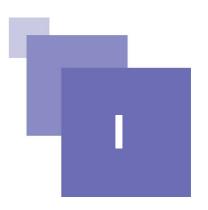
Cours 13 - Les Algorithmes Génétiques

Marie-Hélène Abel

Table des matières

I - I	Introduction	5
	A. Préambule	5
	B. Un peu de génétique	
II -	Les algorithmes génétiques	9
	A. Un peu d'histoire	9
	B. Les Algorithmes Evolutionnaires	10
	C. Les algorithmes génétiques : définition	
	D. Les algorithmes génétiques : mise en œuvre. 1. Fonctionnement simplifié	
	E. Exemple	19
	F. Quelques applications	24

Introduction



Préambule 5
Un peu de génétique 5

A. Préambule



Rappel : L'IA numérique vise la perception et le fonctionnement réflexe.

- But : Etude de la perception (immédiate, implicite) et du réflexe (idem)
- Origine : la physiologie
- · Construit des modèles informatiques
- Demande peu ou pas de programmation explicite
- Exploite des processus d'apprentissage
- · Permet de développer des systèmes auto adaptatifs et évolutifs



Exemple

- Réseaux de neurones
- Algorithmes génétiques



Méthode

Comment acquérir un savoir-faire à partir d'exemples et d'expérimentation sans passer par une étape de formalisation et de raisonnement ?

B. Un peu de génétique



Fondamental: L'évolution dans la nature survient quand:

- des entités ont la capacité de se reproduire,
- qu'il existe une population de ces entités,
- qu'il existe une variété (diversité) à travers ces entités,
- que la survie des entités dépend des différences entre elles.



Remarque

Toute entité vivante possède un génotype et un phénotype.

1. Le génotype



Définition

Le génotype est constitué de gènes situés sur des chromosomes stockés dans le noyau des cellules sous la forme d'une longue chaîne d'acide désoxyribonucléique (ADN).

Dans la nature, l'ADN est un polymère constitué par l'enchaînement de quatre molécules :

- les nucléotides adénine (A),
- la cytosine (C),
- la guanine (G),
- la thymine (T).



Remarque

On peut donc décrire l'ADN par des chaînes de quatre caractères ACGT.



Remarque

L'ADN constitue l'ensemble des chromosomes, ou le génome d'un individu.

2. Le phénotype



Définition

Le phénotype est l'ensemble des protéines et des enzymes qui peuvent être fabriqués à partir de l'ADN.



Remarque

En fait, l'ADN est copiée par un messager (ARN) qui au niveau du ribosome, se traduit en chaînes d'acides aminés formant les protéines et les enzymes.

En général, on compte une protéine (un enzyme) par gène.



Méthode : Ce sont les protéines et les enzymes qui dictent la structure et le comportement des cellules qui permettent à un individu de :

- Réaliser des tâches dans son environnement,
- · Survivre,
- · Se reproduire à des taux différents.

3. La reproduction

La reproduction se traduit par la transmission du génome aux individus de la progéniture, ce qui permet de préserver les gènes menant à des performances supérieures.



Remarque

Occasionnellement, un processus naturel, la mutation génétique, introduit une variation dans les chromosomes.

Principes de survie et reproduction décrit par Charles Darwin « On the Origin of Species By Means of Natural Selection » en 1859

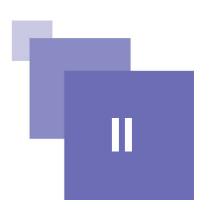
Les individus les mieux adaptés, c'est-à-dire capables de mieux effectuer les tâches nécessaires à leur survie, se reproduisent à des taux les plus élevés, alors que les individus les moins adaptés se reproduisent à des taux plus faibles.



Remarque

Une population ayant une grande variété va, de génération en génération, contenir des individus dont le génotype se traduit par une meilleure adaptation, et ceci à cause de la contrainte de la sélection naturelle.





