

# Evolutionary Neural Network Tools for Life Emergence and More

F. Furfaro

2021

## Abstract

L'idée de vie artificielle remonte au année 1940 par principe d'autoréplication, meme époque ou les concepts sur l'intelligence artificielle sont créés. Nous proposons un modele à définition incomplete, mais suffisante pour etudier la fonctionnalisation des structure cerebrale lors de l'evolution. On combine l'apprentissage par renforcement et la modification aléatoire des connections intercouche. Pour tester notre modele, le jeu du chat et la souris est un exemple de choix pour etudier l'evolution des comportement basique de proie/predateurs, tout en etant assez complet pour une approche évolutive. Ces outils n'ont pas la prétentions de reproduire la vie sur ordinateurs, mais peuvent donner des idée, minime soit elle, pour l'optimisation des structure de reseau de neurones et d'avancer vers la creation de vie articiel. En effet, lors du processus d'apprentissage (evolution + descente de gradient), on on observe par l'apparition de structure et stratégie convergente. Ces structures sont conservé pendant un grand de cycle alors meme que des reseau "challenger" sont toujours ajouter à chaque cycle de reproduction. En dehors, on observe que ce type d'algorithme peut s'averer efficace avec moins de neurone pour la classification MNIST.

## Introduction

La vie est apparu il y a 3,8 milliards d'année et n'a pas cessé d'évoluer et de coloniser l'environnement [5]. Elle a commencer de maniere unicellulaire sans noyau [3, 7], à un large panel de forme avec l'appartition du noyau [4, 6] et de la multicellularité [2]. Pour autant, il est assez difficile de definir ce qui est vraiment vivant et cette question est toujours débattu. L'une des definition les plus simple est celle donnée par la NASA lors des programme de recherche de vie extra-terrestre : Un systeme chimique auto-entretenu capable d'évolution darwinienne. L'évolution darwinienne est un principe fondamentale dans l'étude des systeme complexe disposant d'heredité, de variation et de selection. On ne le retrouve pas seulement en biologie, mais on peut l'observer dans l'evolution des courants musicaux, artistique, philosophique, technologique, etc. Une autres définition de la Vie existe, et elle est plus lié à l'une des transformation thermodynamique fondamentale : L'entropie. La vie serait : Une structure dissipative capable d'auto-catalyse, d'homéostasie et d'apprentissage [1]. Ce qui diminue localement l'entropie en favorisant l'auto-organisation. Mais ces deux definitions se restreignent à un systeme "chimique", est ce qu'un robot dotée d'une intelligence, voient d'emotion humaine serait vivant dans cette définition ? (tout probleme n'est pas définissable). Aujourd'hui, il existe un grand nombre de simulation de vie sur ordinateurs (liste de 2-3) et certaine sont encore en cours d'execution depuis 1990. Et son d'une grande importance pour comprendre l'emergence de la vie (avec les experiences de cellules minimaliste)

**AUTOMATE** : Dans notre cas, nous allons essayer de reproduire un systeme vivant à l'intersection de ses deux definitions. Un systeme auto-entretenu capable d'apprentissage evolutif. Qu'on restreindra dans un premiers lieux à "un systeme capable d'apprentissage évolutif" pour en developper les outils futurs. Pour cela on a creer un environnement de jeu simple : le jeu du chat et la souris, alternant ainsi deux comportements : proie/prédateur. On se base sur deux outils : les automates cellulaires et les reseaux de neurones évolutifs. Les automates sont à l'origine des objets mathématique permettant de resoudre des probleme de decidabilité mathématique [Von Newman]. On en retrouva ensuite des utilisation experimentale comme le plus connu "jeu de la vie" [Conway], qui à partir de 2 regles simple, faire emerger une forte complexité, voir meme reproduire une machine de Turing [ref à revoir]. Le L'état d'une cellule au temps  $t+1$  est fonction de l'état au temps  $t$  d'un nombre fini de cellules appelé son « voisinage ». Le probleme est donc une grille ou l'on a des case "automates" qui se deplace dans la grille, elle a un adversaire qui va chercher soit à "attraper" l'automate, soit la "fuir". La decision de deplacement de l'automate est controler/réguler par un reseau de neurone evolutif par principe d'apprentissage par renforcement. Les propriété de

l'automate changeront ainsi à chaque génération, aussi bien paramètre de "vue" de l'automate que les actions ou la structure du réseau.

**STATISTIQUE** : Les réseaux neuronaux ont été construits à partir du modèle de neurone biologique, le neurone formel [McCulloch & Pitts], où l'on constate qu'un neurone est une fonction "affine" avec une fonction d'activation. Aujourd'hui il s'est rapproché des méthodes statistiques non inductive (descriptive : on ne sait pas la loi de probabilité) pour résoudre des problèmes de régression, classification et partitionnement. Les principaux outils d'analyse de grosse donnée en statistiques sont, pour ne citer que les plus connus : La PCA, analyse de composante principale, qui permet de projeter l'ensemble des points sur les axes de régression multilinéaire des variables corrélées entre elles, ce qui réduit la dimension [Karl Pearson], Les SVM, séparateurs à vaste marge, généralisant les classificateurs linéaires par des vecteurs supports (théorie de la « régularisation statistique » + surapprentissage) [Vapnik] et enfin Le K-mean qui divise un ensemble de points par minimisation d'un certain nombre de graine barycentrique [Steinhaus, Lloyd]. Tous ces outils encore utilisés aujourd'hui, sont peu à peu remplacés par des réseaux de neurones vers 2000s après le succès de l'apprentissage profond en compétition.

**APPRENTISSAGE** : La question de l'apprentissage a été étudiée au cours des années 1950, où il a pu émerger les règles de Hebb [Hebb 1949] qui modifiaient simplement les valeurs des coefficients synaptiques d'un réseau simple couche, il eut un déclin en fin 1970 par l'impossibilité de résoudre le problème non linéaire XOR ou connexité [Minsky 1969]. Mais ce problème fut résolu par le perceptron multicouche, mais cette fois la modification des poids était plus compliquée. C'est le système de rétropropagation du gradient [1984] qu'on retrouve aussi dans le cerveau [Stuart 1997] et l'algorithme du gradient, par minimisation itérative de l'erreur, qui permet de résoudre ce problème. En parallèle, inspiré par les neurones moteurs de la vue, on développa le réseau de neurone convolutif qui réduit fortement le calcul des poids en entrée du réseau, car n'est pas entièrement connecté [Lecun, mais pas que]. Le souci de ce genre de réseau est qu'il faut un grand nombre de données pour adapter les poids, pour résoudre un problème de décision markovien, il y eut l'apparition du Q-learning convergent par l'apprentissage des récompenses [Watkins 1989 Dayan 1992] qui ont encore été améliorés par des réseaux adversariaux "actor-critic" à l'image des réseaux génératifs, il permet d'accélérer la convergence des poids [Konda 2003].

**FONCTIONNALISATION** : Néanmoins, ces réseaux optimisent l'apprentissage du réseau mais ne cherchent pas à maximiser la structure des couches du réseau. Seuls les réseaux de neurones récurrents ont été le plus traités car il y avait des problèmes de "vanishing gradient", surtout pour le traitement du texte. Une autre catégorie de réseau est apparue dans la fin des années 1990, les réseaux NEAT, où les poids et les connexions sont "appris" par un processus de sélection. Hors le cerveau n'a pas des poids qui ont été sélectionnés par l'évolution, mais ce qui est le cas la structure [Ref]. L'idée est donc de s'inspirer des données en neurosciences récentes, où il est montré le principe de fonctionnalisation cérébrale que le cerveau est pré-cablé pour voir certains dangers ou encore marcher à la naissance [araignées Rakison, D. H. & Derringer, J 2008 et sommeil] et il n'y a pas eu besoin d'avoir un temps d'apprentissage long. À cette image, on a donc développé un algorithme où les couches interconnectées de manière aléatoire de façon à fonctionnaliser le problème "chat-souris" aussi bien dans la structure que dans l'apprentissage par renforcement.

## Méthodes

L'ensemble du projet est disponible sur le lien [github.com/fabienfrfr](https://github.com/fabienfrfr). Le projet a été codé intégralement en Python de façon à utiliser efficacement les bibliothèques logicielles : PyTorch pour les réseaux de neurones, Keras-Tensorflow pour la base de données MNIST, Numpy et Scipy pour les calculs scientifiques, Pandas pour l'analyse et le stockage des données, Networkx pour la représentation et l'analyse des graphes et enfin Matplotlib pour la visualisation et l'animation des données. On distingue 6 fichiers d'expérience, 2 fichiers d'analyse de données (DATA\_ANALYSIS et EXTRA\_FUNCTION) et un fichier de débogage modulaire. Dans les fichiers "expérience" on retrouve au plus bas niveau GRAPH\_EAT encapsulant GRAPH\_GEN, générant la liste des connexions pour chaque couche de neurone, ainsi que pRNN, générant la structure du réseau en fonction de la liste d'adjacence. À plus haut niveau, on retrouve AGENT qui contient les propriétés d'entrée/sortie de l'agent et son algorithme d'apprentissage et de mémorisation, ainsi que TAG-ENV contenant cette fois les règles du jeu "chat-souris". Enfin on a le MAIN qui va lister l'ensemble des agents et environnements associés pour lancer les expériences d'évolution. On retrouve aussi un fichier LOG qui permet de stocker au format *csv* les expériences.

# 1 Réseau de neurone évolutif

Le perceptron est un algorithme d'apprentissage fonctionnant comme une fonction de seuillage et décrivant le neurone formel. Une fonction de seuil est un classifieur linéaire qui a pour entrée un vecteur à valeurs réelles (virgule flottante) et une valeur  $f(z)$  en sortie, qui peut être suivant la fonction d'activation : binaire (Heaviside), réel (Sigmoid) ou entre les deux (ReLU). Le perceptron se représente par le produit scalaire entre le vecteur d'entrée et le vecteur poids des neurones, soit l'application multilinéaire suivante :

$$o = f(z) = \begin{cases} 1 & \text{si } \sum_{i=1}^n w_i x_i > \theta \\ 0 & \text{sinon} \end{cases}$$

Ce qui s'écrit dans la plupart des bibliothèques logicielles comme la séquence d'un neurone à “**n**” entrée, suivie d'une fonction d'activation.

Code PyTorch :

```
Layers = nn.Sequential(nn.Linear(N, 1), nn.ReLU())  
x = Layers(x)
```

Le perceptron est un classifieur linéaire, il converge uniquement si l'ensemble des données d'entrée sont linéairement séparables (droite). Mettre à jour les poids “**w**” correspond à l'apprentissage du perceptron, la *regle delta* [Russell Ingrid 2002] est la plus simple et consiste à comparer linéairement la valeur obtenue par la sortie du réseau “**z**” et la valeur attendue par le réseau de neurones “**d**”. Il existe d'autres méthodes plus sophistiquées ou limitant les problèmes d'annulation tout aussi simple (Least Mean Square), mais ici présenté, la règle delta se présente sous la forme :

$$w_i(t+1) = w_i(t) + (d - z)x_i$$

Dans la plupart des bibliothèques logicielles, on précise le type de calcul de perte (loss ou criterion) avant la boucle d'apprentissage, ainsi que l'algorithme de calcul de la descente de gradient (optimizer). Les algorithmes les plus utilisés sont SGD “Stochastic Gradient Descent” plutôt adapté à la classification et “Adam” plutôt adapté au problème de régression [ref car pas sûr].

