

## 상수처리 수질제어를 위한 약품주입 자동연산

°하대원\*, 고백범\*, 황희수\*\*, 우광방\*  
 \*연세대학교 전기공학과, \*\*한국고속철도건설공단

### Optimum Chemicals Dosing Control for Water Treatment

°Daewon Ha\*, Taekbeom Kho\*, Heesoo Hwang\*\*, Kwangbank Woo\*  
 \*Dept. of Electrical Engineering, Yonsei University  
 \*\*Dept. of Korea High-Speed Rail Constr. Authority

#### ABSTRACT

This paper presents a neuro-fuzzy modelling method that determines chemicals dosing model based on historical operation data for effective water quality control in water treatment system and calculates automatically the amount of optimum chemicals dosing against the changes of raw water qualities and flow rate. The structure identification in the modelling by means of neuro-fuzzy reasing is performed by Genetic Algorithm(GA) and Complex Method in which the numbers of hidden layer and its hidden nodes, learning rate and connection pattern between input layer and output layer are identified. The learning network is implemented utilizing Back Propagation(BP) algorithm. The effectiveness of the proposed modelling scheme and the feasibility of the acquired neuro-fuzzy network is evaluated through computer simulation for chemicals dosing control in water treatment system.

#### 1. 서론

본 논문에서는 상수처리 시스템의 효율적인 수질제어를 위하여 과거의 운전실적 데이터를 토대로 약품주입모델을 결정하고, 원수 수질 및 유량 변화에 대응한 최적의 약품 주입율을 자동연산하는 뉴로-퍼지 모델링 방법을 제시한다. 뉴로-퍼지 추론에 의한 모델링에서 네트워크의 중간층 수, 중간층 노드 수, 중요도가 높은 입력 퍼지 뉴런의 출력이 직접 출력층으로 연결되도록 하는 연결 패턴과 최적의 학습율을 결정하는 네트워크의 구조 동정을 위해 유전 알고리즘과 콤플렉스 방법이 사용되었고, 네트워크의 연결강도와 소속 함수의 파라미터 동정을 위해 역전파알고리즘이 사용되었다. 그리

고, 제안된 모델링방법의 효율성과 동정된 뉴로 퍼지 네트워크의 유용성을 정수장의 약품주입 운전 실적과 비교하여 평가분석된다.

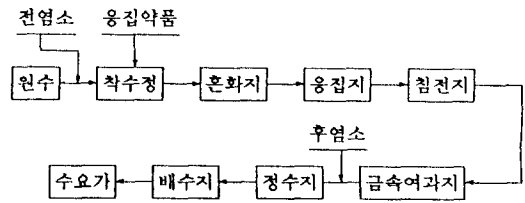


그림 1. 일반적인 정수처리 공정도

#### 2. 정수 처리 공정

현재 국내의 일반적인 정수처리 공정은 그림 1과 같이 취수, 침전, 여과, 살균 소독처리과정을 거치며, 이 가운데 약품 주입에 의한 응집, 침전 및 소독 처리는 상수처리의 수질 제어에서 가장 핵심적인 부분이다. 상수처리 약품 주입공정은 크게 응집제 및 알칼리제 주입과 전염소, 후염소 처리등으로 구분되고, 그 특성으로는 1) 원수 수질 측정( 탁도, 알칼리도, pH, 수온등)의 복잡성과 다양성, 2) 극심한 부하 변화(기후, 시간 및 계절에 따른 원수 수량과 수질의 큰 변화), 3) 비선형 다변수 시스템, 4) 제어량과 조작량의 선택, 조합의 불명확성등을 들 수 있다. 상수 처리 시스템에서 기존의 약품주입 제어는 운전자의 수동 조작에 의한 주입율 설정에 기초하여 주입량을 제어하고 있으며, 수질, 기후등에 상당한 영향을 받으므로 운전원의 기교와 경험에 크게 의존하고 있는 실정이다. 비선형 다변수의 공정특성을 가진 복잡한 약품주입 공정을 대상으로 운전자의 수동조작을 대신하여 약품주입율을 자동으로 연산하는 방법에는 과거의 운전실적

데이터를 토대로한 수학적 모델식을 구하는 방법이 있지만, 수학적 모델식이 기후, 시각, 계절에 따르는 원수유량, 수질의 큰 변화에 정확하고 효과적으로 악플주입 공정을 기술하지 못한다.

### 3. 악플 주입 공정의 뉴로 퍼지 모델링

뉴로-퍼지 네트워크를 이용한 모델링의 전체 구성도는 그림 2와 같다.