Un peu d'histoire	9
Les Algorithmes Evolutionnaires	10
Les algorithmes génétiques : définition	11
Les algorithmes génétiques : mise en œuvre	13
Exemple	19
Quelques applications	24

Les algorithmes génétiques sont des algorithmes d'optimisation s'appuyant sur des techniques dérivées de la génétique et des mécanismes d'évolution de la nature :

- croisements,
- mutations,
- sélections,
- etc...

Ils appartiennent à la classe des algorithmes évolutionnaires.

A. Un peu d'histoire

1859

Charles Darwin publie son livre intitulé "L'origine des espèces au moyen de la sélection naturelle ou la lutte pour l'existence dans la nature".

Dans ce livre, Darwin rejette l'existence «de systèmes naturels figés», déjà adaptés pour toujours à toutes les conditions extérieures, et expose sa théorie de l'évolution des espèces :

Sous l'influence des contraintes extérieures, les êtres vivants se sont graduellement adaptés à leur milieu naturel au travers de processus de reproductions.

20ème siècle : Mise en évidence de l'existence de mutations génétiques.

Les problèmes de traitement de l'information sont résolus de manières figés :

Lors de sa phase de conception, le système reçoit toutes les caractéristiques nécessaires pour les conditions d'exploitations connues au moment de sa conception ce qui empêche une adaptation à des conditions d'environnement inconnues, variables ou évolutives.



Remarque

Les chercheurs en informatique étudient donc des méthodes pour permettre aux systèmes d'évoluer spontanément en fonction de nouvelles conditions.

1966

Programmation évolutionnaire L. J. Fogel.

1973

Stratégie d'évolution I. Rechenberg.

1975

Dans les années 1960, John Holland étudie les systèmes évolutifs et, en 1975, il introduit le premier modèle formel des algorithmes génétiques (the canonical genetic algorithm AGC) dans son livre "Adaptation in Natural and Artificial Systems".



Remarque

Ce modèle servira de base aux recherches ultérieures.

1989

David Goldberg publie un ouvrage de vulgarisation des algorithmes génétiques.

Années 1990

Programmation d'une panoplie d'algorithmes génétiques.

B. Les Algorithmes Evolutionnaires



Définition

L'appellation Algorithmes Evolutionnaires (EAs) est un terme générique pour décrire les approches informatiques reposant sur les principes de l'évolution comme éléments clefs de leurs fonctionnements.



Remarque : Il existe un grand nombre d'algorithmes évolutionnaires :

Toutes ces approches ont en commun la simulation de l'évolution de structures individuelles au travers de méthodes de :

- sélection,
- · mutation,
- reproduction.

Ces processus reposent sur la performance des structures individuelles à leur environnement.

Plus précisément :

- Les EAs maintiennent une population de structures, qui évolue en accord avec des règles de sélection et d'autres opérateurs, appelés opérateurs de recherche.
- Chaque individu possède une mesure de son adéquation avec l'environnement : la fitness.
- La reproduction s'appuie sur cette mesure pour favoriser la prolifération des

individus les mieux adaptés.

• La recombinaison et la mutation perturbent ces individus fournissant par ce biais des heuristiques pour l'exploration de nouvelles solutions.

Bien que relativement simple ces algorithmes sont suffisants pour assurer des mécanismes de recherche robustes et puissants.

C. Les algorithmes génétiques : définition



Définition

- Les algorithmes génétiques sont des algorithmes d'exploration fondés sur la sélection naturelle et la génétique.
- Ils utilisent les principes de la survie des structures les mieux adaptées, les échanges d'informations pseudo aléatoires.
- Ils reposent de manière intensive sur le hasard mais ne sont pas purement aléatoires.

1. Objectif

L'objectif des AG est de déterminer les extrêmes d'une fonction f : X -> R, où :

- X est un ensemble quelconque appelé espace de recherche,
- f est appelée fonction d'adaptation ou fonction d'évaluation ou encore fonction fitness.



Remarque

La fonction agit comme une «boite noire»pour l'AG.

