

# 연속생산공정에서의 유도형 유전알고리즘을 이용한 인과형 예측모델에 관한 연구

정호상\* · 정봉주\*

## A Causal-Forecasting Model using Guided Genetic Algorithm in Continuous Manufacturing Process

Ho-Sang Jung\*, Bong-Ju Jeong\*

### ■ Abstract ■

This paper presents a causal forecasting model using guided genetic algorithm in continuous manufacturing process. The guided genetic algorithm(GGA) is an extended genetic algorithm(GA) using penalty function and population diversity index to increase forecasting accuracy. GGA adds to the canonical GA the concept of a penalty function to avoid selecting the unproductive chromosomes and to make a proper searching direction. Also, GGA modifies the current population using the similarity of chromosomes to avoid falling into the trap of local optimal solution. For investigating GGA performance, we used a set of real data that was collected in local glass melting processes, and experimental results show the proposed model results in the better forecasting accuracy than linear regression model and canonical GA.

## 1. 서 론

연속생산 공정은 일반제조 공정과는 달리 거대 투자가 필요한 장치 기반 산업들의 주요 공정으로 대표적으로는 철강, 유리 산업의 용해공정 등을 들 수 있다. 연속 생산 공정은 공정 라인 주위에 공정

을 제어하거나 관측하는데 사용되는 여러 공정 인자들의 영향을 받기 때문에 신뢰할 만한 공정 예측이 매우 힘들다. 공정인자는 공정 라인에 투입되는 원자재나 반제품 등의 상태에 따라 주입 압력, 속도 등을 제어할 수 있는 제어 인자들과 공정 내 각 부분의 상태를 파악할 수 있는 관측인자를 포

\* 연세대학교 산업시스템공학과

합한 제품의 품질에 영향을 미칠 수 있는 인자를 말한다. 이러한 여러 공정인자들은 미세한 변화에도 제품의 품질에 큰 영향을 끼치는 것들로 데이터의 종류와 양적인 측면에서 매우 방대한 경우가 많다. 따라서 생산 공정의 효율 및 생산성의 가장 큰 지표라 할 수 있는 제품 품질에 영향을 주는 공정 내 불량률 미리 예측할 수 있는 예측 모델의 개발이 필요하다.

본 연구에서는 연속 생산 공정에 있어서 유도형 유전 알고리즘을 사용하여 공정 내 불량률 예측할 수 있는 인과형 예측 모델을 제안하고 실험을 통해 실제 연속 생산 공정에 적용 가능함을 보이는 것을 목적으로 한다. 기존의 예측 모델은 주로 시계열 분석, 회귀 분석 등에 의한 통계적 모형을 통해 이루어졌는데, 이러한 통계적 분석은 과거 일정 양의 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 저하되며, 특히 현실적으로 최근의 공정 데이터들에 대한 가중치를 예측 모델 차원에서 반영하기 어렵다는 단점이 있어왔다. 1990년대 중반부터는 이와 같은 단점을 극복하기 위해 유전 알고리즘을 비롯한 여러 발전적 기법들을 이용한 예측 모델이 개발되기 시작하였다.

유전 알고리즘은 1975년 Holland[7]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 좋은 성능을 보여왔다. 특히, 유전 알고리즘을 이용하여 예측 모델을 구축한 연구들로는 크게 인과형 예측 모델의 계수값을 유전 알고리즘을 사용하여 결정하는 방법에 관한 연구들[2, 8, 10]과 유전 알고리즘을 신경망 기법, 시뮬레이터드 어닐링 기법, 타부 서치 등의 발견적 기법들과 함께 이용해서 예측 모델을 구축한 연구

들[9, 19]이 있었다.

일반적으로 유전 알고리즘을 현상의 다양한 문제의 해결을 위해 적용하는 경우, 특히 공정의 무작위성과 변동이 심한 연속생산 공정의 정확한 예측 모델 구축에 적용하는 데 있어서는 크게 두 가지 고려되어야 할 사항이 있다. 첫째, 해의 조기 수렴 현상의 방지와 둘째, 우수한 해의 선별을 위한 효율적인 적합도 함수의 개발 문제가 존재한다.

먼저 해의 조기 수렴현상이란 탐색할 해공간을 표현하는 유전 알고리즘에 사용되는 모집단이 지역 최적해에 머물며 더 이상 우수한 해를 탐색하지 못하는 상태에 도달함을 뜻하는 것으로서 전체 유전 알고리즘의 성능 향상을 위해 고려되어야 하는 사항이다. 이와 같은 조기 수렴현상을 극복하기 위해 많은 연구들이 이루어져 왔는데, 무질서도를 기반으로 모집단의 다양성 해법을 제시한 연구[16], Markov Chain 분석을 이용하여 유전 알고리즘의 지역 최적해로의 수렴을 분석한 연구[12], 최적의 퍼지(Fuzzy) 제어와 디자인을 위해 모집단의 다양성을 고려한 유전 알고리즘을 제시한 연구[17] 등이 있다.

또한 유전 알고리즘은 내부 단계 중 선별(Selection)을 통해 우수한 해를 산출해 내기 때문에 선별 과정에 필요한 적합도 함수를 보다 효율적이고 정확하게 정의하는 것이 전체 알고리즘의 속도 및 해의 개선정도에 큰 영향을 주는데, 적합도 함수의 성능을 향상시킬 수 있는 벌금 함수(Penalty Function)를 정의하고 도입하는 방법에 관한 연구들[3, 4, 14]이 이루어져 왔다. 특히 Lau[11]는 유전 알고리즘의 매 단계별로 얻어진 우수한 개체들의 형태특성을 관찰하고, 해의 개선이 정체되어 있는 경우에 벌금 함수를 부여하는 방법으로 선별의 정확성을 높이고 보다 우수한 해를 찾아가는 방법을 제안하였다.

본 연구에서는 선별의 정확성을 높이기 위한 벌금 함수와 조기 수렴을 방지하기 위한 모집단 다양성 지표를 동시에 고려한 유도형 유전 알고리즘을 제안하고 그 종류와 양에서 방대한 공정 인자

들에 대한 데이터를 가지고 있는 연속생산 공정을 대상으로 보다 효과적이고 정확한 불량 예측이 가능한 인과형 예측 모델을 제안하고 검증하였다. 인과형 예측 모델 검증을 위한 실험은 대표적인 연속 생산 공정 중의 하나인 실제 유리 용해 공정의 데이터를 사용하였고 유사한 조건에 있는 3개의 공정 라인에 대해서 이루어졌다. 각 공정 라인에 대해 회귀 분석 모델을 이용한 불량 예측 실험과 전통적인 유전 알고리즘을 이용한 불량 예측 실험을 하였고, 본 논문에서 제안된 유도형 유전 알고리즘을 이용한 불량 예측 실험 결과와 비교하였다.

본 논문의 구성은 다음과 같다. 2장에서는 인과형 예측 모델을 제안하며, 3장에서는 인과형 예측 모델의 계수 결정을 위한 유도형 유전 알고리즘의 구조를 제시하고, 4장에서는 제안된 인과형 예측 모델을 실제 공정에 적용한 결과를 제시하고 분석한다. 마지막으로 5장에서는 본 연구에 대한 결론을 맺는다.

