

새로운 유전자 표현 기반 유전 알고리즘을 이용한 OCST 문제에 대한 해법

김종율^{*} · 도경훈^{*} · 정완영^{*}

^{*}부산광역시 사상구 주례2동 산69-1 동서대학교 컴퓨터정보공학부

The Solution of OCST Problems using a New Genetic Representation-based Genetic Algorithm

Jong Ryul Kim^{*} · Kyeong-Hoon Do^{*} · Wan Young Chung^{*}

^{*} Division of Computer and Information Engineering, Dongseo University

E-mail : {xmaskjr khdo}@gdsu.dongseo.ac.kr, wychung@dongseo.ac.kr

요 약

최근 늘어가는 통신 서비스에 대한 관심은 관련 이론 및 기술의 발전을 주도해 왔다. 이러한 통신 시스템에 관련한 문제들 중에 가장 근간이 되는 문제 중 하나는 최적 통신 걸침 나무 (OCST: Optimal Communication Spanning Tree) 문제이다. 본 논문에서는 이러한 최적 통신 걸침 나무 문제에 대한 해법으로 최적 통신 걸침 나무 문제에 적합한 새로운 유전자 표현법을 이용한 유전 알고리즘을 설계하여 구현한다. 제안하는 유전 알고리즘은 n 개의 노드에 대해 $n-2$ 개의 숫자 열로 표현가능하다고 알려진 기존의 Prüfer수 기반의 유전자 표현법을 개선하여 보다 더 최적 통신 걸침 문제에 적합하도록 고안한 새로운 유전자 표현법을 이용한다. 임의로 생성된 예제에 대한 수치 실험을 통해 통신시스템의 기본 문제 중 하나인 최적 통신 걸침 문제의 해법으로서의 제안 알고리즘의 유용성과 효율성을 확인한다.

키워드

OCST (Optimal Communication Spanning Tree), Genetic Algorithm (GA), Prüfer Number (PN)

1. 서 론

인터넷의 출현은 인류에게 많은 도전과 많은 삶의 변화를 촉발하고 있다. 네트워크의 시스템을 위한 고비용의 하드웨어 및 소프트웨어 자원을 효율적으로 공유하고 네트워크상에 있는 원거리로부터의 정보에 대한 접근을 제공하기 위해 통신 네트워크의 이용은 급속히 증가하고 있다. 이러한 인터넷 시장의 규모가 점차 확대됨에 따라 네트워크를 어떻게 설계할 것인가 또는 어떻게 네트워크 서비스를 운용할 것인가에 대한 다양한 연구가 네트워크 설계자, 네트워크 분석가, 네트워크 관리자 등과 같은 많은 관련 연구자들에 의해 폭 넓게 이뤄지고 있다.

일반적으로 텔레커뮤니케이션, 컴퓨터 네트워크, 교통망 시스템, 운송망 최적화 시스템 등과 같은 많은 실제계의 네트워크 설계 문제에서, 어떤 제약조건들을 만족하면서 목적함수들을 최적화하는 네트워크 시스템을 설계하는 것은 매우 중대한 문제 중 하나이다. 이러한 네트워크 설계 문제들 중에 가장 근간이 되는 문제는 어떻게 네트워크 서비스를 운용할 것인가에 대한 방법으로

제일 많이 이용되는 최적 걸침 나무 (OCST : Optimal Communication Spanning Tree) 문제이다 [6]. OCST를 적용한 대형 통신 네트워크 시스템들은 백본으로 여러 네트워크들을 함께 묶어서 구성된다. 또한 근래에 네트워크를 구축하려고 할 때, 무중계 전송 가능 거리 (최대 120Km 이상), 광대역폭 (거의 50 Tbps), 적은 신호 감쇠 (거의 0.2 dB/km), 적은 신호 왜곡, 낮은 전력 요구 등 [1]과 같은 잠재력을 가지고 있어 광케이블이 주로 사용되고 있으며 광케이블 접속의 비싼 비용을 고려할 때 초기의 네트워크의 구조는 주로 걸침 나무 (Spanning Tree) 형태로 구성되는 것이 일반적이다. 또한, 인터넷 수능 방송과 같이 다수의 사용자가 동일한 서비스를 요청할 때 서비스하는 방식으로 그룹 통신을 위한 멀티캐스트 서비스에도 그 그룹을 관리하기 위해 주로 걸침 나무 형태로 구성하여 서비스하는 것이 일반적이며 최근 유비쿼터스 컴퓨팅과 관련하여 주목받고 있는 센서 네트워크에서도 그 배터리 수명을 고려한 현실적인 서비스 형태는 걸침 나무를 기반으로 하고 있다.

