

BULLETIN

DE

L'AFIA



AVRIL 2007

N° 63

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)169 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://afia.lri.fr>**Adhésions, Liens avec les adhérents****Marie-Carol LOPES**LRI, Bâtiment 490
Université Paris-Sud
91405 ORSAY

Mél. : Marie-Carol.Lopes@lri.fr

Parrainage de manifestations**Marc-Philippe HUGET**ESIA-LISTIC
Université de Savoie
B.P. 806

74016 Annecy cedex

Marc-philippe.Huget@univ-savoie.fr

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 FRANCAIS DU PÉTROLE,
PEUGEOT S.A., Université de Savoie, Université PARIS 9 DAUPHINE.

Bureau de l'AFIA

Michèle SEBAG, président**Serge DUPUY, secrétaire****Marie-Carol LOPES, trésorier****Comité :**

Emmanuel d'ANGELO, Jean-Paul BAQUIAST, Jacques BLANC-TALON,
Gilles BISSON, Laurent CANDILLIER, Alain CARDON, Amal EL FALLAH,
Jérôme EUZENAT, Olivier FRANCOIS, Christophe JACQUEMIN, Nicolas
LACHICHE, Vincent LEMAIRE, Patrick MARTY, Eunika MERCIER-LAURENT,
Philippe MORIGNOT, Jean-Denis MULLER, Gérald PETITJEAN, Sylvie
PESTY, Michèle SEBAG, Marc SCHOENAUER, Olivier TEYTAUD, Fabien
TORRE, Samuel WIECZOREK, Jean-Daniel ZUCKER.

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****Rubriques « Sommaires des revues »****et « petites annonces »**

LIMSI — CNRS

B.P. 133, 91403 ORSAY Cedex

grau@limsi.fr**Marc-Philippe HUGET****Rubrique****« Conférences et comptes rendus »**

ESIA-LISTIC

Université de Savoie

B.P. 806

74016 Annecy cedex

Marc-philippe.Huget@univ-savoie.fr**Amedeo NAPOLI****Rubrique « livres »**

LORIA

B.P. 239

F-54 506 Vandœuvre lès Nancy

Amedeo.Napoli@loria.fr**Nicole TOURIGNY****Rubrique « IA au Québec »**

Université Laval - Pavillon Adrien-Pouliot

Québec, Canada, G1K 7P4

tourigny@ift.ulaval.ca

Réa. Numérocopie – 01 46 31 64 53

ISSN 1273-1323

Dépôt légal à parution

L'ordinateur pour décrypter le vivant

Les données que fournissent le décodage du génome sont massives, l'ordinateur est requis pour les traiter et l'Intelligence Artificielle a son mot à dire. Ce numéro de printemps du Bulletin est consacré à la bioinformatique, avec un dossier préparé par François Coste (IRISA) et Christine Froidevaux (LRI).

Jérôme Pierson (Orange R&D) nous propose un compte-rendu de ECAI'06, tenue à Riva del Garda en Italie, du 28 août au 1er septembre dernier.

Ensuite, le cinquantenaire de l'I.A., organisé par Eunika Mercier-Laurent (IAE Lyon 3), a été fêté le 3 novembre dernier au Ministère de la Recherche à Paris. L'amphithéâtre Poincaré était comble, les personnes intéressées par les démonstrations étaient nombreuses, et les discussions autour d'un verre riches et sympathiques. Nous espérons avoir fêté dignement cet événement (un compte-rendu se trouve sur le site de l'AFIA <http://afia.lri.fr>) et vous donnons rendez-vous ... pour le centenaire ! (ou, plus sérieusement, pour le soixantenaire).

Nous renouvelons encore et toujours notre appel à contribution : si vous, cher lecteur, avez un sujet que vous aimeriez faire partager à la communauté française d'I.A. (chants d'amour ou coups de gueule, ou plus sérieusement recension de livres que vous avez aimés, compte-rendus de conférences que vous avez aimées, etc), n'hésitez pas à nous faire parvenir vos textes, nous serions ravis de les publier.

Philippe MORIGNOT

Rédacteur en chef du Bulletin de l'AFIA

Intelligence Artificielle et Bioinformatique

Coordonneurs

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

Introduction

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.prd.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'europpéenne.

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 bio-informatique.

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 médicamenteux : 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.

a) 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 œuvre 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%. Nous développons actuellement une méthode d'apprentissage par agrégation de décision de plusieurs experts. Ce travail est effectué en collaboration avec le service de bioinformatique du centre de biotechnologie de Sfax (Tunisie).

b) 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 F., 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 particulier 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 clés orientant la suite de la procédure, analyse les données pour les corriger ou en déduire de nouvelles informations. (exemple : partitionnement multi-critères 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 Dobrokhotov, chercheur post doctoral, Julie Chabalière, 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:

- (i) 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.
- (ii) 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. National 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:

(i) 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.

(ii) 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. Chabalière). 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.

Chabalière 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 Vandœuvre-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 Guerneur, 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/~guerneur/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 Décrypton <http://www.decrypton.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-méta-

stases", 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

Guerneur 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.

Guerneur 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., Guerneur 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&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

1) 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 tech-

nique 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.

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&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 œuvre, 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).

2) 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'amarrage (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 IMPBio, <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://](http://www.lri.fr/bio/soft/bioguide/bioguide.html)

www.lri.fr/bio/soft/bioguide/bioguide.html

BioRelief : [http://](http://www.lri.fr/~chris/bioinfo/BioRelief)

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. Rouveïrol 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. Rouveïrol, 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&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 nœuds 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 classeur 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 synexprimés en contruisant 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 classeur associé.

Fusion de données et prédiction :

L'une des directions de recherche les plus actives en apprentissage automatique est celle de la combinaison de classeurs dont le but est d'améliorer le pouvoir prédictif. Par ailleurs, des études très récentes ont démontré l'intérêt de combiner des classeurs 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 classeurs pour augmenter le pouvoir prédictif de leur combinaison est un enjeu scientifique clef. L'approche 2KC que nous avons mis en œuvre 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 combinaison 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). 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 multi-couches.

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 Wiczorek (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 (ChémoGénomique 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 multiparamétriques 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 ACCAMBA (<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 toxoplasmose). 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 "génétique 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

1) 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.

2) 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 œuvre 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 façon 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 GF, 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.

Adresse : 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: 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 (génétique, 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. Choisine, 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 & C. A. Boucher, *Genome sequence of the plant pathogen *Ralstonia solanacearum**, *Nature*, 415, 497-502 (2002).

Laboratoire de Biologie Virtuelle

Adresse : UMR 6543 CNRS & Université de Nice (UNSA), Centre de Biochimie, Faculté des Sciences, 06108 Nice cedex 2
Site Web : <http://bioinfo.unice.fr> et <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'œil 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 partagées.

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 F., 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, Palaïos 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, Basavanappa Tallur.

Thème général de l'équipe

Symbiose est un projet de recherche en bio-informatique, 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 macromolécules, 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 macromolécules. 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 également 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 macros-molécules

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 fonctionnelle), 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. Laguarriague,
Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks,
Biosystems, 84, 2006
- O. Radulescu, S. Laguarriague, 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 malades). 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 œuvre 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% des cas » (règle d'association) ou enfin « L'ensemble des gènes 1, 3, 78, 349, 554 est le plus grand ensemble des gènes qui sont toujours sur-exprimés simultanément dans l'ensemble des situations 9, 10 et 11 » (concept formel ou association d'ensembles fermés liés par la connexion de Galois). D'autres codages de propriétés ont été étudiés avec, par exemple, l'enregistrement de la relation booléenne entre des motifs sur les séquences promotrices des gènes (typiquement, des sites d'accrochage de facteurs de transcription) et les gènes où ils apparaissent (Besson et al. 2005).

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 approximation 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

Plateforme Biomathématiques et génomique de Paris 5

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 CISMéF, 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 Gendraul-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 trans-ferts 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/migfr/presentation/publication/>)

Abdel Vetah M. O., Aubin S., Bessières P., Bisson G., Hamon T., Lagarrigue S., Manine A. P., Nazarenko A., Nédellec 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), Genève, 2004.

Cussens, J. and Nédellec, 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/lll/lll05/>.

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.

Nédellec, 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., Nédellec, C. and Boulicaut, J. (2004).

Fouille de données et analyse du transcriptome, Chapitre 4.

101-141. traité IC2. Hermès Lavoisier.

Robin, S., Rodolphe, F. and Schbath, S. (2005). *DNA, Words and Models.*

Cambridge University Press, English version of ADN, mots et modèles, BELIN 2003.

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 email au rédacteur en chef du Bulletin (voir page 2).

