



HAL
open science

Prise en charge des personnes vieillissantes à leur domicile. Monitoring et classification

Emmanuel Conchon, Nicolas Singer, Jean-Marie Pecatte, Sylvie Trouilhet

► To cite this version:

Emmanuel Conchon, Nicolas Singer, Jean-Marie Pecatte, Sylvie Trouilhet. Prise en charge des personnes vieillissantes à leur domicile. Monitoring et classification. *Revue des Sciences et Technologies de l'Information - Série RIA : Revue d'Intelligence Artificielle*, 2013, 27 (3), pp.321-346. hal-03467059

HAL Id: hal-03467059

<https://hal.science/hal-03467059>

Submitted on 6 Dec 2021

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.



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 12367

To link to this article : DOI :10.3166/ria.27.321-346
URL : <http://dx.doi.org/10.3166/ria.27.321-346>

To cite this version : Conchon, Emmanuel and Singer, Nicolas and Pecatte, Jean-Marie and Trouilhet, Sylvie *Prise en charge des personnes vieillissantes à leur domicile. Monitoring et classification.* (2013) Revue d'Intelligence Artificielle, vol. 27 (n° 3). pp. 321-346. ISSN 0992-499X

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Prise en charge des personnes vieillissantes à leur domicile

Monitoring et classification

Emmanuel Conchon¹, Nicolas Singer¹, Jean-Marie Pecatte²,
Sylvie Trouilhet²

1. Université de Toulouse - IRIT
ISIS, Centre universitaire J.-F. Champollion
Rue Firmin Oulès, F-81100 Castre

2. Université de Toulouse - IRIT
Université Paul Sabatier, 118 Route de Narbonne
F-31062 Toulouse Cedex 9
{prénom.nom}@irit.fr

RÉSUMÉ. Le maintien à domicile des personnes âgées est devenu un enjeu important de santé publique. Nos travaux sont destinés à traiter à grande échelle l'information recueillie au domicile des personnes pour en faire une information consolidée décrivant des comportements globaux. Nous proposons pour ce faire des algorithmes de classification pour identifier des profils collectifs et y rattacher les personnes suivies. Ces algorithmes tirent profit des technologies multi-agents afin de gérer l'hétérogénéité des équipements et des services produisant l'information ainsi que leur totale distribution. Ils sont résistants aux déficits d'information pouvant provenir de pannes ou d'interruption de la télé-vigilance, et sont suffisamment anonymes pour assurer les besoins de confidentialité des informations échangées. Les profils obtenus sont utilisés pour estimer l'état des personnes âgées et les efforts à déployer pour qu'elles puissent continuer à vivre à leur domicile, et à un niveau plus global pour évaluer des tendances sanitaires générales (épidémie, fortes chaleurs, etc.).

ABSTRACT. Ageing at home is nowadays a matter of public health in western societies. The presented work proposes to process the huge amount of data collected at elders homes to provide a consolidated view which describes their general behaviors. To achieve this processing, several classification algorithms are presented. These algorithms rely on a multi-agent paradigm and deal with the heterogeneity among the collected data. This heterogeneity is mainly caused by the heterogeneity of the hardware used to collect the data at home. Moreover, these algorithms are resistant to data loss and can provide several general health profiles of elders even with an incomplete set of data. Based on the resulting profiles, changes in health status can be anticipated for a specific elder but it is also possible to use these

profiles to detect some global events such as an epidemic or the impact of high temperature on an ageing population.

MOTS-CLÉS: classification, système multi-agent, maintien à domicile, télé-vigilance.

KEYWORDS: clustering, multi-agent system, elderly homecare, monitoring system.

DOI:10.3166/RIA.27.321-346

1. Introduction

Le processus de vieillissement de la population entraîne une modification profonde et rapide de notre société, en transformant sa structure d'âge. En France, selon les projections démographiques de l'INSEE, la proportion des personnes âgées de 60 ans ou plus passera de un sur cinq en 2005 à un sur trois en 2050 (Rivière *et al.*, 2010).

Par ailleurs, 7 % des plus de 60 ans sont dépendants en raison d'une maladie invalidante, soit un peu plus d'un million de personnes, dont plus des deux tiers ont la maladie d'Alzheimer ou une autre forme de démence. Si on fait l'hypothèse que ce pourcentage reste stable, le nombre de personnes dépendantes pourrait atteindre un premier pic à partir de 2020-2025 (1,2 million en 2020) et devrait se maintenir à un niveau élevé jusqu'en 2050-2060. D'ici 2020, selon l'INSEE, 1,3 million de personnes seront touchées par la maladie d'Alzheimer, soit une personne de plus de 65 ans sur quatre (Gimbert *et al.*, 2010).

Ce processus de vieillissement de la population et de la dépendance entraîne aussi de nombreux problèmes financiers et organisationnels pour les systèmes de santé : apparition du « cinquième risque ». La personne âgée dépendante doit pouvoir rester le plus longtemps possible dans son environnement familial et conserver les liens familiaux, affectifs, les relations de voisinage qui nourrissent sa vie intérieure et sa vie sociale.

On parle alors de maintien à domicile ou MAD. « Le maintien à domicile c'est d'abord et avant tout, l'expression du désir de nombreuses personnes qui ont fait le choix de demeurer dans leur domicile jusqu'au bout de leur âge et de leur vie, c'est-à-dire qu'elles ont souhaité continuer à vivre et mourir chez elles. » (Ennuyer, 2006). L'usage des technologies dans le MAD (les gérontechnologies : les dispositifs médicaux, les aides techniques, les technologies de l'information et de la communication et l'habitat intelligent) peut grandement aider à créer un environnement de vie optimal permettant aux personnes âgées de continuer à vivre chez elles dans un milieu confortable et sécurisé (Carré *et al.*, 2009).

Concernant le suivi de la santé des personnes chez elles, de très nombreux systèmes de télé-vigilance sont d'ores et déjà disponibles sur le marché. Ces systèmes se composent habituellement d'un appareillage installé au domicile des personnes surveillées qui via un ensemble de capteurs font remonter des informations sur un certain nombre d'évènements (appels de détresse, chutes,

entrée/sortie du domicile, déplacements, appétit, etc.) (Chan *et al.*, 2003). Ces systèmes de téléassistance ou télé-vigilance produisent énormément de données que les opérateurs utilisent essentiellement pour justifier leurs activités auprès des donneurs d'ordre. Mais ces mêmes données pourraient aussi être utilisées pour avoir une image sanitaire et sociale d'une partie de la population.

Cet article aborde le maintien à domicile d'une personne dépendante en allant au-delà des technologies de surveillance et des méthodes de traitement des données. En effet, il apparaît que les systèmes actuels dédiés au MAD restent centrés sur un patient ou une juxtaposition de patients. Aucun des projets actuels ne vise à collectiviser un grand nombre de surveillances individuelles. L'hétérogénéité des applications, le cloisonnement et la forte distribution des données de surveillance ainsi que les aspects liés au respect de la vie privée rendent ardue cette tâche. Pourtant, cela permettrait par recoupements et processus de généralisation de faire émerger des catégories de personnes afin de procurer une surveillance adaptative ou encore de contribuer à la prévision et la mesure des évolutions démographiques.

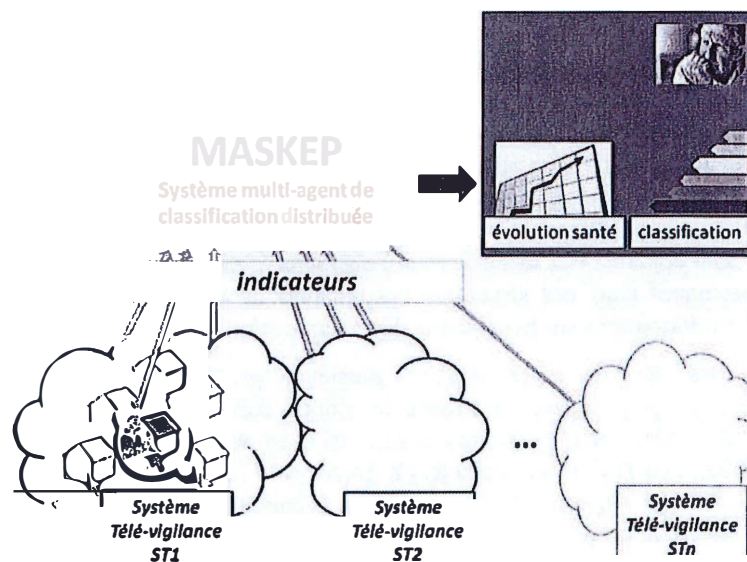


Figure 1. Vue générale du système MASKEP

Nous proposons un système appelé MASKEP destiné à traiter les nombreuses données collectées au domicile des personnes âgées. Ces données sont accessibles soit par des systèmes automatisés de télé-vigilance, soit par des indications humaines fournies par les aidants, voire le corps médical. Le traitement aboutit à la constitution de motifs (catégories) caractérisant un état. Cet état peut être utilisé comme un révélateur, par exemple de l'autonomie des personnes classées dans cette catégorie, et

donc des moyens nécessaires à leur maintien à domicile. Tout changement d'état d'une personne est alors vu comme une réflexion à mener sur l'évolution des moyens à déployer par la famille, le corps médical et les associations de services à la personne. Ces états peuvent aussi être utilisés par les instances de gouvernance et de gestion, pour avoir une vision globale de l'ensemble des personnes analysées par le système. Certaines crises sanitaires (fortes chaleurs, épidémies...) pourraient par exemple être détectées. La figure 1 illustre le fonctionnement global de notre système.

