

BULLETIN

DE

L'AFIA



DÉCEMBRE 2008

N° 67

Association Française pour l'Intelligence Artificielle

Présentation du bulletin

Le **Bulletin de l'Association Française pour l'Intelligence Artificielle** vise à fournir un cadre de discussion et d'échanges au sein de la communauté universitaire et industrielle. Ainsi, toutes les contributions, pour peu qu'elles aient un intérêt général pour l'ensemble des lecteurs, sont les bienvenues. En particulier, les annonces, les comptes rendus de conférences, les notes de lecture et les articles de débat sont très recherchés. Le Bulletin de l'AFIA publie également des dossiers plus substantiels sur différents thèmes liés à l'IA. Le comité de rédaction se réserve le droit de ne pas publier des contributions qu'il jugerait contraire à l'esprit du bulletin ou à sa politique éditoriale. En outre, les articles signés de même que les contributions aux débats, reflètent le point de vue de leurs auteurs et n'engagent qu'eux-mêmes.

Pour contacter l'AFIA

Président

Michèle SEBAG

L.R.I., Bât. 490, Université Paris-Sud,
91405 ORSAY

Tel : +33 (0)1 69 15 76 02

Fax : +33 (0)1 69 15 65 86

Michele.Sebag@lri.fr

<http://www.lri.fr/~sebag>

Contributions au bulletin

Philippe MORIGNOT

voir ci-contre

Serveur WEB

<http://www.afia-france.org>

Adhésions, liens avec les adhérents

Marie-Carol LOPES

LRI, Bâtiment 490

Université Paris-Sud

91405 ORSAY

Mél. : Marie-Carol.Lopes@lri.fr

Relation avec les entreprises

Gérald PETITJEAN

EURODECISION

9a, rue de la porte de Buc

78000 VERSAILLES

Mél. : Gerald.Petitjean@eurodecision.com

Membres d'honneur

Marie-Odile Cordier (1999), Jean-Paul Haton (1999), Jacques Pitrat (1999), Jean-Marc David (2000), Daniel Kayser (2000), Claude Vogel (2000), Henri Farreny (2001), Alain Colmerauer (2002), Jean-Louis Laurière (2002), Gérard Sabah (2003), Jean-Claude Latombe (2004), Yves Kodratoff (2004), Malik Ghallab (2005).

Personnes morales adhérentes à l'AFIA

ADIT, AI*IA, CEA, CNET PARIS-A, ENS Mines, ENST-Paris, ESIEA, IGN, ILOG, INRETS, INRIA, Institut Français du Pétrole, Peugeot S.A., Université de Savoie, Université Paris 9 Dauphine.

Bureau de l'AFIA

Michèle Sebag, président

Olivier Boissier, secrétaire

Marie-Carol Lopes, trésorier

Comité :

Gilles BISSON, Olivier BOISSIER, Serge DUPUY, Amal EL FALLAH, Emmanuel GUERE, Marc-Philippe HUGET, Nicolas LACHICHE, Vincent LEMAIRE, Jérémie MARY, Eunika MERCIER-LAURENT, Philippe MORIGNOT, Jean-Denis MULLER, Gérald PETITJEAN, Jean-François PERROT, Marc SCHOENAUER, Michèle SEBAG, Patrick SAINT-JEAN, Fabien TORRE, Samuel WIECZOREK.

Comité de Rédaction

Philippe Morignot

Rédacteur en chef

AXLOG Ingénierie

19-21, rue du 8 mai 1945,
94110 ARCUEIL

philippe.morignot@axlog.fr

Brigitte GRAU

Rubrique « Sommaire des revues »

LIMSI – CNRS

B.P. 133, 91403 ORSAY Cedex

grau@limsi.fr

Marc-Philippe HUGET

Rubrique « Petites annonces »

Polytech'Savoie-LISTIC

Université de Savoie, B.P.
80439, 74944 Annecy-le-vieux
CEDEX

Marc-Philippe.Huget@univ-savoie.fr

Nicole Tourigny

Rubrique « I.A. au Québec »

LSI-ERICA

Université Laval, Québec, Canada

Nicole.Tourigny@ift.laval.ca

Récapitulatif des dossiers parus

L'Assemblée Générale de l'AFIA a décidé début 2008 qu'il n'y aurait désormais plus de Bulletin papier (tout est au format électronique), sauf un récapitulatif annuel des dossiers parus, qui lui est au format papier (en plus d'être disponible au format électronique, comme depuis des années).

Ce numéro du Bulletin de l'AFIA est donc le premier numéro consacré aux dossiers parus. Nous re-publions deux dossiers : celui coordonné par Flavien Balbo (LAMSADE) et Régine Seidowsky (INRETS) qui avait pour thème "SMA et Transport", et celui du printemps 2007, coordonnée par François Coste (IRISA) et Christine Froidevaux (LRI), et qui avait pour thème "I.A. et Bioinformatique".

A lire et à relire !

Philippe Morignot
Rédacteur en chef du Bulletin de l'AFIA

Dossier : Système multi-agent et Transport

Coordonnateurs :

Flavien Balbo (balbo@lamsade.fr), Régine Seidowsky (regine.seidowsky@inrets.fr)

Le développement des Systèmes Intelligents de Transport illustre l'évolution et plus généralement le rôle grandissant de l'informatique dans les systèmes de transport. L'objectif de ces systèmes est de rendre les réseaux plus sûrs, plus efficaces, plus fiables et plus écologiques sans avoir nécessairement à modifier matériellement l'infrastructure existante.

Les domaines du transport et des Systèmes Multi-Agents entretiennent depuis très longtemps des liens privilégiés. Sur le plan international, la pérennité des ateliers dédiés au transport dans les conférences multi-agents (AAMAS) ou dédiés aux SMA dans les conférences transports (IEEE ITSC) illustrent la vivacité de ces relations. Sur le plan national, le succès du forum *Systèmes & Logiciels pour les NTIC dans le transport*¹ organisé par l'Inrets et dédié à l'utilisation des systèmes multi-agents montre l'intérêt de notre communauté pour ce domaine.

Le domaine du transport favorise l'utilisation d'un paradigme simplifiant la gestion de systèmes physiquement et fonctionnellement distribués. De plus, de très nombreux systèmes sont soumis à des contraintes de gestion en temps réel, et/ou un environnement dynamique. Ces caractéristiques supposent des capacités de pro-activité et d'autonomie qui sont au cœur de la conception de systèmes multi-agents.

Plus d'une dizaine de laboratoires ont participé à ce dossier. Un premier constat concerne la richesse des domaines d'applications. Ainsi pratiquement tous les modes de transport sont pris en compte, et la communauté multi-agent française propose des solutions à de très nombreuses problématiques transports : information voyageur, gestion de trafic, surveillance de réseau, véhicules intelligents, ...

Cette diversité en terme d'application se double de la mise en œuvre d'aussi nombreuses compétences. Par exemple, de nombreux travaux exploitent l'approche bottom-up supportée par le paradigme multi-agent pour la conception de simulations rendant compte de l'aggrégation des comportements des acteurs d'un réseau de transports : simulation de piétons, simulation de réseaux de transport en commun, simulation de trafic, simulation de

véhicules, simulation de crises. Autre exemple, la personnalisation permise par le paradigme agent est mise à profit pour la gestion des informations voyageurs, la différenciation comportementale des véhicules dans les générateurs de trafic ou la modélisation du comportement de navire. Ou encore les travaux sur la négociation sont utilisés pour gérer les comportements aux carrefours ou la simulation du comportement de piétons.

Dans ce dossier, vous trouverez pour chaque laboratoire impliqué les coordonnées de la personne à contacter, la problématique de recherche ainsi que les projets en cours ou réalisés ainsi qu'une bibliographie ciblée.

CITERES

Cités, TERritoires, Environnement, Sociétés

- CONTACT : PH. MATHIS Professeur Émérite, mathis@univ-tours.fr
- ADRESSE : UMR CITERES 6173
- TEL : 02 47 52 86 50

Membres de l'équipe

Ce travail est un contrat de Recherche Predit financé par l'ADEME. L'équipe était composée d'Hélène Bailleul, Benoit Feildel, Jean Baptiste Buguellou, doctorants en Aménagement et de Cédric Moquillon doctorant en Informatique, Anthony TOUET et Kim Hong TAN Stagiaires D.I.4 Ecole Polytechnique de l'Université de Tours, sous la direction de Philippe Mathis Professeur Émérite, pour la simulation SMA des déplacements à partir des agendas des ménages reconstitués. La seconde phase sur les relocalisations et les effets de décisions urbanistiques à partir des catégories socioprofessionnelles de chaque zone est réalisée par Philippe Mathis.

1. <http://www.inrets.fr/services/manif/ForumNTIC>

Le Modèle SIMULH Cohérence entre PLU et SCoT par modélisation SMA multi échelles spatiales et temporelles

L'agglomération de Mulhouse est dans une situation paradoxale développant à la fois un réseau de tramway et un tram-train et voulant simultanément reconquérir son centre ville déstructuré par la disparition des industries.

Objectif Immédiat

L'objectif était donc d'étudier la possible cohérence entre le P.D.U. et le SCoT de Mulhouse. La réponse proposait une démarche prospective basée sur une modélisation multi échelles temporelles et spatiales et sur une prise en compte adoptant une démarche bottom up, SMA. Le modèle est principalement développé à partir des travaux menés avec mes doctorants depuis le début des années 2000 et sur nos travaux sur la cellularisation des graphes. Le modèle a été conçu pour pouvoir s'appliquer à un grand nombre d'agglomérations urbaines.

Originalité de la démarche

- La finesse de la définition spatiale : un graphe de 12 000 sommets, 35 000 arcs.
- La structure SMA : 120 000 ménages sont mobiles, reconstitués avec leurs agendas, d'où des cycles de déplacements quotidiens, pouvant être à l'origine de transferts modaux et/ou de relocalisations...
- L'influence de la gouvernance locale qui peut modifier les règles d'urbanisme SCoT et de circulation PDU et le problème de la cohérence de ces deux aspects.
- La prise en compte d'insatisfactions de localisation, pas seulement économiques (foncier et transport) mais aussi environnementales (naturelles, sociales et urbaines) et de la promotion immobilière sont en cours.

Il est multi échelles spatiales, de la cellule de 1 mètre du graphe au polygone de Voronoï, de l'IRIS à l'Agglomération au sens large, et multi échelles temporelles de la seconde pour la simulation individuelle des déplacements, à l'année pour les relocalisations de logement, et à la mandature locale et au double même pour la production d'infrastructures. Le modèle est prospectif et chaque niveau rétroagit sur les autres, par exemple les délocalisations entraînent une modification des CSP qui elles mêmes etc..

La méthode

Les ménages ont été reconstitués à partir des statistiques de population. Chaque ménage dont la structure, la

localisation, le lieu de travail sont fixés, a défini un programme d'activité (travail, courses, conduite des enfants à l'école...) réparti entre le ou les parent(s). Les activités étant a priori dans des lieux différents les agents doivent donc se déplacer en prenant en compte le temps de déplacement prévu d'où la création d'agendas. Le chemin le plus court (algorithme de Djiskra) n'est valable que s'il n'y a pas de congestion ralentissant la circulation. Un mécanisme d'apprentissage par expérience a été programmé pour qu'ils n'utilisent pas tous en même temps les mêmes trajets bloquant ainsi le système et surtout pour qu'ils puissent réaliser leurs agendas soit en changeant de trajets, d'horaires si cela est possible ou de localisations. La technique des graphes cellulaires mise au point à Tours permet une grande précision : cellules d'1 mètre et une itération par seconde pour simuler le comportement des 120 000 véhicules (accélération, freinage, croisement etc.) respect du code de la route Cette phase est fonctionnelle pour tester des améliorations du système de transport et notamment des transferts de modes vers les transports en commun et sites propres tramway et tram-train. Cette première partie a été réalisée en C++ à partir des travaux de l'équipe du DA (CESA) ce qui représente la contribution partielle de trois thèses.

Les ménages insatisfaits de leurs logements ou de sa localisation - trop petit, trop cher, environnement social peu apprécié, trop éloigné de la ville, des " bons " collèges et lycées, du travail etc. - tentent chaque année de se relocaliser - Chaque ménage insatisfait appartient à une catégorie socioprofessionnelle. Chaque CSP a de la ville une vision particulière. Ses choix sont différents et l'ensemble définit la structure sociale de la ville et donc le marché foncier qui évolue lentement avec la ségrégation et les spécialisations partielles des quartiers. Toutes les demandes de relocalisation ne peuvent pas être satisfaites dans les quartiers choisis ce qui définit une demande de logement par catégorie. L'objet de la promotion immobilière troisième niveau, est précisément de répondre à cette demande tout en faisant un bénéfice (privé) ou en respectant des budgets (financements publics). Les promoteurs sont les agents et l'espace est constitué des zones constructibles et des rénovations urbaines Cette partie du modèle est en cours, les itérations sont annuelles.

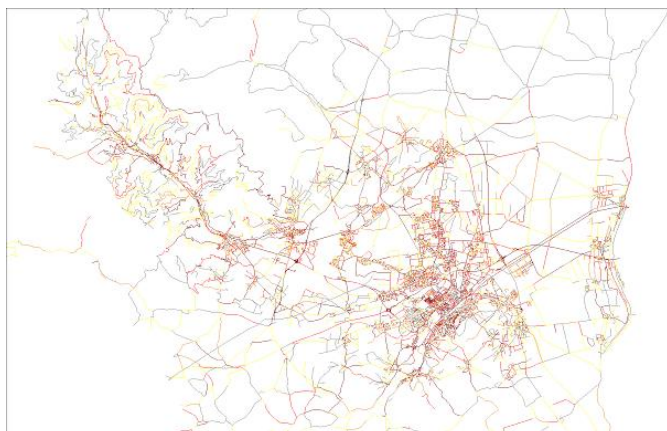
Dans le quatrième niveau, les collectivités locales (gouvernance : communes, regroupements de communes, agence d'urbanisme, SIVOM...) interviennent ou non pour définir ou modifier les règles d'urbanisme, ce qui modifie les capacités de réponses des promoteurs et les types de logements réalisés, d'où certaines évolutions du bâti et corrélativement des CSP sont favorisées ou au contraire freinées suivant les lieux. Les collectivités locales inter-

viennent aussi par les plans de circulation, les règlements (vitesse, parking, tarifs...) et par la création d'infrastructures routes etc. mais aussi tramway, tram-train et leurs caractéristiques : confort, fréquence, coût. Ce niveau dont les itérations sont pluri annuelles (de 1 an - modifications de règlements de POS et de circulation- à 15 ans pour les infrastructures et les plans d'urbanisme) modifie naturellement les données du premier et du second niveau d'où des bouclages dynamiques multiples.

Compte tenu de la lourdeur des calculs nécessaires, les deuxièmes et troisièmes niveaux sont réalisés, dans un premier temps, sur un espace simplifié de 1600 avant d'être transférés sur l'espace développé. La convergence de l'adaptation des conducteurs à l'intensité du trafic, entre trente et cinquante itérations, demande deux jours de calcul sur une station travail HP 9300 en C++. Certains algorithmes sont de types $O(n^3)$ et le remplissage d'une matrice de 120 000 par 120 000 demande beaucoup de temps même si pour accélérer ces calculs au niveau de l'individu sur un graphe plus limité par sa pratique, on utilise l'algorithme de DJISKRA. Ceci permet de ne pas modifier la structure des données qui servent aux différents niveaux. Le second niveau utilise FORTRAN95 et 2003. Ces deux langages interopérables, C++ et Fortran ont été retenus car étant les plus rapides et parallélisables.

Les résultats actuels

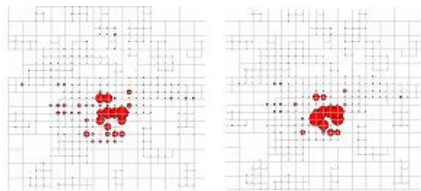
La simulation des déplacements a été mise au point sur l'agglomération de Tours que nous connaissons bien, ce qui a permis de vérifier la cohérence du modèle et son adéquation car le problème du test des modèles SMA basés sur des règles de comportement plus littéraires que statistiques est difficile. Nous avons choisi de vérifier les out put pour juger de l'adéquation du modèle.



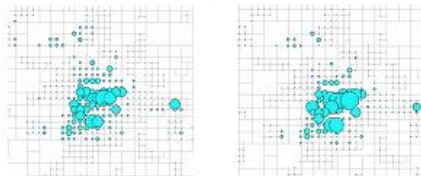
Diffusion des agents dans l'espace de simulation

Les différentes courbes classiques utilisées par les modèles traditionnels en input : débit- vitesse, densité- vitesse et débit densité sont retrouvées. Les flux correspondent aux données classiques sur les horaires de pointe. Les fronts de congestion sur les arcs ont été aussi recréés avec en plus la congestion des divers carrefours qui a été simulée avec les trajectoires et les densités de trafic possibles. Cela a permis à Jean Baptiste Buguellou de montrer l'inapplicabilité en zone urbaine du théorème de Ford-Fulkerson, la contrainte de capacité la plus forte n'étant pas sur les arcs mais sur les carrefours et dépendant entre autres des trajectoires des véhicules, du type de carrefour etc. Leurs capacités pouvant n'être que de 40 à 80% de celles des arcs entrant ! D'où la congestion que tout automobiliste expérimente. De même, l'utilisation individuelle et myope des chemins les plus rapides conduit à la congestion du réseau et seul un apprentissage d'autres parcours par les conducteurs permet à l'ensemble d'éviter la congestion et de réaliser les agendas. Les coûts de déplacements avec simulation des conséquences du coût du pétrole, mais aussi les émissions localisées des pollutions gazeuses ont pu être calculés pour chaque ménage à partir des temps de déplacement des distances des vitesses des types de voitures utilisées en fonction de sa CSP. La simulation d'une limitation générale des vitesses a été faite.

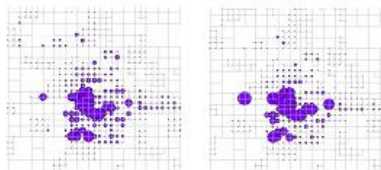
Des résultats sur l'analyse fine de la ségrégation urbaine par CSP et de ses dynamiques à partir du modèle simplifié et émergence d'une structure globale sont en cours d'exploitation : stratégie de concentration avec ou sans limite de densité, de surfaces différentielles utilisées par les CSP, d'offre ponctuelle de logements supplémentaires, existence d'aménités localisées ... stratégies de déconcentration - rurbanisation- stratégies mixtes simultanées et intégration des coûts de déplacement ainsi que l'analyse de l'amplification éventuelle de la ségrégation par la promotion immobilière ainsi que par le facteur amplifiant ou atténuant des politiques urbaines. Les résultats acquis montrent une structure de localisation évoluant sur 3 ou quatre mandatures locales mais de façon complexe avec des aspects contre intuitifs.



Cadres et professions intellectuelles supérieures, itération 12 et 18 dures de deux et trois mandats municipaux, attractions totales surfaces différentielles. L'accroissement des surfaces d'habitat de cette catégorie a pour effet d'accroître la surface occupée et donc le nombre de zones, sans pour autant que la population apparaisse comme ayant augmenté. La conséquence est naturellement une baisse de la densité dans ces zones. Cette CSP, à chaque itération, choisit ses localisations en seconde position.



Ouvriers, itération 12 et 18 dures de deux et trois mandats municipaux au minimum attractions totales surfaces différentielles.



Retraités, itération 12 et 18 dures de deux et trois mandats municipaux, attractions totales surfaces différentielles.

Simulation d'une tendance à la concentration non contrebalancée pour l'étude de la ségrégation spatiale : les différences de localisations apparaissent clairement ; l'analyse de l'amplification éventuelle de la ségrégation par la promotion immobilière ainsi que par le facteur amplifiant

La demande non satisfaite est évaluée mais actuellement le sous programme promotion immobilière n'est pas encore développé sauf de façon très grossière et insuffisante par création exogène de logements par zone, avec retard.

Trois difficultés principales ont été rencontrées : informatique, lourdeur et temps de calcul du modèle développé de simulation de trafic. Le modèle simplifié au niveau spatial reste complexe du fait du nombre de variables et a saturé les deux compilateurs utilisés d'où une restructuration complète en une hiérarchie de sous programmes. Le second problème classique dans ce domaine et dans une approche SMA est celui de l'interdépendance des variables et du manque de données instrumentalisables au niveau des règles de décision des agents et des groupes et enfin en troisième lieu celui de précisions des calculs sur des variables entières (affectation des restes, transfert au niveau spatial des erreurs de calcul, utilisation des probabilités sur de petits nombres répartis dans de nombreuses zones. ...)

Bibliographie sommaire

Rapport "Cohérence entre les politiques de transport et les politiques d'aménagement - le cas de Mulhouse", rapport de recherche Intermédiaire, 2008 PREDIT-ADEME.

Enquête Travail et Emploi du temps de l'INSEE Hall et Al (1992)

(Highway capacity manual), Committee on Highway Capacity and Quality of Service, of the U.S. Transportation Research Board.

Sven Maerivoet - Modelling traffic on motorways : state-of-the-art, numerical data analysis and dynamic traffic assignment. Katholieke Universiteit Leuven Juni 2006

CHOPARD (B.), DUPUIS (A.) - " Cellular Automata Simulation of Traffic : A Model of the City of Geneva ".- Networks and spatial economics, n°3, 2003, pp.9-21. Ceux-ci cependant créent directement le cellulaire sans le déduire d'un graphe et consécutivement ne peuvent pas revenir à un graphe si nécessaire dans un changement d'échelle ou de perspective comme par exemple la prise en compte de surfaces duales.

Buguellou Jean-Baptiste " Analyse multi-niveaux des flux de transport : modélisation multi-agents de l'interaction entre l'offre et la demande de transport " sous la direction de Philippe Mathis, Doctorat en aménagement, Ecole Polytechnique de l'Université de Tours - Département Aménagement Tours, soutenance prévue Février 2009.

Graphes et Réseaux modélisation multi niveaux s/dir Ph. Mathis, Hermès Lavoisier, 2003 Paris 360P.

Graphs and networks Multilevel Modeling edited by Philippe Mathis ISTE Ltd, London, New Beach 2007, traduction anglaise et complements.

Graphs and networks Multilevel Modeling edited by Philippe Mathis ISTE Ltd, and John Wiley & Sons Inc. Second completed edition 2008.

Mihai Tivadar Structures sociales urbaines et ville écologique, Thèse Lille 2006.

INRETS - GRECIA

Génie des REseaux de Transport et Informatique Avancée

- CONTACT : Mr SCÉMAMA (Dr) - scemama@inrets.fr,
- ADRESSE : Inrets - Arcueil 2, Avenue du Général-Malleret Joinville F-94114 Arcueil CEDEX,
- WEB : www.inrets.fr/ur/gretia/,
- TEL : 01 47 40 70 00.

Membres de l'équipe

Fabien Badeig, Flavien Balbo, Gérard Scémama, Mahdi Zargayouna.

Thème général de l'équipe

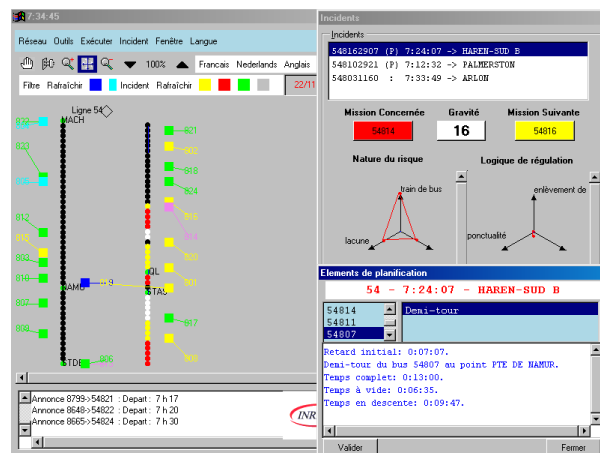
Cette recherche vise à harmoniser des systèmes de transport et concerne le développement d'applications STI multimodales innovantes portant sur la gestion globale des réseaux, la gestion des pôles d'échange, l'information des voyageurs et le transport à la demande. Cette thématique repose sur le développement des Nouvelles Technologies de l'Information et de la Communication pour le transport. Les technologies utilisées concernent les méthodes "Orientées Objet" pour la formalisation de modèles conceptuels de composants et d'outils génériques utiles pour la représentation des réseaux multimodaux. Ces modèles serviront de base à la réalisation d'outils et d'applications de transport. Les technologies de l'Internet (langage Java, XML, services Web) les architectures, les agents intelligents, les interfaces homme-machine sont également étudiées dans ce cadre. Cette recherche s'appuie d'une part sur des expérimentations sur sites en vraie grandeur (Toulouse, Bruxelles), et d'autre part sur un équipement laboratoire permettant l'acquisition de données multi-sites/multi-sources et la simulation pour les recherches sur l'Intermodalité (plateforme SITI)

Description des travaux

Au GRECIA, le paradigme multi-agent a été utilisé selon une approche unifiée [Bad08] développé en collaboration avec le Lamsade (section) et reposant sur le principe de coordination appelé PbC (Property-based Coordination) [Zar06]. Utilisant l'observabilité de l'environnement multi-agent, le principe de la coordination fondée sur les propriétés est le suivant : *Concevoir les composants du système multi-agent comme des composants observables symboliques et gérer leur traitement à des fins de coordination.* Ce principe général a été appliquée pour des applications très différentes décrites ci-dessous.

SATIR (Système Automatique de Traitement des Incidents en Réseau)

Le système d'aide à la décision SATIR [Bal05, Bal07] propose une réponse adaptée (fig.) aux difficultés de gestion en temps réel de réseaux urbains de surface.



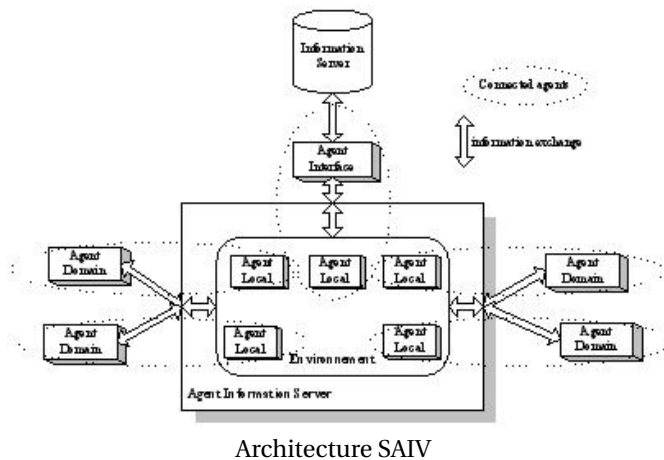
SATIR : Interface utilisateur

Le système multi-agent proposé permet d'assurer une gestion dynamique des horaires ainsi que la gestion des incohérences des données issues des capteurs. Lorsqu'une situation à problème a été détectée (en l'occurrence le retard important d'un véhicule) le système multi-agent se réorganise afin d'assurer un diagnostic puis d'apporter une solution. Pour les phases de diagnostic et de planification, une modélisation formelle et son intégration au sein du système multi-agent de gestion de réseau ont été proposées. Le modèle de diagnostic repose sur la prise en compte explicite des perturbations afin d'identifier les informations nécessaires à leur compréhension. La modélisation doit rendre compte de la dynamique du processus (effet train de bus, création de lacune) afin d'assurer son suivi suivant trois axes : (1) le temps (l'importance du retard évolue); (2) l'espace (les véhicules se déplacent sur le réseau), (3) la forme (la nature du problème se modifie dans le temps). D'un point de vue multi-agent, la représentation d'une perturbation résulte du regroupement dynamique des agents concernés par le problème au sein d'un environnement dédié. Le modèle de planification consiste en une recomposition de l'offre de transport afin de répondre à une modification de la structure de la demande.

SAIV (Serveur Agent d'Information Voyageur)

Le SAIV [Zar05] fait le lien entre les sources d'information et les utilisateurs d'un réseau multimodal donné. L'objectif est de proposer des informations pertinentes selon les intérêts de l'utilisateur et selon sa position dans son trajet. Reposant sur le principe de la PbC, le modèle interactionnel proposé par EASI (section), permet de concevoir un serveur d'information paramétré par ses utilisateurs. Si les systèmes d'information sont adaptés à la satisfaction de besoin ponctuel (modèle requête/réponse), la gestion du suivi d'information suppose la mise en œuvre de traitements supplé-

mentaires. Ces traitements seront d'autant plus complexes que les sources d'information sont distribuées et que la gestion du suivi suppose une comparaison des informations disponibles. Le modèle EASI permet une mise en commun des informations disponibles et mise à jour par les serveurs.



Nous avons identifié trois catégories d'agents. La première concerne les agents *Interface*, son rôle est de faire le lien entre un système d'information existant et le serveur d'information. Cet agent devra collecter l'information nécessaire à l'application et/ou répondre aux demandes issues du serveur. Les informations envoyées par cet agent seront déposées dans l'environnement. La seconde catégorie concerne les agents *Domaine* qui sont spécifiques à l'application. Ces agents peuvent être récepteurs et/ou émetteurs d'informations. Contrairement aux agents *Interface* dont le rôle est d'intégrer au sein de l'application des sources de données existantes, ces agents ont été conçus afin d'exploiter le serveur d'information. A ces deux catégories d'agent qui sont physiquement distants s'ajoute la catégorie agent *Local*. Cet agent est le représentant d'un agent *Interface* ou *Domaine* au sein de l'environnement. Son rôle est de gérer les filtres pour l'agent distant (*Interface* ou *Domaine*) auquel il est associé. L'architecture du serveur d'information doit permettre de distribuer l'information entre la structure émettrices / réceptrices distantes (Agents *Interface* et *Domaine*) et la structure organisatrice (Agents *Local*).

