PRÉSENTATION TRAVAIL EN COURS - APPRENTISSAGE INTERACTIF

Ianis Lallemand, 21 janvier 2013

APPRENTISSAGE INTERACTIF

DEFINITION

Contours encore assez flous dans le champ de l'apprentissage automatique.

Néanmoins, intérêt récent :
e.g. *Workshop on Analysis and Design of Algorithms for Interactive Machine Learning* (NIPS 2009)

APPRENTISSAGE INTERACTIF

DEFINITION (WORKSHOP NIPS)

"The traditional role of the human operator in machine learning problems is that of a batch labeler, whose work is done before the learning even begins. However, there is an important class of problems in which the human is interacting directly with the learning algorithm as it learns. Canonical problem scenarios which fall into this space include active learning, interactive clustering, query by selection, learning to rank, and others. Such problems are characterized by three main factors:

1. the algorithm requires input from the human during training, in the form of labels, feedback, parameter guidance, etc.

2. the user cannot express an explicit loss function to optimize, either because it is impractical to label a large training set or because they can only express implicit preferences.

3. the stopping criterion is performance that's "good enough" in the eyes of the user.

APPRENTISSAGE INTERACTIF

DEFINITION (WORKSHOP NIPS)

"The traditional role of the human operator in machine learning problems is that of a batch labeler, whose work is done before the learning even begins. However, there is an important class of problems in which the human is interacting directly with the learning algorithm as it learns. Canonical problem scenarios which fall into this space include active learning, interactive clustering, query by selection, learning to rank, and others. Such problems are characterized by three main factors:

1. the algorithm requires input from the human during training, in the form of labels, feedback, parameter guidance, etc.

2. the user cannot express an explicit loss function to optimize, either because it is impractical to label a large training set or because they can only express implicit preferences.

3. the stopping criterion is performance that's "good enough" in the eyes of the user.

APPRENTISSAGE INTERACTIF

DEFINITION (WORKSHOP NIPS)

"The traditional role of the human operator in machine learning problems is that of a batch labeler, whose work is done before the learning even begins. However, there is an important class of problems in which the human is interacting directly with the learning algorithm as it learns. Canonical problem scenarios which fall into this space include active learning, interactive clustering, query by selection, learning to rank, and others. Such problems are characterized by three main factors:

1. the algorithm requires input from the human during training, in the form of labels, feedback, parameter guidance, etc.

2. the user cannot express an explicit loss function to optimize, either because it is impractical to label a large training set or because they can only express implicit preferences.

3. the stopping criterion is performance that's "good enough" in the eyes of the user.

SPÉCIFICITÉS

Contraintes imposées par l'utilisation de techniques d'apprentissage automatique dans un contexte musical ou artistique :

1. souvent (très) peu d'exemples d'apprentissage disponibles.

2. difficulté de formuler le problème comme un problème d'optimisation : la qualité des résultats peut-être relative à l'utilisateur (installation interactive), au contexte ou au style (génération automatique de musique), etc.

3. le contexte (données d'entrées, attentes sur résultats, etc) peut varier en cours d'apprentissage : difficulté de séparer la phase d'apprentissage du fonctionnement « normal » du système.

APPRENTISSAGE PAR RENFORCEMENT

INTÉRÊTS

1. Apprentissage de la politique (« comportement ») du système simultanément à son fonctionnement « normal ».

2. Cadre théorique formulable indépendamment du modèle implémentant la politique.

INCONVÉNIENTS

1. Convergence souvent très lente (très grand nombre d'itérations nécessaires).

2. Dans la formulation standard (MDP, *Markov Decision Process*), récompenses attribuées par l'environnement : besoin d'un modèle de récompenses environnementales, intégré de manière fixe.

« FAÇONNEMENT »

DÉFINITION

W. B. Knox and P. Stone, “Interactively Shaping Agents via Human Reinforcement: the TAMER Framework”, Proceedings of the fifth international conference on Knowledge capture (K-CAP ‘09).

“we define shaping as interactively training an agent through signals of positive and negative reinforcement (...), scalar signal of approval or disapproval”

« FAÇONNEMENT »

DÉFINITION FORMELLE

- Un agent reçoit une séquence de description d'états de l'environnement (s_1, s_2, ...).
- L'agent peut choisir une action a_i parmi plusieurs actions.
- Un « entraîneur » humain observe l'agent : cet entraîneur a la compréhension d'un critère de performance (quantitatif ou subjectif).
- L'entraîneur évalue la qualité du comportement récent : renforcement positif ou négatif de paires état-action (s, a) récentes, en accord avec la « vraie » politique à apprendre.

REMARQUE

Analogue à une formulation MDP\R.

MDP sans explicitation de la fonction de récompenses environnementales.