## 2. 인과형 예측 모델

### 2.1 모델의 구조

연속 생산 공정은 하나의 공정 주위에 수많은 공정 인자들이 존재하며 각 공정 인자들에 대한 데이터가 실시간으로 기록, 보존되게 된다. 따라서 기록되는 실시간 공정 인자 데이터를 이용하여 현재 시점으로부터 특정 시점까지의 미래를 예측하게 된다. 본 연구에서의 유도형 유전 알고리즘을 이용한 인과형 예측 모델은 용이하고 모든 고려 가능한 공정 인자들을 예측 모델에 반영 가능한 선형 인과형 모델을 토대로 하였다. 선형 인과형 예측 모델은 예측에 영향을 줄 수 있는 원인 인자인 연속 생산 공정의 공정 인자에 대한 데이터와 인과형 예측 모델을 이용해 예측하고자 하는 연속 생산 공정 내의 제품의 품질 또는 불량을 나타내는 지표에 대한 데이터를 사용해 다음과 같이 구성한다.

$$y(k) = b_0 + b_1x_1(k-l_1) + b_2x_2(k-l_2) + \dots + b_nx_n(k-l_n) + e(k) \quad (1)$$

여기서,  $y(k)$  : 시점  $k$ 에서의 불량 지표치(예 : 불량율)  
 $x_i(k-l_i)$  : 공정인자  $x_i$ 에 대한 시점  $k-l_i$ 에서의 수치  
 $l_i$  : 불량 지표치  $y$ 와 공정인자  $x_i$ 사이의 lag time  
 $b_i(i = 0, \dots, n)$  : 인과형 예측 모델의 계수 값  
 $e(k)$  : 시점  $k$ 에서의 보정치

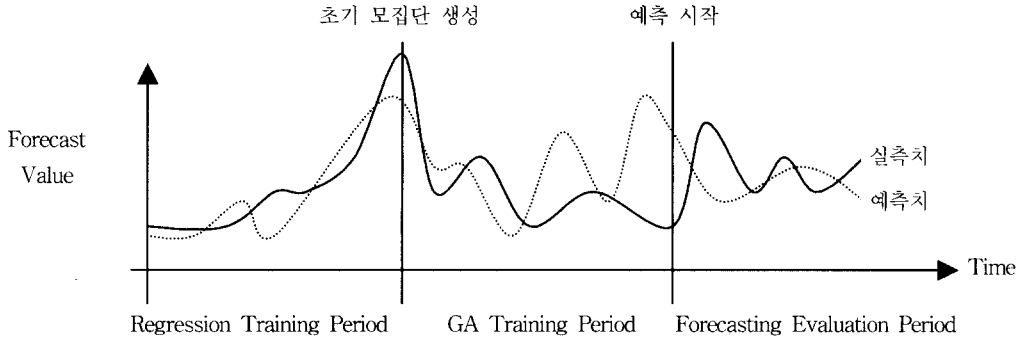
제안된 모델은  $y$ 값으로 사용될 품질 또는 불량 지표와  $x$ 값으로 사용될 공정 인자들에 대한 데이터는 연속 생산 공정 내에 존재하는 실측 데이터들을 사용하고, 모델의 계수 값들은 유도형 유전 알고리즘을 사용하여 결정하고자 한다. 또한 제안된 모델에서 lag time,  $l$ 은 특정 관측 시점에서의 불량 지표치와 공정 인자 수치 사이의 영향을 주고 받는데 소요되는 시간 차이를 뜻한다. 4장의 실험 및 결과 분석에서 연속 생산 공정내의 lag time 을 효율적으로 결정할 수 있는 방법을 제안하였다.

### 2.2 모델에 사용되는 데이터의 기간 설정

일반적인 예측 모델에 사용되는 데이터의 기간은 크게 예측 모델 구성을 위해 사용되는 기간과 실제 예측 기간의 두 부분으로 나누게 된다. 본 연구에서는 인과형 예측 모델을 구성하는데 있어 유도형 유전 알고리즘을 사용하기 위해 전체 데이터 기간을 다음과 같이 세 부분으로 나눈다<그림 1>.

- 회귀 분석을 이용한 모델 성숙 기간(Regression Training Period : RTP)
- 유전 알고리즘을 이용한 모델 성숙 기간(Genetic Algorithms Training Period : GATP)
- 예측 검증 기간(Forecasting Evaluation Period : FEP)

유전 알고리즘은 잠재해들로 구성된 모집단을 운용하며 최적해를 찾아내는 방법으로 초기 모집단의 선정이 전체 해의 품질에 큰 영향을 끼친다.



〈그림 1〉 데이터 기간의 분류

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 개선 효과를 뛰어나게 하는 것으로 알려져 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 기간동안의 데이터에 대해 회귀 분석을 수행하고 그 결과로 얻어진 회귀 분석 모델의 계수값을 이용해 유도형 유전 알고리즘에서 사용할 초기 모집단을 생성한다.

유전 알고리즘을 이용한 모델 성숙 기간의 시작 시점에서는 앞에서 구성된 유전 알고리즘의 초기 모집단을 사용하여 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 유전 알고리즘을 이용한 모델 성숙 기간의 끝 시점까지 모델을 성숙시켜 나간다.

마지막으로 예측 검증 기간에는 앞의 두 기간동안 성숙된 인과형 예측 모델을 사용하여 예측을 수행하고 그 정확성을 검증한다.

### 3. 인과형 예측 모델을 위한 유도형 유전 알고리즘

일반적인 유전 알고리즘은 잠재해를 표현한 개체(chromosome)들로 구성된 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 함수(fitness function)를 이용한 적합도 평가 과정을 거치며 해 공간을 탐색하고, 최적해를 찾는다. 특히 유전 알고리즘은 복잡한 해공간의 탐색 능력이 우수하여 변수와 제약이 많은 대형 수리 문제를 푸는 데 적합한 기법이고, 또한 모형에 대한 유연성 및 강건성이 높아 제약 첨가나 목적 함수의 변경이 용이하다는 장점을 갖는다 [6]. 따라서 연속생산 공정과 같이 제품의 품질에 영향을 미치는 공정 인자들의 종류와 양이 방대한 공정의 예측 모델 구성에는 유전 알고리즘의 적용이 가능하다.

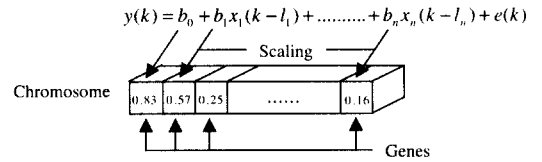
그러나 본 연구에서 다루고자 하는 연속 생산 공정에서의 인과형 예측 모델과 같은 복잡한 해공간의 탐색해야 하는 문제에 있어서는 기존의 유전 알고리즘과 비교해 보다 정확한 선별, 효과적인 해공간의 탐색, 다양한 해 공간에 대한 검토가 가능한 유도형 유전 알고리즘의 적용을 고려해 볼

수 있다. 일반적인 유도형 유전 알고리즘은 현 세대의 모집단으로부터 다음 세대에 생존할 개체를 선택하는 과정인 선별에 있어서 목적 함수만을 사용하는 것이 아니라, 선별된 개체의 구조를 파악해 그 우수함, 열등함의 정도에 따라 벌금을 부과하는 방법을 사용한다. 그러나 해공간의 크기가 넓고 고려 가능한 해공간이 복잡하게 구성되어 있는 경우는 벌금 함수의 도입만으로는 우수한 해의 도출이 힘들다. 따라서 본 연구에서는 기존의 벌금 함수의 개념을 인과형 예측 모델의 구성에 알맞게 변경, 개체 선별을 위한 적합도 함수에 적용하고 아울러 해공간의 효과적인 탐색이 가능하도록 모집단의 다양성을 추구할 수 있는 지표를 산출하여 연속 생산 공정의 인과형 예측 모델에 적합한 유도형 유전 알고리즘을 제안한다.

