

De l'utilisation de l'Analyse de Données Symboliques dans les Systèmes multi-agents

Flavien Balbo*, Julien Saunier*, Edwin Diday**, Suzanne Pinson*

*LAMSADE, Université Paris-Dauphine

**LISE-CEREMADE, Université Paris-Dauphine,

Place du Maréchal de Lattre de Tassigny, Paris Cedex 16

{balbo,saunier,pinson}@lamsade.dauphine.fr

diday@ceremade.dauphine.fr

Résumé. L'exploitation en temps réel de connaissances complexes est un défi dans de nombreux domaines, tels que le web sémantique, la simulation ou les systèmes multi-agents (SMA). Dans le paradigme multi-agents, des travaux récents montrent que les communications multi-parties (CMP) offrent des opportunités intéressantes en termes de réalisme des communications, diffusion des connaissances et sémantique des actes de langage. Cependant, ces travaux se heurtent à la difficulté de mise en oeuvre des CMP, pour lesquelles les supports de communications classiques sont insuffisants. Dans cet article, nous proposons d'utiliser le formalisme de l'Analyse de Données Symboliques (ADS) pour modéliser les informations et les besoins des agents. Nous appuyons le routage des messages sur cette modélisation dans le cadre d'un environnement de communication pour les systèmes multi-agents. Afin d'illustrer notre propos, nous utiliserons l'exemple de la gestion des communications dans un poste d'appels d'urgence. Nous présentons ensuite notre retour d'expérience, et discutons les perspectives ouvertes par la fertilisation croisée de l'ADS et des SMA.

1 Introduction

L'exploitation en temps réel de connaissances complexes est un défi dans le domaine des systèmes multi-agents (SMA) (Pujol et al., 2002; Zha et al., 2003). Les agents négocient et raisonnent à partir de connaissances (Pujol et al., 2002); ils doivent également construire des connaissances communes au niveau multi-agents, par exemple dans les systèmes de réputation (Zha et al., 2003). Plus spécifiquement, le problème principal lié à la gestion des connaissances au niveau multi-agents est celui de la distribution des informations entre agents. De façon à partager les informations, la majorité des travaux utilise des *interactions directes*, fondées sur la communication adressée point-à-point : un émetteur envoie un message à un récepteur localisé par son adresse. Un certain nombre de travaux proposent d'ajouter au niveau de l'infrastructure un environnement logique pour faciliter les échanges d'information. Cette famille de modèles est celle des *interactions indirectes*, qui repose sur un partage de l'information (voir par exemple (Omicini et Zambonelli, 1999)). Ainsi, au lieu de stocker l'information dans

les agents, celle-ci est externalisée, ce qui permet à chaque agent d'émettre des informations accessibles à tous, et de récupérer les données qui l'intéressent.

Le point commun entre ces deux modes d'interaction est la notion de routage des informations : quel agent doit obtenir quelle information dans quel contexte ? La résolution de ce problème nécessite la gestion de méta-informations sur le système multi-agents, à savoir à la fois les agents, les informations et le contexte. Dans cet article, nous montrons comment ces méta-informations sont gérées et exploitées grâce à une formalisation issue de l'analyse de données symboliques. Etant dédié au domaine des SMA et plus particulièrement à la gestion des communications, la nature des connaissances est connue ; il est donc possible de tirer parti de cette spécificité pour optimiser son traitement. Ensuite, puisque ces connaissances sont utilisées pour réaliser le routage des informations, il y a une contrainte d'efficacité : les agents doivent recevoir les informations dans un temps raisonnable.

En section 2, nous définissons plus précisément la problématique multi-agents, et justifions l'utilisation de l'analyse de données symboliques. En section 3, nous présentons le modèle d'environnement de communication pour le routage des informations et la structuration des connaissances. En section 4, nous discutons les intérêts et limites de notre modèle, et nous proposons des pistes de recherche pour une meilleure exploitation de l'ADS dans les systèmes multi-agents.

2 Motivation

2.1 Problématique des communications multi-parties

Dans le paradigme multi-agents, des travaux récents montrent que les communications multi-parties (CMP) (Kumar et al., 2000; Saunier et Balbo, 2007) permettent de tirer parti de la complexité des communications humaines. En particulier, les agents peuvent surveiller le système (Gutnik, 2005), propager les informations (Busetta et al., 2002) et montrer des comportements opportunistes (Kamali et al., 2007).

