

유전 알고리즘을 이용한 유리 용해 공정에서의 불량 예측 시스템†

정호상 · 정봉주

연세대학교 산업시스템공학과

A Quality Forecasting System in Glass Melting Processes using Genetic Algorithms

Ho-Sang Jung · Bong-Ju Jeong

This paper presents a computerized quality forecasting system for glass manufacturing. In forecasting the molten glass quality, we are concerned with three major issues : (1) to find the reasonable time lags between a set of process conditions and the quality measurement of glass products, (2) to find the most significant process variables affecting the quality, and (3) to construct the appropriate causal forecasting models using genetic algorithms. The experimental results show the proposed model results in better forecasting than linear regression model. The suggested forecasting model was implemented successfully and is being currently used in a real manufacturing line.

1. 서론

유리 용해 공정은 일반 제조 공정과는 달리 거대 투자가 필요한 자본 집약적인 설치 기반 산업으로 용해로(Furnace)를 이용해 주원료가 모래인 재료들을 용해시킨 다음, 일정한 온도 및 점도를 유지시키며 양질의 유리 용해 제품을 생산해 내는 공정이다. 유리 용해 공정은 유체흐름의 특성상 공정 주위에 위치한 수많은 공정 인자들의 영향을 받기 때문에 사실상 신뢰할 만한 불량 예측이 매우 힘든 공정 중의 하나이다. 특히 용해로 내의 대류 현상에 의한 유체의 움직임은 예측이 거의 불가능하다고 알려져 있다. 공정 인자는 크게 투입되는 원료의 양 및 배합비, 연료량, 공기량 등의 제어 인자들과 용해로 및 공정 내 각 부분의 온도 및 압력 등을 나타내는 관측 인자 등으로 나뉜다. 이러한 공정 인자들은 미세한 변화에도 제품의 품질에 큰 영향을 끼치며, 공정 제어 시스템에서 수집되는 관련 데이터의 종류와 크기는 매우 방대하다. <그림 1>에서 보는 바와 같이 유리 용해 공정은 우선 주원료인 모래를 비롯한 여러 원료들을 용해로(Furnace)에 투입한 다음 특정 시간 동안 용해시킨다. 용해시키는 과정 중에는 용해 물질의 품질 유지 및 불량 방지를 위해 관측 인자들을 통해 용해로의 상태를 수시로 감시하며 제어 인자를 통해 용해로 내의 관측 인자들, 즉 온도 및 압력 등을 원하는 수준으로 유지한다. 용해로를 빠져 나온 유

리 용해 물질은 용해관(Forehearth)을 따라 흘러가며 일정한 온도 및 점도 등을 유지시킨다. 마지막으로 정제로(Refinery)를 통과한 유리 용해 물질은 제품의 종류 및 규격에 따라 다양한 형태의 반제품 또는 완제품으로 나오게 된다.

본 연구에서는 유리 용해 공정에 있어서 유전 알고리즘을 사용하여 제품의 불량 예측을 할 수 있는 예측 모델을 개발하고 실험을 통해 실제 유리 용해 공정에 적용 가능함을 보이는 것을 목적으로 한다. 기존의 불량 예측은 주로 회귀 분석 등에 의한 통계적 모형을 통해 이루어졌는데, 이러한 통계적 분석은 과거 일정량의 공정 데이터를 이용해 최적화된 모델을 구축하고 앞으로의 예측을 수행하는 데 있어서 비교적 효율적인 방법이기도 하나, 데이터의 무작위성(Randomness) 또는 공정의 변동(Process Variability)이 심할 경우 예측의 신뢰도가 크게 저

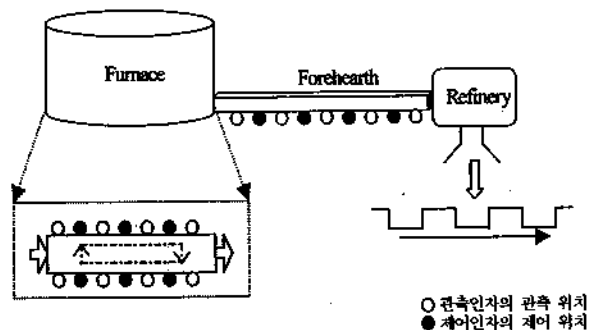


그림 1. 유리 용해 공정의 구조.

†본 연구는 한국 과학 재단 목적 기초 연구(과제번호: 971-1013-069-2) 지원으로 수행되었음.

하되며, 특히 현실적으로 최근의 공정 데이터들에 대한 가치를 예측 모델 차원에서 반영하기 어렵다는 단점이 있어 왔다. 반면, 기존의 여러 발견적 기법들은 그 기법상의 용이함에 비해 복잡하고 넓은 해 공간을 효율적으로 탐색하지 못함으로 인한 제한된 예측만이 가능하다는 단점을 가지고 있다.

유전 알고리즘은 1975년 Holland (1975)가 정립한 발견적 기법으로 생물의 진화 과정을 추상화한 것이다. 유전 알고리즘은 잠재해들로 이루어진 초기 모집단을 세대를 진행하며 목적 함수를 최적화할 수 있는 해를 찾아가는 발견적 기법으로, 지금까지 많은 분야에 적용되어 왔다. 순환방문 판매원 문제 (Traveling Salesman Problem)와 같은 몇몇 유명한 문제에서 유전 알고리즘은 좋은 결과를 얻었으며 (Whitley et al., 1989; Homaifar et al., 1991; Syswerda, 1991), Job-Shop 스케줄링 문제 (Cross et al., 1995; Kobayashi et al., 1995; Dondorf and Pesch, 1995), 통신 네트워크의 설계 (Tanaka and Berlage, 1996), 철도회사의 배송계획의 최적화 (Gabbent et al., 1991) 등에도 적용되고 있다. 유전 알고리즘을 이용하여 예측 모델을 구축한 사례는 Chiraphadhanakul et al. (1997)의 연구가 있는데, 은행의 예금량을 연별로 예측하는데 유전 알고리즘을 사용하였으며, Goto et al. (1999)는 유전 알고리즘의 구조를 이용해 일별 전기 부하량을 예측하는 모델을 제안하였다. 또한, 일반적으로 유전 알고리즘은 신경망 기법 등의 다른 발견적 기법들과 함께 사용되어 왔는데, Kai and Wenhua (1997)의 연구에서는 두 가지 발견적 기법을 혼합하여 주가를 예측할 수 있는 모델을 발견할 수 있다. Yip et al. (1997)은 두 가지 기법을 순차적으로 사용하는 판매 예측 모델을 제안하였으며, Kung et al. (1998)은 신경망이 내재된 유전 알고리즘 모델을 사용하여 전력 부하에 대한 예측을 수행하였다.

본 연구에서는 그 종류와 양에서 방대한 공정 인자들을 가지고 있으며 용해로 내의 복잡한 공정 매커니즘을 가진 유리 용해 공정을 대상으로, 공정의 특성을 반영하고 주요 공정 변수를 선택하며 유전 알고리즘을 이용한 보다 효과적이고 정확한 불량 예측이 가능한 모델을 개발하고 검증하였다. 불량 예측 모델 검증을 위한 실험은 실제 유리 용해 공정의 데이터를 사용하였고 유사한 조건에 있는 4개의 공정 라인에 대해서 이루어졌다. 각 공정 라인에 대해 회귀 분석 모델을 이용한 불량 예측 실험과 유전 알고리즘을 이용한 불량 예측 실험을 하였고 그 결과를 비교하였다.

본 논문의 구성은 다음과 같다. 2장에서는 불량 예측 모델을 제안하며, 3장에서는 불량 예측 모델의 계수 결정을 위한 유전 알고리즘의 구조를 제시하고, 4장에서는 제안된 불량 예측 모델을 실제 라인에 적용한 결과를 분석하고, 구현된 시스템을 제시한다. 마지막으로 5장에서는 본 연구에 대한 결론을 맺는다.

2. 불량 예측 모델

2.1 예측 모델

유리 용해 공정의 제품 불량에 영향을 미치는 수많은 공정

인자들이 존재하나 이 중 실제로 관리하는 공정 인자들은 일부분인 70~80여개에 지나지 않으며, 이들 관리 대상 인자들에 대해서만 데이터가 실시간으로 기록, 보존되게 된다. 따라서 불량 예측은 기록된 과거의 실시간 공정 인자 데이터만을 사용하여 현재 시점으로부터 특정 시점까지의 미래를 예측하게 된다. 본 연구에서의 유전 알고리즘을 이용한 불량 예측 모델은 현재 유리 용해 공정에서 가장 많이 이용되고 있는 선형 인과 모델을 토대로 불량 지표와 공정 인자들의 데이터를 용해 식 (1)과 같이 구성한다.

$$y(k+1) = b_0 + b_1x_1(k) + b_2x_2(k) + \dots + b_nx_n(k) + e(k) \quad (1)$$

여기서, k : 시점 k

$x_1(k) \sim x_n(k)$: 공정인자 1부터 n 에 대한 현재 시점 k 에서의 수치

$y(k+1)$: 시점 $k+1$ 에서의 불량 지표치 (예: 불량율)

$b_i (i = 0, \dots, n)$: 불량 예측 인과 모델의 계수값

$e(k)$: 시점 k 에서의 보정치

제안된 모델은 y 값으로 사용될 불량 지표와 x 값으로 사용될 공정 인자들에 대한 데이터는 공정 내 수집된 데이터를 이용하고, 모델의 계수값들은 유전 알고리즘을 사용하여 결정하고자 한다. 특히 효율적이고도 정확한 불량 예측을 하기 위해 다음의 네 가지 사항에 초점을 맞추어 불량 예측 모델을 구성한다.

첫째, 실제 공정을 가능한 정확히 반영하는 현실적인 불량 예측이 될 수 있도록, 모델에 사용되는 데이터들 간의 공정 Lag Time을 결정하고 반영한다. Lag Time을 결정하는 방법은 그래프를 이용한 경향분석법을 사용한다.

