

## 실수코딩 유전알고리즘을 이용한 시스템 식별

안중갑\* · 이윤형\*\* · 진강규\*\*\* · 소명옥†

(원고접수일 : 2007년 2월 28일, 심사완료일 : 2007년 5월 13일)

### System Identification by Real-Coded Genetic Algorithm

Jong-Kap Ahn\* · Yun-Hyung Lee\*\* · Gang-Gyoo Jin\*\*\* · Myung-Ok So†

**Abstract** : This paper presents a method for identifying various systems based on input-output data and a real-coded genetic algorithm(RCGA). The advantages of this technique are, first, it is not dependent on the deterministic or stochastic nature of the systems and, second, the globally optimized models for the original systems can be identified without the need of a differentiable measure function of linearly separable parameters. Under suitable hypotheses, the estimation error is shown to converge in probability to zero. The performance of the proposed algorithm is demonstrated through several simulations.

**Key words** : System identification(시스템 식별), Parameter estimation(파라미터 추정), Real-coded genetic algorithm(실수코딩 유전알고리즘)

#### 1. 서 론

시스템 식별의 방법에는 모델의 구조를 가정하는 경우와 그렇지 않은 경우로 구분할 수 있다. 전자는 식별하고자 하는 시스템의 구조를 가정하고 파라미터 또는 파라미터와 차수를 추정하는 방법이고, 후자는 입·출력 신호로부터 모델의 임펄스 응답, 주파수 응답 등을 구해 직접적으로 식별하는 방법이다<sup>1)</sup>.

시스템의 구조는 일반적으로 수학적 모델링 등을 통해 알 수 있는 경우가 많으므로, 본 연구에서는 구조가 알려진 시스템의 파라미터를 얼마나 효과적

이고 정확하게 추정할 수 있는지에 초점을 두었다. 다음과 같은 입·출력 시스템을 고려해보자.

$$G : U \rightarrow Y$$

여기서  $G$ 는 시스템,  $U$ 는 입력,  $Y$ 는 출력이다. 식별하고자 하는 시스템을  $\hat{G}$ 라 하면, 우리의 관심사는  $\hat{G}$ 를  $G$ 에 근접시키는 것으로서 (1)로 나타나는 에러의 최소화이다.

$$e(t) = (Gu)(t) - (\hat{G}u)(t) \quad (1)$$

$e(t)$ 를 최소화하기 위한 기존의 방법에는 확률적

† 교신저자(한국해양대학교 선박전자기계공학부), E-mail: smo@hhu.ac.kr, Tel:051)410-4248

\* 한국항공우주연구원

\*\* 한국해양대학교 대학원 메카트로닉스공학과

\*\*\* 한국해양대학교 IT공학부

인 방법, 최소자승법, Maximum-likelihood(ML)방법, Instrumental variable(IV)방법 등이 있다. 그러나 이와 같은 기존의 방법은 완전한 탐색공간에서만 그 성능이 보장되며, 만약 탐색공간이 완전하더라도 식별하고자 하는 시스템이 고차인 경우에는 적용하기 어렵다. 또한 실제 추정 환경에서는 잡음, 모델의 부정확성, 탐색구역의 비선형성 등으로 인해 탐색 공간이 다봉(Multimodal)이 되어 기존의 추정 방법은 지역해(Local solution)에 수렴할 가능성이 높다<sup>[2]</sup>.

따라서 본 연구에서는 최근 강인한 탐색법으로 많이 사용하는 실수코딩 유전알고리즘(Real-coded Genetic Algorithm:RCGA)을 이용하여  $e(t)$ 를 최소화하는 문제를 다루고자 하며, 여러 가지 종류의 시스템에 대해 시뮬레이션을 실시하여 적용한 방법의 유효성을 입증한다.

