



SORBONNE UNIVERSITÉ

2I013 : PROJET (APPLICATION)

Sujet :
IA Football

Fangzhou Ye
Angelo Ortiz

Table des matières

Introduction	3
I Architecture logicielle	4
II Stratégies	5
1 Un joueur	5
2 Deux joueurs	5
3 Quatre joueurs	6
III Méthodes d'optimisation	8
4 Recherche en grille	8
5 Algorithmes génétiques	8
6 Apprentissage automatique	9
Conclusion	11

Introduction

Présentation

Le projet a consisté en le développement d'intelligences artificielles de joueurs de football. Le simulateur du jeu étant déjà fourni, notre travail a été d'implémenter des stratégies de jeu pour mieux réussir les matches.

L'objectif du travail a été d'apprendre à bien mener un projet long sur un nouvel environnement de développement, à savoir Python. Sachant que travail a été fait en binôme, on a eu besoin d'un outil collaboratif, en l'occurrence **git** qui est un logiciel de gestion de versions décentralisé. On a ainsi appris à gérer l'avancement d'un projet à travers la méthode par fonctionnalités.

Aperçu

Voici un aperçu du simulateur de football lors d'une partie de notre équipe *ChPerFusion* en rouge face à l'équipe d'un de nos camarades de classe.

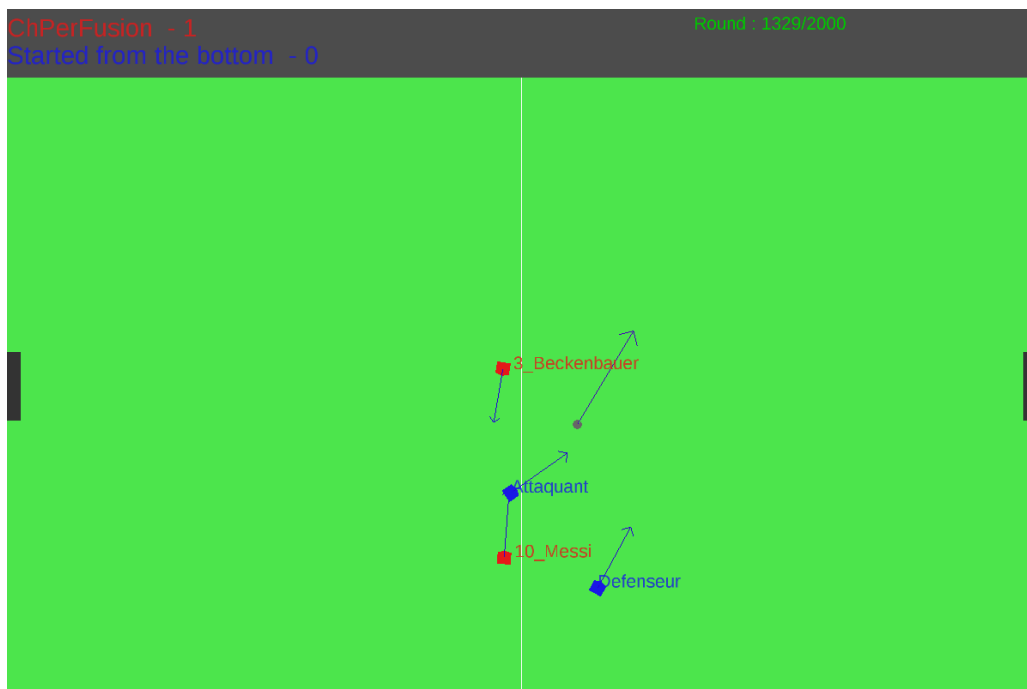
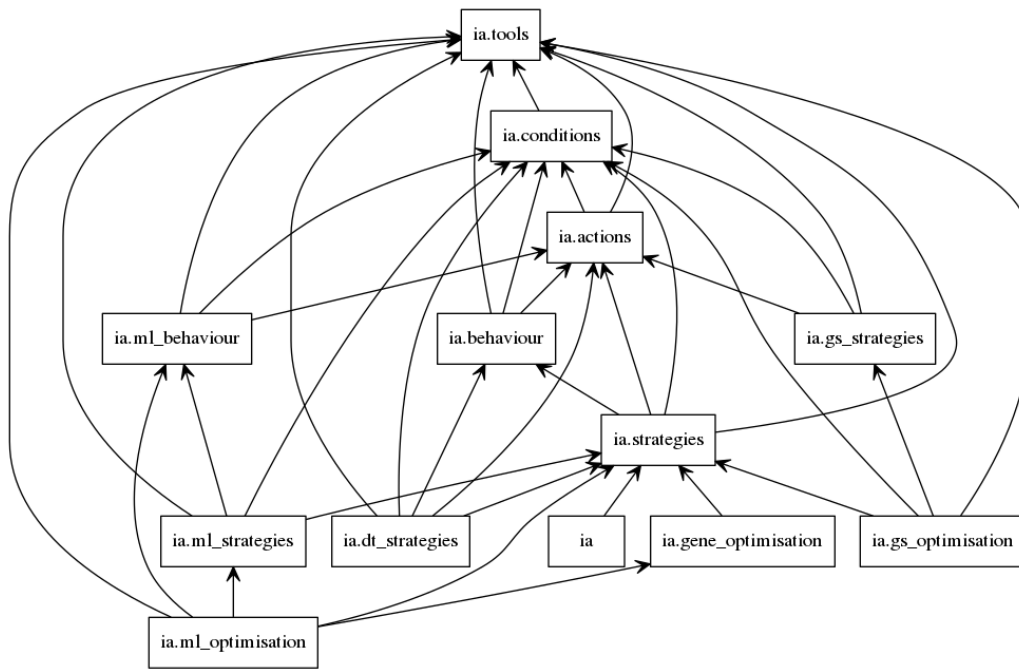