둘째, 불량 예측 모델을 구성하는 공정 인자들의 수를 변수 선택 과정을 통해 적정 수준으로 줄인다. 현장에서 수집되는 모든 인자들에 대한 방대한 양의 데이터들은 사실상 공정 개선을 위한 자료로 그대로 이용하기도 어려울 뿐만 아니라 많은 양의 데이터들은 거의 분석에 이용되지도 않고 저장 공간만 차지하는 경우가 많다. 따라서 예측 모델 구성의 용이성과 효율성 측면에서 적절한 수의 변수들을 선택하는 과정을 도입한다.

셋째, 최종 불량 예측 모델을 결정하는 데 사용되는 데이터 기간을 회귀 분석을 이용한 모델 성숙 기간, 유전 알고리즘을 이용한 모델 성숙 기간의 두 부분으로 구분한다. 회귀 분석을 이용한 모델 성숙 기간을 통해 유전 알고리즘의 초기 모집단을 보다 가능성 있는 잠재해들로 구성하고, 유전 알고리즘을 이용한 모델 성숙 기간을 통해 실제 예측을 위한 최종 예측 모델을 구성한다.

넷째, 유리 용해 공정의 불량 예측을 하기 위한 선형 인과 모델을 구성하는 데 적합한 유전 알고리즘을 제시한다. 특히, 유

전 알고리즘 성숙 기간을 진행해 나감에 따라 우수한 해를 선별해 낼 수 있도록 지난 세대들의 데이터를 함께 고려할 수 있는 적합도 함수(Fitness Function)를 사용하고, 유전 알고리즘을 이용해 결정된 최종 예측 모델의 성능 향상을 위해 보정치를 사용한다. 보정치는 그래프를 이용한 방법과 과거 예측치의 성능을 반영하는 방법을 함께 고려하여 산출한다.

2.2 Lag Time 결정 방법

일반적으로 Lag Time이란 어떤 인자를 변화시켰을 때 그 결과로 다른 인자가 영향을 받을 수 있는 때의 그 시간적인 차이를 말한다. 생산 공정에서 진행중인 재공과 완성된 제품 사이에도 일정한 시간적인 차이, 즉 Lag Time이 존재한다. 특히 공정에서 얻어진 데이터들, 즉 공정 인자의 값들은 관측한 시간이 각각 다르기 때문에 공정 인자 사이에도 <그림 2>와 같은 Lag Time이 존재한다.

일반적으로 Lag Time을 l 이라고 놓았을 때, 시점 t 에서의 공정 조건은 시점 $t+l$ 에서의 유리 용해 제품의 불량 여부에 영향을 준다. 정확하고 보다 현실적인 불량 예측을 위해서는 Lag Time을 데이터에 반영해 주는 것이 필요한데, 일반적인 제조 공정과는 달리 유리 용해 공정과 같은 연속 생산 공정의 경우는 용해로 내의 온도분포에 따른 유체흐름의 특성상 정확한 Lag Time을 산출하는 것이 매우 어렵다. 즉, 특정 유리 용해물의 위치와 그 흐름을 추적하는 것이 사실상 불가능하다. 일반적으로 용해로를 통과하는 공정시간, 즉 원료투입에서 용해로 산출 시점간의 Lag Time을 구하기 위해서는 화학적 추적 물질을 사용하는데, 추적 물질을 원재료와 함께 용해로에 투입하여 그 투입한 시점과 추적 물질이 용해 공정 내의 용해관(Forehearth)의 끝 지점에서 감지되기 시작하는 시점까지를 계산하여 결정한다. 그러나 공정 주위의 수많은 공정 인자들 사이의 Lag Time을 결정하기 위해서는 위와 같은 추적 물질 사용법은 현실적으로 불가능하므로 본 연구에서는 다음과 같은 Lag Time 결정 방법을 사용한다.

제안된 Lag Time 결정 방법은 그래프를 사용한 경향 분석법을 기본으로 한다. 두 개의 서로 다른 인자들 간의 관계를 규명하기 위해서 두 변수의 시간별 데이터의 추이를 동시에 그래

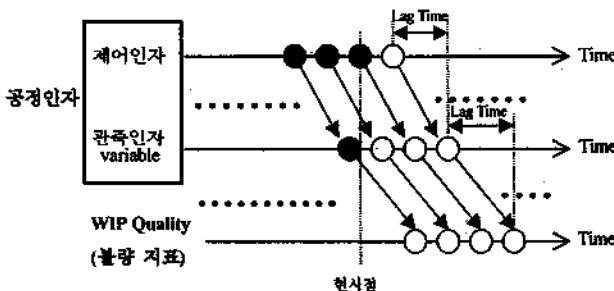


그림 2. 공정 내 Lag Time의 개념.

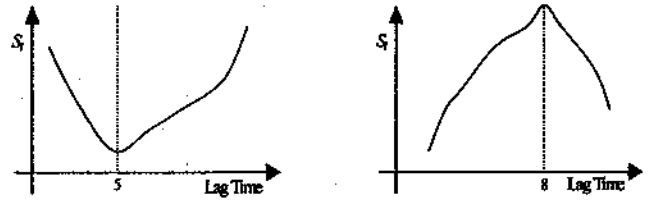


그림 3. 공정 인자간 유사성 지표(SI) 그래프의 형태.

프에 표시한 후 그 경향을 분석하는 방법을 사용한다. 우선, 두 변수가 보여주는 시간별 경향의 유사성(S_l) 지표를 식 (2)를 이용해 산출한 후 두 변수간의 Lag Time을 결정한다.

$$S_l = \sum_{i=1}^n |d_{ii} - \bar{d}_l| / n \quad (2)$$

여기서, $d_{ii} = q_t - q_{t+l}$ and $\bar{d}_l = \sum_{i=1}^n (p_t - q_{t+l}) / n$

p_t : 시점 t 에서의 공정 인자의 값

q_{t+l} : 시점 $t+l$ 에서의 불량지표의 값

산출된 유사성 지표는 두 개의 서로 다른 인자간의 차이의 분산의 평균을 뜻하는 것으로 이 수치가 0이면 두 개의 인자 p 와 q 는 Lag Time l 의 차이를 가지고 정확히 같은 경향을 보인다고 말할 수 있다. 따라서 <그림 3>과 같이 Lag Time l 을 변화시켜 가며 유사성 지표 S_l 수치를 그래프로 도시해 보면 Lag Time의 선택에 따라 두 인자간의 유사성 정도가 어떻게 변화하는지 알 수 있다. 만약 S_l 그래프가 Convex 형태를 그리게 되면 그래프에서 적시된 최소값 시점에서 다른 Lag Time들에 비해 가장 높은 양의 유사성(같은 방향의 경향)을 보인다고 할 수 있으며 S_l 그래프가 Concave 형태를 그리게 되면 그래프에서 적시된 최대값 시점에서 다른 Lag Time들에 비해 가장 강한 음의 유사성(반대 방향의 경향)을 갖는다고 볼 수 있다.

그래프에 따라 특정한 형태가 나타나지 않는 경우는 두 개의 인자가 서로간에 직접적인 영향이 없거나 두 개 이외에도 다른 인자들의 영향을 받을 수 있다는 것을 뜻한다. 이와 같은 인자의 경우는 정확한 Lag Time을 사실상 결정하기가 어려우므로 공정상의 실제 위치를 환산하거나 공정 전문가의 경험적 의견을 반영하였다. 이렇게 결정된 Lag Time의 수정은 주기적으로 새로이 결정하거나, 공정 조건 등의 특이한 변화가 일어났을 때마다 하게 된다.

2.3 주요 변수선택

공정 인자의 종류와 수가 많을 때, 모든 사용 가능한 인자들 을 가지고 적합한 예측 모델을 만든다는 것은 어려운 일이다. 실측 데이터를 계속 실시간으로 수집하고 관리하는 데 많은 노력과 계속 설비가 필요하고 수집된 방대한 데이터를 저장, 분석하는 데 소요되는 시간적, 공간적 손실은 매우 크다. 더구

나 모델에 포함되는 인자의 수와 예측의 성능은 비례관계가 아니라는 사실이다. 따라서 변수 선택 과정을 통한 주요 공정 인자를 선정하여 예측 모델의 효율성을 크게 향상시킬 수 있다.

인자 선정 방법은 일반적으로 Forward Selection, Backward Elimination, Stepwise Method 등의 방법이 널리 사용되는데, 본 연구에서는 모든 변수들에 대한 일차적인 고려가 가능하고 시스템 구축의 용이성을 고려하여, 모든 공정 인자들이 포함된 초기 모델에서 시작하여 기여도가 가장 낮은 변수들을 차례대로 하나씩 소거하여 나가는 방법인 Backward Elimination 방법을 사용하였다 (Flurry and Reidwyl, 1988; Afifi and Clark, 1990). 본 연구에서 사용된 변수 선택 과정은 다음과 같다.

(단계 1) 변수 선택 과정에 사용될 데이터를 크게 Training 기간, $[t_2, t_1]$ 과 Evaluation 기간, $[t_1, t]$ 의 두 부분으로 구분한다. 단, $t_2 < t_1 < t$ 이며 t 는 현재 시점을 뜻한다.

모든 공정 인자들로 포함되는 전체 모델을 구성한다.

(단계 2) 공정 인자들에 대한 소거 순서 리스트를 작성한다. 모든 인자들이 포함된 모델에서 시작하여 하나의 인자만이 포함된 모델이 될 때까지 하나씩 변수를 소거해 나간다. 변수 소거에는 Training 기간 동안의 데이터에 대한 F-test를 이용한다.

