

# Development of evolutionary algorithm for determining the k most vital arcs in shortest path problem<sup>1)</sup>

Hoyeon Chung, Dongju Shin

Department of Industrial Engineering, Jeonju University  
3Ga 1200 Hyoja-Dong Wansan-Gu Chonju-Si, 560-759

## Abstract

The purpose of this study is to present a method for determining the k most vital arcs in shortest path problem using an evolutionary algorithm. The problem of finding the k most vital arcs in shortest path problem is to find a set of k arcs whose simultaneous removal from the network causes the greatest increase in the total length of shortest path. The problem determining the k most vital arcs in shortest path problem has known as NP-hard. Therefore, in order to deal with the problem of real world the heuristic algorithm is needed. In this study we propose to the method of finding the k-MVA in shortest path problem using an evolutionary algorithm which known as the most efficient algorithm among heuristics. For this, the expression method of individuals compatible with the characteristics of shortest path problem, the parameter values of constitution gene, size of the initial population, crossover rate and mutation rate etc. are specified and then the effective genetic algorithm will be proposed. The method presented in this study is developed using the library of the evolutionary algorithm framework (EAF) and then the performance of algorithm is analyzed through the computer experiment.

## 1. 서론

최단경로문제는 특정한 두 교점 사이의 경로 중에서 가장 짧은 길이의 경로를 찾는 문제이다. 이 최단경로문제에서 k개의 호의 제거로 시발점에서 종착점까지의 최단거리가 가장 크게 늘어나는 k개의 호집합을 찾는 문제를 k개의 치명호(k Most Vital Arcs)를 찾는 문제라고 한다.

현재 최단경로문제에서 k개의 치명호를 찾는 문제에 대한 해법은 Corley & Sha[2,6]가 제시한 충분조건과 이 문제가 NP-hard 라는 사실만이 알려져 있다. 안재근[2] 등은 호와 경로와의 관계를 이용한 k-치명호에 대한 수리모형식과 이에 대한 해법을 제시한 바 있으나 실용성에 다소 문제점을 안고 있다. 이 문제는 그러나 Bar-Noy[5] 등에 의해 NP-hard의 복잡도 임이 밝혀졌기 때문에 최적화 알고리즘이 존재하더라도 문제의 크기가 커짐에 따라 컴퓨터 용량과 계산시간의 한계를 갖게 된다. 이러한 유형에 속한 문제는 최적해를 구하기 보다는 발견적 해를 구하는 것이 여러 면에서 경제적이다[1]. 따라서 본 연구에서는 발견적 해법 중에서도 해의 속도나 질적인 면에서 좋은 결과를 제공하는 것으로 알려져 있는 진화 알고리즘을 적용하여 최단경로문제에서 k개의 치명호를 결정하는 알고리즘을 개발하고자 한다.

## 2. k-MVA 결정을 위한 유전알고리즘

• 개체표현 방법 및 모집단 구성  
유전알고리즘을 적용하기 위해서는 먼저 주어진 문제의 특성을 반영하는 개체 표현방법이 필요하다. 본 연구에서는 이진표현 대신 호에 일련번호를 부여하여 호를 표현하는 방법을 사용한다. 이 표현을 사용하면 하나의 개체는 치명호에 해당하는 k개의 호로 구성되는 호 인자로 개체가 표현될 수 있다.

초기 모집단의 크기(population size)는 임의로 생성된 k개의 호 집합으로 구성되는 이러한 개체 10~500개의 범위 내에서 적절히 선택할 수 있게 하였다.

• 적응도 평가 및 계산

본 연구에서 정의한 개체는 치명호에 해당하는 k개의 호로 구성된다. 따라서 이에 대한 적응도는 개체를 구성하는 k개의 호를 제거한 다음 최단경로의 길이를 구한 값과 원문제의 최단경로의 길이의 차로써 계산된다.

최단경로문제를 구할 때 사용한 알고리즘은 Dijkstra 알고리즘을 사용하였다. 따라서 개체에 대한 적응도 값은 그 값이 클수록 k-MVA에 근접해 간다고 말할 수 있다.

