

대화형 유전자 알고리즘의 성능향상을 위한 지식기반 인코딩

김희수, 조성배
연세대학교 컴퓨터과학과

A Knowledge-based Encoding for Performance Improvement of Interactive Genetic Algorithm

Hee-Su Kim and Sung-Bae Cho
Computer Science Department, Yonsei University

요 약

진화 연산은 최적화 및 분류 작업을 필요로 하는 대부분의 응용 분야에서 매우 효율적인 해결 방법을 제시해 주지만, 예술이나 감성 등의 특정 분야에의 적용에 있어서는 그 한계를 드러낸다. 이를 극복하기 위해서 여러 가지 기술들이 제안되었으며, 이 중에서 특히 대화형 유전자 알고리즘이 오늘날 널리 연구되고 있다. 대화형 유전자 알고리즘은 상호작용을 통하여 사용자의 평가치를 개체의 적합도로 받아들이고, 이를 기반으로 집단을 진화시키는 방법이다. 본 논문에서는 이를 의상디자인 지원 시스템에 적용시킴으로써 일반적으로 나타내기 어려운 사용자의 선호도나 감성을 디자인 과정에 반영할 수 있었다. 또한, 이론에 기반한 분석 및 실험적인 결과를 통해, 제안된 인코딩 방법이 유용함을 알 수 있었다.

1. 서론

진화연산은 적자생존, 유전자 교차 등의 자연 진화 메카니즘에 기반한 문제해결 알고리즘으로 기계학습, 최적화 및 분류 작업에 매우 효율적인 해결 방법을 제시해 준다. 이 방법은 다수의 개체들을 문제에 맞게 인코딩한 후 세대를 거듭하면서 이들을 진화시킴으로써 해결구하는 방법으로, 최근 인공지능의 주된 연구 분야 중 하나이다. 그러나 진화연산은 예술이나 감성적인 문제 등에는 적합도 함수를 도출하기 어렵다는 점 때문에 적용이 힘들다는 한계점을 지니고 있다. 이를 해결하기 위해 많은 연구가 진행되고 있으며, 그 중 하나가 사용자의 평가를 적합도 함수로 사용함으로써 그의 선호도나 감성을 진화에 반영시키는 대화형 유전자 알고리즘이다.

대화형 유전자 알고리즘의 응용분야로서 이전 연구에서 의상디자인 지원 시스템을 제안하고 평가한 바 있다[3]. 이 시스템은 사용자의 선호도를 진화에 반영하여 디자인 분야의 전문가가 아니더라도 큰 탐색공간에서 자신이 원하는 디자인을 효율적으로 찾을 수 있게 해 준다. 본 논문에서는 제안된 시스템의 인코딩 방법에 초점을 맞추고, 스키마 정리 등의 이론적인 분석과 실험을 통해 그 유용성을 보이고자 한다.

2. 대화형 유전자 알고리즘

유전자 알고리즘은 John Holland에 의해 1970년대에 제안된 알고리즘으로 다수의 개체들을 교차, 돌연변이, 적자 생존 등의 메카니즘을 사용하여 진화시키는 방법이며, 최적화 및 분류 문제에 매우 효율적으로 적용되고 있다[2].

대화형 유전자 알고리즘은 적합도 함수 대신에 사용자와의 상호작용으로부터 개체의 적합도를 얻는다는 점에서 일반적인

유전자 알고리즘과 구별된다. 이를 통해 사용자의 선호도나 취향 등이 진화과정에 반영되며, 따라서 유전자 알고리즘이 해결하기 어려운 예술, 디자인 등의 문제에도 적용될 수 있다[6].

의상 디자인 지원 시스템의 목적은 사용자가 선호하는 디자인을 찾는 것이며, 이를 위해 적절한 개체를 인코딩하고, '좋은 디자인'이라는 적합도를 극대화시키는 방법으로 유전자 알고리즘을 적용시켜 볼 수 있다. 그러나 좋은 디자인이라는 척도를 수학적인 함수로 나타내는 것은 불가능하므로, 일반적인 유전자 알고리즘 대신 대화형 유전자 알고리즘을 사용하는 것이 바람직하다[6].

