

RAPPORT DE PROJET

Algorithme génétique

Auteurs :

Thibault ODOR

Joan TERIIHOANIA

Jean-Baptiste BES

Lino FAVOT

Superviseur :

Marin BOUGERET

12 mai 2020



Pour l'obtention du DUT Informatique

algorithme - génétique - évolution - simulation - informatique

Table des matières

1	Cahier des charges	6
1.1	Principe général	6
1.1.1	Sélection	7
1.1.2	Croisement	7
1.1.3	Mutation	8
1.2	Application	9
2	Rapport technique	9
2.1	Conceptualisation	9
2.1.1	Environnement	9
2.1.2	Problème	9
2.1.3	Mouvement	9
2.1.4	Individu	9
2.1.5	Sélection	10
2.1.6	Croisement	10
2.1.7	Mutation	10
2.1.8	Evaluation	10
2.2	Architecture du code	10
3	Gestion de projet	12
3.1	Démarche personnelle	12
3.1.1	Méthode de développement	12
3.1.2	Gestion du code	12
4	Manuel	13

Remerciements


Nous souhaitons remercier tout particulièrement notre tuteur, BOUGERET Marin ainsi que l'équipe pédagogique de notre formation qui nous a accompagné tout le long de notre projet.

Glossaire

Au sens d'un algorithme génétique :

- *Individu* : Objet algorithmique correspondant à une solution potentielle.
- *Gène* : Donnée innée à un individu, il s'agit de la solution de l'individu. Sa forme peut varier (String, Integer, etc).
- *Élément d'un gène* : Prenant un gène de type chaîne de caractères, un caractère est un élément du gène.
- *Partie d'un gène* : Une partie d'un gène est une section découpée du gène de \mathbf{x} élément.
- *Valeur sélective* : Représentation numérique de la performance d'un individu à la manière d'une note.

Introduction

Au travers de ce document, des termes et expressions cliquables sont disponibles re-dirigeant vers de la documentation en ligne afin d'apporter un support documentaire éventuel aux idées décrites. Ces éléments cliquables sont indiqués par la présence d'une icône les précédant. Par exemple :  Wikipedia.

Dans ce projet, nous produirons un algorithme génétique et étudierons les résultats obtenus afin d'en mesurer la pertinence. Dans cette optique, nous partirons d'un cas établi afin d'en déduire des observations par rapport aux résultats attendus et obtenus. Observations qui nous permettront de mesurer l'utilité d'un algorithme génétique.

L'algorithme génétique n'est pas un algorithme en soit, mais un modèle permettant de mimer le concept d'évolution génétique par génération. Il permet de créer pour une situation donnée une solution correspondante la plus optimale possible en s'adaptant à l'environnement fourni.

L'objectif de ce projet est d'une première part, de **produire** un algorithme et une structure algorithmique permettant de calculer et de **simuler** l'évolution d'une population d'individus dans un environnement pré-déterminé. Puis, d'une seconde part, d'**observer** leur comportement afin d'en extrapoler des résultats que nous tenterons d'expliquer et d'évaluer ;

Enfin, à partir de nos observations, d'**étudier** la pertinence de l'utilisation d'un algorithme génétique pour la résolution d'un problème donné au regard des résultats obtenus. *Est-ce qu'une résolution à la main ou via un algorithme classique donne de meilleurs résultats au regard de sa simplicité et du temps requis à son implantation ?*

Dans un environnement déterminé surnommé **problème**, des individus surnommés **solutions** sont confrontés afin de déterminer lequel, selon des critères donnés, a le plus de valeur et sera sélectionné pour l'environnement confronté à le surmonter. Pour un problème établi, nous émettons des "*hypothèses*" de solutions qui, au fur et à mesure de la simulation, établiront selon leur évaluation leur efficacité à résoudre le problème édicté. *On peut considérer que la solution de chaque individu est un gène G qui leur est inné et qu'ils peuvent transmettre, croiser et muter.*

Cette situation, bien que relativement simple, peut avoir des applications variées. Qu'il s'agisse d'une simulation naturelle ou d'un environnement économique où les individus sont des entreprises pour en déterminer le comportement le plus efficient pour, par exemple, un résultat lucratif maximal.

Il est important de garder en tête que l'algorithme et toutes les phases qui le composent, notamment les phases de sélection, mutation et croisement, doivent être optimisés. Etant donné qu'un algorithme génétique se base sur une population avec un nombre élevé pour avoir un résultat adéquat.

Le choix des paramètres de la simulation, du nombre d'individus générés et du nombre

de cycle notamment, sont des facteurs importants. Leur choix permet d'obtenir un ratio coût-résultat idéal. Un trop grand nombre d'individus et de cycles peut rallonger la durée de calcul inutilement. Un trop faible nombre peut produire des individus qui ne sont pas assez performants.