PRÉSENTATION DE LABORATOIRES

Présentations de laboratoires dans les bulletins de l'AFIA

LIFIA, Grenoble	Bulletin n°1	LAAS - Toulouse (1).....	Bulletin n°30
LRI, Orsay	Bulletin n°1	Sony CSL.....	Bulletin n°31
Service Systèmes Experts, Renault	Bulletin n°1	LAAS - Toulouse (2).....	Bulletin n°32
CEDIAG,	Bulletin n°2	LIMSI - Département CHM.....	Bulletin n°33
CERT, ONERA, Toulouse	Bulletin n°2	LAMSADE.....	Bulletin n°34
IRIT, Toulouse	Bulletin n°2	Institut autrichien de recherches en IA.....	Bulletin n°36
LAAS, Toulouse	Bulletin n°2	LIP6 – Université Pierre et Marie Curie.....	Bulletins n°37 & 38
HEUDIASYC, UTC	Bulletin n°3	GREYC – Université de Caen.....	Bulletin n°40
IFP, Rueil Malmaison	Bulletin n°3	LIFL – Université de Lille.....	Bulletin n°41
DIAM, INSERM U194.....	Bulletin n°3	LRI (équipes IA et IASI).....	Bulletin n°43
Lab. Math. Info., Fac Médecine de Marseille	Bulletin n°4	IMAG - Grenoble	Bulletin n°44
GMD, St. Augustin (RFA).....	Bulletin n°4	PSI (Perception, Système, Information - Rouen).....	Bulletin n°45
ONERA, Chatillon.....	Bulletin n°4	INRIA – Sophia Antipolis	Bulletin n°46/47
KSL, Université de Stanford (USA).....	Bulletin n°5	LIH – Laboratoire d'Informatique du Havre	Bulletin n°46/47
Dépt Applications de l'IA au CNET, Lannion ...	Bulletin n°5	Tech-CICO – Université de Technologie de Troyes	Bulletin n°51
LAFORIA, Univ. Pierre et Marie Curie	Bulletin n°6	LIFO – Université d'Orléans – Équipe Contraintes et Apprentissage.....	Bulletin n°52
L'institut FAW, ULM (RFA).....	Bulletin n°6	LIIA – Ecole Nationale des Arts et Industries de Strasbourg.....	Bulletin n°52
Institut IIIA, Compiègne.....	Bulletin n°6	LRL – Laboratoire de recherche sur le langage – Université Blaise Pascal Clermont 2.....	Bulletin n°53
LAIR, OHIO State University (USA)	Bulletin n°7	MIG – Mathématique, Informatique et Génome – INRA	Bulletin n°53
ARAMIHS, Labo mixte MATRA-CNRS,	Bulletin n°7	ESIEA Recherche, Laval et Paris	Bulletin n°54
CEA, Service SERMA, Saclay.....	Bulletin n°8	Équipe Intelligence Artificielle et Applications (IAA) du Crip5, Paris 5.....	Bulletin n°55
Société ILOG	Bulletin n°8	LAMIH – Univ. de Valenciennes et du Hainaut-Cambrésis	Bulletin n°55
LAIAC, Université de Caen.....	Bulletin n°9	Université de Lund	Bulletin n°58
Institut Français du Pétrole.....	Bulletin n°10	CRIL – Centre de Recherche en Informatique de Lens – Université d'Artois	Bulletin n°58
DFKI (Centre allemand de recherches en IA)..	Bulletin n°11	LIRIS – Laboratoire d'Informatique en Image et Systèmes d'Information	Bulletin n°60
GRTC, Marseille.....	Bulletin n°11	KIN – Knowledge Innovation Center.....	Bulletin n°60
Inst. d'Analyse des Systèmes, Ac. Russe	Bulletin n°12	LIP6 – Laboratoire d'informatique de l'université Paris 6.....	Bulletin n°62
Georges Mason Univ., Center for AI (USA)....	Bulletin n°13		
IRISA, INRIA et Université de Rennes	Bulletin n°13		
Société INGENIA.....	Bulletin n°14		
LIPN, Université de Paris Nord.....	Bulletin n°14		
Institut EURISCO.....	Bulletin n°15		
LRDC, Université de Pittsburgh (USA)	Bulletin n°15		
Société ISOFT	Bulletin n°16		
Dépt. d'Info de l'Université d'Ottawa.....	Bulletin n°16		
Equipe CHM, Université du Colorado (USA) .	Bulletin n°17		
LIRMM, Montpellier	Bulletin n°19		
Institut autrichien de recherches en I.A.....	Bulletin n°20		
ENST Bretagne.....	Bulletin n°21		
LIA - Université de Savoie	Bulletin n°22		
INRETS	Bulletin n°23		
IRIN Nantes.....	Bulletin n°24		
CRIN - INRIA Lorraine	Bulletin n°25		
DIRO - Université de Montréal	Bulletin n°26		
IRIT - Toulouse (1).....	Bulletin n°28		
IRIT - Toulouse (2).....	Bulletin n°29		



ECAI 2006

17ème Conférence Européenne sur l'Intelligence Artificielle, Riva del Garda (Italie), du 28 Août au 1er Septembre

<http://ecai2006.itc.it/>

*Compte-rendu rédigé par :
Jérôme Pierson (Orange R&D).*

À une petite cinquantaine de kilomètre au sud de Trento, là où les contreforts des Alpes italiennes bordent le Lac de Garde, à Riva del Garda, s'est tenue la 17e conférence européenne sur l'intelligence artificielle (ECAI). Les présidents du comité d'organisation, Anna Perini et Paolo Traverso, accueillirent plus de 800 participants (dont plus d'un tiers d'étudiants) qui ont assisté du 28 août au 1er septembre à la présentation de 131 papiers et de 75 posters. Pour le président du comité de programme, Gerhard Brewka, et son équipe, il aura fallu les sélectionner parmi 550 soumissions (le taux de sélection est de 26,1%) provenant de 43 pays différents, la France étant le plus important contributeur avec 111 soumissions pour 49 papiers acceptés. Parmi tous les thèmes abordés, les plus représentés sont la représentation de connaissances et le raisonnement, l'intelligence artificielle distribuée et ses agents, suivis de près par l'apprentissage. Ils ont été développés pendant les trois jours et demi de conférence au rythme de sept sessions parallèles de trois papiers trois fois par jour, auxquels il convient d'ajouter quatre orateurs invités, deux sessions de posters et démos et deux jours de workshop et tutoriaux en début de semaine. Pour finir sur une très bonne note cette introduction, le prix du meilleur papier a été décerné pour l'article : "**A Real generalization of discrete AdaBoost**", issu d'une collaboration entre **Richard Nock** de l'Université d'Antilles-Guyane et **Franck Nielsen**, un doctorant français des laboratoires Sony à Tokyo.

Ma première journée à ECAI 2006 était la plus importante car je présentais mes travaux de thèse lors du workshop Context and Ontologies : Theory, Practice and Applications (28 soumissions, 9 papiers acceptés) organisé par Pavel Shvaiko, Jérôme Euzenat, Alain Léger, Deborah L. McGuinness et Holger Wache. Plutôt que de vous parler de ma présentation

très moyenne et qui, aux vues des nombreuses questions, manquaient cruellement de détails techniques; j'évoquerais la présentation invitée de **Frank van Harmelen** de l'Université Libre d'Amsterdam. Il nous sensibilise à l'importance que peut avoir le contexte quand on recherche un alignement entre deux ontologies. Plus précisément, il utilise les connaissances fondamentales d'un domaine ("background knowledge") connues par deux agents pour chercher les correspondances entre ces connaissances et l'ontologie source, puis entre ces connaissances et l'ontologie cible; pour trouver les correspondances entre la source et la cible. Puis il présenta un ensemble d'expérience pour quantifier l'apport qualitatif et quantitatif de la prise en compte de ce "background knowledge" dans la recherche de correspondances. Reste à savoir la ou les relations entre "background knowledge" et le contexte, pour peu qu'on ait une définition commune du contexte.

Ce lundi soir se déroulait un exposé invité intitulé "Gender & Science" pour nous sensibiliser à la différence de représentation entre hommes et femmes dans le milieu scientifique et plus généralement dans la société. Cet exposé a été suivi d'une pièce de théâtre : "Mister Buffo" de Dario Fo.

En continuant à m'intéresser au même domaine, j'assistais le mardi au workshop Context Representation & Reasoning organisé par Chiara Ghidini, Luciano Serafini et Paolo Bouquet. La matinée a été consacrée à des exposés traitant exclusivement de représentation de connaissances et notamment de la recherche de correspondances entre ontologies. Les exposés de l'après-midi se sont plus attardés à la notion de contexte dans divers domaine et pour diverses applications. De cette journée, j'ai retenu l'exposé de **Stefano Spaccapietra**

issue des bases de données. Au début de son exposé, **"What would you like to drink?"**, il affirme que le contexte détermine un ensemble de réponses et qu'il est nécessaire à la compréhension de la question. À partir de là, il propose un modèle de base de données où les objets réels peuvent avoir une représentation informatique pour chaque contexte. Il introduit le concept de Multi-context Database.

C'est le mardi en fin d'après-midi que la conférence ECAI a officiellement débuté avec une cérémonie d'ouverture, la remise des prix et le premier exposé invité de la conférence de **Fausto Giunchiglia** intitulé **"Managing diversity in knowledge"**.

La première journée pleine de conférence commença avec l'exposé de **Cynthia Breazeal** du MIT Media Lab intitulé **"Socially Intelligent Robots"**. Elle présente un projet de robot spécialisé dans l'interaction avec l'humain. Elle détaille notamment plusieurs techniques d'apprentissage utilisées pour le même robot : apprentissage des interactions sociales, imitations des expressions du visage, reconnaissance d'une voix, connaissance et reconnaissance de nouveaux objets et de leurs caractéristiques, apprentissage d'une série d'actions. L'effort d'intégrer toutes ces technologies dans un même robot lui confère des possibilités d'interaction riches avec l'humain : communication, jeux, guidage, aide.

Après cet exposé, j'assistais à la session "systèmes multi-agents" où la présentation **"Emergence control in cognitive and autonomous agent systems"** de **Caroline Chapinaud, Amal El Fallah Seghrouchni et Patrick Tailibert** a suscité de nombreuses questions. Il y est question d'éviter l'apparition de comportements non souhaités dans un système où évoluent des agents autonomes et donc par définition non-contrôlables. Les lois de l'environnement sont modélisées dans les agents par des réseaux de petri, les interactions entre les lois sont réglées en fusionnant plusieurs réseaux et les agents peuvent collaborer à l'aide de ces réseaux de petri.

La session suivante était consacrée à la créativité et l'intelligence artificielle. Deux présentations m'ont particulièrement intéressé. La première s'intitule **"Cognitive Situated Agents Learn to Name Action"** par **Julien Poudade, Lionel Landwerlin et Paroubek Patrick**. C'est un travail que j'avais découvert à RFIA 2006. Les auteurs exposent un mécanisme de jeux de langages par lequel des agents cognitifs situés apprennent à désigner de nouveaux objets. On observe l'apparition d'un lexique commun pour désigner ces objets. En ajoutant une mémoire de travail aux agents, on réduit considérablement les ambiguïtés. Enfin, l'expérience est enrichie en ajoutant des actions et leur description aux jeux de langage. La deuxième s'intitule **"An architecture for Interactive Musical Agents"** par **David Murray-Rust, Alan Smaill et Michael Edwards**.