Code PyTorch :

```
LOSS = torch.nn.MSELoss(reduction='sum')  
optimizer = optim.SGD(Layers.parameters())  
#### Loop  
LOSS = LOSS(x_pred, x)  
optimizer.zero_grad()  
LOSS.backward()  
optimizer.step()
```

Un réseau de neurone correspond à une interconnexion entre couche de neurone contenant “**n**” perceptrons. Une unique couche de “**n**” perceptron correspond au réseau le plus simple et permet de résoudre des problèmes linéairement séparables de type hyperplan. Comme toutes les entrées “**x**” sont interconnectées au “**n**” neurone, les vecteurs poids “**w**” forment une matrice “**m\*n**” avec “**m**” le nombre d'entrée au réseau. La sortie s'exprime dans ce cas :

$$\vec{y} = \begin{bmatrix} w_{0,0} & & \\ & \ddots & \\ & & w_{m,n} \end{bmatrix} \cdot \begin{bmatrix} x_0 & \dots & x_m \end{bmatrix} + \begin{bmatrix} b_0 \\ \\ b_n \end{bmatrix}$$

$\vec{s} = f(\vec{y})$

Pour que l'algorithme mette à jour efficacement les poids du réseau, on utilise un certain lot de données d'apprentissage, qu'on appelle “batch”, soit il contient l'ensemble des données de l'expérience, soit il est réparti en “mini-batch” pour éviter la sur-utilisation de la mémoire et augmenter la vitesse de convergence de la vallée de stabilité, la taille du lot ne doit pas être trop petite non plus car augmenter le bruit [ref mini-batch]. Dans ce cas, le vecteur d'entrée devient une matrice “**m\*b**”, et la matrice des poids devient un tenseur d'ordre 3 “**n\*m\*b**”. La plupart des bibliothèques logicielles s'adapte automatiquement à la taille du batch, et n'a pas besoin d'être spécifié.

Code PyTorch :

```
Layers = nn.Sequential(nn.Linear(N, M), nn.ReLU())
x = Layers(x)
```

Pour autant, les reseaux linéaire ne sont pas les seules à être employé et ne sont pas forcément les plus adaptés pour les problèmes d’images [ref]. En effet, une image est une donnée 2D dont les données spatiales sont localement liées, mais les informations ne sont pas interconnectées, c’est pour cela qu’il existe les réseaux de convolution. Ces réseaux ont l’avantage d’être interconnectés mais par fenêtre “spatiale” ce qui réduit fortement le nombre de calculs pour l’actualisation des poids. Elle est composée de “N” perceptrons, mais ne représente plus comme un produit tensoriel, mais comme un produit de convolution où les entrées sont moyennées pondérément par le glissement “C” d’une fenêtre de taille “k” sur les poids “w”. La sortie s’exprime de la sorte :

$$out(N_i, C_{out_j}) = bias(C_{out_j}) + \sum_{k=0}^{C_{in}-1} weight(C_{out_j}, k) \star input(N_i, k)$$

Dans le cas le plus simple, on peut considérer qu’il y a qu’une seule entrée par neurone, pour cela, le nombre d’entrée est égal au nombre de sortie du canal de la couche. Ensuite, la taille de la fenêtre est de “1” et le groupement de connexion doit être égal au nombre de neurones, qui est un nombre divisible de l’entrée et de la sortie par lui-même. Ce type de paramétrage est équivalent à N neurones indépendants à 1 seule connexion chacune. Contrairement au réseau linéaire, la taille du batch doit être spécifiée lorsqu’on donne les données d’entrée à la couche de convolution.

Code PyTorch :

```
Layers = nn.Sequential(nn.Conv1d(I, I, 1, groups=I), nn.ReLU())
x = Layers(x.view(BATCH_SIZE, I, 1)).view((BATCH_SIZE, I))
```

## 1.1 Fonctionnalisation

S’inspirer du cerveau “modèle” pour construire des algorithmes neuronaux a toujours été présent dans le développement de nouvelles méthodes en *Intelligence Artificielle*. Dans ce projet, nous nous sommes inspirés de la spécialisation fonctionnelle du cerveau. En effet, le moyen le plus efficace de réaliser des fonctions différentes c’est de confier la réalisation de chaque fonction à un outil particulier et spécialisé [Tooby, J. & Cosmides, 1992 ; Tooby, J. & Cosmides, 2015]. C’est pour cette raison que la sélection naturelle (survie, reproduction) a organisé notre corps en différents organes qui sont chacun spécialisés, mais aussi dans le domaine de la cognition [Confer, J. C. et al. 2010]. La spécialisation fonctionnelle du cerveau est observée empiriquement par plusieurs troubles neuropsychologiques, comme par exemple avec le syndrome de Capgras, où une défaillance d’une région entraîne la perte de reconnaissance des visages [Thomas Antérion, 2008].

Dans notre cas, nous avons simplifié le plus possible un cerveau, où l’on a en entrée des neurones de “vision”, suivis de couches séquentielles “fonctionnelles” et en sortie, des neurones “moteurs”. On distingue des neurones stables, les neurones d’entrée et de sortie qui ne changeront pas pendant le processus évolutif. Les neurones de “visions” sont à l’image des cônes de la rétine c’est-à-dire des neurones connectés à des batonnets de vue unique [Dale Purves, G-J Augustine, 2005] qu’on représente dans notre cas comme une couche de convolution “simple sparse” : une seule entrée par neurone. Les neurones “moteurs” quant à eux sont analogues aux motoneurones qui activent la contraction musculaire et donc le mouvement des organismes [Fitzpatrick, D. (2001) The Primary Motor Cortex], on les représente dans notre cas comme une couche linéaire de neurones où les sorties seront liées à l’action du réseau de neurone complet.

La vraie particularité du projet est que les couches intermédiaires “fonctionnalisées” ne sont pas construites manuellement, mais adaptées par un processus évolutif de sélection. Ainsi, l’objectif est de faire émerger une structure intermédiaire fonctionnalisée à notre problème, ici le “jeu du chat et la souris”. Les connexions intermédiaires sont initialement aléatoires avec au minimum une connexion entre la couche “n+1” et “n”, pour avoir toujours un lien entre l’entrée et la sortie et ne pas avoir de problème lors des calculs de poids des neurones. Au cours de l’apprentissage, on s’attend à ce que le réseau soit de plus en plus structuré.

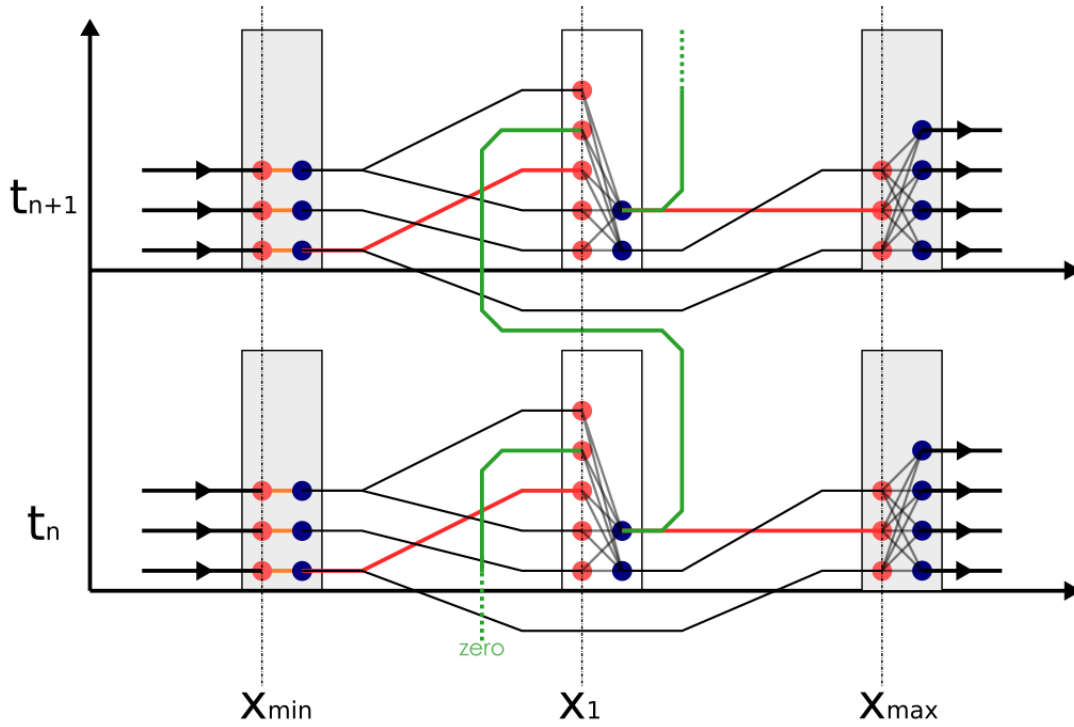


Figure 1: Réseau monocouche à connexion aléatoire; Le même réseau est représenté 2 fois à deux intervalle de temps à la suite, les nombres entrée/sorties sont pour l'exemple. De bas en haut : La commutation entre deux temps. De gauche à droite : L'entrée vers la sortie.

Comme les connexions entre l'entrée d'une couche  $X_n$  et la sortie d'une couche en  $X_{n+m}$ , avec  $m \geq 0$ , n'est pas possible pour le calcul de la rétropropagation à un temps donnée, on se trouve avec la connexion  $X_{n+m,t-1} \rightarrow X_{n,t}$ . Ce type de connexion laisse apparaitre une forme de réseau récurrent pour le calcul du gradient, mais comme nous voulons éviter la structure artificielle comme LSTM qui empêche le *vanishing gradient* des réseaux récurrents [Ref LSTM], nous avons considéré cette entrée à "t-1" comme une *entrée virtuelle*. Cette entrée virtuelle, n'est pas la solution la plus élégante, mais est la plus flexible dans le cas où les connexions sont complètement désordonnées et réduit le temps de calculs car la taille de l'entraînement est linéaire. L'entrée virtuelle consiste au détachement de la sortie du graphes computationnelle à l'algorithme du gradient. On obtient un réseau de "neurones à propagation avant" que l'on qualifiera dans le programme de *pseudo-récurrents* à cause des entrées virtuelles.

## 1.2 Construction du graphe

La topologie du réseau présenté précédemment peut être vue comme un graphe orienté acyclique, en effet, les connexions entre l'entrée  $X_n$  et la sortie  $X_{n+m}$  sont vues comme des entrées virtuelles indépendantes de la rétropropagation. Les connexions intercouche peuvent être négligées pour la construction initiale du graphe, car elles sont complètement connectées (voir code PyTorch *nn.linear*). Pour les connexions entre les couches, on se retrouve face à plusieurs contraintes :

- Quel est le nombre minimal de connexions pour que le réseau soit complètement connecté ?
- Combien de connexions sont attribuées par couche ?
- Comment attribuer les connexions par couche de neurone ?
- Comment stocker ces informations pour reconstruire le graphe ?

Empiriquement, le nombre total de perceptrons dans la couche intermédiaire " $N_h$ " est initialement défini comme la racine entière du nombre d'agent par génération d'entraînement et ne peut dépasser 32.

