

Algorithmes génétiques - Fondements mathématiques

- Il n'existe pas de preuve de convergence asymptotique des algorithmes génétiques vers un optimum global (comme le recuit simulé, par exemple).
- Il existe toutefois des résultats théoriques indiquant que la recherche est "sensée"

Théorie des schémas

- Le traitement explicite des chromosomes, entraîne également le traitement des schémas sous-jacents.
- Nous allons donc étudier l'évolution des schémas dans une population de chromosomes, soumis aux opérateurs de reproduction, mutation et croisement.
- Soit un ensemble de chromosomes A_i , $i=1, \dots, n$ constituant une population $A(t)$ à la génération t .
- Un chromosome est construit à partir de l'alphabet $V = \{0,1\}$
- Un schéma est construit à partir de l'alphabet $V^+ = \{0,1,*\}$, où $*$ est un caractère qui peut représenter 0 ou 1.
- Un schéma (hyperplan, modèle de similarité) représente donc un sous-ensemble de chromosomes possédant des caractéristiques communes.

Exemple

| | |
|----------------------|------------------|
| Chromosome 1: 110001 | Performance: 90 |
| Chromosome 2: 010101 | Performance: 10 |
| Chromosome 3: 111001 | Performance: 100 |
| Chromosome 4: 100101 | Performance: 5 |
| Chromosome 5: 000011 | Performance: 95 |
| Chromosome 6: 010111 | Performance: 90 |
| Chromosome 7: 001100 | Performance: 5 |
| Chromosome 8: 101010 | Performance: 5 |

Les chromosomes 1 et 3 appartiennent au schéma 11****

Le schéma 11**** contient les chromosomes 1 et 3

Théorie des schémas

- 3^l schémas peuvent être construits à partir de chromosomes ou chaîne de bits de longueur l .
- chaque chromosome appartient à 2^l schémas
- Dans une population comptant n chromosomes de longueur l , on retrouve au plus $n2^l$ schémas.
- Tous les schémas ne sont pas égaux:

(a) Certains schémas sont plus spécifiques que d'autres.

Ordre d'un schéma H ou $o(H)$: le nombre de positions fixées (donc, tout schéma H contient $2^{l-o(H)}$ chromosomes)

110011 est un schéma d'ordre 6 qui ne contient qu'un seul chromosome

*1*01* est un schéma d'ordre 3 et contient 2^3 chromosomes

***** est un schéma d'ordre 0 qui contient 2^6 chromosomes

(b) Certains schémas couvrent une plus grande portion du chromosome

Longueur Effective d'un schéma H ou $\delta(H)$: différence entre la première et la dernière position fixée

110011 est un schéma de longueur effective $6-1=5$

*1*01* est un schéma de longueur effective $5-2=3$

***** est un schéma de longueur effective 0

- Ces caractéristiques vont nous permettre d'analyser l'effet des opérateurs de reproduction, croisement et mutation sur l'évolution des schémas dans la population

Théorie des schémas

Effet de la reproduction

- On suppose que dans une population de n chromosomes à la génération t , on a m chromosomes appartenant à un schéma H , noté $m(H,t)$.
- Durant la reproduction, un chromosome i est sélectionné avec probabilité:

$$p_i = f_i / (\sum_{j=1 \dots n} f_j).$$

- Il s'agit donc d'une loi de Bernouilli avec probabilité de succès p_i .
- Puisqu'on répète le processus n fois de façon indépendante, on obtient une loi binomiale $B(n, p_i)$, dont l'espérance est np_i . Le nombre espéré d'enfants pour le chromosome i est donc:

$$n p_i = n f_i / (\sum_{j=1 \dots n} f_j) = f_i / f_{\text{moy}}$$

- Ainsi, le nombre espéré de chromosomes appartenant au schéma H à la génération $t+1$, noté $m(H, t+1)$ devient:

$$\begin{aligned} m(H, t+1) &= \sum_{i=1 \dots m(H,t)} f_i / f_{\text{moy}} \\ &= m(H,t) \sum_{i=1 \dots m(H,t)} f_i / (m(H,t) f_{\text{moy}}) \\ &= m(H,t) f_{\text{moy}}(H) / f_{\text{moy}} \end{aligned}$$

- Donc, le nombre de représentants du schéma croît selon le rapport entre la performance moyenne des chromosomes appartenant au schéma et la performance moyenne des chromosomes dans l'ensemble de la population.
- Si la performance moyenne d'un schéma est supérieure à la moyenne de la population, le nombre de représentants augmente, sinon il diminue (et ceci se produit pour tous les schémas présents dans la population).