Ils présentent un système multi-agent où chaque agent représente un musicien avec un instrument de musique différent, un style et un répertoire de phrases musicales. De plus, ils définissent des interactions entre les agents comme des "music acts" s'inspirant des "speech acts". C'était une présentation que j'ai trouvée très intéressante, et un bon point d'entrée vers une problématique qui personnellement m'intéresse tout particulièrement.

Pour la dernière session d'exposé de la journée, j'assistais à la session "Neuroscience/Human experimentation". Ce fut l'occasion de revoir un autre travail vu à RFIA : **"A similarity and fuzzy logic based approach to cerebral categorisation"** par **Julien Erny, Henri Prade et Josette Pastor**.

Ils proposent un modèle formel pour le processus de catégorisation en essayant de simuler les mécanismes neuronaux du cerveau humain. Le formalisme s'appuie sur un calcul de similarité incluant des règles de pondération et des techniques de fusions empruntées à la logique floue.

Pour finir cette première journée, une visite des stands de démos et des posters accompagnés d'un énième expresso à l'italienne et de deux trois pâtisseries bien choisies s'imposent. L'occasion de revoir un ancien camarade de promotion du DEA 217 de Dauphine, **Mahdi Zargayouna**, qui défend son poster : **"Time Constrained VRP: An Agent Environment-Perception Model"**.

La collation qui s'en suivit au parmesan et jambon de pays généreusement arrosés a permis de faire la démonstration de la convivialité de nos voisins transalpins.

Le jeudi commença par la présentation de **Wolfgang Wahlster** de l'université de la Saare : **"SmartWeb : Getting Answers on the Go"**. Ces travaux s'appuient sur les avancées dans les technologies du web sémantique, dans la construction de web services "sémantique" et autonomes, et dans l'extraction de connaissances. L'infrastructure SmartWeb offre un accès au web avec une interface multimodale et sensible au contexte d'utilisation.

De la première session de l'après-midi consacrée à l'argumentation, j'ai retenu la présentation d'**Isabelle Bloch**: **"Mediation in the Framework of Morphologic"**, travaux réalisés avec **Ramon Pino-Perez et Carlos Uzcategui**. Elle présente un opérateur pour réaliser la médiation entre deux ensembles de préférences, croyances ou besoins.

Pour la dernière session de la journée, j'assiste à une session PAIS (Prestigious Applications of Intelligent Systems). **Sylvain Castagnos** nous présente son travail de thèse supervisé par **Anne Boyer** : **"A Client/Server User-Based Collaborative Filtering Algorithm : Model and Implementation"**. Ils présentent un modèle de recherche

CONFÉRENCES

d'information par filtrage collaboratif décentralisé et une implémentation dans une application de multidiffusion de vidéos.

La dernière journée commença par une présentation d'**Hector Levesque** intitulée "**The truth about defaults**" dans laquelle l'auteur montre comment les logiques de défauts et les logiques autoépistémiques peuvent être vues comme les instances d'un même cadre.

Pour ma dernière session d'ECAI 2006, j'assistais à la session enchères et théorie des jeux. Je fus particulièrement intéressé par la dernière présentation de cette dernière session : "**Cheating is not playing**" par **Bruno Beaufils et Philippe Mathieu**. Ils nous expliquent que lors des compétitions servant à évaluer des algorithmes relatifs à la théorie des jeux, la plupart des évaluations étant faites autour du dilemme du prisonnier, certains biais peuvent être introduit et fausser les

résultats et l'évaluation. Ils qualifient ces biais de tricherie sans communication et de tricherie avec communication (implicite).

Le Bulletin ne pouvant dépêcher 7 reporters sur les lieux, vous comprendrez que cette vue de l'ECAI est une vue partielle dirigée par mes centres d'intérêt. Vous pourrez toutefois retrouver l'intégralité des communications dans les actes, ainsi que de nombreuses photos de la conférence sur <http://ecai2006.itc.it/>.

J'espère avoir montré que RFIA est la porte ouverte à l'ECAI, donc soumettez massivement à ces deux conférences soutenues par l'AFIA !

Pour synthétiser, ECAI 2006 fut très réussie et s'est déroulée dans une atmosphère décontracté mais studieuse.

Préparez-vous d'ores et déjà pour la prochaine édition qui aura lieu à Patras (Grèce) en 2008.

Vos cours sur le PortAI

Les enseignants et chercheurs sont invités à mettre leur cours à disposition des utilisateurs du PortAI :

<http://afia.lri.fr>

Si vous êtes membre de l'AFIA, cela consiste à mettre un fichier (par exemple, ppt) dans votre espace « valise » et à mettre un lien hypertexte dessus à partir de votre page personnelle. Et signalez nous par email votre contribution, pour que nous regroupions tous les cours depuis une page du PortAI.

Evaluation d'une mesure de similitude en classification supervisée : application à la préparation de données séquentielles

Sylvain Ferrandiz
Thèse de Doctorat

Soutenu le 23 octobre 2006 à l'ENST, Paris.

Jury : Samy Bengio, Senior Researcher, IDIAP ; Marc Boulle, Ingénieur, FTR&D ; Bruno Cremilleux, Professeur Université de Caen ; François Kauffmann, Maître de Conférence, Université de Caen ; Yves Lechevallier, Directeur de Recherche, INRIA Rocquencourt ; Marc Sebban, Professeur, Université de Saint-Etienne.

Résumé : En phase de préparation d'un processus de fouille de données, une part importante du travail est consacrée à la construction et à la sélection des variables descriptives. L'approche filtre univariée usuellement adoptée nécessite l'emploi d'une méthode d'évaluation d'une variable. Nous considérons la question de l'évaluation supervisée d'une variable séquentielle. Pour résoudre ce problème, nous montrons qu'il suffit de résoudre un problème plus général : celui de l'évaluation supervisée d'une mesure de similitude. Nous proposons ensuite une telle méthode d'évaluation. Pour l'obtenir, nous formulons le problème en un problème de recherche d'une bonne partition de Voronoi. Nous proposons un nouveau critère d'évaluation supervisée de ces partitions et une nouvelle heuristique de recherche optimisée. Le critère prévient automatiquement le risque de sur-apprentissage et l'heuristique trouve rapidement une bonne solution. Au final, la méthode réalise une estimation non paramétrique robuste de la densité d'une variable cible catégorielle conditionnellement à une mesure de similitude définie à partir d'une variable.

La méthode a été testée sur de nombreux jeux de données. Son utilisation permet de répondre à des questions comme : quel jour de la semaine ou quelle tranche horaire sur la semaine discrimine le mieux le segment auquel appartient un foyer à partir de sa consommation téléphonique fixe? Quelle série de mesures permet de quantifier au mieux l'appétence à un nouveau service?

Mots-clés : Apprentissage - Exploration de Données - Statistique Bayésienne - Analyse Discriminante.

Equipes d'accueil : a. URD Traitement Statistique de l'Information du laboratoire SUSI du CRD Technologies de FTR&D ; b. Equipe Données, Documents, Langues du GREYC, laboratoire d'informatique de l'Université de Caen

Cartographie sémantique : des connaissances à la carte

Christophe Tricot
Thèse de Doctorat

Soutenu le 24 novembre 2006 à l'Université de Savoie, Chambéry.

Jury : Jean-Paul Barthès ; André Manificat ; Guy Melançon ; Christophe Roche (Directeur de thèse); Xavier Briffault ; Daniel Beauchêne ; Anne Dourgnon-Hanoune.

Résumé : Cette étude a débuté avec un constat important : les organisations sont submergées par le nombre toujours croissant d'informations qu'il est nécessaire de maîtriser pour mener à bien leurs activités. En effet, étant obligées de faire face à la mondialisation des flux et des services, les organisations ont découvert de nouveaux besoins et elles ont développé de nouvelles activités : veille stratégique, gestion de l'innovation, capitalisation des savoirs, gestion du capital intellectuel et humain, intelligence économique, etc. Depuis, elles collectent et stockent toujours plus d'informations et leur espace informationnel ne cesse de grandir. Dans ce contexte mondialisé, l'avenir des organisations passe par la maîtrise de leur espace informationnel.

La maîtrise d'un espace est une problématique universelle qui préoccupe l'homme depuis son origine. La cartographie résulte de cette problématique ; elle permet de visualiser un espace pour mieux l'appréhender. Chaque carte joue alors le rôle de support externe de la pensée pour amplifier la cognition des personnes qui l'utilisent.

L'étude de la cartographie nous a appris que pour construire des cartes à partir de l'espace informationnel des organisations, il est nécessaire de mettre en œuvre des paradigmes de représentation, de visualisation et d'interaction. Cette « recette » est maintenant bien connue et maîtrisée. Cependant elle ne répond que partiellement à nos besoins. En effet, il subsiste deux problématiques à résoudre pour y parvenir : comment proposer une cartographie basée sur la sémantique du domaine ? comment déterminer les paradigmes à mettre en œuvre pour fournir des cartes adaptées à l'utilisateur ?

Face à cette double problématique, nous avons défini une nouvelle cartographie capable de construire des cartes adaptées aux besoins des organisations. Nous la nommons « cartographie sémantique ». Elle repose sur nos différentes propositions et sur leurs réalisations. La mise en œuvre de la cartographie sémantique dans le monde des entreprises nous a permis de valider l'ensemble de l'étude mais surtout de s'assurer que la problématique était bien résolue.

Mots-clés : Cartographie d'informations et de connaissances, visualisation d'information, processus de cartographie, ontologies, représentation des connaissances, intelligence artificielle, gestion et recherche d'informations graphiques, architectures à agents.

Laboratoire de recherche : LISTIC – Équipe Condillac
<http://ontology.univ-savoie.fr>

Optimisation et évaluation de performance pour l'aide à la conception et à l'administration des entrepôts de données complexes

**Jérôme Darmont, Maître de conférences
en informatique à l'Université Lumière
Lyon 2
HDR**

Jury : Prof. Danielle Boulanger, Université Jean Moulin Lyon 3 (rapporteur) ; Prof. Claude Chrisment, Université Paul Sabatier Toulouse III (rapporteur) ; Prof. Georges Gardarin, Université de Versailles Saint-Quentin (rapporteur) ; Prof. Stefano Spaccapietra, École Polytechnique Fédérale de Lausanne (examinateur) ; Prof. Abdelkader Zighed, Université Lumière Lyon 2 (coordinateur).