## 2. 실수코딩 유전알고리즘

유전알고리즘(Genetic Algorithm:GA)은 1975년 J. H. Holland<sup>[3]</sup>교수에 의해 처음 소개된 이후, Goldberg가 광범위하게 전파했으며, De Jong이 잠재적인 여러 탐색방향, GA의 성능에 영향을 미치는 식별 인자들의 영향력들을 요약 하였다. Fig. 1은 유전알고리즘의 동작과정을 보여 주고 있다. 유전알고리즘의 탐색 과정은 크게 집단의 초기화, 적합도 평가, 재생산, 교배, 돌연변이의 5단계로 구분된다.

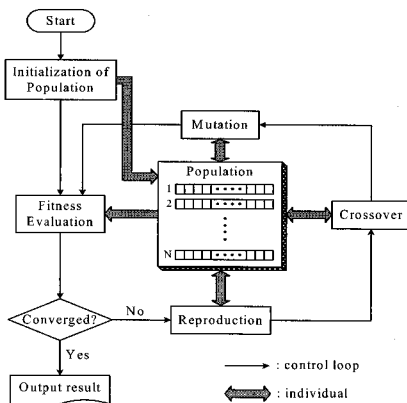


Fig. 1 Basic structure of a GA

### 2.1 파라미터의 표현

본 연구에서는 변수(표현형)와 유전자(유전자형) 간에 일대일 일치로 프로그래밍이 간편하고 부호화 및 복호화 과정이 필요 없어 탐색속도를 높일 수 있는 RCGA를 사용하기로 한다.

Fig. 2는 실수코딩 염색체의 한 예를 보여주고 있다.

$$x = [10 \ 5 \ 2]^T \quad \Leftrightarrow \quad s = (10 \ 5 \ 2)$$

Fig. 2 A Real-coded chromosome

### 2.2 초기 집단의 생성

$k$ 세대에서의 집단  $P(k)$ 는 염색체로 특징 지워지는 다수 개체들의 집합으로 (2)와 같이 정의된다.

$$P(k) = \{s_1(k), s_2(k), \dots, s_i(k), \dots, s_N(k)\} \quad (2)$$

여기서  $s_i(k)$ 는  $i$ 번째의 염색체로서 탐색 공간상의 한 점을 나타내며,  $N(>1)$ 은 집단크기(Population Size)를 의미한다.

### 2.3 유전 연산자

#### 2.3.1 재생산

재생산(Reproduction)은 자연환경에 더 적합한 생물이 더 높은 생존 기회를 갖는 자연도태(Natural selection)현상을 모방하려는 인위적인 메카니즘으로서, 적합도값을 기반으로 현재 집단 내의 개체들을 선택하여 새로운 집단을 형성해준다. 이러한 선택은 약한 개체들을 집단에서 몰아내고 강한 개체를 선택함으로써 유전자들이 이후 세대의 집단에 넓게 퍼지도록 하여 전체집단을 더욱 강하게 해주는 특징을 가진다. 본 논문에서는 개체들의 복제를 가급적 피하면서 집단을 최적 개체쪽으로 유도하는 구배와 유사한 재생산(Gradient-like reproduction)을 사용하였다.

#### 2.3.2 교배

교배(Crossover)는 자연계 생물들의 유성생식(Sexual mating)을 모방하려는 인위적인 메카니

즘이다. 본 논문에서는 이진코딩의 일점교배를 아주 근접하게 묘사한 것으로서  $n$ 번째 요소가  $\lambda$ 에 의해 일차 결합되고 그 후의 요소들은 서로 교환되어 자손이 생산되는 Fig. 3과 같은 수정단순교배 (Modified simple crossover)를 사용하였다.

$$\begin{aligned}
 S_1 &= (c_1^1 \ c_2^1 \ \dots \ c_{n-1}^1 \ c_n^1 \ c_{n+1}^1 \ \dots \ c_l^1) \\
 &\rightarrow S_1' = (c_1^1 \ c_2^1 \ \dots \ c_{n-1}^1 \ d_n^1 \ c_{n+1}^1 \ \dots \ c_l^1) \\
 S_2 &= (c_1^2 \ c_2^2 \ \dots \ c_{n-1}^2 \ c_n^2 \ c_{n+1}^2 \ \dots \ c_l^2) \\
 &\rightarrow S_2' = (c_1^2 \ c_2^2 \ \dots \ c_{n-1}^2 \ d_n^2 \ c_{n+1}^2 \ \dots \ c_l^2)
 \end{aligned}$$

단,  $d_n^1 = \lambda c_n^1 + (1-\lambda)c_n^2$ ,  $d_n^2 = \lambda c_n^2 + (1-\lambda)c_n^1$ ,  $\lambda$ 는 곱인수로서  $[0, 1]$ 사이의 난수이다.

