

불연속면 군 분류를 위한 유전자알고리즘의 응용

정용복^{1)*}, 선우춘¹⁾

The Application of Genetic Algorithm for the Identification of Discontinuity Sets

Yong-Bok Jung and Choon Sunwoo

Abstract One of the standard procedures of discontinuity survey is the joint set identification from the population of field orientation data. Discontinuity set identification is fundamental to rock engineering tasks such as rock mass classification, discrete element analysis, key block analysis, and discrete fracture network modeling. Conventionally, manual method using contour plot had been widely used for this task, but this method has some shortcomings such as yielding subjective identification results, manual operations, and so on. In this study, the method of discontinuity set identification using genetic algorithm was introduced, but slightly modified to handle the orientation data. Finally, based on the genetic algorithm, we developed a FORTRAN program, Genetic Algorithm based Clustering(GAC) and applied it to two different discontinuity data sets. Genetic Algorithm based Clustering(GAC) was proved to be a fast and efficient method for the discontinuity set identification task. In addition, fitness function based on variance showed more efficient performance in finding the optimal number of clusters when compared with Davis - Bouldin index.

KeyWords discontinuity set identification, GA(Genetic Algorithm), fitness, DBI, VI, orientation matrix

초 록 암반 불연속면의 조사 및 분석 과정에서 거쳐야할 필수적인 단계 중 하나는 방대한 불연속면 자료로부터 군을 판별하는 것이다. 불연속면 군 분류는 암반분류, 키블록 해석, 개별요소해석 및 불연속면결합 생성과 같은 암반공학적 업무에 있어서 필수적이다. 일반적으로 등고선도를 이용한 수작업 군 분류가 적용되었으나 이 방법은 수작업에 의존한 주관적인 결과를 제공한다는 단점이 있다. 본 연구에서는 유전자알고리즘을 이용한 불연속면 군 분석기법을 도입하였으며 방향성 자료에 적용하기 위해 기본적인 유전자알고리즘을 변경하였다. 최종적으로 이러한 이론을 적용한 FORTRAN 프로그램 GAC를 개발하였으며 두 가지 형태의 불연속면 자료의 군 분석에 적용하였다. 적용 결과 GAC를 적용한 군 분류는 빠르고 효율적인 군 분석방법임을 확인하였으며 최적의 불연속면 군 수를 결정하는 데 있어서 분산에 근거한 적합도 함수가 Davis-Bouldin 지수에 근거한 적합도 함수보다 효율적인 것으로 나타났다.

핵심어 불연속면 군 분류, 유전자알고리즘, 적합도, DBI, VI, 방향행렬

1. 서 론

암반구조물의 안정성은 주로 불연속면의 방향 및 역학적 특성에 좌우된다. 따라서 불연속면의 조사 및 분석 작업은 아주 중요하며 이 가운데 필수적으로 수행해야 할 작업이 불연속면의 방향에 근거하여 불연속면 군을 판별하는 것이다.

기존의 불연속면 군 판별작업은 평사투영도를 이용한 도식적 방법을 주로 사용하였으며 최근에 통계 및 퍼지

클러스터링 기법 등을 사용한 불연속면 군 분석 자동화가 이루어져왔다(Hammah & Curran, 1998; Marcotte & Henry, 2002; 정용복 과 전석원, 2003). 도식적 방법의 경우 주관적인 결과가 나올 가능성이 있으며 통계 기법의 경우 사전에 불연속면의 방향 분포에 대한 가정이 이루어져야 한다는 단점이 있다(Hammah & Curran, 1998). 또한 퍼지 기법의 경우 전역적 해를 구하는 것이 아니라 국부적 해를 구하는 기법이므로 전역해를 구하지 못할 가능성이 항상 있다는 단점이 있다.

따라서 본 연구에서는 전역적 해를 구할 가능성이 타 기법보다 많은 유전자알고리즘(GA)을 적용한 불연속면 군 분석 프로그램을 개발하였으며, 유전자알고리즘의 핵심 제어변수인 후보 크기, 교차확률 및 돌연변이

¹⁾ 정회원, 한국지질자원연구원 지반안전연구부

* 교신저자 : ybjung@kigam.re.kr

접수일: 2004년 12월 14일

심사 완료일: 2005년 1월 12일

확률에 따른 탐색 능력을 군 수 및 평균(중심)방향을 사전에 알고 있는 모의 자료를 사용하여 비교 및 검토하였다. 또한 적합도 함수(불연속면 군 수를 결정하는 기준)는 일반적인 클러스터링 분야에서 많이 사용되는 Davies-Bouldin Index(DBI)와 절리군의 평균방향에 대한 각 자료의 분산값에 근거한 VI를 사용하였으며 최종 결정된 제어변수를 사용하여 절개사면 현장에서 조사한 불연속면 자료에 대하여 절리군 분석을 실시하였다.