Résumé : Les entrepôts de données forment le socle des systèmes décisionnels. Ils permettent d'intégrer les données de production d'une entreprise ou d'un organisme et sont le support de l'analyse multidimensionnelle en ligne (OLAP) ou de la fouille de données. Avec l'exploitation de plus en plus courante de données complexes dans le cadre des processus décisionnels, de nouvelles approches d'entreposage, qui exploitent notamment le langage XML, sont développées. Dans ce contexte, le problème de la performance des entrepôts de données demeure plus que jamais un enjeu crucial.

Le travail présenté dans ce mémoire vise à proposer des solutions innovantes au niveau de l'optimisation et de l'évaluation des performances des entrepôts de données. Nous avons en effet conçu une approche générique dont l'objectif est de proposer automatiquement à l'administrateur d'un entrepôt des solutions permettant d'optimiser les temps d'accès aux données. Le principe de cette approche est d'appliquer des techniques de fouille de données sur une charge (ensemble de requêtes) représentative de l'utilisation de l'entrepôt de données afin de déduire une configuration quasi-optimale d'index et/ou de vues matérialisées. Des modèles de coût permettent ensuite de sélectionner parmi ces structures de données les plus efficaces en terme de rapport gain de performance/surcharge.

Par ailleurs, l'évaluation de performance peut venir en appui de la conception des entrepôts de données. Ainsi, afin de valider notre approche de manière expérimentale, nous avons également conçu plusieurs bancs d'essais génériques. Le principe directeur qui a présidé à leur élaboration est l'adaptabilité. En effet, pour comparer l'efficacité de différentes techniques d'optimisation des performances, il est nécessaire de les tester dans différents environnements, sur différentes configurations de bases de données et de charges, etc. La possibilité d'évaluer l'impact de différents choix d'architecture est aussi une aide appréciable dans la conception des entrepôts de données. Nos bancs d'essais permettent donc de générer diverses configurations d'entrepôts de données, ainsi que des charges décisionnelles qui s'y appliquent.

Finalement, nos solutions d'optimisation et d'évaluation des performances ont été mises en œuvre dans les contextes des entrepôts de données relationnels et XML.

Mots clés : Bancs d'essais, Données complexes, Entrepôts de données, Evaluation de performance, Fouille de données, Index, OLAP, Optimisation de performance, Vues matérialisées, XML.

Réseaux Bayésiens : apprentissage et modélisation de systèmes complexes.

**Philippe Leray, Maître de Conférences
à l'INSA Rouen.
HDR**

**Soutenu le 28 novembre 2006 à l'Université
de Rouen, Saint-Etienne du Rouvray.**

Jury : P. Aknin, Directeur de Recherche, LTN, INRETS ; S. Benferhat, Professeur des Universités, CRIL, Université d'Artois, Rapporteur ; S. Canu, Professeur des Universités, LITIS, INSA de Rouen ; B. Dubuisson, Professeur des Universités, Heudiasyc, Université Technologique de Compiègne, Rapporteur ; I. Guyon, Ingénieur Conseil, ClopiNet, Berkeley, Rapporteur ; L. Heutte, Professeur des Universités, LITIS, Université de Rouen.

Résumé : La représentation des connaissances et le raisonnement à partir de ces représentations a donné naissance à de nombreux modèles. Les modèles graphiques probabilistes, et plus précisément les réseaux bayésiens, initiés par Judea Pearl dans les années 1980, se sont révélés des outils très pratiques pour la représentation de connaissances incertaines, et le raisonnement à partir d'informations incomplètes.

Dans certains problèmes, l'expert est souvent amené à construire lui-même le graphe du réseau bayésien, en réfléchissant en terme de causalité. A l'opposé, l'apprentissage du graphe à partir de données se fait dans un cadre plus général que celui des réseaux bayésiens causaux, cadre dans lequel plusieurs graphes seront équivalents, mais où un seul capturera éventuellement les relations de causalité du problème.

Depuis 1990, les travaux de recherche dans le domaine se sont essentiellement intéressés à l'apprentissage des réseaux bayésiens lorsque toutes les variables sont connues (pas de variables latentes), et lorsque ces variables sont complètement observées (pas de données

manquantes), problème lui aussi NP-complet. La prise en compte de données incomplètes, comme la découverte de variables latentes, posent encore de sérieux défis en terme de complexité.

La découverte de réseaux bayésiens complètement causaux à partir de données est une question qui a été abordée plus récemment. Les travaux sur le sujet s'accordent sur le fait qu'il est impossible de travailler à partir de données d'observations uniquement. Les plans d'expériences, c'est à dire la façon dont les données ont été obtenues, sont des informations essentielles pour capturer la notion de causalité. Les méthodes proposées dans le cadre de l'apprentissage de structure ont transformé les réseaux bayésiens non seulement en outil de représentation des connaissances, mais aussi en outil de découverte de celles-ci.

L'essentiel de nos travaux de recherche concerne l'apprentissage de structure d'un réseau bayésien, en levant progressivement les différentes hypothèses posées :

- pas de variables latentes, données complètes
- pas de variables latentes, données incomplètes
- comment découvrir des variables latentes ?
- comment arriver à une structure réellement causale ?

Réseaux Bayésiens : de l'identification de structure à la reconnaissance de formes à partir d'informations complètes ou incomplètes

**Olivier François
Thèse de Doctorat**

**Soutenu le 28 novembre 2006 à l'Université de
Rouen, Saint-Etienne du Rouvray.**

Jury : A. Aussem, Professeur des Universités, PRISMa, Université Lyon 1 ; S. Benferhat, Professeur des Universités, CRIL, Université d'Artois, Rapporteur ; S. Canu, Professeur des Universités, LITIS, INSA de Rouen, Directeur ; J.Y. Jaffray, Professeur des Universités, LIP6, Université Paris 6 ; P. Leray, Maître de Conférences, LITIS, INSA de Rouen, Encadrant ; E. Mazer, Directeur de Recherche, CNRS, ProBayes, Rapporteur ; T. Paquet, Professeur des Universités, LITIS, Université de Rouen.

Résumé : Durant ces travaux de thèse, nous avons effectué une comparaison empirique de différentes techniques d'apprentissage de structure. En effet, même s'il existe des comparaisons ponctuelles, il n'y a pas de comparaison plus globale de ces algorithmes. Cette phase de tests nous a permis d'identifier quelles méthodes souffraient d'un problème d'initialisation. Nous avons alors proposé une technique d'initialisation pour ces dernières. Nous avons ensuite adapté différentes méthodes d'apprentissage de structure aux bases de données incomplètes. Nous avons notamment introduit une technique pour apprendre une structure arborescente efficacement. Cette méthode est ensuite adaptée au problème plus spécifique de la classification. Nous avons également introduit un formalisme permettant de générer des bases de données incomplètes ayant des données manquantes MCAR ou MAR. De nombreuses bases synthétiques ou réelles ont été utilisées pour tester ces méthodes d'apprentissage de structure à partir de bases incomplètes.

Web Sémantique et Informatique Linguistique: propositions méthodologiques et réalisation d'une plateforme logicielle

Florence Amardeilh
Thèse de Doctorat

**Soutenu le 10 mai 2007 à l'Université Paris 10,
Nanterre.**

Jury : Mme Nathalie Aussenac-Gilles, Chargée de recherche (HDR) au CNRS, Rapporteur ;
Mr Gilles Kassel, Professeur d'université, Université de Picardie, Rapporteur ;
Mr Benoît Habert, Professeur d'université, Université Paris X-Nanterre, Examineur ;
Mme Maria-Teresa Pazienza, Professeur d'université, Università di Roma Tor Vergata, Examineur ;
Mr Philippe Laublet, Maître de conférences, Université Paris IV-Sorbonne, Co-Directeur de Thèse ;
Mr Jean-Luc Minel, Ingénieur de recherche (HDR), CNRS, Co-Directeur de Thèse ;
Mr Jean Delahousse, PDG de Mondeca.

Résumé : Cette thèse aborde les problématiques liées à l'annotation sémantique et au peuplement

d'ontologies dans le cadre défini par le Web Sémantique (WS). La vision du Web Sémantique initiée en 1998 par Sir Tim Berners-Lee a pour objectif de permettre une meilleure exploitation des informations disponibles sur le Web par les agents logiciels. Pour cela, les ressources, textuelles ou multimédias, doivent être sémantiquement étiquetées par des annotations structurées. Dans ce processus d'annotation sémantique, les ontologies jouent un rôle primordial puisqu'elles modélisent les concepts, attributs et relations utilisées pour annoter le contenu des ressources. Mais, il est tout aussi important que la base de connaissance, associée à cette ontologie, contienne les instances à utiliser pour l'annotation sémantique. C'est pourquoi la tâche de peuplement d'ontologie a pour but d'enrichir (semi-)automatiquement la base de connaissance avec les nouvelles instances de concepts, d'attributs et de relations.

La réalisation de ces deux tâches consiste à combiner les outils d'extraction d'information (EI) avec les outils de représentation des connaissances du WS. En effet, le principal mode de transfert de la connaissance se fait par l'utilisation du langage naturel dans les ressources documentaires. Malgré tout, il existe actuellement un fossé entre les formats de représentation des analyses linguistiques et ceux de représentation des connaissances. Cette thèse propose de combler ce fossé en concevant un médiateur capable de transformer les étiquettes linguistiques générées par les outils d'EI en des représentations plus formelles, annotations sémantiques des textes ou instances d'une ontologie donnée et relations entre celles-ci. L'enjeu consiste aussi bien à proposer une réflexion méthodologique sur l'interopérabilité des différentes technologies qu'une conception de solutions opérationnelles dans le monde des entreprises, et à plus large échelle du Web.