Fig. 3 Modified simple crossover

2.3.3 돌연변이

유전자의 다양성 결핍으로 지역해(Local solution)나 사점(Dead corner)에 빠지게 되는데 이때 원치 않는 해로부터 벗어나기 위한 메카니즘이 돌연변이(Mutation)이다. 본 논문에서는 정밀도를 높이기 위해 미세 조정이 가능하도록 고안된 동적 돌연변이(Dynamic mutation)를 사용하였다. 이것의 특징은 초기에는 전 공간을 균등한 확률로 탐색하다가 세대수가 증가하면 매우 지역적으로 탐색하도록 해준다.

$$\tilde{s} = (\tilde{x}_1 \ \tilde{x}_2 \ \dots \ \tilde{x}_j \ \dots \ \tilde{x}_n) \rightarrow s = (x_1 \ x_2 \ \dots \ x_j \ \dots \ x_n)$$

Fig. 4 Dynamic mutation

Fig. 4에서  $j$ 번째 유전자에서 돌연변이가 일어나면  $x_j$ 는 (3)으로부터 결정된다.

$$x_j = \begin{cases} \tilde{x}_j + \Delta(k, x_j^{(U)} - \tilde{x}_j), & \tau = 0 \\ \tilde{x}_j - \Delta(k, x_j - x_j^{(L)}), & \tau = 1 \end{cases}$$

$$\Delta(k, y) = y \cdot r \cdot (1 - \frac{k}{T})^b \tag{3}$$

여기서  $r$ 은 0과 1사이의 실수 난수,  $T$ 는 최대세

대수,  $b$ 는 불균등 매개변수로서 사용자에 의해 결정된다. 이 함수는 영역  $(0, y)$ 에서 값을 가지는데, 세대수  $k$ 가 증가함에 따라 돌연변이가 발생할 확률이 낮아진다.

2.4 적합도 평가

자연계에서 생물들은 끊임없이 환경에 적응해 가는데 이 적응능력이 유전알고리즘에서는 개체의 적합도를 평가하는 과정을 통해 반영된다. 새로운 집단이 완성되고 개체들의 적합도가 목적함수에 의해 계산되며 적합도의 평가는 새로운 집단이 완성될 때마다 계속된다. 이 과정에서 적합도가 큰 개체가 더 많은 보상을 받을 수 있도록 배려하기 위해서 적합도 함수는 최대화 문제 형태로 표현한다. 본 논문에서는 최소화 문제로 기술되는 목적함수는 (4)와 같이 사상하여 최대화 문제 형태로 표현한다.

$$f(s) = -F(x) - F_{min} \tag{4}$$

여기서  $f(s)$ 는 적합도 함수,  $F(x)$ 는 목적함수,  $F_{min}$ 은  $f(s) \geq 0$ 을 보장하는 상수이다.

또한, 본 논문은 초기 세대에서는 우량 개체의 적합도 영향을 줄일 수 있도록 개체들의 적합도 값을 넓은 범위의 값 위로 분포시키고, 후기 세대에는 더 적합한 개체들이 더 많이 선택될 수 있도록 분포시키는 선형 스케일링(Linear scaling)을 사용한다. 그리고 탐색성능을 높이기 위해 이전 세대의 최적 개체를 저장하고 있다가 재생산-교배-돌연변이 연산을 거쳐 현 세대의 집단이 완성되면 최적 개체의 생존 유무를 확인하고, 만일 최적 개체가 소멸되었으면 저장된 개체를 현 세대 집단 안에서 적합도가 가장 낮은 개체와 교환해주는 엘리티즘 전략을 사용하였다.

