

주요성분분석에 의한 일반회귀 신경망의 성능개선

조 용 현[†]

요 약

본 논문에서는 독립변수들의 특징을 추출하여 패턴층 뉴런의 중앙값으로 이용함으로써 일반회귀 신경망의 성능을 개선하는 방법을 제안하였다. 제안된 방법에서는 적응적 학습 알고리즘의 주요성분분석 기법을 이용하여 회귀분석에 이용되는 데이터의 각 독립변수들의 특징을 추출하였다. 이는 주요성분분석 기법이 가지는 대용량의 입력, 데이터를 통계적으로 독립인 특징들의 집합으로 변환시키는 특징을 살려 일반회귀 신경망의 성능을 더 개선하기 위함이다. 제안된 기법의 일반회귀 신경망을 2개의 독립변수 집합을 가진 Solow의 경제문제와 4개의 독립변수 집합을 가진 국내 유선전화문제에 각각 적용하여 시뮬레이션한 결과, 중심값을 각 독립변수들의 평균이나 가중평균을 이용하는 일반회귀 신경망에 의한 결과와 비교할 때 더욱 우수한 회귀성능이 있음을 확인할 수 있었다. 그리고 신경망의 뉴런 수나 평활요소의 설정 면에서도 우수한 특성이 있음을 확인할 수 있었다.

Performance Improvement of General Regression Neural Network Using Principal Component Analysis

Yong-Hyun Cho[†]

ABSTRACT

This paper proposes an efficient method for improving the performance of a general regression neural network by using the feature to the independent variables as the center for pattern-layer neurons. The adaptive principal component analysis is applied for extracting efficiently the features by reducing the dimension of the given independent variables. It can achieve a superior property of the principal component analysis that converts input data into set of statistically independent features and the general regression neural network, respectively. The proposed general regression neural network has been applied to regress the Solow's economy (2-independent variable set) and the wire telephone(4-independent variable set). The simulation results show that the proposed neural networks have better performances of the regression for the test data, in comparison with those using the means or the weighted means of independent variables. Also, it is affected less by the number of neurons and the scope of the smoothing factor.

1. 서 론

회귀분석은 하나의 종속변수가 다른 독립변수들에 의해 어떻게 설명 또는 예측되는지를 알아보기 위해 적절한 함수로 표현하여 자료분석을 하는 통계적인 기법이다[1-7]. 이러한 회귀분석은 모수 추정을 포함한

관계의 기술이나 새로운 것에 대한 예측, 그리고 통계적 제어 또는 통제 등에 널리 이용되고 있다.

기존의 수치적인 기법에 기초한 회귀분석은 독립변수의 개수에 따라 단순회귀와 다중회귀로 구분되며, 종속변수와 독립변수들의 관계에 따라 선형회귀와 비선형회귀로 다시 나누어진다[7]. 일반적으로 독립변수로부터 다항식의 집합을 계산하여 이용하는 다항회귀 (polynomial regression)와 다중 독립변수의 선형적인 조

[†] 정 회 원 : 대구가톨릭대학교 컴퓨터정보통신공학부 교수
논문접수 : 2000년 8월 24일, 심사완료 : 2000년 11월 6일

함으로 종속변수를 추정하는 다중선형회귀분석(multiple linear regression)이 널리 이용되고 있다. 이들 기법들에서는 주어진 자료들로부터 독립변수와 종속변수의 상관관계에 대한 사전지식을 통하여 회귀분석 방정식의 모델을 설정하든지 회귀계수들의 값을 통계적으로 결정하였다. 하지만 문제의 복잡성에 따라 이러한 모델의 설정이나 계수들의 결정이 매우 어렵다. 이를 위한 대안으로 많은 연구에서 다중선형회귀분석 모델을 이용한 회귀분석이 이루어 졌다. 그러나 여전히 수치적 방법에 의한 회귀분석에는 계수들을 결정하기 위한 지나친 계산시간이 요구되거나 변수 상호간의 비선형 속성을 효과적으로 반영하기 힘든 제약들이 남아 있다.

이러한 기존 기법들이 가지는 제약들을 해결하기 위해서 입력과 출력간의 사상(mapping)을 가능하게 하며, 대규모 분산병렬처리와 학습, 그리고 일반화 등의 속성을 가지는 신경망이 널리 이용되고 있다[1-5]. 신경망을 이용하는 기법은 대규모의 병렬계산과 분산된 국부적 계산의 속성면에서 통계적인 기법이나 계산 이론적인 방법의 문제들도 해결할 수 있다. 특히 입력층과 출력층 사이에 은닉층을 가지는 다층전향신경망은 입출력 데이터 내에 포함된 어떤 비선형 연속함수도 근사화하거나 재구성할 수 있어 대단히 일반적이면서도 융통성 있는 속성을 가지고 있다. 이러한 신경망에는 역전파(backpropagation : BP) 알고리즘의 다층신경망(multilayer perceptron : MLP)과 일반회귀신경망(general regression neural network : GRNN)이 회귀분석에 주로 이용된다[3, 8].

이들 중 MLP는 충분한 뉴런을 가지고 있을 때 어떤 임의의 함수도 근사화할 수 있다고 알려져 있으며, 지금까지 많은 시뮬레이터가 개발되어 있다. 하지만 유용한 모델을 얻기 위해서 많은 학습데이터와 시험데이터가 요구된다. 특히, 기울기하강(steepest descent)의 속성을 이용하는 역전파 알고리즘은 학습파라미터의 설정에 따라 수렴속도와 견실성 중 하나 이상의 문제를 가지며, 전역최소점으로서의 수렴이 보장되어 있지 않다. 이러한 문제를 해결하기 위한 여러 방법들이 많은 연구자들에 의해서 연구되어 왔다[3, 4].