La principale difficulté du support des communications multi-parties est la prise en compte simultanée des communications directes (un à un), des communications de groupe (un à plusieurs) et de l'écoute flottante (plusieurs à un/plusieurs) (Saunier et Balbo, 2007). Dans les CMP, les connaissances de l'émetteur ne sont pas suffisantes car il ne connaît pas nécessairement tous les agents intéressés par son message. Par exemple, un agent peut écouter des messages sans que l'émetteur n'en soit informé (Platon et al., 2007). De plus, une partie des interactions est contextuelle et dépend de l'état du SMA (Tummolini et al., 2004). La seconde difficulté est alors l'intégration des méta-informations se rapportant au contexte. Pour un récepteur, l'utilité d'un message peut dépendre du contexte de l'émetteur, par exemple sa localisation (Julien et Roman, 2004), du contexte du message, par exemple son thème (Busetta et al., 2002), et du contexte du récepteur lui-même, par exemple sa disponibilité. La diffusion des informations, par l'envoi de chaque message à tous les agents, permet à tous les agents intéressés par l'information de la recevoir. C'est la méthode jusqu'ici utilisée pour mettre en oeuvre les CMP (Legras et Tessier, 2004; Kamali et al., 2007). Cependant, la diffusion induit des coûts importants en bande passante, ainsi que des coûts de traitement inutiles pour tous les agents qui ne sont pas intéressés par les messages, ce qui rend la diffusion inapplicable dès que le nombre d'agents et/ou de messages est important.

Pour choisir les récepteurs d'une communication multi-partie, il est nécessaire de pouvoir accéder à deux connaissances : les méta-informations et les besoins des agents. Les besoins des agents sont ceux de l'émetteur, mais également ceux des récepteurs intéressés par l'information. Dans (Platon et al., 2007) et (Saunier et Balbo, 2007), les auteurs montrent que le choix des récepteurs devrait être réalisé au niveau de l'environnement. Celui-ci est considéré comme une abstraction de premier ordre qui encapsule une partie des "responsabilités" (ou services) du système multi-agents (Weyns et al., 2007). L'utilisation du contexte nécessite un mécanisme pour y accéder et le mettre à jour. L'environnement est alors utilisé comme serveur d'informations (Bucur et al., 2005), qui contient les *descriptions* des entités composant le SMA, ainsi que les besoins des agents représentés sous forme de *filtres* logiques.

2.2 Le choix de l'analyse de données symboliques

Le modèle formel d'environnement doit répondre à un certain nombre de critères : (i) l'efficacité de la recherche des récepteurs, (ii) l'expressivité du format de représentation des données, et (iii) la facilité d'exploitation du modèle. Notre modèle vise à regrouper dans l'environnement toutes les informations sur les composants du système multi-agents qui sont nécessaires à l'interaction (critère i). L'environnement peut rapidement contenir un nombre important de descriptions et de filtres. Or, les communications doivent être délivrées rapidement aux récepteurs, il faut donc être capable de décrire et d'organiser des classes de récepteurs de grande taille dans le but de les retrouver de façon efficace. Le critère (ii) est celui de la représentation des données, qui doit être expressive et permettre de gérer de façon unifiée à la fois les descriptions des entités du système multi-agents et celles des filtres. Le dernier critère (iii) concerne l'exploitation du modèle : la modélisation doit être indépendante de l'implémentation, et elle doit être opérationnelle, c'est à dire pouvoir être mise en oeuvre dans le cadre d'applications réelles, en s'appuyant si possible sur des technologies existantes.

La recherche des récepteurs implique un processus de discrimination entre d'une part les descriptions respectant les conditions des filtres, et d'autre part les autres descriptions. En d'autres termes, il s'agit d'une classification des descriptions des composants du système multi-agents selon les filtres présents dans l'environnement. C'est pourquoi nous avons choisi la formalisation proposée par l'Analyse de Données Symboliques (ADS) (Diday et Noirhomme-Fraiture, 2008) pour notre modèle d'environnement EASI (Environnement Actif comme Support de l'Interaction). L'ADS propose un modèle d'analyse de grands ensembles de données dont la formalisation, que nous détaillons ci-dessous, répond à nos critères.

En s'appuyant sur les concepts logiques d'intension et d'extension, elle permet de décrire et de manipuler des classes de données. L'intension est une description logique d'un groupe d'individus, par exemple les hommes qui ont plus de 30 ans. L'extension de cette intension est le sous-ensemble des individus qui satisfont la description. Ces concepts sont détaillés dans la section suivante. La formalisation est expressive car elle permet l'utilisation de tout type de données, qu'elles soient qualitatives, quantitatives ou complexes, c'est à dire formées de plusieurs données simples. Enfin, la formalisation ADS est à la fois indépendante de l'implémentation et exploitable, car les définitions de classes peuvent être traduites en requêtes SQL et en formules de logique du premier ordre.

