

독립변수의 차원 감소에 의한 일반회귀 신경망의 성능개선 Performance Improvement of General Regression Neural Network by Reducing Dimensionality of Independent Variables

조 용 현

Yong-Hyun Cho

대구가톨릭대학교 컴퓨터정보통신공학부

요 약

본 논문에서는 독립변수들의 차원을 감소시켜 일반회귀 신경망의 성능을 개선하는 방법을 제안하였다. 제안된 방법에서는 적응적 학습 알고리즘의 주요성분분석 기법을 이용하여 독립변수 패턴의 특징을 추출하고 이를 일반회귀 신경망의 학습데이터로 이용하였다. 이는 주요성분분석 기법이 가지는 대용량의 입력 데이터를 통계적으로 독립인 특징들의 집합으로 변환시키는 속성을 살려 학습데이터의 차원을 감소시킴으로서 고차원의 학습데이터에 따른 일반회귀 신경망이 가지는 제약을 해결하기 위한 것이다. 제안된 기법의 일반회귀 신경망을 3개의 독립변수 패턴을 가진 암모니아 제조공정문제와 10개의 독립변수 패턴을 가진 자동차 연비문제에 각각 적용하여 시뮬레이션한 결과, 기존의 일반회귀 신경망에 의한 결과와 비교할 때 더욱 우수한 학습성능과 회귀성능이 있음을 확인할 수 있었다. 그리고 커널함수의 평활요소 설정 면에서도 우수한 특성이 있음을 확인할 수 있었다.

ABSTRACT

This paper proposes an efficient method for improving the performance of a general regression neural network by reducing the dimensionality of independent variables. The adaptive principal component analysis is applied for reducing efficiently the dimension by extracting the features 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 Ammonia producing process (3-independent variable) and the mobile oil cost (10-independent variable). The simulation results show that the proposed neural networks have better performances of the regression for test data, in comparison with those using the conventional general regression neural network. It is also affected less by the scope of the smoothing factor for kernel function.

1. 서 론

회귀분석은 하나의 종속변수가 다른 독립변수들에 의해 어떻게 설명 또는 예측되는지를 알아보기 위해 적절한 함수로 표현하여 자료분석을 하는 통계적인 기법이다[1-8]. 이러한 회귀분석은 모수 추정을 포함한 관계의 기술이나 새로운 것에 대한 예측, 그리고 통계적 제어 등에 널리 이용되고 있다.

기존의 수치적인 기법에 기초한 회귀분석에는 독립변수의 개수나 종속변수와 독립변수들의 관계에 따라 여러 가지 회귀기법들로 나누어진다[1,2]. 일반적으로 다항회귀(polynomial regression)와 다중선형회귀분석(multiple linear regression)이 많이 이용되고 있다. 이들 기법들에서는 주어진 자료들로부터 독립변수와 종속변수의 상관관계에 대한 사전지식을 통하여 회귀분석 방정식의 모델을 설정하며 회귀계수들의 값은 통계적으로 결정하였다. 하지만 문제에 따라서는 이러한 모델의 설정이나 계수들의 결정이 매우 힘든 제약들

은 여전히 존재한다.

이러한 기존 수치적 기법들이 가지는 제약들을 해결하기 위해서 입력과 출력간의 사상을 가능하게 하며, 대규모 분산병렬처리와 학습, 그리고 일반화 등의 속성을 가지는 신경망이 널리 이용되고 있다[2-8]. 신경망을 이용하는 기법 중에서 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 이용된다[3-6,8].

이들 중 MLP는 충분한 뉴런을 가지고 있을 때 어떤 임의의 함수도 근사화할 수 있다고 알려져 있지만 유용한 모델을 얻기 위해서는 많은 학습데이터와 시험 데이터가 요구된다. 특히, 기울기하강(steepest descent)의 속성을 이용하는 역전파 알고리즘은 학습파라미터

의 설정에 따라 수렴속도와 견실성 중 하나 이상의 제약을 가지며, 전역최소점으로서의 수렴이 보장되어 있지 않다. 이러한 문제를 해결하기 위한 여러 방법들이 연구되어 왔다[3,6].

한편, GRNN은 입력력간 확률밀도함수의 추정에 기반을 둔 memory-based 전향신경망으로 4개의 층 구조를 가지는 우수한 성능의 회귀분석 신경망이다[4-8]. 특히 패턴층의 출력은 각 뉴런이 가지는 중앙값과 입력값의 차의 제곱이나 절대치를 구하여 활성화 함수를 통해 생성된다. 여기서의 활성화 함수로는 다양한 커널(kernel) 함수들이 이용된다. 이는 GRNN에서 학습데이터의 입력과 출력사이 확률밀도함수가 커널 함수의 가중된 합으로 추정되는 것을 의미하며, 그 함수로는 정규화된 가우스(Gaussian) 함수가 가장 널리 이용되고 있다. GRNN은 학습알고리즘이 대단히 간단하고 구현이 용이하며, 학습시간이 요구되지 않는 매우 빠르면서도 고신뢰성을 가지는 회귀분석 신경망이다. 그러나 여기에도 독립변수 패턴의 고차원 문제, 최적의 평활요소(smoothing factor) 및 중앙값 설정, 회귀분석을 위한 데이터 내 변화를 정확하게 측정하기 위한 많은 학습패턴 요구 등의 제약들이 있다. 특히 독립변수 패턴의 고차원은 학습시간의 증가와 함께 회귀분석 성능의 저하를 유발하게 된다[1,2,5,6]. 이를 해결하기 위해 기존의 방법들에서는 분산팽창인자(variance inflation factor), 상태수, 상태지표, 그리고 분산비를 등의 척도에 의한 다중공선성 분석이나 변수변환 등의 방법들이 이용되고 있다[1,2]. 그러나 척도들을 계산하는데는 표본의 평균과 표준편차, 독립변수 행렬의 고유벡터(eigen vector)와 고유값(eigen value) 등의 계산이 요구되며, 그 계산 또한 매우 복잡하여 계산시간이 오래 걸린다. 따라서 독립변수들간의 상호관계를 잘 나타내는 주요 특징들만 추출하여 회귀분석 데이터로 이용한다면 GRNN이 가지는 독립변수 패턴의 고차원에 따른 문제들을 해결할 수 있을 것이다.

