

Algorithmes Génétiques - EISTI - ING 2

Ecole Internationale des Sciences du Traitement de l'Information

15 novembre 2010

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Sommaire

Présentation Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Sommaire

Présentation Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Pourquoi l'évolution comme principe algorithmique

Mécanismes bien adaptés :

- ▶ aux problèmes nécessitant des ressources de calcul importantes
- ▶ aux programmes qui doivent être adaptatif (trajectoire robot)
- ▶ aux programmes trop complexes pour être programmés à la main

Principes de l'évolution

- ▶ Recherche parmi un nombre de possibilités gigantesque
- ▶ Conception de solutions innovantes à des problèmes complexes
- ▶ Méthodes de recherche massivement parallèles

Définitions

- ▶ Individu/chromosome/séquence : une solution potentielle au problème
- ▶ Population : un ensemble de chromosomes ou de points de l'espace de recherche
- ▶ Environnement : l'espace de recherche
- ▶ Fonction de fitness : fonction (positive) que l'on cherche à maximiser en général (on pourrait très bien minimiser)

Concept

- ▶ Les individus soumis à l'évolution sont des solutions plus ou moins performantes à un problème donné.
- ▶ A chaque génération, une succession d'opérations est appliquée pour engendrer une nouvelle population.
- ▶ Les individus résultants sont les enfants. Les individus utilisés par un opérateur sont les parents (transmission du patrimoine génétique).

Opérateurs de sélection

A chaque génération, les individus se reproduisent, survivent ou disparaissent sous l'action des opérateurs de sélection :

- ▶ La sélection pour la reproduction (combien de fois un individu sera reproduit en une génération)
- ▶ La sélection pour le remplacement (disparition d'individus)

Sélection

Darwin : selection des meilleurs individus :

- ▶ Valeur de performance associée à chaque individu (fitness function) selon l'objectif.
- ▶ Transmission de ses meilleurs caractéristiques à ses descendants.

Opérateurs de variation

Améliorer la qualité des solutions :

- ▶ par mutation : transformation d'un individu en un autre
- ▶ par croisement (crossover) : engendre un ou plusieurs enfants à partir de deux parents (reproduction sexuée)

Algorithme

1. Pour reproduction : sélection des parents parmi une population de μ individus
2. Croisements et mutations à partir des individus sélectionnés pour engendrer λ enfants
3. Evaluation et performance des enfants
4. Sélection pour la survie de μ individus parmi les λ enfants et μ parents ou uniquement parmi les λ enfants selon les paramètres choisis, afin de générer la population suivante.

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Définition

Nous appelons séquence (chromosome, individu) A de longueur $l(A)$ une suite $A = \{a_1, a_2, \dots, a_l\}$ avec $\forall i \in [1, l], a_i \in V$ où V est un codage.

Sommaire

Présentation Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Codage des solutions (chromosomes)

Structure de données codant les gènes :

- ▶ codage binaire : chaîne de bits :
codage historique parfois peu naturel pour certains problèmes.
- ▶ codage à caractères multiples :
codage linéaire constitué de différents types de caractères
- ▶ codage arborescent :
codage autorisant des solutions de taille infinie

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Exemple de codage binaire

Codage classique

Sur 4 bits,

- ▶ on code la valeur 1 par 0001
- ▶ on code la valeur 2 par 0010
- ▶ ces deux valeurs sont proches mais leur codage les éloigne (distance de Hamming de 2) :

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

Codage de Gray

On garde toujours entre 2 valeurs consécutives une distance de 1 :

- ▶ valeur 1 : 0001
- ▶ valeur 2 : 0011
- ▶ valeur 3 : 0010
- ▶ valeur 4 : 0110
- ▶ etc ...

Taille du codage

Précision souhaitée

Si un chromosome doit représenter l'intervalle $[a, b]$ avec une précision Δ , il faut utiliser $l = \log_2 \frac{b-a+\Delta}{\Delta}$ bits.

Exemple

Si $a = -100$, $b = 100$ et $\Delta = 10^{-4}$, il faut 21 bits

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Sommaire

Présentation Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Fonction de fitness

Nous appelons fitness d'une séquence A toute valeur positive que nous noterons $f(A)$ (fonction de fitness).

En général, il s'agit d'une valeur réelle caractérisant l'adaptation de la solution à son environnement. Son calcul dépend fortement du codage choisi pour représenter les solutions.

