

一种计算矩阵特征值特征向量的神经网络方法*

刘怡光¹⁺, 游志胜¹, 曹丽萍², 蒋欣荣¹

¹(四川大学 计算机学院 图像图形研究所, 四川 成都 610064)

²(四川大学 公共管理学院, 四川 成都 610064)

A Neural Network Algorithm for Computing Matrix Eigenvalues and Eigenvectors

LIU Yi-Guang¹⁺, YOU Zhi-Sheng¹, CAO Li-Ping², JIANG Xin-Rong¹

¹(Institute of Image and Graphics, School of Computer Science and Engineering, Sichuan University, Chengdu 610064, China)

²(School of Public Administration, Sichuan University, Chengdu 610064, China)

+ Corresponding author: Phn: +86-28-85412565, Fax: +86-28-85410604, E-mail: lygpapers@yahoo.com.cn

Received 2003-09-22; Accepted 2004-05-08

Liu YG, You ZS, Cao LP, Jiang XR. A neural network algorithm for computing matrix eigenvalues and eigenvectors. *Journal of Software*, 2005,16(6):1064–1072. DOI: 10.1360/jos161064

Abstract: While using continuous time neural network described by the E.Oja. learning rule (Oja-N) for computing real symmetrical matrix eigenvalues and eigenvectors, the initial vector must be on R^n unit hyper-sphere surface, otherwise, the network may produce limit-time overflow. In order to get over this defect, a new neural network (lyNN) algorithm is proposed. By using the analytic solution of the differential equation of lyNN, the following results are received: If initial vector belongs to a space corresponding to certain eigenvector, the lyNN equilibrium vector will converge in this space; If initial vector does not fall into the space corresponding to any eigenvector, the equilibrium vector will belong to the space spanned by eigenvectors corresponding to the maximum eigenvalue. The initial vector maximum space for the lyNN equilibrium vector will fall into space spanned by eigenvectors corresponding to any eigenvalue received. If the initial vector is perpendicular to a known eigenvector, so is the equilibrium vector. The equilibrium vector is on the hyper-sphere surface decided by the initial vector. By using the above results, a method for computing real symmetric matrix eigenvalues and eigenvectors using lyNN is proposed, the validity of this algorithm is exhibited by two examples, indicating that this algorithm does not bring about limit-time overflow. But for Oja-N, if the initial vector is outside the unit hyper-sphere and the matrix is negatively determinant, the neural network will consequentially produce limit-time overflow. Compared with other algorithms based on optimization, lyNN can be realized directly and its computing weight is lighter.

Key words: neural network; symmetric matrix; eigenvalue; eigenvector; limit-time overflow

* Supported by the National Natural Science Foundation of China under Grant No.69732010 (国家自然科学基金); the Innovation Fund for Small Technology-Based Firms of Ministry of Science and Technology of China under Grant No.03C26225100257 (国家科技部科技型中小企业技术创新基金)

作者简介: 刘怡光(1972—),男,四川内江人,博士,工程师,主要研究领域为模式识别,神经网络,遗传算法,小波分解;游志胜(1945—),男,教授,博士生导师,主要研究领域为数据融合,图像处理与模式识别,神经网络,智能交通系统;曹丽萍(1976—),女,硕士生,主要研究领域为信息检索技术,情报分析;蒋欣荣(1972—),男,讲师,主要研究领域为图像处理。

摘要: 当把 Oja 学习规则描述的连续型全反馈神经网络(Oja-N)用于求解矩阵特征值特征向量时,网络初始向量需位于单位超球面上,这给应用带来不便.由此,提出一种求解矩阵特征值特征向量的神经网络(lyNN)方法.在 lyNN 解析解基础上得到了以下结果:初始向量属于任意特征值对应特征向量张成的子空间,则网络平衡向量也将属于该空间;分析了 lyNN 收敛于矩阵最大特征值对应特征向量的初始向量取值条件;明确了 lyNN 收敛于矩阵不同特征值的特征子空间时,网络初始向量的最大取值空间;网络初始向量与已知特征向量垂直,则 lyNN 平衡解向量将垂直于该特征向量;证明了平衡解向量位于由非零初始向量确定的超球面上的结论.基于以上分析,设计了用 lyNN 求矩阵特征值特征向量的具体算法,实例演算验证了该算法的有效性.lyNN 不出现有限溢,而基于 Oja-N 的方法在矩阵负定、初始向量位于单位超球面外时必出现有限溢,算法失效.与基于优化的方法相比,lyNN 实现容易,计算量较小.

关键词: 神经网络;对称矩阵;特征值;特征向量;有限溢

中图分类号: TP183 **文献标识码:** A

1982年,E.Oja对主成分分析(principal component analyzer,简称PCA)提出如下神经元模型^[1]:

$$\frac{d}{dt} \mathbf{X}(t) = \mathbf{A}\mathbf{X}(t) - \mathbf{X}^T(t)\mathbf{A}\mathbf{X}(t)\mathbf{X}(t), \mathbf{A}^T = \mathbf{A}, \mathbf{A} \in \mathbb{R}^{n \times n}, \mathbf{X}(t) \in \mathbb{R}^n \quad (1)$$