본 연구에서는 수치적인 데이터 집합의 차원을 해석하여 이를 감소시키는 것으로 널리 알려진 적응적 주요성분분석(principal component analysis:PCA) 기법[9-11]으로 독립변수 벡터의 특징을 추출한 다음, 추출된 특징을 GRNN 입력층 뉴런의 입력으로 이용하는 새로운 기법을 제안한다. 이렇게 하면 PCA가 가지는 대용량의 입력데이터를 통계적으로 독립인 특징들의 집합으로 변환시키는 장점과 GRNN이 가지는 우수한 속성을 그대로 살릴 수 있다. 제안된 기법의 GRNN을 3개의 독립변수 패턴을 가진 암모니아 제조 공정문제[1]와 10개의 독립변수 패턴을 가진 자동차

연비문제[2]에 각각 적용하여 시뮬레이션하고 그 타당성을 확인하였으며, 기존의 GRNN과 그 성능을 비교 고찰하였다.

2. 주요성분분석 기법을 이용한 독립변수의 차원 감소

2.1 일반회귀 신경망

GRNN은 입력력간의 사전 지식 없이도 어떠한 비선형 관계의 근사화나 제구성고 뉴런간의 상호 연결 및 학습과정 등에서 MLP와 유사한 우수한 성능을 가진 회귀분석 신경망이다[4-8]. 이는 4층의 전향망으로 입력층, 패턴층, 합층, 그리고 출력층으로 구성되며, 층간 뉴런간에는 완전한 연결구조를 이룬다. 그림 1은 각 학습패턴의 독립변수에 대해 하나의 종속변수만을 가지는 GRNN의 구성도를 나타낸 것으로 종속변수의 수가 증가되면 합층의 A 뉴런과 출력층 뉴런만을 동일한 수로 증가된다. 그림에서 점선부분은 종속변수가 하나 이상일 경우의 구조를 나타낸 것이다.

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 경우 각각 2개와 1개의 뉴런으로 구성된다. 입력층은 학습패턴들을 단순히 패턴층으로 분배하는 층이며, 패턴층과의 연결가중치는 입력되는 학습패턴의 독립변수 값이다. 기존의 MLP에서는 연결 가중치가 학습 중에 계속해서 경신되나 GRNN에서는 학습에 따라 오직 한 번씩만 독립변수로 설정된다. 특히 패턴층 뉴런의 데이터 처리과정은 MLP에서 가중된 합을 구해 s-자형(sigmoidal) 함수 등의 비선형 활성화 함수를 거쳐 출력을 생성하는 것과 달리, 각 뉴런이 가지는 중앙값과 입력패턴과의 차의 제곱이나 절대치를 구해 커널함수를 거쳐 출력을

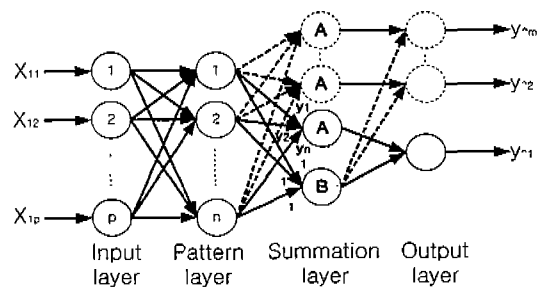


그림 1. 일반회귀 신경망의 구성도

생성한다. p 개의 독립변수와 1개의 종속변수를 가진 n 개의 학습패턴 x 중에서 i 번째 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 j 번째 독립변수의 중심값이며, X_{ij} 는 i 번째 학습패턴의 j 번째 독립변수 값이고, σ_j 는 j 번째 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 결과적인 회귀곡선은 더 평활해 지며 알려진 패턴들 사이의 일반화 특성은 더 커지게 된다. 또한 작은 값일수록 회귀곡선은 더 상세하게 나타나며 학습패턴에 좀더 가까워진다. 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, i 번째 패턴 뉴런의 출력 $f(D_i)$ 는

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

와 같다. 여기서 D_i 는 i 번째 학습패턴으로부터 계산된 거리이다.