Exemple

Si $f_{\text{moy}}(H) = f_{\text{moy}} + c f_{\text{moy}}$, $c > 0$, on obtient:

$$\begin{aligned} m(H, t+1) &= m(H, t) f_{\text{moy}} + c f_{\text{moy}} / f_{\text{moy}} = m(H, t) (1 + c) \\ m(H, t) &= m(H, 0) (1 + c)^t \end{aligned}$$

Théorie des schémas

Effet du croisement

- Le croisement peut détruire certains schémas

Exemple:

| | | | | | | |
|---|---|--|---|---|---|---|
| 1 | 1 | | 0 | 0 | 1 | 1 |
| 1 | * | | * | * | * | 1 |
| * | * | | 0 | 0 | * | * |

- Un schéma survit si le point de coupure survient à l'extérieur des positions fixées extrêmes du schéma (autrement, le schéma risque d'être détruit).
- La probabilité de survie du schéma est donc: $1 - (\delta(H) / l-1)$, où $l-1$ est le nombre de points de coupure possible le long du chromosome.
- Supposant que le taux de croisement est p_c , la probabilité de survie du schéma p_s peut donc s'exprimer ainsi:

$$p_s \geq 1 - p_c (\delta(H) / l-1) \quad \text{où } l-1 \text{ est le nombre de points de coupure possibles le long du chromosome}$$

- Le nombre espéré de représentants du schéma H devient donc:

$$m(H,t+1) \geq (m(H,t) f_{\text{moy}}(H)/f_{\text{moy}}) (1 - p_c \delta(H)/l-1)$$

- L'évolution du nombre de représentants du schéma H dépend maintenant de deux facteurs:
 - (a) la performance moyenne des représentants du schéma H par rapport à la performance moyenne dans l'ensemble de la population
 - (b) la longueur effective du schéma H

Théorie des schémas

Effet de la mutation

- La mutation peut également détruire certains schémas
- Supposant un taux de mutation p_m , un schéma survit si toutes les positions fixées survivent, i.e.

$$(1-p_m)^{o(H)}$$

- Si on suppose p_m petit, l'approximation suivante est valable:

$$(1-p_m)^{o(H)} \approx 1 - o(H)p_m$$

- Le nombre espéré de représentants du schéma H devient donc:

$$m(H,t+1) \geq (m(H,t) f_{\text{moy}}(H)/f_{\text{moy}}) (1 - p_c \delta(H) / l-1) (1 - o(H)p_m)$$

$$m(H,t+1) \geq (m(H,t) f_{\text{moy}}(H)/f_{\text{moy}}) (1 - p_c \delta(H) / l-1 - o(H)p_m)$$

- L'évolution du nombre de représentants du schéma H dépend maintenant de trois facteurs:
 - (a) la performance moyenne des représentants du schéma H par rapport à la performance moyenne de l'ensemble de la population
 - (b) la longueur effective du schéma H
 - (c) l'ordre du schéma H
- "Schema Theorem": les schémas qui croissent le plus rapidement dans la population d'une génération à l'autre sont ceux dont les représentants ont une performance moyenne supérieure à la moyenne de la population, qui sont courts, et qui comptent peu de positions fixées (blocs de construction)
- "Building-block Hypothesis"

Théorie des schémas

Remarques

- Cette théorie suppose que tous les schémas utiles sont dans la population initiale, et qu'une solution optimale peut être obtenue en redistribuant ces schémas entre les chromosomes au cours de l'évolution.
- La théorie "oublie" un aspect fort utile des opérateurs de croisement et de mutation, soit l'introduction de nouveaux schémas au sein de la population (qu'on ne retrouve pas au sein de la population initiale).

