

인공신경망에 의한 생물공정에서 2차원 형광스펙트럼의 분석 II

- 역전파 신경망에 의한 공정의 모델링 -

^{1,5}이금일 · ^{2,4,5}임용식 · ^{2,4,5}손옥재 · ^{1,5}정상욱 · † ^{3,4,5}이종일
전남대학교, ¹산업공학과, ²물질·생물화공과, ³응용화학공학부, ⁴생물공정기술연구실, ⁵바이오광 기반기술개발 사업단
(접수 : 2005. 4. 21., 게재승인 : 2005. 7. 23.)

Analysis of Two-Dimensional Fluorescence Spectra in Biotechnological Processes by Artificial Neural Networks II

- Process Modeling using Backpropagation Neural Network -

Kum-Il Lee^{1,5}, Yong-Sik Yim^{2,4,5}, Ok-Jae Sohn^{2,4,5}, Sang-Wook Chung^{1,5}, and Jong Il Rhee^{3,4,5†}
¹Department of Industrial Engineering, ²Department of Material Chemical and Biochemical Engineering,
³Faculty of Applied Chemical Engineering, ⁴Bioprocess Technology Lab., ⁵Research Center for Biophotonics,
Chonnam National University, YongBong-dong 300, Gwangju 500-757, Republic of Korea
(Received : 2005. 4. 21., Accepted : 2005. 7. 23.)

A two-dimensional (2D) spectrofluorometer was used to monitor various fermentation processes with recombinant *E. coli* for the production of 5-aminolevulinic acid (ALA). The whole fluorescence spectral data obtained during a process were analyzed using artificial neural networks, i.e. self-organizing map (SOM) and feedforward backpropagation neural network (BPNN). Based on the classified fluorescence spectra a supervised BPNN algorithm was used to predict some of the process parameters. It was also shown that the BPNN models could elucidate some sections of the process performance, e.g. forecasting the process performance.

Key Words : 5-aminolevulinic acid; artificial neural networks; fermentation; 2D fluorescence spectra; recombinant *E. coli*

서론

많은 감독신경망 구조와 훈련알고리즘 중에서 피드포워드 역전파 신경망은 (feedforward BPNN)은 단순함과 다양한 사용예제 때문에 생물공정에서 많이 응용되고 있다. 예를 들면 BPNN 알고리즘은 dynamic identification(1); adaptive control(2), 생물시스템의 모의실험과 예측(3) 그리고 생물공정의 모델링에 사용되고 있다(4). 또한, BPNN은 액-액 용매추출을 위한 막 생물반응 공정에서 생물시스템의 전체형광 스펙트럼 데이터로부터 공정수행변수를 추론하기 위해서 사용되기도 하였다(5). 그러나 SOM을 이용하여 2차원 형광스펙트럼 데이터의 양을 줄이고 BPNN 알고리즘을 이용하여 스펙트럼 데이터와 공정변수 사이에 가

능한 관계들을 정량적으로 분석하기 위한 체계적인 연구가 아직 이루어지지 않고 있다. 따라서 본 논문에서는 SOM과 BPNN과 같은 인공신경망들을 이용하여 재조합 대장균의 발효공정에서 수집된 2차원 형광스펙트럼 데이터들을 분류하고 발효공정의 수행변수를 예측하기 위한 실제적인 접근법에 중점을 두어 연구하였다.

재료 및 방법

생물반응기에서의 재조합대장균 발효는 37°C, 1 vvm, 450 rpm의 조건으로 수행하였다. 본 연구에서는 ALA생산을 위해 표 1에서 보는 바와 같이 총 3회의 발효를 수행하였으며, 2차원 형광분광계를 이용하여 공정을 온라인 모니터링하였다.

인공신경망 이론

발효공정에서 수집된 많은 양의 2차원 형광스펙트럼 데이터의 분석을 위해 SOM과 BPNN 알고리즘이 사용되었다. 전체형광 데이터는 SOM 알고리즘을 이용하여 데이터

† Corresponding Author : Faculty of Applied Chemical Engineering, Chonnam National University, YongBong-dong 300, Gwangju 500-757, Republic of Korea
Tel: +82-62-530-1847, Fax: +82-62-530-0846
E-mail : jirhee@chonnam.ac.kr

들을 몇 개의 그룹들로 분류하였으며, 이렇게 분류되어진 형광 데이터는 BPNN에 의한 발효공정 모델링의 입력자료로 사용되었다. 위의 전체적인 작업과정은 Fig. 1에 나타내었다. SOM과 BPNN 알고리즘 훈련은 Matlab Neural Network Toolbox (ver. 6.1, The Mathworks, Inc., USA)를 이용하였으며 512 MB의 팬티엄 V 컴퓨터가 사용되었다(6).