(단계 3) 평가 기간 동안의 데이터를 이용해 (단계 2)에서 작성한 소거 순서 리스트에 의거해 각 모델에 대한 MAD(Mean Absolute Deviation) 수치를 다음과 같이 계산한다 (Winston, 1991).

$$MAD = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (3)$$

여기서, y_i : 시점 i 에서의 실측치, $i=1, 2, \dots, n$

\hat{y}_i : 시점 i 에서의 예측치, $i=1, 2, \dots, n$

(단계 4) MAD 수치 중 가장 작은 수치를 기록한 모델에 포함되어 있는 변수들을 최종 모델에 사용할 공정 인자로 결정한다.

2.4 전체 데이터의 기간 설정

불량 예측에 필요한 기간을 현재 시점을 기준으로 하여 <그림 4>와 같이 세 부분으로 나눈다.

- 회귀 분석을 이용한 모델 성숙 기간 (Regression Training Period)
- 유전 알고리즘을 이용한 모델 성숙 기간 (Genetic Algorithms Training Period)
- 불량 예측 기간 (Forecasting Period)

먼저 회귀 분석을 이용한 모델 성숙 기간을 통해 유전 알고리즘의 초기 모집단을 보다 가능성 있는 잠재해들로 구성한다. 유전 알고리즘은 초기 모집단의 설정이 최종해의 성능에 영향을 미치기 때문에 회귀 분석을 통해 현재 공정의 예측 가능한 여러 해들을 기본 모집단으로 구성함으로써 전체 모델의 성능을 높이는 방법을 사용한다. 우선, 유전 알고리즘을 이용한 모델 성숙 기간의 데이터를 이용, 회귀 분석을 수행하여 그 이후 시점을 예측하는 모델을 구축하고, 구축된 모델에 사용된 계수값들을 중심으로 $\pm 10\%$ 의 범위 내에서 유전 알고리즘에서 사용할 초기 모집단을 구성한다.

유전 알고리즘을 이용한 모델 성숙 기간의 시작 시점에서는 앞에서 구성된 유전 알고리즘의 초기 모집단을 사용하여 교차, 돌연변이, 선별 등의 유전 연산을 수행하게 되며 일정 세대를 반복하여 다음 시점을 위한 모델을 생성하게 된다. 다음 시점에서는 앞서 구축된 모델에서 사용된 계수를 포함해 모집단을 새로이 구성한 후 동일한 방법으로 모델을 구성해 가며 유전 알고리즘을 이용한 모델 성숙 기간의 끝 시점까지 모델을 성숙시켜 나간다.

마지막으로 실제 불량 예측 기간에는 앞의 두 기간 동안 성숙된 불량 예측 모델을 사용하여 실제 예측을 수행하고 그 정확성을 검증한다.

3. 불량 예측 모델을 위한 유전 알고리즘

유전 알고리즘은 잠재해를 표현한 개체(Chromosome)들로 이루어

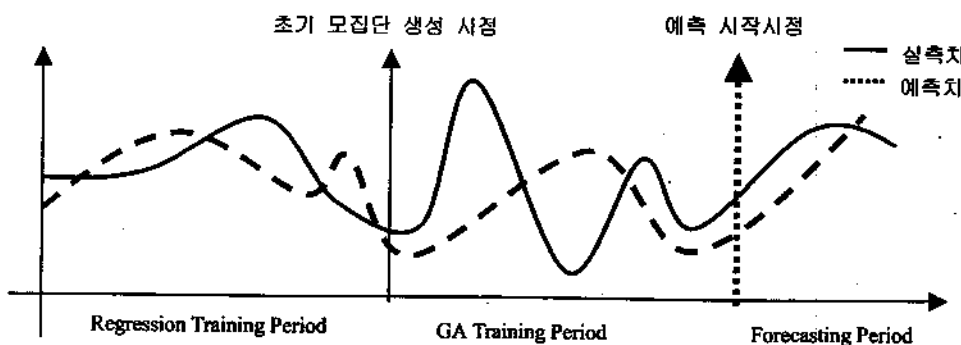


그림 4. 데이터 기간의 분류.

어진 모집단(Population)을 초기화하고, 선별 과정(Selection)을 통해 다음 세대에 생존할 개체들을 선별한다. 선별된 개체들에 대해 교차(Crossover), 돌연변이(Mutation) 등의 유전 연산을 시행하고 그 결과로 산출된 자식 개체들에 대해 적합도 함수를 이용한 적응도 평가 과정을 거치며 해 공간을 탐색하고, 최적해를 찾는다. 특히 유전 알고리즘은 복잡한 해 공간의 탐색 능력이 우수하여 변수와 제약이 많은 대형 수리 문제를 푸는데 적합한 기법이고, 또한 모형에 대한 유연성이 높아 제약 참가나 목적 함수의 변경이 용이하다는 장점을 갖는다. 따라서 유리 용해 공정과 같이 제품의 품질에 영향을 미치는 공정 인자들의 종류와 양이 방대한 공정에는 유전 알고리즘의 적용이 가능할 것으로 보인다. 일반적인 유전 알고리즘의 구조는 다음 < 그림 5 >와 같다(김여근 등, 1997).

```

begin
  t ← 0
  P(t)의 초기화
  P(t)의 적응도 평가
  while (종료조건이 만족되지 않으면) do
    begin
      t ← t+1
      P(t-1)로부터 P(t)를 선별
      P(t)에 대한 유전연산
      P(t)의 적응도 평가
    end
  end
end
    
```

그림 5. 유전 알고리즘의 구조.

3.1 유전 알고리즘의 구성 요소

3.1.1 개체 표현

유전 알고리즘을 특정 문제에 적용함에 있어서 먼저 결정해야 하는 점이 유전 알고리즘에서 사용할 개체의 표현 문제이다. 개체를 어떻게 표현하느냐에 따라 실험의 전체 효율과 문제에 대한 정확한 유전 알고리즘의 적용이 가능하게 된다. 일반적으로 많은 유전 알고리즘에서는 개체를 이진법을 사용하여 표현해왔다. 이진법의 경우 표현과 처리의 용이함으로 인해 많은 응용 문제에 대해 효과적으로 적용되어 왔다. 본 연구에서도 이러한 이진법을 이용한 새로운 표현 방법을 제안한다. 기본 개념은 상이한 범위를 가진 서로 다른 데이터들을 0과 1 사이로 Scaling하는 방법을 이용해 불량 예측 모델에 사용할 개체 표현 기법을 정의한다. 우선 하나의 개체(Chromosome)는

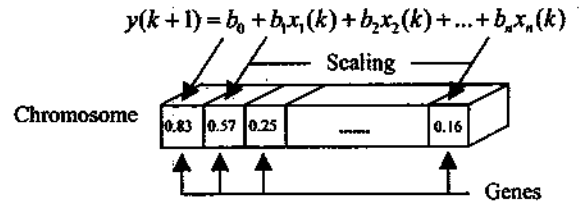


그림 6. 불량 예측 모델을 위한 개체 표현.

개체를 구성하는 여러 인자(Gene)들로 이루어지게 되는데, 본 연구에서는 인자(Gene)들을 각각 선형 인과 모델의 계수들과 일치시킨다. 일반적인 단순 개체 표현법을 사용하게 되면 각기 다른 범위의 수치들로 개체가 구성되는데, 교차, 돌연변이 등의 유전 연산을 일괄적으로 수행하려면 개체 형태를 일치시켜 줄 필요가 생긴다.

본 연구에서는 <그림 6>과 같이 각기 상이한 범위의 데이터들을 연산이 용이하도록 0과 1 사이의 값들 중에서 임의로 발생시켜 하나의 인자를 만들고, 예측 모델에서 필요한 만큼 인자를 발생시키는 과정을 반복하여 하나의 개체를 구성한다.

3.1.2 초기 모집단의 구성

유전 알고리즘은 잠재해들로 구성된 모집단을 운용하며 최적해를 찾아내는 방법으로 초기 모집단의 선정이 전체 해의 품질에 큰 영향을 끼친다. 일반적으로 전통적인 유전 알고리즘은 초기 모집단을 임의의 개체들을 생성하여 구성하는 방법을 많이 사용한다. 그러나, 유전 알고리즘 이외의 다른 발견적 기법을 사용하여 구한 부분 최적해를 초기 모집단에 추가하는 방법이 해의 개선효과를 뛰어나게 하는 것으로 알려져 있다.

임의의 개체들로 초기 모집단을 구성하게 되면, 최적해에 도달하는 데 걸리는 시간이 길게 되고 유전 연산과 선별 과정만으로는 지역 최적해에서 벗어날 수 없는 경우가 생길 수 있다.

본 연구에서는 초기 모집단을 구성하는 데 있어서 회귀 분석을 사용한다. 회귀 분석을 이용한 초기 모집단의 구성은 최적 예측치를 얻을 수 있는 가능성을 높일 수 있으며, 전체 실험 시간면에서도 보다 빠른 시간 내에 적절한 범위 내의 해로 수렴할 수 있다는 장점이 있다. 전체 데이터 기간 중 회귀 분석을 이용한 모델 성숙 기간 동안의 데이터에 대해 회귀 분석을 수행하고 그 결과로 구축된 예측 모델의 계수값들을 Scaling한다. 다음으로 Scaling한 인자값들을 기준으로 $\pm 10\%$ 범위 내에서 임의로 인자를 발생시켜 하나의 개체를 완성한다. 모집단을 구성하는 개체의 적정수가 채워질 때까지 위와 같은 방법을 반복하여 초기 모집단을 구성한다. 회귀 분석을 이용한 초기 모집단의 구성 과정을 간략히 묘사하면 <그림 7>과 같다.

