

대규모 Maximal Covering 문제 해결을 위한 유전 알고리즘

(A Genetic Algorithm for a Large-Scaled Maximal Covering Problem)

박 태 진 [†] 황 준 하 ^{**} 류 광 렬 ^{***}

(Tae Jin Park) (Junha Hwang) (Kwang Ryel Ryu)

요 약 열의 수가 수십만에 이르는 대규모 maximal covering 문제(MCP)를 유전 알고리즘을 통해 해결하는 것에는 한계가 있다. 본 논문에서는 대규모 MCP를 유전알고리즘이 효율적으로 풀 수 있도록 하기 위해 특별히 고안된 교차 연산자와 돌연변이 연산자를 소개한다. 또한, 본 연구에서는 비발현 유전자를 사용하는 새로운 유전 알고리즘을 제시한다. 비발현 유전자는 유전 연산 과정에서 상실될 정보 중 이후의 세대에서 유용할 가능성이 있는 정보를 자손에게 전달하기 위해 보존하는 역할만 할 뿐, 발현되지 않음으로 인해 해의 평가 시에는 반영되지 않는 유전자이다. 비발현 유전자를 사용하는 유전 알고리즘은 집단의 다양성을 유지하는데 유리하여 대규모 MCP를 해결하는데 있어서 보다 효율적으로 탐색을 수행할 수 있다. 현장의 대규모 MCP 데이터로 실험한 결과 비발현 유전자를 가진 유전 알고리즘이 이웃해 탐색 알고리즘인 타부 탐색보다 훨씬 우수한 탐색 성능을 보임을 확인할 수 있었다.

키워드 : 유전알고리즘, maximal covering 문제, 비발현 유전자

Abstract It is very difficult to efficiently solve a large-scaled maximal covering problem(MCP) by a genetic algorithm. In this paper, we present new crossover and mutation operators specially designed for genetic algorithms to solve large-scaled MCPs efficiently. We also introduce a novel genetic algorithm employing unexpressed genes. Unexpressed genes are the genes which are not expressed and thus do not affect the evaluation of the individuals. These genes play the role of reserving information susceptible to be lost by the application of genetic operations but is suspected to be potentially useful in later generations. The genetic algorithm employing unexpressed genes enjoys the advantage of being able to maintain diversity of the population and thus can search more efficiently to solve large-scaled MCPs. Experiments with large-scaled real MCP data has shown that our genetic algorithm employing unexpressed genes significantly outperforms tabu search which is one of the popularly used local neighborhood search algorithms for optimization.

Key words : genetic algorithm, maximal covering problem, unexpressed gene

1. 서 론

Maximal Covering 문제(MCP)는 $n \times m$ 행렬 상에서 p 개의 열을 선택하여 최대한 많은 행을 cover해야 하는 최적화 문제이다. 예를 들어 그림 1과 같은 $n = 5$, $m = 6$, $p = 3$ 인 MCP에서는 열의 집합 $\{c_1, c_2, c_3\}$ 이

하나의 해가 될 수 있다. 이때 이 해에 의해 cover되는 행은 r_1, r_3, r_4 의 3개이며, r_1 은 c_1 과 c_3 에 의해 r_3 은 c_1 에 의해 그리고 r_4 은 c_2 에 의해 각각 cover되었다고 한다. MCP는 버스, 열차, 지하철, 비행기 등의 승무 일정 계획 문제나 facility location 등 고정된 양의 자원으로 최대한의 서비스를 수행해야 하는 문제 등의 모델링에 많이 사용된다.

MCP를 해결하기 위한 종래의 연구에서는 정수계획법[1]과 Lagrangian relaxation 기법[2]들이 많이 적용되었다. 또한 MCP와 매우 유사한 문제라 할 수 있는 Set Covering Problem(SCP)의 해결을 위해서는 유전 알고리즘을 적용한 사례[3]도 있다. 이들 중 정수계획법