Table 1. Three fermentation processes for the production of ALA by recombinant *E. coli*

	FerM1	FerM2	FerM3
Culture medium	MS8	LB	LB
Process operating conditions	batch pH 6.2	batch pH 6.5	batch pH 6.5
Addition of other components	IPTG, LA	----	LA

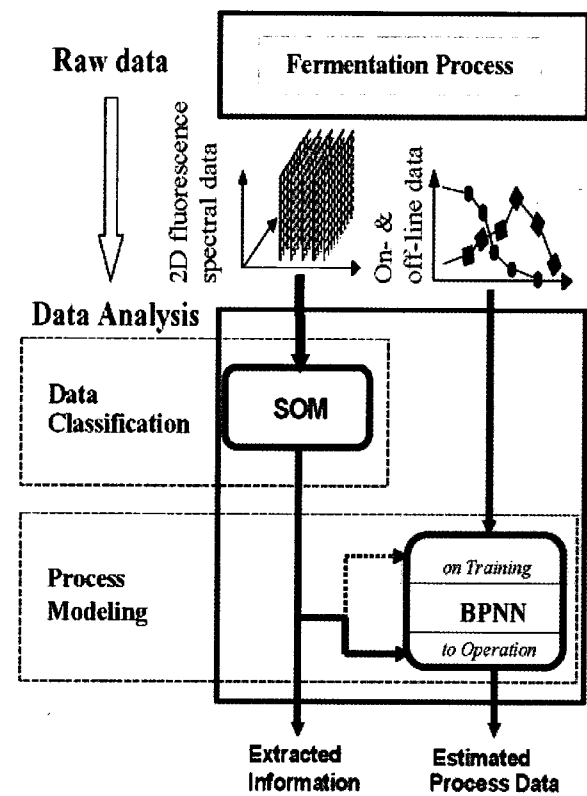


Figure 1. Working procedure for the application of the SOM and BPNN to the analysis of 2D fluorescence spectra.

자기조직화망 (SOM)

이전의 논문(7)에 자세히 설명되었다

역전파 신경망 (BPNN)

BPNN은 3가지 형태의 층으로 구성되는데 보통 한 개의 입력층과 한 개 이상의 은닉층 그리고 한 개의 출력층으로 이루어져 있으며, 각 층은 1개 이상의 뉴런들로 구성된다(8-10).

입력층의 입력뉴런은 SOM 알고리즘 결과 (예를 들어 총 493개의 파장조합에서 98개가 선택됨)에 의하여 결정되는데 열벡터의 형태를 띤다. 출력층의 뉴런에는 공정변수 (CO₂, ALA농도)값을 사용하는데, 이것은 네트워크의 목표값으로 쓰인다. 입력뉴런과 출력뉴런의 자료값들은 식(1)에 의해 0.1과

0.9 사이의 값으로 조정된다(11).

$$Y^* = 0.8 \times \frac{(X^* - X_{\min})}{(X_{\max} - X_{\min})} + 0.9 \tag{1}$$

위식에서 x*는 발효공정에서 수집된 데이터이며, y*는 식(1)에 의하여 조정된 값이다. 그리고 x_{max}와 x_{min}은 각각 x*의 최대값과 최소값을 나타낸다.

BPNN에서 초기 가중치와 상수항은 작은 값으로 무작위로 설정되며 입력력 뉴런은 오류역전파 알고리즘에 의해 계산되어진다. 표준순방향 BPNN알고리즘은 입력값에 가중치를 곱하고 상수값을 더한 후, 다시 앞의 값들을 모두 합산한 결과값을 네트워크 입력값 (net input)으로 사용한다. 그리고 출력값은 각 출력 뉴런으로 들어오는 네트워크 입력값을 전달함수 (여기서는 시그모이드 함수)을 이용하여 계산한다. 신경망의 오류함은 학습 데이터 집합에 대한 오차 ($E = \frac{1}{2} \sum_k \sum_p (t_{pk} - O_{pk})^2$; p번째 입력값에 대해 실제 k번째 출력층에서의 출력값 (O_{pk})과 목표 (실험치) 값 (t_{pk}) 간의 차이)들의 제곱 합을 누적하여 사용하는데, 이것은 역전파신경망의 가중치와 상수항을 갱신하고 보다 정확한 출력값을 얻는데 사용된다. 이 오류함의 충분히 작은 값 (예 : < 0.001)이 되거나 학습횟수가 지정된 값 (예: 5000번)에 도달할 때까지 위의 학습과정은 반복된다.

