Rapport INF8175 : Skypiea

Équipe Challonge : **Orque des Tranchées**
Stephen Cohen - 2412336; Alicia Shao - 2409849

Table des matières

1 Introduction

Dans ce rapport, nous allons présenter le cheminement aboutissant à notre agent final pour le projet du cours Intelligence artificielle : méthodes et algorithmes, centré cette année sur le jeu de société Divercité. Le code est disponible sur notre dépôt Github avec notamment le dossier 'Final'.

2 Réflexion

Dans un premier temps, un algorithme *minimax* a été mis en place, un *alpha-bêta pruning* y a été ensuite ajouté. Enfin, la possibilité d'un Monte-Carlo Tree Search (MCTS) a été également explorée.

Les codes Python du *minimax* et de l'*alpha-bêta pruning* sont disponibles en annexe.

— ***Minimax*** **:** Cette implémentation est disponible sous les noms d'agents : `Water_seven` et `little_garden` sur Abyss. Dans les deux cas, *la profondeur de recherche est fixée à 3* pour **respecter la contrainte de temps imposée**. Avec ce paramètre, il reste environ 100 s à la fin de la partie. La différence entre les deux agents réside dans l'heuristique choisie :

  — L'heuristique de `little_garden` est la suivante : `player_score - opponent_score`. On cherche à maximiser l'écart de points en notre faveur. Cependant, cette heuristique n'est pas assez précise, elle peut favoriser des divercités pour l'adversaire.

  — La seconde, celle de `Water_seven`, tient compte des observations faites lors de nos sessions de jeu. Placer des cités en fin de partie est défavorable, alors que les ressources permettent de finir des divercités ou de bloquer celles de l'adversaire. Ainsi, on pénalise le choix de jouer une cité au fil de la partie et on favorise le choix de ressources. Nous aboutissons à l'heuristique suivante :

  `player_score - opponent_score + (1 - 2 * state.step/40) * nb_cite + (1 + 4 * state.step/40) * nb_ressource`.

  En plus de l'écart entre les points, nous rajoutons donc un paramètre qui tient compte de l'ensemble des coups disponibles (ressources ou cités).

— ***Alpha-bêta pruning*** **:** Cette implémentation est disponible dans les agents : `Enies_Lobby` et `skypiea_vX`. L'avantage de l'*alpha-bêta pruning* réside dans le gain de temps obtenu grâce à l'élagage. Ce temps supplémentaire est réinvesti au profit d'une recherche plus profonde. **Attention :** On remarque qu'une grande profondeur couplée à une mauvaise heuristique propage de mauvais résultats, ce qui entraine une dégradation significative des résultats de l'agent (notamment avec `skypiea` avec l'heuristique classique).

La profondeur est dynamique en fonction de l'avancée de la partie. En effet, une *recherche profonde en début de partie représente un investissement de temps peu rentable pour un coup peu décisif.* Ainsi, le **premier coup est joué de manière aléatoire parmi les coups jouant des cités**, ce qui permet de gagner 100 à 200 s de jeu :

Listing 1 – Code pour le premier coup

```
# Pr -choisir une action si on joue en premier
if current_state.get_step() < 2:
    possible_actions = current_state.get_possible_light_actions()
    city_actions = [action for action in possible_actions
                    if action.data["piece"] in ['RC', 'GC', 'BC', 'YC']]
    return random.choice(city_actions)
```

Puis, la profondeur est gérée comme suit afin de conserver environ 60 s en fin de partie :

Listing 2 – Gestion de la profondeur

```
if nb_pieces_1 + nb_pieces_2 >= 35:
    depth = 4
elif nb_pieces_1 + nb_pieces_2 >= 12:
    depth = 5
else:
    depth = 6

_, best_action = alpha_beta_minimax(current_state, depth, float('-inf'),
                                    float('inf'), True)
```

Cette étape d'ajustement de la profondeur abouti à `skypiea_v2`, l'agent le plus performant de la génération *alpha-bêta pruning*. La profondeur est gérée finalement de la manière suivante :

Listing 3 – Gestion finale de la profondeur

```
if nb_pieces_1 + nb_pieces_2 >= 12:
    depth = 4
else:
    depth = 6
```

et l'heuristique est celle-ci : `player_score - opponent_score + (1 - 2 * state.step/40) * nb_cite + (1 + 4 * state.step/40) * nb_ressource`