3.1.3 유전 연산

유전 알고리즘의 유전 연산에는 크게 교차(Crossover)와 돌연변이(Mutation)가 존재한다. 교차는 부모 형질이 가능한 파괴되지 않고 자손에 상속될 수 있게 하는 역할을 하며 돌연변이의

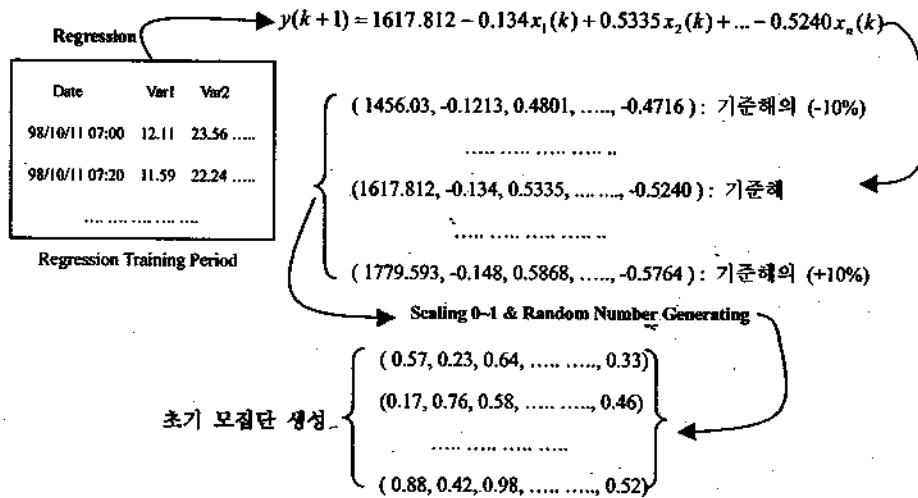


그림 7. 초기 모집단의 구성.

경우는 하나의 개체에서 임의의 인자들을 변화시켜 주는 방법을 사용하여 해 공간을 다양하게 탐색하게 하는 역할을 수행한다. 교차에는 전통적으로 일점교차, 이점교차, 순서교차, 위치교차 등의 방법이 있고, 돌연변이의 경우는 교환, 삽입, 역순, 전위 등의 방법이 사용된다.

본 연구에서는 가장 대표적이고 다양한 문제에 대한 적용성이 뛰어난 이점 교차를 교차 방법으로 사용하였고, 돌연변이의 경우는 개체 내 임의의 한 인자를 다른 임의의 인자와 자리 바꿈하는 방법(Swap)을 사용하였다(Goldberg, 1989). 또한, 선택된 개체들에 대해 교차와 돌연변이를 모두 수행하는 기존의 방법이 아닌 <그림 8>과 같이 선택된 개체들이 교차율(P_c)과 돌연변이율(P_m)에 따라 교차와 돌연변이 중 하나를 수행하게 했으며 교차 과정 내부에 돌연변이 확률을 부여해 교차로 결정된 개체 중 일부는 돌연변이 연산을 수행한 후에 교차를 수행하게 하였다. 전통적인 유전 알고리즘은 개체 내의 인자 단위로 돌연변이를 수행하거나 개체에 따라서는 교차나 돌연변이를 거치지 않고 바로 다음 세대로 넘어가는 경우가 발생하는데, 이런 경우에 전체 유전 연산의 효율을 떨어뜨리거나 세대가 진행되더라도 우수한 개체들과 열등한 개체들간의 생존율의 차이가 크지 않을 수 있다는 단점이 있다. 본 연구에 사용된 유전 연산의 경우는 교차 연산 내부에도 돌연변이 연산을 일정 비율 내에서 수행하게 함으로써 하나의 개체에 대해 교차와 돌연변이를 모두 수행하는 경우와 교차만 수행하는 경우, 돌연변이만 수행하는 경우 등의 여러 가지 형태의 유전 연산을 고려할 수 있는 장점이 있다.

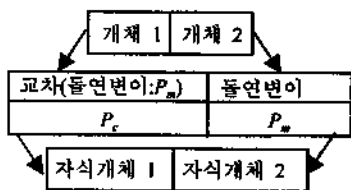


그림 8. 유전 연산의 구조.

3.1.4 선별 방법

선별 과정은 유전 알고리즘에서 현 세대의 모집단으로부터 다음 세대에 생존할 개체를 선택하는 과정으로 해의 산출에 중요한 역할을 하는 부분이다. 일반적으로 선별 규칙은 크게 두 가지로 구분할 수 있다.

- 비례 선별(Proportionate Selection)
- 순위 기반 선별(Ordinal-Based Selection)

비례 선별은 개체를 선정함에 있어 모집단 내의 개체들의 적합도를 기반으로 하며, 순위 기반 선별은 개체들의 모집단 내의 적합도의 상대적 순위에 따라 다음 세대로 갈 개체들을 선별하는 방법이다. 비례 선별 방법에는 Roulette Wheel 선별 방법, Stochastic Universal 선별 방법 등이 있으며, 순위 기반 선별 방법에는 Tournament 선별 방법, Linear Ranking 선별 방법 등이 있다(Miller and Goldberg, 1995). 본 연구에서는 위의 선별 방법과는 다른 MGG (Minimal Generation Gap) 모델 기반 선별 방법을 사용하였다(Shi, 1997). MGG 모델 기반 선별 방법은 유전 알고리즘 내부에 선별 과정을 따로 두지 않고, 유전 연산 단계에서 선택된 개체들에 대해 유전 연산을 수행한 후 나온 자식 개체들과 수행 전의 부모 개체들을 비교하여 적용도가 높은 2개의 개체만을 모집단에 대입하여 새로운 모집단을 구성하는 방법

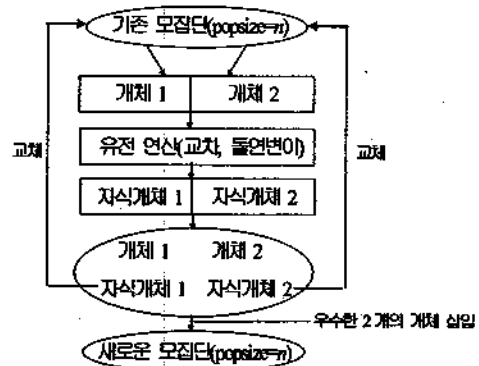


그림 9. MGG 모델 기반 선별 방법.

이다. 기존의 선별 방법들은 선별의 대상이 되는 개체들을 유전 연산을 거친 후에 생성된 새로운 자식 개체들로만 한정하기 때문에 부모 개체들은 성능이 우수하더라도 다음 세대로의 생존이 불가능하다는 단점이 있다. 본 연구에서 선택한 MGG 모델 기반 선별 방법은 위와 같은 단점을 극복한 방법으로, 유전 연산을 통해 나온 개체들을 포함해 모집단 내의 모든 개체들에 대해 선별을 고려할 수 있으며, 따라서 우수 개체들의 손실을 최소화시킬 수 있다는 장점이 있다. 일례로 Job-Shop 스케줄링 문제에 있어서는 MGG 모델 기반 선별 방법이 다른 선별 기법들에 비해 좋은 성능을 보인다는 연구 결과가 나와 있다 (Shi, 1997).

3.1.5 적합도 함수

목적 함수는 유전 알고리즘의 선별에 사용되며 적자 생존의 자연 법칙에 기초하여, 환경에 대한 적응도에 의해 현 세대의 모집단으로부터 다음 세대에 생존할 개체를 선택하는 과정으로 해의 산출에 중요한 역할을 하는 부분이다. 일반적으로 적응도를 평가하는 적합도 함수로는 최적화 문제의 목적 함수 자체를 주로 사용한다.

불량 예측 모델의 경우, 적응도를 평가하는 주된 기준이 개체들로 표현된 예측 모델을 이용해 얻어진 예측값과 실제 관측값과의 차이로부터 얻어지는 예측의 정확성 여부이다. 우선 예측의 정확성을 높이기 위해 적합도 함수를 이용해 얻어지는 예측치의 정확성과 신뢰성을 높이는 것이 중요하다. 본 연구에서는 적합도 함수를 구성하는 데 있어서 크게 두 가지 사항에 초점을 맞추었다. 첫째, 현재 시점에서의 예측의 정확성이다. 유전 알고리즘을 이용해 구축된 예측 모델이 현재 시점에서 어느 정도 예측의 정확성을 보이느냐를 가장 중요한 적합도 함수 구성의 요건으로 보았다. 둘째, 현재 모델을 이용해 과거 시점들의 관측치를 예측하게 함으로써 불량 예측 모델이

유전 알고리즘을 통한 모델 성숙 기간을 진행해 나감에 따라 적합도 함수를 보완할 수 있도록 했다. 적합도 함수의 첫째 항은 현재 시점에서의 예측의 정확성을 판단하는 부분으로 각 시점별로, 구축된 예측 모델을 통해 얻어진 예측치와 그 시점의 관측치 간의 차이를 이용한다. 두 번째 항은 현재 시점을 t , 유전 알고리즘을 통한 모델 성숙 기간의 시작 시점을 1이라고 놓았을 때, 현재 구축된 예측 모델을 이용해 시점 1부터 시점 $t-1$ 까지 $t-1$ 번의 예측을 현재 모델로 수행한다. 즉, 유전 알고리즘을 통한 모델 성숙 기간을 진행해 나감에 따라, 과거 시점의 정확성도 함께 고려할 수 있는 예측 모델을 구축할 수 있도록 한다. 제안된 적합도 함수를 사용하게 되면, 현재 시점의 예측의 정확성에 가장 큰 비중치를 둔 상태에서 과거 유전 알고리즘을 통한 모델 성숙 기간 동안의 예측의 정확성도 각 시점별로 고르게 현재의 선별에 영향을 줄 수 있게 된다. 본 연구에서 제안된 목적 함수, f 는 식 (4)와 같다.

$$f = |y - \hat{y}| + \left(\frac{\sum_{i=1}^k |y - \hat{y}_i|}{k} \right) \tag{4}$$

여기서, y : 실측치, \hat{y} : 예측치

3.1.6 개체 해석

적합도 함수에서 예측의 정확성을 판단하기 위해서는 현재 시점의 실측치와 비교하기 위한 예측치가 필요하다. 선형 인과 모델의 계수가 유전 알고리즘의 개체로 표현되기 때문에 0과 1 사이로 Scaling된 개체의 표현형을 이용해서는 예측치를 직접 산출할 수 없다. 따라서, 유전 알고리즘의 유전 연산 단계를 거친 개체들을 실제 모델의 계수값들로 변환시켜 주는 개체 해석 단계가 필요하다.

