

GA와 T-S 퍼지시스템에 의한 영산강 수질 예측

Forecasting of the water quality in Youngsan river using by GA and T-S Fuzzy system

박성천¹⁾, 오창열²⁾, 김산원³⁾

Sung Chun Park, Chang Ryol Oh, San Won Kim

요지

대상 지점의 수질 예측은 단순한 모델로 설명하는데 쉽지 않을 뿐만 아니라 많은 오차를 내포하고 있다. 그러나 최근, 신경회로망, 퍼지 논리, 전문가 시스템 및 유전자 알고리즘과 같은 인공지능이 대두되면서 복잡한 비선형 과정들을 나타낼 수 있게 되었다. 나아가 진정한 인공 지능을 실현하기 위해서는 신경회로망, 퍼지 논리, 전문가 시스템 및 유전자 알고리즘을 보다 효과적으로 이용하고 통합해야 가능할 것으로 기대된다.

본 연구에서는 유전자 알고리즘(Genetic Algorithm)을 T-S 퍼지시스템(Takagi-Sugeno Fuzzy system)의 삼각형 멤버쉽 함수 형태와 규칙 베이스를 최적화하기 위한 도구로 사용하였으며, 예측은 T-S 퍼지 시스템을 이용하여 실시하였다. 대상지점은 영산강 유역의 나주지점을 선정하여 유량자료 및 수질자료를 이용하여 GA와 T-S 퍼지 시스템의 결합에 의해 수질 예측을 실시한 결과 돌연변이율(P_m) 0.05~0.1에서 우수한 결과를 얻을 수 있었다.

핵심용어 : 유전자 알고리즘, T-S 퍼지 시스템, 수질

1. 서론

물은 지구상에 존재하는 모든 생명의 근원이며, 생물의 생존에 매우 필요한 요소일 뿐 아니라 물이 없었던 인간 문명의 문화생활 또한 처음부터 존재할 수 없었을 것이다. 근래 급속한 산업화와 도시화에 따라 용수의 사용량이 증대되고, 배출되는 산업폐수와 생활 오수 등은 심각한 수질오염을 일으키고 있으며, 오염물질의 부하량 증대 및 하천의 부영양화를 유발하여 하천 수계의 자연정화능력이 한계에 이르렀다. 따라서 하수와 폐수로 인한 수질오염을 방지하고 보다 효과적인 수질관리에 대한 대안이 요구되고 있다.

본 연구에서는 영산강 유역의 나주지점의 하천수 수질 예측을 위하여 Black Box 모형의 하나인 GA와 T-S 퍼지 시스템을 이용하여 수질예측을 실시하였다. 일련의 과정을 거쳐 탐색되어진 모형들은 도식적으로 관측된 값과 모의된 값의 수질곡선을 도시하여 평가하였으며 수치적 기준으로는 상관계수(CC), RMSE를 구하여 모형을 선택하고 평가하였다.

2. 이론적 배경

2.1 유전자 알고리즘을 이용한 T-S 퍼지시스템의 최적화

유전자 알고리즘을 이용하여 문제를 푸는 경우, 각 개체는 풀고자 하는 문제의 해에 관련된 정보를 가지며 이는 일련의 비트, 정수, 혹은 실수 스트링으로 표현된다. 이러한 스트링은 염색체(Chromosome)라고 불려진다.

-
- 1) 정회원 · 동신대학교 토목공학과 부교수(E-mail: psc@dso.ac.kr)
 - 2) 정회원 · 동신대학교 대학원 토목공학과 박사과정(E-mail: new2020@freechal.com)
 - 3) 정회원 · 동신대학교 대학원 토목공학과 석사과정(E-mail: sanwonkim@hanmail.net))

