第 31 卷 第 4 期 电 子 科 技 大 学 学 报 Vol.31 No.4
2002 年 8 月 Journal of UEST of China Aug. 2002

Eigensubspace Estimation Using Discrete Recurrent Neural Networks*

Liang Jinming[1] Zhang Yi[2]

(1. Dept.of Computer Science, Zigong Teacher's College Zigong 643000;

2. College of Computer Science and Engineering, UEST of China Chengdu 610054)

Abstract This paper proposes two models of discrete recurrent neural networks to study the problem of eigensubspace estimation for positive definite symmetric matrix. The first model is a class of nonlinear neural networks. It is used for estimating the largest eigenvalue and one of its corresponding eigenvectors. The second model is a class of linear neural networks which estimates the eigensubspace corresponding to the largest eigenvalue. Dynamic properties of these two classes of discrete recurrent neural network models are studied and used for eigensubspace estimation.

Key words eigenvalue; eigenvector; eigensubspace; recurrent neural network

回复式离散神经网络的特征子空间估值*

梁金明**[1] 章 毅[2]

(1. 自贡师范高等专科学校计算机科学系 自贡 643000; 2. 电子科技大学计算机科学与工程学院 成都 610054)

【摘要】提出了用两种回复式离散神经网络模型研究正定对称矩阵的特征子空间估值问题：第1种模型是非线性神经网络，用于计算最大特征值及其特征向量；第2种模型属于线性神经网络，用于计算相应于最大特征值的特征子空间。详细研究了两种离散神经回路网络模型的动力学性质并用于特征子空间估值。

关 键 词 特征值; 特征向量; 特征子空间; 神经网络

中图分类号 TN711.4; TP393

Given A, an $n \times n$ real positive definite symmetric matrix. We study in this paper the problem of estimating the eigensubspace corresponding to the largest eigenvalue of A. Eigensubspace estimation has a lot of applications, especially in adaptive signal processing. The problem of eigensubspace estimation has been widely studied in recent years. Many algorithms based on neural networks have been proposed, see Ref.[1~13] for examples.

In Ref.[2~4，8~10，12], stochastic learning algorithms for eigensubspace estimation are proposed. These algorithms are expressed in the form of discrete iterative equations. The convergence of these algorithms are proven by studying the convergence properties of the associated deterministic ordinary differential equations of these stochastic algorithms. However, as pointed out in Ref.[11], these

---

Received on April 24, 2002
2002年4月24日收稿
* The Project is supported by the Natural Science Foundation of China, No: 69871005
国家自然科学基金资助项目，编号：69871005
** 男 36岁 大学 讲师

approximations of stochastic algorithms by differential equations may not hold. Hence, the convergence of these stochastic learning algorithms could not be guaranteed.

In Ref.[5, 6], the problem of estimating the eigensubspace corresponding to the smallest eigenvalue is recasted by a constraint optimization problem which is solved by using continuous neural networks model. Later, these results were extended to derive a Newton based adaptive algorithm in Ref.[7].

One of our observations is that most of the existing algorithms are too complicated and inefficient to be used in practice, especially when the dimensions of the input matrix is large. Moreover, continuous neural network models cannot be implemented easily using digital hardware and computer simulation.

In this paper, we propose two simple discrete time recurrent neural networks to address the problem of estimating the eigensubspace corresponding to the largest eigenvalue of A. These networks converge very quickly with great precision. Since the networks are formulated as discrete time system, they favor computer simulations and have advantages over digital simulations of continuous time models. In addition, it can be easily implemented in digital hardware. These neural network models have very clear dynamic behaviors since we can solve the dynamic systems of the neural networks to obtain representations of the solutions.

We organize our work as follows. Preliminaries and problem formulation are presented in Section 2. In Section 3, our nonlinear discrete time recurrent neural network model is presented. We discuss the representation of the solutions of the network and its convergence properties. In Section 4, we propose a linear discrete time neural network model for estimating the eigensubspace that corresponds to the largest eigenvalue. Section 5 concludes the paper.

