[PRESENTATION]

Projet Androide, étudiants, encadrant

[INTRODUCTION]

L'apprentissage par renforcement consiste en une recherche heuristique, dans un environnement donné, d'une stratégie permettant de maximiser une récompense.

Cette recherche étant propre à l'environnement, et par construction s'effectuant dans un espace en grande dimension, on s'intéresse au développement d'outils permettant de mieux visualiser ce processus.

 # APPARITION PAYSAGE DE VALEUR

Le modèle se déplace dans son espace d'apprentissage. En modifiant les paramètres de son réseau de neurones, il cherche a augmenter sa récompense.

 # APPARITION TRAJECTOIRE

On peut enregistrer ses paramètres pas à pas, donnant alors une trace de sa trajectoire.

 # APPARITION TEXTE COTE

Cet espace d'apprentissage étant en grande dimension, il n'est pas possible de se le représenter intuitivement.

C'est pourquoi lors de ce projet, nous nous sommes intéressé au développement d'outils permettant de visualiser la trajectoire du modèle, ainsi que ses alentours.

 # APPARITION PLAN

Dans un premier temps, nous décrirons deux méthodes de visualisations de l'espace puis la structure de notre implémentation.
Ensuite, nous procéderons à un exemple d'exploitation de ces outils.
Enfin, nous développerons des idées de travaux futurs, pour aller plus loin dans le projet.

[PRESENTATION METHODES]

[LIGNE]

 # TEXTE
Nous présentons dans cette partie le principe de méthodes de visualisation de cet espace.

Ces méthodes consistent à visualiser un aperçu de l'espace d'apprentissage, de grande dimension (dimension N), en dimension 2 ou 3.

La première : méthode de l'étude de gradient, permet d'observer un aperçu de la trajectoire d'un modèle ; la seconde : Vignette, permet une visualisation partielle de ses alentours.

Ces méthodes reposent sur un échantillonnage de droites, agencées sous forme de lignes pour former des images.

 # LIGNE
Ces dernières sont crées comme suit :

 # DEUX POINTS
Soient deux points dans l'espace d'apprentissage, on trace la droite les reliant.

 # DROITE
On discrétise ensuite ces droites à une fréquence entrée par l'utilisateur. Plus cette fréquence est grande, plus la visualisation du paysage de valeur est précise.

 # APPARITION IMG
On évalue ensuite chacun de ces échantillons. Leur récompense est alors quantifiée par un code couleur.

[GRADIENT]

 # TEXT
La méthode d'étude de gradient permet de suivre la trajectoire d'un modèle lors de l'apprentissage.

Il permet, grâce à l'échantillonnage de droites présenté précédemment, de connaître les structures qu'il traverse, et de comparer les directions qui ont été prises entre chaque pas.

 # POINT
Soit un point dans l'espace d'apprentissage.

Alors qu'il se déplace, l'utilisateur enregistre sa position à intervalle régulier.

 # TRANSITION SUR LE COTE
L'étude de gradient consiste à, pas à pas, échantillonner une droite entre chacun de ses points.

Pas à pas, on visualise le paysage traversé par le modèle.

 # FIN DE DIRECTIONS
En réalité, on tire une droite plus grande que le simple segment entre deux politiques. On indique alors en rouge et en vert la position relative du modèle entre chaque pas.

 # ANGLE
On peut aussi indiquer une mesure de l'angle pris entre chaque pas en calculant le produit scalaire normalisé entre chaque direction.

 # SWIMMER
Voici un exemple, d'étude de gradient sur 40 étapes pour l'environnement Swimmer, entraîné sous l'algorithme SAC sur 10.000 pas.

On constate que le modèle évolue dans une zone de récompense plutôt uniforme, à vitesse constante.

On remarque que cet outil peut nous renseigner efficacement, sous forme d'image, sur le processus de descente de gradient. Grâce à lui, on connaît les changements d'angle, les tâtonnements du modèle lors de son apprentissage.

[VIGNETTE]

 # TEXT

La méthode précédente est utile pour synthétiser une trajectoire, cependant il peut être pertinent d'obtenir une vision plus globale de l'espace d'apprentissage.

L'outil Vignette donne un aperçu des alentours du modèle.

Il permet notamment de détecter des structures dans les environs proches d'un modèle. De plus, grâce à Vignette, on peut situer un ensemble d'autres points dans l'environnement observé.

