

유전자알고리즘을 이용한 절리군 분석

정용복, 선우춘 (한국지질자원연구원)

1. 서론

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

기존의 절리군 판별작업은 평사투영도를 이용한 도식적 방법을 주로 사용하였으며 최근에 통계 및 퍼지클러스터링 기법 등을 사용한 절리군 분석 자동화가 이루어져왔다(Hammah & Curran, 1998; Marcotte & Henry, 2002). 도식적 방법의 경우 주관적인 결과가 나올 가능성이 있으며 통계기법의 경우 사전에 절리군의 방향분포에 대한 가정이 이루어져야 한다는 단점이 있다. 또한 퍼지 기법의 경우 전역적 해를 구하는 것이 아니라 국부적 해를 구하는 기법이므로 전역해를 구하지 못할 가능성이 있다는 단점이 있다.

따라서 본 연구에서는 전역적 해를 구할 가능성이 타 기법보다 많은 유전자알고리즘을 적용한 절리군 분석 프로그램을 개발하였으며, 유전자알고리즘의 핵심 제어변수인 개체군 크기, 교차 및 돌연변이 확률에 따른 참값의 탐색 능력을 비교 및 검토하였다. 또한 적합도 함수, 즉 본 연구에서는 절리군 수를 결정하는 기준으로서 일반적으로 사용되는 DBI와 절리군의 평균방향에 대한 각 자료의 분산값에 근거한 VI를 적용하였으며 최종 결정된 제어변수를 사용하여 현장사면에서 조사한 불연속면 자료에 대하여 절리군 분석을 실시하였다.

2. 유전자알고리즘

2.1 진화알고리즘

진화알고리즘(EA, Evolution Algorithm)은 자연세계의 진화 과정을 모사함으로써 복잡한 실세계의 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 기법 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 일반적으로 적용될 수 있는 장점을 가지고 있다. 이러한 진화 알고리즘에는 유전자 알고리즘(GA: Genetic Algorithm), 진화전략(ES: Evolution Strategies), 진화프로그래밍(EP: Evolution Programming), 유전자프로그래밍(GP: Genetic Programming) 등이 있다.

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) 이라하며 개체군중에서 환경에 대한 적합도(fitness)가 높은 개체가 높은 확률로 살아남아 재생 할 수 있게 된다. 이 때, 교배(crossover) 및 돌연변이(mutation)로서 다음 세대의 개체군을 형성하게 된다. 이처럼 생물의 진화과정을 모델링한 알고리즘이다.

이러한 일반적인 유전자알고리즘의 구현 절차는 일반적으로 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 변화(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 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 제어 변수이며 본 연구에서는 이러한 구성요소들을 절리와 같은 방향성 자료의 처리에 적합하도록 변경하였다.

3. GA를 이용한 클러스터링

유전자알고리즘을 적용한 절리군 분석 과정을 그림 1에 도시하였으며 2절에서 소개한 주요 항목의 구체적 내용은 다음과 같다.

3.1 개체표현

본 연구에서는 염색체는 각 절리군의 중심방향을 표현하는 실수로 표현된다. 절리군수는 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.

