



**HAL**  
open science

## Conception de simulations multi-agents : vers une modélisation basée sur les observations des comportements réels

Imen Saffar, Arnaud Doniec, Jacques Boonaert, Stéphane Lecoeuche

### ► To cite this version:

Imen Saffar, Arnaud Doniec, Jacques Boonaert, Stéphane Lecoeuche. Conception de simulations multi-agents : vers une modélisation basée sur les observations des comportements réels. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3. hal-00656529

**HAL Id: hal-00656529**

**<https://hal.science/hal-00656529v1>**

Submitted on 17 Jan 2012

**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.

# Conception de simulations multi-agents : vers une modélisation basée sur les observations des comportements réels

I. Saffar<sup>1</sup>

A. Doniec<sup>1</sup>

J. Boonaert<sup>1</sup>

S. Lecoeuche<sup>1</sup>

<sup>1</sup> Université de Lille Nord de France,  
IA, Ecole des Mines de Douai, France

941, rue Charles Bourseul-BP 10838-59508 Douai Cedex  
imen.saffar@mines-douai.fr

## Résumé

*Les simulations multi-agents consistent à utiliser un ensemble d'agents en interaction de manière à reproduire la dynamique et l'évolution des phénomènes que l'on cherche à simuler. Elles sont aujourd'hui une alternative crédible aux simulations classiques basées sur des modèles analytiques, mais leur mise en œuvre reste difficile. Cette tâche est généralement réalisée par le concepteur qui possède une certaine expertise du phénomène à simuler et dispose de données d'observation de ce même phénomène. Dans ce papier, nous proposons une manière originale de traiter l'observation de comportements réels pour la modélisation d'agents simulés en s'appuyant sur des techniques de clustering. La faisabilité de notre approche est démontrée au travers d'un exemple de simulation d'activité humaine.*

## Mots Clef

Simulation Multi-Agents, Fouille de données, Modélisation de comportements, Conception automatique d'agents, Techniques de segmentation.

## Abstract

*The multiagents simulations consist in using a set of interacting agents to reproduce the dynamics and the evolution of the phenomena that we seek to simulate. They are considered now as an alternative to the classical simulations based on analytical models. But, their implementation remains difficult, particularly in terms of behaviors extraction and agents modelling. This task is usually performed by the designer who has some expertises, using extracted observation data from the process. In this paper, we propose an original way to deal with observations to model agent behaviors based on clustering techniques. The feasibility of our approach is demonstrated through a simulation example of human activity.*

## Keywords

Multiagents Simulation, Data Mining, Behavior modelling, Automatic Agent Design, Clustering technique.

## 1 Introduction

Les modélisations et simulations à base d'agents représentent aujourd'hui une alternative crédible aux approches classiques. Les modélisations "agents" permettent de représenter informatiquement des systèmes complexes qui ne peuvent pas être exprimés facilement sous forme d'équations mathématiques [1]. Dans le domaine de la simulation, les approches classiques ont recours à des modèles analytiques reproduisant la dynamique et l'évolution du système à simuler. Les simulations multi-agents consistent à utiliser un ensemble d'agents en interaction de manière à reproduire la dynamique et l'évolution des phénomènes que l'on cherche à simuler. La simulation à partir d'une modélisation "agent" permet de s'affranchir des limites des approches analytiques et de créer un paradigme puissant permettant d'aborder l'étude et l'analyse de systèmes réels complexes [2].

La simulation multi-agents part généralement d'une approche descendante consistant à identifier d'une part les entités composant le phénomène réel à reproduire et d'autre part, les relations entre ces entités. Chaque entité est ensuite modélisée sous la forme d'un agent, généralement en faisant quelques hypothèses simplificatrices. Comme toutes approches de modélisation, les modélisations "agents" restent donc une approximation ne permettant pas de prendre en considération tous les détails du ou des phénomènes réels à simuler. Dans certains cas, cette approximation peut conduire à des écarts importants entre le réel et le simulé induisant ainsi des interprétations, des analyses et des résultats erronés. Il existe donc une vraie problématique autour de l'identification des entités essentielles à modéliser et de la détermination du niveau de détail de leur modélisation.

Du reste, ces approches exigent de gros efforts en matière de modélisation et de programmation. Le concepteur de la simulation doit disposer d'une certaine expertise du système à simuler pour pouvoir le modéliser sous forme d'agents. Pour de nombreuses applications, l'observation

du système à simuler et le recueil de données réelles sont possibles mais dans la plupart des cas ils ne servent qu'à valider la simulation.

Par ailleurs, de nombreux travaux se sont focalisés sur l'identification de comportements en termes de détection d'évènements dans une séquence vidéo [3]. Ces approches se basent, généralement, sur des modèles d'évènements prédéfinis, ou appris automatiquement. Récemment, des études ont proposé de nouvelles techniques capables de modéliser des évènements longs et complexes, par détection, classification et tracking, impliquant plusieurs individus effectuant plusieurs actions [4]. Il existe, également, d'autres approches qui détectent les évènements utilisant des techniques de segmentation et de classification sans connaissances a priori [5]. L'objectif principal de ces recherches est d'aboutir à la reconnaissance des activités relatives aux comportements de personnes dans un cadre particulier.

