

## 인공 신경망을 이용한 생물공정의 규명

이 박 정 식                      이 태 용  
한국과학기술원 화학공학과

### NEURAL NETWORK METHOD FOR BIOPROCESS IDENTIFICATION

PARK CHUNG-SIK                      LEE TAI-YONG  
Dept. of Chemical Engineering  
K A I S T

#### ABSTRACT

It is important to express the specific growth rate of a fermentation reaction as a function of substrate and product concentration in developing bioprocess automation techniques such as modeling of the reactor and controlling it via an advanced control scheme. Typical methods of identification utilize graphical representation of the rate constant data or nonlinear regression with an appropriate noise filter. But the former method fails when the data are erroneous and the latter are mathematically complicated to apply in the field. Neural network is another candidate for the identification from time series data since it is insensitive to the random data error and easy to implement. In this study, we will develop a neural network method of specific growth rate estimation from the time series state variable data and test the performance.

#### 1. 서론

생물 공정 산업에서 생산비의 절감이 중요한 문제가 된지는 이미 오래이지만 이를 위하여 전통적인 화학 공정에서 사용되던 적용 제어나 최적 제어등의 기술이 생물 공정으로 이 전되면서 두 공정의 차이점으로 인한 센서 데이터 (sensor data) 의 해석이나 비정상상태 운전에 대한 문제들은 아직도 많이 미해결된 채로 남아있다. 효소 반응기의 온-라인 (on-line) 최적화와 제어를 위하여 효소 반응기의 동특성을 실시간대에서 계산하는 것이 요구된다. 그러나 효소 반응기의 동적 모델식은 직접 측정할수 없으며 시간에 따라 변하는 비성장속도 (specific growth rate) 와 같은 매개 변수군을

포함하기때문에 반응기의 모델링은 모델의 매개 변수 추산의 문제를 안고 있다. 매개 변수의 추산이란 효소 반응계에서 측정 가능한 상태 변수의 (기질 농도, CO<sub>2</sub> 농도 등) 측정치와 모델식의 계산치가 일치되도록 매개 변수를 역으로 추산하는 과정을 말하며, 이와 같이 추산된 매개 변수는 그 공정의 이후 거동을 예측하는데 사용된다.

지금까지 다루어진 생물 반응기의 규명 문제는 Michaelis-Menten 속도식과 같이 비성장속도의 모델이 주어진 경우에 모델의 매개 변수를 추산하는 문제였으나 이러한 모델이 부정확하거나, 잘 알려지지 않은 경우에는 기존 방법론의 적용이 어려워지며 비성장속도를 기질 농도, 혹은 이를 포함한 다른 상태 변수의 일반적인 함수라 가정하고 추산을 하는 분산 매개 변수 (distributed parameter) 추산의 문제로 바뀐다. 본 연구는 최적화이론을 사용하여 분산 매개 변수를 spline 함수로 추산하고, 이 결과를 생물 공정의 규명 문제에 적용하여 비성장속도등 효소 반응의 속도식 (Kinetic Equation) 모델을 추산하려 하며, 이 때 계산상 복잡한 최적화 이론을 신경망 이론으로 대체하여 알고리즘의 효율을 향상시킨다.

#### 2. 본론

공정 산업에 있어서 기존의 모델링은 대체적으로 두가지로 나뉘어진다. 첫째는 대상 공정에 대한 물질, 에너지 및 모멘텀 (momentum) 수지식을 세우고 반응 속도식, 열역학 평형 관계식, 물성치 등의 구성 (constitutive) 식을 수지식의 상태 변수로 표현하는 방법이다. 이 경우의 모델식은 미분 방정식으로 표현되며, 전산기의 발전으로 많은 생물 반응기 모델식의 계산시간은 그리 문제가 되지 않지만, 속도식등 구성식을 상태 변수로부터 추산하는 과정이 매우 어렵다는 문제가 있다. 여기서는 신경망이론을 이용하여 매개 변수의 추산 작업을 수행하고, 이를 이용하여 모델링을 수행하는 방법을 연구한다.

## 2.1 알고리즘 (algorithm)

신경망 학습까지 고려한 전체적인 알고리즘은 다음과 같다.