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

$$h(x) = \sum_{i=1}^n y_i \exp(-D(x, x_i))$$

$$s(x) = \sum_{i=1}^n \exp(-D(x, x_i)) \quad (3)$$

이다. 또한 출력층 뉴런의 출력 $\hat{y}(x)$ 는 합층의 A 뉴런 출력 $h(x)$ 를 B 뉴런 출력 $s(x)$ 로 단순히 나눈 값으로 계산된다. 즉,

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

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 결정하는데 이용되는 중앙값도 신경망의 성능을 좌우하는 하나의 중요한 파라미터임을 알 수 있다. 이는 각 학습패턴들의 입력값을 패턴 뉴런에 어느 정도 반영할 것인가를 결정하는 것으로 결국 독립변수 벡터 각각을 가장 잘 나타내는 군집의 중심값으로 설정되어야 할 것이다. 지금까지 대부분의 연구에서는 이를 위한 값으로 각 독립변수의 평균이나 종속변수와와의 가중된 평균을 중앙값으로 주로 이용하고 있다.

2.2 적응적 특징추출을 위한 PCA

PCA는 n 차원 입력공간의 데이터를 m 차원 출력공간의 데이터로 투영시키는 것이다. 여기서 $m \ll n$ 이며, 이는 입력데이터 벡터의 대부분 내부정보를 유지하도록 차원의 감소를 얻는 것이다. 자기상관행렬(auto-correlation matrix) $R_{xx} = \langle xx^T \rangle$ 를 가진 평균이 영인 입력벡터 $x = [x_1, x_2, \dots, x_n]^T$ 에 대해서 생각해 보자. 여기서 $\langle x \rangle$ 는 기대치를 나타낸다. 또한 w_1, w_2, \dots, w_m 이 R_{xx} 의 고유벡터와 직교되는 연결가중치 벡터라 할 때, $w_1 = [w_{11}, w_{12}, \dots, w_{1n}]^T$ 는 가장 큰 고유치 λ_1 과 일치하며, $w_2 = [w_{21}, w_{22}, \dots, w_{2n}]^T$ 는 두 번째로 큰 고유치 λ_2 와 $w_n = [w_{n1}, w_{n2}, \dots, w_{nn}]^T$ 는 가장 작은 고유치 λ_n 과 각각 일치한다. 이상의 관계를 행렬방정식으로 나타내면 다음과 같다. 즉,

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

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

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

$$y = W x \quad (6)$$

이다. 여기서 $W = [w_1, w_2, \dots, w_m]^T \in \mathbb{R}^{m \times n}$ 이며, T 는 전치를 나타낸다. 이 식은 행렬 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 m 개 주요 고유벡터에 의해 구성된 부공간으로 정의된다. 결국 PCA는 $\langle \|w_i^T x\|_2^2 \rangle$ 가 최대인 고유벡터 w_1, w_2, \dots, w_m 의 방향을 찾는 것이다.

한편, 입력되는 데이터의 주요특징들을 추출하기 위한 기존의 PCA 방법들을 살펴보면, 수치적 방법에서는 먼저 자기상관행렬 R_{xx} 를 계산한 다음 그것과 일

치하는 고유벡터를 계산하였다. 그러나 입력데이터 벡터가 대규모의 차원을 가진다면, 상관행렬은 대단히 크게 되어 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 적응적 학습알고리즘의 신경망을 이용하는 방법들이 제안되었다[9-11]. 여기서는 입력 데이터 벡터로부터 직접 고유벡터를 추정할 수 있기 때문에 상관행렬의 추정과정이 요구되지 않는다. 이때 이용되는 신경망은 주로 입력층과 출력층으로 구성된 단층구조이다.

Oja[9]는 정규화된 헤비안규칙(normalized Hebbian rule)의 적응학습 방법을 이용한 간단한 뉴런모델을 제안하여 정상입력벡터 계열의 주요 특징으로 연결가중치 벡터가 수렴됨을 증명하였다. 그러나 이 방법에서는 한 개의 중요한 주요특징만을 추출할 수 있으며, 임의의 m 개의 특징추출로의 확장은 불가능하다. Sanger 등[9]은 일반화된(generalized) 헤비안규칙을 이용함으로써 정상과정의 m 개의 가장 중요한 주요특징들을 계산하기 위한 다중 뉴런모델을 제안하였다. 한편, Foldiak[9,10]은 망의 입력과 출력사이의 연결가중치 경신에는 정규화된 헤비안규칙을 이용하고, 망의 출력사이의 측면연결 가중치 경신에는 반 헤비안규칙(anti-Hebbian rule)을 함께 이용한 학습알고리즘을 제안하였다. 일반적으로 Sanger 등에 의해 제안된 학습규칙 보다는 Foldiak에 의해 제안된 학습규칙이 수렴속도 면에서 더 우수한 것으로 알려져 있다[9].

그림 2는 n 개의 입력뉴런과 m 개의 출력뉴런으로 구성된 입력과 출력뉴런간 및 출력뉴런 상호간의 측면연결을 가진 단층신경망의 구조이다.

