

Greedy 선택방법을 적용한 빠른 모의 담금질 방법

이 충 열[†] · 이 선 영^{††} · 이 수 민^{†††} · 이 종 석^{††††} · 박 철 훈^{†††††}

요 약

모의 담금질 방법은 널리 사용되는 최적화 알고리즘들 중의 하나로서, 그 해의 수렴성이 수학적으로 증명되어 있는 장점이 있다. 하지만 원래의 모의 담금질 방법은 수렴 속도가 매우 느리기 때문에 복잡한 문제에 적용하기 힘들고, 이를 해결하기 위해서 빠른 모의 담금질 방법과 같은 다양한 방법이 연구되고 있다. 본 논문에서는, greedy 선택방법을 적용한 모의 담금질 방법을 제안하고, 이 알고리즘이 연속적인 공간에서의 최적화 문제에 대해서 전역 최적점을 찾아낸다는 것을 확률적으로 증명한다. greedy 선택방법은 무조건 좋은 해를 선택하기 때문에, 확률적으로 좋지 않은 해를 선택할 가능성이 있는 Metropolis 선택방법에 비해 빠른 수렴속도를 얻을 수 있다. 컴퓨터 모의 실험 결과, greedy 선택방법을 사용한 모의 담금질 방법이 기존의 빠른 모의 담금질 방법과 비슷한 성능을 보이는 해를 더 빠른 속도로 찾을 수 있음을 보인다. 또한, greedy 선택방법에서는 선택 가능한 상태들의 비용함수 값의 우열관계만을 이용하여 선택하기 때문에 비용 함수의 크기 조정에 무관하게 적용할 수 있다는 장점이 있다.

키워드 : 모의 담금질 방법, greedy 선택방법, 수렴성 증명

Fast Simulated Annealing with Greedy Selection

Chung-Yeol Lee[†] · Sun-Young Lee^{††} · Soomin Lee^{†††} · Jong-Seok Lee^{††††} · Cheol Hoon Park^{†††††}

ABSTRACT

Due to the mathematical convergence property, Simulated Annealing (SA) has been one of the most popular optimization algorithms. However, because of its problem of slow convergence in the practical use, many variations of SA like Fast SA (FSA) have been developed for faster convergence. In this paper, we propose and prove that Greedy SA (GSA) also finds the global optimum in probability in the continuous space optimization problems. Because the greedy selection does not allow the cost to become worse, GSA is expected to have faster convergence than the conventional FSA that uses Metropolis selection. In the computer simulation, the proposed method is shown to have as good performance as FSA with Metropolis selection in the viewpoints of the convergence speed and the quality of the found solution. Furthermore, the greedy selection does not concern the cost value itself but uses only dominance of the costs of solutions, which makes GSA invariant to the problem scaling.

Key Words : Simulated Annealing, Greedy Selection, Convergence Proof

1. 서 론

공학, 경제 등 다양한 영역에서 발생하는 최적화 문제는 선택 가능한 여러 대안들 중에서 만족도를 가장 높이는 해를 찾는 것을 목표로 한다. 주어진 최적화 문제들은 비용함수나 효용함수와 같은 수학적인 함수로 해의 적합도를 표현할 수 있고, 대부분의 최적화 알고리즘에서 최적화 문제는 선택에 따른 해의 비용함수로 정의된다.

연속적인 비용함수 f 와 다차원 실수 공간의 부분집합인 해의 공간 Ψ 가 정의되어 있을 때, Ψ 의 원소들 중 함수 f 를 최소화시키는 해 x^* 를 전역 최적점이라 한다. 최근 들어 목적함수 f 의 복잡도가 증가하고 지역 최적점이 늘어나면서 기존의 지역 최적점을 찾는 수치적인 알고리즘을 적용하기 힘들게 되었다. 이러한 문제점 때문에 지역 최적점에 빠지는 현상을 극복할 수 있는 확률적인 알고리즘들이 개발되고 있다. 이러한 확률적인 알고리즘들은 결과를 예측할 수 없다는 단점이 있지만, 제한된 자원을 이용하여 비교적 우수한 해를 찾아낸다. 또한 비용함수 값 외에 비용함수의 미분값과 같은 복잡한 부가적인 정보를 필요로 하지 않는다는 장점이 있다.

모의 담금질 방법은 다양한 최적화 문제에 있어서 강인하

※ 본 연구는 2007년도 한국과학기술원 BK21 정보기술사업에 의하여 지원되었음.

† 준 회원 : 한국과학기술원 전자전산학과 박사과정 (교신저자)

†† 준 회원 : 한국과학기술원 전자전산학과 박사과정

††† 준 회원 : University of Illinois, 전자전산학과 박사과정

†††† 정 회원 : 한국과학기술원 전자전산학부 연구연구원

††††† 정 회원 : 한국과학기술원 전자전산학부 교수

논문접수 : 2007년 4월 13일, 심사완료 : 2007년 9월 19일

1. 초기 온도 T_0 를 설정하고 초기 상태 벡터 x^0 를 생성한다.
2. 비용함수 f 로 x^0 에 대한 비용을 계산한다.
3. 생성함수 $g(\Delta x, T_k)$ 를 사용하여 x^{k-1} 상태에서부터 새로운 상태 \bar{x}^k 를 발생시킨다.
4. 비용함수 f 로 \bar{x}^k 에 대한 비용을 계산한다.
5. Metropolis 규칙에 의해 x^{k-1} 과 \bar{x}^k 중 하나를 다음 상태 x^k 로 선택한다.
6. 냉각 함수에 의해 온도 T 를 감소시킨다.
7. 원하는 종료 조건이 만족될 때까지, 3에서 6의 과정을 반복한다.

(그림 1) 일반적인 모의 담금질 방법의 개요

고 뛰어난 성능을 보이는 확률적인 알고리즘들 중의 하나로 알려져 있으며[1,2] 개략적인 과정은 (그림 1)에 나타나 있다. 모의 담금질은 그 해의 수렴성이 수학적으로 증명되어 있기 때문에 여러 최적화 문제에서 널리 사용되고 있다.