Des problèmes très complexes peuvent être approchés par programmation génétique sans avoir de compréhension particulière du problème.



Fondamental

La recherche des algorithmes génétiques a pour but l'amélioration de la robustesse et l'équilibre entre la performance et le coût nécessaire à la survie dans un environnement.

Ils sont utilisés comme alternative à l'optimisation de fonction lorsque les méthodes "classiques", telles que :

- les méthodes indirectes qui recherchent à atteindre les extrema locaux en résolvant des systèmes d'équations,
- les méthodes directes par suivi de gradient ou par énumération,

ne sont pas applicables ou ont échoué.

2. AG et autres approches

Ce qui différencie les AG des autres approches peut s'exprimer en quatre points :

- Les AG utilisent un codage des paramètres et non les paramètres euxmêmes.
- Les AG travaillent sur une population de points et non sur un point

particulier.

- Les AG n'utilisent que les fonctions étudiées, pas leurs propriétés (telles que la dérivabilité ou autre).
- Les AG utilisent des règles de transitions probabilistes et non déterministes.

3. Synthèse

Les AG sont des algorithmes inspirés des mécanismes de la sélection naturelle et de la génétique.

=> Ils utilisent à la fois les principes de la survie des individus les mieux adaptés et ceux de la propagation du patrimoine génétique.

De façon très intuitive, on identifie le problème à un environnement donné et les solutions à des individus évoluant dans cet environnement :

- A chaque génération, on ne retient que les individus les mieux adaptés à cet environnement.
- Au bout d'un certain nombre de générations, les individus restants sont particulièrement adaptés à l'environnement donné.



Remarque

On obtient donc des solutions très proches de la solution idéale du problème.

4. Vocabulaire



Définition : Population

Les algorithmes génétiques ne travaillent pas sur un individu, sur une donnée, mais au contraire sur une population de chaînes afin d'effectuer des opérations, des recherches sur un domaine de possibilités plus important.



Remarque

C'est une des grandes forces des AGs.

Une population se compose de chaînes ou chromosomes



Définition : Chromosome ou individu :

Les chaînes des systèmes génétiques artificiels sont analogues aux **chromosomes** des systèmes biologiques :

Ils portent les informations génétiques d'un individu.



Remarque

Ainsi, un individu se compose de gènes.



Définition : Gène :

- Les chromosomes se décomposent en gènes qui peuvent prendre des valeurs différentes que l'on appelle **allèles**.
- La position d'un gène dans un chromosome est identifiée par son locus.
- Un gène est une caractéristique génétique d'un individu.



Remarque

Dans les AGs, les gènes ont des valeurs appartenant à un alphabet qui dépend du

problème à résoudre.

On peut ainsi avoir des allèles binaires, décimaux, ou encore hexadécimaux.

On peut imaginer d'autres sortes de codage.



Définition : Génération :

Une génération est une population à un instant t.

Les AGs faisant évoluer les populations, on crée une nouvelle génération d'individus plus adaptés.

Cette évolution est effectuée par les opérateurs de :

- reproduction,
- croisement,
- · mutation.

Analogie avec la nature

Nature

- Chromosome
- Gène
- Allèle
- Locus
- Génotype
- Phénotype

Algorithme génétique

- Chaîne
- Trait, caractéristique
- Valeur de la caractéristique
- Position dans la chaîne
- Structure
- Ensemble de paramètres, Une structure décodée

image1

D. Les algorithmes génétiques : mise en œuvre

Nous allons nous intéresser au fonctionnement simplifié des AGs puis aux différentes étapes mises en œuvre.

1. Fonctionnement simplifié

Les mécanismes mis en jeu sont très simples :

- On dispose d'une population de chaînes de caractères sur un alphabet que nous supposerons binaire.
- À chaque individu est associé un score (fitness), on en profite pour calculer la somme de ces scores ainsi que les valeurs maximale, minimale et moyenne.