그림에서 입력과 출력의 관계를 나타내면 다음과 같다. 즉,

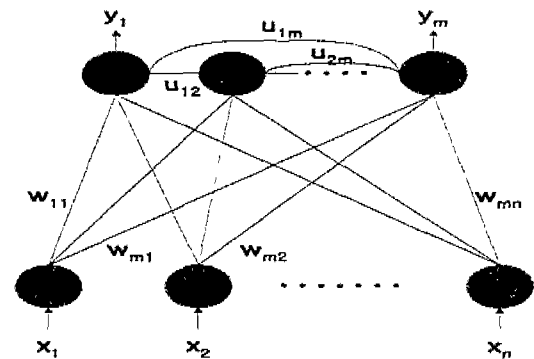


그림 2. m 개의 첫 번째 주요특징 추출을 위한 측면연결의 단층신경망

$$y_i = \sum_{j=1}^n w_{ij} x_j + \sum_{h=1}^{i-1} u_{ih} y_h \quad (i=1, 2, 3, \dots, m) \quad (7)$$

이다. 여기서 w_{ij} 는 입력뉴런과 출력뉴런을 연결하는 연결가중치이고, u_{ih} 는 출력뉴런 상호간의 측면연결 가중치이다. 이때 출력뉴런 간의 상호연결을 보면 뉴런 i 는 다만 $h < i$ 인 뉴런에만 연결된다. 따라서 신경망을 학습시켜 입출력 뉴런간 및 출력뉴런 상호간의 연결가중치를 각각 구함으로써 주어진 n 개의 입력벡터 x 로부터 m 개의 주요특징벡터 즉, 출력벡터 y 를 구할 수 있다. 입력과 출력뉴런 간의 연결가중치 경신에는 정규화된 헤비안규칙을 이용하고 출력뉴런 간의 측면연결가중치 경신에는 반 헤비안규칙을 이용하는 Foldiak의 학습알고리즘은 다음과 같다. 먼저 입출력 뉴런간의 연결가중치 w_{ij} 의 경신규칙을 살펴보면 다음과 같다. 즉,

$$w_{ij}(t+1) = w_{ij}(t) + \eta [y_i(t) x_j(t) y_i(t)^2], \quad (i=1, 2, \dots, m, j=1, 2, \dots, n) \quad (8)$$

이다. 또한 출력뉴런 간의 측면 연결가중치 u_{ih} 의 경신규칙은 다음과 같다. 즉,

$$u_{ih}(t+1) = u_{ih}(t) + \rho y_i(t) y_h(t), \quad (i > h) \quad (9)$$

이다. 식 (8)의 η 와 (9)의 ρ 는 학습율이다. 따라서 식 (8)과 (9)를 이용하여 입력과 출력뉴런 간의 연결가중치 및 출력뉴런 상호간의 측면 연결가중치를 각각 경신시켜 식 (7)에 대입하면 m 개의 주요특징들을 추출할 수 있다.

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 학습데이터의 독립변수 패턴 각각에 적응적 PCA를 이용하여 특징을 구한 다음 이를 GRNN의 입력으로 이용한다. 이렇게 하면 GRNN 입력층 뉴런의 개수가 줄어들어 학습시간이 감소되며, 학습데이터의 고차원에 따른 독립변수들 상호 간의 다중공선성 등으로 인한 회귀성능의 저하도 해결할 수 있어 GRNN이 가지는 우수한 속성을 더욱 더 잘 살릴 수 있을 것이다.

3. 시뮬레이션 결과 및 분석

제안된 기법의 일반회귀 신경망의 성능을 평가하기 위하여 Foldiak의 학습알고리즘의 단층신경망과 회귀

분석을 위한 4층 전향신경망을 각각 구성하였다. 특징 추출을 위한 단층신경망의 입력뉴런과 출력 뉴런사이의 초기 연결가중치는 각각 랜덤시드(random seed)를 이용하여 -1에서 +1 사이의 임의 값으로 설정하였다. 학습은 전체 반복회수가 10,000 이상이거나 전체 오차값이 허용치 10^{-4} 이하일 때 종료되도록 하였다.

제안된 기법을 이용한 GRNN을 21개의 패턴데이터(3개 독립변수)를 가진 암모니아 제조공정문제[1]와 32개의 패턴데이터(10개 독립변수)를 가진 자동차 연비문제[2]를 대상으로 팬티엄III-700 컴퓨터를 이용하여 시뮬레이션한 후 그 타당성을 확인하였으며, 기존의 독립변수를 그대로 이용하는 GRNN과 그 성능을 비교 고찰하였다.

3.1 암모니아 제조공정문제

실험에 이용된 암모니아 제조공정문제는 공정 입력 조건에 대한 공정의 출력특성 관계를 얻기 위한 문제이다. 이 문제는 3개의 공정 입력조건 x_1, x_2, x_3 를 독립변수와 1개의 출력특성 y 를 종속변수로 하는 21개의 패턴들을 가지며, 그 중 실험에서 16개는 학습 패턴으로 나머지 5개는 시험패턴으로 이용하였다. GRNN은 입력조건수의 개수에 해당하는 입력층 뉴런과 학습패턴 수에 해당하는 16개의 패턴층 뉴런으로 구성된 구조를 이용하였다. 다음의 표 1은 문제의 자료 일부와 중앙값으로 이용될 3개의 독립변수 각각에 대한 평균값과 가중평균값을 나타낸 것이다. 여기서 가

중평균은 각 독립변수의 중앙값 결정 시에 종속변수가 가중치로 작용되도록 하는 것이다.

표 2는 적응적 학습알고리즘의 단층신경망을 이용하여 3개의 입력조건인 독립변수 패턴 각각 내에 포함된 특징 2개와 1개를 각각 추출하여 이를 입력조건으로 나타낸 것이다. 이는 원래의 독립변수 패턴의 수를 감소시켜 GRNN이 가지는 독립변수 패턴의 고차원에 따른 제약들을 해결하기 위한 데이터이다.