연속적인 상태의 최적화 문제에서 일반적인 모의 담금질 방법은 다음과 같은 볼츠만 담금질 기법을 사용한다[3].

$$g(\Delta x, T_k) = \frac{\exp[-\|\Delta x\|^2 / (2T_k)]}{(2\pi T_k)^{n/2}}, \tag{1}$$

$$h(\Delta E, T_k) = \min [1, \exp(-\Delta E / T_k)], \tag{2}$$

$$T_k = T_0 / \ln(k+1). \tag{3}$$

여기서 $g(\cdot)$ 는 주어진 상태에서부터 새로운 상태를 만들어 내기 위한 생성함수를 나타내고, $h(\cdot)$ 는 생성된 상태를 새로운 상태로 받아들일 확률을 결정하는 확률함수, 그리고 ΔE 는 현재 상태와 현재 상태에서부터 생성된 상태의 비용 함수 값의 차이를 나타낸다.

이 때, 식 (3)에 따른 담금질은 그 속도가 매우 느리기 때문에 실제 문제에 사용했을 때 비효율적이라는 문제가 있다. 이를 보완하기 위해 여러 가지 변화된 모의 담금질 방법이 개발되고 있으며, 그 중에서 널리 사용되고 있는 것이 Szu 와 Hartly에 의해 개발된 빠른 모의 담금질 방법(FSA: Fast Simulated Annealing)이다[3]. FSA에서는 다음과 같은 생성함수와 선택 확률함수, 냉각 함수를 사용한다.

$$g(\Delta x, T_k) = \frac{a_n T_k}{(\|\Delta x\|^2 + T_k^2)^{(n+1)/2}}, \tag{4}$$

$$h(\Delta E, T_k) = \min [1, \exp(-\Delta E / T_k)], \tag{5}$$

$$T_k = T_0 / k. \tag{6}$$

여기서 a_n 은 확률 밀도 함수 $g(\cdot)$ 을 정규화하기 위한 상수이다. 빠른 모의 담금질 방법은 여러 실제 문제들[4,5]에 적용되어 효과적인 성능을 보이고 있다.

기존에 제안된 일반적인 모의 담금질 방법과 빠른 모의 담금질 방법 모두 새로운 해를 선택하는 과정에서 Metropolis 기법을 사용하는데, 이는 언덕 오름(hill-climbing) 능력을 사용함으로써 해의 조성속 현상을 방지한다. 언덕 오름 능력은 비용 함수 최소화 문제에서 다음 상태의 비용 함수가 더 커지더라도 그 상태를 확률적으로 선택할 수 있도록 하는 것으로서 계속되는 언덕 오름을 통해 보다 쉽게 지역 최적점을 벗어나게 해주는 장점이 있다[6]. 하지만, 비용 함수가 증가하는 방향으로 이동하면서 상태가 수렴하기까지의 탐색 횟수가 증가하기 때문에 시간이 오래 걸리는 단점이 있다. 또한 Metropolis 방법에서는 온도가 다음 상태의 생성과 선택 모두에 영향을 미치기 때문에 알고리즘의 성능이 온도의 선택에 민감하다.

이러한 문제를 해결하기 위하여 지금까지의 연구는 다양한 선택 방법을 여러 가지 문제에 적용하여 실험 결과를 비교 분석함으로써 이루어졌다. 특히, greedy 선택 방법을 모의 담금질에 적용하는 것에 대한 비교 연구[7]에서 실험을 통하여 greedy 선택 방법을 사용하는 것이 기존의 Metropolis 선택방법을 사용하는 것과 큰 차이가 없음을 보였다. 하지만, 기존의 모의 담금질 방법이 갖는 장점 중에 수렴성이 매우 중요한 부분을 차지하고 있기 때문에 실험을 통한 비교 분석만으로 기존의 방법을 대체하지 못 하였다.

본 논문에서는 확률적으로 나쁜 해를 선택할 수 있는 Metropolis 선택방법 대신에, 항상 좋은 해만 선택하는 greedy 선택방법을 사용하는 모의 담금질 방법을 제안한다. 특히 제안하는 방법은 기존의 Metropolis 선택 방법에서와 마찬가지로 충분히 높은 초기 온도를 사용하여 넓은 영역을 탐색할 수 있도록 하면 언덕 오름 능력 없이도 전역 최적점으로 상태가 수렴하는 것을 증명하였다. 본 논문의 구성은 다음과 같다. 다음 장에서 제안하는 greedy 모의 담금질 방법(GSA: Greedy Simulated Annealing)을 설명하고 제안하는 방법의 전역 최적해로의 수렴성을 증명하였다. 3장에서는 컴퓨터를 사용한 모의 실험 결과 및 분석을 하였다. 마지막으로 4장에서 결론과 추후 연구 과제를 정리하도록 한다.

2. Greedy 모의 담금질 방법

일반적으로 greedy 알고리즘은 문제 해결 과정에서, 여러 가지 선택 가능한 해들 중에 항상 가장 좋은 해를 선택하는 알고리즘을 말한다[8]. 본 논문에서는 이러한 greedy 알고리즘을 빠른 모의 담금질 방법에서의 선택과정에 적용하여,

1. 초기 온도 T_0 를 설정하고 초기 상태 벡터 x^0 를 생성한다.
2. 비용함수 f 로 x^0 의 비용을 평가한다.
3. 변이함수 $g(\Delta x, T_k)$ 를 사용하여 x^{k-1} 상태에서부터 새로운 상태 \bar{x}^k 를 발생시킨다.
4. 비용함수 f 로 \bar{x}^k 의 비용을 평가한다.
5. $f(\bar{x}^k)$ 이 $f(x^{k-1})$ 보다 작거나 같으면 \bar{x}^k 를 다음 상태로 선택하고, 그렇지 않으면 x^{k-1} 를 다음 상태로 선택한다.
6. 냉각 함수를 이용하여 온도 T 를 감소시킨다.
7. 원하는 조건이 만족될 때까지, 3에서 6의 과정을 반복한다.