Diagramme de dépendances

Le projet a été organisé en un module **ia** qui contient l'implémentation de nos joueurs et en divers fichiers exécutables permettant de lancer les algorithmes d'optimisation.

On vous présente le graphe de dépendances du module.



Première partie

Architecture logicielle

Dans cette première partie, on vous expliquera les choix faits concernant l'architecture logicielle.

Deuxième partie

Stratégies

Dans cette deuxième partie, on vous détaille le comportement de nos différents joueurs. Compte tenu de la taille du terrain de jeu, le nombre maximum de joueurs par équipe est limité à quatre. Ainsi, il y a trois catégories de matches selon le nombre de joueurs par équipe : un, deux ou quatre.

1 Un joueur

Pour l'équipe composée uniquement d'un joueur, on a développé un seul joueur que l'on appelle *Fonceur*.

Fonceur

Il s'agit du point de départ du projet. Le but des premières semaines étant de se familiariser avec le code source fourni, on a créé une stratégie très simple et naïve. Ainsi, il faisait tout au maximum : il devait se rapprocher de la balle à toute vitesse lorsqu'il ne la contrôlait pas et il frappait avec toute sa puissance dans le cas contraire. Compte tenu que ce comportement ressemblait beaucoup aux autres fonceurs de la classe à une paramétrisation près et que les matches finissaient en nul, on a ajouté les fonctionnalités suivantes : l'avancée balle au pied et le dribble. Ceci a rapporté de bons résultats contre les équipes restant naïves, mais n'a pas suffi pour battre les équipes plus développées. En effet, il est très difficile de devancer un adversaire qui fait une pression sur le ballon à une distance moyenne de sorte que les tentatives de dribbles du fonceur échouent. De plus, ses essais de dribble dans sa propre surface de réparation étaient souvent insuffisants pour échapper au danger que signifie l'adversaire venant le presser. Il dégage à présent la balle le plus loin possible pour pallier cette difficulté.

Cette stratégie est restée assez basique à l'égard de celles de nos camarades de classe vu les résultats des derniers tournois hebdomadaires et de nos tests. Autrement dit, le dribble est très peu réussi et les contre-attaques arrivent lorsque le joueur est dans sa course vers l'avant, ce qui le surprend et fait que l'adversaire se trouve facilement seul face à notre cage.

2 Deux joueurs

Cette catégorie a été celle sur laquelle on a consacré la plupart du projet. On compte trois joueurs fonctionnels que l'on appelle *Attaquant*, *Gardien* et *Défenseur*.

