



# **Evaluation Multi-critères et Distribuée pour l'Apprentissage Collectif de Procédures dans un Jeux Sérieux pour la Gestion de Crise**

Ali Mhammed Oulhaci, Erwan Tranvouez, Sébastien Fournier, Bernard Espinasse

## **► To cite this version:**

Ali Mhammed Oulhaci, Erwan Tranvouez, Sébastien Fournier, Bernard Espinasse. Evaluation Multi-critères et Distribuée pour l'Apprentissage Collectif de Procédures dans un Jeux Sérieux pour la Gestion de Crise. Journée EIAH&IA 2013, May 2013, Toulouse, France. pp.1. <hal-00824333>

**HAL Id: hal-00824333**

**<https://hal.science/hal-00824333v1>**

Submitted on 21 May 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Évaluation Multi-critères et Distribuée pour l'Apprentissage Collectif de Procédures dans un Jeux Sérieux pour la Gestion de Crise

M'hammed Ali Oulhaci<sup>1,2</sup>, Erwan Tranvouez<sup>1</sup>, Sébastien Fournier<sup>1</sup> et Bernard Espinasse<sup>1</sup>

<sup>1</sup> Aix-Marseille Université (AMU), LSIS UMR CNRS 7296  
38, Av. Normandie-Niemen 13397 Marseilles Cedex 20 FRANCE  
<sup>2</sup> Groupe SII, Société d'Ingénieur et de Conseil en Technologies  
Route nationale 7 Le Tholonet 13100 FRANCE  
{ali.oulhaci, erwan.tranvouez, sebastien.fournier,  
bernard.espinasse}@lsis.org

**Résumé.** Les jeux sérieux sont de plus en plus utilisés pour la formation, et en particulier dans le domaine de la gestion de crise. Afin d'améliorer la formation, l'évaluation des apprenants peut fournir des indications sur ce qui a été bien ou mal appris au cours d'une session de formation. Cette appréciation se complexifie lorsqu'elle doit prendre en compte non seulement les actions individuelles des acteurs mais également le résultat de leurs interactions (actions collectives). De plus, selon la nature des phénomènes à émuler (incendie, pollution...) et le nombre d'acteurs impliqués, des besoins de simulations s'ajoutent (propagation incendie, simulation de comportement d'acteurs absents...). L'évaluation des apprenants implique ainsi l'acquisition d'un ensemble hétérogène d'informations à la fois par leur nature (données brutes, connaissances, procédures) et par leur source (niveau logiciel). Cet article propose une approche multi-agents répondant à deux objectifs: i) permettre la production et l'exploitation de ces informations hétérogènes en vue de produire des indicateurs qui alimenteront une évaluation multi-critères; et ii) mettre en œuvre le processus d'évaluation par un système multi-agents accompagnant la production d'évaluations individuelles et collectives. Cette approche est développée et illustrée sur le jeu sérieux SIMFOR dédié à la gestion de crise pour le faire évoluer vers un Système tutoriel.

**Mots-clé:** Jeux Sérieux, Intelligence Artificielle Distribuée, Systèmes Multi-Agents, Simulation Orientée Agent, Évaluation des apprenants

## 1 Introduction

L'intérêt croissant des jeux sérieux, notamment dans le domaine de la formation, a soulevé des besoins nouveaux en termes d'évaluation des joueurs [1] et de simulation de comportement [2]. Un jeu sérieux est généralement défini comme un jeu qui privilégie l'éducation (au sens large) plutôt que le divertissement [3] basé sur

des techniques issues des jeux vidéo (environnement 3D, modes d'interaction). Les jeux sérieux nécessitent également des capacités de simulation pour reproduire des phénomènes complexes ou coûteux (e.g. catastrophes naturelles). Certains jeux sérieux nécessitent de simuler des comportements humains lorsqu'un nombre important d'acteurs est impliqué : certains d'entre eux sont gérés par ordinateur (appelés Personnages Non Joueurs PNJ). Les PNJ sont souvent mis en oeuvre par une Simulation Multi-Agents (SMA) [2].

Dans un jeu sérieux, le but est d'apprendre une compétence. Pour valider l'apprentissage, un processus d'évaluation s'impose, aussi l'évaluation des joueurs constitue un enjeu majeur dans les jeux sérieux [1]. Certains travaux utilisent le concept de tuteur intelligent [4][5] qui permet de suivre les joueurs dans leur formation et les assiste si nécessaire. Ces travaux se rapprochent des systèmes tutoriels intelligents (STI) [6].