최적의 OCST를 찾아내는 것은 매우 복잡한 조

합최적화 문제이며 NP-hard 최적화 문제로 분류된다. 전통적인 방법들을 이용하면, 네트워크의 규모가 지수 적으로 확대되어지는 경우, 네트워크 토폴로지 설계 문제를 풀이하기가 힘들어진다. 이러한 이유 때문에 유전 알고리즘 (GA)을 기초로 하는 휴리스틱 알고리즘이 주목을 받고 있다. GA는 많은 실제계문제들을 풀기 위해 자주 이용되고 있으며 네트워크 최적화 문제들의 최적화 기술로서의 잠재력을 매우 주목받고 있는 기법이다 [2, 3]. 그리고 다목적 문제를 위한 GA 기법으로도 많이 연구되어 지고 있다.

본 논문에서는 이러한 GA를 OCST 문제에 적용하는 해법으로 제안한다. 제안하는 GA는 n 개의 노드에 대해 $n-2$ 개의 숫자 열로 표현가능하다고 알려진 기존의 Prüfer수 (PN)를 개선하여 보다 더 최적 통신 걸침 문제에 적합하도록 고안한 새로운 유전자 표현법을 이용한다. 제안하는 새로운 유전자 표현법은 최상위 노드를 포함한 중간 노드들의 연결 상태는 PN을 이용하여 표현하고 그의 잎노드와 중간노드들과의 연결 상태는 클러스터링 문자열로 표현하여 PN만 사용한 알고리즘보다 더 빠른 GA를 제안한다. 마지막으로 임의로 생성된 예제의 수치 실험을 통해 본 논문에서 제안하는 GA가 OCST문제를 효과적이고 효율적으로 풀이할 수 있는 해법임을 확인한다.

II. OCST 문제

OCST 문제는 일반적인 최소걸침나무 (MST)문제와는 달리 각각 노드들의 쌍과 관련이 있는 통신 요구량 (Communication requirements)과 각 링크의 연결 비용을 고려한다. 제약 조건이 있는 OCST 문제의 경우는 연결비용, 통신 요구량, 네트워크 신뢰도 등과 같은 성능 측도를 최적화하는 최상의 레이아웃 (Layout)을 탐색해 내는 것을 말한다. 이러한 시스템의 성능측도들은 매우 중요하며 주로 네트워크 토폴로지에 의해 좌우되며 클러스터링과 라우팅의 두 가지 주요한 항목들도 OCST 문제에 포함되어 있다. 클러스터링 문제는 얼마나 많은 중간 노드들 (Intermediate Nodes)로 네트워크가 나뉘질 수 있는가와 잎 노드들 (Leaf Nodes)이 클러스터에 연결되어 질 것인가와 같은 두 가지 항목들로 이뤄져 있으며 라우팅 문제는 걸침 나무 토폴로지로 어떻게 상호 연결할 것인가를 결정하는 것이다.

본 논문에서는 다음과 같은 일반적인 OCST문제에 유전 알고리즘을 적용하는 해법을 제시한다. 본 논문에서 고려하는 두 노드간의 통신비용 (Communication cost)은 임의의 i 노드와 j 노드가 있을 때, i 노드와 j 노드의 거리와 통신요구량의 곱으로 정의되며 본 논문에서의 OCST 문제의 해결을 위한 목적 함수는 다음과 같은 통신비용 함수를 계산하여 최소통신비용을 가지는 Spanning Tree를 구성하는 것이다.

$$Cost(T) = \sum_{i, j \in V} d(i, j) \cdot c_t(i, j) \quad (1)$$

III. 제안하는 유전 알고리즘

1. 유전자 표현 및 초기화

유전자 표현은 코딩 공간에서 문제의 후보 해를 표현하기 위한 자료구조를 의미한다. 보통 다른 종류의 문제들은 다른 자료구조 또는 유전자 표현을 사용하여야 유전자 알고리즘의 탐색 능력을 훼손하지 않으면서 준 최적해 또는 최적 해를 빠르게 탐색해 나갈 수 있다. 본 논문에서 제안하는 해법의 첫 번째 단계인 유전자 알고리즘이 주어 진 네트워크 설계 문제 환경에서 스페닝 트리를 찾기 위해서 이용되므로 유전자 표현으로 스페닝 트리를 나타내는 자료구조를 잘 고안해야 한다.