- ① 전형적인 비성장 속도를 근사할 수 있는  $Z_j$  를 가정 (측정) 한다.
- ② 계수  $Z_j$  로 부터 비성장 속도 ( $Y_d$ ) 를 계산한다.
- ③ 효소 반응 모델로 부터 반응기의 행동을 시간의 함수로 정리한다.
- ④ 시간의 함수로 주어진 상태 변수를 spline 근사하여 계수  $X_i$  을 구한다.
- ⑤  $Y_j = f(\sum_i W_{ji} X_i)$  로 부터  $Y_j$  를 구한다. 오차 역류 방법 (error back propagation) 을 사용하여  $\|Y_d - Y_j\|$  을 최소화 하도록 synapses의 강도  $W_{ji}$  를 개선한다.
- ⑥  $W_{ji}$  가 수렴하도록 1 번 부터 반복한다.

여기서 5 번의 단계를 학습이라 부르며 일단 학습이 이루어지면 실험 data  $Z_j$  로 부터 parameter 값들을 쉽고 빠르게 estimation 할 수 있다.

## 2.2 모사 연구 (simulation)

모사 연구를 위한 생물 공정 반응기는 plasmid pKB 58 을 포함하는 *Saccharomyces cerevisiae* SEY2102 를 이용한 연속 배양 실험으로 모델링 목적을 위해 cell mass 농도를  $X$  (g.cell/lit) 로 표현하면 cell mass balance는

$$\frac{dX}{dt} = (\mu - D) X \quad (1)$$

로 나타내어지고 비성장 속도  $\mu$  가 Monod form을 따른다고 밝혀졌다.

$$\mu = \frac{\mu_{max} S}{K_s + S} \quad (2)$$

상수 변수들은 표 1에 정리하였다. 有氣 조건에서의 성장은 전적으로 fermentative 하므로 cell yield를 constant로 보는 것이 타당하다.

$$\frac{dS}{dt} = \frac{-\mu X}{Y} + D(S_f - S) \quad (3)$$

정상 상태 cell mass 와 glucose 농도는 식 (1) (2) (3) 을 사용하여 계산할 수 있다.

|             |                            |
|-------------|----------------------------|
| $\mu_{max}$ | 0.2 ~ 0.4 hr <sup>-1</sup> |
| $K_s$       | 0.028 ~ 0.048 g/lit        |
| $Y$         | 0.16 g.cell/g substrate    |
| $S_f$       | 0.3 g / lit                |
| $X(0)$      | 0.02 ~ 0.2 g cell/lit      |
| $S(0)$      | 0.02 ~ 0.2 g / lit         |

표 1 상수 변수들

많은 실험 및 논문들에서  $\mu_{max}$  와  $K_s$  를 포함한 Monod 식 과 같은 여러식으로 매개 변수를 추산하여 왔다. 그런데 일단 매개 변수가 측정 및 계산되었어도 다른 조건들이 약간만 변하는 경우에 즉시 응답하려면 똑같은 과정을 처음부터 반복하여야 하는 어려움이 있다. 본 연구에서는 한번 매개 변수들이 측정 및 계산되면 그것을 이용하여 학습을 시켜서 그이후에 발생하는 작은 변화에 대하여는 매개 변수들이 즉시 계산되는 방법을 고안하였다. 보통은 기질의 농도  $S$  의 데이터로 부터  $\mu_{max}$  와  $K_s$  를 계산하지만 여기서는 cell mass 농도,  $X$  로 부터도 계산할 수 있다는 것을 보인다. 이것은 효소 반응의 비성장 속도를 기질 농도 뿐만 아니라 product 농도, biomass 농도, power input, broth 의 성질 등의 함수로서 추산할 수도 있다는 것을 의미한다. 이 모델은 사용되는 상황에 따라 다르기때문에 이를 극복하기 위하여 비성장 속도를 기질 농도의 spline 함수의 선형 조합으로 표현하고 각 함수에 대한 계수  $Z_j$  를 추산하는 문제로 변형하였다.

## 2.3 신경회로망

구명 (identification) 문제에 있어서 인공 신경망을 이용하면 목표는 주어진 입력 - 출력 쌍에 대한 신경망의 변수를 adjust 하는 적응(adaptive) 알고리즘이나 규칙을 결정하는 것이 된다. 민일 신경망의 가중치(weights)가 변수 벡터  $\theta$  의 요소라고 고려하면 학습 과정은 출력 오차에 근거한 목적 함수  $J$  를 최소화하는 벡터  $\theta^*$  를 결정하는 것이다. 오차 역류 방법(back propagation) 은 이런 목적에서 가장 흔히 쓰이는 방법이다.  $\theta$  에 대한 목적 함수의 gradient는  $\nabla_{\theta} J$  로 계산되어지고  $\theta$  는 최대 경사법 (steepest descent) 로 adjust 한다.

$$\theta = \theta_{nom} - \eta \nabla_{\theta} J \quad \left| \quad \theta = \theta_{nom} \right.$$