**Propriétés “Unicité des entrée”:** Les liens/connections entre les différentes couches de neurones depend des entrée de celle ci. On peut se connecter plusieurs fois à une sortie, mais “une” entrée n’a qu’une “seule et unique” entrée.

À partir de cette propriétés, si on veut que toute les entrée soit non vide, on peut definir le nombre minimal de connection pour notre reseau comme la somme du nombre de perceptron de la couche intermediaire “ $N_h$ ” et du nombre de perceptron en entrée “ $N_i$ ”. Ainsi, pour que le graphe soit complet, la relations du nombre minimal de connection “ $C_{min}$ ” est :

$$C_{min} = N_h + N_i$$

Le nombre de connection total “ $C_{tot}$ ” est ensuite attribué aléatoirement dans l’interval  $[C_{min}, 2C_{min}]$ . Puis le nombre de couche intermediaire du reseau “ $N_L$ ” est attribué aléatoirement dans l’interval  $[1, C_{tot}]$ . Lorsque que “ $C_{tot}$ ” et “ $N_L$ ” sont défini, il reste à définir le nombre de connection et de perceptron que va avoir chaque couche de neurone.

**Propriétés “Connection limité” :** Il ne peut y avoir moins de “une et unique” connection par couche, mais elle ne peut dépasser l’écart entre “ $C_{tot}$ ” et “ $N_L$ ”.

**Propriétés “Perceptron limité” :** Il ne peut y avoir moins d’“un seule et unique” perceptron par couche, mais elle ne peut dépasser l’écart entre “ $N_h$ ” et “ $N_L$ ”.

À partir de ces deux propriétés, on peut definir par récurrence l’attribution aléatoire du nombre de connection et perceptron par couche intermediaire de neurone. Le schéma de l’évolution de la densité de probabilité uniforme à discrétiser est le suivant :

$$f_{C_n}(x) = \begin{cases} \frac{1}{(C_{tot}-C_{n-1}-N_L-n)-1} & \text{pour } 1 \leq x \leq (C_{tot} - C_{n-1} - N_L - n) \\ 0 & \text{sinon} \end{cases}$$

$$f_{N_n}(x) = \begin{cases} \frac{1}{(N_h-N_{n-1}-N_L-n)-1} & \text{pour } 1 \leq x \leq (N_h - N_{n-1} - N_L - n) \\ 0 & \text{sinon} \end{cases}$$

Une fois l’attribution du nombre de connection réalisé par couche, il est necessaire de positionner ces couches spatialement pour le calcul de la rétropropagation. Dans notre cas, nous avons limité le nombre de couche à 32, avec une seule et unique couche possible par position spatiale discrete. Ainsi la position des couches intermediaire est attribué aléatoirement et sans remplacement dans l’interval  $[1;31]$ , où la position **zero** est reservé à l’entrée et **32** à la sortie. Les positions spatiales des couches de neurone n’influence pas le calcul de la rétropropagation, ainsi, les connection entre couche sont relative à l’ordre positionnel. On stocke ensuite l’ensemble des connection possible dans une liste. Mais une fois le positionnement des couches réalisé, comment connecter les noeud de facon à ce qu’il y ait un lien entre l’entrée et la sortie et qu’il n’y ait pas que des entrée virtuelle ?

Pour cette question, nous avons considéré 3 cas où l’attribution des connections suit des loi uniforme discrétisé avec des densité de probabilité différente. On distingue les 3 cas suivants :

1. La premiere connection : on se connecte à l’une des sortie de la couche la plus proche derriere, si c’est la premiere, on se connecte à l’une des sortie des neurones d’entree. En effet, il est necessaire qu’il y est au moins un chemin qui mene d’entrée vers la sortie pour la retropropagation, ce cas garantis cette condition.

La probabilité est directement défini par  $P(X_k) = \begin{cases} 1 & X_k = X_{k+1} - 1 \\ 0 & \text{sinon} \end{cases}$  pour ce cas.

2. Tant que toute les connections n’ont pas été attribué : On privilégie les connections vers l’arriere à 2/3 des probabilité. De cette facon, on limite l’excess d’entrée virtuelle sans que cela ne soit impossible. Cette regle à été défini empiriquement et ne semble pas avoir d’effet sur le processus de selection sur le long termes (non calculé). La probabilité est calculé à partir de la discretisation de la densité de probabilité discrétisé

$$f_X = \begin{cases} \frac{2}{3} & \text{pour } a \leq x \leq b \\ \frac{1}{3} & \text{pour } b \leq x \leq c \text{ dans ce cas.} \\ 0 & \text{sinon} \end{cases}$$

3. Lorsque toute les connection avec les neurones de sortie ont été au moins attribué une fois, alors il n’y a plus de contrainte spatiale au choix des neurones de sortie : La probabilité suit une loi uniforme sur l’ensemble de l’intervalle de definition spatiale.

Au cours de ce processus, les connections des couches au sortie sont listées dans l'ordre d'attribution par couche, ce qui nous donne une liste d'adjacence par couche. Chacune des listes ont par définition la taille du nombre de connection par couche. En representant l'ensemble des listes dans une matrice d'adjacence, on remarque que la diagonale sépare les connections en amont des connection en aval. La matrice à par définition la structure d'un graph orienté.

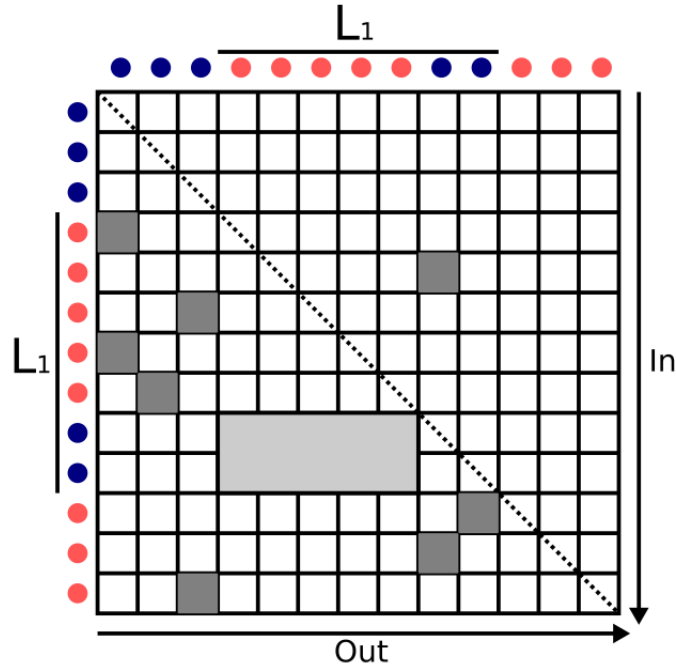


Figure 2: Matrice adjacente des liens du reseau de neurone précédent. En rouge, les noeud de connection (entrée d'une couche) et en bleu, les noeuds de perceptron (sortie de couche). Les case grisé correspondent à une connection de poids constants (1 par défauts).

Cette matrice, n'est pas utiliser dans le processus de restructuration du reseau. La matrice etant creuse, on peut se limiter à un ensemble de liste adjacente par couche de neurone. La representation en liste simplifie la restructuration du reseau et optimise la taille d'utilisation de la mémoire. Cette matrice sera toutefois necessaire pour l'analyse des donnée de graphe, on la reconstruira à posteriori. Les connection intercouches representé en gris clair sur la [Fig 2.] ne sont pas stocké en mémoire car ils sont définit par des opérations tensoriel déjà implémenté.

## 2 Methodes d'apprentissage

Les methodes d'apprentissages standards ont besoin d'un grand nombre de donnée pour que l'ajustement des poids permettent à un modele d'être prédictif [ref]. Avec l'ere du Big Data, cette approche fut concluante, meme si elle risque de favoriser le surapprentissage d'un reseau en classifiant trop avec les données d'entraînement [ref]. Pour autant, dans le vivant, il n'y a pas besoin d'un tres grand nombre d'experience pour qu'un cerveau apprenne efficacement un probleme [ref]. Dans certain cas, le cerveau est meme pré-cablé chez certaine espece pour repondre à certaine fonction, on peut citer comme exemple la marche des enfants à la naissance [ref] ou des mammifere juvénile proie [ref]. C'est le principe de fonctionnalisation cérébrale qu'on essaye ici de reconstituer en selectionnant des reseau "optimum" à la descente de gradient d'un probleme de decision. A partir du graphe vu precedemant, on peut reconstruire les connections entre les différentes couches du *réseau de neurones à propagation avant*. L'algorithme "forward" de la plupart des bibliothèques logiciel se retrouve restructuré.

---

**Algorithm 1** Forward construction by Adjacency List;  $\mathbf{X}$  correspond au Tenseur d'entrée de taille  $I \times \text{Batch}$ .  $\mathbf{NET}$  correspond au information des neurones (index, position, liste adjacente).  $\mathbf{Trace}$  correspond à la mise en mémoire des sorties des couches modulés des liste de neurones  $\mathbf{Layers}$  (PyTorch : *nn.ModuleList*).  $\mathbf{h}$  correspond à la copie détaché de la liste des Tenseurs  $\mathbf{Trace}$  à t-1.

---

```

Batch size, Input size  $\leftarrow \dim(\mathbf{X})$ 
BATCH_  $\leftarrow \text{np.arange}(\text{len}(\mathbf{x}))$ 
Trace[-1]  $\leftarrow \mathbf{Layers}[-1](\mathbf{X}.\text{view}(\text{Batch size}, \text{Input size}, 1)).\text{view}(\text{Batch size}, \text{Input size})$ 
for i, network index ordered by position do :
    tensor  $\leftarrow []$ 
    for j,k in NET[i, -1] :
        if j == 0 :
            tensor += [Trace[-1][BATCH_,None,k]]
        else :
            if (NET[i, 3] >= NET[i, 3]) :
                tensor += [h[j][BATCH_,None,k]]
            else :
                tensor += [Trace[j][BATCH_,None,k]]
    tensor_in  $\leftarrow \text{torch.cat}(\text{tensor}, \text{dim}=1)$ 
    Trace[i]  $\leftarrow \mathbf{Layers}[i](\text{tensor\_in})$ 
    Trace[-2]  $\leftarrow \mathbf{Layers}[-2](\text{tensor\_in})$ 
for t in range(len(Trace)):
    h[t][BATCH_]  $\leftarrow \text{Trace}[t][\text{BATCH\_}].\text{detach}()$ 
return Trace[i], Trace[-2]

```

---

Une fois le reseau completement cablé, celui ci est equivalent à un *réseau à propagation avant* avec plusieurs entrée intermédiaire. Ces entrée intermédiaire corresponde comme on l'a vu dans la partie precedante au *entrée virtuelle*. L'objectif qui suit est d'optimiser le calcul de la fonction de cout  $J(a, b)$  avec  $J(a, b) = \frac{1}{2m} \sum_{i=1}^m (f - y)$ . Dans notre cas, on distingue deux etapes dans l'optimisation du calcul la fonction de perte, d'abords à "**temps courts**" qui correspond à un apprentissage classique de descente de gradient à chaque cycle de reproduction, et à "**temps long**" où l'on va selectionner à chaque étapes les reseau qui auront le mieux reussi pendant de l'entrainement à temps court. La premiere étapes correspond à une methodes classique en apprentissage par renforcement, le Q-learning, la seconde, plus exploratoire, correspond à l'ajout de mutation dans le graphe du reseau qui va changer les connections entre les différents noeuds.