(그림 2) greedy 모의 담금질 방법의 개요

언덕 오름 능력을 사용하지 않고 매번 새로운 상태를 선택할 때마다 무조건 좋은 상태로 천이할 수 있도록 한다. 이 알고리즘을 greedy 모의 담금질 방법이라 정의하고, 개략적인 과정을 (그림 2)에 나타냈다.

GSA에서는 FSA와 같은 생성함수와 냉각함수를 사용하되, 선택 확률 함수에서 나쁜 해를 선택할 확률은 0이 된다. 이들을 식으로 표현하면 아래와 같다.

$$g(\Delta x, T_k) = \frac{a_n T_k}{(\|\Delta x\|^2 + T_k^2)^{(n+1)/2}}, \quad (7)$$

$$h(\Delta E, T_k) = \begin{cases} 1 & \text{if } \Delta E \leq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (8)$$

$$T_k = T_0 / k. \quad (9)$$

Metropolis 선택방법을 사용한 빠른 모의 담금질 방법은 최종적으로 탐색된 해가 전역 최적점에 수렴한다는 것이 수학적으로 증명되어 있다[9]. 이에 기반하여, 이 장에서는 greedy 선택방법을 적용한 GSA의 수렴성을 증명해 보고자 한다.

[정리 1] 전체 탐색 영역을 Ψ 라 할 때, 임의의 $\varepsilon > 0$ 에 대하여 $\Psi_\varepsilon = \{x \in \Psi \mid f(x) < f^* + \varepsilon\}$ 이라 하자.

$\mu(\cdot)$ 은 실수 집합 \mathbb{R}^n 에서의 Lebesgue 척도를 나타낸다고 했을 때, $\varepsilon > 0$ 에 대하여 $\mu(\Psi_\varepsilon) > 0$ 이 성립한다고 가정한다. 이 때, 상수 $M > 0$ 과 모든 $k > 0$ 에 대하여 온도에 기 반하는 상태 생성함수 $g(\cdot)$ 가

$$\min_{x, y \in \Psi} g(y - x, T_k) \geq M/k \quad (10)$$

을 만족시키고, 담금질 과정이 무한히 반복되면서 T_k 가 0

에 가까워지면, 탐색 가능한 집합 Ψ 내의 임의의 원소 x^0 를 초기 상태로 하였을 때, (그림 2)의 greedy 모의 담금질 방법에 의한 최종 상태는 전역 최적점 f^* 에 확률로써 수렴한다.

[증명] 0보다 큰 임의의 ε, δ 값과 탐색 가능한 영역 Ψ 에 속하는 임의의 초기 상태 x^0 , 그리고 양의 정수 k, m 에 대하여 아래 식 (11)를 만족시키는 양의 정수 K 가 존재함을 보인다.

$$P(X^k \in \Psi \setminus \Psi_\varepsilon) < \delta \quad \forall k \geq K \quad (11)$$

X^k 를 k 번째 상태벡터라 정의하고, X^{k+j-1} 에서 X^j , ($j=1, 2, \dots, m$)으로의 상태변화 중 적어도 한 개 이상이 Ψ_ε 으로부터 $\Psi \setminus \Psi_\varepsilon$ 로 이동하는 사건을 $G(k, m)$, $X^k, X^{k+1}, \dots, X^{k+m}$ 중 Ψ_ε 에 속한 상태가 하나도 없는 사건을 $I(k, m)$, 그리고 X^{k+m} 이 $\Psi \setminus \Psi_\varepsilon$ 에 속하는 사건을 $E(k, m)$ 라 하자. 사건의 정의에 의해,

$$E \subset G \cup I, \quad (12)$$

$$P(X^{k+m} \in \Psi \setminus \Psi_\varepsilon) = P(E) \leq P(G) + P(I) \quad (13)$$

가 성립한다.

먼저, greedy 선택방법에서는 나쁜 상태로의 천이가 불가능하므로,

$$P(G(k, m)) = 0 \quad (14)$$

와 같다. 이어서 우리는 식 (15)를 만족시키는 양의 정수 m_0 가 존재함을 보인다.

$$P(I(k, m)) < \delta, \quad \forall m \geq m_0 \quad (15)$$

X^k 와 이로부터 $g(\cdot, T_k)$ 를 거쳐 생성된 새로운 상태 벡터 Y^k 가 있을 때, 0보다 크거나 같은 모든 k 와 $\Psi \setminus \Psi_\varepsilon$ 에 속하는 임의의 x^k 에 대하여, 우리는 다음과 같은 식이 성립함을 보일 수 있다.

$$P(X^{k+1} \in \Psi \setminus \Psi_\varepsilon \mid X^k = x^k) = P(Y^k \in \Psi \setminus \Psi_\varepsilon \mid X^k = x^k) \quad (16)$$

그리고 Markov 속성에 따라, 아래의 식이 성립함을 쉽게 보일 수 있다.