여기서  $\eta$ 라는 스텝 크기는 적절하게 선정된 상수이고  $\theta_{nom}$ 은 gradient가 계산되는 시점에서의  $\theta$ 값이다. 본 연구에서는 그림 1 과 같은 3 layer 인공 신경망을 사용하였고 이 경우 오차 역류 방법은 다음과 같이 정의된다.

$$\text{Min. } J = 0.5 \sum_i (Y_{i^{cal}} - Y_{i^{des}})^2$$

subject to

$$Y_i = f(\sum_k W_{ik}^3 Z_k^2)$$

$$Z_k^2 = f(\sum_j W_{kj}^2 Z_j^1)$$

$$Z_j^1 = f(\sum_n W_{jn}^1 X_n)$$

여기서

- f : 비선형 전달 함수
- $Y_{i^{cal}}$  : 계산 중간과정에서의 결과
- $Y_{i^{des}}$  : 원하는 결과
- $W_{ik}^3, W_{kj}^2, W_{jn}^1$  : weights
- $Z_k^2, Z_j^1$  : 각 마디(node)에서의 값
- $X_n$  : 입력

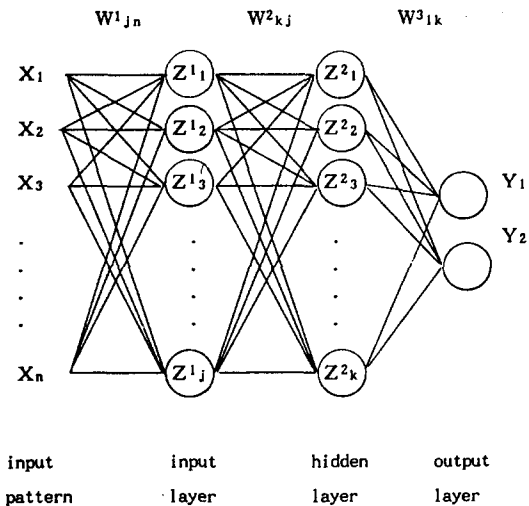


그림 1. 3 layer 인공 신경망

$X_n$  : 시간의 함수로 주어진 상태 변수를 spline 근사하여 나온 계수

$Y_i$  :  $K_s, \mu_{max}$

## 2.4 학습 결과

학습에 쓰인  $K_s$  및  $\mu_{max}$ 의 범위는 그림 2 와 같다.

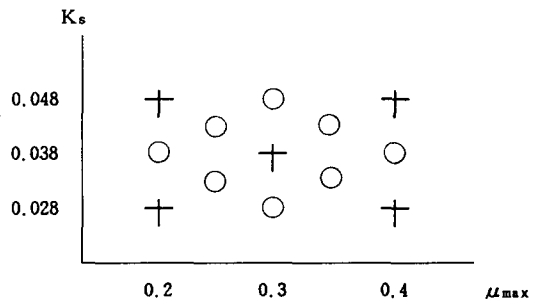


그림 2 학습 시킨 data 범위

- + 학습에 사용된 5 개의 점
- o 시험해 본 8 개의 점

그 결과는 표 2, 표 3 과 같다.

| 목 표 값         | 계 산 값                | 오 차        |
|---------------|----------------------|------------|
| 2.80E-2, 0.20 | 2.79436E-2, 0.200684 | 6.86184E-4 |
| 2.80E-2, 0.40 | 2.80115E-2, 0.400057 | 5.81060E-5 |
| 3.80E-2, 0.30 | 3.80498E-2, 0.299282 | 7.20199E-4 |
| 4.80E-2, 0.2  | 4.80885E-2, 0.199481 | 5.26468E-4 |
| 4.80E-2, 0.4  | 4.78329E-2, 0.400553 | 5.78144E-4 |

표 2 학습에 사용된 5 개의 점

| 목 표 값         | 계 산 값                | 오 차        |
|---------------|----------------------|------------|
| 2.80E-2, 0.30 | 3.79470E-2, 0.357278 | 5.81353E-2 |
| 3.80E-2, 0.40 | 4.46982E-2, 0.408367 | 1.07181E-2 |
| 4.80E-2, 0.30 | 3.60198E-2, 0.249832 | 5.15786E-2 |
| 3.80E-2, 0.20 | 4.03030E-2, 0.191289 | 9.00983E-3 |
| 3.30E-2, 0.25 | 2.76524E-2, 0.235596 | 1.53644E-2 |
| 3.30E-2, 0.35 | 4.31902E-2, 0.392071 | 4.32877E-2 |
| 4.30E-2, 0.25 | 3.04564E-2, 0.207522 | 4.42917E-2 |
| 4.30E-2, 0.35 | 4.55210E-2, 0.361015 | 1.13000E-2 |