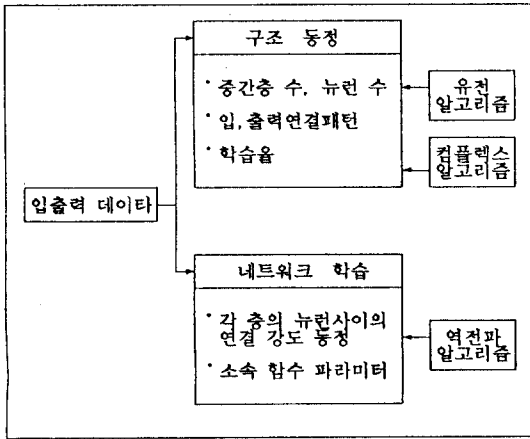


그림 2. 뉴로-퍼지 네트워크를 이용한 모델링

뉴로-퍼지 네트워크는 입력이 네트워크의 입력층 퍼지 뉴런에 전달되고, 퍼지 뉴런 출력은 네트워크 중간층 뉴런에 입력된다. 네트워크의 출력층 뉴런은 중간층 뉴런의 출력과 입력층 뉴런의 출력을 입력으로 받아들이고, 출력층 뉴런의 출력이 뉴로-퍼지 네트워크의 추론값으로 된다. 그림 3에 뉴로-퍼지 네트워크의 추론 구조를 보이며 그 뉴런의 입, 출력관계는 다음과 같다.

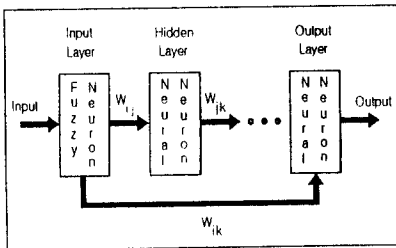


그림 3. 뉴로-퍼지 네트워크의 추론 구조

뉴로-퍼지 네트워크의 중간층과 출력층에서 뉴런 모델의 입, 출력 관계는 다음과 같다.

$$y_i = f(\sum W_{ij}x_j + W_{oi}) \quad (1)$$

$$f(z) = 1 / (1 + e^{-z}) \quad (2)$$

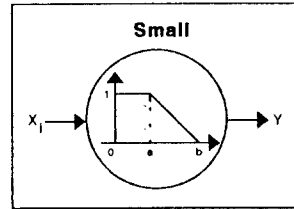
여기서  $W$ 는 연결 강도,  $f(\cdot)$ 는 시그모이드 함수,  $N$ 은 뉴런 수이다.

입력층의 퍼지 뉴런 모델은 그림 4와 같고 그 입, 출력 관계는 다음과 같이 기술된다.

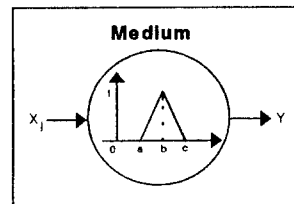
$$\text{Small} \begin{cases} x_j \leq a & y_i = 1.0 \\ x_j > b & y_i = 0.0 \\ a < x_j \leq b & y_i = (x_j - a) / (b - a) + 1.0 \end{cases}$$

$$\text{Medium} \begin{cases} x_j \leq a & y_i = 0.0 \\ x_j > c & y_i = 0.0 \\ a < x_j \leq b & y_i = (x_j - a) / (b - a) \\ b < x_j \leq c & y_i = (b - x_j) / (c - b) + 1.0 \end{cases}$$

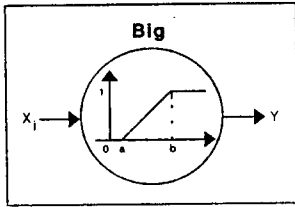
$$\text{Big} \begin{cases} x_j \leq a & y_i = 0.0 \\ x_j > b & y_i = 1.0 \\ a < x_j \leq b & y_i = (x_j - a) / (b - a) \end{cases}$$



a) Small을 정의하는 퍼지 뉴런



b) Medium을 정의하는 퍼지 뉴런



c) Big을 정의하는 퍼지 뉴런

그림 4. 입력층 퍼지 뉴런 모델

예를들어 그림 5는 2층의  $3 \times 5 \times 1$  뉴로 퍼지 네트워크를 나타낸다.