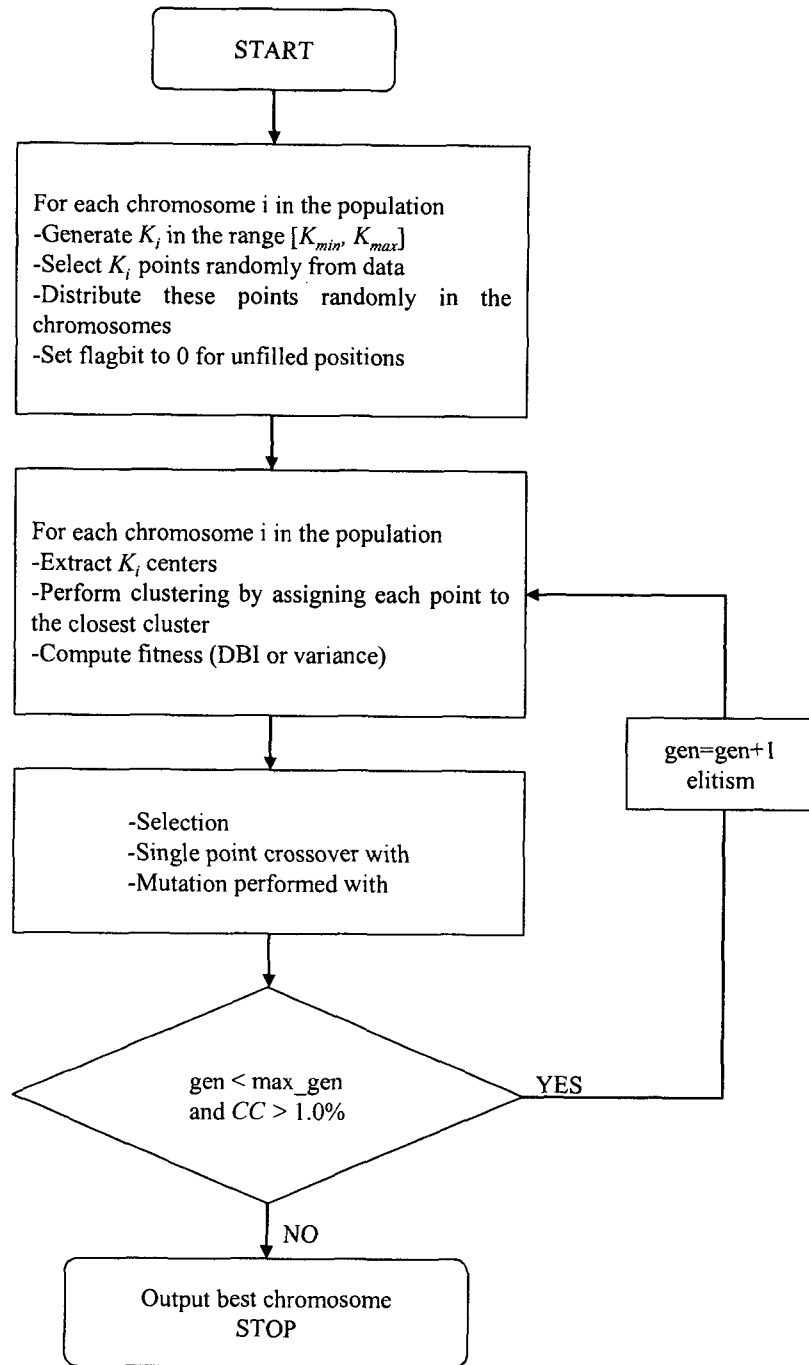


그림 1. 유전자알고리즘을 적용한 절리군 분석 절차

3.2 연산자

유전자알고리즘에 사용되는 연산자는 선택, 교배 및 돌연변이이다. 이러한 연산자가 개체군에 대해서 지정한 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 나눈 나머지가 원래의 값에 할당된다.

3.3 적합도함수

본 연구에서는 적합도함수로서 일반적인 클러스터링 기준으로 많이 사용되는 Davies-Bouldin지수, DBI(Maulik & Bandyopadhyay, 2000)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.

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

여기서 n 은 C_i 내의 자료 개수, S_n 은 평균방향과의 평균거리, d 는 두 벡터사이의 거리이며 절리군 C_i 와 C_j 간의 거리는 다음 식과 같이 계산된다.

$$d_{ij} = \sqrt{1 - (z_i \cdot z_j)^2} \quad (2)$$

VI는 중심방향과 각 자료와의 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.

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

4. GAC 적용

4.1 제어변수에 따른 참값 추정 능력

GA의 경우 개체군 크기(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 찾는 데 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 개체간 다양성의 손실로 최적해를 구하기 전에 수렴할 가능성을 내포한다. 반면에 개체군의 크기가 클 경우 최적해에 도달할 확률은 높아지나 메모리 및 계산시간이 증가한다. 따라서 GA를 이용해 풀고자 하는 문제의 특성에 따라 이러한 제어변수를 결정할 필요가 있다.