将 \mathbf{A} 视为神经网络的连接强度, $\mathbf{X}(t)$ 为神经网络的状态,则式(1)描述了一类连续型全反馈神经网络.E.Oja等人证明式(1)从许多初始向量出发的解都收敛于最大特征值对应的特征向量^[2].罗发龙等人认为,在将式(1)推广用来求解矩阵多个特征矢量时,式(1)的输出只收敛于特征向量张成的子空间而非真正的特征矢量,并提出了一种求解对称正定矩阵特征向量的神经网络^[3,4].实际上,罗发龙等人的认识是有限的,章毅等人用式(1)成功地解决了求解一般实对称矩阵特征向量及其特征值的问题,并给出了严格的数学分析^[5].许多研究人员把求解矩阵特征值特征向量问题转化为优化问题,并用神经网络求解,如 A.Cichocki 等人^[6]把最小特征值问题转化为约束优化问题,通过使用罚函数和 Lagrange 乘子法,用神经网络求解该优化问题,该法求解不易,网络具有最小点,使用非常有限;Y.Tang 等人^[7]设计的神经网络可得到对称半正定矩阵的最小特征值及对应特征向量,并通过构造一个矩阵序列,提出了求解矩阵所有特征向量及特征值的措施^[8].

基于优化思想构造的神经网络,一次收敛只能得到最大或最小特征值对应的特征向量,若要求出全部特征值及特征向量^[6,7],需采取其他措施^[8].近年来,许多研究人员用式(1)求解矩阵所有特征值特征向量^[3-5],文献[5]是这方面研究的成果之一.但直接用式(1)求解矩阵特征值特征向量时存在明显不足,正如文献[5]所述,用式(1)求解矩阵特征值特征向量时,要求系统初始向量 $\mathbf{X}(0)$ 需位于 \mathbb{R}^n 中单位超球面上,否则,当 \mathbf{A} 负定且初始向量位于单位超球面外时,系统一定出现有限溢,算法失效.因此,本文提出一种计算一般实对称矩阵 \mathbf{A} 的特征向量及特征值的神经网络(式(2))方法.式(2)的参数意义同式(1),显然,式(2)也为连续型全反馈神经网络.

$$\frac{d}{dt} \mathbf{X}(t) = \mathbf{A}\mathbf{X}(t) - \frac{\mathbf{X}^T(t)\mathbf{A}\mathbf{X}(t)\mathbf{X}(t)}{\mathbf{X}^T(0)\mathbf{X}(0)}, \mathbf{X}^T(0)\mathbf{X}(0) \neq 0 \quad (2)$$

由模拟电子线路知识得知,在网络(1)模拟电路基础上^[4],增加寄存器存储 $\mathbf{X}^T(0)\mathbf{X}(0)$,增加除法器实现 $\frac{\mathbf{X}^T(t)\mathbf{A}\mathbf{X}(t)}{\mathbf{X}^T(0)\mathbf{X}(0)}$,式(2)完全可由电路实现.

1 符号约定和实对称矩阵的性质及其与网络平衡解向量的联系

1.1 符号约定

\mathbf{A} 的全部特征值设为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, 对应特征向量为 μ_1, \dots, μ_n , 对应特征子空间为 V_1, \dots, V_n , 用 V 表示 $V_1 \oplus V_2 \oplus \dots \oplus V_n$, 并记 $V_0^\perp = V$, V_0^\perp 表示 \mathbb{R}^n 中 V_0 的正交补空间.

假设 \mathbf{A} 有 m 个不同特征值 $\sigma_1 > \dots > \sigma_m$, 前 i 个特征值代数重数之和为 $K_i (i=1, \dots, m)$, 显然 $m \leq n$, $K_m = n$, 并记 $K_0 = 0$.

令 $U_i = V_{k_{i-1}} \oplus \dots \oplus V_{k_i} (i=1, \dots, m)$, 并记 $U_0^\perp = V$.

1.2 实对称矩阵性质

性质 1. 特征值特征向量为实数. 不同特征值对应的特征向量必正交(证略).

性质 2. 对 n 维矩阵, 所有不同特征值对应子空间的规范正交基构成 n 维空间的一组规范正交基(证略).

1.3 矩阵特征值特征向量与网络平衡解向量的关系

由式(2)得知, 网络平衡时存在以下关系

$$AX(\infty) = \frac{X^T(\infty)AX(\infty)}{X^T(0)X(0)} X(\infty), X^T(0)X(0) \neq 0 \quad (3)$$

若式(2)的 $X(\infty)$ 为非零向量, 则 $X(\infty)$ 为 A 的特征向量, 对应 $X(\infty)$ 的特征值为

$$\lambda = \frac{X^T(\infty)AX(\infty)}{X^T(0)X(0)} \quad (4)$$

定理 8 将证明, 对非零初始向量 $X(0)$, 式(2)的 $X(\infty)$ 必为非零向量, 因此式(4)总是有意义的.

2 lyNN 的解析解

定理 1. 令 $S_i = \frac{\mu_i}{\|\mu_i\|}$, 则根据实对称矩阵性质 2, S_1, \dots, S_n 构成 R^n 空间中一组标准正交基, R^n 中任意向量可

表述为 $X(t) = \sum_{i=1}^n z_i(t)S_i$, 对任意非零初始向量 $X(0) = \sum_{i=1}^n z_i(0)S_i$ ($\prod_{i=1}^n z_i(0) \neq 0$), 式(2)的解析解为

$$X(t) = \frac{\sqrt{\sum_{j=1}^n z_j^2(0) \sum_{i=1}^n z_i(0) \exp(\lambda_i t) S_i}}{\sqrt{\sum_{j=1}^n z_j^2(0) \exp(2\lambda_j t)}} \quad (5)$$