La prochaine section de notre contribution présente plusieurs systèmes de télé-vigilance sur lesquels nous nous basons pour obtenir de façon automatisée des indicateurs révélateurs de l'état de la personne suivie. La section 3 présente l'architecture de notre système multi-agent MASKEP destiné à traiter les données récoltées. Nous présentons dans la section 4 les méthodes de classification que notre système utilise et dans la section 5 nous appliquons ces méthodes à un cas concret concernant le maintien à domicile. La section 6 positionne notre approche dans le domaine de la classification multi-agent de données. Enfin nous concluons cette contribution dans la section 7.

2. Les systèmes de télésurveillance

Le vieillissement de la population et la réorganisation des systèmes de santé dans les sociétés des pays développés font du maintien des personnes à domicile un enjeu médical, social et économique majeur. Dans ce contexte, de nombreux systèmes de télésurveillance permettant un suivi des personnes âgées et/ou dépendantes à leur domicile sont apparus. Ces systèmes ont pour objectif commun de retarder l'entrée de ces personnes dans des structures hospitalières ou médicalisées en remontant différentes informations sur la personne dépendante et/ou son environnement.

Au cours des dernières années, plusieurs prototypes de systèmes de télésurveillance sont apparus à travers le monde comme les projets PROSAFE (Chan *et al.*, 2003), TelePat (Boudy *et al.*, 2007) ou AILISA (Rialle *et al.*, 2003 ; Noury, 2005) pour la France, CarerNet (Williams *et al.*, 1998) pour l'Angleterre, les travaux menés par Tagura et Ogawa au Japon (Tamura *et al.*, 1998 ; Ogawa *et al.*, 2000) ou encore le projet de maison intelligente Gator Tech (Helal *et al.*, 2005) aux États-Unis pour ne citer qu'eux.

Ces projets présentent bien sûr des différences au niveau de l'échelle de leur déploiement (local, national etc.) ou de leurs objectifs (asthme, Alzheimer, cardiologie, chutes, activités journalières...). Cependant, chacun remonte des informations pertinentes sur l'évolution de santé de l'occupant et notamment des informations de haut niveau (activité quotidienne, agitation nocturne) exploitables pour classer les patients en fonction de leurs pathologies.

Ainsi, le projet PROSAFE initié par le LAAS de Toulouse repose sur l'utilisation de capteurs de mouvement infrarouges permettant de suivre une personne dans ses déplacements quotidiens de manière non intrusive. Déployé dans

une chambre d'hôpital, il permet de déterminer le temps passé dans chaque zone (lit, fauteuil, salle de bain, etc.) et déclenche une alarme lorsque le temps passé diffère trop par rapport au comportement habituel. Sur le même principe, un capteur positionné au dessus du lit permet de déterminer l'agitation nocturne de l'occupant de la chambre. Ce suivi constant des positions aide également à déterminer les déplacements de l'occupant de la chambre, ce qui peut permettre de détecter des changements d'habitudes à long terme par exemple.

Le projet TelePat développé par l'Institut national des télécommunications (INT) propose une instrumentation du patient grâce à un terminal porté à la ceinture. Ce terminal est équipé de différents capteurs permettant des mesures d'actimétrie et des mesures physiologiques. Par exemple, un capteur de pouls permet de mesurer la fréquence cardiaque de l'utilisateur. L'utilisation de capteurs ECG permet d'effectuer une mesure plus fine, notamment en cas d'antécédents cardiaques. Le terminal embarque en sus des capteurs inertiels donnant des informations sur la posture du patient (allongé, debout, couché). Ce projet peut s'inter-opérer avec d'autres projets antérieurs comme le projet GARDIEN qui, comme PROSAFE, effectue une localisation du patient à l'aide de capteurs infrarouges et tente de détecter les chutes grâce à l'utilisation d'un capteur acoustique.

Le projet AILISA développé par Noury *et al.* vise au déploiement de plateformes de maintien à domicile à travers la France. Ces plateformes situées à Toulouse et Grenoble intègrent trois dispositifs : un dispositif domotique HIS (habitat intelligent pour la santé) issue du projet TISSAD, un gilet intelligent issue du projet VTAMN et un robot déambulateur issue du projet MONIMAD3. Le dispositif domotique se compose de capteurs de mouvement volumétriques infrarouges, d'un pèse-personne sans fil, d'un tensiomètre et d'un détecteur de pouls. Le gilet intègre plusieurs capteurs physiologiques (fréquence cardiaque et respiratoire, température, actimétrie), un système GPS pour la localisation et un capteur de détection de chute. Le robot MONIMAD a été conçu pour aider les personnes âgées à se déplacer en les guidant de manière automatique à travers le logement. Ces différents équipements permettent d'obtenir des informations de plus haut niveau comme les trajets effectués, la vitesse de déplacement, l'agitation ou le temps d'occupation d'une pièce. L'étude sur la durée de ces différentes informations permet de déterminer un comportement standard de l'occupant.

En Angleterre, le projet CarerNet mené par l'Université de Bangor au pays de Galles, fut une des premières propositions de système de télé-vigilance et d'aide à l'hospitalisation à domicile. Il intègre des capteurs physiologiques (pression sanguine, ECG, vitesse de la respiration, température etc.), des capteurs environnementaux (fumée, température des pièces, niveau sonore) et effectue des analyses sur le style de vie de l'occupant. Cette analyse se fait à l'aide de capteurs de mouvement pour mesurer la mobilité, d'accéléromètres pour détecter les chutes ou encore de capteurs de pression pour connaître le taux d'occupation du lit.

Au Japon, Tamura *et al.* ont également travaillé sur une maison intelligente pour permettre le suivi de personnes âgées à leur domicile. Cette maison intègre des toilettes intelligentes permettant de déterminer le poids de la personne et repérer une éventuelle perte de poids, des capteurs ECG intégrés à la baignoire qui permettent de mesurer l'activité cardiaque à chaque bain ou encore des capteurs thermiques disposés sur le matelas pour mesurer l'activité nocturne et détecter l'agitation.

Aux États-Unis, on peut souligner le projet Gator Tech « Smart House » de l'Université de Floride, qui propose un habitat intelligent pour le confort et le suivi des personnes à domicile. Ce projet de grande ampleur utilise de nombreux services intelligents : boîte aux lettres, porte d'entrée, baignoire, sol, lit, etc. dont l'objectif est de faciliter la vie de l'occupant en se servant des murs ou du miroir de la salle de bain comme pense-bête électronique par exemple. Bien que ses objectifs ne soit pas purement médicaux, il dispose d'un ensemble de capteurs ultra-sons permettant de localiser précisément l'occupant et d'avoir également des informations sur son orientation et ses mouvements. Un plancher intelligent permet de détecter les chutes dans la cuisine et peut participer au processus de localisation. Le projet Gator Tech définit une architecture de services sensibles au contexte permettant à l'environnement de s'adapter en fonction des conditions rencontrées. Ainsi, les auteurs prennent pour exemple le four à micro-onde qui refuse de réchauffer une pizza lorsque l'occupant doit suivre un régime.

La détection d'activité est une problématique qui revient très fréquemment dans le domaine du maintien à domicile comme le montrent notamment les travaux de Hoey *et al.*, (2010) à travers le projet COACH. Ce projet vise à aider des patients atteints de démence à se laver les mains correctement. Pour cela, il observe un évier à l'aide d'une caméra et donne des indications régulières à la personne à travers un écran et un haut-parleur. Philipose *et al.* (2004) proposent d'équiper les objets du quotidien de capteurs RFID pour déterminer les activités d'une personne. Celle-ci, équipée d'un gant pourvu d'un lecteur RFID déclenchera un signal en interagissant avec les objets. D'autres exemples de projets sur la détection d'activités peuvent être trouvés dans (Mihailidis *et al.*, 2012).