STAD (Système de Transport à la demande)

Un STAD [Zar08] est un système conçu pour répondre à des clients qui désirent être transportés d'un point du réseau à un autre. Les clients contraignent leurs demandes par une fenêtre de temps pour l'arrivée et une autre pour le départ. Les critères d'évaluation de l'efficacité d'un STAD sont, d'abord, la minimisation du nombre de véhicules uti-

lisés pour servir tous les utilisateurs et ensuite la minimisation de la distance totale parcourue par les véhicules. Un STAD est confronté à trois types de difficultés inhérentes aux applications transports :

- La gestion des connaissances : Les utilisateurs sont inconnus au démarrage du système. Dans notre système, l'utilisation de l'environnement permettra la mise en commun de toutes les informations dès qu'elles deviennent disponibles.
- La dimension spatio-temporelle du problème : La distribution spatio-temporelle des utilisateurs ainsi que la largeur moyenne de leur fenêtre temporelle conditionne le comportement des véhicules et leur coordination, influençant ainsi grandement la qualité des solutions fournies par le système. Dans notre application, l'environnement structure les composants du SMA temporellement et spatialement de telle manière que l'interaction entre les agents est dirigée selon leur perception de cette information.
- La dynamique de l'environnement réel : Un SMA modélisant un STAD est un SMA ouvert puisque les agents (clients et véhicules) rejoignent et quittent le système librement. Dans un tel environnement dynamique, la limitation des communications est très importante car la diffusion de toutes les informations serait très coûteuse.

Nous proposons un modèle de coordination (section) conforme à PbC, dans lequel la communication est découplée dans le temps et l'espace palliant le risque de perte d'information dans un environnement dynamique. Cela permet aux agents d'interagir sans qu'ils s'exécutent au même endroit ni au même instant.

Nous proposons également une nouvelle heuristique fondée sur les champs de perception des agents véhicules [Zar08], permettant une focalisation du processus de résolution, non plus sur l'optimisation incrémentale des critères définis pour le problème (nombre de véhicules utilisés et distance totale parcourue), mais d'optimiser les possibilités qu'aura un véhicule dans le futur de participer à l'insertion de clients dans sa tournée (matérialisées dans ce que nous appelons "champs de perception" d'un agent). Comme prévu, les expérimentations démontrent que cette approche permet de diminuer le nombre de véhicules mobilisés pour la desserte d'un ensemble de clients, par rapport aux travaux de la littérature.

SGC (Système de gestion de crise)

L'expertise concernant la gestion de crise est incomplète, car chaque situation de crise est complexe et unique. Une crise est un phénomène dynamique défini par une situation initiale dépendant du lieu et du moment, et par son

impact sur la population et l'infrastructure. Pour répondre aux situations de crises liées au trafic routier, des Plans de Gestion de Trafic (PGTs) sont développés et spécifient les mesures à prendre en cas de crise pour coordonner les différents services d'intervention (services d'urgence, les médias, les gestionnaires de trafic). Notre proposition [Bad08] est de concevoir un système d'aide à la décision pour l'évaluation et la validation de PGTs. Ce SAD s'appuie sur la conception d'un simulateur de gestion de crise prenant en compte les constats suivants :

- Système multi acteurs. Le partage des responsabilités plus ou moins précis par des acteurs de cultures différentes et la mise en commun des moyens couplés à des intérêts éventuellement divergents doit être pris en compte dans la simulation.
- Espace tridimensionnel d'évolution des risques. Par extension des travaux de traitement des incidents sur un réseau de transport, nous pouvons considérer que les caractéristiques de l'évolution d'une crise peuvent se mesurer suivant trois axes, le temps l'espace et la forme. Le simulateur doit donc pouvoir faire évoluer la crise simulée sur chacun de ces axes.
- Connaissance morcelée. Les connaissances nécessaires au traitement de la crise sont réparties dans le temps et l'espace. Alors qu'un système d'informations devrait envoyer la bonne information au bon acteur au bon moment, le simulateur devrait permettre de tester différents scénarios de gestion des connaissances, d'une organisation totalement distribuée à la simulation d'un poste de gestion de crise.

Afin de répondre aux différents points présentés ci-dessus, nous proposons de fonder notre simulateur sur le paradigme agent et plus particulièrement sur le modèle de simulation EASS (section) prenant en compte l'activation contextuel des agents par le simulateur.

Références

[Bad08] F. Badeig, F. Balbo, G. Scemama, M. Zargayouna, "Agent-Based Coordination Model for Designing Transportation Applications". In 11th International IEEE Conference on Intelligent Transportation Systems (ITSC'08), Pékin (Chine), octobre 2008. IEEE Computer Society Press.

[Bal07] F. Balbo, S. Pinson. A Transportation Decision Support System in Agent-Based Environment, *Journal of Intelligent Decision Technologies* 1(3), pages 97-115, 2007.

[Zar08] M. Zargayouna, F. Balbo, G. Scemama, "A multi-agent approach for the dynamic VRPTW". In International Workshop on Engineering Societies in the Agents World (ESAW 2008), Saint-Etienne (France), septembre 2008.

[Zar05] M. Zargayouna, F. Balbo, J. Saunier, "Agent Information Server : a middleware for traveler information". In O. Dikenelli, M.-P. Gleizes and A. Ricci (Eds) : International Workshop on Engineering Societies in the Agents World (ESAW 2005), Kusadasi, Turquie, pages 3-16, Octobre 2005. Volume 3963 of Lecture Notes in Artificial Intelligence. Springer Verlag.

[Zar06] M. Zargayouna, J. Saunier, F. Balbo, "Property Based Coordination". In J. Euzenat and J. Domingue (Eds) : International Conference on Artificial Intelligence : Methodology, Systems, Applications (AIMSA 2006), Varna, Bulgarie, pages 3-12, Septembre 2006. Volume 4183 of Lecture Notes in Artificial Intelligence. Springer-Verlag.

Imagin

Images et Interactivité Numérique

- CONTACT : Michel Augeraud
- ADRESSE : Equipe Imagin "Images et Interactivité Numérique"
Laboratoire L3I
Pôle Sciences & Technologies, Université de La Rochelle,
Avenue M. Crépeau,
17042 La Rochelle cedex 01
- WEB : l3i.univ-larochelle.fr
- TEL : 0546513973

Membres de l'équipe

- Michel Augeraud (michel.augeraud@univ-lr.fr)
- Marie-Hélène Verrons (marie-helene.verrons@univ-lr.fr)
- Sandra Lemp (sandra.lemp@univ-lr.fr)

Thème général de l'équipe

L'équipe Imagin s'intéresse à l'image et aux comportements avec contraintes temporelles ainsi qu'à l'analyse des comportements implicites. Elle vise à définir les modèles, architectures et outils permettant de mettre en œuvre un environnement paramétrable, personnalisable et reconfigurable pour la conception et l'exécution adaptative d'activités interactives.

introduction

Nous nous sommes intéressés plus particulièrement à la mise en place et à la modification d'environnements ur-

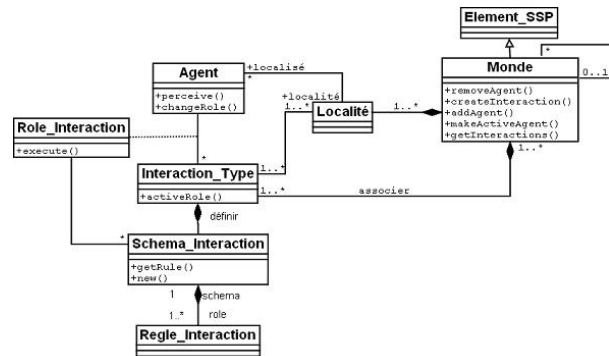
bains. Celui-ci nécessite, pour un décideur, des outils permettant de comprendre l'impact des décisions prises. L'objectif de notre projet est d'offrir des simulations permettant de visualiser l'impact des décisions sur la représentation d'un quartier. Pour cela, l'outil développé permet une représentation "individu centré" et une visualisation 3D. De plus, il offre la possibilité d'interagir en "jouant" avec le simulateur. Un utilisateur peut ainsi prendre la place d'un véhicule ou d'un piéton. L'immersion de l'utilisateur offre un retour sur des critères qui sont difficilement évaluable car difficiles à caractériser. L'approche naturelle utilisée pour modéliser puis simuler un système complexe est de développer un modèle, souvent appelé modèle conceptuel. Ce modèle utilise un formalisme et des concepts liés au domaine qui permettent au modélisateur de "comprendre" son fonctionnement du système. Ce modèle doit ensuite être traduit dans un formalisme qui rendra possible l'exécution du modèle obtenu et donc l'évaluation du "système". Les informations nécessaires à la simulation de systèmes complexes ne sont pas directement liées aux informations qui seront visualisées au moyen de l'environnement virtuel de visualisation. Nous nous plaçons dans le cadre d'une utilisation par un décideur.

Architecture du système

Pour le projet en lui-même, nous nous appuyons sur deux composants, à savoir une plateforme de visualisation 3D et une plateforme SMA permettant de représenter, selon un modèle "individu centré", la logique du système. Pour ce dernier aspect, nous utilisons la plateforme multi-agents Jade afin de représenter par des agents, les différentes entités pouvant exister dans un réseau routier, à savoir les feux, les passages piétons, les carrefours ainsi que les véhicules. La mise en œuvre modélise le caractère localisé des interactions. Nous modélisons en Jade, un système dans lequel les entités peuvent interagir entre elles selon leurs positions relatives. Cependant, afin d'obtenir une visualisation des interactions, nous utilisons une plateforme de visualisation 3D. Conçue par le laboratoire L3I et réalisée par la société Succubus, cette plateforme 3D permet de tester l'impact que des décisions prises peuvent avoir sur l'environnement urbain. Elle fonctionne sur la base d'instructions LUA (<http://www.lua.org/>). On peut constituer un environnement urbain de manière modulaire en créant des plaques où l'on peut ajouter les éléments constituant une ville (routes, feux, bâtiments...) ou en utilisant des plaques prédéfinies. En reliant ces différentes plaques, nous pouvons ainsi créer des environnements urbains représentant les cas d'utilisation que nous souhaitons étudier. L'objectif est de pouvoir utiliser ce système pour représenter un quartier entier d'une ville.

mise en oeuvre du SMA

Le modèle de conception s'appuie sur un modèle générique. Il est fondé sur les unités conceptuelles que sont : le Monde, la Localité, l'Interaction Type et l'Agent.



Ce modèle permet d'associer un rôle à un agent lorsqu'il entre en interaction avec un autre agent. Un rôle est attaché à une interaction. Le modèle permet de faire évoluer le système. La sémantique de l'évolution du système est représentée par un calcul dans lequel un état est un Ambient (Cardelli et al., 1998). L'expression sous forme d'ambients du système permet de valider formellement le modèle sous forme d'agent interagissant. Le modèle des espaces d'interaction permet de représenter les liens entre interaction et localisation. Un agent exécute une boucle de proaction perception, décision, action. La perception consiste d'une part à demander au monde virtuel de fournir la liste des entités perçues, d'autre part à solliciter pour chaque entité la création d'une interaction. La décision est un mécanisme qui se déroule en deux temps. Dans un premier temps, l'entité "récupère" le résultat des interactions actives appelé "influence" selon le principe développé par Ferber. Dans un deuxième temps, un programme de décision synthétise ces influences pour produire la mise à jour des paramètres d'exécution. Les différents types d'agents présents dans notre SMA sont les agents présents dans un environnement urbain, à savoir : les feux, les passages piétons, les carrefours (créés selon les feux présents) et les véhicules.

Interoperabilité SMA/Plateforme de visualisation

Pratiquement, au lancement, le SMA récupère les informations relatives aux composants "fixes" tels que les passages piétons et les feux (nom, positions graphiques...), selon leur présence sur les plaques actives dans le composant de visualisation. Les agents de type "véhicule" sont eux créés à partir de Jade et ils sont ensuite ajoutés à la plateforme 3D. La gestion des comportements et des interactions des agents est définie dans le SMA car celui-ci est indépen-

dant de la plateforme 3D. Cependant, afin que la visualisation soit réelle, les agents tels que les feux ou les véhicules envoient des informations vers la plateforme 3D telles que " je change de couleur " ou encore " je dois avancer ". Actuellement, nous sommes en train de créer un éditeur permettant de faciliter l'utilisation de ces deux plateformes et de renforcer les possibilités offertes jusque là. Notre objectif est de pouvoir généraliser l'utilisation d'autres environnements urbains plus développés pour représenter un quartier afin de valider notre approche et ainsi réussir un passage à l'échelle. Notre éditeur a aussi pour but d'automatiser l'ajout de véhicules. En effet, on veut définir des flux de véhicules afin de représenter réellement ce qui peut se passer dans un quartier urbain selon les horaires de la journée. Dans la version actuelle, les trajets des véhicules sont codés " en dur " en définissant l'ensemble des points de contrôle où ils doivent passer dans la grille. On souhaite pouvoir préciser uniquement le lieu de départ et le lieu de destination des véhicules et que le trajet se calcule de manière automatique via des algorithmes de parcours de graphes. Cette manière de faire pourra nous permettre par exemple de simuler un trafic routier aux abords d'un centre commercial.

Travail en cours et perspectives

Le travail en cours vise à permettre une interactivité de l'utilisateur avec le système et ainsi à intervenir dans le déroulement des scénarios pour modifier les objectifs de l'agent. La manière de mettre en place les interactions du système s'adapte aux problèmes de trafic urbain bien sûr mais aussi à d'autres contextes plus généralistes tels que la scénarisation et l'édition de scénario ainsi que le contrôle et la vérification. Ainsi, les interactions sont programmées de manière externe aux deux plateformes afin que la limite entre le comportement des véhicules et les réactions obtenues soit bien définie. De plus, il existe différentes interactions possibles selon le type de l'agent et selon le contexte telles que la coordination des feux (les feux d'un même carrefour doivent être synchronisés pour obtenir un comportement cohérent vis-à-vis de la circulation) ou encore la réaction d'un véhicule par rapport à un feu rouge (le comportement normal est de s'arrêter au feu rouge, un comportement plus dangereux peut-être de griller le feu). Ainsi, nous avons défini différents types de comportements applicables aux véhicules : conduite normale, conduite dangereuse, ... La définition de ces comportements implique de fixer certains paramètres tels que la distance de sécurité d'un véhicule par rapport à un autre véhicule ou à un feu... Cet environnement devra permettre de prendre en compte les différentes situations pouvant se dérouler dans une ville au niveau des piétons, des véhicules, des flux de véhicules. Ainsi, ce projet doit pouvoir être présenté à une commu-

nauté dans le but que celle-ci puisse s'en servir pour mener à bien des simulations sur la mise en place d'aménagements urbains tels que la réalisation d'un rond point plutôt que d'un carrefour à feux.

Références

Augeraud Michel and Collé Frédéric et Sarramia David - Un modèle et une méthode de conception centrées interaction Revue Ingénierie des systèmes d'information, 2007

J.M. Boussier, P. Estrailier, D. Sarramia et M. Augeraud - Approach to design behavioural models for traffic network users 8th International Conference on Design & Decision Support Systems in Architecture and Urban Planning In proceedings of DDSS2006

M. Augeraud, JM. Boussier, F. Collé et D. Sarramia - Aide à la décision pour la conception de systèmes complexes : Une approche multi agents Revue RTSI série ISI - Avril 2008, vol. 13, pp. 9 à 32

Cardelli, L. et Gordon, A., D. 1998. Mobile Ambients. In Foundations of Software Science and Computation Structures : First International Conference, FOSSACS '98, Berlin Germany

IRIT - Equipe SMAC



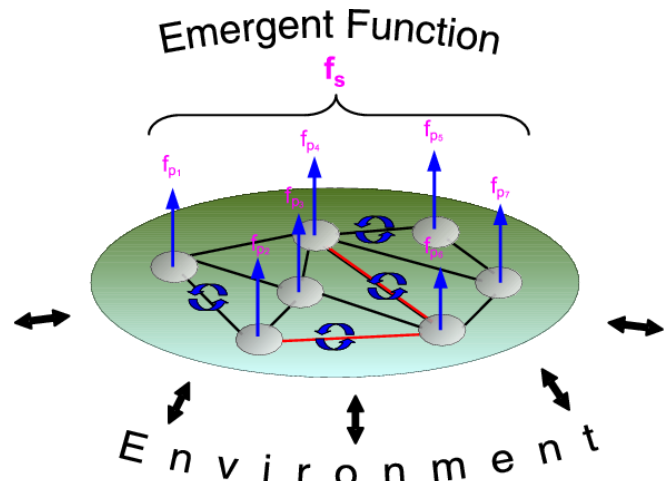
Systèmes Multi-Agents Coopératifs

- CONTACT : Marie-Pierre Gleizes
- ADRESSE 1 (UPS) : IRIT, Université Paul Sabatier, 118 Route de Narbonne, F-31062 TOULOUSE CEDEX 9
- ADRESSE 2 (UT1) : Université Toulouse 1 Sciences Sociales, 2 rue du Doyen-Gabriel-Marty 31042 Toulouse Cedex 9
- ADRESSE 3 (UTM) : Université Toulouse II Le Miral, 5 allées Antonio Machado 31058 TOULOUSE Cedex 9
- Web : www.irit.fr/SMAC
- Tél : 05.61.55.82.94

Membres de l'équipe

L'équipe est constituée de 17 permanents et de 14 doctorants et post-docs, situés sur trois sites : UPS, UT1, UTM (voir adresses ci-dessus).

- Membres permanents
 - Françoise Adreit, MCF, UTM
 - Frederic Amblard, MCF, UT1
 - Eric Andonoff, MCF, UT1
 - Jean-Paul Arcangeli, MCF, UPS
 - Carole Bernon, MCF, UPS
 - Valérie Camps, MCF, UPS
 - Stéphanie Combettes, MCF, UPS
 - Jean-Pierre Georgé, MCF, UPS
 - Marie-Pierre Gleizes (responsable équipe), PR, UPS
 - Pierre Glize, IE, CNRS
 - Chihab Hanachi, PR, UT1
 - André Machonin, IE, INRIA
 - Christine Maurel, MCF, UPS
 - Frédéric Migeon, MCF, UPS
 - Christine Régis, MCF, UPS
 - Christophe Sibertin-Blanc, PR, UT1
 - Marc de Coligny, PR, UTM
- Membres non-permanents
 - Wassim Bouaziz, Doctorant, ATER
 - Issam Bouslimi, Doctorant
 - Paul Chapron, Doctorant
 - Elsy Kaddoum, Doctorant
 - Sylvain Lemouzy, Doctorant
 - Matthias Mailliard, Doctorant
 - Jean-Pierre Mano, Contractuel
 - Emilie Marquois, Doctorant
 - Victor Noel, Doctorant
 - Eliane Pozzebon, Doctorant
 - Sylvain Rougemaille
 - Zied Sellami, Doctorant
 - Ines Thabet, Doctorant
 - Sylvain Videau, Doctorant



Auto-adaptation : le système se modifie en fonction de son environnement

Thème général de l'équipe

L'équipe SMAC s'intéresse à la **modélisation et à la résolution de problèmes dans les systèmes complexes** [DiGKa2005.1] grâce à la technologie multi-agent. Les systèmes complexes sont des systèmes composés d'un grand nombre d'entités en interaction locale dont on ne peut déterminer certains comportements autrement que par l'expérience ou la simulation. Autrement dit, malgré une connaissance parfaite des composants élémentaires du système, on ne connaît pas de programme capable de prévoir son comportement qui soit plus court que celui décrivant le système de manière exhaustive. Maîtriser le fonctionnement de tels systèmes comprenant des non-linéarités provenant de boucles de rétroaction, ainsi que le couplage entre niveaux (le macro-niveau du système est produit par les entités du microniveau, mais il contraint aussi ce micro-niveau) est l'objectif de nos travaux de recherche [GeGIGl2003.1, GlCaGeCa2008.1].

Six axes de recherche

Ces travaux s'articulent autour de six axes fortement liés les uns aux autres par les concepts d'adaptation, d'auto-organisation et d'émergence.

- Axe **auto-adaptation**. Le comportement adaptatif d'un système lui permet de réagir à l'interaction avec un environnement dynamique, soit pour continuer à réaliser sa tâche, soit pour améliorer son fonctionnement. Le défi est alors de concilier l'autonomie d'un système face à des situations imprévues et la maîtrise nécessaire que l'humain doit en avoir.

- Axe **auto-organisation, coopération et émergence**. L'auto-organisation est un mécanisme ou processus qui permet à un système de changer son organisation en cours de fonctionnement et sans contrôle explicite externe. Dans un système multi-agent, l'auto-organisation se traduit par un changement de l'organisation entre les agents. Le défi est alors de trouver les comportements locaux des agents (des lois au niveau individuel), les règles des agents qui mettent en oeuvre l'auto-organisation et qui permettent de concevoir des systèmes fonctionnellement adéquats. L'adéquation fonctionnelle est le jugement positif porté par des utilisateurs finals sur l'activité du système.
- Axe **conception et modélisation**. L'objectif de cet axe est d'améliorer et de fournir une assistance pour le développement d'applications à base d'agents coopératifs.
- Axe **coordination adaptative**. Elle permet de gérer les dépendances entre les activités des agents.
- Axe **résolution collective de problèmes**. L'objectif est ici d'étudier la manière dont les systèmes multi-agents adaptatifs apportent une solution à la résolution de problèmes multi-disciplines, multi-objectifs et multi-échelles. Cette approche permet, en environnement dynamique, de proposer une nouvelle configuration à partir d'une solution précédente en minimisant les modifications endogènes et, surtout, sans recommencer une nouvelle résolution à chaque perturbation.
- Axe **simulation**. La problématique de cet axe est d'étudier et de développer des outils et méthodes permettant d'explorer des modèles. A long terme, on souhaite notamment intégrer à une plate-forme de simulation multi-agent, les outils nécessaires à l'étude de propriétés émergentes.

Le projet ScanMaris

Le projet **ScanMaris - Surveillance et Contrôle des Activités des Navires en Mer** - est un projet ANR-07-CSOSG (Concepts systèmes et outils pour la sécurité globale) qui a débuté le 14 janvier 2008 pour une durée de 24 mois.

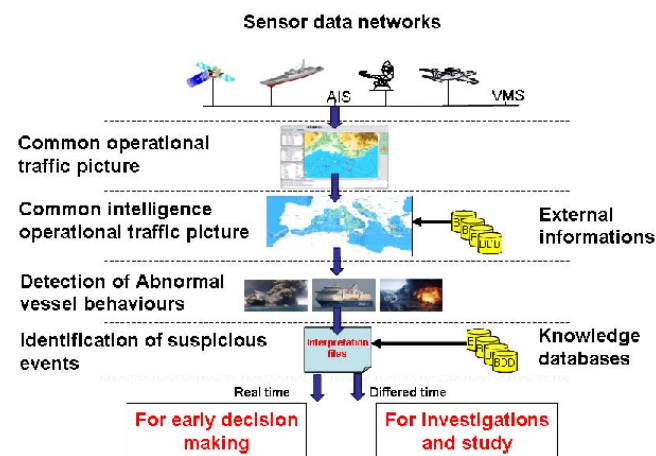
- Contact dans l'équipe : Jean-Pierre Georgé (george@irit.fr)
- Site web : www.irit.fr/ScanMaris

Description des objectifs du projet

Le projet ScanMaris [MoNaLiGeJa2008.1, MoNaLiGIGl2007.1] est un atelier logiciel pour développer et éva-

luer des solutions de prévention et surveillance des frontières maritimes. Ces solutions reposent sur des outils d'exploitation de tenue de situation dite renseignée issue du traitement continu d'importants volumes de données très largement hétérogènes acquises en temps réel (capteurs) et en différé (renseignements).

Les solutions ScanMaris permettent de surveiller les évolutions permanentes du trafic dense dans une zone maritime globale pour suivre les flux de marchandise transportée (vrac, conteneur, énergie, produit chimique, produit de la pêche, passager, etc.) selon différents itinéraires (corridor, autoroute de la mer, cabotage, etc.) et détecter les trafics criminels de produits illicites (stupéfiant, contrebande, immigrant clandestin, ressource halieutique pêchée illégalement, etc.).



Scanmaris : des données jusqu'à la levée d'alerte

Les solutions ScanMaris sont composées d'outils de traitement pour croiser et fusionner des données cinématiques des navires et des renseignements hétérogènes pour établir et maintenir une tenue de situation renseignée complexe et globale du trafic maritime, des méthodes d'apprentissage et des modèles qui exploitent ces tenues de situation renseignées pour acquérir une meilleure connaissance des flux de marchandise transportée pour en améliorer le suivi, et des moteurs de règles d'investigation pour détecter les anomalies comme les flux de produits illicites, les sinistres, les violations à la réglementation, etc.

Cette approche système et transverse de la sauvegarde aux frontières maritime (sécurité du territoire et protection contre les actions de type malveillante) fait appel à des compétences pluridisciplinaires qui sont capitalisées dans un groupement de partenaires complémentaires. Les partenaires sont des industriels (DCNS, SOFRESUD), société de service (ECOMER), des académiques (IRIT, ARMINES, ONERA, Centre de Droit Maritime et des Transports) ainsi que des acteurs opérationnels de la sauvegarde maritime

(Prescripteurs et Opérateurs de la Direction des Affaires Maritimes), Centres Régionaux Opérationnels de Surveillance et Sauvetage (CROSS), Centre Opérationnel de la Marine (COM) et de la Gendarmerie Maritime qui sont membres d'un Comité d'Utilisateurs.

Pour la tenue de situation renseignée du trafic sur l'espace maritime, les données hétérogènes traitées, croisées et fusionnées sont :

- Les observations géo référencées faites en temps réel par des capteurs déployés (radar côtier, radar longue portée, AIS, etc.) et itinéraires des navires simulés.
- Les renseignements extraits en temps différé des bases de données existantes accessibles en ligne comme TF2000, CD-ROM LLOYDS, EQUASIS, TROCS, etc.
- Le zonage de l'espace maritime (eau territoriale, contiguë, économique exclusive, protection écologique, aire de pêche saisonnière, haute mer, etc.).
- Les conditions météorologiques (visibilité, nébulosité, vent de surface, etc.) et océanographiques (spectre de la houle, température de surface de la mer, courant de surface, etc.).

Les fonctions étudiées, développées et évaluées dans le projet ScanMaris permettront d'améliorer l'efficacité d'une surveillance globale et de la détection des contrevenants. Elles contribueront donc à mettre en place des réponses optimales dans le temps et les moyens d'intervention adaptés de lutte contre les activités illicites et les violations aux réglementations.

L'approche SMA dans ScanMaris

La modélisation de comportement de navires est un domaine scientifique peu exploré. Le projet a pour objet aussi de réaliser une modélisation de ces comportements complexes.

Tout navire (marchand, passager, plaisance, pêche, etc...) peut être considéré comme un agent qui cherche à satisfaire un objectif en allant du port de départ à destination. L'activité illicite cherche toujours à se masquer derrière les objectifs déclarés d'un ou plusieurs navires. L'ensemble des transports maritimes est ainsi un système multi-agent dont les nombreux agents sont potentiellement en interaction, et qui ont des comportements évolutifs et plus ou moins nominaux. Détecter la fraude maritime (contrebande, terrorisme, rejets illégaux, pêche interdite, etc...) consiste donc à faire émerger des activités implicites à partir de comportements quasi standards. Les systèmes multi-agents adaptatifs exhibant des propriétés d'auto-organisation, semblent pertinents pour réaliser cette tâche. En effet, la théorie des AMAS (Adaptive Multi-Agent Systems) permet une résolu-

tion émergente de problèmes, de manière que l'apprentissage des agents ne soit pas guidé par la connaissance d'une fonction globale définie préalablement à l'apprentissage par le concepteur. Cette technique d'apprentissage s'avère donc particulièrement adaptée pour la découverte de fraudes, qui par nature sont non explicites et évolutives.

ScanMaris impose aussi une combinaison particulière d'algorithmes d'apprentissage pour la détection d'activités illicites pour deux raisons :

- Le corpus d'apprentissage est contraint par les cas réels connus. Il faudra segmenter en type de fraude (pêche, pollution, trafics illicites, immigration clandestine, menaces asymétriques, etc...) et classer les comportements selon les activités des navires (citerne, ferry, pêche, marchand, porte-conteneur, plaisance, etc...) pour éviter d'avoir un surapprentissage et pour bien observer les fraudes.
- Les activités illicites sont en constante évolution, ce qui impose la détection sans apprentissage supervisé. Pour cela nous emploierons un apprentissage émergent fondé sur la technologie des systèmes multi-agents adaptatifs.

Ainsi, l'enjeu scientifique est d'importance et très actuel, car il faudra pouvoir combiner en temps réel l'emploi de méta-heuristiques avec un algorithme d'apprentissage émergent qui n'a pas de fonction objectif.

LAMIH - RAIHM

Raisonnement Automatique et Interaction Homme-Machine

- CONTACT : R. Mandiau
- ADRESSE :
LAMIH UMR CNRS 8530
équipe RAIHM - thème SMA
Université de Valenciennes
59300 Valenciennes Cedex 9
- WEB : www.univ-valenciennes.fr/LAMIH
- TEL : 03.27.51.14.38

Membres de l'équipe

Permanents : Emmanuel Adam, Veronique Delcroix, Sylvia Estivié, Emmanuelle Grislin-Le Strugeon, René Mandiau, Sylvain Piechowiak.

Thèses en cours : David Gacquer, Amine Hamila, Yoann Lebrun, Pierre Monier.

Thème général de l'équipe

Les travaux s'inscrivent au LAMIH (Laboratoire d'Automatique, de Mécanique et d'Informatique, Industrielles et Humaines). L'équipe RAIHM (Raisonnement Automatique et Interaction Homme-Machine) du Laboratoire est actuellement dirigé par le PR C. Kolski. Cette équipe a pour vocation d'entreprendre des recherches dans le domaine des Interactions Hommes-Machines et de l'Intelligence Artificielle/Intelligence Artificielle Distribuée - Systèmes multi-agent (IAD-SMA). Les travaux dans le domaine des SMA visent plus particulièrement à la conception de systèmes ouverts dynamiques ; dont la modélisation se caractérise par la description de comportements de différents agents. Nous nous intéressons également à des problématiques de décision et de résolution de situations compétitives entre ces derniers.

Description des travaux ou projets en lien avec le thème du dossier

Les travaux en lien avec le thème du dossier sont effectués dans le thème "IAD -SMA". A la fin des années 90, les travaux se sont orientés vers des applications dans le contexte du transport (l'une des problématiques transversales du LAMIH). Nous avons effectué des recherches nous permettant d'acquérir une bonne expérience dans des applications de simulation pour les transports : régulation de bus (jusqu'à 2002), comportements émergents dans des modèles de piétons (à partir de 2002), des aspects liés à l'information multimodale (à partir de 2001) ainsi que dans la simulation de trafic routier (à partir de 1999).