그래프 이론에서 오래된 정리들 중 하나는 k 개의 노드로 이뤄진 완전 그래프에서 $k-2$ 개의 구별되는 트리들이 존재한다는 Cayley 정리이며 Prüfer는 그러한 트리들과 k 개의 숫자들로 이뤄지는 $k-2$ 길이의 문자열들의 집합사이의 일대일 대응 관계를 확증하였다. 이것은 k 개의 다른 숫자들로 구성되는 $k-2$ 개의 숫자 조합의 열로 트리를 표현할 수 있다는 것을 의미한다. 왜냐하면 임의의 트리에 대해 적어도 2개의 잎(leaf) 노드가 존재하기 때문이다. 이러한 숫자 조합의 열을 Prüfer수(PN)이라고 한다[2, 12]. 이러한 아이디어를 기초하여 다음의 프로시저를 통해서 PN으로부터 트리를 생성할 수 있다:

Procedure: Decoding of Prüfer Number

- Step 1:** P 를 PN이라고 P' 를 P 에 속하지 않는 노드들의 집합이라고 둔다.
- Step 2:** P' 에서 가장 작은 숫자(가장 작은 숫자로 레이블된 노드)를 i 라고 하고 P 의 가장 왼쪽에 있는 숫자를 j 라고 하자. 그러면 i 로부터 j 에로의 간선 (i, j)을 트리에 추가한다. 추가한 후 P' 로부터는 i 를, P 로부터는 j 를 제거한다. 만약 P 로부터는 삭제되는 j 가 삭제된 후 P 에 j 가 더 이상 나타나지 않으면 j 를 P' 에 입력한다.
- Step 3:** P 에 숫자가 남지 않을 때까지 Step 2를 반복한다. P 에 숫자가 남지 않으면 P' 에 숫자 두개, 즉 r 과 s 가 남는다. 그러면 간선 (r, s)을 트리에 추가하면 최종적인 트리가 생성된다.

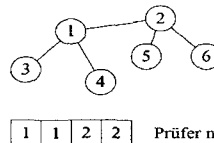


그림 1. 트리과 그에 대한 Prüfer수(PN)

그림 1에 6개의 노드로 구성되는 스페닝 트리를 PN (1 1 2 2)으로 표현하는 예를 나타내었다. PN으로부터 트리를 생성하는 디코딩 과정을 살펴보면, $P=(1\ 1\ 2\ 2)$ 에 대해 노드 3, 4, 5, 6이 PN에 나타나지 않는 노드 집합을 $P'=\{3, 4, 5, 6\}$ 으로

로 나타낼 수 있다. 노드 3이 제일 낮은 라벨을 가진 노드고 P의 가장 왼쪽에 있는 숫자가 1이므로 노드 3과 노드 1을 연결한 후 P에서 노드 3을 PN에서 노드 1을 제거한다. 그 다음으로 노드 4와 P의 가장 왼쪽 숫자 1을 고려하여 노드 4와 노드 1을 연결한 후 각각 제거를 한다. 이때 P에서 숫자 1을 제거할 때 이제 더 이상 숫자 1이 나타나지 않기에 숫자1을 집합 P에 포함시킨다. 그럼, P는 {1, 5, 6}이 된다. 이러한 과정을 P가 비어지고 P에 두개의 숫자가 남을 때까지 반복하고 P에 남은 두 숫자의 노드를 서로 연결하면 그림 1의 트리가 생성된다.

PN을 이용한 인코딩 방법은 가능한 모든 스페닝 트리를 표현할 수 있을 뿐만 아니라 PN에 노드가 d-1번 나타나면 그 노드의 차수는 d가 된다는 정보도 나타낼 수 있다. 즉, PN에 노드의 라벨이 d-1번 나타나면 트리에서 그 노드는 d개의 연결을 가지게 된다. PN은 운송문제, 최소 걸침 나무 문제 등과 같은 연구 분야에서 걸침 나무를 표현하기 위한 방법으로 적당하다는 연구들의 보고가 있다[2, 3].

그림 2에 표시된 것과 같이 본 논문에서는 n개의 노드에 대하여 PN에 기반을 둔 표현 방법은 n-2개의 메모리만으로도 트리를 표현 할 수 있기 때문에 서비스 센터들의 연결을 결정하는 의사 결정자를 위한 유전자 표현으로 이용하고 센터와 터미널(사용자)과의 연결을 결정하는 의사 결정자를 위해서는 클러스터 문자열로 유전자 표현법을 이용한다.

