le cnam

Sujet UE RCP211
Intelligence Artificielle Avancée

Année universitaire 2020–2021

Examen 2ème session : septembre 2021

Responsable : Nicolas THOME

Durée : 2h

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

Sujet de 4 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 **(7 points)**
   (a) Un Husky est une race de chien qui présente un certain nombre de caractéristiques communes avec les loups. À quel type d'incertitude décisionelle un modèle ayant été entraîné à classer des images de chiens et de loups doit-il faire face ? **(1 point)**
   (b) En quoi consiste l'approximation variationnelle de la distribution postérieure ? Donner la formulation de la fonction de coût ELBO et interpréter les deux termes résultants **(2 points)**
   (c) Quel est l'utilité d'une mesure d'incertitude décisionnelle pour une tâche d'apprentissage actif ? **(1 point)**
   (d) En quoi consiste la calibration des probabilités ? En quoi cette propriété est-elle importante pour la robustesse d'un système décisionnel ? Est-elle garantie avec les réseaux de neurones modernes ? **(3 points)**

Correction :

   (a) Aleatoric, confusion de classe
   (b) On minimise la KL entre une distribution postérieure approximée et la postérieure réelle. La fonction de coût ELBO fait apparaître un terme de bonne classification et un terme de divergence KL entre le distribution approximée et une distribution a priori (à fixer)
   (c) Déterminer les exemples les plus incertains comme ceux nécessitant une annotation manuelle
   (d) Calibration : l'incertitude décisionnelle correspond au pourcentage d'erreur du modèle. C'est important pour fixer un seuil de décision permettant de garantir une erreur bornée. Les réseaux de neurones profonds modernes ne sont pas calibrés.

2. Modèles génératifs **(6 points)**
   (a) Comment peut-on passer d'un modèle génératif à un modèle discriminatif ? **(1 point)**
   (b) Donner deux exemples d'espaces latents. **(1 point)**
   (c) En quoi peut-on dire que la fonction de coût d'un GAN est une fonction de coût *perceptuelle* ? Justifier brièvement. **(2 points)**
   (d) Qu'appelle-t-on le *reparametrization trick* dans les VAE ? Quelle est son utilité ? **(2 points)**

Correction :

   (a) Théorème de Bayes, cf. cours.

(b) Représentation intermédiaire d'un auto-encodeur, d'un VAE, espace $\mathcal{Z}$ d'échantillonnage d'un GAN, espace résultant d'une projection par t-SNE…

(c) Le générateur tente de tromper le discriminateur, c'est-à-dire de produire des données qui sont *perçues* par le discriminateur comme appartenant à la même distribution que les données réelles. Cette fonction de coût n'est pas explicitement définie et compare des différences de plus haut niveau sémantiques qu'une fonction de coût au niveau pixel, comme la MSE.

(d) Le *reparametrization trick* consiste à échantillonner ε plutôt que z en écrivant : $z = g(\phi, x, \varepsilon)$ plutôt que $z\ q(z|\phi, x)$. Cela permet de rétropropager à travers z, puisqu'il n'est sinon pas possible de rétropropager à travers l'opération d'échantillonnage.

3. Apprentissage par renforcement **(7 points)**

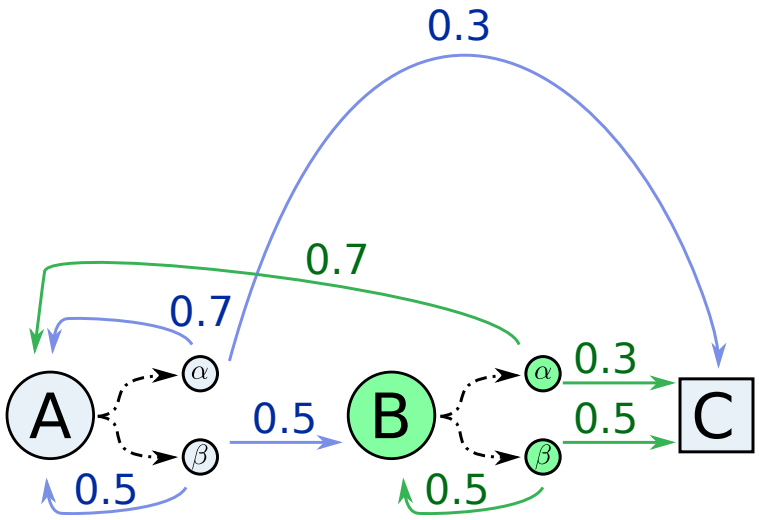

FIGURE 1 – Diagramme représentant les états accessibles par l'agent ainsi que ses actions et la dynamique de l'environnement. L'état C est terminal et deux actions sont possibles depuis A et B : α ou β. On voit par exemple que depuis l'état A, en sélectionnant l'action α, on a une probabilité de 0.3 d'atteindre directement C et 0.7 de revenir en A.