Dans le cadre de cette thèse, nous avons donc conçu une démarche, nommée OntoPop pour «Ontology Population», qui met en place une passerelle reposant sur un ensemble de règles, dites «d'Acquisition de Connaissance» et sur un langage d'implémentation de ces règles, OPAL (Ontology Population and Annotation Language). Nous montrons comment cette passerelle peut être utilisée dans un cycle complet d'extraction d'information, d'enrichissement des ressources terminologiques et ontologiques, d'annotation sémantique et de mise à jour des lexiques utilisés par l'outil

d'EI. L'accent est porté sur la résolution des problèmes soulevés par un tel cycle de vie, notamment à propos de la consolidation des nouvelles annotations et instances vis-à-vis du modèle de l'ontologie. Enfin, nous soumettons des propositions pour l'opérationnalisation de la démarche OntoPop à travers une méthodologie et une plateforme logicielle basée sur l'outil de représentation des connaissances ITM de la société Mondeca. La méthodologie a pour objectif de fournir un mode d'emploi simple et efficace pour la réalisation d'une application concrète d'annotation sémantique ou de peuplement d'ontologie au sein d'une entreprise. La plateforme logicielle offre des exemples de composants logiciels modulaires, autorisant un maximum de flexibilité vis-à-vis des besoins et objectifs de chaque nouvelle application d'annotation sémantique ou de peuplement d'ontologie.

Construire une ontologie de la pneumologie : aspects théoriques, modèles et expérimentations

Audrey Baneyx
Thèse de Doctorat

Soutenue le 6 février 2007 à l'INSERM, Paris.

Jury : Madame Nathalie AUSSENAC-GILLES (Rapporteur, IRIT-CNRS Toulouse) ; Monsieur Stefan Darmoni (Rapporteur, Pr INSA Rouen) ; Madame Anne Doucet (Examinateur, Professeur Univ. Paris 6) ; Madame Marie-Christine Jaulent (Examinateur, INSERM UMR_S 872) Monsieur Bruno Bachimont (Examinateur, Pr UTC) ; Monsieur Jean Charlet (Dir. thèse, AP-HP, INSERM UMR_S 872).

Résumé : Depuis une vingtaine d'années, l'accès aux connaissances médicales est un enjeu majeur pour les professions de santé comme pour le grand public. Les limites actuelles des outils de traitement de l'information ne proviennent pas de leurs performances pour stocker et traiter rapidement des gros volumes, mais de leur incapacité à prendre en compte les spécificités des vocabulaires métier des utilisateurs. Le développement de ressources terminologiques et ontologiques pour faciliter l'usage des terminologies nationales et internationales, disponibles notamment dans le domaine de la médecine, revêt par conséquent

une importance particulière. Il faut également souligner la pertinence de telles recherches dans la mouvance des Sciences et Technologies de l'Information et de la Communication, dans le cadre de la société de l'information et dans le contexte du Web sémantique.

Dans ce contexte, notre réflexion a porté sur la collecte, l'organisation, la représentation et la formalisation des connaissances en médecine, tout particulièrement, dans le domaine de la pneumologie. Nous avons été amenés à considérer le problème dans son ensemble, afin de comprendre les mécanismes qui sous-tendent la constitution de ressources terminologiques et ontologiques à partir de textes. Nous avons également considéré chaque tâche séparément, afin de proposer, pour chaque étape, si ce n'est une solution, au moins un savoir-faire personnel susceptible d'apporter des éléments de réponse. L'objectif principal de cette thèse consiste à mettre au point une ontologie dans le domaine de la pneumologie pour faciliter, d'une part, l'aide au codage médico-économique des pathologies et, d'autre part, la représentation des connaissances pertinentes relatives au patient, dans ce domaine de spécialité. Nos recherches couvrent l'ensemble du cycle de vie d'une ontologie, de la mise au point d'une méthodologie de construction à partir de textes à son utilisation dans un système opérationnel. Nous contribuons aux recherches dans les domaines de l'Ingénierie des connaissances et de l'Informatique médicale.

La méthode de travail adoptée est une démarche expérimentale ascendante qui consiste à partir des problématiques concrètes rencontrées pour aller vers la résolution des questions scientifiques sous-jacentes. Selon cette démarche, nous avons tout d'abord cerné les besoins des pneumologues en termes de représentation des connaissances. Ensuite, nous avons mis au point une méthodologie, destinée à l'ingénieur des connaissances, fondée sur la méthode ARCHONTE définie par B. Bachimont. L'enchaînement des processus d'extraction, de sélection et de choix des candidates termes du domaine ainsi que l'aide fournie par les patrons lexico-syntaxiques pour renseigner les principes différentiels la rende relativement facile d'emploi (ou moins difficile qu'une autre) pour un ingénieur des connaissances. L'ontologie construite compte à ce jour 2 260 concepts primitifs. Enfin, nous avons développé un outil de codage semi-automatique

proposant deux types de codages : (1) un codage médical qui représente graphiquement les informations pertinentes relatives aux pathologies du patient et qui, à terme, servira de descripteur pour indexer intelligemment les comptes rendus d'hospitalisation ; (2) un codage médico-économique pour lequel nous obtenons un rappel de 80% et une précision de 87%. Nos résultats concernant l'ontologie et l'outil nous encouragent à poursuivre nos recherches et à améliorer les solutions proposées.

Mots-clés : ontologie, thésaurus, aide au codage, sémantique différentielle, Ingénierie des connaissances, Informatique médicale, Pneumologie.

Apprentissage pour le traitement de données séquentielles : Application à la modélisation d'utilisation et aux interfaces stylo

Thierry Artières
HDR

**Soutenu le 14 Décembre 2006 au Laboratoire
d'Informatique de Paris 6 (LIP6), Université Pierre
et Marie Curie, Paris, France.**

Jury : Gérard Chollet, Examineur ; Bernadette Dorizzi , Rapporteur ; Patrick Gallinari, Examineur ; Eric Gaussier, Rapporteur ; Yann LeCun, Rapporteur ; Maurice Milgram, Examineur.

Résumé : Le traitement de données séquentielles est au cœur de nombreux domaines de recherche, comme le traitement du signal, les statistiques, l'économétrie, l'intelligence artificielle, la reconnaissance des formes. Aujourd'hui le domaine de l'apprentissage automatique en général dispose de nombreux modèles de séquence plus ou moins adaptés à différents types de données et différentes tâches. Une famille de modèle s'est distinguée depuis une vingtaine d'années, il s'agit de la famille des modèles Markoviens. Ces modèles constituent un bon compromis entre richesse d'expression et facilité de mise en œuvre si bien que le cadre Markovien se révèle un cadre assez pratique pour concevoir de nouveaux modèles. Nous présentons ici nos travaux dans le domaine du traitement de

séquences, réalisés sur divers types de données parole, écrit, gestes, traces d'interactions) et de tâches (classification, clustering, segmentation, interprétation, diagnostic). Les contributions concernent d'une part la proposition de nouveaux modèles de séquences adaptés à divers types de données, d'autre part l'apprentissage de ces modèles et notamment de leurs structures.

Nous nous sommes intéressés à une famille de modèles Markoviens en particulier appelés modèles segmentaux ou modèles de trajectoires car ils opèrent sur des segments d'observations plutôt que sur des observations isolées. Nous présenterons tout d'abord les travaux que nous avons réalisés dans la proposition, la mise au point, l'analyse de divers types de modèles segmentaux, des modèles Markoviens multi-streams, des modèles neuro-prédicatifs et des modèles de trajectoires non linéaires. Nous discuterons également de modèles génératifs permettant de prendre en compte encore plus de dépendances entre observations. Egalement, nous présenterons des approches discriminantes obtenues soit par ajout d'information discriminante dans les systèmes génératifs soit par exploitation de modèles naturellement discriminants tels que les champs Markoviens aléatoires.

Nous discuterons ensuite de méthodes permettant de réaliser l'apprentissage de la structure et des paramètres de modèles Markoviens au sein d'une famille restreinte mais suffisante pour la plupart des domaines d'applications que nous avons rencontrés. Il s'agit de mélanges de modèles Markoviens à structure contrainte gauche-droite. L'apprentissage de ces modèles est très lié à la tâche de partitionnement de séquences guidé par les modèles, et nous présenterons nos travaux à la lumière des travaux réalisés dans ce domaine. Nous discuterons également de l'apprentissage de modèles hiérarchiques et présenterons différents types de modèles hiérarchiques, des cascades de modèles aux modèles véritablement hiérarchiques.

Enfin, nous présenterons des travaux réalisés applicatifs en modélisation d'utilisateurs et en reconnaissance de gestes écrits en ligne.

Analyse et extraction de connaissances des bases de données spatiotemporelles

Karine Zeitouni
HDR

**Soutenue le 1er décembre 2006 au laboratoire
PRISM, Versailles.**

Jury : Jiawei Han, Professeur à l'Université de l'Illinois, rapporteur ; Robert Laurini, Professeur à l'INSA de Lyon, rapporteur ; Philippe Rigaux, Professeur à l'Université de Paris Dauphine, rapporteur ; Georges Gardarin, Professeur à l'Université de Versailles-St-Quentin ; Philippe Pucheral, Professeur à l'Université de Versailles-St-Quentin ; Michel Scholl, Professeur au CNAM ; Stefano Spaccapietra, Professeur à l'EPFL.

Résumé : Ces dernières années ont vu une croissance phénoménale dans la production et la diffusion des données spatiales de sources aussi variées qu'hétérogènes. Cela a généré des besoins d'intégration dans des entrepôts de données et des perspectives d'analyse exploratoire et de fouille de données spatiales et spatiotemporelles. Nos travaux se placent dans ce contexte visant l'analyse et l'extraction des connaissances depuis les bases de données spatiotemporelles. Ils traitent différents aspects allant de la modélisation avancée des données spatiales, à la fouille de ces données en passant par leur intégration dans un entrepôt, l'optimisation des requêtes et l'analyse en ligne. Ainsi, nous décrivons nos approches pour la modélisation 3D, puis pour la modélisation spatiotemporelle d'objets mobiles. Ensuite, l'intégration de données spatiales est traitée selon deux aspects : l'intégration de formats et l'intégration de données par l'appariement géométrique. Une architecture d'entrepôt de données spatiales basée sur les standards XML et GML est proposée, puis dotée d'une technique d'optimisation de requêtes spatiales basée sur un cache sémantique.