## 2.1 Temps courts : Q-Learning

A temps courts on cherche un algorithme de descente de gradient adapté à notre probleme. La plupart du temps, on donne un jeu de donnée d'entrainement adapté à un probleme, comme par exemple *MNIST* pour la reconnaissance des chiffres au format image, ou encore, *ImageNet* pour la detection et classification d'image d'objet. Dans notre cas, on a essaie de reproduire un systeme "vivant" caractérisé par un automate cellulaire [Voir partie suivante]. Ainsi, le système suit une succession d'états et d'action distincts dans le temps et ceci en fonction de probabilités de transitions. L'évolution du systeme correspond à un processus de decision markovien, représenté par une chaine de markov. Une chaine de markov est une suite de variable aléatoire  $(X_n)$  dans l'espace probabilisé  $(E, B, P)$ , où pour chaque  $n$ , sachant  $X_n, X_{n+1}$  indépendant de  $X_k$ , on a la probabilité de transition (Hypothèse de Markov) :

$$P(X_{n+1} = i_{n+1} | X_1 = i_1, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n)$$

Dans notre cas, on associe à chaque transition, des action et des récompense associé à l'agent de facon à le guider dans le temps. On peut représenter cela suivant le couple de matrices  $(T, R)$ , où  $\mathbf{T}$  correspond à la matrice de transition et  $\mathbf{R}$ , la matrice des récompense. La complexité est que l'on ne connait pas la probabilité de transition.



Le but dans un processus décisionnel markovien est de trouver une bonne « politique » pour le décideur : une fonction  $\pi$  qui spécifie l'action  $\pi(s)$  que le décideur choisira lorsqu'il sera dans l'état  $s$ . Une politique décrit les choix des actions à jouer par l'agent dans chaque état. L'agent choisit une politique à l'aide de la fonction de récompense  $R$ . Lorsqu'une politique et un critère sont déterminés, deux fonctions centrales peuvent être définies :  $V$ , la fonction valeurs des états qui représente le gain engrangé par l'agent s'il démarre à l'état  $s$  et  $Q$ , la fonction de valeur des états-actions qui représente le gain engrangé par l'agent s'il démarre à l'état  $s$  et commence par effectuer l'action  $a$ . Les expressions de  $V$  et  $Q$ , sont ainsi déterminées par les relations de récurrence :

$$V_{k+1}(s) = (1 - \alpha)V_k(s) + \alpha[r + \gamma V_k(s')]$$

$$Q_{k+1}(s, a) = (1 - \alpha)Q_k(s, a) + \alpha[r + \gamma \max_{a'} Q_k(s', a')]$$

Cette dernière correspond à l'équation de Bellman et c'est elle qui est utilisée pour calculer la fonction de coût du réseau de neurone lorsqu'on réalise une action. La première équation ne peut être utilisée seule car elle nous donne l'efficacité de l'agent pour une action donnée, mais peut être utilisée en cas d'utilisation de réseau adversariaux [Ref actor-critic]. Dans notre cas, on réalise initialement une suite d'événements avec une matrice de transition aléatoire (réseau non entraîné), puis on calcule la fonction de perte entre la prédiction des valeurs  $Q$  obtenue au cours de l'expérience avant l'action et le résultat de l'équation de Bellman  $Q$  après l'action d'un batch d'entraînement donnée, puis on répète  $N_{cycle}$  fois cette étape avant le prochain cycle de reproduction.

Code PyTorch :

```
old_state, action, new_state, reward, DONE = MEMORY
actor = MODEL(old_state)
pred_q_values_batch = torch.sum(actor.gather(1, action), dim=1).detach()
pred_q_values_next = MODEL(new_state)
target_q_values_batch = reward + (1-DONE)*GAMMA*torch.max(pred_q_values_next, 1)[0]
MODEL.zero_grad()
loss = criterion(pred_q_values_batch, target_q_values_batch)
```

L'action "acteur" correspond à la sortie du réseau, et correspond à la probabilité d'action optimale. Normalement, à chaque pas de temps, on choisit la valeur maximale en sortie du réseau et l'on attribue une probabilité d'action aléatoire, c'est le **dilemme exploration-exploitation**. Ce dilemme permet au réseau de neurone d'explorer des paramètres et de ne pas se stabiliser dans une vallée non optimale. Mais dans notre cas, nous avons choisi d'attribuer une probabilité d'action à chaque "*pas*" de l'entraînement à partir de la sortie du réseau, qui est équivalent à un dilemme d'exploration-exploitation.

Code Python :

```
DILEMNA = np.squeeze(action_probs.detach()).numpy()
p_norm = DILEMNA/DILEMNA.sum()
next_action = np.random.choice(self.IO[1], p=p_norm)
```

## 2.2 Temps longs : Structure neuronale

À partir des différents modèles de réseau entraînés à temps courts, l'objectif qui suit est de sélectionner ceux qui ont eu le meilleur score d'entraînement, puis de modifier légèrement la structure du graphe neuronal. Le calcul des scores est relatif à l'environnement d'entraînement et est très utilisé dans les modèles évolutifs. L'attribution des points sera plus détaillée dans la partie 3. Le principe est le suivant : Lors du premier entraînement à temps court, on a  $N_r$  réseaux en parallèle et distincts, chacun va suivre le processus d'entraînement vu dans la sous-partie précédente. Ensuite, les  $N_b$  réseaux ayant le meilleur score sont sélectionnés,  $N_b$  réseaux gardent la même topologie pour le cycle suivant,  $N_m$  réseaux par  $N_b$  meilleurs réseaux héritent d'une mutation aléatoire. Enfin  $N_c$  nouveaux réseaux aléatoires sont introduits dans le processus d'entraînement suivant, ceux-ci peuvent hériter de paramètres optimisés des réseaux précédents, mais pas de la structure du réseau, c'est les réseaux "compétiteurs". Les nombres d'agents sont définis comme  $N_r = (N_b + 1)^2$ , de façon à ce que  $N_r$  est une racine entière,  $N_m = N_b$ , tel que le nombre total de mutations soit  $N_{mtot} = N_b^2$  et enfin  $N_c = N_b + 1$ , de cette façon, on a bien  $N_r = N_b + N_{mtot} + N_c$ . Dans notre modèle, on distingue 5 types de mutations sur le graphe neuronal :

1. Ajouter une connexion : Ne change pas le nombre de couche neuronale, mais ajoute une connexion à l'une d'entre elles. Comme on n'ajoute pas de neurone, la liste des connexions possibles ne change pas et le nouveau nœud va être connecté aléatoirement à un neurone pré-existant.

2. Ajouter un neurone à l'une des couches : Comme on ajoute une connection de sortie possible, on renouvelle la liste des connection. Mais vu que le reseau doit etre complet, on rajoute également une connection dans l'une des couches qui va se connecter exclusivement à cette derniere connection de sortie.
3. Ajouter une couche 1\*1 : L'ajout d'une couche n'est composé que d'une connection d'entrée et d'une seule et unique sortie, ce qu'y est equivalent à un seul neurone. La position de la couche doit etre différent de celle precedante et compris entre [1,31]. L'entrée de cette nouvelle couche se connecte uniquement vers l'arriere (pas forcement ? juste position différent ?), par contre, il n'est pas necessaire que la sortie soit connecté vers l'avant, le reseau etant complet, seul lui meme est interdit. Si ce neurone se connecte vers l'arriere, il devient equivalent à un générateur d'entrée virtuelle. Comme on ajoute une connection de sortie, la liste de connection et une nouvelle connection est ajouté tout comme l'ajout d'un neurone. **(necessite correction)**
4. Enlever une connection doublon : Enlever une connection quelconque n'a pas été envisagé dans notre cas, car ils changerait radicalement la structure du reseau. En effet, elle supprimerai possiblement plusieurs connection, ce qui rendrait fortement le reseau incomplet. Par contre, comme le reseau peut etre connecté plusieurs fois à la meme sortie, il est possible d'enlever une connection doublon. Cela ne changerait pas la liste des connections de sortie possible. **(necessite correction)**
5. Enlever un neurone : Cette opération contient deux contrainte, on ne peut pas supprimer un neurone qui correspond au premier lien d'une couche donnée et que le décalage des connections soit possible. Corrolaire des définitions : on ne peut pas supprimer une couche qui ne contient qu'un seul et unique neurone et de meme pour la connection d'entrée. Tout comme l'ajout de neurone, la liste des connections de sortie possible est mise à jours. **(necessite correction)**

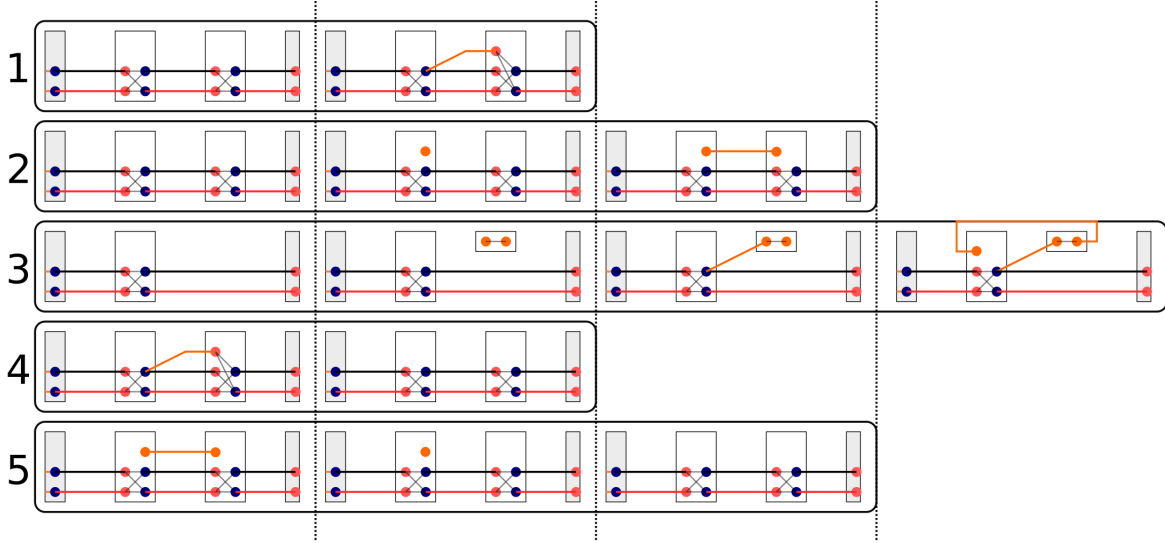


Figure 3: Algorithm visual des mutations. Respectivement les 5 types de mutations citer en paragraphes. De gauche à droite, l'etat initial vers l'etat final. Les propriétés du graphe sont à titre indicatives, cette representation tres simplifié ne reflète pas l'optimum d'une fonction. Les opérations  $1 \leftrightarrow 4$  et  $2 \leftrightarrow 5$  sont symétriques, l'opération 3, n'a pas de symetrie stable simple.