$$\begin{aligned}
 & P(I(k, m)) \\
 &= P(X^{k_0} \in \Psi \setminus \Psi_\epsilon, X^{k_0+1} \in \Psi \setminus \Psi_\epsilon, \dots \\
 & \quad \dots, X^{k_0+m} \in \Psi \setminus \Psi_\epsilon | X^{k_0-1} = x^{k_0-1}) \\
 &= P(X^{k_0} \in \Psi \setminus \Psi_\epsilon | X^{k_0-1} = x^{k_0-1}) \times P(X^{k_0+1} \in \Psi \setminus \Psi_\epsilon | X^{k_0} = x^{k_0}) \times \\
 & \quad \dots \times P(X^{k_0+m} \in \Psi \setminus \Psi_\epsilon | X^{k_0+m-1} = x^{k_0+m-1}) \\
 &\leq \prod_{j=k_0+1}^{k_0+m} \max_{x^{j-1} \in \Psi \setminus \Psi_\epsilon} P(X^j \in \Psi \setminus \Psi_\epsilon | X^{j-1} = x^{j-1}) \\
 &= \prod_{j=k_0}^{k_0+m-1} \max_{x^j \in \Psi \setminus \Psi_\epsilon} P(X^{j+1} \in \Psi \setminus \Psi_\epsilon | X^j = x^j) \\
 &= \prod_{j=k_0}^{k_0+m-1} \max_{x^j \in \Psi \setminus \Psi_\epsilon} P(Y^j \in \Psi \setminus \Psi_\epsilon | X^j = x^j) \\
 &= \exp \left\{ \log \sum_{j=k_0}^{k_0+m-1} \max_{x^j \in \Psi \setminus \Psi_\epsilon} [1 - P(Y^j \in \Psi_\epsilon | X^j = x^j)] \right\} \\
 &= \exp \left\{ \sum_{j=k_0}^{k_0+m-1} \log \left[1 - \min_{x^j \in \Psi \setminus \Psi_\epsilon} P(Y^j \in \Psi_\epsilon | X^j = x^j) \right] \right\} \quad (17)
 \end{aligned}$$

이 때, $[0, 1]$ 사이에 존재하는 모든 y 에 대하여 $\log(1-y) \leq -y$ 이 성립함을 적용시키면, 위의 식은 아래와 같이 정리된다.

$$P(I(k, m)) \leq \exp \left\{ - \sum_{j=k_0}^{k_0+m-1} \min_{x^j \in \Psi \setminus \Psi_\epsilon} P(Y^j \in \Psi_\epsilon | X^j = x^j) \right\} \quad (18)$$

또한, $j > 0$ 에 대하여

$$\begin{aligned}
 \min_{x^j \in \Psi \setminus \Psi_\epsilon} P(Y^j \in \Psi_\epsilon | X^j = x^j) &\geq \min_{x^j \in \Psi \setminus \Psi_\epsilon} \int_{\Psi_\epsilon} g(y - x^j, T_j) dy \\
 &\geq \mu(\Psi_\epsilon) \min_{x^j \in \Psi \setminus \Psi_\epsilon, y \in \Psi_\epsilon} g(y - x, T_j) \geq \mu(\Psi_\epsilon) \min_{x, y \in \Psi} g(y - x, T_j) \\
 &\geq \mu(\Psi_\epsilon) M / j \quad (19)
 \end{aligned}$$

가 성립하므로,

$$\sum_{j=k}^{\infty} \min_{x^j \in \Psi \setminus \Psi_\epsilon} P(Y^j \in \Psi_\epsilon | X^j = x^j) \geq \mu(\Psi_\epsilon) M \sum_{j=k}^{\infty} 1/j = \infty \quad (20)$$

이다.

따라서 m_0 보다 크거나 같은 모든 m 에 대하여 아래 식이 성립하는 양의 정수 m_0 가 존재한다.

$$\sum_{j=k_0}^{k_0+m-1} \min_{x^j \in \Psi \setminus \Psi_\epsilon} P(Y^j \in \Psi_\epsilon | X^j = x^j) > \log(1/\delta) \quad (21)$$

따라서

$$P(I(k, m)) < \exp(-\log(1/\delta)) = \delta \quad (22)$$

를 얻어 식 (15)가 성립한다.

$K = k_0 + m_0$ 일 때, 식 (11)이 성립하므로, 식 (13), (14), (15)에 의해

$$P(X^{k_0+m} \in \Psi \setminus \Psi_\epsilon) \leq P(G(k_0, m)) + P(I(k_0, m)) < \delta \quad (23)$$

를 얻을 수 있다.

위의 정리는 알맞은 담금질 함수 T_k 와 상태 생성에 사용되는 확률 밀도 함수 $g(\cdot)$ 를 사용하면, greedy 선택방법을 사용한 모의 담금질 방법이 전역 최적점으로 수렴하는 성질이 보장된다는 것을 보여준다. 이어지는 따름정리는 GSA의 수렴성을 보장하기 위한 담금질 함수와 상태 생성함수의 형태를 보여 준다.

[따름정리 1] 정리 1에서의 모든 가정이 성립한다고 가정하고, r 은 각각의 원소가 $r_i = \max_{x, y \in \Psi} |x_i - y_i|$, $1 \leq i \leq n$ 을 만족시키는 n 차원 실수벡터라고 하자. 이 때, 온도에 기반한 상태 생성 확률 밀도 함수 $g(\cdot, T_k)$ 가 식 (7)와 같이 주어지고, 냉각 함수 T_k 가 식 (9)를 만족시키면, 탐색 가능한 영역에 속한 임의의 상태 x^0 는 전역 최적점에 확률적으로 수렴한다.

[증명] 따름정리를 증명하기 위해 우리는 식 (10)의 조건이 주어진 상태 생성 함수와 냉각 함수에 대해 성립하는지를 보인다. 식 (7), (9)를 이용하면 1보다 크거나 같은 모든 k 에 대하여,

$$\begin{aligned}
 \min_{x, y \in \Psi} g(y - x, T_k) &= \min_{x, y \in \Psi} \frac{a_n T_k}{[\sum_{i=1}^n (y_i - x_i)^2 + T_k^2]^{(n+1)/2}} \\
 &\geq \frac{a_n T_k}{(\sum_{i=1}^n r_i^2 + T_k^2)^{(n+1)/2}} \geq \frac{a_n T_k}{(\sum_{i=1}^n r_i^2 + T_0^2)^{(n+1)/2}} \\
 &= \frac{1}{k} \frac{a_n T_k}{(\sum_{i=1}^n r_i^2 + T_0^2)} \quad (24)
 \end{aligned}$$

가 성립하므로, 정리 1로부터 따름정리 1을 얻을 수 있다.