— **MCTS :** Les résultats donnés par notre agent basé uniquement sur un MCTS ne sont pas assez satisfaisants pour mériter une présentation complète du code, il ne bat aucun des agents présentés plus haut. Cependant, nous avons développé un agent combinant *alpha-bêta pruning* et MCTS de la manière suivante :

— *Sur les 30% du début de jeu, jouer avec un MCTS moins optimal qu'un alpha-bêta pruning mais très rapide.*

— Puis, un *alpha-bêta pruning de profondeur progressive de type 5 / 7.*

L'idée première était d'obtenir rapidement des premiers coups, cohérents, voirs bons pour libérer du temps afin de réaliser une recherche plus profonde dans un deuxième temps. Dans le cas où le MCTS ne propose que des coups incohérents et mauvais, l'effet pourrait être compensé par la deuxième phase, qui engendrait déjà à ce stade des résultats probants.

L'ajustement final des paramètres n'a pas pu être réalisé avant le début du tournoi. Nos agents étaient soit **trop lents** (mais assez compétitifs, sans battre nos agents précédents) ou **trop peu efficaces** (mais rapides). Le code de notre agent hybride est disponible en annexe sous le nom d'`Impel Down`.

Nous avons donc choisi de perfectionner pour ce projet `skypiea_v2`, un agent *minimax alpha-bêta pruning* pur.

3 Notre agent : Skypiea_v5

3.1 Heuristique

La base de l'heuristique Skypiea_v5 est celle décrite dans la partie *Alpha-bêta Pruning*. Après avoir réalisé des tests pour affiner les coefficients de pénalisation, nous obtenons l'heuristique suivante : `player_score - opponent_score + (1 - 24 * state.step / 40) * nb_cite + (1 + 27 * state.step / 40) * nb_ressource`

3.2 Implémentation

Le code de notre agent est disponible sur Github .

4 Performances

4.1 Performances de jeu

Ces statistiques ont été obtenues à partir d'Abyss. On y présente tous les agents qui ont réalisé au moins un match.

TABLE 1: Performances des agents

Agents	**Notes**	**Elo**	**Total matches**	**W/L**	**Divercités/ Game**	**Marqués/ Concédés**	**Points marqués/- Game**	**Points concédés/- Game**
Skypiea v5	Alpha Beta Pruning	1282	13	3.33	2.31	1.29	19.92	15.46
Impel Down 25000	MCTS puis AB	937	9	2.00	2.44	1.36	22.22	16.33
Impel Down 15000	MCTS puis AB	1006	1	-	5.00	1.81	29.00	16.00
Water Seven	Minimax	971	1721	1.10	1.15	1.05	17.17	16.41
Skypiea v2.2	Alpha Beta Pruning	1165	25	2.13	1.04	1.16	17.88	15.44
Skypiea v2	Alpha Beta Pruning	1275	489	2.52	1.30	1.22	18.58	15.28
Little Garden	Minimax	713	707	0.91	0.69	1.03	16.78	16.22
Skypia	AB	960	5	0.67	0.00	1.03	18.20	17.60

Remarques : Le nombre de matches joués par un agent est assez inhomogène, les conclusions suivantes en prennent compte et sont à nuancer.

Le tableau met en évidence la compétitivité des différents agents dans l'environnement Abyss. Voici un commentaire détaillé des résultats :

- **Agent Skypiea_v5 (*Alpha-bêta pruning*)**
  - Performance globale : Avec un Elo de 1282, l'agent Skypiea_v5 domine la majorité des parties, affichant un ratio victoires/défaites de 3,33 (pour seulement 13 matches cependant).
  - Divercités : Il obtient en moyenne 2,3 divercités par partie, témoignant d'une capacité à exploiter les ressources efficacement, bien que cela soit perfectible.
  - Ratio points marqués/concédés : Le ratio de 1,29 montre une capacité à marquer plus de points que l'adversaire, reflet d'une bonne gestion des ressources et de l'heuristique employée.
- **Agent Impel Down (MCTS puis *Alpha-bêta pruning*)**
  - Deux versions : Les versions avec 15 000 et 25 000 simulations aléatoires affichent des performances variables. La version 25 000 obtient un Elo de 937, mais l'agent peine à être décisif (ratio W/L de 2) et est davantage soumis au risque de *time out*. (Il est actuellement inactif pour cette même raison).