POINT

Soit un point dans l'espace d'apprentissage, représentant la politique d'un modèle.

On souhaite le situer dans celui-ci par rapport à un ensemble d'autres points.

Vignette consiste à, dans la boule unité centrée en ce point, *#DROITES*, tirer aléatoirement un ensemble de droites.

Pour pouvoir inclure les autres points dans la Vignette, la prochaine étape consiste à ajuster sa portée.

On remarque que cela à pour conséquence de réduire la fréquence d'échantillonnage des droites tirées.

Tous les points sont alors situés dans la boule correspondant à Vignette. Il suffit alors de faire passer les droites, correspondant à la sortie, en ces points. Par chaque point, on fait passer la direction tirée aléatoirement qui en était la plus proche.

L'exemple suivant correspond à la Vignette de l'environnement Pendulum sous l'algorithme SAC à 5.000 pas. Nous verrons plus tard une version 3D de cette sortie.

TEXT
On rappelle que l'espace d'apprentissage est un espace en dimension N, dimension du réseau de neurones.

On remarque sur l'exemple que cette méthode, bien que proposant une visualisation partielle de l'espace, permet de faire apparaître les structures environnantes au modèle.

Nous reviendrons sur cet exemple dans une prochaine partie.

[STRUCTURE]

 # TITRE
Nous développons à présent la structure de l'implémentation de nos outils.

Lors de notre projet, nous avons fait le choix d'offrir à l'utilisateur toutes les fonctionnalités nécessaires pour effectuer une analyse complète du paysage de valeurs d'un environnement et d'un algorithme.

Ainsi, nous proposons un ensemble de fonctionnalités permettant d'aller de l'entraînement d'un modèle jusqu'à l'affichage des résultats.

Nous permettons à l’utilisateur, de rentrer ses propres politiques, à condition d’utiliser les bons formats.

Le programme est facilement modulable. L’utilisateur peut implémenter ses propres fonctionnalités, ou développer ses propres outils.

SB3
Pour cela, nous avons complètement réécris le code des années précédentes.

Nous avons procédé à son portage sous la librairie stable-baselines-3, librairie aux diverses implémentations d’apprentissage par renforcement, et au code documenté.

Ainsi, l’utilisateur peut choisir son environnement et l’algorithme d’apprentissage qu’il souhaite utiliser.

PROCESSUS
L’utilisation des outils s’effectue en 3 phases :
Une phase de préparation, durant laquelle l’utilisateur prépare les entrées ;
Une phase de calcul des outils ;
Une phase de sauvegarde des résultats.

PREPARATION
Lors de la phase de préparation, l’utilisateur prépare les entrées des outils. Grâce aux fichiers disponibles dans le projet, il peut entraîner son modèle ou mettre au bon format ses propres politiques.

Ensuite, ses entrées sont traitées par les outils, étude de gradient ou Vignette. Cette étape peut durer plusieurs heures.

Enfin, les données sont sauvegardées dans des objets SavedVignette ou SavedGradient.

[COULEUR]

 # TEXT
Ces objets peuvent être chargés à posteriori pour traitement.

Grâce à cette implémentation, l’utilisateur peut rajouter ses propres méthodes, ou personnaliser celles existantes.

Cependant les données sauvegardées sont lourdes en mémoire. Elles peuvent atteindre la centaine de mégas. Au vu de la durée d’exécution de la phase de calcul, et des nombreux essaies que pourrait être amené à faire l’utilisateur pour ses méthodes, nous avons choisi de privilégier un format lent en compression mais rapide en décompression.

 # GRADIENT
L’utilisateur peut par exemple changer le code couleur utilisé pour lors de l’affichage.

Nous avons constaté que cela était une fonctionnalité utile, car en fonction de l’écran utilisé, ou dans le cas de personnes malvoyantes, la lisibilité influait sur l’analyse des résultats.

Comme nous l'avons vu précédemment, dans l'environnement Swimmer sous l'algorithme SAC de 250 à 10.000 pas, le modèle avance en tâtonnant dans une zone homogène. Grâce aux différentes palettes de couleur, on constate de plus que celui-ci avance dans une mauvaise direction car il réduit sa récompense.

 # VIGNETTE
De même ici, on aperçoit bien les structures autour du modèle. On distingue clairement les zones à fortes ou à basses récompense.