2. 유전자알고리즘

2.1 진화알고리즘

진화알고리즘(Evolution Algorithm, EA)은 자연세계의 진화 과정을 모사함으로써 복잡한 실제계의 문제를 해결하고자 하는 계산 모델이다. 진화알고리즘은 구조가 간단하고 방법이 일반적이어서 응용범위가 매우 넓다. 특히 적응적 탐색과 학습 및 최적화를 통한 공학적인 문제의 해결에 많이 이용되고 있다. 또한 많은 다양한 탐색 알고리즘 중 진화알고리즘은 전통적인 방법에 비해 비선형적이고 미분이 불가능하거나 힘들며 많은 극값을 가지는 문제에 좋은 성능을 보여 왔다. 또한 특정 문제에 대한 사전 지식을 필요로 하지 않으므로 많은 문제에 효과적이고 일반적으로 적용될 수 있는 장점을 가지고 있다.

이러한 진화 알고리즘에는 유전자 알고리즘(Genetic Algorithm, GA), 진화전략(Evolution Strategies, ES), 진화프로그래밍(Evolution Programming, EP), 유전자 프로그래밍(Genetic Programming, GP) 등이 있다. 본 연구에서는 문자열이나 벡터열 탐색에 주로 응용되는 유전자알고리즘을 사용하였다.

2.2 유전자알고리즘

유전자알고리즘은 생물의 진화과정을 인공적으로 모델링한 알고리즘으로써 John Holland(1975)에 의해 창안되었다. 자연계에 있는 생물의 진화과정 즉 어떤 세대를 형성하는 개체들의 집합을 개체군(population)이라 하며 개체군중에서 환경에 대한 적합도(fitness)가 높은 개체가 높은 확률로 살아남아 재생할 수 있게 된다. 이때, 교배(crossover) 및 돌연변이(mutation)로서 다음 세대의 개체군을 형성하게 된다. 이처럼 유전자알고리즘은 생물의 진화과정을 모델링한 알고리즘이다.

이러한 일반적인 유전자알고리즘의 구현 절차는 다음과 같다(조영임, 1999).

1 단계 : 초기집단생성

- 탐색공간 상에서 원하는 개수의 초기집단을 생성시킨다.

2 단계 : 적합도 평가

- 생성된 모집단(Parent population)에 대한 적합도(Fitness)를 계산한다.

3 단계 : 자손의 생성

- 2 단계에서 계산된 적합도를 기준으로 여러 가지 선택방법을 이용하여 모집단에서 두개의 개체를 선택한 후 교배(Crossover)와 돌연변이(Mutation) 연산을 통하여 자손을 생성시킨다.

4 단계 : 종료

- 목적함수(Objective function)의 적합도가 원하는 수준을 만족할 때까지 또는 미리 지정한 최대 반복횟수에 도달할 때까지 2, 3단계를 반복한다.

이러한 유전자알고리즘은 한 점이 아닌 다점 군 탐색 방법이며 탐색에 적합도 함수를 이용한 Blind 탐색(미분값이나 다른 부가적인 지식을 요구하지 않는다)을 수행한다. 또한 결정론적인 규칙이 없고 확률적 연산자를 사용하며 다른 탐색 또는 최적화 방법에 비하여 전역적해를 구할 가능성이 높으며 다른 여러 탐색 방법에 비하여 효율적이다.

유전자알고리즘의 기본 구성요소는 개체표현방법, 연산자, 적합도 함수 및 알고리즘 제어변수이며 본 연구에서는 이러한 구성요소들을 절리와 같은 방향성 자료의 처리에 적합하도록 수정하였다(Treleaven, J. F., 1994).

3. GA를 이용한 클러스터링

유전자알고리즘을 적용한 불연속면 군 분석 과정을 Fig. 1에 도시하였으며 주요 항목의 구체적 구현 절차는 다음과 같다.