Dans cet article, nous formulons l'idée que les données réelles peuvent être utilisées lors de la phase de la modélisation du système multi-agents en ayant recours à des techniques de fouille de données. L'intégration des méthodes de fouille de données aux systèmes multi-agents est un axe de recherche assez récent et actuellement en plein essor [6]. Dans ce papier, nous proposons une méthodologie originale basée sur l'utilisation de techniques de segmentation pour la conception des agents. Nous montrons que l'utilisation de telles techniques peut faciliter non seulement l'identification des entités à modéliser en tant qu'agents mais également la caractérisation de leurs comportements. Dans la section 2, afin de positionner notre contribution, nous exposons des approches relatives à la modélisation multi-agents et nous présentons quelques travaux significatifs faisant état de l'intégration de techniques de fouille de données dans les simulations. Dans la section 3, nous exposons notre démarche qui vise à s'appuyer sur des techniques de fouille de données pour faciliter la phase de modélisation lors de la construction d'une simulation multi-agents. Cette proposition est illustrée en sections 4 et 5 au travers une expérimentation visant à simuler différentes activités humaines au sein d'un hall de bâtiment. Nous exposons en particulier la méthodologie qui nous a permis d'extraire les comportements d'agents à partir de séquences vidéo des activités à simuler. Dans la section 6, nous discutons du travail effectué et des travaux futurs.

## **2 Simulations multi-agents et fouille de données**

### **2.1 Simulations basées sur les agents**

Un processus de simulation consiste à reproduire le comportement dynamique d'un phénomène réel en utilisant des modèles et différentes séquences d'entrée. Un modèle est une construction simplifiée de la réalité qui peut être utilisée pour reproduire, selon un certain niveau de détail, le comportement du phénomène à simuler. Dans la plupart

des outils de simulation, les comportements dynamiques des phénomènes simulés sont exprimés en utilisant des approches mathématiques, telles que les équations différentielles, par exemple. Ces approches tirent parti de l'universalité des mathématiques, mais atteignent rapidement leurs limites pour des phénomènes complexes impliquant de nombreuses interactions.

Les simulations multi-agents offrent une réponse aux limites de ces approches classiques de simulation. Chaque entité du système à simuler est modélisée en agent : une entité logicielle évoluant dans un environnement et exhibant (d'une façon autonome) un comportement approprié et motivé par ses propres croyances et désirs. L'idée centrale des simulations multi-agents repose sur le concept d'émergence : la dynamique globale du système et les phénomènes à reproduire sont le résultat de comportements individuels et les interactions entre les agents.

La mise en œuvre d'une modélisation "agent" n'étant pas triviale, de nombreux efforts ont été fournis pour proposer des méthodologies facilitant la conception des systèmes multi-agents. Nous trouvons dans la dernière décennie, de nombreuses méthodologies de modélisation agent ; pour ne citer que quelques-unes : Gaia [7], O-MAE [8], INGENIAS [9], ADELFE [10]. Tous ces travaux donnent un cadre général pour concevoir un système multi-agents, y compris : des modèles prédéfinis d'agents abstraits, des schémas d'organisation, des protocoles de communication, des actes de langage, etc. Les directives associées à ces méthodologies permettent de raffiner itérativement ces modèles abstraits en des modèles concrets, allant même jusqu'à l'implémentation des agents.

Bien que ces méthodologies puissent être appliquées à la conception des simulations multi-agents, elles ne sont pas bien adaptées à la construction de modèles fidèles pour la simulation, principalement parce qu'elles ont un point de vue global et fonctionnel du système multi-agents. Des approches spécifiques pour la conception des simulations multi-agents ont été introduites offrant une abstraction plus appropriée et des architectures pour exprimer les comportements et les interactions des agents. Par exemple, dans IODA [11], les auteurs proposent de construire une simulation reposant sur les différentes interactions entre les agents. Il est clair qu'une telle approche facilite la modélisation et la mise en œuvre des agents dans des simulations multi-agents, mais ne dispense pas le concepteur d'une étape ardue : l'observation et l'analyse des phénomènes à simuler.

Cette étape est habituellement traitée par des experts du domaine (comme les sociologues ou psychologues) pour obtenir les modèles abstraits dont les implémentations seront traitées par les développeurs. Pour l'instant, cette étape est très coûteuse en temps, car elle nécessite des études pouvant reposer par exemple sur le traitement de données statistiques, l'analyse de séquences vidéo, le dépouillement de questionnaires, l'interview d'experts,... En plus d'être fastidieuse, cette tâche est très subjective et les modèles ré-

sultants peuvent varier d'un expert à l'autre. Par ailleurs, les méthodes présentées ci-dessus, n'écartent pas les problèmes de compréhension entre experts et développeurs. Une façon d'éviter tous ces désagréments est d'automatiser autant que possible l'étape d'observation et d'analyse. Pour ce faire, les outils de fouille de données semblent être une bonne approche pour fournir quelques règles et des résultats facilitant la conception des simulations multi-agents.