표 3은 3개의 독립변수에 대해 GRNN에서 평균값과 가중평균값을 각각 중앙값으로 이용할 때, 5개의 시험패턴과 16개의 학습패턴을 포함한 21개의 패턴에 대해 평활요소 σ 의 변화에 따른 원하는 종속변수 y 와 학습 후 출력 \hat{y} 와의 절대오차(absolute error: AE, $AE = \sum_{i=1}^n |y_i - \hat{y}_i|$)를 나타낸 것이다. 표에서 각 중앙값에 따른 평활요소의 변화범위를 보면 전체적으로 10에서 0.1 사이에서 절대오차 값이 작음을 알 수 있다. 이는 σ 가 큰 값일수록 얻어지는 회귀곡선은 더 평활해지며, 알려진 패턴들 사이의 일반화 특성은 더 커지게 되어 절대오차가 커짐을 확인할 수 있다. 또한 σ 가 0.1 이하에서는 절대오차가 증가되어 회귀곡선은 더 상세하게 나타나나 잡음에 의해서 데이터가 오염되어 원하는 출력이 아님을 추측할 수 있다. 특히, 평활요소 σ 가 1인 경우에 중앙값으로 평균값을 이용할 때의 절대오차 값은 0.003397로 가중평균값을 이용할 때의 52.619118보다 매우 작은 절대오차 값을 가져 평균값의 경우가 우수한 회귀성능이 있음을 알 수 있다. 이는 가중평균을 중앙값으로 이용하는 경우 각 독립변수들이 가지는 공통적인 속성이 중앙값에 충분히 반영되지 않았기 때문이다. 결국 표로부터 GRNN에서는 평활요소 뿐만 아니라 중앙값도 회귀성능에 많은 영향을 미치는 요소임을 확인할 수 있다.

표 1. 암모니아 제조공정자료 및 3개 독립변수 각각의 중앙값

날짜	입력조건			출력특성
	x_1	x_2	x_3	y
1	80	27	89	42
2	80	27	88	37
⋮	⋮	⋮	⋮	⋮
15	50	18	89	8
16	50	18	86	7
⋮	⋮	⋮	⋮	⋮
21	50	20	91	15
평균값 \bar{x}	62.0625	21.5625	88	
가중평균값 weighted \bar{x}	20.95971803	20.9913	19.64915	

$$\bar{x}_1 = \left(\sum_{i=1}^{16} x_{1i} \right) / 16, \bar{x}_2 = \left(\sum_{i=1}^{16} x_{2i} \right) / 16, \bar{x}_3 = \left(\sum_{i=1}^{16} x_{3i} \right) / 16, \text{weighted } \bar{x}_1 = \left(\sum_{i=1}^{16} x_{1i} * y_i \right) / \sum_{i=1}^{16} x_{1i}, \text{weighted } \bar{x}_2 = \left(\sum_{i=1}^{16} x_{2i} * y_i \right) / \sum_{i=1}^{16} x_{2i}, \text{weighted } \bar{x}_3 = \left(\sum_{i=1}^{16} x_{3i} * y_i \right) / \sum_{i=1}^{16} x_{3i}$$

표 2. 3개의 입력조건으로부터 추출된 특징 및 각각의 중앙값

날짜	입력조건(2개)		입력조건(1개)	출력특성
	x_1	x_2	x	y
1	-100.556213	70.27377	-122.678429	42
2	-99.893005	69.9599	-121.95491	37
⋮	⋮	⋮	⋮	⋮
15	-86.96698	56.40694	-103.658089	8
16	-84.659462	55.25192	-101.094009	7
⋮	⋮	⋮	⋮	⋮
21	-97.406082	63.97698	-116.537544	15
평균값 \bar{x}	-91.3558468	61.35964	-110.05892	
가중평균값 weighted \bar{x}	20.05492698	20.30781	20.13371134	

표 3. 평활요소 σ 의 변화에 따른 절대오차 AE의 변화

σ	1000	500	100	50	10	1	0.1	0.01	0.001
평균값	52.615013	52.602634	52.211304	51.008224	26.242371	0.003397	52.619118	52.619118	52.619118
가중 평균값	52.677132	52.851906	58.803448	81.428551	52.619118	52.619118	52.619118	52.619118	52.619118