证明: 将 $X(t) = \sum_{i=1}^n z_i(t)S_i$ 和 $X(0) = \sum_{i=1}^n z_i(0)S_i$ 代入式(2), 有

$$\frac{d}{dt} z_i(t) = \lambda_i z_i(t) - \frac{\sum_{j=1}^n \lambda_j z_j^2(t) E_i(t)}{\sum_{j=1}^n z_j^2(0)} \quad (6)$$

类似文献[5], 分析有

$$\frac{z_i(t)}{z_r(t)} = \frac{z_i(0)}{z_r(0)} \exp[(\lambda_i - \lambda_r)t] \quad (7)$$

由式(6)、式(7)可得

$$\frac{d}{dt} \frac{\sum_{j=1}^n z_j^2(0)}{z_r^2(t)} = -2\lambda_r \frac{\sum_{j=1}^n z_j^2(0)}{z_r^2(t)} + 2 \sum_{j=1}^n \lambda_j \frac{z_j^2(0)}{z_r^2(0)} \exp[2(\lambda_j - \lambda_r)t],$$

等价形式为

$$\exp(2\lambda_r t) \frac{d}{dt} \frac{\sum_{j=1}^n z_j^2(0)}{z_r^2(t)} + 2\lambda_r \exp(2\lambda_r t) \frac{\sum_{j=1}^n z_j^2(0)}{z_r^2(t)} = 2 \sum_{j=1}^n \lambda_j \frac{z_j^2(0)}{z_r^2(0)} \exp[2\lambda_j t],$$

上式两边对 t 积分可得

$$\exp(2\lambda_r t) \frac{\sum_{j=1}^n z_j^2(0)}{z_r^2(t)} - \frac{\sum_{j=1}^n z_j^2(0)}{z_r^2(0)} = \sum_{j=1}^n \frac{z_j^2(0)}{z_r^2(0)} \exp(2\lambda_j t) - \sum_{j=1}^n \frac{z_j^2(0)}{z_r^2(0)},$$

故得到

$$z_r(t) = \frac{z_r(0) \exp(\lambda_r t) \sqrt{\sum_{j=1}^n z_j^2(0)}}{\sqrt{\sum_{j=1}^n z_j^2(0) \exp(2\lambda_j t)}} \tag{8}$$

由式(7)、式(8)可得

$$z_i(t) = \frac{z_i(0) \exp(\lambda_i t) \sqrt{\sum_{j=1}^n z_j^2(0)}}{\sqrt{\sum_{j=1}^n z_j^2(0) \exp(2\lambda_j t)}} \tag{9}$$

所以

$$\mathbf{X}(t) = \sum_{i=1}^n z_i(t) \mathbf{S}_i = \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \sum_{i=1}^n z_i(0) \exp(\lambda_i t) \mathbf{S}_i}{\sqrt{\sum_{j=1}^n z_j^2(0) \exp(2\lambda_j t)}} \tag{10}$$

□

3 lyNN 的渐近行为

定理 2. 如果非零初始向量 $\mathbf{X}(0) \in V_i$, 则式(2)的平衡解向量 $\mathbf{X}(\infty) \in V_i$, 得到的特征值为 λ_i .

证明: 因为 $\mathbf{X}(0) \in V_i$, 对任意 $V_j (j \neq i, j=1, \dots, n)$, $V_i \perp V_j$, 故 $\mathbf{X}(0)$ 在 V_j 的投影为 0, 由定理 1 可得

$$\mathbf{X}(t) = \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \sum_{i=1}^n z_i(0) \exp(\lambda_i t) \mathbf{S}_i}{\sqrt{\sum_{j=1}^n z_j^2(0) \exp(2\lambda_j t)}} = \frac{\sqrt{z_i^2(0)} z_i(0) \exp(\lambda_i t) \mathbf{S}_i}{\sqrt{z_i^2(0) \exp(2\lambda_i t)}} = z_i(0) \mathbf{S}_i \in V_i.$$

所以, $\mathbf{X}(\infty) \in V_i$, 再由式(4)可知特征值为 λ_i . □

定理 3. 若非零初始向量 $\mathbf{X}(0) \notin V_i (i=1, \dots, n)$, 式(2)的平衡解向量属于 \mathbf{A} 的最大特征值对应的特征向量张成的子空间.

证明: 因为非零初始向量 $\mathbf{X}(0) \notin V_i (i=1, \dots, n)$, 所以 $\mathbf{X}(0)$ 对空间 $V_i (i=1, \dots, n)$ 的投影 $z_i(0) \neq 0$, 由定理 1 得

$$\begin{aligned} \mathbf{X}(t) &= \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \sum_{i=1}^n z_i(0) \exp(\lambda_i t) \mathbf{S}_i}{\sqrt{\sum_{j=1}^n z_j^2(0) \exp(2\lambda_j t)}} = \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \left(\sum_{i=K_0+1}^{K_1} z_i(0) \mathbf{S}_i + \sum_{j=1}^{m-1} \sum_{i=K_j+1}^{K_{j+1}} z_i(0) \exp((\sigma_{j+1} - \sigma_j) t) \mathbf{S}_i \right)}{\sqrt{\sum_{i=K_0+1}^{K_1} z_i^2(0) + \sum_{j=1}^{m-1} \sum_{i=K_j+1}^{K_{j+1}} z_i^2(0) \exp(2(\sigma_{j+1} - \sigma_j) t)}} \\ &\xrightarrow{t \rightarrow \infty} \pm \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \sum_{i=K_0+1}^{K_1} z_i(0) \mathbf{S}_i}{\sqrt{\sum_{i=K_0+1}^{K_1} z_i^2(0)}} \in V_{K_0+1} \oplus \dots \oplus V_{K_1}. \end{aligned} \tag{11}$$

□

定理 4. 若非零初始向量 $\mathbf{X}(0)$ 属于 $\bigcap_{i=0}^{k-1} U_i^\perp - \bigcap_{i=0}^k U_i^\perp (k=1, \dots, m)$, 则式(2)的平衡解向量属于 $U_i (i=1, \dots, m)$.