L'exploration des données spatiotemporelles a donné lieu à des solutions originales extension de l'OLAP. Enfin, différentes approches sont proposées pour la fouille de données spatiales. Nous avons ouvert le spectre de nos recherches à la fouille d'autres données complexes, telles que les données séquentielles et textuelles. Ces travaux ainsi que les développements futurs sont exposés dans ce mémoire.

Résumé Automatique de Parole pour un Accès Efficace aux Bases de données Audio

Benoit Favre
Thèse de Doctorat

Soutenue le 19 mars 2007 à l'IUP d'Avignon.

Jury : Guy Lapalme, Professeur, RALI, Montréal, rapporteur ; François Yvon, Maître de Conférences, ENST, Paris, rapporteur ; Patrick Gallinari, Professeur, LIP6, Paris ; Catherine Berrut, Professeur, LIG, Grenoble ; François Capman, Ingénieur, Thales, Colombes ; Jean-François Bonastre, Maître de Conférences, LIA, Avignon, Directeur de Thèse ; Patrice Bellot, Maître de Conférences, LIA, Avignon.

Résumé : L'avènement du numérique permet de stocker de grandes quantités de parole à moindre coût. Malgré les récentes avancées en recherche documentaire audio, il reste difficile d'exploiter les documents à cause du temps nécessaire pour les écouter. Nous tentons d'atténuer cet inconvénient en produisant un résumé automatique parlé à partir des informations les plus importantes. Pour y parvenir, une méthode de résumé par extraction est appliquée au contenu parlé, transcrit et structuré automatiquement. La transcription enrichie est réalisée grâce aux outils Speeral et Alize développés au LIA. Nous complétons cette chaîne de structuration par une segmentation en phrases et une détection des entités nommées, deux caractéristiques importantes pour le résumé par extraction. La méthode de résumé proposée prend en compte les contraintes imposées par des données audio et par des interactions avec l'utilisateur. De plus, cette méthode intègre une projection dans un espace pseudo-sémantique des phrases. Les différents modules mis en place aboutissent à un démonstrateur complet facilitant l'étude des interactions avec l'utilisateur. En l'absence de données d'évaluation sur la parole, la méthode de résumé est évaluée sur le texte lors de la campagne DUC 2006. Nous simulons l'impact d'un contenu parlé en dégradant artificiellement les données de cette même campagne. Enfin, l'ensemble de la chaîne de traitement est mise en oeuvre au sein d'un démonstrateur facilitant l'accès aux émissions radiophoniques de la campagne ESTER. Nous proposons, dans le cadre de ce démonstrateur, une frise chronologique interactive complémentaire au résumé parlé.

Conception d'un modèle Web sémantique appliqué à la génomique fonctionnelle

Fleur Mougin
Thèse de Doctorat

Soutenue le 1er décembre 2006 au CHU de Rennes.

Jury : Mr Pierre LE BEUX, Directeur ; Mr Olivier LOREAL, Co-directeur ; Mr Philippe BESSIERES, Rapporteur ; Mme Christine FROIDEVAUX, Rapporteur ; Mr Dominique LAVENIER, Examineur ; Mme Anita BURGUN, Membre invité.

Résumé : Ce travail de thèse s'inscrit dans la problématique du Web sémantique et plus précisément l'utilisation de ses technologies dans le domaine de la génomique fonctionnelle. Au cours de leurs travaux de recherche, les biologistes et médecins doivent disposer de données concernant l'existant et ces informations sont généralement accessibles sur Internet. Le problème est qu'elles sont réparties dans des sources distribuées, autonomes et hétérogènes à de multiples niveaux.

C'est dans ce cadre que nous proposons le développement d'un système d'intégration basée médiateur. Celui-ci vise à offrir aux biologistes et médecins une interface unique permettant d'accéder aux différentes sources de manière centralisée et homogène. Des systèmes suivant cette approche existent dans la littérature mais présentent des limites en terme de conception et d'évolution. En effet, de nombreuses tâches restent manuelles, ce qui les rend particulièrement fastidieuses.

Pour répondre à ces limites, nous avons conçu un système médiateur au moyen de méthodes semi-automatiques. Tout d'abord, nous avons développé une méthode d'acquisition automatique des schémas des sources de données à intégrer. Ensuite, nous avons utilisé une ressource terminologique existante, l'UMLS, afin de définir un schéma global. Enfin, nous avons proposé différentes approches pour mettre en correspondance les éléments des schémas locaux avec ceux du schéma global. Celles-ci sont situées aux niveaux schéma et instances et sont automatiques, avec une possible validation manuelle par des experts.

Le système implémenté à partir de ces méthodes est opérationnel et accessible sur Internet et permet de

traiter des requêtes simples qui impliquent plusieurs sources de données. Le processus de requêtes effectuée une expansion de celles-ci en exploitant la hiérarchie des concepts identifiés comme pertinents, et rend finalement aux utilisateurs les valeurs des éléments des sources associés à ces concepts. La maintenance du système est facilitée par les méthodes développées pour sa conception, permettant ainsi de la gérer de manière semi-automatique.

Mots clés : bioinformatique, Web sémantique, représentation des connaissances, bio-ontologies, système d'intégration, hétérogénéités syntaxiques et sémantiques, mises en correspondance de schémas.

Utilisation de normes et de réputations pour détecter et sanctionner les contradictions - Contribution au contrôle social des interactions dans les systèmes multi-agents ouverts et décentralisés

Guillaume Muller
Thèse de Doctorat

Soutenue le 11 décembre 2006 à l'École des Mines de Saint-Étienne.

Jury : Yves Demazeau (institut IMAG, Grenoble), rapporteur ; Carles Sierra (Université Autonome de Barcelone), rapporteur ; Olivier Boissier (ÉNS des Mines de Saint-Étienne), co-directeur ; Laurent Vercoeur (ÉNS des Mines de Saint-Étienne), co-directeur ; Salima Hassas (Université Claude Bernard, Lyon) ; Andreas Herzig (Université Paul Sabatier, Toulouse) ; Juliette Rouchier (Université de la Méditerranée, Marseille).

Résumé : Les Systèmes Multi-Agents Ouverts et Décentralisés (SMAOD) sont particulièrement vulnérables à l'introduction d'agents mal conçus ou malveillants. Il est donc nécessaire de contrôler ces systèmes.

Dans cette thèse, nous proposons le modèle LIAR, permettant aux agents eux-mêmes de mettre en place un contrôle des interactions des autres agents, à l'aide d'un modèle de réputation.

Ce modèle permet d'abord aux agents de représenter les interactions qu'ils perçoivent grâce à des engagements sociaux, ainsi que de modéliser les règles que chaque agent doit respecter à l'aide de normes sociales. En comparant les comportements qu'ils ont observés aux normes dont ils ont connaissance, les agents sont capables d'évaluer leurs pairs et d'estimer les niveaux de réputation qu'ils leur associent. Ensuite, les agents peuvent décider des sanctions à appliquer en s'appuyant sur les niveaux de réputation ainsi estimés.

Grâce à l'intégration des deux phases : évaluation des comportements et décision des sanctions à appliquer, le modèle LIAR permet de mettre en place un contrôle social des interactions entièrement automatisé.

Diverses expérimentations ont été menées avec ce modèle dans le cadre d'un réseau pair-à-pair, afin de montrer comment les agents contrôlent les interactions de leurs pairs.

Méthodologie, formalismes et outils de modélisation-simulation des systèmes complexes de mobiles : application à la mobilité géographique

Nicolas Marilleau
Thèse de Doctorat

Soutenue le 15 novembre 2006 à l'UFR des Sciences et Techniques, Besançon.

Jury : Alexis Drogoul, Directeur de recherche, IRD, Bondy, rapporteur ; Jacques Ferber, Professeur, LIRMM, Montpellier, rapporteur ; Christophe Claramunt, Professeur, IRENav, Brest ; Olga Kouchnarenko, Professeur, LIFC, Besançon ; Laurent Philippe, Professeur, LIFC, Besançon, co-directeur ; Christophe Lang, Maître de conférences, LIFC, Besançon ; Pascal Chatonnay, Maître de conférences, LIFC, Besançon.

Résumé : Les principaux apports de cette thèse sont dans le domaine de l'informatique avec des passerelles vers la géographie. La modélisation-simulation de la mobilité est au coeur du problème de cette thèse. L'objectif principal de ce travail est de concevoir une méthodologie destinée à représenter et à simuler les

mobilités qui s'opèrent sur un territoire quelconque. Nous l'avons appelé RAFALE-SP.

Cette méthode permet, de manière progressive, selon un méta-modèle, la modélisation d'un espace et des mobiles qui évoluent dessus. Les modèles de mobilité ainsi générés sont ensuite intégrés au sein de simulateurs distribués. Leurs implémentations sont facilitées grâce à l'utilisation d'une bibliothèque multi-agents prévue à cet effet.

Afin de valider notre méthodologie, nous l'appliquons dans le cadre de divers projets de recherche notamment le projet MIRO visant à simuler la ville en mouvement (travail effectué dans le cadre du programme national PREDIT).

Présentation Multimodale de l'Information

Yacine Bellik
Thèse de Doctorat

Soutenue le 14 décembre au LIMSI-CNRS, Orsay.

Jury : COUTAZ Joëlle (Rapporteur), Professeur, CLIPS-IMAG, Université Joseph Fourier, Grenoble ; BEAUDOUIN-LAFON Michel (Examinateur), Professeur, LRI, Université d'Orsay Paris Sud XI ; MARIANI Joseph (Examinateur), Directeur de Recherches, LIMSI-CNRS, Orsay ; NIGAY Laurence (Examinateur), Professeur, CLIPS-IMAG, Université Joseph Fourier, Grenoble ; PALANQUE Philippe (Rapporteur), Professeur, IRIT, Université Paul Sabatier, Toulouse ; VANDERDONCKT Jean (Rapporteur), Professeur, BCHI, Université Catholique de Louvain, Belgique.