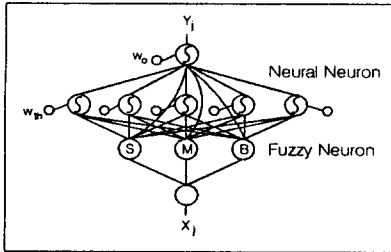


그림 5. 2층  $[3 \times 5 \times 1]$ 의 뉴로 퍼지 네트워크

뉴로-퍼지 네트워크에 의해 시스템을 모델링하는 것은 입력 출력 데이터를 최적으로 학습할 수 있는 네트워크의 구조와 그 네트워크를 구성하는 뉴런의 연결강도를 동정하는 것이다. 네트워크의 구조 동정에서는 자연 도태와 유전적 성질에 기초한 유전 알고리즘에 의해 (1) 중간층 수, (2) 중간층 뉴런수(이는 퍼지 규칙수에 해당됨), (3) 역전파 알고리즘의 학습을 및 모멘텀 계수, (4) 퍼지 뉴런에 직접 연결되는 연결패턴을 찾아낸다.

유전 알고리즘은 복잡한 최적화 문제에서 뛰어난 강인성을 발휘하여 기존의 최적화 기법에 비해서 다음과 같은 점에서 차이가 난다.

(1) 유전 알고리즘은 최적화 문제의 파라미터 집합을 알파벳으로 구성된 유한길이의 문자열로 코딩한다. 그 결과 파라미터 영역의 연속성이나 미분치의 존재와 같은 제약이 받지 않는다.

(2) 대부분의 최적화방법에서 사용되는 점에서 점으로의 이동 방법은 다수 극점을 갖는 탐색영역에서는 국부적인 극점을 찾을 위험이 있지만 유전 알고리즘은 점들의 집단으로부터 동작하여 다수의 최고점을 병렬로 찾기때문에 국부적인 최고점을 찾을 확률이 줄어든다.

(3) 많은 탐색기법이 효과적인 수행을 위해 보조적 정보로 필요로 하는데 비해서 유전 알고리즘은 이러한 보조적인 정

보를 필요치 않으며, 단지 개별적인 문자열과 관련된 목적함수값만을 필요로 한다.

(4) 유전 알고리즘은 탐색영역에 대한 보다 효율적인 탐색 방향을 유도하기위해 확률적 천이규칙을 사용한다.

초기 집단으로부터 시간이 경과함에따라 개선된 집단을 발생시키는 유전 알고리즘은 유전 연산자로 구성되는데, 많은 실용적인 문제에 있어서 좋은 결과를 낳는 단순 유전 알고리즘은 재생, 교배, 돌연변이 연산자로 구성된다.

1) 재생 연산자 : 재생은 적합도 함수값에 의해서 개별적인 문자열이 복제되는 과정이다. 이 연산자를 통해 보다 높은 적합도를 갖는 문자열이 이후의 세대에서 더 많은 수의 자손을 갖게된다. 일단 문자열이 재생을 위해 선택되면, 문자열의 정확한 복제가 이루어지고, 복제된 문자열은 다음의 유전 연산을 위해 임시적 집단인 교배풀 안으로 삽입된다.

2) 교배 연산자 : 재생에 이어서 교배가 진행된다. 교배풀에서 새로이 재생된 문자열이 임의로 짝을 이루면, 각 쌍의 문자열은 임의로 선택된 정수 위치  $K([1, L-1])$  사이,  $L$ 은 문자열의 길이)에 대해  $K+1$  과  $L$ 사이의 모든 문자열이 교환된 새로운 문자열로 형성된다.

3) 돌연변이 연산자 : 2진 코드상에서 돌연변이는 어떤 문자열의 위치에서 1을 0으로, 0을 1로 바꾸는 것을 의미한다. 본질적으로 돌연변이는 문자열 공간을 통한 랜덤 작업으로, 재생 및 교배와함께 사용될 때 중요한 개념의 조기 손실을 막아준다.

본 논문에서 사용된 단순 유전 알고리즘의 순차는 다음과 같다.

단계 1 : 진화 세대수, 집단의 크기, 교배 및 돌연변이율, 문자열의 크기와 같은 초기 파라미터를 설정한다.

단계 2 : 초기 집단을 랜덤하게 발생하고, 문자열로 구성된 각 개체의 적합도를 평가한다.

단계 3 : 유전 연산자에 의해 새로운 개체 집단을 발생한다. 적합도에 비례하여 새로운 개체가 재생되어 교배풀안에 할당되고, 교배풀에서 랜덤하게 선정된 2개의 개체가 교배되며, 그 가운데 랜덤하게 선정된 개체가 돌연변이된다.