## 2.2 Intégration de la fouille de données dans les simulations multi-agents

L'objectif de la fouille de données (*data mining*) est d'extraire, à partir d'un volume important de données, des connaissances ou des informations, susceptibles d'être pertinentes, et jusque-là inconnues ou cachées. Une difficulté bien connue dans les simulations multi-agents consiste à analyser correctement le comportement émergent qui peut contenir des informations implicites et cachées. L'application de techniques d'exploration de données peut révéler les tendances et donc des connaissances non triviales à partir des données de la simulation. L'analyse de ces résultats permet de corriger et d'améliorer la simulation.

À notre connaissance, peu d'études ont mis en œuvre les méthodes de fouille de données dans les simulations multi-agents. Dans [12], les auteurs ont classé l'usage de la fouille de données dans les simulations basées sur l'agent en deux catégories :

- application endogène : lorsque la fouille de données est appliquée pour modifier et améliorer les comportements des agents en utilisant les données, issues des précédentes exécutions de la simulation [6] ;
- application exogène : lorsque la fouille de données est appliquée sur le résultat final de la simulation afin d'analyser et révéler des modèles pertinents facilitant son interprétation [13].

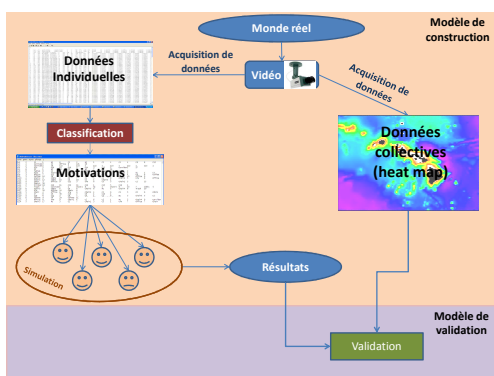


FIGURE 1 – Modèle de l'approche proposée

Dans ce papier, une nouvelle catégorie est proposée, où les techniques de segmentation (clustering) sont utilisées pour aider à concevoir des agents. Nous développons cette proposition dans la prochaine section.

## 3 Modélisation de comportement d'agent conduit par l'observation et la segmentation

Comme nous l'avons vu précédemment, l'intégration de la fouille de données aux simulations multi-agents est essentiellement restreinte aux phases d'amélioration et de validation de modélisations existantes. Nous aimerions aller plus loin et utiliser les techniques de fouille de données pour faciliter la modélisation du comportement des agents.

- Sous l'hypothèse de disposer de données d'observation du ou des phénomènes à simuler, les techniques de fouille de données peuvent en extraire des informations utiles pour :
- l'identification des différentes entités composant le ou les phénomènes à simuler, en fonction de leur importance ;
  - la conception des différentes fonctions des agents ;
  - la répartition de ces agents au sein de la simulation.

Les techniques de segmentation semblent être un bon candidat pour l'obtention de ces trois items. Parmi les techniques utilisées pour l'exploration de données, les techniques de segmentation (classification non supervisée) sont utilisées pour regrouper les données non étiquetées en fonction de leurs similarités. Ainsi, à partir d'un ensemble de données, ces techniques sont capables de créer des clusters (groupes) de telle sorte que les données appartenant au même groupe ont un potentiel de grande similarité, et les données appartenant à deux groupes différents ont des similarités très faibles.

Ainsi, les clusters, déterminés à partir des données d'observations réelles, peuvent être associés à des composantes de phénomènes à modéliser. En d'autres termes, plutôt que de se baser sur sa propre expérience pour identifier les entités essentielles à modéliser, le concepteur peut s'appuyer sur les clusters ainsi obtenus.

A chaque cluster, sont attachées diverses mesures telles que l'indice, le centre (ou prototype du groupe), la cardinalité, la variance, la densité ... Certaines de ces valeurs peuvent être interprétées dans le contexte de la modélisation des agents :

- l'indice d'un cluster identifie un type d'agent particulier ;
- le prototype du groupe fournit un modèle de comportement d'agent ;
- la cardinalité (les distributions des données au sein de chaque cluster) donne la répartition des agents par type au sein de la simulation.

Ces différents points seront illustrés de manière plus précise au travers de l'exemple développé dans la section 4.

Dans un processus de conception de simulations multi-agents, nous distinguons deux phases principales, où les techniques de clustering peuvent être utilisées (voir figure 1) :

- la construction des modèles d'agent.