한편, GRNN은 입출력간 확률밀도함수의 추정에 기반을 둔 memory-based 전향신경망으로 4개의 층 구조를 가지는 우수한 성능의 회귀분석 신경망이다. 특히 2번째 층인 패턴층의 출력은 각 뉴런이 가지는 중앙값과 입력 값의 차의 제곱이나 절대치를 구하여 활

성화 함수를 통해 생성된다. 여기서의 활성화 함수로는 다양한 커널(kernel) 함수들이 이용된다. 이는 GRNN에서 학습데이터의 입력과 출력사이 확률밀도함수는 커널함수의 가중된 합으로 추정되는 것을 의미하며, 그 함수로는 정규화된 가우스(Gaussian) 함수가 가장 널리 이용되고 있다. 이때 커널함수의 폭인 평활요소(smoothing factor)는 회귀성능에 많은 영향을 미치는 중요한 파라미터이다. 하지만 패턴층의 각 뉴런 출력을 결정하는 과정에 이용되는 중앙값도 평활요소와 함께 신경망의 성능을 좌우하는 중요한 파라미터이다. 이는 각 학습 패턴들의 입력 값을 가장 잘 나타내는 군집 값으로 설정되어야 한다. 지금까지 대부분의 연구에서는 이를 위한 값으로 각 입력의 평균값이나 출력과의 가중된 평균값을 주로 이용하고 있다[1-5]. 따라서 입력변수 각각의 군집특성을 좀 더 잘 나타내는 값을 중앙값으로 이용한다면 GRNN의 성능은 더욱 개선될 수 있을 것이다.

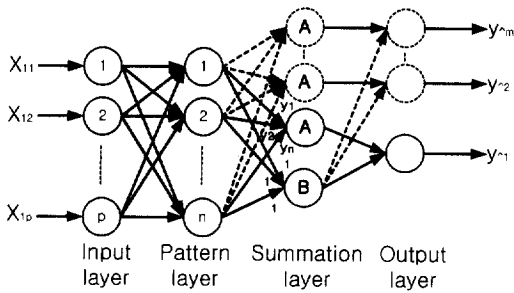
본 연구에서는 수치적인 데이터 집합의 차원을 해석하여 이를 감소시키는 것으로 널리 알려진 적응적 주요성분분석(principal component analysis : PCA) 기법[8-10]을 이용하여 하나의 특징을 추출한 다음, 추출된 특징을 GRNN 각 패턴층 뉴런의 중앙값으로 이용하는 새로운 기법을 제안한다. 결국 이렇게 하면 PCA가 가지는 대용량의 입력데이터를 통계적으로 독립인 특징들의 집합으로 변환시키는 장점과 GRNN이 가지는 우수한 속성을 그대로 살릴 수 있다. 제안된 기법의 GRNN을 41개의 학습패턴(2개 독립변수)을 가진 Solow의 경제문제[6]와 15개의 학습패턴(4개 독립변수)을 가진 국내 유선전화문제[7]에 적용하여 시뮬레이션하고 그 타당성을 확인하였으며, 기존의 평균이나 가중평균을 중앙값으로 이용하는 GRNN과 그 성능을 비교 고찰하였다.

2. 주요성분분석 기법을 이용한 커널 함수의 중심설정

2.1 일반회귀 신경망

일반회귀 신경망은 입출력간의 사전 지식 없이도 어떠한 비선형 관계의 근사화나 재구성과 뉴런간의 상호 연결 및 학습과정 등에서 MLP와 유사한 우수한 성능을 가진 회귀분석 신경망이다[1-5]. 이는 4층의 전향망으로 입력층(input layer), 패턴층(pattern layer), 합층

(summation layer), 그리고 출력층(output layer)으로 구성되며, 층간 뉴런간에는 완전한 연결구조를 이룬다. (그림 1)은 각 학습패턴의 독립변수에 대해 하나의 종속변수만을 가지는 GRNN의 구성도를 나타낸 것으로 종속변수의 수가 증가되면 합층의 A 뉴런과 출력층 뉴런만을 동일한 수로 증가된다. 또한 그림에서 점선 부분은 종속변수가 하나 이상일 경우의 구조를 나타낸 것이다.



(그림 1) 일반회귀 신경망의 구성도

그림에서 입력층의 뉴런 수는 각 학습패턴의 독립변수의 개수와 동일하며, 패턴층의 뉴런 수는 종속변수의 군집수와 일치하게 설정하나 일반적으로 학습패턴의 개수와 동일하게 한다. 또한 합층과 출력층의 뉴런 수는 학습패턴에서 종속변수의 개수에 따라 변하지만 1개의 종속변수일 경우는 각각 2개와 1개이다. 입력층은 학습패턴들을 단순히 패턴층으로 분배하는 층이며, 패턴층과의 연결가중치는 입력되는 학습패턴의 독립변수 값이다. 기존의 MLP에서는 연결가중치가 학습 중에 계속해서 경진되나 GRNN에서는 학습에 따라 오직 한 번씩만 독립변수로 설정된다. 특히 패턴층 뉴런의 데이터 처리과정은 MLP에서 가중된 합을 구해 비선형의 s-자형(sigmoidal) 함수 등의 활성화 함수를 거쳐 출력을 생성하는 것과는 달리, 각 뉴런이 가지는 중앙값과 입력패턴과의 차의 제곱이나 절대치를 구해 커널함수를 거쳐 출력을 생성한다. p개의 독립변수와 1개의 종속변수를 가진 n개의 학습패턴 x 중에서 i번째 학습패턴 x_i 에 대한 제곱이나 절대치를 이용한 거리함수 $D(x, x_i)$ 는 다음과 같이 각각 구할 수 있다. 즉,

$$D(X, X_i) = \sum_{j=1}^p \left(\frac{X_j - X_{ij}}{\sigma_j} \right)^2$$

$$D(X, X_i) = \sum_{j=1}^p \left| \frac{X_j - X_{ij}}{\sigma_j} \right| \quad (i=1, 2, \dots, n) \quad (1)$$