Pour résumer, ces projets diffèrent essentiellement sur les moyens technologiques mis en œuvre. Les informations récoltées se recoupent assez souvent. Ainsi, le suivi des déplacements et la mesure de l'agitation sont proposés par quasiment tous les projets précédents comme le montre le tableau 1. Ces quelques projets ne sont bien sûr qu'un petit aperçu des efforts menés au niveau mondial. Pour les personnes qui le souhaitent, une présentation plus exhaustive de systèmes de télésurveillance peut être retrouvée dans (Chan *et al.*, 2003).

Dans le cadre de MASKEP, nous nous intéressons à ces informations collectées et notamment aux informations de haut niveau issues des interprétations fournies par chaque système. Ces informations, dont un exemple peut être retrouvé dans le tableau 1, vont permettre de suivre l'évolution de l'état d'individus au cours du temps et de constituer des catégories de population présentant les mêmes

pathologies dans le but de pouvoir anticiper l'évolution de l'état de santé d'un individu. Ainsi, MASKEP se veut être un complément des systèmes de télésurveillance existants qui s'avèrent indispensables pour la détection de situation d'urgence. La section suivante explique comment notre système traite ces données relevées de façon automatique, en les associant avec des données issues d'autres sources.

Tableau 1. Indicateurs remontés par quelques systèmes de télé-vigilance

Projet	Équipement	Informations remontées	Info. de haut niveau
PROSAFE	- Détecteur de mouvements IR	- Mouvements de l'occupant - Date de détection	- déplacements - occupation des zones - Agitation nocturne
TéléPat + GARDIEN	- Capteurs d'actimétrie - Capteurs physiologiques - Capteurs inertiels - Détecteurs acoustiques - Détecteur de mouvements IR	- rythme cardiaque - Position - Bruits - Mouvements - Date	- Posture de la personne - Déplacements - Détection de chute - Déplacements - Vitesse de déplacement
AILISA	- Pèse personne sans fil - Gilet intelligent + capteurs physiologiques - Robot déambulateur	- Poids - rythmes cardiaque et respiratoire - tension	- Niveau d'agitation
CarerNet	- Capteurs physiologiques (ECG...) - Capteurs environnementaux (fumée, bruits...) - Capteurs de style de vie	- pression sanguine - Rythme cardiaque et respiratoire - température corporelle - Température de la pièce, bruit - Mouvements, occupation du lit	- Temps passé au lit - Déplacements - Taux de mobilité - Fréquence de chute - Agitation nocturne
Tamura <i>et al.</i>	- Capteurs de température - Toilettes intelligentes - Capteurs ECG - Capteurs domotiques - Plancher avec capteur de pression	- Poids - Rythme cardiaque - Position dans le lit - Activités - Positions	- Evolution du poids - Problèmes cardiaques - Agitation nocturne - Localisation - Détection de chute
Gator Tech	- Capteurs de positions à ultrasons		- Contextes utilisateurs

3. Modèle multi-agent

Les données collectées au domicile des personnes suivies sont de plusieurs natures. Certaines sont des indicateurs médicaux auxquels seul du personnel médical a accès, d'autres sont des informations générales sur l'état de la personne et sur son comportement, auxquelles ont accès d'autres types d'intervenants (aide-ménagère, famille, personnel aidant...). Parmi ces indicateurs, certains sont collectés automatiquement par des systèmes de télé-vigilance, d'autres sont le fruit de rapports textuels rédigés par les aidants, voire par la personne surveillée elle-même. Si l'on ajoute à ce tableau que les personnes surveillées ont la liberté de s'abonner à plusieurs systèmes de télé-vigilance, ainsi que de faire appel aux professionnels de leur choix, nous obtenons une grande diversité de données potentiellement récoltées, distribuées entre plusieurs systèmes et intervenants.

Pour gérer cette large distribution des données, les systèmes multi-agents sont particulièrement pertinents (El Fallah Seghrouchni *et al.*, 2009). Leur architecture permet de décomposer le système en plusieurs agents, chacun s'occupant d'analyser la partie des données portée à sa connaissance, et d'utiliser les capacités de communication et d'auto-organisation des agents pour obtenir une vue consolidée des données.

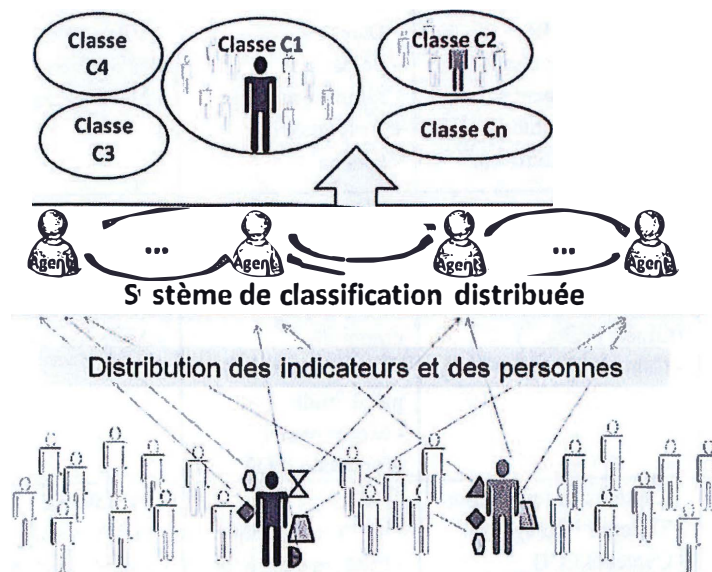


Figure 2. Classification distribuée

Le modèle que nous avons développé est basé sur des agents ayant la capacité d'utiliser des méthodes de classification destinées à classer les données dont ils

disposent. Nous appelons cela une classification locale. Les agents ont également la capacité de mettre en commun leurs données et de réaliser une classification des agents eux-mêmes. Cette classification rend compte des similitudes entre agents dans leur façon d'évaluer leurs données. Elle n'a de sens que quand les agents évaluent des données communes. Dans cet article, nous associons les agents à des individus et nous montrons comment cette dernière étape nous permet de comparer des personnes entre elles. La figure 2 illustre ce fonctionnement.

Une classification multi-agent en trois étapes

Soit A_1, \dots, A_n les agents de notre système. Chaque agent A_i a connaissance d'un ensemble d'objets P_1, \dots, P_m évalués selon un certain nombre d'attributs numériques X_1, \dots, X_r . Notre méthode de classification repose sur trois étapes.

La première consiste pour chaque agent à réaliser une classification locale de ses objets. En utilisant l'algorithme de classification non supervisé ISODATA, il partitionne l'ensemble de ses objets en k ensembles appelés clusters, chaque cluster étant caractérisé par un vecteur-moyen. Pour les lecteurs non familiers des algorithmes des classifications, notons que l'algorithme ISODATA a les mêmes objectifs que l'algorithme k -means mais contrairement à ce dernier où le nombre de clusters est fixé *a priori*, il tente d'optimiser le nombre de clusters en fonction des objets à classer. Pour une description de l'algorithme ISODATA on peut par exemple se référer à Memarsadeghi *et al.*, (2006).

La seconde étape consiste pour un agent ayant établi une classification locale, à se mettre en contact avec d'autres agents ayant classé les mêmes données (ou un sous-ensemble de ces données) suivant des critères proches des siens (mais non forcément identiques) et à former un groupe avec eux. Nous appelons cette étape l'appel à participation.

Les agents répondant positivement à l'appel d'un autre agent sont ceux ayant classé des objets communs aux siens. Plus formellement, chaque agent constitue au départ un groupe auquel lui seul appartient. Sont intégrés à ce groupe, tous les agents ayant au moins un objet commun avec lui. Notons donc qu'au final, un agent peut faire partie de plusieurs groupes : Celui qu'il a lui-même constitué, et tous ceux auquel il a été intégré dans le cadre de l'appel à participation.

La dernière étape est l'étape de la classification inter-agents. Les agents d'un même groupe calculent une nouvelle classification à partir des données suivantes :

- Les objets à classer sont les agents du groupe
- Ces objets sont évalués selon le vecteur [$P_1(X_1), \dots, P_1(X_r), P_2(X_2), \dots, P_2(X_r), \dots, P_m(X_1), \dots, P_m(X_r)$]

La classification ainsi constituée aboutit à une classification des agents. Cette classification prend tout son sens dans le cas où un agent s'occupe de relever les variations des indicateurs d'une personne au cours du temps. En effet dans un tel cadre :

– Les objets classés localement par un agent sont des unités de temps représentant la valeur des indicateurs de la personne suivie à cet instant. La classification locale permet donc de regrouper les périodes de temps où la personne suivie est dans un certain état. Cette information permet de suivre l'évolution de la personne en fonction du temps qui passe.

– Un groupe d'agents est constitué des agents ayant suivi différentes personnes sur au moins une même unité de temps telle que définie ci-dessus. Dans notre contexte, le groupe sera d'autant plus pertinent qu'il y aura d'unité de temps en commun dans le groupe.