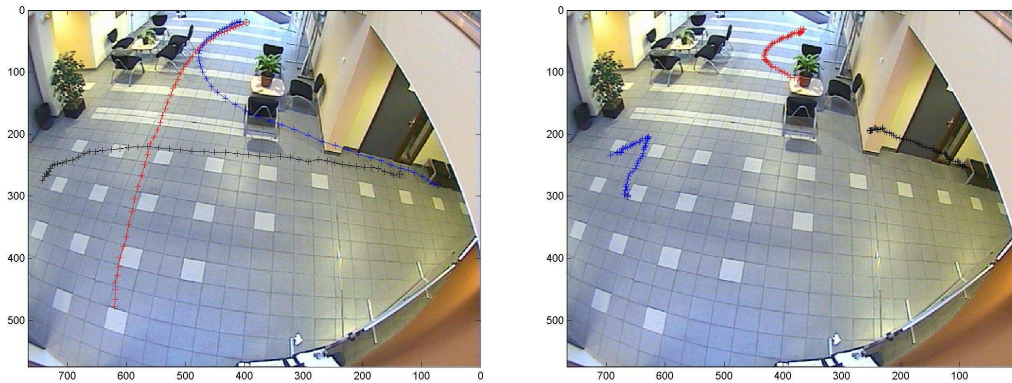


FIGURE 2 – Exemples de trajectoires moyennes

– leurs validations.

La première étape permet principalement l'utilisation des données individuelles sur lesquelles il peut être appliqué des techniques de clustering tel que discuté ci-dessus. Ces données peuvent ensuite contribuer à définir la représentation sous forme d'agents de toutes les personnes observées dans un environnement spécifique.

A l'opposé, la deuxième étape nécessite en particulier des données collectives. Ces données peuvent être par exemple une carte de chaleur mettant en valeur les régions les plus fréquentées et les plus désertes, ou encore des mesures caractérisant les flux (débit)... En les comparant avec les résultats de simulation, ces données sont classiquement utilisées dans l'étape de validation.

Selon le domaine d'application, l'étape de validation peut également impliquer un expert dans le domaine. Basé sur son expérience et ses connaissances, il valide ou non le réalisme et la pertinence de la simulation obtenue.

Notre approche est relativement générique visant, en particulier, la simulation de comportements humains spatialisés dans un environnement connu. Ces derniers peuvent être caractérisables en termes de trajectographies, de postures, de gestuelles, etc.

## 4 L'implémentation de l'approche proposée

L'approche décrite ci-dessus a été mise en œuvre expérimentalement sur un cas d'étude simple impliquant la simulation d'activités humaines dans un hall d'un bâtiment universitaire. Il se compose de trois entrées, deux salles de classe, plusieurs bureaux, un escalier, un espace détente et une machine à café.

L'objectif est de reproduire, par simulation, les activités se déroulant au sein de cet espace : la pause pour les étudiants et le personnel, les passages dans le hall, entrée / sortie des classes, etc .. Une caméra placée au-dessus du hall d'entrée nous permet d'avoir accès à des séquences vidéo. Le but de cette simulation est de reproduire la densité d'occupation dans le hall.

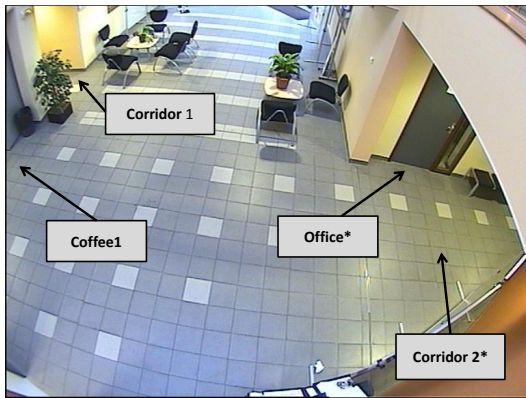
La première étape de notre approche est de traiter les sé-

quences vidéo en extrayant les trajectoires individuelles qui serviront d'entrée pour l'algorithme de clustering. À partir du flux vidéo, l'algorithme de traitement extrait les formes en mouvement de l'image. Cette première opération est effectuée sur la base d'une soustraction de fond. Pour chaque blob extrait et identifié (un ensemble de pixels représentant un corps humain), un vecteur de trajectoire est alors construit, à partir du déplacement de la personne. Chaque trajectoire est caractérisée par un index unique et se compose d'une liste de points estampillés par un numéro de trames et différents paramètres décrivant le blob (taille, forme, vecteur d'orientation, etc).

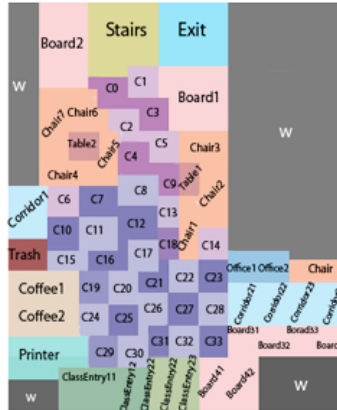
Les trajectoires obtenues sont de taille variable. Les plus courtes des trajectoires qui correspondent à un déplacement non pertinent (par rapport à notre étude) ont été éliminées. Nous sélectionnons les trajectoires qui sont composées d'au moins 30 points. Pour être en mesure d'effectuer le clustering sur l'ensemble de ces trajectoires, nous avons été amenés à appliquer une normalisation en considérant la taille moyenne des trajectoires égale à 50 points.

