

一种新的特征结构提取方法及其神经网络实现¹

虞水俊 梁旬农

(国防科技大学电子技术系 长沙 410073)

摘 要 本文首先建立了特征结构问题的代价函数表示,通过对代价函数求极小可以求得原始数据协方差矩阵的最大特征向量。为了求得其他特征向量,特构造了一个协方差矩阵序列。为实现对代价函数求极小,可把高阶神经网络引入特征结构提取中。这种方法比较直观,它将网络稳定时的输出与所求协方差矩阵的主特征向量的各个分量相对应。理论分析和计算机仿真均验证了这种方法的正确性。

关键词 特征结构, 代价函数, 神经网络

中图分类号 TN911.7, TN-052

1 引言

基于数据协方差矩阵的特征结构之类的算法在信号处理领域具有极其广泛的应用,而特征结构的提取是一项运算量大、计算复杂的计算过程,这限制了特征结构类算法的应用和发展。为减小特征结构提取算法的计算量,降低运算的复杂度,科学工作者做了大量的工作,提出了很多有用的方法。所有这些方法,根据使用技术的不同,可分为分解法和能量函数法(代价函数法)两大类。能量函数法是近十几年发展起来的一类新的方法,它的应用前景是十分广阔的。现有的特征结构提取的最优化方法^[1,2]、神经网络方法^[3,4]均属此范畴。但目前,最优化方法主要集中于纯算法的研究;而特征结构提取的神经网络方法未能就特征结构提取问题建立一个合适的代价函数以对问题进行描述。这给我们分析、理解和使用算法带来了困难。

本文中,我们针对数据协方差矩阵 R 的大特征向量的提取问题,建立了该问题的代价函数表示。代价函数的极小值点对应的解向量正好为 R 的最大特征向量。为求取其它特征向量,特构造一个协方差矩阵序列:使系列中的第 j 个矩阵的最大特征向量与原协方差矩阵的第 j 个最大特征向量一致。建立了所求问题的代价函数表示,为求取代价函数的极小值点,特将神经网络的能量函数与所建立的代价函数进行对比分析,将高阶关联非线性神经网络引入特征结构提取问题中。理论分析和计算机仿真均验证了方法是正确可行的。

2 特征结构提取的代价函数表示

设 R 为一 $N \times N$ 的数据协方差矩阵,其特征值按降序排列为: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$, 与之对应的特征向量为 $q_1, q_2, \cdots, q_N$, W 为一 N 维的列向量。

¹ 1994-04-14 收到, 1995-08-29 定稿
国家 863 高技术计划资助项目

文献 [5,6] 中, 关于 R 的特征向量的提取建立了如下的代价函数:

(1) 若计算从 R 的最小特征向量的提取开始,

$$\left. \begin{aligned} J(W_j, \mu) &= W_j^T R_j W_j + \mu(W_j^T W_j - 1)^2, \\ R_j &= R + \beta \sum_{k=1}^{j-1} W_k^* W_k^{*T}, \quad R_1 = R, \quad j = 1, 2, \dots, L. \end{aligned} \right\} \quad (1)$$

(2) 若计算从 R 的最大特征向量的提取开始,

$$\left. \begin{aligned} J(W_j, \mu) &= W_j^T R_j^{-1} W_j + \mu(W_j^T W_j - 1)^2, \\ R_j^{-1} &= R^{-1} + \beta \sum_{k=1}^{j-1} W_k^* W_k^{*T}, \quad R_1^{-1} = R^{-1}, \quad j = 1, 2, \dots, L. \end{aligned} \right\} \quad (2)$$

(1) 中的 $\mu > \text{trace}(R)/2$, $\beta \geq \text{trace}(R)$, L 为所提小特征向量的个数。(2) 中的 $\mu > \text{trace}(R^{-1})/2$, $\beta \geq \text{trace}(R^{-1})$, L 为所提大特征向量的个数。 W_j 为一 N 维的列向量, W_k^* 为 $J(W_k, \mu)$ 取极小时的最优解向量。