证明:I. 当 $k=1$ 时,非零初始向量 $X(0) \in U_0^\perp - U_1^\perp = V - U_1^\perp$, $\sum_{i=K_0+1}^{K_1} z_i(0) \neq 0$, 由定理 1 易证:

$$X(t) = \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \left(\sum_{i=K_0+1}^{K_1} z_i(0) S_i + \sum_{j=1}^{m-1} \sum_{i=K_j+1}^{K_{j+1}} z_i(0) \exp((\sigma_{j+1} - \sigma_1) t) S_i \right)}{\sqrt{\sum_{i=K_0+1}^{K_1} z_i^2(0) + \sum_{j=1}^{m-1} \sum_{i=K_j+1}^{K_{j+1}} z_i^2(0) \exp(2(\sigma_{j+1} - \sigma_1) t)}} \xrightarrow{t \rightarrow \infty} \pm \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \sum_{i=K_0+1}^{K_1} z_i(0) S_i}{\sqrt{\sum_{i=K_0+1}^{K_1} z_i^2(0)}}$$

故 $X(\infty) \in V_{K_0+1} \oplus \dots \oplus V_{K_1} = U_1$, 结论成立.

II. 当 $k=2, \dots, m$ 时,非零初始向量 $X(0) \in \bigcap_{i=0}^{k-1} U_i^\perp - \bigcap_{i=0}^k U_i^\perp$, 由于 $X(0) \in U_i^\perp (i=1, \dots, k-1)$, 所以 $z_i(0)=0, (i=1, \dots, K_{k-1})$,

又由于 $X(0) \in \bigcap_{i=0}^{k-1} U_i^\perp - \bigcap_{i=0}^k U_i^\perp$, 所以 $\sum_{i=K_{k-1}+1}^{K_k} z_i(0) \neq 0$. 由定理 1 可得

$$\begin{aligned} X(t) &= \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \sum_{i=1}^n z_i(0) \exp(\lambda_i t) S_i}{\sqrt{\sum_{j=1}^n z_j^2(0) \exp(2\lambda_j t)}} = \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \left(\sum_{j=k-1}^{m-1} \sum_{i=K_j+1}^{K_{j+1}} z_i(0) \exp(\sigma_{j+1} t) S_i \right)}{\sqrt{\sum_{j=k-1}^{m-1} \sum_{i=K_j+1}^{K_{j+1}} z_i^2(0) \exp(2\sigma_{j+1} t)}} \\ &= \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \left(\sum_{i=K_{k-1}+1}^{K_k} z_i(0) S_i + \sum_{j=k}^{m-1} \sum_{i=K_j+1}^{K_{j+1}} z_i(0) \exp((\sigma_{j+1} - \sigma_k) t) S_i \right)}{\sqrt{\sum_{i=K_{k-1}+1}^{K_k} z_i^2(0) + \sum_{j=k}^{m-1} \sum_{i=K_j+1}^{K_{j+1}} z_i^2(0) \exp(2(\sigma_{j+1} - \sigma_k) t)}} \xrightarrow{t \rightarrow \infty} \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \left(\sum_{i=K_{k-1}+1}^{K_k} z_i(0) S_i \right)}{\sqrt{\sum_{i=K_{k-1}+1}^{K_k} z_i^2(0)}} \in U_k. \end{aligned}$$

综合 I 和 II 可知结论成立. □

定理 5. 若非零初始向量 $X(0)$ 分别属于 $V_0^\perp - V_1^\perp, \dots, \bigcap_{i=0}^{k-1} V_i^\perp - \bigcap_{i=0}^k V_i^\perp, \dots, \bigcap_{i=0}^{n-1} V_i^\perp - \bigcap_{i=0}^n V_i^\perp$, 则式(2)的平衡解向量分别属于 $V_1, \dots, V_k, \dots, V_n$.

证明:类似定理 4 的证明过程,此定理不难得证.

本定理是文献[5]中定理 3 的更一般结果,该定理相当于本定理 $k=1$ 的情况.

定理 6. 要求式(2)的平衡解向量 $X(\infty)$ 分别属于 $U_1, \dots, U_k, \dots, U_m$, 则初始向量 $X(0)$ 在 R^n 中的最大取值空间分别为 $U_0^\perp - U_1^\perp, \dots, \bigcap_{i=0}^{k-1} U_i^\perp - \bigcap_{i=1}^k U_i^\perp, \dots, \bigcap_{i=0}^{m-1} U_i^\perp - \bigcap_{i=0}^m U_i^\perp$.

证明:利用反证法.假设 $X(0) \notin \bigcap_{i=0}^{k-1} U_i^\perp - \bigcap_{i=0}^k U_i^\perp (1 \leq k \leq m)$, 而 $X(\infty) \in U_k$, 下面证明这是矛盾的.