Enfin, comme la base de données initiale découle d'une analyse vidéo d'une journée, nous avons 682 trajectoires représentées dans une matrice où chaque ligne contient une observation (un chemin) composée de 100 attributs : 50 éléments de l'axe des abscisses et de 50 éléments de l'axe des ordonnées, alternativement. Cette matrice est ensuite traitée par l'algorithme choisi pour cette étude : EM (Espérance-Maximisation) [14].

Le choix de l'algorithme EM reflète le fait que celui-ci est considéré, dans la littérature, comme étant simple, efficace et capable d'estimer les modèles dans le cas de données bruitées et / ou manquantes. Dans notre cas, l'existence de certaines données erronées est principalement due aux limites des algorithmes de traitement vidéo en présence de problèmes d'occultation ou de séquences incomplètes. L'algorithme EM est une technique de clustering qui cherche un modèle statistique représentant un modèle de mélange, qui décrit le mieux le jeu de données à segmenter, estimant ses paramètres grâce à une optimisation du maximum de vraisemblance. Il s'en suit un processus itératif qui cherche à maximiser la vraisemblance du modèle



(a) L'environnement réel



(b) Grille de l'environnement de la simulation

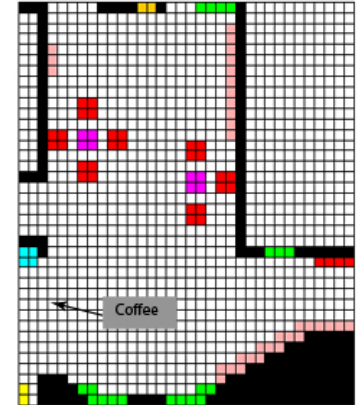


FIGURE 3 – L'espace des motivations correspondant aux zones d'intérêt

estimé jusqu'à convergence vers l'optimum global dans le meilleur des cas.

Chaque itération effectue deux étapes :

- Étape E (Espérance) qui calcule les probabilités a posteriori étant donné le cas à classer. Cela revient à déterminer l'espérance conditionnelle de la vraisemblance à l'égard du modèle existant.
- Etape M (maximisation) qui, comme son nom l'indique, maximise l'espérance de la vraisemblance et estime les paramètres du modèle.

Pour déterminer le nombre de groupes optimal par rapport aux données, la technique de validation croisée (VC) [15] a été appliquée. Cette méthode statistique permet l'évaluation de l'apprentissage effectué en divisant l'ensemble des données en deux parties. Une partie pour l'apprentissage du modèle et une autre pour la validation du modèle. Ce processus fait en sorte que toutes les données devraient être validées.

Considérant notre base de données, le résultat de la validation croisée est constitué de 24 groupes. Ils décrivent des trajectoires moyennes ou des parcours prototypes utilisés pour distinguer les différents profils des personnes dans le hall. Six des trajectoires moyennes<sup>1</sup> sont tracées dans la figure 2. Pour chaque cluster, le cardinal (i.e. le nombre de trajectoires appartenant à ce groupe) est connu et aidera à définir le pourcentage d'agents pour chaque profil connu.

Pour obtenir les comportements des agents à partir de ces profils, nous avons traduit chaque trajectoire moyenne, en une liste de motivations définissant les différentes régions au sein du hall.

Ces motivations résultantes représentent les emplacements successifs auxquels l'agent doit se rendre jusqu'à la destination finale, reproduisant le plus fidèlement possible la trajectoire moyenne. Cela nécessite le partitionnement de

1. Afin d'expliquer pourquoi les trajectoires semblent traverser les meubles, il est important de noter que la trajectoire suit le centre de gravité du corps.

l'espace en différentes zones. Ayant à disposition le plan du hall et connaissant les secteurs de fréquentation, nous sommes capables de reconnaître les régions d'intérêt (RI) (café, couloir, bureau ...). Nous associons par la suite, à chaque région un label qui la caractérise. La figure 3 représente le fond de la séquence vidéo annoté de quelques labels. Pour simplifier l'implémentation, l'environnement a été discrétisé sous forme d'une grille avec un pas relativement fin par rapport à l'occupation moyenne de l'espace par un individu. Par ailleurs, nous avons réalisé la correspondance entre cette grille et le plan caméra utilisé dans les séquences vidéo (voir figure 3). Nous attribuons les étiquettes préalablement sélectionnées à chaque ensemble de cellules correspondantes de la grille (comme représenté sur le côté droit de la figure 3). Par exemple, le label "café" est attribué à l'ensemble des cellules où la machine à café est située. Pour les emplacements des RI dans l'espace de la grille, une transformation a été définie en fonction de la connaissance de la carte du bâtiment et du champ visuel de la caméra. Les trajectoires sont donc sémantiquement interprétées sur la base de ces étiquettes et rapportées dans un fichier texte.

