개체군 크기 2 이상인 베이지안 진화 알고리즘의 수렴 특성

이시은[0] 장병탁
천안외국어대학 컴퓨터정보과 서울대학교 컴퓨터공학부
selee@mail.chonan-c.ac.kr, btzhang@scai.snu.ac.kr

Convergence Properties of Bayesian Evolutionary Algorithms with Population Size Greater Than 1

Si Eun Lee[0] Byoung-Tak Zhang

Department of Computer Information, Chonan College of Foreign Studies[0]
School of Computer Science and Engineering, Seoul National University

요 약

진화 연산의 확률적 모델인 베이지안 진화 알고리즘이 개체군의 크기를 1 로 제한하고 고정된 차원의 탐색 공간을 갖는 경우, 목표 확률분포에 수렴함이 이전 연구[2]를 통해 증명되었다. 본 논문에서는 개체군의 크기가 2 이상인 경우의 베이지안 진화 알고리즘을 개체군 자체를 하나의 상태로 보는 단일 체인의 베이지안 입자 필터(particle filter) 로 변환하여, 입자 필터의 수렴 특성을 이용하여 목표 확률분포에 수렴함을 증명한다.

1.서론

베이지안 진화 알고리즘 (Bayesian Evolutionary Algorithm, 이하 BEA)[3] 은 진화 연산의 확률적 모델로 사전(prior) 확률분포로부터 얻어진 개체군으로 시작하여 부모 개체들의 사후(posterior) 적합도(fitness) 분포를 평가한 후 그 분포로부터 변형(variation)과 선택 연산자들을 사용하여 자손 개체들을 샘플링 해 나간다. 기본적인 진화 알고리즘이 단순히 존재하는 개체들의 변이나 교배 등의 재결합으로부터 자손들을 생성하는 것에 반하여 BEA 는 개체들의 확률분포로부터 자손들을 생성하므로 탐색공간의 탐험(exploration)에 유리하고 에러에도 강하다(robust)고 할 수 있다. 또한 기존의 시뮬레이션 기반의 방법과는 달리 탐색이 개체군 단위로 진행됨으로써 보다 효율적임이 실험적으로 보여졌다. 따라서 BEA 는 단순 진화 알고리즘과 MCMC (Monte Carlo Markov Chain) 방법의 결합으로 인해 탐색 공간의 효율적인 탐험이 가능한 알고리즘이라고 할 수 있다.

이전 연구[2]에서는 개체군의 크기를 1 로 제한하고 고정된 차원의 탐색 공간을 갖는 BEA 에 대해 수렴 특성을 살펴보았다. 먼저 BEA 가 단일 체인 MCMC 로 변환됨을 보여 수렴 결과를 얻었다. 온도에 대한 스케쥴링을 갖는 고정된 데이터 집합에 대하여 BEA 는 시뮬레이티드 어닐링과 같은 수렴 특성을 가진다. 또한 데이터가 계속 증가하면 온도를 감소시키는 것과 같은 효과가 있어 온도를 고려하지 않는 상태에서도 BEA 가 수렴함을 증명하였다.

본 논문에서는 개체군의 크기가 2 이상인 canonical BEA[2]의 수렴 특성에 대해 살펴본다. 개체군 자체를 하나의 상태로 하는 단일 체인의 베이지안 입자 필터로 변환하여 목표 확률분포에 수렴함을 증명한다.

본 논문의 구성은 다음과 같다. 2 장에서는 간략히 진화 알고리즘에서의 베이지안 틀(framework)과 canonical BEA에 대해 알아본다. 3 장에서는 입자 필터에 대해 살펴본다. 또한 canonical BEA 의 수렴 특성을 고려하는데 사용되어질 입자 필터의 수렴 특성을 알아본다. 4 장에서는 canonical BEA 의 입자 필터 버전인 베이지안 입자 필터를 제안하고 제한한 알고리즘이 목표 확률분포에 수렴함을 보인다. 5장에서는 결론과 향후 과제에 대해 언급한다.

