

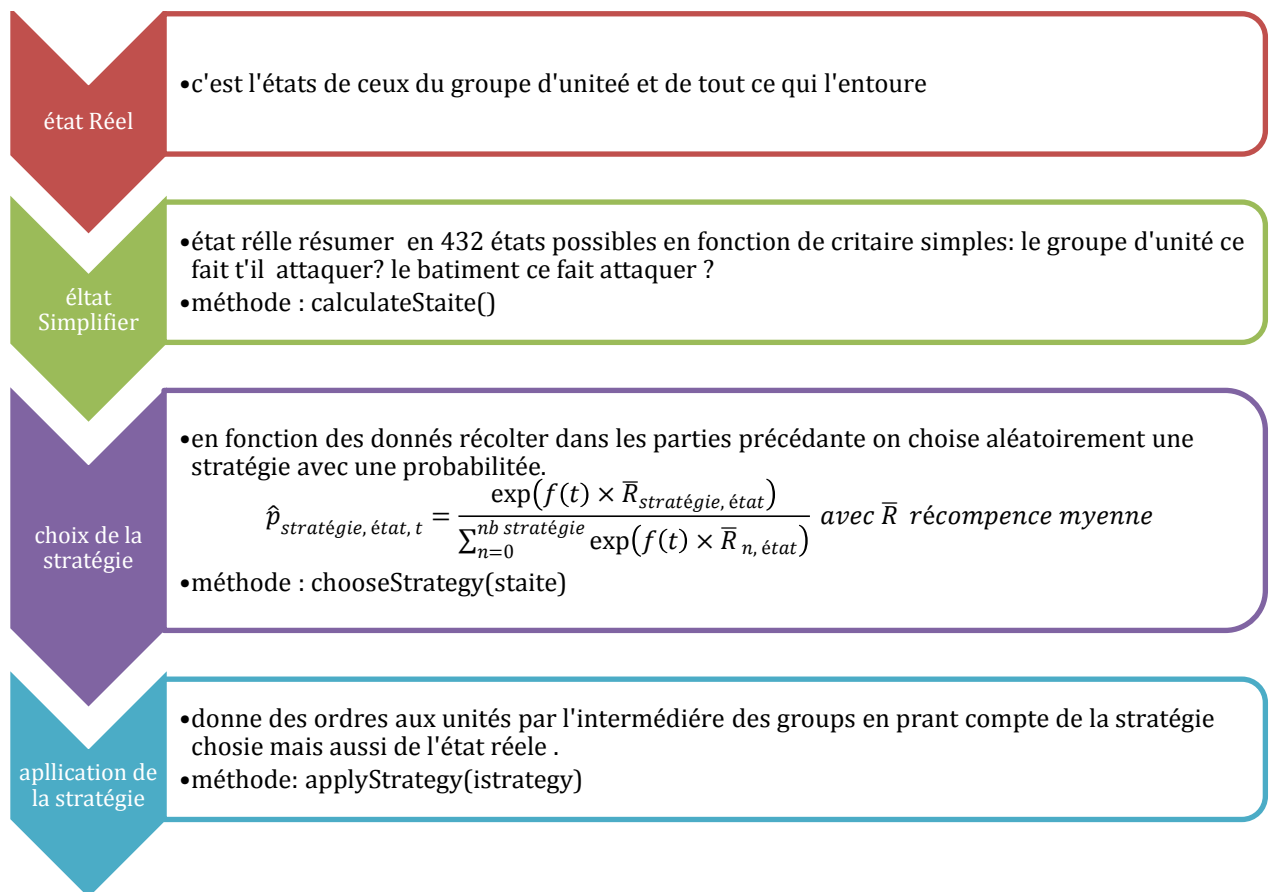
Principe de fonctionnement de l'intelligence artificielle.

L'objectif intelligence artificielle est de lier le plus efficacement possible un état à une action. Pour ce faire, nous avons principalement mis en place de processus Processus d'application et processus d'apprentissage.

Le processus d'application.

Le processus d'application est le processus central du fonctionnement de l'intelligence artificielle. C'est donc lui qui lit à chaque État une action. On applique ce processus à des groupes d'unités à intervalle de temps court et régulier l'objectif étant de choisir des actions qui maximisent la récompense. On définit pour cela un petit nombre de stratégies (5 pour les Su 6 pour les So) qui sont des ensembles d'actions. On s'appuie pour cela sur des récompenses moyennes obtenues lors des parties précédentes.

Figure 1 : processus d'application



Processus d'apprentissage.

Ce processus a pour objectif de mettre à jour les tableaux de récompenses moyennes à la fin de chaque partie afin de faire progresser l'intelligence artificielle.

À chaque fois que l'intelligence artificielle est appelée chaque État, stratégie, et récompenses obtenues, sont sauvegardés. À la fin d'une partie, la mémoire reparcouru afin de mettre à jour le tableau de récompenses moyennes.

Etats s_t

Stratégie a_t

Et $Q_t(s, a)$ étant la récompense moyenne obtenue lorsque l'intelligence artificielle lit un état s à une stratégie a .

Avec $Rp = \sum_{n=t+2}^T R_n \times \gamma^{n-t-2}$ et $\alpha_t(s_t, a_t) = \frac{1}{nb \text{ Mis à jours de } Q(s_t, a_t)}$

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha_t(s_t, a_t) \times (R_{t+1} + \gamma/2 (\max_a Q_{t+1}(s_{t+1}, a) + Rp) - Q_t(s_t, a_t))$$

Ce mode de calcul est très similaire au Q Learning, nous y avons pourtant ajouté le Rp qui rend la récompense actuelle dépendante des récompenses futures.