Trois opérateurs vont être utilisés :

- Un opérateur de sélection (select) dont le rôle est de choisir les individus les plus adaptés (ceux possédant un score élevé).
- Un opérateur de croisement (cross-over ou hybridation)
- Un opérateur de mutation (mutation).

L'algorithme génétique peut s'exprimer sous la forme suivante :

T < -0

P <- initPopulation

Loop

- Evaluate(P,t)
- · Exit when critere d'arrêt
- Q <- select(P,t)
- R <- cross-over(Q)
- P <- mutation(R)
- t <- t + 1

EndLoop



Attention : Avant de pouvoir utiliser un AG, il nous faut définir :

- Un codage adapté au problème (les chromosomes),.
- Une fonction (fitness) qui va caractériser l'adéquation de la solution (représentée par son chromosome) au problème.

2. Chromosomes

- On suppose dans cette approche qu'une solution du problème peut être représentée par un ensemble de paramètres.
- Ces paramètres sont regroupés sous la forme d'une chaîne de caractères.



Remarque

- J. Holland a le premier montré pourquoi une représentation sous forme binaire était efficace.
- Il existe cependant d'autres approches : décimale, hexadécimale, alphanumérique...



Rappel

- En génétique, un ensemble de paramètres représentés par un chromosome particulier est appelé son génotype.
- Le phénotype est une instance particulière d'un génotype.
- La fonction d'évaluation dépend du phénotype.

3. Fitness ou fonction d'évaluation

- Cette fonction est déterminée en fonction du problème à résoudre et du codage choisi pour les chromosomes.
- À un chromosome particulier, elle attribue une valeur numérique, qui est supposée proportionnelle à l'intérêt de l'individu en tant que solution.

- L'évaluation d'un individu ne dépend pas de celle des autres individus.
- Le résultat fourni par la fonction d'évaluation permet de sélectionner ou de refuser un individu pour ne garder que les individus ayant le meilleur coût en fonction de la population courante : c'est le rôle de la fonction **fitness**.



Remarque

Cette méthode permet de s'assurer que les individus performants seront conservés, alors que les individus peu adaptés seront progressivement éliminés de la population.

4. Sélection

- Pour cette phase, nous supposons connu : les individus de la population au temps t , la fonction d'évaluation.
- Il va falloir trouver une manière de choisir les individus répondant le mieux à notre problème.
- Il s'agit donc de privilégier les individus ayant un score au-dessus de la moyenne, et de pénaliser ceux en dessous de cette moyenne.
- · Plusieurs approches sont possibles!

a) Sélection par roulette (wheel)



Définition

Les parents sont sélectionnés en fonction de leur performance.



Remarque

Meilleur est le résultat codé par un chromosome, plus grandes sont ses chances d'être sélectionné.



Méthode

Il faut imaginer une sorte de roulette de casino sur laquelle sont placés tous les chromosomes de la population, la place accordée à chacun des chromosomes étant en relation avec sa valeur d'adaptation.

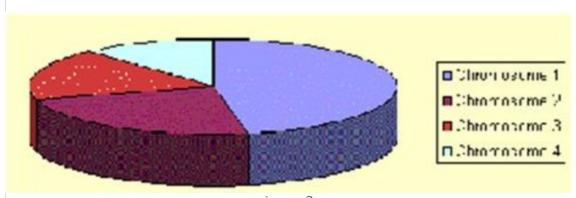


image2



Méthode

- Ensuite, la bille est lancée et s'arrête sur un chromosome.
- Les meilleurs chromosomes peuvent ainsi être tirés plusieurs fois et les plus mauvais ne jamais être sélectionnés.



Remarque : Cela peut être simulé par l'algorithme suivant :

- On calcule la somme S1 de toutes les fonctions d'évaluation d'une population.
- On génère un nombre r entre 0 et S1.
- On calcule ensuite une somme S2 des évaluations en s'arrêtant dès que r est dépassé.
- Le dernier chromosome dont la fonction d'évaluation vient d'être ajoutée est sélectionné.



