

유전알고리즘을 이용한 유가식 효모 배양 최적화

나 정 결 · †장 용 근 · ¹정 봉 현

한국과학기술원 생물공정연구센터 및 화학공학과

¹한국과학기술연구원 생명공학연구소

(접수 : 1999. 6. 25., 게재승인 : 1999. 7. 30.)

Optimization of Fed-Batch Yeast Culture by Using Genetic Algorithm

Jeong-Geol Na, Yong Keun Chang†, and Bong Hyun Chung¹

BioProcess Engineering Research Center and Department of Chemical Engineering, Korea Advanced Institute of Science and Technology, Taejon 305-701, Korea

¹Korea Research Institute of Bioscience and Biotechnology, KIST, Taejon 305-600, Korea

(Received : 1999. 6. 25., Accepted : 1999. 7. 30.)

The optimization of fed-batch yeast fermentation process has been performed using genetic algorithm(GA). Three strategies were designed and applied to obtain the optimal feed rate profiles. Genes in the chromosome (input variables for optimization) included feed rates on fixed time intervals (strategy I), or switching times t_{s1} and t_{s2} , and feed rates on singular arc (strategy II), or feed rates and the length of time interval (strategy III). Strategy III showed the best results for all initial conditions due to efficient utilization of genetic information. Simulation results using GA showed similar or better performance compared with previous results by variational calculus and singular control approach.

Key Words : genetic algorithm, fed-batch optimization

서 론

20세기 초반, 효모 발효공정에서 기질인 맥아즙을 한꺼번에 넣는 것보다 시간을 두고 조금씩 넣어 줄 때 생산성이 더 좋다는 결과가 나온 이후, 많은 항생제, 효소, 유기산 발효에서 유가식 발효공정이 이용되었다. 유가식 발효는 특히 기질저해나, catabolite repression이 있는 공정에 효과적이다.

유가식 발효에서 시간에 따르는 최적 feed rate profile을 구하고자 할 때에는 조작변수인 feed rate가 시스템이나 목적함수값에 선형적인 관계를 가지게 되므로 Pontryagin(1)의 maximum (minimum) principle로는 구할 수 없는 구간이 생기게 된다 (singular control problem). 이러한 singular problem을 풀기 위해 많은 연구가 있었다. Ohno(2) 등은 Green's theorem을 이용하여 lysine 생산 공정의 최적 feed rate를 구하였다. 그러나 이 방법은 물질수지식이 세 개 이상일 경우에는 해를 찾을 수 없다는 한계가 있다. Fishman과 Biryukov는 페니실린 생산 공정에서 최적 포도당 공급속도를 구하는 문제를 연구하였는데, feed가 들어감으로써 생기는 부피 변화를 고려해 주지 않았다.

Modak(3)은 feed rate의 최적해가, 균체농도를 극대화한 후 product 생산성을 극대화하는 방향으로 결정된다는 추측에 따라 singular problem을 풀었으나 문제가 간소화됐음에도 불구하고 해를 얻기 위해서는 매우 많은 계산량이 필요하였다. Singular problem의 여러 가지 어려움을 해결하기 위해 문제 자체를 non-singular 문제로 바꾸어 문제를 해결하려는 시도가 있어 왔다(4). 즉 조작변수를 feed rate가 아닌 부피나 비성장속도로 바꿔줌으로써 singular 구간을 없애는 방법이다. 그러나 이 경우 최적화결과를 실제 배양시 반영하는 과정에서 조작변수를 온라인으로 측정하여 최적 profile을 따라 제어하는 것이 매우 어렵거나 불가능한 경우가 대부분이다. 이를 피하기 위해서는 구해진 부피나 비성장속도 profile을 다시 feed rate로 바꾸어야 하는 번거롭고도 어려운 과정이 따르게 된다.

유전알고리즘은 인공지능 기법을 사용한 검색방법으로 멘델의 유전법칙과 다윈의 '적자생존의 원리'라는 자연 현상을 모델링한 것이다. 일반적으로 어떠한 과제는 해가 될 가능성이 있는 것들을 탐색하는 것이며 '최적화'문제 역시 최적해를 검색하는 과정으로 축소될 수 있다. 보통 좁은 탐색영역에서는 특수한 기법이 사용되어야 하며 유전알고리즘은 그러한 기법들 중 하나이다(5). 유전알고리즘은 방향성 있는 탐색과 확률 탐색의 두 요소를 결합하여 매우 강건하며 빠른 속도로 최적해로 수렴해가므로 선로 라우팅, 스케줄링, 순회판매원 문제 등과 같은 최적화 문제에 성공적으로 적용되어 왔다.

본 연구에서는 유전알고리즘을 이용하여, 효모 유가식 발효시의 다양한 초기조건에 대한 최적 feed rate profile을 계산하고, 이

† Corresponding Author : Bioprocess Engineering Research Center and Department of Chemical Engineering, Korea Advanced Institute of Science and Technology, Taejon 305-701, Korea

Tel : 042-869-3927, Fax : 042-869-8800

e-mail : ychang@sorak.kaist.ac.kr

를 이전의 연구에서 구한 해와 비교함으로써 유전 알고리즘의 우수한 성능을 입증하고자 하였다.

이 론

Problem Formulation

모델 시스템은 포도당으로부터 효모 균체를 생산하기 위한 유가식 배양이다. 에탄올을 제외한 다른 부산물의 생성은 무시할 만 하며 maintenance에 의한 기질소비도 없는 것으로 가정하였다. 이 때의 물질수지식은 다음과 같다.

$$\begin{aligned} \frac{d(XV)}{dt} &= (\mu_G + \mu_E)XV \\ \frac{d(GV)}{dt} &= G_F F - \sigma X V \\ \frac{d(EV)}{dt} &= (\pi - \eta)XV \\ \frac{dV}{dt} &= F \end{aligned} \tag{1}$$

여기서 X, G, E 는 각각 균체, 포도당, 에탄올농도이다. 또, F 는 포도당공급속도 (feed rate), V 는 발효액 부피, G_F 는 feed중의 포도당 농도이고 μ_G, μ_E 는 포도당과 에탄올에 대한 비성장속도이다. σ 는 포도당 소모속도, π 는 에탄올 생산속도, η 는 에탄올 소모속도를 나타낸다. $\mu_G, \mu_E, \sigma, \pi, \eta$ 들은 X, G, E 의 함수이다.

