

# 다중 마르코프 체인의 베이저안 진화 알고리즘

이시은<sup>0</sup>  
천안외국어대학 컴퓨터정보과  
selee@ccfs.ac.kr

장병탁  
서울대학교 전기 컴퓨터공학부  
btzhang@bi.snu.ac.kr

## A Bayesian Evolutionary Algorithm with Multiple Markov Chains

Si-Eun Lee<sup>0</sup>  
Dept. of Computer Information  
Cheonan College of Foreign Studies

Byoung-Tak Zhang  
School of Computer Science And Engineering  
Seoul National University

### 요 약

진화 연산의 확률적 모델인 베이저안 진화 알고리즘의 수렴 특성에 대한 이전 연구를 통해 개체군 크기가 1인 경우에 대해 베이저안 진화 알고리즘을 단일 체인 MCMC로 변환하여 수렴 특성을 보였다. 본 논문에서는 개체군 크기가 1로 제한 되지 않는 경우 베이저안 진화 알고리즘을 다중 체인의 개체군으로 생각하여 수렴 특성을 살펴본다.

### 1. 서 론

자연계의 진화 현상에 기반한 진화 알고리즘이 전역 최적해에 수렴하느냐, 만약 수렴한다면 얼마나 빠르게 수렴하느냐 하는 등의 연구는 진화 알고리즘 분야의 가장 기본적인 이론적 문제로 여러 특별한 형식의 진화 알고리즘, 엘리트 선택을 하는 유전자 알고리즘과 특정한 목적 함수를 갖는 진화 전략 등, 에 관해 연구되어 오고 있다 [ 1, 2]. 진화 알고리즘의 수렴 여부를 알아보기 위한 연구들은 대체적으로 시뮬레이티드 어닐링 접근법, Vose\_Liepins 모델 접근법, 확률 모델 접근법 등으로 요약된다.

시뮬레이티드 어닐링 접근법에서는 진화 알고리즘의 어떤 인자를 온도로 간주하고 기존의 시뮬레이티드 어닐링 수렴 특성을 이용한다. Vose\_Liepins 모델 접근법에서는 진화 알고리즘 (특히 유전자 알고리즘) 을 비확률적인 결정적 동력 시스템으로 생각하고 시스템이 평형 집합으로 수렴함을 보인다. 가장 널리 쓰이는 확률 모델 접근법에서는 진화 알고리즘을 마르코프 체인으로 모델링하여 에르고딕성 (ergodicity) 에 기반한 분석을 한다.

베이저안 진화 알고리즘이란 학습과 최적화를 위한 진화 연산의 확률적 모델이다 [3]. 베이저안 진화 알고리즘 (Bayesian Evolutionary Algorithm, 이하 BEA) 은 개체들의 사후 확률 분포를 평가한 후 그 분포로부터 자손들을 샘플링 해나간다. 기본적인 진화 알고리즘이 단순히 존재하는 개체들의 변이나 교배 등의 재결합으로부터 자손들을 생성하는 것에 반하여 BEA는 개체들의 확률분포로부터 자손들을 생성하므로 문제 공간의 탐색에 유리하고 에러에도 강하다고 할 수 있다. 또한 기존의 시뮬레이션 기반의 방법과는 달리 탐색이 개체군 단위로 진행된다. 따라서

BEA는 단순 진화 알고리즘과 마르코프 체인 몬테 카를로 (Monte Carlo Markov Chain, 이하 MCMC) 방법의 결합으로 인해 탐색 공간의 효율적인 탐색이 가능한 알고리즘이라고 할 수 있다. BEA의 실용성은 여러 실험적 결과들을 통해 보여 왔고 이론적 안정성을 보이기 위한 수렴 특성에 대한 연구 [4, 5] 또한 계속 되어 왔다.

먼저 개체군의 크기를 1로 제한하고 고정된 차원의 탐색 공간을 갖는 경우에 대하여 어닐링 기법을 사용하여 수렴 특성을 보였다. 또한 개체군의 크기를 1로 고정하지 않은 경우, BEA를 베이저안 입자 필터로 변환할 수 있다. 베이저안 입자 필터는 입자들을 적합도에 따라 중복하거나 버림으로써 자손들을 생성한다. 사후 확률 분포는 BEA와 베이저안 입자 필터 모두에서 시간이 지나 데이터가 축적됨에 따라 진화한다. 따라서 입자 필터의 수렴 특성을 이용하여 개체들의 수가 증가할 때 베이저안 입자 필터가 사후 확률 분포에 수렴함을 보였다. 그러나 개체군의 크기가 작을 경우에는 실제 분포와 경험적 평가치와의 평균 제곱 오차가 클 수도 있는 문제점이 있다.