개체 해석에는 개체 표현시에 얻은 초기값(Init Value)과 차

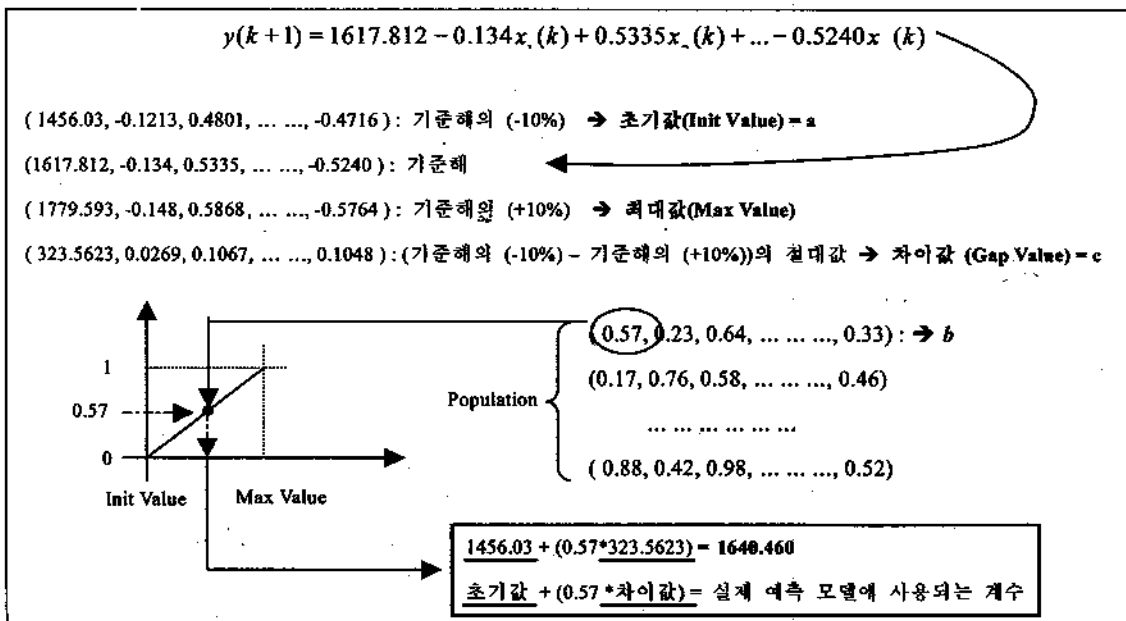


그림 10. 불량 예측 모델을 위한 개체 해석.

이값(Gap Value)을 이용한다.

모집단 내에 임의의 개체 b 가 있을 때 b 는 개체 표현 단계에서 기준해를 이용해 산출된 초기값 집합 a 와 차이값 집합 c 를 갖는다. 개체 b 를 예측에 사용 가능한 실제 계수값으로 바꿔주기 위해서는 <그림 10>과 같은 전환 과정을 거친다.

3.2 유전 알고리즘의 구조

이상의 유전 알고리즘의 구성 요소를 반영한 전체 유전 알고리즘의 구조는 <그림 11>과 같다. 우선, 전체 예측 기간을 세 기간으로 구분한 뒤, 회귀 분석을 이용한 모델 성숙 기간의 데이터를 이용해 회귀 분석을 수행하고, 회귀 모델을 이용해 유전 알고리즘의 초기 모집단을 구성한다. 구성된 초기 모집단을 사용해 각 시점별로 100세대를 반복해서 초기 모집단으로부터 최적해를 구하기 위한 유전 연산을 반복해 나간다. 일반적으로 세대수는 사전 실험을 통해 결정되는데, 세대수의 결정에 가장 큰 영향을 주는 요인은 세대에 따른 실험소요 시간과 해의 수렴성이다. 지나치게 세대수가 짧은 경우는 해 공간에 대한 충분한 탐색이 이루어지지 않아 해의 품질이 나빠질 수 있으며, 반대로 지나치게 세대수가 길게 설정되어 있

```

begin
  divide full period into 3 partial period (1 ~  $t_1$ ,  $t_1 + 1$  ~  $t_{99}$ ,  $t_{99} + 1$  ~  $t$ )
  regression for 1 ~  $t_1$ 
   $t \leftarrow t_1 + 1$ 
  while ( $t < > t_{99}$ ) do
    begin
       $i = 1$ 
       $P(i)$ 의 초기화
      repeat
        begin
          while ( $P(i + 1)$ 의 개체수 < > 100)
            begin
               $P(i)$ 에서 두 개체(부모 개체) 선택
              선택된 개체들에 대한 유전 연산
              부모 개체와 자식 개체들에 대한 적합도 평가
              적합도가 높은 두 개체를  $P(i + 1)$ 에 추가
            end
           $P(i)$ 에 대한 적합도 평가
           $i = i + 1$ 
        end
      until ( $i = 101$ )
       $P$ 에 대한 적합도 평가
      시점  $t$ 에서의 최적 모델 산출
       $t \leftarrow t + 1$ 
    end
  end
  
```

그림 11. 불량 예측 모델을 위한 유전 알고리즘의 구조

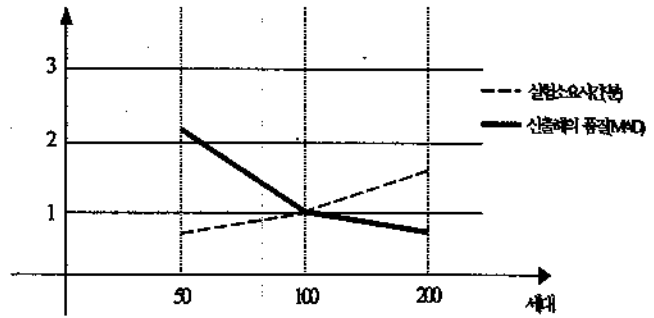


그림 12. 세대수 결정 실험 결과.

는 경우는 전체 실험 시간이 길어지게 되므로 알고리즘의 효율이 떨어지게 된다. 실험 시간과 해의 품질에 대한 최적 trade-off를 위해서는 좀더 다양한 실험이 필요할 것으로 보이나, 본 연구에서는 세대수를 결정하기 위해 대표적으로 50, 100, 200 세대에 대한 사전 실험을 통해 그 경향을 보고 세대수를 결정하였다. 200 세대 이상은 지나치게 많은 실험 시간 소요로 알고리즘의 효율이 현저히 떨어지기 때문에 제외하였다. 세대수 결정 실험을 통해 세대수를 100으로 결정하였으며 세대수 결정 실험 결과 그래프는 <그림 12>와 같다. <그림 12>의 그래프는 실험 소요 시간과 산출해의 품질을 세대수가 100인 경우를 기준으로 Scaling한 것이다.

100세대가 진행된 후에는 그 시점의 최적해를 기록하고 다음 세대로 진행한다. 이전 세대의 최적해는 다음 세대의 초기 모집단에 우수해로서 포함된다. 본 연구에서는 각 세대별로 총 101개의 개체로 구성된 모집단을 운영한다. 모집단의 크기도 세대수와 마찬가지로 사전 실험에 의해 결정되는데, 현 세대에서 임의로 발생시킨 100개의 개체와 이전 세대의 최적해를 우수 개체로 포함시켜 총 101개의 개체를 모집단의 크기로 결정한다. 이와 같은 방법은 이전 세대의 최적해를 현 세대에서의 모집단 탐색에 이용할 수 있으므로 세대간 연결성을 확보하고 새로이 생성된 개체들과의 유전 연산을 거쳐서 보다 좋은 해를 찾을 수 있다는 장점이 있다. 결정된 세대수와 모집단 크기를 가지고 유전 알고리즘을 이용한 모델 성숙 기간이 끝날 때까지 각 시점별로 앞에서 언급된 대로 100세대 동안 모집단 내의 해의 품질을 높여가며 최적 예측치를 찾을 수 있는 불량 예측 모델 계수들의 조합을 찾는다. 유전 알고리즘을 이용한 모델 성숙 기간이 끝나면 최종적으로 하나의 최적해가 도출되며, 얻어진 최적해를 실제 불량 예측 모델의 계수들로 사용하기 위해 개체 해석 과정을 거쳐서 최종 불량 예측 모델을 완성한다. 다음으로는 완성된 불량 예측 모델을 이용해 수 예측 기간을 예측하게 된다 <그림 13>.

3.2.1 예측의 정확성 향상을 위한 보정치 산출

최종 예측을 위한 불량 예측 모델을 구축할 때, 예측의 정확성 향상을 위해 보정치를 산출하고 모델에 반영하여 준다. 보정치의 산출은 그래프를 이용한 방법과 예측치와 실측치 간의

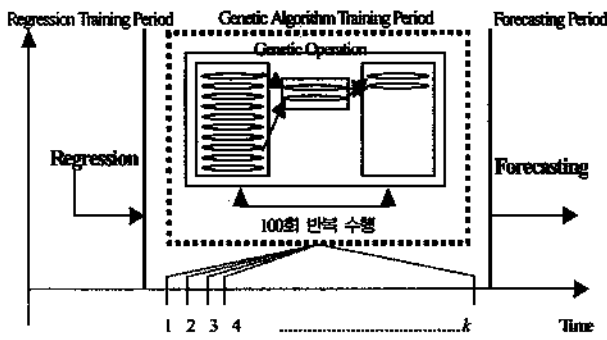


그림 13. 불량 예측 모델의 전체 구조.

