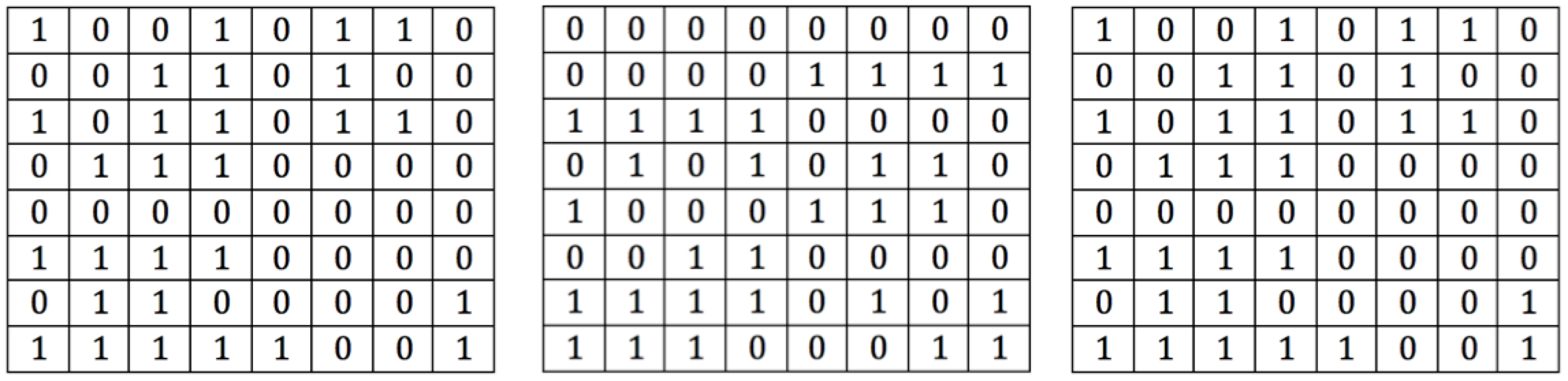
**2. 유전자 알고리즘을 활용한 빈티지 포인트 선택 방법**

**2.1 유전자 알고리즘을 위한 모델링**

최적 분포된 빈티지 포인트 집합을 생성하기 위해서 메타 휴리스틱 기법중에 하나인 유전자 알고리즘을 활용한다. 그 전에 용어를 정의한다.

**정의 1**. *n*차원에서 *m*개의 빈티지 포인트 집합 를 크기가 m인 유전자 *g*라 정의하고, *k*개의 유전자 집합 라고 정의한다. [그림 1]은 *k*=3, *n*=8, *m*=8일 때의 유전자 집합 G를 나타낸다.



[그림 1] *k*=3, *n*=8, *m*=8 일 때의 유전자 집합

[그림 1]에서는 3개의 행렬로 크기가 8인 3개의 유전자를 보여주고 있으며 각 행렬의 8개의 행은 8차원의 빈티지 포인트 8개를 나타낸다. 생성할 빈티지 포인트는 각 차원의 코너 포인트에서 선택하므로 0과 1로 구성되어 진다.

**정의 2**. *n*차원에서 크기가 *m*인 유전자 *g*의 성능을 평가하기 위한 적합도 함수 는 빈티지 포인트 집합이 최적 분포 상태에 가까울수록 우수한 해로 평가되므로 다음 식 (1)과 같다.

빈티지 포인트들이 얼마나 고르게 분포하고 있는지를 측정하기 위하여 빈티지 포인트 집합에서, 상호간 거리를 이용하여 편향성을 측정한다. 만약 전체 공간의 크기에 비해 너무 작은 거리를 갖거나 너무 큰 거리를 갖는 빈티지 포인트 쌍이 존재한다면 이는 고르게 분포하지 않은 빈티지 포인트 집합이다[1]. 따라서 공간상에서 발생할 수 있는 상호간 거리의 평균값인을 기준으로 하여 편차가 적은 상호한 거리를 갖는 빈티지 포인트 집합일 수록 최적 분포이고, 적합도 함수 를 식 (1) 로 정의한다.