Feed rate와 발효조 부피에는 다음과 같은 제한조건이 적용된다.

$$\begin{aligned} V_0 \leq V(t) \leq V_{\max} = V(t_f) \\ 0 = F_{\min} \leq F(t) \leq F_{\max} \end{aligned} \tag{2}$$

적절한 목적함수의 설정은 최적화 연구에서 중요한 부분을 차지하는데 본 연구에서처럼 효모균체생산이 주목적인 경우 부산물인 에탄올생성을 줄이기 위해 feed rate를 줄일 경우 조업시간이 늘어나게 되고 조업시간을 줄이기 위해 feed rate를 빨리 하면 에탄올이 많이 생겨 균체수율이 떨어지는 결과를 가져온다. 따라서 성능지수는 최종 효모농도와 조업시간을 동시에 고려한 값이 되어야 한다. 본 연구에서의 목적을 수학적으로 표현하면 다음과 같다.

$$\max_{F(t)} IP(t_f) = (XV)_f - \epsilon t_f \tag{3}$$

여기서 F 는 feed rate, IP 는 성능지수(performance index), t_f 는 배양종료시간을, 하첨자 f 는 종료시간에서의 값을 의미한다. 또 ϵ 은 발효를 통해 생산된 균체단위당 당 얻을 수 있는 이득을 단위시간당 드는 조업비용으로 나눈 값이다. 상기 목적함수는 본 연구에서의 최적화 목적이 feed rate를 시간에 따라 적절히 조절함으로써 배양종료시점에서의 성능지수의 값을 극대화함에 있으며 성능지수는 배양을 통해 얻을 수 있는 산물의 양(또는 가치)로부터 배양시간에 따른 비용($\epsilon \times t_f$)을 제외한 값으로 정의됨을 말해준다.

전산모사에 사용한 발효 모델은 Modak(7)이 제안한 *S. cerevisiae* LBG H1022에 대한 모델이다(Table 1). 이 모델을

Table 1. An unstructured model for aerobic yeast growth

Mass balance Equations

$$\begin{aligned} \text{Cell Mass} \quad \frac{dX}{dt} &= -DX + \mu X \\ \text{Glucose} \quad \frac{dG}{dt} &= D(G_F - G) - \sigma X \\ \text{Ethanol} \quad \frac{dE}{dt} &= -DE + (\pi - \eta)X \end{aligned}$$

Rate Expressions

$$\begin{aligned} \text{Specific growth rate} \quad \mu &= \mu_G + \mu_E \\ \text{on glucose} \quad \mu_G &= \frac{0.0747G + 0.4G^2}{0.0011 + 0.351G + G^2} \\ \text{on ethanol} \quad \mu_E &= \frac{0.114E}{(0.005 + 0.59\sigma + E)(1 + 1.43\sigma)} \end{aligned}$$

Fraction of glucose fermented

$$R(G) = \frac{1.0 + 20712.9G^2}{100.0 + 20712.9G^2}$$

$$\begin{aligned} \text{Specific glucose consumption rate} \quad \sigma &= \frac{\mu_G}{Y_X} \\ \text{Specific ethanol production rate} \quad \pi &= 0.4590\sigma R \\ \text{Specific ethanol consumption rate} \quad \eta &= \frac{\mu_E}{0.6250} \\ \text{Cellular yield} \quad Y_X &= 0.5435(1 - R) + 0.1776R \end{aligned}$$

사용한 이유는 매우 정확히 효모의 성장 kinetics를 모사한다는 것 외에도 Modak이 구한 singular problem의 해가 있어서 유전알고리즘의 성능을 비교 평가할 수 있기 때문이다.

Genetic Algorithm

유전알고리즘은 다음과 같은 요소를 가지고 있다.

1. 해가 될 가능성이 있는 변수값들의 조합 또는 한가지 변수의 profile (총칭하여 입력변수)의 이진 스트링 표현을 통한 유전자적 표현방식
2. '적합도'에 의해 해를 평가하는 평가함수
3. 다음 세대의 유전자를 창출하는 유전 연산자들
4. 유전연산자들이 작동할 확률, 개체(chromosome)와 집단(population) 크기 등의 매개변수값

따라서 유전알고리즘은 일반적인 최적화방법과 몇가지 차이점을 지니고 있다. 첫째는 최적화를 위해 변수값들의 조합 또는 profile이 아니라 이들을 유전자적 표현으로 바꾼 변수코드를 사용한다는 것이고 둘째는 point별로 탐색하는게 아니고 여러 point들을 동시에 비교평가하며 셋째는 일반적인 최적화 알고리즘이 성능지수의 미분값을 사용하는데 비해 성능지수 자체를 사용한다는 점이다. 또 확실적인 변환과정을 거침으로써 deterministic한 방법보다 비선형적인 함수의 최적화에 강점을 지니고 있다.