2. 베이지안 진화 연산

진화 연산의 베이지안 확률 모델[3]에서는 개체들의 사후 확률분포를 평가한 후 그 분포로부터 자손들을 샘플링해 나간다. 탐색 공간 Θ의 각 개체 θ 에 대하여 $\pi(\theta)$ 를 사전 확률분포라 하고 $f(D|\theta)$ 를 데이터 D 에 대한 가능도 라고 할 때 각 개체 θ 의 적합도를 사후 확률분포 $\pi(\theta \mid D)$ 로 표시하고 다음과 같이 베이즈 정리를 이용하여 매 세대마다 평가해 나간다.

$$\pi(\theta \mid D) = \frac{f(D \mid \theta)\pi(\theta)}{f(D)} = \frac{f(D \mid \theta)\pi(\theta)}{\int f(D \mid \theta)\pi(\theta)d\theta}$$

$$\approx \frac{f(D \mid \theta)\pi(\theta)}{\sum_{\theta' \in \Theta} f(D \mid \theta')\pi(\theta')}$$

Canonical BEA의 기본 알고리즘은 다음과 같다.

알고리즘 2.1 (Canonical BEA)

1. **(초기화)** 초기 개체 $\Theta^0 = \{\theta_1^0, \ldots, \theta_M^0\}$ 를 사전 확률분포 $\pi_0(\theta)$ 로 부터 생성한다. 데이터의 크기 N_0 와 온도 T_0 를 초기화 한다. 세대 계수 $t \leftarrow 0$.
2. **(단계-D)** 크기 N_t 인 데이터 D^t 를 생성하고 가능도 $f(D^t \mid \theta_i^t)$ 를 계산한다.
3. **(단계-P)** 사후 확률분포 $\pi_t(\theta_i^t \mid D^t)$ 를 평가한다. 최적의 개체 θ_{best}^t 를 선정한다.
4. **(단계-V)** $\pi_t(\theta)$ 로 부터 샘플링하여 L 개의 변형 $\Theta' = \{\theta_1', \ldots, \theta_L'\}$ 를 생성한다.
5. **(단계-S)** 가능도 $f(D^t \mid \theta_i')$ 에 의해 Θ' 로부터 M 개의 개체들을 선택, $\Theta^{t+1} = \{\theta_1^{t+1}, \ldots, \theta_M^{t+1}\}$ 를 구성한다.
6. **(단계-R)** 사전 확률분포 $\pi_t(\theta)$ 를 재고(revise) 한다. 온도 T_t 를 갱신한다.
7. **(루프)** 만약 종료조건을 만나면 중단, 그렇지 않으면 단계 2로 간다. $t \leftarrow t+1$

알고리즘은 실제적으로 5 개의 단계로 이루어진다. 단계 D (데이터), P (사후 확률), V (변형), S (선택) , R (재고)이다. 단계 R, D, P 에서는 사전 확률 분포, 가능도, 사후 확률 분포를 각각 계산한다. 단계 V 와 단계 S 에서는 사후 확률 분포로부터의 샘플링을 구현한다.

3. 입자 필터(Particle Filter)

필터링(filtering)이란 시스템의 상태(state)를 평가하는 문제로 시스템의 상태를 샘플들을 사용한 확률 분포로 근사하는 샘플 기반의 알고리즘이라고 할 수 있다. 일반적인 입자 필터는 2 단계의 Markov 프로세스로 구성된다. 시간 t 에서 다음 상태로의 전이는 다음과 같이 이루어진다.

먼저 첫 단계에서 시간 t 에서의 입자 필터 $\Theta^t = (\theta_1^t, \cdots, \theta_N^t)$ 는 N 개의 입자들로 구성되며 각 입자 θ_i^t 에 가중치 w_i^t 를 할당한다. 할당된 가중치에 비례하여 각 입자는 자신을 복사하여 자손들을 생성하며 가중치 0을 가지는 입자는 소멸되기도 하여 새로운 입자 필터 $\overline{\Theta^t}$ 를 얻는다. 시간이 흐름에 따라 샘플들의 다양성이 퇴화되는 경향이 크므로 두 번째 단계에서 각 입자들은 시스템의 역학(dynamics)에 의한 Markov 전이 확률에 따라 서로 독립적으로 전이하여 다음 단계의 입자 필터 Θ^{t+1} 를 얻고 입자들의 가중치를 재조정한다. 입자 필터는 위와 같은 과정을 통해 한 상태의 입자들의 사후 확률 $p(\theta^t, D^t)$ 를 시간 t 에서의 데이터 D^t 를 관찰하며 근사한다. 이러한 입자들은 시간에 따라 진화해 나간다. 이러한 점에서 개체들의 사후 확률을 진화 시켜 나가는 canonical BEA 와 유사하다.