• 선별 방법(selection method)

본 연구에서는 현재 모집단의 우수한 개체를 지속적으로 유지하기 위해 Goldberg에 의해 제안된 우수개체 보호 전략(elitism)을 사용한다. 이를 위해 먼저 모집단  $P(t)$ 의

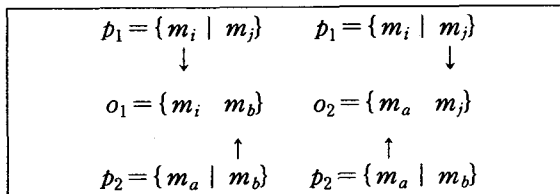
1) 본 연구는 한국과학재단 특정기초연구(과제번호 : 98-0200-09-01-3)지원으로 수행되었음

적응도 값을 내림차순으로 정렬한 후, 최상위 개체 하나를 선택하여 복사(copy)해 놓는다. 현재의 모집단  $P(t)$ 를 대상으로 토너먼트 선별을 실시하여 다음 세대  $P(t+1)$ 를 생성시킨 다음 적응도 값에 따라 내림차순으로 정렬한다. 이 때의 최상위 개체 값과 그 이전 세대에서 복사해 놓은 최상위 개체 값과의 비교를 통해 더 우수한 개체를  $P(t+1)$ 에 그대로 유전시킨다.

• 유전연산자

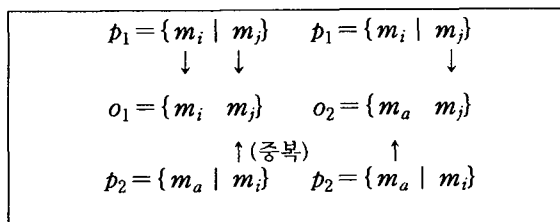
유전 연산자는 교차와 돌연변이로 구성된다. 본 논문에서는 교차나 돌연변이를 실시할 때 발생할 수 있는 중복을 방지하기 위하여 순서교차(order crossover)를 사용한다. 순서교차는 한 부모로부터 부분적인 순서(인자)를 받아들이고, 다른 부모로부터 상대적인 순서를 유지하여 받아들이는 교차방법이다[1]. 본 연구에서는 상대적인 순서를 받아들이는 개념을 중복이 발생할 경우에 한해 해당하는 부모개체를 그대로 상속하게 하는 개념으로 변형하여 적용하였다. 순서교차에서 절단점은  $k=2$  일 때에는 일점(one-point) 절단점을 사용하고,  $k=3$  이상일 때는 이점(two-point) 절단점을 임의로 선택하여 교차를 실시하였다.

$k=2$ 인 경우 순서교차를 수행한 예가 다음 [그림 1]에 나타나 있다.



[그림 1] 순서교차 방법 (중복이 발생되지 않는 경우)

그러나 [그림 2]와 같이 부모  $p_1$ 과  $p_2$ 가 주어졌을 경우에는 부모  $p_1$ 과  $p_2$ 로부터 각 인자를 상속받아 자손  $o_1$ 과  $o_2$ 가 생성되게 되는데 이 때 자손  $o_1 = \{m_i \mid m_j\}$ 가 되어 중복이 발생되게 된다. 이 때에는 부모  $p_2$ 로부터 인자  $m_i$ 를 상속받지 않고 대신 부모  $p_1$ 의 인자를 자손  $o_1$ 에 그대로 상속되게 함으로써 중복을 방지한다.



[그림 2] 순서교차 방법 (중복이 발생될 경우)

돌연변이(mutation)는 개체에 완전히 새로운 유전자를 생성시키는 단계로써 해 공간을 다양하게 탐색하게 해주는 역할을 한다[11]. 본 연구에서 사용한 돌연변이는 원소

단위로 행하여지며, 중복이 발생할 경우에는 다른 원소를 발생시켜 중복이 되지 않는 원소로 돌연변이 시킨다.

• 종료조건

본 연구에서는 새로 생성된 개체수가 1000개 이상이면 더 이상 해를 개선시킬 수 없다고 보고 실험을 종료하도록 하는 종료조건을 사용한다.

이상의 내용을 참고하여 최단경로문제에서  $k$ -MVA를 결정하기 위한 유전 알고리즘을 정리하면 다음과 같다.

단계1. (초기 모집단 구성)

임의 생성법을 사용하여  $P(t)$ 를 구성.