이다. 식에서 X_j 는 j번째 독립변수의 중앙값이며, X_{ij} 는 i번째 학습패턴의 j번째 독립변수 값이고, σ_j 는 j번째 독립변수에 대한 평활요소이다. σ_j 는 커널함수의 폭을 결정하는 요소로 GRNN의 성능을 결정하는 유일한 설정변수로 알려져 있으며, 대부분의 연구에서는 모든 독립변수에 대해 동일한 값 σ 로 가정하고 실험을 통해 결정한다. 일반적으로 σ 가 큰 값일수록 회귀곡선은 더 평활해지며 알려진 패턴들 사이의 일반화 특성은 더 커지게 된다. 또한 작은 값일수록 회귀곡선은 더 정확하게 나타나며 일반화 특성은 더 작아지게 된다. 식 (1)의 계산에서 어느 거리함수의 이용도 가능하나 본 연구에서는 제곱을 이용한 거리함수를 기준으로 설명한다. 또한 계산된 거리값을 이용하여 정규화된 가우스 함수의 커널함수를 거치면 다음과 같은 패턴 뉴런의 출력을 얻을 수 있다. 즉, i번째 패턴 뉴런의 출력 $f(D_i)$ 는

$$f(D_i) = \exp(-D_i) \quad (2)$$

와 같다. 여기서 D_i 는 입력층 i번째 학습패턴으로부터 계산된 거리이다.

한편, 2개의 뉴런으로 구성되는 합층에서 A는 패턴층 뉴런과의 연결가중치로 학습패턴에서 각 독립변수들에 대한 종속변수 값 y 를 가지는 뉴런이며, B는 각 연결가중치가 1로 설정된 뉴런이다. 여기서 A 뉴런의 출력 $h(x)$ 는 패턴층 뉴런의 출력과 종속변수의 가중된 합으로 계산되며, B 뉴런의 출력 $s(x)$ 는 단순한 패턴층 뉴런 출력의 합으로 계산된다. 즉,

$$\begin{aligned} h(x) &= \sum_{i=1}^n y_i \exp(-D(x, x_i)) \\ s(x) &= \sum_{i=1}^n \exp(-D(x, x_i)) \end{aligned} \quad (3)$$

이다. 또한 출력층 뉴런의 출력 $y^*(x)$ 는 합층의 A 뉴런 출력값을 B 뉴런 출력값으로 단순히 나눈 값으로 계산된다. 즉,

$$y^*(x) = \frac{h(x)}{s(x)} \quad (4)$$

이다. 이런 구조의 GRNN은 빠른 학습시간과 더불어

선형 및 비선형의 데이터 관계를 모델링할 수 있으며, 과학습(overtraining)을 줄이기 위해서는 오직 하나의 파라미터인 평활요소만을 잘 조정하면 되는 장점을 가진다. 하지만 식 (1)에서 보면 패턴층의 각 뉴런 출력을 결정하는 과정에 이용되는 중앙값도 신경망의 성능을 좌우하는 하나의 중요한 파라미터임을 알 수 있다. 이는 각 학습패턴들의 입력값을 패턴 뉴런에 어느 정도 반영할 것인가를 결정짓는 것으로 결국 독립변수 각각을 가장 잘 나타내는 군집의 중앙값으로 설정되어야 할 것이다. 지금까지 대부분의 연구에서는 이를 위한 값으로 각 독립변수의 평균이나 종속변수와의 가중된 평균을 중앙값으로 주로 이용하고 있다. 따라서 입력변수 각각의 특징을 가장 잘 나타내는 값을 중앙값으로 이용한다면 GRNN이 가지는 우수한 특성을 더욱더 잘 살릴 수 있다. 이를 위한 대안으로 본 연구에서는 수치적인 데이터 집합의 차원을 해석하여 이를 감소시키는 것으로 널리 알려진 주요성분분석기법의 이용을 제안한다

2.2 적응적 특징추출을 위한 PCA

PCA는 n 차원 입력공간의 데이터를 m 차원 출력공간의 데이터로 투영시키는 것이다. 여기서 m < n이며, 이는 입력데이터 벡터의 대부분 내부정보를 유지하도록 차원의 감소를 얻는 것이다. 자기상관행렬(auto-correlation matrix) $R_{xx} = \langle \mathbf{x}\mathbf{x}^T \rangle$ 를 가진 평균이 영인 입력벡터 $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ 에 대해서 생각해 보자. 여기서 $\langle \mathbf{x} \rangle$ 는 기대치를 나타낸다. 또한 $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ 이 R_{xx} 의 고유벡터(eigenvector)와 직교되는 연결가중치 벡터라 할 때, $\mathbf{w}_1 = [w_{11}, w_{12}, \dots, w_{1n}]^T$ 는 가장 큰 고유치(eigenvalue) λ_1 과 일치하며, $\mathbf{w}_2 = [w_{21}, w_{22}, \dots, w_{2n}]^T$ 는 두 번째로 큰 고유치 λ_2 와 $\mathbf{w}_n = [w_{n1}, w_{n2}, \dots, w_{nn}]^T$ 는 가장 작은 고유치 λ_n 과 일치한다. 이상의 관계를 행렬방정식으로 나타내면 다음과 같다. 즉,

$$R_{xx} \mathbf{w}_j = \lambda_j \mathbf{w}_j \quad (j = 1, 2, \dots, n) \quad (5)$$

이다. 여기서 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ 이다.