유전자 알고리즘에 의해 퍼지 제어를 설계하고 최적화하는 것은 퍼지 제어기의 구성요소 중 규칙이나 입-출력변수의 소속함수를 적절히 0과 1의 스트링(비트 스트링 표현의 경우) 즉, 염색체로 코딩하는 것이 필요하다. 예를 들면 삼각형의 멤버십 함수를 정의하기 위해서는 3개의 점이 필요하며 매개변수는 쉽게 이진 스트링으로 나타낼 수 있다. 이 때 이진 스트링을 나타내기 위해 몇 개의 비트를 사용할 것인가가 문제가 되는데, 이는 많은 수의 비트를 사용하면 분해능(resolution)은 커지지만 전체 파라미터를 나타내기 위해서 많은 메모리를 필요로 하고 효율적인 탐색이 어렵기 때문이다. 8비트가 사용된다면 삼각형의 멤버십 함수 하나를 나타내는데 24 비트가 필요하고 이러한 매개변수들은 하나의 염색체 안에 연속해서 코딩되며 코딩된 스트링 값은 미리 정해진 범위 안에서 식 (1)에 의해 실제 값으로 맵핑된다.

$$x = x_{\min} + \frac{b}{(2^m - 1)}(x_{\max} - x_{\min}) \quad \text{식 (1)}$$

여기서, x_{\max} 와 x_{\min} 는 매개변수의 최대값과 최소값, m 은 비트수, b 는 m 비트에 의한 값을 의미한다.

퍼지 제어기의 규칙과 멤버십 함수의 매개변수를 최적화하기 위해, n 입력 1출력의 MISO 퍼지 제어를 고려한다. i 번째($1 \leq i \leq n$) 입력변수의 멤버십 함수 개수(분할 수)가 m_i 개, 각 멤버십 함수를 나타내는데 필요한 매개변수의 수가 p_i 개라 하고, 출력변수의 소속함수 개수가 m_0 개, 멤버십 함수의 매개변수 수가 p_0 개라 하면 i 번째 입력변수와 출력변수의 소속함수에 사용되는 염색체의 길이는 각각 $m_i \times p_i$, $m_0 \times p_0$ 이 된다. 또한 규칙을 나타내기 위한 스트링은 각 규칙의 결론부의 퍼지 변수를 해당되는 값으로 코딩한다. 출력변수의 퍼지 변수가 {NB, NM, NS, ZO, PS, PM, PB}로 구성된다면 각 퍼지 변수가 {1, 2, 3, 4, 5, 6, 7}에 대응되도록 코딩한다. 이 때 비트나 정수 또는 실수 스트링을 사용할 수 있으며 규칙을 위한 스트링의 길이는 $\prod_{i=1}^n m_i$ 이다. 따라서 염색체 스트링의 전체길이는 식 (2)과 같으며 구조는 그림 1과 같다.

$$\text{염색체 } (C_j) \text{ 길이} = \prod_{i=1}^n m_i p_i + m_0 p_0 + \prod_{i=1}^n m_i \quad \text{식 (2)}$$

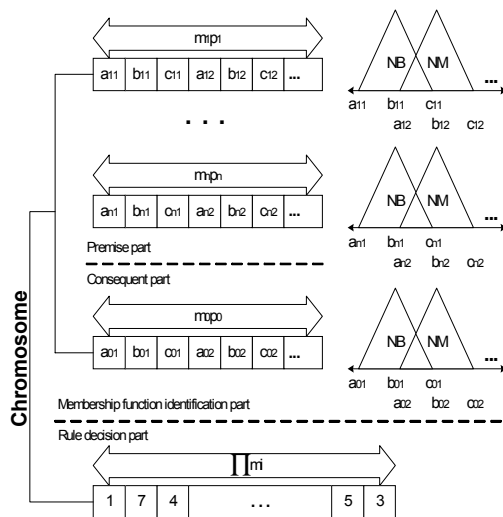


그림 1. 퍼지 규칙 및 멤버십 함수의 최적화를 위한 염색체의 구조

3. 대상지점 및 자료

본 연구의 대상지점은 영산강 수계의 나주지점으로 유역 면적은 2,058.72km²이며 전 영산강 유역의 약 60%에 해당되고, 유로연장이 66.2km인 영산강 본류의 중류부를 대표하는 지점이다. 나주지점의 하천 수질항목인 생물학적산소요구량 (BOD) 예측을 위해 1990년~2001년 사이에 영산강환경관리청에서 조사한 수질자료 및 영산강 홍수통제소에서 조사한 유량자료를 중심으로 입력자료를 구성하였다.