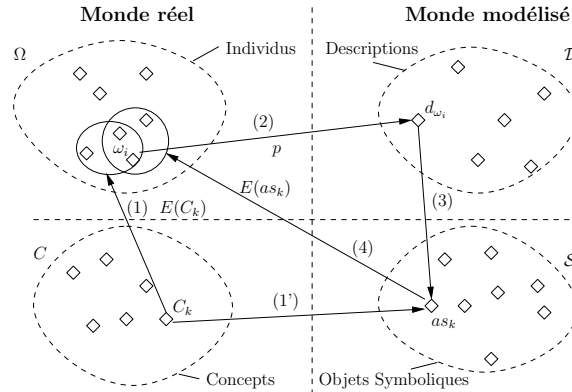


FIG. 1 – Schéma général de la modélisation par l'analyse de données symboliques

3 L'environnement de communication

Afin d'illustrer notre propos, nous utiliserons l'exemple de la gestion des communications dans un poste d'appels d'urgences tel qu'il est décrit dans Dugdale et al. (2000). Les auteurs ont mis en évidence la complexité de la gestion des communications pour produire une réponse coordonnée à un appel d'urgence. La conception d'un support des communications pour un tel environnement doit prendre en compte des communications directes (entre les protagonistes du centre de secours et les moyens mis à leur disposition) et indirectes (à cause du bruit certaines communications sont transmises par des intermédiaires). La gestion de communication de groupes (identification du récepteur selon la disponibilité par exemple) ou d'écoute flottantes (anticipation sur le partage des moyens) doivent également être prises en compte. A partir de l'exemple de la gestion d'un appel concernant une situation d'incendie dans un bâtiment, nous introduisons maintenant les définitions de base de l'Analyse de Données Symboliques et comment le modèle EASI (Environnement Actif comme Support de l'Interaction) est une adaptation au cadre des systèmes multi-agents. Nous faisons l'hypothèse que tous les messages transitent par notre système et que les opérateurs possèdent des informations sur les ressources (en l'occurrence les pompiers).

3.1 Principes de modélisation de l'analyse de données symboliques

Le monde réel (voir figure 1, inspirée de (Diday et Noirhomme-Fraiture, 2008)) est composé de n individus $\omega_i \in \Omega = \{\omega_1, \dots, \omega_n\}$, qui sont chacun caractérisés par r propriétés et une application p donnant la valeur de ces propriétés p_j , avec $j = 1, \dots, r$. Chacune des propriétés p_j possède un domaine de description D_j . Par exemple, l'individu ω_1 possède quatre propriétés : *id*, *compagnie*, *dispo*, et *lieu*. Leurs valeurs sont $p_1(\omega_1) = "a1"$, $p_2(\omega_1) = "J"$, $p_3(\omega_1) = true$, $p_4(\omega_1) = "Batiment A"$. Nous utilisons indifféremment la numérotation de la propriété ou son nom, par exemple $p_3(\omega)$ est équivalent à $dispo(\omega)$. Les domaines de description sont $D_1 = D_2$ deux chaînes de caractères non-vides, $D_3 = \{true, false\}$, et $D_4 = L$ avec L l'ensemble des localisations valides.

Soit $p_j(\omega_i)$ la valeur particulière prise par la propriété p_j pour le $i^{\text{ème}}$ individu, d_{ω_i} est la description de l'individu ω_i par l'ensemble des valeurs de ses propriétés dans le monde modélisé (Fig. 1, relation (2)). Dans l'exemple précité, d_{ω_1} est égal à (“a1”, “J”, true, “Batiment A”). L'ensemble des descriptions d'individus est noté $\mathcal{D} = \{d_{\omega_1}, \dots, d_{\omega_n}\}$

En pratique, les données correspondant à l'ensemble des individus sont représentées par une matrice $n \times r$, dont les colonnes sont les propriétés et les lignes les individus. Dans la figure 2, nous illustrons les premières cases du tableau avec un exemple. Lorsqu'une propriété n'est pas définie pour un individu donné, cette case n'est pas remplie.

	p_1	p_2	p_3	p_4	...	p_r
ω_1	“a1”	“J”	true	“Batiment A”	...	$p_r(\omega_1)$
ω_2	“a2”	“J”	false	“Batiment A”	...	$p_r(\omega_2)$
ω_3	“a3”	“L”			...	$p_r(\omega_3)$
ω_4	“a4”	“K”	true	“Batiment D”	...	$p_r(\omega_4)$
...
ω_n	$p_1(\omega_n)$	$p_2(\omega_n)$	$p_3(\omega_n)$	$p_4(\omega_n)$...	$p_r(\omega_n)$

FIG. 2 – Exemple de données initiales classiques

Les concepts du monde réel C_k , $k = 1, \dots, N$ sont des abstractions associées à un ou plusieurs individus. Si chaque individu est un concept, alors $n = N$. Un concept est modélisé par une description appelée *intension*, son extension notée $E(C_k)$ contient tous les individus satisfaisant cette description (Fig. 1, relation (1)). Par exemple, un concept peut être associé à tous les pompiers dont la localisation est le bâtiment A. Ainsi, l'extension de ce concept sera composée de l'ensemble des individus dont la valeur de la propriété *lieu* est “Batiment A”.