단계 4 : 발생된 개체의 적합도가 평가되고, 정해진 진화세대수가 될때까지 단계 3으로 복귀한다. 이 과정에서 적합도가 가장 높은 개체의 문자열은 보존되며, 구하고자 하는 파라미터 집합은 이 문자열을 디코딩함으로써 얻어진다.

뉴로-퍼지 네트워크의 동정은 식(4)의 목적함수  $E$ 를 최소화시키는 것이고, 유전 알고리즘에서는 적합도 함수를 최대화시키도록 진화가 이루어지므로 최소화 문제를 최대화 문제로 변경해야한다. 그 변환 방법에는 여러가지가 있지만, 본 논문에서는 식(5)와 같이 목적 함수의 역수를 적합도 함수로 결정한다.

$$E = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - y_i^*|}{y_i} \quad (4)$$

$$f = 1 / E \quad (5)$$

여기서 N은 데이터 수,  $y_i$ 는 교사 데이터 출력이고,  $y_i^*$ 는 동정된 뉴로-퍼지 네트워크로부터 추론된 값이다.

소수 집단으로 구성된 유전 알고리즘에서 만족할만한 결과를 얻기 위해서는 집단내의 경쟁수준을 조정하는 것이 필요하며, 이는 적합도 환산을 통해 복제되는 문자열 수를 조정함으로써 이루어진다. 본 논문에서는 식(6)과같은 선형환산 [3]을 사용하여 적합도를 환산하며 식(6)은 식(7)을 만족해야 한다.

$$f' = a * f + b \quad (6)$$

$$f'_{max} = C_{mult} * f_{avg} \quad (7)$$

여기서  $f$ 는 본래 적합도,  $f'$ 는  $f$ 의 환산된 적합도이고,  $C_{mult}$ 는 집단내의 최적 구성원에 대한 복제수로 본 논문에서는 2.0이다.

계수 a, b는 평균 환산 적합도  $f'_{avg}$ 와 본래의 평균 적합도  $f_{avg}$ 가 같도록 결정된다. 이러한 적합도 환산은 바람직하지 않은 조기 수렴을 방지할 수 있으나 유전 알고리즘이 상당히 진화되어 집단의 적합도 평균치와 최대치가 상대적으로 비슷해지면, 낮은 값을 갖는 적합도가 환산된후에 오히려 우위가 발생하는데, 이를 방지하기 위하여 본래의 적합도 최소화  $f_{min}$ 을 환산된 적합도  $f'_{min} = 0$ 에 대응하도록 선형환산한다.

다수 파라미터로 구성된 최적화 문제를 풀기위해 다수의 단일 파라미터를 직렬로 연결하는 코딩방식을 사용한다. 각각의 파라미터는 자신의 코드길이와 최대값과 최소값 사이에서 선형적으로 디코딩된다.

$$P = P_{min} + \frac{P_{max} - P_{min}}{2^l - 1} \quad (8)$$

유전 알고리즘은 정점 근방까지는 빨리 수렴하지만, 최종 수렴치를 보증하지는 못한다. 본 논문에서는 정점근방에서 뛰어난 탐색기능을 가지며, 수렴을 보증하는 수정된 킴플렉스법에 의해 뉴런의 연결 강도와 문턱치에 대한 최적의 학습율과 모넨 계수를 찾아낸다.

수정된 킴플렉스법은 기존에 있던 심플렉스법의 개념을 확장한 것이다. 심플렉스법에는 다음과 같은 3가지의 연산자가 있다.

$$\blacksquare \text{ 반사(Reflection) : } X_r = X_0 + \alpha(X_0 - X_h) \quad (9)$$

$$\blacksquare \text{ 확장(Expansion) : } X_e = X_0 + \beta(X_r - X_0) \quad (10)$$

$$\blacksquare \text{ 축소(Contraction) : } X_c = X_0 + \gamma(X_h - X_0) \quad (11)$$

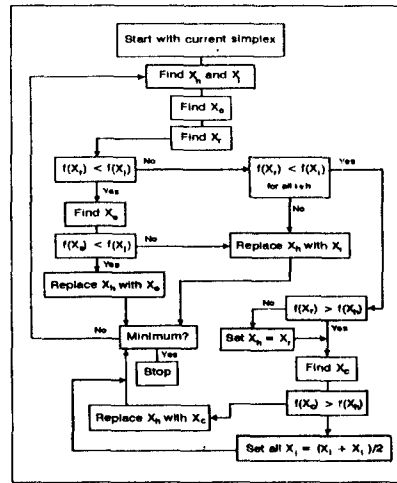


그림 6. 수정된 킴플렉스법의 순서도

심플렉스법에서 이러한 연산을 반복적으로 수행함으로써 식 (4)의 값을 최소로 하게하는 최적의 해를 구하게된다.