3. 시스템 식별

여기서는 시스템을 식별하기 위해 2장에서 설명한 RCGA를 사용한다. RCGA의 제어변수들은 문헌<sup>(4)</sup>에서 알려진 일반적인 결과들을 바탕으로 실험과 경험을 통해 선정되었다.

(5)와 같은 일반적인 비선형 IIR 시스템을 고려하자.

$$y(k) = f(W, Y, U) \tag{5}$$

$W = [w_1, \dots, w_h]$  :  $h$  고정 파라미터  
 $Y = [y(k-1), \dots, y(k-n)]$  :  $n$  자동회귀 간격  
 $U = [u(k-1), \dots, u(k-m)]$  :  $m$  시간 전 입력값

여기서  $f(\cdot)$ 는 비선형 시스템으로서 리아푸노프 관점에서 안정하다.

(6)과 같이 선정된 조정모델의 파라미터는 비선형 시스템 (5)의 출력과 조정모델의 출력이 일치하도록 Fig. 5와 같은 방법에 의해 추정된다.

$$\hat{y}(k) = f(\hat{W}, \hat{Y}, U) \tag{6}$$

$\hat{W} = [\hat{w}_1, \dots, \hat{w}_h]$  :  $h$  파라미터 추정  
 $\hat{Y} = [\hat{y}(k-1), \dots, \hat{y}(k-n)]$  :  $n$  추정된 출력  
 $U = [u(k-1), \dots, u(k-m)]$  :  $m$  시간 전 입력값

RCGA로 추정하고자 하는 유전자 파라미터가  $\hat{w}_i$ 이고,  $\hat{W}$ 가 염색체다.

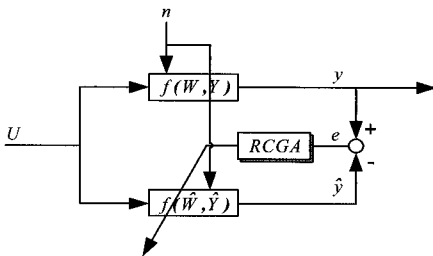


Fig. 5 RCGA-based system identification

개체의 적합도를 평가하기 위해 (7)과 같은 목적함수를 사용하였으며,

$$J = \frac{1}{d} \sum_{k=1}^d \{y(k) - \hat{y}(k)\}^2 \tag{7}$$

RCGA는 궁극적으로 시스템과 조정모델 간에 출력차이가 최소가 되도록 시스템의 파라미터  $[\hat{w}_1, \hat{w}_2, \dots, \hat{w}_h]$ 를 탐색하게 된다. 목적함수  $J$ 의 가장 이상적인 값은 0이 되는 것이다. 이제 몇 가지 예를 통해 지금까지 설명한 방법의 유효성을 살펴

보기로 한다.

### 3.1 선형시스템 식별

(8)과 같은 알려지지 않은 선형시스템<sup>[5]</sup>의 파라미터 추정에 RCGA를 사용한다.

$$y(k) = -0.3y(k-1) + 0.4y(k-2) + 1.25u(k-1) - 2.5u(k-2) + n(k) \tag{8}$$

시스템의 입력신호는 난수 발생기를 통해 [-2.5, 2.5]사이의 임의의 난수로 하고, 잡음은 평균이 0이고 분산이 1인 가우시안 백색 잡음으로 한다.