### 3.1 유도형 유전 알고리즘을 위한 개체 표현 및 해석

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 연구에서는 상이한 범위를 가진 서로 다른 데이터들을 0과 1사이로 Scaling 하는 방법을 이용해 불량 예측 모델에 사용할 개체 표현 기법을 제안하였다. 우선 하나의 개체(Chromosome)는 개체를 구성하는 여러 인자(Gene)들로 이루어지게 되는데 본 연구에서는 인자(Gene)들을 각각 선형 인과형 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 2>과 같이 각기 상이한 범위의 데이터들을 연산이 용이하도록 0과 1사이의 값들 중에서 발생시켜 하나의 인자를 만들고, 예측 모델에서 필요한 만큼 인자를 발생시키는 과정을 반복하여 하나의 개체를 구성한다.



<그림 2> 유도형 유전 알고리즘에서의 개체 표현

개체의 표현과 더불어 특정 개체로부터 실제 예측치를 산출하기 위한 개체 해석 단계가 필요하다. 실제 예측치는 적합도 함수에서 예측의 정확성을 판단하기 위해서 필요한 것인데 선형 인과형 모델의 계수가 유전 알고리즘의 개체로 표현되기 때문에 0과 1사이로 Scaling된 개체의 표현형을 이용해서는 예측치를 직접 산출할 수 없다. 따라서, 유전 알고리즘의 유전 연산 단계를 거친 개체들을 실제 모델의 계수 값들로 변환시켜 주는 개체 해석 단계가 필요하다. 유전 알고리즘을 이용한 모델 성숙 기간내의 시작 시점에서의 모집단이 있다고 가정할 때, 회귀 분석을 이용한 모델 성숙 기간을 통해 얻어진 초기 최적해를 모든 인자들이 0.5로 이루어진 하나의 개체로 표현된다. 또한, 모집단 내의 나머지 개체들은 앞에서 생성된 개체에 대해서 인자별로  $\pm 100\alpha$  %범위 내에서 개체들이 생성된다.

현재 모집단의 특정 개체  $u$ 는 다음과 같은 수식을 사용해 실제 예측에 사용할 수 있는 계수 값들의 집합으로 변환시켜 준다.

$$b_i(u) = b_i^{\min} + c_i(u) \cdot |b_i^{\min} - b_i^{\max}| \quad (2)$$

여기서,  $b_i(u)$  = 개체  $u$ 의  $i$ 번째 계수 값,  $i = 0, 1, \dots, n$

$$b_i^{\min} = \begin{cases} (1 - \alpha) \cdot b_i(s) & \text{if } b_i(s) > 0 \\ (1 + \alpha) \cdot b_i(s) & \text{if } b_i(s) < 0 \end{cases}$$

$$b_i^{\max} = \begin{cases} (1 + \alpha) \cdot b_i(s) & \text{if } b_i(s) > 0 \\ (1 - \alpha) \cdot b_i(s) & \text{if } b_i(s) < 0 \end{cases}$$

$b_i(s)$  = 현재 모집단 내의 최적 개체  $s$ 에 대한  $i$  번째 계수 값  
 $c_i(u)$  = 개체  $u$ 에 대한  $i$ 번째 인자의 Scaling된 수치

### 3.2 벌금함수를 이용한 적합도 함수

적합도 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.

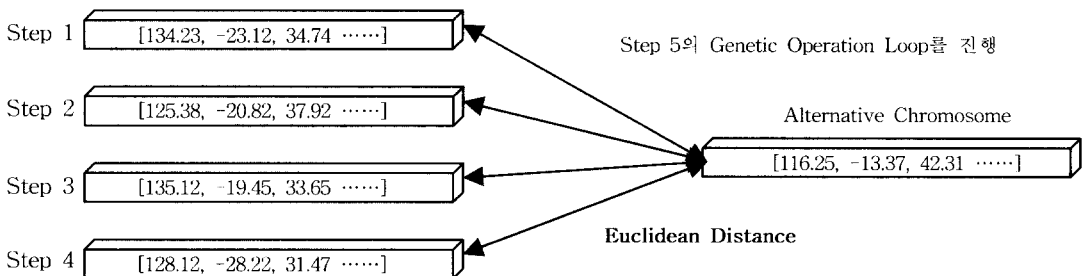
인과형 예측 모델의 경우, 적응도를 평가하는 주된 기준이 개체들로 표현된 예측 모델을 이용해 얻어진 예측값과 실제 관측값과의 차이로부터 얻어지는 예측의 정확성 여부이다. 우선 예측의 정확성을 높이기 위해 적합도 함수를 이용해 얻어지는 예측치의 정확성과 신뢰성을 높이는 것이 중요하다. 본 연구에서는 적합도 함수를 구성하는 데 있어서 크게 세 가지 사항에 초점을 맞추었다. 첫째, 현재 시점에서의 예측의 정확성이다. 유도형 유전 알고리즘을 이용해 구축된 예측 모델이 현재 시점에서 어느 정도 예측의 정확성을 보이느냐를 가장 중요한 적합도 함수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

예측하게 함으로써 인과형 예측 모델이 유도형 유전 알고리즘을 통한 모델 성숙 기간을 진행해 나감에 따라 적합도 함수를 보완할 수 있도록 했다. 셋째, 유전 알고리즘을 통한 모델 성숙 기간을 진행해 나감에 따라 개체를 선별하는 데 있어 이전 시점들에서의 우수 개체들의 구조를 파악해 현재 시점의 모델 선별에 영향을 줄 수 있는 벌금 함수를 사용하였다.

본 논문에서 제안하는 벌금 함수는 크게 두 가지 기준에 따라 벌금(Penalty)을 부여한다. 하나는 유전 알고리즘을 이용한 모델 성숙 기간의 과거 매 시점별로 최적해로 선정된 예측 모델 계수의 집합들과 현재 선택하려는 후보 예측 모델 계수 집합간의 이론적 거리(Euclidean Distance)가 떨어져 있을수록 벌금 값(Penalty Value)을 크게 부여한다.

<그림 3>에서, 유전 알고리즘을 이용한 모델 성숙 기간을 지나며 유도형 유전 알고리즘을 이용해 얻어진 매 시점별 최적 계수 집합들이 있을 때 현재 진행하고 있는 시점 5에서의 선별의 대상이 된 후보 개체는 과거 각 시점별로 얻어진 최적 계수 집합들과의 이론적 거리가 떨어져 있을수록 벌금을 크게 부여 받는다.

또 다른 하나는 현재 선택하려는 후보 예측 모델 계수가 과거 매 시점별로 최적해로 선정된 예측 모델 계수들의 최대값과 최소값으로부터 구한 허용치 범위를 초과하는 경우에는 벌금 카운터



Step  $k$  : 유전 알고리즘을 이용한 모델 성숙 기간(GATP) 내의 시점  $k$

<그림 3> 벌금 함수의 원리(1) : Penalty Value

(Penalty Counter)를 증가 시켜서 벌금을 크게 부여한다.

<그림 4>에서도 <그림 3>와 마찬가지로, 유전 알고리즘을 이용한 모델 성숙 기간을 지나며 유도형 유전 알고리즘을 이용해 얻어진 매 시점별 최적 계수 집합들이 있을 때 현재 진행하고 있는 시점 5에서의 선별 대상이 된 후보 개체는 과거 매 시점별로 선정된 최적 예측 모델 계수들의 최소값의 정해진 허용치 (-)10% 미만이거나 최대값의 (+)10%를 초과하는 경우에는 벌금을 부여한다. 실제 모델을 구축하는데 사용되는 허용치(Tolerance)는 사전 실험에 의해 결정할 수도 있고, 연속 생산 공정 관리자의 의견 또는 목표치에 따라 결정할 수도 있다.