Dans notre modele, on conserve au minimum un neurone par couche et toujours le premier, et on ne supprime pas les couches de neurone. Ce choix plus simple permet de conserver la symetrie des opérations  $1 \leftrightarrow 4$  et  $2 \leftrightarrow 5$  [voir Fig 3], mais aussi de maintenir une structure vestigiale du reseau. En effet, au cours de l'evolution certaine structure sont maintenu, mais ne sont pas les choix optimums por remplir certaine fonction. Par exemple, le chemin qu'emprunte les vaisseau sanguin entre la tete et le coeur chez les mammifere emprunte un chemin non optimal, pourtant il est maintenu car serait trop couteux evolutivement pour changer de trajectoire [ref necessaire]. La vallée de stabilité evolutive lié à ce choix criticable est plus abordé en discussion.

### 3 Regles du jeu : TAG-GAME

Dans le cadre des processus de decision markovien, les donnée sont généré au cours de l’entrainement, c’est l’environnement d’entrainement. Cet environnement contient des regles et s’apparente à un jeu, la bibliotques la plus utilisé en python est *OpenGym* et sa structure à servi d’exemple pour realiser l’environnement de notre probleme. Dans notre cas, on souhaite reproduire certaine stratégie de survie que l’on observe dans le vivant, en particulier, les stratégies de prédatons et de fuites d’une proie. On a choisi pour cela le **jeu du chat et de la souris**, ou *Tag game* en anglais, en effet, on alterne entre proie lorsqu’on est souris, et prédateurs lorsqu’on est chat, ce qui pousse le modele à s’adapter à deux configurations différente. Le chats et la souris ont des “couleurs” différentes suivant les états chat/souris/adversaire/agent. Les étapes d’une partie sont les suivantes :

1. Initialement, l’agent est une proie “la souris”. Il doit eviter de se faire attraper par le prédateur, “le chat”. Le prédateur n’est pas un agent “intelligent”, celui-ci calcul uniquement le déplacement optimal qui minimise la distance euclidienne entre le chat et la souris pour l’etat suivant. *Chat = 1, Souris = 2.*
2. Lorsque l’agent se fait “attraper” par le prédateur, il en devient lui meme un et les role s’inverse. L’agent doit maintenant attraper la proie, la nouvelle souris. La souris cette fois n’est pas un agent “intelligent”, celui ci calcul uniquement le déplacement optimal qui maximise la distance euclidienne entre le chat et la souris pour l’etat suivant. *Chat = 3, Souris = 4.* La nouvelle position des agents/adversaire est calculer comme la collision entre les deux pour ne pas rester bloquer sur la meme position au temps  $t+1$ . La collision est définit comme la symmetrie axiale des deux vecteurs de deplacement à l’instant  $t$ , soit l’équation (**à implementer**) :

$$\bullet \begin{cases} Thales \\ Pythagore \end{cases}$$

3. On revient à l’étapes 1 si l’agent attrape la souris, la partie s’arrete apres  $N_g$  pas de temps.

On a dans ce cas, un agent qui interagit avec un environnement de jeu. L’agent contient les informations de **vue** et d’**action** qui va transmettre à l’environnement. L’environnement contient les information de la position de l’agent et son adversaire, ainsi que les regles de jeu et comptage de points. L’environnement peut etre vue comme une grille à limite périodique (CLP) de dimension  $N \times M$  où les agents et adversaire sont des automates cellulaires avec des fonctionnement différents. L’agent dispose de 9 entrée “états”, 3 sortie “action” et un reseau de neurone entre les deux, où les propriété positionnel entrée/sortie sur la grille sont généré aléatoirement en debut d’experience, puis maintenu à la descendance. Ainsi, l’agent lit 9 case sur une grille  $5 \times 5$  centré sur la position de l’agent, soit 36% de son environnement local, et l’agent se déplace d’une seul case parmit 3 mouvement d’une grille  $3 \times 3$  centré sur la position de l’agent, soit 33,3% de son environnement local. Par contre, l’adversaire voit uniquement la position de l’agent et de lui-meme, mais n’est qu’une fonction mathématique se deplacant suivant 4 trajectoires, haut (0,1), bas (0,-1), gauche (-1,0) et droite (1,0), minimisant ou maximisant la distance entre elle et l’agent suivant son état “chat” ou “souris”. La informations positionnel sont discretisé (grille), le calcul du déplacement de l’adversaire revient à :

$$\begin{cases} d_2 = \sqrt{(v_{adv} - v_{agent})^2} \\ m_{vt} = \begin{cases} m_{vt} - \min(d_2) & \text{if IT} \\ m_{vt} + \max(d_2) & \text{else} \end{cases} \end{cases}$$

Enfin, l’environnement attribue des points à chaque étapes du jeu, l’ensemble des points donne le score de l’agent. L’attribution des points à l’agent est un parametre important en apprentissage par renforcement, un desequilibre des points peut rendre soit un agent tres “agressif” ou encore à l’opposé “passif” [ref necessaire et importante]. Pour cela, les points ont été reflechi en fonction de la taille de grille de jeu, pour une grille de jeu  $16 \times 16$ , la moyenne des déplacement pour atteindre le centre quelque soit la position est d’environ 8 déplacement. Dans cette configuration, pour un calcul équilibré, les comptages des points sont les suivants :

- -1 par “pas” de temps où l’agent est un prédateur “chat”
- +10 si l’agent attrape la proie “souris” et devient la proie à son tour.
- +1 par “pas” de temps où l’agent est une proie “souris”.

- -10 si l'agent est attrapé par le prédateur "chat" et devient le prédateur à son tour.^

Corrolaire : L'agent restant le plus longtemps sous l'état proie (souris) et le moins longtemps sous l'état prédateur (chats) à le plus grand score de jeu.

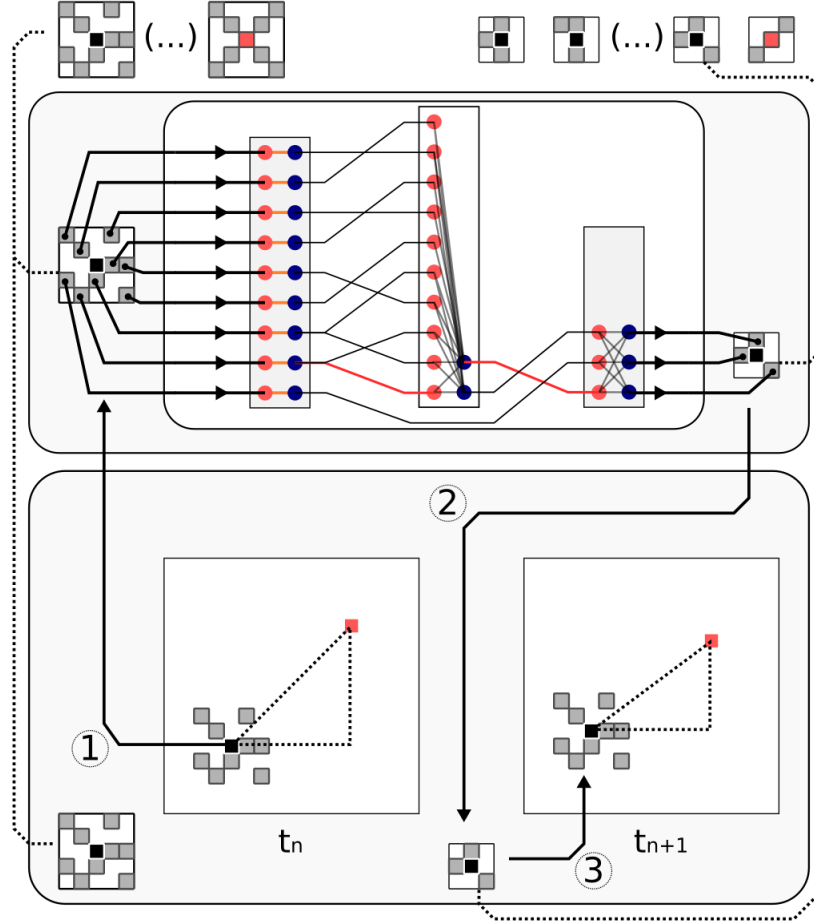


Figure 4: Les interactions entre un agent et un environnement de jeu. En haut, la génération aléatoire des positions relative possible des entree et sortie. Suivi des interactions entre l'agent en haut et l'environnement de jeu en bas. 1 : envoi de l'environnement des informations d'entrée (pixel) à l'agent; 2 : envoi de l'information de sortie de l'agent (mouvement) à l'environnement; 3 : mise à jour de la carte de jeu, l'adversaire effectue son déplacement à ce moment.

Les case des positions relative d'entrée et de sortie sont défini aléatoirement, mais comme le pourcentage des case utilisé sur la grille est au alentour de 35%, les combinaisons possibles sont de l'ordre factoriel. Pour la sortie, on a 3 positions parmi un total de 9 cases, ce qui donne  $\binom{9}{3} = 84$  et comme le probleme est symetrique au 4 mouvement de l'adversaire (exemple :  $(1, 2, 3) \iff (7, 8, 9)$ ), le nombre de combinaison est de l'ordre du nombre du nombre d'agent par génération. Par contre, pour l'entrée, on a 9 positions parmi un total de 25 cases, ce qui donne  $\binom{25}{9} = 2042975$  ce qui pose un probleme de convergence de la vision optimal. Mais, comme les quadrant d'observation sont equivalent, on peut considerer uniquement les 9 cases par quadrant, ce qui nous donne par ratio de 36%  $\binom{9}{3} = 84$ , cela est presque un ordre au dessus du nombre de génération et ne pose moins probleme de stabilité. Les cases ayant obtenu les meilleurs scores sont classé dans l'ordre et l'on attribue la somme des points par case sur les grilles d'observation et d'action. La normalisation de la grille donne la probabilité de choisir l'une des cases pour les etapes suivantes des "challenger" et sera plus abordé dans la partie résultats. Pour

les action, on observe quelque cas typique : 3-cyclique (Exemple :  $(1, 6, 8) \Rightarrow (\nwarrow, \downarrow, \rightarrow)$ ), 4-cyclique (Exemple :  $(1, 3, 8) \Rightarrow (\nwarrow, \downarrow, \downarrow, \nearrow)$ ), semi-2-cyclique (Exemple :  $(2, 6, 8)$ ), asymétrique (Exemple :  $(1, 2, 3)$ ) et statique (Exemple :  $(1, 5, 8)$ ). Les positions des entrée et des sorties sont donc aussi selectionné dans le processus evolutif des agents.