Régulation de Trafic de Bus

L'objectif de cette étude [Laichour02a, Laichour02b] vise à améliorer la qualité des transports en fournissant une aide à la régulation de trafic de bus. Elle modélise l'application sous forme d'une organisation multi-agents. Pour respecter au mieux les horaires théoriques annoncés aux clients/usagers, il est nécessaire pour l'exploitant d'assurer une régulation en temps réel. Ce processus permet de compenser les incidents du réseau (se traduisant par des retards) en effectuant certaines manoeuvres sur les autobus. La régulation étant complexe, elle est réalisée par un opérateur humain qui travaille essentiellement par expérience (le processus de régulation n'est pas formalisé). Afin de répondre aux besoins du régulateur et à la satisfaction des usagers, il a été proposé le développement d'un outil d'aide à base d'agents regroupés dans une organisation en fonction des flux d'informations. Les interactions influencent les comportements des agents, tout en respectant des lo-

giques de régulation. Les principes décrits précédemment ont fait l'objet d'applications pour la CTPM (Compagnie des Transports du Pays de Montbéliard) et pour la Ste Semurval (Valenciennes).

Conception et réalisation de l'animation comportementale de piétons virtuels

Développé dans le cadre d'un projet PREDIT, un mécanisme de sélection d'action basé sur le vote [Hanon07, Hanon06, Hanon05] a été appliqué à la commande de piétons virtuels dans le logiciel RESPECT (Route Empruntée en Sécurité par le Piéton-Enfant Confronté au Trafic). Ce projet visait à concevoir un simulateur éducatif permettant de placer l'enfant-joueur dans des situations accidentogènes engendrées par des interactions entre les piétons et les véhicules. Dans ce cadre, nous avons contribué à la conception et à la réalisation de l'animation comportementale des piétons virtuels, chacun d'eux étant "piloté" par un agent. Les actions à modéliser par les agents correspondent aux commandes de mouvement à générer. Dans le contexte de ce projet, les actions sont appliquées dans un premier temps à un mobile qui correspond à la représentation des caractéristiques physiques de l'agent-piéton, puis répercutées aux niveaux géométrique et graphique. L'architecture proposée est une architecture hybride, intégrant des propriétés cognitives et réactives. Le raisonnement cognitif est utilisé lors d'une phase initiale de recherche d'itinéraire, et en navigation, c'est-à-dire dans la décision concernant l'adoption d'un nouvel objectif. Le raisonnement réactif est le comportement principalement actif en cours de simulation. Il permet de réagir, de manière appropriée et à chaque pas de la simulation, à l'environnement dynamique. Le modèle proposé a également été adapté pour modéliser le comportement des conducteurs. Cela permet de généraliser partiellement la représentation des acteurs, qu'ils soient piétons ou conducteurs de véhicules.

Information multimodale

Dans le domaine de la conception de systèmes interactifs basés sur un SMA, nous nous sommes intéressés en particulier au SMA en tant que système central d'un système d'information [Anli06, Petit06]. En effet, l'approche multi-agent permet à la fois une adaptation à un environnement distribué, dynamique et ouvert tel que l'internet et les bases de données réparties, et une adaptation aux besoins en information de différents utilisateurs. Dans ce cadre, nous avons proposé une architecture multi-agents et une méthode de personnalisation des informations à l'aide d'agents.

Un premier projet, le projet AgenPerso, financé par le PREDIT, nous a permis d'appliquer nos travaux à un prototype de système de recherche d'itinéraire combinant plusieurs modes de transport. L'utilisateur fait une demande de recherche d'itinéraire classique ; le système retourne un ensemble de solutions de déplacement multi-modes de transport et classées selon les préférences supposées de l'utilisateur.

Ces travaux ont été poursuivis dans le cadre du projet MOUVER.PERSO financé par le PREDIM, dans le but de recommander des itinéraires adaptés afin de faciliter les déplacements au quotidien des étudiants du Valenciennois. L'application visée est plus exactement un agenda (sur pages web et sur PDA) capable de proposer une assistance à la planification liée à l'offre de transport. Son originalité réside en la proposition de solutions de transport correspondant aux rendez-vous planifiés. En effet, l'ajout d'un événement lié à un déplacement, dans l'agenda, déclenche le calcul d'un itinéraire correspondant. De plus, cet itinéraire est choisi en fonction du profil de l'utilisateur enregistré dans le système.

Simulation de trafic routier

Ce travail fait l'objet d'une collaboration avec l'INRETS (équipe MSIS, dirigée par S. Espié (section)) sur le simulateur ArchiSim. Les études proposées visent à modéliser de manière "réaliste" les comportements humains. Dans ce contexte, un agent humain est plongé dans un environnement de simulation (routier en l'occurrence) mettant en jeu des interactions entre agents artificiels d'une part, agents artificiels/agents humains d'autre part. La modélisation de chaque agent et de ses comportements s'avère une entreprise délicate, puisque les comportements sont différents (voire égoïstes). Des solutions dans des situations critiques (par exemple, des carrefours) ont été proposées et comparées avec des mesures de trafic. Notons que le problème des carrefours est en fait un problème difficile et mal traité dans les outils actuels (souvent considérés comme une 'boîte noire' dans laquelle les véhicules entrent et sortent. Dans la suite, nous avons considéré que les véhicules caractérisent les agents.

Une première étude en collaboration avec l'INRETS et la société CS-SRILOG, vise à modéliser les interactions entre des agents. En fonction de l'observation du contexte, les interactions entre les agents sont décrits sous forme de gains définis dans une matrice de jeu [Champion04, Mandiau08a]. Chaque agent joueur sélectionne son action en fonction de son gain potentiel et du gain des autres joueurs. A partir de cette évaluation et d'un comportement de type

égoïste ou altruiste, les agents décident de sélectionner leurs propres actions.

Une seconde étude (également une collaboration avec l'INRETS) consiste à modéliser l'environnement perçu par chaque agent comme un réseau de contraintes [Doniec06, Doniec08a, Doniec08b]. Chaque agent essaie d'anticiper le comportement des autres agents et de détecter des situations de blocage par détection d'incohérences du réseau de contraintes. L'agent peut ainsi déduire s'il y aura une situation de blocage inter-véhicules. Le travail a également mis en évidence l'importance de la notion d'agents non normatifs dans l'évolution du trafic global (c'est-à-dire, dans le contexte de la conduite, ces derniers ne respecteront plus obligatoirement le code de la route, considérée jusqu'à présent comme une norme "inviolable").

Références

- [Anli06] A. ANLI (2006). *Méthodologie de développement des systèmes d'information personnalisés - Application à un système d'information au service des usagers des transports terrestres de personnes*. Thèse de Doctorat, Université de Valenciennes et du Hainaut-Cambrésis, octobre.
- [Champion04] A. CHAMPION (2004). *Mécanisme de coordination multi-agent fondé sur des jeux : Application à la Simulation Comportementale de Trafic Routier en situation de Carrefour*. Thèse de Doctorat, Université de Valenciennes, Dec.
- [Doniec06] A. DONIEC (2006). *Prise en compte des comportements anticipatifs dans la coordination multi-agent : application à la simulation de trafic en carrefour*. Thèse de Doctorat, Université de Valenciennes, Oct.
- [Doniec08a] A. DONIEC, R. MANDIAU, S. PIECHOWIAK, S. ESPIÉ (2008). Anticipation based on constraint processing in a multi-agent context. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, 17, pp. 339-361.
- [Doniec08b] A. DONIEC, R. MANDIAU, S. PIECHOWIAK, S. ESPIÉ (2008). Controlling non-normative behaviors by anticipation for autonomous agents. *Web Intelligence and Agent Systems : An International Journal*, 6, pp. 29-42.
- [Hanon05] D. HANON, E. GRISLIN-LE STRUGEON, R. MANDIAU (2005). A behaviour based decisional model using vote. *M. Mohammadian (Ed.), Proceedings IAW-TIC'2005 International Conference on Intelligent Agents, Web Technologies and Internet Commerce (28-30 November 2005, Vienna, Austria)*, novembre, ISBN 1740-88247-4.
- [Hanon06] D. HANON (2006). *Modèle décisionnel orienté comportement fondé sur le vote : Application à la navigation d'agents autonomes en environnement simulé*.

Mémoire de doctorat, Université de Valenciennes et du Hainaut-Cambrésis, décembre.

[Hanon07] R. MANDIAU, D. HANON, E. GRISLIN-LE STRUGEON (2007). Modèles comportementaux appliqués à la navigation de piétons. *M.a. Granié, J.m. Auberlet (Ed.), Le piéton et son environnement - Quelles interactions ? Quelles adaptations ?*, Actes du 1er colloque francophone de la plate-forme intégrative COPIE - Comportement du Piéton dans son Environnement (Paris, Nov. 2007), Les Collections de l'INRETS, Paris, pp. 49-60, novembre, ISBN 978-2-85782-661-3.

[Laichour02a] H. LAICHOIR (2002). *Modélisation multi-agent et Aide à la Décision : Application à la Régulation des correspondances dans les réseaux de Transport Urbain*. Thèse de Doctorat, Villeneuve d'Ascq, 20 Décembre 2002.

[Laichour02b] H. LAÏCHOIR, S. MAOUCHE, R. MANDIAU (2002). Un système multi-agent pour la régulation des correspondances. *P. Mathieu, J.P. Müller (Ed.), Systèmes multi-agents et systèmes complexes, ingénierie, résolution de problèmes et simulation (JFIADSMA'02)*, Hermes Science Publications, Paris, pp. 257-259, janvier.

[Mandiau08a] R. MANDIAU, A. CHAMPION, J-M AUBERLET, S. ESPIÉ, C. KOLSKI (2008). Behaviour based on decision matrices for a coordination between agents in urban traffic simulation. *Applied Intelligence*, 28, pp. 121-138.

[Petit06] C. PETIT-ROZÉ, E. GRISLIN-LE STRUGEON (2006). MAPIS, a multi-agent system for information personalization. *Information and Software Technology*, 48, pp. 107-120.

LAMSADE - AIMC

Agents Intelligents, Modèles Coopératifs

- CONTACT : Suzanne Pinson (Pr) - pinson@lamsade.dauphine.fr,
- ADRESSE : Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16
- WEB : www.lamsade.dauphine.fr/aimc/interaction.html
- TEL : 01 44 05 45 51.

Membres de l'équipe

Fabien Badeig, Flavien Balbo, Tristan Cazenave, Suzanne Pinson.

Thème général de l'équipe

Fondée sur le principe PbC (Property-based Coordination (section)), nos travaux reposent sur la modélisation d'un environnement multi-agent pour la conception de systèmes d'aide à la décision dédiés au domaine du transport. Sur ce principe, nous avons proposés trois modèles : 1) le modèle EASI (Environnement comme Support Actif de l'Interaction) pour la gestion des interactions dans un SMA ; 2) le modèle EASS (Environnement comme Support Actif de la Simulation) pour la conception de simulations multi-agents ; 2) le modèle ACIOS et le langage LACIOS pour la génération d'applications multi-agents.

En collaboration avec le laboratoire GRECIA de l'INRETS (section), ces travaux servent de support au développement de systèmes d'aide à la décision dédiés au transport : 1) SATIR pour l'aide à la décision des régulateurs de réseaux de bus (section) ; 2) SAIV pour l'aide à la décision du voyageur (section) ; 3) Lacios-TAD pour le transport à la demande (section) ; 4) SGC pour l'aide à la décision en situation de crises (section).

Description des travaux

EASI (Environnement comme Support Actif de l'Interaction)

EASI propose une modélisation de l'environnement permettant un partage des interactions. La difficulté est de permettre à chaque agent de trouver "efficacement" les interactions qui l'intéressent. Pour cela nous utilisons une approche fondée sur l'Analyse de Données Symboliques (ADS). L'ADS a pour objectif de découvrir des connaissances parmi un grand ensemble de données qualitatives ou quantitatives en procédant au regroupement dans ce qui est appelé un objet symbolique de données liées par un même concept. Dans nos travaux, nous considérons que l'environnement contient la description symbolique de ses composants (agents, messages, traces ou objets). Une interaction peut alors être modélisée comme la définition en intensification des objets symboliques la composant. Le problème est alors de permettre aux agents d'utiliser ces descriptions pour construire leur domaine de perception des interactions [Bal08, Sau08].

En collaboration avec le GRECIA-INRETS, ce travail a permis la réalisation d'un système multi-agent comme support d'un système d'aide à la décision destiné aux régulateurs d'un réseau d'autobus (SATIR (section [Bal08b, Bal07, Bal05])) et d'un système d'aide à la décision temps réel pour le voyageur (SAIV (section [Zar05])). Le choix du mode d'interaction du système multi-agent a constitué un défi. Le contrôle et la connaissance nécessaire au fonctionnement

du réseau étant répartis entre les différents agents, il est nécessaire de proposer un protocole d'interaction permettant aux agents de coopérer. Alors que dans de nombreux systèmes l'identification des agents partenaires repose sur la recherche de compétences, dans nos applications, l'identification d'un partenaire repose essentiellement sur sa position dans le réseau. De plus, l'information *position* est mise à jour régulièrement et représente un flux de données important. Ces deux caractéristiques rendent peu efficace l'utilisation des technologies classiquement mises en œuvre dans les SMA afin de permettre à un agent de trouver les partenaires adéquats. En effet, l'utilisation d'agents intermédiaires (matchmaker; facilitator; mediators; brokers) ou d'acointances serait très coûteuse en terme de communications supplémentaires nécessaires à la mise à jour des informations. Avec EASI, l'environnement n'est plus considéré comme une entité passive mais comme une entité active dont la fonction est de trouver pour chaque message les récepteurs adéquats. Ce modèle d'interaction se distingue des modèles du type tableau noir où les agents doivent volontairement déposer et surtout lire les messages les concernant. Le modèle d'interaction proposé enrichit les possibilités d'interaction entre les agents :

- La recherche des participants de l'interaction repose sur une représentation logique et relative de leurs caractéristiques au moyen de filtres de communication.
- Les agents peuvent intercepter des communications : un agent peut ainsi recevoir des communications qui ne lui sont destinées mais dont le contenu peut l'intéresser. Cette possibilité est exploitée dans la gestion d'un réseau de transport afin de permettre aux agents de gérer les incohérences relatives aux données.

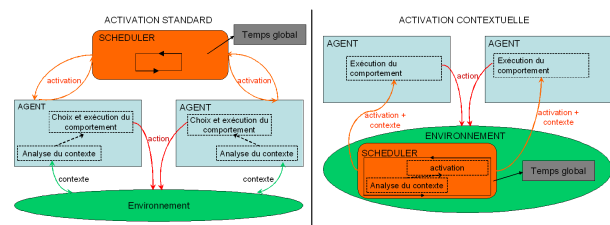
EASS (*Environnement comme Support Actif de la Simulation*)

Nous nous sommes intéressés à la mise en œuvre du processus d'activation des agents dans les plates-formes de simulation à temps discret d'agents situés. Dans ce cadre, les agents évoluent dans un environnement et ne peuvent agir qu'une fois par cycle de temps. La phase d'activation consiste à déclencher l'agent pour qu'il participe au cycle courant de la simulation. La phase d'action correspond alors au processus durant lequel un agent analyse son contexte local pour déterminer l'action la plus pertinente à exécuter en fonction de son contexte local, et exécute l'action. Ainsi, chaque phase d'action est débutée par une analyse du contexte de l'agent pour déterminer l'action adéquate à exécuter pour l'agent. Lorsque la politique d'ordonnancement est globale (i.e. le scheduler est responsable de la politique d'ordonnancement), le processus classique consiste à activer chaque agent à chaque cycle de temps.

Durant cette activation, l'agent récupère et analyse localement son contexte pour choisir l'action qu'il va exécuter.

La procédure liée à l'analyse et à la récupération du contexte (*context computation*) implique de récupérer et d'analyser les données accessibles et réparties dans l'environnement pour chaque agent. Le problème est que cette procédure locale à chaque agent augmente la complexité de la modélisation de la simulation. Par ailleurs, le concepteur de la simulation dispose de deux moyens pour modifier les paramètres de la simulation. Le premier est de modifier l'ordonnancement global de la simulation et donc la manière d'activer les agents. La seconde consiste à modifier le comportement des agents, et plus précisément leur réaction au contexte local. Cette méthode implique souvent de modifier en partie la programmation de l'agent. Notre objectif est de lier explicitement l'activation de l'agent à un contexte particulier. La reification du lien entre le contexte et l'activation de l'agent est appelée *activation contextuelle* et est gérée par l'environnement. Ainsi, nous externalisons la partie relative à la récupération et analyse du contexte (Figure).

Notre proposition est d'appliquer le principe de PBC au processus d'activation des agents. L'objectif est que les agents choisissent dynamiquement leur contexte d'activation et laisse la gestion de la simulation à l'environnement. De plus, l'environnement est responsable du temps de la simulation et les agents gèrent leur propre échelle de temps interne. L'approche est orientée environnement, car nous considérons et plaçons l'environnement au centre de notre modélisation. La figure résume le rôle de l'environnement dans notre modélisation par rapport à l'approche classique.



Différence entre activation classique et contextuelle

Le résultat escompté est une amélioration du contrôle de la simulation. Le contrôle est centralisé, car il est assuré par l'environnement et les agents de la simulation peuvent modifier dynamiquement leur lien à la simulation.

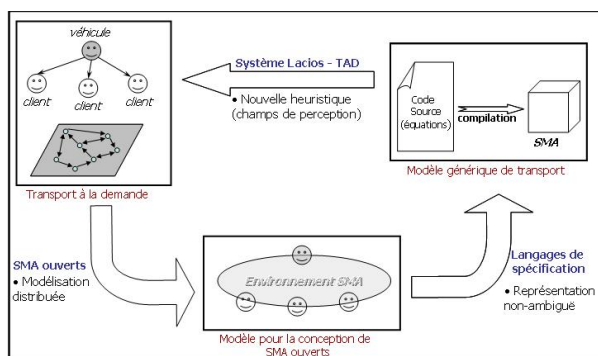
Ce modèle est mis en œuvre dans la conception d'un système d'aide à la décision pour la gestion de crise (section). Le modèle EASS va être mis à profit afin de rendre compte par le simulateur des difficultés liées à la répartition des connaissances dans l'espace le temps et les services impliqués dans la gestion d'une crise.

Acios - Lacios

Nous avons proposé Acios (Agent Contextual Interaction in Open Systems), un modèle de coordination permettant de concevoir des SMA ouverts conformes à PbC [Zar07]. Lacios (Language for Acios) est le langage associé à Acios. Il permet de spécifier formellement le comportement d'un SMA adhérant à Acios, grâce à une syntaxe et une sémantique formelle. Nous proposons également le langage de programmation Java-Lacios, qui est un langage de script au dessus de Java permettant l'implantation de SMA écrits en Lacios et exécutés en Java.

Avec un ensemble limité de primitives et grâce à une syntaxe permettant l'écriture d'expressions d'une grande complexité, Acios permet la conception et Lacios la réalisation d'applications multi-agents dédiées au transport d'une manière plus simple. Les contraintes du problème sont transformées directement en contraintes d'interaction entre agents, et leur comportement est spécifié avec un code compact et plus aisément contrôlable par le programmeur.

Le travail sur le modèle de coordination Acios et le langage Lacios s'est fait en parallèle avec le développement de l'application de STAD (section . Le schéma récapitule le cheminement depuis le problème des STAD jusqu'à la définition d'une solution multi-agent au problème, en passant par la définition du modèle Acios, du langage Lacios et du langage de programmation Java-Lacios.



Interaction entre modèle et application

Références

[Bal08] F. Balbo, J. Saunier. On the Use of Symbolic Data Analysis to Model Communication Environments, M. Klusch, M. Pechoucek, and A. Polleres (Eds.) : CIA 2008, LNAI 5180, pp. 230—244, 2008.

[Bal08b] F. Balbo, S. Pinson. An Agent oriented approach to Transportation Regulation Support Systems, In Proce-

dings 5th Workshop on Agents in Traffic and Transportation, Autonomous Agents and Multiagent Systems, AAMAS 2008.

[Bal07] F. Balbo, S. Pinson. A Transportation Decision Support System in Agent-Based Environment, Journal of Intelligent Decision Technologies 1(3), pages 97-115, 2007.

[Bal05] F. Balbo, S. Pinson, Dynamic modeling of a disturbance in a multi-agent system for traffic regulation, International Journal of Decision Support System, vol.41 (1), pages 131-146, 2005.

[Sau08] J. Saunier, F. Balbo. Regulated Multi-party Communications and Context Awareness through the Environment, Multi-Agent and Grid Systems 00, 2008.

[Zar05] M. Zargayouna, F. Balbo, J. Saunier, "Agent Information Server : a middleware for traveler information". In O. Dikenelli, M.-P. Gleizes and A. Ricci (Eds.) : International Workshop on Engineering Societies in the Agents World (ESAW 2005), Kusadasi, Turquie, pages 3-16, Octobre 2005. Volume 3963 of Lecture Notes in Artificial Intelligence. Springer Verlag.

LIFL - SMAC

Systèmes Multi-Agents et Comportements

- CONTACT : Ph MATHIEU (Pr) - mathieu@lifl.fr,
- ADRESSE : LIFL, UMR 8022 CNRS, Bat M3, Université de Lille 1, 59655 Villeneuve d'Ascq Cedex,
- WEB : www.lifl.fr/SMAC,

Membres de l'équipe

Philippe Mathieu, Sébastien Picault, Benoit Lacroix, Laetitia Bonte

Thème général de l'équipe

L'équipe SMAC travaille sur le domaine de l'Intelligence Artificielle essentiellement sous l'aspect Génie Logiciel et plus précisément sur le thème de la modélisation de comportements entre entités virtuelles. L'équipe s'est spécialisée dans la modélisation de comportements, la mise en oeuvre de ces comportements dans les environnements nécessitant des entités autonomes, l'évaluation de comportements entre agents et la réalisation de plateformes ouvertes d'expérimentation. Les objectifs actuels concernent la théorie des jeux et l'économie computationnelle ; les comportements dans les plateformes de simulations à entités géographiquement situées comme les jeux vidéo ou les simulateurs de trafic et simulateurs de conduite ; ainsi que

les plateformes de E-Commerce via les systèmes de négociation automatique.

Projets

Notre équipe travaille sur les projets suivants :

- Collaboration avec Renault SA pour la différenciation comportementale des véhicules dans les générateurs de trafic. Ce travail est développé dans le cadre de la thèse de B Lacroix en collaboration avec l'équipe d'A Kemeny.
- Collaboration avec l'INRETS pour la modélisation du comportement des deux roues motorisés. Ce travail est développé dans le cadre de la thèse de L Bonte en collaboration avec l'équipe de S Espié (section).

Références

[Bon06] Laetitia Bonte and Philippe Mathieu and Stéphane Espié, Modélisation et Simulation des Usagers des deux-roues motorisés dans ARCHISIM, Article publié dans Actes des 14e Journées Francophones sur les Systèmes Multi-Agents (JFSMA'2006), pages 31–44, en 2006.

[Bon07] Laetitia Bonte and Stéphane Espié and Philippe Mathieu, Virtual Lanes Interest for Motorcycles Simulation, Article publié dans Proceedings of the fifth European Workshop on Multi-Agent Systems, pages 580–596, en 2007.

[Kub08] Yoann Kubera and Philippe Mathieu and Sébastien Picault, Interaction-Oriented Agent Simulations : From Theory to Implementation, ECAI 08, Patras, July 2008, Greece.

[Lac07] Benoit Lacroix and Philippe Mathieu and Andras Kemeny, Towards Traffic Generation with Individual Driver Behavior Model Based, Article publié dans DSC Driving Simulation Conference — North America, Iowa City, USA, en Sept 2007.

[Lac08] Benoît Lacroix and Philippe Mathieu and Andras Kemeny, The Use of Norms Violation to Model Different Agents Behaviors, Workshop Coin@aamas, Estoril, May 2008, Portugal.

Pour plus de références : www.lifl.fr/SMAC/publications/

LIV

Laboratoire Image et Ville.

- CONTACT : Arnaud Banos (CR1 CNRS) - arnaud.banos@lorraine.u-strasbg.fr, Directeur adjoint du GDR-E Spatial Simulation for Social Sciences (s4.parisgeo.cnrs.fr/)

Membre de l'Institut des Systèmes Complexes-PIF (www.iscpif.fr/)

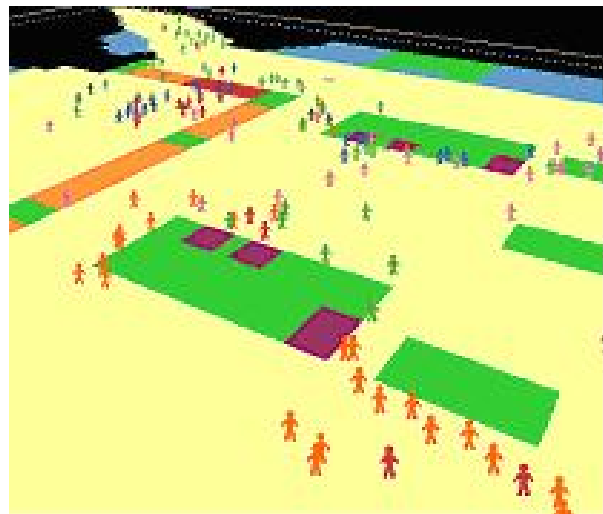
- ADRESSE : Image et Ville, UMR 7011 ULP/CNRS, 3 rue de l'argonne, 67000 Strasbourg
- WEB : web.univ-pau.fr/banos/banos.html,
- TEL : 0684380155.

Projet : Simuler les déplacements de piétons dans les espaces de transport

Equipe :

- Arnaud Banos (Image et Ville, UMR 7011 CNRS/ULP, Strasbourg) : chercheur CNRS, géographe
- Cyrille Bertelle (LITIS, EA4108 Université du Havre), Professeur d'Informatique
- Laurence Bornand (RATP) Responsable d'Unité
- Jérémie Fiegel (RATP / LITIS), doctorant Cifre en informatique
- Angèle Paré (RATP) Ingénieure

Objectifs Scientifiques



Des agents dans le métro (au premier plan, file d'attente au distributeur de tickets)

Au cœur de toute chaîne de déplacements, la marche à pied reste paradoxalement le mode de déplacement le moins bien connu. Diffus, quasi isotropes, souvent de courte portée et dans une grande mesure stochastiques, les déplacements de piétons sont aussi difficiles à observer qu'à modéliser. Pourtant, leur étude est aujourd'hui favorisée par deux phénomènes contingents. L'émergence progressive de politiques pro-environnementales, en revalorisant les modes de transports doux, a ainsi fait prendre conscience de la rareté des connaissances disponibles.

Mieux comprendre les stratégies de déplacement des piétons est alors progressivement devenu l'un des enjeux forts de la planification urbaine. Par ailleurs, les travaux de recherche menés sur les phénomènes d'auto-organisation ont trouvé dans les déplacements de piétons une thématique de recherche privilégiée. Les trajectoires des piétons peuvent en effet être vues comme le résultat émergent d'un système composés d'éléments en interaction : un espace, porteur d'opportunités et de contraintes, qui configure et oriente les itinéraires ; des piétons dotés de buts et mobilisant des stratégies intégrant leur environnement spatial et humain. Pour le gestionnaire des espaces publics de transports, où les expériences in vivo sont difficiles à mener, la simulation des déplacements piétons est moyen avantageux d'explorer les flux, tant dans des perspectives sécuritaires que commerciales. Deux objectifs parfois difficiles à concilier guident en effet l'aménagement de ces espaces : rendre possible la circulation libre et sécurisée des flux dans des espaces où se rassemble parfois une foule importante, et offrir des activités et services imposant des arrêts et des temps d'attente. Les méthodes de simulation des déplacements piétons se sont considérablement développées au cours des 15 dernières années. Les progrès ont été tels que plusieurs modèles sont aujourd'hui commercialisés et exploités par de grandes entreprises, telles que la SNCF. Toutefois, la situation actuelle est caractérisée par une grande diversité de modèles concurrents, parmi lesquels aucun ne semble encore en mesure de jouer un rôle réellement dominant. Par ailleurs, dans un contexte opérationnel, il est difficile d'implémenter et de comparer ces différents modèles, qui répondent à des logiques différenciées et nécessitent en général des données d'entrée différentes. Dans le cadre d'une collaboration avec la RATP, nous avons alors cherché à construire des modèles géographiques minimalistes (MAGE et SIMTRAP), qui soient d'une part suffisamment simples pour pouvoir être construits et implémentés en un temps réaliste, puis modifiés au gré des besoins rencontrés et exprimés, et d'autre part suffisamment légers pour pouvoir être manipulés et testés par des non modélisateurs.