본 논문에서 제안된 최종 벌금 함수는 다음과 같이 나타낼 수 있다.

Penalty Function =

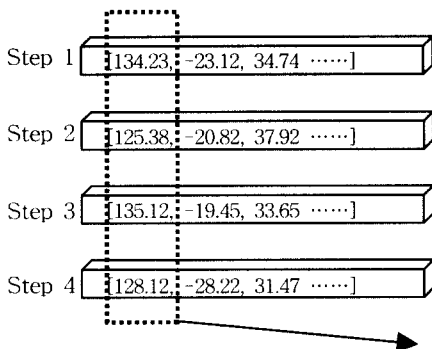
$$\sum_{j=1}^n (1 + P_j) \times \left( \frac{\sum_{i=1}^k \sqrt{\sum_{j=1}^n (F_i(j) - A(j))^2}}{k} \right) \quad (3)$$

$F_i$  = 현재 시점  $i$ 에서의 최적해로 선정된 예측 모델 계수의 집합

$F_i(j)$  =  $F_i$ 의  $j$ 번째 인자,  $j=1, \dots, n$

$n$  : 예측 모델에 사용되는 계수의 총개수

$A_j$  = 현재 선택하려는 후보 예측 모델 계수 집합의  $j$ 번째 인자



$$0.9 * \text{Min}(134.23, 125.38, 135.12, 128.12) < < 1.1 * \text{Max}(134.23, 125.38, 135.12, 128.12)$$

Step  $k$  : 유전 알고리즘을 이용한 모델 성숙 기간(GATP) 내의 시점  $k$

- $k$  = 유전 알고리즘을 이용한 모델 성숙 기간 내의 현재 시점(time)
- $P_i$  = 현재 선택하려는 후보 예측 모델 계수 집합의  $j$ 번째 인자의 Penalty Counter
- $= \begin{cases} 0 & \text{if } A_i \in R_i \\ 1 & \text{otherwise} \end{cases}$
- $R_i = [F_i^l, F_i^u] = i$ 번째 인자 값의 허용범위
- $F_i^l = \min_j F_i(j)$  값의 100  $\sigma$ % lower value
- $F_i^u = \max_j F_i(j)$  값의 100  $\sigma$ % upper value
- $\sigma$  = 허용치

벌금함수를 이용한 본 연구에서 사용된 적합도 함수는 식 (4)와 같다.

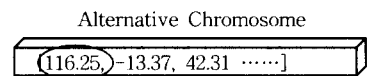
$$f = \alpha |y_k - \hat{y}_k| + \beta \left( \frac{\sum_{i=1}^{k-1} |y_i - \hat{y}_i|}{k-1} \right) + \gamma \left\{ \sum_{j=1}^n (1 + P_j) \times \left( \frac{\sum_{i=1}^k \sqrt{\sum_{j=1}^n (F_i(j) - A(j))^2}}{k} \right) \right\} \quad (4)$$

여기서,  $y$  = 실측치

$\hat{y}$  = 예측치

$\alpha, \beta, \gamma$  = 가중치 인자,  $\alpha + \beta + \gamma = 1$

적합도 함수의 첫째 항은 현재 시점에서의 예측의 정확성을 판단하는 부분으로 각 시점별로, 구축된 예측 모델을 통해 얻어진 예측치와 그 시점의



범위 내에 속하지 않으면 Penalty Counter를 증가 시킨

<그림 4> 벌금 함수의 원리(2) : Penalty Counter

관측치 간의 차이를 이용한다. 두번째 항은 현재 시점을  $k$ , 유전 알고리즘을 이용한 모델 성숙 기간의 시작 시점을 1이라고 놓았을 때, 현재 구축된 예측 모델을 이용해 시점 1부터 시점  $k-1$ 까지  $k-1$ 번의 예측을 현재 모델로 수행한다. 즉, 유전 알고리즘을 이용한 모델 성숙 기간을 진행해 나갈 때 따라, 과거 시점의 정확성도 함께 고려할 수 있는 예측 모델을 구축할 수 있도록 한다. 세번째 항은 현재 시점  $k$ 에서의 벌금 함수로써 과거 시점에서 얻어진 최적해들에 대한 정보를 현시점의 선별에 반영될 수 있도록 했다. 또한, 각 항에 대해 가중치를  $\alpha, \beta, \gamma$ 로 부여함으로써 인과형 예측 모델이 적용될 라인의 특성에 따라 보다 정확한 적합도 함수를 구성할 수 있게 하였다.

본 연구에서는 각 라인별로 사전 실험을 반복해 적합도 함수의 각 항에 대한 가중치를 0.5, 0.3, 0.2로 결정함으로써 현재 시점의 예측의 정확성에 가장 큰 가중치를 둔 상태에서 과거 유전 알고리즘을 이용한 모델 성숙 기간동안의 예측의 정확성도 각 시점별로 고르게 현재의 선별에 영향을 줄 수 있게 하였으며 제안된 벌금 함수를 사용해 예측의 정확성 및 과거 우수 개체에 대한 정보도 함께 고려할 수 있도록 했다.

### 3.3 모집단 다양성 지표

기존의 유전 알고리즘을 사용해 주어진 문제를 해결하는 과정에서 부분 최적해를 벗어나지 못하는 문제는 최종해의 품질과 알고리즘의 효율 차원에서 극복해야 할 가장 큰 문제 중의 하나이다. 잠재해를 표현하는 모집단 내에 존재하는 개체들이 알고리즘을 진행해 나가며 특정 개체 구조로 수렴해 나가는 것은 유전 알고리즘이 해를 찾는 기본적인 원리이지만, 고려 가능한 해영역을 지나치게 좁게 설정하거나, 넓고 복잡한 해공간을 가진 문제에서 모집단이 고려 가능한 해 공간을 모두 포함하지 못한다면 기존의 유전 연산자들로는 더 이상 새로운 해영역을 탐색할 수 없게 되고, 조기에 부

분 최적해로 수렴해 버릴 가능성이 크다. 특히, 모집단 내 개체의 다양성을 유지하기 위해 이용되는 유전 알고리즘의 주요 구성 요소인 돌연변이 연산자가 제 기능을 발휘하지 못하는 결과를 초래한다. 연속 생산 공정의 인과형 예측을 구축하는 문제에서는 유전 알고리즘을 사용해 인과형 예측 모델의 계수를 예측하기 때문에 가능한 해 공간을 정의하는 것도 힘들뿐 아니라, 회귀 분석을 이용해 초기 모집단을 생성해 일정 범위내의 해 공간을 표현할 수 있다 하더라도 최종 예측 시작 시점에서의 최적 모델을 위한 해를 탐색하는 데 한계가 있을 수밖에 없다. 따라서 효율적으로 해 공간을 재정의 하며 최적 해를 찾는 방법을 필요로 한다.

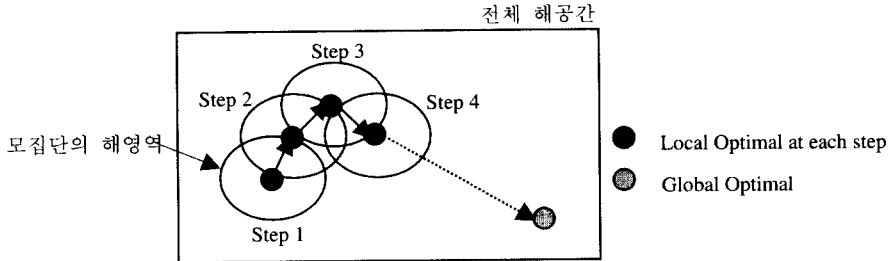
본 연구에서는 모집단의 다양성을 확보하기 위해 현재 시점에서 인과형 예측 모델의 계수값들을 유전 알고리즘을 이용해 찾아나가는데 있어서 모집단의 다양성 여부를 파악할 수 있는 모집단 다양성 지표(Population Diversity Index)를 제안하여 사용하였다. 유전 알고리즘을 이용한 모델 성숙 기간의 매 시점별로 얻어진 최적해들을 이용해 현재 시점을 진행해 나가기 앞서 모집단의 다양성을 검사하고 그 결과에 따라 기존의 해영역을 모집단 차원에서 변경해 준다 <그림 5>.