Dans ce papier nous présentons le projet SIMFOR, un jeu sérieux pour la formation des non professionnels à la gestion de crise, ainsi que les enjeux de formation qu'il soulève. Nous montrons en quoi les SMAs, relevant de l'Intelligence Artificielle Distribuée, peuvent améliorer l'apprentissage humain dans un contexte collectif, au travers de la simulation de comportements humains et de l'évaluation des joueurs apprenants. Dans la section suivante, nous présentons le projet SIMFOR à l'origine de ce travail de recherche, en montrant les enjeux d'évaluation qu'il pose et en les positionnant vis à vis de travaux existants. Les contributions proposées, soit la méthode d'évaluation proposée et les considérations architecturales qu'elle implique, sont introduites respectivement dans les sections 3 et 4. Les dernières sections illustrent notre approche sur un scénario simplifié et concluent sur les résultats obtenus et le travail restant à faire.

## 2 Présentation du projet SIMFOR

Le projet SIMFOR (figure 1) est un jeu sérieux développé par la société SII<sup>3</sup> en partenariat avec la société Pixxim<sup>4</sup> en réponse à l'appel à projet "*Serious Gaming*" lancé par le secrétaire d'état chargé de la Prospective et du Développement de l'économie numérique. Implémenté en C++, SIMFOR permet une approche ludique de l'apprentissage de la gestion de crise sous la forme d'un jeu sérieux. SIMFOR permet aux apprenants de se former à la gestion de crise en intégrant l'aspect multi-intervenants. L'objectif de SIMFOR est de proposer aux utilisateurs un contexte de gestion de crise en temps réel proche de la réalité en termes d'environnement, de scénarios auto-évolutifs et d'acteurs. Reposant initialement sur une évaluation humaine des compétences et des PNJ aux comportements simplifiés, nous visons à faire évoluer SIMFOR en développant :

- La simulation de comportement humain des acteurs non joués.
- L'accompagnement et l'évaluation des apprenants dans leur formation.

<sup>3</sup> <http://www.groupe-sii.com>

<sup>4</sup> <http://www.pixxim.fr>



**Fig. 1.** Captures d'écran du projet SIMFOR

SIMFOR est un jeu multi-acteurs permettant à plusieurs personnes d'acquérir une compétence (différente ou identique), avec la particularité de ne pas cibler uniquement les spécialistes du domaine mais également des non professionnels. La gestion d'une crise majeure peut mobiliser plusieurs centaines d'intervenants, du Préfet dans son bureau au pompier sur le terrain. Ces intervenants sont amenés à communiquer et à collaborer pour rétablir la situation à un état normal. Devant le nombre d'intervenants, le nombre d'absents à une session d'entraînement est non négligeable: il faut alors pouvoir remplacer les apprenants (joueur humain) par des joueurs virtuels. Si une approche de simulation orientée agent s'est rapidement imposée pour développer ces joueurs virtuels, l'approche agent est apparue également pertinente pour assurer l'évaluation des apprenants.

En tant que jeu sérieux, SIMFOR vise à immerger l'apprenant dans un monde virtuel qui lui permet d'agir comme il le ferait dans une situation d'urgence. Les connaissances et les compétences impliquées dans une telle situation sont diverses dans leur nature et leur source. Ce dernier point relève de considérations en terme d'architecture logicielle (comment gérer l'intégration de source d'informations hétérogènes par ex.), nous étudions tout d'abord dans cette section la conséquence de cette hétérogénéité en terme d'évaluation. Ainsi, se pose la question de comment évaluer et certifier (ou non) les connaissances et les compétences acquises par les joueurs. Afin de mieux comprendre l'aspect hétérogène des besoins d'évaluation, la section suivante décrit un exemple de mise en situation.

## 2.1 Problématique d'évaluation soulevée par le projet SIMFOR

Considérons le scénario suivant qui commence par un camion TDM (Transport des Matières Dangereuses) qui s'est renversé après un accident de la circulation. Le réservoir est endommagé et de l'hydrocarbure se répand sur la chaussée. Un témoin de l'accident donne l'alerte en appelant le CODIS (Centre Départemental

d'Incendie et de Secours) qui à son tour doit effectuer quatre missions pour répondre à l'alerte. Tout d'abord, le CODIS doit envoyer un pompier sur les lieux pour récupérer les informations sur le sinistre ("*Envoyer pompier*"). Une fois les informations sur le sinistre reçues (transmises par le pompier sur les lieux) et confirmant l'accident TDM, le CODIS doit donner ensuite des instructions à un officier (pompier) sur les mesures à prendre. Le CODIS doit ensuite remplir une fiche de renseignements sur le sinistre et l'envoyer par fax au maire, au préfet et au sous-préfet (l'ordre d'envoi n'est pas important). Enfin, la dernière mission est d'informer l'officier du lieu du PCO (Poste de Commandement Opérationnel) une fois ce dernier établi et communiqué par le préfet.

