

수목구조 지능시스템을 이용한 고차원 공간에서의 비선형근사

Nonlinear Approximation in High-Dimensional Spaces Using Tree-Structured Intelligent Systems

길 준 민*, 정 창 호**, 강 성 훈***, 박 주 영****, 박 대 회*****
Joonmin Gil*, Changho Jung**, Seonghoon Kang***,
Jooyoung Park****, Daihee Park*****

요 약

기존의 RBF 신경망 및 퍼지 시스템을 고차원 입력 공간에서의 비선형 근사에 적용할 경우 은닉 노드의 수 혹은 퍼지 IF-THEN 규칙의 수가 기하급수적으로 증가한다. 본 논문에서는 이러한 문제점을 개선하기 위해 반국소 유닛을 기본 요소로 하는 수목구조 지능 시스템을 제안하고, 이를 효과적으로 학습하기 위하여 수정 유전자 알고리즘 및 LMS 규칙에 기반을 둔 학습 알고리즘을 개발한다. 제안된 시스템에 대한 근사 능력 해석이 수행되고, 실험적 고찰을 통하여 개발된 방법론의 유용성이 입증된다.

ABSTRACT

Conventional radial-basis-function networks and fuzzy systems have serious problems in dealing with the nonlinear approximations on high-dimensional spaces due to the explosive increase of the number of hidden nodes or fuzzy IF-THEN rules. In order to avoid such problems, this paper proposes a tree-structured intelligent system in which semi-local basis functions form its basic elements, and develops a training algorithm for the proposed system based on the modified genetic algorithm and LMS rule. Theoretical analysis is performed on the approximation capability of the proposed system, together with experimental studies which demonstrate the effectiveness of the developed methodology.

I. 서 론

신경망을 여러 가지 공학 문제에 응용하는 과정에 서, 어떠한 종류의 신경망을 어떻게 학습시킬 것인가 를 결정하는 것은 매우 어려운 선택인 동시에 문제

*고려대 대학원 컴퓨터과학과
**고려대 대학원 전산학과
***고려대 대학원 정보공학과
****고려대 제어계측공학과
*****고려대 전산학과

해결의 성패를 좌우하는 중대한 사안이기도 하다. 이러한 선택에 있어서 상호 경쟁 관계에 있는 주요한 전방향 신경망(feedforward neural networks)으로는 MLP(multi-layered perceptrons)와 RBF(radial basis function) 신경망이 있다. MLP와 RBF 신경망은 각각 임의의 함수가 주어졌을 때 이를 충분히 가깝게 근사할 수 있다는 점에서 그 근사 능력이 이론적으로 검증된 상태인데¹¹⁻¹², 실제적인 문제에 대해 둘 중 어느 것을 선택하는 것이 유리한가는 주어진 문제가 갖는 특성에 따라 달라질 수 있다. RBF 신경망을 선호하는 학자들은 학습이 MLP에 비해 빠르게 수행될 수 있음¹³을 내세운다. 그러나, 이러한 장점은 해결하고자 하는 문제가 고차원의 입력 공간을 갖는 경우에는 은닉노드(hidden nodes) 개수의 폭발적 증가라는 결정적인 단점에 의해 상쇄될 수밖에 없다. 이를 좀 더 구체적으로 설명하면 [8-9] 등에서 사용된 전형적인 RBF 신경망의 경우, 입력 공간이 r -차원의 큐브 $[-L, L]^r$ 이고 각 축(coordinate)이 N 개의 베이스 함수(basis functions)로 표현되면 전체 입력 공간은 N^r 개의 RBF를 필요로 한다. 이에 따라 우리가 필요로 하는 RBF 신경망의 전체 크기는 입력 공간의 차원인 r 이 증가함에 따라 매우 빠른 속도로 증가하게 된다. 따라서 이론적으로는 [4-6]에서 증명된 바와 같이 고차원 공간에서 정의된 임의의 함수에 대해서 이를 충분히 가깝게 근사할 수 있는 RBF 신경망이 존재하지만, 실제적인 근사해로써의 RBF 신경망을 찾아내는 것은 공학적으로 다루기 힘든 문제가 되는 것이다. 이와 같은 어려움은 고차원의 입력 공간을 갖는 문제를 다룰 때 불가피하게 만나게 되는 현상으로써 흔히 curse of dimensionality라고 불린다(RBF 신경망을 이용한 고차원 공간 위에서의 근사를 논할 때 부각되는 이러한 문제점은 MLP의 경우에도 정도의 차이는 있지만 예외는 아니다¹⁴). 이러한 문제점으로 인하여, 함수근사가 고차원 공간에서 이루어져야 할 때에는 신경망의 크기가 다루기 힘든 규모로 증가하는 것을 막기 위해 RBF 신경망이 갖는 함수의 국소적 표현 능력을 어느 정도 희생하는 것이 필요하다. 이러한 현상은, 최근 들어 RBF 신경망과 함수적 동치(functional equivalence) 관계에 있음¹⁰⁻¹¹이 밝혀진 퍼지 시스템의 설계에서도 부딪히게 되는 문제이다.

퍼지 논리는 기존의 논리 체계보다 인간 사고의 기

본인 언어적 정보를 더 적절하게 표현할 수 있으므로, 실세계의 근사적이고 불확실한 현상을 기술하는데 효과적으로 이용될 수 있다. 이러한 퍼지 논리에 기초한 방법론은 시스템의 특성이 복잡하여 기존의 정량적인 방법으로는 해석할 수 없거나, 획득된 정보가 정성적이고 부정확하며 불확실한 경우에 우수한 성능을 보인다.