1 Preliminaries and Problem Formulation

Let $l_i (i = 1,2,\cdots,n)$ be all the eigenvalues of A with $l_1 \geq l_2 \geq \cdots \geq l_n > 0$. Since A ia a symmetric matrix, there exists an orthogonal basis $S_i(i=1, 2,\cdots, n)$ in R^n and $S_i(i=1,2,\cdots, n)$ are eigenvectors of A corresponding to the eigenvalues $l_i (i = 1,2,\cdots,n)$. Let $s_i (i = 1,2,\cdots,m)$ be all the distinct eigenvalues of A ordered with $s_1 > s_2 > \cdots > s_m$. Let $b_i (i = 1,2,\cdots,m)$ be corresponding multiplicities of s_i $(i=1,2,\cdots, m)$. We denote the algebraic sum of the multiplicities of $s_1, s_2, \cdots, s_i$ by $c_i(1 \leq i \leq m)$, i.e.

$$c_i = \sum_{j=1}^{i} b_j \qquad \forall i \in [1, m]$$

Obviously, we have $c_m = n$. To ease the following discussion, we introduce also $c_0 = 0$. It is easy to see that

$$l_i \equiv s_r \qquad \forall i \in [c_{r-1} + 1, c_r] \qquad r \in [1, m]$$

and all $S_i(c_{r-1} + 1 \leq i \leq c_r)$ belong to the same eigensubspace is corresponding to the eigenvalues s_r. We denote the eigensubspace of an eigenvalue s by V_s.

For any $X \in R^n$, since $S_i(i=1,2,\cdots, n)$ forms a basis of R^n, X can be uniquely expressed as $X = \sum_{i=1}^{n} z_i S_i$, where $z_i \in R(i = 1,2,\cdots,n)$.

The dimension of the eigensubspace V_{s_1} is corresponding to the largest eigenvalue s_1 is c_1. The problem of eigensubspace estimation in this paper is to find c_1 orthogonal vectors $W_i(i=1,2,\cdots,c_1)$ in R^n such that each of these vectors is approximately equal to an eigenvector corresponding to the largest eigenvalue s_1. Then we have

$$span\{S_1, S_2, \cdots, S_{c_1}\} \approx span\{W_1, W_2, \cdots, W_{c_1}\}$$

We denote

$$\overline{V}_{s_1} = span\{W_1, W_2, \cdots, W_{c_1}\}$$

To solve this problem of eigensubspace estimation, first we estimate the largest eigenvalue and one of its corresponding eigenvectors. Using this approximate largest eigenvalue we then estimate the eigensubspace corresponding to the largest eigenvalue.

2 Largest Eigenvalue and Eigenvectors Estimation

We now propose a discrete time recurrent nonlinear neural network model to estimate the largest eigenvalue and one of its corresponding eigenvectors.

The proposed neural network consists of a single layer of n neurons interconnected with each other. The dynamics of this network can be described as follows:

$$X(k+1) = \frac{AX(k)}{\sqrt{X^T(k)X(k)}} \tag{1}$$

Where A is a real positive definite symmetric matrix of dimensions $n\times n$, $X(k)$ is a column vector of length n with $X^T(k)$ being its transpose.

In this model, synchronous update of neurons is assumed. Each equilibrium point of the proposed network, if it is non-zero, is an eigenvector of the matrix A.

Since A is positive definite, all its eigenvalues are positive. It is easy to verify that if $X(0)\neq 0$, then $X(k)\neq 0$ for all $k\geq 0$. Moreover, if $X(0)\neq 0$, then we have

$$\sigma_m \leq X^T(k+1)X(k+1) = \frac{X^T(k)AX(k)}{X^T(k)X(k)} \leq \sigma_1 \qquad k\geq 0$$

It shows that any solution of the network starting from a non-zero point in R^n will stay between two concentric hyper-spheres with radiuses σ_1 and σ_m after the first iteration.