Le rôle de séparation des flux joué par les points d'entrée (tourniquets) et de sortie (portes)

Références

- Banos A, Paré A. (2009) Simuler les déplacements de piétons dans une station de métro : une approche à base d'agents, *Revue Internationale de Géomatique*, N°4-2009, 25 p., accepté pour publication
- Banos A, Paré A. (2008) Des agents dans le métro : le modèle MAGE, *Actes de la conférence INFORSID 2008*, Fontainebleau, 27-28 juin, 13 p.
- Banos A, Charpentier A. (2007) Simulating pedestrian behaviour in subway stations with agents, *Proceedings of the 4th European Social Simulation Association*, Toulouse, France, September 10-14, p. 611-621
- Fiegel J., Banos A., Bertelle C. (2008) Modelling And Simulation Of Pedestrian Behaviours In Transport Areas : The Specific Case Of Platforms/Trains Exchanges, In *Proceedings Of The Agent Based Spatial Simulation (ABS2) International Workshop*, 24-25 November, Paris

Accroître l'accessibilité à la ville pour maîtriser la mobilité urbaine : le projet MIRO2

Financement :

- cadre : ANR Villes Durables, 2008-2012.
- Coût total du projet : 1 734 720 euros
- Aide alloué par l'ANR : 345 785 euros

Responsable scientifique : Arnaud Banos (arnaud.banos@lorraine.univ-strasbourg.fr)

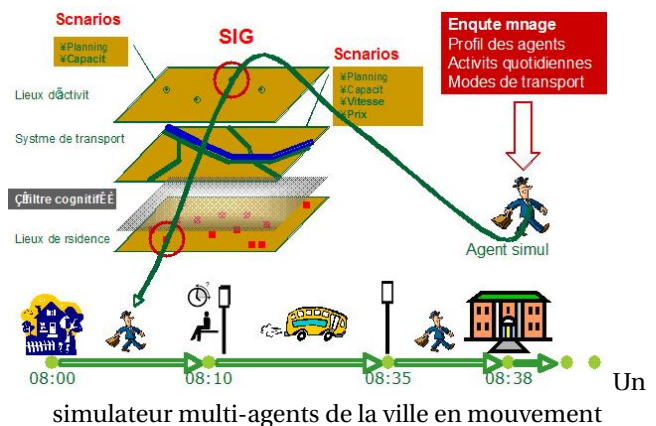
Equipes :

- IMAGE ET VILLE (LIV), UMR 7011 CNRS / Université Louis Pasteur, Strasbourg 1
 - Responsable : Arnaud Banos (Image et Ville, Strasbourg) : chercheur CNRS, géographe
 - Annabelle Boffet (CEDETE, Orléans) : maître de conférences en géographie
 - Arnaud Piombini (Image et Ville, Strasbourg) : maître de conférences en géographie
- LABORATOIRE D'INFORMATIQUE DE L'UNIVERSITE DE FRANCHE-COMTE, (LIFC), EA4157, Besançon
 - Responsable : Christophe Lang (LIFC, Besançon) : maître de conférences en informatique
 - Fabrice Bouquet (LIFC, Besançon) : Professeur d'Informatique
 - Nicolas Marilleau (IRD, Bondy) : ingénieur de recherche, informaticien
 - Pierre-Henri Morand (CRESE, Besançon) : Professeur de sciences économiques
- POLITIQUES PUBLIQUES, ACTION POLITIQUE, TERRITOIRES (PACTE), UMR5194 CNRS / Université Joseph Fourier, Grenoble
 - Responsable : Sonia Chardonnel (PACTE, Grenoble) : chercheuse CNRS, géographe
 - Elise Beck, (PACTE, Grenoble) : maître de conférences en géographie
 - Isabelle Bouhet-André-Poyaud (PACTE, Grenoble) : ingénieure d'études CNRS, statisticienne
 - Laure Charleux (PACTE, Grenoble) : maître de conférences en géographie
- THEORISER ET MODELISER POUR AMENAGER (THEMA), UMR6049 CNRS / Université de Bourgogne, Dijon
 - Responsable : Thomas Thévenin (Théma, Dijon) : maître de conférences en géographie
 - Benjamin Motte (Théma) : maître de conférences en géographie

Objectifs Scientifiques

Ce projet de recherche, qui s'inscrit dans la continuité d'un programme PREDIT, propose d'interroger la ville durable au travers d'une de ses composantes centrales, la mobilité quotidienne. En effet, si l'on peut admettre que la ville est une forme d'organisation spatiale qui favorise les interactions sociales, force est d'admettre qu'un nombre toujours plus grand de déplacements quotidiens est nécessaire pour atteindre cet objectif de mise en relation. L'étalement urbain, la spécialisation fonctionnelle des territoires urbains et la sur-valorisation sociale de la mobilité sont autant de facteurs qui participent à cette tendance en l'amplifiant et nous amènent à questionner la durabilité des orien-

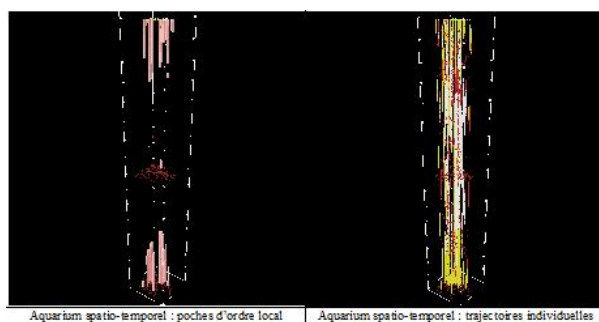
tations en cours. Le droit à une mobilité choisie, en relation avec l'aspiration à des modes de vie plus individualisés et autonomes, est susceptible de réduire considérablement un droit tout aussi fondamental, celui de l'accès à la ville pour tous. Si " trop de mobilité tue la ville ", l'enjeu d'un développement urbain harmonieux et durable est alors d'identifier des conditions de mobilité quotidienne permettant à la ville de continuer à jouer son rôle de mise en relation, tout en conservant la maîtrise de la croissance d'externalités négatives qui se posent aujourd'hui avec une acuité particulière : étalement urbain, congestion, consommation et production énergétique, risques et nuisances ... En replaçant les mobilités urbaines quotidiennes au cœur du débat sur la ville durable, ce projet vise à questionner un certain nombre de pré-supposés actuels, en les ré-interprétant du point de vue des systèmes complexes. En premier lieu, l'accessibilité spatio-temporelle réelle des citoyens aux territoires urbains ne peut être obtenue simplement en travaillant sur l'accessibilité des lieux, et implique de reconstruire les programmes d'activités et les trajectoires spatio-temporelles des individus. Des données fiables existent (enquêtes ménages-déplacements), qu'il s'agit d'exploiter et de valoriser dans une perspective renouvelée. Toutefois, compte tenu des interactions en jeu et de la forte dépendance des comportements individuels aux situations localement rencontrées, une approche fondée sur la simulation individus-centrée semble constituer une alternative privilégiée. Une telle approche permettrait en effet d'explorer aussi bien l'influence des comportements individuels sur le fonctionnement global de la ville que les possibles modifications comportementales induites par différentes politiques publiques.



Le prototype Multi-Agents MIRO

Le prototype MIRO, développé au cours du précédent projet financé par le PREDIT, constitue le socle sur lequel nous proposons de construire ce nouveau projet. Combinaison

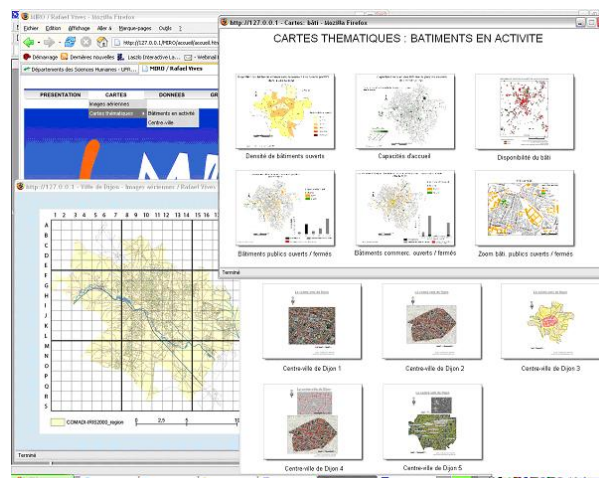
nant des outils formels de modélisation de type UML et Ploom-unity à l'utilisation de bibliothèques multi-agent de type Madkit, le prototype MIRO permet d'évaluer, par simulation, l'influence sur le fonctionnement de l'espace urbain de comportements individuels orientés notamment vers la construction et la réalisation de programmes quotidiens d'activités (fig. 11). Il se positionne ainsi à l'intersection de deux socles théoriques aujourd'hui considérés comme classiques, la Time Geography d'un côté et l'intelligence artificielle distribuée de l'autre.



Visualisation des mobilités urbaines grâce à l'aquarium spatio-temporel (en X et Y l'espace, en Z le temps (début de la journée en bas, fin de la journée en haut))

Les principaux apports du projet financé par le PREDIT peuvent être résumés de la manière suivante :

- Le premier point concerne la définition d'une méthodologie de construction d'un environnement virtuel, peuplé d'agents géographiques dotés de caractéristiques socio-démographiques et de programmes d'activités ;
- Au cours d'une simulation, les agents cherchent à réaliser leur programme sur la journée en tenant compte des heures d'ouverture des bâtiments (travail, loisirs) et du trafic ambiant. A la fin de chaque journée, ils modifient si besoin est leur programme pour le lendemain en fonction de l'expérience de la journée ;
- Des outils et interfaces de visualisation spécifiques ont enfin été créés pour analyser les résultats des simulations :
 - Actogrammes permettant d'analyser et de comparer les programmes d'activités affectés aux agents et modifiés par eux ;
 - Visualisation des trajectoires observées dans l'espace et le temps (aquarium spatio-temporel (fig.)) ;
 - Représentations cartographiques multi-niveaux (fig.), des trajectoires individuelles à la ville en mouvement.



Une plateforme de visualisation intégrée

Enjeux méthodologiques du projet

Plusieurs enjeux ont été plus particulièrement identifiés : 1) intégrer les mécanismes complexes d'arbitrage et de choix mobilisés lors du processus de prise de décision, dans un environnement riche, complexe et mouvant porteur d'opportunités mais également de contraintes, ce qui implique notamment de formaliser et intégrer la notion d'information imparfaite dans les processus de choix ; 2) créer des environnements urbains virtuels aussi réalistes que possible, alimentés par des données multi-sources gérées par système d'information géographique et détaillant autant que faire se peut l'offre de service (publique comme privée) et de transport (multimodale) urbain, dans ses composantes spatiales et temporelles ; 3) créer des populations d'agents synthétiques aussi représentatives que possible des populations étudiées, en combinant des données d'enquêtes socio-démographiques et de mobilité ; 4) construire et tester des indicateurs d'accessibilité spatio-temporelle désagrégés permettant de caractériser les situations localement vécues par les acteurs et de préciser, sur cette base, les véritables pertes et gains d'accessibilité engendrés par des projets d'aménagement, ainsi que leurs retombées sur les programmes d'activités des individus ; 5) construire et mettre en oeuvre des protocoles de vérification et de validation permettant, à toutes les étapes de la démarche (conceptualisation, formalisation, modélisation, implémentation, simulation) de déterminer la confiance que l'on peut accorder au modèle construit et aux résultats de simulation produits ; 6) dépasser le seuil des 80 000 agents rencontré lors du projet PREDIT, de manière à se rapprocher de conditions expérimentales plus réalistes ; 7) créer une plateforme de simulation collaborative permettant de réunir différents acteurs autour d'un modèle et de les faire participer au processus de simulation, en leur don-

nant la capacité d'interagir avec certains paramètres du modèle, dans le cadre de scénarios pré-définis. Il s'agit ici de tester la capacité des systèmes multi-agents à devenir des systèmes multi-acteurs, aptes à révéler et mettre en perspective les stratégies individuelles et leurs impacts globaux ; 8) adapter et appliquer les modèles développés à deux agglomérations françaises significativement différenciées (Dijon et Grenoble), dans une perspective comparative.

Références

BANOS Arnaud, BOFFET-MAS A., CHARDONNEL Sonia, LANG Christophe, MARILLEAU Nicolas et THEVENIN Thomas, 2005 : Modéliser les programmes d'activités par les systèmes multi-agents, Actes du 16e colloque de l'ASRDFL, Dijon, septembre 2005, 12 p.

BANOS Arnaud, CHARDONNEL Sonia, LANG Christophe, MARILLEAU Nicolas et THEVENIN Thomas, 2005 : Simulating the swarming city : a MAS approach, Proceedings of the 9th International Conference on Computers in Urban Planning and Urban Management, London, June 29-30, 15p.

MARILLEAU, C. LANG, P. CHATONNAY, AND L. PHILIPPE (2005) : A Meta-Model of Group for Urban Mobility Modeling. In procs. of The Third Int. Conf. on Active Media Technology, AMT 2005, Takamatsu, Japon, pages 397-400, May 2005

MARILLEAU, C. LANG, P. CHATONNAY, AND L. PHILIPPE (2005) : Cognitive Perception in RAFALES-SP Methodology. In Int. Conf. on Computational Intelligence for Modelling Control and Automation, CIMCA'05, Vienna, Austria, pages 594-600, November 2005.

MARILLEAU, C. LANG, P. CHATONNAY, AND L. PHILIPPE (2006) : An Agent-Based Framework for Urban Mobility Simulation. In Procs of the 14th IEEE Euromicro Conference on Parallel, Distributed and Network based Processing (PDP 2006), Montbéliard, France, pages 355-361, February 2006

LORIA - MAIA

MAchine Intelligente Autonome

Coordonnées

- Directeur : François Charpillet
- Adresse : LORIA - bâtiment C Campus scientifique B.P. 239
54506 Vandoeuvre-lès-Nancy

- Web : www.loria.fr/equipes/maia
- Tél : +33 3 83 59 20 81

Membres de l'équipe

- François Charpillet Chercheur INRIA Directeur de Recherche, HDR
- Olivier Buffet Chercheur INRIA (CR)
- Alain Dutech Chercheur INRIA (CR)
- Nazim Fatès Chercheur INRIA (CR)
- Bruno Scherrer Chercheur INRIA (CR)
- Christine Bourjot MCF, U. Nancy 2
- Vincent Chevrier MCF HDR. U. Henri Poincaré (Nancy 1)
- Alexis Scheuer MCF U. Henri Poincaré (Nancy 1)
- Olivier Simonin MCF U. Henri Poincaré (Nancy 1)
- Vincent Thomas MCF, U. Nancy 2
- Maan El-Badaoui-El-Najjar MCF associé, Univ. Lille.

Thème général de l'équipe

Notre groupe s'intéresse à la modélisation, la construction et la simulation de systèmes composés d'un ou plusieurs agents. Un agent peut prendre différentes formes : physique (comme un robot mobile, ou un robot d'anesthésie), et ou logiciel (softbot, par exemple un assistant intelligent) ou biologique (humains, insectes sociaux). Dans ce cadre, nos recherches s'articulent autour des thèmes suivants :

- Apprentissage par renforcement, Planification (décision dans l'incertain),
 - Intelligence collective (modèles d'auto-organisation).
- L'objectif étant de modéliser et interpréter des comportements collectifs ou individuels et de produire des comportements permettant à des agents artificiels de réaliser des tâches complexes.

Description des travaux

Véhicule intelligent

- Problématique du Platooning : convoi automatique par accrochage immatériel de proche-en-proche (voir projets CRISTAL et TACOS).
- Localisation et Vision pour véhicules intelligents (expérimentations avec le véhicule Cycab au LORIA).

Robotique mobile et swarm intelligence

- Navigation multi-robots (l'équipe dispose d'une plate-forme de six robots mobiles Khepera III) : pla-

nification, contrôle de trajectoires, déplacements en formation, re-configuration automatique, etc.

- Supervision et partage d'autorité d'essaims de drones (projet SMAART)

Organisation et systèmes de transports

- Contrôle aérien (projet PAUSA)
- Simulation de réseaux de transport.

Projets

Projet "CRISTAL"

- Pôle de compétitivité "Véhicule du Futur" Alsace - Franche-Comté
- "Véhicule intra-urbain à usage public"
- Web : <http://projet-cristal.net/>
- 2007-2009

Projet ANR "TACOS" (ANR Sécurité et Informatique)

- "Trustworthy Assembling of Components : from requirements to Specification"
- Web : <http://tacos.loria.fr>
- 2007-2008

Projet PEA DGA "SMAART"

- "Etude des Systèmes multi-agents adaptés à la reconnaissance de théâtre et l'auto-organisation des drones"
- Web : www.loria.fr/~simoniol/OSimonin/smaart/smaart.html
- 2006-2008

Projet "PAUSA"

- "Partage d'Autorité dans le Système Aéronautique"
- Web : <http://www.eurisco.org/pausa/>
- Le projet PAUSA (Partage d'Autorité dans le Système Aéronautique), vise à étudier le partage et la distribution d'autorité dans le système aérien, d'un point de vue Facteurs Humains. Partage et distribution sont envisagés à la fois entre le sol et le bord et entre les humains et les machines.

Projet "Modélisation multi-agent de réseaux de transport"

- Contact : Vincent.Chevrier@loria.fr
- La CUGN (communauté urbaine du grand nancy) dispose de données décrivant les différents événements

se produisant dans une journée sur le réseau de transport public. Nous avons développé un premier outil qui permet de "rejouer" en temps différé une journée d'exploitation modulo quelques restrictions (une ligne et un sous ensemble des événements).

Références

– C. Smaili, M. El Najjar and F. Charpillet, "A road matching method for precise vehicle localization using hybrid Bayesian network" Special Issue of the International Journal in Intelligent Transportation Systems edited by Francis and Taylor October 2008. (à paraître)

– C. Capelle, M. El Najjar, D. Pomorski and F. Charpillet "Intelligent geo-localisation in urban areas using GPS, 3D-GIS and vision" Special Issue of the International Journal in Intelligent Transportation Systems edited by Francis and Taylor October 2008. (à paraître)

– A. Glad, O. Simonin, O. Buffet and F. Charpillet "Theoretical Study of Ant-based Algorithms for Multi-Agent Patrolling" Proceedings of the Eighteenth European Conference on Artificial Intelligence (ECAI-08).

– O. Simonin, O. Grunder "A cooperative multi-robot architecture for moving a paralyzed robot" International Journal MECHATRONICS The Science of Intelligent Machines (IFAC) Elsevier, accepted for publication in 2009.

– F. Legras, A. Glad, O. Simonin and F. Charpillet "Authority Sharing in a Swarm of UAVs : Simulation and Experiments with Operators" SIMPAR'08 International Conference on Simulation, Modeling and Programming for Autonomous Robots to appear in proceedings SIMPAR 2008 LNCS Springer.

– E. Gechter, V. Chevrier and F. Charpillet "A Reactive Agent-Based Problem-Solving Model : Application to localization and Tracking", Transactions on Autonomous and Adaptive Systems, vol1, N° 2, 2007.

– O. Simonin, A. Lanoix, S. Colin, A. Scheuer and F. Charpillet. "Generic expression in B of the Influence/Reaction model : specifying and verifying situated multi-agent systems". RR INRIA 6304.

SET - ICAP

Informatique : Communication, Agents et Perception

- CONTACT : A. Koukam (Pr) - abder.koukam@utbm.fr,
- ADRESSE : Équipe ICAP du Laboratoire Systèmes et Transports (SET - EA 3317), Université de Technologie de Belfort-Montbéliard,
- WEB : set.utbm.fr,

– TEL : 03 84 58 33 19.

Membres de l'équipe

Les membres de l'équipe ICAP concernés par l'activité SMA et transports sont : Abderrafiâa Koukam, Yassine Ruichek, Jean-Charles Créput, Stéphane Galland, Franck Gechter, Pablo Gruer, Amir Hajjam El-Hassani, Vincent Hilaire et Fabrice Lauri (enseignants chercheurs) ; Achraf Ben Miled, Jean-Michel Contet, Nicolas Gaud, Pawel Kmiotek, David Meignan, Davy Monticcolo, Sergio Nogueira et Elie Tagnefute (doctorants).

Thème générale de l'équipe

La thématique commune de l'équipe ICAP s'articule autour des systèmes composés d'entités autonomes et parfois mobiles, dotées de capacités de communication et/ou de perception. Les entités autonomes peuvent être des agents logiciels ou physiques. La thématique initiale, qui portait sur les systèmes multi-agents, s'est étendue dans une deuxième phase au domaine de la communication mobile, comme résultat d'une collaboration avec le CNET de Belfort (actuellement division d'Orange Recherche). Finalement, une action portant sur la perception de l'environnement a été initiée, dans le cadre des activités relatives aux problématiques du domaine des véhicules intelligents.

Description des travaux

Les activités du groupe Systèmes Multi Agents de l'équipe ICAP se déclinent, dans le domaine des systèmes de transport, aussi bien au niveau des composants tels que les véhicules et les équipements, qu'au niveau des systèmes considérés globalement. Dans le premier cas le groupe s'intéresse à l'intelligence des composants, essentiellement des véhicules. Cette intelligence est conçue dans ses aspects liés à l'autonomie, la navigation, la perception et la coopération entre véhicules. Dans le deuxième cas, le groupe s'intéresse aux approches de modélisation orientée agents pour la simulation et l'évaluation, ainsi qu'aux approches collaboratives pour l'optimisation dans les systèmes de transport.

Véhicule intelligent et platooning

Il s'agit d'une activité de recherche qui vise le développement des technologies nécessaires à la mise en place de services innovants de transport urbain de passagers. Ces services reposent sur l'utilisation de véhicules intelligents capables de coopérer en s'auto-organisant dans des formations (platoons) à configuration linéaire (train), avec un

accrochage immatériel entre unités. Ces véhicules devront posséder toutes les capacités nécessaires à un fonctionnement autonome. Ceci inclut le suivi d'itinéraire par géopositionnement, la détection d'obstacles, l'interprétation de signalisations, l'accrochage/décrochage automatiques. L'approche explorée repose sur des systèmes d'agents réactifs dont les interactions sont guidées par des modèles d'inspiration physique [Contet et al. 2007]. Une attention particulière est accordée à la spécification formelle et la vérification par model-checking des propriétés de sûreté de fonctionnement caractéristiques de l'application visée (transporter des passagers). La recherche sur les capacités de perception des agents-véhicules s'articulent autour de concepts tels que la fusion multi-capteurs, la vision artificielle stéréoscopique pour la détection d'obstacles [Nogueira et al. 2007] et la navigation par détection d'indices visuels. Cette activité a conduit à l'obtention de trois distinctions pour jeunes chercheurs : un premier prix en 2007 et un premier prix en 2008 au concours A'Doc, organisé par l'Ecole Doctorale SPIM de Franche-Comté et une médaille d'argent au concours YEAR'2008, de la Commission Européenne. Une partie de cette action est engagée dans le projet CRISTAL, du pôle de compétitivité Alsace-Franche Comté Véhicule du Futur, en collaboration avec l'équipe MAIA du LORIA Nancy (responsable : F. Charpillat).

Approches heuristiques et méta heuristiques pour l'optimisation

Les réseaux de transport posent de nombreux problèmes (dimensionnement, positionnement, allocation, ordonnancement, routage). Les outils d'optimisation, permettant de réduire les coûts d'aménagement, d'augmenter la qualité des services ou d'en réduire l'impact environnemental, sont devenus essentiels. Les approches heuristiques et méta heuristiques sont adaptées à la complexité des réseaux de transports et des problèmes d'optimisation associés. Dans ce cadre, nous étudions et développons des méthodes d'optimisation orientées-agents. Nous abordons la résolution de problèmes de positionnement mono et multi-niveaux par des heuristiques multi-agents, et des problèmes de tournées de véhicules par des méta-heuristiques multi-agents.

Le positionnement est un problème complexe qui consiste à optimiser le placement de ressources par rapport à un ensemble de demandes. Pour ce problème, nous avons proposé une heuristique d'inspiration physique à base de champs de potentiels [Moujahed et al. 2007]. Dans cette approche, chaque ressource est un agent subissant un ensemble de forces d'attraction et de répulsion. Ces forces amènent les agents à modifier leur position et entraînent globalement le système vers une configuration plus

stable, conformément aux principes d'auto-organisation et d'émergence. Les avantages de cette heuristique sont sa robustesse, c'est à dire sa tolérance aux variations des paramètres du problème traité, sa flexibilité et sa capacité à traiter plusieurs variantes du problème. En plus des résultats très prometteurs évalués sur des benchmarks, cette approche a été appliquée à un problème de positionnement d'arrêts de bus ainsi qu'à un problème de patrouille.

Dans le cadre de l'optimisation de tournées de véhicules, nous avons développé une méta heuristique fondée sur le principe de coalition d'agents [Meignan et al. 2008], [Créput et al. 2007]. Plusieurs agents organisés en coalition traitent simultanément un problème d'optimisation. Chaque agent dispose d'un ensemble d'opérateurs permettant d'explorer l'espace des solutions du problème. Ils améliorent individuellement leur stratégie de recherche par apprentissage par renforcement et coopèrent par mimétisme et échange de solutions. A travers cette étude, nous illustrons les apports de l'approche multi-agent dans le cadre des méta heuristiques qui reposent, notamment, sur la distribution et l'intégration des techniques d'apprentissage.

A partir de cette problématique, des travaux récents ont abordé le problème de la patrouille, par une approche multi-agents. Une approche en deux étapes a été proposée. La première consiste à positionner les agents suivant une couverture optimale des noeuds et fait appel à un algorithme évolutionniste. La deuxième étape détermine, pour chaque agent, l'ensemble de noeuds à visiter et fait appel à un algorithme de type colonie de fourmis [Lauri et al. 2008].

Approches multi-agents pour la simulation de réseaux de transports

Les réseaux de transport en commun sont un élément clé de la mobilité. Face à la demande grandissante, les autorités en charge de leur mise en place ont besoin d'outils d'aide à la décision permettant d'observer et d'évaluer un réseau réel ou virtuel. Les expérimentations directes étant souvent impossibles, il devient pertinent de développer des outils de simulation pour l'évaluation.

L'approche multi-agent est adaptée à la modélisation de tels systèmes, composés d'un grand nombre d'entités liées par de multiples interactions. Dans ce cadre, nous nous intéressons à la simulation multi-agents de réseaux de transport en commun. Ces travaux ont donné lieu au projet SeT-GIS, un logiciel spécialisé dans l'aménagement de réseaux de transport urbains, permettant de simuler le mouvement des bus au sein des flux routiers. Cette application est actuellement utilisée par le Syndicat Mixte des Transports en Commun du Territoire de Belfort et a motivé le dépôt en

2006 d'une Enveloppe Solo à l'Institut National de la Protection Industrielle.

L'équipe explore également les approches de modélisation et simulation multi-niveaux fondées sur les systèmes multi-agents holoniques. Ces travaux ont été appliqués à la modélisation d'un site industriel, pour évaluer des indicateurs relatifs aux flux de matières et de personnels [Rodriguez et al. 2007]. De même a été abordée la simulation multi-niveaux holonique de piétons dans un environnement urbain [Gaud et al. 2007].

Références

- [Contet et al. 2007] J.-M. Contet, F. Gechter, P. Gruer, A. Koukam. "Application of reactive multiagent systems to linear vehicle platoons". In proc. IEEE International Conference on Tools with Artificial Intelligence - ICTAI 2007, Patras, Grece, October 2007.
- [Créput et al. 2007] J.C. Créput, A. Koukam, A. Hajjam, "Self-Organizing Maps in Evolutionary Approach for the Vehicle Routing Problem with Time Windows" International Journal of Computer Science and Network Security, vol. 7, no. 1, January 2007.
- [Gaud et al. 2007] N. Gaud, F. Gechter, S. Galland and A. Koukam, "Holonomic multiagent multilevel simulation : Application to real-time pedestrians simulation in urban environment", In Proc. of International Joint Conference on Artificial Intelligence (IJCAI'07), pp.1275-1280, 2007.
- [Lauri et al. 2008] F. Lauri, A. Koukam, "A two step evolutionary and ACO approach for solving the multi-agent patrolling problem". In proc. of IEEE World Congress on Computational Intelligence, Hong-Kong, China, 2008.
- [Moujahed et al. 2007] S. Moujahed, N. Gaud and D. Meignan, "A Self-Organizing and Holonic Model for Optimization in Multi-Level Location Problems", In Proc. of IEEE International Conference on Industrial Informatics, INDIN 2007, Vol. 2, pp 1053-1058, 2007.
- [Meignan et al. 2008] D. Meignan, J.-C. Créput and A. Koukam, "A Coalition-based Metaheuristic for the Vehicle Routing Problem", In Proc. of IEEE World Congress on Evolutionary Computation, 2008.
- [Nogueira et al. 2007] S. Nogueira, Y. Ruichek, F. Gechter, A. Koukam, F. Charpillat. "An artificial vision based environment perception system", Advances for In-Vehicle and Mobile Systems, Springer Ed., 2007.
- [Rodriguez et al. 2007] S. Rodriguez, V. Hilaire, and A. Koukam, "Towards a holonic multiple aspect analysis and modeling approach for complex systems : Application to the simulation of industrial plants". Journal Simulation Modelling Practice and Theory, 15(5) :521-543, May 2007.

INRETS-MSIS

Modélisations, Simulations et Simulateurs de conduite

- CONTACT : S. Espié (DR) - espie@inrets.fr,
- ADRESSE : Département INRETS-MSIS de L'Institut National de Recherche sur les Transports et leur Sécurité, 58 Bd Lefebvre, F75015 Paris,
- WEB : www.inrets.fr/ur/msis/simus/archisim.htm,
- TEL : +33 (0)1 40 43 65 54

Membres de l'équipe

Permanents : Espié S., Auberlet JM.

Thème général de l'équipe

Notre thématique générale concerne la Simulation comportementale du trafic routier.

Il s'agit de modéliser les comportements des différents acteurs du système de trafic (usagers divers mais aussi opérateurs) à partir, notamment, de travaux en psychologie de la conduite, et d'obtenir en simulation, par émergence, un comportement collectif reproduisant celui observé dans la réalité. Les travaux initiaux de l'équipe remontent au tout début des années 90. Ces travaux ont été conduits conjointement par F. Saad (chercheur en psychologie, INRETS) et par S. Espié (chercheur en informatique, INRETS).

L'architecture informatique développée permet de simuler le comportement individuel d'agents divers représentant soit des individus (automobilistes, piétons, motocyclistes), soit des dispositifs interagissant avec ces derniers (contrôleur de carrefours à feux, systèmes télématiques, ...), dispositifs opérés par les gestionnaires des infrastructures routières. Une particularité des travaux conduits par MSIS est la capacité de l'architecture à accueillir des simulateurs de conduite qui permettent de faire interagir des sujets humains avec les agents simulés. Le modèle de trafic ainsi constitué a été validé pour des situations autoroutières diverses, et pour quelques situations urbaines simples, les travaux de recherche portant sur la simulation des comportements des conducteurs au niveau des carrefours étant toujours en cours. Les travaux portant sur la simulation du comportement spécifique des conducteurs de deux roues motorisés ont permis de produire des comportements "réalistes". Une première validation a été réalisée par comparaison des temps de parcours auto/moto pour un trajet domicile travail en zone urbaine dense (grande banlieue parisienne - centre de Paris).

Description des travaux

Simulation du trafic autoroutier

Thèses achevées :

- S. El Hadouaj - encadrement INRETS S. Espié, LIP6 A. Drogoul

Le modèle "économique" développé dans le cadre de cette thèse s'appuie sur les travaux réalisés en psychologie de la conduite par F. Saad, notamment pour les situations autoroutières et les situations d'approche sur bretelles d'autoroute. Il permet de reproduire des comportements cohérents tant du point de vue des individus que du point de vue du trafic.