– La classification inter-agent représente quant à elle un classement des agents et à travers eux des personnes suivies. Elle permet de regrouper les personnes dont l'évolution est similaire et de les attacher à des catégories dont un expert peut avoir fait une interprétation sémantique.

En d'autres termes, la classification locale permet de suivre l'évolution d'une personne au cours du temps, quand la classification inter-agent permet de comparer les personnes les unes aux autres. La section 4.2. montre un exemple détaillé du fonctionnement des deux étapes de la classification, tandis que la section 5 illustre leur utilisation concrète dans le cas du suivi de patients souffrant de la maladie d'Alzheimer.

4. Classification

La principale difficulté qu'ont à résoudre nos algorithmes de classification vient de l'hétérogénéité des données à classer. En effet les indicateurs remontant des différentes plateformes de télé-vigilance sont de natures différentes, ce qui amène les personnes à être évaluées selon des critères différents. De plus, toutes les personnes ne sont pas évaluées suivant la même échelle de temps, ce qui peut mener à des attributs manquants lors des classifications inter-agents.

Dans cette section nous décrivons comment nous adaptons les algorithmes traditionnels de classification pour les rendre capables de gérer des données qui ne sont pas toutes décrites avec les mêmes attributs.

4.1. Classification hétérogène

Nous considérons ici le problème consistant à classer un ensemble de points dans un espace à plusieurs dimensions dans lequel certaines coordonnées de ces points ne sont pas définies. Cela peut se produire en cas de données incomplètes (du fait de la méthode de collecte par exemple) ou plus simplement quand les points ont des dimensions différentes, autrement dit quand on essaie de classer des points d'un espace à n dimensions avec des points issus d'un espace à m dimensions, n étant différent de m . Les méthodes classiques de classification, comme k -means ou ISODATA, ont besoin de toutes les coordonnées des points pour fonctionner. Le cœur de leurs algorithmes est basé sur l'évaluation de distances entre points et sur le

calcul de vecteurs moyens. Nous adaptons la façon dont ces méthodes effectuent leur calcul pour prendre en compte le fait que certaines données ne sont que partiellement décrites. Nous appelons cela une classification hétérogène dans le sens où les données à classifier proviennent d'espaces de dimensions différentes.

Plus formellement, soit un ensemble de points (x_1, x_2, \dots, x_n) où chaque observation est un vecteur réel de dimension d . Classifier ces points consiste à partitionner les n points en k ensembles ($k < n$) $S = \{S_1, S_2, \dots, S_k\}$ avec certains objectifs à atteindre, comme par exemple minimiser la somme des distances entre les points appartenant au même ensemble (MacQueen, 1967). Nous nous focalisons sur deux algorithmes de classification célèbres : k -means et ISODATA. Ces deux algorithmes se basent sur le calcul de distances et de vecteurs centres. La littérature décrivant ces deux algorithmes est très fournie, on pourra par exemple se référer à (Hartigan, 2002 ; Kanungo *et al.*, 2002 ; Memarsadeghi *et al.*, 2007).

Pour la compréhension de cet article, voici une vue générale de l'algorithme k -means :

Étape 1. Choisir k , le nombre de classes finales.

Étape 2. Choisir n'importe quelle partition répartissant les données dans ces k classes.

Étape 3. Pour chaque point, calculer sa distance au point-centre (centroïde) de chacune des classes. Si ce point n'appartient pas à la classe où cette distance est la plus petite, retirer le point de sa classe actuelle et le mettre dans la classe dont il est le plus proche. Mettre alors à jour le centroïde des deux classes ayant respectivement gagné et perdu le point.

Étape 4. Répétez l'étape précédente jusqu'à ce que l'algorithme converge, c'est-à-dire jusqu'à ce que l'étape 3 ne génère plus de changement de classes.

Les deux principales opérations de cet algorithme sont le calcul de la distance entre un point et le centroïde d'une classe, et la mise à jour de ce centroïde dès lors que la classe est modifiée.

Pour ce qui est de la distance entre deux points : Soit $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$ un point représenté comme un vecteur dans un espace réel à d dimensions. Nous désignons par x_{in} la $n^{\text{ième}}$ coordonnée de x_i . La distance entre deux points est définie par :

$$d(x_i, x_j) = \sum_{k=1}^d (x_{ik} - x_{jk})^2$$

La distance ne peut pas être calculée si un des x_{ik} ou des x_{jk} n'est pas connu. Nous proposons une adaptation de cette méthode de calcul pour la rendre capable de prendre en compte le cas où les x_i et x_j possèdent des coordonnées non définies.

Soit c_i l'ensemble de coordonnées définies de x_i et soit c_j l'ensemble des coordonnées définies de x_j . Notre nouvelle façon de définir la distance entre deux points est :

$$d(x_i, x_j) = \sum_{k \in c_i \cap c_j} (x_{ik} - x_{jk})^2$$

Autrement dit, la nouvelle distance est calculée en prenant uniquement en compte les coordonnées communes de x_i et de x_j . Si l'ensemble des coordonnées communes est vide, la distance n'est pas définie. Pour cette raison, pour que la classification puisse avoir lieu, les points doivent tous avoir en commun au moins une coordonnée. Plus formellement, il doit exister une coordonnée i , où x_i est définie pour tous les points.

Pour ce qui est du calcul du centroïde d'une classe : Pour calculer le centroïde V d'un ensemble S de n points, le calcul est le suivant :

$$V_i = \frac{1}{n} \sqrt{\sum_{x \in S} x_i} \text{ où } V_i \text{ est la } i^{\text{ème}} \text{ coordonnée du vecteur } V.$$

L'adaptation au fait que certains des x_i ne sont pas définis se fait en changeant la méthode de calcul de la façon suivante : soit n' le nombre de points pour lesquels la coordonnée i est définie, et soit S' l'ensemble de ces points. La coordonnée i du centroïde est désormais calculée par :

$$V_i = \frac{1}{n'} \sqrt{\sum_{x \in S'} x_i}$$

Autrement dit, pour calculer la coordonnée i du centroïde, nous faisons la moyenne des coordonnées i de tous les points où cette coordonnée est définie. Si l'ensemble S' est vide, la coordonnée i du centroïde n'est pas définie. Étant donné que nous avons imposé, pour que la classification puisse avoir lieu, qu'une des coordonnées soit définie pour tous les points, nous savons qu'au moins une des coordonnées du centroïde est définie.

Afin de résumer les adaptations proposées, quand un calcul a besoin d'être effectué sur une coordonnée qui n'est pas définie, le point correspondant est neutralisé. S'il s'agit d'un calcul de distance, cela signifie que la distance à ce point compte pour zéro. S'il s'agit d'un calcul de moyenne, cela signifie que le point n'est pas pris en compte en tant que membre de l'échantillon.

Cette technique travaillant sur un sous-ensemble des données initiales nous permet de réaliser une classification impossible à calculer avec les algorithmes traditionnels. La fiabilité des résultats obtenus dépend ensuite du niveau d'interdépendance entre les coordonnées retenues et celles mises de côté. Nous détaillons ce point dans la section 5.2.

Illustrons dans un premier temps, via un exemple simple, quel est l'impact de nos modifications sur une classification.

Soit un ensemble de données composé de deux points P_1 et P_2 définis dans un espace à trois dimensions et de deux points P_3 et P_4 définis dans un espace à deux dimensions. Les coordonnées de ces points sont : $P_1 \{0, 1, 1\}$, $P_2 \{0, 3, 1\}$, $P_3 \{0, \text{undef}, 1\}$ and $P_4 \{0, \text{undef}, 4\}$.

Nous pouvons constater que la seconde coordonnée des points P_3 et P_4 n'est pas définie et que les coordonnées un et trois sont définies pour tous les points. Notre contrainte imposant qu'au moins une des coordonnées soit définie pour tous les points est par conséquent vérifiée. La classification de ces quatre points selon l'algorithme k-means paramétré pour obtenir deux clusters, produit les résultats suivants (obtenu 22 fois sur 30 exécutions pour éviter le problème des optima locaux) :

Cluster 1	Cluster 2
centroïde $\{0, 2, 1\}$	centroïde $\{0, \text{undef}, 4\}$
$P_1 [0, 1, 1], P_2 [0, 3, 1], P_3 [0, \text{undef}, 1]$	$P_4 [0, \text{undef}, 4]$

La deuxième coordonnée du centroïde du cluster 2 est indéfinie car cette coordonnée est également indéfinie pour le seul point contenu dans ce cluster. Toutes les coordonnées du centroïde du cluster 1 sont définies car les trois points faisant partie de ce cluster définissent au moins une fois chacune des coordonnées.