Ajouter annonce du plan ici

1 Cahier des charges

1.1 Principe général

L'algorithme génétique se base sur une évolution par sélection naturelle. Le principe de *sélection naturelle* est basé sur la théorie de *Darwin*. Ce principe repose sur la base d'une valeur sélective définie à chaque individu parmi une masse plus ou moins diffuse pour obtenir, sur une période donnée, une souche détenant les paramètres les mieux adaptés à un environnement sélectionné sur des critères d'évaluation déterminés. L'individu *a* faisant parti d'une population d'une génération à un instant *t* et pouvant passer ses "gènes" *G* à un individu *b* fils de la génération à l'instant *t+1* en passant par une phase d'évaluation, sélection, croisement, mutation puis réinsertion.

Nous pouvons représenter ce schéma de réalisation par le suivant :



Important : Afin de préserver un potentiel d'évolutivité, l'algorithme ne doit ni être eugéniste (*Eugénisme*) ni élitiste. Les individus de la population, qu'ils soient faiblement ou moyennement évalués ne doivent pas être supprimés pour conserver une certaine diversité au sein de la population et éviter une stagnation. Par la même occasion, la sélection des individus à croiser et muter ne doit pas totalement ignorer les individus faiblement ou moyennement évalués.

1.1.1 Sélection

L'algorithme sélectionne un certain nombre d'individu (ou part de la population définie) arbitrairement. Les individus sélectionnés subiront les phases suivantes de croisement et mutation. La façon dont les individus sont sélectionnés peut varier :

1. **Sélection aléatoire simple** : On sélectionne les individus de manière totalement aléatoire. Elle permet d'obtenir une sélection difforme et variée et d'ainsi varier les individus qui suivront.
2. **Sélection aléatoire pondérée** : On attribue à chaque individu une **valeur sélective** puis, en se basant sur celle-ci, on sélectionne les individus de manière aléatoire. Aussi, un individu dont la valeur est plus élevée aura plus de chance d'être sélectionné. *Il est important de garder en tête que l'algorithme ne doit pas être "sur-sélectif". Il n'est donc pas pertinent de rendre la valeur des individus les moins évalués nulle afin de leur donner une chance d'être sélectionné, évoluer et éventuellement s'améliorer.* Elle permet d'obtenir une sélection difforme et variée tout en privilégiant les gènes les plus efficaces sans ignorer les autres.
3. **Sélection prédéterminée** : On attribue un nombre (ou pourcentage) déterminé lors du lancement de l'algorithme qui sélectionnera (en fonction de ce nombre ou pourcentage) les individus parmi une part préalablement déterminée de la population (en fonction de leur évaluation). *Entre autre, il sera opportun d'effectuer notre sélection en grande majorité parmi la population la mieux évalués, et parmi la population dont l'évaluation est moyenne et faible.*

Note : La méthode prédéterministe peut être mixée avec les méthodes aléatoires. La méthode de sélection peut différer en fonction du résultat escompté. On privilégiera dans ce projet la méthode de **sélection aléatoire pondérée**.

1.1.2 Croisement

Parmi la population sélectionnée par la phase précédente, l'algorithme effectue un croisement de gène. Celui-ci peut se faire entre deux individus ou plus (principalement 2). Cette phase permet de produire une population (dont les individus sont principalement bien évalués) dont les gènes sont croisés et diversifiés par rapport à la population de départ. *La persistance des individus dont les performances sont faibles ou moyennes leur permet de s'améliorer grâce aux individus les mieux évalués.*

Note : Usuellement, la population d'entrée **a** est supprimée et remplacée par la population de sortie **b**. Il faut donc que cette phase produise pour des croisements entre **n** individus, **n** individus croisés afin de garder une population au même nombre que celle d'entrée.

Les méthodes de croisement d'individus existent sous plusieurs formes :