오류역전파 신경망 모델의 학습은 gradient descent (기울기 급감) 방법을 이용하여 네트워크의 가중치 (w)들을 변화시키면서 이루어지며 다음과 같이 표현할 수 있다.

$$\Delta W(r+1) = \alpha(r) \times \frac{dE(r)}{dW} + \eta \times \Delta W(r) \tag{2}$$

식(2)에서 r, α (r) 및 E(r)은 각각 학습횟수, 학습률 그리고 오차 제곱합의 누적치이며, η 은 모멘텀을 나타낸다. 그리고 이 알고리즘에서 학습률을 다음과 같이 나타낼 수도 있다.

$$(r+1) = (r) \times 0.7, \text{ if } E(r) > \xi^* E(r-1)$$

$$(r+1) = (r) \times 0.5, \text{ if } E(r) < E(r-1)$$

$$(r+1) = (r), \text{ otherwise}$$

위 식에서 ξ는 미리결정 된 요소이다.

BPNN모델은 교차확인 (cross-validation) 기법을 이용하여 타당성을 확인하는데, 이 알고리즘은 네트워크의 일반화 성능을 최적화하는데 사용된다 (예를 들어 은닉층의 뉴런 수를 최적화하거나 학습 싸이클의 최적수를 정하는데 쓰인다)(12). 본 연구에서는 온라인 또는 오프라인 측정 데이터 (이동평균과 내삼법을 이용하여 얻은 426개 데이터)에서 입력벡터의 행요소인 발효시간 중 100개를 골라 학습용과 검증용 데이터로 나누었다. 학습용 데이터 70%와 검증용 데이터 30%는 무작위로 선택하였으며, 학습용 데이터를 이용하여 신경망을 학습하였다. 그리고 네트워크의 일반화 성능을 확인하기 위해 정기적으로 학습 데이터로 학습된 네트워크에 검증용 데이터를 대입하여 오류율을 계산하였고, 그 중 가장 작은 오류함 값을 갖는 네트워크구조를 선택하였다(13).

BPNN 알고리즘으로 학습을 마친 출력뉴런은 본래 공정의 목표값과 동일한 단위의 값으로 바꾸어준다. 이러한 과정에 의해 만들어진 모델의 예측력은 실제상관계수 (R_{xy})와 표준오차 (SE)의 결과값을 이용하여 나타낸다.

$$R_{xy} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (3)$$

$$SE = \sqrt{\frac{1}{n(n-2)} \left[n \sum y^2 - (\sum y)^2 - \frac{n \sum xy - (\sum x \sum y)^2}{n \sum x^2 - (\sum x)^2} \right]} \quad (4)$$

위 식에서 \bar{x} 와 \bar{y} 는 각각 x와 y의 평균값을 나타내며, y값은 모델에 의해 예측된 x의 값이다.

상관계수는 0과 1사이의 값을 나타내고, 상관계수 (R_{xy}) 값이 더 클수록 더 정확하게 예측된 모델임을 의미한다. 그리고 표준오차 (SE) 값은 x와 y에 대응되는 예측값 y와의 오차를 나타낸다.

결과 및 고찰

SOM을 이용한 형광스펙트럼의 분류

재조합 발효공정에 대해 이전의 논문(7)에서 설명한 것처럼 SOM에 의해 전체 스펙트럼 데이터는 2개에서 10개의 그룹들로 분류할 수 있었다(14). 분류된 스펙트럼 데이터로부터 유용한 스펙트럼 데이터들을 찾아내기 위해서 형광세기 벡터와 가중치 벡터 사이에 역유클리디안 거리를 계산하였다. 각 그룹에서 가장 큰 요소들 중 10% 내의 여기 방출파장의 조합을 각 그룹의 중심영역으로 포함시켰다. FerM1에서 전체형광영역은 5개로 분류되었으며 그룹 1은 17, 그룹 2는 7, 그룹 3은 11, 그룹 4는 5, 그룹 5는 8개의 중심영역을 나타내었다. 또한 그룹 분포표에서 각 그룹에 대한 스펙트럼요소들의 가중치를 계산하였다. 전체 스펙트럼 데이터가 5개 그룹으로 분류될 경우 5개의 시간에 따른 가중치의 경로를 얻을 수 있었으며, 이것을 이용하여 발효공정을 정성적으로 분석하였다. 발효공정에 대한 정보는 그룹분포표에서 각 그룹에 대한 분산과 정규화된 가중치의 변화를 온라인 및 오프라인 데이터와 비교함으로써 얻을 수 있다.