[†] 비 회 원 : 부산대학교 정보컴퓨터공학부

parktj@pusan.ac.kr

^{**} 중 신 회 원 : 금오공과대학교 컴퓨터공학부 교수

jhhwang@cespc1.kumoh.ac.kr

^{***} 중 신 회 원 : 부산대학교 전자전기정보컴퓨터공학부 교수

kr Ryu@hyowon.pusan.ac.kr

논문접수 : 2003년 12월 26일

심사완료 : 2004년 2월 26일

	C1	C2	C3	C4	C5	C6
r1	1		1		1	
r2				1		
r3	1					
r4		1		1		1
r5					1	

그림 1 maximal covering 문제의 예

은 최적해를 보장해 주는 대신 문제의 규모가 조금만 커져도 해를 구하는 것 자체가 어려워지기 때문에 현실적 규모의 문제에는 적용이 불가능하다. 따라서 Lagrangian relaxation이나 유전알고리즘과 같은 휴리스틱 기법들이 그 대안으로 제시되기는 했지만 이들 또한 행의 수가 수십만에 이르는 대규모 MCP(혹은 SCP)가 주어질 경우에는 그다지 만족스러운 성능을 보이지 못했다. 최근 들어서는 simulated annealing이나 타부 탐색(tabu search)과 같은 이웃해 탐색 기법들을 이용하거나 이들을 정수계획법과 하이브리드 형태로 결합하여 적용함으로써 대규모 MCP를 효율적으로 풀 수 있음을 보인 연구 결과가 보고된 바 있다[4,5].

본 논문에서는 대규모 MCP를 해결하기 위해 유전알고리즘을 이용하는 새로운 접근 방안을 소개한다. 대부분의 최적화 문제에 대해 유전알고리즘이 이웃해 탐색 기법에 비해 계산시간이나 해의 질 측면에서 더 우수한 성능을 보이는 경우는 드문 것으로 알려져 있다. 대신 유전알고리즘의 전역적 탐색 능력과 이웃해 탐색의 신속한 수렴성을 결합한 하이브리드 형태의 해법이 많이 시도되어 성공을 거두어 왔다[6,7]. 본 논문의 유전알고리즘에서도 교차와 돌연변이 연산이 그리디(greedy) 휴리스틱 탐색에 의해 이루어지도록 함으로써 탐색 효율의 향상을 시도하였다. 또한, 본 논문에서는 집단(population) 구성원의 다양성을 유지하기 위한 수단으로 비발현 유전자(unexpressed gene)를 포함하는 새로운 염색체 구조를 제안한다. 비발현 유전자란 표현되지 않는 유전자로서 각 개체를 평가할 때에는 반영되지 않는 유전자를 말한다. 비발현 유전자는 유전연산(선택, 교차, 돌연변이)의 적용을 통한 진화 과정에서 상실될 정보 중 미래에 유용하게 사용될 가능성이 있는 정보들을 별도로 보존해 두는 역할을 한다. 비발현 유전자의 도입은 집단의 조기 수렴을 막아 탐색을 다각화하는데 크게 기여한다. 대규모 MCP에 대한 실험 결과 제안된 유전 알고리즘이 이웃해 탐색 기법보다도 더 좋은 성능을 보임을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 대규모 MCP를 위한 염색체 구조와 교차연산자에 대해서 설명한다. 3장에서는 대규모 MCP를 효율적으로 해결하기 위해 본 논문에서 도입한 비발현 유전자에 대해서 설명한다. 4장에서는 대규모 MCP를 위해 특별히 제안된 돌연변이 연산자에 대해서 설명하며, 5장에서는 실제 대규모 MCP에 대해서 타부 탐색과 성능 비교 실험을 한다. 6장에서는 결론을 맺고 향후 연구방향에 대해서 토의한다.

2. 염색체 표현과 그리디 교차

MCP에서 하나의 후보해는 문제에서 주어진 전체 열의 집합 $S = \{s_1, s_2, \dots, s_m\}$ 으로부터 선택된 p 개의 열로 이루어진 집합이다. 각 후보해는 기본적으로 p 개의 열이 cover하는 행의 수에 비례하는 목적함수에 의해 평가될 것이다. 문제에 따라서는 cover되는 행의 수 외에도 해를 구성하는 각 열에 대한 질적 평가치와 같은 다른 요소들도 목적함수에 포함될 수 있다. 유전알고리즘은 이러한 평가를 바탕으로 한 선택 그리고 교차 및 돌연변이를 적용하여 세대를 변천시킴으로써 탐색을 수행하게 된다.

그림 2는 전체 m 개의 열들을 대상으로 선택 여부에 따라 1과 0을 부여한 2진 부호 방식의 염색체 표현을 보여주고 있다. 이 표현 방식은 대규모 MCP에서처럼 p 에 비해 m 이 매우 클 경우 염색체의 길이가 지나치게 길어져서 비효율적이다.

