Une approche multi-agent adaptative pour la simulation de schémas tactiques

Aydano Machado*, Yann Chevaleyre**, Jean-Daniel Zucker*

* Laboratoire d'Informatique de Paris VI (LIP6) – Université Paris 6
Boîte 169 – 4 Place Jussieu
75252 PARIS CEDEX 05
{Aydano.MACHADO, Jean-Daniel.ZUCKER}@lip6.fr
http://www-poleia.lip6.fr/~{machado, zucker)
** LAMSADE – Université Paris-Dauphine
place du Maréchal de Lattre de Tassigny
75775, Paris
Yann.Chevaleyre@lamsade.dauphine.fr
http://www.lamsade.dauphine.fr/~chevaley

Résumé. Ce papier est consacré à la simulation ou à la réalisation automatique de schémas tactiques par un groupe d´agents footballeurs autonomes. Son objectif est de montrer ce que peuvent apporter des techniques d'apprentissage par renforcement à des agents réactifs conçus pour cette tâche. Dans un premier temps, nous proposons une plateforme et une architecture d'agents capable d'effectuer des schémas tactiques dans des cas relativement simples. Ensuite, nous mettons en œuvre un algorithme d'apprentissage par renforcement pour permettre aux agents de faire face à des situations plus complexes. Enfin, une série d'expérimentations montrent le gain apporté aux agents réactifs par l'utilisation d'algorithmes d'apprentissage.

1 Introduction

Dans le domaine des sports en équipe, de plus en plus d'entraîneurs font appel à des outils informatiques durant leur activité pédagogique, en particulier de logiciels de simulation afin d'enseigner aux joueurs à améliorer leur tactique. Jusqu'à présent, ces logiciels qui permettaient essentiellement à l'entraîneur de faire se déplacer sur un écran des agents joueurs, nécessitaient de sa part de spécifier quasiment trame par trame la position des agents. Par voie de fait, un entraîneur souhaitant montrer le déploiement d'un schéma tactique particulier doit effectuer un important travail avant que la simulation puisse être lancée.

Dès lors, rendre les agents plus autonomes, améliorer le réalisme de leur comportement et leur capacité de prendre des décision allégerait le travail de l'entraîneur, et lui permettrait de n'avoir qu'à spécifier des schémas tactiques relativement abstrait pour voir comment des agents joueurs déploieraient ce schéma « intelligemment » sur le terrain.

Notre objectif est donc d'utiliser diverses techniques d'intelligence artificielle pour améliorer l'autonomie des agents devant déployer un schéma spécifié par l'entraîneur. Cette tâche peut être considéré comme un sous-ensemble du problème de la simulation sportive

(par exemple la RoboCup), du fait que les agents se voient indiqués la route à suivre (le schéma tactique), mais doivent pouvoir en dévier s'ils croisent un adversaire qui leur prend la balle.

Dans un premier temps, un système multi-agents a été construit, dans lequel les agents suivent un comportement décrit dans une base de règle. Nous avons montré sur quelques schémas que les comportements obtenus étaient parfois insuffisants, du fait que les agents ne s'autorisaient pas à dévier suffisamment des indications de l'entraîneur pour faire face à l'adversaire. Ensuite, nous avons implémenté un algorithme d'apprentissage par renforcement qui a permis aux agents de se comporter correctement dans les cas ou ils échouaient avec le système à base de règles. Enfin, nous avons créé une plateforme logicielle intégrant ces différents algorithmes et permettant à l'entraîneur de faire des simulations aisément.

Dans ce papier, après avoir introduit les schémas tactiques, nous présentons essentiellement les résultats des expérimentations de l'algorithme d'apprentissage par renforcement, que nous comparons sur différents schémas aux performances des agents basés sur des règles. Nous montrons en particulier que l'apprentissage converge rapidement malgré la dimension importante du problème.

2 La simulation de schémas tactiques

En regardant l'état de l'art du sujet en question nous avons trouvé deux axes principaux : les outils commerciaux et les travaux scientifiques.