Résumé : Nos travaux prennent place dans le domaine de l'interaction Homme-Machine et plus particulièrement dans celui de l'interaction multimodale. Ils concernent l'analyse, la conception, le développement et l'évaluation de systèmes interactifs disposant de différents moyens d'interaction et sortant du cadre classique des interfaces graphiques. Dans ce contexte, nous nous intéressons plus spécifiquement à la présentation multimodale de l'information. Les récentes évolutions technologiques et sociales (avènement de la société de l'information, mobilité des utilisateurs en

situation d'interaction, apparition de l'informatique ambiante,...) créent de nouveaux besoins en termes d'interaction. La diversité des utilisateurs, des systèmes et des environnements fait qu'il n'est plus possible de continuer à proposer des interfaces au comportement prédéfini et figé. Face à ce problème nous adoptons une approche qui consiste à intégrer dans le système interactif des mécanismes d'adaptation qui lui permettent de modifier dynamiquement son comportement pour être en adéquation avec un contexte d'interaction en perpétuelle évolution. Nous envisageons l'adaptation sous l'angle de la multimodalité. La diversité des interactions qu'offrent les interfaces multimodales, leur flexibilité et leur caractère intuitif et naturel, les rendent aptes à cibler différentes catégories d'utilisateurs. Ces propriétés leur confèrent d'importantes capacités à s'accommoder des fréquentes modifications que peut subir l'environnement physique de l'interaction et à exploiter de façon optimale les ressources physiques des systèmes qui les accueillent.

C'est pourquoi nous pensons que la multimodalité est particulièrement intéressante à exploiter dans le cadre d'un contexte sujet à de fréquentes évolutions. Nous cherchons dans nos travaux à élaborer des modèles et outils logiciels qui permettent d'exploiter de façon optimale toutes les modalités dont peut disposer le système à un instant donné, pour communiquer une information à l'utilisateur.

Nous déclinons nos travaux selon trois axes : la présentation contextuelle de l'information, la présentation opportuniste de l'information dans l'ambient et enfin la présentation non visuelle de l'information.

Contrôleurs neuronaux dynamiques et apprentissage : Du perfectionnement des algorithmes à leur application temps réel sur des systèmes robotiques

Vincent Scesa
Thèse de Doctorat

Soutenue le 20 octobre 2006, à Vélizy.

Jury : Prof. Fethi BenOuedzou, Université de Versailles, Directeur de thèse ; Prof. Gérard Dreyfus, ESPCI, rapporteur ; M. Etienne Dombre, Directeur de recherche CNRS, LIRMM, rapporteur ; M. Faycal

Namoun, Directeur de l'entreprise BIA ; M. Patrick Hénaff, Maître de Conférences, Université de Cergy ; Prof. Claude Timsit, Université de Versailles.

Résumé : Comment contrôler un système dynamique complexe lorsqu'il est soumis à des couplages, qu'il contient des non linéarités, et dont le comportement peut évoluer par son interaction avec l'environnement ? De manière plus générale, comment contrôler un système dont la connaissance est incomplète ou délicate à synthétiser ? Dans ces situations, les contrôleurs basés sur des modèles de connaissances atteignent leurs limites. Une réponse possible à ces questions peut consister à combler les lacunes de la modélisation théorique par une synthèse empirique. L'observation et l'expérience se substituent alors à la connaissance mathématique des phénomènes rencontrés. Les techniques d'apprentissages alliées à des modèles universels peuvent alors apporter une solution à ces difficultés.

Cette thèse, qui s'inscrit dans cette démarche, donne un aperçu complet de la réalisation du contrôle de systèmes dynamiques complexes, depuis le choix, la mise en place, et l'amélioration des algorithmes, jusqu'à leur application sur des systèmes réels, en passant par la description de leur implémentation temps réel. Notre travail a consisté à étudier la mise en œuvre de modèles neuronaux dynamiques, de type Continuous Time Recurrent Neural Networks (CTRNN), et de techniques d'apprentissage basées sur l'algorithme BackPropagation Through Time (BPTT) pour le contrôle temps réel de deux systèmes dynamiques polyarticulés : le simulateur de route BIA et le robot bipède ROBIAN du LISV.

Le premier chapitre de cette thèse présente, tout d'abord, un état de l'art des méthodes de contrôle en proposant une taxinomie de ces techniques. Il expose, ensuite, un aperçu des plateformes bipèdes et des hexapodes hydrauliques existants, et présente en détail celles du LISV et de BIA, utilisées dans cette étude.

OFFRE ADHERENTS

OFFRE PRÉFÉRENTIELLE D'ABONNEMENT RESERVEE AUX MEMBRES DE L'ASSOCIATION FRANCAISE POUR L'INTELLIGENCE ARTIFICIELLE

Offrez-vous chaque mois le meilleur de l'actualité scientifique
Abonnez-vous à *La Recherche* !

OFFRE 1 AN

11 numéros (dont 4 numéros spéciaux)
+ 4 numéros des Dossiers La Recherche
pour seulement **60€** au lieu de 92€¹⁰
soit 32€¹⁰ d'économie

+ RECEVEZ EN CADEAU DE BIENVENUE 2 NUMEROS HORS-SERIE

Au choix : n° 15 *Le Soleil, les secrets d'une étoile* – n° 16 *Les molécules du bonheur*
N°17 *Le risque climatique* – n° 18 *L'héritage Einstein* –
n° 19 *L'Histoire de la vie* – n° 20 *Mathématiques « nouveaux défis et vieux casse tête »*

+ BENEFICIEZ DE L'ACCES GRATUIT AUX ARCHIVES DEPUIS 1996 SUR WWW.LARECHERCHE.FR

OFFRE SPECIALE RESERVEE AUX MEMBRES DE L'ASSOCIATION FRANCAISE DE L'INTELLIGENCE ARTIFICIELLE

A retourner sous enveloppe affranchie à : Service abonnements LA RECHERCHE - B 604 - 60732 Ste-Geneviève cedex

OUI, je souhaite souscrire un abonnement d'1 an à La Recherche : AFIA06
11 n^{os} + 4 n^{os} des Dossiers de La Recherche au tarif préférentiel de **60€** soit **35%** d'économie

Je choisis en cadeau les 2 numéros hors-série suivants :

n° 15 n° 16 n° 17 n° 18 n° 19 n° 20

Je certifie être adhérent et je joins une photocopie de ma carte de membre.

Je joins mon règlement :

par chèque bancaire ou postal à l'ordre de LA RECHERCHE

par carte bancaire dont voici le numéro : |_|_|_| |_|_|_| |_|_|_| |_|_|_|

IMPORTANT : notez aussi les 3 derniers chiffres du numéro inscrit au dos de votre carte bancaire, au
niveau de la signature : |_|_|_| Date d'échéance : |_|_|_|_|

Signature obligatoire :

Voici mes coordonnées :

NOM :PRENOM :

ADRESSE :

CODE POSTAL : |_|_|_|_| VILLE

SOMMAIRES DES REVUES

N'hésitez pas à envoyer un message à Brigitte Grau (grau[at]limsi.fr) pour lui indiquer toute suggestion permettant d'améliorer cette rubrique. Les revues figurant régulièrement au sommaire mais n'ayant pas de nouveau numéro apparaissent seulement avec leur nom et leur adresse WEB.

REVUE D'INTELLIGENCE ARTIFICIELLE

<http://ria.e-revues.com/>

RIA V20 N°2-3 2006

Numéro spécial

Décision et planification dans l'incertain, eds. F. CHARPILLET, F. GARCIA, P. PERNY, O. SIGAUD

Processus décisionnels de Markov décomposés et factorisés pour l'optimisation de stratégies d'exploration, F. TEICHTAIL-KÖNIGSBUCH, P. FABIANI

Recherche en ligne pour les processus décisionnels de Markov, L. PÉRET, F. GARCIA

Prise de décision en temps-réel pour des POMDP de grande taille, S. PAQUET, L. TOBIN, B. CHAIB-DRAA

Exploration guidée en apprentissage par renforcement. Connaissances a priori et relaxation de contraintes, P. GARCIA

Apprentissage par renforcement dans le cadre des processus décisionnels de Markov factorisés observables dans le désordre. Étude expérimentale du Q-Learning parallèle appliqué aux problèmes du labyrinthe et du New York Driving, G. LAURENT, E. PIAT

Étude de différentes combinaisons de comportements adaptatives, O. BUFFET, A. DUTECH, F. CHARPILLET

Algorithmes de co-évolution pour la résolution approchée de PDM multi-agent, I. CHADÈS

Apprentissage de la coordination multiagent. Une méthode basée sur le Q-learning par jeu adaptatif, O. GIES, B. CHAIB-DRAA

Processus de décision markoviens et préférences non classiques, P. WENG

RIA V20 N°4-5 2006

Numéro spécial

Interaction émotionnelle, ed. C. PÉLACHAUD

Représentation et détection des émotions dans des dialogues enregistrés dans un centre d'appel. Des émotions complexes dans des données réelles, L. DEVILLERS, L. VIDRASCU

Du corpus vidéo à l'agent expressif. Utilisation des différents niveaux de représentation multimodale et émotionnelle, J. MARTIN, S. ABRILIAN, L. DEVILLERS, M. LAMOLLE, M. MANCINI, C. PELACHAUD

De E-Wiz à C-Clone. Recueil, modélisation et synthèse d'expressions authentiques, V. AUBERGÉ, N. AUDIBERT, A. RILLIARD

De la construction du corpus émotionnel au système de détection. Le point de vue applicatif de la surveillance dans les lieux publics, C. CLAVEL, I. VASILESCU, G. RICHARD, L. DEVILLERS

Perception d'états affectifs et apprentissage, G. LOOSLI, S. LEE, A. RAKOTOMAMONJY

Le paradigme MAUI pour des agents multimodaux d'interface homme-machine socialement intelligents, C. LISETTI

Expressions intelligentes des émotions, M. OCHS, R. NIEWIADOMSKI, C. PELACHAUD, D. SADEK

Conception et évaluation d'un modèle d'expressivité pour les gestes des agents conversationnels, S. BUISINE, B. HARTMANN, M. MANCINI, C. PELACHAUD

Développement et évaluation de jeux dialogiques émotionnels. Application à des personnes autistes, O. GRYNSZPAN, J. MARTIN, J. NADEL