Si nous appliquons à présent nos méthodes de calcul à l'algorithme ISODATA paramétré avec une distance minimale entre cluster de 1, une distance maximale intra-cluster de 0.3 et un nombre maximal de cluster de 4, nous obtenons :

Cluster 1	Cluster 2	Cluster 3
centroïde $\{0, 1, 1\}$	centroïde $\{0, \text{undef}, 4\}$	centroïde $\{0, 3, 1\}$
$P_1 [0, 1, 1], P_3 [0, \text{undef}, 1]$	$P_4 [0, \text{undef}, 4]$	$P_2 [0, 3, 1]$

On constate que les paramètres fournis à ISODATA ont produit une classification à trois clusters dans laquelle P_4 et P_2 ont été isolés.

Nous n'avons pas testé nos méthodes de calcul sur d'autres types d'algorithmes de classification car dans le contexte de cet article, les algorithmes k-means et ISODATA satisfont nos besoins. Cependant, la méthode peut être adaptée à toute méthode de classification reposant sur des approximations de problèmes de type k-centres, k-médiane ou k-moyen.

4.2. Exemple détaillé

Dans cette section, nous appliquons notre algorithme de classification sur un jeu de test composé de 4 agents classant chacun trois objets P_1 , P_2 et P_3 évalués selon

trois attributs X_1 , X_2 , et X_3 (tableau 2). Nous considérons que contrairement aux trois autres, l'agent A_4 n'a pas eu connaissance de l'attribut X_3 .

Tableau 2. Quatre agents et leurs objets

Agent A1				Agent A2			
Objets/ Attributs	P ₁	P ₂	P ₃	Objets/ Attributs	P ₁	P ₂	P ₃
X ₁	1	2	6	X ₁	1	3	7
X ₂	2	2	5	X ₂	2	6	5
X ₃	3	1	4	X ₃	3	7	1

Agent A3				Agent A4			
Objets/ Attributs	P ₁	P ₂	P ₃	Objets/ Attributs	P ₁	P ₂	P ₃
X ₁	1	6	6	X ₁	2	3	7
X ₂	3	4	1	X ₂	5	4	1
X ₃	4	5	7	X ₃	<i>undef</i>	<i>undef</i>	<i>undef</i>

En appliquant une méthode de classification locale (ISODATA dans notre cas), chaque agent construit sa propre partition. Chaque classe de cette partition est caractérisée par un vecteur centre (le centroïde) calculé par ISODATA. Le résultat de cette classification locale est :

Agent A ₁ : C ₁ ={P ₃ }, centroïde {X ₁ =6; X ₂ =5; X ₃ =4} C ₂ ={P ₁ , P ₂ }, centroïde {X ₁ =1.5; X ₂ =2; X ₃ =2}	Agent A ₂ : C ₁ ={P ₁ }, centroïde {X ₁ =1; X ₂ =2; X ₃ =3} C ₂ ={P ₂ }, centroïde {X ₁ =3; X ₂ =6; X ₃ =7} C ₃ ={P ₃ }, centroïde {X ₁ =7; X ₂ =5; X ₃ =1}
Agent A ₃ : C ₁ ={P ₂ , P ₃ }, centroïde {X ₁ =6; X ₂ =2.5; X ₃ =6} C ₂ ={P ₁ }, centroïde {X ₁ =1; X ₂ =3; X ₃ =4}	Agent A ₄ : C ₁ ={P ₁ , P ₂ }, centroïde {X ₁ =2.5; X ₂ =4.5; X ₃ =-1} C ₂ ={P ₃ }, centroïde {X ₁ =7; X ₂ =1; X ₃ =-1}

L'étape de l'appel à participation de notre algorithme décrit dans la section 3.1 mène à la constitution de quatre groupes d'agents identiques, les agents ayant tous les mêmes objets à classer. Chaque groupe calcule une classification basée sur les données du tableau 3.

Tableau 3. Matrice utilisée pour la classification inter-agents

Attributs/ Points	X ₁ P ₁	X ₁ P ₂	X ₁ P ₃	X ₂ P ₁	X ₂ P ₂	X ₂ P ₃	X ₃ P ₁	X ₃ P ₂	X ₃ P ₃
A ₁	1	2	6	2	2	5	3	1	4
A ₂	1	3	7	2	6	5	3	7	1
A ₃	1	6	6	3	4	1	4	5	7
A ₄	2	3	7	5	4	1	<i>undef</i>	<i>undef</i>	<i>undef</i>

La classification calculée produit alors trois classes, la première regroupant les agents A₃ et A₄:

$$C_1 = \{A_3, A_4\} \text{ centroïde } \{1.5, 4.5, 6.5, 4, 4, 1, 4, 5, 7\}$$

$$C_2 = \{A_2\} \text{ centroïde } \{1, 3, 7, 2, 6, 5, 3, 7, 1\}$$

$$C_3 = \{A_1\} \text{ centroïde } \{1, 2, 6, 2, 2, 5, 3, 1, 4\}$$

Cette classification indique que les caractéristiques des objets P₁, P₂ et P₃ connues par les agents A₃ et A₄ présentent certaines concordances que l'on ne retrouve pas chez A₁ et A₂. La section suivante illustre comment ces concordances peuvent être exploitées dans un cas concret.

5. Illustration

5.1. Suivi de personnes atteintes par la maladie d'Alzheimer

L'objet de cette section est d'illustrer le fonctionnement de notre système en prenant pour contexte le suivi à domicile de personnes atteintes de la maladie d'Alzheimer. Cet exemple va en particulier nous permettre de montrer l'intérêt de nos deux types de classification : locale et multi-agent.

La maladie d'Alzheimer est la plus répandue d'un groupe important de maladies appelées maladies neuro-dégénératives. Ces maladies affectent le cerveau et sont caractérisées par une détérioration des facultés cognitives et de la mémoire causée par la dégénérescence progressive des cellules du cerveau. Comme toute maladie neuro-dégénérative la maladie d'Alzheimer est caractérisée par des pertes de mémoire, des modifications du jugement et du raisonnement, des changements d'humeur et de comportement.

L'évolution de la maladie d'Alzheimer présente certains stades d'évolution. Cependant la maladie affecte chaque personne différemment : les symptômes, l'ordre de leur apparition et la durée de chacun des stades de la maladie varient selon les personnes. Dans la plupart des cas, la maladie d'Alzheimer progresse lentement et les symptômes peuvent chevaucher plus d'un stade. La durée de la maladie est généralement de sept à dix ans mais peut être plus longue chez certaines personnes.

Elle est traditionnellement découpée en trois stades selon l'échelle de détérioration globale (EDG). Les différentes phases se distinguent par le degré de détérioration de certaines capacités : mentales (perte de mémoire, reconnaissance des visages, orientation dans l'espace), émotionnelles (sautes d'humeur, anxiété, dépression), comportementales (concentration, agressivité, passivité) et physiques (coordination, appétit, sommeil). Dans notre exemple, les personnes sont évaluées selon trois indicateurs :

– **I₁ : Le nombre d'oublis par jour.** La perte de mémoire est caractéristique de l'évolution de la maladie. Les oublis bénins et peu fréquents de la phase initiale s'aggravent dans la phase intermédiaire où la personne peut progressivement oublier son passé, ne plus reconnaître ses amis et sa famille, jusqu'à ne plus être capable de communiquer dans la phase avancée.

– **I₂ : La mobilité.** Un patient atteint de la maladie d'Alzheimer peut souffrir de désorientation spatiale et temporelle. La détermination du trajet de la personne dans son environnement quotidien permet d'observer sa désorientation spatiale : en effet une personne dont l'étude du trajet montre qu'elle passe de pièce en pièce, qu'elle « tourne en rond », permet de suspecter une désorientation spatiale (Larab *et al.*, 2008). La phase la plus avancée de la maladie se caractérise au contraire par une augmentation de la passivité et de l'alitement, qui produit une perte de mobilité. La valeur de cet indicateur sera dans notre exemple le nombre de franchissements de seuil pendant la période considérée.

– **I₃ : Le niveau de bêta-amyloïde en pg/ml.** La mesure de ce marqueur biologique dans le liquide céphalo rachidien ou dans le sérum des patients atteints de la maladie d'Alzheimer permet d'augmenter la précision du diagnostic précoce mais aussi celle du suivi de l'évolution de la pathologie. Plus le taux est faible (< 192 pg/ml) plus la personne est susceptible d'être dans une phase avancée de la maladie.

On notera la nature très différente de chacun de ces indicateurs dont les valeurs seront tantôt obtenues par les aidants en ce qui concerne le nombre d'oublis, des capteurs de mouvement pour la mobilité, et le corps médical pour les marqueurs biologiques. Dans ce cas de figure, nous simulons le suivi de six personnes pendant 12 ans en faisant la moyenne sur l'année des indicateurs décrits ci-dessus. Quatre sont atteintes de la maladie d'Alzheimer (P1 à P4). Les deux dernières ne le sont pas, et P6 se distingue de P5 par une mobilité qui augmente avec le temps. Notre système affecte un agent à chaque personne pour suivre tous ses indicateurs.