Simulation du comportement des automobilistes dans les carrefours

Thèses achevées :

- A. Champion - encadrement INRETS S. Espié, LAMIH R. Mandiau, C. Kolski (section) Le modèle de comportement développé dans cette thèse s'appuie sur la théorie des jeux pour définir, pour chaque conducteur simulé dans un environnement de carrefour (en approche ou dans le carrefour), la solution optimale du point de vue individuel entre avancer ou s'arrêter.
- A. Doniec - encadrement INRETS S. Espié, LAMIH R. Mandiau, S. Piechowiak (section) Les phénomènes d'anticipation, au centre des décisions de tout conducteur, ont été pris en compte dans cette thèse par le biais de l'utilisation de réseaux de contraintes et par un raffinement de l'identification des agents en interaction dans le carrefour.

Simulation du comportement des piétons

Thèses en cours :

- V. Airault - encadrement INRETS S. Espié, CRIP5 N. Cot L'hypothèse, au coeur de cette thèse, est que les motivations qui sous-tendent le comportement des piétons lors de leurs déplacements (hors "baguenaudage") est sensiblement le même que celui qui meut les conducteurs, et que c'est principalement la représentation qu'ils se font de l'environnement dans lequel ils se trouvent (environnement très peu structuré) qui explique les phénomènes observés.

Post-doc 2007 :

- A. Tom - encadrement INRETS JM. Auberlet Les interactions entre les piétons en situations de traversée en carrefour et le trafic automobile ont peu été modélisées dans le cadre de simulation du trafic. Ce post-doc a consisté à réaliser un travail de recherche bi-

bibliographique afin 1) de déterminer les manques les plus importants dans les modèles actuels de simulation de piétons, et 2) de déterminer les modèles psychologiques valides et " implémentables " pour modéliser le comportement des piétons aux carrefours.

Exemple de projets passés :

- DIATS, 4ème PCRD, simulation trafic et ITS en autoroutier (S. Espié)
- STARDUST, 5ème PCRD, simulation trafic et ITS en urbain (S. Espié, JM. Auberlet)

Simulation du comportement des deux roues motorisés

Thèses en cours :

- L. Bonte - encadrement INRETS S. Espié, LIFL P. Mathieu (section)
- L'hypothèse sous-jacente à cette thèse est globalement identique à celle émise pour le comportement des piétons. On considère ici que la notion de voie de circulation " utilisable " est variable et qu'elle dépend du trafic automobile. Cette notion de voies " virtuelles " permet de rendre compte des comportements de remontée de file que l'on peut observer dans la réalité.

Projets

Exemple de projets en cours :

- SICAP, Fondation Sécurité routière, comportement des piétons en traversée de rue (JM. Auberlet)
- PRE-DRIVE C2X, 7ème PCRD, simulation trafic et ITS (S. Espié)
- TRAIN-ALL, 6ème PCRD, usage du modèle de comportement conducteur pour la conception d'un tuteur intelligent pour la formation à la conduite (S. Espié)

Bibliographie

Airault V., Espié S. (2005) - Behavioural model of the pedestrian interaction with road traffic. In European Transport Conference (ETC) proceedings, October 3-5, 2005, Strasbourg, France. ISBN 978-1-905701-00-1.

Bonte L., Espié S., Mathieu P. (2007) - Virtual lanes interest for motorcycles - European Workshop on Multi-Agent Systems (EUMAS 07), December 13-14, 2007, Hammamet, Tunisia.

Doniec A., Mandiau R., Piechowiak S., Espié S. (2008). Anticipation based on constraint processing in a multi-agent context. Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS) (accepté le 07/04/08, sous presse).

Espie S., Auberlet J.M. (2007). ARCHISIM : a behavioural multi-actors traffic simulation model for the study of a traffic system including ITS aspects. International Journal of ITS Research. Volume 5 n° 1/2007. pp 1-6. Éditions ITS Japan.

Mandiau R., Champion A., Auberlet J.-M., Espié S., Kolski C. (2008). Behaviour based on decision matrices for a coordination between agents in urban traffic simulation. Applied Intelligence. Volume 28 n° 2/2008. pp 121-138. Éditions Springer.

Dossier : I.A. et Bioinformatique

Coordonneurs :

François Coste (IRISA, coste@irisa.fr) et Christine Froidevaux (LRI, chris@lri.fr)

La bioinformatique a connu ces dernières années un essor considérable suite à l'explosion des technologies (e.g. puces à ADN ou protéomiques) permettant d'obtenir à haut débit une masse de données en génomique considérable, nécessitant des méthodes automatiques de traitement. L'outil informatique est alors devenu indispensable pour le stockage et la manipulation de ces données. Après cette phase d'acquisition des données, le nouveau défi de la post-génomique consiste à extraire de la connaissance à partir de cette masse de données. Face au changement d'échelle du nombre de données disponibles, l'informatique a un rôle à jouer pour assister le biologiste dans sa quête de la compréhension du vivant, l'élaboration de nouveaux concepts et la formulation de nouvelles hypothèses. Dans ce contexte, l'intelligence artificielle peut aider le biologiste à modéliser des systèmes complexes composés d'objets de types différents en fortes interactions.

En réponse à l'explosion des résultats de la recherche en bioinformatique, la communauté de chercheurs a rapidement augmenté en taille et s'est progressivement structurée. Au niveau national, les " Journées Ouvertes de Biologie Informatique Mathématiques " (JOBIM), créées en 2000, réunissent chaque année de 350 à plus de 500 participants francophones. De même que la communauté Intelligence Artificielle est rassemblée dans l'AFIA, la communauté Bioinformatique a ressenti tout récemment le besoin de se regrouper au sein d'une société savante. C'est ainsi que la Société Française de Bioinformatique (SFBI, <http://sfbi.impg.prn.fr/>) a été fondée en 2005. Enfin, au cours des deux dernières décennies, la communauté Bioinformatique a été successivement soutenue financièrement par diverses structures. La dernière en date est le GDR CNRS 3003 Bioinformatique Moléculaire (<http://www.gdr-bim.u-psud.fr/>), créé en 2006, qui a pour principale mission de favoriser les rencontres et les synergies entre les différentes disciplines biologie, informatique, mathématiques et physique, afin de susciter la conception de projets de recherche véritablement pluridisciplinaires tant à l'échelle nationale qu'euro-péenne.

La bioinformatique moléculaire profite des avancées de divers champs de l'informatique, tels que l'algorithmique et la combinatoire des séquences, des arbres (phylogénie,

structures d'ARN), des graphes (réseaux biologiques, structures moléculaires, réarrangements chromosomiques ...), la programmation logique et les réseaux de contraintes et les diverses techniques d'optimisation. L'informatique apporte aussi ses techniques en analyse d'images (puces pour le transcriptome ou le protéome), en visualisation et manipulation de données massives et complexes (molécules, réseaux biologiques), ses outils pour leur traitement efficace sur des grandes grilles de calcul, ainsi que ses méthodes de l'Intelligence Artificielle.

L'intelligence artificielle peut être utilisée avantageusement dans différents domaines clés de la bioinformatique, par exemple : la structuration et l'évolution des génomes, la prédiction de la topologie de protéines et de leurs sites actifs, la compréhension du contrôle du niveau d'expression, la modélisation des réseaux d'interactions et de leur niveau d'organisation, la découverte de voies thérapeutiques, etc. Toutes les composantes de l'intelligence artificielle sont concernées. Le choix judicieux d'une bonne représentation de connaissances est crucial pour une modélisation pertinente. Le recours à des méthodes d'apprentissage permet de découvrir des régularités dans ces masses de données (séquences et structures moléculaires, pour les réseaux biologiques, pour les données d'expression) et de proposer des modèles à partir d'exemples pouvant conduire à l'identification de nouveaux éléments satisfaisant le modèle. Enfin, l'interprétation des phénomènes biologiques passe par l'intégration de données très hétérogènes pouvant bénéficier de l'utilisation d'ontologies pour leur mise en correspondance.

Les contributions présentées dans ce dossier résultent d'un appel à participation diffusé par le biais de mailing lists, à la fois dans la communauté Intelligence Artificielle et dans la communauté Bioinformatique. Nous avons sollicité la présentation des équipes de recherche de la communauté francophone impliquées dans des actions de recherche et de développement en IA et bioinformatique. Nous tenons à remercier les quinze équipes qui nous ont envoyé un descriptif de leurs activités. Sans avoir la prétention d'être exhaustif, le panorama offert par ces présentations apparaît représentatif de l'ensemble des activités à l'intersection de l'IA et de la bioinformatique. La couverture

thématique de ce dossier est très large, allant de la biologie, du biomédical ou des statistiques à l'informatique, et implique à chaque fois une ou plusieurs facettes de l'intelligence artificielle : ontologies, web sémantique, raisonnement dans des bases de connaissances, apprentissage automatique, fouille de textes et extraction de connaissances à partir de textes... Cette couverture thématique sera encore appelée à s'élargir au fur et à mesure de l'évolution de ce domaine encore jeune, peut-être même à la suite de votre lecture que nous vous souhaitons très bonne...

François Coste et Christine Froidevaux.

Liste des équipes :

Centre de Bioinformatique Ecole des Mines de Paris
CRIL-Lens EA 3781-Marseille EA 3888 Laboratoire d'Informatique Médicale-Pontchaillou Equipe ABC LORIA-Nancy
Equipe Bioinformatique LRI-Orsay Equipe Fouille de données et Nutrigénomique (Paris 6 et Paris13) Equipe IACG
BIM-Grenoble Equipe PLIAGE de l'IMAG et équipe LCM de l'IBS-Grenoble
Equipe SIAGBM (MIA) INRA-Toulouse Laboratoire Biologie Virtuelle-Nice
Projet Symbiose IRISA-Rennes Thème DM et BDI du LIRIS-Lyon U729 Inserm
SPIM-Paris Unité MIG INRA-Jouy

Centre de Bioinformatique, Ecole des Mines de Paris

- Web : <http://cbio.ensmp.fr>
- Contact : Jean-Philippe Vert, directeur (courriel : Jean-Philippe.Vert@ensmp.fr ;
- Tel : 01 64 69 47 82)

Membres de l'équipe concernés par le projet

Véronique Stoven, Christian Lajaunie, Pierre Mahé, Martial Hue, Franck Rapaport, Joannès Vermorel, Laurent Jacob, Mikhail Zaslavskiy, Jean-Philippe Vert

Thème général de l'équipe

Apprentissage statistique et applications en bioinformatique

Travaux en cours en lien avec le thème du dossier :

Le fil directeur de nos projet de recherche est notre volonté de développer des méthodes mathématiques et informatique innovantes, notamment en apprentissage statistique, pour répondre aux défis nouveaux de la biologie et

de la médecine post-génomique. Cette stratégie se décline à travers différents projets de recherche :

- Analyse de données biologiques hétérogènes : nous développons des méthodes innovantes pour l'analyse et la classification de données biologiques structurées et hétérogènes (notamment les séquences génomiques et structures protéiques), dans le cadre des méthodes d'apprentissage dites à noyau. Ces méthodes sont appliquées notamment à l'annotation fonctionnelle des gènes (collaborations avec UC Berkeley, l'Université de Washington, et l'Université de Kyoto) et à la modélisation de siRNA (collaboration avec le CEA).
- Biologie des systèmes : nos efforts portent sur la modélisation et la reconstruction de réseaux génétiques par des méthodes d'apprentissage. Nous travaillons notamment sur la reconstruction de réseaux métaboliques chez la levure et les bactéries (avec l'Université de Kyoto), la modélisation de progression tumorale pour plusieurs cancers (avec l'Institut Curie), et la modélisation du parasite *P. falciparum* responsable du paludisme (avec l'Institut Pasteur).
- Criblage virtuel et recherche de médicament : nous étudions de nouvelles méthodes pour le criblage virtuel par apprentissage statistique et intégrations de la modélisation de la cible thérapeutique et des ligands potentiels.
- Théorie et algorithmes d'apprentissage : nous maintenons une activité de recherche méthodologique forte en apprentissage statistique, notamment sur les méthodes d'apprentissage à noyau (y compris les machines à vecteur de support) et les modèles graphiques.

Bibliographie

(bibliographie complète disponible sur <http://cbio.ensmp.fr>)

R. Vert and J.-P. Vert, Consistency and convergence rates of one-class SVM and related algorithms, *Journal of Machine Learning Research*, 7 :817-854, 2006.

H. Saigo, J.-P. Vert and T. Akutsu, Optimizing amino acid substitution matrices with a local alignment kernel, *BMC Bioinformatics* 7 :246, 2006.

Y. Yamanishi, J.-P. Vert and M. Kanehisa, Supervised enzyme network inference from the integration of genomic data and chemical information, *Bioinformatics*, vol. 21, p. i468-i477, 2005.

M. Cuturi, K. Fukumizu and J.-P. Vert, Semigroup kernels on measures, *Journal of Machine Learning Research*, vol. 6, p. 1169-1198, 2005.

P. Mahé, N. Ueda, T. Akutsu, J.-L. Perret and J.-P. Vert, Graph kernels for molecular structure-activity relationship analysis with support vector machines, J. Chem. Inf. Model., vol. 45, n. 4, 939 -951, 2005.

Centre de Recherche en Informatique de Lens - CRIL CNRS FRE 2499

- Adresse : Faculté des sciences, rue Jean Souvraz SP 18, 62307 Lens cedex
- Web : <http://www.cril.univ-artois.fr>
- Directeur : Eric Gregoire (Tel : +33 3 21791785, courriel : gregoire@cril.univ-artois.fr)

Personnes participantes

Engelbert Mephu Nguifo (mephu@cril.univ-artois.fr),
Hugues Delalin

Thème général du laboratoire

Le thème fédérateur du laboratoire est l'intelligence artificielle symbolique et ses applications. Plus précisément, les activités du laboratoire sont structurées selon deux axes principaux : d'une part, le traitement des informations imparfaites, dynamiques, contextuelles et multi-sources, et d'autre part, l'algorithmique pour l'inférence et la prise de décision.

Description des travaux

Nous nous intéressons à deux problèmes en bioinformatique : la prédiction de motifs, et l'alignement de séquences.

Prédiction de motifs

Aujourd'hui, les avancées tant dans le domaine bioinformatique que dans le domaine biologique nécessitent souvent une annotation "in silico" des génomes. L'annotation des génomes est une étape primordiale de la compréhension des fonctions biologiques. C'est pourquoi depuis l'essor pris par le séquençage des génomes complets, l'élaboration d'algorithmes de prédiction d'éléments génomiques fiables et robustes représente une composante dynamique de la bioinformatique. Prédire "in silico" la position des gènes et des promoteurs est devenu de nos jours une analyse facilement réalisable grâce à de nombreux algorithmes utilisant des approches différentes. Certains permettent d'obtenir des résultats de qualité. Néanmoins les

usages pratiques de ces algorithmes montrent que le problème reste ouvert, notamment dans le cas de la prédiction des opérons, et des promoteurs.

Nous nous intéressons à la prédiction de structures opéroniques chez les procaryotes, par une approche logique de décomposition basée sur une combinaison de méthodes. De nombreux éléments génomiques comme les gènes, les régions codantes, les promoteurs ou même les terminateurs sont largement étudiés mais il existe peu d'études sur des structures plus complexes comme les opérons réunissant à la fois les promoteurs, les gènes, les terminateurs en un système organisé et ceci malgré l'intérêt que représentent les structures opéroniques pour l'étude des voies métaboliques et des réseaux des régulations. En effet, souvent les gènes d'un même opéron codent pour des protéines impliquées dans une même voie ou appartenant à une même cascade de réaction et sont fréquemment soumis à un système intégré de régulation. Les quelques études déjà effectuées sur le sujet requièrent un nombre important de données à partir desquelles de bons résultats "ab initio" peuvent être obtenus, mais qui, malheureusement, sont restreints à un organisme étudié. Nous avons débuté la mise en place d'une plate-forme de prédiction suivant la logique biologique. Pour cela l'utilisation d'algorithmes de prédiction de gènes a été combinée avec des algorithmes de recherche de motifs consensus de promoteurs d'une part et de terminateurs d'autre part. Ce travail sera appliqué à l'analyse du génome de la bactérie *Bordetella Pertussis*, dans le cadre d'une collaboration avec le laboratoire BioPuces de l'Institut Pasteur de Lille et la société GENOSCREEN.

Nous nous intéressons aussi à la prédiction de promoteurs par la mise en oeuvre d'une méthode de combinaison d'experts. La prédiction des éléments de régulation (promoteurs, terminateurs, ...) dans les génomes constitue un défi important de la bioinformatique moderne. En effet la performance des programmes actuels, qui se basent sur différentes approches, reste bien en deçà des espérances et leur précision ne dépasse guère les 50

Alignement de séquences

La plupart des algorithmes d'alignement de séquences sont basés sur l'utilisation des matrices de substitution. Cependant, compte tenu du nombre de matrices existantes, le choix de la plus appropriée aux données de l'utilisateur est difficile. En effet, bien que la qualité d'un alignement dépende des propriétés des acides aminés conservés au cours de l'évolution, on ne sait pas quelles sont les propriétés supportées par une matrice. C'est pourquoi nous proposons une méthode de représentation des acides aminés

par leurs propriétés permettant à l'utilisateur d'ajouter ses propres connaissances.

Nous avons développé une méthode basée sur un changement de représentation des acides aminés. Les acides aminés sont représentés par un vecteur de bits où chaque bit représente une propriété des acides aminés. Une fonction permet le calcul d'un niveau de similarité entre deux acides aminés. Actuellement nous effectuons une comparaison de la méthode en utilisant deux algorithmes d'alignement de séquences classiques, Smith et Waterman, et, Needleman et Wunsch, par rapport aux programmes CLUSTAL, FASTA et BLAST.

Collaborations

- Laboratoire BioPuces de l'Institut Pasteur de Lille, et Société GENOSCREEN. Co-encadrement d'une thèse en bioinformatique sur la prédiction de structures opéroniques.
- Service de BioInformatique du Centre de Biotechnologie de Sfax (Tunisie).

Bibliographie

Hubans C., Hot D., Lemoine Y., Mephu Nguifo E., 2004, Plate-forme de prédiction logique des structures opéroniques, in Poster Session, 12th Journées Ouvertes de Biologie, Informatique et Mathématiques (JOBIM), Juin, Montréal (Canada).

Delalin H., Mephu Nguifo E., 2004, Uses of bit-vector representation of amino-acids for sequence alignment, in Poster Session, 12th Intl. Conference on Intelligent Systems for Molecular Biology - 3rd European Conf. on Computational Biology (ISMB-ECCB), July, Glasgow (UK).

Lemoine E., Mephu Nguifo E., Merceron D., Sallantin J., 1999, Improving the efficiency of a User-Driven Learning System With Reconfigurable Hardware. Application to DNA Splicing, Proc. of Intl. Conf. on BioComputing (PSB), World Scientific Press, pp 290-301, Mauna Lani, Hawaï, January.

Vignal L., D'Aubenton-Carafa Y., Lisacek E., Mephu Nguifo E., Rouze P., Quinqueton J., Thermes C., 1997, Exon prediction in eucaryotic genomes, Biochimie, vol. 78, pp 327-334.

Mephu Nguifo E., Sallantin J., 1993, Prediction of Primate Splice Junction Gene Sequences, in Proc. of 1st Intl. Conf. on Intelligent Systems for Molecular Biology (ISMB), 292-300, Washington DC (USA), July 7-9th, AAAI Press.

EA 3781 - Evolution Biologique - Equipe Phylogénomique (dirigé par Pierre Pontarotti)

- Université de Provence - Centre St-Charles - 3 Place Victor Hugo - 13331 Marseille
- Web : <http://www.up.univ-mrs.fr/evol>
- Contact : Philippe Gouret (Tel : 04-91-10-64-89 ; courriel : philippe.gouret@up.univ-mrs.fr)

Membres de l'équipe concernés par le sujet

Philippe Gouret - Ingénieur ; Virginie Lopez - Doctorante

Thème générale de l'équipe

Etude de l'histoire évolutive des génomes

Description des travaux

L'activité de recherche du laboratoire concernant les problématiques liées à l'évolution des espèces et en particuliers de leurs génomes à travers le temps, impose de disposer d'outils d'annotation génomique performants et fiables. On parle ici d'annotation structurale, c'est-à-dire de la détection et la prédiction de la structure intron / exons des gènes présents dans les séquences ADN et aussi d'annotation fonctionnelle, c'est-à-dire de la prédiction de la fonction de ces gènes grâce aux relations d'homologies déterminées par phylogénie, et aux bases de données fonctionnelles existantes.

Le laboratoire a depuis plusieurs années mis au point de nouvelles procédures d'annotation originales et complexes, la qualité et la fiabilité des résultats ne pouvant être obtenues que par l'intervention d'une expertise de très haut niveau.

Pour analyser les énormes volumes d'ADN, issus du séquençage des génomes de différentes espèces, une automatisation complète de ces procédures est indispensable. Une telle automatisation implique donc de réussir la synthèse entre les outils mathématiques habituellement utilisés par le biologiste dans son travail d'annotation et son expertise qui intervient directement entre les phases de calculs mathématiques. Le biologiste décide en effet des outils mathématiques à utiliser, des paramètres à leur associer ; il intervient sur le " jeu de données " pour le construire, le filtrer, le modifier...

Nous avons ainsi réalisé la plate-forme informatique FIGENIX qui automatise complètement toutes nos procédures d'annotation à travers plusieurs modèles de pipeline bioinformatique. A de nombreuses étapes de ces pipelines intervient un module expert (fonctionnant en chaînage arrière), qui remplace le biologiste tout au long de la procédure. Ainsi il mémorise des informations cruciales, répond à des questions clefs orientant la suite de la procédure, analyse les données pour les corriger ou en déduire de nouvelles informations. (exemple : partitionnement multicritères d'un arbre phylogénétique en groupe d'orthologues (sous-arbres)).

Il s'agit d'un serveur logiciel autonome développé en langage PROLOG. Un ensemble de règles logiques modélise les connaissances et les raisonnements des biologistes du laboratoire. Chaque tâche (instance de pipeline) en cours d'exécution possède un monde PROLOG associé, c'est à dire un ensemble d'assertions représentant le contexte de la tâche.

Le choix de PROLOG se justifie par plusieurs faits : la simplicité, la souplesse et la concision de la description des connaissances et de la mise en oeuvre des raisonnements, la facilité de manipulation des structures de liste et d'arbre (notamment phylogénétiques) souvent présentes en bioinformatique, le mécanisme de backtracking natif à PROLOG qui facilite la recherche de solution multiples à un problème. (Cela est souvent utile en bioinformatique (exemple trivial : épissage alternatif pour la définition de protéines à partir d'un gène)) L'outil FIGENIX, publié dans la revue BMC Bioinformatics (voir références plus bas), est accessible par internet.

Par ailleurs l'analyse de génomes à grande échelle amène à l'étude de régions composées d'ensemble de gènes. Nous étudions en particuliers la conservation de ces régions à travers le processus de l'évolution et cherchons à reconstruire ces régions pour des espèces ancestrales. Nous développons un logiciel nommé C.A.S.S.I.O.P.E (Clever Agent System for Synteny Inheritance and Other Phenomena in Evolution) qui est un système multi-agents dédié à ce type d'analyse. En généralisant, on peut dire qu'il s'agit d'un système qui sera capable de répondre à des questions scientifiques de haut niveau.

Parmi les agents, on distingue les agents chargés du rapatriement de données biologiques disponibles à travers le Web (ftp, Web services, API dédiées (Ensembl, NCBI, ...)), les agents responsables de la persistance de l'information (ontologie persistante), les agents pilotant des systèmes informatiques externes (ex : FIGENIX), des agents " intelligents " à base de règles logiques (moteur de résolution en chaînage avant). Ces derniers réagissent à des impulsions (initiales : les questions, et événementielles : des faits venant

des autres agents (souvent suite aux requêtes qu'ils ont reçues)) en déduisant de nouveaux faits et en émettant de nouvelles requêtes vers les autres agents. Des résultats sont produits à chaque nouvelle déduction.

Une ontologie, centrée sur les concepts de l'évolution, définit l'univers du discours entre les agents. La construction de cette ontologie constitue en elle-même un challenge, puisqu'il s'agit d'intégrer des concepts de génomique, de phylogénie et ceux du comportement fonctionnel des gènes.

Les systèmes DAG (Direct Acyclic Graph) comme celui défini par Gene Ontology (<http://www.geneontology.org>), basés sur un unique type de relation entre concepts ne permettent pas de décrire toute la richesse sémantique dont nous avons besoin. L'unique type de relation, entre deux concepts, utilisée dans Gene Ontology est la relation " du plus générale vers le plus particulier ". Comment dans ce cas décrire plusieurs types de relation (exemple : espèce S appartient à la famille F et l'espèce S dérive d'un ancêtre A) ?

Nous utilisons donc la logique de description (DL), dont les capacités de modélisation sont plus larges, et certaines technologies actuelles (langage OWL) qui s'y rapportent.

La mise en place de cette ontologie est évidemment une condition incontournable pour la poursuite des réalisations informatiques nécessitant l'introduction de techniques d'intelligence artificielle basées sur la logique mathématique. C.A.S.S.I.O.P.E est aujourd'hui à l'état de prototype, non encore publié. Les principales technologies utilisées sont PROLOG, JAVA, JADE, OWL, PROTEGE, JENA.

En conclusion nous pouvons donc dire qu'en plus des techniques " classiques " de bioinformatique, nous introduisons dans nos systèmes informatiques, dès que cela est possible et naturel, une modélisation des concepts et des raisonnements faisant partie de l'expertise des biologistes. Nous utilisons et appliquons pour cela des techniques et des langages liés à l'intelligence artificielle, qui facilitent grandement cette tâche.

Bibliographie ciblée

Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, Danchin E.G.J. FIGENIX : Intelligent automation of genomic annotation : expertise integration in a new software platform. BMC Bioinformatics, 2005 ; 6 : 198.

<http://www.biomedcentral.com/1471-2105/6/198/>

EA 3888 Laboratoire d'Informatique Médicale

- CHU - Pontchaillou 2, rue Henri Le Guilloux- 35033 RENNES
- web : <http://www.med.univ-rennes1.fr/ea3888/index.php>
- Personne à contacter : Anita Burgun (Tel : 02 99 28 42 15 ; courriel : anita.burgun@univ-rennes1.fr)

Membres de l'équipe concernés par le sujet

Pavel Dobrokhoto, chercheur post doctoral, Julie Chabrier, chercheur post doctoral, Fleur Mougin, en thèse, Gwénaëlle Marquet, en thèse, Merc Cuggia, AHU, Olivier Dameron, maître de conférences, Anita Burgun, MCU-PH

Thème général de l'équipe

Les thèmes de recherche principaux de l'EA 3888 sont :

- Intégration de données et de connaissances
- Terminologies du domaine biologique et médical
- Ontologies biomédicales
- Etude des systèmes d'information en santé

Description des travaux ou projets en lien avec le thème

La thématique principale de l'équipe concerne les ontologies biomédicales. Elle aborde les aspects formels et terminologiques des ontologies :

- ontologies formelles, avec en particulier les travaux d'Olivier Dameron sur le Foundational Model of Anatomy. Les ontologies permettent de réutiliser des connaissances d'un domaine pour exploiter intelligemment des données. Deux facteurs importants sont donc d'une part la richesse descriptive d'une ontologie, et d'autre part la puissance du langage dans lequel elles sont représentées. Les ontologies formelles s'appuient sur des opérateurs ayant une définition mathématique précise. Cela permet ainsi d'utiliser des outils génériques pour effectuer des raisonnements complexes. Le travail d'Olivier Dameron porte principalement sur l'utilisation de formalismes de représentation issus du Web Sémantique comme OWL pour exploiter et enrichir des ontologies complexes.
- larges corpus terminologiques biomédicaux, en particulier le Unified Medical Language System (UMLS). Notre équipe a acquis une compétence particulière en ce qui concerne l'UMLS, développé par la U.S. Na-

tional Library of Medicine (NLM), que nous avons expérimenté dès 1992 pour la description des actes médicaux. L'UMLS est une vaste base de termes médicaux issus de plus de cent terminologies, organisés en plus de 1 million de concepts reliés par environ 22 millions de relations. De ce fait, l'UMLS est à ce jour la plus importante ressource de concepts biomédicaux au monde. L'EA 3888 a participé avec les chercheurs de la NLM à des travaux de recherche dans le cadre du projet MOR (Medical Ontology Research) initié en 2001 par la NLM. Le projet MOR explore les méthodologies d'acquisition et de validation d'ontologies du domaine biomédical et a pour objet l'évolution de l'UMLS vers une ontologie du domaine biomédical. Notre participation a consisté en travaux portant sur les co-occurrences en tant que composants des espaces sémantiques, sur l'alignement avec d'autres ontologies (Cyc, WordNet), et sur la cohérence de l'UMLS.

Les méthodes développées comprennent l'alignement d'ontologies, les méthodes liées aux logiques de description, et l'enrichissement d'ontologies à partir d'autres connaissances ou bases de données. Nous nous intéressons, entre autres, au rôle des ontologies de référence pour le raisonnement et l'acquisition de relations entre objets complexes, et aux méthodes basées sur les données pour acquérir des relations entre objets. Ces aspects seront renforcés par notre participation à partir d'avril 2006 au projet européen (FP6-2004-IST-4) BOOTStrep (Bootstrapping Of Ontologies and Terminologies STRategic REsearch Project) qui concerne la construction d'ontologies en génomique fonctionnelle. Dans ce projet, nous serons plus particulièrement responsables des méthodes visant à enrichir l'ontologie.

Un champ d'application concerne l'utilisation des ontologies pour l'intégration d'informations biomédicales :

- intégration des sources (travaux de F. Mougin). La recherche dans le domaine biomédical nécessite d'accéder à une multitude de bases de données et de connaissances accessibles via le Web, mais qui sont disparates et hétérogènes dans leur structure et leur terminologie. La collecte manuelle de ces informations ne peut être que partielle ou incroyablement fastidieuse, étant donné le nombre de sources disponibles et leur diversité. Nous cherchons donc à automatiser cette collecte et à la rendre " intelligente ". Dans ce cadre, nous avons ainsi conçu un système d'intégration de sources de données hétérogènes basé médiateur. Ce dernier s'appuie sur l'utilisation d'ontologies et terminologies existantes pour

disposer d'un vocabulaire commun aux différentes sources.