— Divercités : Le score moyen de divercités par partie (2,44) est légèrement meilleur que Skypiea_v5, suggérant que l'exploration initiale en MCTS réussit à choisir de meilleurs coups en début de partie que les agents de type Skypiea.

— **Agents intermédiaires (Skypiea_v2)**

— Ces deux agents de type *Alpha-bêta pruning* purs montrent des résultats stables sur un grand nombre de matches, obtenant respectivement 1275 et 1108 d'Elo.

— Ces résultats confirment qu'un agent basé *alpha-bêta pruning* représente une base solide pour obtenir des résultats probants rapidement et facilement. Cela conforte le choix d'avoir poursuivi l'affinage des agents de type Skypiea.

4.2 Nombre d'états visités et profondeur dynamique

Ci-dessous, un graphique du nombres d'états évalués par Skypiea_v5 en fonction de l'avancée de la partie.

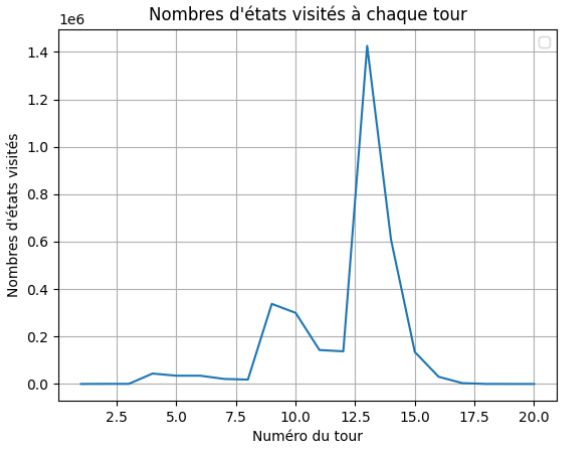

FIGURE 1 – États visités en fonction de l'avancée de la partie

Sur le graphique des états visités, nous observons bien les différentes profondeurs de l'arbre du *minimax*. De plus, notre stratégie qui vise à avoir un milieu de partie très profond est rendu possible grâce à un début de partie très peu profond. Voici le nombre d'états visités au fil de la partie : `[0, 452, 402, 43888, 34809, 34825, 21231, 18401, 337614, 299769, 143075, 137567, 1425242, 608967, 134754, 30259, 3387, 169, 25, 3]`

— Le premier coup est aléatoire donc 0

— La phase de recherche en profondeur 2 : `[452, 402]`

— La phase de recherche en profondeur 4 : `[43888, 34809, 34825, 21231, 18401]`

— La phase de recherche en profondeur 5 : `[337614, 299769, 143075, 137567]`

— La phase de recherche en profondeur 7 : `[1425242, 608967, 134754, 30259, 3387, 169, 25, 3]`

5 Limites

Outre des *time out* dès le coup 1 que nous ne comprenons pas, voici les pistes d'amélioration que nous avons identifiées :

— **L'heuristique :** Notre heuristique ne tient pas compte de **la présence de divercités** ou d'**opportunités de divercités** pour l'agent ou l'agent adverse. En effet, l'agent ne sait à aucun moment combien de ses pièces ou des pièces adverses sont encore capables de réaliser une divercité, ni si elles en sont éloignées. C'est pourquoi, nous pouvons juger une situation plus favorable alors qu'elle nous empêche de réaliser une divercité ou rapproche l'adversaire d'en faire une. C'est d'ailleurs ce que l'on peut observer dans l'étude des performances, nos agents réalisent en moyenne moins de 1,5 divercité par partie.

— **La connaissance des pièces adverses :** Dans Divercité, les joueurs ont connaissance des pièces adverses. Par exemple, savoir que l'adversaire n'a plus de ressources rouges permet de savoir qu'il ne pourra plus faire de divercité. S'il n'a plus de cité, les places de cité disponibles nous reviennent. Il n'y aurait donc plus besoin de les pénaliser, il faudrait même construire des divercités autour des emplacements de cités vides.

