

UNIVERSITÉ PIERRE ET MARIE CURIE

PLAN DE DÉVELOPPEMENT

PIAD DE MASTER1 D'INFORMATIQUE EN INTELLIGENCE
ARTIFICIELLE ET DÉCISION

Semi-supervised Learning Agents

Auteurs :

Lan ZHOU

Matthieu ZIMMER

Superviseurs :

Paolo VIAPPIANI

Paul WENG

9 avril 2013

Version 1.1

Table des matières

1	Aperçu du PIAD	2
1.1	Objectifs	2
1.2	Références	2
1.3	Définitions et acronymes	2
2	Organisation du PIAD	4
2.1	Participants	4
2.2	Activités globales	4
2.3	Tâches jalonnées	4
2.4	Rôles et responsabilités	6
2.5	Calendrier	7

1 Apercu du PIAD

1.1 Objectifs

Nous devons développer une bibliothèque indépendante du domaine dans lequel l'agent évolue. Elle fournira plusieurs algorithmes de base de l'apprentissage par renforcement (Q-Learning, SARSA, ...), plusieurs critères de performance, ainsi qu'une ouverture sur l'apprentissage semi-supervisé : c'est à dire avec les retours de l'expert pris en compte.

Dans un premier temps, l'expert pourra simplement compenser la fonction de récompense en précisant si l'agent a bien ou mal agit. Dans un second temps, si le temps le permet, l'expert pourra également agir sur le choix de l'action à entreprendre lors de l'exploration de l'agent, ou encore dire à l'agent s'il est temps d'exploiter ou d'explorer.

Parallèlement au développement de la bibliothèque, afin d'avoir une application pratique de la théorie, on utilisera ces algorithmes dans le simulateur TORCS sur l'apprentissage automatique de la conduite de voiture sur circuit. Nous modifierons également l'interface TORCS pour intégrer les retours positifs ou négatifs de l'expert, voire des retours plus complexes cités précédemment Zhou and Zimmer (2013).

1.2 Références

Groupe PDMIA (2008). *Processus Décisionnels de Markov en intelligence artificielle*. Laboratoire Informatique Paris 6, Equipe MAIA - INRIA (LORIA).

Russell, S. and Norvig, P. (2006). *Intelligence Artificielle*. PEARSON Education.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning : An Introduction*. MIT Press, Cambridge, MA.

TORCS Team (2001). The open racing car simulator. <http://torcs.sourceforge.net/>.

Zhou, L. and Zimmer, M. (2013). *Cahier des charges - PIAD Semi-supervised Learning Agents*.

1.3 Définitions et acronymes

Env. Réduit Il s'agit de l'environnement de TORCS avec des perceptions limitées pour l'agent conducteur :

- son angle (par rapport à la droite au milieu de la route)
- la distance entre son centre de gravité et le milieu de la route

ainsi que des actions limitées pour la voiture :

- son accélérateur
- son freinage
- sa direction

On s'attend donc à ce que l'agent dans cet environnement roule toujours avec la première vitesse.

Env. Complexe Il s'agit toujours de l'environnement de TORCS mais agrémenté de nouvelles perceptions

- coincé ou non (dans un mur)
- vitesse (pour savoir comment accélérer/freiner)
- RPM (pour savoir quelle vitesse mettre)
- friction du sol (pour déterminer s'il est sur la route ou en dehors, quel type de terrain, ...)
- distance du prochain virage (pour déterminer l'accélération et la vitesse maximale)
- angle du prochain virage (pour le prendre plus sec)

et de nouvelles actions

- boîte à vitesse

CdC Cahier des charges

PdD Plan de développement

APR Apprentissage par Renforcement

MDP Processus de Décision Markovien En plus de ces définitions, les documents permettant de bien comprendre le plan sont les suivants : Groupe PDMIA (2008) et Sutton and Barto (1998).

2 Organisation du PIAD

2.1 Participants

Encadrants :

- Paolo VIAPPANI (Paolo.Viappiani@lip6.fr)
- Paul WENG (Paul.Weng@lip6.fr)

Etudiants :

- Lan ZHOU (lan_612zhou@yahoo.cn)
- Matthieu ZIMMER (contact@matthieu-zimmer.net)

2.2 Activités globales

Scripts Maintenir et développer un ensemble de script permettant de compiler, installer, nettoyer, lancer des tests et faciliter l'ensemble du développement du projet.