3.1 개체표현

본 연구에서 염색체는 각 군의 중심방향을 표현하는 실수로 표현된다. 군의 수는 최소 2(K_{min})개에서 최대 5(K_{max})개까지 가질 수 있으며 한 개체의 크기는 K_{max} 이다. 한 개체내의 각 염색체는 실제 평균방향일 수도 있고 널(null) 값일 수도 있다.

3.2 연산자

유전자알고리즘에 사용되는 연산자는 선택, 교배 및 돌연변이이다. 이러한 연산자가 개체군에 대해서 지정한

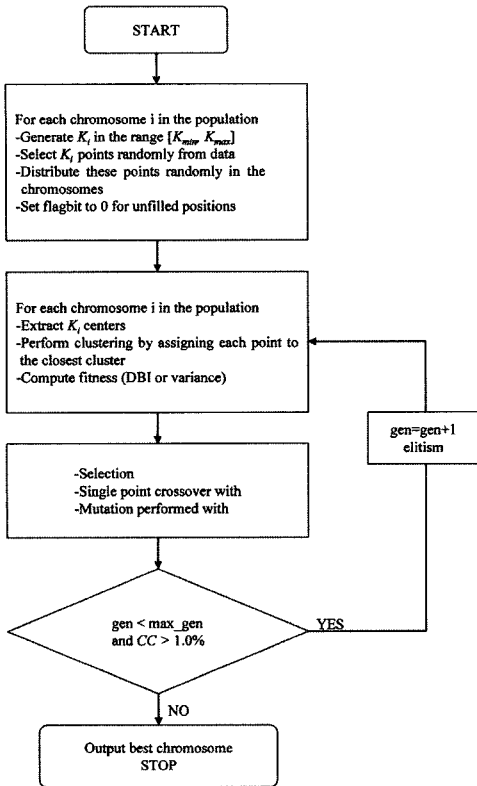


Fig. 1. Flowchart of GAC program

횟수(세대)만큼 적용된다.

선택 - 기존의 비례선택법을 적용하였다. 즉, 개체군 중에서 적합도가 클수록 교배연산에 선택될 확률이 높아진다.

교배 - 일점교배법이 교배확률(P_c)에 따라 임의로 발생한다.

돌연변이 - 각각의 유효한 염색체는 돌연변이확률(P_m)에 따라 다음과 같은 방식으로 변이를 일으킨다. 두개의 난수가 각각 [0,360]과 [0,90]범위에서 발생하며 원래의 방향자료에 더해진다. 이 값을 각각 360과 90으로 나누어 나머지가 원래의 값에 할당된다.

3.3 적합도함수

본 연구에서는 적합도함수로서 일반적인 클러스터링 기준으로 많이 사용되는 DBI(Maulik & Bandyopadhyay, 2002)와 새로운 개념의 VI 지수를 사용하였다.

DBI는 군 내부의 분산과 군과 군 사이의 분리정도의 함수로서 DBI가 작을수록 클러스터링이 최적화되었다는 의미이다. 따라서 본 연구에서는 DBI의 역수를 적합도로 사용하였으며 DBI의 정의는 다음과 같다.

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{j, j \neq i} \left(\frac{S_n(C_i) + S_n(C_j)}{d(C_i, C_j)} \right) \quad (1)$$

여기서 n 은 군 C_i 에 속하는 자료 개수이며 S_n 은 군 C_i 의 평균방향과 각 자료 사이의 평균거리, d 는 두 벡터 사이의 거리를 표시하는 함수이다. 군 C_i 와 군 C_j 간의 거리는 다음 식을 사용하여 계산하였다.

$$d(C_i, C_j) = \sqrt{1 - (z_i \cdot z_j)^2} \quad (2)$$

여기서 z_i, z_j 는 군 C_i, C_j 의 평균방향을 표시하는 단위 벡터이다. VI는 평균방향에 대한 각 자료의 거리에만 근거한 것이며 다음과 같이 정의된다.

$$VI = \left(\sum_{k=1}^K \sum_{x \in C_k} d(x, z_k) \right)^{-1} \quad (3)$$

3.4 수렴기준

본 연구에서는 최대 500번의 세대반복이 일어나며 이 과정 내에서 개체군 중 최대적합도와 평균적합도의 차이(CC)가 1%미만일 경우 연산을 종료한다. 또한 각 반복에서 최대적합도를 가지는 개체는 따로 저장하여 최종 연산이 끝난 경우, 최대적합도를 가진 개체가 해(solution)로서 선택된다. 수렴기준으로 사용된 CC는 다음과 같이 정의된다.