퍼지 시스템을 설계하는 과정은 구조 식별(structure identification)과 매개변수 조정(parameter tuning)의 두 단계로 나눌 수 있다¹²⁻¹³. 구조 식별 단계에서는 시스템의 입력력 변수를 가려내고 입력력 관계를 퍼지 규칙 베이스로 표현한다. 그리고, 매개변수 조정 단계에서는 목표 시스템과 모델 사이의 오차를 줄이기 위해 각 매개변수를 조정한다. 즉, 퍼지 소속 함수를 위한 매개변수와 규칙 결론부의 실수 항을 조정한다. 퍼지 시스템의 설계에 있어서, 주어진 전문가의 지식을 퍼지 규칙의 형태로 나타내는 것은 어렵지 않다. 하지만, 퍼지 소속 함수의 개수 및 형태를 결정하는 일, 퍼지 규칙의 개수 및 조합을 결정하는 일, 그리고 규칙 결론부의 실수 항을 정밀하게 조정하는 일 등을 설계자의 주관에만 의존하여 해결하고자 하는 경우에는, 많은 노력과 시간이 요구되곤 하였다. 이러한 단점은 기본적으로 학습 기능(learning mechanism)의 결여에 의해 파생된다고 할 수 있겠는데, 최근에는 학습 능력을 갖고 있는 신경망 이론과 유전자 알고리즘을 이용하여 서로의 장점을 결합한 융합(fusion) 연구가 활발하게 이루어져 왔다. 이러한 방향의 연구들은 퍼지 소속 함수의 조정¹⁴⁻¹⁵과 구조 식별 단계에 대한 적용¹⁶⁻¹⁷ 등으로 이루어지는데, 아직은 구조 식별과 매개변수 조정이 종합적으로 이루어지는 완성된 방법론의 출현과는 거리가 있는 형편이다. 더구나 퍼지 시스템이 고차원 입력 공간을 갖는 경우엔 만나게 되는 퍼지 규칙 개수의 폭발적 증가는 기존의 융합 연구로는 대처하기 힘든 측면을 갖는다.

본 논문에서는 이와 같은 문제점에 대한 인식에 근거하여, 함수 공간 $C[-L, L]^r$ 의 근사를 위해 국소적 특성의 표현에 적절한 풀린 국소 베이스 함수(local basis functions)의 일차 결합(linear combination)으로 구성되는 RBF 신경망 혹은 퍼지 시스템의 기본 구조를 학습에 따라 그 구조 및 매개변수가 적응적으로 변하는 반국소(semi-local) 베이스의 일차 결합 형

태의 구조로 바꾸어서 고차원 공간에서의 함수근사를 효과적으로 도모하고자 한다. 본 논문에서 제안하는 지능형 시스템의 구조는 초기 단계에서 변수 x_1, \dots, x_r 중 하나에만 의존하는 베이스스 함수의 일차 결합을 사용하고, 학습이 진행되는 과정에서 필요에 따라 새로운 변수가 기존의 반국소 베이스스에 추가되어 점차로 다변수에 의존하는 수목(혹은, 트리) 형태의 베이스스 함수가 형성되도록 한다. 그리고 수목 형태의 형성에 있어서 반국소 베이스스 함수의 모양과 연결 강도는 유전자 알고리즘과 LMS(least mean square) 규칙을 이용하여 결정한다. 여기서 제안하는 수목구조 시스템이 고차원 공간 위에서의 비선형 근사 문제에 어느 정도 역량을 발휘할 수 있을지는, 근사 능력에 관한 이론적 분석과 시뮬레이션을 통해 확인할 수 있다. 본 논문에서는, 제안된 수목구조의 근사 능력을 수목의 깊이(depth)가 충분히 깊은 경우와 그렇지 못한 경우로 구분하여 살펴보기로 한다. 여기에서 수목의 깊이가 충분하다 함은 모든 입력 변수를 포함하는 국소 베이스스가 출현할 정도로 수목의 확장이 반복된 경우를 의미하고, 그렇지 못한 경우는 수목의 확장이 이러한 베이스스의 출현 이전에 중단되는 경우를 의미한다. 깊이가 충분한 경우에는 EBF 신경망^[23] 또는 Wang-Mendel 퍼지 모델^[24]이 전체 수목의 부분 구조를 형성하므로, 이들에 대한 근사 능력 분석 결과가 곧바로 활용될 수 있다. 반면에 수목의 깊이가 불충분할 경우에는, 수목구조 시스템이 갖는 구조적 특성으로부터 변수 분리 현상을 관찰한 후 이를 이용한 분석을 수행하기로 한다. 이와 같은 근사 능력 해석은 수목구조 시스템에 관한 중요한 이론적 기초를 제공하게 된다. 반면에 시뮬레이션을 통한 실험적 고찰은, 본 논문에서 제안하는 뉴로퍼지 시스템의 구조 및 학습 방법론이 고차원 입력 공간 위에서 정의된 함수가 간단한 형태의 부분 구조의 조합으로 만족스러운 정도의 정확성을 가지고 근사될 수 있을 때, curse of dimensionality를 얼마나 효과적으로 완화시킬 수 있는지를 보여주게 된다. 실험적 고찰 과정에서는, 예측 알고리즘의 성능 평가를 위한 적용 문제로 널리 활용되고 있는 Mackey-Glass 시계열 문제를 대상으로 본 연구의 방법론과 관련 연구간의 성능 비교가 이루어진다.

본 논문의 순서는 다음과 같다: 2장에서는 본 논문

을 통해 제안되는 수목구조 시스템을 자세히 설명한다. 3장에서는 수목구조 시스템의 근사 능력에 논한다. 그리고, 4장에서는 시계열 예측 문제를 위한 시뮬레이션을 통해 그 결과를 관련 연구와 비교 분석한다. 마지막으로 5장에서는 결론을 제시한다.

II. 수목구조 시스템

본 논문에서는, 기존의 RBF 신경망 및 퍼지 시스템을 고차원 공간에서의 함수 근사에 적용할 경우 은닉노드의 수 혹은 퍼지 IF-THEN 규칙의 수가 급속히 증가함을 피할 수 없다는 문제점에 대한 인식으로부터 출발하여, 반국소 유닛(semi-local units)을 기본으로 하는 뉴로퍼지 시스템의 구조를 제안하고 이를 효과적으로 학습하기 위하여 유전자 알고리즘을 활용하고자 한다.

본 논문의 방법론을 구체적으로 설명하기 위하여, 수목구조 시스템에 관한 내용을 초기 수목의 구조, 유전자 알고리즘을 이용한 학습, 그리고 수목의 확장으로 나누어서 기술한다.