본 연구에서는 참값을 아는 모의 절리 자료 및 절리군에 대해서 최적의 제어변수를 구하고자 표 1과 같이 제어변수를 변화시키면서 총 36가지 경우에 대해서 절리군 분석을 수행하였다.

총 36회 분석 결과 DBI 및 VI 모두 3개의 군으로 분석하였으며 적합도 값도 동일하였다. 다만 반복회수에 있어서 DBI의 경우 평균 352회, VI의 경우 256회로 VI적합도 기준을 사용할 경우 수렴속도가 약간 빨랐다. 표 2는 분석 결과 구한 3개의 절리군에 대한 경사/경사방향이다. 각각 1.5도 이내에서 참값을 추정할 것을 알 수 있다. 그림 2는 계산과정 중 각 개체의 절리군 수 및 적합도의 변화를 표시한 것이다. 초기에는 절리군 수가 2에서 5까지 다양하지만 대부분 소멸하고 200세대까지 2개 및 3개의 절리군을 가진 개체끼리 경쟁하다가 이후 3개의 절리군을 가진 개체가 우세하게 됨을 알 수 있다. 적합도의 경우 평균치가 최대 적합도에 점진적으로 수렴하고 있음을 알 수 있다. 그림 3은 절리군 분석결과를 표시한 것이며 참값과 동일하게 군 분석이 이루어졌다.

표 1. GA제어변수 결정을 위한 실험 계획

제어변수	적용값				기타
개체군 크기 (PS)	50	100	150	-	3x4x3 = 36 cases
교차확률 (P_c)	0.2	0.4	0.6	0.8	
돌연변이확률 (P_m)	0.001	0.005	0.01	-	

표 2. 모의 절리 자료의 평균절리방향 및 추정 결과

절리군	참값	추정값	자료 수
	경사/경사방향	경사/경사방향	
1	60/120	60.5/120.0	100
2	25/220	25.5/218.5	120
3	85/65	83.4/65.5	130

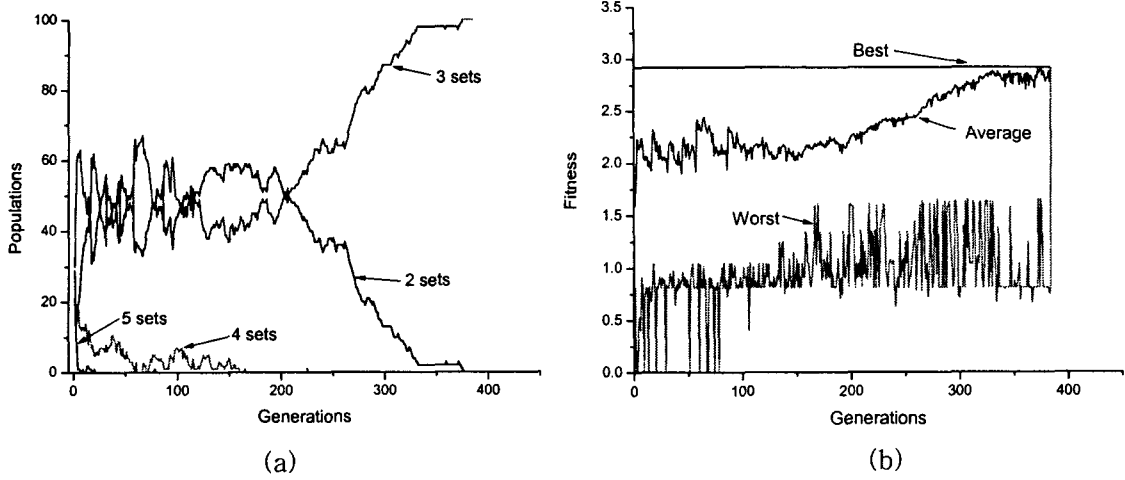


그림 2. 진화과정 중 (a)적합도 및 (b) 질리군 수 변화 ($P_c = 0.6$, $P_m = 0.01$, $PS = 100$, DBI).