유전알고리즘의 기본적인 연산자는 재생산, 교배, 돌연변이이

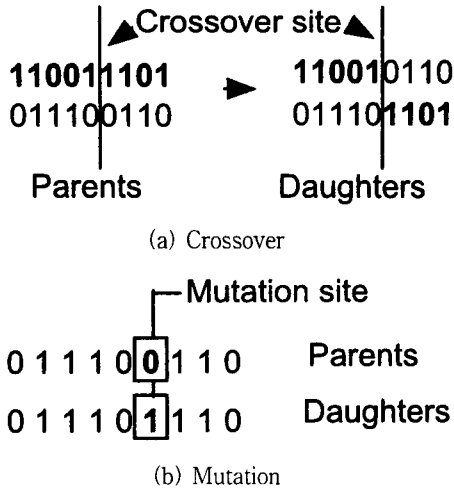


Figure 1. Operations in genetic algorithm.

다(6). 재생산(reproduction)은 현세대의 어떤 염색체(변수코드)가 주어진 환경에 얼마나 적합인지(성능지수값이 얼마나 크지)에 따라 우성염색체는 많이 재생산시키고 열성염색체는 도태시키는 역할을 한다. 단순한 재생산만으로는 더 적합한 염색체를 찾을 수 없으므로 우성염색체들간의 교배(crossover)를 통해 새로운 염색체를 찾게 된다(Figure 1(a)). 재생산이나 교배는 기본적으로 처음 주어진 유전정보만을 사용하는 연산과정이므로 local minimum에 빠질 우려가 있다. 이 점을 보완하기 위해 돌연변이가 연산자가 사용된다(Figure 1(b)).

유전알고리즘을 이용한 최적화는 대개 다음과 같은 방식으로 작동된다. 우선 입력변수값들을 이진수 스트링으로 변환시킴으로써 각 입력변수가 유전자(gene)처럼 표현되도록 한다. 하나의 염색체(개체)는 이러한 유전자들의 집합이고 한 집단은 고정된 수의 염색체들로 이루어진다(Figure 2). 변환과정이 끝나면 각 염색체들의 적합도(성능지수)에 따라 재생산과정을 거치게 된다. 재생산을 통해 골라낸 염색체들은 확률에 따라 교배나 돌연변이를 거치게 되고 다음 세대에 최적해에 보다 가까운 유전정보를 물려주게 된다. 따라서 새로운 세대의 집단이 될 수록 우성유전자를 지닌 개체들이 더 많게 되므로 성능지수는 최적값으로 수렴하게 되는 것이다.

유전알고리즘의 이론적인 기초는 해의 이진 스트링 표현과 스키마의 개념에 의존한다. 스키마는 염색체들 사이의 유사성을 나타낼 수 있는 하나의 전형으로 각 스트링에서 '0'이나 '1'의 유전정보 대신 '*'를 도입하여 만들어진다. 예를 들어 길이가 10인 스트링과 스키마를 생각했을 때, 스키마 (*1*1100100)은 다음 4개의 스트링 {(0101100100), (0111100100), (1101100100), (1111100100)}과 일치한다. 어떤 스키마가 r개의 '*'를 가진다면, 그 스키마는 정확히 2^r개의 스트링과 일치하게 되고 길이가 m인 스트링의 경우 2^m개의 스키마와 일치하게 된다. 따라서 크기가 n이고 스트링의 길이가 m인 개체집단에서는 n · 2^m개의 다른 스키마들이 표현될 수 있다. 스키마에는 차수와 정의길이라는 두 가지의 중요한 성질이 있는데, 스키마 정리는 이러한 성질들의 기초 위에 정립되어 있다. 스키마 S의 차수(O(S))는 스키마에 존재하는 고정된 0과 1의 수이다. 다시 말하면 전체길이에서 '*' 심볼의 수를 뺀 것이다. 차수는 스키마의 특별함을 정

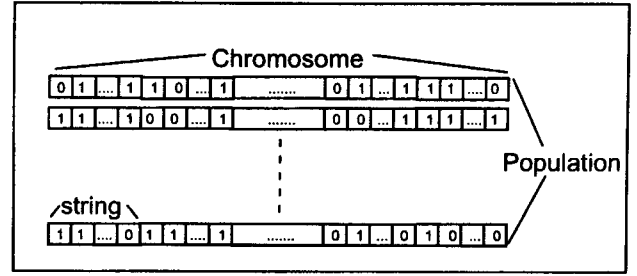


Figure 2. Composition of a population.

의하는데 예를 들어 길이가 10인 스키마 S₁ = (**001*11*)의 차수는 5이다. 또 다른 특성인 정의길이(δ(S))는 첫 번째와 마지막의 고정된 스트링 위치 사이의 거리이고 이는 스키마에 포함되어 있는 정보의 밀집성을 나타낸다. 예를 들면, δ(S₁)=9-4=5이다. 유전알고리즘에서 진화과정(최적화 탐색과정)은 적합도에 의한 선택과 유전알고리즘의 두 연산자인 교배와 돌연변이를 통해 일어난다. 시간 t에서 스키마 S와 일치하는 개체집단내의 스트링 수를 ξ(S,t)라 정의하고 이 스트링들의 평균 적합도를 val(S,t)라고 정의하면 val(S,t)는 다음과 같다.

$$val(S, t) = \frac{\sum_{i=1}^{\xi(S, t)} val(v_i)}{\xi(S, t)} \quad (4)$$

선택단계 동안 각 스트링은 적합도에 따라 한번도 선택되지 않거나 한번, 혹은 여러번 선택되며 한번의 스트링 선택에서 스트링 v_i는 p_i=val(v_i)/F(t)의 확률로 선택된다(F(t)는 시간 t에서 개체집단내 모든 스트링들의 적합도의 합이다). 선택단계 이후에 스키마 S와 일치하는 스트링의 개수 ξ(S,t+1)는 다음과 같다.

$$\xi(S, t+1) = \xi(S, t) \cdot pop_size \cdot eval(S, t) / F(t) \quad (5)$$

개체집단의 평균적합도를 고려하면, 위 수식을 다음과 같이 쓸 수 있다.

$$\xi(S, t+1) = \xi(S, t) \cdot eval(S, t) / \overline{F(t)} \quad (6)$$

다시 말하면, 개체집단내의 스트링의 수는 평균적합도에 대한 스키마의 적합도의 비율로 증가한다. 이것은 다음 세대에서 “평균 이상의” 스키마는 스트링의 수가 지수 증가하고, “평균 이하의” 스키마는 지수 감소하며, 평균적인 스키마는 같은 수준에 머물러 있다는 것을 의미한다.