$$CC = \left| 1 - \frac{fitness_{best}}{fitness_{average}} \right| \times 100(\%) \quad (4)$$

3.5 방향행렬

각 불연속면 군의 중심방향은 방향행렬을 사용하여 계산하였다. 군 i 에 대하여 N 개의 방향코사인인 주어진 경우 방향행렬, S_i 는 다음과 같이 정의된다(Markland, 1973).

$$S_i = \begin{bmatrix} \sum_{j=1}^N x_j x_j & \sum_{j=1}^N x_j y_j & \sum_{j=1}^N x_j z_j \\ \sum_{j=1}^N x_j y_j & \sum_{j=1}^N y_j y_j & \sum_{j=1}^N y_j z_j \\ \sum_{j=1}^N x_j z_j & \sum_{j=1}^N y_j z_j & \sum_{j=1}^N z_j z_j \end{bmatrix} \quad (5)$$

이러한 방향행렬의 고유치 및 고유벡터 중 최대고유치에 대응하는 고유벡터가 평균방향의 추정치가 된다.

이러한 방법은 각 벡터성분별 합을 통해 평균방향을 구하는 기존의 방법과 비교할 때 방향코사인인 부호에 대해서 고려하지 않아도 되는 장점이 있다.

4. GAC 적용

4.1 제어변수에 따른 탐값 추정 능력

GA의 경우 개체군 크기(PS , population size), 교배확률(P_c , probability of crossover) 및 돌연변이확률(P_m , probability of mutation)에 따라 탐색능력이 좌우된다. 일반적으로 교차확률과 돌연변이확률이 클 경우 탐색능력을 향상시켜 초기단계에서 적합도가 높은 탐색공간을 찾는데 유리하지만 동시에 어느 정도 좋은 해를 찾은 후에는 이 탐색공간 내에서 최적해로 수렴하는데 있어서 수렴속도의 저하를 초래한다. 작은 교차 및 교배확률의 경우 이와는 반대의 특성을 가진다. 또한 개체군의 크기가 작을 경우 적합도 계산에 필요한 시간이 절약되나 개체군 다양성의 손실로 최적해를 구하기 전

에 수렴할 가능성을 내포한다. 반면에 개체군의 크기가 클 경우 최적해에 도달할 확률은 높아지나 메모리 및 계산시간이 증가한다. 따라서 GA를 이용해 문제를 풀고자 할 경우 문제의 특성에 따라 이러한 제어변수를 결정할 필요가 있다.

본 연구에서는 기지의 군 개수와 평균방향을 가지도록 DIPS 프로그램에 포함된 JTDist를 사용하여 모의 자료를 생성하였으며 최적의 제어변수를 구하고자 Table 1과 같이 제어변수를 변화시키면서 총 72가지 경우에 대해서 모의 자료에 대한 군 분석을 수행하였다. 총 72회 분석 결과 제어변수의 값에 관계없이 DBI 및 VI 모두 3개의 군으로 분석하였으며 적합도 값도 동일하였다. 이는 각각의 군이 시각적으로도 명확하게 구별이 되기 때문에 동일한 결과를 얻은 것으로 판단된다. 다만 각 조건별로 반복회수에 있어서는 차이를 보였으며 DBI의 경우 평균 352회, VI의 경우 256회로 VI적합도 기준을 사용할 경우 수렴속도가 약간 빨랐다. Table 2는 분석 결과 구한 3개의 군에 대한 경사/경사방향이며

Table 1. The variation of GAC control parameters

Control variables	Values				Remark
Population size (PS)	50	100	150	-	3x4x3x2 = 72 cases
Crossover probability (P_c)	0.2	0.4	0.6	0.8	
Mutation probability (P_m)	0.001	0.005	0.01	-	
Fitness function	DBI	VI	-	-	

Table 2. Statistics of synthetic joint sets and estimated results

Set	True mean direction(DIP/DD*)	Estimated mean direction(DIP/DD*)	No. of data
1	60/120	61/120	100
2	25/220	26/219	120
3	85/65	83/66	130