차이의 평균치를 함께 사용한다. 보정치를 산출하기 위해 사용하는 데이터는 유전 알고리즘을 이용한 모델 성숙 기간 동안의 예측치와 실측치이며, 이 중 예측치의 경우는 완성된 최종 예측 모델을 이용해 유전 알고리즘을 이용한 모델 성숙 기간의 불량지표를 예측한 값을 뜻하며, 실측치의 경우도 유전 알고리즘을 이용한 모델 성숙 기간 동안의 실제 불량지표를 사용하게 된다. 보정치 산출방법은 유전 알고리즘을 이용한 모델 성숙 기간 동안의 시점별 예측치와 실측치 간의 차이를 오름 차순으로 정렬한 뒤 그래프로 표시한다. 완성된 그래프에는 <그림 14>와 같이 그래프의 영역이 모두 양의 범위에 속하는 것, 모두 음의 범위에 속하는 것, 음의 범위와 양의 범위에 걸쳐서 그래프가 존재하는 세 가지 경우가 가능하다.

유전 알고리즘을 이용한 모델 성숙 기간 내의 임의의 시점을 t 라고 보았을 때 그림의 그래프는 시점별로 얻어진 예측치와 실측치의 차이인 $(\hat{y}_t - y_t)$ 값들을 오름차순으로 정렬하여 그린 것이다.

먼저, 그래프의 영역이 모두 양의 범위에 속하는 유형 1의 경우는 b 만큼을 예측 모델에 더해줌으로써, 유형 3의 경우는 a 만큼을 예측 모델에서 빼줌으로써 손쉽게 예측의 정확성을 높이는 것이 가능하다. 유형 2의 경우는 값들이 음과 양의 범위에 모두 걸쳐 있으므로 직관적으로 보정치를 결정하기가 어렵다. 따라서, 유전 알고리즘을 이용한 모델 성숙 기간 동안의 예측치와 실측치 간의 평균적인 차이인 $\frac{\sum_{i=1}^k (\hat{y}_i - y_i)}{k}$ 값을 보정치로 결정한다. 단, 여기서 k 는 유전 알고리즘을 이용한 모델 성숙 기간의 길이를 뜻한다. 또한, 그래프가 유형 1과 유형 3의 경우를 따르더라도 직관적인 보정치 산출이 힘든 경우는

$\frac{\sum_{i=1}^k (\hat{y}_i - y_i)}{k}$ 값을 보정치로 결정한다. 결정된 보정치는 순수 예측을 하기 위해 최종 결정된 불량 예측 선형 모델의 마지막 항으로 추가 시켜줌으로써 실제 예측시보다 정확한 예측이 가능할 수 있도록 한다.

새로이 제안된 보정치 산출방법의 경우, 전체 연산의 속도에 영향을 주지 않으면서 실제 예측 기간의 예측 정확성을 높일 수 있으며, 불량 예측 모델과 예측에 사용되는 데이터의 변경 등에 관계없이 예측 모델에 적합한 보정치를 산출할 수 있다.

4. 실험 분석 및 시스템 구현

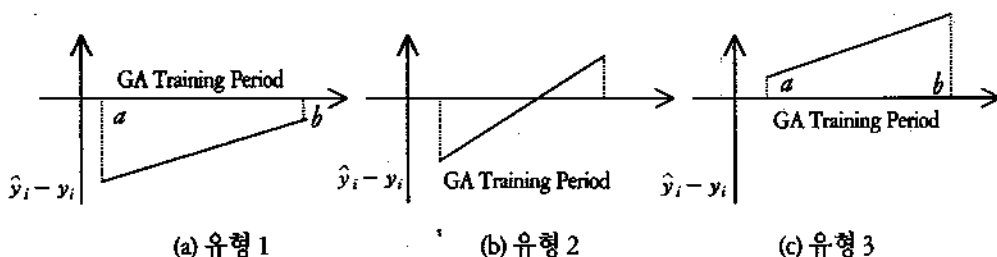
본 연구에서는 실제 유리 용해 공정의 데이터를 이용해 제안된 Lag Time 결정법의 유효성과, 변수 선택 과정을 이용해 얻은 주요 인자들로 구성된 불량 예측 모델의 성능을 실험을 통해 검증하였다. 또한, 유전 알고리즘을 이용한 불량 예측 모델의 성능을 회귀 분석을 이용한 불량 예측 모델과 비교하였으며, 실험의 검증과 실제 현업에의 적용을 위해 본 연구에서 제안된 여러 기법들을 중심으로 유리 용해 공정에서 사용 가능한 불량 예측 시스템을 구현하였다.

4.1 실험 분석

4.1.1 실험 설계

실험에 사용된 데이터는 실제 유리 용해 공정에서 사용된 실측치를 이용하였으며, 네 개의 유사 유리 용해 공정을 대상으로 하였다. 각 공정별 데이터들은 불량 지표와 이에 영향을 줄 수 있는 공정 인자로 구분할 수 있으며, 불량 지표에는 유리 용해 제품의 품질을 좌우하는 제품 내 기포(Blister)의 비율을 사용하였고, 초기 공정 인자에는 기포 발생에 영향을 줄 수 있는 용해 공정 내 각 위치의 온도, 압력, 전압, 기름의 양 등에 해당하는 77개의 인자를 사용하였다. 데이터들은 1시간 간격으로 공정 내 77개 인자들에 대해 관측되거나 조작된 수치들이 기록되며, 실험에 사용된 데이터는 약 10개월 간의 실제 공정별 실측치를 이용하였다.

실험은 우선 그래프 경향법을 이용해 77개 각 공정 인자별로 불량 지표와의 Lag Time을 산출한 후 데이터 전처리 과정을



(a) 유형 1

(b) 유형 2

(c) 유형 3

그림 14. 보정치 산출을 위한 예측치와 실측치 간의 차이 그래프.

통해 결정된 Lag Time을 데이터에 반영해 주었다. 또한, 중점적인 관리를 필요로 하는 불량 예측 모델에 포함될 중요 인자들을 선별해 내기 위해 변수 선택 과정을 통해 라인별로 총 77개의 공정 인자들을 48개의 인자로 줄였다. 유전 알고리즘을 적용하기 위한 데이터 기간의 구분은, 전체 데이터 기간 중 회귀 분석을 이용한 모델 성숙 기간은 6개월간으로 설정했으며, 유전 알고리즘을 이용한 모델 성숙 기간은 유전 알고리즘을 이용한 모델과 회귀 분석을 이용한 모델의 성능 비교를 위해 각각 12시간, 24시간의 두 가지로 설정해 실험을 진행하였다.

순수 예측 기간은 유전 알고리즘을 이용한 모델 성숙 기간 별로 현재 시점으로부터 하루 동안과 한달 동안의 두 경우에 대해 완성된 불량 예측 모델로 예측하였다. 네 개의 각 공정별로 회귀 분석을 사용한 예측 모델과 유전 알고리즘을 사용한 예측 모델 간의 예측의 정확성을 MAD (Mean Absolute Deviation) 수치를 이용해 비교하였다.

유전 연산에 사용된 교차의 경우 교차율을 0.7, 돌연변이율은 0.3, 그리고 교차로 선택된 개체들 중에서 0.16의 확률로 돌연변이를 수행한 후 교차를 수행하게 하였다. 따라서 선택된 모든 개체들은 교차나 돌연변이 중 하나의 유전 연산을 수행하게 된다. 교차율과 돌연변이율은 파라미터 결정 실험을 통해 교차율의 경우 0.5, 0.6, 0.7, 0.8, 0.9의 다섯 가지 대안에 대해 표본 유리 용해 공정의 데이터를 사용해 10회의 반복 실험을 통해 결정하였으며, 교차로 선택된 개체들 중에서 돌연변이 연산을 수행할 확률은 기초 실험을 통해 0.16으로 결정하였다. 제안된 유전 알고리즘 및 검증 실험, 그리고 불량 예측 모델의 시스템 구현은 Pascal 언어와 Delphi를 사용하였고, Pentium II 400 MHz의 IBM-PC를 사용하였다.

4.1.2 실험 결과

■ Lag Time과 변수 선택 과정 검증 실험

각 라인별로 총 77개의 인자에 대해 Lag Time을 적용하였을 경우와 적용하지 않았을 경우의 예측 성능을 회귀 분석 모델을 이용해 비교하였다. Lag Time을 적용하지 않은 경우라 함은 각각의 공정 인자들과 불량 지표 사이에 개별적인 Lag Time이 없이 모든 공정 인자들과 불량 지표 사이에 일괄적으로 12시간의 공정 시간 차이를 주는 경우이다. 이는 실제 대상이 되는 유리 용해 공정에서 사용하는 Lag Time이다. 또한, 변수 소거 과정을 이용해 라인별로 77개 인자들 중에서 48개의 주요 인자를 선정하고 전체 공정 인자가 모두 포함된 경우와 주요 인자만 포함된 경우의 예측 성능을 회귀 분석 모델을 이용해 비교하였다. <표 1>의 실험 결과는 시간축을 이동해 가며 10회의 반복 실험을 수행하였을 때 얻어진 MAD 수치의 평균값들이다.