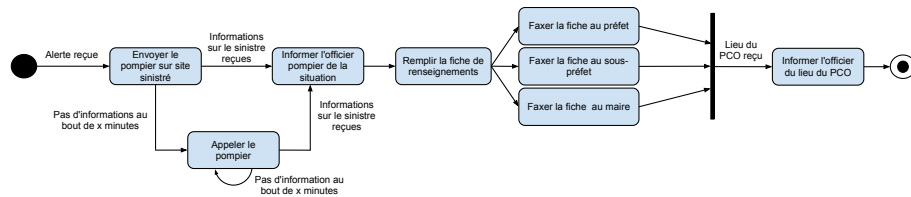


Fig. 2. Diagramme d'activité UML

Ce scénario implique plusieurs acteurs (CODIS, pompier, ...), chacun avec des procédures et des connaissances différentes à évaluer ainsi que l'orchestration des interactions qui doit être respectée. Comme ces informations diffèrent par leur nature, chacune nécessite une représentation et une méthode d'évaluation spécifique. Le diagramme d'activité UML de la figure 2 représente, de manière simplifiée, les connaissances procédurales de l'acteur jouant le rôle du CODIS, sous la forme de mission à réaliser. Les interactions peuvent être évaluées par le temps écoulé entre la détection du sinistre et sa résolution ainsi que la zone impactée (par ex. zone détruite par l'incident). Chaque connaissance nécessaire à l'évaluation peut utiliser un modèle de représentation différent, ainsi qu'un calcul d'évaluation spécifique pour produire un score d'apprentissage pour valider ou non l'objectif d'apprentissage d'un scénario. Ces connaissances peuvent être produites individuellement ou collectivement (interactions entre acteurs et/ou joueurs simulés).

Ainsi, l'évaluation des apprenants nécessite la collecte d'informations provenant de différentes sources en fonction du rôle incarné par l'apprenant dans le scénario. Par exemple, pour évaluer la première action de l'acteur CODIS (téléphoner), nous avons besoin de données de l'environnement pédagogique (environnement 3D). Nous avons également besoin de l'historique des actions de l'apprenant pour ensuite les comparer avec la séquence d'actions prévue dans le modèle du domaine (défini par un expert du domaine, figure 2). Les informations nécessaires pour l'évaluation varient donc dans leur type ainsi que leur nature (selon le rôle joué par l'acteur), et devront être acquises auprès de sources de connaissances ou de données distribuées sur des composants logiciels variés.

## 2.2 Travaux existants

L'accompagnement des apprenants et leur évaluation est un sujet déjà traité dans la littérature [7][4]. Les besoins d'évaluation des jeux sérieux ciblent plus particulièrement l'évaluation durant l'apprentissage (aide en temps réel) et après l'apprentissage (diagnostic sur les compétences apprises) impliquant éventuellement un "débriefing" entre l'apprenant et le moniteur. Dans ce contexte, les SMA interviennent principalement dans la simulation de comportement d'acteurs virtuels (PNJ) [2] et dans une moindre mesure dans le processus d'évaluation.

L'évaluation dans les jeux sérieux est une problématique de recherche en soit [1][8], à laquelle peut répondre les Systèmes Tutoriels Intelligents (STI), pour qui l'évaluation des apprenants constitue un enjeu théorique et expérimental majeur. En combinant les possibilités d'évaluation d'un STI avec un jeu sérieux, nous pouvons ainsi améliorer les résultats de formation utilisant un jeu sérieux. Les travaux de Lourdeaux et al. [4], proposent un outil basé sur un STI pour la formation des conducteurs de TGV avec le concept de HAL (Help Agent for Learning). L'environnement de formation de HAL est basé sur la réalité virtuelle avec la reproduction d'un cockpit de TGV via un écran géant affichant l'environnement 3D. Par définition HAL est mono-apprenant et ne permet pas une évaluation finale afin de diagnostiquer l'apprenant.

Dans le domaine de la gestion des risques, le projet PEGASE (PEDagogical Generic and Adaptive SystEm) est un STI pour la gestion des risques à bord des porte-avions, mobilisant néanmoins un nombre de rôles joués limités (uniquement les membres d'équipage) et n'aborde pas l'évaluation collective. Toujours dans la gestion des risques, Amokrane [9] propose un outil de formation pour les sites SEVESO : HERA (Helpful agent for safeTy leaRning in virtuAl environment). Si l'évaluation des apprenants est traitée, elle reste individuelle (mono-apprenant) et ne traite pas des procédures collaboratives.