Intégration Linux Maintenir la compatibilité des sources sous le système Linux.

Intégration Mac Maintenir la compatibilité des sources sous le système Mac.

2.3 Tâches jalonnées

Prise de connaissances de l'APR Lire les 2 premiers chapitres de Groupe PDMIA (2008) et se familiariser avec l'apprentissage par renforcement. Sans oublier le chapitre dédié de Russell and Norvig (2006).

Compréhension du Robot TORCS Comprendre la structure des les fichiers nécessaire pour concevoir un robot dans TORCS. Déterminer les perceptions disponibles et les actions possibles pour le robot.

Etablissement du CdC/PdD Définir le cahier des charges avec les encadrants, découper le projet en plusieurs tâches, assigner les rôles de chacun.

Analyse et Conception Réfléchir aux différentes classes et méthodes nécessaire pour concevoir une bibliothèque réutilisable. Concevoir les diagrammes de classe.

Implémentation du Q-learning Implémenter l'algorithme Q-learning dans la bibliothèque de façon générique et réutilisable.

Implémentation de SARSA Implémenter l'algorithme SARSA dans la bibliothèque de façon générique et réutilisable.

Implémenter la fonction de Récompenses Coder des fonctions qui permettront de donner une valeur à un état, pour ça il faut retrouver les informations données par TORCS et y appliquer quelques formules mathématiques.

Interfacer Q-learning - Env. Réduit Il s'agit de faire le lien entre la fonction générique Q-learning codée dans la bibliothèque et le robot dans l'environnement réduit de TORCS.

Interfacer SARSA - Env. Réduit Il s'agit de faire le lien entre la fonction générique SARSA codée dans la bibliothèque et le robot dans l'environnement réduit de TORCS.

Implémentation des méthodes approchés de l'APR Etant donné que dans les environnements complexes il y a une explosion combinatoire, il nous faut développer des moyens pour réduire le nombre d'états de nos agent. C'est le rôle de cette tâche.

Interfacer Q-learning - Env. Complexe Il s'agit de faire le lien entre la fonction générique Q-learning codée dans la bibliothèque et le robot dans l'environnement réduit de TORCS.

Interfacer SARSA - Env. Complexe Il s'agit de faire le lien entre la fonction générique SARSA codée dans la bibliothèque et le robot dans l'environnement réduit de TORCS.

Implémenter les mesures de Perf. Pour permettre de comparer les différents algorithmes et les différents

Modifier l'interface de TORCS Intégrer les retours positifs/négatifs de l'expert.

Fonctionnalités additionnelles Continuer le développement parmi les fonctions facultatives dont nous avons parlé dans Zhou and Zimmer (2013).