*DD : Dip direction

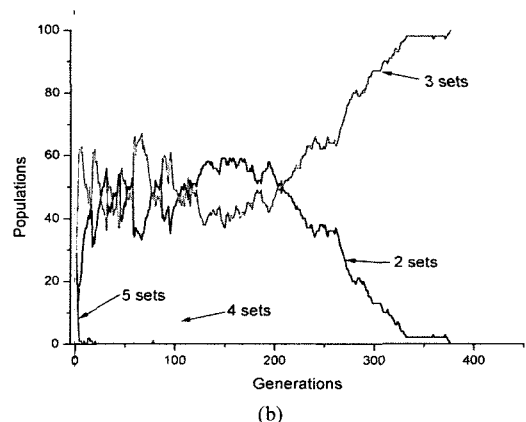
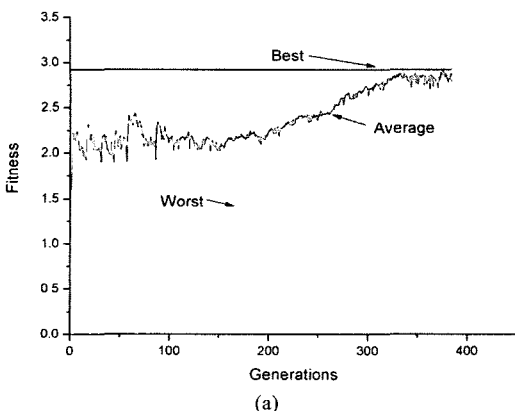


Fig. 2. The variation of (a) fitness and (b) cluster number during evolution($P_c = 0.6$, $P_m = 0.01$, $PS = 100$, DBI)

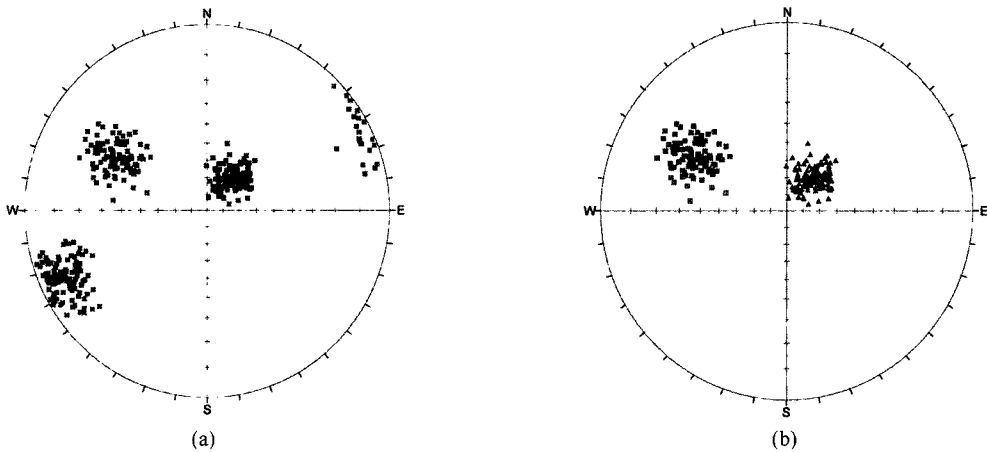


Fig. 3. The lower hemispherical projection of (a) original joint data and (b) clustered joint data($P_C = 0.6$, $P_m = 0.01$, $PS = 100$, DBI)

각각 2도 이내에서 참값을 추정할 것을 알 수 있다. Fig. 2는 계산과정 중 개체군 내의 군 개수 및 적합도의 변화를 표시한 것이다. 초기에는 군 개수가 2에서 5까지 다양하지만 대부분 소멸하고 200세대까지 2개 및 3개의 군을 가진 개체끼리 경쟁하다가 이후 3개의 군을 가진 개체가 우세하게 됨을 알 수 있다. 적합도의 경우 평균치가 최대적합도에 점진적으로 수렴하고 있음을 알 수 있다. Fig. 3은 군 분석결과를 표시한 것이며 참값과 동일하게 군 분석이 이루어졌음을 알 수 있다.