Les représentants du premier axe sont destinés aux professionnels du football et traitent effectivement les schémas tactiques, mais ils proposent une solution simple pour le déploiement d'un schéma tactique. En fait, ils ne font qu'une animation d'une séquence de positions prédéfinies en utilisant des techniques d'interpolation d'images où les objets ont une trajectoire rectiligne entre deux positions successives. Le résultat final est donc une animation dont la qualité dépend de l'intervalle entre chaque image. Plus l'intervalle augmente plus la fiabilité diminue. Cette approche laisse tout le travail fastidieux et répétitif à l'utilisateur du programme qui doit prévoir et décrire en détail le déplacement des objets à chaque instant. En particulier si un changement se présente ce travail est à refaire.

La RoboCup est le plus grand représentant du deuxième axe. Il s'agit d'un projet de coopération international destiné à encourager le développement de l'intelligence artificielle (IA), de la robotique et d'autres domaines connexes. Du point de vue des systèmes multi-agents (SMA) le modèle footballistique crée par la RoboCup est un défi intéressant car il regroupe plusieurs caractéristiques (Noda et al., 1997), dont un environnement qui évolue dynamiquement, une nécessité pour les agents de communiquer et se coordonner pour atteindre leurs objectifs.

Actuellement, il n'existe pas encore d'équipe utilisant le déploiement de schémas tactiques tel quel le font les joueurs de football. Les équipes ont déjà suffisamment de problèmes à résoudre avec les contraintes et les définitions imposées par le modèle en question.

Nous proposons une solution qui s'inscrit dans le premier axe (pour les professionnels du sport) pour la simulation de schémas tactiques basé sur les SMA comme la RoboCup mais sans toutes ses contraintes et déterminations. Nous avons opté pour l'utilisation des méthodes d'apprentissage automatique pour la conception des agents, évitant ainsi la tâche complexe de programmer les comportements des joueurs. Dotés de la faculté d'apprendre, les agents

gagneront en autonomie, et seront capables de s'adapter à leur adversaire ou à leur environnement.

En entraînant des agents à jouer avec certains schémas tactiques, l'entraîneur pourra développer chez ces agents des aptitudes particulières liées à ces schémas. Par exemple : il pourra créer un défenseur en mettant un agent joueur face à des attaquants avec le ballon, avec pour objectif apprendre à récupérer le ballon.

2.1 Algorithme d'apprentissage

Nous avons choisi l'apprentissage par renforcement parce que l'agent apprend par interaction avec l'environnement sans avoir besoin d'exemples. Dans les sections suivantes, nous définirons les récompenses et l'agent devra découvrir par un processus d'essais et d'erreurs, l'action optimale à effectuer pour chacune des situations afin de maximiser ses récompenses (Sutton, 1998). Pour modéliser un problème en utilisant de l'apprentissage par renforcement, on doit se poser les questions suivantes : quelles actions peuvent être effectuées par les agents ? Quelle représentation de l'environnement employer ? Quelles récompenses leur donner ?

2.1.1 Les espaces d'actions et d'état

En ce qui concerne l'espace d'état, nous avons choisi une représentation basée sur des distances, afin que les comportements des agents ne dépendent que des positions relatives des uns par rapport aux autres, et non de leur position absolue sur le terrain (voir le *TAB. 1*).

	Caractéristiques de l'état
1	distance entre l'objectif de l'agent et le but de l'adversaire
2	distance entre l'objectif de l'agent et son but
3	distance entre l'agent et son objectif (-1 s'il n'a pas d'objectif)
4	distance entre l'agent et le but de l'adversaire
5	distance entre l'agent et le son but
6	distance entre l'agent et l'adversaire le plus proche (-1 s'il n'a pas d'adversaire)
7	distance entre l'agent et le ballon
8	si l'agent a le ballon ou pas
9	quelle équipe a le ballon (-1=adversaire ; 0=personne ; 1=son équipe)
10	si l'agent est sensé arriver à l'objectif avec le ballon ou pas
11	distance entre l'agent et le compagnon Ti (-1 s'il n'a pas d'adversaire)
12	angle[1] entre l'axe x et le compagnon Ti ([0 - 2π[ou -1 si Ti n'existe pas)

TAB. 1 – *Description des caractéristiques des états*

En utilisant comme base les comportements de navigations primaires décrites par Reynolds (1999), nous avons créé les actions suivantes :

- **Déplacement vers un point** : elle combine les comportements *seek* et *unaligned collision avoidance* afin d'aller vers un point en évitant les collisions ;
- **Déplacement en groupe vers un point** : Cette action permet au joueur de se déplacer tout en restant proche du groupe, grâce au comportement *cohesion*.
- **Positionnement** : inspiré de l'action se positionner de Veloso et al. (1999), elle combine les comportements de *cohesion*, *separation* et *unaligned collision avoi-*

[1] en sens inverse des aiguilles d'une montre

dance. Cette action va déplacer l'agent vers une zone stratégique qui est en même temps proche de ses compagnons et loin de ses adversaires, en évitant les collisions.