## Résultats

Pour générer l'ensemble des données expérimentale, il est nécessaire d'avoir plusieurs agents par génération pour la convergence de la densité de probabilité des positions d'entre/sortie et plusieurs génération pour verifier s'il y a une convergence des structures neuronales. Pour que le nombre d'agent par génération soit de l'ordre du nombre de combinaison d'entree/sortie, nous avons choisi 25 agents par cycle de reproduction. 25 est le carré de 5, ce qui compatible avec l'algorithme de mutation de la structure neuronale. Ce qui donne, 4 meilleurs agents par cycle, ceux ci sont maintenu au prochain cycle et chacun recoit 4 mutation, ce qui fait 20 agents ayant des propriétés analogue à la génération precedente, les 5 restants sont des nouveaux agents avec des structures neuronale aléatoire, mais avec informations d'entrée/sortie hérité de la densité de probabilité des evement precedent. Les grilles de jeu sont de  $16 \times 16$ , ce qui permet à l'agent de pas voir tout le plan de jeu, mais ne pas que l'adversaire soit trop loin lorsque l'agent est un prédateur. L'avantage de cette configuration est que l'on peut représenter l'ensemble des environnement à un temps donnée par une grille  $(5 \times 16) \times (5 \times 16)$ , ce qui facilité la visualisation et l'interpretation des resultats. Il est possible aussi de représenter cette grille par décomposition de deux nombre premier si le nombre d'agents par cycle n'est pas le carré d'un nombre. Cette methode correspond à la verification de la parité puis recurrence des nombre impaire jusqu'à la racine entiere supérieur du module du nombre d'agent par génération :

$$n \% \begin{cases} 2 \\ \text{impair}(\sqrt{n}) \end{cases} . \text{ Cette derniere n'a pas été utilisé ici, mais est inclu dans le programme.}$$

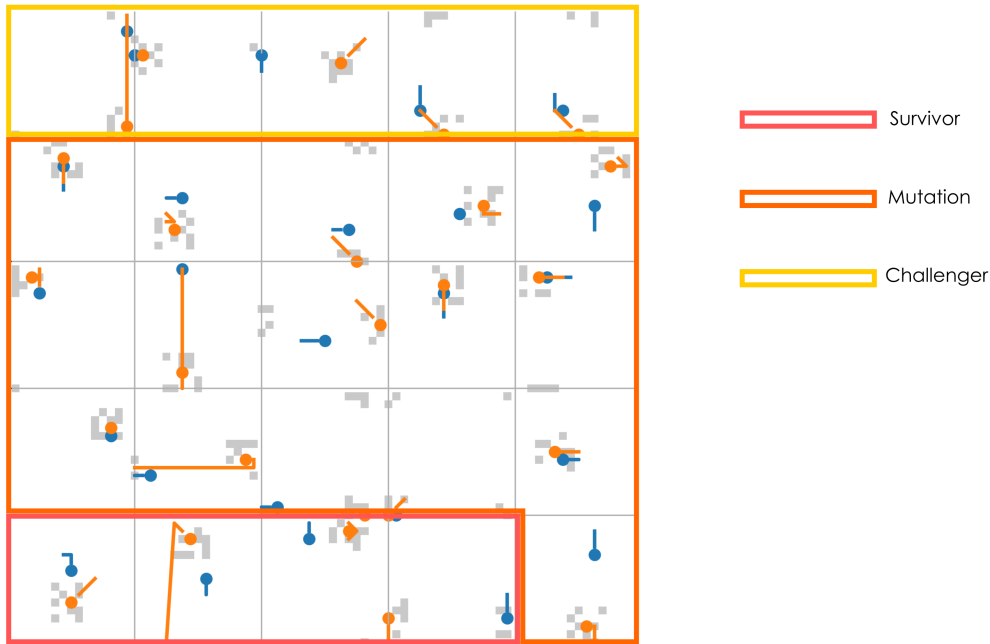


Figure 5: Representation d'une donnée expérimentale à un instant  $t$ . En encadré, les différentes catégories d'agent (survivant, mutation et compétiteur). En points bleu, les adversaires et en orange, l'agent intelligent. Les cases grisé correspondre à la vue de l'agent à un instant  $t$ .

Le nombre de génération n'est pas clairement défini, mais peut, tout comme le nombre d'agent par génération, être justifié par la loi des grands nombres, si l'on considère que la loi de probabilité à une espérance. On va donc verifier, pour 25 agents par génération, combien, il faudrait pour que la densité moyenne des entrée et des sortie

converge. Soit  $N_a = 25$ , le nombre de d'agent par génération,  $N_c = 84$  le nombre de configuration de combinaison maximal entre entrée et sortie et  $N_g$  le nombre de génération que l'on cherche. Soit  $X_n$ , la variable aléatoire indépendante nous donnant une combinaison à chaque tirage, la limite à l'infini de l'inégalité de markov devient :

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{1}{N_g} X - p\right| \leq \varepsilon\right) = 0$$

De cette façon, la moyenne est un estimateur de l'esperance. On trouve pour un  $\varepsilon$  de 1% et d'un calcul par itération directe, une valeurs de 10 générations pour que la moyenne converge. Cette approximation n'est pas valable pour la convergence du reseau de neurone, ainsi, on considere que les données experimentale seront exploitable qu'a partir de  $N_g > 10$ .

## 4 Caractérisation de la convergence de la densité positionnelle des entrées/sorties

L'entrée de l'agent est défini comme la combinaison de 9 cases parmi 25, sa sortie est défini comme la combinaison de 3 cases parmi 9 [Section §3]. On cherche à verifier l'existence ou non des configurations "vue" et "déplacement" optimales au probleme du jeu du chat et la souris. Dans notre cas, à chaque cycle de reproduction, on ordonne les scores pour chacun des agents, et on additionne les scores obtenu par case de la grille de l'agent sur la grille complete. Comme le probleme est symetrique suivant les 4 directions, on considere les 4 rotation de  $k\pi/2$  dans le comptage des points. On réalise cela par le produit matriciel des positions centré avec la matrice de rotation

$\begin{pmatrix} x_0 + x_c & y_0 + y_c \\ \vdots & \vdots \\ x_n + x_c & y_n + y_c \end{pmatrix} \cdot \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$ . La normalisation de la grille nous donne la proportion des meilleurs

scores obtenus des événement aléatoire. Si l'on considère que la distribution suit une loi avec esperance, alors le *théoreme de transfert* nous permet d'exprimer l'esperance d'une fonction d'une variable aléatoire  $X_n$  en fonction d'une intégrale convergente. Numériquement, il correspond à l'agorithme de *Monte-Carlo*, décrit comme :

$$G = E[g(X)] = \int g(x)f_X(x)dx$$

A partir de ce calcul, on obtient la densité de probabilité de succes lorsqu'on génère une nouvelle grille d'entrée "vue" et de sortie "déplacement". Par la loi des grands nombres, on a vue que cette densité convergent au moins à partir de 10 cycle de reproduction. On peut ainsi représenter ces deux densités sur des grilles à deux dimensions.

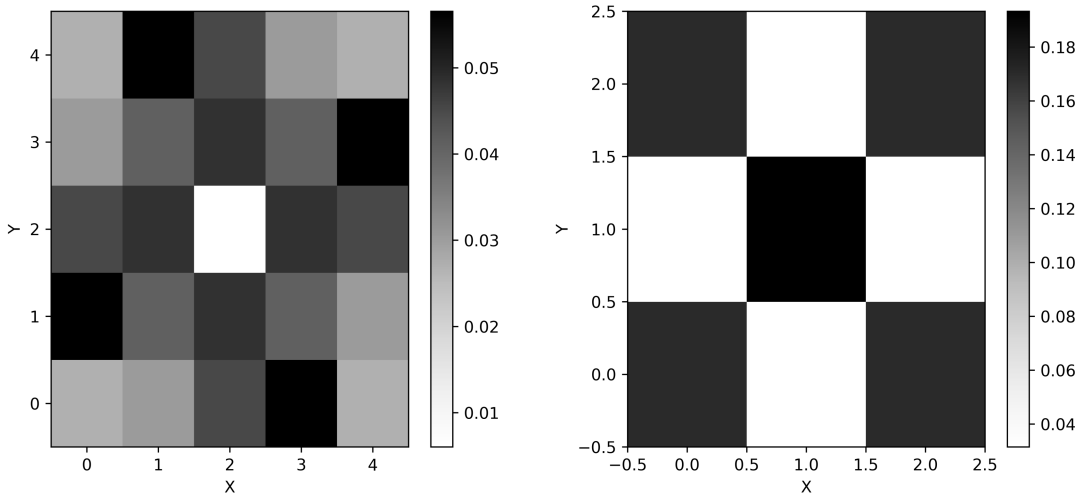


Figure 6: Mesure de la densité positionnel de succes des entree/sorties. A gauche, l'entree correspondant à la vue de l'agent et à droite la sortie, correspondant au mouvement relatif de l'agent. En nuance de gris le niveau de probabilité d'avoir plus de succes avec cette case.

On obtient pour l'entrée et la sortie deux distributions spatiale différente [voir Figure 6]. Pour l'entrée on observe 4 positions dominantes et 1 position centrale faible. Pour la sortie, on observe 4 positions en diagonal forte. En entree, l'ensemble des positions  $(0, 1)$ ,  $(1, 4)$ ,  $(3, 0)$  et  $(4, 3)$  à la probabilité la plus forte et est symétrique à l'ensemble  $(1, 0)$ ,  $(0, 3)$ ,  $(4, 1)$  et  $(3, 4)$ . Cet ensemble d'entrée correspond au position au quasi-extremité de la grille  $5 \times 5$ , ce qui pourrait permettre d'anticiper le mouvement de l'adversaire. Ensuite, les positions d'entree  $(1, 2)$ ,  $(2, 3)$ ,  $(2, 1)$  et  $(3, 2)$  corresponde au case qui entoure le centre de la grille, celle-ci permettent de "voir" l'adversaire lorsque celui ci est au plus proche en terme de mouvement. De plus, ces dernieres prolonge les cases en diagonale precedante. En sortie, on l'ensemble des position  $(0, 0)$ ,  $(0, 2)$ ,  $(2, 0)$  et  $(2, 2)$  ont par somme, la probabilité la plus forte. Cette ensemble de sortie correspond au position en diagonale. On remarque que les positions qui entoure le centre  $(0, 1)$ ,  $(1, 0)$ ,  $(1, 2)$  et  $(2, 1)$  ont les probabilités les plus faibles. Ces deux observations sur la sortie implique que le mouvements est privilégié en diagonale et semi-2-cycliques. Ces cases pourrait etre privilégier car l'adversaire n'a que 4 déplacement ( $\uparrow, \downarrow, \leftarrow, \rightarrow$ ) non diagonaux, ce qui donne un avantage en cas de déplacement diagonal. Par contre, l'absence de symetrie complete des mouvements de l'agents ne semble pas etre un désavantage pour l'obtention d'un meilleurs score.

On remarque que les positions optimales en entrée pourrait etre corrélé avec les mouvements optimaux de sortie. En effet, les mouvements diagonaux necessite de savoir si un adversaire est positionné à sa diagonale avant que l'adversaire est fait le mouvement à  $t+1$ . Par exemple, si l'agent peut faire les mouvements  $(1, 6, 9) \iff (\nwarrow, \emptyset, \swarrow)$ , si l'adversaire est en position relative  $(3, 0)$ , alors en  $t+1$ , l'adversaire sera en  $(3, 1)$ , si l'agent fait le mouvement  $(\swarrow)$ , dans ce cas, il sera à la meme position que l'adversaire ce qui lui fera perdre des points, alors qu'il sera à la position a plus éloigné si il fait le mouvement  $(\nwarrow)$ . Pour vérifier la correlation entre l'entrée et la sortie, on peut regarder la convergence des familles de densité des entrées et sorties. Pour cela, on mesure la densité projeté suivant une seule dimension spatiale entre  $t = 0$  et  $t = N_g$ , ainsi que sa variation entre chaque génération.