Un objet symbolique est une description en intension d'une classe d'individus (Fig. 1, relation (3)). Son extension est l'ensemble des individus (la classe) qui satisfont cette description (Fig. 1, relation (4)). Une assertion est un objet symbolique particulier. Soit $v = (v_1, \dots, v_r)$ la description requise d'un individu ω , avec v_j une valeur appartenant à l'ensemble de description D_j , l'assertion a la forme générale suivante :

$$as = [p_{j_1} R_{j_1} v_{j_1}] \wedge \dots \wedge [p_{j_v} R_{j_v} v_{j_v}] \text{ avec } 1 \leq j_1, \dots, j_v \leq r$$

R_j est l'opérateur de comparaison entre la valeur $p_j(\omega)$ de la propriété p_j et la valeur v_j , et $[p_j R_j v_j]$ un test de la valeur de p_j . Les indices j_1 à j_v sont inclus dans les indices $1, \dots, r$ des propriétés, ce qui signifie que toutes les propriétés ne sont pas nécessairement testées. Cela revient aussi à dire que les autres propriétés sont autorisées à prendre une valeur quelconque de leur domaine respectif. Par exemple, l'assertion $as_1 = [p_2 = \text{“J”}] \wedge [p_4 = \text{“Batiment A”}]$ décrit les pompiers de la compagnie J situés dans le bâtiment A. Les tests sont effectués sur les propriétés p_2 (la propriété *compagnie*) et p_4 (*lieu*). Avec les individus de la figure 2, l'extension $E(as_1)$ est une classe d'individus qui contient les individus ω_1 et ω_2 .

Une assertion correspond à des conditions sur les valeurs des propriétés décrivant un individu. Lorsqu'on évalue une assertion as pour un individu particulier $\omega \in \Omega$, elle prend la valeur $as(\omega) = \text{vrai}$ si l'assertion est vérifiée, et $as(\omega) = \text{faux}$ sinon. L'ensemble des individus pour lesquels la valeur est vraie forme l'extension de l'objet symbolique et représente donc une approximation de l'extension du concept que cet objet symbolique modélise.

Il est possible de réunir les conditions sur plusieurs objets au sein d'un même objet symbolique qui est alors appelé une *horde*. Par exemple, la horde $h = [\text{compagnie}(u) = \text{“L”}] \wedge$

$[dispo(v) = true]$ décrit deux individus. Son extension est l'ensemble des couples d'individus (u, v) respectant les conditions. Dans la figure 2, il s'agit des deux couples (ω_3, ω_1) et (ω_3, ω_4) . L'individu ω_3 est le seul individu satisfaisant la première condition $[compagnie(u) = "L"]$, et les individus ω_1 et ω_4 satisfont la seconde condition $[dispo(v) = true]$.

Une horde peut décrire des individus dont les descriptions ne sont pas homogènes, dans le sens où ils ne sont pas décrits par les mêmes propriétés. Par exemple, les individus ω_1 et ω_3 ont seulement deux propriétés en commun.

3.2 Le modèle Environnement Actif comme Support de l'Interaction

Nous donnons dans cette section les éléments de formalisation liés à la gestion des connaissances que nous illustrons dans la figure 3. Une description complète de la formalisation est donnée dans (Saunier et Balbo, 2007). Selon la modélisation issue de l'ADS, un SMA est composé de l'ensemble des individus (appelé entités) Ω (agents, messages, objets) et de l'environnement \mathcal{E} . Les besoins des agents (informations et récepteurs) sont des concepts. Nous appelons objet tout ce qui n'est pas un agent ou un message mais qui permet de donner de l'information sur le contexte d'une communication. Par exemple, chaque pièce d'un bâtiment peut être représentée par un objet donnant des informations (localisation, capacité, ...). L'environnement contient le monde modélisé, autrement dit l'ensemble des descriptions \mathcal{D} des entités et l'ensemble des objets symboliques associés à des concepts, dont le sous-ensemble \mathcal{F} (Fig. 3) correspond aux filtres des agents sur les communications échangées.

Chaque entité est décrite dans l'environnement par un ensemble de couples \langle propriété, valeur \rangle qui est utilisé pour le routage des informations. Dans l'exemple Fig. 3, on cherche à adresser un message aux agents disponibles situés dans le bâtiment A (Fig. 3, extension (0)). La valeur d'une propriété est dynamique, mais l'ensemble de propriétés par lequel est décrit une entité ne varie pas. Par exemple, la disponibilité d'un agent varie dans le temps mais la propriété donnant la valeur de la disponibilité existe toujours dans la description d'un agent. Un individu n'a pas une description complète dans l'environnement, par exemple les agents ne divulguent que certaines informations.

Nous utilisons trois types de concepts afin de classifier les données. Le premier type de concept, appelé **Pdescription**, concerne les entités décrites par un même ensemble de propriétés. Soit C une catégorie et P_C sa *Pdescription*, l'assertion $cat_C = [P_C \subset P_\omega]$ est vraie pour une entité $\omega \in \Omega$ si P_C est inclus dans la *Pdescription* de ω . Par exemple figure 3, un pompier est une entité $po \in \Omega$ et la relation (2) donne sa description d_{po} dont la valeur des propriétés évolue avec le temps, alors que la *Pdescription* de la catégorie à laquelle po appartient est $P_{po} = \{id, compagnie, lieu\}$ et ne varie pas dans le temps. En fonction de cette *Pdescription* et du concept initial (Fig. 3, relation (1)), nous obtenons l'assertion cat_p . Son extension dans Ω contient les entités $\{\omega_1, \omega_2, \dots\}$ (Fig. 3, relation (4)).