Attaquant

Il a une vocation plutôt offensive. Son comportement initial était celui d'un attaquant classique. Sa conduite sans la balle dépendait de la zone dans laquelle se trouvait le porteur de la balle : il le pressait si l'adversaire était éloigné de sa surface et se décalait vers les côtés pour faciliter les contre-attaques dans le cas contraire. Quant à son style de jeu avec la balle au pied, il avançait de petites distances sur le terrain et frappait la balle lorsqu'il se trouvait dans la surface de réparation adverse. Tout comme pour le fonceur, il perdait la balle trop facilement. C'est pourquoi on a ajouté le dribble aléatoire, ce qui l'a rendu imprévisible face aux tentatives adverses de lui arracher le ballon. Lorsque le contrôle de la balle avait lieu en sa zone défensive et qu'un adversaire était si proche de lui qu'il pouvait aussi la frapper, la

puissance de son dribble était dépassée par celle de la frappe. On a ajouté le dégagement vers un point quelconque tel que le tir soit le moins risqué possible pour éviter ce genre de situations.

Puis on s'est rendu compte qu'il se trouvait en infériorité numérique face à l'équipe adverse dans la plupart des matches. Cette contrainte a motivé l'idée de faire monter le coéquipier dans le terrain et, par conséquent, le développement de l'action de passe de la balle. Ce comportement s'est révélé très efficace en situation d'attaque vu que les autres équipes utilisent seulement un attaquant, mais les contre-attaques s'avéraient très dangereuses. Celles-ci arrivaient comme résultat des dribbles échoués. Par ailleurs, on a remarqué que les adversaires restaient fixés dans l'axe tout en laissant dans les côtés des voies libres. Tout ceci nous a amené à mettre en place la toute dernière modification : le dribble latéral. Dans le milieu de terrain, le joueur essaie systématiquement de dribbler le joueur lui bloquant la trajectoire vers les côtés. Le comportement final est encore plus dangereux en attaque et les possibilités de contre-attaques sont à présent très mitigées.

Gardien

La prise de décisions initiale du gardien avait pour vocation de défendre la cage à tous les coups. En effet, il se positionnait dans le point central sur la ligne de sa cage et n'en sortait que pour dégager la balle lorsqu'un adversaire se rapprochait avec elle. Ce dégagement consistait en fait à faire une pseudopasse vers son coéquipier, c'est-à-dire qu'il l'adressait vers l'un de deux points axialement symétriques selon si le coéquipier se trouvait en la partie haute ou basse du terrain. Le point faible de cette conduite était que certains attaquants adverses tiraient vers notre cage depuis des longues distances et on essayait ainsi d'attraper la balle avec des trajectoires circulaires. On a donc décidé d'améliorer la trajectoire d'interception de la balle avec une certaine prévision de sorte à y arriver en ligne de droite et plus rapidement. De plus, sa position de base passerait à une certaine distance de sa cage alignée avec une potentielle trajectoire de la balle vers la cage.

On n'est jamais parvenus à perfectionner l'interception de la balle et on compte ainsi jusqu'à présent un taux non négligeable d'échec. De plus, comme dit précédemment, une attaque à un n'offrait pas d'avantage à l'égard des équipes de nos camarades. Ceci a impulsé le développement de la déjà mentionnée montée du gardien. Lorsque son coéquipier possède la balle, il monte dans le terrain pour proposer des possibilités de passe. Son comportement avec la balle a fait aussi objet de modifications : c'est exactement le même que celui de l'attaquant. On peut donc résumer le comportement de notre équipe à deux joueurs en l'expression footballistique « la meilleure défense, c'est l'attaque. »

Défenseur

Il est à noter que ce joueur ne fait pas partie de l'équipe préconisée. Il correspond à la dernière stratégie implémentée dans cette catégorie, à savoir celle d'un défenseur, qui n'a par contre qu'une vocation défensive. En effet, s'il a la balle, il s'en défait le plus rapidement possible soit à l'aide d'une passe vers son coéquipier si démarqué, soit via le dégagement le moins risqué possible. Quant à son effort défensif, il demeure à une certaine distance radial à partir de sa cage et s'approche de la balle en vue d'une interception s'il en suffisamment proche.

3 Quatre joueurs

Il s'agit de la catégorie que l'on a trouvé la plus intéressante en raison de la difficulté de la gestion des espaces et la répartition des fonctions. Tout comme pour la catégorie précédente, on compte trois joueurs fonctionnels que l'on repère d'ailleurs par les mêmes noms.