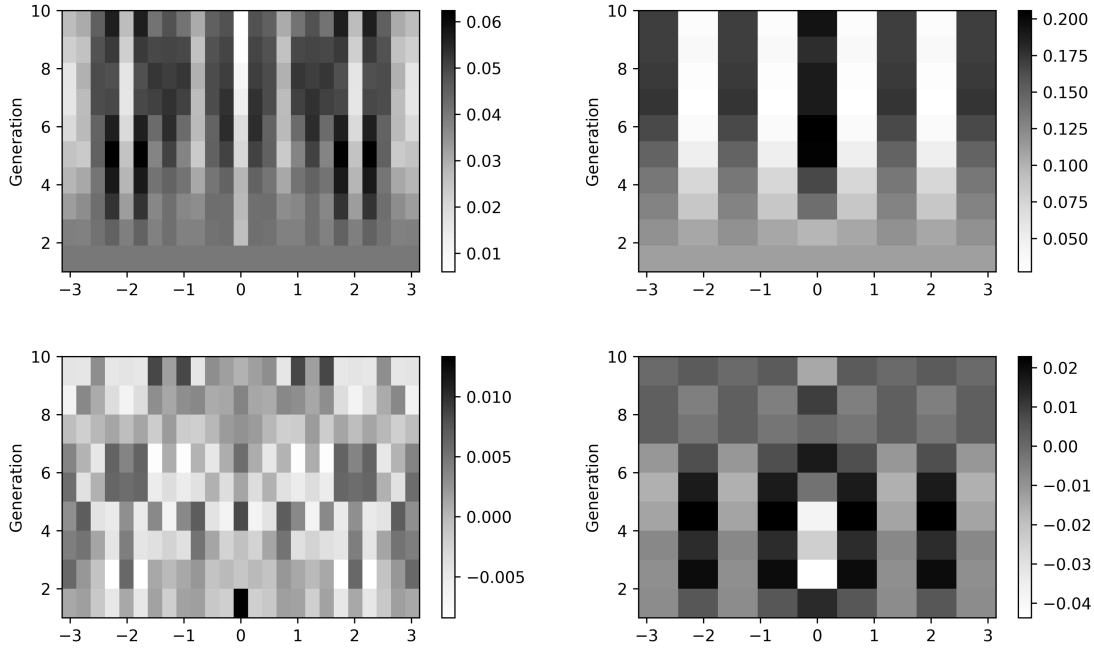


Figure 7: Mesure de la densité positionnel projeté de succes des entree/sorties en fonction du temps. À gauche, l'entrée et à droite la sortie. En haut, la densité de probabilité avec en nuance de gris le niveau de probabilité d'avoir plus de succes avec cette case. En bas, la variation de probabilité avec en nuance de gris le niveau de variation.

On obtient l'evolution de la distribution spatiale des entrée et des sorties [Voir figure 7]. On observe que pour l'entrée, certaine position se stabilise à partir de la 4eme génération, mais qu'il peut y avoir une alternance des positions en quasi-limite avec des positions plus centrale. Par contre, pour la sortie, les positions sont plus stable et se stabilise complètement à partir de la 5eme générations, la variation de la distribution devient quasi-homogène et

semble tendre à être constant. La variation de l'entrée est bien plus importante, on n'atteint pas d'homogénéité, même après 10 cycles, ce qui semble montrer qu'il y a des positions dominantes, mais celles-ci oscillent tout le temps entre plusieurs positions : la suite d'entrée semble être semi-alternée convergente. La convergence de la suite semi-alternée pour chaque position d'entrée  $u_n$  est vérifiée par le rayon de convergence  $R_n$  du critère de Leibnitz :

$$u_n = (-1)^n \epsilon_n \iff R_n \leq |a_{n+1}|$$

Comme la densité de sortie semble être un invariant, et que l'entrée oscille entre quelques positions d'instabilité, on peut considérer qu'il n'influencera pas trop fortement le processus de sélection de réseau une fois stabilisé. Ainsi les challengers qui hériteront de la densité de probabilité d'I/O ne devraient pas influencer leur victoire par leurs propriétés d'entrée/sortie, mais de la nouvelle structure aléatoire. Cette hypothèse forte, sera vérifiée dans les parties suivantes et plus discutée en fin de partie.

## 5 Caractérisation de la connectivité du réseau de neurone

Le réseau de neurone est défini comme un graphe orienté acyclique [1.2]. Comme celui-ci est généré aléatoirement, il est nécessaire d'explorer l'étude de certains paramètres de graphes pour caractériser s'il existe ou non des propriétés dominantes au cours de la sélection.

Contrairement à la mesure de la densité, au cours de la sélection, les agents ayant les meilleurs scores sont sélectionnés, mais, les nouvelles structures sont définies aléatoirement comme à l'instant initial. Les agents sélectionnés subissent 5 mutations aléatoires, on cherche à mesurer l'effet de la mutation sur la structure du réseau et la sélection. Pour cela, on peut représenter l'ensemble des matrices adjacentes ayant eu les meilleurs résultats au cours des générations.

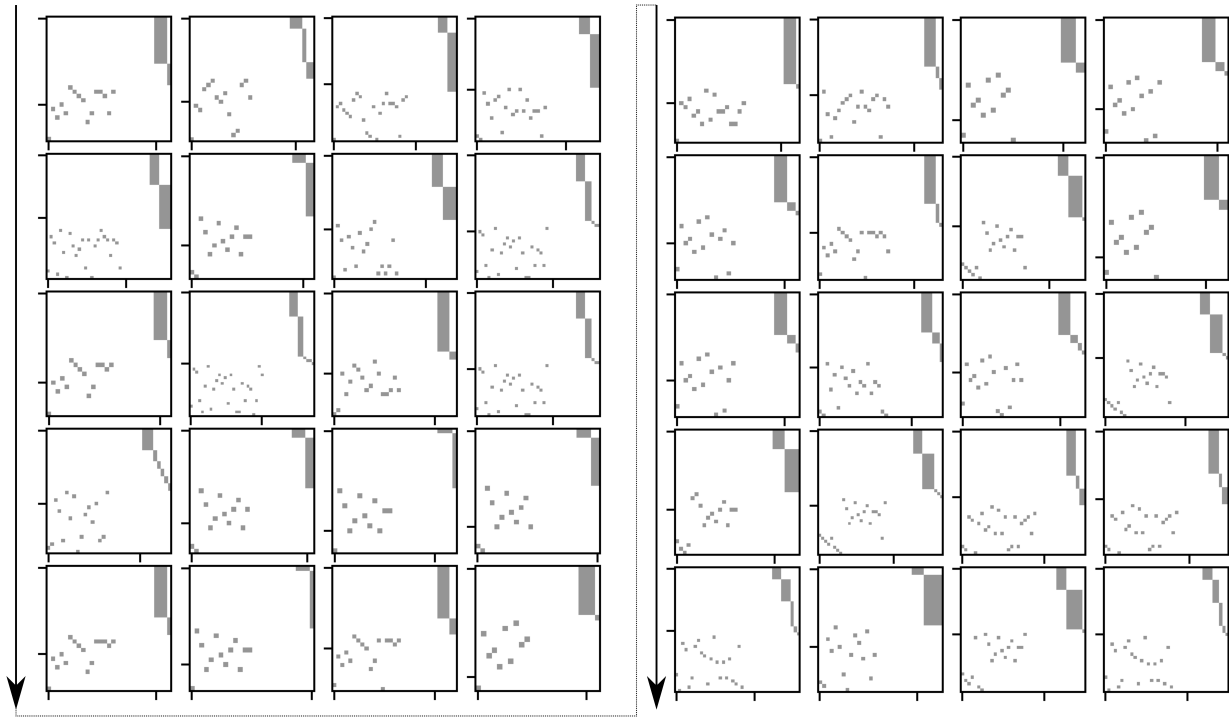


Figure 8: Evolution des réseaux des 4 meilleurs agents par générations. À gauche, de la première génération à la 5ème et à droite, de la 5ème à la dernière génération.

On obtient ainsi un ensemble de grilles de matrices adjacentes de taille  $N_b \times N_g$ , avec  $N_g$ , le nombre de génération et  $N_b$ , le nombre de meilleurs agents par génération. On n'observe pas de structure très organisée dominante au dixième cycle, le réseau ne semble pas converger vers une structure d'équilibre. Pour autant, la matrice adjacente



n'est pas suffisante pour avoir les informations des graphes, il est nécessaire de vérifier certains paramètres si l'on veut caractériser le réseau. Pour cela, on va comparer les paramètres de "degrés", "proximité", "intermédiarité" et "orthogonalité" entre les 4 meilleurs réseaux et les 4 pires réseaux au cours du temps.

Le degré (ou valence) d'un sommet d'un graphe est le nombre de liens (arêtes ou arcs) reliant ce sommet, avec les boucles comptées deux fois. Il est défini pour un graphe orienté comme (...), tel que  $deg(s) = d^+(s) + d^-(s)$ . L'intermédiarité est une mesure de centralité d'un sommet d'un graphe. Elle est égale au nombre de fois que ce sommet est sur le chemin le plus court entre deux autres nœuds quelconques du graphe  $g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$  (...). L'excentricité d'un nœud  $x$  est définie comme la somme des distances à tous les autres nœuds, et la proximité est définie par Bavelas comme l'inverse de l'éloignement  $C(x) = \frac{1}{\sum_y d(y,x)}$ . **Définir l'ensemble des paramètres étudié !**