Since the matrix A is symmetrical, A is associated with an orthogonal basis $S_i(i=1,2,\cdots,n)$ in R^n and each element of the basis is an eigenvector of A. Using this property, we can solve the nonlinear system of Eq.(1) to get a representation of the solutions.

Theorem 1 For any non-zero vector $X(0)\in R^n$, if $X(0)=\sum_{i=1}^{n} z_i(0)S_i$, the solution of Eq.(1) starting from $X(0)$ is represented as

$$X(k) = \sum_{i=1}^{n} \frac{\lambda_i^k z_i(0)}{\sqrt{\sum_{i=1}^{n} \lambda_j^{2(k-1)} z_j^2(0)}} S_i \qquad k\geq 0$$

Proof For any $k\geq 0$, let $X(k)$ be the solution of Eq.(1) starting from $X(0)$. Since $S_i(i=1,2,\cdots,n)$ is an orthogonal basis of R^n, $X(k)$ can be represented as

$$X(k) = \sum_{j=1}^{n} z_i(k)S_i \tag{2}$$

where $z_i(k)(i=1,2,\cdots,n)$ are some real discrete functions. Then, it follows from Eq.(1) that

$$z_i(k+1) = \frac{1}{\sqrt{\sum_{j=1}^{n} z_j^2(k)}} \lambda_i z_i(k) \qquad i=1,2,\cdots,n \tag{3}$$

Since $l_i > 0 (i=1,2,\cdots,n), z_i(k)(i=1,2,\cdots,n)$ have the same sign as $z_i(0)$ for all $k\geq 0$. It is easy to see that $z_i(k)=0$ for all $k\geq 0$ if and only if $z_i(0)=0$. Let $I=\{i \mid z_i(0)\neq 0, (i=1,2,\cdots,n)\}$. Then we have

$$\frac{z_j(k+1)}{z_i(k+1)}=\left(\frac{l_j}{l_i}\right)\frac{z_j(k)}{z_i(k)} \qquad i\in I;\ j=1,2,\cdots,n;\ k\geq 0$$

Solving this equation we get

$$\frac{z_j(k)}{z_i(k)}=\left(\frac{l_j}{l_i}\right)^k\frac{z_j(0)}{z_i(0)} \qquad i\in I;\ j=1,2,\cdots,n;\ k\geq 0 \tag{4}$$

For each $i\in I$, from Eq.(3) and Eq.(4), it follows that

$$z_i(k+1)=\frac{l_i}{\sqrt{\sum_{j=1}^{n}\left[\frac{z_j(k)}{z_i(k)}\right]^2}}sign(z_i(k))=$$

$$\frac{l_i sign(z_i(0))}{\sqrt{\sum_{j=1}^{n}\left(\frac{l_j}{l_i}\right)^{2k}\left[\frac{z_j(0)}{z_i(0)}\right]^2}}=$$

$$\frac{l_i^{(k+1)}z_i(0)}{\sqrt{\sum_{j=1}^{n} l_j^{2k} z_j^2(0)}} \qquad k\geq 0$$

That is

$$z_i(k)=\frac{l_i^k z_i(0)}{\sqrt{\sum_{j=1}^{n} l_j^{2(k-1)} z_j^2(0)}} \qquad i\in I,\ k\geq 0 \tag{5}$$

Obviously, for each $i\in\{1,2,\cdots,n\}-I$, Eq.(5) still holds. This completes the proof .

Theorem 2 Each solution of Eq.(1) starting from a non-zero point in R^n converges to an eigenvector of A.