Sommaire

Présentation Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Roulette

Choix aléatoire : individu a une probabilité d'être choisi proportionnelle à sa qualité. La probabilité d'un chromosome x sur une population π vaut :

$$p(x) = \frac{f(x)}{\bar{f}}$$

avec

$$\bar{f} = \sum_{x \in \pi} f(x)$$

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Sélection par tournoi

Tournoi déterministes

Sélection aléatoire de k individus, sélection du meilleur

- ▶ autant de tournois que d'individus sélectionnés
- ▶ tirage avec ou sans remise
- ▶ pression de sélection ajustée par le paramètre k

Tournoi binaire stochastique

Sur 2 individus en compétition, le meilleur gagne avec une probabilité comprise entre 0,5 et 1.

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Pression de sélection

- ▶ Algorithme dans un optimum local.
- ▶ Pression sélective quasi-inexistante : dérive génétique.
- ▶ Ajustements dynamiques de la fonction de performance

Sélections pour le remplacement

Remplacement générationnel

Génération $g + 1$ constituée uniquement des enfants $\mu = \lambda$.

En général, sélection des meilleurs, $\lambda > \mu$

Sommaire

Présentation Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Exemple avec un codage binaire

Création d'un nouvel individu à partir de 2 individus par échange de chromosomes entre individus

- ▶ Sélection des deux individus parents
- ▶ Détermination de 1 ou plusieurs points de croisement (entre 1 et $longueur(individu) - 1$)
- ▶ Selon une valeur probabiliste, échange des segments finaux des 2 parents (dans le cas d'un seul point de croisement) afin de générer 2 nouveaux individus (enfants)

Sommaire

Présentation Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Opérateurs de mutation

Exemple avec un codage binaire

Tirage aléatoire d'un (ou plusieurs) gène du chromosome et remplacement par son inverse selon une certaine probabilité.

Rôle dans l'algorithme

- ▶ Permet d'effectuer une recherche locale
- ▶ Permet de sortir d'un optimum local
- ▶ Permet d'atteindre potentiellement tous les points de l'espace de recherche

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Sommaire

Présentation Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Paramétrisation de l'algorithme génétique

- ▶ Taille de la population N : si N est trop grand le temps de calcul peut être trop long, s'il est trop petit il y a un risque de convergence trop rapide vers un mauvais chromosome.
- ▶ Probabilité de croisement p_c . Elle dépend de la forme de la fonction de fitness. Plus elle est élevée, plus la population subit des changements importants (en général entre 0,5 et 0,9)
- ▶ Probabilité de mutation p_m . Taux généralement faible afin d'éviter de converger vers une solution sous-optimale.

Sommaire

Présentation Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Algorithme

- ▶ Initialisation : une population initiale de N chromosomes est tirée aléatoirement.
- ▶ Evaluation : chaque chromosome est décodé, puis évalué.
- ▶ Création d'une nouvelle population de N chromosomes :
 - ▶ Sélection : création d'une nouvelle population de M chromosomes : $M < N$ (souvent $M = N/2$) par la méthode de sélection appropriée.
 - ▶ Reproduction : possibilité de croisement et mutation au sein de la nouvelle population. On retrouve une population de N chromosomes.
- ▶ Retour à la phase d'évaluation jusqu'à l'arrêt de l'algorithme (critères identiques aux autres heuristiques : temps, nombre d'itération, convergence ...).

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base

Pseudo-Code

POP, *POP'* et *QUALITE*: tableaux de taille *N*

POP := initialiser population

$$f(\mathbf{x}) := \min_{1 \leq i \leq N} [f(\mathbf{x}_i)]$$

$$f_{min} := f(\mathbf{x})$$

$$\mathbf{x}_{min} := \mathbf{x}$$

REPETER

QUALITE (évaluer la population *POP*)

REPETER (phase de reproduction génétique)

sélection

croisement

mutation

JUSQU'A *POP'* remplie

POP := sélectionner nouvelle population (*POP*, *POP'*)

$$f(\mathbf{x}) := \min_{1 \leq i \leq N} [f(\mathbf{x}_i)]$$

SI $f(\mathbf{x}) < f_{min}$

$$f_{min} := f(\mathbf{x})$$

$$\mathbf{x}_{min} := \mathbf{x}$$

FIN DE SI

JUSQU'A conditions d'arrêt satisfaites

Présentation
Générale

Codage

Fitness

Sélection

Croisement

Mutation

Autres paramètres

Cycle de Base