3. 지식기반 인코딩

3.1 인코딩 방법

진화연산을 이용한 기존의 디자인 지원 시스템은 의상디자인 분야의 지식을 고려하지 않았기 때문에 실제 의상 디자인과는 동떨어진 결과를 도출하는 경우가 많았다[5]. 이를 해결하기 위해 본 연구에서 제안한 시스템에서는 의상디자인 분야의 지식에 기반한 디테일 모델을 사용하였다[4]. 이들을 특성에 따라 목과 몸통, 팔과 소매, 허리선 및 스커트의 세 부분으로 분류하고 각각을 비트스트링으로 인코딩하였다. 여기에 각 부위의 색상을 결정하는 3비트씩을 더하여 23비트의 유전자형을 결정하였다. 그림 1은 제안된 유전자형 및 인코딩된 디테일 요소들을 보여주고 있으며, 이 유전자형에서 얻을 수 있는 디자인 조합의 수, 즉 탐색공간 크기는 $34 \times 8 \times 12 \times 8 \times 9 \times 8 = 1,880,064$ 와 같이 계산될 수 있다. 디자인 개체들은 이 유전자형에 따라 각 디자인 요소들을 조합함으로써 만들어진다. 그림 2는 이 유전자형에서 디자인 개체가 만들어지는 예이다.

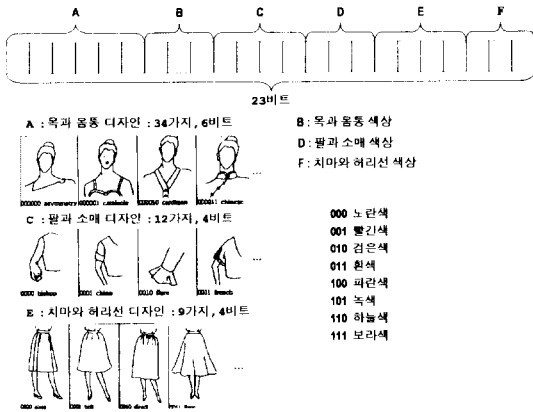


그림 1. 유전자 인코딩

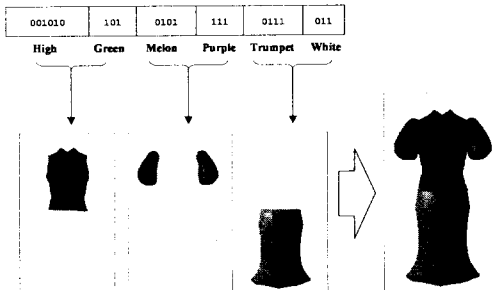


그림 2 유전자형으로부터 개체가 조합되는 예

3.2 이론적 배경

John Holland는 유전자 알고리즘에서 각 개체들이 어떻게 진화되는지를 설명하기 위해 스키마라는 개념을 소개하였다. 그는 이를 이용하여 적합한 개체들간의 유사성이 어떻게 유전자 알고리즘을 이용한 탐색의 방향을 결정하는지 설명하였다. 그에 따르면, $t+1$ 번째 세대의 스키마 H 의 발현수 $m(H, t+1)$ 은 $m(H, t)$ 의 향으로 나타내어질 수 있다[2]. 재생산(비례 선택 및 세대 교체), 재결합(1점 교차), 돌연변이 등의 고전적 유전자 연산자를 고려하면 $t+1$ 번째 세대의 스키마 H 의 발현수는 다음과 같다.

$$\begin{aligned}
 m(H, t+1) &\geq m(H, t) \cdot \frac{f(H)}{\bar{f}} \left[1 - p_c \cdot \frac{\delta(H)}{l-1} \right] [1 - p_m]^{o(H)} \\
 &\cong m(H, t+1) \geq m(H, t) \cdot \frac{f(H)}{\bar{f}} \left[1 - p_c \cdot \frac{\delta(H)}{l-1} - o(H) \cdot p_m \right] \quad (1)
 \end{aligned}$$

- l : 스트링의 길이
- $\delta(H)$: 스키마 길이, 즉 스키마 H 의 첫 번째와 마지막 고정값 사이의 거리
- $o(H)$: 스키마 H 의 차수, 즉 스키마 H 의 고정값의 수
- $f(H)$: 스키마 평균, 세대 t 에서 스키마 H 에 속하는 스트링들의 적합도 평균
- \bar{f} : 집단 평균, 집단 전체의 적합도의 평균
- p_c : 교차 확률, 교차 연산이 발생할 확률
- p_m : 돌연변이 확률, 돌연변이 연산이 일어날 확률

(1)로부터, 길이가 짧고 낮은 차수와 평균 이상의 적합도를 갖는 스키마들은 세대를 거듭할수록 그 발현수가 지속적으로 증가한다는 결론을 얻을 수 있다. 이러한 스키마를 빌딩 블록이라고 한다. 스키마 H 가 교차 연산에 의해 파괴될 확률은

$$p_c \frac{\delta(H)}{l-1}$$

와 같다. 하나의 유전자형을 겹치지 않게 분할하는 n 개의 빌딩 블록들을 가정하면 일반적인 1점 교차 후에 생존할 빌딩 블록의 수는

$$\begin{aligned}
 &((n-1) \sum_{k=1}^n p_c \cdot \frac{\delta(H_k)}{l-1} + n \cdot p_c \cdot \frac{n-1}{l-1}) + n(1-p_c) \\
 &= p_c \cdot \frac{n-1}{l-1} (\sum_{k=1}^n \delta(H_k) + n) + n(1-p_c)
 \end{aligned}$$