사후 확률 분포로부터 샘플링하여 얻은 가중치를 갖는 입자 필터의 샘플들로 적분을 이산 합으로 근사 할 수 있다. 즉 $\{\theta_i^t; i = 1, \ldots, N\}$ 를 사후 확률 분포로부터 얻은 임의의 샘플들이라 하고 $\{D^t\}$ 를 관측, $\delta(d\cdot)$ 는 Dirac delta 함수라고 할 때 사후 확률 분포는 다음과 같은 경험적 추정치(empirical estimate)에 의해 근사된다.

$$\hat{p}(\theta \mid D^t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\theta_i^t}(d\theta)$$

따라서 다음과 같은 기대값

$$E(g_t(\theta)) = \int g_t(\theta)\, p(\theta \mid D^t)\, d\theta$$

은 아래 식으로 근사 된다.

$$\overline{E(g_t(\theta))} = \frac{1}{N}\sum_{i=1}^{N} g_t(d\theta_i^t)$$

그러나 사후 확률 분포로부터 직접 샘플링 하는 것이 실제적으로 어려우므로 샘플링 하기가 용이한 중요도 제안(importance proposal) 함수 $q(\theta \mid D)$ 를 사용한다. 이러한 입자 필터의 수렴 특징에 대해서는 이론적으로 잘 증명되어 있다[4]. $B(\mathfrak{R}^n)$ 을 $\mathfrak{R}^n$ 상에서 정의된 상하한을 갖는 Borel −measurable 함수들의 공간이라 하고 $\|f\| \approx \sup_{x \in \mathfrak{R}^n} |f(x)|$ 라고 표시할 때, 정리 3.1 은 [4]의 Theorem1 의 직접적인 결과이다.

정리 3.1 [4 : Theorem1]
만약 중요도 가중치

$$w_t \propto \frac{p(D^t \mid \theta^t)\, p(\theta^t \mid \theta^{t-1})}{q(\theta^t \mid \theta^{0:t-1}, D^{1:t})}$$

가 어떠한 (θ^{t-1}, D^t) 에 대해서도 상한을 갖고, 중요도 가중치에 비례하여 자손들을 생성한다면, 모든 $t \geq 0$ 에 대하여 N 과는 독립적인 c_t 가 존재하여 임의의 $f_t \in B(\mathfrak{R}^{n_x \times (t+1)})$ 에 대해서도 다음이 성립한다.

$$E\left[\left(\frac{1}{N}\sum_{i=1}^{N} f_t(\theta_i^t) - \int f_t(\theta)\, p(d\theta \mid D)\right)^2\right]$$

$$\leq c_t \frac{\|f_t\|^2}{N}$$

4. 베이지안 입자 필터와 수렴 특성

먼저 canonical BEA 가 입자 필터로 변환됨을 보인 후 입자 필터의 수렴 성질을 이용하여 경험적 추정치가 목표 확률분포에 수렴함을 증명한다.

Canonical BEA와는 달리 온도 스케쥴링이나 데이터의 점진적인 증가는 고려하지 않는다. 제안하는 베이지안 입자 필터의 기본적인 알고리즘은 다음과 같다.

알고리즘 4.1 (베이지안 입자 필터)

1. **(초기화)** 초기 입자들 $\Theta^0 = \{\theta_1^0, \dots, \theta_M^0\}$ 를 사전 확률 분포 $\pi_0(\theta)$ 로 부터 샘플링한 임의의 M 개의 샘플들로 초기화하고 가중치를 $1/M$ 로 할당한다. 세대 계수 $t \leftarrow 0$.
2. **(단계-D)** 데이터 D^t 를 관찰하여 가능도 $f(D^t|\theta_i^t)$ 를 계산한다.
3. **(단계-P)** 적합도 사후 확률 분포 $\pi(\theta_i^t|D^t)$ 를 평가 한다.

4-1. **(단계-V1)** L 개의 새로운 변형 입자들 $\Theta' = \{\theta_1', \dots, \theta_L'\}$ 을 생성한다. 즉 Θ^t 의 입자 θ_i^t 는 m_i^t 개의 복사본을 생성한다. 여기서 $m_i^t = Lw_i^t$ 이고 가중치 w_i^t 는 다음과 같다.

$$w_i^t = \frac{\pi(\theta_i^t \mid D^t)}{\sum_{\theta_j^t \in \Theta^t} \pi(\theta_j^t \mid D^t)}$$

4-2. **(단계-V2)** 각 입자들은 시스템의 역학에 따른 Markov 전이 확률에 따라 서로 독립적으로 전이한다.