단계2. (적응도 평가)

$P(t)$ 에 있는 모든 개체의 적응도를 평가.

단계3. (선별)

$P(t)$ 로부터  $P(t+1)$ 를 선별한다. 이를 위해  $P(t)$ 의 적응도 값을 내림차순으로 정렬했을 때 나타나는 최상위 개체 값과 그 이전 세대에서 복사해 놓은 최상위 개체 값과의 비교를 통해 더 우수한 개체를  $P(t+1)$ 에 유전시키고, 나머지 개체들은 토너먼트 선별을 통해 세대를 구성한다.

단계4. (교차)

현재의 모집단에서 임의로 선정된 두 개체로부터 순서교차를 사용하여 자손 개체를 생성한다.

단계5. (돌연변이)

모집단을 구성하고 있는 개체 중에서 돌연변이를에 따라 임의의 개체를 돌연변이 시킨다.

단계6 (종료조건 검사)

종료조건을 만족하면 끝내고, 그렇지 않으면  $t = t + 1$ 로 두고 단계2로 간다.

3. 실험 및 분석

최단경로문제에서  $k$ -MVA를 결정하기 위한 유전 알고리즘의 실험은 DIMACS의 표준문제 생성기에서 생성된 문제를 대상으로 실험하였다. 제안한 유전알고리즘은 Visual C++ 컴퓨터 언어를 사용하였으며, 128M RAM을 가진 350MHz의 Pentium-II CPU를 장착한 IBM PC 호환기종에서 수행되었다.

• 유전 파라미터 값 설정을 위한 예비실험

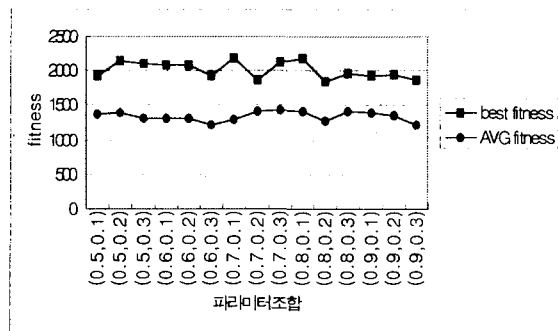
제안된 유전 알고리즘으로 좋은 해를 찾기 위해서는 유전 알고리즘에 민감하게 영향을 미치는 유전 파라미터의 값을 최적으로 설정하여야 한다. 최적 유전 파라미터의 값을 결정하기 위하여 본 연구에서는 DIMACS의 표준문제 생성기에서 생성된 spagrid3 문제를 대상으로 예비실험을 수행하였다. spagrid3 문제는 마디(node)의 수가 661개이고 호의 수가 1980개인 최단경로문제이며, 실험에서는 치명호의 수가 3인 3-MVA를 구하기로 하였다. 예비실험은 먼저 모집단의 크기를 100으로 설정한 후 교차율은 0.5에서 0.9까지 돌연변이율은 0.1에서 0.3까지 0.1 단위로 값을 변경해 가면서 각 조합당 20회씩의 반복실험을 통해 최적 교차율과 돌연변이율을 탐색하였다. 이 때 각각의 교호작용은 무시하였으며, 20회의 반복실험 중에서 가장 높은 적응도의 평균값을 기준으로 최적 유전 파라미터의 값을 설정

하였다. 각 조합 조건에 대한 실험의 종료조건은 실험을 통해 새롭게 생성된 개체의 수가 1000개 이상이면 더 이상 해가 개선될 가능성이 없기 때문에 새로 생성된 개체의 수가 1000개 이상일 때 실험을 종료하는 종료조건을 사용하였다. 예비 실험계획에 따라 5가지의 교차율과 3가지의 돌연변이율에 따른 15가지의 조합에 대하여 각각 20회씩 반복 실험을 수행한 결과가 아래 [표 1]에 정리되어 있다.