1) 초기 수목의 구조

본 논문에서 다루는 수목구조 시스템에서는, 고차원 공간 위에서 정의된 베이스스 함수가 여러 변수를 동시에 포함할 경우에는 필요한 함수의 수가 매우 커질 수밖에 없음에 착안하여, 전체 시스템이 초기에는 깊이가 1 이고 각 베이스스 함수가 하나의 변수에만 의존하는 초기 형태를 취하다가 학습이 진행되어 감에 따라 수목이 확장되어 가는 방식을 취하도록 한다. 따라서, 초기 수목에서는 우리가 근사하고자 하는 함수의 출력 y 를, 다음의 (1)식에서 보여진 바와 같이 하나의 변수에만 의존하는 반국소 베이스스 함수의 일차 결합 $y[1]$ 로 근사한다.

$$y[1] = \sum_{i=1}^r \sum_{j=1}^{J_i} w_{ij} \phi_{ij}(x_i) \quad (1)$$

(1) 식에서 소개된 깊이 1의 구조는 베이스스 함수 ϕ_{ij} 의 일차 결합 꼴이므로, 일정 조건 아래서 LMS 규칙을 사용하면 최적 값으로 수렴해 가는 연결 강도 w_{ij} 를 얻을 수 있다. 여기에서 제시하는 수목구조의 원형이라고 할 수 있는 Sanger의 LMS 트리^[18-19]에서는

비슷한 구조로부터 수목 형성이 시작되는데, 이때 베이스 함수 $\phi_{ij}(x_i)$, $j=1, \dots, J_i$ 로는 일정한 간격으로 이동되어진 가우시안, sin 등의 함수가 사용된 바 있다. 그러나, LMS 트리에서와 같이 각 축마다 미리 정해진 베이스 함수(fixed basis functions)를 사용하는 경우에는 각 베이스의 모양이나 개수를 결정하는 작업이 매우 어려워진다. 즉, 함수의 완만한 변화와 급격한 변화가 입력 공간의 부분에 따라 각각 존재하는 경우에, 변화가 급격한 부분에 초점을 맞추면 완만한 부분에서도 조밀한 베이스를 사용할 수밖에 없는 낭비적 요소를 지니게 되고, 반면에 완만한 변화를 갖는 부분에 초점을 맞추면 변화가 급격한 부분은 충분히 정확하게 표현할 수 없는 문제점을 갖게 된다. 본 논문에서 제안하는 수목구조 시스템은 초기의 수목구조가 (1)식의 응답 특성으로부터 출발하되, 각 베이스 함수의 모양, 개수 및 연결 강도를 유전자 알고리즘을 통해 최적에 가깝게 결정하는 특징을 갖는다. 즉, 각 변수 x_i 에 관한 베이스 함수 ϕ_{ij} 가 미리 정해지지 않고, 유전자 알고리즘을 이용한 학습 과정 중 평가 함수를 극대화하는 방향으로 결정되는 것이다. 이러한 과정에서 각 축을 담당하는 베이스 함수의 완만 또는 급격한 변화 정도가 조절될 뿐만 아니라 일정 변수에 대한 의존도를 0으로 하는 경우도 등장할 수 있도록 한다. 지금까지 소개한 깊이 1의 초기 수목이 가우시안 베이스를 갖는 경우에는, 출력 노드의 응답은 다음 식들에서 보여지는 바와 같이 가우시안의 가중합(weighted sum) 또는 가중평균(weighted average)으로 정의할 수 있다:

$$y[1] = \sum_{i=1}^r \sum_{j=1}^{J_i} w_{ij} \exp\left[-\frac{1}{2} \left(\frac{x_i - \mu_{ij}}{\sigma_{ij}}\right)^2\right] \quad (2-1)$$

$$y[1] = \frac{\sum_{i=1}^r \sum_{j=1}^{J_i} w_{ij} \exp\left[-\frac{1}{2} \left(\frac{x_i - \mu_{ij}}{\sigma_{ij}}\right)^2\right]}{\sum_{i=1}^r \sum_{j=1}^{J_i} \exp\left[-\frac{1}{2} \left(\frac{x_i - \mu_{ij}}{\sigma_{ij}}\right)^2\right]} \quad (2-2)$$

(2-1), (2-2)식의 정의에 있어서, 가중합은 RBF 신경망 분야에서 그리고 가중평균은 퍼지 시스템 분야에서 널리 쓰이는 형태이다. 다음에서는, 가우시안 베이스 함수를 이용한 수목구조 시스템에 대하여, 베이스 함수의 개수 J_i , 가우시안 베이스의 중심

(centroids) μ_{ij} , 유연계수(smoothing factors) σ_{ij} , 및 연결 강도 w_{ij} 가 유전자 알고리즘에 의하여 최적에 가깝게 결정되는 과정을 자세히 설명한다.

2) 유전자 알고리즘을 이용한 학습

(2-1), (2-2)식으로 표현되는 초기 수목에 있어서 우리가 결정해야 할 부분은

$$J_1, \dots, J_r, \mu_{ij}, \sigma_{ij}, w_{ij}, i=1, \dots, r, j=1, \dots, J_i$$

이다. 본 논문에서는 이들 중 연결 강도 값 w_{ij} 를 제외한 전부분을 베이스의 기하학적 특성에 따라 일정 길이 L의 r배인 rL 비트의 이진 스트링(binary string) 개체로 부호화 한다. 이들 중 (i+1)번째 자리부터 (i+L)번째 자리까지의 스트링은 매개변수 $J_i, \mu_{ij}, \sigma_{ij}, j=1, \dots, J_i$ 를 결정하는데 사용되며, 그 의미는 다음과 같다: 0이 개체 내의 입력 변수 부분에 하나 이하로 존재하는 경우에는 해당 변수에 의존하는 베이스를 삭제한다. 이러한 정의는, 우리가 근사하고자 하는 함수가 특정 입력 변수와는 무관한 경우를 효과적으로 다룰 수 있는 가능성을 제공해 준다.

다음에서는 깊이 1의 초기 수목에 대하여, 학습 데이터에 대한 오차와 일반화 능력(generalization capability)을 반영한 평가 함수를 최대화시키는 탐색 과정을 설명한다. 이 탐색 과정에는 기존의 유전자 알고리즘보다 우수한 성능을 가진 것으로 알려진 수정 유전자 알고리즘 modGA^[20]를 사용한다(그림 1).

```

procedure modGA
begin
  t := 0
  initialize P(t)
  evaluate P(t)
  while (not termination-condition) do
    begin
      t := t+1
      select-parents from P(t-1)
      select-dead from P(t-1)
      form P(t): reproduce the parents
      evaluate P(t)
    end
  end

```

그림 1. 수정 유전자 알고리즘 modGA

그림 1의 수정 유전자 알고리즘은 선택 과정에서 개체 집단의 개체를 세 부류로 나눈다: 부모 개체 (parent individual), 소멸 개체 (dead individual), 나머지 개체 (remaining individual). 부모 개체는 재생산 과정에 사용될 개체이고, 소멸 개체는 낮은 적합도 값을 갖기 때문에 다음 세대로 선택되지 않는 개체들이다. 그리고, 나머지 개체는 유전 연산을 거치지 않고 다음 세대의 개체집단으로 복제되는 부류이다. 부모 개체로 선택된 개체는 현재 세대의 개체 집단들 중 우수한 적합도를 갖는 개체들로 구성되며 유전 연산을 통하여 더 좋은 개체로 진화할 수 있는 기회를 갖는다. 또한, 나머지 개체는 유전 연산을 거치지 않기 때문에, 유전 연산에 의하여 우수한 개체가 소멸되는 현상을 방지해 주게 된다. 개체의 분류는 적합도의 크기에 따라 행해지는데, 본 논문에서는 여기에 사용되는 개체 s_m 의 적합도 $E(s_m)$ 를 다음과 같이 정의하였다:

$$E(s_m) = \frac{1}{\alpha_E(\text{학습오차}) + \alpha_B(\text{베이스스 함수의 갯수})} \quad (3)$$