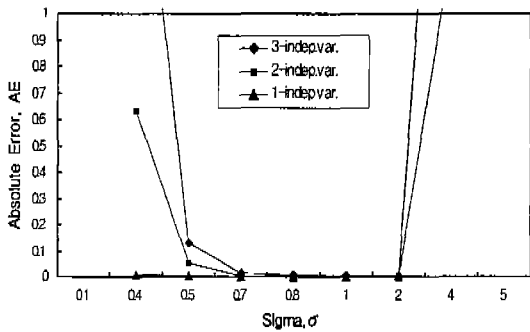


그림 3. 평활요소 σ 의 변화에 따른 절대오차 AE의 변화

그림 3은 패턴층 뉴런 수를 16개로 하고 각 독립 변수의 평균값을 중앙값로 하여 평활요소 σ 를 0.4에서 4까지 변화시키면서 원 문제의 3개 독립변수와 제안된 방법으로 추출된 특징인 2개 및 1개의 독립변수를 입력조건으로 각각 이용할 때 절대오차의 변화를 나타낸 결과이다. 그림에서 σ 의 범위를 제한한 것은 이 범위 밖에서는 3가지 입력조건 모두 상대적으로 많은 절대오차를 가져 좀 더 정확한 σ 의 값을 추정하기 위함이다. 여기서 σ 는 0.8에서 2 사이의 값을 가질 때 3가지 입력조건 모두에서 가장 작은 절대오차를 나타낼 수 있다. 또한 그림에서 보면, 제안된 기법에 의해 구해진 2개와 1개의 특징들을 독립변수로 이용하는 경우가 3개의 독립변수를 입력조건으로 이용하는 경우보다 절대오차가 작아 우수한 회귀 성능이 있음을 알 수 있다. 이는 문제에 이용되는 3개의 독립변수 상호간에는 선형적인 연관성이 많음을 확인할 수 있다. 특히, 제안된 기법에서는 독립변수가 2개인 경우보다 1개만을 이용하는 경우가 약간 우수한 회귀성능을 가졌다. 이는 암모니아 제조공정자료에

서 입력조건인 독립변수 상호간에는 많은 상관성이 있음을 다시 한번 더 확인할 수 있으며, 결국 독립변수 상호간의 상관성은 회귀성능을 저하시키는 결과를 초래함을 확인할 수 있다. 따라서 GRNN에서 독립변수의 고차원에 따른 회귀성능 저하와 같은 제약들은 제안된 기법의 PCA를 이용한 특징추출로 해결될 수 있을 것이다. 또한 GRNN의 학습시간 t_{CPU} 를 비교하기 위해 정수로 표현되는 컴퓨터의 CPU 시간을 측정 한 결과, 3가지 입력조건 모두에 대해서 t_{CPU} 는 0으로 상호 비교가 불가능하였다. 하지만 제안된 기법을 이용하면 GRNN 학습과정에서 입력조건 차원 감소로 계산시간이 줄어들어 상대적으로 학습시간도 줄어 들 것이다. 한편 그림에서는 독립변수가 3개일 때보다 2개와 1개로 감소시킨 제안된 방법의 GRNN이 넓은 범위에서 작은 절대오차 값을 가져 평활요소의 설정 범위도 용이함을 확인할 수 있다.

표 4는 21개의 패턴에 대해서 평균값을 중앙값으로 평활요소 σ 를 2로 하여 3가지 입력조건에 대해서 패턴층 뉴런의 개수에 따른 절대오차를 나타낸 것이다. 표에서 보면, 입력조건에 무관하게 패턴층 뉴런의 개수가 줄어들면 절대오차는 증가하여 이에 따른 회귀 성능도 나빠짐을 알 수 있다. 이는 패턴층 뉴런의 개수도 GRNN의 성능에 영향을 미치는 파라미터임을 의미하며, 일반적으로 널리 이용되는 개수 설정 방법에서는 학습패턴의 개수 즉, 종속변수의 개수와 동일 하게 한다. 이는 학습패턴 각각을 하나의 군집으로 간주하는 것으로 GRNN의 패턴층 뉴런 각각이 하나의 군집에 대응하도록 한 것이다. 결국 패턴층 뉴런 수를 증가시키면 회귀성능은 개선되나 그에 따른 계산시간의 증가로 GRNN의 학습시간은 증가될 것이다. 따라서 학습패턴 상호간의 유사성을 찾아서 그에 따라 군

표 4. 패턴층 뉴런의 개수에 따른 절대오차 AE의 변화

σ	16	14	12	10	8	4	2	1
3	0.000697	0.000805	0.001511	0.066033	1.05236	248.031464	418.999695	513.999939
2	0.001855	0.002425	18.822041	25.669941	37.553993	219.903122	418.999817	514.000061
1	0.001881	0.002841	28.607803	40.495136	55.028484	213.654358	418.999542	514.000061

집화시킨 후 군집의 수를 궁극적으로 패턴층 뉴런의 수로 설정하는 것도 하나의 해결방안이 될 수 있을 것이다. 한편, 표에서 보면 독립변수 개수를 줄인 경우 회귀성능은 상대적으로 패턴층 뉴런의 수에 더욱 더 많은 영향을 받음을 알 수 있다. 하지만 뉴런의 수를 종속변수의 개수로 설정하는 일반적인 방법을 따를 경우 고차원의 입력변수를 저차원으로 줄여 이용하는 잇점은 충분히 얻을 수 있다.

3.2 자동차 연비문제

자동차 연비문제는 10개의 독립변수에 대하여 1개의 종속변수를 가진 회귀분석 문제이다. 이 문제는 실린더 수(x_1), 입방인치 단위의 배기량(x_2), 마력(x_3), 최종기어비인 최종구동장치비(x_4), 자동차 무게(x_5), 순간 가속도(x_6), V자형(0)이나 직선형(1)의 엔진형태(x_7), 자동(0)이나 수동(1)이나의 기어의 종류(x_8), 기어속도의 수(x_9), 그리고 기화기의 밸브수(x_{10})에 대한 자동차 연비(y)와의 관계를 나타낸 것이다. 본 실험에서는 32개의 패턴데이터 중에서 25개는 학습패턴으로 나머지 7개는 시험패턴으로 이용하였다. 또한 이용된 GRNN은 입력조건수의 개수와 동일한 입력층 뉴런과 학습데이터에 대한 종속변수의 수인 25개의 패턴층 뉴런으로 구성된 구조를 이용하였다. 다음의 표 5는 문제의 자료 일부와 중앙값으로 이용될 10개의 독립변수 각각에 대한 평균값과 가중평균값을 나타낸 것이다. 여

기서도 평균값과 가중평균값은 각각 표 1의 실험에서와 동일한 방법으로 얻어진 값들이다.