Le second type de concept est lié aux besoins en communication des agents, par exemple transmettre le message aux pompiers de la compagnie J dans le bâtiment A. Un **filtre** est un objet symbolique donnant la description en intension des contraintes sur les individus (Fig. 3 relation (3)) concernés par ce besoin. Un filtre est composé de deux assertions décrivant respectivement quels agents doivent recevoir le message (assertion notée f_a) et quels messages sont concernés (assertion notée f_m), ainsi que d'une horde optionnelle décrivant des contraintes sur d'autres objets du contexte (horde notée f_C). L'extension d'un filtre (Fig. 3, extension (5)) est donc un ensemble de triplets $\langle agent, message[, contexte] \rangle$ qui mettent en relation

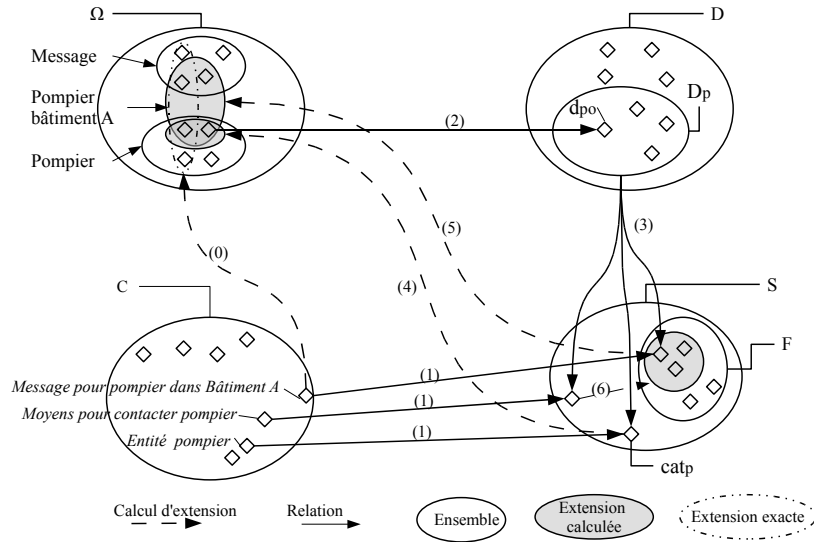


FIG. 3 – Modélisation des composants de l’interaction

les messages avec les agents. Dans l’exemple, si le filtre concerne les agents de la compagnie J situés dans le bâtiment A, alors l’assertion concernant les agents est $[compagnie(\omega) = "J"] \wedge [lieu(\omega) = "Batiment A"]$. De façon à restreindre le nombre de descriptions à tester, nous utilisons la *Pdescription* $P_{f_a} = \{compagnie, lieu\}$ qui contient l’ensemble des propriétés que les agents doivent posséder pour pouvoir satisfaire les conditions de l’assertion f_a . Le regroupement des entités selon cette catégorie est fait *a priori* et lorsqu’un message concerné par ce filtre doit être affecté, la recherche des récepteurs peut être limité à cette catégorie.

Le dernier type de concept concerne la gestion des relations entre catégorie d’entités (Fig. 3 relation (3)) et objets symboliques. L’idée fondamentale est de structurer les informations en stockant le lien entre les *Pdescriptions* des entités et les *Pdescriptions* des descriptions des filtres. Il s’agit d’une application sa telle que $sa : S \rightarrow \{true, false\}$. Par exemple, la relation entre un pompier et les filtres qui le concernent est un concept. Ce concept est exprimé grâce à l’application $sa_p(f) = [P_{f_a}(f) \subset P_p]$; et son extension dans \mathcal{F} est $Reception_p = \{f \in \mathcal{F} | as_p(f) = true\}$ (Fig. 3, extension (6)). Seules les filtres de $Reception_p$ peuvent déclencher la réception de messages pour la catégorie d’agent de p . Le même raisonnement pour une entité de la catégorie message nous permet d’évaluer les ensembles $Channel_m$ (ensemble de filtres relatif à la catégorie de message m), $Recepteur_m$ (ensemble des agents pouvant recevoir m) et $Context_m$ (ensemble des contextes dans lesquels la réception est possible). Ainsi, lorsqu’un nouveau message est émis, il n’est pas nécessaire de traiter l’ensemble des filtres \mathcal{F} , mais seulement ceux de $Channel_m$. Pour ces filtres, il est alors possible de rechercher les récepteurs et contextes dans $Recepteur_m$ et $Context_m$ au lieu de l’ensemble des agents et Ω .

4 Discussion

Dans cette section, nous discutons comment en s'appuyant sur une modélisation commune, l'approche multi-agent proposée dans EASI et l'approche statistique supportée par l'ADS abordent un problème similaire d'affectation de messages. Nous profiterons de cette comparaison pour mettre en évidence les apports respectifs des deux domaines, explorés et à venir.

