: 特征值为 2.614261, 0.330443, 0.144516, 0.001452; 而 MATLAB 的计算结果为 2.6227, 0.3309, 0.1445, 0.0012。

另外, 当 R 为 Hermite 矩阵时亦进行了计算机仿真, 取:

$$R = \begin{bmatrix} 2 & 1-j \\ 1+j & 3 \end{bmatrix} \quad \text{则 } R' = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 3 & -1 & 0 \\ 0 & -1 & 2 & 1 \\ 1 & 0 & 1 & 3 \end{bmatrix}$$

这样可算得 R 的特征值为 0.999 和 3.987, 与真值 1 和 4 非常接近。同时可形成 R 的 $2^2 = 4$ 个特征向量矩阵, 它们分别为:

$$\begin{bmatrix} -0.035 + j0.817 & 0.528 - j0.218 \\ 0.423 - j0.388 & 0.753 + j0.314 \end{bmatrix}, \begin{bmatrix} -0.035 + j0.817 & 0.218 + j0.529 \\ 0.423 - j0.388 & -0.312 + j0.753 \end{bmatrix}$$

$$\begin{bmatrix} 0.817 - j0.035 & 0.528 - j0.218 \\ -0.388 - j0.422 & 0.753 + j0.314 \end{bmatrix}, \begin{bmatrix} 0.817 + 0.035 & 0.218 + j0.529 \\ -0.388 - j0.422 & -0.312 + j0.753 \end{bmatrix}$$

通过观察不难发现, 这四个矩阵均为其中之一经正交变换所得, 它们均可作为 R 的特征向量矩阵。

结果表明该方法是正确的, 切实可行的。但是也暴露出一个问题: 网络的稳定性与参数 c 的选取密切相关。如果选用 (18) 中的 $c=1$, 始时 u_1 是一随机量。若 μ 取得很大, 则 u_1 的修正量较大, 这导致 $v_1 = \tanh t(u_1) \approx \pm 1$, 反过来 v_1 又影响下一时刻的 u_1 。如此反复相互作用, 使 v_1 在 ± 1 两数上变化, 导致网络不能正常运行。为保证网络能顺利地收敛到我们所需要的稳定平衡点, 应恰当地选择 c 。实验表明: c 的选取与 μ 紧密相关, 且保证网络能稳定运行的最小 c 与 μ 的关系为: $c \geq 2\mu$ 且大于 1, 一般地取 c 为三倍的 μ 较为合适。

4 结束语

文章分析了特征结构的提取高阶关联神经网络实现,第一次将高阶关联非线性神经网络引入到特征结构的提取问题中,该方法简单、明了,为特征结构的提取提供了一种新的途径。

参 考 文 献

- 1 G W Stewart. An Updating Algorithm for Subspace Tracking. IEEE T-SP Vol. 40. 1992,1535~1541
- 2 Chan. I. F. Alternative to the SVD: Rank Revealing QR-factorization SPIE Vol. 696 1986.
- 3 G Mathew, V U Reddy and S. Dasgupta Gauss-Newton Adaptive Subspace estimation. ICASSP, IV-205, 1992
- 4 B Champagne. A New Adaptive Eigendecomposition Algorithm Based on a first Order perturbation Criterion. INRS Telecomm. (Canada)
- 5 J F Yang, M Kaveh Adaptive Eigensubspace Algorithms for Direction or Frequency Estimation and Tracking. ICASSP'1988
- 6 J F Yang etc. A Robust and High resolution Adaptive Eigensubspace Algorithm for Nonstationary Environments. ICASSP'1990

Neural-network Based for Eigenstructure Extraction

Yu Shuijun Kong Tieshen Liang Diannong

(Department of Electronic Technology)

Abstract

In the paper, The expression of the cost function of the eigenstructure extraction problem is discussed in detail. Through a proper transformation of the expression of the cost function, high-order neural network is introduced into the eigenstructure extraction problem. According to the method, the eigenstructure to be extracted is connected with the output of the network directly when the network is stable. The method is simple and direct, and can be used to track the eigenstructure adaptively. It opens a new area for the extraction of the eigenstructure.