본 논문에서는 [4]에서의 개체군 크기가 1이라는 가정을 벗어나 즉 단일체인이 아닌 다중 체인들의 개체군을 유지하여 BEA가 MCMC방법에서와 같이 사후 확률 분포에 수렴함을 증명한다. 본 논문의 구성은 다음과 같다. 2장에서는 지금까지 연구된 베이저안 진화 알고리즘의 수렴 특성에 대해 정리해본다. 3장에서는 MCMC 방법과 그의 수렴 특성에 대해 살펴본다. 4장에서는 베이저안 진화 알고리즘을 다중 체인의 개체군으로 생각하여 수렴 특성을 살펴본다. 5장에서는 결론 및 향후 과

제를 언급한다.

## 2. 단일 체인 베이저안 진화 알고리즘

BEA는 진화 연산의 확률적 모델로 탐색 공간  $\Theta$ 의 각 개체  $\theta$ 에 대하여  $\pi(\theta)$ 를 사전 확률 분포라 하고  $f(D|\theta)$ 를 데이터  $D$ 에 대한 가능도라고 할 때 각 개체  $\theta$ 의 사후 확률 분포를  $\pi(\theta|D)$ 로 표시한 후 다음과 같이 베이지 정리를 이용하여 매 세대마다 사후 확률 분포  $\pi(\theta|D)$ 를 평가해 나간다.

$$\begin{aligned}\pi(\theta|D) &= \frac{f(D|\theta)\pi(\theta)}{\int f(D|\theta)\pi(\theta)d\theta} \\ &\approx \frac{f(D|\theta)\pi(\theta)}{\sum_{\theta' \in \Theta} f(D|\theta')\pi(\theta')}.\end{aligned}$$

그런 후 사후 확률 분포로부터 변형과 선택 연산자들을 사용하여 자손 개체들을 샘플링 해 나간다. 표준 베이저안 진화 알고리즘 (canonical BEA)은 다음과 같다.

### 알고리즘 2.1 (표준 BEA)

1. (초기화)  $\pi_0(\theta)$ 로부터  $\Theta^0 = \{\theta_1^0, \dots, \theta_M^0\}$ 를 생성한다. 온도  $T_0$ 를 초기화한다. 세대 카운트  $t \leftarrow 0$ 으로 한다
2. (D-단계) 데이터  $D$ 를 생성 (관찰) 한 후 가능도  $f(D|\theta_i')$ 를 계산한다.
3. (P-단계) 사후 확률 분포  $\pi_i(\theta_i'|D)$ 를 평가한다.
4. (V-단계)  $\pi_i(\theta)$ 로부터 샘플링하여  $L$ 개의 변형들  $\Theta' = \{\theta_1', \dots, \theta_L'\}$ 를 생성한다.
5. (S-단계)  $\Theta'$ 으로부터 가능도  $f(D|\theta_i')$ 에 의해  $M$ 개의 개체들을 선택하여  $\Theta^t = \{\theta_1^t, \dots, \theta_M^t\}$ 를 구성한다.
6. (R-단계) 사전 확률 분포  $\pi_i(\theta)$ 를 재고한다. 온도  $T_t$ 를 갱신한다.
7. (루프) 세대 카운트  $t \leftarrow t + 1$ 로 한 후 단계 2로 이동한다.

알고리즘은 기본적으로 D (data), P (posterior), V (variation), S (selection), R (revision)의 5개의 단계로 구성된다. R, D 그리고 P 단계는 각각 사전 확률 분포, 가능도와 사후 확률 분포의 계산을 포함한다. V와 S 단계에서는 시간  $t$ 에서 사후 확률 분포로부터의 샘플링을 구현한다. 진화 연산의 수렴 성질을 보이기 위해 마르코프 체인 분석이 사용될 수 있다. 일반적으로 마르코프 체인이 동질적이거나 문제 탐색 공간이 미리 가정된대로의 잘 정의된 기하 구조를 갖는다는 정칙 조

건을 가정을 한다. 그러나 알고리즘 2.1에서 생성되는 개체  $\theta_i^t$ 는  $\pi_i(\theta_i^t)$ 로부터 생성되는데 그 분포는 시간에 따라 변화한다. 따라서 BEA의 완벽한 이론적인 분석은 그것에 의해 생성되는 마르코프 체인이 비동질적이므로 실제적으로는 상당한 어려움이 있다. 그러나 어닐링 기법을 사용하여 비동질적 마르코프 체인의 점근적 수렴성을 보일 수 있음이 최근 많이 연구되고 있다. 만약 개체군의 크기를 1로 제한하고 고정된 차원의 상태 공간만을 고려하기로 가정하면 BEA는 기하적 수렴 결과들이 잘 알려진 MCMC 방법으로 변환할 수 있다. 먼저  $T_t = T$ 와  $M = 1$ 로 가정하면 다음의 단일 체인 BEA를 얻는다.