由 (1) 式和 (2) 式可知, 如果要求 R 的最大特征向量则不可避免地会遇到两个问题:

①若用 (1) 式的代价函数就会存在冗余计算; ②若用 (2) 式的代价函数, 虽无冗余计算, 但需要求矩阵的逆。为克服这些不足, 在计算 R 的最大特征向量时, 不妨考察下式:

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

关于该式有如下定理成立:

定理 1 数据协方差矩阵 R 的最大特征向量为 (3) 式的全局极小值点, 而 R 的其它特征向量为 (3) 的鞍点。

证明 设 $W = \sum_{i=1}^N a_i q_i$, $\sum_{i=1}^N a_i^2 = b^2 > 0$, 将它代入式 (3) 有

$$J(W, \mu) = -2 \sum_{i=1}^N a_i^2 \lambda_i + b^2 \sum_{j=1}^N a_i^2 \lambda_i + b^4 \mu - 2b^2 \mu + \mu. \quad (4)$$

这里 W 为一 N 维的列向量。在上式中, 由于 R 为对称正定矩阵, λ_i 按降序排列, 因此有: ①当 $b^2 - 1 > 0$ 时, 当且仅当 $a_1^2 = \dots = a_{N-1}^2 = 0$, $a_N^2 = b^2$, (3) 式才取得极小值; ②当 $b^2 - 2 < 0$ 时, 当且仅当 $a_1^2 = b^2$, $a_2^2 = \dots = a_{N-1}^2 = 0$, (3) 式才取得极小值。由此可知, 在情形①下, 对 (4) 取极小所得的最优解向量与 q_N 为同一向量; 而在情形②时, 对 (4) 式取极小所得的最优解向量与 q_1 为同一向量。因此, 在判断对 (4) 取极小时, 其最优解向量是否与 R 的最大特征向量为同一向量之前, 先判断出 (3) 式取极小值时 b^2 的大小是必要的。

首先假设 $b^2 > 2$, 此时 (3) 式可写为

$$J(W, \mu) = (b^2 - 2)b^2 \lambda_N + b^4 \mu - 2b^2 \mu + \mu. \quad (5)$$

显然 (5) 式为一关于 b 的函数关系式, 对 (5) 式关于 b 求导, 并令之为零有

$$-4b\lambda_N + 4b^3\lambda_N + 4b^3\mu - 4b\mu = 0. \quad (6)$$

由于 $b \neq 0$, $\lambda_1 + \mu > 0$, 因此有

$$b^2 = 1. \quad (7)$$

这与假设相矛盾, 因此, 当 (3) 取得极小值时其最优解向量对应于 R 的最大特征向量。按照前面相同的讨论方法, 可推得当 (3) 式的最优解向量对应于 R 的最大特征向量时, 同样有 $b^2 = 1$ 。关于 R 的最大特征向量是否是 (3) 式的全局极小值, 可从 $J(W, \mu)$ 的 Hessian 的矩阵的正 (负) 定性判断得出。为此, 对 (3) 式关于 W 求得

$$D(W, \mu) = -4RW + 2W^T RWW + 2W^T WRW + 4\mu(W^T W - 1)W. \quad (8)$$

显然 R 的所有特征向量均为 $D(W, \mu) = 0$ 的解向量。对 $D(W, \mu)$ 关于 W 求导, 得 $J(W, \mu)$ 的 Hessian 矩阵:

$$H(W, \mu) = -4R + 8RWW^T + 2W^T WR + 2W^T RWI + 8\mu WW^T + 4\mu(W^T W - 1)I, \quad (9)$$

$$H(W)|_{W=q_i} = -2R + 2\lambda_i I + 8(\lambda_i + \mu)q_i q_i^T = -2 \sum_{i=1}^N \lambda_j q_j q_j^T + 2\lambda_i I + 8(\lambda_i + \mu)q_i q_i^T, \quad i = 1, 2, \dots, N. \quad (10)$$