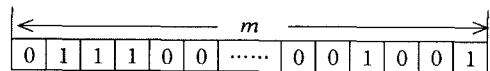


그림 2 MCP를 위한 염색체의 이진 표현

m 이 매우 클 경우 보다 적합한 염색체 표현 방식으로는 그림 3과 같이 선택된 p 개 열의 인덱스들을 스트링 형태로 나열하는 것을 생각할 수 있다. 여기서 열들의 스트링은 실제로 그 나열 순서에는 아무런 의미가 없는 열들의 집합일 뿐이다. 이러한 염색체 표현 하에서 보통의 일점교차(one-point crossover)나 이점(two-point)교차 연산을 적용할 경우에는 생성된 자손의 염색체가 동일한 열 두 개를 포함하는 불법(illegal) 스트링이 될 수가 있다. 따라서 본 연구에서는 부분집합 선택(subset selection)을 위한 새로운 교차 연산자를 도입한다. 선택된 2개의 부모 염색체를 P_1, P_2 라 하면, 부분집합 선택 교차는 집합 $P_1 \cup P_2$ 로부터 p 개의 열을 선택하여 부분 집합 P_c 를 자손으로 생성하는 연산이다. 이 교차 연산은 두 부모로부터 하나의 자손만 생성된다는 점에서도 보통의 교차 연산과 다르다.

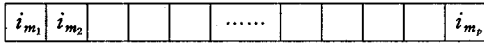


그림 3 MCP를 위한 염색체의 집합표현

그러나 부분집합 선택 교차 과정에서 부모들로부터 p 개의 열을 단순히 임의로 선택할 경우 좋은 자손이 생성될 가능성은 아주 낮다. 보다 좋은 자손을 생성하기 위해서는 문제에서 주어진 행렬의 행들이 최대한 많이 cover 되도록 p 개의 열들이 선택되어야 할 것이다. 그림 4는 두 부모의 열들을 모두 포함한 합집합으로부터 좋아 보이는 열들부터 하나씩 선택하여 자손을 만들어 내는 greedy-adding 휴리스틱에 의한 교차 연산의 수행과정을 보이고 있다. 부모의 열들로부터 가장 좋은 열 j 를 선택하는 기준이 되는 평가값 $\delta(j)$ 는 아래 수식 1과 같이 계산된다.

$$\delta(j) = Eval(P_c \cup \{j\}) - Eval(P_c) \quad (1)$$

여기서 $Eval(P)$ 는 열의 집합 P 가 cover하는 행의 개수이다. 이 수식에 의하면 추가함으로써 cover할 수 있는 행이 가장 많아지는 열이 가장 좋은 평가값을 받게 된다. $\delta(j)$ 값이 가장 큰 열이 2개 이상 존재하는 경우에는 $P(= P_1 \cup P_2)$ 의 열들 중 나머지 열들에 대한 유사도가 가장 낮은 열을 선택한다. 열의 집합 P 중 임의의 열 $j \in P$ 의 다른 나머지 열들(j 도 포함)에 대한 유사도 $S_P(j)$ 는 다음 수식 2와 같이 계산된다.

$$S_P(j) = \sum_{v \in P} j \cdot v = j \cdot \sum_{v \in P} v \quad (2)$$

이진 벡터 j 와 v 는 서로 유사할수록 그 내적의 값이 커지므로 j 와 다른 모든 벡터간의 내적의 합을 구함으로써 유사도를 추정할 수 있는 것이다. 유사도 $S_P(j)$ 는 그림 4의 알고리즘에서 while loop가 시작되기 전에 P 의 모든 열에 대해 미리 계산된다. $\delta(j)$ 값이 같더라도 유사도 $S_P(j)$ 의 값이 가장 낮은 열을 선택하는 것이 좋은 이유는 차후의 연이은 열 선택 과정에서 계속 이미 선택

```

Algorithm GreedyCrossover
input: parent chromosomes  $P_1, P_2$ 
output: child chromosome  $P_c$ 
Begin
 $P := P_1 \cup P_2;$ 
 $P_c := \emptyset;$ 
while  $|P_c| \neq p$  Do
    find  $j$  with maximum  $\delta(j)$  where  $j \in P;$ 
     $P_c := P_c \cup \{j\};$ 
     $P := P - \{j\}$ 
Return  $P_c$ 
End
    
```

그림 4 그리디 교차

된 열들이 cover하지 않는 행들을 많이 cover할 수 있을 가능성이 커지기 때문이다. 만약 이 때 반대로 유사도가 가장 높은 열을 선택한다고 가정해 보면 그 다음에 선택될 열이 이전에 이미 선택된 열들과 유사할 가능성이 높아지고 따라서 새로이 cover되는 행의 수가 늘어날 가능성은 낮아질 것임을 알 수 있다.