4.1 Approche SMA du problème d'affectation de messages

Comme nous l'avons vu en section 3.2, le modèle EASI permet d'associer à chaque besoin d'un agent un filtre dans l'ensemble des objets symboliques \mathcal{S} . Ainsi EASI respecte l'approche multi-agent classique de modélisation des intervenants par des agents et de leurs relations par des protocoles d'interaction (suite ordonnée de messages). Les agents sont représentés dans le système par des entités et les protocoles d'interaction s'appuient sur la mise en oeuvre de filtres. Les messages et autres objets sont également décrits par des entités. L'existence d'un nouveau message dans Ω doit donner lieu au dépôt d'une description dans D et ainsi entraîner le protocole d'interaction adéquat. Ainsi contrairement à l'ADS qui travaille à partir d'une matrice de données initiales, avec le modèle EASI le système peut fonctionner sans données initiales. La conception de ce protocole est issue de la connaissance des experts du domaine et/ou de l'observation du fonctionnement du système modélisé.

EASI permet la modélisation de la dynamique du système en intégrant dans son fonctionnement le processus de mise à jour des données dans Ω qui se répercute sur D puis sur le calcul des extensions des objets symboliques contenus dans \mathcal{S} . Ainsi, l'évolution des valeurs des entités est directement prise en compte dans l'évaluation des messages. Cette dynamique est également prise en compte dans la gestion même des objets symboliques qui peuvent être ajoutés ou retirés dynamiquement en fonction de l'évolution des besoins des agents. Ainsi des pompiers peuvent être plus ou moins à l'écoute des messages diffusés par le poste de contrôle en fonction de leur activité, la notion floue de "plus ou moins" pouvant être traduite par la présence de filtres adéquats (et de leur priorité (Saunier et al., 2007)).

EASI permet un méta-contrôle du fonctionnement du système en introduisant un nouveau type d'objet symbolique dont l'extension est calculée au sein même de l'ensemble des objets symboliques \mathcal{S} . Il s'agit d'obtenir une connaissance sur le fonctionnement du système et notamment sur la manière dont il est utilisé par les agents. Ainsi, il est par exemple possible de détecter un agent isolé (qui ne peut pas être contacté) ou les messages qui ne peuvent pas être adressés (car mal conçus), parce qu'il n'existe aucun objet symbolique les prenant en compte. Actuellement, ces relations sont utilisées afin d'optimiser le traitement des données mais une utilisation de monitoring du système est également possible.

Symétriquement, l'utilisation d'un système multi-agent peut permettre de "donner vie" à un modèle de données statiques. Ainsi la modélisation multi-agent issue de l'analyse d'un problème dynamique génère des données qui peuvent alors être comparées à la matrice de données initiales à des fins de calibrage et validation du modèle multi-agent. Le système multi-agent peut alors être utilisé à des fins de simulation de nouvelles données.

4.2 Approche statistique du problème d'affectation de messages

Un modèle statistique comme l'ADS peut aborder la problématique d'affectation de messages de la façon suivante : à partir d'une matrice de données décrivant les précédents échanges, l'ADS utilise une fonction d'agrégation pour générer le filtre puis son extension est calculée. L'analyse des erreurs induites par l'extension conduit à une amélioration récursive du filtre.

Si la formalisation as d'un concept C_k est parfaite et les descriptions des individus sont correctes, leurs extensions sont identiques ($E(as) = E(C_k)$). Une problématique de l'Analyse de Données Symboliques est de vérifier que le processus de modélisation est correct, par la comparaison entre ces deux extensions. Deux types d'erreur peuvent apparaître : soit l'extension $E(as)$ contient des individus qui n'appartiennent pas à $E(C_k)$, soit l'extension $E(as)$ ne contient pas certains individus qui appartiennent à $E(C_k)$.

Par exemple, le message "éteindre le feu" peut être associé aux pompiers de la compagnie J dans le bâtiment A. La description est ici généralisante, fondée sur une information statistique issue des données : la compagnie J est la plus importante à proximité du lieu, et le bâtiment A possède le plus de salles atteintes. L'objet symbolique calcule alors l'extension de ce concept dans l'espace des individus, mais des erreurs apparaissent lorsque la généralisation a fait perdre de l'information pertinente, par exemple des pièces atteintes par le feu dans le bâtiment B, ou des salles non atteintes dans le bâtiment A (erreurs de premières et de secondes espèces). Il est donc nécessaire d'envoyer un second message, en affinant la description des lieux atteints et donc de l'objet symbolique les représentant. Ici, il s'agira d'adopter une granularité plus fine des lieux décrits. Le processus est donc cyclique : description des individus, généralisation, calcul de l'extension de l'objet symbolique, observation des erreurs, et ainsi de suite jusqu'à disparition des erreurs.