Le développement des stratégies suivantes a commencé lorsque leurs pairs de la catégorie précédente étaient proches de leurs versions finales. En effet, il n'y avait que le dribble latéral qui n'était pas encore mis en place lors du début de leur conception. Cette modification dans le dribble a été ajoutée à ces stratégies au même temps que pour leurs pairs.

Attaquant

Il y a deux différences en son comportement par rapport à son pair de la catégorie ci-dessus et celles-ci concernent toutes les deux la situation où il se trouve dans sa surface de réparation. La première fait référence à la récupération de la balle : il ne se précipite plus et fait de nouveau le contrôle, le dribble ou la passe nécessaire de la même manière que dans le milieu de terrain. La deuxième correspond à l'effort de marquer les adversaires sans marquage. En effet, les avancées adverses comportent plus de participants et il se révèle crucial de marquer les adversaires susceptibles de recevoir une passe pour intercepter la balle et lancer des contre-attaques rapides et efficaces.

Gardien

Ce joueur combine le comportement de son pair de la catégorie ci-dessus et de l'attaquant de cette catégorie à des valeurs des paramètres près. En effet, il suit les mêmes indications que l'attaquant lorsqu'il contrôle et réalise les mêmes mouvements sans ballon que le défenseur de l'équipe à deux joueurs.

Défenseur

Ce défenseur prend les mêmes décisions que celui de l'équipe à deux joueurs, à l'exception de sa prise de risque lors d'une interception. Effectivement, si un coéquipier se trouve dans une meilleure position pour intercepter la balle, i.e. il en est plus proche que le défenseur, celui-ci continue de positionner radialement à une certaine distance de sa cage. Cette modification est le produit de la constatation que l'effort défensif des autres joueurs de cette catégorie n'égale pas celui du défenseur et que l'abandon de position défensive entraîne des terribles risques pour le score.

Troisième partie

Méthodes d'optimisation

Dans cette troisième partie, on vous détaille les différents algorithmes appris tout au long du semestre pour l'amélioration des stratégies initialement proposées.

4 Recherche en grille

La première méthode d'optimisation présentée en cours a été celle de la recherche en grille. Celle-ci est utilisée pour l'optimisation d'une action.

Tout d'abord, on doit définir l'action à optimiser et le critère d'optimalité, et repérer les paramètres associés. Puis on obtient un ensemble de valeurs pour chaque paramètre. Pour les paramètres discrets, il suffit de prendre toutes les valeurs possibles, alors que pour les paramètres continus il faudra faire en amont une discrétisation, i.e. on divise l'intervalle délimité par des bornes définies selon un pas de précision. On compte ainsi des vecteurs dont les coordonnées correspondent chacune à un paramètre. On fait ensuite une recherche exhaustive sur tous les vecteurs obtenus : on teste chaque vecteur sous différentes conditions environnementales et on moyenne les évaluations du critère dans lesdites conditions. Finalement, on prend la valeur optimale et son vecteur associé.

Il est important de remarquer que la plupart de nos paramètres sont continus. Pour obtenir des valeurs plus précises et un meilleur comportement, on a besoin d'un pas de discrétisation très petit. Ainsi, on réalise qu'une augmentation du nombre de paramètres à optimiser a une très forte influence sur le temps d'exécution de l'algorithme mis en place. On en a déduit que le minuscule gain relatif nombre de paramètres-temps d'exécution ne convenait pas à moyen terme.

5 Algorithmes génétiques

Dans un deuxième temps, on est passés aux algorithmes génétiques. On en a mis en place une implémentation pour optimiser les équipes composées d'un ou deux joueurs.

Cette classe d'algorithmes se base sur la théorie de l'évolution. En effet, la notion de sélection naturelle est le mécanisme permettant de choisir des solutions potentielles de plus en plus meilleures.

La toute première chose à faire est la définition de la fonction $f : \mathbb{R}^n \rightarrow \mathbb{E}$ à optimiser, où n est le nombre de paramètres concernés. Un élément $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ est appelé *candidat* ou *chromosome*, et chacune de ses coordonnées x_i avec $i = 1, \dots, n$, appelée *propriété* ou encore *gène*, correspond à un paramètre à optimiser. Elles forment le *génotype*. Autrement dit, un candidat correspond à une équipe définie par les valeurs de ses propriétés.