FerM2에서 수집된 전체 형광스펙트럼 데이터들은 SOM 알고리즘을 이용하여 총 8개 그룹으로 분류하였으며, 가장 작은 유클리디안 거리 중 20% 내의 요소들을 중심영역에 포함시켰다. 그 결과 그룹 1은 19, 그룹 2는 13, 그룹 3은 11, 그룹 4는 13, 그룹 5는 6, 그룹 6은 15, 그룹 7은 7, 그룹 8은 15개의 중심영역을 각각 나타내었다.

BPNN을 이용한 발효공정의 모델링

그룹 분포표에서 각 그룹의 중심영역과 함께 여기파장과 방출파장의 조합은 BPNN 알고리즘에 의한 발효공정 모델링의 입력뉴런으로 사용되었다. 발효공정 모델링에서 입력뉴런의 숫자는 매우 중요한데 이러한 입력뉴런의 숫자는 그룹중심영역의 비율 (percentage)에 의존한다(15). 그러므로 모델 성능에 대한 입력뉴런숫자의 영향성을 조사하기 위해 그룹분포표에서 중심영역 숫자를 변화시켜 그 영향성을 조사하였다. 예를 들면 가장 작은 유클리디안 거리 중 5% 내의 요소들을 중심영역으로 포함시킬 경우 25개의

중심영역을 얻을 수 있고, 가장 작은 유클리디안 거리 중 30% 내의 요소들을 중심영역에 포함시킬 경우 147개의 중심영역을 나타내었다.

ALA농도에 대한 예측오차 값은 교차확인(cross-validation) 알고리즘을 이용하여 계산하였다. 예측오차 값은 훈련횟수가 증가할수록 감소하지만 일정수준이상 증가하면 다시 증가하게 된다. 은닉뉴런의 숫자가 5개인 네트워크 모델에서 학습알고리즘은 팬티엄 V 컴퓨터로 약 0.5시간의 훈련 후에 최소 예측오차 값에 도달하였다. 그러나 최적 훈련횟수의 정확한 숫자는 네트워크의 초기화에 의존한다.

네트워크모델에 의한 ALA농도의 예측값을 실제 오프라인 데이터와 비교하였다. Fig. 2는 모델에 의해 예측된 ALA농도와 실제 오프라인 측정값들의 결과가 잘 일치함을 보여주고 있다($R_{xy} = 0.983$, $SE = 0.0547$).

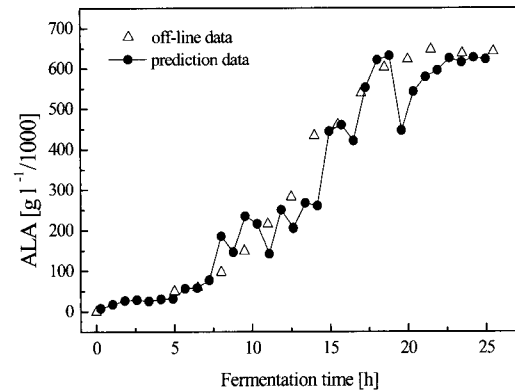


Figure 2. Comparison of the predicted ALA concentrations predicted by the network model (98-5-1) with the off-line data in FerM1.

FerM1에서 몇몇 공정수행변수에 대한 상관계수 (R_{xy})와 표준오차 (SE) 값을 최적의 네트워크 구조를 이용하여 계산하였으며 그 결과를 표 2에 나타내었다.