### 알고리즘 2.2 (단일 체인 BEA)

1. (초기화)  $\pi_0(\theta)$ 로부터  $\theta^0$ 를 생성한다.  $t \leftarrow 0$
2. (V-단계) 사전 분포  $\pi(\theta)$ 로부터  $\theta'$ 를 생성한다.
3. (S-단계) 가능도  $f(D|\theta_i')$ 를 평가하고
 
$$\theta^{t+1} = \begin{cases} \theta' & \text{w.p. } \min\left\{\frac{f(D|\theta')}{f(D|\theta^t)}, 1\right\} \\ \theta^t & \text{otherwise.} \end{cases}$$
를 결정한다.
4. (루프)  $t \leftarrow t + 1$ 로 한 후 단계 2로 이동한다.

이 때 Metropolis-Hastings 알고리즘에 대한 수렴 결과들이 단일 체인을 갖는 BEA의 Metropolis-Hastings 버전에 적용된다.

### 3. 마르코프 체인 몬테 카를로

MCMC [6]는 다양한 범위의 통계적 모델링을 제공하는 마르코프 체인을 사용한 몬테 카를로 적분이라고 할 수 있다. 높은 차원의 확률분포에 대한 적분이 필요할 때 몬테 카를로 적분은 그 분포로부터 샘플들을 추출하고 기대값을 근사하기 위하여 샘플들의 평균을 취한다. MCMC는 이러한 샘플들을 잘 고안된 마르코프 체인으로부터 추출한다. 이러한 마르코프 체인을 만드는 방법에는 여러가지가 있는데 그 중 대표적인 Metropolis-Hastings 알고리즘은 다음과 같다. 마르코프 체인  $\{X_0, X_1, \dots\}$ 을 생각해 보자. 주어진  $X_t$ 에 대해 다음 상태  $X_{t+1}$ 의 후보는 제안 분포  $q(\cdot|X_t)$ 로부터 샘플링한  $Y$ 라고 하자.  $Y$ 는 확률

$$\alpha(X_t, Y) = \min\left(1, \frac{\pi(Y)q(X_t|Y)}{\pi(X_t)q(Y|X_t)}\right)$$

로 수락되어 다음 상태가 되거나 거절된다면 체인은 이동하지 않는다. 기본 알고리즘은 다음과 같다.

### 알고리즘 3.1 Metropolis-Hastings

1.  $X_0$ 를 초기화 한다.  $t \leftarrow 0$ .

2. 제안분포  $q(\cdot | X_t)$  로 부터  $Y$  를 샘플링한다.
3. 균등분포(0,1)로 부터 확률 변수  $U$  를 샘플링한다.
4. 만약  $U \leq \alpha(X_t, Y)$  라면  $X_{t+1} = Y$ ,  
그렇지 않으면  $X_{t+1} = X_t$ .
5. 세대 카운트  $t \leftarrow t + 1$  로 한 후 단계 2로 이동한다.

#### 4. 다중 체인 BEA

단일 체인 BEA에서 개체군의 크기가 1이라는 제한에서 벗어나 같은 정상 분포  $\pi$  를 갖는  $N$  개의 독립적인 대칭 분포인 Metropolis-Hastings 샘플러  $\{\pi_i^t\}_{i=1}^N$  를 고려하자. 이러한  $N$  개 체인들의 집합의 결합 전이 분포에 대한 detailed balance를 만족하도록 조정해 나간다. 또한  $T_i = T$  이고 고정된 차원의 상태 공간만을 고려하기로 가정한다.