유전자 알고리즘을 수행하기에 앞서, 초기 유전자 집합 G를 생성한다. 이 때, 초기 세대에서 우수한 해로 시작하기 위하여 휴리스틱 방법과 무작위 방법을 이용한다. 세대 교체를 진행할 때에는, 한번의 세대에서 선택, 교차, 변이 세 가지 연산이 수행되고, 원하는 성능이 최적 분포에 근접했을 때, 해당 세대에서 종료한다.

**2.2 효율적인 세대 교체 방법**

**2.2.1 초기 유전자 집합 구성 방법**

2의 제곱 차원에서는에 대한 편차가 0인 최적 분포를 가지는 빈티지 포인트 집합이 존재한다[1]. 모든 자연수는 서로 다른 2의 제곱수의 합으로 표현할 수 있으므로, 저차원의 최적 분포를 갖는 빈티지 포인드 집합들을 합하여 *n*차원의 초기 유전자를 구성한다. 이 유전자의 칼럼을 기준으로 shift 하여도 적합도 함수값은 동일하므로, 이 방법을 통해 생성한 적합도가 높은 유전자와 무작위 방법으로 유전자를 생성하여 초기 유전자 집합 *G*를 생성한다.

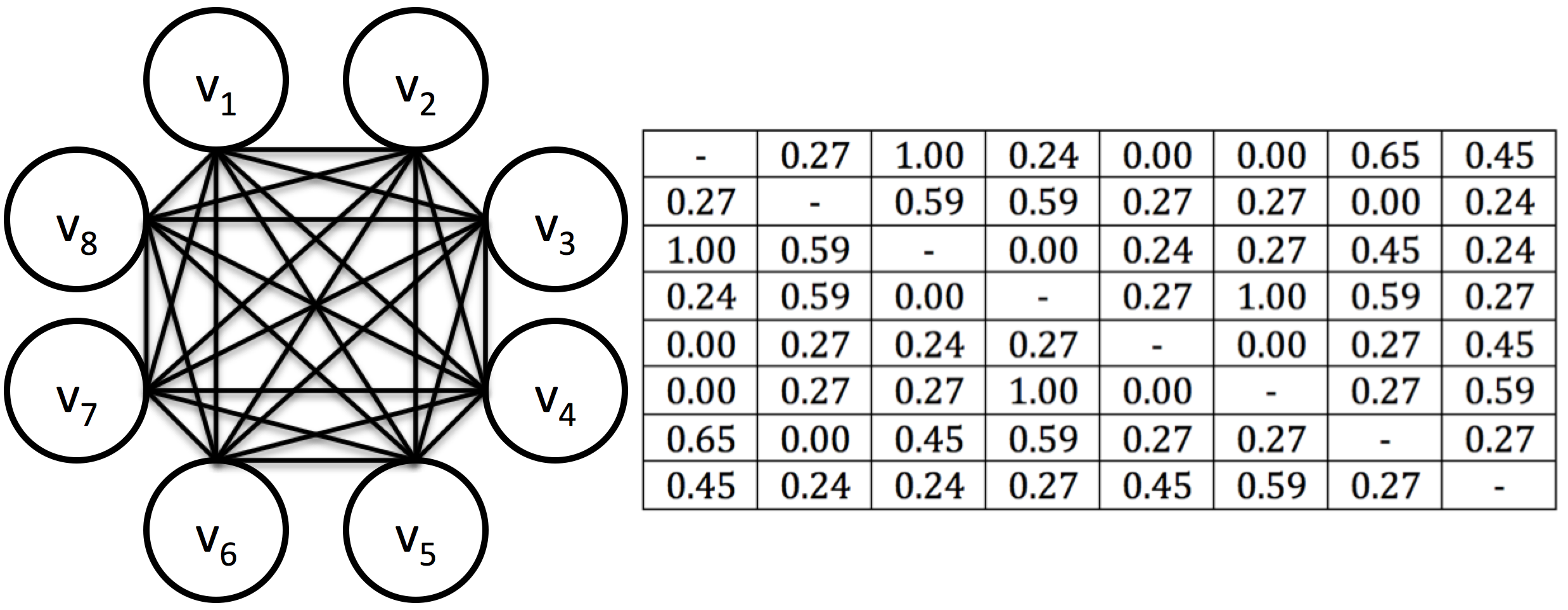
**2.2.2 선택**

엘리트 보존 전략을 사용하였고, 우성 유전자에게는 다른 모든 유전자들과 교차할 수 있는 기회를 준다. 우성 유전자는 교차 후 적합도가 교차 전 적합도보다 높지 않다면, 교차하지 않고 다음 세대로 보존하여 반드시 생존할 수 있도록 한다. 그 외 유전자들은 교차할 유전자가 남아있을 때까지 탐욕적 알고리즘 방법으로 항상 적합도가 높은 두 유전자를 선택한다.

**2.2.3 교차**

교차는 두 부모가 갖는 유전자를 조합하여 자손을 생성하는과정으로, 선택된 두 유전자 의 적절한 교차점을 선정하는것이 가장 중요하다. 하나의 유전자에서 적절한 교차점을 찾는 방법은 유전자에 있는 빈티지 포인트들을 두 집합 로 나누는데, 각각의 적합도가 가장 높게 분할하는 것이다.