기본 원리는 유전 알고리즘을 이용한 모델 성숙 기간 내에서 현재 시점을 기준으로 바로 이전 시점에서 얻어진 최적 해와 그 이전 시점들에서 얻어진 최적 해들과의 계수들의 유사성 정도를 이용해 모집단의 다양성 지표를 산출하고 그 값을 이용해 모집단을 재정의 할 필요가 있는 경우는 이전 시점에서 얻어진 최적 해를 기준으로 새로이 모집단을 구성한다

최근 시점의 최적 계수 집합들 간의 유사성 정도에 높은 가중치를 주며 최대 가중치는 1, 최소 가중치는 항상 0 보다 큰 수를 부여하기 위해 Log 함수를 사용하였다.

본 논문에서 제안된 모집단 다양성 지표는 다음과 같이 나타낼 수 있다.





〈그림 5〉 모집단 재정의의 개념

Population Diversity Index =

$$\frac{\sum_{i=1}^{k-2} [\log_{k-1}(1+i) \times \{ \frac{\sum_{j=1}^n |C_{F_{k-1}}(j) - C_{F_i}(j)|}{n} \}]}{\sum_{i=1}^{k-1} \log_{k-1}(1+i)}$$

(5)

$k \geq 3$

- $C_{F_i}$  : 시점  $i$ 에서의 최적해로 선정된 예측 모델 계수의 집합을 표현하는 개체(Chromosome)
- $C_{F_i}(j)$  :  $C_{F_i}$ 의  $j$ 번째 인자(Gene)
- $n$  : 예측 모델에 사용되는 계수의 총개수
- $k$  : GATP중 현재 시점

식 (5)에서 제시된 모집단 다양성 지표를 이용해 기존의 모집단을 새로이 재정의 하는 과정은 다음과 같다.

[단계 1] 유전 알고리즘을 이용한 모델 성숙기간 (GATP) 내의 Step  $n$ 을 진행하기 앞서 모집단 다양성 지표(Population Diversity Index)를 산출한다.  
모집단 다양성 지표와 비교할 난수(Random Number)를 생성한다.

[단계 2] (1) 단계 1의 난수가 모집단 다양성 지표보다 크거나 같으면 현재 Step  $n$ 에서 사용할 그림 4의 개체 해석 기준해를 바로 이전 시점에서 얻은 최적해로 변경한다. : 개체 해석 기준해가 바뀌므로써 개체 해석 범위가 바뀌게 되고 해 탐색의 대상이 되는 모집단이 변경된다.

(2) 단계 1의 난수가 모집단 다양성 지표보다 작으면 기존의 기준해를 계속 사용한다.

[단계 3] 새로운 시점 Step  $n$ 을 진행한다.

### 3.4 유전 연산

유전 알고리즘의 유전 연산에는 크게 교차(crossover), 돌연변이(mutation), 선별(selection)이 존재한다. 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 교차(Two-Point Crossover)를 교차 방법으로 사용하였고, 돌연변이의 경우는 개체 내 임의의 한 인자를 다른 임의의 인자와 자리바꿈하는 방법(Swap)을 사용하였다[6].

선별방법에는 크게 비례선별, 순위 기반 선별 방법이 있는데, 비례 선별은 개체를 선정함에 있어 모집단 내의 개체들의 적합도를 기반으로 하며, 순위 기반 선별은 모집단 내의 적합도의 상대적 순위에 따라 다음 세대로 갈 개체들을 선별하는 방법이다[13].

본 연구에서는 위에 언급된 선별 방법과는 다른 MGG(Minimal Generation Gap)모델 기반 선별방법을 사용하였다[15]. MGG 모델 기반 선별방법은 유전 알고리즘 내부에 선별 과정을 따로 두지 않고, 유전 연산 단계에서 선택된 개체들에 대해 유전 연산을 수행한 후 나온 자식 개체들과 수행 전

의 부모 개체들을 비교하여, 벌금 함수가 도입된 적합도 함수를 이용해 얻어진 적응도가 높은 2개의 개체만을 모집단에 대입하여 새로운 모집단을 구성하는 방법이다. 기존의 선별방법 들은 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 선별방법은 우수한 개체들의 손실을 최소화 할 수 있다는 장점을 가지며, 일례로 Job-Shop 스케줄링 문제에 있어서는 MGG 모델 기반 선별방법이 다른 선별 기법들에 비해 좋은 성능을 보인다는 연구 결과가 나와있다[15].

### 3.5 유도형 유전 알고리즘의 구조

이상의 유전 알고리즘의 구성 요소들, 벌금 함수를 이용한 적합도 함수, 모집단 다양성 지표들을 반영한 전체 유도형 유전 알고리즘의 절차는 다음과 같다.

#### 유도형 유전 알고리즘의 절차

[단계 0] 초기화  
 [단계 0.1] 파라미터 값 설정 : 모집단의 크기  $pop\_size$ , 교차율  $P_c$ , 돌연변이율  $P_m$ , 세대수  $gen\_size$   
 [단계 0.2] 데이터 기간 분류 :  $1 \sim t_{RTP}$ ,  $t_{RTP} + 1 \sim t_{GATP}$ ,  $t_{GATP} + 1 \sim t$   
 [단계 1]  $1 \sim t_{RTP}$ 기간에 대해 회귀분석 수행  
 시점 초기화 :  $t = t_{RTP} + 1$   
 [단계 2] 모집단의 초기화  
 난수 발생  
 If (난수  $\geq$  PDI) then  
     초기 해영역의 재정의  
     임의의 개체를  $pop\_size$ 개 만큼 생성  
 Else  
     임의의 개체를  $pop\_size$ 개 만큼 생성  
 [단계 3] 유전 연산  
 [단계 3.1] 모집단 내에서 임의의 두 개체 선정  
 [단계 3.2] 교차연산 적용  
 $P_c$ 의 확률로 선정된 개체에 대해 이점 교차를 수행.