3. 실험 및 결과 분석

이 장에서는 GSA와 FSA를 매개변수 최적화 문제 8개와 외원원 순환문제에 대해 탐색된 해의 정확도와 수렴 속도 측면에서 비교한다.

3.1 매개변수 최적화 문제에 관한 실험결과 및 분석

본 논문에서 제안한 GSA의 성능을 평가하고 FSA의 성능과 비교하기 위해 먼저 잘 알려진 파라미터 최적화 문제

들 중 다음과 같은 8개의 검사함수를 사용하여 모의 실험을 수행하였다[10,11].

1) Sphere model

$$f_1(x) = \sum_{i=1}^n x_i^2, \quad -5.12 \leq x_i \leq 5.12 \text{ and } n = 100 \quad (25)$$

2) Generalized Rosenbrock's function

$$f_2(x) = \sum_{i=1}^{n-1} \{100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2\}, \quad -5.12 \leq x_i \leq 5.12 \text{ and } n = 100 \quad (26)$$

3) Step function

$$f_3(x) = 6n + \sum_{i=1}^n [x_i], \quad -5.12 \leq x_i \leq 5.12 \text{ and } n = 100 \quad (27)$$

4) Quadratic function with noise

$$f_4(x) = \sum_{i=1}^n nx_i^4 + rand(0,1), \quad -1.28 \leq x_i \leq 1.28 \text{ and } n = 100 \quad (28)$$

5) Shekel's function

$$f_5(x) = \frac{1}{K} + \sum_{i=1}^{25} \frac{1}{j + \sum_{i=1}^2 (x_i - a_{ij})^6}, \quad -65.356 \leq x_i \leq 65.356, K = 500 \text{ and } n = 2 \quad (29)$$

6) Generalized Rastrigin's function

$$f_6(x) = nA + \sum_{i=1}^n (x_i^2 - A \cos(wx_i)), \quad -5.12 \leq x_i \leq 5.12, A = 10, w = 2\pi \text{ and } n = 100 \quad (30)$$

7) Ackley's function

$$f_7(x) = -a \exp(-b \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}) - \exp(\frac{1}{n} \sum_{i=1}^n \cos(cx_i)) + a + e, \quad a = 20, b = 0.2, c = 2\pi, -32.768 \leq x_i \leq 32.768, \text{ and } n = 50 \quad (31)$$

8) Weighted sphere model

$$f_8(x) = \sum_{i=1}^n ix_i^2, \quad -5.12 \leq x_i \leq 5.12 \text{ and } n = 100 \quad (32)$$

Metropolis 선택방법을 사용한 빠른 모의 담금질 방법과 본 논문에서 제안한 greedy 선택방법을 적용한 모의 담금질 방법 모두 n-dimensional Cauchy 함수[12]를 사용하여 새로운 상태를 생성하고, 식 (9)로 온도를 담금질 하였다. 주어진 각 문제에 대하여 적절한 초기 온도 T_0 를 정하기 위해 W. Ben-Ameur가 제안한 방법을 이용하였다[13]. 이를 위해, 먼저 1000개의 상태를 임의로 생성한 후, 그들의 비용함수를 평가하고 Metropolis 선택방법에서 비용함수가 가장 낮은 상태에서 가장 높은 상태로 천이할 확률을 0.99로 하는 온도를 T_0 로 설정하였다. 모든 검사함수에 대해 담금질의 반복 횟수를 2천만번으로 제한하였다. 우리의 목표는 제

<표 1> 매개변수 최적화의 최종 비용 (괄호 안은 표준편차)

| function | f* | dim. | FSA | GSA |
|----------|-----|------|------------------|-------------------|
| $f_1(x)$ | 0 | 100 | 1.02e-2(4.28e-3) | 1.09e-5(3.89e-6) |
| $f_2(x)$ | 0 | 100 | 3.00e-3(1.17e-3) | 1.62e-2(7.72e-3) |
| $f_3(x)$ | 0 | 100 | 0.00e+0(0.00e+0) | 0.00e+0(0.00e+0) |
| $f_4(x)$ | 0 | 100 | 3.12e-2(5.31e-3) | 5.89e-4(2.63e-4) |
| $f_5(x)$ | ≈ 1 | 2 | 9.98e-1(1.62e-5) | 9.98e-1(2.22e-16) |
| $f_6(x)$ | 0 | 100 | 2.94e-2(4.47e-3) | 2.64e-5(4.80e-6) |
| $f_7(x)$ | 0 | 50 | 2.10e-1(4.33e-2) | 3.40e-2(3.29e-3) |
| $f_8(x)$ | 0 | 100 | 6.67e-2(1.00e-2) | 8.07e-2(2.87e-2) |

안한 알고리즘의 수렴성을 보이는 것이므로 초기 전이 확률을 0.99로 비교적 큰 값으로 하고, 담금질 또한 비교적 많은 횟수를 반복하도록 하였으며 이와 같은 변수 값은 여러 번의 실험을 거쳐 결정하였다.

실험 결과는 <표 1>에 정리되어 있으며, 비교를 위해 모든 검사함수들의 전역 최적값을 명시하였다. 각 함수마다 10번씩 실험을 반복하였으며, 최종적으로 찾은 해의 비용함수 값들의 평균과 표준편차를 <표 1>에 나타내었다. 공정한 비교를 위해 빠른 모의 담금질 방법과 greedy 모의 담금질 방법 모두 같은 랜덤 시드를 사용하여 같은 초기 상태에서 탐색을 시작하고 상태의 변이량도 같도록 해 주었다. 따라서 두 알고리즘은 오로지 새로 생성된 상태를 선택할 확률 값에서만 차이를 가진다.