这里 I 为单位阵。由上式可知 $H(W)|_{W=q_i}$ 的、与特征向量 q_j 对应的特征值为 λ'_j , 而

$$\lambda'_j = \begin{cases} 2(\lambda_i - \lambda_j), & i \neq j; \\ 8\lambda_i + 8\mu, & i = j. \end{cases} \quad (11)$$

因此, 当 $\mu \geq 0$ 时,

$$H(W)|_{W=q_1} \text{ 的 } N \text{ 个特征值为 } \lambda'_j = \begin{cases} 2(\lambda_1 - \lambda_j) > 0, & j > 1; \\ 8\lambda_1 + 8\mu > 0, & j = 1. \end{cases} \quad (12)$$

$$H(W)|_{W=q_i(i \neq 1)} \text{ 的 } N \text{ 个特征值为 } \lambda'_j = \begin{cases} 2(\lambda_i - \lambda_j) < 0, & j < i; \\ 8\lambda_i + 8\mu > 0, & j = i; \\ 2(\lambda_i - \lambda_j) > 0, & j > i. \end{cases} \quad (13)$$

由上知, $H(W)|_{W=q_1}$ 为正定矩阵, $H(W)|_{W=q_i(i \neq 1)}$ 为不定矩阵。因此, R 的最大特征向量为 (3) 式的全局唯一极小值点, 而 R 的其它特征向量为 (3) 式的鞍点。证毕

由定理可知, 我们可通过对 (3) 求极小来求得 R 的最大特征向量 q_1 。设求得了 q_1 , 若要求 q_2 , 可按如下思想来构造求取 q_2 的协方差矩阵 R_2 : 使 R_2 的最大特征向量与 R 的第二个最大特征向量相同。这样用 R_2 代替 (3) 中的 R 并对 (3) 求极小, 求得的最优

解向量 $\mathbf{W}_2^*$ 即为 q_2 。设求得 $\mathbf{W}_k^*, k=1, 2, \dots, j-1$, 现要求 R 的第 j 个最大特征向量, 可按如下方法来构造 R_j

$$R_j = R - \beta \sum_{k=1}^{j-1} \mathbf{W}_k^* \mathbf{W}_k^{*T}, \quad R_1 = R, \quad j = 2, 3, \dots, L. \quad (14)$$

其中 L 为所提大特征向量的个数, $\mathbf{W}_k^*$ 为 $\min -\mathbf{W}^T R_k \mathbf{W} + \mathbf{W}^T R_k \mathbf{W} \mathbf{W}^T \mathbf{W} + \mu(\mathbf{W}^T \mathbf{W} - 1)^2$ 的解。当 $\beta > \lambda_1$ 时, 易证 R_j 的所有特征向量与 R 的相同, R_j 的最大特征值为 λ_j , 其对应的特征向量为 q_j 。实际中, R 的特征值不可能预先求得, 但 $\lambda_1 < \text{trace}(R)$, 因此取 $\beta \geq \text{trace}(R)$ 。需要指出的是, 按 (14) 式所形成的 R_j 是一个不定矩阵, 它有 $j-1$ 个特征值小于零, 而有 $N-j+1$ 个特征值大于零, 如果以该 R_j 代替 (3) 中的 R , 利用 Hessian 矩阵的正定性同样可以证明, 当以 R_j 代替 (3) 中的 R 时, q_j 为此时的 (3) 式的全局唯一极小值点。因此, 若要计算 R 的第 j 个最大特征向量, 可构造出如下的能量函数:

$$J(\mathbf{W}, \mu) = -2\mathbf{W}^T R_j \mathbf{W} + \mathbf{W}^T R_j \mathbf{W} \mathbf{W}^T \mathbf{W} + \mu(\mathbf{W}^T \mathbf{W} - 1)^2. \quad (15)$$