이러한 정의는, 학습 데이터에 대한 오차와 일반화 능력에서 동시에 뛰어난 개체가 계속 진화해 갈 수 있도록 해준다. 이상에서 설명한 수정 유전자 알고리즘을 초기 수목의 학습에 적용하면 다음과 같은 계산 순서가 얻어진다:

• modGA를 이용한 수목의 학습

- [1] 이진 난수 발생에 의해, 길이 rL을 갖는 이진 스트링 개체들로 이루어지는 초기 집단 $S(t) = \{s_1(t), \dots, s_M(t)\}$ 를 구성함($t=0$).
- [2] 집단 내의 각 개체가 갖는 이진 스트링을 앞에서 설명한 방법에 따라 베이스스 함수로 변화시킴.
- [3] 학습 데이터를 이용하여 각 개체가 표현하는 수목구조 시스템을 LMS 규칙으로 학습시킴.
- [4] LMS 규칙에 따른 학습이 미리 정해진 횟수에 도달하거나, 학습 오차의 변화량이 정해진 값 이하가 되면, LMS 규칙에 따르는 학습을 종료하고 단계 [5]로 넘어감.
- [5] 집단 내의 개체 $s_m(t)$, $m=1, \dots, M$ 에 대하여 적

도 $E(s_m)$ 을 계산.

[6] 각 개체가 갖는 적합도를 이용하여 개체 집단 $S(t)$ 의 개체들을 부모 개체, 소멸 개체, 및 나머지 개체로 분류.

[7] 부모 개체에 대하여 교배 연산 및 돌연변이 연산을 수행하여 자식 개체들을 생성하고 여기에 나머지 개체들을 그대로 복제하여 추가함으로써, 다음 세대의 개체 집단 $S(t+1)$ 를 구성함.

[8] 정해진 반복 세대수에 이를 때까지 단계 2에서 단계 7까지를 반복함.

위의 과정 중 단계 [2]에서는 각 개체에 해당하는

$$J_1, \dots, J_r, \mu_{11}, \dots, \mu_{rj}, \sigma_{11}, \dots, \sigma_{rj},$$

를 결정하게 되고, 단계 [3]-[4]에서는 LMS 규칙에 의한 학습을 통해 연결 강도 w_{ij} 를 구한다. 이같은 과정을 거치면 전체 시스템의 구조와 매개변수가 모두 결정된다. 각 개체가 나타내는 수목구조 시스템이 갖는 적합도는 식 (3)에서와 같이 학습 오차와 베이스스 함수의 개수를 반영하여 결정하는데, 이 값은 modGA의 개체 분류 과정인 단계 [6]에서 사용된다. 이상에서 설명한 학습 과정을 통하여 최고의 적합도를 갖는 개체를 선택할 경우에는, 적합도의 정의가 갖는 특성으로부터 학습 데이터에 대한 오차의 최소화와 overfitting의 방지를 동시에 기할 수 있다.

3) 수목의 확장

앞부분에서 설명한 방법을 통하여 얻은 깊이 1의 수목구조 시스템의 성능이 만족스럽지 못할 경우에는 연결 강도에 대한 변화 압력이 가장 큰 베이스스를 따라 수목을 확장시킨다. 베이스스 함수 ϕ 와 출력을 잇는 연결 강도가 갖는 변화 압력은 LMS 규칙을 참조하여 다음과 같이 정의한다. (4)식에서 위첨자 p 는, p 번째 학습 데이터를 의미하는 index이다.

$$\sum_{p=1}^P |(y^p - \hat{y}^p)\phi(x^p)|^2 \quad (4)$$

이러한 수목 확장은 그림 2와 같이, 이차원 공간 위에 정의된 베이스스의 effective support(각 베이스스의 출력 값이 $\exp(-1/2)$ 이상인 부분)를 이용하면 효과

적으로 설명할 수 있다. 초기 수목을 구성하는 반국소 베이스 함수(semi-local basis functions)의 effective support가 그림 2와 같을 때, 수목은 이들 중 변화 압력이 가장 큰 부분을 따라 확장된다.

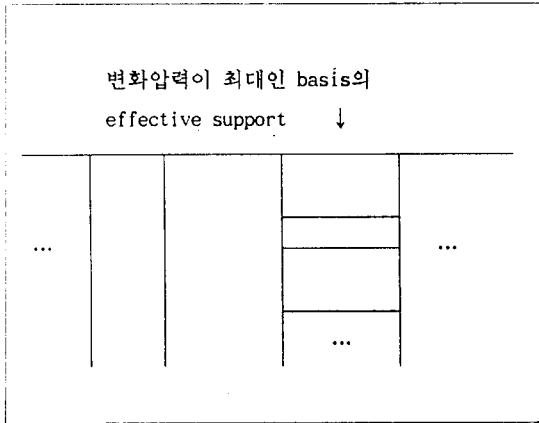


그림 2. Effective support를 이용한 수목 확장 방법 설명

따라서 새롭게 확장된 수목구조 시스템은

(기존의 수목구조) + (변화 압력이 가장 큰 베이스)
* (새롭게 추가된 부수목) (5)

형태의 식으로 표현될 수 있다. (5)식을 전개하면, 각 기본 항은 1개 이상의 변수를 포함하는 가우시안 형태를 띠게 된다.