8개의 함수 모두에 대해 두 알고리즘은 항상 전역 최적점에 성공적으로 도달하였으며, greedy 선택방법이 Metropolis 선택방법에 비해 조금 더 좋은 결과를 보였다. 2번 함수와 8번 함수를 제외하면, 비용함수 값들의 평균은 GSA가 더 작았으며, 이는 GSA를 이용하여 찾은 해들의 정확도가 FSA를 이용하여 찾은 해들의 정확도보다 높음을 의미한다. 또한 6번과 7번 함수의 경우, 그 문제가 복잡하여 지역 최적점이 많음에도 불구하고 GSA가 더 정확한 해를 찾았다는 것은 언덕 오름 능력 없이도 충분히 높은 초기 온도를 가지고 지역 최적점으로부터 벗어날 수 있음을 보여준다.

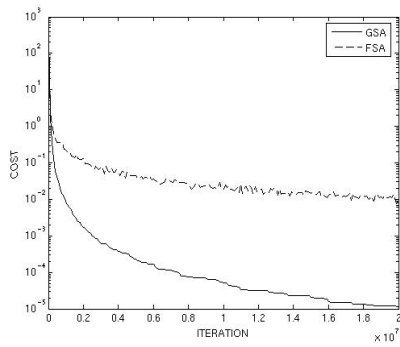
표준편차 역시 평균값과 비슷한 경향을 보인다. 즉, 함수 2와 8을 제외한 모든 함수에서 greedy 선택방법을 통해 찾은 해들의 표준편차가 Metropolis 선택방법을 통해 찾은 해들의 표준편차보다 작았는데, 이는 greedy 선택방법을 통해 찾은 해들이 더 균일하다는 것을 의미한다.

(그림 3)은 GSA와 FSA의 수렴 속도를 보여준다. 함수 3과 5를 제외한 나머지 함수들에 대해 비용함수 값의 변화를 잘 나타내기 위해 y축에 로그 스케일을 적용하였다. 함수 3의 경우, 두 알고리즘에서 모두 최종 비용함수 값이 0으로 수렴하여, 로그 스케일을 적용할 수 없었다. 함수 5는 문제의 복잡도가 낮아 다른 함수들에 비해 매우 빠르게 수렴하였기 때문에 초기 5만 번까지의 결과만을 나타내었다. 그리고 일반 스케일을 사용하였을 때, 비용함수 값의 변화가 더 잘 나타나 로그 스케일을 사용하지 않았다.

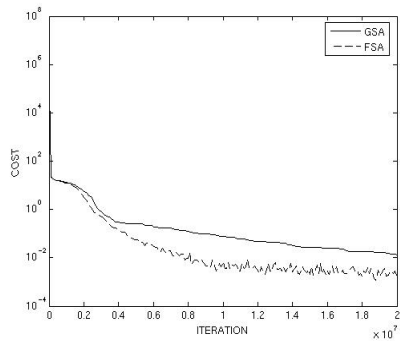
(그림 3)에서 볼 수 있듯이, 함수 1, 4, 6, 7의 경우 시작부터 끝까지 GSA가 FSA보다 빠르게 수렴하였고, 그 외의 함수에 대해서 GSA는 FSA와 거의 동일한 속도로 수렴하였다. 함수 1, 4, 6에서 FSA는 2천만 번까지 반복했음에도 불구하고

비용함수의 값이 10^{-1} 근처에 수렴한 반면, GSA는 500만 번 만에 이미 10^{-3} 이하의 값으로 수렴하였다. 마찬가지로 함수 7에서도 FSA는 2천만 번 반복 후에도 10^{-1} 보다 큰 비용함수 값에 수렴하였지만, GSA는 500만 번 반복했을 때 10^{-1} 보다 작은 값에 도달하여 계속해서 낮은 비용함수 값으로 수렴하는 경향을 보인다. 함수 2와 8에서는 전반적인 탐색시간 동안 FSA로 탐색한 해의 비용함수가 GSA로 탐색한 해의 비용함수에 비해 적은 값을 가지고 있으며 FSA가 좀 더 빠르게 수렴하는 경향을 보였다. 두 검사함수는 식(26)과 식(32)에서 볼 수 있듯이 해의 각각의 매개변수가 비용에 미치는 영향이 현저하게 차이가 난다. 기존의 FSA에서는 비용의 차이 크기가 선택에 영향을 미치고 비용 차이가 큰 쪽을 선호하는 경향이 나타나게 된다. 하지만 제안하는 GSA에서는 비용의 차이 크기에 관계없이 비용의 어느 쪽이 비용이 적은 지만을 이용하여 선택하기 때문에 기존의 방법보다는 상대적으로 느린 수렴을 보였다. 하지만 그 차이가 다른 검사함수에서 GSA가 보인 장점보다는 적은 것을 알 수 있다. 이로써 우리는 greedy 선택방법을 적용한 모의 담금질 방법이 해의 정확도와 수렴속도의 측면에서 대부분의 경우 기존의 빠른 모의 담금질 방법보다 성능이 좋다는 것을 알 수 있다.

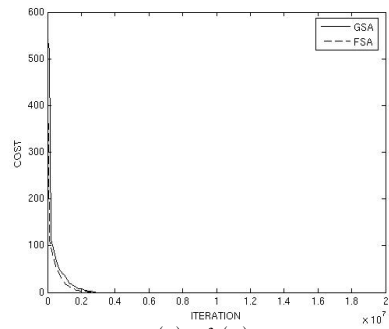
제안하는 GSA 알고리즘과 기존의 FSA 알고리즘의 시간 복잡도는 n번의 반복 횟수에 대해 각각 비용 함수 계산과 식(7)에서 (9) 또는 식(4)에서 (6)의 계산을 위한 상수 시간을 필요로 하므로, $O(n)$ 으로 동일하다. 하지만 기존의 FSA에서는 식(5)의 Metropolis 방법을 이용하여 전이 확률을 구하는 반면, 제안하는 GSA 방법에서는 식(8)과 같이 단순한 비교를 통하여 전이 여부를 판별하기 때문에 실제 수행 시간은 더 줄어든다.