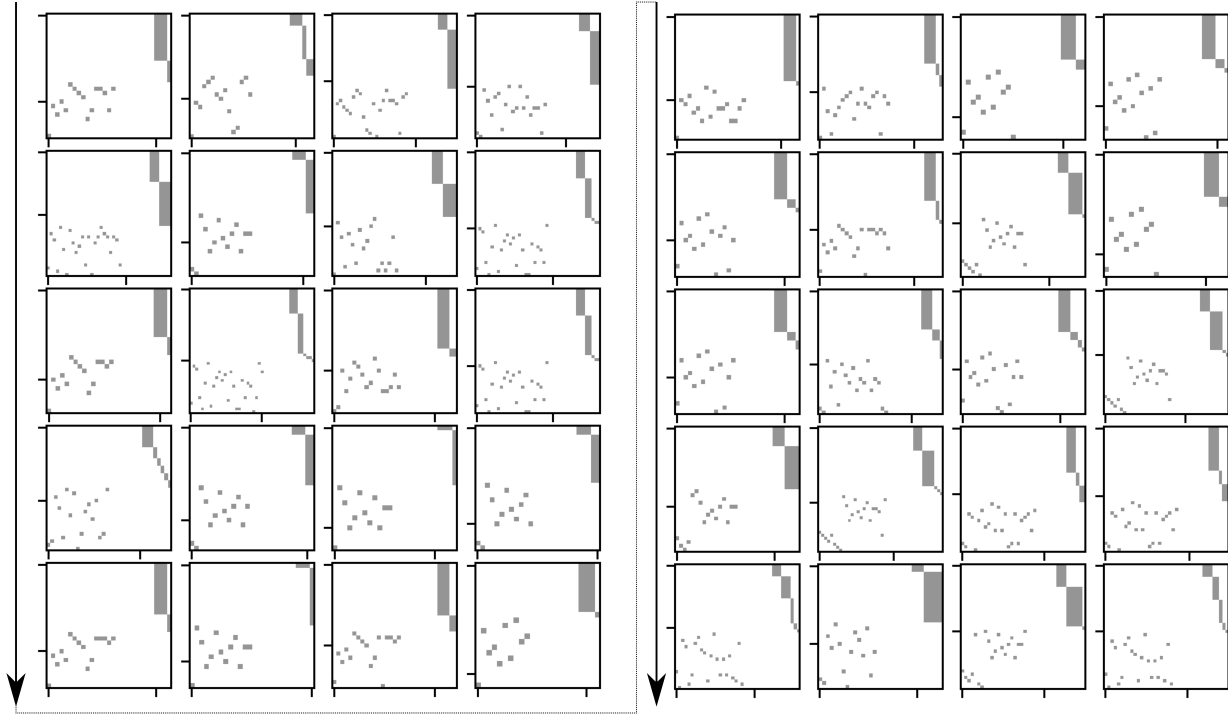


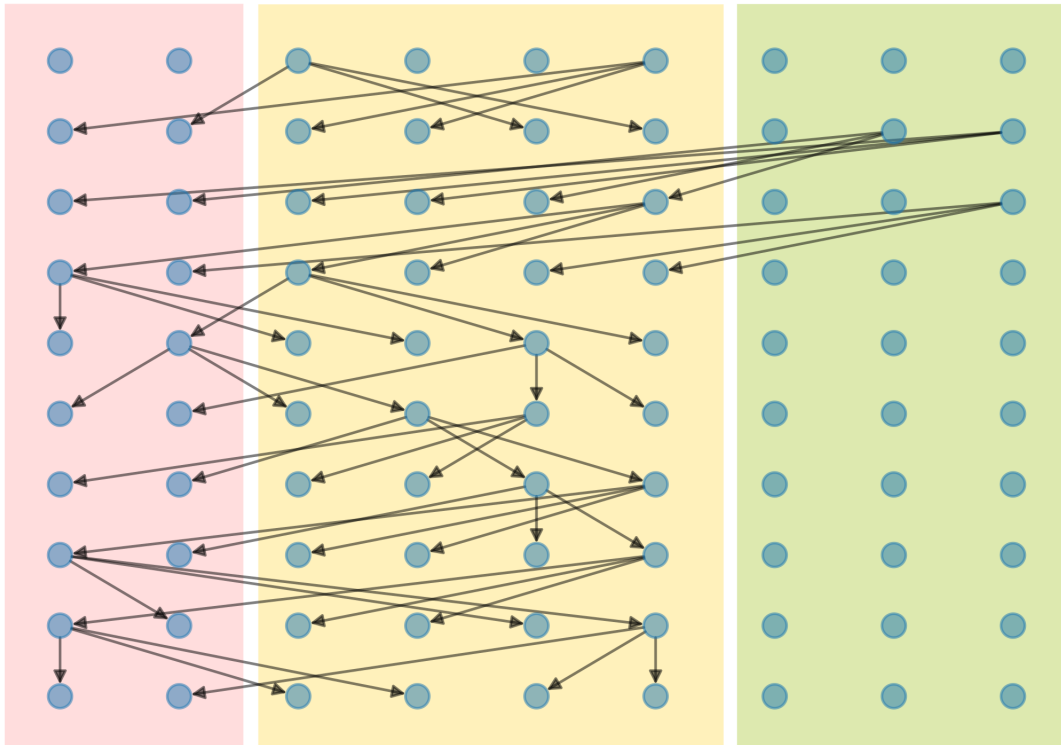
Figure 9: Comparaison des paramètres de centralité entre les meilleurs et les pires agents. Pour chaque graphe, en abscisse le temps, en ordonnée le paramètre "degree", "betweenness", "closeness" et "orthogonality", respectivement ...

Convergence de la série de graphes par l'évolution ?

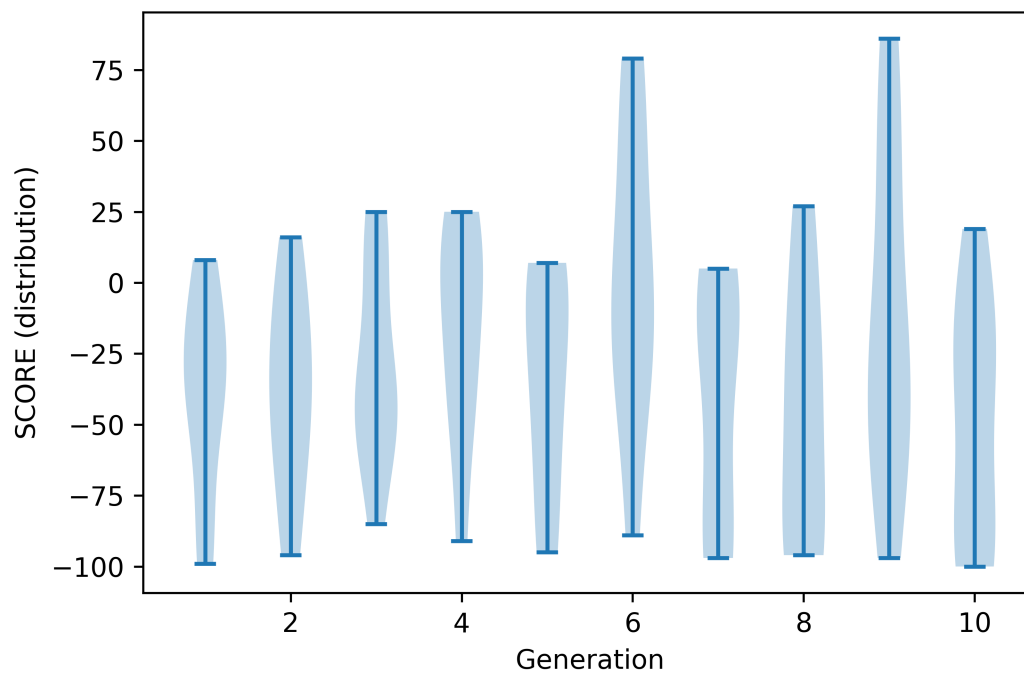
On n'observe pas de stabilité, mais les graphes les plus asymétriques marchent le mieux à court terme, mais pas à long terme (l'entraînement q-learning marche moins bien ?) → oui enfin, c'est par construction, et puis je ne sais pas comment calculer l'écart type entre deux graphes ??

## 6 Efficacité selective

Conclure un peu les deux données précédentes sur la convergence des I/O et de la structure du réseau.



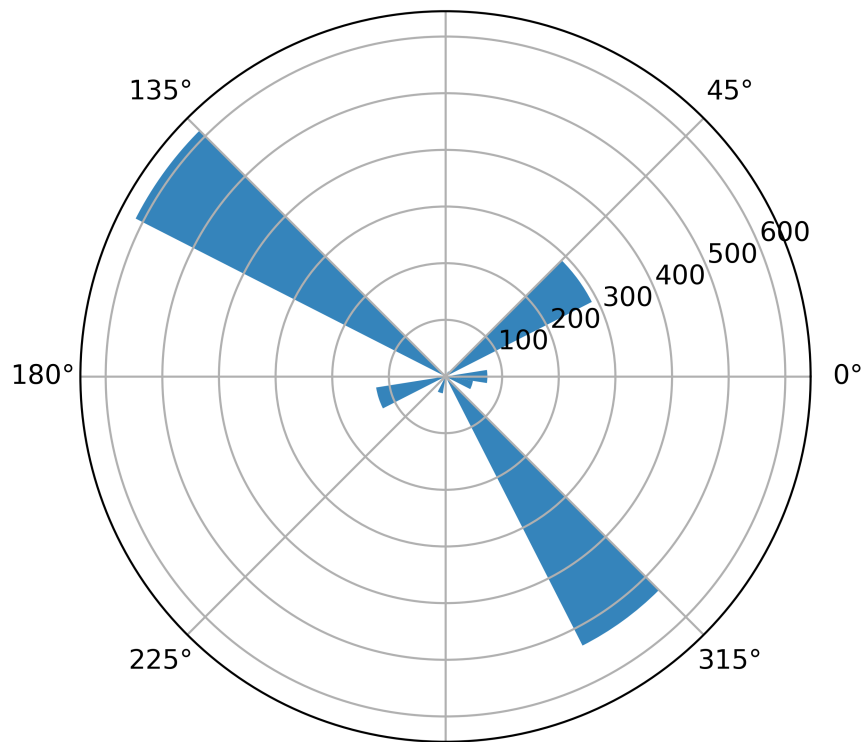
Idée ici est d'approfondir la vitesse d'apprentissage, un ce qu'un reseau tres adapté à un probleme apprend vite au debut, mais atteint un plafond moins optimale qu'un reseau moins adapté au debut, mais à la suite devient plus performant ?



Temps pour que le reseau soit adapté, heritabilité, comparaison entre aléatoire totale et avec choix de densité

I/O.

Mersure de la dispersion des agent :



On observe quel type de trajectoire domine suivant condition limite périodique ou non.

## Discussion

Paragraphe critique 1 : case carré ! pourquoi pas un cercle ? car carré problème de symétrie

Paragraphe critique 2 : réseau vestigiale n'est peut être pas optimal et empêche d'aller vers une vallée plus optimum (voir pb en bio-evo → illustration vallée d'optimalité)

Paragraphe critique 3 : Artefact de trajectoire, le réseau est avantagé si trajectoire uniquement en diagonal (+ CLP)

Paragraphe ouverture 1 : Optimisation de la structure du réseau en même temps que l'apprentissage. (citer les autres méthodes de "self-learning", NEAT, GNN, etc, mais ici dans le plus élémentaire)

Paragraphe ouverture 2 : 1 seul environnement partagé (coop vs compét)

Paragraphe ouverture 3 : Utilisation hors contexte sur base MNIST : comparaison efficacité réseau simple et réseau structuré.

IMAGE BATCH\_SIZE comparaison et réseau classique (temps d'adaptation par réseau) (en fonction des cycles d'entraînement) → (comparé entre cycle d'hérédité et cycle d'entraînement)

## References

- [1] Stuart Bartlett and Michael L. Wong. Defining life in the universe: From three privileged functions to four pillars. 10(4):42. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.
- [2] John Tyler Bonner. The origins of multicellularity. 1(1):27–36.
- [3] Roger Buick. When did oxygenic photosynthesis evolve? 363(1504):2731–2743. Publisher: Royal Society.

- [4] S. Blair Hedges, Jaime E. Blair, Maria L. Venturi, and Jason L. Shoe. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. 4(1):2.
- [5] J. Huxley. Evolution. the modern synthesis. Publisher: London: George Alien & Unwin Ltd.
- [6] Andrew Knoll, Malcolm Walter, Guy Narbonne, and Nicholas Christie Blick. The ediacaran period: a new addition to the geologic time scale. 39(1):13–30. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1080/00241160500409223>.
- [7] S. J. Mojzsis, G. Arrhenius, K. D. McKeegan, T. M. Harrison, A. P. Nutman, and C. R. L. Friend. Evidence for life on earth before 3,800 million years ago. 384(6604):55–59. Number: 6604 Publisher: Nature Publishing Group.