le cnam

Sujet UE RCP211
Intelligence Artificielle Avancée

Année universitaire 2020–2021

Examen 1ère session : juin 2021

Responsable : Nicolas THOME

Durée : 2h

Seuls documents autorisés : 2 feuilles A4 recto-verso, manuscrites.

Sujet de 7 pages, celle-ci comprise.

Vérifiez que vous disposez bien de la totalité des pages du sujet en début d'épreuve et signalez tout problème de reprographie le cas échéant.

---

1. Robustesse décisionnelle.
   (a) Incertitude décisionnelle (**1 points)**
      — Un système de détection d'objets a été entraîné pendant l'été sur un site montagnard. Si on déploie le système pendant la période hivernale, à quelle type d'incertitude le système est-il sujet ? Expliquer.

      Correction :

      — Incertitude épistémique, la distribution des données a changé, les exemples de test sont dans une zone de l'espace différente des données d'apprentissage.

   (b) Modèles Bayésiens (**3points)**
      — Quelle est la différence entre un modèle (*e.g.*réseau de neurones) déterministe et un modèle Bayésien ?
      — Comment un modèle Bayésien permet-il de quantifier l'incertitude décisionnelle ? quelles sont les deux grandeurs clés à estimer ?

      Correction : 1.5 + 1.5

      — Estimation de la distribution prédictive vs estimation ponctuelle (pointwise)
      — La forme de la distribution prédictive donne une information sur le niveau d'incertitude décisionnelle : distribution prédictive plate $\Rightarrow$ incertitude forte, distribution prédictive piquée $\Rightarrow$ incertitude faible.
         — Étape 1 : estimation de la distribution postérieure $p(\mathbf{w}|\mathcal{D})$
         — Étape 2 : estimation de la distribution prédictive $p(\mathbf{y}|\mathbf{x}^*, D) = \int p(\mathbf{y}|\mathbf{x}^*, \mathbf{w}) p(\mathbf{w}|\mathcal{D}) d\mathbf{w}$

   (c) Réseaux de neurones Bayésiens (**2points)**
      — Quelle est la différence entre l'approximation de Laplace et l'approximation variationnelle de la distribution postérieure ?
      — Quelle est la distribution postérieure des paramètres associée à la méthode MC-dropout ?

      Correction :

      — Approximation locale du postérieure en w_{MAP} avec l'approximation de Laplace vs approximation globale de la distribution en minimisant la KL entre la distribution empirique et la distribution variationnelle
      — Bernouilli : $q(\mathbf{W}_l) = \text{diag}(\hat{\boldsymbol{\varepsilon}}_l)\mathbf{M}_l, \varepsilon_{l,i} \sim Bernoulli(1 - p_i)$

(d) Stabilité décisionnelle (**1 point)** : qu'est-ce qu'un exemple adversaire pour un réseau de neurones ?

Correction :

— Fonction de décision non stable par rapport à des petites variations de l'entrée

2. Modèles génératifs

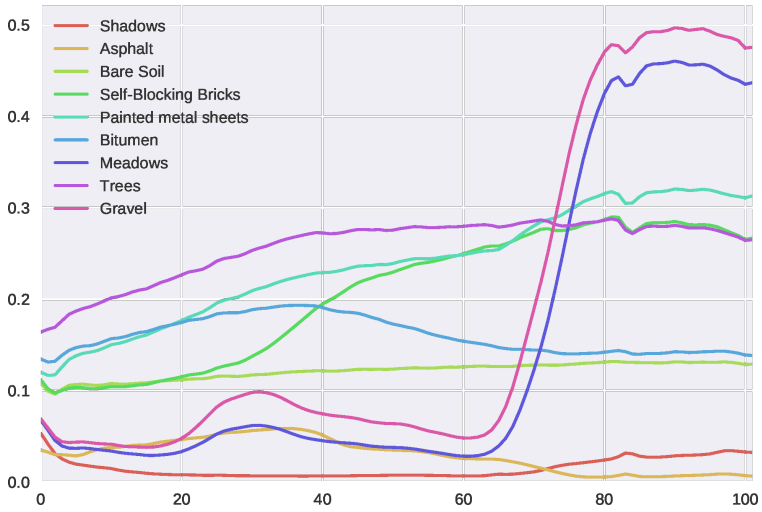

FIGURE 1 – Spectres lumineux de différentes matériaux (intensité réfléchie en fonction de la longueur d'onde). Chaque spectre un vecteur unidimensionnel de longueur ≈ 100.