이때 새롭게 추가되는 부수목(subtree)은, 초기 수목과 마찬가지로 가우시안의 일차 결합(linear combination) 형태를 띠게 된다. 따라서 rL 비트로 이루어지는 이진 스트링을 이용하여, 새롭게 추가되는 부수목이 갖는 베이스 함수의 모양과 개수를 나타낼 수 있다. 추가된 부수목 부분을 최적화 시키는 과정은, 초기 수목을 최적화 시키는 데 사용되었던 알고리즘(modGA를 이용한 수목의 학습)이 사실상 그대로 적용될 수 있다. 다만 단계 3에서 등장하는 “각 개체가 표현하는 수목구조 시스템”은 깊이 1의 초기 수목 혹은 새롭게 추가되는 부수목을 의미하는 대신에, 식 (5)에서 설명된 전체 시스템으로 해석되어야 한다. 그리고, 전체 수목구조 시스템의 출력 노드가 갖는 응답은 (2)식에서와 같이 전체 수목구조가 갖는 기본항 즉 가우시안 베이스에 대한 가중합 혹은 가중평균

형태로 얻어진다. 이와 같은 수목 확장은, 주어진 데이터를 이용한 학습 과정에서 주기적으로 적용하되 전체 시스템의 학습 오차가 충분히 만족스러운 값을 갖게 될 때까지 계속한다.

III. 수목구조 시스템의 근사 능력

RBF 신경망과 퍼지 시스템에 대해서는 제각각, 임의의 주어진 함수에 대하여 충분히 가까운 원소를 제공할 수 있는 근사 능력을 소유하고 있음이 밝혀진 바 있다^{4-6, 21-22}. 본 논문에서는, 제안된 수목구조의 근사 능력을 수목의 깊이가 충분히 깊은 경우와 그렇지 못한 경우로 구분하여 다루도록 한다. 여기에서 수목의 깊이가 충분하다 함은 모든 입력 변수를 포함하는 국소 베이스가 출현할 정도로 수목의 확장이 반복된 경우를 의미하고 그렇지 못한 경우는 수목의 확장이 그 이전에 중단된 경우를 의미한다. 수목의 깊이가 충분히 깊은 경우에는, (5)식의 수목 확장이 반복되면서 EBF 신경망(elliptic-basis-function networks)^[23] 혹은 Wang-Mendel 퍼지 모델^[24]이 전체 수목의 일부분으로 자리잡게 된다. EBF 신경망 및 Wang-Mendel 퍼지 모델의 기본 식은 각각 식 (6), (7)과 같다.

$$\sum_{i=1}^M w_i K\left(\frac{x_1 - z_{i1}}{\sigma_{i1}}, \frac{x_2 - z_{i2}}{\sigma_{i2}}, \dots, \frac{x_r - z_{ir}}{\sigma_{ir}}\right) \quad (6)$$

$$y[1] = \frac{\sum_{l=1}^M y^{-l} \left[\prod_{i=1}^r \exp\left(-\left(\frac{x_i - \bar{x}_i^l}{\sigma_i^l}\right)^2\right) \right]}{\sum_{l=1}^M \left[\prod_{i=1}^r \exp\left(-\left(\frac{x_i - \bar{x}_i^l}{\sigma_i^l}\right)^2\right) \right]} \quad (7)$$

(6)식으로 표현되는 EBF 신경망은 커널함수 K가 적분 가능하고 연속이며 K⁻¹(0)가 입력 변수 축을 포함하지 않는 경우에 주어진 연속 함수를 긴밀한(compact) 입력 공간 위에서 충분히 가깝게 근사할 수 있음^[23]이 보여진 바 있는데, 가우시안 함수는 이 조건을 만족하므로 (6)식의 EBF 신경망을 부분 구조로 갖는 수목구조 시스템은 universal approximator가 된다. 한편, (7)식의 모델 역시 근사 능력을 소유함이 [24]에서 Stone-Weierstrass 정리에 의하여 보여졌다. 따라서, 수목구조의 깊이가 충분히 깊은 경우에는, 가우시안 베이스를 사용하는 수목구조 시스템은 근사 능력

을 소유함을 확인할 수 있다. 반면에, 수목의 깊이가 충분하지 않은 경우에는 근사 능력의 보장 여부가 다분히 회의적이라고 할 수 있겠는데, 이는 본 논문에서 제시하는 수목구조 시스템이 애당초 고차원 공간 위에서 주어진 함수가 입력 변수에 대한 부분 구조의 조합으로 충분히 가깝게 근사될 수 있는 경우를 효과적으로 다루는 것을 목적으로 하고 있기 때문이다. 다음에서는 이러한 예측을 뒷받침하는 이론적 해석을 제공함으로써, 수목구조 시스템이 그 깊이가 충분히 크지 않은 경우에는 universal approximator가 될 수 없음을 밝힌다. 다만, 수목구조의 깊이가 끝까지 즉 입력 공간의 차원인 r 까지 확장될 경우에는, 전체 수목구조는 앞서 설명한 바와 같이 근사 능력을 회복하게 된다.

수목의 깊이가 불충분한 경우에 대한 근사 능력의 결핍을 증명하기 위해서는 counter-example의 제시로 충분하므로 이론 전개의 편의상 입력 공간의 차원이 2이고 수목구조의 깊이가 1인 경우를 생각해 보자. 깊이가 1의 수목구조 시스템이 출력 응답으로 국소 베이시스의 가중합을 사용하는 경우에, 전체 시스템의 출력은 두 입력 변수 x_1, x_2 에 대해 항상 다음과 같은 형태로 변수 분리될 수 있다:

$$f(x_1, x_2) = g(x_1) + h(x_2)$$

이러한 변수 분리는 다음의 보조정리의 성립을 가능하게 해준다.

보조정리: 연속 함수 $g: R \rightarrow R$ 과 $h: R \rightarrow R$ 에 대해서, $f(x_1, x_2) = g(x_1) + h(x_2)$ 를 만족하는 연속 함수 $f: R^2 \rightarrow R$ 를 생각하자. 만약 임의의 $x_1, x_2 \in R$ 에 대해서, $|f(x_1, 0)| < \epsilon$ 과 $|f(0, x_2)| < \epsilon$ 이 만족되면 다음 식은 반드시 성립한다: $|f(x_1, x_2) - f(0, 0)| < 2\epsilon$.

(증명) $|f(x_1, 0)| = |g(x_1) + h(0)| < \epsilon$ 과 $|f(0, x_2)| = |g(0) + h(x_2)| < \epsilon$ 이 성립하므로, triangle inequality에 의하여 다음 식이 성립한다.

$$|f(x_1, x_2) - f(0, 0)| = |g(x_1) + h(x_2) + g(0) + h(0)| \leq |g(x_1) + h(0)| + |g(0) + h(x_2)| < 2\epsilon. \text{ (증명끝)}$$

위의 보조정리를 이용하면 다음의 정리를 얻을 수 있다:

정리: 식 (2-1)로 표현되는 깊이가 1인 수목구조 시스템의 집합 S 는 $C(R^2)$ 위에서 조밀(dense)하지 않다.