至此, 大特征向量的提取问题已转变为求解无约束最小问题。要求无约束最小问题可以应用最优化方法, 而这里我们则讨论它的神经网络求解方法。

3 特征结构提取的神经网络实现

上一节讨论了特征结构提取问题的代价函数表示, 如果可用高阶关联非线性连续时间神经网络来实现对 (3) 式求极小, 就可以用高阶关联非线性连续时间神经网络来对 R 的特征结构进行估计。首先讨论提取最大特征向量的情形, 然后再将它推广。为简单起见, 设矩阵 R 、特征向量等均为实值。考虑一三阶神经网络, 设网络的输入向量为 U , 输出向量为 V , 且设 $V = W$, 通过将网络的能量函数与所建立的代价函数表示式 (3) 式进行对比分析, 则可得出网络的各阶连接强度为

$$\left. \begin{aligned} T_{ij} &= \begin{cases} 4(r_{ij} + \mu), & i = j; \\ 4r_{ij}, & i \neq j. \end{cases} \\ T_{ijk} &= 0, \quad \text{无论 } i, j, k \text{ 为何值.} \\ T_{ijkl} &= \begin{cases} -4(r_{ij} + \mu), & i = j \text{ 且 } k = l; \\ -4r_{ij}, & i \neq j \text{ 且 } k = l; \\ 0, & \text{其它.} \end{cases} \end{aligned} \right\} \quad (16)$$

若取网络的输入输出关系和动态方程为

$$\begin{aligned} V &= f(U) \\ \frac{dU}{dt} &= -\frac{dJ(\mathbf{W}, \mu)}{d\mathbf{W}} = -D(\mathbf{W}, \mu) \\ &= -\{-4R\mathbf{W} + 2\mathbf{W}^T R \mathbf{W} \mathbf{W} + 2\mathbf{W}^T \mathbf{W} R \mathbf{W} + 4\mu(\mathbf{W}^T \mathbf{W} - 1)\mathbf{W}\}. \end{aligned} \quad (17)$$

当 $f(\cdot)$ 为一单调增函数时, 可推得

$$\frac{dJ(W, \mu)}{dt} \leq 0. \quad (18)$$

上式说明, 按上述方法构造的神经网络是稳定的, 随着时间的推移, 网络向其稳定状态演变, 当网络达到稳定状态时, 系统能量取得极小值, 此时网络的输出正好对应于 R 的最大特征向量。相应地, 以 R_j 代替 R 亦有同样的结论成立, 这样便可求得 R 的大特征向量。上面均假设 R , W 为实值, 当它们为复值时, 设

$$R' = \begin{bmatrix} \text{Re}(R) & -\text{Im}(R) \\ \text{Im}(R) & \text{Re}(R) \end{bmatrix}, \quad (19)$$

$$\begin{aligned} W' &= [(\text{Re}(W))^T (\text{Im}(W))^T]^T \\ &= [w'_1, w'_2, \dots, w'_{2L}]^T. \end{aligned} \quad (20)$$

显然, R' 为一实对称正定矩阵, R 的特征值与 R' 的特征值按大到大, 小到小以一对二的关系相对应, 且特征向量也对应起来, 这样 R 的特征结构提取问题就转变为 R' 的特征结构提取问题。 R' 的特征结构提取可用前面讨论的方法, 设求得了 R' 的第 j 个最小特征向量, 则可按

$$W_{[(j+1)/2]} = [w'_{j,1} + jw'_{j,L+1}, \dots, w'_{j,1} + jw'_{j,L+i}, \dots, w'_{j,L} + jw'_{j,2L}]^T \quad (21)$$

来形成 R 的第 $[(j+1)/2]$ 个特征向量, $[\cdot]$ 表示取整。

4 仿真和结论

为验证方法的正确性, 对本文提出的方法进行了计算机仿真。取

$$R = \begin{bmatrix} 1627.9 & -729.4 & -398.3 \\ -729.4 & 1682.8 & -756.3 \\ -398.3 & -756.3 & 1707.7 \end{bmatrix}.$$