4.2 현장 적용

현장 적용성을 검토하기 위해서 고속도로 절개사면에서 조사한 148개의 불연속면 자료에 대하여 군 분석을 수행하였다(Fig. 4). 제어변수는 4.1의 결과 및 일반적

인 관례를 기초로 하여 개체군 크기는 100, 교차확률은 0.6, 그리고 돌연변이 확률은 0.01로 정하여 분석을 수행하였다. 적합도 함수는 DBI 및 VI 를 모두 적용하였다.

분석 결과 두 가지 적합도 함수에 관계없이 모두 동일하게 5개의 군으로 분류하였으며 각 군의 평균방향도 전문가가 결정한 값과 3도 이내의 오차를 보여 양호한 결과를 얻었다.

최종 군 분류 결과는 동일하였지만 Fig. 5에 표시한 것과 같이 DBI 를 적합도 함수로 사용하였을 경우 개체간 경쟁이 전체 세대에 걸쳐 계속되었으나 VI 를 사용하였을 경우에는 세대 초기부터 5개의 군을 가진 개체가 급격히 증가하였다. 또한 DBI 를 사용한 경우 평균적합도도 최대적합도에 수렴하지 않고 거의 일정한 값을 보였으나 VI 를 사용한 경우 평균적합도가 점진적으로 최

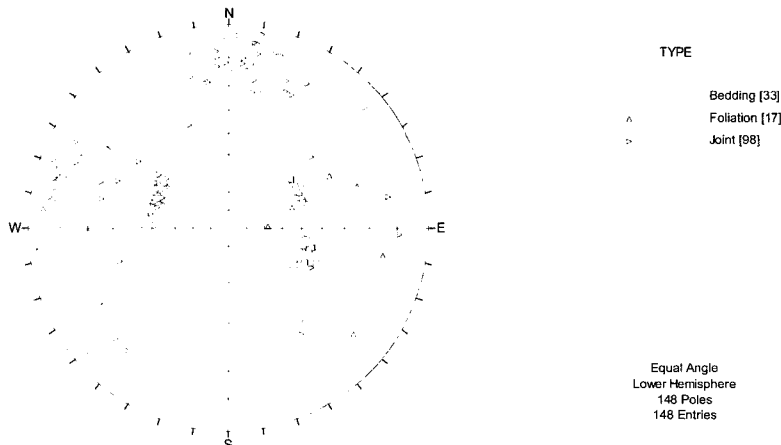


Fig. 4. The distribution of field surveyed orientation data

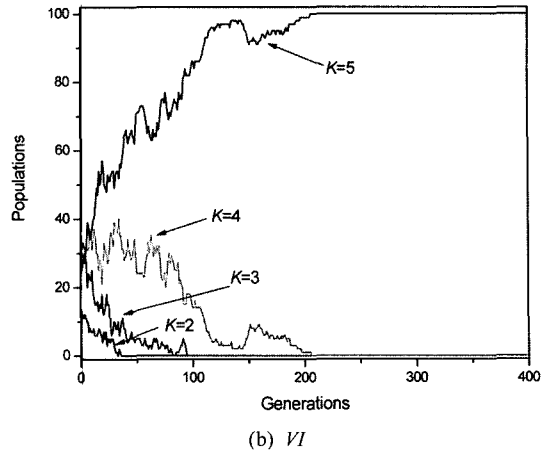
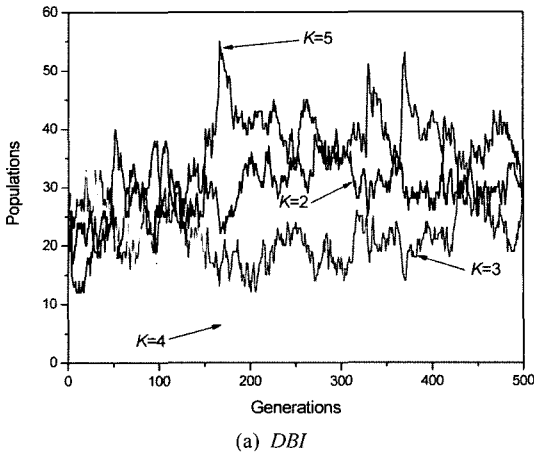