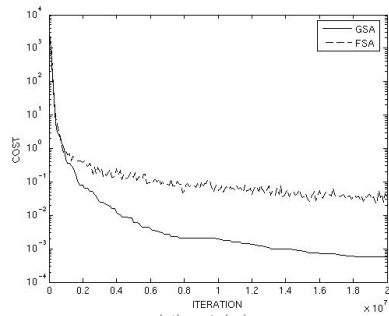
(a) $f_1(x)$



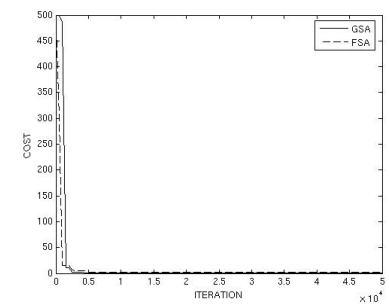
(b) $f_2(x)$



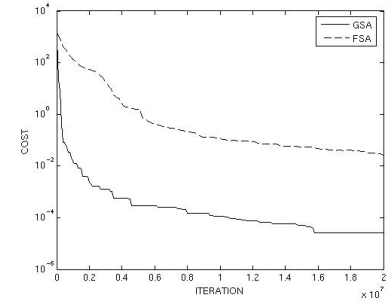
(c) $f_3(x)$



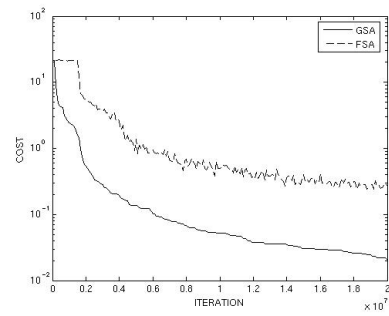
(d) $f_4(x)$



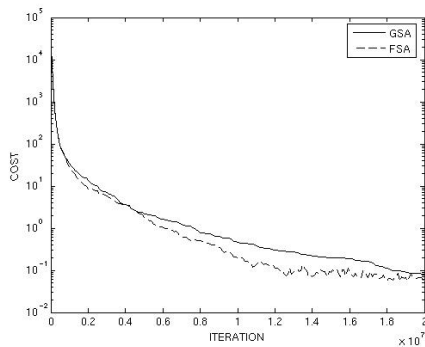
(e) $f_5(x)$



(f) $f_6(x)$



(g) $f_7(x)$



(h) $f_8(x)$

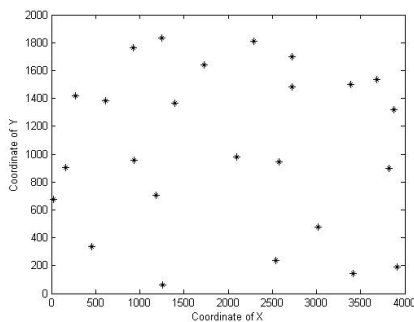
(그림 3) 매개변수 최적화 문제에 대한 GSA와 FSA의 수렴곡선

3.2 외관원 순환문제에 관한 실험결과 및 분석

외관원 순환문제 (TSP: Traveling Salesman Problem)는 유명한 조합 최적화 문제들 중의 하나로서, 주어진 도시들을 모두 순환하는 최단 경로를 찾아내는 것이 목표이다[14]. 이는 NP-hard 클래스에 속하는 문제로서 정확한 해를 찾는 것이 매우 어려운 것으로 알려져 있다. 본 논문에서는 다음과 같은 방법으로 TSP의 해를 찾으려 하였다.

각각의 상태 벡터는 방문해야 할 도시의 개수와 같은 개수의 원소들을 가지며, 각 원소들은 0에서 1사이의 값을 가지고, 이들의 값이 방문할 도시들의 순서를 결정한다. 예를 들어, 순회해야 할 도시가 3개인 경우 각각의 상태 벡터는 3개의 원소를 가지게 된다. 만약 어떠한 상태 벡터가 (0.15, 0.83, 0.72)의 값을 가지면, 각각의 원소들을 오름차순으로 정렬한 순서 “1-3-2-1”이 도시의 방문 경로가 된다.

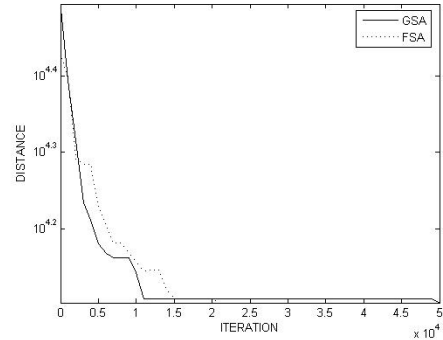
25개의 도시[10]를 방문하는 문제에 대해 모의 실험을 수행하였으며, 도시들의 좌표는 (그림 4)에 나타나 있다. 모의 실험 수행시, GSA와 FSA의 초기 온도는 50도로 하였으며, 담금질 횟수는 100만 번으로 제한하였다. 이 값은 매개변수 최적화의 경우와 마찬가지로 실험을 통해 얻어진 것이다. <표 2>는 주어진 25개 도시를 최단 경로로 순환할 때의 거리, 즉 전역 최적점을 포함하여, 100번의 실험 결과를 통해 찾은 최종 해들의 평균값과 표준편차를 보이고 있다. FSA의 경우는 100번 중 28번, GSA의 경우는 100번 중 25번 정확한 최단 경로를 찾아냈다. FSA를 통해 찾은 경로의 평균 거리가 GSA에 비해 조금 더 짧았지만 표준편차 측면에서 FSA는 탐색된 해들 간의 편차가 크게 나타났다.