Au vu de la différence d'appréciation, nous avons décidé d'implémenter une visualisation interactive, en 3D des Vignettes. Nous montrerons son utilisation dans la partie suivante.

[DEMONSTRATION]

 # TEXT
Grâce à la phase de sauvegarde, il est facile d'implémenter de nouvelles fonctionnalités.

Comme évoqué précédemment, nous avons développé une version interactive de l'outil Vignette, en trois dimensions. Cette visualisation permet une meilleur appréhension des structures de l'environnement.

[DEMO DIRECTE 3D]

 # ROTATION
L'utilisateur peut se déplacer dans cette visualisation.

VUE DE DESSUS
On remarque que en vue de dessus, on retrouve bien la Vignette 2D.

SLIDERS
De plus, l'utilisateur peut définir des curseurs, pour pouvoir transformer le paysage en temps réel.

Il peut par exemple régler l'opacité des surfaces pour laisser apparaître les politiques d'entrée. Immédiatement, on peut observer le processus de montée de gradient, de 500 pas en 500 pas.

Grâce aux droites tirées dans la boule autour du modèle n°5000, on constate que celui-ci est monté lors de son apprentissage vers des zones à haute récompense.

[FAISCEAUX]

 # TITRE
Nous présentons maintenant quelques idées de développement futurs à nos travaux.

 # VITESSE
Tout d'abord, concernant la vitesse de calcul des outils.

Dans le cas de politiques stochastiques, les modèles sont soumis au problème de généralisation. Pour les évaluer, il convient alors de calculer leur performance moyenne sur un certain nombre d'échantillons.

Plus le nombre d’évaluations nécessaire est grand, plus les calculs des outils sont longs. L’utilisateur est alors contraint à un compromis entre précision et puissance de calcul disponible.

Pour pallier à ce problème, quelques solutions sont envisageables :

L’échantillonnage des différentes lignes étant indépendants entre eux, il est possible de les effectuer sur plusieurs coeurs.

De plus, l’utilisateur devrait pouvoir disposer de fonctionnalité de pause de calcul. Il devrait pouvoir sauvegarder des points de contrôle, pour reprendre des exécutions plus tard.

Nous pensons qu’il pourrait lui être utile de pouvoir effectuer les calculs en plusieurs passes, à différents niveaux de précision. Ainsi, les calculs seraient effectués à un niveau de précision croissant, permettant à l’utilisateur d’observer des résultats de plus en plus précis. Il pourrait alors observer des résultats préliminaires.

 # FAISCEAUX

Nous présentons maintenant notre idée de fonctionnement pour un autre outil, la méthode des faisceaux.

POINT

L'outil Vignette permet à l'utilisateur d'obtenir un aperçu global de l'espace d'apprentissage autour d'une politique. Par ailleurs, l'outil d' étude de gradient permet d'observer un aperçu du chemin pris par celle-ci lors de la descente de gradient.

Cependant, plus les sauvegardes du processus d'apprentissage sont espacées dans le temps, moins les visualisations proposées sont pertinentes. En effet, les outils ne donnent qu'un aperçu partiel de l'environnement : plus le nombre de pas entre deux politiques est grand, moins ceux-ci nous informent sur l'environnement dans lequel la descente de gradient a été effectuée.

ROTATION

On constate par exemple ici, que du fait de la faible fréquence de sauvegarde, ni la descente de gradient, ni Vignette *# VIGNETTE* ne permettent de calculer efficacement le paysage de valeur autour du modèle.

Pour combler ce problème, on imagine le fonctionnement d’un outil s’en inspirant. On cherche à comparer la direction globale prise par le modèle avec le paysage de valeur autour de celle-ci.

On trace le segment entre deux points, puis on y calcule les deux hyperplans orthogonaux au segment.

On les échantillonne ensuite, pour obtenir un ensemble de politiques similaires aux deux points.

Selon la méthode d’échantillonnage des droites, on peut alors échantillonner un faisceaux entre les deux hyperplans.

 # TRANSITION

Cette méthode revient à concentrer Vignette selon une direction, pour obtenir une visualisation plus pertinente de l’environnement rencontré par le modèle.

De plus, du fait de la concentration des droites, on aura moins de discontinuités au niveau de la détection de structures.

[CONCLUSION?]