- **Marquage d'un adversaire** : une heuristique permet déterminer un point proche d'un adversaire situé entre le but et le ballon, vers lequel le joueur se dirige.
- **Interception du ballon** : nous utilisons l'algorithme de Stone et McAllester (2001) pour trouver le temps nécessaire à l'interception du ballon. Avec cette information l'agent peut savoir vers où il doit aller pour attraper le ballon.

Il y a d'autres actions également importantes, mais qui n'utilisent pas les comportements de navigation, on peut lister : **faire une passe** et **prendre le contrôle du ballon**.

2.1.2 Le renforcement

Pour définir les récompenses d'une manière plus facile nous avons déterminé un ordre de priorités. Tout d'abord il faut réaliser le schéma tactique (arriver aux objectifs avec les conditions satisfaites), ensuite ne pas perdre le ballon et enfin ne pas mettre le ballon hors du terrain. D'après les priorités définies, nous avons attribué les valeurs du *TAB. 2* comme récompense.

Situation	Récompense
Réaliser le schéma	100
Réussir un objectif	10
Perdre le ballon	-100
Ballon hors du jeu	-50
Déplacement	[–1 ; 1]*

TAB. 2 – *Récompenses (*variation de la distance entre l'agent et son objective normalisé).*

2.1.3 Fonction d'évaluation

Nous avons créé une fonction pour évaluer la situation actuelle d'une équipe par rapport à la réalisation d'un schéma tactique. Cette fonction est donc la somme des évaluations individuelles de tous les joueurs de l'équipe.

Pour évaluer individuellement un joueur *j* nous prenons en compte le nombre d'objectifs qui ont été déjà réussis (noté or_i), la distance normalisée (entre 0 et 1) au prochain objectif *o* (notée dn_{oj}), et nous calculons la valeur $e(j) = (or_j + dn_{oj})/no_j$, où no_j est le nombre total d'objectifs du joueur *j*. Dès lors, pour évaluer une équipe *t*, nous définissons la fonction *s(t)* égale à la moyenne des évaluations des joueurs de l'équipe.

3 Expérimentation et résultats

Pour interagir avec l'entraîneur, réaliser nos expérimentations et voir les résultats, nous avons réalisé une plateforme comportant différents modules dont : un module d'interaction avec l'entraineur, un module d'apprentissage par renforcement. En plus des agents basés sur l'apprentissage par renforcement, nous avons implémenté un autre type d'agents dont le comportement dérive d'un système à base de connaissances (SBC), ce qui constitue une approche plus classique. Ainsi, en comparant les deux approches, nous pouvons mesurer le gain apporté par un algorithme d'apprentissage. Pendant nos expérimentations, nous avons utilisé un CMAC avec 32 tableaux de 9 cases (pour les paramètres continus et 1, 2 ou 3 cases pour les valeurs discrètes). Les configurations utilisées sont résumées par le *TAB. 3*.

Configuration	(a)	(b)	(c)	(d)
Epsilon	0,3	0,3	0,2	0,2
Alpha	0,2	0,2	0,3	0,3
Lambda	0,9	0,9	0,9	0,9
Gamma	0,9	0,9	0,8	0,8
Algorithme	SARSA	QL	SARSA	QL

TAB. 3 – *Paramètres utilisés par l'algorithme d'apprentissage*

Nous avons commencé les expérimentations avec le schéma présenté par la Fig. 1. L'équipe noire (les défenseurs) utilise le SBC implémenté. Le schéma détermine que le joueur 0 doit arriver à son objectif sans le ballon et que le joueur 1 doit arriver à son objectif avec le ballon. Selon le SBC, un agent va se déplacer vers son objectif sans faire attention au ballon et l'autre va le chercher. L'équipe grise (les attaquants) utilise l'apprentissage et le joueur 0 n'a pas un objectif défini, mais par contre le joueur 1 doit arriver sur la zone définie avec le ballon. Il faut souligner que les flèches sont uniquement illustratives, l'entraîneur ne précise que les zones et les conditions pour la zone (si le joueur doit passer avec le ballon ou si la zone doit être prise en compte).