4.3 Bénéfices croisés d'une approche conjointe

La figure 4 représente les différents flux d'information qui sont le résultat de l'utilisation croisée de modèles statistiques issues de l'ADS et de EASI. Le premier (Fig. 4, *Modélisation*) représente l'approche multi-agent classique de résolution d'un problème dynamique, le résultat étant un SMA avec en plus dans EASI la partie description des composants et un algorithme de matching. Les objets symboliques sont issus de l'analyse du problème et directement mis à disposition dans l'ensemble des objets symboliques \mathcal{S} pour une utilisation immédiate ou dans les agents pour une utilisation ultérieure en fonction de l'évolution de leurs besoins. Le résultat du fonctionnement du SMA est un artefact du problème initial (*création*). Un ensemble de données a donc été créé et peut être analysé par les techniques issues de l'ADS (*Evolution*). L'avantage ici pour l'ADS est d'accéder à des données pour un problème qui pouvait ne pas en avoir initialement, ces données étant au format propre au modèle. Sur ces données, les techniques issues de l'approche statistique permettent une analyse dont le résultat est une amélioration du modèle initial (*Amélioration*). Cette analyse peut s'appuyer sur une comparaison entre l'artefact et des données réelles (si elles existent), sur l'avis des experts ou sur la recherche de nouveaux objets symboliques. Pour le modèle multi-agent, le bénéfice est un retour au concepteur qui peut être de deux types. Le premier concerne la détection des erreurs de modélisation des filtres et descriptions aboutissant à des différences entre l'extension d'un concept et l'extension de l'objet symbolique le représentant. Le second type permet l'identification de relations de communication privilégiées émergentes entre les agents lors de l'exécution, et plus

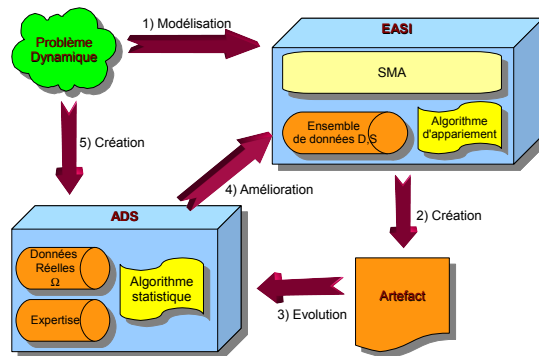


FIG. 4 – *Processus d'amélioration croisée*

généralement l'extraction de l'organisation du système. Cette connaissance est un atout pour mettre en place de nouvelles règles de l'environnement, soit pour favoriser les communications ayant déjà été observées, soit pour empêcher des comportements non souhaités. Le retour opéré permet au concepteur d'améliorer le fonctionnement du SMA itérativement, par exemple en ajustant la granularité des descriptions ou en modifiant les protocoles de communication.

Ce cycle peut commencer par le flux *création* si le problème possède déjà des données. Dans ce cas le modèle multi-agent peut directement être issu du résultat de l'analyse statistique. Ainsi si pour l'instant, les filtres sont créés par le concepteur du système multi-agents, l'approche statistique est une piste intéressante pour la génération et modification automatique de filtres en temps réel par les agents, s'appuyant sur la connaissance courante stockée dans l'environnement. Ce processus complexe serait une adaptation pour le domaine multi-agent des travaux réalisés pour passer de la statistique sur des données classiques à celle sur des données symboliques (Billard et Diday, 2003; Diday et Noirhomme-Fraiture, 2008).

5 Conclusion et perspectives

Dans cet article, nous avons montré comment l'utilisation d'une formalisation issue de l'analyse de données symboliques peut être utilisée dans le cadre de la gestion des méta-informations d'un système multi-agents. La modélisation, appelée EASI (Environnement Actif comme Support de l'Interaction) est celle d'un environnement actif, les agents lui déléguant les tâches d'adressage des données et de recherche d'information.

En s'appuyant sur les concepts logiques d'intension et d'extension, cette formalisation permet de décrire et de manipuler des classes de données. Cette formalisation est expressive car elle permet l'utilisation de tous types de données, qu'elles soient qualitatives, quantitatives ou complexes. Enfin, le modèle est à la fois indépendant de l'implémentation et exploitable, car les définitions de classes peuvent être traduites directement en formules de logique du premier ordre et implémentées sous différentes formes.

Nous avons implémenté le modèle EASI et réalisé des tests sur les algorithmes de recherche des récepteurs dans le cadre d'un exemple d'intelligence ambiante (Saunier et Balbo, 2007). Ceux-ci ont été comparés à la diffusion, qui est le modèle utilisé jusqu'ici pour la mise en oeuvre des communications multi-parties. En terme de temps d'exécution, les tests ont montré un gain allant de 67% pour 10 agents à 80% pour 200 agents. Lorsque la fréquence de mise à jour des descriptions augmente, le gain atteint jusqu'à 95%. Ces expériences montrent que le modèle EASI permet de gérer les communications d'un nombre significativement plus grand d'agents que la diffusion. Des expérimentations complémentaires sont en cours de réalisation.

L'intégration de techniques issues de la représentation des connaissances et de la fouille de données ouvre de nombreuses opportunités pour l'amélioration de la gestion des connaissances dans les systèmes complexes, et plus particulièrement dans les systèmes multi-agents. Réciproquement, nous pensons que l'analyse de données symboliques peut tirer partie des spécificités de mise en oeuvre dans ce contexte, en particulier grâce à la structuration des données et la mise en oeuvre automatique en temps réel de décisions fondée sur ces connaissances (en l'occurrence le routage).