- intégration des données liées à l'annotation des gènes (projet BioMeKE, G. Marquet) ou plus généralement biomédicales (M. Cuggia), Ces projets visent en particulier à combiner pour les patients atteints de cancer des informations cliniques, des informations sur les échantillons tumoraux (tumorothèque), des données recueillies dans le cadre de protocoles thérapeutiques nationaux et des données de génomique (incluant transcriptome et protéome). Dans le projet BioMeKE, G. Marquet a développé une méthode d'annotation biomédicale des gènes ou produits de gènes, basée sur les connaissances de l'UMLS.

Un autre domaine est l'utilisation des ontologies pour la modélisation des réseaux de gènes dans le cadre de l'analyse transversale du transcriptome (J. Chabaliér). L'analyse transversale regroupe les gènes suivant plusieurs types d'information : 1) biologique, 2) médicale, et 3) transcriptomique (données d'expression). A partir de ces différents groupes, les relations fonctionnelles entre les gènes sont prédites. Ces relations correspondent à la similarité sémantique entre les différents gènes calculée par la méthode des modèles d'espace vectoriel (Vector Space Model). A partir d'un ensemble de gènes donnés, l'approche transversale fournit un ensemble de réseaux de gènes fonctionnellement liés. La validation de cette approche a été réalisée sur des gènes impliqués dans la différenciation entérocytaire. Cette application fait l'objet d'un contrat PRIR de la Région Bretagne.

Par ailleurs, dans le projet BIOTEXT, P. Dobrokhotov va étudier l'impact des ontologies, de l'UMLS et d'autres informations structurées sur le " bio-text mining " utilisé pour l'analyse et l'annotation des puces.

Courte bibliographie ciblée

Dameron O, Rubin DL, and Musen MA. Challenges in converting frame-based ontology into OWL : the Foundational Model of Anatomy case-study. In American Medical Informatics Association Conference AMIA05, 2005.

Chabaliér J, Garcelon N, Aubry M, Burgun A. A transversal approach to compute semantic similarity between genes. Workshop on Biomedical Ontologies and Text Processing - European Conference on Computational Biology (ECCB'05), Madrid, 28 sept - 1 oct 2005

Burgun A. Desiderata for domain reference ontologies in biomedicine. J. Biomed Inform 2005 Oct 17 ;.

Marquet G., Golbreich C., Burgun A. From an ontology-based search engine towards a more flexible integration for medical and biological information, Semantic Integration

workshop, ISWC 2003, Sanibel, FL, oct 2003, CEUR Vol 82, 2003

Mougin F., Bodenreider O. (2006) Éliminer les cycles dans les systèmes terminologiques : comparaison de deux approches. Actes du congrès francophone Reconnaissance des Formes et Intelligence Artificielle (RFIA 2006).

Equipe Apprentissage et Biologie Computationnelle (ABC) du LORIA UMR 7503

- LORIA, Campus Scientifique, BP 239, 54506 Vandoeuvre-lès-Nancy Cedex
- Site Web : <http://modbio.loria.fr/>
- Contact : Yann Guermeur (Tel : 03 83 59 30 18 ; courriel : Yann.Guermeur@loria.fr)

Membres de l'équipe concernés par le sujet

Yannick Darcy, Doctorant à l'Université Nancy 1 ; Yann Guermeur, CR CNRS ; Emmanuel Monfrini, Chercheur post-doctorant à l'Université Nancy 1 ; Fabienne Thomarat, MCF à l'Ecole des Mines de Nancy (INPL)

Thème général de l'équipe

L'objectif de notre équipe est double :

- effectuer des recherches dans le domaine de la classification et de manière plus spécifique développer la composante à la fois la plus utile et la moins avancée de sa théorie, la théorie statistique de la discrimination multi-classe,
- utiliser les résultats obtenus pour mettre au point des méthodes à noyau dédiées à la reconnaissance des formes et appliquées de manière privilégiée à des tâches relevant de la biologie moléculaire.

Description des travaux ou projets en lien avec le thème du dossier

L'implication d'ABC en bioinformatique se situe essentiellement au niveau régional et national. Elle s'organise autour de trois collaborations principales.

1. Projet "Apprentissage automatique appliqué à la prédiction de la structure tertiaire des protéines" (GENOTO3D) financé par l'ACI "Masses de Données" : <http://www.loria.fr/~guermeur/ACIMD/>

2. Projet "Développement et utilisation d'approches informatiques et théoriques pour l'analyse des liens existant entre défauts d'épissage et maladies génétiques". Ce projet, une collaboration avec le "Laboratoire Maturation des ARN et Enzymologie Moléculaire" (MAEM), UMR 7567 à Nancy, est financé pour 18 mois par le programme Décryphon <http://www.decryphon.fr/>. Il fait également l'objet d'une opération du thème "Bioinformatique et applications à la génomique" du "PRST Intelligence Logicielle".
3. Projet "Modélisation de la protéine FAK (Focal Adhesion Kinase) en vue de l'identification de molécules anti-métastases", en collaboration avec l'"équipe de Dynamique des Assemblages Membranaires" (eDAM) du laboratoire "Structure et Réactivité des Systèmes Moléculaires Complexes", UMR 7565 à Nancy. Ce projet est une opération du thème "Bioinformatique et applications à la génomique" du "PRST Intelligence Logicielle".

Courte bibliographie ciblée

Guermeur Y., Lifchitz A. et Vert R. (2004). A kernel for protein secondary structure prediction. Dans *Kernel Methods in Computational Biology*, édité par B. Schölkopf, K. Tsuda et J.-P. Vert, the MIT Press, Chap. 9, 193-206.

Guermeur Y., Pollastri G., Elisseeff A., Zelus D., Paugam-Moisy H. et Baldi P. (2004). Combining protein secondary structure prediction models with ensemble methods of optimal complexity. *Neurocomputing*, Vol. 56C, 305-327.

Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F., Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P., Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M., Weissenbach J. et Vivares C.P. (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*, Vol. 414, 450-453.

Sapay N., Guermeur Y. et Deléage G. (2006). Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. *BMC Bioinformatics*, Vol. 7, 255.

Thomarat F., Vivares C.P. et Gouy M. (2004). Phylogenetic analysis of the complete genome sequence of *Encephalitozoon cuniculi* supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes. *J Mol Evol*, Vol. 59, 780-791.

Equipe Bioinformatique du LRI CNRS UMR 8623

- Bât 490 - Centre Scientifique d'Orsay - 91 405 Orsay Cedex
- Web : <http://www.lri.fr/bio/indexbioinfo.html>
- Responsable : Christine Froidevaux (Tél : 01 69 15 65 07 ; courriel : chris@lri.fr)

Membres de l'équipe concernés par le sujet

Contacts : Jérôme Azé, Christine Froidevaux, Céline Rouveirol aze, chris, celine@lri.fr

Thème général de l'équipe

L'équipe Bioinfo du LRI comporte 6 enseignants-chercheurs et 8 doctorants aux compétences informatiques complémentaires : algorithmique, combinatoire, visualisation, intégration de données hétérogènes, représentation des connaissances et apprentissage, ce qui en fait sa richesse. Elle travaille en étroite collaboration avec des biologistes, principalement, à l'IGM (Institut de Génétique et de Microbiologie, Orsay), l'Institut Curie, MIG (Mathématiques, Informatique, Génome, INRA Jouy) et l'IBBMC (Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, Orsay). Les recherches sont effectuées selon trois axes :

1. analyse des structures et séquences biologiques (essentiellement structures d'ARN) ;
2. modélisation et la simulation de processus biologiques (modélisation de réactions biochimiques au niveau macromoléculaire) ;
3. analyse et l'intégration de données biologiques massives : voir ci-dessous.

Description plus détaillée des travaux ou projets en lien avec le thème du dossier

Intégration de données hétérogènes et ontologies

Le projet HKIS (projet européen R et D, Integrated Software Platform for Biological and Biomedical Data Processing in Cancerology, <http://www.hkis-project.com>), a élaboré une plate-forme basée sur le logiciel Amadea (produit Isoft), afin de fournir aux cancérologues un espace de travail où ils puissent analyser leurs données, en se basant sur des scénarios d'analyse et en interrogeant diverses sources de données dans un format unifié. Dans ce

contexte, afin d'aider l'utilisateur à choisir les sources à interroger à une étape donnée d'un scénario d'analyse, nous avons conçu (Cohen et al. 2004) le logiciel BioGuide, en collaboration avec Susan Davidson (Université de Pennsylvanie). BioGuide est un logiciel convivial qui permet au biologiste de choisir les sources et outils pouvant répondre à ses requêtes, tout en respectant ses préférences, et sa stratégie de navigation à travers les sources, et flexible (on peut aisément intégrer de nouveaux critères de préférences, ou de nouvelles sources).

Dans le cadre du projet RAFALE (voir projets), nous cherchons à apprendre des règles d'annotation à partir des cas connus, qui permettront d'inférer les classes fonctionnelles de protéines de fonctions inconnues. Pour cela, nous travaillons sur l'élaboration de "mappings" (correspondances) entre les classifications fonctionnelles et ontologies utilisées pour l'annotation des génomes existants.

Apprentissage et fouille de données biologiques

Analyse de données biologiques à grande échelle

Data mining dans des données CGH

Collaboration : Institut Curie Paris (équipe de F. Radvanyi et service Bioinformatique, E. Barillot)

Une puce CGH permet, pour un patient donné, de mesurer le nombre d'occurrences moyen d'un ensemble de séquences fixées sur le génome humain dans un prélèvement donné. Cette technique permet donc de connaître les régions altérées dans un échantillon, c'est-à-dire les régions amplifiées et perdues. Nous avons développé un outil d'extraction de régions minimales d'altérations génomiques dans un ensemble de profils (Rouveirol et al., 2006). Le problème se ramène à un problème de recherche de motifs séquentiels contraints, dans le paradigme des Bases de Données Inductives. Nous avons également développé un algorithme d'apprentissage supervisé, permettant de rechercher des combinaisons d'altérations génomiques associées avec un attribut de classe, la difficulté dans ce cas étant d'obtenir un classifieur robuste et fondé sur des règles statistiquement validées.

Ce thème va évoluer vers l'analyse de données hétérogènes (analyse de données de puces d'expression, CGH, protéiques, etc.), toujours en collaboration avec l'Institut Curie. Extraction de réseaux de régulation à partir de données d'expression Collaboration : équipe de F. Radvanyi (Institut Curie)

Un des défis majeurs de l'ère post-génomique est la construction, à partir d'informations telles que les données d'expression, de réseaux de régulation transcriptionnelle. Le but est de connaître, pour chaque gène et dans un

contexte cellulaire donné, quels facteurs de transcription influencent sa transcription, et surtout comment plusieurs facteurs se coordonnent pour accomplir certaines régulations. Les puces à ADN permettent de mettre en évidence des relations entre facteurs de transcription et leurs gènes cibles, à travers l'observation des valeurs d'expression simultanées de milliers de gènes et de facteurs de transcription. Nous avons proposé une approche à base de recherche de motifs fréquents (Elati et al., 2005) dédiée à l'apprentissage de relation de régulation complexes. Cette méthode permet de réduire efficacement l'espace des régulateurs candidats pour un gène cible et a été évaluée sur des données de transcriptome de tumeurs de vessie, ainsi que sur des données de transcriptome de levure.

Analyse de données de transcriptome

Collaboration : équipe de Marie Dutreix (Institut Curie) et équipe I et A du LRI

Le but est d'étudier les différences d'expression des gènes de cellules de levure dans deux conditions (environnement normal et exposition à de faibles doses de radiation). Nous avons conçu l'algorithme BioRelief qui est une technique d'estimation d'attributs qui cherche les gènes les plus significativement corrélés à la classe prédite (irradiés vs non irradiés). BioRelief est peu coûteux à mettre en oeuvre, ne fait pas d'hypothèse d'indépendance sur l'expression des gènes et reste valide en présence de données manquantes. Nous avons ainsi pu mettre en évidence la réponse transcriptionnelle des cellules de levure exposées à de faibles doses de radiation (Mary et al. 2004).

Apprentissage de fonctions de tri de complexes protéines-protéines

Collaboration : équipe Génomique Structurale (IBBMC).

Ce travail (Bernauer et al. 2005) vise à déterminer la fonction des protéines à partir de leur conformation 3D.

La fonction de la plupart des protéines est subordonnée à l'interaction avec un ou plusieurs partenaires qui peuvent être aussi bien de petites molécules que des macromolécules. Les complexes protéines-protéines sont impliqués dans beaucoup de processus cellulaires et physiologiques, c'est pourquoi leur détection présente un grand intérêt dans l'étude des fonctions cellulaires et biochimiques de leurs composants.

Actuellement, nous disposons d'algorithmes d'amarage (ou "docking") pouvant engendrer, à partir des conformations 3D des protéines, un nombre relativement important de complexes protéines-protéines, qui doivent être analysés par des biologistes pour vérifier s'ils peuvent être réellement observés expérimentalement. Cette vérification

étant très coûteuse, il est capital de disposer d'une fonction efficace de tri de ces complexes candidats.

La modélisation géométrique retenue pour générer les complexes candidats est réalisée à l'aide de diagrammes de Voronoï. A l'heure actuelle, de telles fonctions sont obtenues par apprentissage supervisé reposant sur un algorithme évolutionnaire et sur des données fournies par une analyse géométrique sur les conformations issues des algorithmes de Docking.

Projets en cours

RAFALE (LRI, MIG et IBBMC) : Règles d'Annotation Fonctionnelle semi-Automatique pour la Levure, ACI IMP-Bio, <http://www.lri.fr/RAFALE/>, But : aider les biologistes dans la tâche d'annotation (ou ré-annotation) en leur proposant un module semi-automatique d'annotation des protéines basé sur un système de règles apprises automatiquement, et doté d'une capacité de raisonnement. Actuellement, les règles sont apprises sur des génomes bactériens par de la programmation logique inductive. RAFALE s'appuie sur l'environnement intégré AGMIAL de MIG.

MICROBIOGENOMICS (IGM, MIG, LRI) : Extraction optimisée d'informations pertinentes à partir de données complexes et hétérogènes issues de comparaisons génomiques, ANR Masse de Données, <http://microbiogenomics.u-psud.fr>) But : rassembler un maximum de données pertinentes relatives aux génomes microbiens dans un entrepôt de données pour effectuer différentes techniques de fouille de données ; concevoir une interface graphique facilitant l'extraction d'informations pertinentes et l'analyse interactive de ces données. Application à l'étude des arbres évolutifs de familles de modules et des blocs de synténie. Logiciels BioGuide : <http://www.lri.fr/bio/soft/bioguide/bioguide.html> BioRelief : <http://www.lri.fr/~chris/bioinfo/BioRelief>

Bibliographie

(liste complète : <http://www.lri.fr/bio/indexbioinfo.html>)

J. Bernauer, A. Poupon, J. Aze, J. Janin., A docking analysis of the statistical physics of protein protein recognition, *Physical Biology*. 2005, 2 :S17-S23.

S. Cohen Boulakia, S. Davidson and Ch. Froidevaux, A User-centric Framework for Accessing Biological Sources and Tools, In Proc. of DILS'05 (Data Integration in the Life Sciences), Springer-Verlag, Lecture Notes in Bioinformatics (LNBI), Vol. 3615, 3-18, 2005.

M. Elati, C. Rouveirol et F. Radvanyi, Apprentissage de signatures de facteurs de transcription à partir de données d'expression, 5èmes journées d'Extraction et Gestion des Connaissances (EGC 2005), 667-672, 2005.

G. Mercier, N. Berthault, J. Mary, J. Peyre, A. Antoniadis, J-P. Comet, A. Cornuéjols, Ch. Froidevaux and M. Dutreix, Biological detection of low radiation by combining results of two microarray analysis methods, *Nucleic Acids Research*, Vol 32 (1), e12, 2004.

C. Rouveirol, N. Stransky, Ph. Hupé, Ph. La Rosa, E. Viara, E. Barillot, and F. Radvanyi, Computation of recurrent minimal genomic alterations from array-CGH data, *Bioinformatics* 22(7), 849-56, 2006.

Equipe Fouille de données et Nutrigénomique (P6 et P13)

- Université Paris 6-INSERM UMR 755 " Nutriomique ", Hôpital de l'Hotel-Dieu, 75004 Paris.
- Université Paris 13-LIM et BIO EA3969 "Laboratoire Informatique Médicale et BioInformatique ", 93000 Paris.
- Web : <http://www.limbio-paris13.org/Thematiques/Vise2.html>
- Responsable : Jean-daniel Zucker (jdzucker@nutriomique.org)

Membres informaticiens de l'équipe concernés

Arriel Benis, Florence Combes, Mélanie Courtine, Blaise Hanczar, Corneliu Henegar, Mustapha Lebbah, Jean-Daniel Zucker.

Thème général de l'équipe

Le thème de recherche de l'équipe concerne la bioinformatique du transcriptome en général, et en particulier l'analyse et la fouille de données génomiques, cliniques et transcriptomiques liées à l'obésité. Les méthodes utilisées sont d'une part des algorithmes d'apprentissage automatique supervisé et non supervisé mais aussi ceux changements de représentation (réduction de dimension, fusion de données et construction d'attribut). Les données utilisées proviennent principalement d'expérimentations menées dans le cadre de l'UMR Nutriomique dont le but est de comprendre les mécanismes physiopathologiques impliqués dans la régulation du poids et la résistance à l'amaigrissement chez l'obèse, qui caractérise la chronicisation de la maladie.

De nombreuses ressources biocliniques ont été accumulées et nous développons des approches informatiques nous permettant d'aborder les aspects multifactoriels de l'obésité en intégrant un grand nombre d'informations issues de la génétique, de la clinique et de la génomique fonctionnelle (puces à ADN notamment).

Cette recherche se décline en trois axes :

- La construction de modèles prédictifs à partir de données d'expression : c'est là l'un des problèmes clefs en bioinformatique du transcriptome. Il s'agit de construire des modèles permettant de se baser sur les données d'expression pour prédire l'évolution de paramètres cliniques. Dans le cas des données qui nous préoccupent, il s'agit de prédire l'évolution du poids (sa diminution ou sa reprise) à la suite de modifications (avant, pendant ou après un régime). Méthodes IA utilisée : Réduction de dimension, fusion de données cliniques et transcriptomiques.
- L'analyse fonctionnelle des données d'expression : considéré comme un défi majeur, la translation des données d'expression génique vers les mécanismes biologiques correspondants (appelée couramment analyse fonctionnelle), est une étape incontournable pour la compréhension des phénomènes biologiques sous-tendus par les modèles expérimentaux étudiés. Sur les données d'expression génique du tissu adipeux le but est de contribuer à la caractérisation des mécanismes moléculaires et des réseaux de régulation métabolique associés à l'évolution de l'obésité et à la survenue des complications. Méthodes IA utilisée : annotation fonctionnelle, Analyse de la surreprésentation des catégories fonctionnelles.
- L'analyse des réseaux d'interaction transcriptomique et fonctionnelle : l'analyse intégrative des interactions transcriptomiques est une problématique de recherche en plein essor dans la bioinformatique du transcriptome. Des travaux récents ont montré la structuration de ces interactions dans des réseaux modulaires dont l'architecture est fortement conditionnée par la fonction biologique des transcrits. Au sein de ces réseaux les modules d'interaction transcriptomique sont centrés et reliés par l'intermédiaire de gènes du développement ou régulateurs clés (appelés " hubs " ou noeuds transcriptomiques) qui assurent des rôles essentiels dans l'initiation et l'enchaînement des processus biologiques au niveau cellulaire. Dans notre contexte de recherche, ces analyses ont comme but principal de contribuer à la caractérisation des mécanismes moléculaires impliqués dans le développement de l'obésité et de ses complications afin d'identifier les meilleures cibles géniques

pour un futur développement pharmacologique. Méthodes IA utilisées : Apprentissage non supervisé, Représentation Multi-instance.

Description plus détaillée de certains travaux en lien avec le thème du dossier

Les axes de recherches indiqués ci-dessous correspondent à des collaborations étroites entre biologistes, cliniciens et informaticiens. Les paragraphes suivants détaillent la composante IA et bioinformatique de chacun de ces axes.

Réduction de dimension pour l'apprentissage de classifieur à partir de données puces

Dans le cas des données puces la disproportion entre le nombre d'exemples d'apprentissages et d'attributs est caricaturale. Le cas typique est celui où l'on dispose d'une cinquantaine d'exemples pour plusieurs dizaines de milliers d'attributs continus (les ratios d'expression). Une réduction de dimension est donc requise pour affiner les performances en classification et une interprétation biologique correctes. Deux méthodes de réduction de dimensions ont été développées. La première, ProGene, effectue une forme de compression de l'expression des gènes synexpressés en construisant de nouveaux attributs. La seconde FeatKNN mesure les interactions d'information entre les gènes et identifie des paires de gènes synergiques pour construire des nouveaux attributs permettant d'améliorer la précision du classifieur associé.

Fusion de données et prédiction :

L'une des directions de recherche les plus actives en apprentissage automatique est celle de la combinaison de classifieurs dont le but est d'améliorer le pouvoir prédictif. Par ailleurs, des études très récentes ont démontré l'intérêt de combiner des classifieurs utilisant des descriptions distinctes des mêmes exemples (Pavlidis 2002 ; Berrar 2004 ; Lanckriet 2004). Cet aspect est tout particulièrement intéressant en biomédecine où des patients peuvent être décrits à de multiples niveaux de détails (génétique, transcriptomique, protéomique, clinique, environnemental, etc.). L'intégration de tels classifieurs pour augmenter le pouvoir prédictif de leur combinaison est un enjeu scientifique clef. L'approche 2KC que nous avons mis en oeuvre dans l'équipe combine des noyaux sur les données biopuces et les données cliniques et permet d'améliorer significativement les performances. La validation empirique de cette approche est faite dans le cadre d'un projet de recherche européen DIOGENES " Diet, Obesity an Genes ".

Algorithmes d'apprentissage non supervisé pour l'analyse fonctionnelle des profils d'expression et des réseaux d'interaction transcriptomique

L'analyse des interactions biologiques au niveau transcriptomique se heurte à des difficultés particulières liées à la structure des bases d'annotations génomiques (i.e. GO, KEGG) sur lesquelles repose l'analyse fonctionnelle, qui relie chaque catégorie fonctionnelle à l'ensemble des transcrits qu'elle annote. Afin de contourner ces difficultés nous avons construit un modèle de représentation des catégories fonctionnelles en tant qu'objets multi-instance (i.e. chaque transcrit est considéré comme étant une instance du processus biologique dans lequel il est impliqué) qui nous a permis de développer une approche originale d'apprentissage non-supervisé permettant de quantifier les interactions fonctionnelles et d'inférer la structure des classes fonctionnelles (i.e. modules d'interaction fonctionnelle) exprimées au niveau transcriptomique. Ce développement a reposé à la fois sur une adaptation des approches conventionnelles d'apprentissage non-supervisé pour le modèle multi-instance, ainsi que sur une extension des approches multi-instance dans le cadre non-supervisé.

Algorithme de clustering pour l'analyse de combinaisons de Polymorphismes

Aujourd'hui, les biologistes savent que les maladies multifactorielles reposent sur le rôle de multiples combinaisons de variants génétiques qui agissent en interaction entre eux ou avec l'environnement. Ces interactions mettent en cause aussi bien des allèles fréquents que des allèles rares. Les études à large échelle des relations existantes entre les S.N.P. (Single Nucleotide Polymorphism) et d'autres données (phénotypiques et/ou environnementales) sont très peu étudiées, car elles engendrent de larges espaces de recherche avec de nombreuses relations peu "pertinentes". L'approche ObeLinks conçu permet de gérer ce type de données en combinant un algorithme de classification basé sur le treillis de Galois et des méthodes statistiques. La définition de tests de multiplicité adéquats a été un enjeu important pour filtrer les relations "faussement" significatives.

Publications Récentes

Henegar, C., Clément K. and Zucker J.-D. (2006). Unsupervised Multiple-Instance Learning for Functional Profiling of Genomic Data. ECML 2006.

Henegar, C., Canello R., Rome S., Vidal H., Clément K. and Zucker J.-D. (2006). "Clustering biological annotations and gene expression data to identify putatively co-regulated biological processes." *Journal of Bioinformatics and Computational Biology (JBCB)* 4(4).

Lebbah M, Thiria S, Badran F. Les perceptrons multicouches. A paraître en 2006 aux Editions Hermès.

Taleb, S., Haaften R. V., Henegar C., Hukshorn C., Canello R., Pelloux V., Hanczar B., Viguerie N., Langin D., Evelo C., J. Zucker J.-D., Clément K. and Saris W. H. M. (2006). "Microarray profiling of human white adipose tissue after exogenous leptin injection" *European Journal of Clinical Investigation* 36.

Henegar C, Bousquet C, Lillo-Le Louet A, Degoulet P, Jaulent MC. Building an ontology of adverse drug reactions for automated signal generation in pharmacovigilance. *Int J Med Inform.* 2005 Aug;74(7-8) :563-71.

Taleb S, Lacasa D, Bastard JP, Poitou C, Canello R, Pelloux V, Viguerie N, Benis A, Zucker JD, Bouillot JL, Coussieu C, Basdevant A, Langin D, Clement K. Cathepsin S, a novel biomarker of adiposity : relevance to atherogenesis. *FASEB J.* 2005 Sep;19(11) :1540-2.

Poitou C, Viguerie N, Canello R, De Matteis R, Cinti S, Stich V, Coussieu C, Gauthier E, Courtine M, Zucker JD, Barsh GS, Saris W, Bruneval P, Basdevant A, Langin D, Clement K. Serum amyloid A : production by human white adipocyte and regulation by obesity and nutrition. *Diabetologia.* 2005 Mar;48(3) :519-28.

Equipe Intelligence Artificielle pour la ChémoGénomique (IACG, laboratoire BIM)

- Laboratoire de Biologie Informatique Mathématiques Direction des Sciences de la Vie (DSV) Département Réponse et Dynamique Cellulaires (DRDC) CEA - Grenoble 17, avenue des Martyrs 38027 Grenoble Cedex 9
- Site Web : en cours de création
- Contact : Sylvaine Roy (04 38 78 59 21, sylvaine.roy@cea.fr)

Composition

Samia Aci (post-doctorant) Samuel Wieczorek (ingénieur-chercheur CEA et doctorant) Sylvaine Roy (ingénieur-chercheur CEA, responsable)

Contexte biologique

Les travaux de cette équipe, née il y a quelques mois, portent sur le développement de méthodes et d'outils adaptés au traitement et à l'analyse de données issues d'une approche globale en biologie, la chémoGénomique. Cette discipline vise à explorer les espaces chimiques et biologiques

et à comprendre leurs interrelations. Elle recouvre un ensemble d'approches dans lesquelles les petites molécules sont utilisées pour l'étude des systèmes biologiques.

Elle est explorée par plusieurs équipes au CEA, en particulier l'équipe d'Eric Maréchal (Chemogénomique comparative des plantes et des apicomplexes, DRDC/PCV) et l'équipe de Laurence Lafanechère responsable du centre de Criblage pour des Molécules Bio-Actives (CMBA), biologistes avec qui IACG développe d'étroites collaborations.

Eric Maréchal s'intéresse notamment à l'étude des cibles biologiques, et les travaux menés par IACG en collaboration avec son équipe ont concerné des outils destinés à l'étude des génomes biaisés.

Un objectif du CMBA est de découvrir des molécules présentant des mécanismes d'actions biologiques nouveaux, qui puissent servir d'outils pour la recherche. Le principal outil utilisé pour trouver ces molécules est le criblage de larges collections de molécules chimiques (chimiothèques) grâce à des automates (criblage à haut débit). Les criblages sont réalisés soit sur des protéines isolées, soit sur des cellules dans leur entier. On cherche alors à mesurer une modification spécifique de la physiologie de la cellule, se traduisant par un changement de son phénotype (criblage phénotypique). Les criblages phénotypiques peuvent être utilisés pour cibler des protéines qui ne peuvent pas être analysées sous une forme pure ou des protéines inconnues. Les molécules sélectionnées peuvent ensuite servir d'"hameçons" pour purifier la protéine cible. Du fait de sa similarité conceptuelle avec la génétique classique, ce genre d'approche est appelé "génétique chimique directe".

Objectifs généraux et travaux actuels

Dans ce contexte, la masse de données générées est importante, et pour répondre aux questions biologiques posées, il faut pouvoir intégrer, modéliser, analyser des informations tant au niveau des structures moléculaires, qu'au niveau des cibles, ou encore au niveau de l'interaction molécule-cible.

L'Intelligence Artificielle a d'abord été présente de façon indirecte à travers la modélisation-objet du système d'information PhenoScreen conçu pour analyser et gérer les résultats de criblage (représentation des molécules chimiques, réalisation des analyses multi-paramétriques d'activité, identification des molécules bio-actives, aide à la détection des artefacts expérimentaux). Pour répondre aux besoins d'exploration globale des interactions molécule-cible, nous privilégions actuellement l'apprentissage artificiel. Nous collaborons ainsi étroitement avec Gilles Bisson et Mirta Gordon de l'équipe "Apprentissage" du Laboratoire Leibniz, IMAG, Grenoble.

Notre projet principal actuellement est le projet AC-CAMBA (<http://accamba.imag.fr/>) dont l'objectif est de développer des outils permettant d'évaluer la bio-activité potentielle des chimiothèques utilisables et de guider la synthèse de molécules plus actives. La partie modélisation des SAR (Structure-Activity Relationship) est basée sur une utilisation intensive de l'apprentissage artificiel.

Notre contribution porte plus particulièrement sur les questions suivantes :

- recherche et extraction de descripteurs moléculaires pertinents : Des méthodes nommées "descripteurs moléculaires" ont été développées ces dernières années afin de décrire des corrélations existantes entre certaines propriétés d'une petite molécule et sa structure. Il existe des centaines de descripteurs moléculaires différents et il est nécessaire d'évaluer leur pertinence et leur intérêt afin d'éviter de saturer le programme d'apprentissage d'informations récurrentes d'une part, et, d'autre part, de passer à côté d'une corrélation entre bio-activité et structure moléculaire du fait de l'absence d'un descripteur traduisant la propriété adéquate. Les travaux réalisés ont permis la conception d'un premier jeu de données qui sera utilisé très prochainement par les apprentistes. Par ailleurs, les axes de réflexion dégagés autour de ce travail nous amènent à nous interroger sur les notions d'ontologie de la molécule. La formalisation de ces notions est actuellement en cours et nécessite d'être approfondie dans un contexte de représentation des connaissances.
- travaux sur des algorithmes d'apprentissage relationnel : Les méthodes d'apprentissage relationnel semblent prometteuses pour représenter et manipuler des données structurées comme les molécules (que l'on peut représenter grossièrement par un graphe). Cependant, ces méthodes ont besoin d'être adaptées pour faire face à la réalité des données de criblage :
- les molécules des chimiothèques étudiées (dont la taille peut atteindre 20000 composés pour une petite chimiothèque académique) comportent en moyenne une quarantaine d'atomes; ceci peut compliquer leur analyse dans le cadre de méthodes telles que la Programmation Logique Inductive,
- les résultats positifs (molécules bio-actives) représentent une très faible proportion par rapport aux résultats négatifs, ce qui déséquilibre fortement les jeux d'apprentissage,
- il peut exister plusieurs concepts à apprendre dans un même jeu d'apprentissage : plusieurs éléments structuraux peuvent expliquer une bio-activité,

l'interaction de molécules avec plusieurs cibles totalement différentes dans le cadre d'un criblage cellulaire peuvent donner une activité similaire,

- enfin, le bruit dans les données, inhérent aux expérimentations sur du matériel vivant, ajoute encore à la difficulté de l'apprentissage.