과 같다. 가정에서 모든 스키마 길이의 합은 $(l-n)$ 과 같으므로, 위 식은 다음과 같이 쓸 수 있다.

$$\begin{aligned}
 &p_c \cdot \frac{n-1}{l-1} \cdot l + n(1-p_c) \\
 &= p_c \cdot \frac{(n-1) \cdot l - n \cdot (l-1) + n}{l-1} + n \\
 &= p_c \cdot \frac{nl - l - nl + n + n}{l-1} + n \\
 &= p_c \cdot \frac{n-1}{l-1} + n \quad (2) \\
 &(\textcircled{*} \sum_{k=1}^n \delta(H_k) = (l-n))
 \end{aligned}$$

한편, 스키마 H 가 돌연변이 연산 후에 생존할 확률은 (1)로부터 $(1 - p_m)^{o(H)}$ 으로 쓸 수 있다. 따라서, 스키마 H 가 돌연변이 연산에 의해 파괴될 확률은 $1 - (1 - p_m)^{o(H)}$ 이며, 이로부터 돌연변이 연산에 의해 파괴될 빌딩 블록의 수를 계산해 보면 다음 식을 얻는다.

$$\begin{aligned}
 &\sum_{k=1}^n 1 \cdot (1 - (1 - p_m)^{o(H_k)}) \\
 &= \sum_{k=1}^n (1 - (1 - p_m)^{o(H_k)}) \\
 &= n - \sum_{k=1}^n (1 - p_m)^{o(H_k)} \quad (3)
 \end{aligned}$$

(2), (3)에서 교차 및 돌연변이 연산 이후 생존할 빌딩 블록의 수는

$$\begin{aligned}
 &(p_c \cdot \frac{n-1}{l-1} + n) - (n - \sum_{k=1}^n (1 - p_m)^{o(H_k)}) \\
 &= p_c \cdot \frac{n-1}{l-1} + n - n + \sum_{k=1}^n (1 - p_m)^{o(H_k)} \\
 &= p_c \cdot \frac{n-1}{l-1} + \sum_{k=1}^n (1 - p_m)^{o(H_k)} \quad (4)
 \end{aligned}$$

과 같다. 마지막으로, 적합도 값에 비례하여 선택되는 재생산을 고려하면 $f(I)$ 를 개체 I 에 대한 적합도라고 할 때, (4)는

$$\frac{f(I)}{\bar{f}} (p_c \cdot \frac{n-1}{l-1} + \sum_{k=1}^n (1 - p_m)^{o(H_k)})$$

처럼 쓰여질 수 있다. 본 논문에서 제안된 인코딩 방법은 가까운 디테일 요소들을 3부분으로 재분류하여, 이러한 정보를 각각의 짧은 블록에 인코딩하였으므로, 진화과정에서 위에서 언급한 장점을 취할 수 있다. 다음 장에서 실험에 의해 이를 좀더 명확하게 확인할 수 있다.

4. 실험 및 결과 분석

집단은 8개의 개체로 구성되어 있으며 0.5 확률의 1점 교차와 0.05 확률의 돌연변이 연산, 각 세대의 최적할 개체를 다음 세대까지 그대로 보존하는 전략을 적용하였다. 대화형 유전자 알고리즘은 사용자의 평가에 기반하여 진화의 방향이 결정되므로, 수렴성을 정량적으로 보이는 표준화된 방법은 아직 없다. 우선, 제안된 시스템과 동일한 인터페이스에 인코딩 방법만을 다르게 하여 원래의 시스템과 비교하여 보았다.

이 시스템에는 가능한 디자인 조합과 유전자형을 의미없이 순서대로 매핑시키는 순차 인코딩을 사용하였으며, 20대 남, 여 피험자 10명에게 두 시스템을 사용하여 시원한 느낌의 디자인을 만들 것을 요구하였다. 사용자가 직접 개체에 대한 평가를 내리야 한다는 대화형 진화연산의 근본적인 한계로 인하여 세대 수는 10번으로 제한하였다 [3]. 그림 3은 각 시스템을 사용한 경우의 평균적합도 변화를 나타낸 것이다. 지식기반 인코딩 시스템의 적합도는 꾸준히 오르고 있는데 반해 순차 인코딩 시스템은 세대를 거듭하여도 적합도가 거의 변하지 않음을 알 수 있다.