Le tableau 4 donne les valeurs des indicateurs collectées par l'agent 1 pendant 12 ans. Après normalisation des valeurs entre zéro et un, l'agent 1 effectue une classification des différentes années de la personne suivie. Cette classification locale mène aux résultats donnés dans le tableau 5.

Cette classification fait apparaître trois classes. La classe 1 regroupe les quatre premières années, la classe 2 concerne les huit qui suivent et enfin la classe 3 contient les deux dernières. L'interprétation qui peut en être faite consiste à identifier trois phases dans l'évolution temporelle de la valeur des indicateurs,

phases qui peuvent alerter sur une possible évolution de la maladie. Ici, la classification locale est utilisée pour mesurer l'évolution de l'état de la personne en fonction du temps, selon les variations de la valeur des indicateurs suivis.

Tableau 4. Indicateurs de la personne 1

Années/Indicateurs	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆
I ₁	0	1	1	0	2	2
I ₂	210	205	220	207	180	181
I ₃	1500	1599	1400	1600	1800	1860
Années/Indicateurs	A ₇	A ₈	A ₉	A ₁₀	A ₁₁	A ₁₂
I ₁	1	2	2	2	5	4
I ₂	183	182	180	185	149	151
I ₃	1700	1760	1700	1680	380	250

Tableau 5. Classification locale de la personne 1

	Agent 1		
	I ₁	I ₂	I ₃
C ₁ (A ₁ ,A ₂ ,A ₃ , A ₄)	0.08	0.61	0.65
C ₂ (A ₅ ,A ₆ ,A ₇ , A ₈ ,A ₉ , A ₁₀)	0.31	0.33	0.76
C ₃ (A ₁₁ ,A ₁₂)	0.75	0.02	0.14

Par comparaison, les tableaux 6 et 7 présentent les données de la personne numéro 2 et la classification qui résulte de ces données. On notera que là encore, trois phases sont identifiées par la classification, mais que ces phases ne sont pas réparties de la même façon, témoignant du rythme différent d'évolution de la maladie par rapport à la personne 1.

Tableau 6. Données de la personne numéro 2

Années/Indicateurs	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆
I ₁	1	0	3	2	3	4
I ₂	210	220	191	190	192	190
I ₃	1400	1450	1650	1700	1720	1730
Années/Indicateurs	A ₇	A ₈	A ₉	A ₁₀	A ₁₁	A ₁₂
I ₁	3	4	5	6	5	5
I ₂	188	187	150	148	148	147
I ₃	1725	1729	1400	1300	400	200

Tableau 7. Classification locale de la personne numéro 2

	Agent 2		
	I ₁	I ₂	I ₃
C ₁ (A ₁ ,A ₂)	0.08	0.66	0.62
C ₂ (A ₃ ,A ₄ ,A ₅ ,A ₆ ,A ₇ ,A ₈)	0.53	0.41	0.74
C ₃ (A ₉ , A ₁₀ ,A ₁₁ ,A ₁₂)	0.88	0.01	0.36

La prochaine étape, à savoir la classification interagents, a pour but de pouvoir comparer entre elles les évolutions des personnes suivies. Les six agents vont chercher à se regrouper suivant la méthode donnée section 3.1. Une fois le groupe constitué, les objets à classer sont représentés comme une entité évaluée selon 36 (12 * 3) indicateurs. Les douze premiers indicateurs correspondent aux valeurs de I₁ pour chacune des douze années. Les douze qui suivent correspondent aux valeurs de I₂ pour chacune des douze années, jusqu'aux douze derniers qui correspondent aux valeurs de I₃ pour chacune des douze années. Chaque agent ayant surveillé ici une personne pendant douze ans, les groupes constitués sont identiques pour tous les agents. Ils auraient été différents si, par exemple, certaines personnes avaient été suivies tous les mois sur un an, au lieu de tous les ans sur douze ans. Auquel cas les différents groupes auraient été constitués des agents ayant suivi des personnes selon les mêmes unités de temps.

Pour illustrer ce qui vient d'être dit, le tableau 8 présente une partie des données connues par les six agents une fois le regroupement réalisé.

Tableau 8. Données des personnes mesurées selon 36 indicateurs

	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆
I _{1A1}	0	1	1	1	0	0
I _{1A2}	1	0	0	1	1	1
...
I _{1A12}	4	5	6	6	1	0
I _{2A1}	210	210	250	245	250	250
...
I _{2A12}	151	147	148	144	252	250
I _{3A1}	1500	1400	1600	1450	1500	1490
...
I _{3A12}	250	200	480	550	1590	2200

Le tableau complet n'apportant rien à la compréhension de l'article, nous avons préféré dans un but de concision, nous en tenir à ce tableau partiel. Pour clarifier les choses I_1A_1 représente le nombre d'oublis journaliers pour l'année 1, quand I_3A_{12} représente le nombre de franchissement de seuil dans l'année douze. Chaque agent est alors capable de calculer une classification des six personnes de l'exemple. Cette classification partitionne nos données en trois classes de deux personnes chacune, dont la composition et une partie du vecteur-centre sont données dans le tableau 9.

Tableau 9. Classification des six personnes selon l'évolution de leurs indicateurs

	I_{1A1}	I_{1A2}	...	I_{1A12}	I_{2A1}	...	I_{2A12}	I_{3A1}	...	I_{3A12}
$C_1(P_1, P_3)$	0.08	0.08	...	0.83	0.78	...	0.06	0.67	...	0.16
$C_2(P_2, P_4)$	0.17	0.08	...	0.92	0.75	...	0.02	0.62	...	0.16
$C_3(P_5, P_6)$	0	0.17	...	0.08	0.96	...	0.96	0.65	...	0.82

Cette classification est obtenue en utilisant l'algorithme ISODATA paramétré avec un nombre initial de cluster de 5, un nombre de points minimum dans un cluster de 1, un nombre maximal d'itérations de 10, une distance maximale dans un cluster de 0,8, et une distance minimale entre deux clusters de 1. Elle regroupe P_1 avec P_3 (les deux sont restés 4 ans en phase initiale, 8 ans en phase intermédiaire, et 2 ans en dernière phase), P_2 avec P_4 (phase initiale de 2 ans, phase intermédiaire de 6 ans et dernière phase de 8 ans) et P_5 avec P_6 (non atteints par la maladie).

En définitive, chaque agent est capable de distinguer les différentes phases de l'évolution de l'état d'une personne. La communication entre agents permet quant à elle, de situer cette évolution par rapport à un ensemble de personnes.

5.2. Données incomplètes

Essayons à présent de voir ce que peut donner la classification multi-agent en cas de données incomplètes et en utilisant la classification hétérogène décrite section 4.

Prenons le cas de P_1 et posons comme hypothèse que l'évolution de son état est en cours d'analyse et que nous en sommes à l'année 6. Les données des années 7 à 12 sont donc inconnues. En ce qui concerne la classification multi-agent, cela se traduit par l'absence de valeurs pour la série d'indicateurs I_{1A7} - I_{1A12} , I_{2A7} - I_{2A12} , I_{3A7} - I_{3A12} .

Pour autant la classification hétérogène donne la classification décrite dans le tableau 10. Cette classification opère les mêmes regroupements que la classification où toutes les données sont connues, ce qui est un bon résultat.

Tableau 10. Classification hétérogène avec P_1 ayant uniquement les 6 premiers ans connus

	I_{1A1}	I_{1A2}	...	I_{1A12}	I_{2A1}	...	I_{2A12}	I_{3A1}	...	I_{3A12}
$C_1 (P_1, P_3)$	0.08	0.08	...	1	0.78	...	0.04	0.67	...	0.21
$C_2 (P_2, P_4)$	0.17	0.17	...	1	0.91	...	0.01	0.63	...	0.16
$C_2 (P_5, P_6)$	0	0.17	...	0.08	0.96	...	0.96	0.65	...	0.82

Pour aller plus loin, nous avons testé le nombre de différences entre les deux classifications en fonction du nombre d'années inconnues de P_1 . Les résultats sont donnés dans le tableau 11.

Tableau 11. Différences entre classifications en fonction du nombre de données inconnues

Nombres d'années inconnues	0	1	2	3	4	5	6	7	8	9	10	11
Personnes mal classées	0	0	0	0	0	0	0	1	1	1	1	1

On peut voir qu'à partir de sept années inconnues, P_1 n'est plus correctement classé. Les résultats le classe alors avec P_2 et P_4 (au lieu de P_3). Cela signifie que dès l'année 6, le système fait correctement le rapprochement entre P_1 et P_3 . L'évolution de P_3 peut donc être utilisée pour prévoir celle de P_1 et par conséquent déclencher une action si cette évolution n'est pas souhaitable.