주어진 입력벡터 \mathbf{x} 를 위한 첫 번째 m 개의 주요 특징벡터 \mathbf{y} 는 다음의 선형변환식으로 나타낼 수 있다. 즉,

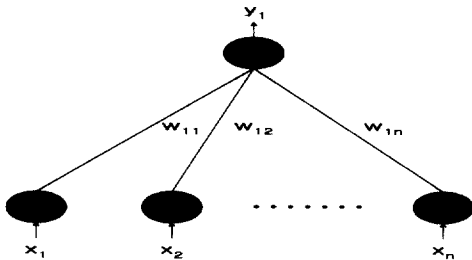
$$\mathbf{y} = \mathbf{W} \mathbf{x} \quad (6)$$

이다. 여기서 $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m]^T \in \mathbb{R}^{m \times n}$ 이며, T는 전치를 나타낸다. 이 식은 행렬 \mathbf{W} 의 행이 가장 큰 고유치와 일치하는 상관행렬 R_{xx} 의 고유벡터임을 의미한다. 다시 말하면, 입력 데이터 공간에서 m 차원의 주요특징을 나타내는 부공간은 R_{xx} 의 m개 주요 고유벡터에 의해 구성된 부공간으로 정의된다. 결국 PCA는 $\langle \|\mathbf{w}_j^T \mathbf{x}\|_2^2 \rangle$ 가 최대인 고유벡터 $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ 의 방향을 찾는 것이다.

한편, 입력되는 데이터의 주요특징들을 추출하기 위한 기존의 방법들을 살펴보면, 수치적 방법에서는 먼저 자기상관행렬 R_{xx} 를 계산한 다음 그것과 일치하는 고유벡터를 계산하였다. 그러나 입력데이터 벡터가 대규모의 차원을 가진다면, 상관행렬은 대단히 크게 되며 또한 요구되는 고유벡터를 찾는 것도 매우 복잡한 일이다. 결국 기존의 수치적 기법들은 대규모의 실시간 처리가 요구되는 응용문제에서는 매우 비효율적이다. 이를 해결하기 위한 대안으로 상관행렬의 고유벡터를 실시간으로 추정하기 위하여 적응학습 알고리즘의 신경망을 이용하는 방법들이 제안되었다[8-10]. 여기서는 입력 데이터 벡터로부터 직접 고유벡터를 추정할 수 있기 때문에 상관행렬의 추정과정이 요구되지 않는다. 이때 이용되는 신경망은 주로 입력층과 출력층으로 구성된 단층구조이다. Oja[8]는 정규화된 헤비안규칙(normalized Hebbian rule)의 적응학습 방법을 이용한 간단한 뉴런모델을 제안하여 정상입력벡터 계열의 주요 특징으로 연결가중치 벡터(weights vector)가 수렴됨을 증명하였다. 그러나 이 방법에서는 한 개의 중요한 주요특징만을 추출할 수 있으며, 임의의 m 개의 특징추출로의 확장은 불가능하다.

GRNN에서는 학습패턴의 독립변수 벡터 각각에 대한 군집의 중앙값이 패턴층 뉴런에서 이용되므로 m개의 특징을 추출할 필요가 없다. 이는 회귀분석을 위한 학습패턴에서 각 독립변수 벡터마다 하나의 특징만을 추출하여 중앙값으로 이용하면 되기 때문이다. 따라서 정규화된 헤비안 규칙의 적응적 PCA를 그대로 이용할 수 있다. (그림 2)는 1개의 주요특징 추출을 위해 n개의 입력뉴런과 1개의 출력뉴런으로 구성된 단층신경망의 구조이다.

그림에서 입력과 출력의 관계를 행렬방정식으로 나타내면 식 (6)과 동일하다. 여기서 각 뉴런의 활성화함수는 선형함수이고, 연결가중치 벡터 $\mathbf{w}_1 = [w_{11}, w_{12}, \dots, w_{1n}]^T$ 이다. 따라서 신경망을 학습시켜 연결가중



(그림 2) 1 개의 주요특징 추출을 위한 단층신경망

치를 구함으로써 주어진 n개의 입력벡터 x로부터 1개의 주요 특징 즉 출력 y₁은 식 (7)과 같이 구할 수 있다.

$$y_1 = \omega_1^T x = \sum_{j=1}^n \omega_{1j} x_j \quad (7)$$

또한 정규화된 헤비안규칙에 기반을 둔 (그림 2)의 단층신경망을 학습시키는 알고리즘은 다음과 같다.

$$w_{1j}(t+1) = w_{1j}(t) + \eta [y_1(t) x_j(t) - w_{1j}(t) y_1^2(t)], \quad (j = 1, 2, \dots, n) \quad (8)$$

여기서 η는 학습율이다. 식 (8)을 이용하여 연결가중치를 구한 다음 입력벡터와 함께 식 (7)에 대입하면 주요특징을 구할 수 있다. 즉, 학습은 상관행렬의 고유벡터를 추정하기 위하여 임의의 초기연결 가중치로부터 주요특징을 내포하는 연결 가중치를 구하는 과정이다. 그러므로 신경망은 입력데이터 내에 포함된 상호 독립인 정보를 부호화시키는 검출기이다.

결국 PCA는 높은 차원의 입력공간을 더 낮은 차원의 표현공간으로 사상시켜 입력데이터가 가지는 두드러진 특징들을 추출하는 기법이다. 이는 입력 데이터 내에 존재하는 주요특징들을 추출함으로써 입력의 개수를 감소시키는데 이용될 수 있다. 따라서 GRNN에 이용되는 학습패턴 데이터의 독립변수 벡터 각각에 적용적 PCA를 이용하여 특징을 구한 다음 이를 패턴층 뉴런의 각 중앙값으로 이용한다. 이렇게 하면 GRNN 패턴층 뉴런의 적당한 중앙값이 설정되어 이 신경망이 가지는 우수한 속성을 그대로 살릴 수 있을 것이다.