(증명): x_1 축과 x_2 축에서 각각 0 값을 갖고, 이들 축 위에 있지 않는 점 (x_1^0, x_2^0) 에서 4를 갖는 임의의 함수 $p \in C(R^2)$ 를 생각하자. 즉,

$$p(x_1, 0) = p(0, x_2) = 0 \quad \forall x_1, x_2. \quad p(x_1^0, x_2^0) = 4.$$

그러면, 함수 p 에 대하여 $\|p - f\|_\infty < 1$ 를 만족하는 수목구조 시스템 $f \in S$ 는 존재하지 않는다. 이 명제를 귀류법을 이용하여 증명하기 위하여, 그러한 f 가 집합 S 안에 존재한다고 가정하면 f 는 x_1 과 x_2 축 그리고 원점 위에서 다음을 만족한다:

$$\begin{aligned} |f(x_1, 0) - p(x_1, 0)| &= |f(x_1, 0)| < 1, \quad \forall x_1 \\ |f(0, x_2) - p(0, x_2)| &= |f(0, x_2)| < 1, \quad \forall x_2 \\ |f(0, 0) - p(0, 0)| &< 1 \end{aligned}$$

이 식들은 보조정리로부터, 다음 식을 제공한다.

$$|f(x_1^0, x_2^0) - p(x_1^0, x_2^0)| < 2.$$

이식에 따르면, $f(x_1^0, x_2^0)$ 의 절대값은 3 이상이 될 수 없는데, 이 사실은 $f(x_1^0, x_2^0) = 4$ 에 모순이 된다. 따라서 $\|p - f\|_\infty < 1$ 를 만족하는 $f \in S$ 는 존재하지 않는다. (증명끝)

위의 정리는 결론적으로 가중합을 이용하는 수목구조는 깊이가 충분하지 못한 경우에, RBF 신경망과 퍼지 시스템이 소유하는 근사 능력을 가질 수 없음을 분명히 보여준다. 따라서, 우리가 근사하고자 하는 함수가 매우 까다로운 경우에는 수목구조 시스템의 깊이가 입력 공간 차원 r 까지 확대될 수밖에 없음을 알 수 있다. 하지만, 많은 공학 문제에서 볼 수 있는 바와 같이 주어진 함수의 입력 변수들이 어느 정도 중복적인 정보들을 포함하고 있는 경우에는, 적절한 정도의 수목 확장으로 충분히 가까운 근사가 가능해지게 된다. 본 논문의 방법론은, 바로 이러한 종류의 문제들

을 curse of dimensionality를 피해 가면서 해결하는 것을 주목적으로 한다.

IV. 시뮬레이션

식 (8)의 지연 미분 방정식의 해로부터 얻어지는 시계열 데이터의 예측 문제인 Mackey-Glass time-series는, 예측 알고리즘의 성능 평가를 위한 주요한 적용 문제의 하나로 인식되어 왔다:

$$\frac{dx(t)}{dt} = \frac{0.2 x(t-\tau)}{1 + x^{10}(t-\tau)} - 0.1 x(t) \quad (8)$$

본 장에서는 수목구조 시스템의 성능 분석을 위하여, τ 가 30일 때 Mackey-Glass 시계열의 입력 벡터 $\{x(t), x(t-1), \dots, x(t-8)\}$ 로 부터 다음 시간 스텝의 출력값 $x(t+1)$ 을 예측하는 문제를 생각한다. 이 문제는 참고문헌 [24-25]에서 퍼지 시스템의 성능 분석을 위하여 사용된 바 있는데, 입력 공간의 차원이 높고 각 입력 변수들이 어느 정도 중복적인 정보를 포함하는 문제를 효과적으로 다루고자 하는 본 논문의 방법론을 평가하는데 적절한 예가 될 것으로 기대된다. 미분방정식 (8)의 해는 4차 Runge-Kutta 방법을 이용한 수치 해석적 풀이로부터 얻어졌으며, 이 과정에서 사용된 시간 단계(time step)와 초기 데이터 $x(0)$ 값으로는 각각 0.1과 0.8이 사용되었다. t 가 100일 때부터 1099일 때까지 1000개의 입출력 쌍이 본 논문의 수목구조 시스템 및 비교 대상 방법론을 위한 학습 및 시험용의 데이터로 추출되었다. 이들 중 $x(100)$ 부터 $x(799)$ 에 이르는 700개의 실수 값을 출력으로 하는 9차원 입력 벡터와 출력 값의 쌍들이 학습용 데이터(training data)로 사용되었으며 $x(800)$ 부터 $x(1099)$ 까지를 출력으로 하는 300개의 입출력 쌍들이 시험용 데이터(checking data)로 사용되었다. 본 논문의 방법론을 적용한 실험에서 사용한 주요 파라미터들은 표 1과 같다.

그리고, 성능 비교의 대상으로는 같은 문제를 다룬 [24-25]의 퍼지 시스템이 선택되었다. 비교 대상으로 사용된 퍼지 시스템의 경우에는, 9개의 입력변수 각각이 7개의 영역을 가지는, 즉 7^9 개의 cell로 구성되는 fuzzy look-up table을 통하여 퍼지 소속함수(membership functions)들이 구성되어 있다. 그림 3에서는 학습 및 시험용 데이터들을 추출하는데 사용한 곡선이

표 1. 수목구조의 학습에 사용된 주요 파라미터

최대 세대 반복 횟수	10
개체 집단의 크기	15
부모 개체의 개수	10
각 변수에 할당된 유전인자의 비트수	10
교배 연산자의 확률	0.3
돌연변이 연산자의 확률	0.03
LMS 규칙을 위한 학습률	0.1
LMS 규칙을 위한 최대 학습 반복 횟수	100
학습 오차를 위한 가중치	1
베이스스 개수를 위한 가중치	0.001
수목 확장 횟수	9