표 6은 10개의 입력뉴런과 25개의 패턴 뉴런을 가진 GRNN에서 평균값과 가중평균값을 각각 중앙값으로 이용할 때, 25개의 시험패턴과 7개의 학습패턴을 포함한 32개의 패턴에 대해 평활요소 σ 의 변화에 따른 원하는 종속변수 값과 학습 후 출력과의 절대오차를 나타낸 것이다. 여기서도 표 3에서처럼 평균값을 중앙값으로 이용할 때의 절대오차가 가중평균값을 이용할 때보다 작은 값으로 중앙값이 회귀성능에 영향을 미침을 알 수 있다. 특히, 가중평균값을 이용할 경우에 절대오차의 변화가 적음을 알 수 있다. 이는 이 문제에서도 가중평균의 경우에 출력특성이 포함됨으로써 중앙값 설정에 있어 각 독립변수들의 속성이 충분히 반영되지 않음으로 추측된다. 또한 표에서는 σ 의 값이 50 근처일 때 상대적으로 낮은 절대오차를 보이며, σ 의 변화에 따른 회귀성능의 영향도 확인할 수 있다. 결과적으로 암모니아 제조공정 문제에서와 유사한 결과를 알 수 있다.

그림 4는 평균값을 중앙값으로 하여 평활요소 σ 를 100에서 1까지 변화시키면서 원 문제의 10개 독립변수와 제안된 방법으로 추출된 특징들인 9개, 8개, 5개, 그리고 2개의 독립변수를 입력조건으로 각각 이용할 때 절대오차의 변화를 나타낸 결과이다. 그림에서 제안된 기법에 의해 구해진 9개와 8개의 특징들을 독

표 5. 자동차 연비자료 및 10개 독립변수 각각의 중앙값

개수	입력조건										출력특성 y
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	
1	6	160	110	3.9	2.62	16.46	0	1	4	4	21
2	6	160	110	3.9	2.875	17.02	0	1	4	4	21
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
24	8	350	245	3.73	3.84	17.05	0	0	3	4	13.3
25	8	400	175	3.08	3.845	18.9	0	0	3	2	19.2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
32	4	121	109	4.11	2.78	18.6	1	1	4	2	21.4
평균값 \bar{x}	6.4	246.83	141.64	3.49	3.4	18.22	0.44	0.24	3.4	2.6	
가중평균값 weighted \bar{x}	18.18	17.02	17.45	20.16	18.07	19.76	24.06	26.92	20.09	17.81	

표 6. 평활요소 σ 의 변화에 따른 절대오차 AE의 변화

σ	1000	500	100	50	10	1	0.1	0.01	0.001
중앙값									
평균값	32.286507	31.802691	13.577857	0.602318	23.992434	32.440567	32.440567	32.440567	32.440567
가중 평균값	32.999462	34.627811	44.038193	28.161028	32.440567	32.440567	32.440567	32.440567	32.440567

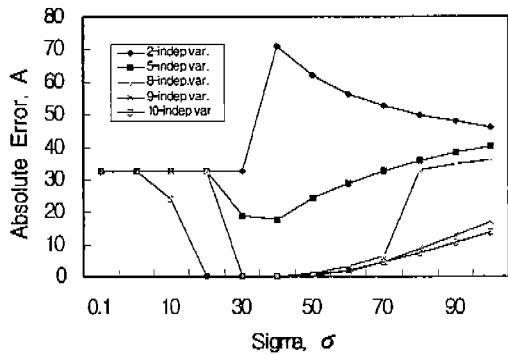


그림 4. 평활요소 σ 의 변화에 따른 절대오차 AE의 변화

립변수로 이용하는 경우 10개의 독립변수를 입력조건으로 이용하는 경우와 σ 가 50에서 30 사이에서는 절대오차가 비슷하나 그 외의 범위에서는 다소 큰 값을 알 수 있다. 또한 입력조건을 감소시킬 경우에 평활요소의 설정범위가 상대적으로 줄어드는 것도 확인할 수 있다. 특히, 특징 5개와 2개를 입력조건으로 이용할 경우는 상대적으로 더욱 더 큰 절대오차를 보여 이는 지나치게 입력변수를 줄일 경우 추출된 특징들이 독립변수들 각각의 속성을 충분히 잘 반영하지 못하기 때문으로 추측된다. 한편 암모니아 제조공정문제의 결과인 그림 2와 비교하면, 그림 2에서는 특징을 추출하여 입력조건을 줄임으로써 평활요소의 설정범위는 더욱 더 넓어지나 그림 4에서는 상대적으로 그 설정범위가 좁아짐을 알 수 있다. 또한 제안된 기법에 의한 4가지 입력변수 상호간에서도 입력변수가 줄어들수록 절대오차는 많아지며 그에 따른 회귀성능도 나빠짐을 알 수 있다. 이는 자동차 연비문제의 경우에 입력조건인 각 독립변수들 사이에는 상호 연관성이 거의 없으며, 암모니아 제조공정문제에서는 3개의 입력조건 상호간에 선형적인 관계가 더욱 많이 존재함을 보여준 것이다. 결국 제안된 기법은 각 독립변수 상호간에 더욱 더 많은 다중공선성을 가지는 회귀분석 문제에서는 우수한 성능이 있음을 추측할 수 있으며, 특