Key words eigenstructure cost function neural network

附录: β , μ 的范围选定

对式(3), 令: $W = \sum a_i q_i$, 且 W 的模的平方为 b^2 , 则

$$\begin{aligned} J(W, \mu) &= J(a_1, \dots, a_L, \mu) \\ &= \sum a_i^2 \lambda_i + \mu(b^2 - 1)^2 \end{aligned}$$

$$= b^2 \sum \eta_i^2 \lambda_i + \mu(b^2 - 1)^2 \quad (A-1)$$

其中 $\eta_i^2 = a_i^2/b^2$, λ_i 及 q_i 的意义同前, 对式(A-1), 由于 $\mu(b^2-1)$ 与 a_i 无关, 且 λ_i 按降序排列, 因此当且仅当 $a_1 = a_2 = \dots = a_{L-1} = 0$ 时 J 最小, 此时 $a_L = b$, 将之代入(A-1), 有

$$J = b^2 \lambda_{\min} + \mu(b^2 - 1)^2 \quad (A-2)$$

J 是 b 的函数, 对 J 关于 b 求导有:

$$dJ/db = 2b\lambda_{\min} + 4\mu b(1 - b^2) \quad (A-3)$$

令(A-3)为零, 且 $b \neq 0$ 有:

$$b^2 = 1 - \lambda_{\min}/2\mu \quad (A-4)$$

由于 $b^2 > 0$, 因此

$$1 - \lambda_{\min}/2\mu > 0 \quad (A-5)$$

即

$$\mu > \lambda_{\min}/2 \quad (A-6)$$

这样便选定了提取最小特征向量时的 μ 值的范围。当 $\mu > \lambda_{\min}/2$ 时, $b = \pm [1 - \lambda_{\min}/(2\mu)]^{1/2}$, $W^* = bq_L$. 相应地当 β 选择合适时, 用 R_j 代替(4)中的 R , 可推得:

$$\mu > \lambda_{L-j+1}/2 \quad (A-7)$$

考虑 $j=L$ 时, μ 的范围适用于所有情形。此时有: $\mu > \lambda_{\max}/2$

选定了 μ 的范围, 我们现在来推导 β 的范围选定。对(9)式, 设 $W_k^* = b_{L-k+1}q_{L-k+1}$, 则

$$\begin{aligned} R_j &= \sum_{i=1}^L \lambda_i q_i q_i^T + \beta \sum_{i=L-j+1}^L b_i^2 q_i q_i^T \\ &= \sum_{i=1}^{L-j+1} \lambda_i q_i q_i^T + \sum_{i=L-j+2}^L (\lambda_i + \beta b_i^2) q_i q_i^T \end{aligned} \quad (A-8)$$

要使 R_j 的最小特征向量为 q_{L-j+1} , 则必须有:

$$\lambda_i + \beta b_i^2 > \lambda_{L-j+1} \quad i = L-j+2, \dots, L \quad (A-9)$$

考虑 $j=L$ 时, β 的范围适用于所有情形, 此时必须有

$$\lambda_i + \beta b_i^2 > \lambda_{L-j+1} \quad i = 2, \dots, L \quad (A-10)$$

将 b_i^2 的值代入经过一序列的运算有

$$\beta > (\lambda_{\max} - \lambda_{\min}) / (1 - \lambda_{\min}/2\mu) \quad (A-11)$$

又由于 $\lambda_{\max}, \lambda_{\min}$ 无法确定, 且 $\text{tr}(R) > \lambda_{\max}, \text{tr}(R)/L \geq \lambda_{\min}$, 因此可这样来选取 β, μ :

$$\mu > \text{tr}(R)/2 \quad (A-13)$$

$$\beta \geq \text{tr}(R) \quad (A-14)$$

同理, 若要计算 R 大特征向量, 则 μ, β 的范围选取为

$$\mu > \text{tr}(R^{-1})/2 \quad (A-15)$$

$$\beta \geq \text{tr}(R^{-1}) \quad (A-16)$$