2 ORGANISATION DU PIAD

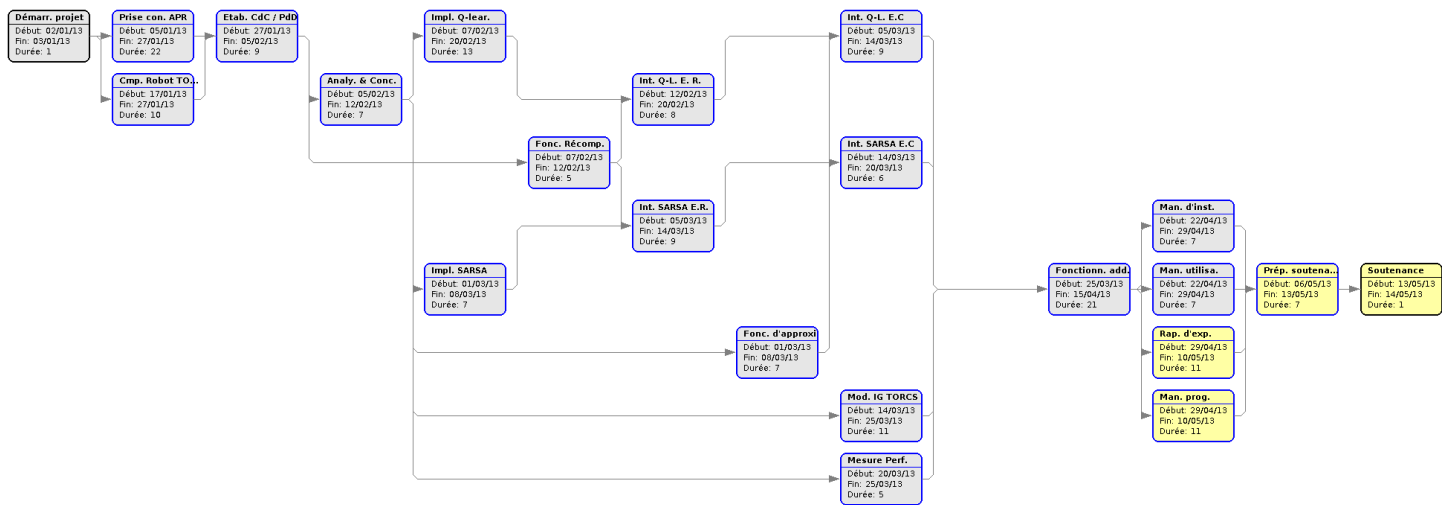


FIGURE 1 – Diagramme de PERT (zoomer pour meilleur qualité)

2.4 Rôles et responsabilités

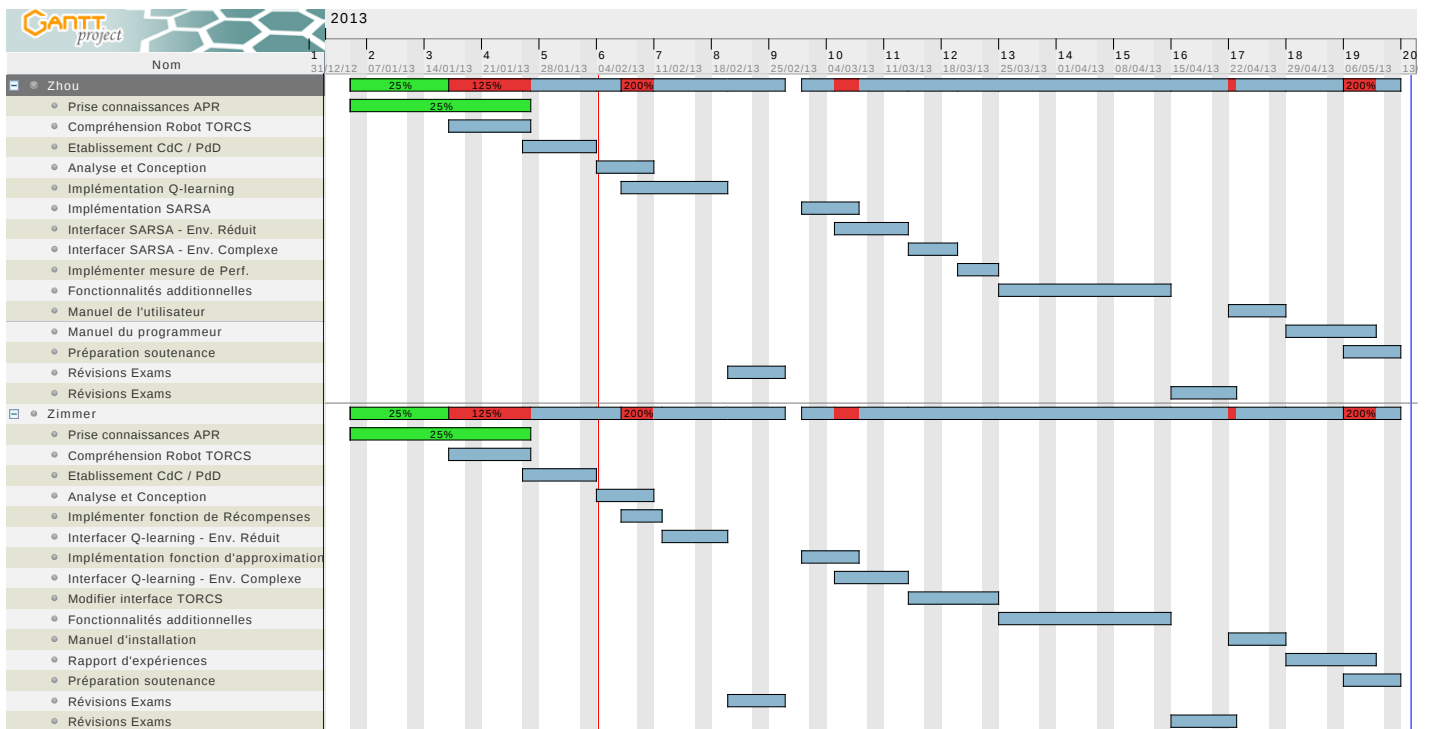


FIGURE 2 – Diagramme de Gantt par Ressource (zoomer pour meilleur qualité)