Dans le cadre du projet, on a défini $f(x) = (V, N, D, P, C)$ comme le bilan après avoir disputé plusieurs matches avec les mêmes valeurs des paramètres données par x , i.e. le tuple composé du nombre de victoires, matches nuls, défaites, buts marqués et buts encaissés. La relation d'ordre suit les règles du classement d'un tournoi de football.

Cette classe d'algorithmes comportent quatre étapes.

1. Initialisation : Tout d'abord, il faut générer un ensemble de candidats $\{x_1, \dots, x_m\}$. Le but est d'obtenir une population initiale la plus diverse possible de sorte à pouvoir atteindre le plus de maxima relatifs. C'est pourquoi cette première génération est obtenue de manière aléatoire.

2. Évaluation : Ensuite, on évalue la fonction en chacun des candidats, i.e. on calcule $\{f(x_1), \dots, f(x_m)\}$.
3. Sélection : Suivant la notion de sélection naturelle, on ne conserve que les candidats avec les meilleurs résultats. Autrement dit, on dresse un classement des équipes associées aux candidats et n'en garde qu'une partie.
4. Reproduction : Finalement, on attribue les places libres aux candidats nés du brassage génétique des candidats les plus performants. Pour ce faire, on compte sur deux méthodes, à savoir le croisement et la mutation. Étant donné deux parents, on commence par les cloner en deux enfants. Puis on choisit une ligne de coupe divisant le génotype en deux parties. Ensuite, les enfants s'échangent une partie de leur génotype. Le croisement s'arrête à ce stade-là, tandis que la mutation ajoute du bruit aléatoirement dans un gène pour chacun des nouveaux chromosomes. On a décidé de faire une petite modification à cette étape-là lors de l'implémentation pour avoir plus de diversification : une minorité des places libres est attribuée à des candidats générés aléatoirement.

Cet algorithme étant itératif, on recommence à la deuxième phase avec la nouvelle génération. Au bout d'un nombre raisonnable d'itérations, on se retrouve avec une génération composée de chromosomes très proches des maxima relatifs. On finit donc l'algorithme par prendre le candidat optimal de la dernière population.

Cette méthode a permis d'améliorer nos équipes progressivement et les valeurs de tous les paramètres mis en place ont été définies au moyen de nombreuses exécutions de cette méthode.

6 Apprentissage automatique

Dans la dernière partie du semestre, on s'est concentré sur l'apprentissage automatique. On nous a présenté les deux grandes familles d'algorithmes d'apprentissage : supervisés et non supervisés, ainsi que l'apprentissage par renforcement. Deux méthodes ont été implémentées lors du projet.

Arbres de décision (de classification)

Il s'agit d'une technique d'apprentissage supervisé. Pour l'appliquer, on a besoin d'un *espace de représentation* \mathcal{X} , d'un ensemble d'*étiquettes* ou *classes* Y et d'une liste de m exemples d'apprentissage (x^i, y^i) , où $x^i \in \mathcal{X}$, $y^i \in Y$, $i = 1, \dots, m$, appelée *ensemble d'apprentissage*. Le but de cette technique est de trouver une fonction $f : \mathcal{X} \rightarrow Y$ telle que l'on puisse prédire à quelle classe appartient une future variable $x \in \mathcal{X}$.

Concrètement, on veut associer une certaine action à chaque état du jeu. Comme le nombre d'états possibles est trop grand, on utilisera une représentation par n caractéristiques généralisées qui précisent la configuration du terrain. On a ainsi que $\mathcal{X} = \mathbb{R}^n$, avec n la dimension de l'espace de représentation. De plus, l'ensemble d'étiquettes devient l'ensemble d'actions possibles A . On veut donc trouver une fonction $f : \mathbb{R}^n \rightarrow A$ qui fournit la meilleure action possible pour tout état du jeu.

Le problème auquel on est confrontés se réduit alors à partitionner l'espace de représentation en $p = |A|$ parties $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_p$ de sorte que $f([\mathcal{X}_i]) = \{a_i\} \subset A$, $i = 1, \dots, p$. On utilise les arbres de décisions pour représenter ce partitionnement. Ceci implique que chaque nœud interne correspond à un test sur une des n dimensions de \mathcal{X} et que chaque feuille correspond à l'action à entreprendre pour l'état donné par le chemin suivi depuis la racine.