Bandit armé à K bras (K-armed bandit problem)

Bandit armé à 2 bras (slot machine)

- Le joueur peut actionner l'un ou l'autre bras
- Un bras donne une récompense moyenne μ_1 , avec une variance σ_1 . L'autre bras donne une récompense moyenne μ_2 , avec une variance σ_2 . De plus, on suppose $\mu_1 \geq \mu_2$.
- Évidemment, le joueur ne connaît pas l'identité du bras le plus payant en moyenne.
- Si on dispose d'un certain nombre d'essais n , quelle est la meilleure stratégie en termes de perte espérée minimale?
- Exemple de stratégie: on actionne les deux bras à tour de rôle pour $2m$ essais (exploration), et on retient le plus payant pour les $(n-2m)$ essais restants (exploitation).
- deux sources de perte:
 - (1) utilisation du bras le moins payant en phase d'exploration (m essais)
 - (2) choix du bras le moins payant en phase d'exploitation ($n-2m$ essais).
- Espérance de la perte:

$$P(m,n) = (\mu_1 - \mu_2) [(n-m)q(m) + m(1-q(m))]$$

où $q(m)$ est la probabilité que le bras le moins payant est le meilleur bras observé après m essais sur chaque bras

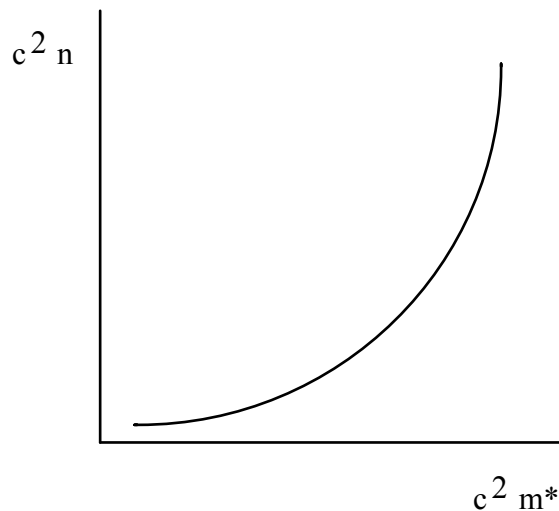
première composante de la somme: on a choisi le mauvais bras à la fin de l'exploration.

deuxième composante de la somme: on a choisi le bon bras à la fin de l'exploration

Bandit armé à K bras (K-armed bandit problem)

Bandit armé à 2 bras

- La valeur optimale pour m est l'objet d'un compromis:
 - plus m est grand, plus la perte durant l'exploration est grande. Par contre, les probabilités de perte durant l'exploitation diminuent.
 - plus m est petit, plus la perte durant l'exploration est petite. Par contre, les probabilités de perte durant l'exploitation augmentent.
- La valeur optimale m^* dépend de n , et du ratio entre le signal et le bruit, i.e. $c = (\mu_1 - \mu_2) / (\sigma_1^2 + \sigma_2^2)^{0.5}$
- Si on dessine la courbe $c^2 n$ en fonction de $c^2 m^*$, on observe que le nombre total d'essais n augmente exponentiellement par rapport à m^* . Donc, lorsque n augmente, le nombre d'essais accordé au meilleur bras observé ($n - m^*$) croît exponentiellement par rapport au nombre d'essais m^* accordé à l'autre bras.



Bandit armé à K bras (k-armed bandit problem)

- Une généralisation du problème est le bandit à K bras.
- La même conclusion s'impose dans ce cas: le nombre d'essais alloué au meilleur bras augmente exponentiellement par rapport aux autres bras.
- Supposons qu'un ensemble de schémas sont en compétition lors de la reproduction:

définition: deux schémas H_1 et H_2 sont en compétition si pour toute position i , $i = 1, \dots, l$, on a:

$((H_{1i} = H_{2i} = *) \text{ ou } (H_{1i} \neq *, H_{2i} \neq *))$ et

$(H_{1i} \neq H_{2i}, \text{ pour au moins une position } i).$

Exemple:

| | | | | | |
|---|---|---|---|---|---|
| * | 0 | 0 | * | 0 | * |
| * | 0 | 0 | * | 1 | * |
| * | 0 | 1 | * | 0 | * |
| * | 0 | 1 | * | 1 | * |
| * | 1 | 0 | * | 0 | * |
| * | 1 | 0 | * | 1 | * |
| * | 1 | 1 | * | 0 | * |
| * | 1 | 1 | * | 1 | * |

Ces 8 schémas sont en compétition, car tout chromosome n'appartient qu'à un et un seul de ces schémas.

On a donc ici un problème de bandit armé à 8 bras.

- Il est important de noter que dans un algorithme génétique, plusieurs problèmes de bandit armé à K bras sont résolus en parallèle.

Par exemple, pour tout $i = 1, \dots, l$, on a:

(choisir i parmi l) problèmes de bandit armé à 2^i bras.