Attention

La sélection par roulette rencontre des problèmes lorsque la valeur d'adaptation des chromosomes varie énormément.

Si la meilleure fonction d'évaluation d'un chromosome représente 90% de la roulette alors les autres chromosomes auront très peu de chance d'être sélectionnés et on arriverait à une stagnation de l'évolution.

b) Sélection par rang



Méthode

- La sélection par rang trie d'abord la population par fitness.
- Chaque chromosome se voit associé un rang en fonction de sa position.



Remarque

Le plus mauvais chromosome aura le rang 1, le suivant 2, et ainsi de suite jusqu'au meilleur chromosome qui aura le rang N (pour une population de N chromosomes).

Avec cette méthode de sélection, tous les chromosomes ont une chance d'être sélectionnés.

Cependant, elle conduit à une convergence plus lente vers la bonne solution.

Ceci est dû au fait que les meilleurs chromosomes ne diffèrent pas énormément des plus mauvais.

c) Sélection steady-state

L'idée principale est qu'une grande partie de la population puisse survivre à la prochaine génération.



Méthode : L'algorithme génétique marche alors de la manière suivante :

- A chaque génération sont sélectionnés quelques chromosomes (parmi ceux qui ont le meilleur coût) pour créer des chromosomes fils.
- Les chromosomes les plus mauvais sont retirés et remplacés par les nouveaux.
- Le reste de la population survie à la nouvelle génération.

d) Elitisme



Remarque

A la création d'une nouvelle population, il y a de grandes chances que les meilleurs chromosomes soient perdus après les opérations d'hybridation et de mutation.



Méthode : Pour éviter cela, on utilise la méthode d'élitisme :

- Copier un ou plusieurs des meilleurs chromosomes dans la nouvelle génération.
- Générer le reste de la population selon l'algorithme de reproduction usuel.

Cette méthode améliore considérablement les algorithmes génétiques, car elle permet de ne pas perdre les meilleurs solutions.

5. L'hybridation (ou cross-over ou croisement)

Le cross-over est une méthode permettant de construire de nouveaux individus à partir de leurs parents.

Il existe deux grandes approches :

- le cross-over simple,
- le cross-over double.



Méthode : Le cross-over simple

Consiste à :

Choisir un point de croisement compris entre 1 et (L - 1) où L est la longueur de la chaîne,

Recombiner les enfants comme étant le début d'un chromosome et la fin de l'autre.



Exemple

- Considérons les deux chromosomes suivants 00110011 et 10101010.
- Choisissons comme point de croisement la valeur 3.
- On obtient les deux nouveaux individus suivants : 001 01010 et 101 10011.



Méthode : Le cross-over double

Consiste à sélectionner deux points de croisement et à échanger les parties intermédiaires.



Exemple : Reprenons l'exemple précédent :

Choisissons comme points de croisement 3 et 6.

On obtient les deux nouveaux individus : 001 010 11 et 101 100 10.



Remarque

le cross-over simple peut être considéré comme un cross-over double si on "voit" :

- les chromosomes comme des chaînes circulaires,
- le deuxième point de croisement comme étant en position L.

6. Mutation

- La mutation est appliquée individuellement à tous les "enfants" après le croisement.
- Elle a pour but d'altérer aléatoirement une information avec une faible probabilité (de l'ordre de 0.001).



Remarque

L'intérêt de la mutation est de pouvoir explorer aléatoirement une autre portion de l'espace de recherche, garantissant ainsi que tous les points de l'espace ont une probabilité non nulle d'être essayés.

7. Itération



Méthode

A partir d'une population initiale de M individus, l'AG sélectionne une population intermédiaire de M individus en faisant une sélection sur la population initiale.



Remarque

Un même individu peut être sélectionné plusieurs fois ou peut ne pas être sélectionné du tout, en fonction de la valeur de sa fonction d'évaluation



Méthode

Les M individus de la population se croisent deux à deux (les couples se forment aléatoirement) pour construire M nouveaux individus.