— **Le temps de calcul :** L'agent présenté utilise presque l'intégralité du temps disponible (il reste 35 secondes à la fin), sans pourtant présenter des résultats excellents. Cela suggère qu'il est possible d'élaguer davantage d'états, toujours dans l'optique d'augmenter la profondeur de recherche.

6 Avantages

— **L'implémentation :** L'algorithme *minimax* avec *alpha-bêta pruning* n'est pas difficile à implémenter, ce qui facilite le développement et conduit à une phase de tests beaucoup plus longue, permettant ainsi d'ajuster l'heuristique plus finement que par rapport à un algorithme plus complexe à mettre en place (MCTS, ici).

— **Les performances :** Le ratio entre la performance et la difficulté d'implémentation est excellent pour Skypiea_v5 et les autres agents de sa famille. A titre de comparaison, les agents hybrides avec un MCTS sont prometteurs mais nécessitent l'ajustement complexe de plusieurs autres paramètres (nombre de simulations totales, choix entre simulations aléatoires ou guidées par une heuristique, choix du moment de passage à l'*alpha-bêta pruning*) sans garantie d'amélioration substantive.

7 Pistes d'améliorations

— **L'heuristique :** L'heuristique actuelle peut-être enrichie avec des facteurs comme le contrôle des zones stratégiques, la proximité de réalisation des divercités, et les blocages potentiels pour l'adversaire. Cela permet une prise de décisions plus stratégiques en fonction du contexte du jeu, surtout lors de phases plus critiques.

— **La réduction du temps de calcul :**
  — On pourrait **stocker les évaluations des états déjà visités** pour éviter de recalculer des scénarios redondants. L'utilisation d'une table de transposition permet d'accélérer les recherches, notamment dans les configurations de jeu symétriques. Pour se faire, on peut utiliser des hashcodes d'états pour les identifier efficacement ; tout en tenant compte des symétries du plateau pour réduire de manière significative les répétitions de calcul.
  — Une stratégie de ***Late Move Reduction*** peut également être mise en place. Elle consiste à réduire la profondeur de recherche sur les coups explorés tardivement dans une branche, en supposant que les premiers coups sont généralement les meilleurs.
  — Une autre piste d'amélioration est de **gérer le temps de manière adaptive en fonction du temps restant**. Il s'agit de limiter la profondeur ou d'interrompre des branches si le temps alloué pour un coup est proche de l'expiration.

— **Tenir compte de l'historique de nos parties :** Via une exploration bayésienne des actions, on pourrait ajuster la probabilité de sélection des actions en fonction de leur performance dans des simulations précédentes.

8 Conclusion

Ainsi, notre équipe a développé un agent pour Divercité de type *alpha-bêta pruning* plutôt performant, fruit d'un processus itératif formé de trois générations d'agents.

L'heuristique employée met l'accent non seulement sur le score de la partie, mais prend également en compte des stratégies développées lors des sessions de jeu réelles en favorisant les coups impliquant des cités et en pénalisant les coups impliquant des ressources en *early game*.

L'heuristique et le type d'algorithme choisis ayant obtenu des résultats probants très tôt, le principal enjeu et défi de ce projet aura donc été d'augmenter la profondeur de recherche dans l'algorithme afin d'améliorer ces résultats, tout en tenant compte de la contrainte de temps.

Plusieurs approches ont été employées pour y répondre, dont une profondeur dynamique en fonction de l'avancée de la partie, pour favoriser la recherche d'états favorables en milieu de jeu, identifié comme le moment décisif d'une partie.

L'idée d'un agent hybride MCTS-*alpha-bêta pruning*, qui apporte une réponse différente à ce problème, n'a pas été retenue. Elle demeure cependant une approche prometteuse mais plus complexe, nécessitant davantage d'investissement pour l'ajuster, l'affiner et l'améliorer.

Finalement, à travers ce projet, nous avons mis en pratique plusieurs algorithmes de recherche adversariale dans un cadre applicatif stimulant et ludique. Cette expérience mettant à profit nos capacités analytiques a nécessité la mise en place d'une réelle démarche de recherche collaborative, enrichissant ainsi notre bagage scientifique et nos compétences interpersonnelles.