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

— Donner une méthode pour résoudre exactement un MDP de petite taille tel que présenté dans le diagramme 1. **(0,5 point)**

— On choisit une politique aléatoire entre les action α et β. Quelle est alors la matrice de transition conditionnée par cette politique ? **(1,5 point)**

— Donner la valeur des différents états. On supposera ici que la récompense associée à chaque transition est de -1 et que le facteur de réduction $\gamma = 1$. Avec cette politique, quel état est le pus avantageux entre A et B ? **(1 point)**

(b) Optimisation de la valeur **(2 point)**

— Justifier que l'échantillonnage de Monte-Carlo minimise l'erreur moyenne quadratique (*Mean Square Error*) dans le calcul de la valeur. **(1 point)**

— Cet estimateur est-il stable ? Donner le rapport entre l'erreur moyenne quadratique et la variance du retour. **(1 point)**

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

— Justifier que la dynamique du modèle (*ie.* les probabilités de transition) n'intervient pas dans le calcul du gradient de la valeur. **(1 point)**

— Un schéma basé *policy gradient* est-il garanti de converger ? Si oui vers une solution optimale ? **(1 point)**

Correction :

(a) Markov Decision Process (MDP)

— Les valeurs des états ainsi que la politique optimale peuvent être trouvé par programmation dynamique en suivant par exemple un schéma glouton d'amélioration de la politique ou d'itération sur la valeur.

— On veut trouver $p(S_{t+1} = s_j | S_t = s_i)$ pour $(i, j) \in \{A, B, C\}^2\}$. Pour cela nous devons calculer les marginales de $p(S_{t+1} = s_j, A_t = a | S_t = s_i)$, *ie.* intégrer par rapport à la politique :

$$\begin{aligned} p(S_{t+1} = s_j | S_t = s_i) &= \sum_{a \in \mathcal{A}} p(S_{t+1} = s_j, A_t = a | S_t = s_i) \\ &= \sum_{a \in \mathcal{A}} p(A_t = a | S_t = s_i) p(S_{t+1} = s_j | S_t = s_i) \text{ (indépendance)} \end{aligned}$$

Ces valeurs se calculent facilement étant donné que $\pi(\alpha) = \pi(\beta) = 0,5$. On a alors la matrice de transition suivante :

$$\begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0.6 & 0.25 & 0.15 \\ 0.35 & 0.25 & 0.4 \\ 0 & 0 & 1 \end{pmatrix} \begin{array}{c} A \\ B \\ C \end{array} \end{array}$$

— Pour trouver la fonction de valeur du MDP associée à une politique aléatoire, il faut résoudre le système :

$$\begin{cases} v(A) = R + \pi(0.7v(A) + 0.3v(C)) + \pi(0.5v(B) + 0.5v(A)) \\ v(B) = R + \pi(0.7v(A) + 0.3v(C)) + \pi(0.5v(B) + 0.5v(C)) \\ v(C) = 0 \end{cases}$$

avec $\pi = 0.5$ et $R = -1$. On trouve $v(A) = -4,7$ et $v(B) = -3,5$. L'état B est donc plus avantageux.

(b) Optimisation de la valeur.

— Par définition l'erreur quadratique moyenne est donnée par :

$$MSE = \mathbb{E}[(v(s) - \hat{v}(s))^2]$$

À partir de N trajectoires, cette quantité est minimisée par la moyenne des N retours obtenus pour s, ce qui correspond à l'estimateur de Monte-Carlo.

— Cet estimateur, bien que non biaisé, présente ne grande variance. Par définition, on a :

$$MSE = \frac{\sigma^2}{N},$$

avec σ la variance du retour associé à l'état s.

(c) Optimisation de la politique

— Les probabilités de transitions constituant la dynamique du modèle ne dépendent pas des paramètres de la politiques. Dans le calcul du gradient de l'objectif, seuls les termes faisant intervenir la politique interviennent (*cf.* vidéo 4 $\sim$ 1h).

— Une approche purement *policy gradient* est garantie de converger mais vers un minimum local.