교차연산을 수행한 개체에 대해  $P_m$ 의 확률로 돌연변이를 수행  
 [단계 3.3] 돌연변이  
 $1 - P_c$ 의 확률로 선정된 개체에 대해 자리바꿈 돌연변이를 수행  
 [단계 3.4] 새로운 모집단 생성을 위한 선별 적합도 함수를 이용해 2개의 부모개체와 2개의 자식개체 중 적합도가 높은 2개의 개체를 선별  
 [단계 3.5] If (새로운 모집단의 개체수 =  $pop\_size$ ) then  
     Go to [단계 4]  
 Else  
     Go to [단계 3.1]  
 [단계 4] 모집단에 대한 평가  
 식 (4)의 첫번째 항을 이용해 모집단 내의 각 개체들에 대해 예측의 정확성 평가  
 최적 개체 선정  
 [단계 5] If (세대수 =  $gene\_size$ ) then  
     Go to [단계 6]  
 Else  
     Go to [단계 3] with 새로운 모집단  
 [단계 6] 세대 평가  
 식 (4)의 첫번째 항을 이용해 각 세대별 최적 개체에 대해 예측의 정확성 평가  
 최적 개체 선정  
 [단계 7] If ( $t = t_{GATP}$ ) then  
     Go to [단계 8]  
 Else  
      $t = t + 1$   
     Go to [단계 2] with [단계 6]에서 얻은 시점  $t - 1$ 의 최적 개체  
 [단계 8] 인과형 예측 모델 생성  
 유전 알고리즘을 이용한 모델 성숙 기간을 거치며 얻어진 최적 개체를 이용해 인과형 예측 모델 생성  
 [단계 9] 인과형 예측 모델 평가  
 $t_{GATP} + 1 \sim t$  기간 동안에 생성된 인과형 예측 모델을 사용해 예측  
 식 (7)을 이용해 모델의 성능 평가

## 4. 실험 및 결과 분석

본 연구에서는 대표적인 연속 생산 공정 중의 하나인 실제 유리 용해 공정의 데이터를 이용해 제안된 유도형 유전 알고리즘을 이용한 인과형 예측 모델의 성능을 회귀 분석을 이용한 인과형 예

측 모델, 전통적인 유전 알고리즘을 이용한 인과형 예측 모델과 비교하였다.

#### 4.1 실험 설계

실험에 사용된 데이터는 대표적인 연속생산공정 중의 하나인 실제 유리 용해 공정에 사용된 실측치를 이용하였으며, 세개의 유사 유리 용해 공정을 대상으로 하였다. 그림 6에서 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 7나온 유리 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.

각 공정별 데이터들은 불량 지표와 이에 영향을 줄 수 있는 공정 인자로 구분할 수 있으며, 불량 지표에는 유리 용해 제품의 품질에 가장 큰 영향을 주는 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 인자들

에 대해 관측되거나 조작된 수치들이 기록되며, 실험에 사용된 데이터는 약 10개월 간의 실제 공정별 실측치를 이용하였다.

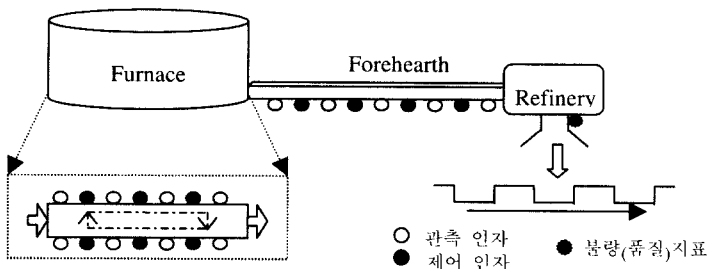
유도형 유전 알고리즘을 이용해 인과형 예측 모델을 구성하기 앞서 실제 연속 생산 공정의 특성을 반영하는 현실적인 인과형 예측이 가능하도록, 모델에 사용되는 데이터들 간의 공정 lag time을 결정하고 반영하였다. 일반적으로 lag time이란 어떤 인자를 변화 시켰을 때 그 결과로 다른 인자가 영향을 받을 수 있을 때의 그 시간적인 차이를 말한다. 생산 공정에서 진행중인 재공과 완성된 제품 사이에도 일정한 시간적 차이, 즉 lag time이 존재한다. 특히 공정에서 얻어진 데이터들, 즉 공정 인자의 값들은 관측한 시간이 각각 다르기 때문에 공정 인자 사이에도 lag time이 존재한다. 그래프를 사용한 경향 분석법은 두 개의 서로 다른 인자들 간의 관계를 규명하기 위해서 두 변수의 시간별 데이터의 추이를 동시에 그래프에 표시한 후 그 경향을 분석하는 방법을 사용한다. 우선, 두 변수가 보여주는 시간별 경향의 유사성( $S_l$ ) 지표를 식 (6)를 이용해 산출한 후 두 변수간의 lag time을 결정한다.

$$S_l = \sum_{t=1}^n |d_{it} - \bar{d}_l| / n \quad (6)$$

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

$p_t$  = 시점  $t$ 에서의 공정 인자의 값

$q_{t+l}$  = 시점  $t+l$ 에서의 불량지표의 값



〈그림 6〉 유리 용해 공정의 구조

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 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 concave 형태를 그리게 되면 그래프에서 적시된 최대값 시점에서 다른 lag time들에 비해 가장 높은 음의 유사성(반대 방향의 경향)을 보인다고 할 수 있으며  $S_l$  그래프가 convex 형태를 그리게 되면 그래프에서 적시된 최소값 시점에서 다른 lag time들에 비해 가장 높은 양의 유사성(같은 방향의 경향)을 갖는다고 볼 수 있다.

또한 불량 예측 모델을 구성하는 공정 인자들의 수를 변수 선택 과정을 통해 적정 수준으로 줄였다. 현장에서 수집되는 모든 인자들에 대한 반대한 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 적정한 수의 변수들을 선택하는 과정을 도입하였다. 변수 선택 방법으로는 모든 변수들에 대한 일차적인 고려가 가능하고 적용의 용이성을 고려하여, 모든 공정 인자들이 포함된 초기 회귀 분석 모델에서 시작하여 기여도가 가장 낮은 변수들을 차례대로 하나씩 소거하여 나가는 방법인 Backward Elimination 방법을 사용하였다[1, 5].

실험은 우선 그래프 경향법을 이용해 77개 각 공정 인자별로 불량 지표와의 lag time을 산출한 후 데이터 전처리 과정을 통해 결정된 lag time을 데이터에 반영해 주었다. 또한, 중점적인 관리를 필요로 하는 인과형 예측 모델에 포함될 중요한 인자들을 선별해 내기 위해 변수 선택 과정을 통해 라인별로 총 77개의 공정 인자들을 48개의 인자로 줄였다. 유도형 유전 알고리즘을 적용하기 위한 데

이터 기간의 구분은, 전체 데이터 기간 중 회귀 분석을 이용한 모델 성숙 기간은 6개월간으로 설정하였으며, 유전 알고리즘을 이용한 모델 성숙 기간은 12시간으로 설정해 실험을 진행하였다.

순수 예측 기간은 현재 시점으로부터 하루 동안과 한달 동안의 두 경우에 대해 완성된 인과형 예측 모델로 예측하였다. 세 개의 각 공정별로 회귀 분석을 사용한 예측 모델과 전통적인 유전 알고리즘을 사용한 인과형 예측 모델을 유도형 유전 알고리즘을 사용한 인과형 예측 모델과 비교하였다. 모델간의 예측의 정확성은 MAD(Mean Absolute Deviation) 수치를 이용해 비교하였다.

MAD 수치는 다음과 같이 계산한다[18].

$$MAD = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (7)$$

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

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.

신뢰성 있는 모델의 검증을 위해 모델간의 비교는 시간 축(time window)을 3일 간격으로 10회 이동해 가며 10개의 독립적인 모델들의 성능을 서로 비교하였다.

제안된 유전 알고리즘 및 검증 실험, 그리고 불량 예측 모델의 시스템 구현은 Pascal 언어와 Delphi를 사용하였고, Pentium II 400 MHz의 IBM-PC를

사용하였다.

### 4.2 실험 결과

유도형 유전 알고리즘을 이용한 인과형 예측 모델과 회귀 분석을 이용한 예측 모델, 전통적인 유전 알고리즘을 이용한 인과형 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.

유도형 유전 알고리즘과 전통적인 유전 알고리즘을 이용한 인과형 예측 모델의 경우는 10회의 반복 실험의 평균을 사용하였다.