Fig. 5. The variation of cluster numbers during evolutions

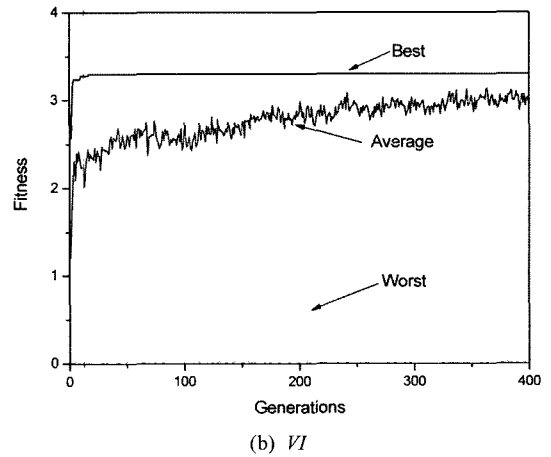
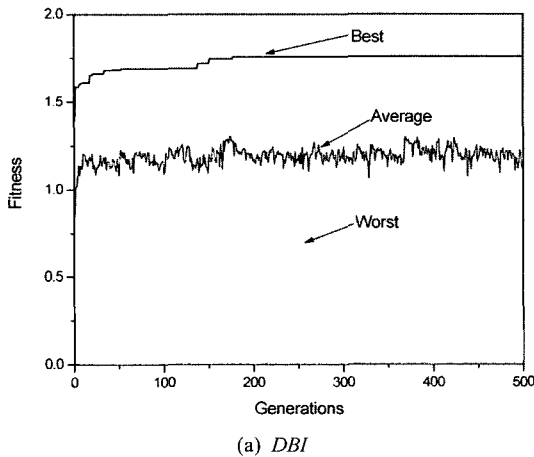


Fig. 6. The variation of fitness during evolutions

대적합도에 수렴하였다(Fig. 6). 이러한 차이는 DBI가 군 내의 분산과 더불어 군과 군 사이의 분리 정도를 함께 고려하기 때문에 인접한 2개의 군이 있을 경우 한 개의 군으로 여기는 개체와 분리된 2개의 군으로 간주하는 개체간 경쟁이 계속되는 것이다. 반면 VI는 군내의 분산의 합을 고려하며 군과 군 사이의 분리는 고려하지 않기 때문에 전체 자료에 있어서 분산이 최소화시키는 군의 수를 결정한다. 따라서 밀집된 2개의 군이 서로 인접할 경우 가능하면 별개의 군으로 처리하는 경향이 있다. 따라서 방향성 자료에 대한 군 분류에 있어서는 모의 자료 및 현장자료에 대한 분석 결과 군 내부의 분산만을 고려하는 VI가 DBI에 비해 더 좋은 적합도 기준이라고 판단된다.

최종적으로 분류된 5개의 군에 대한 평균 방향은 Table 3과 같으며 Fig. 7에 분류된 각 군에 속하는 극점 자료를 서로 다른 기호를 사용하여 표시하였다.

Table 3. Mean directions of five discontinuity sets

SET	Experts(DIP/DD)	GAC(DIP/DD)
1	44/112	44/111
2	45/243	45/244
3	44/288	45/291
4	78/186	79/188
5	80/109	83/107