[표 1] 예비실험 결과

파라미터조합 (교차율/돌연변이율)	best fitness	average fitness
(0.5/0.1)	1915	1374
(0.5/0.2)	2138	1392
(0.5/0.3)	2109	1313
(0.6/0.1)	2080	1319
(0.6/0.2)	2081	1319
(0.6/0.3)	1915	1216
(0.7/0.1)	2190	1283
(0.7/0.2)	1863	1407
(0.7/0.3)	2116	1430
(0.8/0.1)	2180	1401
(0.8/0.2)	1846	1270
(0.8/0.3)	1968	1412
(0.9/0.1)	1924	1390
(0.9/0.2)	1951	1341
(0.9/0.3)	1862	1216

위의 표는 교차율과 돌연변이율의 각 조합에 대하여 20회씩의 반복실험을 통해 구해진 가장 좋은 적응도(best fitness)값과 평균 적응도(average fitness)의 값을 보여준다. 이를 그림으로 나타내면 다음 [그림 3]과 같다.



[그림 3] 예비실험 결과

본 연구에서는 위의 예비실험 결과에서 가장 좋은 평균 적응도 값을 나타낸 (0.7/0.3) 파라미터 조합을 최적 유전파라미터로 설정하였다. 최고의 적응도(best fitness) 보다 평균적응도(average fitness) 값이 높은 파라미터의 조합을 선택한 이유는 우연성을 배제하기 위함이며 안정적으로 해를 산출할 수 있는 파라미터의 조합을 설정하기 위함이다. 최적 모집단의 크기는 앞의 예비실험에서 구했던 (교차율/돌연변이율)을 (0.7/0.3)으로 설정한 상태에서 모집단의 크

기를 몇 단계로 변경시켜 가면서 20회씩의 반복실험을 통해 탐색하였다. 모집단의 크기에 따른 평균 적응도 값의 변화가 다음 표에 나타나 있다.

[표 2] 모집단 크기에 따른 평균적응도

모집단크기	average fitness
100	1394
200	1516
300	1086
400	973
500	1179

위의 표에서 보는 바와 같이 모집단의 크기가 200일 때 평균적응도 값이 가장 높게 나타났다. 따라서 예비실험 결과 가장 좋은 알고리즘의 성능을 보인 최적 파라미터의 조합은 모집단의 크기는 200, 교차율은 0.7, 그리고 돌연변이율은 0.3으로 최종 결정되었다. 이러한 유전 파라미터의 값으로 spagrid3 문제에 대한 실험 결과 3-MVA는 (1317,1365,1320)로 결정되었다.

최적 유전 파라미터값(모집단의 크기=200, 교차율=0.7, 돌연변이율=0.2)을 사용하여 호의 수가 5~18개인 작은 규모의 최단경로문제 5개에 대해 3-MVA를 찾은 결과 100% 최적해를 찾을 수 있었다. 이에 대한 결과가 [표 3]에 나타나 있다.

[표 3] 작은 규모의 문제 적용결과

문제명 (마디수, 호의 수)	best fitness	3- MVA	최적 여부
Test1 (5,7)	∞	{7,4,6},{3,6,1},{3,7,6},{2,1,3},{6,1,7}, {3,1,6},{5,6,7},{7,6,2},{3,4,2}	최적해
Test2 (6,10)	∞	{6,10,8},{10,7,8},{2,8,10},{5,1,2}, {6,8,5},{3,8,5},{8,10,4},{5,8,4}, {10,6,7},{5,7,10},{2,7,3}	최적해
Test3 (7,10)	∞	{2,1,9},{9,5,1},{10,1,9},{10,3,9}, {4,2,3},{10,9,6},{4,5,9},{9,5,10}, {9,3,8},{10,9,2},{4,9,10},{9,10,4}, {8,9,5},{9,7,8},{5,6,1}	최적해
Test4 (8,13)	∞	{7,11,1},{7,1,4},{6,9,7},{7,6,2}, {7,6,10},{3,1,7},{11,7,10},{1,7,12}, {7,5,1},{2,7,1},{6,2,7},{13,7,1}, {1,7,10},{3,6,7},{5,6,7}	최적해
Test5 (10,18)	∞	{3,18,17},{18,9,17},{2,3,1},{18,17,15}, {1,17,18},{17,10,18}, {17,7,18},{18,4,17},{17,8,18}, {2,18,17},{14,18,17}	최적해

DIMACS의 표준문제 생성기에서 생성된 문제에 대하여 실험을 수행한 결과는 다음 [표 4]에 나타나 있다.