Le fichier texte est organisé comme un tableau, où chaque ligne définit un cluster (trajectoire moyenne) avec la liste des motivations. Un échantillon est montré dans le tableau I. La première colonne représente l'identifiant du cluster, la seconde est le pourcentage d'individus appartenant à ce cluster. Puis, dans la dernière colonne : les motivations (la prochaine case) et le temps passé sur celles-ci. Les six lignes dans le tableau I correspondent aux six trajectoires de la figure 2. Le fichier résultant est utilisé comme paramètre d'entrée de la simulation.

## 5 La simulation multi-agents et les résultats

Concernant la simulation multi-agents, nous avons choisi, parmi un certain nombre d'outils de simulation agent, la

Cluster	Distribution	{(Motivation, Temps)}
CL1	3	(Office, 7),(C21, 2),(UTW, 7),(C5, 4),(C6, 5),(C7, 4),(C8, 6),(Coffee, 15)
CL2	5	(Chair, 4),(Office, 6),(UTW, 6),(Chair1, 4),(C2, 6),(C1, 5),(C0, 6),(Cbis0, 3),(Stairs, 10)
CL3	3	(Exit, 1),(Stairs, 5),(C0bis, 5),(C0, 5),(C1, 2),(Chair5, 2),(NC, 4),(C3, 6),(C7, 5),(C8, 1),(C12, 3),(C16, 4),(C20, 5)
CL4	5	(UTW, 20),(Office, 24),(Chair, 6)
CL5	6	(Table1, 6),(C2, 2),(C1, 13),(C0bis, 9),(Board2, 17),(Exit, 2)
CL6	6	(Coffee, 9),(C8, 29),(Coffee, 11)

TABLE 1 – L'organisation du fichier de motivations

plate-forme Repast Simphony<sup>2</sup>. Notre choix s'explique par le fait que c'est un outil libre et ouvert pour la modélisation des simulations multi-agents offrant une riche variété de fonctions. Utilisant cet outil de simulation, l'agencement du hall d'entrée a été construit comme décrit dans la section 4 et illustré dans la figure 3 (b).

Notre modèle est structuré autour de deux concepts essentiels de Repast :

- Le contexte qui est une sorte de conteneur qui inclut tous les composants du modèle, y compris les agents.
- Les projections qui peuvent créer une structure rattachée à un contexte, permettant de définir les relations entre les agents appartenant à ce contexte. Ils définissent, en général, la nature de l'environnement dans lequel les agents opèrent (grille ou espace métrique continu ...).

La structure de la simulation s'appuie sur trois éléments principaux :

- une classe pour le contexte général, qui gère le lancement de la simulation ;
- une classe définissant les agents ;
- une classe qui représente le gestionnaire de l'environnement.

Cette dernière est conçue pour maintenir une certaine généralité en faisant abstraction de la nature de l'environnement (discret ou continu). Toutefois, les agents ont accès à leurs motivations à partir d'un fichier contenant les résultats du clustering traduits en objectifs.

La génération d'un agent dans la simulation est faite par le gestionnaire de la simulation après avoir défini le type de l'environnement. Le nombre et le type d'agents créés proviennent du fichier contenant les motivations et le pourcentage de chaque type obtenu à l'issue de l'étape de clustering. Nous avons choisi de modéliser l'environnement comme une grille pour faciliter le déplacement des agents. Chaque motivation de l'agent est traduite en coordonnées d'une cellule au sein de la région correspondante de la grille de simulation (l'un des RI comme représenté sur le côté droit de la figure 3). Le gestionnaire de l'environnement calcule le plus court chemin entre la position actuelle d'un agent et sa prochaine destination. Il est à noter que l'agent choisit sa cellule de destination de façon aléatoire dans la région correspondante. Ceci permet d'obtenir une certaine variété<sup>3</sup> dans le déplacement des agents.

2. [http://repast.sourceforge.net/repast\\_simphony.html](http://repast.sourceforge.net/repast_simphony.html)

3. A terme, il est prévu d'utiliser la variance de chaque classe pour obtenir une variabilité des comportements plus fine

Pour terminer, le gestionnaire de simulation construit incrémentalement une carte de chaleur résultant des déplacements et des interactions (évitements de collision principalement) des agents. Bleu au début, les cellules de la carte virent graduellement au rouge chaque fois qu'un agent les visite. Cette carte permet de visualiser les zones les plus fréquentées du hall. Les résultats obtenus à partir de la simulation sont représentés au niveau de la carte de chaleur dans la figure 4.c. La figure 4.b présente la carte de chaleur des données collectives évoquées dans la section 3. Ces données décrivent toutes les trajectoires parcourues dans le hall d'entrée du département pendant la même période traitée dans la simulation.

Nous constatons que les zones B, C, E, F sont similaires dans les deux cartes (réalité et simulation), et correspondent aux RI de l'environnement. En effet, nous retrouvons l'espace café (E), la sortie des classes (F) et l'entrée principale (B) comme étant les régions les plus convoitées dans la simulation. Cette constatation semble naturelle pour valider cet aspect de la simulation. Par ailleurs, la carte 4.b vient corroborer cette observation, démontrant des concentrations au niveau de ces mêmes régions. De même pour la région centrale (C), qui représente une zone de croisement entre le flux orienté nord-sud et celui orienté est-ouest, se retrouve dans les deux cartes.