<표 1>에서 모델에 포함된 공정 인자의 수에 상관없이 거의 모든 경우에서 Lag Time을 적용한 경우가 적용하지 않은 경우보다 좋은 성능을 보였다. 공정 인자는 일반적으로 주요 인자를 기준으로 공정을 설명해 줄 수 있는 많은 인자들이 있는 경우에 실제 공정을 잘 반영한다고 볼 수 있지만, 관리와 분석

표 1. Lag Time과 변수 선택 과정 검증 실험 결과: MAD 비교

	77 variables (full model)		48 variables (reduced model)	
	Without Lag Time	With Lag Time	Without Lag Time	With Lag Time
Line 1	7.308	4.247	6.382	6.262
Line 2	20.306	26.782	25.051	21.154
Line 3	17.623	17.043	18.649	15.166
Line 4	2.730	2.729	3.022	2.341

표 2. Full Model과 Reduced Model간 MDR비교

	MAD Deviation Ratio
Line 1	-0.47
Line 2	0.21
Line 3	0.11
Line 4	0.14

의 효율이 떨어진다는 단점이 있다. 따라서, 모델에 포함된 인자의 수는 적으면서도 전체 공정 인자를 포함한 모델의 경우와 유사하거나 더 뛰어난 성능을 보일 수 있는 주요 인자를 선정하는 것이 중요하다. <표 2>는 라인별로 77개 전체 공정 인자를 사용한 예측 모델(Full Model)과 변수 선택 과정을 거처 얻어진 48개의 공정 인자를 사용한 예측 모델(Reduced Model) 간의 성능을 Lag Time을 적용한 데이터를 사용하여 비교한 결과이다. 결과 비교를 위해 식 (5)와 같은 MDR(MAD Deviation Ratio)을 사용하였다.

$$MDR = \frac{\text{Full Model의 MAD} - \text{Reduced Model의 MAD}}{\text{Full Model의 MAD}} \quad (5)$$

<표 2>에서 라인 1의 경우를 제외하고는 모두 10% 이상 좋은 결과를 보였다. 1라인의 경우는 다른 라인에 비해 변수 선택 과정의 효과보다는 Lag Time의 적용으로 인한 해의 성능 향상이 두드러짐을 알 수 있었다. 이것은 1라인이 실제 공정에서 안정된 라인이므로 주요 변수들뿐만 아니라 다른 여러 공정 인자들이 복합적으로 해의 불량에 영향을 주기 때문에 모든 변수들이 고려된 불량 예측 모델을 사용한 경우가 해의 품질을 높인다고 생각할 수 있다.

■ 예측의 정확성 비교 실험

유전 알고리즘을 이용한 불량 예측 모델과 회귀 분석을 이용한 불량 예측 모델을 현재 시점으로부터 하루 동안과 한달 동안의 순수 예측 결과로 비교하였다. 두 모델에 사용된 데이터는 Lag Time을 적용한 실제 유리 용해 공정의 실측치이며, 변수 선택 과정을 통해 결정된 48개 인자들로 모델을 구성하였다. 유전 알고리즘을 이용한 불량 예측 모델의 경우 10회의 반

표 3. GA 모델-Regression 모델 간 예측 성능 비교 실험 결과: MAD 비교

Line	Forecasting Period	Model		
		Regression	GA (12 hrs Training)	GA (24 hrs Training)
1 Line	1 day	1.116	0.355	0.460
	1 month	7.611	8.295	7.537
2 Line	1 day	10.477	0.490	2.035
	1 month	22.660	18.242	18.335
3 Line	1 day	21.548	18.111	20.892
	1 month	19.767	21.391	19.750
4 Line	1 day	1.060	0.260	0.230
	1 month	4.325	6.395	4.325

복 실험의 평균을 사용하였으며 유전 알고리즘을 이용한 모델 성숙 기간은 12시간(Model 1)과 24시간(Model 2)의 두 가지로 나누어서 실험하였다.

<표 3>에서 보는 바와 같이 모든 경우에 있어서 유전 알고리즘을 이용한 불량 예측 모델이 회귀 분석을 이용해 구성된 불량 예측 모델보다 좋은 성능을 보였다. 회귀분석을 이용한 모델과 유전 알고리즘을 이용한 모델간의 성능 차를 순수 예측 기간의 크기별로 알아보기 위해 MDI를 이용해 오차를 비교하면 <표 4>와 같다. 사용된 MDR은 식 (6)과 같이 나타낼 수 있다.

$$MDR = \frac{\text{Regression Model의 MAD} - \text{GA Model의 MAD}}{\text{Regression Model의 MAD}} \quad (6)$$

유전 알고리즘을 이용한 모델 성숙 기간의 길이에 따라 성능차가 있었으나, 모든 경우에서 단기 예측의 경우 장기 예측보다 좋은 결과를 얻을 수 있었다.

4.2 시스템 구현 및 적용

제안된 유전 알고리즘을 이용한 불량 예측 모델을 실제 공정에 적용 가능하도록 시스템으로 구현하였다. 구현된 시스템

표 4. GA 모델-Regression 모델 간 MDR 비교

Line	Forecasting Period	MAD Deviation Ratio
1 Line	1 day	0.68
	1 month	0.01
2 Line	1 day	0.95
	1 month	0.19
3 Line	1 day	0.16
	1 month	0.001
4 Line	1 day	0.78
	1 month	0.00

중에서 Lag Time 산출 부분과 주요 인자 선정 부분을 중심으로 한 회귀 분석 예측 모델은 이미 국내 기업의 실제 유리 용해 공정 사이트에 성공적으로 적용되어 사용되고 있으며 기존의 불량 예측 시스템에 비해 좋은 성능을 보이고 있다. 또한, 유전 알고리즘을 이용한 개선된 불량 예측 시스템의 경우는 현장 적용성을 검토중이다. 전체 시스템은 크게 두 부분으로 나누어져 있다. 우선, 공정 내에서 수집된 데이터들을 이용해 기초적인 통계 분석을 할 수 있으며, 불량 예측 모델에 사용될 수 있도록 잘못되거나 누락된 데이터가 있을 경우에는 이동 평균법 등을 사용해 데이터의 전처리를 하는 데이터 관리 부분이 있다. 다음으로는 불량 예측 모델을 구성하는 부분이 있는데, 이 부분은 크게 Lag Time을 결정하고 주요 변수를 설정하는 부분과 유전 알고리즘을 이용한 불량 예측 모델 구축 부분으로 나눌 수 있다. 데이터 관리 부분으로부터 불량 예측 모델에 사용되는 데이터를 받아서 각 인자별로 Lag Time을 결정하며, 결정된 Lag Time을 데이터에 반영해 준다. 주요 변수 선정 부분에서는 현재 공정 내의 주요 인자를 선정하고 최종적으로 유전 알고리즘을 사용해 불량 예측 모델을 완성하게 된다. 이상의 기능이 반영된 전체 시스템의 구조는 <표 5>와 같다.

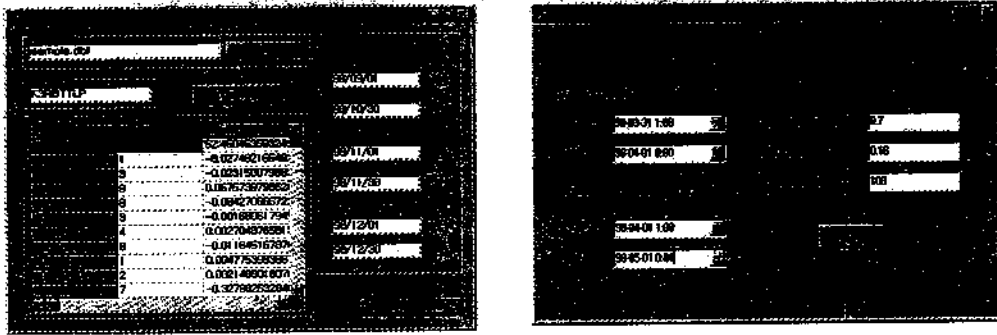
시스템의 기능별 주요 화면은 <그림 15>에 예시된 바와 같다.

5. 결론

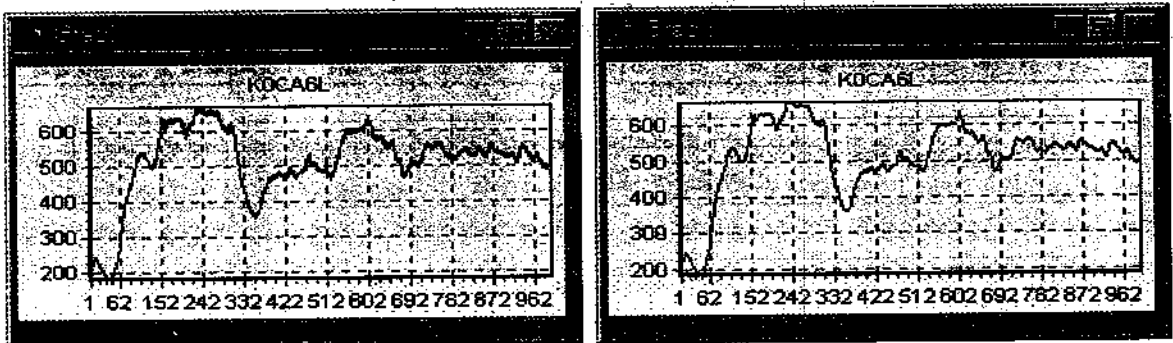
발전적 기법 중의 하나인 유전 알고리즘은 잠재해들로 이루어진 모집단을 이용해 해를 찾지 때문에 복잡한 수리 문제를 해결하는 데 많이 사용되고 있으며, 다양한 분야에 적용되고 있다. 본 연구에서는 유리 용해 공정의 불량 지표를 예측하는 데 있어, 실제 공정 조건을 반영해 주기 위해 Lag Time의 개념을 도입하였으며 그래프 경향법에 근거한 효율적인 Lag Time 결정 방법을 제안하였다. 또한, 공정 내의 방대한 데이터들의 관리 및 분석에서 오는 비효율성을 막고 공정 내 주요 인자들을 선정하기 위해 Backward Elimination 방법을 이용해 불량 예측 모델에 사용될 공정 인자를 결정하였으며 아울러 예측에 사용