(a) Auto-encodeurs variationnels
On souhaite générer des visages synthétiques en apprenant un Variational Auto-Encoder (VAE) sur le jeu de données CelebA. En plus des images, ce jeu de données présente pour chaque éléments des descripteurs indiquant la composition du visage (ex. : lunettes, chapeaux, couleur de peau, pilosité faciale...).

— Un VAE classique permet-il de générer des visages avec des caractéristiques particulières ? Justier **1 point**

— Proposer une approche permettant de traîter ce problème. **1 point**

Correction :

— Un VAE classique ne permet pas de faire de choix dans l'espace lattent (à moins de procéder par gridd search). En effet, l'espace lattent est structuré pour que sa distribution soit proche d'une gaussienne centrée réduite permettant ainsi de l'échantillonner mais l'information y est mélangée.

— L'approche la plus directe pour permettre d'apprendre la dépendance d'un espace lattent en fonction de caractéristiques sur les données est d'utiliser un VAE conditionnel. Le modèle apprend alors à encoder les données sachant d'autres données (labels, imaes, texte...). D'autres approches permettent également de structurer l'espace lattent pour favoriser la répartitions de ces caractéristiques sur des sous-espaces disjoints : β-VAE, *Normalizing flow*...

(b) Réseaux génératifs antagonistes
On dispose d'un jeu de données de spectres lumineux (intensité réfléchie en fonction de la longueur d'onde). Les spectres sont annotés et groupés dans 9 catégories différentes (quelques exemples sont illustrés dans la Fig. 1). On souhaite entraîner des réseaux génératifs antagonistes (GAN) afin de générer de nouvelles données synthétiques. On ne cherche pas à conditionner le modèle au type de matériau.

— Proposer une métrique d'évaluation afin de mesurer les performances du générateur. **1 point**

— Comment se manifesterait le *mode collapse* dans ce cas ? Quelles sont les façons de l'éviter en pratique ? **2 points**

— On choisit finalement un Wasserstein-GAN plutôt que la formulation du GAN habituel. En quoi la minimisation de la distance de Wasserstein est-elle plus intéressante que celle de la divergence de Kullback-Leibler ? **1 point**

Correction :

i. Les métriques FID et Inception Score pourraient être envisagées si l'on disposait d'un classifeur profond (par exemple, un CNN 1D). Une alternative est la suivante : apprendre un classifieur (profond ou non) sur les exemples réels et mesurer ses performances sur les données synthétiques. Si les performances synthétiques sont nettement inférieures, alors les données générées sont a priori peu réalistes. Si les performances synthétiques sont nettement supérieurs, alors les données générées sont "stéréotypiques" et très concentrées aautour de certains points. On peut aussi examiner les statistiques des spectres générés et les comparer à celles des spectres réels.

ii. Le *mode collapse* s'observerait par la répétition de spectres très similaires, ou de spectres d'une seule classe. On peut l'éviter en passant par un W-GAN, en utilisant du *label smoothing* ou des mélanges ("MixUp").

iii. La divergence de Kullback-Leibler est nulle lorsque les supports des distributions ne s'intersectent pas. Cela rend l'apprentissage du GAN plus difficile. Ce n'est pas le cas de la distance de Wasserstein qui augmente avec la distance entre les distributions, y compris si leurs supports sont disjoints.