(그림 4) TSP 문제에서 25개 도시의 좌표

<표 2> TSP 문제의 최종 비용 (괄호 안은 표준편차)

| f^* | dim. | FSA | GSA |
|------------------|------|-----------------|-----------------|
| $\approx 12,674$ | 25 | 13,480(9,264.3) | 13,708(1,167.2) |



(그림 5) TSP에 대한 GSA와 FSA의 수렴곡선

(그림 5)는 TSP 문제에 대한 FSA와 GSA의 일반적인 수렴 속도를 보여주고 있다. 비용함수 값의 변화를 잘 나타내기 위해 x축, y축에 로그 스케일을 사용하였다. Metropolis 선택방법을 사용하면 해가 수렴하는 데까지 평균적으로 12만 번의 반복횟수가 필요한 데 반해 greedy 선택방법을 사용하면 평균적으로 10만 번의 반복횟수가 필요하였다. 최종 해의 결과와 수렴 속도를 비교했을 때, GSA 방법은 기존의 FSA 방법보다 보다 안정적으로 비슷한 수준의 해를 빨리 찾는다는 것을 알 수 있다.

4. 결론 및 추후 과제

본 논문에서는 빠른 모의 담금질 방법에서 언덕 오름 능력을 사용하는 기존의 Metropolis 선택방법 대신에 greedy 선택방법을 사용하는 방법을 제안하였다. 그리고 새로운 상태를 발생시키는 함수와 담금질 방법을 올바르게 선택하면 greedy 선택방법을 적용한 빠른 모의 담금질 방법이 전역 최적점에 수렴한다는 것을 수학적으로 증명함으로써, GSA의 활용 가능성을 보였다. 이러한 greedy 선택방법은 선택 가능한 상태들의 비용함수 값의 우열관계만을 이용하여 선택하기 때문에, 비용함수 값의 크기에 무관하다는 장점이 있다. 컴퓨터 모의 실험 결과 매개변수 최적화 문제나 외관원 순환문제에 대해 GSA는 기존의 FSA보다 좀 더 빠른 수렴 속도를 보이면서도 비슷한 성능의 해를 안정적으로 찾았다. 즉, greedy 선택방법을 사용함으로써 지역 최적점에 빠지지 않으면서 빠른 시간 안에 주어진 문제에 적합한 해를 찾을 수 있다.

참고 문헌

[1] P. J. M. Laarhoven and E. H. L. Aarts, ‘Simulated Annealing: Theory and Applications’, Kluwer Academic Publishers, 1987.
 [2] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization

by Simulated Annealing," Science, Vol.220, pp.671-680, 1983.

[3] H. H. Szu and R. L. Hartley, "Fast Simulated Annealing," Physics Letters A, Vol.122, pp.157-162, 1987.

[4] H. H. Szu and R. L. Hartley, "Nonconvex optimization by fast simulated annealing," Proceedings of the IEEE, Vol 75, No.11, Nov. 1987.

[5] T.-C. Chen and Y.-W. Chang, "Modern floorplanning based on fast simulated annealing," Proceedings of the ACM International Symposium on Physical Design, pp.104-112, San Francisco, California, Apr. 2005.

[6] S. A. Kravitz and R. A. Rutenbar, "Placement by Simulated Annealing on a Multiprocessor," IEEE trans. on computer-aided design of Integrated Circuits and Systems, Vol.6, No.4, pp.534-549, 1987.

[7] P. Moscato and J.F. Fontanari, "Stochastic versus Deterministic Update in Simulated Annealing," Physics Letters A, Vol.146, No.4, pp.204-208, 1990.

[8] http://en.wikipedia.org/wiki/Greedy_algorithm.

[9] R. L. Yang, "Convergence of the simulated annealing algorithm for continuous global optimization," Journal of Optimization Theory and Applications, Vol.104, pp.691-716, 2000.

[10] K. D. Jong, 'An analysis of the behaviour of a class of genetic adaptive systems', Ph.D. Dissertation, University of Michigan, 1975.

[11] <http://www.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/genetic/ga/systems/genesys/0.html>

[12] D. Nam, J.-S. Lee and C. H. Park, "N-dimensional Cauchy neighbor generation for the fast simulated annealing," IEICE Trans. on Information and Systems, Vol.E87-D, No.11, pp.2499-2502, Nov. 2004.

[13] W. Ben-Ameur, "Computing the Initial Temperature of Simulated Annealing," Computational Optimization and Applications, Vol.29, No.3, pp.369-385, 2004.

[14] D. Mitra, F. Romeo and A. L. Sangiovanni-Vincentelli, "Convergence and finite-time behavior of simulated annealing," Advances in Applied Probability, Vol.18, No.3, pp.747-771, 1986.



이 충 열

e-mail : chungyeol.lee@gmail.com
 2000년 한국과학기술원 전기및전자공학과 학사
 2003년 한국과학기술원 전자전산학과 공학석사
 2003년~현재 한국과학기술원 전자전산학과 박사과정

관심분야: 신경회로망, 최적화



이 선 영

e-mail : lsy1116@kaist.ac.kr
 2005년 이화여자대학교 정보통신학과 학사
 2006년 한국과학기술원 전자전산학과 공학석사
 2006년~현재 한국과학기술원 전자전산학과

박사과정
 관심분야: 최적화, 센서 네트워크



이 수 민

e-mail : soom@kaist.ac.kr
 2003년 경북대학교 전자전산학과 학사
 2005년 한국과학기술원 전자전산학과 공학석사
 2005년~현재 University of Illinois, 전자전산학과 박사과정

관심분야: 최적화



이 종 석

e-mail : jslee@nnmi.kaist.ac.kr
 1999년 한국과학기술원 전기및전자공학과 학사
 2001년 한국과학기술원 전자전산학과 공학석사
 2006년 한국과학기술원 전자전산학과 공학박사

2006년~현재 한국과학기술원 전자전산학부 연수연구원
 관심분야: 시청각 음성인식, 멀티모달 인터페이스



박 철 훈

e-mail : chpark@kaist.ac.kr
 1984년 서울대학교 전자공학과 학사
 1985년 Caltech 전자공학과 공학석사
 1990년 Caltech 전자공학과 공학박사
 1991년~현재 한국과학기술원 전자전산학부 교수

관심분야: 지능시스템, 신경회로망, 최적화, 지능제어