유전 알고리즘에 의해 동정된 구조를 갖는 네트워크는 일정한 학습 횟수내에서 역전파 알고리즘에 의해 네트워크의 각 층을 연결하는 연결강도를 조정한다. 그림 7에 뉴로 퍼지 네트워크를 역전파 알고리즘으로 학습시키는 학습 흐름도를 보인다.

#### 4. 시뮬레이션 및 결과고찰

악몽주입 모델의 입력변수로는 원수의 탁도, 수온, pH, 알칼리도를 사용하고 출력변수로는 응집제인 P.A.C의 투입량을 사용한다. 입, 출력 교사 데이터를 사용해서 구조 동정을 하기전에 고려할 사항이 몇 가지가 있다. 첫째로, 공정이 복잡하고 비선형성이 강하기때문에 유전 알고리즘에 의한 최적 네트워크 구조 동정시 한 문자열로 표현된 구조에 대한 적합도를 얻는데 많은 시간이 걸린다. 둘째로, 입력변수 수의 증가는 뉴로-퍼지 입력층의 퍼지 뉴런을 정의하는 멤버십 파라미터 수의 증가와 연결 패턴을 결정하기 위하여 고려해야할 퍼지 뉴런 수를 증가시키게 된다. 이는 유전 알고리즘의 총 문자열 길이를 길어지게 하여 보다 큰 세대수를 필요로 하게 되고 결국 구조 동정에 많은 시간이 걸리게 된다. 이러한 이유로 구조 동정을 하기전에 교사 데이터로부터 출력 P.A.C와 각 입력 변수와의 상관관계를 나타내는 그림에 의해서 퍼지 뉴런의 멤버십 함수 파라미터와 연결 패턴을 미리 결정함으로써 네트워크의 중간층 수, 중간층 노드 수와 학습율을 찾는 데 구조 동정을 국한시킨다. 이로써 구조동정에 걸리는 시간을 줄일수 있다.

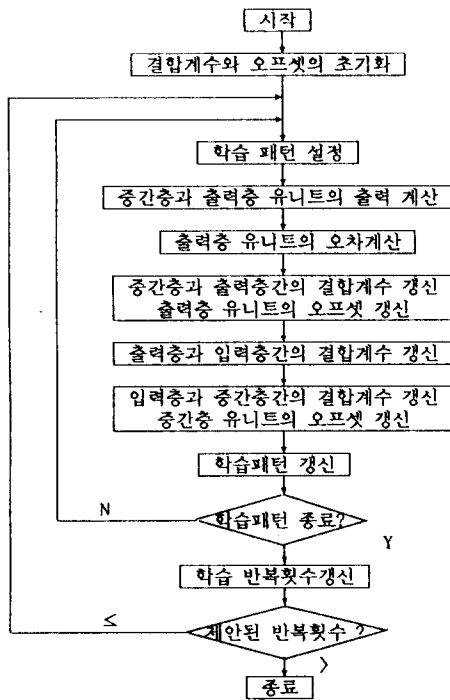


그림 7. 뉴로-퍼지 네트워크를 위한 역전파 알고리즘의 학습 흐름도

뉴로-퍼지 네트워크의 구조동정을 위한 유전 알고리즘의 초기 파라미터는 표 1.에 보여지며, 네트워크의 구조를 표현하는 문자열은 그림 8.에 보여진다.

표 1. 유전 알고리즘의 초기 파라미터

진화 세대수	5
집단의 크기	12
총 문자열의 길이	11
교배율	0.95
돌연변이율	0.01
역전파 알고리즘의 학습횟수	500

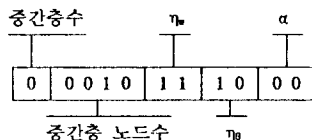


그림 8. 뉴로-퍼지 네트워크 구조 동정을 위한 문자열

그림 8 에서 중간층수를 표현하는 문자열의 길이는 1이고, 그 코드는 중간층이 1이거나 2를 의미하며, 중간층 뉴런수를

표현하는 문자열의 길이는 4이고 코드 0000은 뉴런수가 31인 경우이고 코드 1111은 뉴런수가 46인 경우를 의미하고 0000과 1111사이의 코드는 식(6)에 의해서 선형적으로 디코딩된다.  $\eta_0$ 는 연결강도의 학습율로서 문자열의 길이가 2이고 0.1 ~ 0.9 사이의 값을 가지며,  $\eta_a$ 는 문턱치에 대한 학습율로서 문자열의 길이가 2이고 0.1 ~ 0.9 사이의 값을 가지며,  $\alpha$ 는 모멘텀 계수로서 문자열의 길이가 2이고 0.1 ~ 0.9 사이의 값을 가진다.