수질 예측을 위해 나주 지점의 GA와 T-S 퍼지시스템의 훈련을 위한 자료는 1990년 1월~1999년 12월까지의 월평균 자료를 이용하였고 훈련 과정에서 구해진 T-S 퍼지시스템의 규칙의 전체부, 멤버십 함수의 형태와 규칙 베이스를 이용하여 2000년 1월~2001년 6월까지의 월평균 자료에 적용하여 검증을 실시하였다.

4. 입력자료의 처리 및 모형의 구성

4.1 입력 자료의 처리

GA와 T-S 퍼지시스템을 위한 각 입력 자료는 효율을 높이기 위해서 전처리과정(pre-processing)을 거쳐 scaling하여 이용하였고, 출력 자료는 후처리과정(post-processing)을 거쳐 수질 자료로 복원하여 검증을 실

시하였다. 전처리 방법에는 여러 가지가 있는데 본 연구에서는 식 (3)과 같이 선형 처리하였다.

$$Y_i = \frac{1}{X_{\max} - X_{\min}}(X_i - X_{\min}) \quad \text{식 (3)}$$

여기서, Y_i : 정규화된 변수값, X_i : 실제 변수값
 X_{\max} : 변수의 최대값, X_{\min} : 변수의 최소값

4.2 모형의 구성

나주지점의 GA와 T-S 퍼지시스템에 의한 수질 예측을 위한 모형은 시행착오법에 의해 다음과 같이 3가지로 구성하였다. BOD 농도는 시차 3으로 고정하였으며 TN, TEMP, QMIN은 각각 시차 1에서부터 3까지 달리하며 Model I, Model II, Model III로 구성하였다.

$$\text{Model I} : B_q = \Phi(\text{bod}_{t-3}, \text{tn}_{t-3}, \text{tempt}_{t-3}, \text{qmin}_{t-3}) \quad \text{식 (4)}$$

$$\text{Model II} : B_q = \Phi(\text{bod}_{t-3}, \text{tn}_{t-2}, \text{tempt}_{t-2}, \text{qmin}_{t-2}) \quad \text{식 (5)}$$

$$\text{Model III} : B_q = \Phi(\text{bod}_{t-3}, \text{tn}_{t-1}, \text{tempt}_{t-1}, \text{qmin}_{t-1}) \quad \text{식 (6)}$$

여기서 B_q : 나주 지점에서의 BOD

bod , tn , temp : 나주 지점에서의 월평균 수질 자료

qmin : 나주 지점에서의 일평균 월최저 유출량

5. 모의결과 및 고찰

본 연구에서 실시한 유전자 알고리즘과 T-S 퍼지시스템에 의한 홍수유출예측 모의 결과의 정확성을 평가하기 위해서 상관계수(CC), RMSE를 이용하였다. CC는 모형의 효율성 판단기준, RMSE는 모형의 수행체도의 판단기준으로 각각 사용하였으며 식 (7), (8)과 같다.

$$CC = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}} \quad (7) \quad RMSE = \sqrt{\frac{\sum_{i=1}^N (\text{obs}_i - \text{exp}_i)^2}{N}} \quad (8)$$

본 연구에서 실시한 GA와 T-S 퍼지시스템의 결합에 의한 모형의 수행은 최적 모형의 선별을 위해 subpop는 4, maxgen은 800, 교배율은 1.0, 돌연변이율은 0.1의 동일한 조건에서 실행하여 선별하였다. 그 결과 Model I에서 가장 좋은 결과를 얻을 수 있었으며 Model I에 대하여 돌연변이율(P_m)을 0.01, 0.05, 0.1, 0.2, 0.3, 0.4로 변화시켜 훈련을 실시한 결과 $P_m(0.1)$ 에서 가장 우수한 예측력을 보였고, 각 모형에 대하여 1개월 후 수질 예측을 모의한 결과는 다음과 같다.