회귀 분석(Regression)을 이용한 예측 모델, 전

<표 1> 인과형 모델 간 예측 성능 비교 실험 결과 : MAD 비교

Line 1	1 Day			1 Month		
	Regression	Canonical GA	GGA	Regression	Canonical GA	GGA
Experiment 1	1.12	0.36	0.39	7.61	8.30	3.86
Experiment 2	5.54	1.64	0.73	6.36	6.23	2.39
Experiment 3	2.46	0.97	1.39	6.99	7.33	2.62
Experiment 4	0.86	0.47	0.83	8.03	8.55	2.98
Experiment 5	7.15	1.63	0.81	8.88	13.90	4.08
Experiment 6	8.32	4.00	2.52	12.75	15.70	3.97
Experiment 7	1.96	1.16	1.74	16.77	17.90	4.99
Experiment 8	10.49	2.36	1.10	12.33	5.59	2.32
Experiment 9	1.31	1.47	0.86	6.64	6.61	3.92
Experiment 10	0.24	0.88	1.10	14.17	13.90	5.36
Average	3.945	1.494	1.147	10.053	10.401	3.649
Line 2	Regression	Canonical GA	GGA	Regression	Canonical GA	GGA
Experiment 1	10.48	0.49	0.62	22.66	18.20	15.20
Experiment 2	14.70	10.70	4.67	25.47	25.70	15.30
Experiment 3	9.72	14.70	17.50	30.00	28.70	19.60
Experiment 4	11.94	5.93	4.46	27.68	42.30	16.70
Experiment 5	5.43	1.02	1.09	32.62	31.00	26.50
Experiment 6	3.19	1.86	3.02	46.54	24.90	39.00
Experiment 7	6.99	16.40	2.44	20.35	41.10	19.40
Experiment 8	36.97	25.50	23.40	17.06	14.10	32.80
Experiment 9	5.73	6.11	6.59	62.61	68.30	40.70
Experiment 10	16.46	18.80	16.08	80.11	62.10	59.20
Average	12.161	10.151	7.987	36.510	35.640	28.440
Line 3	Regression	Canonical GA	GGA	Regression	Canonical GA	GGA
Experiment 1	21.55	18.10	16.10	19.77	21.40	18.00
Experiment 2	6.22	3.41	4.30	15.96	14.50	15.90
Experiment 3	8.22	2.04	3.67	18.90	16.20	18.00
Experiment 4	12.20	18.40	17.70	19.49	17.20	15.70
Experiment 5	28.30	26.40	24.60	26.95	25.70	4.23
Experiment 6	15.71	4.46	1.09	26.64	39.90	14.80
Experiment 7	15.67	5.27	8.02	21.41	14.40	20.70
Experiment 8	4.13	0.88	4.04	34.74	40.30	20.60
Experiment 9	3.26	10.20	4.22	25.31	25.00	16.20
Experiment 10	1.20	3.48	0.97	38.51	57.40	21.00
Average	11.646	9.264	8.471	24.768	27.200	16.513

통적인 유전 알고리즘(Canonical GA : Canonical Genetic Algorithm)을 이용한 인과형 예측 모델, 유도형 유전 알고리즘(GGA : Guided Genetic Algorithm)을 이용한 인과형 예측 모델간의 세 개의 실제 유리 용해 공정에 대한 예측 성능 비교 실험 결과는 <표 1>과 같다.

<표 1>에서 보는 바와 같이 모든 경우에 있어서 유도형 유전 알고리즘을 이용한 인과형 예측 모델이 회귀 분석을 이용한 예측 모델과 전통적인 유전 알고리즘을 이용한 인과형 예측 모델보다 좋은 성능을 보였다. 전통적인 유전 알고리즘을 이용한 인과형 예측 모델의 경우에는, 순수 예측 기간이 하루인 단기 예측의 경우에 회귀 분석을 이용한 예측 모델보다 좋은 성능을 보였으나 순수 예측 기간이 한 달인 장기 예측의 경우에는 오히려 좋지 않은 성능을 보였다. 이것은 전통적인 유전 알고리즘을 이용한 인과형 예측 모델의 경우 순수 예측을 시작하는 시점으로부터 가까운 과거의 데이터들에 단순히 가중치를 주는 개념으로 가까운 미래의 예측에는 좋은 성능을 보이지만 예측 기간이 늘어나게 되면 예측의 성능이 떨어진다고 볼 수 있다. 그러나, 유도형 유전 알고리즘을 이용한 인과형 예측 모델의 경우는 전통적인 유전 알고리즘의 장점을 살리면서 벌금 함수와 모집단의 다양화를 통해 인과형 예측 모델의 정확성 및 강건성(Robustness)을 높이기 때문에 회귀 분석을 이용한 예측 모델보다 뛰어난 예측 성능을 보인다고 할 수 있다.

유도형 유전 알고리즘을 이용한 인과형 예측 모델의 예측 성능 향상 정도를 보다 쉽게 파악하기 위해 MDR(MAD Deviation Ratio)을 이용해 오차를 비교하면 <표 2>와 같다. 사용된 MDR은 식 (8), 식 (9)과 같이 나타낼 수 있다. 식 (8)은 회귀 분석을 이용한 예측 모델과 유도형 유전 알고리즘을 이용한 인과형 예측 모델간의 비교를 위한 MDR 식이며 식 (9)는 전통적인 유전 알고리즘을 이용한 인과형 예측 모델과 유도형 유전 알고리즘을 이용한 인과형 예측 모델간의 비교를 위한

MDR 식이다.

$$MDR 1 =$$

$$\frac{\text{Regression Model의 MAD} - \text{GGA Model의 MAD}}{\text{Regression Model의 MAD}} \quad (8)$$

$$MDR 2 =$$

$$\frac{\text{Canonical GA Model의 MAD} - \text{GGA Model의 MAD}}{\text{Canonical GA Model의 MAD}} \quad (9)$$

라인별로 MDR 수치를 이용해 오차를 비교하면 <표 2>와 같다.

<표 2> 인과형 예측 모델 간 MDR(MAD Deviation Ratio) 비교

	MDR 1		MDR 2	
	1 Day	1 Month	1 Day	1 Month
Line 1	0.709	0.637	0.232	0.649
Line 2	0.343	0.221	0.213	0.202
Line 3	0.273	0.333	0.086	0.393
Average	0.442	0.397	0.177	0.415

<표 2>에서의 각 수치는 라인별로 시간 축을 이동하며 생성된 10개의 독립적인 모델들에 대한 MDR수치들의 평균값이다. <표 2>에서 보는 바와 같이 거의 모든 경우에서 유도형 유전 알고리즘을 이용한 인과형 예측 모델이 다른 모델들보다 좋은 예측 결과를 산출함을 손쉽게 알 수 있다.

또한 유도형 유전 알고리즘의 성능을 보다 자세히 분석하기 위해 다음의 네 가지 유전 알고리즘을 이용한 변형 예측 모델들 간의 성능을 표 3과 같이 비교하였다.

- Canonical GA
- Penalty-GA : 벌금함수를 이용한 적합도 함수만을 전통적인 유전 알고리즘에 적용한 인과형 예측 모델
- PDI-GA : 모집단 다양성 지표만을 전통적인 유전 알고리즘에 적용한 인과형 예측 모델
- GGA

<표 3>에서의 각 수치는 위에 언급된 네 가지 모델들을 사용해 1라인에 대해 시간 축을 이동하며 생성된 10개의 독립적인 모델들을 사용해 예측 기간이 한 달인 경우에 얻어진 MAD 수치들이다. <표 3>에서 보는 바와 같이 유도형 유전 알고리즘의 성능은 벌금 함수의 사용보다는 보다 넓은 해공간을 탐색할 수 있는 기능을 제공하는 모집단 다양성 지표의 사용에 주로 기인한다는 것을 알 수 있다.