L'apprentissage collaboratif vise à apprendre à un groupe d'apprenants à travailler ensemble pour améliorer l'apprentissage individuel et les résultats de leur collaboration. Woolf [10] fait un diaporama des environnements informatiques contribuant à ce domaine selon quatre niveaux de soutien à la collaboration [11]:

- *Structuration* de la collaboration: soit comment constituer l'équipe qui va collaborer (choix d'un partenaire par ex.).
- *Retour d'information (mirror)* sur la collaboration: ou comment rendre compte aux acteurs de la dynamique de leurs interactions au travers de tableaux de bord graphiques (par ex. un graphe d'interaction).
- *Analyse métacognitive* : il s'agit d'analyser voire diagnostiquer ces interactions et l'avancement des apprenants. Dans le projet COMET [12] des étudiants en médecine doivent diagnostiquer un malade via un ensemble de symptômes. Durant leur diagnostic, le système agrège les données d'interaction des apprenants en indicateurs de haut niveau et les en informe.
- *Guidage dans la collaboration* : ceci peut revenir à exploiter les résultats précédents pour expliquer et guider les acteurs sur leur processus collaboratif. Ainsi le projet COLER ([13]) (dans le cadre de l'apprentissage de la conception de base de données) détecte les opportunités d'interaction dans

un groupe (différence de résultats entre groupes, niveau d'implication, et qualité des solutions) et peut faire des recommandations d'interaction.

L'apprentissage collaboratif cible majoritairement l'apprentissage de compétences homogènes (à l'image d'une classe où les élèves apprennent le même cours), alors que la gestion de crise est un processus collaboratif qui passe par la réalisation de plusieurs tâches différentes (dépendant du rôle de l'acteur ainsi que du sinistre). Donc SIMFOR vise un apprentissage de tâches collaboratives hétérogènes, réalisées par des acteurs différents ayant un objectif commun.

Ainsi SIMFOR présente des besoins spécifiques du fait de son caractère multi-acteurs qui requiert deux types d'évaluation : une évaluation individuelle et une évaluation collective. La résolution d'une crise (incendie, pollution...) passe par la résolution de toutes les procédures des intervenants : dès lors l'évaluation individuelle peut avoir une influence sur l'évaluation collective, et réciproquement. Considérons, par exemple, un apprenant qui a bien effectué ses procédures, mais dont le but global n'a pas été atteint (e.g. pertes matérielles et humaines). Cet apprenant doit être évalué sur sa performance individuelle et collective pour déduire la raison de l'échec (manque de communication, procédure manquante d'un autre apprenant, ...). La section suivante décrit comment nous entendons répondre à ces besoins.

### 3 Amélioration de l'évaluation dans les jeux sérieux : le concept d'espace d'évaluation

Dans cette section nous proposons d'intégrer l'évaluation des apprenants dans un jeu sérieux, illustré avec le projet SIMFOR, en prenant en compte la nature et les origines diverses des éléments nécessaires à la production d'indicateurs alimentant cette évaluation.

#### 3.1 Définition du concept d'espace d'évaluation

Comme vu dans la section précédente, intégrer l'évaluation dans un jeu sérieux implique l'utilisation de connaissances, d'informations ou de données, produites ou transformées en permanence jusqu'à la fin du déroulement du scénario de jeu. Chaque élément d'information nécessite une manipulation spécifique (voire un raisonnement sur ces connaissances) afin d'en extraire une évaluation. Une façon naturelle de faire face à la complexité de la gestion de ces informations (au sens large) est de diviser et d'organiser ces informations en groupes homogènes dans lesquels des primitives dédiées peuvent être utilisées pour produire une évaluation. La notion *d'espace d'évaluation* s'inscrit dans cette démarche en englobant tous les éléments nécessaires pour produire des évaluations, en considérant un scénario de jeu à travers des vues différentes, chacune correspondant à un objectif d'évaluation particulier.