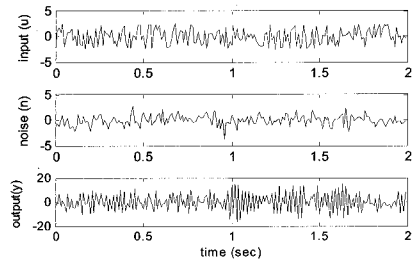


Fig. 6 Plots of linear system

파라미터 추정에 사용된 조정모델은 (9)와 같다.

$$\hat{y}(k) = a_1 \hat{y}(k-1) + a_2 \hat{y}(k-2) + b_1 u(k-1) + b_2 u(k-2) + n(k) \tag{9}$$

개체의 적합도 평가를 위한 목적함수는 (10)과 같다.

$$J = \frac{1}{d} \sum_{k=1}^d [y(k) - a_1 \hat{y}(k-1) - a_2 \hat{y}(k-2) - b_1 u(k-1) - b_2 u(k-2) - n(k)]^2 \tag{10}$$

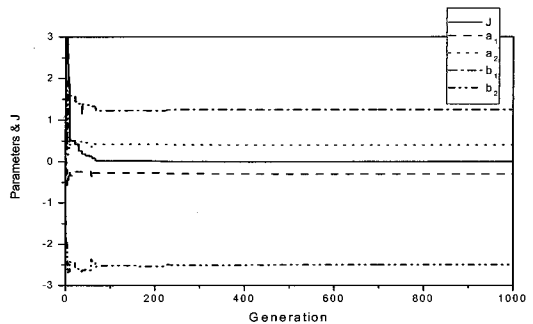


Fig. 7 Parameters estimation of linear system

**Table 1** Estimated parameters of linear system

Parameters	Mean value	Standard deviation
$a_1$	-0.300000	0
$a_2$	0.400000	0
$b_1$	1.250000	0
$b_2$	-2.500000	0

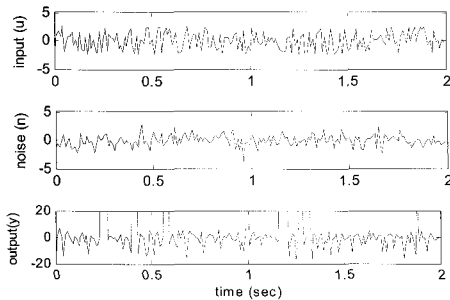
미지의 파라미터는 구간  $-3 \leq a_i \leq 3$ ,  $-3 \leq b_i \leq 3$ 에서 탐색된다. 이때 RCGA의 제어변수로는 집단의 크기  $N=50$ , 재생산계수  $\eta=1.8$ , 교배확률  $P_c=0.9$ , 돌연변이 확률  $P_m=0.2$ 가 사용되었다. Table 1은 독립된 씨드(Seed)로 프로그램을 5번 실행하고 구한 값들의 평균과 표준편차를 나타낸다. Fig. 7은 RCGA가 파라미터를 추정하는 하나의 예를 나타낸 것이다.

**3.2 비선형시스템 식별**

다음은 (11)과 같은 알려지지 않은 비선형시스템<sup>[5]</sup>의 파라미터 추정 문제를 고려해 본다.

$$y(k) = \left( \frac{3 - 0.3y(k-1)u(k-2)}{5 + 0.4y(k-2)u^2(k-1)} \right)^2 + (1.25u^2(k-1) - 2.5u^2(k)) \ln(|1.25u^2(k-2) - 2.5u^2(k)|) + n(k) \tag{11}$$

시스템의 입력신호는 난수 발생기를 통해 [-2.5 2.5]사이의 임의의 난수로 하고, 잡음은 평균이 0이고 분산이 1인 가우시안 백색 잡음으로 한다.



**Fig. 8** Plots of nonlinear system

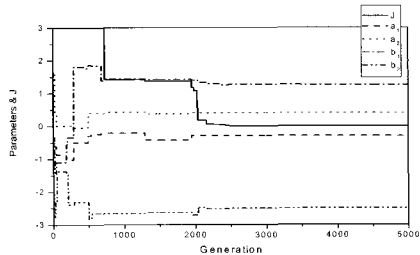
파라미터 추정에 사용된 조정모델은 (12)와 같다.

$$\hat{y}(k) = \left( \frac{3 + a_1 \hat{y}(k-1)u(k-2)}{5 + a_2 \hat{y}(k-2)u^2(k-1)} \right)^2 + (b_1 u^2(k-1) + b_2 u^2(k)) \ln(|b_1 u^2(k-2) + b_2 u^2(k)|) + n(k) \tag{12}$$