그러나 선택만으로는 탐색공간에서 고려해야 할 새로운 점들을 도입하지 못하므로 교배 연산자와 돌연변이 연산자에 의해 탐색공간의 영역을 넓히게 된다. 교배 연산자를 통해 교배가 일어나는 위치는 스트링의 길이가 m이라 할 때 m-1개의 가능한 위치중에서 균등하게 선택된다. 이것은 스키마 S의 파괴확률이

$$p_d(S) = \frac{\delta(S)}{m-1} \quad (7)$$

이라는 것을 암시하고, 결국 스키마의 생존확률은

$$p_s(S) = 1 - \frac{\delta(S)}{m-1} \quad (8)$$

이다. 교배의 작동확률이 p_c 라는 것에 주목하면 스키마의 생존 확률은 다음과 같다.

$$p_s(S) \geq 1 - p_c \cdot \frac{\delta(S)}{m-1} \quad (9)$$

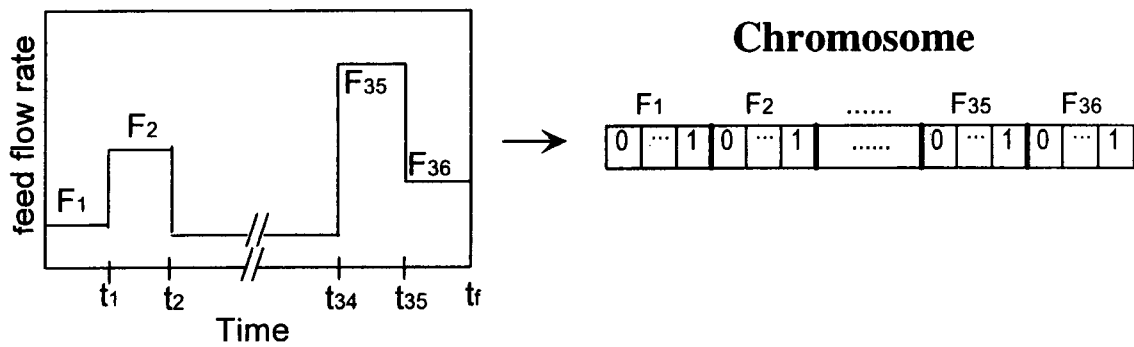
돌연변이 연산자는 확률 p_m 으로 염색체내의 한 비트를 0 또는 1로 바꾼다. 만일 돌연변이의 영향을 받지 않았다면 스키마의 모든 고정된 위치가 변화되지 않아야 한다. 한 비트가 변화하는 확률이 p_m 이므로, 비트의 생존확률은 $1-p_m$ 이다. 한 돌연변이는 다른 돌연변이와 무관하고, 따라서 스키마 S 가 돌연변이되지 않을 확률은

$$p_s(S) = (1 - p_m)^{\alpha(S)} \approx 1 - O(S) \cdot p_m \quad (10)$$

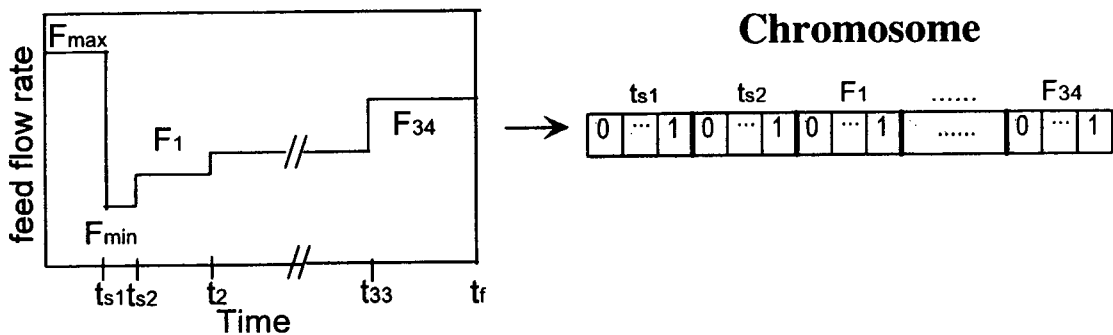
이다. 선택, 교배, 그리고 돌연변이의 결합된 효과에 의해 다음과 같은 스키마 증가 방정식을 얻을 수 있다.

$$\xi(S, t+1) \geq \xi(S, t) \cdot \frac{eval(S, t)}{\overline{F(t)}} [1 - p_c \cdot \frac{\delta(S)}{m-1} - O(S) \cdot p_m] \quad (11)$$