3. 시뮬레이션 결과 및 분석

제안된 기법의 일반회귀 신경망의 성능을 평가하기

위해서 특징추출을 위한 단층신경망과 회귀분석을 위한 4층 전향신경망을 각각 구성하였다. PCA를 위한 단층신경망의 입력뉴런과 출력 뉴런사이의 초기 연결가중치는 각각 랜덤시드(random seed)를 이용하여 -1에서 +1 사이의 임의의 값으로 설정하였다. 학습은 전체 반복회수가 10,000이상이거나 전체 오차값이 허용치 10⁻⁴이하일 때 종료되도록 하였다. 실험은 팬티엄III-700 시스템 환경하에서 비주얼 c++ 언어로 구현하였다.

제안된 PCA 기법을 이용한 GRNN을 41개의 학습패턴(2개 독립변수)을 가진 Solow의 경제문제와 15개의 학습패턴(4개 독립변수)을 가진 국내 유선전화문제에 적용하여 시뮬레이션하고 그 타당성을 확인하였으며, 기존의 평균이나 가중평균을 중앙값으로 이용하는 GRNN과 그 성능을 비교 분석하였다.

3.1 Solow의 경제문제

실험에 이용한 Solow의 경제문제는 1909년부터 1949년까지 미국경제에 있어서 노동시간당 총괄생산량 q(t), 총괄적인 자본과 노동의 비율 k(t), 그리고 기술지표 A(t)의 자료를 수집하고, 이를 통하여 자료상호 간의 관계에 대한 여러 모형을 제시한 문제이다. 이 문제는 2개의 독립변수 q(t)와 k(t) 및 1개의 종속변수 A(t)로 구성된 41개의 패턴들을 가진 문제이다. 여기서는 41개의 패턴 중에서 31개는 학습패턴으로 나머지 10개는 시험패턴으로 이용하였다. GRNN은 2개의 입력층 뉴런과 31개의 패턴층 뉴런으로 구성된 구조를 이용하였다. 다음의 <표 1>은 경제자료 일부와 중앙값으로 이용될 2개의 독립변수 각각에 대한 평균값, 가중평균값, 그리고 제안된 기법으로 구해진 값을 각각 나타낸 것이다.

<표 2>는 GRNN에서 평균값, 가중평균값, 그리고 제안된 기법에 의한 값을 각각 중앙값으로 이용할 때, 10개의 시험패턴과 31개의 학습패턴을 포함한 41개의 패턴에 대해 평활요소 σ의 변화에 따른 원하는 종속변수 값 A(t)와 학습 후 출력 Â(t)와의 절대오차(absolute error : AE, AE = ∑_{i=1}ⁿ | A_i(t) - Â_i(t) |)를 나타낸 것이다. 표에서 각 중앙값에 따른 평활요소 σ의 변화범위를 보면 전체적으로 0.5에서 0.08 사이의 절대오차 값이 작음을 알 수 있다. 이는 σ가 큰 값일수록 얻어지는 회귀곡선은 더 평활 해지며, 알려진 패턴들 사이의 일반화 특성은 더 커지게 되어 절대오차

<표 1> Solow의 경제자료 및 2개 독립변수 각각의 구해진 중앙값

연도	총괄생산량 q(t)	자본과 노동비율 k(t)	기술지표 A(t)
1909	0.623	2.06	1.000
1910	0.616	2.10	0.983
1911	0.647	2.17	1.021
⋮	⋮	⋮	⋮
1938	1.000	2.78	1.445
1939	1.034	2.66	1.514
1940	1.082	2.63	1.590
⋮	⋮	⋮	⋮
1949	1.275	2.70	1.809
평균값	$\overline{q(t)} = 0.81070968$	$\overline{k(t)} = 2.64129$	
가중평균값	<i>weighted</i> $\overline{q(t)} = 1.21090844$	<i>weighted</i> $\overline{k(t)} = 1.20305$	
제안된 기법값	0.828284	2.693789	

$$\overline{q(t)} = (\sum_{t=1}^{31} q(t))/31, \overline{k(t)} = (\sum_{t=1}^{31} k(t))/31, \textit{weighted } \overline{q(t)} = (\sum_{t=1}^{31} q(t) \cdot A(t)) / \sum_{t=1}^{31} q(t), \textit{weighted } \overline{k(t)} = (\sum_{t=1}^{31} k(t) \cdot A(t)) / \sum_{t=1}^{31} k(t)$$

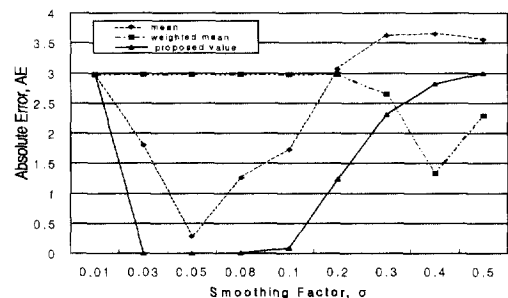
<표 2> 평활요소 σ의 변화에 따른 절대오차 AE의 변화

중앙값 \ σ	100	10	5	1	0.5	0.3	0.1	0.08	0.01
평균값	2.977115	2.780747	2.991505	3.219541	3.557127	3.630611	1.724396	1.258476	2.977061
가중 평균값	2.977203	2.989188	3.024665	3.541450	2.293280	2.650761	2.977061	2.977061	2.977061
제안된 기법값	2.977096	2.979368	2.985987	3.076269	2.995272	2.309283	0.086442	0.011441	2.977061

가 커짐을 확인할 수 있다. 또한 σ가 0.01 이하에서는 절대오차가 증가되어 회귀곡선은 더 상세하게 나타나나 잡음에 의해서 데이터가 오염되어 원하는 출력이 아님을 추측할 수 있다. 한편, 3가지 방법 중에서 제안된 기법인 PCA에 의한 중앙값이 전체적으로 가장 작은 오차를 가지며, 특히 σ가 0.3에서 0.08의 경우 절대오차가 상대적으로 매우 적어 학습패턴의 회귀성능이 우수함을 알 수 있다. 이는 패턴뉴런의 중앙값도 평활요소와 함께 회귀성능에 영향을 미치는 요소임을 보여주는 것이다.