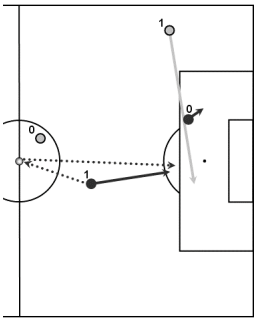

FIG. 1 – *Schéma tactique.*

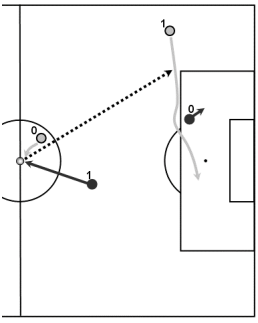

FIG. 2 – *Déploiement du schéma tactique.*

Le graphe qui montre l'évolution de l'apprentissage de nos agents selon la quantité d'épisodes (axe des abscisses) et notre fonction d'évaluation (axe des ordonnées) pour la configuration a est présentés par la Fig. 3.

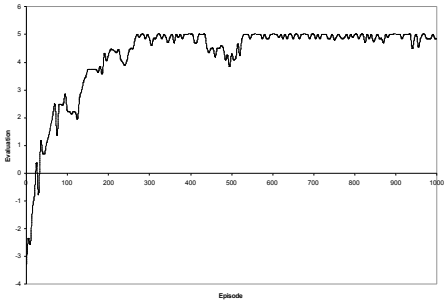

FIG. 3 – *Evaluation de l'apprentissage.*

Pour toutes les configurations, nous avons un résultat final similaire, montré par la Fig. 2, où l'agent qui n'a pas d'objectif apprend qu'il faut chercher le ballon (sinon l'équipe adverse va le prendre) pour ensuite faire une passe à son coéquipier pour pouvoir accomplir le schéma tactique donné.

4 Conclusion et travaux futurs

Le travail présenté ici montre comment une approche SMA adaptative pour la simulation de schémas tactiques peut être mise en œuvre, quel type de résultats on peut en attendre et quels sont les apports vis-à-vis des autres solutions existantes.

Nous allons poursuivre la recherche en variant les composants de l'architecture du SMA et les composants de la méthode d'apprentissage (e.g. types d'agents, de mémoire, etc.). Nous nous intéressons notamment à l'emploi de techniques pour améliorer l'évolution et la coordination de l'apprentissage. De plus nous étudions l'utilisation de l'apprentissage par imitation pour apprendre à partir de séquences vidéo numérisées, en accélérant donc l'apprentissage.

A travers ce travail, nous espérons ouvrir une voie nouvelle dans les approches de la simulation numérique dans le milieu de tactiques sportives et contribuer à la conception de nouveaux outils d'aide aux entraîneurs et autres professionnels du sport.

Références

Noda, I., H. Matsubara, K. Hiraki et I. Frank (1997). *Soccer server : a tool for research on multi-agent systems*.

Reynolds, C. W. (1999). *Steering Behaviors For Autonomous Characters*. Présenté à Game Developers Conference GDC1999, San Jose, California.

Stone, P. et D. McAllester (2001). *An architecture for action selection in robotic soccer*. Montreal, Quebec, Canada.

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

Veloso, M., P. Stone, et M. Bowling (1999). *Anticipation as a key for collaboration in a team of agents: A case study in robotic soccer*. Présenté à SPIE.

Summary

This paper is about simulation of tactical schemes by a group of autonomous soccer agents. It shows what advantages are obtained if we apply reinforcement learning techniques to a group of reactive agents designed for this task. Firstly, we propose a framework and an agent architecture to carry out simple tactical schemes. Then, we implement a reinforcement-learning algorithm to allow the agents to learn how to act in face of more complex and uncrowned situations. Finally, the results of our experiments are showed to confirm the learning algorithm benefits and success.