2.5 Calendrier

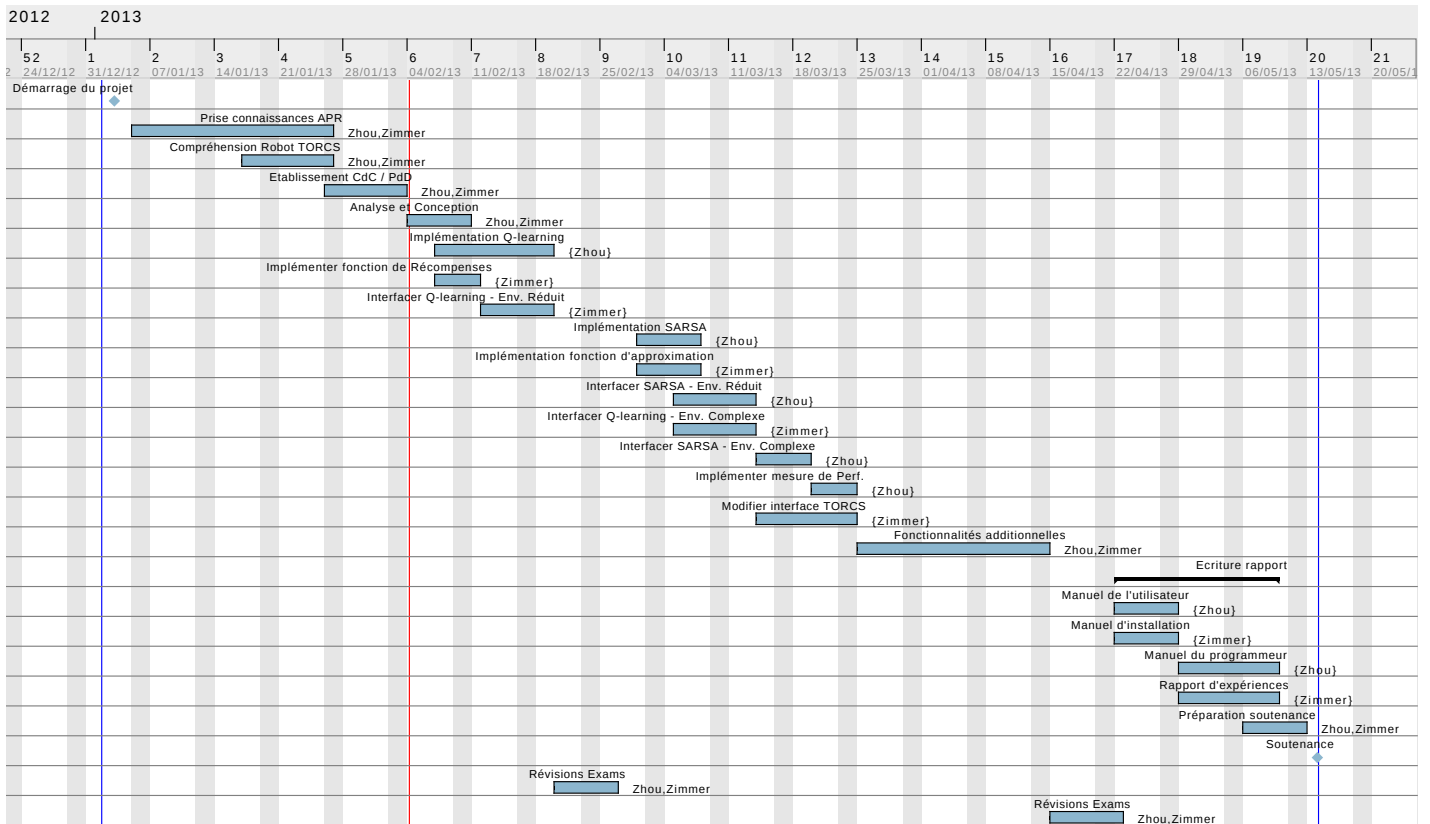


FIGURE 3 – Diagramme de Gantt (zoomer pour meilleur qualité)