On nous a fourni une stratégie spéciale `KeyboardStrategy` comportant les outils de base nécessaires à la mise en place de cette méthode, allant de la génération de l'ensemble d'apprentissage au partitionnement de l'espace de représentation, en passant par l'application de la

fonction lors des matches. Elle permet d'associer des touches clavier aux différentes stratégies de sorte que l'on puisse préciser la stratégie préconisée pour les différentes phases du jeu. Ceci conforme ce que l'on appelle l'ensemble d'apprentissage. Cet ensemble est traité à l'aide des n caractéristiques généralisées définies auparavant et, à l'issue de ceci, on compte un partitionnement de \mathbb{R}^n . Cette stratégie utilise alors la fonction f lors des matches suivants, ce qui donne un comportement plus intelligent à nos joueurs.

Cette méthode d'optimisation n'a pas été utilisée pour améliorer la prise de décisions sur les actions parce que l'on a trouvé difficile l'interaction à travers le clavier dès que le nombre d'actions possibles dépasse 5.

Q-learning

Cet algorithme d'apprentissage par renforcement repose sur un ensemble d'états S , un ensemble d'actions A , et une fonction $Q : S \times A \rightarrow \mathbb{R}$, appelée *politique*, pondérant une paire formée par un *état* donné et une *action* à entreprendre pour cet état-ci. Cette pondération est nommée *récompense*. La notion d'état correspond à celle donnée dans la section des arbres binaires. Il utilise aussi la notion d'*épisode* que l'on associe à un match dans le cadre du projet. Son principe est le suivant : on apprend une politique comportementale qui optimise l'espérance des récompenses.

La première phase consiste à initialiser Q à une valeur fixe arbitraire, ici 0. La deuxième phase concerne un épisode, ou encore un match, et est répétée autant de fois qu'il y a d'épisodes. Pour un état s_t donné, on choisit une action a_t selon la politique Q et observe la récompense r_t associée à cette paire. Cette action génère un nouvel état s_{t+1} . Il est à noter qu'il n'y a pas de choix concernant l'état initial s_0 , puisqu'il est toujours le même pour un match de football : ceci correspond au coup d'envoi. Ensuite, on met à jour la politique selon la formule ci-dessous.

$$Q(s_t, a_t) \leftarrow (1 - \alpha) \cdot \underbrace{Q(s_t, a_t)}_{\text{valeur actuelle}} + \alpha \cdot \underbrace{\left(r_t + \gamma \cdot \max_a (Q(s_{t+1}, a)) \right)}_{\text{valeur apprise}}$$

Ici, α représente la vitesse d'apprentissage et γ le facteur d'actualisation. D'un côté, la vitesse d'apprentissage détermine l'équilibre entre exploration et exploitation. En effet, une valeur de 0 implique l'utilisation exclusive des choix actuels, alors qu'une valeur 1 ne ferait que considérer le dernier résultat. D'un autre côté, le facteur d'actualisation détermine l'importance des récompenses futures. Une valeur de 0 ne prend en considération que les récompenses courantes, tandis qu'une valeur de 1 met en valeur les récompenses plus lointaines. On avait précisé qu'un épisode correspondait à un match ; par conséquent, la réussite d'une succession d'actions dépend du résultat final du match. De ce fait, on a intérêt à retarder au maximum les récompenses. Ceci se traduit par le choix d'un facteur γ proche de 1.

Juste après la mise à jour de la politique Q , on détermine si le nouvel état s_{t+1} est terminal, i.e. qu'il correspond au coup de sifflet final. Dans le cas affirmatif, on passe à l'épisode suivant ou on s'arrête selon s'il y en a encore à faire ; on réitère cette phase à partir du choix d'une action pour s_{t+1} dans le cas contraire.

À la fin de l'algorithme, on se retrouve avec une politique qui nous rend l'action la plus adéquate pour tout état du jeu. Étant donné que les caractéristiques utilisées pour définir un état sont continues, on a défini des intervalles d'équivalence. Cela n'empêche que le nombre total d'états est énorme et fait donc exploser la mémoire. On pourrait mettre en place un réseau de neurones pour pallier cette difficulté. Ceci n'a pas été le cas pour ce projet et donc, bien qu'implémenté, on s'est pas servi de cet algorithme pour améliorer nos stratégies.

Conclusion

...