보다 구체적으로 두 시스템에 대한 만족도를 비교하기 위해 각각의 마지막 세대에서 최적할 개체들을 선별하여 10명의 피험자들에게 비교하도록 하였다[1]. 그림 4는 이 상대적 만족도를 -3에서 3까지의 구간에서 보인 결과로, 사용자들이 대체적으로 지식기반 인코딩 방법의 결과를 더 선호하였음을 알 수 있다.

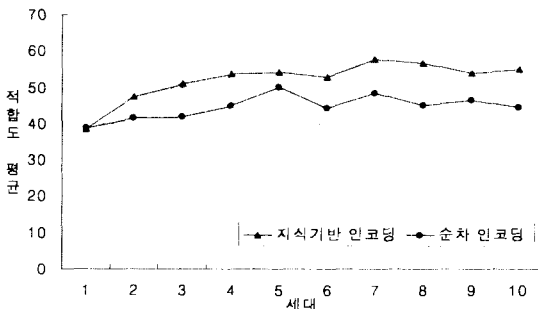


그림 3. 인코딩 방식에 따른 세대별 적합도 변화

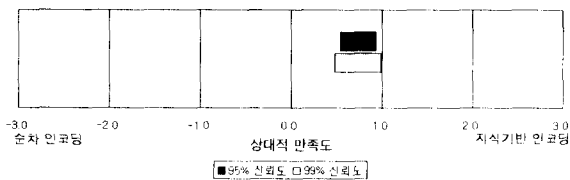


그림 4. 인코딩 방법에 따른 상대적 만족도

선호되는 스키마 수가 증가하고 있음을 보이기 위해서 시원한 디자인을 찾는 탐색 하나를 무작위로 선택하여 마지막 세대의 최적할 개체를 근사해라고 가정하고, 여기에 포함된 빌딩 블록들의 빈도를 각 세대별로 분석하였다. 선택된 근사해는 흰색 slit 몸통 디자인, 푸른색 scooter 스커트 디자인과 sleeveless 디자인의 조합이었으며, 그림 5는 근사해를 이루는 스키마들의 집단 내 빈도가 세대를 거듭하

면서 독립적으로 증가하고 있음을 보여준다. 중요한 점은 이들의 초기 빈도가 0에 가깝다는 사실로, 세대수가 10으로 제한되어 있었음에도 불구하고 진화 과정이 초기 집단에 크게 구애받지 않고 세대를 거듭하면서 근사해를 탐색하게 됨을 알 수 있다.

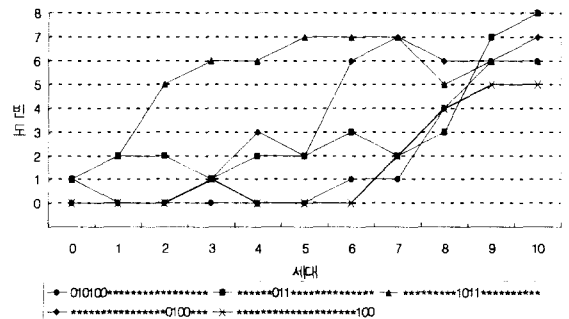


그림 5. 탐색과정에서의 스키마 해의 빈도 변화

5. 결론

본 논문에서는 대화형 유전자 알고리즘을 응용한 의상디자인 지원 시스템의 성능을 개선하기 위한 지식기반 인코딩 방법을 제안하였다. 실험 결과 세대를 거듭하면서 집단 내에서 해에 가까운 스키마의 출현 빈도가 높아진다는 것과 지식기반 인코딩 방법이 순차 인코딩 방법보다 더 만족스러운 결과를 도출해 내었음을 알 수 있었다. 반면, 집단의 크기가 작기 때문에 일어날 수 있는 해의 풀림은 문제점이라 할 수 있으며, 현재 이를 보완하기 위한 연구가 진행 중이다.

참고 문헌

- [1] David, H.A., *The Method of Paired Comparison*, Charles Griffin and Co. Ltd., 1969.
- [2] Goldberg, D.E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Publishing Co. Inc., 1989.
- [3] Kim, H.-S. and Cho, S.-B., "Development of an IGA-based fashion design aid system with domain specific knowledge," *Proc. of IEEE Int. Conf. on Systems, Man, and Cybernetics*, Interactive Evolutionary Computation session, 1999.
- [4] McKelvey, K., *Fashion Source Book*, Blackwell Science, 1996.
- [5] Nakanishi, Y., "Capturing preference into a function using interactions with a manual evolutionary design aid system," *Genetic Programming 1996 Late-Breaking Papers*, pp. 133~138, 1996.
- [6] Takagi, H., "Interactive evolutionary computation: Cooperation of computational intelligence and human KANSEI," *Proc. of Int. Conf. on Soft Computing*, pp. 41~50, 1998.