먼저 유전자 을 고려해보면 의 빈티지 포인트들을 정점으로 하고, 빈티지 포인트들간의 거리들의 에 대한 절대 편차를 가중치로 하는 무방향 완전 그래프를 만들 수 있다.



[그림 2] [그림 1]의 첫번째 유전자로 구성한 무방향 완전

그래프와 간선들의 가중치

이 그래프를 가질 수 있는 최적의 적합도를 갖는 두 집합으로 분할하기 위해서 다음의 특성을 정리한다.

**정리 1**. 유전자 *g*가 나타내는 빈티지 포인트 집합을 두 집합 로 분할할 때, 유전자 *g*로 구성한 무방향 완전 그래프에서 Max Cut으로 분할한 두 집합 의 적합도 평가 가 가장 높다.

(증명) 유전자 *g*에서 만든 두 집합을 , 라고 하자. 먼저 함수 를 두 점 와 사이 거리의 에 대 한 절대 편차로 정의한다.

또한, 집합의 절대 편차를 라 하고 집합 의 절대 편차를 라 했을 때 나누기 전의 편차 D는 다음 식과 같다.

편차 D는 일정하므로의 크기를 낮추기 위해서는의 크기가 커야 한다. 따라서 Max Cut으로 분할한 두 집합 의 적합도 평가가 가장 높음을 알 수 있다.

이 특성을 이용해서 유전자 을 두 집합 로 분할 하고, 유전자 를 두 집합 로 분할한다. 이제 분할 된 두 집합을 교차시켜서 새로운 유전자 과 를 만드는데, Max Cut을 해서 분할된 두 집합의 크기는 일정하지 않다. 따라서집합과 집합의 크기가 다를 수 있으므로 항상 교차할 수는 없다. 이 때, 항상 교차할 수 있도록 원하는 집합의 크기로 분할하기 위하여 다음의 특성을 정리한다.