Pour conclure cette partie sur les données incomplètes, précisons que notre système travaille essentiellement sur des données de haut niveau issues d'une première interprétation effectuée en local. Cette interprétation locale a pour effet de calibrer les capteurs de bas niveau, de consolider leurs données en données de plus haut niveau et de les normaliser. Cette phase de prétraitement est essentielle car, si la classification locale n'est pas impactée par la diversité des capteurs, la pertinence de la classification interagents dépend d'un nombre minimum d'indicateurs communs ou interdépendants.

En effet les indicateurs non communs génèrent des données incomplètes lors de la classification interagents. Au-delà d'un certain nombre de données incomplètes non contrebalancées par des données corrélées, et même si nos algorithmes rendent possible la classification, la pertinence des résultats se dégrade fortement. Il est difficile de placer un seuil limite au-delà duquel le nombre de données incomplètes rend inintéressante toute tentative de classification, ce seuil étant fortement lié au contexte.

6. Positionnement de notre approche dans le paradigme de classification : critères discriminants et originalité

La méthode de classification multi-agent que nous avons proposée permet de comparer des classifications distribuées entre elles, en étant en particulier capable de travailler sur des données partiellement décrites. Dans le domaine des multi-agents, d'autres travaux se sont intéressés à la problématique de la classification. Pour les comparer, nous avons déterminé cinq critères :

- 1– le type de distribution possible : la répartition des données chez les agents peut être verticale (distributions des attributs) et/ou horizontale (distribution des objets) ;
- 2– les méthodes de classifications prises en charge ;
- 3– les compétences des agents chargés de la classification ;
- 4– les types de coopération mis en œuvre : comment la réponse est construite collectivement à partir des réponses partielles des agents ;
- 5– le domaine d'application.

La distribution est une des capacités intrinsèques des systèmes multi-agents (SMA). Nous nous sommes donc intéressés aux SMA dédiés aux problèmes de classification dynamique. Nous avons sélectionné quatre systèmes finalisés se rapprochant de nos objectifs : Samarah qui utilise une méthode non supervisée de collaboration multistratégie, NeurAge qui aborde le problème de la distribution verticale, le système de S. Mukhopadhyay à base de listes d'accointances et enfin le système de Quteishat qui implémente la négociation entre classifieurs.

Samarah est un système d'apprentissage hybride qui utilise une méthode de classification collaborative à plusieurs stratégies (Wemmert *et al.*, 2002 ; Gançarski *et al.*, 2007). Il est basé sur l'éventuelle complémentarité des différents classifieurs. Ainsi, les auteurs postulent que la combinaison de différentes méthodes de classification peut accroître l'efficacité de la classification et la précision des résultats. Le système gère donc différentes méthodes de classification non supervisée. Chaque agent a sa méthode. Le système combine ensuite les réponses des agents et produit une réponse commune qui est le résultat d'un consensus entre tous les agents. En cas de conflit, deux agents peuvent utiliser des opérateurs comme diviser, fusionner ou reclasser (Forestier *et al.*, 2008).

Comme dans notre système, les agents coopèrent pour affiner leurs partitions respectives. Mais le but est différent : Samarah peut effectuer une classification d'objets complexes ayant beaucoup d'attributs ; un exemple de classification d'images hétérogènes pour la compréhension de scène est donné. Cependant, les agents doivent connaître tous les attributs de chaque objet.

NeurAge, et son successeur ClassAge, sont des systèmes multiclassifieurs, composés de plusieurs agents neuronaux, tous poursuivant un même but (Santana *et al.*, 2006). Quand une observation est soumise au système, les agents produisent leur sortie. Ils communiquent ensuite entre eux et utilisent une méthode de négociation

qui se déroule sur plusieurs tours et qui est fondée sur la notion de confiance. Un agent A vérifie l'information donnée par un autre agent B pour une observation-test. Après cette vérification, le degré de confiance de A vers B peut diminuer. L'agent contrôlé peut alors être amené à quitter le tour. L'agent ayant le plus haut degré de confiance est désigné comme étant le plus approprié pour classer l'observation-test. Le système étendu ClassAge peut aussi utiliser des agents non neuronaux. Les caractéristiques des deux systèmes sont les mêmes (Canuto *et al.*, 2008).

Cette méthode permet l'utilisation d'une distribution de données verticale dans laquelle chaque agent doit classer une observation inconnue à partir d'un sous-ensemble d'attributs. Contrairement à notre système, l'objectif premier de ClassAge est de donner la meilleure classification possible à partir de celles calculées par les différents agents. Tout comme notre système, il classe des objets évalués selon des attributs différents, mais ne cherche pas à comparer les classifications entre elles.

Peng met en œuvre la notion de rationalité limitée pour la classification de documents (Minsky, 1986). Ce type de classificateurs utilise classiquement un thésaurus représentatif des termes propres au domaine auquel les documents peuvent appartenir. La méthode de Peng repose sur la création d'un ensemble d'agents interconnectés (Peng *et al.*, 2001). Chaque agent procède à un classement de ses propres documents. Il envoie ses documents à d'autres agents si son classement est infructueux, par exemple si le thésaurus de l'agent ne concerne pas tous les mots de ses documents. Pour les communications, des listes d'accointances sont constituées dynamiquement et permettent au système de choisir l'agent le plus pertinent pour un service donné. Pour démontrer que le temps de calcul est moindre et que la classification est meilleure, il effectue une étude comparative entre une classification mono-agent et une classification multi-agent. Les tests montrent que le temps de classification multi-agent est meilleur dans le cas d'un très grand thésaurus.

De plus, cette approche offre plus de souplesse en facilitant l'introduction de nouveaux thésaurus. Enfin, elle prend en compte la tolérance aux pannes et la confidentialité des thésaurus (Mukhopadhyay *et al.*, 2003). Toutefois, comme dans ClassAge, le classement final ne cherche pas à faire apparaître les différences entre les classifications de chaque agent ; chaque agent propose sa propre classification complète et le meilleur est choisi.

L'équipe de Quteishat a développé un système multi-agent de classification basé sur le modèle confiance-négociation-communication (Quteishat *et al.*, 2010). Ce système TNC-based MAC est à base d'un réseau de neurones artificiels. Les agents sont organisés hiérarchiquement : on trouve l'agent parent, les chefs d'équipe et les membres de l'équipe. Les agents utilisent une méthode de négociation pour affecter une classe à un échantillon d'entrée. Chaque agent au sein de l'équipe donne une prévision de la classe d'affectation avec une valeur de confiance. Ensuite, le chef d'équipe sélectionne la prévision ayant la plus grande valeur et donne sa prédiction à l'agent père. Une prédiction a une valeur de confiance, une valeur de réputation et un facteur de confiance. L'agent père rend une décision définitive et attribue une

classe à l'échantillon d'entrée. L'expérimentation est composée de deux équipes d'agents : la première est l'équipe d'agents Fuzzy Min-Max et la deuxième celle des agents de type Fuzzy ARTMAP (Quteishat *et al.*, 2009).

Tableau 12. Analyse comparative des systèmes étudiés

	Samarah	NeurAge	Système de Mukhopadhyay	Système TNC-based MAC	MASKEP
Type de distribution	Pas de distribution des données	Distribution verticale	Distribution du thesaurus	Pas de distribution des données	Distribution verticale (étape 1) et horizontale (étape 2)
Méthode de classification	Méthode de classification non supervisée	Perceptron multi-couche et fonctions « radial basis »	Algorithme de classification non supervisée avec une étape d'apprentissage	Classification supervisée en réseau	Non supervisées avec possibilité de classer des données hétérogènes
Compétences des agents	Algorithme du K-mean	Distribution des méthodes : chaque agent à une méthode différente	Agents ayant les mêmes compétences à l'exception du thesaurus	Méthode d'apprentissage incrémental (comme FMM)	Chaque agent dispose d'une méthode de classification locale
Type de coopération	Réponse collective	Négociation entre agents	Sélection de la réponse du plus compétent	Négociation avec méthode d'offres	Classification des agents entre eux
Type d'application	Interprétation d'images	Système générique	Classification de textes	Applications industrielles (comme la génération d'énergie)	Maintien à domicile

Comme dans notre système, les agents forment des groupes et utilisent une coopération multi-étape à l'intérieur d'un groupe et entre groupes. Mais là encore, aucune information n'est obtenue en comparant les différences de choix des agents, seule la réponse du meilleur agent est considérée.

Le tableau 12 donne une vision synthétique de ces systèmes en fonction des six critères retenus. La dernière colonne permet de présenter les caractéristiques de notre système MASKEP selon ces mêmes critères.