3. 비발현 유전자의 도입

앞에서 소개한 그리디 교차는 부모로부터 좋은 자손을 생성시키는 데 상당한 효과를 발휘한다. 그러나 세대변천이 거듭되어 감에 따라 그리디 교차 과정에서 선택되지 못한 열들이 집단으로부터 급격히 제거되면서 집단 전체 구성원의 다양성이 떨어져 결국에는 조기수렴을 하게 된다. 본 논문에서 제안하는 비발현 유전자를 가진 염색체 구조는 이러한 조기수렴 문제를 효과적으로 해결하기 위해 고안된 것이다. 비발현 유전자는 그리디 교차 과정에서 선택되지 못한 열들 중 이후 세대에서 유용하게 사용될 수 있을 것으로 기대되는 부분들을 별도로 보존하는 역할을 한다.

그림 5는 비발현 유전자를 가진 염색체의 구조를 보인 것으로 이 염색체 C 는 각각 p 개 유전자로 이루어진 발현부 $E(C)$ 와 비발현부 $U(C)$ 로 구성된다. 염색체를 평가할 때에는 비발현부가 무시되고 발현부만 반영되지만, 교차 연산을 통해 자손을 생성할 때에는 발현부와 비발현부가 모두 사용된다. 부모 염색체 P_1, P_2 가 선택된 경우, 자손 염색체 P_c 의 발현부 $E(P_c)$ 는 그림 4의 그리디 교차와 거의 동일한 교차 연산에 의해 생성된다. 다만, 부모 염색체의 발현부와 비발현부 전체 약 $4p$ 개의 서로 다른 열들 중 $E(P_c)$ 를 구성할 p 개만 선택하면 된다는 점이 다를 뿐이다. 비발현부 $U(P_c)$ 는 $E(P_c)$ 가 생성된 다음에 다시 P_1, P_2 로부터 $E(P_c)$ 와 유사도가 가장 낮은 열들을 우선적으로 선택하여 만들어진다.

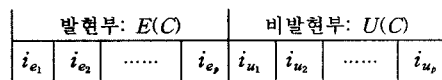


그림 5 비발현 유전자를 가진 염색체

그림 6은 자손의 비발현부를 만드는데 사용되는 그리디 교차 연산 과정을 보인 것이다. 그림 4의 교차 연산과 다른 점은 부모 염색체에 발현부와 비발현부가 모두 포함되어 있다는 것과 매 번 새로운 열 j 를 추가할 때 $\delta(j)$ 를 보지 않고 대신 $E(P_c)$ 와의 유사도인 $S_{E(P_c)}(j)$ 가 최소인 것을 고른다는 점이다. $S_{E(P_c)}(j)$ 는 수식 2와 유사하게 다음 수식 3과 같이 계산된다.

```

Algorithm GreedyCrossover-UnexpressedGene
input: parent chromosome  $P_1, P_2$ 
output: child unexpressed genes  $U(P_c)$ 
Begin
 $P := P_1 \cup P_2;$ 
 $U(P_c) := \emptyset;$ 
while  $|U(P_c)| \neq p$  Do
    find  $j$  which has minimum  $S_{EP_c}(j)$  where  $j \in P;$ 
     $U(P_c) := U(P_c) \cup \{j\};$ 
     $P := P - \{j\}$ 
Return  $U(P_c)$ 
End
    
```

그림 6 비발현부를 위한 교차 연산자

$$S_{EP_c}(j) = \sum_{v \in E(P_c)} j \cdot v = j \cdot \sum_{v \in E(P_c)} v \quad (3)$$

이렇게 $S_{EP_c}(j)$ 가 최소인 것을 우선적으로 선택하여 자손의 비발현부를 만드는 것은 이 자손들이 다음 세대에 부모가 되어 다시 그들의 자손을 생성할 때, 보다 다양한 열들 중에서 그 자손의 발현부가 선택될 수 있기를 기대하기 때문이다. 뒤에 실험 결과와 함께 설명되겠지만, 이러한 비발현 유전자의 도입은 세대가 변천되어 가더라도 집단의 다양성이 계속 유지되어 조기 수렴을 피할 수 있도록 하는데 크게 기여하게 된다.