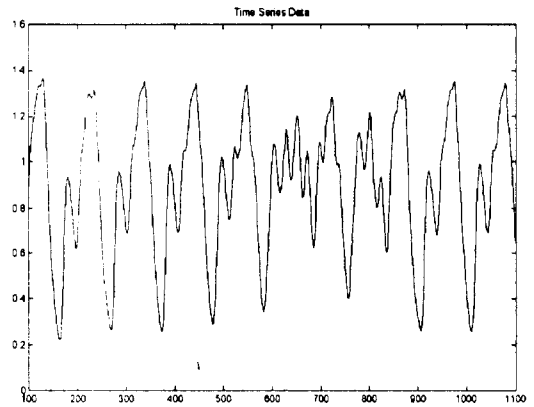


그림 3. 학습 및 시험용 데이터를 추출하는 데 사용한 곡선

보여졌다. 그리고, 가중평균을 이용한 수목구조 시스템과 비교 대상으로 선택된 퍼지 시스템이 제공하는 예측 능력이 각각 그림 4와 5를 통해 요약되었다. 예측 결과들과 그림 3의 실제 데이터 사이에는 매우 미세한 차이만이 존재하므로, 이들의 성능을 시각적으로 비교하는 데에는 이들 그림에서와 같이 잔류오차(residual error), 즉 예측값과 실제값의 차이를 이용하는 것이 효과적이다. 그리고, 전반적인 예측 성능을 하나의 지수로 나타낼 경우에는 다음에서 정의하는 NMSE(normalized mean square error) 값이 널리 쓰인다.

$$NMSE = \frac{E[\hat{x}(t) - x(t)]^2^{1/2}}{E[x(t) - E\{x(t)\}]^2^{1/2}} \quad (9)$$

이들 그림은, 수목구조 시스템이 학습 및 시험 데이

타 구간에서 모두 다 Wang과 Mendel에 의해 제안된 방법 [24-25] 보다 작은 크기의 NMSE와 현격하게 저하된 잔류오차를 제공한다(그림 4의 trn NMSE=0.048104, chk NMSE=0.050159, 그림 5의 trn NMSE=0.084468, chk NMSE=0.079153). 따라서, 본 논문에서 제안한 수목구조 시스템은, 고차원의

입력공간을 갖는 비선형 근사 문제를 다룸에 있어서 주어진 입출력 데이터를 기억하는 학습 능력과 학습에 사용되지 않은 입력이 들어왔을 때 출력을 정확하게 예측할 수 있는 일반화 능력(*generalization capability*)을 고루 갖추고 있는 방법론임을 알 수 있다.

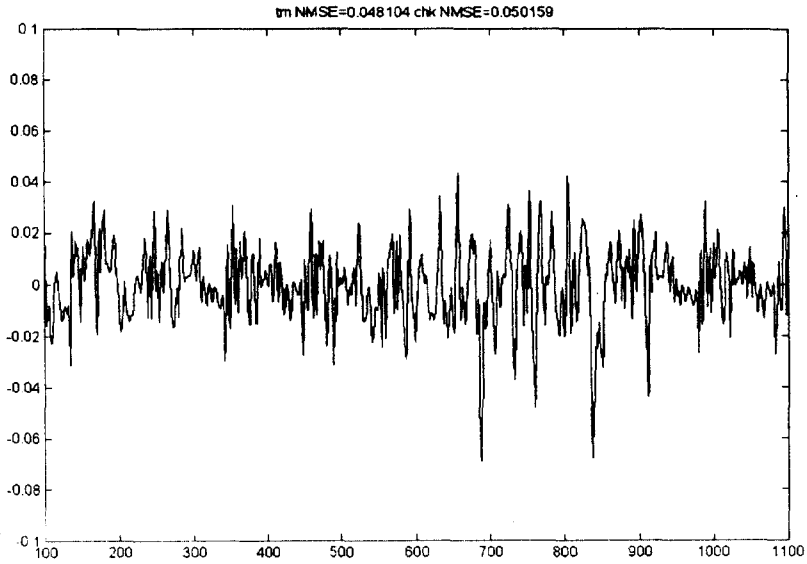


그림 4. 수목구조 시스템을 이용할 때의 잔류오차와 NMSE 값

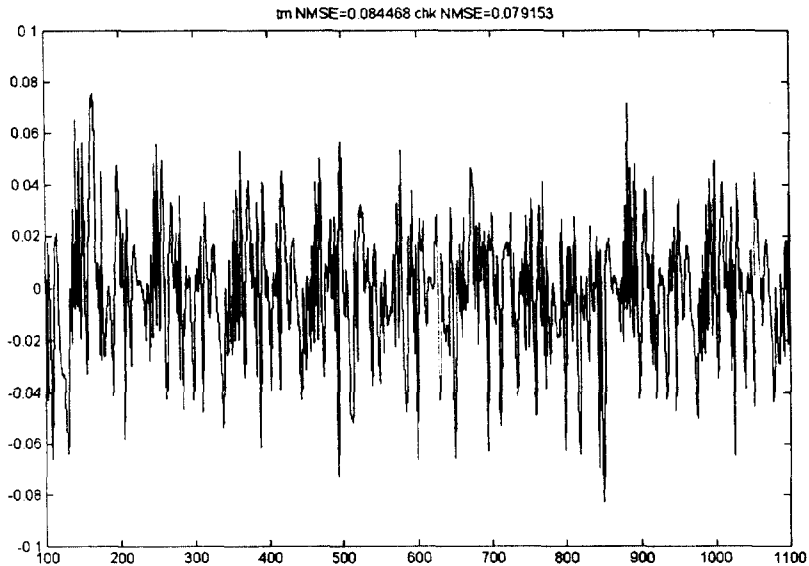


그림 5. [24-25]의 퍼지 시스템을 이용할 때의 잔류오차와 NMSE 값

V. 결 론

본 논문에서는, 고차원 입력 공간 위에서의 비선형 근사를 효과적으로 수행하기 위하여 1) 학습이 진행됨에 따라 구조가 확장되는 수목구조 지능 시스템을 제안하고, 2) 수정 유전자 알고리즘과 LMS 규칙을 이용하여 수목 확장 및 학습을 조절하는 방법론을 제안하였다. 그리고, 제안된 시스템에 관한 이론적 기초를 제공하기 위하여 수목구조 시스템에 관한 근사 능력 분석을 수행하였다. 이에 따르면 수목의 깊이가 충분한 경우, 즉 모든 입력 변수를 포함하는 국소 베이스가 출현할 수 있을 때까지 수목의 확장이 반복될 경우에는 가우시안 베이스를 갖는 수목구조는 universal approximator가 되고, 수목의 깊이가 불충분하고 출력 응답이 베이스 함수의 가중합인 경우에는 변수 분리 현상으로 인하여 그 근사 능력에 한계가 발생함을 확인하였다. 마지막으로, 개발된 방법론이 본 논문에서 목표로 삼는 실제적 문제, 즉 고차원의 입력 변수들이 어느 정도 중복성이 있는 정보를 취하는 경우에 충분한 효과를 발휘하는지를 살펴보기 위하여, Mackey-Glass 시계열 예측 문제를 대상으로 시뮬레이션을 행하였다. 그 결과, 본 논문의 방법론이 주어진 문제를 효과적으로 다룰 수 있음을 알 수 있었다. 앞으로 시도해야 할 향후 연구 과제로, 학습 과정에서 구조를 바꾸는 여타 방법론에 대한 연구 및 비교 분석으로 추가적인 성능 향상을 기하는 것이 필요하다.