<표 3> 유전 알고리즘을 이용한 예측 모델들 간의 성능 비교 : MAD 비교

Experiment Number	Canonical GA	Penalty-GA	PDI GA	GGA
1	8.30	8	4.03	3.86
2	6.23	6.33	2.25	2.39
3	7.33	7.32	2.51	2.62
4	8.55	8.36	2.89	2.98
5	13.90	10.7	4.21	4.08
6	15.70	14.1	4.81	3.97
7	17.90	17.9	5.46	4.99
8	5.59	5.57	2.49	2.32
9	6.61	6.61	4.22	3.92
10	13.90	13.9	5.34	5.36
Average	10.401	9.879	3.821	3.649

## 5. 결 론

전통적인 유전 알고리즘은 발견적 기법 중의 하나로서 잠재해들로 이루어진 모집단을 이용해 해를 찾기 때문에 복잡하고 대형의 수리 문제를 해결하는 데 많이 사용되고 있으며, 다양한 분야에 적용되고 있다. 본 연구에서는 연속 생산 공정의 인과형 예측 모델을 구축하는데 있어 전통적인 유전 알고리즘에 벌금 함수를 도입함으로써 유전 알고리즘 내의 선별 과정의 효율을 높이며, 유전 알고리즘에서 사용되는 모집단 내의 개체들을 다양화 시킴으로써 부분 최적 해에서 벗어나지 못하는 경우를 방지함으로써 보다 효율적으로 정확한 최

적해를 산출 가능하게 하는 유도형 유전 알고리즘을 제안하였고, 공정 내 불량을 예측 할 수 있는 인과형 예측 모델을 구축하였다. 제안된 인과형 예측 모델은 유도형 유전 알고리즘을 기반으로 실제 공정 조건을 반영해 주기 위해 lag time의 개념을 도입하여 사용하였고, 공정 내의 방대한 데이터들의 관리 및 분석에서 오는 비효율성을 막기 위해 공정 내 주요 인자들을 선정하여 인과형 예측 모델에 사용하였다. 아울러 유도형 유전 알고리즘을 인과형 예측 모델의 구축에 적용하기 위해 데이터 기간을 새로이 분류하였다.

제안된 모델은 기존의 회귀 분석을 이용한 예측 모델 및 전통적인 유전 알고리즘을 이용한 인과형 예측 모델과 비교해 시험 대상으로 삼은 대표적인 연속 생산 공정 중의 하나인 유리 용해 공정의 세 개의 라인에서 모두 우수한 성능을 보였으며, 유도형 유전 알고리즘의 초기 모집단을 임의의 개체가 아닌 회귀 분석을 이용한 현실적인 개체들로 사용함으로써 유전 알고리즘의 수렴 속도를 향상시키고 예측 결과의 신뢰성을 높일 수 있도록 했다. 제안된 인과형 예측 모델은 실제 연속 생산 공정의 데이터를 검증하였으므로 유사 다른 연속 생산 공정에의 적용도 가능할 것으로 보인다.

향후 추가적인 연구가 필요한 부분은 유전 알고리즘을 이용한 모델 성숙 기간의 길이와 예측의 정확성간의 관계를 규명하는 것과 유도형 유전 알고리즘을 이용한 인과형 예측 모델을 연속 생산 공정만이 아닌 다른 여러 공정 및 분야에 적용시키는 연구도 진행되어야 할 것이다.

## 참 고 문 헌

- [1] Afifi, A.A. and V. Clark, *Computer-Aided Multivariate Analysis*, Second Edition, Van Nostrand Reinhold Co., 1990.
- [2] Chiraphadhanakul, S., P. Dangprasert and V. Avatchanakorn, "Genetic Algorithms in Forecasting Commercial Banks Deposit," *Pro-*

- ceedings of the 1997 IEEE International Conference on Intelligent Processing Systems*, 1997.
- [3] Coit, D.W. and A.E. Smith, "Penalty Guided Genetic Search For Reliability Design Optimization," *Computers and Industrial Engineering*, Vol.30, No.4(1996), pp.895-904.
- [4] Coit, D.W, A.E. Smith and D.M. Tate, "Adaptive Penalty Methods for Genetic Optimization of Constrained Combinatorial Problems," *INFORMS Journal of Computing*, Vol.8(1996), pp.173-182.
- [5] Flury, B. and H. Riedwyl, *Multivariate Statistics : A Practical Approach*, Chapman and Hall, 1998.
- [6] Goldberg, D., *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesely, Reading, Massachusetts, USA, 1989.
- [7] Holland, J., *Adaptation In Natural And Artificial Systems*, Univ. of Michigan Press, Ann Arbor, 1975.
- [8] Ju, Y.J., C. Kim and J.C. Shim, "Genetic-Based Fuzzy Models : Interest Rate Forecasting Problem," *Computers and Industrial Engineering*, Vol.33, No.3(1997), pp.561-564.
- [9] Kai, F. and X. Wenhua, "Training Neural Network with Genetic Algorithms for Forecasting the Stock Price Index," *Proceedings of the 1997 IEEE International Conference on Intelligent Processing Systems*, 1997.
- [10] Kim, D. and C. Kim, "Forecasting Time Series with Genetic Fuzzy Predictor Ensemble," *IEEE Transactions on Fuzzy System*, Vol.5, No.4(1997), pp.523-535.
- [11] Lau, T.L., *Guided Genetic Algorithms*, Ph.D Thesis, Department of Computer Science, Univ. of Essex, UK, 1998.
- [12] Leung, Y., Y. Gao, and Z.B. Xu, "Degree of Population Diversity-A Perspective on Prerequisite Convergence in Genetic Algorithms and its Markov Chain Analysis," *IEEE Transactions on Neural Networks*, Vol.8, No.5 (1997), pp.1165-1176.
- [13] Miller, B., D. Goldberg, "Genetic Algorithms, Selection Schemes, and the Varying Effects of Noise," *IlligAL Report*, No.95009, 1995.
- [14] Richardson, J.T., M.R. Palmer, G. Liepins, and M. Hilliard, "Some Guidelines for Genetic Algorithms with Penalty Functions," *Proceedings of the 3<sup>rd</sup> International Conference on Genetic Algorithms*, 1989.
- [15] Shi, G., "A Genetic Algorithms Applied to a Classic Job Shop Scheduling Problem," *International Journal of Systems Science*, Vol.28, No.1(1997), pp.25-32.
- [16] Tsujimura, Y. and M. Gen, "Entropy-Based Genetic Algorithms for Solving TSP," *1998 2<sup>nd</sup> International Conference on Knowledge-Based Intelligent Electronic Systems*, 1998.
- [17] Wen, J.Y., Q.H. Wu, D.W. Shimmin, D.R. Turner and S.J. Cheng, "Population Diversity Based Genetic Algorithm for Fuzzy Control of Synchronous Generators," *Proceedings of the IEEE International Symposium on Computer Aided Control System Design*, 1999.
- [18] Winston, W.L., *Operational Research Applications and Algorithms*, Second Edition, PWS-KENT Publishing CO., 1991.
- [19] Yip, D.H.F., E.L. Hines, and W.W.H Yu, "Application of Artificial Neural Networks in Sales Forecasting," *Proceedings of the 1997 IEEE International Conference on Neural Networks*, 1997.
- [20] Zitzler, E. and L. Thiele, "Multiobjective Evolutionary Algorithms : A Comparative Case Study and the Strength Pareto Approach," *IEEE Transactions on Evolutionary Computation*, Vol.3, No.4(1999), pp.251-271.