1. **Croisement aléatoire simple** : Chaque élément (ou partie) du gène des n individus sont parcourus afin de produire un individu. L'algorithme choisi aléatoirement l'élément (ou partie) de l'individu A, B, etc. Cela produit alors un nouveau gène mêlant les éléments (ou parties) de l'individu A, B, etc. Cette opération est effectuée n fois afin de produire une population de quantité équivalente.
2. **Croisement aléatoire hiérarchisé (ou pondéré)** : On considère que chaque type de gène dispose d'un coefficient (ou valeur sélective à la manière d'un individu). Ce dernier est utilisé comme base afin de pondérer le choix entre les gènes des individus croisés. Un gène à coefficient élevé a des chances plus élevées d'être sélectionné. Il en existe deux formes :
 - **Statique** : Ici, les gènes disposent d'une hiérarchie (dont les valeurs sont fixées arbitrairement au départ de la simulation). Chaque élément du gène des individus possèdent un coefficient qui identifie son niveau par rapport à un autre. Par exemple, la hiérarchie $G > B > N$ (dont les coefficients sont $3 > 2 > 1$) où $B=2$ est supérieur à $N=1$ mais inférieur à $G=3$. De la même manière qu'une méthode de sélection aléatoire pondérée, l'algorithme choisi un élément aléatoirement entre l'individu A et B où l'élément dont la valeur hiérarchique est supérieure a plus de chance d'être sélectionnée.
 - **Dynamique** : Cette méthode reprend le fonctionnement de la méthode de croisement aléatoire pondéré statique mais dans laquelle la hiérarchie des gènes n'est pas statique mais dynamique. C'est-à-dire qu'elle évolue à chaque génération et chaque population en fonction des gènes de cette dernière. Ainsi, une population dans laquelle le gène G est dominant verra sa hiérarchie telle que $G > \dots$. Bien entendu, il est possible d'effectuer un *décalage* entre une génération n et une génération $n+k$ tel que : $h(n)$ de $p(n)$ est utilisée sur $p(n+1)$ où h (hiérarchie), p (population), n (numéro de génération) et k (nombre de génération de décalage). Le dynamisme d'une hiérarchie peut se décliner sous deux formes :
 - **Dynamisme général** : Une population possède une hiérarchie commune. Celle-ci est recalculée au début de la phase de croisement en fonction de la quantité du gène au sein de la population totale ou sélectionnée.
 - **Dynamisme unitaire** : Pour une population avec un gène de k élément, il existe k hiérarchie commune recalculée au début de la phase de croisement pour chaque k ème élément e du gène G (noté $e[k]$) qui dispose d'une hiérarchie de l'élément k notée $h(e[k])$ dont le dynamisme évolue en fonction de la quantité des gènes de la population du k ème élément. *Soit la hiérarchie $h(e[1])$, la hiérarchie de l'élément 1.*

1.1.3 Mutation

La population résultant de la phase de croisement est alors soumise à la phase de mutation. Durant cette dernière, chaque élément (ou partie) de leur gène a la possibilité de muter (*Les chances de mutation sont par usage très faibles dans les alentours de 0.05%*) en un autre gène déterminé aléatoirement. Cette phase a pour but de donner une chance d'évolution et de diversification à la population.

1.2 Application

2 Rapport technique

2.1 Conceptualisation

Afin d'obtenir les résultats escomptés par l'objectif de ce projet, il est opportun de réaliser un exercice. Avant de produire un algorithme génétique généralisant les cas d'utilisation d'une évolution, nous pouvons nous consacrer à la réalisation d'un cas concret et spécifique afin de l'utiliser comme base du cas général. Comme cas concret, nous prendrons le suivant :

2.1.1 Environnement

Nous disposons d'un plateau de $W \times W$ cases. Chaque case peut contenir **Pièce** ou pas (*Ces pièces sont au nombre $nbPièce$ réparties aléatoirement sur le plateau*). D'une position initiale **X** du pion **Pion** et d'un entier **n**.

2.1.2 Problème

*Quel est l'enchaînement de mouvement/pas **M** à partir de la position **X** qui permet de récupérer le plus de **Pièce** avec **n** pas ?*

2.1.3 Mouvement

Un mouvement **M** peut être **Haut**, **Bas**, **Gauche**, **Droite** relatif à la position actuelle de **Pion** et à l'axe du plateau. Un mouvement **M** est valide tant qu'il ne fait pas sortir le pion en dehors des limites du plateau.

2.1.4 Individu

Chaque *individu* de la population contiendra une solution/gène **G** : une suite de **n** caractère(s) chacun représentant un mouvement/pas de type **M**.

2.1.5 Sélection

2.1.6 Croisement

2.1.7 Mutation

2.1.8 Evaluation

Chacun *individu* sera évalué et obtiendra un capital d'évaluation ou **valeur sélective** qui sera calculée en fonction du nombre de **Pièce** qu'il aura récupéré, de la distance parcourue (nombre de pas). Au cours du développement, la question des enchaînements de mouvement menant dans les murs du plateau. Trois hypothèses sont envisageables :

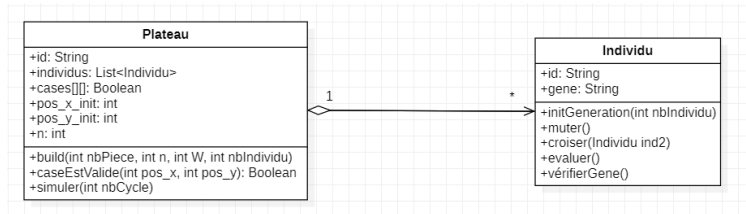
- Le sanctionnement par un score -infini. *Qui aura pour effet de rendre les individus ayant les dits enchaînements comme des individus ayant des chances d'être sélectionnées nulles. Néanmoins, comme précisé plus haut dans ce document, il est opportun de conserver tout individu comme conserver le potentiel évolutif de la population.*
- Le sanctionnement par un score -1. *Qui aura pour effet de rendre les individus concernés moins efficaces. De plus, ils ne seront pas écartés de la sélection.*
- Aucun sanctionnement. *En effet, selon notre méthode d'évaluation, le score des individus est retiré du **nombre de pas**. Ceci est en soit une sanction, étant donné qu'un mouvement dans un mur équivaut à un pas perdu que l'individu n'utilisera pas pour obtenir une pièce.*