감사의 글

이 논문은 1995년도 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

참 고 문 헌

1. G. Cybenko, "Approximation by superposition of a sigmoidal function," *Mathematics of Control, Signals, and Systems*, vol. 2, pp. 303-314, 1989.
2. K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359-366, 1989.

3. Y. Ito, "Approximation of continuous functions on R^d by linear combinations of shifted rotations of a sigmoid function with and without scaling," *Neural Networks*, vol. 5, no. 1, pp. 105-115, 1992.
4. J. Park and I. W. Sandberg, "Universal approximation using radial basis function networks," *Neural Computation*, vol. 3, pp. 246-257, 1991.
5. J. Park and I. W. Sandberg, "Approximation and radial basis function networks," *Neural Computation*, vol. 5, pp. 305-316, 1993.
6. T. Poggio and F. Girosi, "Networks for approximation and learning," *Proceedings of the IEEE*, vol. 78, pp. 1481-1497, 1990.
7. J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Computation*, vol. 1, no. 2, pp. 281-294, 1989.
8. S. Chen, S. A. Billings, C. F. N. Cowan, and P. M. Grant, "Nonlinear systems identification using radial basis functions," *International Journal of Systems Science*, vol. 21, no. 12, pp. 2513-2539, 1990.
9. S. Chen, S. A. Billings, C. F. N. Cowan, and P. M. Grant, "Practical identification of NARMAX models using radial basis functions," *International Journal of Control*, vol. 52, no. 6, pp. 1327-1350, 1990.
10. J. R. Jang and C. Sun, "Functional equivalence between radial basis function networks and fuzzy inference systems," *IEEE Transactions on Neural Networks*, vol. 4, pp. 156-159, 1993.
11. J. Nie and D. A. Linkens, "Learning control using fuzzified self-organizing radial basis function network," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 4, pp. 280-287, 1993.
12. M. Sugeno and T. Yasukawa, "A fuzzy logic based approach to qualitative modeling," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 7-31, 1993.
13. R. Yager, "Structures for building flexible neural fuzzy systems," *IFSA '93, Plenary Talk*, 1993.
14. D. Park, *Intelligent fuzzy reasoning models with*

application to fuzzy control, Ph. D Dissertation, Florida State University, Tallahassee, Florida, 1992.

15. D. Park, A. Kandel, and G. Langholz, "Genetic-based new fuzzy reasoning models with application to fuzzy control," IEEE Transactions on Systems, Man, and Cybernetics, vol. 24, no. 1, pp. 39-47, 1994.

16. M. A. Lee and H. Takagi, "Integrating design stages of fuzzy systems using genetic algorithms," Second IEEE International Conference on Fuzzy Systems, vol. 1, pp. 612-617, 1993.

17. K. C. Ng and Y. Li, "Design of sophisticated fuzzy logic controllers using genetic algorithms," Proceedings of Third IEEE International Conference on Fuzzy Systems, pp. 1708-1712, 1994.

18. T. D. Sanger, "A tree-structured algorithm for reducing computation in networks with separable basis functions," Neural Computation, vol. 3, no. 1, pp. 67-78, 1991.

19. T. D. Sanger, "A tree-structured adaptive network for function approximation in high-dimensional spaces," IEEE Transactions on Neural Networks, vol. 2, no. 2, pp. 285-293, 1991.

20. Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, Second, Extended Edition, Springer-Verlag, New York, 1994.

21. J. J. Buckley, "Sugeno type controllers are universal controllers," Fuzzy Sets and Systems, vol. 53, pp. 299-303, 1993.

22. J. J. Buckley and Y. Hayashi, "Fuzzy input-output controllers are universal approximators," Fuzzy Sets and Systems, vol. 58, pp. 273-278, 1993.

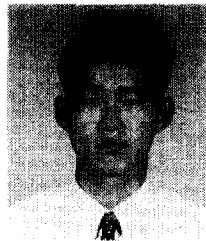
23. J. Park and I. W. Sandberg, "Nonlinear approximation using elliptic basis function networks," Circuits, Systems and Signal Processing, vol. 13, no. 1, pp. 99-113, 1993.

24. L. Wang, Adaptive fuzzy systems and control: Design and stability analysis, Prentice Hall, Englewood Cliffs, New Jersey, 1994.

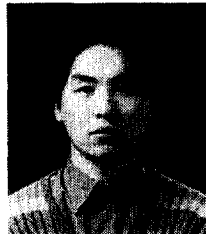
25. L. Wang and J. M. Mendel, "Generating fuzzy rules from numerical data, with applications," IEEE Transactions on Systems, Man, and Cybernetics, vol. 22, no. 6, pp. 1414-1427, 1992.



길 준 민(Joonmin Gil) 정회원
 1994년:고려대 전산학과(학사)
 1996년:고려대 대학원 전산학과(석사)
 1996년~현재:고려대 대학원 컴퓨터과학과 박사과정 재학중
 ※주관심분야: 퍼지시스템, 신경망, 유전자 알고리즘 등임



정 창 호(Changho Jung) 정회원
 1996년:고려대 전산학과(학사)
 1996년~현재:고려대 대학원 전산학과 석사과정 재학중
 ※주관심분야: 퍼지시스템, 신경망, 유전자 알고리즘 등임



강 성 훈(Seonghoon Kang) 정회원
 1995년:고려대 제어계측공학과(학사)
 1995년~현재:고려대 대학원 정보공학과 제어계측전공 석사과정 재학중
 ※주관심분야: 제어, 퍼지이론 및 유전자 알고리즘 등임



박 주 영(Jooyoung Park) 정회원
 1983년:서울대 전기공학과(학사)
 1992년:텍사스대 전기 및 컴퓨터공학과(박사)
 1993년~현재:고려대 제어계측공학과 조교수
 ※주관심분야: 신경망, 제어 및 시스템 이론 등임



박 대 희(Daihee Park) 정회원

1982년: 고려대 수학과(학사)

1984년: 고려대 수학과(석사)

1989년: 플로리다 주립대 전산학
과(석사)

1992년: 플로리다 주립대 전산학
과(박사)

1993년~현재: 고려대 전산학과 조
교수

※주관심분야: 퍼지시스템, 신경망, 유전자 알고리즘,
데이터베이스 등임