Méthode

Ces individus passent par un opérateur de mutation (qui agit aléatoirement avec une possibilité faible 2-3% de bits) pour former une nouvelle population.

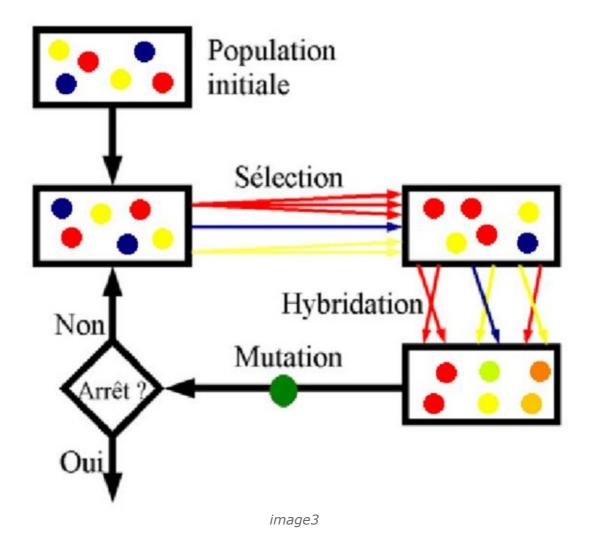


Méthode

On réitère ensuite le procédé à partir de cette population jusqu'à obtenir une solution que l'on juge satisfaisante.

8. Synthèse

Schéma résumant le fonctionnement des AG



E. Exemple

Source

http://wwwsi.supelec.fr/yb/projets/algogen/box.htm

Problème de la boîte à n interrupteurs

- Il se compose d'une boîte avec n interrupteurs dont on ne connaît absolument pas le contenu et les résultats du positionnement des interrupteurs.
- Nous voulons connaître la position des interrupteurs qui optimise la sortie de la boîte, sachant que l'on peut mesurer et évaluer la sortie.

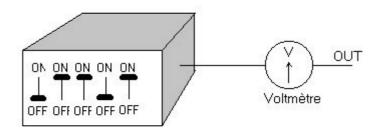


image4



Exemple: Prenons une boîte noire avec cinq interrupteurs.

- Par une méthode classique, il faudrait parcourir toutes les positions possibles des interrupteurs afin de déterminer la position optimale, soit pour n interrupteurs 2^n positions.
- Le nombre de positions possibles croît de façon exponentielle.



Exemple: Exemples d'équivalences:

- 5 interrupteurs : 2^5 équivaut à 32 positions possibles.
- 16 interrupteurs: 2^16 équivaut à 65 536 positions possibles.
- 32 interrupteurs: 2^32 équivaut à 4 294 967 296 positions possibles.



Remarque

Pour cinq interrupteurs, le parcours de toutes les positions serait faisable mais audelà, le temps imparti à ce travail serait inconcevable.

- Les AGs vont nous permettrent de trouver rapidement une solution qui se rapproche de l'optimum.
- La solution trouvée ne sera pas forcément la solution optimale mais les AGs la trouveront dans un temps raisonnable.



Méthode : Dans un premier temps, il faut tout d'abord coder le problème.

- Nous allons travailler sur une population de chaînes contenant les informations sur la position des interrupteurs de la boîte.
- Ce codage est très simple :

Une chaîne de cing bits suffit pour coder les cing positions :

- un bit positionné à 1 code un interrupteur sur ON
- un bit positionné à 0 est équivalent à un interrupteur sur OFF.



Remarque

Les AGs vont donc travailler sur des individus binaires.



Exemple : Il s'agit alors de tirer au hasard une population initiale, une batterie de test :

10010

11000

01101 10001



Remarque

Nous avons une population composée de cinq chaînes contenant toutes les informations sur la position des cinq interrupteurs.



Méthode : Il nous faut alors tester ces chaînes afin de connaître leur adaptation.

On mesure la sortie qui nous donne la valeur d'adaptation de chaque chaîne.