En revanche, les zones A et D diffèrent. Cela peut s'expliquer par la conjonction de plusieurs constatations. D'une part, à l'issue du clustering, bon nombre des trajectoires passant par les zones A et D se sont retrouvées regroupées avec des trajectoires, certes proches, mais plutôt situées dans l'espace entre les zones B et C. D'autre part, du point de vue de la simulation les algorithmes utilisés pour les évitements d'obstacles ont également contribué à éloigner les agents des zones A et D très obstruées par des chaises. Ces différences peuvent s'interpréter comme une simplification du monde réel (qui est un des buts de la simulation) plutôt que comme un défaut de modélisation. Notons qu'il aurait été plus gênant de faire apparaître en simulation des points chauds qui n'existent pas en réalité, chose que nous n'avons pas constatée à l'issue de nos expérimentations.

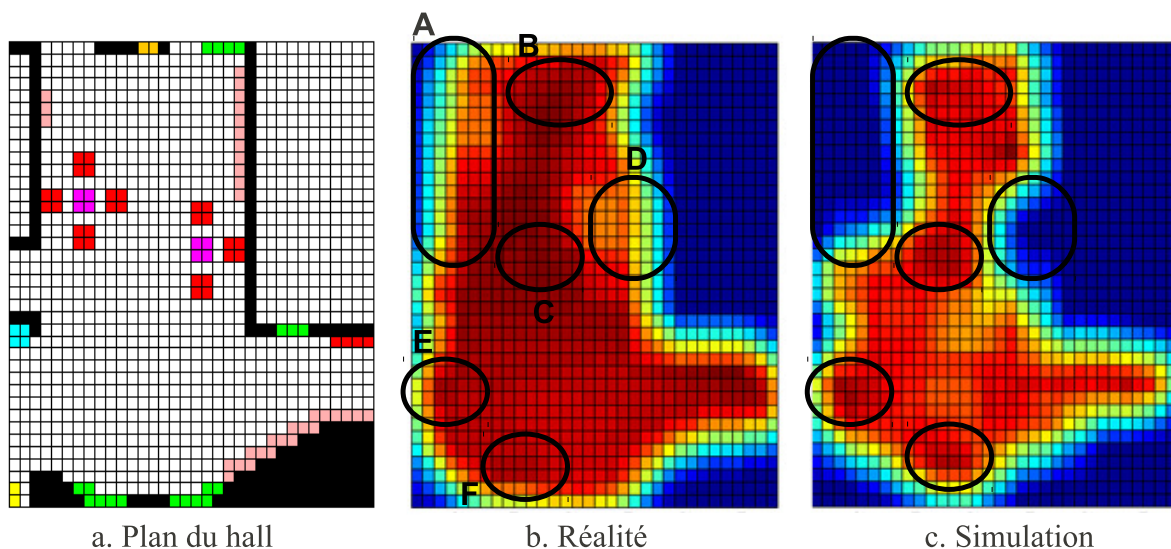


FIGURE 4 – Comparaison des cartes des points chauds

## 6 Perspectives

### 6.1 Perspectives d'ordre général

L'algorithme EM a été appliqué pour extraire des trajectoires moyennes. À partir de ces prototypes de déplacement, les types principaux des agents, leurs comportements et leurs motivations ont été isolés. Parmi les trajectoires moyennes obtenues à partir de l'algorithme EM, nous remarquons que certaines sont atypiques et peuvent être grossièrement considérées comme du bruit, causé par des aberrations éventuelles du traitement d'image, ou comme un comportement irrégulier dépourvu de motivations ou de but précis. Nous avons tenu compte, jusqu'à présent, de ces trajectoires dans notre simulation. Il doit être remarqué, cependant, qu'un moyen d'améliorer la simulation est de traiter ces trajectoires atypiques différemment afin d'en obtenir une modélisation plus fine. En ce qui concerne la simulation, la génération d'un agent dépend fortement du cluster auquel il appartient, et donc de la liste des motivations. Si nous générons trois agents à partir de trois groupes différents, nous avons à mettre en œuvre trois agents de types différents. Il serait intéressant d'explorer une méthode qui, à partir des données, rassemble tous les groupes dans une structure de graphe définissant un agent générique unique basée sur une chaîne de décision. Chaque chemin dans cette structure (composé d'une succession de nœuds (états)/arc (transitions)) définirait, ainsi, le comportement d'un agent spécifique.