유전 알고리즘의 초기 수행에서는 랜덤한 문자열이 정해진 집단의 크기만큼 발생된다. 각 문자열로부터 네트워크의 구조가 생성되며 정해진 학습횟수만큼 학습이 이루어진다. 적합도는 식(5)에 의해서 이루어지고, 학습의 평가는 식(6)에 의해서 이루어진다.

결국 유전 알고리즘이 진화될수록 집단의 문자열은 학습을 효과적으로 수행할 수 있는 구조를 나타내게 된다. 유전 알고리즘의 마지막 세대의 집단중에서 일정한 값이상의 적합도를 갖는 문자열을 수정된 컴플렉스법의 초기해로 하여서 최적의 네트워크 구조를 찾아낸다. 표 2.는 유전 알고리즘과 컴플렉스법에 의해서 동정된 뉴로-퍼지 네트워크의 구조를 나타낸다.

표 2. 동정된 뉴로-퍼지 네트워크의 구조

중간층 수	1
중간층 뉴런수	39
학습율 ( $\eta_w$ )	0.11904763
학습율 $\eta_0$	0.49365079
학습율 $\alpha$	0.36111113

표 3.과 그림 9.에 식(4)에 의해서 계산된 약품주입 공정의 뉴로-퍼지 모델의 오차가 보여진다.

표 3. 동정된 약품주입 공정의 뉴로-퍼지 모델의 오차

학습 횟수	2500	5000	7500	10000
제안된 뉴로-퍼지 모델	0.00222	0.00195	0.00179	0.00169

## 5. 결론

본 논문에서는 뉴로-퍼지 네트워크를 이용하여 비선형 다변수의 공정특성을 가진 상수처리 약품주입공정을 모델링하고, 원수 수질 및 유량변화에 대응한 최적의 약품 주입율을 자동 연산했다. 뉴로-퍼지 추론에 의한 모델링에서 네트워크의 구조 동정을 위해 유전 알고리즘과 수정된 컴플렉스 방법을 사용했고, 파라미터 동정을 위해서 역전파 알고리즘을 사용함

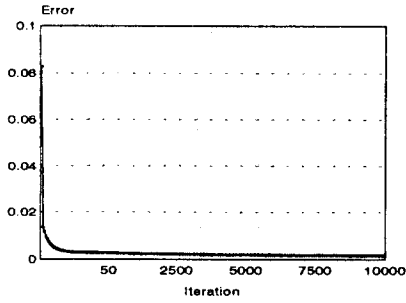


그림 9. 최적 구조를 갖는 약품주입 공정의  
뉴로-퍼지 모델의 오차

으로써 공정모델의 성능을 향상시킬 수 있었고, 정수장의 약품주입 운전실적과 비교하여 뉴로-퍼지 네트워크의 유용성을 평가 분석했다.

## 6. 참고 문헌

1. M. Sugeno and T. Yasukawa, "A Fuzzy-Logic-Based Approach to Qualitative Modeling", IEEE Trans. on Fuzzy Systems, Vol.1, No.1, Feb, 1993.
2. M. Sugeno and G.T. Kang, "Structure identification of fuzzy model", Fuzzy Sets and Systems, Vol.24, pp. 15 ~ 33, 1988.
3. Goldberg, D.E., Genetic Algorithms in Search, Optimization, and Machine Learning, Addison Wesley, 1989.
4. J.M. Keller & D.J. Hunt, "Incorporating Fuzzy Membership Function into the Perceptron Algorithm", IEEE Trans. Pattern Anal. & Machine Intell., Vol.PAMI-7, No.6, pp. 693 ~ 699, 1985.
5. 황 희수, 오 성근, 우 광방, "유전 알고리즘과 퍼지 추론 시스템의 합성", 대한 전기학회 논문지, 41권 9호, pp. 1095 ~ 1103, 1992년 9월.