Chaine	Adaptation		
10010	5832		
11000	13824		
00010	8		
01101	2197		
10001	4913		





Remarque : Les individus de notre population sont plus ou moins bien adaptés.

- La chaîne 11000 est bien adaptée.
- La chaîne 00010 est très mal adaptée.
- Le maximum de l'adaptation est 13824 et la moyenne vaut 5354,8.



Méthode : Il s'agit maintenant de faire évoluer la population pour obtenir une nouvelle génération d'individus plus adaptés.

- Le but est de faire mourir les individus les moins adaptés, tout en gardant les meilleurs.
- Nous allons croiser les informations des meilleurs afin de créer de nouveaux individus qui seront (on l'espère) mieux adaptés.



Remarque

Si on élimine directement l'individu le moins adapté, soit 00010, nous avons un taux de mortalité de 20% environ.



Méthode

Nous tirons au hasard deux individus en fonction de la valeur de leur adaptation (tirage pseudo aléatoire pondéré par la fonction d'adaptation) afin d'effectuer un croisement.



Exemple: Après tirage au sort:

- l'individu 1 est apparié avec l'individu 2,
- l'individu 4 est apparié avec l'individu 5.



Exemple : Il s'agit maintenant de les croiser.

Pour cela, il faut tirer, pour chaque paire, un nombre qui représentera la frontière de croisement. (tirage : 3 et 2).

100|10

110|00

01|101

10|001

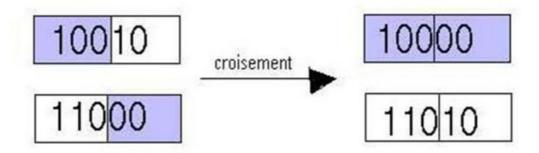


image6



Exemple : Après croisement, nous obtenons deux nouvelles paires d'individus avec leur valeur d'adaptation correspondante :

Chaîne	Adaptation		
10000	4096		
11010	17576		
01001	729		
10101	9261		

image7



Remarque : De nouveaux individus ont été créés

La population est maintenant composée de 8 individus (parents + enfants).



Méthode

Il nous faut donc choisir aléatoirement (toujours en fonction de l'adaptation) cinq individus parmi les huit, afin de garder une population constante.



Exemple

Un individu adapté a toutes les chances de se retrouver dans la nouvelle génération.

On obtient la génération suivante :

Chaîne	Adaptation			
11010	17576			
10000	4096			
10101	9261			
10010	5832			
11000	13824			

image8



Remarque

- La valeur maximale de l'adaptation a augmenté et est passée à 17576.
- La valeur moyenne est égale à 10117.
- => La nouvelle génération est globalement mieux adaptée.



Méthode : Afin de ne perdre aucune possibilité de positions, on introduit l'opération de mutation.

- En fonction d'une probabilité de mutation (0,001% par exemple), les individus vont muter de façon plus ou moins probable.
- · La probabilité de mutation doit être faible.



Exemple

Aléatoirement, l'individu 3 mute et devient 10111.

On obtient la nouvelle génération :

Chaîne		Adaptation				
	11010	17576				
	10000	4096				
	10111	9261	10101	Mutation	•	10111
	10010	5832	10101			
	11000	13824				

image9

Bilan

- Nous avons obtenu, après ces opérations simples et sans se soucier de la sémantique de ces chaînes codées, une nouvelle génération plus adaptée que la précédente.
- La valeur moyenne de l'adaptation est égale à 10699.



Remarque

- En renouvelant ces opérations N fois, on obtient une solution qui s'approche de l'optimum.
- Le test d'arrêt peut être la stagnation de la population sur plusieurs générations ou un nombre total de génération.

F. Quelques applications

Les algorithmes génétiques ont été appliqués avec succès à :

- · la recherche de solution dans un jeu de stratégie,
- l'optimisation de réseaux de neurones,
- · l'optimisation de fonctions (travaux de De Jong),
- · la simulation de cellules biologiques,
- · la reconnaissance de formes.