(그림 3)은 평활요소 σ를 0.5에서 0.01까지 변화시키면서 절대오차의 변화를 나타낸 결과이다. 이는 주어진 문제를 위한 좀 더 정확한 σ의 값을 추정하기 위함이다. 그림에서 보면, 제안된 기법에 의해 구해진 중앙값이 평균값이나 가중평균값을 이용할 경우보다 절대오차가 작아 우수한 회귀성능을 가진다. 특히, 제안된 기법에서 σ가 0.03에서 0.1까지 범위에서 절대오차의 값은 거의 0으로 매우 양호한 회귀성능을 가짐을 알 수 있다. 그러나 평균값을 이용할 경우는 0.05, 가중평균을 이용할 경우에는 0.4일 때 다른 값보다 우수한 성능을 가진다. 또한 그림에서 제안된 기법이 다른

2개의 방법보다 넓은 범위에서 작은 절대오차 값을 가져 평활요소의 설정도 용이함을 확인할 수 있다.



(그림 3) 평활요소 σ의 변화에 따른 절대오차 AE의 변화

한편, <표 3>은 41개의 패턴에 대해서 평활요소 σ 값을 0.08로 고정하고 패턴층 뉴런의 개수에 따른 절대오차를 나타낸 것이다. 표에서 보면, 중앙값에 무관하게 패턴층 뉴런의 개수가 줄어들면 절대오차는 증가하며, 이에 따른 회귀성능도 나빠짐을 알 수 있다. 이는 패턴층 뉴런의 개수가 GRNN의 성능에 영향을 미치는 파라미터임을 의미하며, 일반적으로 널리 이용되는 뉴런 개수의 설정 방법에서는 학습패턴의 개수, 즉 종속

〈표 3〉 패턴층 뉴런수에 따른 절대오차 AE의 변화

중양값 \ 뉴런수	31	25	20	15	10	5	1
평균값	1.258476	4.663394	4.634554	8.310066	9.485885	11.173926	13.327983
가중 평균값	2.977061	3.337759	4.353666	5.494248	7.172335	9.811333	13.002929
제안된 기법값	0.011441	0.4674	0.464538	8.337306	9.485898	11.173931	13.327989

변수의 개수와 동일하게 한다. 이는 학습패턴 각각을 하나의 군집으로 간주하는 것으로 GRNN의 패턴층 뉴런 각각이 하나의 군집에 대응하도록 한 것이다. 하지만 이 방법은 뉴런 수의 증가에 따른 GRNN의 학습시간 증가를 가져온다. 이를 해결하기 위해서는 학습패턴 상호간의 유사성에 따라 동일한 군집으로 분류하면 궁극적으로 패턴층 뉴런의 수를 감소시킬 수 있다. 표에서도 제안된 기법이 다른 방법에 의한 것보다 뉴런의 수가 20개까지는 상대적으로 적은 절대오차 값을 가져 우수한 성능이 있다. 그러나 15개 이하에서는 가중평균값을 이용하는 것이 다소 우수하지만, 평균값과 가중평균값, 그리고 제안된 기법값 다 많은 오차를 포함하고 있다. 따라서 제안된 기법은 패턴층 뉴런의 설정에 대해서도 우수한 회귀성능이 있다.

3.2 유선전화 문제

유선전화 문제는 4개의 독립변수에 대하여 1개의 종속변수를 가진 회귀분석 문제이다. 본 실험에서는 국내 인구수, 가격지수, GDP, 이동전화 가입자 수에 대

한 유선전화 가입자의 수요예측을 하였다. 여기에서 가격지수는 전화 가입비와 전화사용료를 합한 자료이며, 실험에서는 1984년부터 1998까지의 15개의 자료를 회귀분석 자료로 이용하였다. 이용된 GRNN은 4개의 입력층 뉴런과 15개의 패턴층 뉴런으로 구성된 구조를 이용하였다. 다음의 <표 4>는 유선전화 자료일부와 중양값으로 이용될 4개의 독립변수 각각에 대한 평균값, 가중평균값, 그리고 제안된 기법으로 구해진 값을 각각 나타낸 것이다. 실험에서는 각 변수들의 규모가 달라 자연대수 Ln을 취하여 구한 값을 이용하였다. 표에서 평균값, 가중평균값, 그리고 제안된 기법값은 각각 <표 1>의 실험에서와 동일한 방법으로 얻어진 값들이며, 1999년부터 2003년까지의 독립변수 자료는 예상자료들을 나타낸다. 예상자료 중 인구수, GDP, 그리고 이동전화 가입자 수는 1984년부터 1998년까지의 년 평균 증가에 따른 수치이며, 가격지수는 100으로 하였다. 또한 유선전화 가입자 수는 KISDI의 추정값이다.