Table 2. Empirical correlation coefficients (R_{xy}), standard errors (SE) and the optimal network topology for a few process parameters in FerM1

Parameters	R_{xy}	SE	Optimal topology (In-Hid-Out)
Glucose	0.973	0.0734	98-5-1
CO ₂	0.991	0.0291	98-8-1
DO	0.995	0.0219	98-5-1
O ₂	0.989	0.0265	98-3-1
DCW	0.978	0.0674	98-5-1

FerM2에서 몇몇 공정변수에 대한 네트워크 모델링은 SOM을 이용하여 얻어진 여기 및 방출파장의 조합들을 이용하였다. 공정변수에 대한 모델의 기하학적구조는 전체 입력데이터 중 70%는 시험 데이터로 30%는 검증데이터로 분리한 후에 결정하였다. Fig. 3(a)에서는 98-3-1 구조의 BPNN 모델을 이용하여 예측된 DCW 값과 오프라인 측정값을 비교하였으며 상관계수와 표준오차는 각각 0.997와 0.0247이었다. 위의 결과는 FerM2에서 98-3-1 구조가 DCW 값 예측을 위한 가장 최적의 모델구조임을 의미하는 것이

다. Fig. 3(b)는 BPNN 모델에 의한 CO₂ 농도의 예측값과 실제 CO₂ 측정값을 비교한 것이다. CO₂ 농도 예측을 위해서 98-8-1의 BPNN 모델구조를 이용하였으며, 그 결과 상관계수와 표준오차는 각각 0.970과 0.0327을 나타내었다.

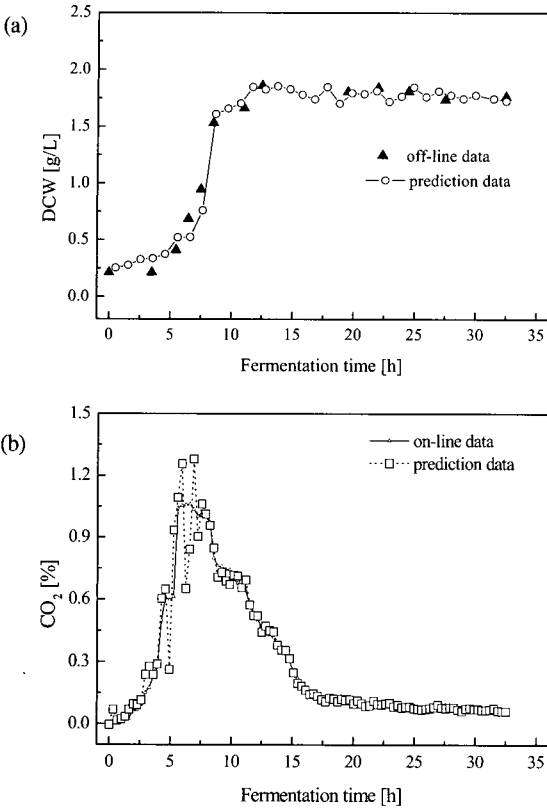


Figure 3. (a) Comparison of the DCW predicted using a model with the off-line data in FerM2, (b) Comparison of the CO₂ concentrations predicted by a model with the on-line data in FerM2.

FerM2에서 다른 공정변수에 대한 최적모델구조, 상관계수 및 표준오차 값은 표 3에 나타내었다. FerM2에서 몇몇 공정변수에 대한 최적모델구조는 FerM1에서의 모델구조와 조금 다르게 나타났다. 그 이유는 배양액의 첨가요소를 다르게 하였으며, 그러한 결과로 발효액의 형광스펙트럼세기가 다르게 나타났기 때문일 것이다.

Table 3. The empirical correlation coefficients (R_{xy}), standard errors (SE) and optimal topology for some process parameters in FerM2

Parameters	R_{xy}	SE	Optimal topology (In-Hid-Out)
ALA	0.984	0.0584	98-8-1
DO	0.992	0.0322	98-5-1
O ₂	0.993	0.0331	98-5-1

많은 발효공정변수에 네트워크 모델의 적합성을 시험하는 것은 매우 중요하다. 즉 특정 발효공정을 위해 세워진 인공신경망모델이 다른 유사한 발효공정에서도 공정변수들을 예측할 수 있는가를 검증하는 것이 필요하다는 것이다. 이를 위해서 FerM3에서는 LA를 지수성장기 종반에 첨한 것을 제외하고는 FerM2와 모든 조건을 동일하게 수행