Proof Let $X(0)$ be any of a non-zero point in R^n, then $X(0)$ can be represented as $X(0)=\sum_{i=1}^{n} z_i(0)S_i$. Let $l=\min\{l \mid z_l(0)\neq 0, 1\leq l\leq n\}$ then there exists a $r\in\{1,2,\cdots,m\}$ such that $c_{r-1}\leq l\leq c_r$. From Eq.(1) we have $X(k)=\sum_{i=1}^{n} z_i(k)S_i$. It is easy to see that if $0\leq i<l$, $z_i(k)=0$ for all $k\geq 0$. Suppose that $I\geq l$, then we have

$$z_i(k)=\frac{l_i z_i(0)}{\sqrt{\sum_{j=l}^{c_r} z_j^2(0)+\sum_{j=c_r+1}^{n}\left(\frac{l_j}{l_i}\right)^{2(k-1)} z_j^2(0)}}$$

If $l\leq i\leq c_r$, then $l_i > l_j$ for all $c_r<j\leq n$ and it follows that

$$z_i(k)\to\frac{l_i z_i(0)}{\sqrt{\sum_{j=l}^{c_r} z_j^2(0)}}=\frac{l_i z_i(0)}{\sqrt{\sum_{j=c_{r-1}+1}^{c_r} z_j^2(0)}} \qquad k\to+\infty$$

If $i\geq c_r+1$, obviously $z_i(k)\to 0$ as $k\to 0$. Then we have

$$X(k) \to \sum_{i=c_{r-1}+1}^{c_r} \frac{l_i z_i(0)}{\sqrt{\sum_{j=c_{r-1}+1}^{c_r} z_j^2(0)}} \qquad S_i \in V_{s_r}, k \to +\infty$$

This completes the proof.

Theorem 3　The solution of Eq.(1) starting from $X(0)$ converges to an eigenvector corresponding to the largest eigenvalue s_1 of A if and only if $X(0)$ is not orthogonal to the eigensubspace V_{s_1}.

Proof　Let $X(0)=\sum_{i=1}^{n} z_i(0)S_i$. Since $S_i(i=1,2,\cdots,c_1)$ consists of a basis of the eigensubspace V_{s_1}, $X(0)$ is not orthogonal to the eigensubspace if and only if there exists an $i(1 \le i \le c_1)$ such that $z_i(0) \ne 0$. The result now follows from the proof of Theorem 2 and the proof is completed.

3　Eigensubspace Estimation

Consider the following simple linear discrete recurrent neural network model:

$$X(k+1)=\frac{1}{s_1} AX(k) \tag{6}$$

Where s_1 is the largest eigenvalue of A. This network consists of a single layer of n neurons interconnecting with each other. Synchronous update of neurons is assumed. It is easy to see that if $X(0) \ne 0$, then $X(k) \ne 0$ for all $k \ge 0$.

Theorem 4　The set of equilibrium points of Eq.(6) is equal to the eigensubspace V_{s_1}.

Proof　A vector x is an equilibrium point of Eq.(6) and only if $x=(1/s_1)Ax$. That is $Ax = s_1 x$, and so x is a equilibrium if and only if $x \in V_{s_1}$. The proof is completed.

Theorem 5　For any $X(0) \in R^n$, if $X(0)=\sum_{i=1}^{n} z_i(0)S_i$, the solution of Eq.(6) starting from $X(0)$ can be represented as

$$X(k) = \sum_{i=1}^{c_1} z_i(0)S_i + \sum_{i=c_1+1}^{n} \left(\frac{l_i}{s_1}\right)^k z_i(0)S_i \qquad k \ge 0$$

$$X(k) \to \sum_{i=1}^{c_1} z_i(0)S_i \in V_{s_1} \qquad k \to +\infty$$

Proof　Suppose that $X(k)=\sum_{i=1}^{n} z_i(k)S_i$ then it follows from Eq.(6) that

$$z_i(k+1) = \begin{cases} z_i(k) & 1 \le i \le c_1 \\ \left(\dfrac{l_i}{s_1}\right)^{k+1} z_i(k) & \text{otherwise} \end{cases}$$

and so

$$z_i(k+1) = \begin{cases} z_i(0) & 1 \le i \le c_1 \\ \left(\dfrac{l_i}{s_1}\right)^{k+1} z_i(0) & \text{otherwise} \end{cases}$$