개체의 적합도 평가를 위한 목적함수는 (13)과 같다.

$$J = \frac{1}{d} \sum_{k=1}^d [y(k) - \left( \frac{3 + a_1 \hat{y}(k-1)u(k-2)}{5 + a_2 \hat{y}(k-2)u^2(k-1)} \right)^2 - (b_1 u^2(k-1) + b_2 u^2(k)) \ln(|b_1 u^2(k-2) + b_2 u^2(k)|) - n(k)]^2 \tag{13}$$

미지의 파라미터는 구간  $-3 \leq a_i \leq 3$ ,  $-3 \leq b_i \leq 3$ 에서 탐색된다. 이때 RCGA의 제어변수로는 집단의 크기  $N=50$ , 재생산계수  $\eta=1.8$ , 교배확률  $P_c=0.9$ , 돌연변이 확률  $P_m=0.05$ 가 사용되었다. Fig. 9는 파라미터를 추정하는 하나의 예를 보여주고 있다.



**Fig. 9** Parameters estimation of nonlinear system

Table 2는 독립된 씨드(Seed)로 프로그램을 5번 실행하고 구한 값들의 평균과 표준편차를 나타낸다.

**Table 2** Estimated parameters of nonlinear system

Parameters	Mean value	Standard deviation
$a_1$	-0.2999702	3.2018745e-5
$a_2$	0.3999986	7.6026311e-6
$b_1$	1.2459396	0.0068607
$b_2$	-2.4969376	0.0046420

3.3 불안정시스템 식별

마지막으로 (14)와 같은 불안정시스템<sup>(6)</sup>의 파라미터 추정 문제를 생각해 보자.

$$y(k) = \frac{1}{1 - 1.5y(k-1) + 1.5y(k-2)} u(k) + e(k) \quad (14)$$

시스템의 입력신호는 (15)와 같은 제어입력을 사용한다.

$$u(k) = r(k) + 0.95y(k-2) \quad (15)$$

여기서  $r(k)$ 는 평균이 0이고 분산이 1인 백색잡음이고,  $e(k)$ 는 평균이 0이고 분산이 0.01인 백색잡음이다.

추정하고자하는 시스템은 (16)과 같다.

$$\hat{y}(k) = \frac{b_1}{1 + a_1 \hat{y}(k-1) + a_2 \hat{y}(k-2)} u(k) + e(k) \quad (16)$$

이때 사용한 목적함수는 (17)과 같다.

$$J = \frac{1}{d} \sum_{k=1}^d [y(k) - \frac{b_1}{1 + a_1 \hat{y}(k-1) + a_2 \hat{y}(k-2)} u(k) - e(k)]^2 \quad (17)$$

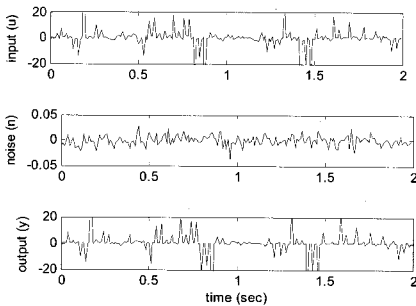


Fig. 10 Plots of unstable system

미지의 파라미터는 구간  $-2 \leq a_i \leq 2$ ,  $-2 \leq b_i \leq 2$ 에서 탐색된다. 이때 RCGA의 제어변수로는 집단 크기  $N=50$ , 재생산계수  $\eta=1.8$ , 교배확률  $P_c=0.9$ , 돌연변이 확률  $P_m=0.2$ 가 사용되었다.