하였다. FerM2에서 ALA농도 예측을 위한 98-8-1구조의 BPNN 모델을 FerM3에 적용하여 그 적합성을 검증하였다. FerM3에서 수집된 형광스펙트럼을 입력데이터로 사용하고, FerM2의 모델링을 위해 사용된 가중치를 이용하여 FerM3에서의 ALA농도를 예측하였다. Fig. 4에서는 FerM3에서 ALA농도의 오프라인 측정값과 FerM2에서 사용된 가중치를 이용하여 예측된 ALA농도 값을 비교하였다. 12시간 이전에는 예측값과 오프라인 측정값 사이에 좋은 연관성을 보였다 ($R_{xy} = 0.845$, $n = 50$). 그러나 12시간 이후에 모델에 의한 ALA농도 값은 증가하지만 오프라인 측정값들은 감소하였다. 이렇게 두 결과가 일치하지 않는 이유는 FerM3에서 재조합 대장균의 플라스미드 안정성이 발효 12시간에 90%에서 발효 종료시점에 10%로 감소하였기 때문일 것이다. 결과적으로 재조합대장균의 회분발효에 대한 BPNN 모델은 매우 적합하며 ALA농도와 같은 공정변수들의 예측을 위해서도 사용될 수 있을 것이다. 그러나 이러한 BPNN 모델의 적합성도 2차원 형광분광계의 photomultiplier 세기변화와 같은 측정조건에 많은 제약을 받을 수 있다.

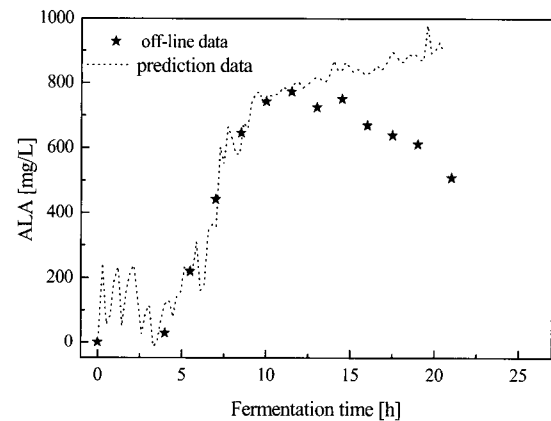


Figure 4. Comparison of the off-line ALA data measured in the FerM3 with the ALA data predicted by the network with the same weights as the network model for the FerM2.

전체형광스펙트럼 데이터를 사용한 BPNN 모델은 훈련데이터의 범위를 넘어서는 다른 공정데이터를 예측하는 경우에도 사용될 수 있다. 98-5-1구조의 BPNN 모델을 이용하여 FerM3에서의 DCW값을 6.9, 9.2, 11.6, 16.3 시간에 각각 예측하였다. 그리고 위의 모든 경우에서 예측은 공정이 종료되는(21시간) 시점까지 계속되었다. 그림 5는 각각의 경우에서 몇 가지 변화들을 나타내고 있으며 특히 예측이 빨리 시작된 경우(6.9시간)에 변화가 비교적 크게 나타내고 있음을 보이고 있다. 그러나 11.6시간 이후에 예측된 DCW값은 오프라인 데이터의 DCW값과 잘 일치하였다. 즉, 각 모델의 예측능력을 나타내는 RMSEP (the root means square error of the prediction,

$$RMSEP = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_{j,forecast} - y_{j,off})^2}$$

를 살펴보면 예측된 DCW값 ($y_{j,forecast}$)과 11.6시간과 23시간 사이에 더하여진 오프라인 데이터 값 ($y_{j,off}$)을 이용한 결과 0.112 g/L이었으며 9.2시간과 16.3시간 이후에 RMSEP 값은 각각 0.165 g/L과 0.143 g/L이었다. 이러한 결과는 발효공정에 대해 완전히 훈련된 BPNN 모델은

좋은 예측능력을 가지고 있다는 것을 의미하는 것이다.

2차원 형광분광계를 이용하여 발효공정을 모니터링할 경우 많은 양의 형광스펙트럼 데이터를 얻을 수 있으며, 이렇게 얻어진 많은 양의 데이터에서 유용한 정보를 갖는 스펙트럼요소를 가려내기 위해서 chemometric 방법이 이용되고 있다. 그러나 미리 정의된 경험상의 모델을 요구하는 선형 보정기법 (선형다중회귀분석, 부분최소자승)들과는 다르게, 신경망은 일단 훈련데이터 쌍들이 얻어지면 입력변수와 이에 상응하는 출력변수간의 비선형관계에서 네트워크 스스로 패턴을 인식하고 학습할 수 있다. 이러한 경우로 Wolf 등(16)이 BPNN을 이용하여 전체 형광데이터를 “지문”화하는 유형인식 접근방법으로 공정수행변수를 잘 추론할 수 있다는 것을 발견하였다.