**정리 2**. 크기가 *m*인 유전자 는 를 만족하는 크기가 와 인 유전자 과 로 최대 번 Max Cut을 적용하여 구할수 있다.

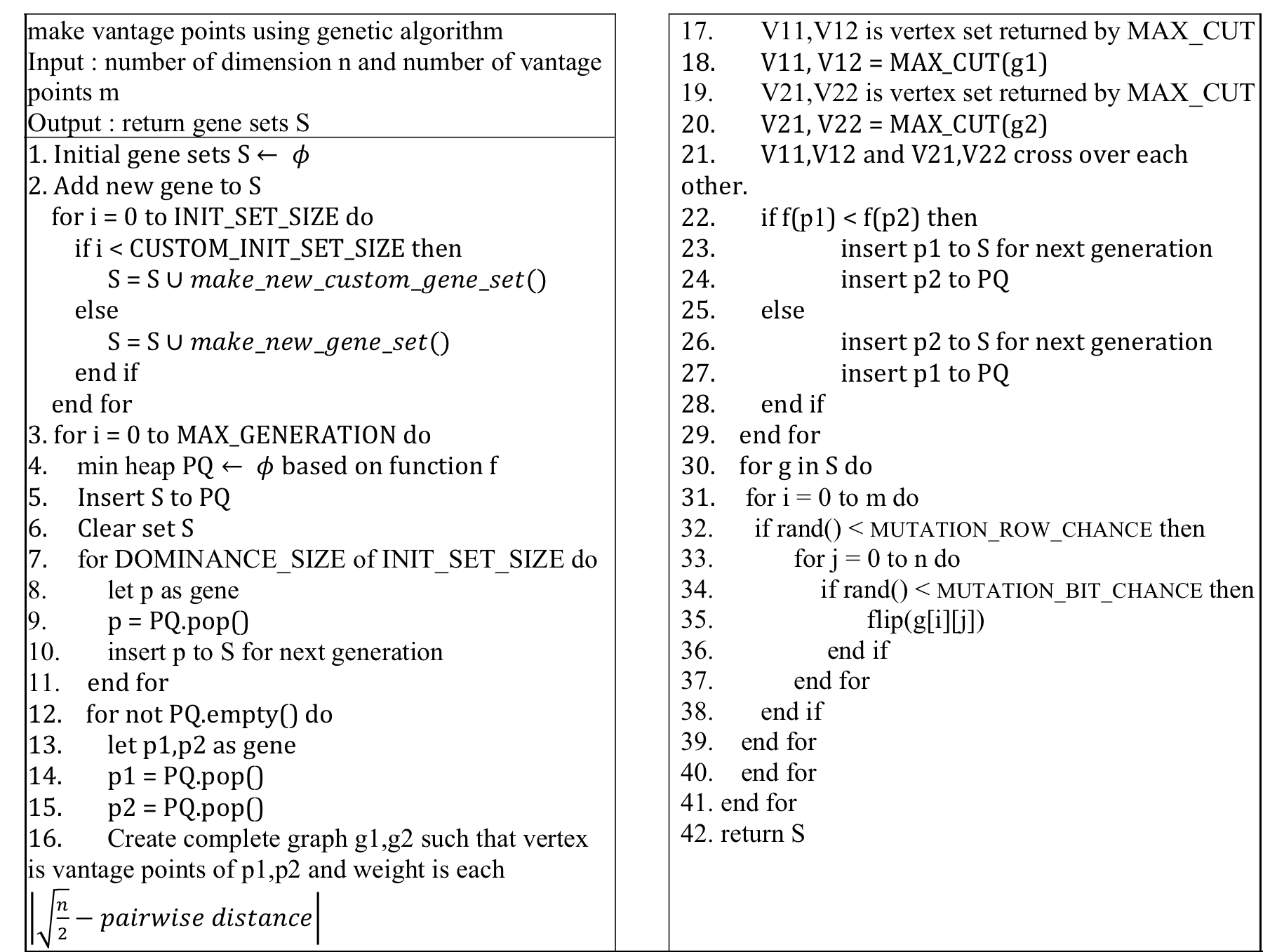
(증명) 유전자 의 총 빈티지 포인트의 개수를 라고 하고, 분할 할 집합의 크기를 라고 하자. 그래프를 구성하고 Max Cut을 유한번해서 분할 한다면, 최종적으로 더이상 분할 할 수 없는 상태의 집합은 하나의 빈티지 포인트를 가진 집합이다. 더이상 분할 할 수 없는 상태의 집합의 크기를 전부 더하면 총 빈티지 포인트의 개수 이고, 이는 보다 크므로 항상 원하는 크기의 집합으로 분할 할 수 있다.

따라서 선택된 두 유전자를 Max Cut을 이용하여 같은 크기의 집합으로 분할 한 다음 교차하여 다음 세대의 유전자를 만든다.