히 입력조건 상호간에 선형적 연관성이 많을수록 더욱 더 우수한 성능이 있을 것으로 추측된다. 이는 제안된 기법은 데이터 상호간의 선형성을 바탕으로 주요성분인 특징을 추출하는 PCA 기법에 기반을 두고 있기 때문이다. 그림에서는 평활요소 σ 의 값이 30일 때 각 독립변수를 입력조건으로 하는 5가지 모두에서 가장 낮은 절대오차를 나타내고 있다. 또한 여기에서도 GRNN의 학습시간 t_{CPU} 를 비교하기 위해 컴퓨터의 CPU 시간을 측정한 결과, 5가지 입력조건 모두에 대해서 t_{CPU} 는 0으로 상호 비교가 불가능하였다. 이는 실험에 이용된 컴퓨터의 성능에 비해 문제의 규모가 너무 작기 때문으로 추측되나 여기에서도 제안된 기법은 입력조건을 감소시켜 상대적으로 빠른 학습속도를 얻을 수 있을 것이다.

표 7은 32개의 패턴에 대해서 평균값을 중앙값으로 평활요소 σ 를 30으로 하여 5가지 입력조건에 대해서 패턴층 뉴런의 개수에 따른 절대오차를 나타낸 것이다. 여기서도 표 4에서처럼 입력개수에 무관하게 패턴층 뉴런의 개수가 줄어들면 절대오차는 증가하여 이에 따른 회귀성능도 나빠짐을 알 수 있다. 또한 절대오차의 변화는 입력조건이 줄어들수록 더욱 더 많은 영향을 받음도 알 수 있다. 따라서 일반적으로 널리 이용되는 뉴런층 뉴런 개수는 학습패턴의 개수, 즉 종속변수의 개수와 동일하게 설정하는 것이 가장 적합한 방법으로 생각되나 고차원의 입력변수를 저차원으로 줄여 이용하는 장점을 더욱 더 잘 살리기 위해서는 그 설정방안도 충분히 연구되어야 할 것이다.

4. 결 론

본 논문에서는 적응적 학습 알고리즘의 주요성분분석 기법으로 독립변수 패턴의 특징을 추출하고 이를 일반회귀 신경망의 학습데이터로 이용함으로써 입력조건을 감소시켜 일반회귀 신경망의 성능을 개선하는 방법을 제안하였다. 이는 주요성분분석 기법이 가지는 대용량의 입력 데이터를 통계적으로 독립인 특

표 7. 패턴층 뉴런의 개수에 따른 절대오차 AE의 변화

σ	25	22	20	18	15	10	5	1
10	0.002005	6.43364	12.737302	16.300236	18.147196	44.572617	86.240463	154.482788
9	0.004357	3.041022	7.34145	9.011165	9.06415	9.497796	148.024002	154.236008
8	0.005844	32.425346	59.466438	86.540298	96.085686	96.65258	101.278442	151.683975
5	18.90683	76.007149	85.744064	99.203705	130.67168	156.541245	198.521332	219.666016
2	32.44057	33.301432	53.093578	147.86557	149.01515	154.468582	308.874542	347.681061

정들의 집합으로 변환시키는 속성을 살려 학습데이터의 차원을 감소시킴으로써 고차원의 학습데이터에 따른 일반회귀 신경망이 가지는 제약을 해결하기 위함이다.

제안된 기법의 일반회귀 신경망을 3개의 독립변수를 가진 21개 학습패턴의 암모니아 제조공정문제와 10개의 독립변수를 가진 32개 학습패턴의 자동차 연비문제에 각각 적용하여 시뮬레이션한 결과, 기존의 일반회귀 신경망에 의한 결과와 비교할 때 학습시간과 절대오차로 나타나는 회귀성능에서 더욱 우수함을 확인할 수 있었다. 또한 평활요소의 설정 면에서도 우수한 특성이 있음을 확인할 수 있었다.

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.

참고문헌

[1] 허명희 외 1, 'SAS 회귀분석,' 자유아카데미, 1996.
 [2] 강명옥 외 3, '회귀분석 : 모형개발과 진단,' 을국출판사, 1997.
 [3] S. Haykin, 'Neural Networks : A Comprehensive Foundation', IEEE Press, New York, 1994
 [4] D. F. Specht, "A General Regression Neural Network", IEEE Trans. on Neural Networks, Vol. 2, No. 6, pp. 568-576, Nov. 1991.
 [5] M. Caudill, "GRNN and Bear It", AI Expert, vol. 5, pp.28-33, May 1993.
 [6] T. Masters, 'Advanced Algorithms for Neural Networks : A C++ Sourcebook', John Wiley & Sons, Inc., New York, 1995.

[7] 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.
 [8] Y. Chtioui, S. Panigrahi, and L. Francl, "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.
 [9] A. Cichock and R. Unbehau, "Neural Networks for Optimization and Signal Processing", John Wiley & Sons., New York, 1993.
 [10] 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.
 [11] 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 (Yong-Hyun Cho)

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 교수
 관심분야 : 신경망, 병렬분산처리, 신호처리, 영상인식, 교환기 등