### 6.2 Perspectives applicatives

La vocation souhaitée de ce travail est de reproduire le plus fidèlement possible des environnements connus en simu-

lation. Notamment, un espace de vente de type magasin, une galerie commerciale, etc. Nous envisageons d'utiliser l'outil de simulation obtenu sur des configurations d'environnements virtuels pour pouvoir faire de la prédiction de comportements. Nous souhaitons, à partir de séquences vidéo d'un magasin de vêtement par exemple, modéliser les agents en réutilisant la méthode présentée dans cet article. Les techniques de clustering permettraient d'obtenir une segmentation de la population en regroupant les clients ayant des trajectoires de déplacement semblables. Chaque groupe serait alors associé à la trajectoire moyenne caractérisant un type de client : le client pressé, le client adepte de lèche-vitrine, le client indécis, etc. L'analyse de ces trajectoires moyennes fournirait les caractéristiques de chaque type de client : le temps passé par rayon, le temps passé dans les cabines d'essayage, etc. L'utilisation de séquences vidéo datées (différentes périodes pendant la journée, différents jours de la semaine ou le mois ou la saison) pourrait également fournir des caractéristiques plus spécifiques en fonction du contexte d'observation. Connaissant le type, le nombre et les caractéristiques de plusieurs comportements de clients extraits à partir d'observations réelles, des simulations plus précises pourraient être construites.

En terme de validation, en plus des cartes de chaleur nous prévoyons d'utiliser d'autres indicateurs de comparaison tels que la liste de produits préférés, le temps passé dans chaque rayon, le temps écoulé dans la queue devant les caisses,...

## 7 Conclusions

Dans cet article, nous définissons une méthodologie basée sur l'intégration des techniques de clustering dans la conception de simulations multi-agents. Dans le contexte de la simulation d'activités humaines, l'objectif de notre



approche est de simplifier le travail du concepteur en automatisant certaines tâches telles que l'identification d'entités et l'extraction de leur modèle de comportement. La méthodologie a été implémentée et testée avec une expérimentation réelle qui a consisté à modéliser les interactions dans le hall d'un bâtiment. Cet espace a été équipé d'une caméra pour observer les comportements des individus. Les observations réelles sous forme de séquences vidéo ont été traitées en utilisant un algorithme de clustering afin d'en extraire les modèles des agents. La simulation résultante est apparue réaliste au vu d'une comparaison entre les densités d'occupation de l'espace en réel et en simulé.

## Remerciements

Nous tenons à remercier la Région Nord Pas De Calais pour le soutien qu'elle apporte à ce projet.

## Références

- [1] H. V. D. Parunak, R. Savit, and R. L. Riolo, "Agent-based modeling vs. equation-based modeling : A case study and users' guide," in *MABS'98*, 1998, pp. 10–25.
- [2] V. Grimm and S. F. Railsback, *Individual-based Modeling and Ecology*, 2005.
- [3] Y. Benabbas, N. Ihaddadene, C. Djeraba, J. Boonaert, and A. Chaari, "Abnormal behaviors modeling and detection. some of the approaches developed in the canada project," juin 2011.
- [4] A. Hakeem and M. Shah, "Learning, detection and representation of multi-agent events in videos," *Artif. Intell.*, vol. 171, no. 8-9, pp. 586–605, 2007.
- [5] J. Nascimento, M. Figueiredo, and J. Marques, "Segmentation and classification of human activities," pp. 79–86, 2005.
- [6] P. Fournier-Viger, R. Nkambou, U. Faghihi, and E. M. Nguifo, *Data Mining and Multiagent Integration*. Springer, 2009, ch. Mining Temporal Patterns to Improve Agents Behavior : Two Case Studies, pp. 77–92.
- [7] M. Wooldridge, N. R. Jennings, and D. Kinny, "The Gaia methodology for agent-oriented analysis and design," *Autonomous Agents and Multi-Agent Systems*, vol. 3, no. 3, pp. 285–312, 2000.
- [8] J. C. García-Ojeda, S. A. DeLoach, Robby, W. H. Oyenan, and J. Valenzuela, "O-MaSE : A customizable approach to developing multiagent development processes," in *proc. of AOSE*, 2007, pp. 1–15.
- [9] J. Pavón and J. J. Gómez-Sanz, "Agent oriented software engineering with INGENIAS," in *CEEMAS*, 2003, pp. 394–403.
- [10] C. Bernon, V. Camps, M. P. Gleizes, and G. Picard, "Designing agents' behaviors and interactions within the framework of ADELFE methodology," in *ESAW*, 2003, pp. 311–327.
- [11] Y. Kubera, P. Mathieu, and P. Sébastien, "IODA : An interaction-oriented approach for multi-agent based simulations," *Journal of Autonomous Agents and Multi-Agent Systems*, pp. 303–343, 23 (3) 2011.
- [12] M. Remondino and G. Correndo, "MABS validation through repeated execution and data mining analysis," *International Journal of Simulation : Systems, Science & Technology*, vol. 7, no. 6, 2006.
- [13] O. Baqueiro, Y. J. Wang, P. McBurney, and F. Coenen, "Integrating data mining and agent based modeling and simulation," in *ICDM*, 2009, pp. 220–231.
- [14] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
- [15] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining : Practical Machine Learning Tools and Techniques*, 3rd ed. Burlington, MA : Morgan Kaufmann, 2011.