3. Apprentissage par renforcement

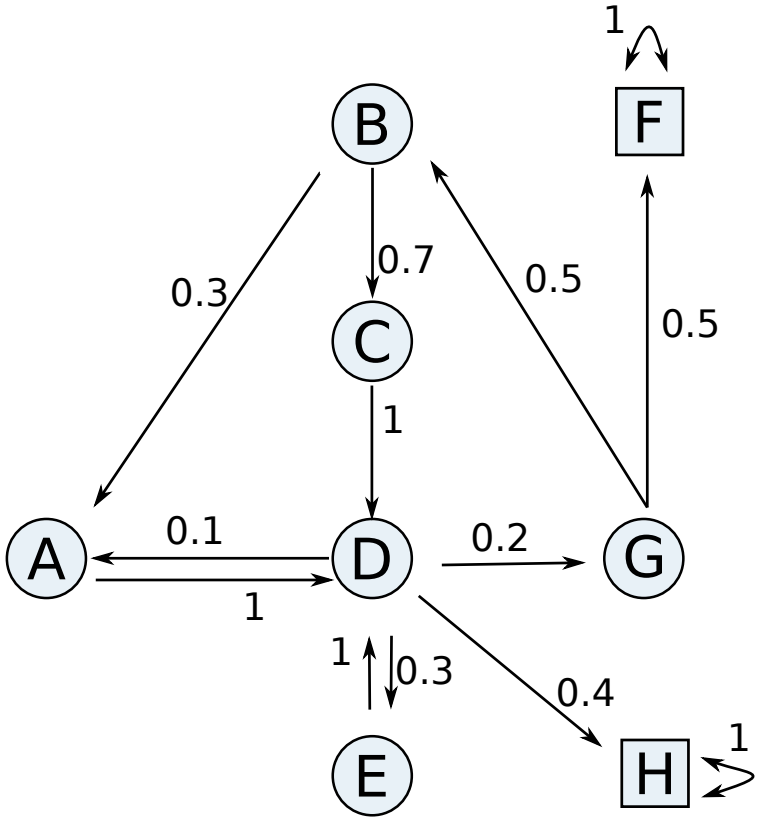

FIGURE 2 – Diagramme représentant les états accessibles par l'agent ainsi que la dynamique de l'environnement (*ie.* les probabilités de transition). Les états F et H sont terminaux étant donné qu'on ne peut pas les quitter une fois atteints.

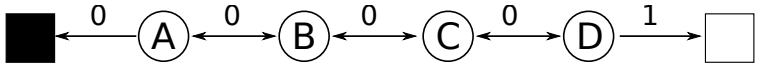

FIGURE 3 – Digramme d'une marche aléatoire entre 4 états de transitions A, B, C et D et deux états terminaux ■ et □. La probabilité p d'aller à gauche ou à droite est de $p = 0.5$. La récompense associée à chaque transition est de 0 et de 1 pour $D \rightarrow \square$.

(a) Markov Decision Process **3 points**

— Le schéma Fig. 2 présente un processus Markovien avec les états que peut ateindre un agent. Donnez la matrice de transition associée à cette modélisation.

— On suppose à présent que la récompense associée à chaque transition est de -1 en général et de 0 lorsque un état terminal est atteint *ie.* pour les transitions $G \rightarrow F$ et $D \rightarrow H$. Quelle serait la politique optimale de l'agent dans

ces condition ? *NB :* ici, l'environnement est déterministe. On cherche donc pour chaque état la probabilité des transitions possibles (on ne change pas les flèches juste leur "valeur").

— Le schéma Fig. 3 présente le diagramme associé à une marche aléatoire entre états : $A, B, C, D, \square$ et $\blacksquare$. L'agent peut aller à gauche ou à droite avec une probalité uniforme de $p = 0.5$ pour ces deux actions. Les deux états $\square$ et $\blacksquare$ sont des états terminaux. Toutes les récompenses de transition sont nulles exceptée celle de D à $\square$ valant 1. Donner la valeur de chacun des états en supposant que l'agent démarre en C.