#### 알고리즘 4.1 (다중 체인 BEA)

1. (초기화)  $\{\pi_i^0\}_{i=1}^N$  로부터  $\Theta^0 = \{\theta_1^0, \dots, \theta_M^0\}$  를 생성한다. 세대 카운트  $t \leftarrow 0$  으로 한다
2. (D-단계) 임의의 확률로 체인의 인덱스  $i$  를 선택한다. 데이터  $D$  를 관찰한 후 가능도  $f(D | \theta_i^t)$  를 계산한다.
3. (P-단계) 사후 확률 분포  $\pi_i(\theta_i^t | D)$  를 평가한다.
4. (V-단계)  $\pi_i^t$  로 부터  $\theta'$  을 샘플링하여  $N$  개의 변형들  $\Theta' = \{\theta_1', \dots, \theta_N'\}$  를 생성한다.
5. (S-단계) 가능도  $f(D | \theta_i^t)$  에 의해
 
$$\theta^{t+1} = \left\{ \begin{array}{l} \theta' \text{ w.p. } \min\left\{\frac{f(D|\theta')}{f(D|\theta^t)}, 1\right\} \\ \theta^t \text{ otherwise.} \end{array} \right\}$$
 을 선택하여  $\Theta^t = \{\theta_1^t, \dots, \theta_N^t\}$  를 구성한다.
6. (R-단계) 사전 확률 분포  $\pi_i(\theta)$  를 재고한다.
7. (루프) 세대 카운트  $t \leftarrow t + 1$  로 한 후 단계 2로 이동한다.

알고리즘의 D 단계에서는 임의의 동일한 확률로 다중 체인 중의 인덱스  $i$  를 선택하고 선택된  $\theta_i^t$  에 대한 가능도를 평가한다. P 단계에서는 선택된 체인의 사후 확률 분포를 평가한다. V 단계에서는  $\pi_i$  로 부터 샘플링한  $\theta'$  로 대체하여  $\Theta'$  를 생성한다. S-단계에서는 수락 확률을 계산하여  $\theta'$  을 수락하거나 거절하여  $\Theta^t$  을 구성한다. R-단계에서는 다음 단계를 위해 사전 확률을 재고한다.

#### 정리4.1 베이저안 진화 알고리즘은 최적해에 수렴한다.

(증명)  $T_k$  를 임의의 체인  $k$  에서의 전이 확률이라 할 때 다음과 같이 detailed balance를 만족한다.

$$\begin{aligned} \pi(\theta)T_k(\theta, \theta') &= \pi(\theta)\pi_k(\theta, \theta')\alpha(\theta, \theta') \\ &= \pi_k(\theta, \theta') \min(f(D | \theta), f(D | \theta')) \\ &= \pi_k(\theta', \theta) \min(f(D | \theta'), f(D | \theta)) \\ &= \pi(\theta')\pi_k(\theta', \theta)\alpha(\theta', \theta) = \pi(\theta')T_k(\theta', \theta) \end{aligned}$$

따라서  $\pi$  는 정상 분포이다. 또한  $\pi_k(\theta, \theta') \neq 0$  이고  $\pi(\theta) \neq 0$  이므로 마르코프 체인은 에르고딕성을 갖는다.

□

#### 5. 결론

본 논문에서는 개체군의 크기가 1로 제한 되지 않는 경우 베이저안 진화 알고리즘을 다중 체인의 개체군으로 생각하여 수렴 특성을 살펴 보았다. 향후 과제로 좀 더 일반적인 조건 하에서의 베이저안 진화 알고리즘의 수렴 특성과 기존의 진화 알고리즘의 수렴 특성과의 비교 연구를 진행하고 있다.

#### 6. 참고문헌

- [1] Rudolph, G., "Convergence properties of canonical genetic algorithms", *IEEE Trans. On Neural Networks*, vol.5, no. 1, pp. 96-101, 1994
- [2] Francois, O., "An evolutionary strategy for global minimization and its Markov chain analysis", *IEEE Trans. On evolutionary Computation*, vol. 2, no. 3, pp77-90, 1998
- [3] Byoung-Tak Zhang. "A Bayesian Framework for Evolutionary Computation", *Proc CEC, Special Session On Theory and Foundations of Evolutionary Computation*, 1999
- [4] Byoung-Tak Zhang, Paass, G., and Muehlenbein, H., "Convergence Properties of Incremental Bayesian Evolutionary Algorithm with Single Markov Chains", *Proc. CEC, Special Session on Theory and Foundations of Evolutionary Computation*, 2000
- [5] Si-Eun Lee, Byoung-Tak Zhang, and Doucet, A., "Convergence Properties of Bayesian Evolutionary Algorithms with Population Size Greater Than 1", *Proc. CEC, Special Session on Theory and Foundations of Evolutionary Computation*, 2001
- [6] Gilks, W.R., Richardson, S., and Spiegelhalter, D.J., *Markov Chain Monte Carlo in Practice*, Chapman & Hall, 1996