나주지점의 1개월 후 예측에서 수치적 기준에 의한 결과를 표 1에 도식적 기준은 그림 2~7에 나타내었다. NJ1($P_m0.1$) 모형의 수치적 결과를 살펴보면 훈련과정에서 상관계수(CC)가 0.77을 나타냈고, 검증에서는 0.71의 결과를 나타내었다. 모형의 전체적인 예측의 판단기준인 RMSE 값은 훈련과정에서 10.03을 보였고, 검증에서는 7.11을 나타내어 모형의 전체적 예측에 있어서도 우수한 결과를 나타내었다.

표 1. 나주지점의 1개월 후 BOD 예측 결과 (훈련, 검증)

Criteria	Model	모형	Training verification	
CC	Model I	NJ1($P_m0.01$)	0.74	0.62
		NJ1($P_m0.05$)	0.76	0.67
		NJ1($P_m0.1$)	0.77	0.71
		NJ1($P_m0.2$)	0.74	0.64
		NJ1($P_m0.3$)	0.74	0.56
		NJ1($P_m0.4$)	0.72	0.55
		Model II	NJ2($P_m0.1$)	0.73
Model III	NJ3($P_m0.1$)	0.77	0.65	
RMSE (m ³ /sec)	Model I	NJ1($P_m0.01$)	10.70	7.93
		NJ1($P_m0.05$)	10.23	7.40
		NJ1($P_m0.1$)	10.03	7.11
		NJ1($P_m0.2$)	10.56	8.03
		NJ1($P_m0.3$)	10.60	8.46
		NJ1($P_m0.4$)	11.08	8.71
		Model II	NJ2($P_m0.1$)	10.84
Model III	NJ3($P_m0.1$)	9.93	7.59	