Un *espace d'évaluation* rassemble ainsi des informations ainsi que des primitives de manipulation de ces dernières, dans le but de produire une évaluation. Un espace est défini par les éléments suivants :

$$Space = \{Kw, I, M, CM\} \quad (1)$$

- *Représentation des connaissances (Kw)*: Comme expliqué dans la section 2.1, différents types de connaissances sont impliqués et chacun est basé sur un paradigme de modélisation spécifique (modélisation de données, base de règles, réseau Bayésien, ...). Pour assurer l'homogénéité de l'évaluation, un espace est associé à un ensemble de modèles de représentation des connaissances similaires.
- *Indicateurs (I)*: Un indicateur est une information associée à un phénomène défini pour observer ses variations périodiques. Ainsi, un indicateur représente une donnée quantitative qui caractérise une situation en évolution (une action ou les conséquences d'une action). L'usage d'indicateurs pour l'évaluation est récurrente dans les travaux sur les jeux sérieux [8] et les EIAH [10].
- *Métriques (M)*: Les métriques représentent les méthodes et les unités de mesure utilisées pour comparer les résultats attendus des acteurs et leurs actions réelles (comportements, décisions, ...). La métrique est utilisée pour quantifier l'indicateur, en d'autre terme, donner une valeur à l'indicateur pour alimenter une évaluation.
- *Modèle de calcul des indicateurs (CM)*: Il y a différents modèles d'évaluation en fonction de l'espace et sa représentation de connaissances. L'évaluation peut porter sur une action, sur une procédure ou bien globalement sur une session d'entraînement (cf. les différents types d'évaluation dans la section 3.3). Le calcul d'indicateurs repose sur un modèle d'évaluation spécifique ainsi que sa métrique. Selon le modèle de représentation des connaissances, la production d'indicateurs peut requérir des méthodes issues de l'IA.

### 3.2 Illustration du concept d'espace d'évaluation dans le projet SIMFOR

La section 2.1 montre le caractère hétérogène de l'évaluation dans un scénario de jeu SIMFOR. Pour faire face à cette hétérogénéité, nous avons défini trois différents *espaces d'évaluation* : Comportemental, Physique et Social.

**Espace Comportemental.** cet espace comprend les actions et les connaissances des acteurs, ainsi que les différentes informations sur les compétences et procédures à apprendre. Ces informations correspondent au modèle du Domaine et au modèle de l'Apprenant dans un système tutoriel intelligent. Le modèle du domaine est statique (composé par exemple d'une ontologie et de plans<sup>5</sup> [8] qui n'évoluent pas au cours du jeu) et est défini par un expert du domaine. Le modèle de l'apprenant est dynamique (évolue durant l'expérience de jeu de l'apprenant) et est alimenté par les actions réalisées par l'apprenant ainsi que les connaissances acquises. L'évaluation nécessite donc des capacités de raisonnement sur les modèles de connaissance composant le modèle du domaine. Dans cette espace nous aurons à évaluer des actions ainsi que des missions, via des

<sup>5</sup> eux mêmes décrits par des réseaux de Petri ou des diagrammes d'activité UML.



indicateurs dédiés. Par exemple pour l'action *téléphoner*, les indicateurs seront la *durée de communication*, mesurée en secondes (métrique), les *informations échangées*, ainsi que l'*acteur appelé*. Ainsi l'évaluation se basera sur ces indicateurs pour attribuer un score à l'action.

**Espace Physique.** L'espace physique représente le monde virtuel, avec les avatars des acteurs, les moyens mis à disposition des joueurs (voitures, téléphone, fax, ...) et l'environnement (bâtiments, routes, arbres...). Ces données peuvent évoluer dans le temps comme l'état d'un sinistre, la position d'un acteur... Le monde virtuel est représenté par l'interface 3D du jeu sérieux. Le traitement et la manipulation de ces données est effectués durant la partie par la simulation du jeu ainsi que par les différents mécanismes du moteur de jeu (interaction, animation, ...). L'évaluation consiste en l'agrégation de ces données via des expressions mathématiques comme par exemple l'indicateur *pertes matérielles et humaines* qui va permettre l'évaluation de la performance globale des joueurs. Cet indicateur peut avoir comme métrique le montant des pertes (en euro par exemple) ainsi que le nombre de blessés et de morts.

**Espace Social.** L'espace social représente l'interaction sociale entre les différents acteurs intégrant ainsi la dimension collective de l'apprentissage. L'espace social est représenté par un graphe qui reprend chaque interaction entre les acteurs et permet un calcul d'une force d'interaction entre chaque acteur ainsi que le degré d'intégration des acteurs du réseau. L'évaluation sera basée sur ces mesures. On retrouve ici les considérations de [2] sur la représentation et l'exploitation des interactions pour l'évaluation des apprenants. Les indicateurs représentés dans cet espace sont le couplage du réseau et la force d'interaction entre les acteurs et sont calculés comme dans [14]. Les échanges (entre les acteurs) courts et fréquents produisent un indice de couplage fort, les échanges longs et rares traduisent un couplage faible. Cet indicateur va déterminer si les acteurs ont bien communiqué avec les bonnes personnes.