Les travaux actuels visent à se doter d'une mesure de ressemblance efficace, qui permet d'identifier plusieurs sous-structures communes (ou différentes) entre deux molécules ; ces sous-structures disjointes ou partiellement recouvrantes sont autant de graines pour les algorithmes d'apprentissage. Ces travaux font l'objet de la thèse de S. Wiczorek, encadré par le laboratoire Leibniz.

L'équipe IACG collabore également aux projets suivants :

- le projet TULIP : porté par Eric Maréchal, ce projet dont le coeur a constitué la thèse d'Olivier Bastien, s'attache à développer des méthodes numériques pour l'étude des génomes biaisés notamment ceux des Apicomplexes (agents de la malaria, de la toxoplasmosse). Il s'élargit sur l'exploration génomique comparée des Angiospermes et des Apicomplexes.
- le projet DOCK ou " Conformation Sampling and Docking on Grids " : ce projet est soutenu par l'ANR dans le cadre du Programme " Calcul Intensif et Grilles de Calcul ". Dragos Horvath (IBL Lille) y développe de nouveaux algorithmes de docking qui s'appuieront notamment sur des approches d'algorithmes génétiques. Ces algorithmes intrinsèquement parallèles tourneront sur une plate-forme de grilles de calcul, dont l'architecture sera construite par l'équipe d'El Ghazali Talbi (LIFL, Université de Lille). Les résultats de ces algorithmes seront validés " in vivo " (par le CMBA et IACG), sur des cibles intéressantes pour la recherche de composés anti-cancéreux.

Axes futurs envisagés

Les projets de nos collaborateurs biologistes montrent un besoin croissant d'analyse et d'intégration de données hétérogènes. Le CMBA, par exemple, envisage d'établir le profil cytologique des molécules reconnues comme actives sur un phénotype, par analyse d'images des cellules soumises à l'action de ces molécules. Ce profil sera comparé avec ceux obtenus en parallèle pour un ensemble de composés de référence, présentant des effets biologiques connus et ciblant différentes voies de signalisation. Leur analyse comparée devrait permettre de classer les molécules sélectionnées et de prédire la voie ciblée. Il est également prévu d'enrichir les données sur les interactions

molécules-cibles par des techniques classiques de criblage virtuel (docking) et par l'extraction de données de la littérature. Dans ce contexte, l'utilisation de techniques comme l'analyse automatique de la littérature, la reconnaissance et la classification des phénotypes analysés par imagerie, sera sans doute pertinente, les choix de bonnes représentations de connaissances adaptées pour l'intégration des données sur les molécules, sur les cibles et sur leurs interactions seront certainement indispensables.

Références

E. Sans-Soleilhac, C. Barette, S. Wiczorek, S. Roy, E. Maréchal et L. Lafanechère, Criblages phénotypiques et "généétique chimique directe" : une approche innovante pour la découverte de molécules bio-actives et/ou de candidats médicaments, *Spectra Analyse* 32(233), 33-37 (2003).

O. Bastien, J.-C. Aude, S. Roy and E. Maréchal. Fundamentals of massive automatic pairwise alignments of protein sequences : theoretical significance of Z -value statistics *Bioinformatics* 20 (4) 534-537 (2004)

O. Bastien, P. Ortet, S. Roy, and E. Maréchal. A configuration space of homologous proteins, conserving mutual information and allowing a phylogeny inference based on pair-wise Z-score probabilities *BMC Bioinformatics* 6 : 49 (2005)

Contribution à l'édition d'un ouvrage collectif :

Le criblage pharmacologique automatisé : du haut-débit au haut-contenu d'information, Edited by E. Maréchal, S. Roy and L. Lafanechère, EDP Sciences, sous presse.

Equipe PLIAGE du laboratoire LSR (Logiciels, Systèmes, Réseaux) de l'IMAG (Informatique et Mathématiques Appliquées de Grenoble) et équipe LCM du laboratoire IBS (Institut de Biologie Structurale, Grenoble)

Membres

Fabien Corblin (doctorant, LSR), Eric Fanchon (CNRS, IBS), Nicolas Thierry-Mieg (CNRS, LSR), Laurent Trilling (Enseignant, LSR)

Thème général

pooling pour la planification d'expériences en biologie haut-débit, modélisation, inférence et simulation de réseaux biologiques.

Description des travaux en lien avec le dossier*Pooling pour la planification d'expériences en biologie haut-débit*

Dans le cadre d'expériences où l'objectif est d'identifier une petite fraction de positifs parmi un grand nombre d' " objets " (clones, drogues, peptides, ...), l'utilisation de méthodes de pooling permet de diminuer le nombre d'expériences à réaliser tout en luttant efficacement contre le bruit expérimental. L'idée consiste à construire un petit ensemble de pools astucieusement choisis et redondants d'objets ; ces pools sont ensuite criblés et le motif des pools positifs permet d'identifier les objets positifs. Comme chaque objet est présent dans de nombreux pools, les faux-positifs et les faux-négatifs peuvent être identifiés et corrigés. Un problème crucial réside dans le choix de la composition des pools. Pour y répondre, nous avons développé un nouvel algorithme efficace et robuste : le Shifted Transversal Design (STD) (Thierry-Mieg 2005).

Un pool est modélisé par une clause positive dont chaque atome est un objet : par exemple, l'interaction, vraie ou fausse, d'une protéine avec un appât. Une telle clause est vraie si le criblage contre l'appât du pool associé a donné un résultat positif, fausse sinon. Il s'agit de fournir un ensemble " personnalisé " de telles clauses à la fois de taille réduite et suffisamment redondant pour détecter un nombre intéressant d'erreurs expérimentales. Corriger ces erreurs revient à déterminer un modèle logique " le moins distant " pour un ensemble de clauses incohérent. Cette détermination requiert des algorithmes particulièrement optimisés qui d'ailleurs ne sont pas spécifiques à STD et peuvent être adaptés à d'autres systèmes de pooling.

Le développement de STD a été motivé en particulier par les projets de cartographie des interactomes par double-hybride dirigés par Marc Vidal au CCSB (Center for Cancer Systems Biology) auxquels nous participons depuis 1998 (Walhout et al. 2000). Une première expérience pilote reposant sur STD et concernant 940 protéines humaines est en cours depuis décembre 2004 : nous avons construit 169 pools suivant la méthode STD, et les résultats initiaux du criblage de 100 protéines appât sont prometteurs. Nous souhaitons assurer le passage de l'approche STD, au sein du CCSB, à l'échelle d'un interactome : il s'agit de standardiser les protocoles et d'améliorer les algorithmes, en tirant parti

des résultats de l'expérience pilote en cours et des spécificités de STD.

Modélisation, inférence et simulation de réseaux biologiques

Dans la mesure où la quantité de données au niveau moléculaire explose, la recherche en biologie aborde une nouvelle étape. Le but à atteindre est de découvrir l'organisation de ces composants moléculaires sous la forme de système de régulation. Il s'agit de bâtir des modèles formels à partir de descriptions souvent informelles, incomplètes, imprécises et qualitatives (fournies par les biologistes).

L'objectif consiste à développer un outil informatique à base de PLC (Programmation Logique avec Contraintes) permettant à la fois d'inférer un (ou des) modèle(s) d'un phénomène biologique et de procéder à des simulations ou à des prédictions à partir de ce (ou ces) modèle(s).

Les modèles inférés doivent l'être bien sûr en cohérence avec les données fournies par les biologistes. Ces données concernent des indications sur les possibilités d'interaction des composants et sur les comportements du phénomène tels que des inégalités de seuil, des états stationnaires ou des propriétés concernant des suites d'états caractéristiques. La représentation d'un réseau adoptée est une extension des " réseaux logiques " AMLM (Asynchronous Multivalued Logical Model) proposée par R.Thomas . Ces réseaux logiques, adaptés au manque de données quantitatives précises, permettent une approximation discrète très raisonnable de réseaux définis par des équations différentielles. Pour permettre à la fois l'inférence, la simulation ou la proposition d'expériences destinées à confirmer des prédictions ou des hypothèses (à partir des caractéristiques obligatoirement observables), l'approche adoptée est déclarative et utilise la PLC.

Une première mise en oeuvre a été réalisée et testée avec succès sur le phénomène de l'adhésion cellulaire en liaison avec des biologistes de l'IBS (Institut de Biologie Structurale, Grenoble) (Fanchon et al. 2005). L'outil employé pour le logiciel est Prolog IV, très utile grâce aux combinaisons possibles de contraintes booléennes et numériques (construction de contraintes " réifiées ") et à ses possibilités de définition de négations et de prise en compte de données quantitatives.

Les études sont menées dans plusieurs directions :

- développement du formalisme AMLM pour prendre en compte des états singuliers (ayant au moins une composante sur un seuil) en étendant la notion d'équations focales.
- détermination de propriétés statiques du réseau et étude de décompositions possibles du réseau de fa-

çon à améliorer les performances en introduisant des contraintes redondantes

- conception d'un langage logique (inspiré par exemple de CTL, Computation Tree Logic), avec lequel les biologistes pourront exprimer les propriétés constatées et/ou à vérifier, ainsi que les spécifications des objectifs d'expériences à proposer.
- représentation en intention de modèles. L'idée consiste d'une part à établir des classes d'équivalence de modèles du point de vue des comportements, d'autre part à apprendre les modèles intéressants d'une façon intelligible par les biologistes, c'est-à-dire sous la forme d'un ensemble de contraintes simples dont ils sont solution. Un des moyens d'apprentissage envisagé est la PLI (Programmation Logique Inductive).
- évaluation de l'approche et élaboration d'une méthodologie en ré-examinant les constructions ad hoc de réseaux biologiques de la littérature.
- utilisation d'autres types de contraintes, stochastiques (par exemple pour l'inférence de modèles), faibles (pour déterminer le modèle " le plus proche " en cas d'incohérence), différentielles (pour intégrer des données quantitatives). Etude de contraintes spécialisées pour les modèles AMLM.

Bibliographie

Thierry-Mieg N. A new pooling strategy for high-throughput screening : the Shifted Transversal Design, BMC Bioinformatics 2006, 7 :28.

Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GE, Brasch MA, Thierry-Mieg N, Vidal M. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. Science. 2000 Jan 7;287(5450) :116-22.

Fanchon E, Corblin F, Trilling L, Hermant B, Gulino D. Modeling the Molecular Network Controlling Adhesion Between Human Endothelial Cells : Inference and Simulation Using Constraint Logic Programming. Computational Methods in Systems Biology 2004, Danos V, Schachter V (eds). Lecture Notes in Bioinformatics 3082, 114-118, 2005.

Fanchon E, Corblin F, Trilling L. Modeling, inference and simulation of biological networks using Constraint Logic Programming (CLP). ECCS05, European Conference on Complex Systems, Paris, 2005.

Equipe Statistique et informatique appliquées à la génétique et à la biologie moléculaire. Département de mathématiques et informatique appliquées (MIA), Institut National de la Recherche Agronomique (INRA), Toulouse.

- Adress : INRA. Chemin de Borderouge, BP 52627, Castanet-Tolosan 31326 Cedex, France
- Tel : (+33) 5.61.28.54.28
- Fax : (+33) 5.61.28.53.35
- web : <http://www.inra.fr/mia>
- Responsable : Thomas Schiex
(Thomas.Schiex@toulouse.inra.fr)
- Composition : une douzaine de permanents dont 6 chercheurs.

Domaine d'activité

L'équipe développe et applique des méthodes issues de l'informatique, de l'intelligence artificielle et de la statistique à des problèmes issus de la biologie. L'équipe comprend quatre chercheurs informaticiens et deux chercheurs statisticiens. Cette palette de compétences permet de répondre de façon pertinente à des problèmes variés qui peuvent être issus de la génétique ou de la biologie moléculaire et en particulier de l'analyse de séquences nucléotidiques (ADN et ARN). De façon plus spécifique, l'équipe se focalise à l'heure actuelle sur des problèmes de localisation de régions fonctionnelles au niveau génétique (marqueurs génétiques) et au niveau des séquences (génome).

Dans le domaine de la génétique, l'équipe a une longue expérience qui va de la construction de cartes génétiques et d'hybrides irradiés à l'identification de régions du génome dont le contenu est statistiquement lié à un caractère quantitatif d'intérêt (QTL : quantitative trait loci). L'activité va de l'élaboration de modèles mathématiques pertinents et traitables jusqu'à la construction de logiciels permettant l'utilisation directe des méthodes développées par les biologistes (logiciels CarthaGène (de Givry et al. 2005) - www.inra.fr/mia/T/carthagene - et MCQTL - www.genoplante.com). L'équipe occupe donc un nombre non négligeable d'ingénieurs informaticiens. Les méthodologies mises en oeuvre et développées sont, pour ce qui concerne l'intelligence artificielle, liées à l'optimisation (programmation dynamique, recherche locale...) ou à la modélisation probabiliste (HMM, réseaux bayésiens ou réseaux de contraintes pondérés très proches)

Dans le domaine de l'analyse de séquences nucléotidiques (ADN/ARN), l'équipe a également une longue histoire dans l'analyse des séquences d'ARN et en particulier dans l'étude et l'exploitation de leur structure (tridimensionnelle) qui définit leur fonction. Appréhendé à un niveau simplifié, appelé structure secondaire, il peut s'agir de déterminer la structure à partir de la séquence et de connaissances thermodynamiques et biologiques ou de rechercher les régions d'un génome qui permettent de synthétiser un gène d'ARN soit à partir de propriétés générales des séquences de gènes d'ARN (détection " *ab initio* "), soit à partir de l'analyse de la conservation de ces régions entre génomes (analyse comparative) soit en connaissant de façon grossière la structure secondaire adoptée. Les méthodes et outils développés doivent répondre à plusieurs critères : les utilisateurs souhaitent un résultat bien caractérisé (une réponse simple à analyser) et la taille des séquences analysées (des centaines de millions de caractères dans certains cas) nécessite une efficacité algorithmique suffisante. Les travaux de l'équipe dans le domaine, combinant algorithmique des textes (combinatorial pattern matching) et réseaux de contraintes ont ainsi permis de construire un outil de recherche de gènes d'ARN efficace et capable en particulier de prendre en compte des interactions entre molécules (logiciel MilPat - carlit.toulouse.inra.fr/MilPaT/MilPat.pl, thèse de Patricia Thébault). La découverte récente de nombreuses familles de gènes d'ARN en interaction (Gaspin et al. 2005) a fortement motivé ces travaux.

L'équipe a également une activité sur la localisation de gènes de protéines aussi bien sur des organismes procaryotes (bactéries) qu'eucaryotes (plantes, animaux). Ici, programmation dynamique, modèles de Markov, apprentissage (SVM...) et optimisation stochastique ont pu être combinés pour construire deux outils dédiés à ces deux grandes classes d'organismes et appliqués à différents génomes bactériens et végétaux. Dans tous les cas, la nécessité de fusionner des informations de natures très diverses est un obstacle méthodologique aussi bien que pratique important.

Plus récemment, l'étude des remaniements chromosomiques entre espèces a suscité un intérêt important qui a mené à la définition de problème de calcul de distances entre organismes basées sur d'autres mécanismes que les mutations ponctuelles. L'équipe s'est intéressé à la définition d'algorithmes efficaces (Bérard et al. 2005) pour le calcul de telles distances, contribuant à l'étude de l'évolution des espèces.

Enfin, du côté de l'intelligence artificielle, l'équipe maintient depuis longtemps une activité purement méthodologique, parfois motivée par les problèmes appliqués (gé-

nétique, gènes d'ARN), autour des réseaux de contraintes (Cooper et al. 2004).

Partenariat, finalisation

les problèmes issus de la biologie sont généralement traités en collaboration avec des biologistes, permettant un travail pertinent, sur une problématique bien identifiée et motivée par des données récentes. Le caractère générique des problèmes traités est important car il permet de répondre, via la construction d'un outil dédié, à une demande générale et donc conduit à une meilleure visibilité. Ces travaux sont parfois menés très loin dans la direction de la finalisation pour en vérifier la pertinence, menant en général à des co-publications avec des biologistes dans des journaux de biologie parfois prestigieux (Salanoubat et al. 2002) ou de génétique en sus des publications plus méthodologiques (Cooper et al. 2004). Naturellement, la localisation de l'équipe favorise l'identification de tels partenaires, mais l'équipe collabore bien sûr aussi avec des biologistes issus d'autres organismes (CNRS, universités).

Pour certains problèmes ou projets d'ampleur (séquençage, annotation ; projets européens), une collaboration avec d'autres équipes impliquées dans la bioinformatique est nécessaire. L'équipe a ainsi des collaborations avec les principaux centres internationaux (TIGR, MIPS...) ainsi qu'en France (INRIA).

Enfin, l'équipe entretient des collaborations nationales et internationales (Espagne, Italie, Irlande...) sur la principale thématique méthodologique côté IA : les réseaux de contraintes.

Support et service

L'équipe héberge des structures orientés vers le service au biologiste et au bioinformaticien via la mise à disposition d'outils, de banques de séquences et de données et de logiciels (Plateforme bioinformatique de la génopole de Toulouse, dirigée par C. Gaspin et D. Allouche) ou via le traitement et la mise à disposition de données biologiques (EST, puces ...) internes INRA (projet SIGENAE, C. Klopp).

Références

S. de Givry, M. Bouchez, P. Chabrier, D. Milan, et T. Schiex, CARTHAGENE : multipopulation integrated genetic and radiated hybrid mapping, *Bioinformatics*, 21(8) : 1703-1704 (2005).

C. Gaspin et P. Thébault, Identification d'ARN non codants, *Biofutur* 251, (33-36), 2005.

S. Bérard, A. Bergeron, C. Chauve, C. Paul, Perfect Sorting by Reversals Is Not Always Difficult, WABI 2005, Mallorca, Spain.

M. Cooper, T. Schiex, Arc consistency for soft constraints, Artificial Intelligence, Volume 154 (1-2), Avril 2004, 199-227.

M. Salanoubat, S. Genin, F. Artiguenave, J. Gouzy, S. Mangenot, M. Arlat, A. Billault, P. Brottier, J. C. Camus, L. Cattolico, M. Chandler, N. Choisne, C. Claudel-Renard, S. Cunnac, N. Demange, C. Gaspin, M. Lavie, A. Moisan, C. Robert, W. Saurin, T. Schiex, P. Siguier, P. Thébault, M. Whalen, P. Wincker, M. Levy, J. Weissenbach et C. A. Boucher, Genome sequence of the plant pathogen *Ralstonia solanacearum*, Nature, 415, 497-502 (2002).

Laboratoire de Biologie Virtuelle

- Adresse : UMR 6543 CNRS et Université de Nice (UNSA), Centre de Biochimie, Faculté des Sciences, 06108 Nice cedex 2
- Site Web : <http://bioinfo.unice.fr> <http://thea.unice.fr>
- Contact : Richard Christen ou Claude Pasquier (Tel : 04 92 07 6947; courriel : richard.christen@unice.fr; claud.pasquier@unice.fr)

Membres permanents de l'équipe concernés par le sujet

Richard Christen, Claude Pasquier

Thème général de l'équipe

Intégration, modélisation et exploitation des connaissances; application aux données génomiques et post-génomiques

Description des travaux ou projets en lien avec le thème du dossier

Le laboratoire de Biologie Virtuelle est composé d'une équipe pluridisciplinaire appartenant à l'Institut de Signalisation, Biologie du Développement et Cancer (UMR6543). Dans le cadre des contributions "IA et Bioinformatique" nos travaux s'inscrivent dans deux types d'études complémentaires : d'une part l'intégration et la modélisation des données génomiques et post-génomiques dans un système à base de connaissances permettant la représentation de données hétérogènes et complexes, et d'autre part l'exploitation "intelligente" de ces connaissances en utilisant des modèles de raisonnement.

Une proportion non négligeable des connaissances dans les sciences du vivant sont incomplètes, imprécises, incertaines, seulement valides dans un contexte particulier, et, le plus souvent, évolutives. Dans l'équipe, nous accordons une grande importance à la modélisation des connaissances contextuelles. Nous développons pour cela un réseau sémantique dans lequel chacune des relations est associée à un contexte. Ce contexte représente les propriétés pour lesquelles la relation est vraie (au minimum, la source de la relation et l'espace temporel durant lequel cette relation est valide). Ce travail a donné lieu au développement d'Allonto, un ensemble d'outils permettant de construire et d'exploiter une base de connaissances. Allonto inclut un module d'import des données au format RDF/OWL, des fonctions de raisonnement RDFS et OWL, une API pour la navigation dans le graphe et un langage de requêtes basé sur RDQL. La base de connaissances est utilisée dans THEA, une application, diffusée aux biologistes, qui constitue une aide à l'analyse du transcriptome. A partir de groupes de transcrits (mis en évidence, par exemple, grâce à une classification hiérarchique des transcrits en fonction de leurs niveaux d'expression mesurés par des expériences de puces à ADN), THEA fournit des informations se rapportant aux fonctions ou processus biologiques, aux localisations cellulaires, aux stades de développement ou encore à la position physique des gènes sur le génome.

Les raisonnements sur des très gros jeux de données, permettent de mettre en évidence des invariants ou des singularités, qui peuvent eux-mêmes être des gros jeux de données. Nous portons une attention particulière à la mise au point de modes de représentations qui puissent être exploitées visuellement par l'oeil humain. THEA, par exemple, effectue des analyses statistiques puis en fournit le résultat sous une forme graphique, permettant de détecter facilement d'éventuelles caractéristiques significativement taguées.

Courte bibliographie ciblée

(Liste complète sur http://bioinfo.unice.fr/publications_en.htm)

Martinez R., Christen R., Pasquier C. and Pasquier N. Exploratory Analysis of Cancer SAGE Data. Proceedings of the PKDD'2005 conference, Discovery Challenge, october 2005.

Pasquier C, Girardot E, Jevardat de Fombelle K. and Christen R. THEA : Ontology driven analysis of microarray data. (2004) Bioinformatics, 20 16, Pages 2636-2643.

Iliopoulos I, Tsoka S., Andrade MA, Enright AJ, Carroll M, Pouillet P, Promponas V, Liakopoulos T, Palaaios G, Pasquier C., Hamodrakas S, Tamames J, Yagnik AT, Tramontano A, Devos D, Blaschke C, Valencia A, Brett D, Martin D, Leroy

C, Rigoutsos I, Sander C, Ouzounis CA. Evaluation of annotation strategies using an entire genome sequence. *Bioinformatics*. 2003 Apr 12;19(6):717-26.

Projet Symbiose, IRISA, campus de Beaulieu 35042 Rennes Cedex

- Web : <http://www.irisa.fr/symbiose/>
- Responsable : Jacques Nicolas (tél : 02.99.84.73.12, Jacques.Nicolas@irisa.fr)

Membres de l'équipe concernés par le sujet

François Coste, Michel Leborgne, Israël-César Lerman, Jacques Nicolas, Ovidiu Radulescu, Anne Siegel, Basavanneppa Tallur.

Thème général de l'équipe

Symbiose est un projet de recherche en bioinformatique, qui s'intéresse d'une manière générale à la modélisation des données génomiques et post-génomiques, dans le but d'assister le biologiste moléculaire dans la formulation, la découverte et le raisonnement sur ses connaissances.

Le projet s'articule autour de 3 grands axes :

- L'analyse de séquences par des langages formels : cet axe concerne la recherche des structures pertinentes (i.e. fonctionnelles) dans les macro-molécules, en s'appuyant sur la théorie des langages. Le premier objectif est d'offrir au biologiste le moyen de construire des modèles expressifs sur les macro-molécules. Nous proposons un environnement graphique, un langage à variables de chaînes et un calcul sur des arbres de suffixes dans ce but. Le second objectif est d'apprendre automatiquement les modèles à partir d'ensembles de séquences. Nous développons des recherches en inférence d'automates et de grammaires capables de caractériser un ensemble ou de discriminer un ensemble d'un autre dans ce but.
- L'étude des déplacements d'équilibre de systèmes biologiques : il s'agit d'identifier des produits avec un effet significatif dans des voies métaboliques et de signalisation, en interaction avec un réseau génétique. Notre approche repose sur la synthèse des connaissances d'interaction sous forme d'un modèle qualitatif et sa dérivation en modèles différentiels et/ou en systèmes d'équations dans l'algèbre des signes +, -, ?, . Le but est d'expliquer les observations mais égale-

ment de prédire le comportement du système biologique.

- L'optimisation combinatoire et l'usage du parallélisme : cet axe pose le problème de l'accélération des calculs complexes en génomique (comparaison, recherche de motifs intensive, prédiction de structures ...). Nous proposons deux types d'études pour dépasser les limitations actuelles : (1) accélérer le temps de traitement en utilisant des architectures reconfigurables pour accélérer le filtrage (FPGA) et en modélisant finement les problèmes à résoudre dans le cadre de l'optimisation discrète (en particulier programmation en nombres entiers pour la prédiction de la structure des protéines) ; (2) accélérer l'accès aux données en utilisant une architecture parallèle de disques et des techniques d'indexation.

L'ensemble de ces travaux est appliqué sur des données biologiques dans le cadre de collaborations favorisées par notre participation à Ouest-Genopole, un regroupement d'une cinquantaine de laboratoires du grand Ouest.

Description plus détaillée de travaux ou projets en lien avec le thème du dossier

Apprentissage de grammaires sur les macromolécules

Les projets de séquençage permettent de générer un grand nombre de séquences génomiques (ADN/ARN, Protéines) qui contraste avec le petit nombre de séquences dont la fonction a été établie expérimentalement. Le défi est alors d'établir la fonction des nouvelles séquences obtenues. Celle-ci peut être prédite par analogie si l'on trouve une séquence suffisamment proche dont la fonction est connue. Des outils de comparaison de séquences tels que Blast ont été développés et utilisés avec succès pour la recherche de séquences similaires. Cette approche reste néanmoins limitée à des séquences très fortement similaires, ceci sur une grande longueur. Une autre approche consiste à établir des bibliothèques de signatures des fonctions (voir par exemple pour les protéines la base Interpro, <http://www.ebi.ac.uk/interpro/>). La signature d'une fonction représente les parties caractéristiques de la séquence par rapport à la fonction. La prédiction de la fonction d'une nouvelle séquence se fait alors par la reconnaissance de parties caractéristiques d'une fonction par une signature de la base. Cette approche permet de se focaliser plus finement sur les zones fonctionnelles mais nécessite de constituer les signatures.

Dans l'équipe, nous nous intéressons à l'apprentissage automatique de ces signatures à partir d'exemples de séquences partageant une même fonction (famille fonc-

tionnelle), et éventuellement de séquences proches mais n'appartenant pas à la famille. La nature séquentielle des macros-molécules nous a amenés à formaliser le problème dans le cadre de l'inférence grammaticale : nous considérons la famille de séquences comme un langage (ensemble éventuellement infini de mots) dont on cherche à apprendre une représentation finie (grammaire, automate, ...) à partir d'un échantillon du langage. Nous accordons une grande importance à ne pas apprendre seulement un langage mais aussi une représentation explicite de celui-ci permettant de modéliser la famille. Contrairement à des approches de type boîte noire comme les réseaux de neurones, on ne veut pas seulement savoir si une séquence appartient à la famille mais aussi savoir pourquoi elle appartient à la famille. Les travaux de l'équipe sur ce sujet portent sur les classes de représentations du type systèmes de réécriture adaptées aux différents types de macro-molécules et sur les algorithmes permettant de les apprendre à partir d'exemples, les connaissances du domaine étant intégrées dans l'apprentissage que ce soit pour restreindre l'espace de recherche ou pour guider l'algorithme.

Etude des déplacements d'équilibre dans les systèmes biologiques

Les techniques développées par les biologistes pendant les 10 dernières années, du type puces à ADN, puces à cellules ou production en masse de données concernant les interactions protéine-protéine ont fait naître le besoin d'une approche globale ou analytique en biologie moléculaire. Le domaine de la "biologie systémique" développe actuellement un certain nombre de réponses à ces besoins. Dans ce domaine, il s'agit d'abord de modéliser des interactions au sein d'une cellule, puis de comprendre le comportement d'un système comme résultat de l'action de chacun de ses composants. Dans ce cadre, on peut utiliser des simulations ou des prédictions. On peut ensuite confronter les prédictions des modèles aux données massives disponibles, ce qui permet de valider le modèle, ou de l'affiner via un processus d'inférence de modèle.

Des difficultés se posent à chacune de ces étapes : comment choisir les éléments clés d'un système à modéliser ? Quelles interactions prendre en compte ? Quels formalismes utiliser pour simuler le comportement du modèle puis établir ses prédictions (formalismes continus, discrets, linéaire par morceaux) ? Enfin, lorsque les prédictions et les données ne sont pas compatibles, comment proposer des modifications au modèle qui rendent les données compatibles, et permettre ainsi de mieux comprendre le phénomène étudié ?