因为 $X(0) \notin \bigcap_{i=0}^{k-1} U_i^\perp - \bigcap_{i=0}^k U_i^\perp$, 所以 $X(0) \in V - \left(\bigcap_{i=0}^{k-1} U_i^\perp - \bigcap_{i=0}^k U_i^\perp \right)$, 即

$$X(0) \in V - \bigcap_{i=0}^{k-1} U_i^\perp + \bigcap_{i=0}^k U_i^\perp = (U_0^\perp - U_1^\perp) + (U_1^\perp - U_1^\perp \cap U_2^\perp) + \dots + \left(\bigcap_{i=0}^{k-2} U_i^\perp - \bigcap_{i=0}^{k-1} U_i^\perp \right) + \bigcap_{i=0}^k U_i^\perp.$$

根据鸽巢原理, $X(0)$ 必属于 $\bigcap_{i=0}^k U_i^\perp, \left(\bigcap_{i=0}^\lambda U_i^\perp - \bigcap_{i=0}^{\lambda+1} U_i^\perp \right) (\lambda=0, \dots, k-2)$ 的其中之一.

I. 若 $X(0) \in \bigcap_{i=0}^k U_i^\perp$, 则 $z_i(0)=0 (i=1, \dots, K_k)$, 不失一般性, 设第 1 个非零值为 $z_L(0) \neq 0$, 并设 $l > k, K_l + 1 \leq L \leq K_{l+1}$, 由定理 1 可得

$$\begin{aligned} \mathbf{X}(t) &= \frac{\sqrt{\sum_{j=1}^n z_j^2(0) \sum_{i=1}^n z_i(0) \exp(\lambda_i t) \mathbf{S}_i}}{\sqrt{\sum_{j=1}^n z_j^2(0) \exp(2\lambda_j t)}} = \frac{\sqrt{\sum_{j=1}^n z_j^2(0) \sum_{j=0=K_{j+1}}^{m-1} \sum_{i=K_{j+1}}^{K_{j+1}} z_i(0) \exp(\sigma_{j+1} t) \mathbf{S}_i}}{\sqrt{\sum_{j=0=K_{j+1}}^{m-1} \sum_{i=K_{j+1}}^{K_{j+1}} z_i^2(0) \exp(2\sigma_{j+1} t)}} \\ &= \frac{\sqrt{\sum_{j=1}^n z_j^2(0) \left(\sum_{i=L}^{K_{j+1}} z_i(0) \mathbf{S}_i + \sum_{j=l+1=K_{j+1}}^{m-1} \sum_{i=K_{j+1}}^{K_{j+1}} z_i(0) \exp((\sigma_{j+1} - \sigma_l) t) \mathbf{S}_i \right)}}{\sqrt{\sum_{i=L}^{K_{j+1}} z_i^2(0) + \sum_{j=l+1=K_{j+1}}^{m-1} \sum_{i=K_{j+1}}^{K_{j+1}} z_i^2(0) \exp(2(\sigma_{j+1} - \sigma_l) t)}} \xrightarrow{t \rightarrow \infty} \frac{\sqrt{\sum_{j=1}^n z_j^2(0) \left(\sum_{i=L}^{K_{j+1}} z_i(0) \mathbf{S}_i \right)}}{\sqrt{\sum_{i=L}^{K_{j+1}} z_i^2(0)}} \in U_l. \end{aligned}$$

所以,式(2)的平衡解向量 $\mathbf{X}(\infty) \notin U_k$, 与假设矛盾.

II. 若 $\mathbf{X}(0) \in \left(\bigcap_{i=0}^{\lambda} U_i^\perp - \bigcap_{i=0}^{\lambda+1} U_i^\perp \right) (\lambda = 0, \dots, k-2)$, 则由定理 4 可知, 式(2)的平衡解向量 $\mathbf{X}(\infty)$ 属于 $U_\lambda (1 \leq \lambda \leq k-1)$, 不属于 U_k , 也与假设矛盾. 综合 I 和 II 可知, 结论成立. \square

定理 7. 假设特征向量 S_h 已知, 若 $\mathbf{X}(0) \perp S_h$, 则式(2)的平衡解向量 $\mathbf{X}(\infty) \perp S_h$.

证明: 因为 S_1, \dots, S_n 构成 R^n 空间中一组标准正交基, $\bigcap_{i=0}^n V_i^\perp = \{0\}$, 又由于

$$(V - V_1^\perp) + (V_1^\perp - V_1^\perp \cap V_2^\perp) + (V_1^\perp \cap V_2^\perp - V_1^\perp \cap V_2^\perp \cap V_3^\perp) + \dots + \left(\bigcap_{i=0}^{n-1} V_i^\perp - \bigcap_{i=0}^n V_i^\perp \right) = V - \bigcap_{i=0}^n V_i^\perp = V - \{0\}.$$