Then, we have

$$X(k) = \sum_{i=1}^{c_1} z_i(0)S_i + \sum_{i=c_1+1}^{n} \left(\frac{l_i}{s_1}\right)^{k+1} z_i(0)S_i \qquad k \ge 0$$

Since $s_1 > l_i\,(c_1+1 \le i \le n)$, it is easy to see that $(l_i/s_1)^{k+1}\ z_i(0) \to 0$ as $k \to +\infty$. Then $X(k) \to \sum_{i=1}^{c_1} z_i(0)S_i \in V_{s_1}$ as $k \to +\infty$. The proof is completed.

Theorem 6 A solution $X(k)$ of Eq.(6) starting from $X(0)$ converges to zero if and only if $X(0)$ is orthogonal to V_{s_1}.

Proof By Theorem 5 we have $\lim_{k\to+\infty} X(k) = \sum_{i=1}^{c_1} z_i(0)S_i$. Then, $X(k)$ converges to zero if and only if $z_i(0) = 0 (i = 1, 2, \cdots, c_1)$. This is equivalent to $X(0)$ is orthogonal to V_{s_1}. This completes the proof.

Theorem 7 Let $W \in V_{s_1}$ and $X(0) \perp W$, if the solution of Eq.(6) starting from $X(0)$ does not converge to zero, it must converge to an eigenvector in V_{s_1} and this eigenvector is still orthogonal to W.

Proof Suppose $X(0)=\sum_{i=1}^{n} z_i(0)S_i$, it follows from Theorem 5 that the solution starting from $X(0)$ converges to $\sum_{i=1}^{c_1} z_i(0)S_i$. Obviously, if it is non-zero, it must be an eigenvector in V_{s_1}. Suppose $W=\sum_{i=1}^{c_1} w_i S_i$, then by $X(0) \perp W$ we get $\sum_{i=1}^{c_1} z_i(0)w_i = 0$. This shows that $W \perp \sum_{i=1}^{c_1} z_i(0)S_i$. The proof is completed.

In above, we have got an approximate largest eigenvalue $\bar{s}_1$ but it is not really the largest eigenvalue s_1. Though $\bar{s}_1$ can be close to s_1 in arbitrary precision if the error parameter goes to zero, we still need to know what will happen if we replace s_1 by $\bar{s}_1$ in the network described by Eq.(6). To achieve this aim, let us consider the linear neural network

$$\bar{X}(k+1) = \frac{1}{\bar{s}_1} A\bar{X}(k) \tag{7}$$

Theorem 8 If $|s_1 - \bar{s}_1| < s_1 - s_2$, for any non-zero vector $X(0) \in R^n$, the solution $X(k)$ of Eq.(6) and the solution $\bar{X}(k)$ of Eq.(7) starting from the same initial value $X(0)$ satisfy

$$\lim_{k\to+\infty} \left\| N(X(k)) - N(\bar{X}(k)) \right\| = 0$$

where $N(\cdot)$ is the normalization of $\cdot$.

Proof By Theorem 5, we have $X(k) \to \sum_{i=1}^{c_1} z_i(0)S_i$ as $k \to +\infty$, and so

$$N(X(k)) = \frac{X(k)}{\sqrt{X^T(k)X(k)}} \to \frac{1}{\sqrt{\sum_{j=1}^{c_1} z_j^2(0)}} \sum_{i=1}^{c_1} z_i(0)S_i \qquad k \to +\infty$$

Suppose $\bar{X}(k) = \sum_{i=1}^{n} \bar{z}_i(k)S_i$ is the solution of Eq.(7) starting from $X(0)$, then it follows from Eq.(7) that

$$\bar{z}_i(k+1) = \frac{l_i}{\bar{s}_1} \bar{z}_i(k)$$

Note that $\bar{z}_i(0) = z_i(0) (i = 1, 2, \cdots, n)$, so

$$\bar{z}_i(k+1) = \left(\frac{l_i}{\bar{s}_1}\right)^{k+1} z_i(0) \qquad k \ge 0$$

and thus

$$\overline{X}(k)=\sum_{i=1}^{n}\left(\frac{l_i}{\overline{s}_1}\right)^k z_i(0)S_i \qquad k\geq 0$$