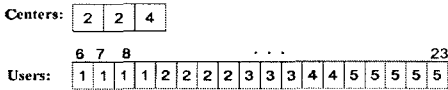


그림 2. 제한하는 유전자 표현

초기화 과정에서 상위 문제의 결정 변수들이 나타내는 백본 네트워크의 연결 상태는 [1, n]의 영역에서 임의로 n-2개의 숫자(PN)로 생성하여 나타내고 하위 문제의 결정 변수들이 의미하는 분산네트워크의 연결 상태는 영역 [1, n]에서 임의로 m개 숫자(클러스터 문자열)로 생성한다. 여기서 클러스터 문자열은 어떻게 사용자를 서비스 센터에 할당되는지에 대한 정보를 나타낸 것으로 사용자가 어떤 특정 서비스 센터에 속하는 것을 의미한다. 본 논문에서의 유전자 표현은 n+m-2의 길이를 가진다.

2. 평가 함수

본 논문에서는 다음과 같이 의사결정자를 위한 목적 함수치로 염색체들의 적합도를 계산한다.

$$eval(V_k) = Cost(V_k)$$

여기서 V_k^i 는 k번째 세대에서 집단내의 i번째 염색체를 의미하고 $Cost(\cdot)$ 는 수식 (1)에 의해 결정되는 목적함수를 의미한다.

3. 선택

GA에서 선택연산은 매우 중요한 역할을 한다.

교차 및 돌연변이 같은 유전자 연산들은 해 공간에서의 탐색을 위한 탐험연산이라고 한다면 선택 연산은 진화과정을 인도하는 GA를 위한 개척연산이라고 할 수 있다. 본 논문에서 사용된 선택 연산은 자유로이 해 공간을 탐색하도록 하기 위해 토너먼트 선택법과 엘리트리스트 (elitist)방법을 혼합한 형태이다.

토너먼트 선택은 n개의 염색체를 임의로 선택하고 이들 중에서 가장 적응도가 높은 개체를 다음세대의 집단에 포함시킨다. 다음 세대의 염색체들을 모두 얻을 때까지 이 과정을 반복한다. 이때 선택된 염색체를 현재 집단에서 제거하지 않는다. 즉, 한 염색체가 여러 번 선택될 수 있다. 본 논문에서는 n을 5로 설정하여 이용한다. 선택 압력은 토너먼트 크기인 n의 값에 따라 좌우되는데 n이 크면 적합도가 낮은 염색체들이 다음 세대에서 선택되는 기회가 적어진다. 이러한 토너먼트 선택은 최근에 주로 많이 사용되는 선택법 중에 하나이고 구현이 쉬우며 병렬 구조로 수행될 때 용이하고 선택 압력을 쉽게 조절할 수 있다는 장점을 가지고 있다. 본 논문에서 사용하는 선택법을 다음의 프로시저로 설명할 수 있다.

프로시저: 선택법

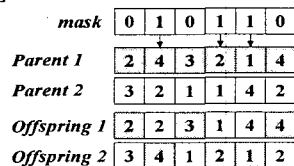
- Step 1: 먼저 엘리트리스트 (elitist)방법으로 이전 세대의 집단의 최선해를 다음세대의 집단으로 선택한다.
- Step 2: 먼저 이전 세대의 집단 중에서 n개를 선택한다.
- Step 3: 선택된 n개 중에서 가장 높은 적합도록 가진 하나의 염색체를 선택한다.
- Step 4: 다음 세대의 염색체들을 모두 얻을 때까지 Step 2와 3을 반복한다.

이런 선택연산을 이용하여 최선의 염색체를 현재의 세대에서 다음세대로 유지할 수 있다.

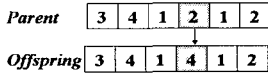
4. 유전연산자들

균등교배 (uniform crossover)로도 불리는 다점 교배 (multi-point crossover)를 이용한다. 이러한 종류의 교배는 그림 3에서 보는 바와 같이 두개의 부모 염색체를 임의로 선택하여 하나의 부모 염색체에서 임의로 유전자 표현형들을 선택하여 자손염색체를 구성한다 [2].

또한, 본 논문에서의 돌연변이 연산은 하나의 염색체에서 임의로 선택된 한 개의 위치에 있는 유전자 표현형들을 변환하는 단순 돌연변이를 사용한다 [2].