所以, $(V - V_1^\perp) + (V_1^\perp - V_1^\perp \cap V_2^\perp) + (V_1^\perp \cap V_2^\perp - V_1^\perp \cap V_2^\perp \cap V_3^\perp) + \dots + \left(\bigcap_{i=0}^{n-1} V_i^\perp - \bigcap_{i=0}^n V_i^\perp \right)$ 为非零初始向量 $\mathbf{X}(0)$ 的完备取值空间, 根据鸽巢原理, 随机确定的初始向量 $\mathbf{X}(0)$ 必属于 $V - V_1^\perp, V_1^\perp - V_1^\perp \cap V_2^\perp, V_1^\perp \cap V_2^\perp - V_1^\perp \cap V_2^\perp \cap V_3^\perp, \dots, \bigcap_{i=0}^{n-1} V_i^\perp - \bigcap_{i=0}^n V_i^\perp$ 的其中之一; 假设 $\mathbf{X}(0) \perp S_h$, 则 $\mathbf{X}(0) \in V_h^\perp$; 因为 $\left(\bigcap_{i=0}^{k-1} V_i^\perp - \bigcap_{i=0}^k V_i^\perp \right) \cap V_k^\perp = \bigcap_{i=0}^k V_i^\perp - \bigcap_{i=0}^k V_i^\perp$ 为空集, 所以 $V_h^\perp \subseteq V - \left(\bigcap_{i=1}^{h-1} V_i^\perp - \bigcap_{i=1}^h V_i^\perp \right)$, 故有:

$$\mathbf{X}(0) \in V_h^\perp \subseteq V - \left(\bigcap_{i=0}^{h-1} V_i^\perp - \bigcap_{i=0}^h V_i^\perp \right) = (V - V_1^\perp) + \dots + \left(\bigcap_{i=0}^{h-2} V_i^\perp - \bigcap_{i=0}^{h-1} V_i^\perp \right) + \dots + \left(\bigcap_{i=0}^h V_i^\perp - \bigcap_{i=0}^{h+1} V_i^\perp \right) + \dots + \left(\bigcap_{i=0}^{n-1} V_i^\perp - \bigcap_{i=0}^n V_i^\perp \right).$$

再根据鸽巢原理, 垂直于 S_h 的 $\mathbf{X}(0)$ 必然属于 $V_0^\perp - V_1^\perp, V_1^\perp - V_1^\perp \cap V_2^\perp, \dots, \bigcap_{i=0}^{h-2} V_i^\perp - \bigcap_{i=0}^{h-1} V_i^\perp, \bigcap_{i=0}^h V_i^\perp - \bigcap_{i=0}^{h+1} V_i^\perp, \dots,$

$\bigcap_{i=0}^{n-1} V_i^\perp - \bigcap_{i=0}^n V_i^\perp$ 的其中之一, 又据定理 5, 式(2)必然收敛到 $V_1, V_2, \dots, V_h, V_{h+1}, \dots, V_n$ 的其中之一. 所以平衡解向量 $\mathbf{X}(\infty) \notin V_h$, 即 $\mathbf{X}(\infty) \perp S_h$. \square

定理 8. 式(2)从任意非零初始向量 $\mathbf{X}(0)$ 出发的平衡解向量 $\mathbf{X}(\infty)$ 都位于半径为 $\|\mathbf{X}(0)\|$ 的超球面上.

证明: 对任意非零初始向量 $\mathbf{X}(0)$, 由定理 1 得

$$\mathbf{X}(t) = \frac{\sqrt{\sum_{j=1}^n z_j^2(0) \sum_{i=1}^n z_i(0) \exp(\lambda_i t) \mathbf{S}_i}}{\sqrt{\sum_{j=1}^n z_j^2(0) \exp(2\lambda_j t)}} = \frac{\sqrt{\sum_{j=1}^n z_j^2(0) \sum_{j=0=K_{j+1}}^{m-1} \sum_{i=K_{j+1}}^{K_{j+1}} z_i(0) \exp(\sigma_{j+1} t) \mathbf{S}_i}}{\sqrt{\sum_{j=0=K_{j+1}}^{m-1} \sum_{i=K_{j+1}}^{K_{j+1}} z_i^2(0) \exp(2\sigma_{j+1} t)}}.$$

不失一般性, 假设第 1 个非零值为 $z_p(0) \neq 0, K_q < p < K_{q+1}$, 则由上式得

$$\mathbf{X}(t) = \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \left(\sum_{i=K_q+1}^{K_{q+1}} z_i(0) \mathbf{S}_i + \sum_{j=q+1}^{m-1} \sum_{i=K_{j+1}}^{K_{j+1}} z_i(0) \exp((\sigma_{j+1} - \sigma_{q+1})t) \mathbf{S}_i \right)}{\sqrt{\sum_{i=K_q+1}^{K_{q+1}} z_i^2(0) + \sum_{j=q+1}^{m-1} \sum_{i=K_{j+1}}^{K_{j+1}} z_i^2(0) \exp(2(\sigma_{j+1} - \sigma_{q+1})t)}} \xrightarrow{t \rightarrow \infty} \frac{\sqrt{\sum_{j=1}^n z_j^2(0)} \left(\sum_{i=K_q+1}^{K_{q+1}} z_i(0) \mathbf{S}_i \right)}{\sqrt{\sum_{i=K_q+1}^{K_{q+1}} z_i^2(0)}}$$

所以, $\|\mathbf{X}(\infty)\| = \sqrt{\sum_{j=1}^n z_j^2(0)} = \|\mathbf{X}(0)\|$. □

此定理表明 $\mathbf{X}(\infty)$ 不是零向量,用式(4)计算特征值总是有意义的.