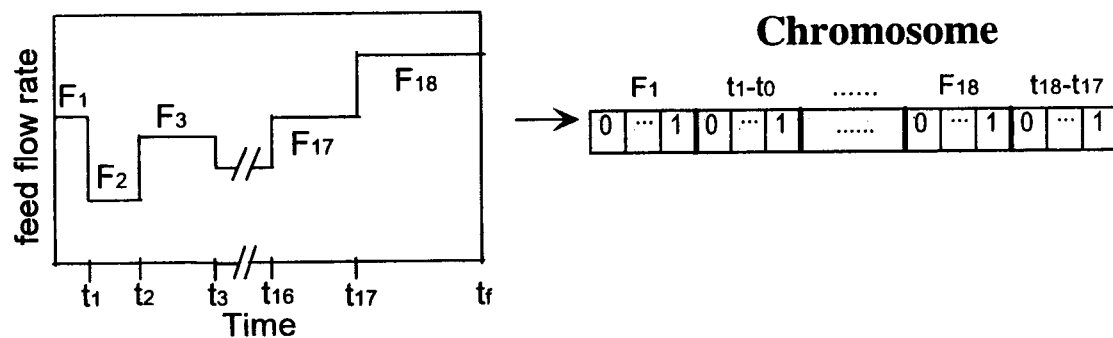
위와 같이 유전알고리즘에서의 선택은 평균이상의 스키마를 늘려주고, 그 변화는 지수적이라는 것을 보여준다. 또한, 교배연산자가 도입되어 구조적이고 동시에 임의적인 정보교환을 가능하게 해준다. 게다가 돌연변이연산자를 통해 개체집단에 큰 다양성이 도입된다. 따라서 유전알고리즘은 표본 중에서 우수한



(a) Strategy I



(b) Strategy II



(c) Strategy III

Figure 3. Strategies for the computation of optimal feed rate profile.

적합도를 지닌 스트링은 살아남고 적합도가 낮은 스트링은 지속적으로 감소하여 최적값으로 수렴하며 그 과정은 짧고, 낮은 차수의 고성능 스키마의 병렬처리를 통해 이루어진다.

Genetic Algorithm을 이용한 유가식 배양 최적화

최적화 과정은 성능지수(식 (3))가 최대화되도록 조작변수인 feed rate의 profile $F(t)$ 를 구하는 것이다. 이 문제는 다시 전체 배양시간을 여러 구간으로 나누고 목적을 만족시키도록 각 구간의 F_i 를 찾는 최적화문제로 전환될 수 있다.

본 연구에서는 다음과 같은 세 가지 전략을 가지고 유전알고리즘을 이용하였다(Figure 3).

1. 각 구간의 크기는 고정하고 $F_i(t)$ 를 유전정보로 사용
2. Switching times, t_{s1} , t_{s2} 와 singular arc에서 $F_i(t)$ 를 유전정보로 사용
3. 구간의 크기와 각 구간의 feed rate F_i 를 유전정보로 사용

첫 번째 전략은 가장 간단한 방식으로 구간의 크기를 고정시켜 놓고 각각의 구간에 대해 feed rate를 구하는 방식이다.

두 번째 전략은 Modak(3)의 연구결과에 착안한 것이다. Modak은 유가식 배양의 최적 feed rate profile은 초기에는 균체농도를 높이는 방향으로 진행되고 공정후반에는 product의 농도를 극대화시키는 방향으로 진행될 것이라는 추측에 따라 복잡한 variational problem을 최적 switching time을 얻는 문제로 간소화시켰다(3). Figure 4를 보면 각각의 초기조건에 따른 최적 feed rate profile의 형태를 알 수 있다. Figure 4(a)같이 초기포도당농도와 균체농도가 모두 낮을 경우 F_{max} 로 포도당을 공급시켜 일단 균체농도를 높이고 그 다음에는 product의 농도를 극대화시키는 방향, 즉 본 연구에서는 에탄올 생산을 억제하는 포도당농도를 유지시키기 위해 feed를 잠시 끊어주었다(F_{min}) F_s 의 형태로 포도당을 공급하여야 한다. Figure 4(b)의 경우에는 초기 균체농도가 높으므로 에탄올 생산을 억제하는 포도당농도를 유지하기 위해 $F_{min} \rightarrow F_s$ 의 형태로 포도당을 공급하여야 하고 Figure 4(c)에서와 같이 적절한 초기농도에서 발효가 시작될 경우에는 바로 F_s 의 속도로 포도당을 공급하면 된다. Figure 4(d)의 경우는 균체농도는 높으나 포도당농도가 낮으

로 F_{max} 로 발효조안의 포도당 농도를 일정한 수준까지 올린 다음 F_s 로 에탄올 생산을 억제하면서 포도당을 공급하여야 한다. 따라서, 최적 feed rate profile을 얻는 문제는 F_{max} 로 포도당을 공급하는 시간 ($0 \rightarrow t_{s1}$), 포도당공급을 중단하는 시간 ($t_{s1} \rightarrow t_{s2}$), 발효조의 부피제한조건을 만족시킬 때까지 F_s 로 넣어주는 시간 ($t_{s2} \rightarrow t_{s3}$)과 singular 구간에서의 feed rate F_s 를 구하는 문제로 바꿀 수 있다. 이 중 singular 구간의 포도당 공급시간은 발효조의 부피제한조건에 의해 쉽게 얻을 수 있는 값이다. 따라서 두 번째 전략은 switching time t_{s1} , t_{s2} 와 나머지 구간(singular 구간)을 일정하게 나뉘었을 때 각 구간에서의 $F_i(t)$ 를 구하는 방식이다.

세 번째 전략은 첫 번째 것을 보완한 것으로 구간이 고정될 경우 생길 수 있는 정확도상의 문제를 배제하고 인접한 여러 구간의 feed rate가 같을 때는 하나의 구간으로 합함으로써 유전정보의 낭비를 막기 위한 것이다.

결과 및 고찰

여러 초기조건에 대한 전산모사

유전알고리즘을 사용하여 유가식 효모 발효공정 최적 feed rate profile을 구하였다. 알고리즘의 강건성을 살펴 보기 위해 다음의 네 가지 초기 조건에 대해 전산모사를 수행하였다(Table 2).