5. **(단계-S)** 새로운 입자 필터를 얻기 위해 M 개의 입자들을 선택한다.
6. **(단계-R)** 사전 확률분포를 재고한다. 새로운 입자들의 가중치를 조정한다.
7. **(루프)** $t \leftarrow t+1$, 단계 D 로 이동.

베이지안 입자 필터의 1, 2, 3 단계는 초기화, 가능도 및 사후 확률에 대한 계산을 포함한다. Canonical BEA의 단계 V 에서는 가우스 분포로 가정한 사후 확률 분포로 부터 샘플링을 하여 L 개의 변형 입자들을 생성한다. 반면 베이지안 입자 필터의 단계 V 는 2 단계로 구성된다. 먼저 단계 V1 에서 각 입자 θ_i^t 은 단계 P 에서 평가된 사후 확률에 의해 정의된 가중치에 비례하는 m_i^t 개 만큼 복사된다. 즉 적합도가 낮은 입자들의 소멸로 인한 자리를 적합도가 높은 입자들로 대치하는 것이다. 단계 V2 에서는 Markov 체인의 전이 확률에 따라 각 입자들은 전이한다. 단계 S 에서는 각 입자의 가능도에 근거하여 M 개의 입자들을 선택한다. 단계 R 에서는 단계 V1 에서의 가중치가 Markov 체인 전이 후의 가중치를 반영한 임의의 입자가 되도록 가중치를 재조정한다.

정리 4.1

알고리즘 4.1 을 통해 얻은 베이지안 입자 필터의 경험적 추정치는 입자의 수가 무한대에 가까이 갈 때 목표 확률 분포에 수렴한다.

(증명)

우리는 단계 P 에서 경험적 사후 확률 $\pi_t(\theta)$ 로 부터 직접 샘플링을 하였으므로 정리 3.1 의 제안함수는 다음과 같다.

$$q(\theta^t \mid \theta^{0:t-1}, D^{1:t}) = p(\theta^t \mid \theta^{0:t-1}, D^{1:t})$$

따라서 중요도 가중치의 분산을 최소화하며 중요도 가중치 w_t 는 상한을 가진다. 그러므로 정리 3.1 에 의해 독립적인 임의의 샘플들인 입자의 수가 무한대에 가까이 갈 때 입자들이 목표 확률분포에 수렴함을 알 수 있다.

5. 결론

본 논문에서는 베이지안 입자 필터를 제안하고 개체군의 크기가 1 이 아닌 canonical BEA 를 개체군 자체를 하나의 상태로 보는 단일 체인의 베이지안 입자 필터로 변환하여 개체군 기반의 canonical BEA 가 목표 확률분포에 수렴함을 보였다. 향 후 과제로 직접 사후 확률분포로부터 샘플링 하기가 어려운 경우, 중요도 비율 (importance ratio) 의 분산을 최소화 할 수 있는 제안 (proposal) 분포로부터 샘플링하고 또한 문제 상태의 차원을 감소 시켜 고차원 샘플링의 어려움을 해결하는 베이지안 진화 알고리즘으로 확대하려 한다.

감사의 글

본 연구는 과학 기술부 뇌 연구 개발사업(BR-2-1-G-06)에 의해 일부 지원되었음.

참고 문헌

[1] Gilks, W.R., Richardson, S., and Spiegelhalter, D.J., *Markov Chain Monte Carlo in Prac tice*, Chapman & Hall, 1996

[2] Byoung-Tak Zhang, Paass, G., and Muehlenbein, H., "Convergence Properties of Incremental Bayesian Evolutionary Algorithm with Single Markov Chains ", *Proc. Congress on Evolutionary Computation,* Special Session on Theory and F oundations of Evolutionary Computation, 2000

[3] Byoung-Tak Zhang. "A Bayesian Framework for Evolutionary Computation ", *Proc. Congress on Evolutionary Computation*, Special Session On Theory and Foundations of Evolutionary Computation, 1999

[4] Crisan, D. and Doucet, A. "Convergence of Generalized Particle Filters ", *Technical report CUED/F-INFENG/TR 381* , Cambridge University Engineering Department, 2000