(a) 다점 교배 연산



(b) 교환 돌연변이 연산자

그림 3. 유전자 연산

IV. 수치실험

수치 실험을 통해 본 논문에서 제안하는 새로운 방식의 유전자 표현법을 이용한 유전 알고리즘의 성능에 대한 우수성을 나타낸다.

수치실험은 5 (최상위 및 중간노드 수:3, 잎노드: 2)개의 노드를 가진 문제(Example 1)와 12 (최상위 및 중간노드 수:4, 잎노드:8)개의 노드를 가진 문제(Example 2)로 수행하였으며 각각은 집단의 크기를 100, 최대 세대수를 1000, 교배율을 0.7, 돌연변이율 0.6으로 설정하여 수행되었다.

Example 1의 통신 요구량 행렬과 비용행렬은 참고문헌 [6]에 나타내져 있는 것을 이용하였고 Example 2의 통신 요구량 행렬과 비용행렬은 각각 [0, 200]사이의 영역에서 임의로 생성한 행렬을 이용하였다. 수치 실험은 각 예제에 대해 20번씩 수행하였으며 평균을 내어 그 결과들을 표 1에서 요약하여 나타내었다.

표 1. 수치 실험 결과

Numerical Examples	New Encoding		Old Encoding	
	objective	time	objective	time
1	283	1	285	1.05
2	31348	19.1	32354.4	19.9

표 1에서 볼 수 있듯이 제안한 새로운 유전자 표현법이 PN만 사용한 표현법보다 계산시간은 적게 요구하면서 더 최적해에 가까운 값을 찾아낸다는 것을 확인할 수 있다.

V. 결론

본 논문은 실세계의 네트워크 문제들 중에 가장 기본이 되는 OCST 문제를 위해 유전 알고리즘을 새로운 유전자 표현법을 이용하여 Prüfer수를 이용한 유전 알고리즘 보다 계산시간과 탐색해낸 해의 최적성의 측면에서 더 좋은 성능을 가진 해법으로 개선하였고 수치 실험을 통해 그에 대한 증명을 하였다. 앞으로의 연구는 보다 더 많은 수치예제에 적용해 보고 메시지 지연시간 및 네트워크 신뢰도에 대한 제약이 있는 OCST 문제에 대한 해법을 고안하는 것에 열중할 것이다.

참고문헌

[1] Mukherjee, B., Optical Communication Networks, McGraw-Hill, Inc., New York,

1997.
 [2] Gen, M. and R. Cheng, Genetic Algorithms and Engineering Design, John Wiley & Sons, New York, 1997.
 [3] Gen, M. and R. Cheng, Genetic Algorithms and Engineering Optimization, John Wiley & Sons, New York, 2000.
 [4] 임용택, 임강원, "게임 이론과 교통망 문제", 서울대학교 환경대학원학술지 환경논총, 제 42권, pp.107-121, 2004.
 [5] 김동환, 현창희, "정보통신정책의 게임구조와 그 역동성", 국가기간전산망저널 (한국전산원), 제2권, 제4호 pp.36-48, 1995.
 [6] 석상문, 장석철, 변성철, 안병하, "최적통신 결집나무 문제를 해결하기 위한 진화 알고리즘", 정보과학회논문지: SA, 제32권, 제4호, pp.268-276, 2005.4.
 [7] J. F. Wang and J. Periaux. "Multi-point optimization using GAs and Nash/Stackelberg games for high lift multi-airfoil design in aerodynamics," Proc. of the 2001 Congress On Evolutionary Computation, Vol.1, pp.552-559, 2001.
 [8] M. Sefrioui and J. Periaux, "Nash genetic algorithms: examples and applications," Proc. of the 2000 Congress On Evolutionary Computation, pp.509-516, 2000.
 [9] Elbaum, R. and M. Sidi, "Topological design of local-area networks using genetic algorithms," IEEE/ACM Transactions on Networking, Vol.4, No.5, pp.766-778, 1996.
 [10] Skiena, S., Implementing Discrete Mathematics Combinatorics and Graph Theory with Mathematica, Addison-Wesley, Reading, MA, 1990.
 [11] G. R. Raidl and B. A. Julstrom, "Edge sets: an effective evolutionary coding of spanning trees," IEEE Transaction on Evolutionary Computation, Vol.7, No.3, pp.225-239, June, 2003.
 [12] Gen, M. and R. Cheng, "Evolutionary network design: hybrid genetic algorithms approach," International Journal of Computational Intelligence and Applications, Vol.3, No.4, pp.357-380, 2003.