- A) 초기 포도당농도와 균체농도가 낮을 경우($S_0 = 0.0005$ g/L, $X_0 = 0.2$ g/L)
- B) 초기 포도당농도와 균체농도가 높을 경우($S_0 = 5$ g/L, $X_0 = 1$ g/L)
- C) 초기 포도당농도와 균체농도가 적절할 경우($S_0 = 0.065$ g/L, $X_0 = 0.68$ g/L)
- D) 초기 포도당농도는 낮고 균체농도는 높을 경우($S_0 = 0.0005$ g/L, $X_0 = 1$ g/L)

모든 경우 초기 배양액 부피(V_0)는 2 L이었고 발효조 최대 부피는 5 L이었다. 발효조에 공급되는 포도당은 45 g으로 고정하였다. 전산모사에 사용한 이진 염색체의 길이는 360비트이고 이중 각 유전자는 10비트의 크기를 가졌다. 집단 내 개체수는 100개이었다. 또, 교배율은 0.6, 돌연변이확률은 0.003으로 고정하였다.

유전알고리즘의 수렴성을 측정하기 위해 먼저 A)번의 초기조건을 가질 경우에 대해서 100,000번의 iteration을 수행하였다. Figure 5에 나타나 있듯이 약 10,000번의 iteration 후에는 전략의 종류에 상관없이 성능지수값이 수렴하는 것을 알 수 있었다. 이 결과를 감안하여 다른 초기조건에 대해서는 10,000번의 iteration을 하였는데 Figure 6은 계산된 최적 feed rate profile을 사용했을 경우(초기조건 (B))의 유가식 배양 전산모사 결과이다. 한번의 최적화 계산에 걸린 CPU 시간은 50분 정도였다(Pentium 133 MHz PC, RAM 32M).

각 전략의 비교

각 전략을 통해서 얻은 최적화결과들이 Table 2에 정리되어 있다. 모든 초기조건에 대해 세 번째 전략이 가장 우수한 성능을 보였고 초기 균체농도와 기질농도가 낮을 때를 제외하고는 첫 번째 전략이 두 번째 전략보다 나은 결과를 나타내었다. 두 번째 전략은 초기조건에 따른 최적 feed rate profile 형태를 미

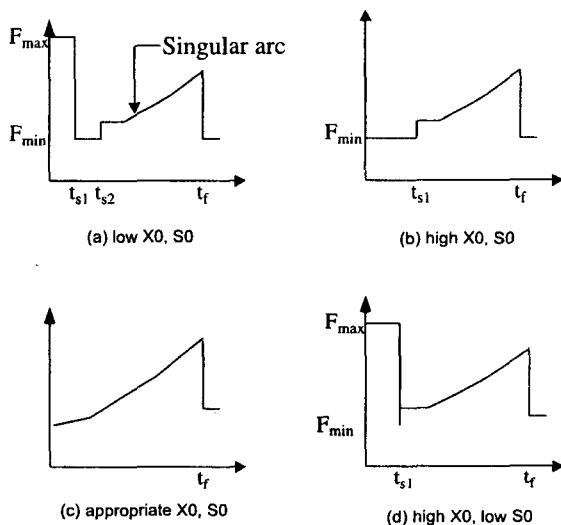


Figure 4. Feeding strategies for different initial conditions (ref. (7)).

리 t_{s1} , t_{s2} 에 대한 유전정보로 제공했음에도 불구하고 단순한 형태인 첫 번째 전략보다 나쁜 결과가 나왔다. 이는 다소 예상 밖의 결과이지만 다음과 같이 추론해 볼 수 있다. 두 번째 전략의 경우 switching time(t_{s1} , t_{s2})과 feed rate profile의 형태가 미리 제공되는데 이진 염색체 길이 360비트 중 switching time에 대해서는 20비트만이 제공되고 나머지 구간인 singular arc부분의 feed rate값에 340비트가 제공된다. 따라서 상대적으로 switching time값의 계산이 정확해지지 못할 가능성이 있다. 실제로 Modak(7)이 singular control approach로 구한 switching time과 두 번째 전략에 의해 유전알고리즘이 얻은 switching time의 값을 비교하면 많은 차이가 있음을 알 수 있었다(Table 3). 초기 균체농도와 초기 기질농도가 낮을 경우에 두 번째 전략이 나은 결과를 보인 이유는 Figure 5에 나타나듯이 iteration 초기에 이미 높은 적합도를 가지고 있었기 때문이라고 분석할 수 있다. 세 번째 전략의 경우 feed interval값에 변화를 줌으로써 구간이 고정될 때 생기는 쓸모 없는 유전정보의 발생을 상당

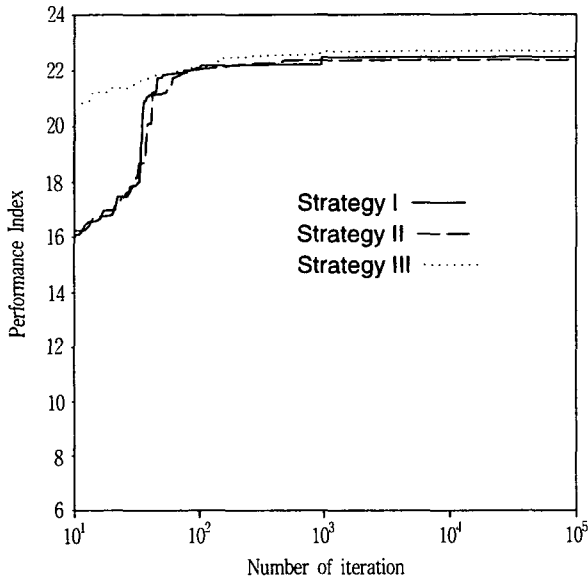


Figure 5. Results of convergence test.
 Strategy I : _____
 Strategy II : _____
 Strategy III : _____

히 막아주므로 전체 염색체의 유전자를 고르게 사용하여 좋은 결과를 나타낸다는 걸 알 수 있다.