4. K-exchange 돌연변이

본 논문에서 제안하는 돌연변이 연산은 타부 탐색을 이용하여 MCP를 해결한 기존 연구[4,5]에서 사용한 k -exchange 이웃해 생성 방법을 유사하게 따르고 있다. k -exchange 돌연변이 연산의 구체적 과정은 그림 7에 나타나 있다. 보통의 돌연변이가 개별 유전자에 대해 소정의 확률로 변이를 시도하는 것과는 달리 k -exchange 돌연변이에서는 각 염색체 단위로 소정의 확률로 변이를 시도한다. 일단 돌연변이를 하는 것으로 결정이 되면

```

Algorithm K-ExchangeMutation
input: chromosome  $C$ , integer  $k$ 
output: mutated chromosome  $C$ 
Begin
if MutationTest( $C$ ) = pass then
    RemoveGenes( $C$ ,  $k$ );
    InsertGenes( $C$ ,  $k$ );
end if
return  $C$ 
End
    
```

그림 7 k -exchange 돌연변이

대상 염색체로부터 k 개의 유전자를 제거한 다음 다른 k 개의 유전자를 다시 추가함으로써 돌연변이 연산이 완료된다. 단, 비발현부를 포함하고 있는 염색체에 대해서는 k -exchange 돌연변이를 발현부에만 적용한다.

염색체로부터 k 개의 유전자를 제거하는 과정은 그림 8에 설명되어 있다. 임의로 선택된 유전자 즉 열 j 의 제거는 확률적으로 이루어지는데 그 제거 확률은 수식 4와 같이 계산된 $v(j)$ 값을 기반으로 결정된다.

$$v(j) = Eval(C) - Eval(C - \{j\}) \quad (4)$$

여기서 $Eval(C)$ 는 앞의 2절에서 설명된 수식 1에서와 마찬가지로 열의 집합인 염색체 C 가 cover하는 행의 개수이다. 따라서, $v(j)$ 는 열 j 를 염색체로부터 제거할 경우 cover되는 행의 수가 얼마나 줄어드는지의 정도를 나타내게 된다. $v(j)$ 의 값이 크다는 것은 j 가 그만큼 중요한 역할을 하는 열이라는 의미이므로 제거 확률 P_r 을 낮추어 주고 $v(j)$ 의 값이 작을수록 P_r 이 커지도록 한다. 본 연구에서는 $v(j)$ 의 각 값에 대해 실험적으로 미리 결정해 둔 확률을 사용한다.

k 개의 유전자가 제거된 후에는 앞에서 설명한 그리디 교차 연산에 도입되었던 greedy-adding 휴리스틱을 사용하여 원래 문제에서 주어진 전체 열의 집합 $S = \{s_1, s_2, \dots, s_m\}$ 으로부터 k 개의 열을 선택하여 삽입한다. 전체 열의 집합 S 로부터 k 개의 열을 그리디하게 선택하는 과정은 $|S|$ 가 클 경우 상당히 많은 비용을 필요로 하지만 돌연변이 확률 자체가 매우 낮기 때문에 유전알고리즘 전체 수행 과정에 큰 부담으로 작용하지는 않는다.

비발현 유전자를 가진 염색체의 경우 발현부에 대해서는 이상과 같은 k -exchange 돌연변이를 적용하는 반

```

Algorithm RemoveGenes
input: chromosome  $C$ , integer  $k$ 
output: removed chromosome  $C$ 
Begin
while  $|C| > p - k$  Do
    select a random gene  $j$ ;
    calculate  $v(j)$ ;
    calculate remove probability  $P_r$  from  $v(j)$ ;
    remove  $j$  from  $C$  with probability  $P_r$ ;
Return  $C$ 
End
    
```

그림 8 RemoveGenes 알고리즘

면, 비발현부에 대해서는 보통의 돌연변이 방법을 적용한다. 즉, 개별 유전자에 대해 돌연변이 확률이 따라 대상 유전자를 전체 열 집합 S로부터 임의로 선택한 다른 열로 교체하는 것이다. 실험 결과 비발현부에 대한 돌연변이는 발현부의 경우보다 훨씬 높은 확률로 실시하는 것이 탐색 성능의 향상에 도움이 되는 것으로 나타났다.