2.2 Architecture du code

L'architecture du code se base sur la logique orientée objet du langage Java qui sera utilisé pour le développement de l'algorithme génétique. Afin de rendre la structure du code générale à toute sorte d'application de l'algorithme, il sera opportun d'utiliser des classes **abstraites** ou des **interfaces** (vues en cours de *Programmation orientée objet*) et l'intégration des fonctions essentielles à l'exécution de l'algorithme. Visiblement, nous aurons besoin d'au moins deux classes pour l'environnement (**Plateau**) et l'individu (**Pion**). A partir de cette situation, il nous est possible d'identifier plusieurs classes permettant la réalisation de cet algorithme :

- **Plateau**;
- Et **Individu**.

Les relations des classes peuvent être exprimées avec le schéma suivant :



Au début du projet, il avait été jugé opportun de créer une classe intermédiaire qui se chargerait du contrôle de chaque étape de vérification : **Mouvement**. Mais après réflexion, la création de cette classe rendait la classe **Individu** inutile, étant donné que presque toutes, si ce n'est toutes, les méthodes de **Individu** appelaient les méthodes de **Mouvement**. Il est donc préférable de se tenir aux deux classes ci-dessus pour le moment.

3 Gestion de projet

3.1 Démarche personnelle

3.1.1 Méthode de développement

Nous avons opté pour la méthode agile qui se déroule sur une période bouclée d'une semaine. A la fin de chaque sprint, nous effectuons une réunion avec notre tuteur afin de discuter de nos avancées, des éventuelles modifications et améliorations à faire et de la suite du travail. Cette méthode de travail correspond à notre contexte de développement où nous développons dans le but d'observer un résultat plutôt que d'essayer d'obtenir un résultat. Ainsi, nous pouvons, afin nos observations, faire des modifications et réfléchir sur l'algorithme afin de l'améliorer.

3.1.2 Gestion du code

Qu'il s'agisse de la phase de versionning ou de travail collaboratif, plusieurs plateformes ont été utilisées afin de permettre un partage efficace du code. En outre, nous avons utilisé, utilisons ou avons tenté d'utiliser des applications telles que :

- **Github** (principalement) ;
- **Floobits** (plateforme de codage collaboratif en temps réel) ;

Néanmoins, après avoir essayé la plateforme *Floobits*, nous avons observé que cette méthode de travail et de fonctionnement n'était pas adapté à notre méthode de fonctionnement et n'incluait pas un suivi d'historique des changements qu'offrait Github.

Nous avons utilisé également un package implémenté dans Java dans notre code ou afin de tester l'application : *Junit* pour produire des tests unitaires et vérifier le fonctionnement de nos méthodes et fonctions.

Au départ du projet, nous avons débuté le code en créant des dossiers séparés au sein du dépôt afin de pouvoir créer une version de notre projet de manière individuelle. Après quelques semaines, nous avons rapidement conclu qu'il s'agissait d'une mauvaise idée, car menant à de trop grandes différences et un travail de synchronisation supplémentaire qui aurait pu être évité. Après cette conclusion, nous avons donc refactorisé le code dans une version unique et commune sur laquelle chaque membre aura des tâches à faire spécifiques, en profitant de cette refactorisation pour mettre en commun nos idées.

4 Manuel

5 Bibliographie

Il est possible de coupler cet algorithme à un réseau de neurones artificiels ou ↗ *Neural Network* basé sur les idées du psychologue ↗ *Franck Rosenblatt*. Notamment pour le calcul des poids de neurones.

Quelques lectures peuvent être trouvées à ce sujet dont (en anglais) :

- ↗ *"Deep Learning with Darwin: Evolutionary Synthesis of Deep Neural Networks"* de M. J. Shafiee, A. Mishra et A. Wong ;
- ↗ *"Evolutionary Robot Behaviors Based on Natural Selection and Neural Network"* dans *Artificial Intelligence Applications and Innovations* de M. Bramer et V. Devdizic ;
- ↗ *"Evolving Neural Networks"* et ↗ *"A biology-inspired neural network evolving through natural selection"* de Souvik Das ;
- Ou encore, un jeu de simulation ↗ *"The Bibites"* développé par Léo Caussan dont le développement peut être suivi depuis sa chaîne ↗ YouTube.