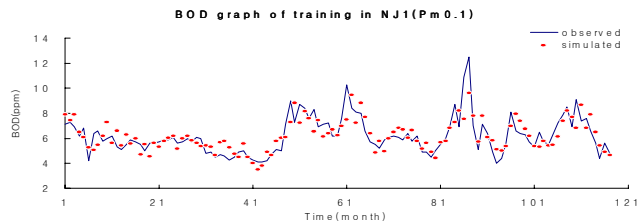


그림 2. NJ1($P_m0.1$)에서 훈련에 대한 관측값과 모의값

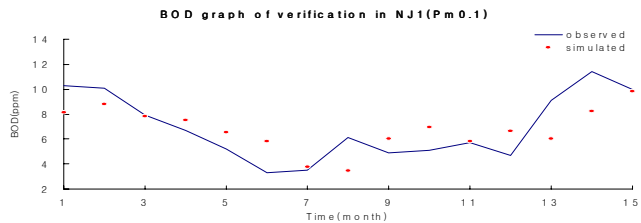


그림 3. NJ1($P_m0.1$) 모형의 검증

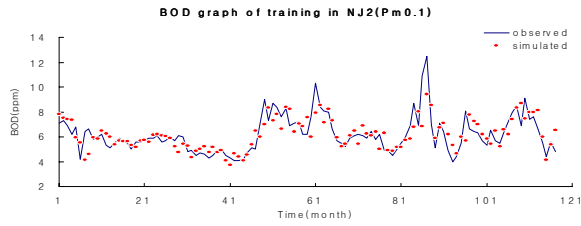


그림 4. NJ2(P_m0.1)에서 훈련에 대한 관측값과 모의값

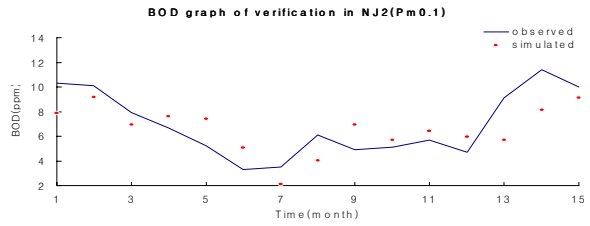


그림 5. NJ2(P_m0.1) 모델의 검증

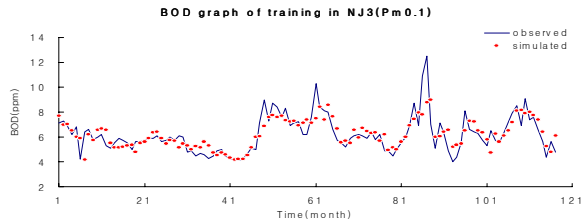


그림 6. NJ3(P_m0.1)에서 훈련에 대한 관측값과 모의값

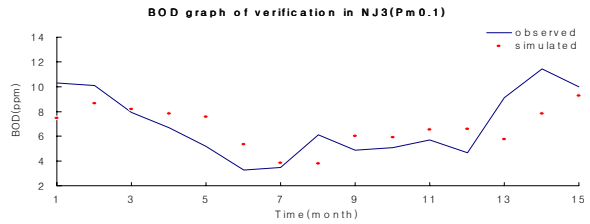


그림 7. NJ3(P_m0.1) 모델의 검증

나주지점에 대하여 BOD를 예측한 결과 Model I에 대하여 가장 우수한 예측력을 보였으며, Model II와 Model III의 경우 비교적 우수한 결과를 나타내었으나 모델의 검증에 있어서 Model I에 비해 다소 떨어지는 예측력을 보였다.

6. 결 론

본 연구에서는 영산강을 유역의 나주 지점을 대상으로 하여 하천수의 BOD 농도 예측을 위해 유전자 알고리즘과 T-S 퍼지 시스템을 이용하였다. 입력자료로는 나주지점의 유량 및 수질자료 이용하였으며 모델의 수행은 유전자 알고리즘의 돌연변이율(P_m)을 변화시켜 가장 우수한 모형을 선별하여 수질예측을 실행하였다. 모형은 시행착오를 통해 구성하여, 각 모형들의 상관계수(CC)값, 및 RMSE(Root Mean Square Error)값을 비교하여 선택한 결과 다음과 같은 결론을 얻었다.

1. 나주 지점의 수질 예측에서 모델 선별 결과 BOD, T-N, TEMP, QMIN을 시차 3을 적용하여 구성된 Model I이 가장 뛰어난 결과를 보여 주었다. 또한 돌연변이율에 따라 구분된 모형에서는 P_m(0.1)에서 가장 우수한 예측을 보였다.
2. 나주 지점의 돌연변이율(P_m)을 변화시킨 후 Model I에서 모형의 예측력을 평가한 결과 P_m 0.01에서는 낮은 돌연변이율로 인해 초기 유전자 조합 이외의 공간의 탐색범위가 축소되어 초기 수렴의 형태로 나타나 좋지 않은 결과를 나타내었고, P_m(0.4)의 경우 초기 유전자 조합의 고유 특성을 상실하여 임의 탐색(random search)의 성격으로 변질되어 결국 찾고자 하는 해의 질에도 한계가 드러났다. 본 연구에서는 돌연변이율의 변화에 따른 결과를 비교한 결과 P_m 0.05 ~ 0.1에서 우수한 결과를 얻을 수 있었다.
3. 유전자 알고리즘과 T-S 퍼지 시스템을 통하여 얻어진 멤버쉽 함수의 형태와 규칙 베이스를 수질예측을 실시한 결과 우수한 결과를 나타내어 이를 바탕으로 유전자 알고리즘과 T-S 퍼지 시스템을 이용하여 수질관리 시스템의 개발 등에 응용할 수 있을 것으로 판단된다.

참고문헌

1. 오창렬(2001). 신경망 이론을 이용한 하천의 수질예측. 석사학위 논문, 동신대학교.
2. 노경범(2001). 유전자 알고리즘과 신경망의 결합에 의한 유출량 예측. 석사학위 논문, 전남대학교.
3. 박성천외 3인(2002). 유전자 알고리즘과 T-S 퍼지시스템에 의한 하천의 유출량 예측, 2002년도 대한토목학회 학술발표회, pp. 50-53
4. 홍창오(2002). GA와 T-S 퍼지시스템에 의한 홍수유출 예측, 석사학위논문, 전남대학교.
5. Frank Hoffmann, Oliver Nelles(2001) Genetic Programming for model selection of TSK-fuzzy system. Information Science 136, pp. 7-28.