5. 실험 결과

실험에 사용된 데이터는 국내 모 도시의 지하철 일일 승무일정계획 데이터이다. 지하철 일일 승무일정계획 문제는 고정된 숫자의 승무원을 하루에 운행할 전체 열차에 배정하는 문제로서, MCP로 모델링할 경우 각 행은 승무원이 담당할 최소 단위의 열차 운행 구간이고, 각 열은 한 승무원이 하루 동안 실제로 담당할 열차 운행 구간들의 집합으로 소위 말하는 일일 근무계획에 해당한다. 실험 대상 문제의 경우 $n = 814$, $m = 179,514$, $p = 83$ 로서 대략 18만개의 열중 83개를 뽑아서 814개의 행을 최대한 많이 cover해야 하는 MCP이다. 실험 데이터에서 각 열은 근무조건 제약에 따라 공히 10개의 행을 cover하게 되어 있다. 본 데이터의 경우에는 특별히 야간 근무와 관련된 18개의 열이 미리 결정되어 있기 때문에 실제로는 65개의 열들을 추가로 선택하여 나머지 행들을 cover하는 문제이다.

모든 실험은 Pentium IV 2G Hz, 1G RAM 서버 상에서 1시간의 시간제한을 두고 수행하였다. 유전 알고리즘은 비발현 유전자를 가진 것(GAUG)과 가지지 않은 것(GA)을 모두 구현하였다. GAUG의 경우 집단의 크기는 1,500으로 하였고 3-exchange 돌연변이를 사용하였으며 발현 유전자와 비발현 유전자의 돌연변이 확률은 각각 0.01과 0.1로 하였다. GA의 경우에는 비발현 유전자가 없으므로 인해 염색체의 길이가 절반 밖에 되지 않음을 감안하여 집단의 크기를 GAUG의 두 배인 3,000으로 하였고 역시 3-exchange 돌연변이를 사용하였으며 돌연변이 확률은 0.01로 하였다. 적합도 평가함수로는 선택된 열의 집합이 cover하는 행의 수를 사용했고, 비발현 유전자를 가진 염색체의 경우는 발현부에 포함된 열만 해의 평가에 반영하였다. 실험 결과 무작위로 생성한 초기 집단보다 greedy-adding 휴리스틱을 사용하여 생성한 비교적 좋은 초기 집단을 가지고 시작하는 것이 유리한 것으로 나타났다.

제한한 유전 알고리즘과의 성능 비교 대상으로 최적화 문제에 널리 활용되는 이웃해 탐색기법 중 하나인 타부 탐색을 구현하여 실험하였다. 타부 탐색에서도 유전 알고리즘의 초기 집단 생성 시와 마찬가지로 greedy-adding 휴리스틱을 사용하여 비교적 좋은 초기해를 생성하여 시작하였으며, 유전 알고리즘의 k-

exchange 돌연 변이와 유사한 k-exchange 이웃해 생성 방법을 사용하되 k의 범위 1부터 5 사이에서 골고루 이웃해들을 생성하였다. 특히, 100회의 iteration 동안 해의 개선이 없으면 탐색의 다각화[8]를 시도하게 함으로써 타부 탐색의 효율을 극대화하였다.

표 1은 유전 알고리즘과 타부 탐색을 매회 1시간의 시간제한을 가지고 10회 실험한 결과 여전히 cover되지 못한 행의 개수의 평균, 최대, 최소 값을 보이고 있다. 따라서, 이 수치가 작을수록 더 좋은 결과에 해당하는 것이다. 보통의 유전알고리즘은 타부 탐색에 비해 훨씬 좋지 못한 결과를 보이고 있지만 비발현 유전자를 사용할 경우 집단의 다양성이 유지됨으로 인해 그 성능이 타부 탐색을 뚜렷이 능가하는 수준으로 향상됨을 알 수 있다.

표 1 지하철 승무일정계획 문제 적용 결과

	최대	평균	최소
GAUG	10	9.0	8
GA	17	15.9	15
타부 탐색	12	10.7	9

그림 9는 탐색 시간에 따른 최적해의 변화 추이를 그래프로 나타낸 것이다. 두 유전 알고리즘이 모두 초기의 일정 기간 동안 해의 질을 개선시키지 못하는 것은 초기집단을 그리디 하게 생성했기 때문에 나타나는 현상이다. 초기해를 무작위로 생성하면 이와 같은 현상은 사라지게 되나 최종적인 해의 질은 그리디한 초기해를 가지고 시작하는 경우보다 못한 것으로 관찰되었다. 그림에서 보듯이 탐색의 초반에는 타부 탐색이 급격히 해의 개선을 이루고 있지만 얼마 지나지 않아 비발현 유전자를 사용하는 GAUG에 역전된 이후 따라 잡지 못하게 된다. 또한 비발현 유전자를 사용하지 않는 GA의 경우에도 초기 수렴 현상을 뚜렷이 보여주는 것을 확인할 수 있다. 이는 앞에서도 언급하였듯이 집단의 다양성이