표 3 시험해 본 8 개의 결

### 3. 결론

매개 변수 추산은 모델식의 특성에 따라 도해법을 사용하거나 해석적인 최소자승법과 같은 방법들을 사용하여 왔다. 도해법은 추산결과의 신뢰도가 매우 낮으며, 최소자승법은 복잡한 최적화의 이론과 측정 오차에 대한 통계 처리등 계산 과정이 복잡하여 실제 공정에 적용하기 어려운 경우가 많다. 신경망 이론은 적용 대상 모델식의 구체적인 형태에 영향을 적게 받으므로 본 연구의 결과는 서로 다른 여러가지 효소 반응 속도식 모델의 적용이 가능하다. 또한 여기에서 개발된 알고리즘 (algorithm) 은 공정의 모델을 따로 구하지 않고 초기의 신경망 학습 과정을 통하여 얻어진 모델을 이용하여 현장에서도 쉽게 사용이 가능하다. 본 연구에서 개발한 신경망을 이용한 "software sensor" 는 생물 반응계의 모델 형태의 제한을 적게 받고 계산 시간이 짧아서 생물 공정의 온-라인 최적화 및 제어를 가능하게 하리라 기대된다.

### 참고 문헌

- 1) 채 회정, 김 지현, 차형준, 유 영재. "효소 반응 모델식에서의 매개 변수 추정," 한국생물공학회지, 5(2), pp. 133 - 140, 1990.
- 2) Anderson, J. A. and E. Rosenfeld, (ed.) "Neuro-computing," MIT Press, 1988.
- 3) Antsaklis, P. J., "Neural Networks in Control Systems," IEEE Control Systems Magazine, pp. 3 - 5, April, 1990.

- 4) Bailey, J. E. and D. F. Ollis, Biochemical Engineering, McGraw-Hill, 1986.
- 5) Bhat, N. and T. J. McAvoy, "Use of Neural Nets for Dynamic Modeling and Control of Chemical Process Systems Magazine, pp. 24 - 30, April, 1990
- 6) Bhat, N. V., P. A. Moderman, Jr., T. J. McAvoy, and N. S. Wang, "Modeling Chemical Process Systems Magazine, pp. 24 - 30, April, 1990.
- 7) Grosz, R., G. Stephanopoulos, and K. -Y. San, "Studies on On-line Bioreactor Identification. III. Sensitivity Problems with Respiratory and Heat Evolution Measurements," Biotechnol. Bioeng., 26, pp. 1198 - 1208, 1984.
- 8) Lee, H. and I. S. Kang, "Neural Algorithm for Solving Differential Equations," J. Comp. Phys., 91(1), pp. 110 - 131, 1989.
- 9) Lee, T. and J. H. Seinfeld, "Estimation of Absolute and Relative Permeabilities in Petroleum Reservoirs," Inverse Problems, 3, pp. 711 - 728, 1987
- 10) Miller, W. T., R. S. Sutton, and P. J. Werbos, (ed.) "Neural Networks for Control," MIT Press, 1990.
- 11) Naidu, S. R., E. Zafiriou, and T. J. McAvoy, "Use of Neural Networks for Sensor Failure Detection in a Control System," IEEE Control Systems Magazine, pp. 49 - 55, April, 1990.
- 12) Narendra, Y. and K. Parthasathy, "Identification and Control of Dynamic Systems Using Neural Networks," IEEE Trans. Neural Networks, 1(1), 1990.
- 13) Pomerleau, Y. and M. Perrier, "Estimation of Multifl Specific Growth Rates in Bioprocesses," AIChE J., 36(2), pp. 207 - 215, 1990.
- 14) Rumelhart, D. E., J. L. McClelland, and the PDP Research Group, "Parallel Distributed Processing," Vol. 1, Mit Press, 1986.
- 15) San, K.-Y. and G. Stephanopolous, "Studies on On-line Bioreactor Identification. II. Numerical and Experimental Results," Biotechnol. Bioeng., 26, pp. 1189 - 1197, 1984.
- 16) Stephanopolous, G. and K.-Y. San, "Studies on On-line Bioreactor Identification IV. Utilization of pH Measurements for Production Estimation," Biotechnol. Bioeng., 26, pp. 1209 -1218, 1984.
- 17) Stephanopolous, G. and K.-Y. San, "Studies on On-line Bioreactor Identification I. Theory," Biotechnol. Bioeng., 26, pp. 1176 -1188, 1984.