**2.2.4 돌연변이**

돌연변이는 개체에 새로운 유전자를 추가하는 것으로, 한 유전자에서 아주 작은 확률로 임의의 비트를 변형하여 더 많은 빈티지 포인트를 탐색할 수 있게 하는 방법이다. 크기가 *m*인 유전자 *g*에서 하나의 빈티지 포인트가 바뀌었을 때, 새로 생기는 쌍의 개수는 -1개 이다. 돌연변이 시킬 행을 제한하지않고, 무분별하게 돌연변이를 시킨다면 모든 빈티지 포인트 쌍이 새로 생기므로, 아주 작은 확률로 행을 선택하고 선택한 행에서 아주 작은 확률로 열을 선택하여 돌연변이 시킨다.

**2.3 유전자 알고리즘의 의사코드**

2.2에서 제시한 효율적인 세대 교체의 방법을 적용한 유전자 알고리즘의 의사코드는 [그림 3]과 같다. 먼저 휴리스틱 방법과 무작위 생성 방법을 이용하여 초기 유전자 집합 S를 생성한다. 이후 세대 교체를 진행하게 되는데, 유전자 집합 S에서 적합도가 높은 유전자를 효율적으로 선택할 수 있도록, 적합도를 우선으로 하는 최소 힙을 구성한 뒤 선택을 수행한다. 선택 된 두 유전자를 이용하여 각각 무방향 완전 그래프를 만들고, [2] 에서 제시된 Max Cut 알고리즘을 이용하여 분할한 다음 교차한다. 그 후, 아주 작은 확률로 행을 선택한 다음 아주 작은 확률로 열을 선택하여 돌연변이를 수행한다. 

[그림 3] 빈티지 포인트 생성 유전자 알고리즘

**2.4 성능 평가**

**2.4.1 유전 파라메터**

본 연구에서 사용한 유전 파라메터는 기존 유전 알고리즘 연구에서 제시된 모집단의 수, 종료 세대 수와 우성 유전자의 비율, 행의 돌연변이 비율과 열의 돌연변이 비율을 설정하였다. 여러 파라메터를 적용한 반복된 실험을 통하여 알고리즘과 문제의 특성에 맞는 파라메터를 적용하였다. 본 연구에서 적용한 파라메터는 다음 [표 1]과 같다.

|  |  |
| --- | --- |
| 파라메터 | 값 |
| Number of population | 100 |
| Number of generation | 10000 |
| Dominance rate | 0.1 |
| Mutation row rate | 0.1 |
| Mutation column rate | 0.3 |

[표 1] 유전 파라메터

2.4.2 실험 결과 및 성능 평가

이 장에서는 제안된 유전자 알고리즘의 성능 평가를 위하여 [1,3] 의 Spacing Variance와 Correlation Coefficient를 기준으로 측정한다. Spacing Variance는 공간 변환 후 거리의 편차를 나타내고, Correlation Coefficient는 데이터 객체들과 빈티지 포인트 쌍의 유사도를 나타내기 때문에, 두 측정 기준 모두 0에 가까울수록 더 좋은 분포를 나타낸다.

|  |
| --- |
| Macintosh HD:Users:sisobus:Desktop:스크린샷 2015-02-02 오전 11.41.44.pngMacintosh HD:Users:sisobus:Desktop:스크린샷 2015-02-02 오전 11.42.07.png |

[그림 4] 차원 수의 증가에 따른 m개의 빈티지 포인트의 Spacing variance와 Correlation Coefficient

[그림 4]는 제안된 기법의 빈티지 포인트와 [3]의 데이터 기반 빈티지 포인트의 측정 결과이다. 두 측정 모두 제안된 기법의 빈티지 포인트가 우수함을 보인다.

참고문헌

[1] S. Pramanik, A. Watve, S. Jung, and C. Lim, “Database Independent Vantage Point Selection for Range Queries”, Technical Report, MSU, 2014.

[2] Michel X. Goemans and David P. Williamson, “Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming”, Journal of the ACM., 42:6:1115-6:1145, Nov 1995.

[3] R. H. Van Leuken and R. C. Veltkamp, “Selecting vantage objects for similarity indexing”, ACM Trans. Multimedia Comput. Commun. Appl., 7:16:1–16:18, September 2011.