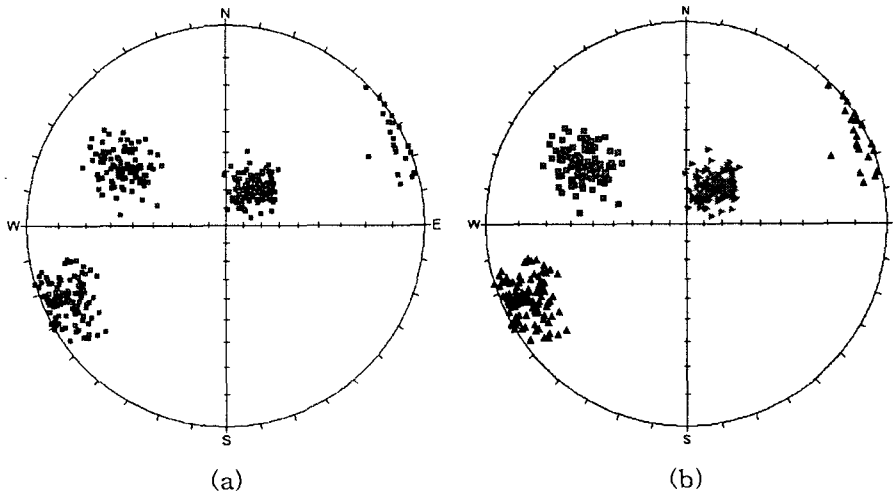


그림 3. 질리군 분석 결과 (a) 참값, (b) 추정값 ($P_c = 0.6$, $P_m = 0.01$, $PS = 100$, DBI).

4.2 현장 적용

고속도로 절개사면에서 조사한 148개의 불연속면 자료를 이용하여 군 분석을 수행하였다. 제어변수는 4.1의 결과 및 일반적인 관례를 기초로 하여 개체군 크기는 100, 교차확률은 0.6, 그리고 돌연변이 확률은 0.005로 정하여 분석을 수행하였다. 적합도 함수는 DBI 및 VI를 모두 적용하였다.

분석 결과 두 가지 적합도 모두 동일한 군 분석 결과를 보여주었으며 다만 VI를 적합도 함수로 사용할 경우 수렴속도에 있어서 좀더 좋은 결과를 보여주었다.

그림 4는 현장조사 자료의 평사투영도이며 절리, 층리 및 엽리로 구분하여 표시하였다. 그림 5 및 표 3은 분석 결과이며 지질전문가 및 현장조사자의 군 분석 결과와 3도 이내에서 일치함을 알 수 있다.

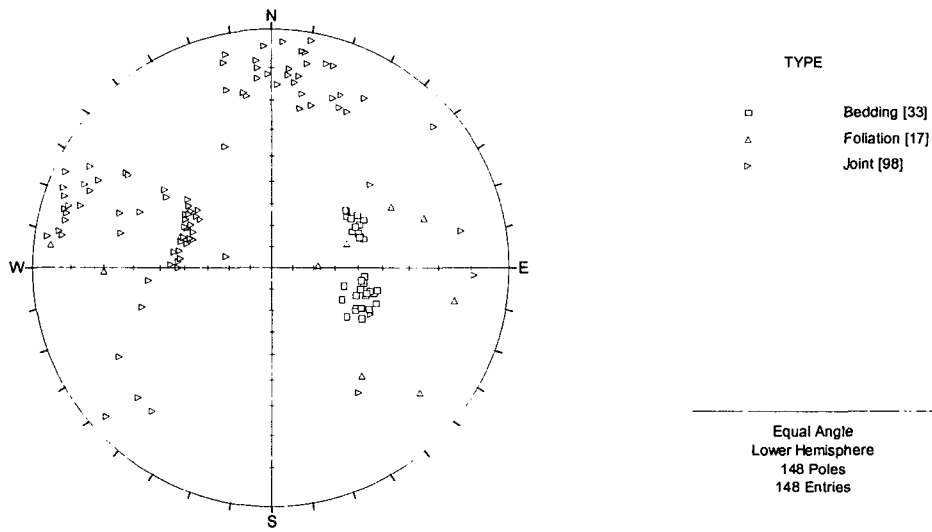


그림 4 고속도로 절개사면내 불연속면 조사자료의 방향 분포.