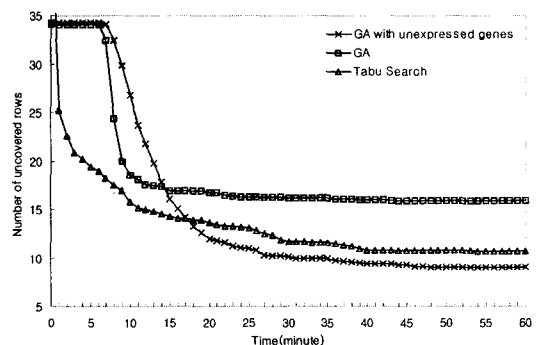


그림 9 시간에 따른 최적해 변화

급격히 떨어지는데 기인하는 것이다.

표 2는 본 논문에서 제안한 유전 알고리즘이 타부 탐색에 비해 일관되게 더 좋은 성능을 나타내는지 확인할 목적으로 추가 실험용을 하기 위해 인공적으로 생성한 다양한 규모의 MCP 데이터를 보여주고 있다. 표 3은 이 데이터에 비발현 유전자를 사용하는 유전 알고리즘과 타부 탐색을 적용시켜 본 결과를 정리한 것이다. 모든 문제에 대해 1시간의 시간 제약을 두고 10회씩 실험 하였으며, 결과 수치는 앞에서와 마찬가지로 cover되지 못한 행의 수이다. mcp453을 제외하고는 모든 경우 유전 알고리즘 GAUG가 더 좋은 결과를 보여 주고 있으며, 특히 문제의 규모가 가장 큰 mcp634에 대해서는 현격한 차이로 유전 알고리즘이 더 우수함을 확인 할 수 있다. 반면, 타부 탐색이 더 좋은 결과를 보여준 mcp453의 경우에는 사실 타부 탐색과 유전 알고리즘의 성능 차이가 극히 미미한 수준에 그치고 있다.

표 2 추가 실험을 위해 사용한 MCP 데이터

	행의 수	열의 수	p	각 열이 cover 하는 행의 수
mcp634	634	142265	65	10
mcp520	520	92139	53	9~10
mcp453	453	72094	45	9~10
mcp384	384	54738	42	8~9
mcp312	312	55807	39	7~8

표 3 추가 실험 결과

데이터	GAUG			타부 탐색		
	최소	평균	최대	최소	평균	최대
mcp634	4	6.3	8	8	8.8	10
mcp520	4	5.3	6	5	6.2	7
mcp453	12	12.3	13	11	12.0	13
mcp384	6	7.3	8	7	7.9	9
mcp312	2	2.7	3	3	4.0	5

타부 탐색에 비해 GAUG가 더 좋은 결과를 보여주는 이유는 다음과 같이 분석된다. 타부 탐색의 이웃해 생성 방법과 유전 알고리즘의 세대 교체를 위한 교차 연산은 다 같이 greedy-adding 휴리스틱에 기반을 두고 있다. 그러나 타부 탐색에서는 이웃해를 생성하기 위해 현재해로부터 k 개의 열을 제거한 후 18만 개에 가까운 전체 열들을 모두 검색하여 그 중 좋아 보이는 k 개의 이웃해를 선택하여 교체해야 하므로 계산시간의 부담이 상당하다. 이에 비해 유전 알고리즘에서는 새로운 자손을 생성하기 위해 단지 부모 염색체에 있는 약 $2p$ 또는 $4p$ 개의 열만을 대상으로 그 중 좋아 보이는 p 개만 선택하면 되므로 계산시간의 부담이 극히 적다. 말하자면 유

전 알고리즘은 좋은 열들의 집합을 선택하기 위해 다수의 소규모 문제를 병렬적으로 (집단이 다수의 후보해로 구성되므로) 푸는 방법을 사용하는데 비해, 타부 탐색은 대규모의 문제 하나를 연속적으로 반복하여 푸는 방법을 사용하는 점에서 다르다고 볼 수 있는 것이다. 유전 알고리즘 성공의 관건은 각 소규모 문제를 구성하는 약 $2p$ 혹은 $4p$ 개의 열들의 집합 즉, 부모 염색체들이 질적으로 우수하면서도 다양해야 한다는데 있다. 유전 연산을 이용한 세대 변천은 개별 염색체들이 우수해지도록 진화시켜 주며, 비발현 유전자는 이 과정에서 다양성을 잃지 않고 유지할 수 있게 하는 역할을 수행한다.