### 3.3 L'évaluation dans SIMFOR

L'évaluation des joueurs d'un scénario de jeu, s'appuie sur des capacités d'évaluation mobilisées à des moments variés du jeu à partir d'indicateurs situés dans des espaces différents. Ainsi, trois types d'évaluation intermédiaires ont été définis : en temps réel, individuelle et collective.

**Évaluation en temps réel.** Cette évaluation porte sur un diagnostic évoluant tout au long de la partie (scénario de jeu). L'évaluation en temps réel recouvre l'évaluation d'une action ou d'une mission (*Espace des comportements*). Pour évaluer une action, le calcul d'un score se base sur les indicateurs correspondant. Par exemple, pour l'action "*Fax*" (figure 2), l'évaluation consiste à vérifier qu'aucun interlocuteur n'a été oublié. Pour évaluer une mission, différents indicateurs sont mobilisés (dépendant de la mission à réaliser : temps, pré-conditions, respect de l'ordre des actions, ...). Le score de la mission est la moyenne des différents scores produits par ces indicateurs. Les indicateurs sont calculés de manière régulière (mise à jour dirigée par le temps ou des événements) et sont basés sur des variables qui peuvent elles-mêmes exiger un calcul spécifique.

**Évaluation finale individuelle.** L'évaluation finale individuelle correspond à une évaluation sommative qui évalue et certifie l'apprentissage de l'apprenant à la fin du scénario de jeu. Cette évaluation compare l'état final du modèle de l'apprenant (actions de l'apprenant ainsi que ses connaissances) avec le modèle du domaine et établit un diagnostic sur les compétences acquises et les compétences qui restent à apprendre.

**Évaluation collective.** Utilisant l'*Espace social*, un graphe d'interaction est construit, représentant les différentes communications et les interactions entre les différents acteurs (les apprenants ainsi que les acteurs simulés). Avec ce graphe, nous savons qui a contacté qui, quand, et pour combien de temps, ainsi que les informations échangées. Combinant les informations provenant des différents espaces, nous avons la possibilité de déduire un lien de causalité entre les missions des différents acteurs (l'acteur A a échoué dans sa mission parce que l'acteur B n'a pas envoyé les bonnes informations) et ainsi obtenir une évaluation complète et précise de tous les acteurs.

L'*évaluation globale*, qui ne peut être déterminée qu'à la fin d'une session de jeu, va intégrer l'évaluation à la fois individuelle et collective. La certification d'une compétence ou des connaissances de l'apprenant peut être obtenue par négociation comme dans [15].

## 4 Un SMA pour l'évaluation des apprenants

Les sections précédentes ont présenté en détail les besoins en termes d'information (au sens large) nécessaires pour produire une évaluation complète et précise dans le jeu sérieux SIMFOR. Afin de faire évoluer SIMFOR en un jeu sérieux orienté STI (pour prendre en compte l'aspect suivi et évaluation des apprenants), l'architecture logiciel doit s'adapter aux différents composants logiciels (environnement 3D, simulation multi-agents, simulation des interactions humaines, outils de base de données, etc) tout en respectant l'hétérogénéité des connaissances nécessaires à l'évaluation des apprenants.

L'aspect distribué et hétérogène du problème nous a conduit à opter pour une architecture multi-agents pour la mise en œuvre de la solution. Les SMA permettent une architecture logicielle décentralisée et distribuée d'autant plus que les agents possèdent des capacités sociales [16] (communication entre les agents). Ils peuvent ainsi également intervenir à un niveau d'intégration logicielle [17]. Notre objectif est de combiner les différentes technologies des jeux sérieux (3D, mécanisme de jeu, animation, ...) et un STI (représentation des connaissances, pédagogie, ...) pour obtenir un outil d'apprentissage optimal. Les agents composant le système multi-agents ont pour mission de collecter les données liées aux apprenants, de traiter et d'évaluer ces données et fournir un support aux apprenants. Pour réaliser ces missions, nous avons défini cinq types d'agent :

- *Agent Source de Données (AgSD)*: ce type d'agent est responsable de la collecte des données (couche logicielle).
- *Agent Indicateur (AgI)*: ce type d'agent agrège/transforme les données en indicateurs pour l'évaluation des apprenants. Cette transformation peut nécessiter des techniques IA.

- *Agent Apprenant (AgA)*: ce type d'agent est responsable de la mise à jour du modèle de l'apprenant (et peut informer les AgIs de changements notables).
- *Agent Évaluateur (AgE)*: ce type d'agent calcule et maintient l'évaluation d'un apprenant en exploitant les informations fournies par les AgIs.
- *Agent Pédagogique (AgP)*: cet agent analyse la situation pédagogique (retours des AgEs), et sélectionne la stratégie adaptée pour aider l'apprenant.