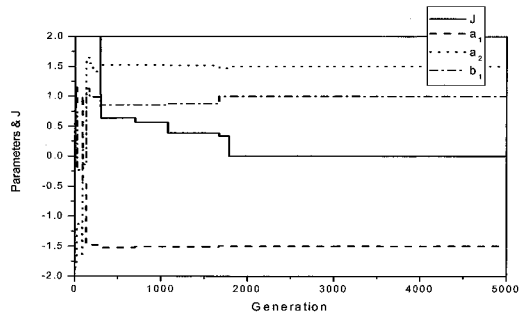


Fig. 11 Parameters estimation of unstable system

Fig. 11은 파라미터를 추정하는 하나의 예를 보여주고 있다.

Table 3 Estimated parameters of unstable system

Parameters	Mean value	Standard deviation
$a_1$	-1.4999974	2.3021728e-6
$a_2$	1.4999894	1.0876580e-5
$b_1$	0.9999770	3.1937438e-5

Table 3은 독립된 씨드(Seed)로 프로그램을 5번 실행하고 구한 값들의 평균과 표준편차를 나타낸다.

4. 결 론

본 연구에서는 유전알고리즘의 한 종류인 RCGA를 사용하여 선형시스템, 비선형시스템 그리고 불안정시스템의 파라미터를 추정하는 문제를 다루었다. 선형시스템의 경우 목적함수가 0으로, 실제시스템과 동일한 파라미터를 탐색하였고 비선형시스템과 불안정시스템의 경우에도 목적함수가 거의 0으로 실제시스템에 가까운 파라미터를 탐색하였다.

이 후에는 파라미터 탐색에 필요한 적절한 집단수, 적절한 교배 및 돌연변이 확률에 대한 연구, 그리고 실제 플랜트에서 데이터를 취득하여 시스템을 식별하는 연구가 진행되어야 할 것이다.

참고문헌

- [1] 하주식, 시스템동정과 제어, 해양문화사, 2002.
- [2] K.Kristinsson, G.A. Dumont, System identification and control using genetic algorithms, IEEE Transactions on Systems, Man, and Cybernetics 22 (5) (1992) 1033 - 1046.
- [3] J. H. Holland, "Adaptation in Natural and Artificial Systems," The University of Michigan Press, Michigan, 1975.
- [4] 진강규, 유전알고리즘과 그 응용, 교우사, 2004.
- [5] Leehter Yao, Willian A. Sethares, "Nonlinear Parameter Estimation via the Genetic Algorithm", IEEE Transactions on signal processing, Vol.42, no.4, April, 1994.
- [6] Urban Forssell, Lennart Ljung, "Identification of Unstable Systems Using Output Error and Box- Jenkins Model Structures", IEEE Transactions on Automatic Contrl, Vol.45, no.1, January, 2000.



**진강규(陳康奎)**

1977년 한국해양대학교 기관학과(학사). 1985년 Florida Institute of Technology, 전기·전자·컴퓨터공학과(석사). 1996년 University of Wales Cardiff, 전기·전자·시스템공학과 박사. 현재 한국해양대학교 IT 공학부 교수.  
 관심분야 : 유전알고리즘, 퍼지제어, 최적제어



**소명옥(蘇明玉)**

1980년 한국해양대학교 기관학과(학사). 1989년 한국해양대학교 대학원 기관학과 석사. 1997년 동 대학원 박사. 한국 항만·운송노동연구원 연구위원. 2005~2006년 University of Wales Cardiff 파견교수. 현재 한국해양대학교 선박 전자기계공학부 교수.  
 관심분야 : 퍼지제어, 최적제어

저 자 소 개



**안중갑(安鍾甲)**

2000년 한국해양대학교 기관시스템공학부(공학사). 2006년 한국해양대학교 대학원 메카트로닉스공학과(석사). 2006~동 대학원 메카트로닉스 공학과 박사과정. 현재 한국항만연수원 부교수.  
 관심분야 : 제어이론, 유전알고리즘 응용



**이윤형(李潤炯)**

2002년 한국해양대학교 기관시스템공학부(공학사). 2004년 한국해양대학교 대학원 기관시스템공학과(석사). 현재 한국해양대학교 대학원 메카트로닉스 공학과 (박사과정).  
 관심분야 : 제어이론, 유전알고리즘 응용