Since $\left|s_1-\overline{s}_1\right|<s_1-s_2$, we have $\overline{s}_1>l_i\,(c_1+1\leq i\leq n)$. Then

$$N(\overline{X}(k))=\frac{\overline{X}(k)}{\sqrt{\overline{X}^T(k)\overline{X}(k)}}=\frac{\left[\sum_{i=1}^{c_1}\left(\frac{s_1}{\overline{s}_1}\right)^k z_i(0)S_i+\sum_{i=c_1+1}^{n}\left(\frac{l_i}{\overline{s}_1}\right)^k z_i(0)S_i\right]\sum_{i=1}^{c_1} z_i(0)S_i}{\sqrt{\sum_{j=1}^{c_1}\left(\frac{s_1}{\overline{s}_1}\right)^{2k} z_j^2(0)+\sum_{j=c_1+1}^{n}\left(\frac{l_j}{\overline{s}_1}\right)^{2k} z_i^2(0)\sum_{j=1}^{c_1} z_j^2(0)}}$$

Hence, $\lim_{k\to+\infty}\left\|N(X(k))-N(\overline{X}(k))\right\|=0$. This completes the proof.

The normalization in Theorem 8 is important. Since solutions of Eq.(7) may not converge, we use the normalization of the solutions instead. This theorem allows us to replace s_1 by $\overline{s}_1$ in Eq.(6) during computation.

4 Conclusion

In this paper, we proposed some discrete recurrent neural network models to study the problem of eigensubspace estimation for real positive definite symmetric matrix. Dynamic properties of these networks were studied in detail.

References

1 Diamantaras K I, Hornik K, Strintzis M G. Optimal linear compression under unreliable representation and robust PCA neural models. IEEE Trans.: Neural Networks, 1999, 10(5): 1 186-1 195

2 Luo F, Unbehauen R, Cichocki A. A minor component analysis algorithm. Neural Networks, 1997, 10(2): 291-297

3 Luo F, Unbehauen R. A minor subspace analysis algorithm. IEEE Trans.: Neural Networks, 1996, 8(5): 1 149-1 153

4 Luo F, Unbehauen R, Li Y D. A principal component analysis algorithm with invariant norm. Neurocomputing, 1995, 8: 213-221

5 Mathew G, Reddy V U. Development and analysis of a neural network approach to Pisarenko's harmonic retrieval method. IEEE Trans.: Signal Processing, 1994, 42(3): 663-667

6 Mathew G, Reddy V U. Orthogonal eigensubspace estimation using neural networks. IEEE Trans.: Signal Processing, 1994, 42(7):1 803-1 811

7 Mathew G, Reddy V U, Dasgupta S. Adaptive estimation of eigensubspace. IEEE Trans.: Signal Processing, 1995, 43(2): 401-411

8 Oja E, Karhunen J. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. J. of Math. Anal. Appl., 1985, 106: 69-84

9 Oja E. Principal components, minor components, and linear neural networks. Neural Networks, 1992,5: 927-935

10 Reif K, Lou F, Unbehauen R. The exponential stability of the invariant norm PCA algorithm.: IEEE Trans.: Circuits and Sys.-II, 1997, 44(10): 873-876

11 Taleb A, Cirrincione G. Against the convergence of the minor component analysis neurons. IEEE Trans.: Neural Networks, 1999, 10(1): 207-210

12 Xu L, Oja E, Suen C. Modified hebbian learning for curve and surface fitting. Neural Networks, 1992,5: 441-457

13 Yang J F, Kaveh M. Adaptive eigensubspace algorithms for direction or frequency estimation and tracking. IEEE Trans.: Acoust., Speech, Signal Processing, 1988, 36(2): 663-667