Correction :

— Matrice de transistion :
NB : Une erreur s'était glissée dans le diagramme du partiel avec un lien en trop de G vers E.

$$\begin{array}{c} \begin{matrix} A & B & C & D & E & F & G & H \end{matrix} \\ \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0.7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0.1 & 0 & 0 & 0 & 0.3 & 0 & 0.2 & 0.4 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \end{matrix}$$

— Le plus simple est de procédé par itération sur la valeur pour trouver la politique optimale *ie.* on part de la fin du problème :

$$\begin{array}{c} \begin{matrix} A & B & C & D & E & F & G & H \end{matrix} \\ \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0/0.5/1 & 0 & 1/0.5/0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \end{matrix}$$

— Pour obtenir la valeur associée aux différents états, on peut construire un système linéaire à résoudre en se basant sur la récurrence de Bellman :

$$v(s') = \mathbb{E}\left[r + \gamma v(s)\right].$$

En se rappelant que dans le cas discret l'espérance de $x = \{x_1, ..., x_N\}$ est donnée par $\mathbb{E}[x] = \sum_{i=1}^{N} p_i x_i$, on obtient alors :

$$\begin{cases} v(A) = 0.5(0 + v(\blacksquare)) + 0.5(0 + v(B)) = 0.5[v(\blacksquare) + v(B)] \\ v(B) = 0.5(0 + v(A)) + 0.5(0 + v(C)) = 0.5[v(A) + v(C)] \\ v(C) = 0.5(0 + v(B)) + 0.5(0 + v(D)) = 0.5[v(B) + v(D)] \\ v(D) = 0.5(0 + v(C)) + 0.5(1 + v(\square)) = 0.5[v(C) + 1 + v(\square)] \end{cases}$$

Comme depuis les états terminaux, on n'accumule plus de récompense $v(\blacksquare) = v(\square) = 0$ et on se retrouve avec un système de 4 équations à 4 inconnues. Au final, on a : $v(A) = \frac{1}{5}, v(B) = \frac{2}{5}, v(C) = \frac{3}{5}$ et $v(D) = \frac{4}{5}$.

(b) Contrôle basé sur la valeur **2 points**

— Quelles sont les principales différences entre estimation par échantillonnage de Monte-Carlo et par Différence Temporelle (TD) ?

— Dans le cas où la fonction de valeur d'état-action est approchée, quelle est la fonction de coût (*loss*) à minimiser dans une approche TD ? Proposez au moins deux méchanismes généralement introduits pour améliorer la stabilité d'approches basées sur des Deep Q-Networks.

Correction :

— L'échantillonnage de MC consiste à lancer des trajectories, compter le nombre de fois qu'un état est visité et calculer le retour moyen pour chaque état. TD consiste à corriger l'erreur entre la prédiction à l'instant $t+1$ et l'instant t. Le grandes différences sont :

  — MC est non biaisé / TD est biaisé

  — MC a une variance bien plus grande que TD

  — TD converge plus vite que MC (conséquence du point précédent)

  — MC est *offline* / TD est *online*

— Dans le cas où la fonction de valeur est approchée, on cherche à minimiser la différence entre la valeur estimée à l'instant t et celle à l'instant $t+1$. En général, la fonction de coût choisie est la RMSE.

  — Une Huber Loss permet en pratique une meilleure convergence

  — Pour éviter d'apprendre sur des échantillons corrélés, le modèles est généralement entraîné sur un ensemble de séquences enregistrées périodiquement (*experience replay*).

— L'apprentissage du modèle est généralement plus stable avec l'utilisation d'un réseau auxiliaire chargé de calculer la fonction de valeur à l'instant $t+1$ (*target network*) et mis à jour moins régulièrement que le réseau entraîné.

— Enfin, l'utilisation d'un *Double DQN* améliore généralement la stabilité de l'apprentissage tout en réduisant le biais de maximisation.

(c) Optimisation de la politique **2 point**

— Dans le schéma Acteur-Critique que réprésentent l'acteur et la critique ? Comment sont-il estimés ?

— Quels moyens a-t-on pour réduire la variance du gradient de la fonction objectif par rapport à la politique ?

Correction :

— L'acteur correspond à la maximisation de la fonction objectif par montée de gradient sur la politique. La critique correspond à l'estimation de la fonction de valeur.

— Les moyens possibles pour réduire la variance du gradient sont :

  — Structuration temporelle *ie.* hyppothèse de causalité

  — Utilisation d'une baseline *eg.* la fonction de valeur d'état

  — *Boostraping ie.* estimation de la critique par une approche TD