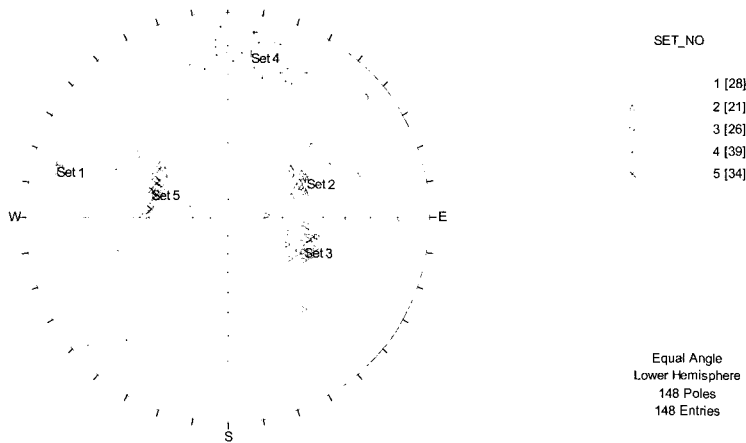


Fig. 7. Automatically identified five discontinuity sets using GAC

4.3 토의사항

본 연구에서는 군 분석시 불연속면의 방향정보만을 사용하였으며 랜덤절리에 대한 처리는 이루어지지 않은 상태이다. 즉, 어떤 불연속면이든 어느 군에는 속하도록 한 상태이기 때문에 랜덤절리는 고려할 수 없다. 군과의 거리가 특정값(cut-off distance) 이상일 경우 랜덤절리로 간주할 수 있으나 이 경우 최종적인 랜덤절리 군은 랜덤분포가 아니다. 즉, 기존의 군이 차지한 영역은 빈 공간으로 남게 된다. 따라서, 어떠한 기법을 사용하더라도 방향정보만을 고려할 경우에는 랜덤절리를 명확하게 처리할 수 있는 방법은 없다. 방향정보 외에 불연속면의 거칠기나 연장성과 같은 추가적인 불연속면의 특성을 함께 고려한다면 랜덤절리 처리도 가능할 것으로 판단된다(Hammah & Curran, 1989).

또한 유전자알고리즘을 이용한 클러스터링 기법은 확립된 이론이지만 클러스터링의 기준이 되는 적합도 함수는 문제의 성격에 따라 정의되어야 하며 불연속면 군 분석에 일반적으로 적용할 수 있는 새로운 기준을 마련하기 위해서는 군에 대한 엄밀한 정의가 선행되어야 하며 이에 대해서는 추가적인 연구가 필요하다.

5. 결 론

본 연구에서는 유전자알고리즘을 이용하여 불연속면 군 분석을 수행하는 프로그램을 개발하였고 이를 사용하여 모의 자료 및 현장 불연속면 자료의 군 분류에 적용하였다.

모의 자료에 대한 실험 결과 수렴속도에서 차이를 보일 뿐 최종 해는 개체군 크기, 교차 확률 및 돌연변이 확률에 관계없이 모두 동일한 결과를 보여 주었으며 참값

과 2도 이내의 오차범위에서 평균방향을 유사하게 추정하였다.

개발된 프로그램을 현장 불연속면 자료에 대하여 적용한 결과 각 군의 평균방향의 경우 전문가의 군 분석 결과와 3도 이내에서 유사하였으며 군 개수도 동일하게 추정하였다.

끝으로, DBI 및 VI 적합도 기준은 불연속면 군 분석 및 최적의 군 개수를 판별하는 적절한 기준이 될 수 있지만 방향만을 고려한 불연속면 군 분류에 있어서 군 내부의 분산만을 고려하는 VI가 DBI에 비하여 효율적이며 합리적인 결과를 제공하였다.

참고문헌

1. 조영임, 1999, 최신 인공지능, 학문사, 364.
2. 정용복, 전석원, 2003, 절리군 분석을 위한 퍼지 클러스터링 기법, 터널과 지하공간, 13(4), 294-303.
3. Bandyopadhyay, S. and U. Maulik, 2002, Genetic clustering for automatic evolution of clusters and application to image classification, *Pattern Recognition* 35(6): 1197-1208.
4. Hammah, R. E. and Curran, J. H., 1998, Fuzzy cluster algorithm for the automatic identification of joint sets, *Int. J. Rock Mech. Min. Sci.*, 35(7), 889-905.
5. Holland J.H., 1975, *Adaptation in natural and artificial system*, Ann Arbor, The University of Michigan Press.
6. Marcotte D. & Henry E., 2002, Automatic joint set clustering using mixture of bivariate normal distribution, *Int. J. Rock Mech. Min. Sci.*, 39, 323-334.
7. Markland, J., 1973, The analysis of principal components of orientation data, *Int. J. Rock Mech. Min. Sci.*, 11(3), 157-163.
8. Treleaven, J. F., 1994, Genetic-algorithm programming environments, *IEEE Computer* 27(6):28-43.



정용복

1994년 서울대학교 공과대학 자원공학과
공학사

1996년 서울대학교 대학원 자원공학과
공학석사

2000년 서울대학교 대학원
지구환경시스템 공학부 공학박사

Tel: 042-868-3097

E-mail: ybjung@kigam.re.kr

현재 한국지질자원연구원 지반안전연구부
선임연구원



선우준

1978년 서울대학교 공과대학 자원공학과
공학사

1984년 서울대학교 대학원 자원공학과
공학석사

1985년 Paris VI 대학(France) DEA

1988년 Paris VI 대학 지구구조학과
(France) 공학박사

Tel: 042-868-3235

E-mail: sunwoo@kigam.re.kr

현재 한국지질자원연구원 지반안전연구부
자원개발안전연구실장, 책임연구원