Références

- Billard, L. et E. Diday (2003). From the statistics of data to the statistics of knowledge : Symbolic data analysis. *Journal of the American Statistical Association* 98(462), 470–487.
- Bucur, O., P. Beaune, et O. Boissier (2005). Steps towards making contextualized decisions : How to do what you can, with what you have, where you are. In T. Roth-Berghofer, S. Schulz, et D. B. Leake (Eds.), *MRC*, Volume 3946 of *Lecture Notes in Computer Science*, pp. 62–85. Springer.
- Busetta, P., A. Donà, et M. Nori (2002). Channeled multicast for group communications. In *AAMAS '02 : Proceedings of the First International Joint Conference on Autonomous Agents and Multi-agent Systems*, New York, NY, USA, pp. 1280–1287. ACM Press.
- Diday, E. et M. Noirhomme-Fraiture (2008). *Symbolic Data Analysis and the SODAS Software*. New York, NY, USA : Wiley-Interscience.
- Dugdale, J., J. Pavard, et B. Soubie (2000). A pragmatic development of a computer simulation of an emergency call center. In *Designing Cooperative Systems : The Use of Theories and Models*, pp. 241–256. IOS Press.
- Gutnik, G. (2005). Monitoring large-scale multi-agent systems using overhearing. In *AAMAS '05 : Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, New York, NY, USA, pp. 1377–1377. ACM Press.
- Julien, C. et G.-C. Roman (2004). Supporting context-aware interaction in dynamic multi-agent systems. In D. Weyns, H. V. D. Parunak, et F. Michel (Eds.), *E4MAS*, Volume 3374 of *Lecture Notes in Computer Science*, pp. 168–189. Springer.
- Kamali, K., X. Fan, et J. Yen (2007). Towards a theory for multiparty proactive communication in agent teams. *International Journal of Cooperative Information Systems* 16(2), 271–298.
- Kumar, S., M. J. Huber, D. McGee, P. R. Cohen, et H. J. Levesque (2000). Semantics of agent communication languages for group interaction. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*, pp. 42–47. AAAI Press / The MIT Press.

- Legras, F. et C. Tessier (2004). Lotto : Group formation by overhearing in large teams. In *Advances in Agent Communication*, Volume 2922 of *LNAI*, pp. 254–270. Springer.
- Omicini, A. et F. Zambonelli (1999). Tuple centres for the coordination of internet agents. In *SAC '99 : Proceedings of the 1999 ACM symposium on Applied computing*, New York, NY, USA, pp. 183–190. ACM Press.
- Platon, E., N. Sabouret, et S. Honiden (2007). Tag interactions in multiagent systems : Environment support. In *Proceedings of Environment for Multi-Agent Systems, Workshop held at the Fifthourth Joint Conference in Autonomous Agents and Multi-Agent Systems*, Volume 4389 of *LNAI*, pp. 106–123. Springer Verlag.
- Pujol, J. M., R. Sangüesa, et J. Delgado (2002). Extracting reputation in multi agent systems by means of social network topology. In *AAMAS '02 : Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, New York, NY, USA, pp. 467–474. ACM.
- Saunier, J. et F. Balbo (2007). An environment to support multi-party communications in multi-agent systems. In H.-D. Burkhard, G. Lindemann, R. Verbrugge, et L. Z. Varga (Eds.), *CEEMAS*, Volume 4696 of *Lecture Notes in Computer Science*, pp. 52–61. Springer.
- Saunier, J., F. Balbo, et F. Badeig (2007). Environment as active support of interaction. In *Proceedings of Environment for Multi-Agent Systems, Workshop held at the Fifth Joint Conference on Autonomous Agents and Multi-Agent Systems*, Volume 4389 of *Lecture Notes in Artificial Intelligence*, pp. 87–105. Springer Verlag.
- Tummolini, L., C. Castelfranchi, A. Ricci, M. Viroli, et A. Omicini (2004). "exhibitionists" and "voyeurs" do it better : A shared environment approach for flexible coordination with tacit messages. In *Proceedings of Workshop on Environments for Multi-Agent Systems*, Volume 3374 of *LNAI*, pp. 215–231. Springer Verlag.
- Weyns, D., A. Omicini, et J. Odell (2007). Environment as a first class abstraction in multiagent systems. *Autonomous Agents and Multi-Agent Systems* 14(1), 5–30.
- Zha, X. F., S. Y. E. Lim, et W. F. Lu (2003). A knowledge intensive multi-agent system for cooperative/collaborative assembly modeling and process planning. *J. Integr. Des. Process Sci.* 7(1), 99–122.

Summary

The use of complex knowledge in real-time is a challenge in many areas such as semantic web, simulation and multi-agent systems (MAS). In the multi-agent paradigm, recent works have shown that multi-party communications (MPC) present interesting opportunities in terms of communication realism, information spread and speech acts semantics. However, classic communication infrastructures are not adequate to support MPC. In this article, we propose the use of the symbolic data analysis (SDA) formalism to model the information and the agents needs. We base the message routing on this modeling, in the framework of a communication environment for MAS. We illustrate our proposal with an example of communication management for an emergency call center. We present our experience feedback, and discuss the perspectives of cross-fertilization between SDA and MAS.