En plus des agents définis ci dessus, s'ajoutent les agents du SMA régissant les Personnages Non Joués (PNJ). À chaque PNJ, est associé un *Agent de Jeu* qui reproduit le comportement de l'acteur simulé (rôle incarné). Les Agents de jeux sont basés sur l'architecture BDI [18], afin de disposer de comportements adaptatifs en fonction des actions des joueurs humains, avec quelques modifications pour fournir une plus grande souplesse au concepteur du scénario. Ils relèvent donc d'une approche classique de simulation orientée agent.

Une évaluation globale et collective peut être effectuée par un processus collaboratif (ou processus de négociation) entre les différents agents évaluateurs. Le processus d'évaluation est basé sur une transformation graduelle des données de jeu jusqu'à la production d'une évaluation des apprenants. Chaque étape est assurée par un agent dédié, de manière asynchrone (ie au fur et à mesure du déroulement du jeu). Il s'agit de conserver le caractère hétérogène de ces informations (connaissances, procédure, données brutes) en proposant des méthodes d'évaluation adaptées (techniques dites IA ou simple aggrégation mathématique) ainsi que la variété de leur source. La spécialisation des agents ainsi que leurs interactions à un niveau de langage indépendant des composants logiciels "bas niveau", permet d'assurer une certaine genericité et ainsi faciliter l'évolution du système en cas de modification de processus d'évaluation ou d'ajout de nouveaux indicateurs et/ou sources d'information. Il s'agit également ici de proposer une approche générique permettant tout Jeu Sérieux existant d'évoluer vers un STI.

## 5 Fonctionnement général: Exemple de scénario

Pour illustrer l'évaluation des apprenants, nous reprenons l'exemple de scénario présenté dans la section 2.1 (scénario défini à l'aide d'un expert du domaine), et montrons l'interaction entre les différents agents mettant en oeuvre ce scénario.

L'*agent Apprenant (AgA)* a pour mission de mettre à jour le modèle apprenant (un *agent apprenant* est associé à chaque acteur), pour cela l'AgA s'abonne à l'*agent Source de Données*<sup>6</sup> (AgSD) pour obtenir les actions et les connaissances de l'apprenant.

Lorsque l'apprenant effectue une action (dans l'exemple, l'acteur CODIS appelle un acteur pompier pour l'avertir de l'accident), l'AgSD envoie les données de l'action à l'*agent Apprenant* (id de l'acteur, cible, ...). Une fois que les données sont reçues, l'*agent Apprenant* met à jour le modèle de l'apprenant et informe l'*agent Pédagogique (AgP)* d'un nouvel événement (nouvelle action effectuée). Une fois que l'événement est reçu, l'*AgP* procède à l'accompagnement

<sup>6</sup> connecté à l'environnement 3D de SIMFOR

de l'apprenant. Tout d'abord, l'AgP demande une évaluation de l'apprenant. L'*agent Évaluateur* (AgE) récupère les informations pertinentes du modèle de l'apprenant (niveau, rôle, procédure en cours de réalisation, ...), il fait ensuite une requête à l'*agent Indicateur* (AgI) pour obtenir la valeur des indicateurs concernant l'évaluation. Dans l'exemple, l'apprenant effectue l'action *téléphoner*, donc l'AgE effectue l'évaluation d'une action. Pour calculer l'efficacité de cette action, l'AgE reçoit la valeur de l'indicateur *actionPhoneIndicator*, qui contient les informations nécessaires pour l'évaluation. Cet indicateur est spécifique à l'action *téléphoner* et a pour paramètres le temps d'exécution de l'action et l'échange d'informations lors de l'appel. Une fois l'évaluation calculée, l'AgE envoie le résultat à l'AgP. Ce dernier analyse la situation (évaluation de l'apprenant, modèle du domaine et modèle de l'apprenant) et sélectionne une stratégie de support pour l'apprenant.

Lorsque l'AgE détecte la fin d'une mission, l'AgE procède à l'évaluation de cette dernière (évaluation de mission, section 3.3). Pour cela, l'AgE envoie une requête à l'AgI pour obtenir les indicateurs adéquats. Selon la mission, l'AgI envoie à son tour une requête à l'AgSD pour obtenir les données manquantes pour compléter les indicateurs, par exemple: pour évaluer une des missions du pompier "*intervention sur le sinistre*", l'AgI a besoins des : *pertes matériels et humaines*, et du *temps nécessaire pour maîtriser le sinistre*. Comme mentionné dans la section 3.3, l'AgE calcule un score d'évaluation combinant les différents scores en utilisant les données issues des trois différents espace (espace Comportemental, espace Physique et l'espace Social).