표 5. 불량 예측 시스템의 구조

기능별 영역	세부 기능	
데이터 관리	변수 추가 / 삭제	공정 내 인자들 중에서 불량 예측 모델에 사용될 인자들을 추가시키거나 삭제함
	데이터 입력	기존의 데이터 파일(Excel 또는 Text 파일 형식)을 불러들임
		사용자가 직접 변수별 데이터를 입력함
	데이터 유효성 검증 (Validation)	기존의 데이터 파일을 검색해 보았을 때 누락된 데이터가 있거나, 공정이상 등으로 발생한 한계 범위 이상 또는 이하의 잘못된 데이터가 있는 경우에 데이터 유효성 검증을 통해 사용자로부터 하여금 잘못된 부분의 데이터를 수정하게 함
	데이터 보정 (Filtering)	데이터 유효성 검증을 통해 데이터의 수정이 필요하거나, 지나치게 데이터의 변동이 심해 정확한 불량 예측이 힘든 경우 이동 평균법, 지수 평활법 등을 사용하여 데이터를 보정함
	기초 통계 분석	공정 내 수집된 데이터를 이용해 상관 분석, 회귀 분석, 요인 분석 등의 기초 통계 분석을 함
	데이터 조회	라인별로 또는 기간별로 공정으로부터 수집된 데이터들을 검색하고 조회함
데이터베이스 관리	라인별로 데이터 베이스의 상태를 파악함	
	데이터 베이스의 시스템 내 저장 기간을 설정하거나 불필요한 데이터 베이스를 삭제함	
모델 관리	기간 설정	불량 예측 모델에서 사용되는 회귀 분석을 이용한 모델 성숙 기간, 유전 알고리즘을 이용한 모델 성숙 기간, 순수 예측 기간 등을 설정함
	Lag Time 산출 및 적용	제안된 그래프 경향법을 사용해 불량 예측 모델에 사용되는 인자들에 대해 Lag Time을 산출하고 데이터에 적용함
	주요 변수 선정	라인별로 전체 공정 인자들을 대상으로 불량 예측 모델에 적합한 최적 공정 인자를 Backward Elimination을 이용해 선정함
	모델 I/O 구조 설정	불량 예측 모델의 입력과 출력 구조를 설정함
	유전 알고리즘 초기값 지정	유전 알고리즘에 사용되는 세대수, 모집단수, 교차율, 돌연변이율 등의 초기값을 지정함
	Test	Lag Time과 주요 변수 선정 방법을 적용한 불량 예측 모델에 대한 테스트 및 검증작업
	모델 저장 및 조회	결정된 Lag Time들과 주요 변수들을 저장하고 조회함
불량 예측	불량 지표 선택	불량 예측을 하기 앞서 라인별로 불량 지표에 사용될 인자를 선택함
	적용 Line 선택	불량 예측을 할 Line을 결정함
	모델 확정	유전 알고리즘과 보정치 산출 방법을 사용해 최종 불량 예측 모델을 확정함
	결과 그래프 조회 및 저장	불량 예측 결과 그래프를 저장하고 조회함
		결과 분석
	결과 분석 조회 및 저장	결과 분석에서 분석된 자료들과 테이블들을 저장하고 조회함
공정 Simulation	Line/Output 선택	실제 공정을 미리 Simulation해보기 위해 대상 Line과 결과 출력 형태를 선택함
	결과 화면 조회 및 저장	Simulation 결과를 저장하고 조회함
조회/보고 관리 (Documentation)	예측결과 집계	라인별 또는 기간별로 예측 결과 그래프와 테이블, 분석 자료들을 집계함
	모델 정보 집계	라인별 또는 예측 성능별로 구축되어 있는 불량 예측 모델들에 대한 정보를 집계함



(a) 불량 예측 시스템의 입력 화면



[보정 전 데이터]

[보정 후 데이터]

(b) 이동평균법을 이용한 데이터 보정 화면

KJFB3L	4.285451657
KJFBGR	4.116310892
KJFBGL	5.378853379
KJCA6R	6.00515953
KJ7A	5.751952572
KJCASPT	5.46611962
KJCAT	4.637757907
KJBLT	4.943703185
KJ52R	7.224826656
KJMC1	8.1405392
KJMSFL	8.15798361
TSPARE4	9.07620564

(c) 주요 변수 선정 화면



(d) 불량 예측 시스템의 결과 화면

그림 15. 유전 알고리즘을 이용한 불량 예측 시스템의 구현 화면: 예시.

되는 데이터 기간을 새로이 분류하였다. 제안된 불량 예측 모델은 유전 알고리즘을 기반으로 현재 시점의 데이터와 과거 데이터들을 함께 고려할 수 있는 적합도 함수를 포함하며 예측의 정확성을 높이기 위해 보정치를 사용하였다.

제안된 모델은 기존의 회귀 분석 모델과 비교해 시험 대상으로 삼은 네 개의 라인에서 모두 우수한 성능을 보였으며, 유전 알고리즘의 초기 모집단을 임의의 개체가 아닌 회귀 분석을 통한 현실적인 개체들로 사용함으로써 유전 알고리즘의 수렴 속도를 향상시키고 예측 결과의 신뢰성을 높일 수 있도록 했다. 또한, 유전 알고리즘을 이용한 불량 예측 모델의 경우 중, 장기 예측보다는 현재 시점으로부터 가까운 미래를 예측하는 단기 예측에서 회귀분석과 비교해 보았을 때 보다 좋은 성능을 보였다. 제안된 불량 예측 모델은 실제 유리 용해 공정의 데이터를 사용하여 검증하였으므로 유사 다른 연속 생산 공정의 적용도 가능할 것으로 보인다.

앞으로 추가적인 연구가 필요한 부분은 유전 알고리즘의 적합도 함수를 보다 효율적이고도 실제 공정의 추이를 반영할 수 있도록 개발하는 것이며, 유전 알고리즘을 이용한 모델 성숙 기간의 길이와 예측의 정확성간의 관계를 규명하는 연구도 진행되어야 할 것이다.

참고문헌

김여근, 윤복식, 이상복 (1997), *패타 휴리스틱*, 영지 문화사.
 Afifi, A. A. and Clark, V. (1990), *Computer-Aided Multivariate Analysis*, Second Edition, Van Nostrand Reinhold Company, 186-218.
 Chiraphadhanakul, S., Dangprasert, P., and Avatchanakorn, V. (1997), Genetic Algorithms in Forecasting Commercial Banks Deposit, *Proceedings of the 1997 IEEE International Conference on Intelligent Processing Systems*.
 Cross, F., Tadei, R., and Volta, G. (1995), A Genetic Algorithms for the Job Shop Problems, *Computers and Operations Research*, 22, 15-24.
 Dondorf, U., and Pesch, E. (1995), Evolution Based Learning in A Job Shop Scheduling Environment, *Computers and Operations Research*, 23, 74-81.
 Flurry, B. and Riedwyl, H. (1988), *Multivariate Statistics : A Practical Approach*, Chapman and Hall, 54-74.

Gabbent, P., Brown, D., Huntley, C., Markowicz, B., and Sappington, D. (1991), A System for Learning Routes and Schedules with Genetic Algorithms, *Proceedings of ICGA 91*.
 Goldberg, D. (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesely, Reading, Massachusetts, USA.
 Goto, Y., Yukita, K., Mizuno, K., and Ichianagi, K. (1999), Daily Peak Load Forecasting by Structured Representation on Genetic Algorithms for Function Fitting, *Transactions of the Institute of Electrical Engineers of Japan B, A Publication of Power and Energy Society*, 119(B)(6), 735-736.
 Holland, J. (1975), *Adaptation In Natural And Artificial Systems*, Univ. of Michigan Press, Ann Arbor.
 Homayfar, A., and Guan, S. (1991), A New Approach on the Traveling Salesman Problem by Genetic Algorithms, *Technical Report*, North Carolina A&T State University.
 Kai, F., Wenhua, X. (1997), Training Neural Network with Genetic Algorithms for Forecasting the Stock Price Index, *Proceedings of the 1997 IEEE International Conference on Intelligent Processing Systems*.
 Kobayashi, S., Ono, I., and Yamamura, M. (1995), An Efficient Genetic Algorithm for Job Shop Scheduling Problems, *Proceedings of 6th ICGA*, 506-511.
 Kung, C. H., Devaney, M. J., Huang, C. M., and Kung, C. H. (1998), An Adaptive Power System Load Forecasting Scheme Using a Genetic Algorithms Embedded Neural Network, *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, 1, 308-311.
 Miller, B., Goldberg, D. (1995), Genetic Algorithms, Selection Schemes, and the Varying Effects of Noise, *IlligAL Report*, No 95009.
 Shi, G. (1997), A Genetic Algorithms Applied to a Classic Job Shop Scheduling Problem, *International Journal of Systems Science*, 28(1), 25-32.
 Syswerda, G. (1991), *Scheduling Optimization Using Genetic Algorithms*, *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York, 332-349.
 Tanaka, Y., and Berlage, O. (1996), Application of Genetic Algorithms to VOD Network Topology Optimization, *IEICE Transaction on Communication*, E79(B)(8).
 Whirely, D. (1989), Starkweather, T., and Furuay, D., Scheduling Problems and Traveling Salesman: The Genetic Edge Recombination Operator, *Proceedings of ICGA 89*.
 Winston, W.L. (1991), *Operational Research Applications and Algorithms*, Second Edition, PWS-KENT Publishing Co., 11-65.
 Yip, D.H.F., Hines, E.L., and Yu, W.W.H. (1997), Application of Artificial Neural Networks in Sales Forecasting, *Proceedings of the 1997 International Conference on Neural Networks*.



정봉주

1986년 서울대학교 산업공학 학사
 1988년 서울대학교 산업공학 석사
 1989년 Anderson Consulting, 시스템 컨설턴트
 1993년 미국 Pennsylvania State University 산업공학 박사
 1993년 삼성전자(주) 반도체 부분 CIM팀 선임연구원
 현재: 연세대학교 산업공학과 조교수
 관심분야: 생산시스템 분석 및 모델링, 반도체 생산시스템, SCM



정호상

1998년 연세대학교 산업시스템공학 학사
 현재: 연세대학교 산업시스템공학 석사과정
 관심분야: Evolutionary Algorithm, 생산시스템 분석 및 모델링