<표 5>는 4개의 입력뉴런과 15개의 패턴 뉴런을 가진 GRNN에서 평균값, 가중평균값, 그리고 제안된 기

〈표 4〉 국내 유선전화 자료 및 4개 독립변수 각각의 구해진 중양값

연 도	인구수		가격지수		GDP		이동전화 가입자수		유선전화 가입자수	
	실제값	Ln값	실제값	Ln값	실제값	Ln값	실제값	Ln값	실제값	Ln값
1984	40405956	17.51448776	274	5.613128106	1758157	14.37977666	2658	7.885329239	5594973	15.53737907
1985	40805744	17.52433341	279.6	5.63336001	1943354	14.4799259	4685	8.452121195	6517395	15.68998531
1986	41213674	17.53428065	240.39	5.482262604	2254349	14.6283718	7093	8.866863661	7520699	15.83316964
⋮										
1991	43295704	17.58356397	170.25	5.137267945	4996877	15.42412353	166198	12.02093513	14572585	16.49465258
1992	43747962	17.59395559	170.85	5.140785979	5380594	15.49830933	271868	12.51307193	15593454	16.56236177
1993	44194628	17.6041133	141.3	4.95088529	6270219	15.65132184	471784	13.06427653	16632593	16.62687476
⋮										
1998	46429817	17.65345242	100	4.605170186	9544082	16.07143183	13982919	16.45334707	21373020	16.87763994
평균값		17.58402111		5.101632852		15.33596355		11.93920332		
가중평균값		16.3658627		16.33565419		16.38047021		16.45536784		
제안된 기법값		17.584123		5.129157		15.357646		12.53249		
1999	46831408	17.66206465	100	4.605170186	10063144	16.12439018	14914936	16.5178737	20857000	16.85320018
2000	47232999	17.67060334	100	4.605170186	10582205	16.1746844	15846954	16.57848786	21439000	16.88072225
2001	47634590	17.67906974	100	4.605170186	11101267	16.2225698	16778971	16.63563694	21957000	16.90459655
2002	48036181	17.68746506	100	4.605170186	11620329	16.2682666	17710989	16.68989582	22417000	16.92533016
2003	48437772	17.69579048	100	4.605170186	12139390	16.31196611	18643006	16.74098162	22828000	16.94349841

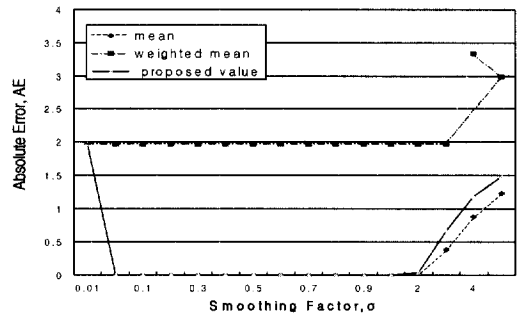
<표 5> 평활요소 σ 의 변화에 따른 절대오차 AE의 변화

중양값 \ σ	100	10	5	1	0.8	0.5	0.1	0.08	0.01
평균값	1.96832	1.833286	1.226199	0.000029	0.000041	0.000176	0.000231	0.000558	1.964898
가중 평균값	1.970087	2.33026	2.990574	1.964898	1.964898	1.964898	1.964898	1.964898	1.964898
제안된 기법값	1.969107	1.916357	1.47755	0.000023	0.000024	0.000144	0.000195	0.000537	1.964898

법에 의한 값을 각각 중양값으로 이용할 때, 5개의 시험패턴과 15개의 학습패턴을 포함한 20개의 패턴에 대해 평활요소 σ 의 변화에 따른 원하는 종속변수 값과 학습 후 출력과의 절대오차를 나타낸 것이다. 제안된 기법에 의해 구해진 값을 패턴층 뉴런의 중양값으로 이용할 때가 다른 두 값을 이용할 때보다 절대오차가 적어 우수한 회귀성능을 가지며, 가중평균을 이용할 때 가장 큰 절대오차값을 가짐을 알 수 있다. 이는 가중평균의 경우 각 독립변수의 중양값 결정에 종속변수의 가중치도 함께 고려한 것으로 독립변수가 가지는 완전한 군집의 특성을 반영하지 못하기 때문이다. 여기서도 <표 2>에서처럼 σ 의 값이 상대적으로 크거나 작은 값일 경우에 절대오차 값은 커져 원데이터에 대한 지나친 평활화나 잡음 등에 따른 영향을 알 수 있다. 또한 3개의 중양값 모두에 대해서 σ 의 값은 1에서 0.08사이일 때가 적당하며, 그 중에서도 제안된 기법의 값을 중양값으로 할 때 가장 작은 절대오차를 보인다. 따라서 Solow의 경제 문제에서처럼 패턴층 뉴런의 중양값도 평활요소와 더불어서 중요한 성능 결정요인임을 확인할 수 있으며, 이들 2개의 요소 상호간에도 의존성이 있음을 확인할 수 있다.

(그림 4)는 평활요소 σ 의 값변화에 따른 절대오차의 변화를 나타낸 것이다. 여기서도 (그림 3)에서처럼 제안된 기법에 의해 구해진 중양값이 다른 값들을 이용할 경우보다 절대오차가 작아 우수한 회귀성능이 있다. 특히, 제안된 기법의 경우 0.08에서 1까지에서 절대오차 값은 거의 0으로 매우 양호한 회귀성능을 가짐을 알 수 있다. 그러나 평균값을 이용할 경우도 거의 제안된 기법을 이용할 경우와 동일한 결과를 보이나 그림의 데이터를 보면 제안된 기법이 약간 우수한 성능이 있음을 확인할 수 있었다. 여기에서도 제안된 기

법은 평활요소의 범위가 넓어 그 설정이 용이함을 알 수 있다.



(그림 4) 평활요소 σ 의 변화에 따른 절대오차 AE의 변화

한편, <표 6>은 5개의 시험패턴을 포함한 20개의 패턴에 대해서 평활요소 σ 를 0.8로 고정하고 패턴층 뉴런의 개수에 따른 절대오차의 변화를 나타낸 것이다. 표에서도 패턴층 뉴런의 개수가 줄어들면 절대오차는 상대적으로 증가함을 알 수 있다. 이는 <표 3>에서처럼 패턴층 뉴런의 개수도 GRNN의 성능에 영향을 미치는 요소임을 알 수 있다. 여기서도 제안된 기법이 다른 방법에 의한 것보다 뉴런의 수가 10개까지는 상대적으로 적은 절대오차 값을 가져 우수한 성능이 있다. 그러나 8개 이하에서는 가중평균값을 이용하는 것이 다소 우수하지만, 평균값과 가중평균값, 그리고 제안된 기법값 다 많은 오차를 포함하고 있다. 따라서 제안된 기법은 패턴층 뉴런의 설정에 대해서도 상대적으로 우수한 성능을 다시 확인할 수 있다.