Dans le cadre du projet Symbiose, nous développons une approche qualitative qui permet de comparer les déplacements d'équilibre prédits par le graphe d'interaction

associé à un modèle avec le signe des variations de certaines molécules réellement observées pendant des expérimentations (typiquement, via des analyses transcriptomiques). La méthode employée consiste à coder le graphe d'interaction en un système d'équations qualitatives dans l'algèbre des signes +, -, ?. Pour être compatibles, des observations doivent pouvoir s'étendre à une solution du système d'équations. La recherche des solutions des équations qualitatives est un problème NP-complet que nous résolvons pour des modèles comportant plusieurs centaines de gènes à l'aide de diagrammes de décision ternaires. Ceci permet d'abord de tester la compatibilité entre des données et un modèle. Lorsqu'il y a incompatibilité, nous pouvons proposer des corrections du modèle (ou des données) minimales en terme du nombre de modifications. On entre alors dans le cadre de l'inférence de modèles. Lorsqu'il y a compatibilité, on peut identifier les sommets clés du modèle, au sens où leur variation influe sur une très grosse partie du système. Ces sommets sont à observer en priorité lors d'expérimentations futures pour comprendre le fonctionnement du système.

Références récentes

F. Coste and G. Kerbellec, A Similar Fragments Merging Approach to Learn Automata on Proteins, ECML 2005 (European Conference on Machine Learning).

F. Coste, D. Fredouille, C. Kermorvant, and C. de la Higuera, Introducing Domain and Typing Bias in Automata Inference, ICGI 2004 (International Colloquium on Grammatical Inference).

I. Jacquemin and J. Nicolas, Disulfide bonds prediction using inductive logic programming, WCB 2005, Workshop on Constraint Based Methods for Bioinformatics.

A. Siegel, O. Radulescu, M. Le Borgne, P. Veber, J. Ouy, S. Laguarrigue, Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks, Biosystems, 84, 2006

O. Radulescu, S. Laguarrigue, A. Siegel, M. Le Borgne, P. Veber, Topology and linear response of interaction networks in molecular biology, Journal of the Royal Society Interface, 3, 2006.

Thème Data mining et Bases de Données Inductives du LIRIS CNRS UMR 5205

– INSA Lyon Bâtiment Blaise Pascal F-69621 Villeurbanne cedex

– Contact : Jean-François Boulicaut (courriel : Jean-Francois.Boulicaut@insa-lyon.fr)

Participants aux recherches sur l'interrogation inductive de données post-génomique

Jérémy Besson, Sylvain Blachon, Ieva Mitasiunaitė, Ruggero Pensa, Christophe Rigotti, Céline Robardet

Principales collaborations en biologie moléculaire

Olivier Gandrillon (équipe " Signalisations et identités cellulaires " du CGMC CNRS-UCBL UMR 5534), Sophie Rome (équipe " Mécanismes Moléculaires du Diabète " de l'UMR INRA/INSERM 1235).

L'extraction de connaissances à partir de données (ECD) est un domaine de recherche au carrefour de multiples thématiques (intelligence artificielle et raisonnement, induction et apprentissage automatique, bases de données, statistiques, etc.). Un processus ECD est intrinsèquement interactif, itératif et complexe. Il faut travailler sur les données, extraire des motifs ou des modèles à l'aide d'algorithmes qui peuvent être coûteux (e.g., le calcul de toutes les règles d'association valides) puis disposer d'outils de post-traitement des extractions pour assister l'interprétation et l'explicitation de connaissances utiles aux propriétaires des données. Dans le cadre des bases de données inductives (BDI), un tel processus revient à une séquence de requêtes qui traitent les données mais aussi les motifs/modèles qui tiennent dans les données. Le contrat européen cInQ IST/FET-2000-26469 (" consortium on discovering knowledge with Inductive Queries ", 2001-2004) a débuté l'étude de ce cadre théorique. Nos travaux se poursuivent maintenant dans le cadre de l'ACI Masse de Données Bingo (" Bases de données inductives pour la génomique ", 2004-2007) et du contrat européen IQ IST/FET FP6-516169 (" Inductive Queries for patterns and models ", 2005-2008). Nous privilégions des applications en biologie moléculaire, et plus particulièrement en analyse du transcriptome. En effet, les collaborations avec des équipes de biologistes nous permettent à la fois de progresser sur le cadre théorique BDI (e.g., l'identification de primitives pour la spécification déclarative des contraintes sur les motifs et modèles, la démonstration de l'importance du concept de représentation condensée), de concevoir de nouveaux outils et méthodes ECBD (algorithmes, logiciels), mais aussi de participer à des découvertes de connaissances en biologie (voir, e.g., (Thierry-Mieg 2005)). Plus précisément, nous avons fouillé des données expérimentales comme les données humaines SAGE (expression de dizaines de milliers de gènes dans une centaine de tissus ou organes) ou bien

l'analyse de données Puces à ADN avec une application particulière à l'étude de mécanismes de régulation par l'insuline chez l'homme.

Les données d'expression peuvent être vues comme de grandes matrices de nombres qui quantifient un niveau d'expression de chaque gène dans chaque situation ou expérience biologique. Disons que l'un des défis consiste à trouver des modules sous la forme d'ensembles de gènes qui fonctionnent de concert (co-régulation) dans un ensemble de situations biologiques particulier (e.g., pour des maladies). L'une des principales techniques utilisées est la classification non supervisée (" clustering ") pour calculer des groupes de gènes et/ou de situations ayant des profils d'expression similaires. D'une part les limites de ces méthodes sont bien connues (voir les chapitres d'état de l'art dans (Boulicaut et al. 2002)). D'autre part, les sources de données pertinentes pour étudier le transcriptome sont clairement hétérogènes : en marge des valeurs d'expressions, on s'intéresse aussi aux fonctions des gènes, aux motifs qui apparaissent dans leurs séquences promotrices, ou encore à la caractérisation des types de situations expérimentales ou à diverses observations cliniques. Notre spécificité est de développer de nouvelles méthodes de découverte de connaissances basées sur (a) un codage booléen des propriétés des gènes étudiés, (b) l'extraction complète sous contrainte de collections de motifs locaux satisfaisant une certaine contrainte dans les contextes booléens construits, et (c) diverses propositions pour des usages multiples de tels motifs locaux.

Ainsi, il est possible de pré-traiter les données numériques d'expression pour coder des propriétés comme " tel gène est sur-exprimé ou pas dans telle situation " (Becquet et al. 2002). Une fois que de telles matrices sont construites, on peut mettre en oeuvre les techniques d'extraction de motifs ensemblistes. On peut alors trouver des motifs du type " Les gènes 1, 3, 78 et 349 sont souvent sur-exprimés simultanément " (ensemble fréquent) ou bien " Quand les gènes 1 et 3 sont sur-exprimés dans une situation, alors les gènes 78, 349 et 512 le sont également dans 98

La découverte de tels motifs au moyen d'algorithmes complets (on veut tous les ensembles fréquents pour un seuil de fréquence donné ou bien toutes les règles valides ou encore tous les concepts formels) pose des problèmes algorithmiques difficiles. Nous avons d'abord appliqué nos algorithmes de calculs d'ensembles fréquents et de règles d'association initialement développés pour l'analyse de données transactionnelles (Boulicaut et al. 2003). Pour traiter les contextes éventuellement petits en taille mais très denses, nous avons dû développer de nouveaux algorithmes permettant l'extraction de modules de transcription potentiels. Sur un plan informatique, il s'agit alors soit

du calcul sous contraintes de concepts formels (Boulicaut et al. 2003) soit de plusieurs extensions des concepts formels vers la tolérance aux exceptions et donc une meilleure adéquation à des données fondamentalement bruitées. Lorsque des collections, éventuellement grandes, de motifs locaux désignent des régularités dans les données et, par exemple, de possibles mécanismes de régulation, il faut faciliter leurs post-traitements par les biologistes. L'idée est, notamment, de les aider dans leurs navigations pour détecter le plus rapidement possible les faux positifs. L'une des voies très prometteuses consiste à étudier la complémentarité entre motifs locaux (e.g., des concepts formels) et des motifs globaux (e.g., des bi-partitions). Par exemple, nous venons de proposer un cadre de bi-partitionnement générique basé sur la connaissance de motifs locaux : le partitionnement est calculé à partir des singularités locales. De façon duale, nous pouvons aussi calculer des bi-partitions et exploiter des motifs locaux pour caractériser les groupements obtenus.

Pour conclure, une vision synthétique de notre approche pour l'analyse de données d'expression de gènes est décrite dans (Pensa et al. 2006) et les outils logiciels associés devraient être très bientôt accessibles pour la recherche. De nouvelles opportunités apparaissent pour l'étude des aspects dynamiques des processus de régulation (e.g., la disponibilité de données d'expression cinétiques). On assiste ainsi à une possibilité de rapprochement entre des travaux sur la modélisation et la simulation de systèmes complexes (e.g., la régulation des gènes au sein de la cellule) et des développements sur l'extraction de connaissances à partir de données. L'émergence programmée d'un nouvel axe au LIRIS intitulé " Connaissance et Systèmes Complexes " va dans cette direction.

Sélection de publications

(voir liste complète sur <http://liris.cnrs.fr/~jboulica/index.html#Publications>)

C. Becquet, S. Blachon, B. Jeudy, J-F. Boulicaut, O. Gandrillon. Strong-association-rule mining for large-scale gene-expression data analysis : a case study on human SAGE data. *Genome Biology journal* 3(12), December 2002.

J-F. Boulicaut and O. Gandrillon Coordinateurs. Informatique pour l'analyse du transcriptome, Hermes Science Publishing, Traité IC2, 310 pages. Juillet 2004.

J. Besson, C. Robardet, J-F. Boulicaut, S. Rome. Constraint-based concept mining and its application to microarray data analysis. *Intelligent Data Analysis* 9(1) :59-82, 2005.

J-F. Boulicaut, A. Bykowski, C. Rigotti. Free-sets : a condensed representation of boolean data for the approxi-

mation of frequency queries. *Data Mining and Knowledge Discovery* 7(1) 2003. pp. 5-22.

R. Pensa, J. Besson, C. Robardet, J-F. Boulicaut. Contribution to gene expression data analysis by means of set pattern mining. *Constraint-based mining and Inductive Databases*. Springer-Verlag LNCS 3848, pp. 328-347, 2006.

U729 Inserm Santé Publique et Informatique Médicale (SPIM)

- 15, rue de l'Ecole de Médecine 75006 Paris
- Web : <http://www.spim.jussieu.fr/>
- Contacts : Marie-Christine Jaulent (Tel : 01 42 34 69 86, courriel : marie-christine.jaulent@spim.jussieu.fr)
Natalia Grabar (Tel : 01 53 10 92 13, courriel : natalia.grabar@spim.jussieu.fr)

Membres de l'équipe concernés par le sujet

Marie-Christine Jaulent, 01 42 34 69 86, marie-christine.jaulent@spim.jussieu.fr Natalia Grabar, 01 53 10 92 13, natalia.grabar@spim.jussieu.fr Cédric Bousquet, 01 53 10 92 13, cedric.bousquet@spim.jussieu.fr Inès Jilani, 01 42 34 69 83, ines.jilani@spim.jussieu.fr

Thème général de l'équipe

L'U729 Inserm, créée en janvier 2002, est une équipe de recherche en ingénierie des connaissances pour l'e-santé qui s'appuie sur les Sciences et Technologies de l'Information et de la Communication (STIC) pour développer des systèmes à base de connaissances et des réseaux de soins en biomédecine.

La recherche menée au sein de l'U729 vise à construire des serveurs d'expertise, accessibles sur le réseau Internet, qui diffusent une connaissance biomédicale contrôlée et adaptée. Ils offrent des services d'aide à la décision et de mise en réseau des connaissances permettant le travail collaboratif.

Les objectifs de la recherche sont de proposer des méthodes innovantes pour développer ces serveurs d'expertise biomédicaux, les mettre en place et mesurer l'impact de leur utilisation sur la pratique de soins. Trois thèmes sont associés à cette recherche :

1. Thème " ingénierie des connaissances " : la modélisation et la représentation des connaissances biomédicales dans des structures informatiques,
2. Thème " génie logiciel " : la spécification et le développement de systèmes informatiques partagés qui

utilisent ces connaissances (systèmes experts, sites web, services, etc.),

3. Thème " évaluation " : la recherche des barrières et facilitations pour l'implantation de ces systèmes dans la pratique des soins et, finalement, l'évaluation du bénéfice de leur utilisation et leur impact.

Description des travaux ou projets en lien avec le thème du dossier

5 Plateforme Biomathématiques et génomique de Paris

Collaboration avec EBI / Hinxton / UK, Coling / Jena / Allemagne dans le cadre du Réseau d'Excellence européen Semantic Mining

Problématique :

Une des tâches les plus ardues dans l'activité des biologistes travaillant sur les expériences biopuces et en comparaison de séquences est la validation et l'interprétation des résultats obtenus. Cette tâche consiste à faire des recherches dans des bases des documents ou des bases de données génétiques avec des mots clés correspondant aux gènes et au phénomène biologique étudié afin de trouver des résultats qui argumentent ou complètent ceux des biologistes. Cette recherche par mots clés engendre un deuxième travail pour le biologiste à savoir l'analyse des documents trouvés par le moteur de recherche classique afin d'identifier les connaissances pertinentes.

La notion de Web sémantique repose sur une description des ressources disponibles sur le Web. L'une des ambitions du web sémantique est de faciliter la tâche de recherche en essayant d'automatiser le traitement de l'information sur le web et ce en associant à chaque document une annotation dite sémantique basée sur une ontologie décrivant le domaine. Cette annotation va décrire le contenu sémantique des documents à savoir dans le domaine des biopuces, le type de gènes intervenant dans l'expérience décrite par l'article scientifique et les interactions qui peuvent exister entre eux, avec des composants cellulaires ou des processus biologiques.

Les enjeux pour les secteurs de recherche du domaine biomédical sont particulièrement importants car la recherche dans ce domaine nécessite d'accéder à une multitude de bases de connaissances accessibles via le Web, mais hétérogènes dans leur structure et leur terminologie. Parmi ces ressources, citons les bases de données comme Swissprot ou FlyBase où les produits de gènes sont annotés par GeneOntology, les bases documentaires comme Medline ou le catalogue CISMef, dont les articles sont indexés par MeSH.

Expertise de l'U729 :

Notre mission dans ce projet consiste à analyser automatiquement de gros volumes de la littérature scientifique afin d'en extraire les éléments pour l'interprétation et la validation des clusters de gènes obtenus suite aux comparaisons de séquences et aux expériences avec les biopuces. Notre expérience en construction des ontologies et en ingénierie des connaissances nous permet ensuite de concevoir des outils d'intégration des connaissances extraites sur les gènes, de les modéliser et de les faire collaborer avec les travaux en biologie.

Apport à la plateforme :

- apport de contenu lié aux projets en cours aux Cordeliers et au niveau international
- apport de compétence en bioinformatique (sur le text mining, les algorithmes d'apprentissage automatiques, l'intelligence artificielle en général)
- apport original de compétence en ingénierie des connaissances de santé (sur le traitement des connaissances de nature linguistique en particulier)
- apport de contenu et de structuration aux termes de GeneOntology.

Courte bibliographie ciblée

Voir <http://www.spim.jussieu.fr/> pour une liste complète de références

Grabar N, Sillam M, Jaulent MC, Lefebvre C, Henrion E, Néri C. 2005. From likelihoodness between words to the finding of functional profile for ortholog genes. In RANLP Text Mining WS.

Grabar N, Hamon T. 2004. Les relations dans les terminologies structurées : de la théorie à la pratique. Revue d'Intelligence Artificielle, 18, pp. 57-85.

Bousquet C, Jaulent MC, Le Bozec C, Degoulet P. 2003. New terminology services based on term comparison using semantic definitions and similarity computation. Stud Health Technol Inform. 2003 ;95 :385-90

Grabar N, Jaulent MC, Chambaz A, Néri C. 2006. Sifting abstracts from PubMed and evaluation of their relevance for the molecular biology. Soumis à MIE 2006

Grabar N, Jaulent MC, Chambaz A, Néri C. 2006. Using PubMed abstracts and log likelihood ratio associating gene names with their functional profile. Soumis à SMBM 2006

Unité Mathématique, Informatique et Génome (MIG) - INRA

- Nom : MIG
- Adresse : Centre de Jouy, Domaine de Vilvert, 78352 Jouy-en-Josas
- Site Web : <http://migale.jouy.inra.fr/mig>
- Contact : Claire Nédellec (Tél : 01 34 65 28 78 ; courriel : claire.nedellec@jouy.inra.fr)

Membres de l'unité concernés par le sujet

François Rodolphe (directeur), Philippe Bessières, Robert Bossy, Hélène Chiapello, Vincent Fromion, Jean Garnier, Jean-François Gibrat, Annie Gendrault-Jacquemard, Christelle Hennequet-Antier, Valentin Loux, Alain-Pierre Manine, Elodie Marchadier, Antoine Marin, Claire Nédellec, Pierre Nicolas, Etienne Roquain, Sophie Schbath, Jean-François Taly, Stéphane Teletchea, Nicolas Turenne, Karel Zimmermann

Présentation de MIG

Créée en 2000 à Jouy-en-Josas, l'unité MIG rassemble des biologistes, des physico-chimistes, des informaticiens et des mathématiciens issus de trois départements de l'INRA : MICA (MICrobiologie et Alimentation), PHASE (Physiologie Animale et Systèmes d'Élevage) et MIA (Mathématiques et Informatique Appliquées). L'unité est placée sous la responsabilité de ce dernier, est en charge de la Bioinformatique, et a trois types de missions :

- développer de nouvelles méthodes d'analyse de génomes qui constituent les axes de recherche de l'unité ;
- contribuer à l'annotation de génomes d'intérêt pour l'INRA en collaboration avec des biologistes ;
- développer le système d'information capable de supporter les deux activités précédentes.

Le noyau des recherches de MIG concerne l'analyse des séquences biologiques et des structures des protéines :

1. analyse des séquences d'ADN par des modèles de chaînes de Markov permettant la détection de gènes procaryotes, l'identification de gènes fortement exprimés ou issus de transferts horizontaux interspécifiques, la détection de motifs contribuant à la stabilité des génomes ou à la régulation de l'expression des gènes ;
2. recherche de similitudes structurales entre protéines par protein threading (comparaison des repliements

des séquences sur les domaines structuraux connus et recherche de repliements "compatibles").

Mais d'autres sujets sont rapidement apparus :

- l'analyse des résultats d'expériences à grande échelle déployées par des programmes de génomique fonctionnelle, pour l'étude de l'expression des gènes (puces à ADN) ;
- la comparaison des génomes par des méthodes d'alignements et de l'analyse de contexte des gènes ;
- l'extraction des connaissances biologiques décrites dans la littérature scientifique dans un contexte d'exploitation de l'intégration des connaissances appliquée à la "biologie intégrative".

L'ensemble des outils et bases de données réalisées par l'unité sont accessibles sur la page Web.

Apprentissage appliqué à l'extraction d'information

C'est à travers ce dernier sujet d'extraction de connaissances à partir du texte que MIG s'est investi dans les méthodes d'apprentissage artificiel. Le constat de départ était qu'un certain nombre de connaissances biologiques cruciales pour la compréhension du fonctionnement d'une cellule, typiquement les régulations des gènes et des protéines, ne sont décrites que dans la littérature scientifique. C'est donc un enjeu important pour l'analyse des génomes que d'arriver à extraire automatiquement ces connaissances et à les structurer pour pouvoir les croiser systématiquement avec les résultats expérimentaux et les prédictions bioinformatiques.

Une chaîne d'extraction d'information sur les régulations biologiques à partir des résumés bibliographiques MedLine a été progressivement élaborée, composée d'un filtrage des fragments de texte pertinents, de reconnaissance d'entités nommées (noms de gènes et de protéines), d'extraction de termes, d'analyse syntaxique, de construction d'ontologies et finalement d'apprentissage des règles d'extraction. L'idée centrale est d'appliquer des méthodes d'apprentissage, notamment à la constitution des ressources linguistiques propres à un domaine, pour une meilleure exhaustivité et une adaptabilité de la chaîne à d'autres types d'informations à extraire. Le classifieur filtrant le texte et les règles d'extraction des informations sont appris de manière supervisée, et reposent respectivement sur des méthodes statistiques (Bayésien naïf, Support Vector Machines) et de la programmation logique inductive. Les classes sémantiques composant l'ontologie sont apprises à partir du corpus de texte par itération d'étapes de classification conceptuelle automatique et de validation manuelle par un expert.

Ce thème fait l'objet de collaborations depuis 2000 entre MIG et le LIPN (Laboratoire d'Informatique Paris-Nord à Villetaneuse, U. Paris 13) comme partenaire principal, ainsi que les laboratoires LEIBNIZ (IMAG, Grenoble), LRI (Laboratoire de Recherche en Informatique à Orsay, U. Paris XI), le projet Symbiose (IRISA, Rennes), avec la participation de l'unité de Génétique Animale de l'INRA-ENSAR de Rennes. Ces collaborations ont fait l'objet de différents financements, Caderige (ACI Bioinformatique), Extraplodocs (RNTL), ALVIS (STREP, 6ème PCRD) .

Apprentissage appliqué à l'annotation de la fonction des gènes

D'autres domaines de recherche de MIG constituent un terrain d'application à l'apprentissage artificiel. Ainsi, l'unité ayant mis en place une plate-forme d'annotation de génomes séquencés par l'INRA, de nouveaux projets envisagent l'utilisation des méthodes d'apprentissage à l'annotation automatique de la fonction des gènes. Ces projets (ACI IMPBio RAFALE et ANR MicrobioGenomics) impliquent des collaborations avec l'équipe bioinformatique du LRI, l'IBBMC et l'IGM (Institut de Génétique et Microbiologie) de la faculté des Sciences d'Orsay.

Bibliographie

(voir aussi http://migale.jouy.inra.fr/mig/mig_fr/presentation/publication/)

Abdel Vetah M. O., Aubin S., Bessi res P., Bisson G., Hamon T., Lagarrigue S., Manine A. P., Nazarenko A., N dellec C., Poibeau T., Weissenbacher D., Event-based Information Extraction for the biomedical domain : the Caderige project", in Proceedings of International Joint workshop on Natural Language Processing in Biomedicine and its Applications (NLPBA/BioNLP) 2004, Collier N., Ruch P. and Nazarenko A. (Eds), Gen ve, 2004.

Cussens, J. and N dellec, C. (august, 2005). Extracting Relations from Biomedical Texts, Proceedings of the Learning Language in Logic ICML joint workshop and the Challenge task, <http://www.cs.york.ac.uk/aig/111/11105/>. Springer Verlag, Bonn, Germany.

Page web : http://migale.jouy.inra.fr/mig/mig_fr/presentation/publication/

Martin, J., Gibrat, J.-F. and Rodolphe, F. (2005a). Choosing the optimal hidden markov model for secondary-structure prediction. /IEEE Intelligent Systems/. 20 (6) 19-25.

N dellec, C. and Nazarenko, A. (2005). Ontology and Information Extraction : A Necessary Symbiosis. (B. M. P. Buitelaar, P. Cimiano, ed.). Volume 123 Frontiers in Artificial Intelligence and Application. IOS Press.

Robardet, C., N dellec, C. and Boulicaut, J. (2004). Fouille de donn es et analyse du transcriptome, Chapitre 4. 101-141. trait  IC2. Herm s Lavoisier.

Robin, S., Rodolphe, F. and Schbath, S. (2005). DNA, Words and Models. Cambridge University Press, English version of ADN, mots et mod les, BELIN 2003.

Les dossiers du bulletin

SMA et transport	n° 66	IA et Terminologie	n° 32
IA et entreprises (2)	n° 64	Décision et IA	n° 31
Bioinformatique	n° 63	IA et Image	n° 30
IA et entreprises (1)	n° 62	Raisonn. temporel et spatial	n° 29
Ingénierie du Web et I.A.	n° 61	Systèmes Multi-agents	n° 28
IA et image	n° 60	IA et robotique	n° 27
Raisonnement à partir de cas	n° 59	I.A. et biologie moléculaire	n° 26
Multimédia et I.A.	n° 58	I.A. et droit	n° 25
Planification	n° 56/57	I.A. et fusion de données	n° 24
Systèmes d'information	n° 55	I.A. et musique	n° 23
Web sémantique	n° 54	Apprentissage	n° 22
L'IA dans le RNTL	n° 53	Les explications dans les SBC	n° 20
IA et diagnostic	n° 52	Pétrole-Chimie	n° 19
Temps, espace et évolutif	n° 51	Raisonnement à partir de cas	n° 18
Equipes d'IA en France	n° 49/50	I.A. et temps-réel	n° 17
IA et Médecine	n° 48	Planification et action	n° 16
Fouille de données	n° 46/47	Traitement automatique des langues	n° 15
IA et document	n° 44	I.A. et médecine	n° 14
IA et connexionnisme	n° 43	Diagnostic à base de modèles	n° 13
IA et Vie Artificielle	n° 42	Validation des SBC	n° 12
IA et CHM	n° 41	Le connexionnisme	n° 11
IA et EIAH	n° 40	I.A. et jeux	n° 10
Plates-formes multi-agents	n° 39	E.I.A.O.	n° 9
IA et WEB	n° 38	I.A. et gestion	n° 8
Mémoires d'entreprises	n° 36	Conception et I.A.	n° 7
IA et logique	n° 35	Intelligence artificielle distribuée	n° 6
Ingénierie des connaissances	n° 34	Acquisition des Connaissances	n° 5
IA et Télécommunications	n° 33	IA et ordonnancement	n° 4

L'AFIA recrute

Nous recherchons une personne qui pourrait tenir la rubrique « **Présentation de laboratoires** » pour le Bulletin de l'AFIA. Cette personne fera partie du Comité de Rédaction du Bulletin, et est capable de joindre des chercheurs/enseignants/industriels dans les mondes académiques ou industriels, afin de récolter une présentation de laboratoire tous les 3 mois. Toute personne intéressée est invitée à envoyer un courrier électronique au rédacteur en chef du Bulletin (voir page 2).

L'AFIA recrute

Nous recherchons également une personne qui pourrait tenir la rubrique « **Débats** ». Cette personne fera partie du Comité de Rédaction du Bulletin. Comme ci-dessus, toute personne intéressée est invitée à envoyer un courrier électronique au rédacteur en chef du Bulletin voir page (2).

Entreprises actives en IA, Bulletins 62 et 64

Dossiers réalisés par Gérard Petitjean.

A2IA
AKIO SOFTWARE
ALCTRA
ALDEBARAN Robotics
APODIS
ARDANS
ARTELYS
AXLOG ingénierie
BAYESIA
BOUYGUES E-LAB
CANTOCHE
CODEAS
CO-DECISION TECHNOLOGY SAS
DAUMAS AUTHEMAN et Associés
EUROBIOS
EURODECISION
EVITECH
EXALEAD
FACING-IT
FIRCOSOFT
FRANCE TELECOM R&D
GLOBAL VISION SYSTEMS
GOSTAI
IBM
I-KM
IObjects
I-NOVA

INOVIA
INTELLITECH
KLS OPTIM
KOALOG
KXEN
KYNOGON
MASA Group
MONDECA
NORMIND
ONTOLOGOS Corp
OSLO
PACTE NOVATION
PERTIMM
PERTINENCE
PROBAYES
RENAULT
ROBOSOFT
ROSTUDEL
SEMANIA
SKYRECON
SOLLAN
Spir. Ops.
SYNAPSE
SWORD
TREE LOGIC
TRUE WORLD
VECSYS
VIRTU'OZ

Adhésion individuelle et abonnement☐ **Demande**☐ **Renouvellement**

Nom : _____ Prénom : _____
 Affiliation : _____
 Adresse postale : _____
 N° de téléphone : _____ N° de télécopie : _____
 Adresse électronique : _____
 Activité (à titre professionnel / à titre privé (*rayez la mention inutile*)) : _____

Type d'adhésion	Bulletin électronique	+ Dossiers imprimés
☐ Simple :	30 €	40 €
☐ Étudiant (sur justificatif) :	15 €	20 €
☐ Soutient :	Sans objet	90 €
☐ Bulletin sans adhésion :	Sans objet	55 €

- ☐ Adhésion au collège IAD-SMA : ajouter 7,5 € pour les étudiants, 15 € pour les autres
☐ Adhésion au collège *Cafe (Apprentissage)* : ajouter 7,5 € pour les étudiants, 15 € pour les autres

Adhésion personne morale☐ **Demande**☐ **Renouvellement****Organisme :****Adresse postale commune aux bénéficiaires couverts par cette adhésion :**

Nom et prénom du représentant : _____ Fonction : _____
 Mél : _____ Tél : _____ Fax : _____
 Adresse postale : _____

Le tarif d'adhésion comprend une partie fixe et une partie par bénéficiaire.

Coordonnées des bénéficiaires (10 maximum) :

Nom, prénom	Mél.	Tél.	Fax

	Tarif de base fixe :	Tarif par bénéficiaire :
☐ Laboratoire universitaires	150 €	Gratuit limité à 5
☐ Personnes morales non universitaires	450 €	Gratuit
☐ Adhésion de soutien	600 €	Sans objet

☐ j'accepte que les renseignements ci-dessus apparaissent dans l'annuaire de l'AFIA☐ j'accepte que les renseignements ci-dessus soient transmis à l'ECCAI pour constituer un fichier européen**Veillez trouver un règlement (à l'ordre de l'AFIA) de Euros****Trésorier AFIA :** Marie-Carol LOPES, LRI, Bâtiment 490, Université Paris-Sud, 91405 Orsay, France.**Mode d'adhésion :**De préférence, en ligne via le site Internet de l'AFIA : <http://www.afia-france.org>

A défaut, cette page doit être envoyée au trésorier.

Modes de paiement :

1. par chèque, à l'ordre de l'AFIA, envoyé au trésorier ;
2. par bon de commande administratif, à l'ordre de l'AFIA, envoyé au trésorier ;
3. par virement bancaire sur le compte de l'AFIA : Société Générale, 4 T rue de Paris, 91400 Orsay, France. Code banque 30003, code guichet 01902, numéro de compte 00037283856 clef RIB 39.

SOMMAIRE DU BULLETIN N° 67

Editorial	3
Dossier SMA et Transport	4
Dossier I.A. et Bioinformatique	31
Les dossiers du bulletin de l'AFIA	56
Entreprises actives en IA	58

CALENDRIER DE PARUTION DU BULLETIN DE L'AFIA

<i>Hiver</i>	<i>Été</i>
Réception des contributions: 15 décembre	Réception des contributions: 15 juin
Sortie le 31 janvier	Sortie le 31 juillet
<i>Printemps</i>	<i>Automne</i>
Réception des contributions: 15 mars	Réception des contributions: 15 septembre
Sortie le 30 avril	Sortie le 31 octobre