A Annexes

A.1 Code du minimax

Listing 4 – Code de l'implémentation Minimax

```
def minimax(state: GameState, depth: int, maximizing_player: bool) -> float:
    if depth == 0 or state.is_done():
        return self.evaluate_state(state), None

    if maximizing_player:
        max_eval = float('-inf')
        for action in state.get_possible_light_actions():
            next_state = state.apply_action(action)
            eval, _ = minimax(next_state, depth - 1, False)
            if eval > max_eval:
                max_eval = eval
                best_action = action
        return max_eval, best_action
    else:
        min_eval = float('inf')
        for action in state.get_possible_light_actions():
            next_state = state.apply_action(action)
            eval, _  = minimax(next_state, depth - 1, True)
            if eval < min_eval:
                min_eval = eval
                best_action = action
        return min_eval, best_action

if current_state.get_step() < 2:
    possible_actions = current_state.get_possible_light_actions()
    return random.choice(list(possible_actions))
else:
    # Ajustement de la profondeur en fonction du nombre de pi ces restantes
    players = current_state.players
    players_id = [p.get_id() for p in players]
    dic_player_pieces = current_state.players_pieces_left
    dic_pieces_1 = dic_player_pieces[players_id[0]]
    dic_pieces_2 = dic_player_pieces[players_id[1]]
    pieces = ['RC', 'RR', 'GC', 'GR', 'BC', 'BR', 'YC', 'YR']
    nb_pieces_1 = sum(dic_pieces_1[p] for p in pieces)
    nb_pieces_2 = sum(dic_pieces_2[p] for p in pieces)

    _ , best_action = minimax(current_state, 3, True)

    return best_action
```

A.2 Code de l'alpha-beta pruning

Listing 5 – Code de l'implémentation Alpha-Beta Pruning

```
def compute_action(self, current_state: GameState, **kwargs) -> Action:
    def alpha_beta_minimax(state: GameState, depth: int,
                           alpha: float, beta: float,
                           maximizing_player: bool) -> float:
        if depth == 0 or state.is_done():
            return self.evaluate_state(state), None

        if maximizing_player:
            max_eval = float('-inf')
            best_action = None
            actions = state.get_possible_light_actions()

            if len(actions) > 5:
                actions = sorted(actions,
```

```
                                key=lambda a: self.evaluate_state(state.apply_action(a
                                    )),
                                reverse=True)

            for action in actions:
                next_state = state.apply_action(action)
                eval, _ = alpha_beta_minimax(next_state, depth - 1,
                                             alpha, beta, False)
                if eval > max_eval:
                    max_eval = eval
                    best_action = action
                alpha = max(alpha, eval)
                if beta <= alpha:
                    break
            return max_eval, best_action
        else:
            min_eval = float('inf')
            best_action = None
            actions = state.get_possible_light_actions()

            if len(actions) > 5:
                actions = sorted(actions,
                                key=lambda a: self.evaluate_state(state.apply_action(a
                                    )))

            for action in actions:
                next_state = state.apply_action(action)
                eval, _ = alpha_beta_minimax(next_state, depth - 1,
                                             alpha, beta, True)
                if eval < min_eval:
                    min_eval = eval
                    best_action = action
                beta = min(beta, eval)
                if beta <= alpha:
                    break
            return min_eval, best_action

    if current_state.get_step() < 2:
        possible_actions = current_state.get_possible_light_actions()
        return random.choice(list(possible_actions))
    else:
        players = current_state.players
        players_id = [p.get_id() for p in players]
        dic_player_pieces = current_state.players_pieces_left
        dic_pieces_1 = dic_player_pieces[players_id[0]]
        dic_pieces_2 = dic_player_pieces[players_id[1]]
        pieces = ['RC', 'RR', 'GC', 'GR', 'BC', 'BR', 'YC', 'YR']
        nb_pieces_1 = sum(dic_pieces_1[p] for p in pieces)
        nb_pieces_2 = sum(dic_pieces_2[p] for p in pieces)

        if nb_pieces_1 + nb_pieces_2 >= 35:
            depth = 2
        elif nb_pieces_1 + nb_pieces_2 >= 12:
            depth = 5
        else:
            depth = 7

        _, best_action = alpha_beta_minimax(current_state, depth,
                                            float('-inf'), float('inf'), True)
        return best_action
```