4 计算特征向量及特征值

根据 lyNN 的渐近行为分析,设计计算方法如下:

- 1) 在 R^n 中任取非零初始向量代入式(2)得到向量 \mathbf{S}_1 (这里,下标表示求解的先后关系),依据式(4)得到 λ_1 .
- 2) 在垂直 \mathbf{S}_1 的任意方向取非零初始向量代入式(2)得到向量 \mathbf{S}_2 ,依据式(4)得到 λ_2 .
- 3) 在垂直 $\mathbf{S}_1, \mathbf{S}_2$ 的任意方向取非零初始向量代入式(2)得到向量 \mathbf{S}_3 ,依据式(4)得到 λ_3 .
- 4) 在垂直 $\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_k$ 的任意方向取非零初始向量代入式(2)得到向量 \mathbf{S}_{k+1} ,依据式(4)得到 λ_{k+1} .反复如此操作,就可以算出所有特征向量及特征值.

下面以两个算例示意上面算法.

例 1:计算文献[5]所示矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$ 的特征值及特征向量.

解:把 \mathbf{A} 代入式(2),得到 $\frac{d}{dt} \mathbf{X}(t) = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \mathbf{X}(t) - \frac{1}{\mathbf{X}^T(0)\mathbf{X}(0)} \mathbf{X}^T(t) \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \mathbf{X}(t) \mathbf{X}(t)$,任取 $\mathbf{X}_1(0) = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$,则网

络将收敛到 $\mathbf{X}(\infty) = \frac{\sqrt{5}}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$,把 \mathbf{S}_1 代入式(4)得到 $\lambda_1 = \frac{\mathbf{X}^T(\infty)\mathbf{A}\mathbf{X}(\infty)}{\mathbf{X}_1^T(0)\mathbf{X}_1(0)} = 5$;任取垂直于 \mathbf{S}_1 的初始向量 $\mathbf{X}_2(0) = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$,

则式(2)收敛到 $\mathbf{X}(\infty) = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$,有式(4)得到 $\lambda_2 = \frac{\mathbf{X}^T(\infty)\mathbf{A}\mathbf{X}(\infty)}{\mathbf{X}_2^T(0)\mathbf{X}_2(0)} = -1$;任取垂直于 $\mathbf{S}_1, \mathbf{S}_2$ 的初始向量 $\mathbf{X}_3(0) = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$,

则式(2)收敛到 $\mathbf{X}(\infty) = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$,从式(4)得到 $\lambda_3 = -1$.

例 2:计算矩阵 $\begin{bmatrix} -1 & -2 & -1 \\ -2 & -5 & -1 \\ -1 & -1 & -2 \end{bmatrix}$ 的特征值及特征向量.

解:把 $\begin{bmatrix} -1 & -2 & -1 \\ -2 & -5 & -1 \\ -1 & -1 & -2 \end{bmatrix}$ 代入式(2)得

$$\frac{d}{dt} \mathbf{X}(t) = \begin{bmatrix} -1 & -2 & -1 \\ -2 & -5 & -1 \\ -1 & -1 & -2 \end{bmatrix} \mathbf{X}(t) - \frac{1}{\mathbf{X}^T(0)\mathbf{X}(0)} \mathbf{X}^T(t) \begin{bmatrix} -1 & -2 & -1 \\ -2 & -5 & -1 \\ -1 & -1 & -2 \end{bmatrix} \mathbf{X}(t) \mathbf{X}(t).$$

$$\text{任取 } X_1(0) = \begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}, \text{ 则式(2)收敛到 } X(\infty) = \sqrt{11} \begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}, \text{ 由式(4)得 } \lambda_1 = 0; \text{ 任取垂直于 } S_1 \text{ 的初始向量}$$

$$X_2(0) = \begin{bmatrix} -1 \\ \sqrt{5} \\ -(3+\sqrt{5}) \end{bmatrix}, \text{ 则 } X(\infty) = \sqrt{(20+6\sqrt{5})} \begin{bmatrix} -1 \\ \sqrt{5} \\ -(3+\sqrt{5}) \end{bmatrix}, \text{ 从式(4)得 } \lambda_2 = -4 + \sqrt{5}; \text{ 任取垂直于 } S_1, S_2 \text{ 的初始向量}$$

$$X_3(0) = \begin{bmatrix} 1 \\ \sqrt{5} \\ 3-\sqrt{5} \end{bmatrix}, \text{ 则 } X(\infty) = \begin{bmatrix} 1 \\ \sqrt{5} \\ 3-\sqrt{5} \end{bmatrix}, \text{ 借助式(4)得 } \lambda_3 = -4 - \sqrt{5}.$$

5 讨论

lyNN 与基于优化的方法相比,依据式(2)可直接构建神经网络.而基于优化的方法必须是把求解矩阵特征值特征向量的问题转化为优化问题,再用神经网络求解,计算量一般较大.

与基于 Oja-N 的方法相比,lyNN 具有初始向量 $X(0)$ 取值广泛的特点.Oja-N 要求初始向量必须位于单位超球面上,否则在矩阵特征值为负且 $X(0)$ 位于单位超球面外时,网络存在有限溢.而 lyNN 只要初始向量非零,无论矩阵特征值如何,都不存在有限溢.例 1 所用矩阵与文献[5]举例所用矩阵相同,文献[5]所取初始向量全部位于单位超球面上,例 1 所取初始向量都不位于单位超球面上,同样得到全部特征值和特征向量.例 2 待求矩阵的特征值 2 负 1 零,如果初始向量位于单位超球面外,文献[5]所示方法存在有限溢,不能得到所求结果;而利用 lyNN 时,由例 2 可见,单位超球面外的初始向量不影响收敛结果的准确性.因此,本文提出的神经网络完全克服了文献[5]所示 Oja-N 神经网络有限溢的缺点.

总体而言,下面是本文方法与其他方法各种性能的对比.