## 6 Conclusion

Avec l'intérêt croissant des jeux sérieux pour la formation, la question de l'évaluation des joueurs (apprenants) est de plus en plus pertinente. Dans cet article, nous avons présenté les caractéristiques ainsi que les enjeux liés à cette évaluation en nous appuyant sur le projet SIMFOR : un jeu sérieux pour la formation des non professionnels à la gestion de crise. Nous avons alors proposé d'ajouter aux jeux sérieux des capacités d'évaluation via le concept d'*espace d'évaluation*. Ensuite, une architecture SMA a été proposée pour mettre en œuvre l'évaluation des apprenants tout en prenant en compte les besoins d'intégration que nous avons identifiés. Cette intégration permet à SIMFOR d'améliorer *a posteriori* ses capacités pédagogiques grâce à un système d'évaluation multi-critères (évaluation des métiers et des compétences différentes : pompier, policier, préfet ...) et distribué (via des *espaces d'évaluation* dédiés), basée sur l'architecture d'un système tutoriel intelligent. Cette approche permet une évaluation individuelle et collective en temps réel ainsi qu'une évaluation à la fin du jeu qui produit un diagnostic des tâches individuelles et collectives.

Le cadre intégratif ayant été défini, nos travaux futurs dans le projet SIMFOR portent sur l'affinement des critères collaboratifs dans le domaine de la gestion de crise, sur la base d'une analyse en temps réel du graphe d'interaction devant permettre un meilleur support pédagogique.

## References

1. Nieborg, D.: America's army: More than a game. In: Proceedings of 35th Annual Conference of the International Simulation And Gaming Association (ISAGA) and Conjoint Conference of SAGSAGA. (2004)
2. Mathieu, P., Panzoli, D., Picault, S.: Serious games et sma - application à un supermarché virtuel. In: JFSMA. (2011) 181–190
3. Michael, D.R., Chen, S.L.: Serious Games: Games That Educate, Train, and Inform. Muska & Lipman/Premier-Trade (2005)
4. Lourdeaux, D.: Réalité virtuelle et formation: conception d'environnements virtuels pédagogiques. PhD thesis, Paris (2001)
5. Buche, C., Bossard, C., Querrec, R., Chevaillier, P.: Pegase: A generic and adaptable intelligent system for virtual reality learning environments. *International Journal of Virtual Reality* **9**(2) (2010) 73
6. Burns, H., Capps, C.: Foundations of intelligent tutoring systems: An introduction. *Foundations of intelligent tutoring systems* (1989) 1–18
7. Johnson, W., Rickel, J., Stiles, R., Munro, A.: Integrating pedagogical agents into virtual environments. *Presence* **7**(6) (1998) 523–546
8. Thomas Benjamin, P., Labat, J.M., Muratet, M., Yessad, A.: How to Evaluate Competencies in Game-Based Learning Systems Automatically? In: 11th International Conference in Intelligent Tutoring Systems. *Lecture Notes in Computer Science*, Springer (June 2012) 168–173
9. Amokrane, K.: Suivi de l'apprenant en environnement virtuel pour la formation à la prévention des risques sur des sites SEVESO. PhD thesis, Université de Technologie Compiègne (December 2010)
10. Woolf, B.P.: Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning. Morgan Kaufmann (2010)
11. Soller, A.: Artificial intelligence methods for modeling and assessing. In: Collaborative Distance Learning Conference CILVR. (2006)
12. Suebnukarn, S., Haddawy, P.: A collaborative intelligent tutoring system for medical problem-based learning. In: Proceedings of the 9th international conference on Intelligent user interfaces, ACM (2004) 14–21
13. de los Angeles Constantino-Gonzalez, M., Suthers, D.D., de los Santos, J.G.E.: Coaching web-based collaborative learning based on problem solution differences and participation. *International Journal of Artificial Intelligence in Education* **13**(2) (2003) 263–299
14. Miller, J.G.: *Living systems*. (1978)
15. Chadli, A., Tranvouez, E., Bendella, F.: Cooperative skills evaluation and participative multiagent simulation for rodent control training. In: Machine and Web Intelligence (ICMWI), 2010 International Conference on, IEEE (2010) 41–47
16. Wooldridge, M., Jennings, N., et al.: Intelligent agents: Theory and practice. *Knowledge engineering review* **10**(2) (1995) 115–152
17. FIPA: Agent software integration specification. Technical Report XC00079B, Foundation for Intelligent Physical Agents (2001)
18. Rao, A., Georgeff, M., et al.: Bdi agents, from theory to practice. In: Proceedings of the first international conference on multi-agent systems (ICMAS-95), San Francisco (1995) 312–319