7. Conclusion et perspectives

Dans cet article nous avons présenté MASKEP, un nouveau système multi-agent pour la classification de données hétérogènes et distribuées. Ce système est appliqué dans le cadre du maintien à domicile de personnes âgées et/ou dépendantes. En s'appuyant sur des systèmes de télé-vigilance qui font remonter des indicateurs médicaux ou comportementaux propres à chaque personne, il utilise une nouvelle

méthode de classification pour évaluer l'évolution de l'état de santé de la personne suivie, à travers l'obtention de classes. Cette méthode originale permet d'une part, de suivre l'évolution d'une personne au cours du temps, et d'autre part, de comparer entre elles des personnes qui ne sont pas forcément évaluées suivant les mêmes indicateurs, en exploitant ceux qui sont communs.

Corrélées via une expertise médicale à des pathologies et des états de forme, les classes obtenues et leurs évolutions sont caractéristiques d'une évolution de l'état de santé de la personne suivie, ce qui permet au corps médical d'anticiper une réponse adéquate, ou aux organisations d'aide à la personne de renforcer ou d'alléger le dispositif au domicile.

Au fil du temps, le nombre de personnes et de données les concernant va augmenter. Cette augmentation du nombre de données disponibles sur différentes échelles de temps, nous invite à nous pencher de manière plus approfondie sur la constitution, le filtrage et la sélection des groupes d'agents. En effet, de par le mécanisme actuel de constitution des groupes, leur nombre a tendance à se multiplier dès lors que les personnes sont surveillées pendant des durées différentes. Des méthodes sont donc à développer pour éliminer les groupes porteurs de classifications non pertinentes.

Afin de faciliter les étapes d'intégration du système, en réduisant les coûts et le temps nécessaire, des travaux sont également en cours pour mettre au point des simulateurs concernant le maintien à domicile. Ils doivent permettre d'alimenter notre système en données caractérisant des personnes virtuelles dont l'état de santé et le comportement évoluent en fonction du temps. L'intérêt par rapport à un système réel est la quantité de données immédiatement disponible pour tester les capacités de traitement de masse de notre système, et également la possibilité de rejouer des situations réelles déjà médicalement analysées, en changeant la nature des indicateurs disponibles.

Bibliographie

- Boudy J., Baldinger J.L., Delavault F., Belfeki I., Dorizzi B., Farin I. (2007). Télévigilance médicale au service du patient à domicile. *1^{re} conférence ASSISTH07*, vol. 1, p. 397-40.
- Canuto A., Santana L.E., Abreua M., Xavier Jr J. (2008). An analysis of data distribution in the ClassAge system: An agent-based system for classification tasks. *Neurocomputing*, vol. 71, Issues 16-18, p. 3319-3325.
- Carré M. et al., (2009). *Domicile, autonomie et technologies*, Ed. DOC'éditions.
- Chan M., Campo E., Estève D. (2003). PROSAFE, a multisensory remote monitoring system for the elderly or the handicapped in Independent Living for Persons with Disabilities and Elderly People. *Proceedings of the 1st International Conference on Smart Homes and health Telematics (ICOST'2003)*, September, p. 89-95.

- Chan M., Estève D., Escriba C., Campo E. (2008). A review of smart homes-Present state and future challenges. *Computer Methods and Programs in Biomedicine*, vol. 91, n° 1, p. 55-81, July.
- Ennuyer B. (2006, septembre). *Repenser le maintien à domicile : enjeux, acteurs, organisation*, Ed. Dunod.
- Forestier G., Wemmert C. and Gançarski P. (2008). Multi-source Images Analysis Using Collaborative Clustering. *EURASIP Journal on Advances in Signal Processing, Special issue on Machine Learning in Image Processing*, p. 11, Hindawi.
- Gançarski P., Wemmert C. (2007). Collaborative Multi-step Mono-level Multi-strategy Classification. *Journal on Multimedia Tools and Applications*, vol. 35, n° 1, p. 1-27, Springer Ed., ISSN: 1380-7501.
- Gimbert V., Godot C. (2010, juillet). *Vivre ensemble plus longtemps : enjeux et opportunités pour l'action publique du vieillissement de la population française*, Département Questions sociales du Centre d'analyse stratégique, <http://www.strategie.gouv.fr/>.
- Hartigan J.A. (2002). *Clustering Algorithms*. Wiley. MR0405726.
- Helal S., Mann W., El-Zabadani H., King J., Kaddoura Y., Jansen E. (2005). The Gator Tech Smart House: A Programmable Pervasive Space. *Computer*, vol. 38, p. 50-60.
- Hoey J., Poupart P., von Bertoldi A., Craig T., Boutilier C., Mihailidis A. (2010). Automated handwashing assistance for persons with dementia using video and a partially observable Markov decision process. *Journal of Computer Vision and Image Understanding*, vol. 114, p. 503-519.
- Kanungo T., Mount D.M., Netanyahu N.S., Piatko C.D., Silverman R., Wu A.Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Analysis and Machine Intelligence*.
- Larab A., Rigaud B., Pecatte J.M. (2008). Detection of the space confusion problem at the Alzheimer patient. *International Symposium on Bio- and Medical Informatics and Cybernetic*, BMIC-2008, Florida (USA).
- MacQueen J.B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, University of California Press, vol. 1, p. 281-297.
- Memarsadeghi N., Mount D.M., Netanyahu N.S., Le Moigne J. (2007). A Fast Implementation of the ISODATA Clustering Algorithm. *International Journal of Computational Geometry and Applications*, vol. 17, p. 71-103.
- Mihailidis A., Boger J., Czarnuch S., Jiancaro T., Hoey J. (2012). Ambient Assisted Living Technology to Support Older Adults with Dementia with Activities of Daily Living: Key Concepts and the State of the Art. *Handbook of Ambient Assisted Living*.
- Minsky M. (1986). *The Society of mind*, Simon & Schuster.
- Mukhopadhyay S., Peng S., Raje R., Palakal M., Mostafa J. (2003). Multi-agent information classification using dynamic acquaintance lists. *Journal of the American Society for Information Science and Technology*, vol. 54, p. 966-975. doi: 10.1002/asi.10292.
- Noury N. (2005 June). AILISA : Plateformes d'évaluations pour des technologies de télésurveillance médicale et d'assistance en gérontologie. *Gérontologie et Société*, vol. 113, p. 97-119.

- Ogawa R., Togawa T. (2000). Attempts at monitoring health status in the home. *1st Annual International Conference on Microtechnologies in Medicine and Biology*, vol. 1, p. 552-556.
- Peng S., Mukhopadhyay S., Raje R., Palakal M., Mostafa J. (2001). A comparison between single-agent and multi-agent classification of documents. *Proceedings of 15th International Parallel and Distributed Processing Symposium*, p. 935-944, San Francisco.
- Philipose M., Fishkin K., Perkowitz M., Patterson D., Fox D., Kautz H., Hahnel D. (2004). Inferring Activities from Interactions with Objects. *IEEE Pervasive Computing*, vol. 3, p. 50-57.
- Quteishat A., Peng Lim C., Tweedale J., Jain L.C. (2009). A Multi-Agent Classifier System Based on the Trust-Negotiation-Communication Model. *Advances in Soft Computing*, vol. 52, p. 97-106, Applications of Soft Computing.
- Quteishat A., Peng Lim C., Saleh J.M., Tweedale J., Jain L.C. (2010). A neural network-based multi-agent classifier system with a Bayesian formalism for trust measurement. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*.
- Rialle V., Lamy J.-B., Noury N., Bajolle L. (2003). Telemonitoring of patients at home : a software agent approach. *Computer Methods and Programs in Biomedicine*, vol. 72, n° 3, p. 257-268.
- Rivière A.C., Brugière A. (2010, juin). *Bien vieillir grâce au numérique. Autonomie, qualité de vie, lien social*. FYP éditions.
- Santana L., Canuto A., Abreu M. (2006). Analyzing the Performance of an Agent-based Neural System for Classification Tasks Using Data Distribution among the Agents. *Proceedings of International Joint Conference on Neural Networks*, p. 2951-2958, Vancouver.
- El Fallah Seghrouchni A., Briot J.-P. (2009). Technologies des systèmes multi-agents et applications industrielles. *IC2 Series*, Hermès Lavoisier.
- Tamura T., Togawa T., Ogawa M., Yoda M. (1998). Fully automated health monitoring system in the home. *Medical Engineering & Physics*, vol. 20, n° 8, p. 573-579.
- Wemmert C., Gancarski P. (2002, september). A multi-view voting method to combine unsupervised classifications. *Artificial Intelligence and Applications*, p. 447-452, Malaga, Spain.
- Williams G., Doughty K., Bradley D.A. (1998). A systems approach to achieving CarerNet - an integrated and intelligent telecare system. *IEEE Transactions on Information Technology in Biomedicine*, vol. 2, n° 1, p. 1-9.