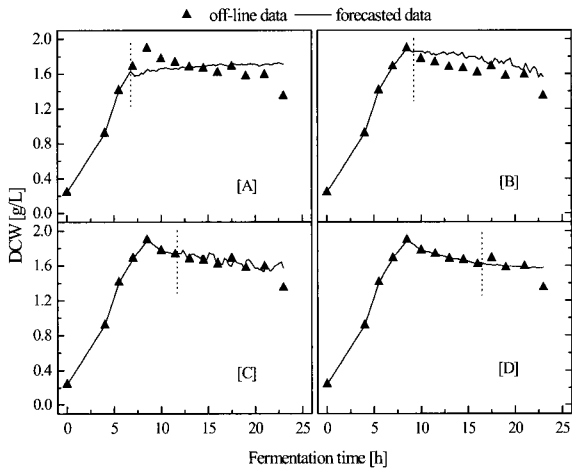


Figure 5. Comparison of the off-line DCW data with the forecast of the DCW data in the FerM3 [A]: forecast of the DCW from 6.9 h; [B]: forecast of the DCW from 9.2 h; [C]: forecast of the DCW from 11.6 h; [D]: forecast of the DCW from 16.3 h.

본 연구에서는 인공신경망을 이용하여 2차원 형광 스펙트럼데이터를 분류, 분석하고 모델식을 이용하여 발효공정을 예측하였다. SOM 알고리즘은 많은 양의 형광스펙트럼 데이터를 소수의 특징그룹으로 분류하고 전체 형광 데이터에서 중요한 스펙트럼요소들을 가려내었다(493개의 스펙트럼 요소 중 49개). 그리고 SOM을 이용하여 분류한 데이터들은 공정모델링을 위한 BPNN의 입력자료로 사용되었다. 몇몇 공정변수들은 SOM으로부터 얻은 분산 및 정규화된 가중치와 연관시켜서 정성적인 분석을 하였다. 세포의 대사상태나 공정의 수행변수들과 같은 정량적인 정보들은 BPNN모델을 이용하여 예측 가능하였다. 네트워크의 최적 구조는 발효공정의 공정변수뿐만 아니라 관심의 대상이 되는 다른 부분에 대해서도 시험하였다. 위의 결과들을 종합하여볼 때 BPNN모델은 공정변수의 예측에 매우 유용하게 이용될 수 있다는 것을 알 수 있었다. 네트워크모델은 매우 강력하며 다른 유사한 발효공정에서의 공정변수 예측에 쓰일 수 있다. 또한 이러한 네트워크모델이 충분히 훈련될 경우 훈련데이터의 범위를 넘어서는 공정의 예측도 가능하였다.

요 약

본 연구에서는 인공신경망 알고리즘을 이용하여 생물공정에서 수집된 형광스펙트럼 데이터를 분류, 분석하고 공정변수들을 예측하기 위한 공정의 모델링에 대해서 논의하였다. SOM에 의해 분류된 전파장 스펙트럼 데이터들은 발효공정의 변수와 형광데이터 사이에 비선형관계를 설명하기 위하여 사용되었다. BPNN알고리즘은 SOM에서 분류된 데이터들을 입력자료로 이용하여 공정에 대한 모델식을 세우고, 이를 이용하여 배출가스 내 CO₂ 농도 및 발효액 중 세포농도와 같은 공정변수들을 예측하는데 사용되었다. 또한 BPNN 모델은 강력하면서도 훈련데이터의 범위를 넘어서는 공정의 데이터들을 예측할 수 있기 때문에 폭넓은 활용가능성을 가지고 있다.

감 사

본 연구는 한국과학재단 (KOSEF)의 2002년도 특정기초 연구과제 (과제번호: R01-2002 -000-00027-0)에 의해 이루어졌으며 이에 감사드립니다.