5. 결론 및 향후 연구 방향

본 논문에서는 대규모 MCP를 효율적으로 해결하기 위한 유전 알고리즘을 제안하였다. 제안된 유전 알고리즘에서는 MCP에 적합하도록 설계된 교차 연산자와 돌연변이 연산자를 사용한다. 특히, 비발현 유전자를 도입한 새로운 기법을 제시함으로써 집단의 다양성을 유지할 수 있도록 하여 보통의 유전 알고리즘이 가진 조기 수렴 문제를 해결하였다. 실제 현장의 대규모 MCP에 대한 실험 결과, 본 논문에서 제시한 진화 연산자들과 새로운 염색체 구조가 탐색의 효율화에 크게 기여함으로써 타부 탐색과 같은 이웃해 탐색 기법에 비해서도 더 우수한 탐색 성능을 보임을 확인하였다. 본 연구에서 제안한 새로운 염색체 구조가 MCP 외에도 다양한 대규모 최적화 문제의 해결에 도움을 줄 수 있을 것이라 생각된다. 향후 이러한 새로운 염색체 구조를 채택한 유전 알고리즘을 다양한 최적화 문제에 적용해 봄으로서 그 효과를 검증해 볼 필요가 있을 것으로 사료된다.

참고 문헌

- [1] B.T. Downs, J.D. Camm, *An Exact Algorithm for the Maximal Covering Problem*, Naval Research Logistics, 43, 435-461, 1996.
- [2] R.D. Galvao, C. ReVelle, *A Lagrangean Heuristic for the Maximal Covering Location Problem*, European Journal of Operational Research, 88, 114-123, 1996.
- [3] J.E. Beasley and P.C. Chu, *A Genetic Algorithm for the Set Covering Problem*, European Journal of Operational Research, 94, 392-404, 1996.
- [4] J.H. Hwang, C.S. Kang, K.R. Ryu, Y.H. Han, H.R. Choi, *A Hybrid of Tabu Search and Integer Programming for Subway Crew Scheduling Optimization*, IASTED-ASC, 72-77, 2002.
- [5] 황준하, 박춘희, 이용환, 류광렬, 정수계획법과 휴리스틱 탐색기법의 결합에 의한 승무원정체계획의 최적화, 한국정보과학회 논문지, 8(2), 195-205, 2002.

- [6] T. Kido, H. Kitano, M. Nakanishi, *A Hybrid Search for Genetic Algorithms: Combining Genetic Algorithms, TABU Search, and Simulated Annealing*, Proceedings of the Fifth International Conference on Genetic Algorithms, 641, 1993.
- [7] D. Levine. *Application of a hybrid genetic algorithm to airline crew scheduling*. Computers & Operations Research, 23(6), 547-558, 1996.
- [8] F. Glover, M. Laguna, *Tabu Search*, Kluwer Academic Publishers, 1997.
- [9] 박태진, 이용환, 류광렬, 유전 알고리즘을 이용한 Maximal Covering 문제의 해결, 한국지능정보시스템 학회 2002년 추계학술대회 논문집, 502-509, 2002.
- [10] J.E. Beasley, *An Algorithm for Set Covering Problems*, *European Journal of Operational Research*, 31, 85-93, 1987.
- [11] L.A.N. Lorena, L.S. Lopes, *Computational Experiments with Genetic Algorithms Applied to Set Covering Problems*, *Pesquisa Operacional*, 16, 41-53, 1996.



박 태 진

2001년 부산대학교 컴퓨터공학과 학사
 2001년~2003년 부산대학교 컴퓨터공학과 석사. 2003년~현재 부산대학교 컴퓨터공학과 박사과정. 관심분야는 인공지능, 최적화, 유전 알고리즘, 최적화, 기계학습, 데이터 마이닝 등



황 준 하

1995년 부산대학교 컴퓨터공학과 학사
 1997년 부산대학교 컴퓨터공학과 석사
 2002년 부산대학교 컴퓨터공학과 박사
 2002년 9월~현재 금오공과대학교 컴퓨터공학부 교수. 관심분야는 인공지능, 최적화, 일정계획, 기계학습 등

류 광 렬

정보과학회 논문지 : 소프트웨어 및 응용
 제 31 권 제 2 호 참조