유전알고리즘을 사용한 최적화 제어에서는 각 구간의 길이를 너무 작게 줄일 경우 구간수가 이에 반비례하여 늘어나며 따라서 계산량이 크게 증가하게 되므로 미분가능한 곡선형태가 아닌 계단형태의 feed rate profile을 얻을 수 밖에 없다. 따라서 각각의 유전정보들(정해진 구간에서의 feed rate)은 이론적으로는 최적점의 필요조건을 완벽하게 만족시켜 줄 수 없으며 유전정보가 모인 전체 염색체를 통해서만 이 최적점에 가까운 값을 구현하는 방향으로 수렴하게 된다. 이러한 면에서 볼 때 본 연구에서 사용한 유전알고리즘을 일종의 suboptimal approach라 할 수 있다.

유전알고리즘과 singular control approach의 비교

Modak(7)이 singular control approach에 얻은 결과는 문제의 간소화에도 불구하고 최적 feed profile과 switching time을 얻기 위해서는 adjoint variable이 도입된 복잡한 formulation이 필요하게 된다. 또한 최적화 과정이 두가지 루프 (1. switching

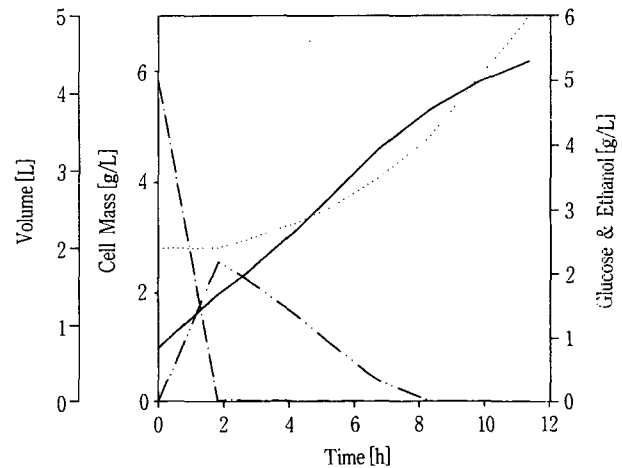


Figure 6. Optimal profiles of cell, glucose and ethanol concentrations, and volume for initial condition B ($X_0 = 1$ g/L, $S_0 = 5$ g/L).
 (Cell mass: _____ Glucose: _____ Ethanol: _____
 Volume: _____)

Table 2. Results of fed-batch yeast culture optimization for different initial conditions.

Case	Initial Cell Mass Conc. g/L	Initial Glucose Conc. g/L	Performance Index $IP(t_f)$			
			Genetic Algorithm			Singular control approach (ref. (7))
			Strategy I	Strategy II	Strategy III	
A	0.2	0.0005	21.67	21.86	22.42	21.52
B	1.0	5.0	24.18	24.10	24.45	23.39
C	0.68	0.065	23.42	23.42	23.75	25.22
D	1.0	0.0005	24.28	23.91	24.66	25.06

$V(0) = 2$ L, $V_{max} = 5$ L, $F_{max} = 1.0$ L/hr, $\epsilon = 0.1$ g/hr

total amount of glucose = 45 g, $G_F = \frac{45 - GV(0)}{V_{max} - V(0)}$ g/L

Table 3. Times for reaching the singular arc in strategy II.

Case	Initial Cell Mass Conc.	Initial Glucose Conc.	Arrival time at singular arc(hr)	
	g/L	g/L	Genetic Algorithm, Strategy II	Singular control approach (ref. (7))
A	0.2	0.0005	4.728	3.75
B	1.0	5.0	3.730	2.40
D	1.0	0.0005	0.070	0.007

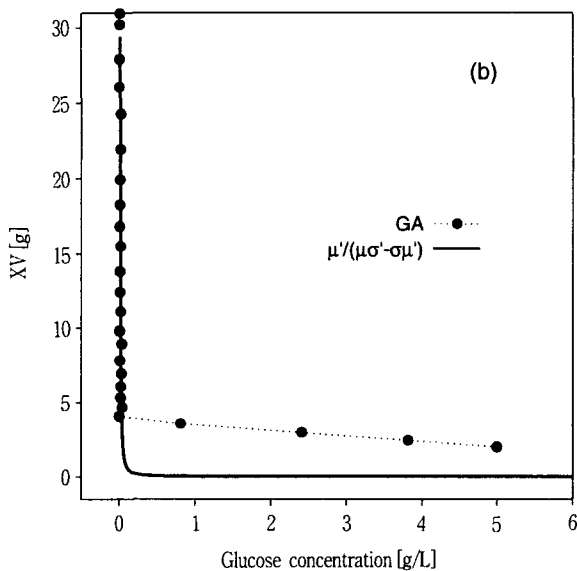
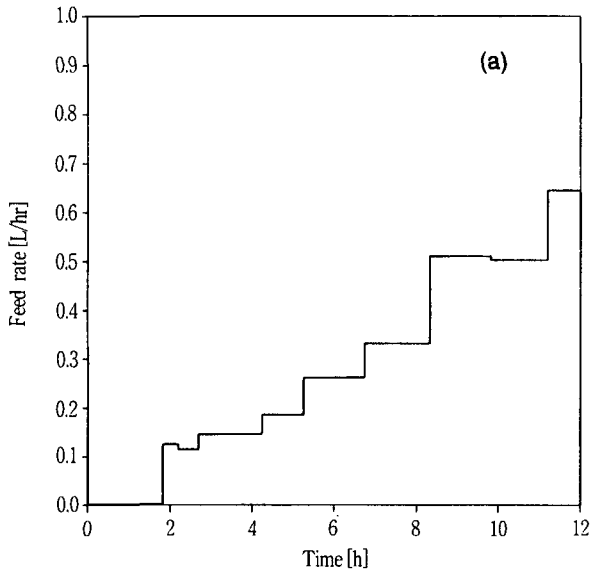


Figure 7. Comparison of optimal feed rate profiles and phase plane curves(genetic algorithm vs. singular control approach for initial condition B: $X_0 = 1.0$ g/L, $S_0 = 5.0$ g/L).