[표 4] 문제적용결과

적용문제 (마디수, 호의수)	best fitness	average fitness	3- MVA
Spagrid3_1980	4631	1516	{1317,1365,1320}
Spagrid7_5100	1556	949	{1156,2349,3045}
Spagrid10_6000	9935	4935	{436,907,4109}
Sprand3_2000	2746	995	{1510,1354,1799},{1510,1469,308},{1731,25,1510}
Sprand5_5000	3958	1900	{222,997,2802}
Sprand11_10000	3531	1412	{3355,5974,3918},{4266,1508,3910},{7493,2953,7432},{7432,3457,7528}, {4284,6363,1577},{7432,9865,6056},{5252,8485,5974},{2335,4266,3221}
Spacyc7_7025	2359	429	{2061,5180,3720}
Spacyc8_14364	7135	356	{4676,1076,6393}
Spacyc10_20140	2232	223	{4169,6574,1598},{1299,19032,6574}

#### 4. 결론

본 논문은 유전알고리즘을 사용하여 최단경로문제에서  $k$ -MVA를 결정하는 방법을 다루고 있다. 최단경로문제에서  $k$ -MVA를 결정하는 문제는 조합 최적화 문제로써 문제의 크기가 커지면 시간과 기억 용량이 지수적으로 증가하는 NP-hard 문제에 속한다. 이 문제에 대한 해법은 현재까지 알려져 있지 않으며, 최근에 제시된 방법도 현실적으로 실용화되기 어려운 단점을 갖고 있다. 그러나 본 논문에서 제시한 해법은 대형문제에 적용한 결과 호의 수가 20,000개 정도까지도 3-MVA를 찾을 수 있었다. 문제에 대한 최적해법이 정확히 알려져 있지 않기 때문에 대형문제에 대한 최적해를 알 수 없었지만, 최적해를 알 수 있는 작은 규모의 문제에 본 연구에서 제안한 방법을 적용한 결과 100%의 해의 정확도를 보여 양질의 좋은 해를 구할 수 있음을 알 수 있었다.

#### 참 고 문 헌

- [1] 김여근, 윤복식, 이상복, 메타휴리스틱 - 유전알고리즘, 시뮬레이티드 어닐링, 타부서치, 초판, 영지문화사, 1997
- [2] 안재근, 정호연, 박순달, "최단경로문제에서 k개의 치명호를 찾는 방법", 한국경영과학회지 제23권 제4호, 한국경영과학회, (1998.12), pp.11-20
- [3] 정호연, "최대유통문제에서 k-MVA를 결정하는 방법", 한국군사운영분석학회지, 제25권 제2호, 한국군사운영분석학회, (1999.12), pp.106-116
- [4] Ahuja.R.J., Magnanti.T.L., Orin.J.B., Network Flows : Theory, Algorithms, and Applications, Prentice-Hall, 1992,
- [5] Bar-Noy A., S. Khuller, B. Schieber, "The Complexity of Finding Most Vital Arcs and Nodes", Univ. of Maryland Technical Reports, CS-TR-3539 (1995), pp 1 - 7.
- [6] Corley JR. H. W., D. Y. Sha, "Most Vital Links and Nodes in weighted Networks", Operations Research Letters, Vol.1 No.4 (1982), pp 157 - 160.
- [7] Goldberg.A.V.,<http://www.neci.nj.nec.com/homepages/avg>
- [8] Lubore S.H.,H.D.Ratliff, G.T.Sicilia, Determining The Most Vital Link in a Flow Network, NRLQ, Vol.18, No.4 (1971), pp 497-502
- [9] Malik K., A. K. Mittal, S. K. Gupta, "The  $k$  Most Vital Arcs in the Shortest Path Problem", Operations Research Letters, Vol.8 (1989), pp 223 - 227.
- [10] M.O. Ball, B.L. Golden, and R.V. Vohra. Finding the most vital arcs in a networks. Operations Research Letters, Vol.8 (1989), pp 73-76.
- [11] Munakata and Hashier.D.J.,A Genetic Algorithm Applied to the Maximum Flow Problem, ICGA, 1993, pp 488-493
- [12] Ratliff H.D., S.H.Lubore, G.T.Sicilia, Finding the n Most Vital Links in a Flow Network, Management Science, Vol.21, No.5 (1975), pp 531-539