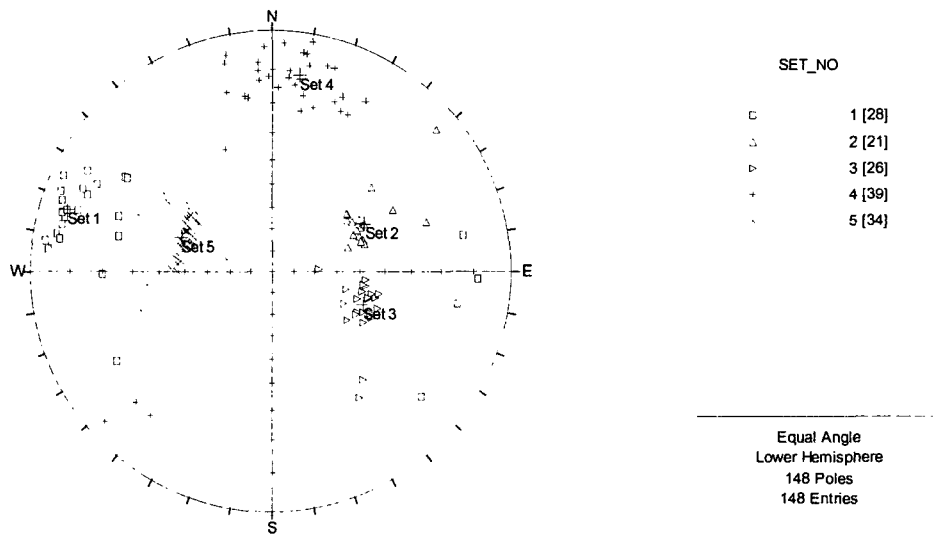


그림 5 유전자알고리즘을 적용한 절리군 분석 결과(+ : 각 절리군의 평균방향).

표 3 현장불연속면 자료 특성 및 절리군 분석 결과.

절리군	경사/경사방향	
	전문가	유전자알고리즘
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

5. 결론

본 연구에서는 유전자알고리즘을 이용하여 절리군 분석을 수행하는 프로그램을 개발하였으며 모의 절리 자료 및 현장 불연속면 자료에 대해 절리군 분석을 수행하였다.

모의 자료에 대한 실험 결과 수렴속도에서 차이를 보일 뿐 최종 해는 개체군 크기, 교차 확률 및 돌연변이 확률에 관계없이 모두 동일한 결과를 보여 주었으며 참값과 2도 이내의 오차범위에서 평균방향을 추정하였다.

현장 불연속면 자료에 대하여 개발된 프로그램을 적용한 결과 전문가의 군 분석결과와 3도 이내의 오차범위에서 각 절리군의 평균방향을 추정하였으며 절리군 수도 정확하게 추정하였다.

끝으로, DBI 및 VI 적합도 기준은 절리군 분석 및 최적 절리군 수를 판별하는 적절한 기준이 됨을 확인하였다.

참고문헌

1. 조영임, 1999, 최신 인공지능, 학문사, p.364
2. Bandyopadhyay, S. and U. Maulik, 2002, Genetic clustering for automatic evolution of clusters and application to image classification, *Pattern Recognition* 35(6): pp.1197-1208.
3. Grossman, N. F., 1985, The bivariate normal distribution on the tangent plane at the mean attitude: A keynote lecture, *Proc. Int. Sympo. on Fundamentals of Rock Joints, Björkliden*, pp.15-20.
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), pp.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, pp.323-334.
7. Treleaven, J. F. a. P., 1994, Genetic-algorithm programming environments, *IEEE Computer* 27(6): pp.28-43.