NOTATION

ALA	5- Aminolevulinic acid (mg/L)
ANN	Artificial neural networks
BMU	Best matching unit
BPNN	Backpropagation neural network
DCW	Dry cell weight (g/L)
DO	Dissolved oxygen concentration (%)
E, E(r)	Cumulative sum-squared-error
EGFP	Enhanced green fluorescent protein
FerM	Fermentation process
IPTG	Isopropylthiogalactoside
LA	Levulinic acid
LB	Luria Bertini
M	Number of classes
n	Number of total data points
O _{pk}	Actual output of the kth output layer neuron for the pth input pattern
PCA	Principal component analysis
PLS	Partial least square regression analysis
r	Present number in the fluorescence spectrum
R _{xy}	Empirical correlation coefficient
RMSEP	Root mean square error of prediction
SE	Standard error
SOM	Self-organizing map
T	Number of spectral components in a class
t _{pk}	Target output of the kth output layer neuron for the pth input pattern
VAR _i (r)	Time-dependent variance of the fluorescence intensities of all spectral components in a class
VAR _m	Mean variance of the fluorescence intensities for all spectral components of all classes in a distribution card
x*	Real value of a variable
x _{max}	Maximum value of a variable
x _{min}	Minimum value of a variable
y*	Rescaled value of a variable
y _j	output neuron value at j neuron
Greek letters	
(r)	learning rate at iteration of r
η	momentum
ξ	predefined factor

REFERENCES

1. Alaradi A. A., S. Rohani (2002), Identification and control of a riser-type FCC using neural networks, *Com. & Chem. Eng.* **26**, 401-421.
2. Kobayashi H., R. Ozawa (2003), Adaptive neural network control of tendon-driven mechanisms with elastic tendons, *Automatica* **39**, 1509-1519.
3. Gontarski C. A., P. R. Rodrigues, M. Mori, L. F. Prenem (2000), Simulation of an industrial wastewater treatment plant using artificial neural networks, *Com. & Chem. Eng.* **24**, 1719-1723.
4. Chen L. Z., S. K. Nguang, X. D. Chen, X. M. Li (2005), Modeling and optimization of fed-batch fermentation processes using dynamic neural networks and genetic algorithms, *Biochem. Eng. J.*, in print.
5. Wolf G., J. S. Almeida, C. Pinheiro, V. Correia, C. Rodrigues, M. A. M. Reis, J. G. Crespo (2001), Two-dimensional fluorometry coupled with artificial neural networks: a novel method for on-line monitoring of complex biological processes, *Biotechnol. Bioeng.* **72**, 297-306.
6. Matlab manual, vers. 6.1, The Mathworks, Inc., USA (2002).
7. Lee K. I., Y. S. Yim, Kim C. K., S. H. Lee, S. W. Chung, J. I. Rhee (2005), Analysis of two-dimensional fluorescence spectra in biotechnological processes by artificial neural networks. 1. Classification of fluorescence spectra using self-organizing maps, *Korean J. Biotechnol. Bioeng.*, submitted.
8. Baughman D. R., Y. A. Liu (1995), *Neural Networks in bioprocessing and chemical engineering*, Academic Press, San Diego, USA.
9. Rhee J. I., K. I. Lee, C. K. Kim, Y. S. Yim, S. W. Chung, J. Wei, K. H. Bellgardt (2005), Classification of two-dimensional fluorescence spectra using self-organizing maps, *Biochem. Eng. J.* **22**, 135-144.
10. Tambe S. S., B. D. Kulkarni (1996), *Elements of artificial neural networks with selected applications in chemical engineering, chemical & biological sciences*, Simulation & Advanced Controls, Inc. Louisville, USA.
11. Basheer I. A., M. Hajmeer (2000), Artificial neural networks: fundamentals, computing, design, and application, *J. Microbiol. Methods* **43**, 3-31.
12. Bhat N. V., T. J. McAvoy (1992), Determining model structure for neural network stripping, *Com. & Chem. Eng.* **16**, 271-281.
13. Servercan M., F. Severcan, P. I. Haris (2001), Estimation of protein secondary structure from FTIR spectra using neural networks, *J. Molecul. Structure* **565**, 383-387.
14. Chung S. Y., K. H. Seo, J. I. Rhee (2005), Influence of culture conditions on the production of extra-cellular 5-aminolevulinic acid (ALA) by recombinant *E. coli*, *Process Biochem.* **40**, 385-394.
15. Nicholson D. J., D. B. Kell, C. L. Davey (1996), Deconvolution of the dielectric spectra of microbial cell suspensions using multivariate calibration and artificial neural networks, *Bioelectrochem. Bioener.* **39**, 185-193.
16. Mukherjee J., C. Lindermann, T. Scheper (1999), Fluorescence monitoring during cultivation of *Enterobacter aerogenes* at different oxygen levels, *Appl. Microbiol. Biotechnol.* **52**, 489-494.