仿真实验时取非线性函数为双曲正切函数, $\mu = 0$ 。利用本文提出的方法可求得 R 的特征向量的估计为

$$q_{1,e} = \begin{bmatrix} 0.4126 \\ -0.7678 \\ 0.4902 \end{bmatrix}, \quad q_{2,e} = \begin{bmatrix} -0.7304 \\ 0.0426 \\ 0.6816 \end{bmatrix}, \quad q_{3,e} = \begin{bmatrix} -0.5442 \\ -0.6393 \\ -0.5432 \end{bmatrix}$$

同时, 亦绘制了不同时刻算法的性能曲线, 如图 1 所示。图中纵轴为 $|q_{j,e}^T(t)q_j|$, 其中 $q_{j,e}(t)$ 为 t 时刻 q_j 的估计。该图不仅反映了算法的学习过程, 同时亦反映了网络的动态演化过程。由图可知, 当网络参数选择合适时, 网络可以较高的速度演变到其稳定平衡点, 当网络达到稳定平衡状态时, 网络的输出正好对应于所要提取的特征向量。另外, 我们亦就 R 为 Hermite 矩阵的情形进行了仿真, 仿真结果验证了所提方法的正确性。

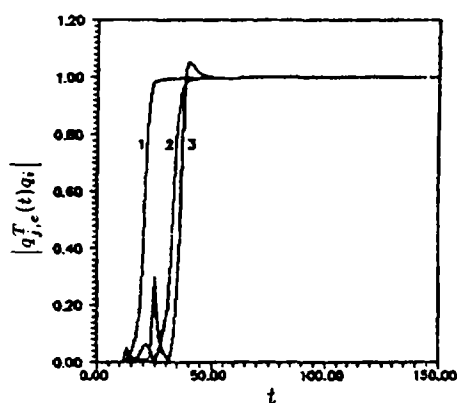


图 1 算法性能曲线
1: q_1 的提取, 2: q_2 的提取, 3: q_3 的提取

5 结 束 语

本文从一个新的角度对数据协方差矩阵的大特征向量的提取问题进行了探讨, 建立了求解该问题的代价函数表示。为求代价函数的极小, 引用了高阶神经网络。理论分析和计算机仿真均表明所提方法是可行的。

参 考 文 献

- [1] Yang Bin. Subspace Tracking Based on the Projection Approach and the Recursive Least Squares Method. Minnesota: Proc. IV of ICASSP'93, 1993, 145-148.
- [2] Mathew G, et al. Gauss-Newton Adaptive Subspace Estimation. California: Proc. IV of ICASSP'92, 1992, 205-208.
- [3] Luo Fa-Long, Li Yan-Da. Real-Time Neural Computation of the Noise Subspace for the MUSIC Algorithm. Minnesota: Proc. I of ICASSP'93, 1993, 485-487.
- [4] Kung San Yuan, et al. A Neural Network Learning Algorithm for Adaptive Principal Component Extraction(APEX). New Mexico: Proc. of ICASSP'90, 1990, 861-864.
- [5] 虞水俊, 等. 国防科技大学学报, 1995, 17(1): 16-21.
- [6] 虞水俊, 孔铁生, 梁甸农. 电子学报, 1995, 23(8): 109-111.

A NEW METHOD FOR EIGENSTRUCTURE EXTRACTION AND ITS NEURAL NETWORK IMPLEMENTATION

Yu Shuijun Liang Diannong

(National University of Defense Technology, Changsha 410073)

Abstract The cost function for eigenstructures extration is discussed in detail, one can obtain the largest eigenvector by minimizing the cost function. In order to obtain the other eigenvectors, a covariance matrix series is constructed. If one compares the cost function with the energy function of a neural network, the neural network can be introduced to extract the eigenvectors. Theoretical analysis and simulations show that the proposed method is reasonable and feasible.

Key words Eigenstructure, Cost function, Neural networks

虞水俊: 男, 1965 年生, 博士生, 信号与信息处理专业.

梁甸农: 男, 1936 年生, 教授, 博士生导师, 信号与信息处理专业.