<표 7>은 패턴층 뉴런을 15개로 하고 중심값을 제안된 기법으로 구한 다음 1999년부터 2003년까지 향후 5년간 유선전화 가입자 수를 예측한 결과이다. 표에서 보

<표 6> 패턴층 뉴런수에 따른 절대오차 AE의 변화

중양값 \ 뉴런수	15	13	10	8	6	5	1
평균값	0.000041	0.000041	0.364187	5.463024	8.788733	12.275886	19.232632
가중 평균값	1.964898	2.520171	3.442823	4.371309	6.15526	9.820766	11.584711
제안된 기법값	0.000024	0.000039	0.357565	7.002859	8.7931	12.275864	19.232581

면 제안된 기법의 예상치와 KISDI에서의 예상치는 유사한 값을 가진다. 그러나 표에서 약간의 차이를 보이는 것은 상호 간의 예측방법의 차이와 예측환경의 차이로 생각된다. 하지만 이들 두 예측치 중에서 어느 예측치가 더 타당한지에 대한 검증은 향후 계속되어야 하나 KISDI에서는 회귀분석용 모형으로 자연대수 모형을 이용한 기존의 수치적 방법이며, 본 연구에서의 GRNN을 이용한 차이로 추측된다.

〈표 7〉 제안된 기법의 GRNN과 KISDI에서의 유선전화 가입자 수 예측치

기법 \ 연도	1999	2000	2001	2002	2003
제안된 기법의 GRNN	20896975	21451895	22252988	22896929	23278354
KISDI	20857000	21439000	21957000	22417000	23828000

4. 결 론

본 연구에서는 적용적 주요성분분석 기법을 이용하여 회귀분석 패턴의 각 독립변수 벡터에 대한 하나의 특징을 추출한 다음, 이를 GRNN 각 패턴층 뉴런의 중앙값으로 이용하는 새로운 기법을 제안하였다. 이는 주요성분분석 기법이 가지는 대용량의 입력데이터를 통계적으로 독립인 특징들의 집합으로 변환시키는 장점을 살려 GRNN의 회귀성능을 증가시키기 위함이다.

제안된 기법의 GRNN을 2개의 독립변수와 1개의 종속변수를 가진 41개의 학습패턴의 Solow의 경제문제와 4개의 독립변수와 1개종속변수를 가진 15개의 학습패턴을 가진 국내 유선전화 가입자문제를 대상으로 시뮬레이션한 결과, 평균이나 가중평균을 이용한 기존의 GRNN에 의한 결과와 각각 비교할 때 절대오차로 표현된 회귀성능에서 더욱 우수한 성능이 있음을 확인하였다. 특히, GRNN의 뉴런 수나 평활요소의 설정 면에서도 우수한 특성이 있음도 확인하였다.

향후 제안된 기법의 GRNN을 좀 다양한 회귀분석 문제에 적용하고 다중공선성과 같은 회귀검증도 함께 고려되는 연구가 계속되어야 할 것이다. 또한 효율적인 응용을 위해서는 제안된 기법의 GRNN에서 다른 성능결정 요소들에 대한 연구도 뒤따라야 할 것이다.

참 고 문 헌

[1] D. F. Specht, "A General Regression Neural Network," IEEE Trans. on Neural Networks, Vol. 2, No.6, pp.568-576, Nov., 1991.

[2] M. Caudill, "GRNN and Bear It," AI Expert, Vol. 5, pp.28-33, May, 1993.

[3] T. Masters, 'Advanced Algorithms for Neural Networks : A C++ Sourcebook', John Wiley & Sons, Inc., New York, 1995.

[4] M. Catusus, W. Branagh, and E. D. Salin, "Improved Calibration for Inductively Coupled Plasma-Atomic Emission Spectrometry Using Generalized Regression Neural Networks," Applied Spectroscopy, Vol.49, No. 6, pp.798-807, 1995.

[5] Y. Chtioui, S. Panigrahi, and L. Franci, "A Generalized Regression Neural Network and Its Application for Leaf Wetness Prediction to Forecast Plant Disease," Elsevier, Chemometrics and Intelligent Laboratory Systems, Vol.48, pp.47-58, 1999.

[6] 강명옥 외 3, "회귀분석 : 모형개발과 진단", 을국출판사, 1997.

[7] 성낙일 외 1, "한국통신의 통신요금지수 개발 및 추세 분석", 한국통신보고서, 1996.

[8] A. Cichock and R. Unbehau, 'Neural Networks for Optimization and Signal Processing', John Wiley & Sons., New York, 1993.

[9] P. Foldiak, "Adaptive Network for Optmal Linear Feature Extraction," International Joint Conference on Neural Networks, Washington D.C., Vol.1, pp. 401-406, June 1989.

[10] K. I. Diamantaras and S. Y. Kung, 'Principal Component Neural Networks : Theory and Applications, Adaptive and Learning Systems for Signal Processing, Communications, and Control', John Wiley & Sons, Inc., 1996.



조 용 현

e-mail : yhcho@cuth.cataegu.ac.kr

1979년 경북대학교 공과대학 전자공학과(공학사)

1981년 경북대학교 대학원 전자공학과(공학석사)

1993년 경북대학교 대학원 전자공학과(공학박사)

1983년~1984년 삼성전자(주) 사원

1984년~1987년 한국전자통신연구소 연구원

1987년~1997년 영남전문대학 전자과 부교수

1997년~현재 대구가톨릭대학교 컴퓨터정보통신공학부 조교수

관심분야 : 신경망, 병렬분산처리, 신호처리, 영상인식, 교환기 등