CADRE TAMER

TAMER : *Training an Agent Manually via Evaluative Reinforcement*

ALGORITHME GÉNÉRAL

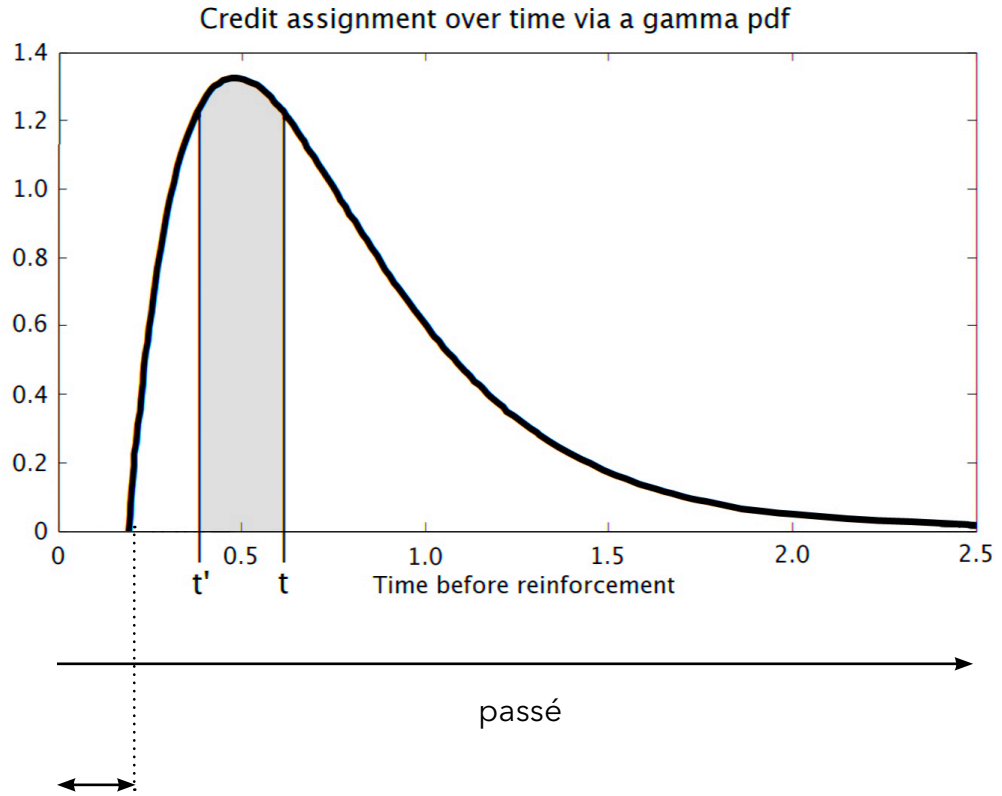
Algorithm 1 A general greedy TAMER algorithm

Require: *Input: stepSize*

ReinfModel.init(stepSize)
$\overrightarrow{s} \leftarrow \overrightarrow{0}$
$\overrightarrow{f} \leftarrow \overrightarrow{0}$
while *true* **do**
  $h \leftarrow getHumanReinfSincePreviousTimeStep()$
  if $h \neq 0$ **then**
    $error \leftarrow h$ - $ReinfModel.predictReinf(\overrightarrow{f})$
    $ReinfModel.update(\overrightarrow{f}, error)$
  end if
  $\overrightarrow{s} \leftarrow getStateVec()$
  $a \leftarrow argmax_a(ReinfModel.predict(getFeatures(\overrightarrow{s}, a)))$
  $\overrightarrow{f} \leftarrow getFeatures(\overrightarrow{s}, a)$
  $takeAction(a)$
  wait for next time step
end while

CRÉDIT

Attribution ou prédiction du renforcement sur une séquence de plusieurs paires état-action : mécanisme de fenêtrage, pondération.



TRAVAIL EN COURS

MODÈLE D'ENVIRONNEMENT	Oracle des facteurs (*Factor Oracle*) probabilisé.
QUANTITÉ À RENFORCER	Renforcement des probabilités de transition.
CRÉDIT	Via densité de probabilité Gamma et étalonnage.
PRÉDICTION DU RENFORCEMENT	Modèle linéaire . « renforcement précédent de la transition» **x** « crédit attribué à transition »
RENFORCEMENT	Calculé par descente de gradient.

IMPLÉMENTATION

ALGORITHME

- Tamer (Java)
- Oracle des Facteurs (Java)

REPRÉSENTATION

- Construction de représentations symboliques à partir de données audio par k-moyennes (Java, weka)
- Construction de représentations symboliques à partir de séquences MIDI (gestion de la polyphonie, cf. Assayag *et al.*) (Java)

INTERFACES

- Pour données audio segmentées via CataRT (Max, mxj)
- Pour séquences MIDI (Max for Live, mxj)

ÉVALUATION

- « Niveau 0 »

TRAVAIL À RÉALISER (COURT TERME)

DONNÉES MIDI

- Implémentation d'exemples musicaux réels (MIDI).
- Apprentissage interactif de la représentation musicale adéquate en fonction des attentes de l'utilisateur (pitch, rythme, etc) : plusieurs Oracles en parallèle.

DONNÉES AUDIO

- Abandon de l'approche de quantification *a priori* par k-moyennes.
- Approche « Oracle Audio » : transitions entre trames similaires vs. quantification *a priori*.
- Segmentation interactive : plusieurs oracles audio en parallèle avec des seuils de similarité différents.
- Application / évaluation possible de la mesure de similarité présentée à SMC 2012.