time에 대한 계산 2. adjoint variable 계산을 들게 되어 수렴이 극도로 어려워지는 문제가 발생한다. 이에 비해 유전알고리즘을 이용한 최적화 계산은 별도의 adjoint variable이나 미분이

필요없어 쉽게 formulation할 수 있다. Table 2를 보면 유전알고리즘에 의해 얻은 성능지수와 singular control approach로 구한 성능지수값이 나와 있다. 초기조건 A)와 B)에서는 유전알고리즘의 결과가 더 좋았고 C)와 D)에서는 singular control approach로 구한 값이 더 높았다. Singular control approach 결과보다 suboptimal approach인 유전알고리즘의 결과가 A)와 B)에서 더 좋은 이유는 두가지로 생각해 볼 수 있는데 첫째, singular control approach에 의해 얻은 결과가 local minimum에 빠져서 더 이상의 개선이 이루어지지 않았거나 둘째, 계산의 복잡성으로 인한 수렴성의 저하로 최적값을 찾지 못했을 가능성이 있다. 초기 조건이 바로 singular arc와 만나는 C)의 경우에는 switching time을 계산하는 과정이 없어져서 local minimum에 빠질 가능성이 적어지고 수렴이 빨라지므로 singular control approach가 더 나은 결과를 보인다고 생각된다. D)의 경우에는 singular arc와 도달하는데 걸리는 시간이 매우 짧아 singular control approach에 의한 방법으로 계산할 경우, 일찍 수렴값에 도달한 것으로 보인다. 그러나 C)의 경우는 singular arc와 만나는 초기조건을 실제 공정에서 맞추기가 거의 불가능하기 때문에 유전알고리즘과 singular control approach의 성능 비교를 위한 예로서 부적합한 면이 있다. 또한 유전알고리즘의 경우 formulation이 간단하고 초기조건에 영향을 거의 받지 않는 장점이 있어 실질적인 측면에서는 singular control approach에 비해 훨씬 유리하다는 걸 알 수 있다.

최적제어가 이루어질 때 singular 구간에서 균체질량과 포도당 농도사이의 관계(phase plane curve)는 다음과 같은 식으로 표현되어진다(3).

$$X = a \frac{\mu'}{\mu' \sigma - \sigma' \mu} \tag{12}$$

여기서 μ' , σ' 는 비성장속도와 포도당 소비속도를 포도당 농도에 대해 미분한 값이고 a 는 미지의 상수이다. Figure 7에서 볼 수 있듯이 유전알고리즘에 의한 결과중 feed rate는 singular control approach에 의해 계산된 profile과는 상당한 차이가 있는 것으로 보이나 포도당농도-균체질량 phase plane상의 궤적은 매우 유사함을 알 수 있다.

요 약

유전알고리즘을 사용하여 유가식 효모배양의 최적화를 수행하였다. 최적 포도당 공급속도 profile을 얻기 위해 세 가지 전략을 세웠다. 첫 번째 전략은 전체 배양시간을 일정한 크기의 구

간들로 나누고 각각의 구간에 대해 feed rate를 구하는 방식이고 두 번째 전략은 singular control approach에서 얻은 결과를 응용한 것으로 switching time들과 singular 구간에서의 포도당 공급속도를 구하는 방식이다. 세 번째 전략에서는 각 구간의 포도당 공급속도 뿐만이 아니라 구간의 크기도 계산하였다. 이 중 세 번째 전략이 효율적인 유전정보의 사용으로 가장 좋은 결과를 나타내었다. 유전알고리즘으로 얻은 성능지수값은 간단한 수식화과정에도 불구하고 variational calculus와 singular control approach를 통해 얻은 값과 거의 유사하거나 더 높은 값을 보였다.

사 용 기 호

Upper case

X	: Cell mass concentration (g/L)
S	: Substrate concentration (g/L)
E	: Ethanol concentration (g/L)
V	: Volume (L)
F	: Feed rate (L/hr)

Lower case

t_{s1}, t_{s2}	: Switching times (hr)
------------------	------------------------

Greek

ε	: Relative operating cost (g/hr)
μ	: Specific grown rate (1/hr)
σ	: Specific glucose consumption rate (1/hr)
π	: Specific ethanol formation rate (1/hr)
η	: Specific ethanol consumption rate (1/hr)

Subscripts

E	: Ethanol
-----	-----------

F	: Feed
f	: Final condition
G	: Glucose
min	: Minimum
max	: Maximum
s	: Singular
0	: Initial condition

참 고 문 헌

1. Pontryagin, L. S., V. G. Boltyanskii, R. V. Gramkrelidze, and E. F. Mischenks(1962), The Mathematical Theory of Optimal Process, Wiley, New York.
2. Ohno, H, E. Nakanishi, and T. Takamatsu(1976), Optimal Control of a Semibatch Fermentation, Biotech. Bioeng., 18, 847-864.
3. Modak, J. M., H. C. Lim and Y. J. Tayeb(1986), General Characteristics of Optimal Feed Rate Profiles for Various Fed-Batch Fermentation Processes, Biotech. Bioeng., 28, 1396-1407.
4. Modak, J. M. and H. C. Lim(1989), Simple Nonsingular Control Approach to Fed-Batch Fermentation Optimization, Biotech. Bioeng., 33, 11-16.
5. Michalewicz, Zbigniew(1992), Genetic Algorithm + Data Structures = Evolution Programs, Springer-Verlag, Tokyo.
6. Goldberg, D. E.(1989), Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, New York.
7. Modak, J. M.(1988), A Theoretical and Experimental Optimization of Fed-Batch Fermentation Processes, Ph. D. Thesis, Dept. of Chem. Eng., Univ. of Purdue., West Lafayette, Indiana, USA.