方法	求解途径	得到全部值的措施	限制条件	实现难易
基于 Oja-N 的方法 ^[1-5]	直接构造神经网络	改变初始向量取值	初始向量须位于单位超球面	较易
基于优化的方法 ^[6-8]	化为优化问题,再构造神经网络	构造矩阵序列等措施	初始向量不能为零向量	计算量一般较大,使用有限
本文方法 lyNN	直接构造神经网络	改变初始向量取值	初始向量不能为零向量	较易

6 结束语

本文提出了一种用于求解矩阵特征向量特征值的神经网络.探讨了网络平衡解向量与矩阵特征值特征向量的关系,严格的理论证明了有关该神经网络的以下结果:

- 1) 非零初始向量属于任意特征子空间,则网络平衡解向量也属于该空间,若初始向量不属于任意特征子空间,则平衡解向量属于最大特征值对应的特征向量张成的子空间;
- 2) 网络收敛于各特征值对应的特征向量张成的子空间时,初始向量的取值空间,并证明该空间是 R^n 中符合条件的最大初始向量取值空间;
- 3) 在与已知特征向量垂直的子空间取非零初始向量,网络平衡解向量也将垂直于该特征向量;
- 4) 平衡解向量位于由非零初始向量确定的超球面上.

基于以上分析,本文设计了用该神经网络求解实对称矩阵特征向量特征值的计算步骤,并用两个例子验证了该求解方法的有效性.同基于 Oja-N 的方法比,lyNN 具有初始向量不受单位超球面限制、不存在有限溢等特点,两个举例也体现了这种特性.与基于优化的方法相比,lyNN 可直接构造神经网络,计算量较小.

致谢 作者衷心感谢审稿人和编辑老师对本文提出的修改建议.

References:

[1] Oja E. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 1982,15:267-273.

[2] Oja E, Karhunen J. On stochastic approximation of eigenvectors and eigenvalues of the expectation of a random matrix. *Journal of Mathematical Analysis and Applications*, 1985,106:69-84.

[3] Luo FL, Li YD. Neural network approach to computing the eigenvectors corresponding to the largest and smallest eigenvalues of a positive matrix. *Acta Electronica Sinica*, 1994,22(1):13-19 (in Chinese with English abstract).

[4] Luo FL, Li YD, Wang Z. Neural network algorithm for solving eigenvalues and eigenvectors of a type positive matrix. *Science in China (Series A)*, 1994,24(1):83-88 (in Chinese with English abstract).

[5] Zhang Y, Wang PA, Zhou MT. Computing eigenvalues and eigenvectors of matrix by neural networks. *Chinese Journal of Computers*, 2000,23(1):71-76 (in Chinese with English abstract).

[6] Cichocki A, Unbehauen R. Neural networks for computing eigenvalues and eigenvectors. *Biological Cybernetics*, 1992,68:155-164.

[7] Tan Y, He ZY. Neural network approaches for the extraction of the eigenstructure. In: *Neural Networks for Signal Processing VI-Proc. Kyoto: IEEE Workshop*, 1996. 21-32.

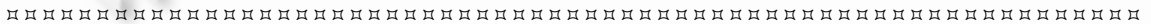
[8] Tan Y, Liu Z. On matrix eigendecomposition by neural networks. *Neural Network World*, 1998,3:337-352.

附中文参考文献:

[3] 罗发龙,李衍达.求解正定矩阵最小和最大特征值对应的特征矢量.电子学报,1994,22(1):13-19.

[4] 罗发龙,李衍达,王哲.一类正定矩阵特征矢量的神经网络求解.中国科学(A辑),1994,24(1):83-88.

[5] 章毅,王平安,周明天.用神经网络计算矩阵的特征值和特征向量.计算机学报,2000,23(1):71-76.



2005 中国计算机大会 征文通知

由中国计算机学会主办的 2005 中国计算机大会(China National Computer Conference 2005, CNCC 2005)将于 2005 年 10 月 13 日至 15 日在武汉举行。它将提供中国计算机界交流最新研究成果的良好舞台。CNCC 2005 是继 2003 年 11 月 26-28 日在北京成功举办 2003 中国计算机大会之后的中国计算机界又一次盛会。CNCC 2005 的议题涉及计算机领域各个方面。本次大会将举办若干分会进行相关议题的讨论，同时将举办讲座和展览。与会议同时举行的还有中国计算机学会理事会议。CNCC 2005 诚请广大计算机界研究人员投稿。

会议的议题主要包括（但不限于此）：

计算机体系结构	集成电路设计	数据库与信息检索
高性能计算和网络计算	嵌入式计算与系统	人机接口、多媒体和虚拟实现
互联网和移动计算	人工智能	生物信息学
信息安全与可信计算	算法与算法复杂性	计算机应用技术
外存储系统	软件工程和中间件	

投稿须知

作者投往本届大会的稿件必须是原始的、未发表的研究成果、研究经验或工作突破性进展报告。稿件请以 pdf 格式提交。所有稿件将依据统一的原则进行审理。大会根据稿件的审理结果决定稿件是否录用。所有录用稿件将收录在本届大会论文集中。此外，本届大会的优秀稿件将推荐往《计算机学报》、《软件学报》、《计算机研究与发展》、《小型微型计算机系统》、《高技术通讯》上发表。

重要日期

征稿截止	2005 年 6 月 15 日
论文处理结果通知	2005 年 7 月 15 日
注册截止	2005 年 8 月 15 日