RIA V20 N°6 2006

Numéro spécial

New Methods in Machine Learning. Theory and Applications, eds. F. DENIS, L. RALAIVOLA

Asymmetrical Semi-Supervised Learning and Prediction of Disulfide Connectivity in Proteins, C. MAGNAN

A Bayesian Approach for the Clustering of Short Time Series, L. BRÉHÉLIN

Bayesian Networks: a Non-Frequentist Approach for Parametrization, and a more Accurate Structural Complexity Measure Bayesian Networks Learning, S. GELLY, O. TEYTAUD

Kernel Basis Pursuit, V. GUIGUE, A. RAKOTOMAMONJY, S. CANU

Learning Recursive Automata from Positive Examples, I. TELLIER

Universal Consistency and Bloat in GP Some theoretical considerations about Genetic Programming from a Statistical Learning Theory viewpoint, S. GELLY, O. TEYTAUD, N. BREDECHE, M. SCHOENAUER

Processus décisionnels de Markov décomposés et factorisés pour l'optimisation de stratégies d'exploration, F. TEICHTAIL-KÖNIGSBUCH, P. FABIANI

SOMMAIRES DES REVUES

ARTIFICIAL INTELLIGENCE

<http://www.elsevier.nl/inca/publications/store/5/0/5/6/0/1/>

AI V170 N°12-13 SEPTEMBER 2006

Planning as satisfiability: parallel plans and algorithms for plan search, J. RINTANEN, K. HELJANKO, I. NIEMELÄ

AI V170 N°14-15 OCTOBER 2006

Sequential inference with reliable observations: Learning to construct force-dynamic models, A. FERN, R. GIVAN
Corrective feedback and persistent learning for information extraction, A. CULOTTA, T. KRISTJANSSON, A. MCCALLUM, P. VIOLA

AI V170 N°16-17 NOVEMBER 2006

Maximizing over multiple pattern databases speeds up heuristic search, R. C. HOLTE, A. FELNER, J. NEWTON, R. MESHULAM, D. FURCY
Controlled generation of hard and easy Bayesian networks: Impact on maximal clique size in tree clustering, O. J. MENGSHOEL, D. C. WILKINS, D. ROTH
Learning adaptation knowledge to improve case-based reasoning, S. CRAW, N. WIRATUNGA, R. C. ROWE

AI V170 N°18 DECEMBER 2006

Special Review Issue, ed. by P. NORVIG, D. PERLIS
Complex systems: Network thinking, M. MITCHELL
Book reviews
Strike while the iron is, M. L. ANDERSON
David Cope, *Computer Models of Musical Creativity*, MIT Press (2005), R. B. DANNENBERG
Marcus Hutter, *Universal Artificial Intelligence*, Springer (2004), T. OATES, W. CHONG
Kurzweil's argument for the success of AI, D. MCDERMOTT
Three books on language: pro-Noam, anti-Noam, net-Noam, W. J. IDSARDI
Jon Doyle, *Extending Mechanics to Minds: The Mechanical Foundations of Psychology and Economics*, Cambridge University Press (2006), T. COSTELLO, J. MCCARTHY
Knowledge representation and commonsense reasoning: Reviews of four books, L. MORGENSTERN

AI V171 N°1 JANUARY 2007

Iterated belief revision, revised, Y. JIN, M. THIELSCHER
A representation theorem for minmax regret policies, S. LI
A generalization of discrete AdaBoost, R. NOCK, F. NIELSEN

Audiences in argumentation frameworks, T. J.M. BENCH-CAPON, S. DOUTRE, P. E. DUNNE

AI MAGAZINE

<http://www.aaai.org/Magazine/magazine.html>

AI MAGAZINE V27 N°3 FALL 2006

Innovative AI Applications: Introduction to This Special Issue, N. JACOBSTEIN, B. PORTER
TEXTAL: Crystallographic Protein Model Building Using AI and Pattern Recognition, K. GOPAL, T. D. ROMO, E. W. MCKEE, R. PAI, J. N. SMITH, J. C. SACCHETTINI, T. R. IOERGER
NESTA: NASA Engineering Shuttle Telemetry Agent, G. S. SEMMEL, S. R. DAVIS, K. W. LEUCHT, D. A. ROWE, K. E. SMITH, R. O'FARRELL, L. BÖLÖNI
Automating the Underwriting of Insurance Applications, K. S. AGGOUR, P. P. BONISSONE, W. E. CHEETHAM, R. P. MESSMER
Intelligent Multiobjective Optimization of Distribution System Operations, R. J. SÁRFI, A. M. G. SOLO
A Multiagent Simulator for Teaching Police Allocation, V. FURTADO, E. VASCONCELOS
Automatically Generating Game Tactics through Evolutionary Learning, M. J. V. PONSEN, H. MUÑOZ-AVILA, P. SPRONCK, D. W. AHA
The AAI 2005 Mobile Robot Competition and Exhibition, P. E. RYBSKI, S. TEJADA, D. BLANK, A. STROUPE, M. BUGAJSKA, L. GREENWALD
Report on the First Conference on Human-Robot Interaction (HRI), M. A. GOODRICH, A. C. SCHULTZ, D. J. BRUEMMER
Modeling Decision for Artificial Intelligence (MDAI 2006), V. TORRA, Y. NARUKAWA

AI MAGAZINE V27 N°4 WINTER 2006

A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, J. MCCARTHY, M. L. MINSKY, N. ROCHESTER, C. E. SHANNON
Happy Silver Anniversary, AI!, E. A. FEIGENBAUM
AI@50: We Are Golden!, A. K. MACKWORTH
(AA)AI – More Than the Sum of Its Parts, R. J. BRACHMAN
What Do We Know About Knowledge?, B. G. BUCHANAN
AI Meets Web 2.0: Building the Web of Tomorrow, TODAY, J. M. TENENBAUM

SOMMAIRES DES REVUES

A Personal Account on the Development of Stanley, the Robot That Won the DARPA Grand Challenge, S. THRUN
Review of Thinking About Android Epistemology, R. MORRIS
The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years, J. MOOR

COGNITIVE SCIENCE

<http://www.leaonline.com/loi/cog>

COGNITIVE SCIENCE V30 N°1 2006

Animal Foraging and the Evolution of Goal-Directed Cognition, T. T. HILLS
Idiomatic Syntactic Constructions and Language Learning, M. P. KASCHAK, J. R. SAFFRAN
The Role of Functionality in the Mental Representations of Engineering Students: Some Differences in the Early Stages of Expertise, J. MOSS, K. KOTOVSKY, J. CAGAN
A Model of Plausibility, L. CONNELL, M. T. KEANE
Modeling Recognition Memory Using the Similarity Structure of Natural Input, J. P. W. LACROIX, J. M. J. MURRE, E. O. POSTMA, H. J. VAN DEN HERIK
Formalization and Analysis of Reasoning by Assumption, T. BOSSE, C. M. JONKER, J. TREUR
Deferred Interpretations: Why Starting Dickens is Taxing but Reading Dickens Isn't, B. MCELREE, S. FRISSON, M. J. PICKERING

COGNITIVE SCIENCE V30 N°2 2006

Processing Reflexives and Pronouns in Picture Noun Phrase, J. T. RUNNER, R. S. SUSSMAN, M. K. TANENHAUS
The Isolation, Primacy, and Recency Effects Predicted by an Adaptive LTD/LTP Threshold in Postsynaptic Cells, S. SIKSTRÖM
Modeling the Development of Children's Use of Optional Infinitives in Dutch and English Using MOSAIC, D. FREUDENTHAL, J. M. PINE, F. GOBET
Aging and the Use of Context in Ambiguity Resolution: Complex Changes From Simple Slowing, K. S. DAGERMAN, M. C. MACDONALD, M. W. HARM
Diversity and Unity of Modularity, B. SEOK
Culture and Change Blindness, T. MASUDA, R. E. NISBETT

COGNITIVE SCIENCE V30 N°3 2006

With the Future Behind Them: Convergent Evidence From Aymara Language and Gesture in the Crosslinguistic Comparison of Spatial Construals of Time, R. E. NÚÑEZ, E. SWEETSER
The Recognition of Phonologically Assimilated Words Does Not Depend on Specific Language Experience, H. MITTERER, V. CSÉPE, F. HONBOLYGO, L. BLOMERT
The Coordinated Interplay of Scene, Utterance, and World Knowledge: Evidence from Eye Tracking, P. KNOEFERLE, M. W. CROCKER
Memory and Mystery: The Cultural Selection of Minimally Counterintuitive Narratives, A. NORENZAYAN, S. ATRAN, J. FAULKNER, M. SCHALLER
A Hierarchical Bayesian Model of Human Decision-Making on an Optimal Stopping Problem, M. D. LEE
Arrows in Comprehending and Producing Mechanical Diagrams, J. HEISER, B. TVERSKY
Is the Brain a Quantum Computer?, A. LITT, C. ELIASMITH, F. W. KROON, S. WEINSTEIN, P. THAGARD

COGNITIVE SCIENCE V30 N°4 2006

Bringing Cognitive Science into Education, and Back Again: The Value of Interdisciplinary Research, D. S. MCNAMARA
Relations, Objects, and the Composition of Analogies, D. GENTNER, K. J. KURTZ
Uncertainty About the Rest of the Sentence, J. HALE
An Embodied Model for Sensorimotor Grounding and Grounding Transfer: Experiments With Epigenetic Robots, A. CANGELOSI, T. RIGA
The Interpretation of Classically Quantified Sentences: A Set-Theoretic Approach, G. POLITZER, J-B. VAN DER HENST, C. DELLE LUCHE, I. A. NOVECK
Unsupervised Discovery of Nonlinear Structure Using Contrastive Backpropagation, G. HINTON, S. OSINDERO, M. WELLING, Y-W. TEH
Perception of Auditory Motion Affects Language Processing, M. P. KASCHAK, R. A. ZWAAN, M. AVEYARD, R. H. YAXLEY
Flexible Conceptual Projection of Time Onto Spatial Frames of Reference, A. TORRALBO, J. SANTIAGO, J. LUPIÁÑEZ
Metacognitive Control and Optimal Learning, L. K. SON, R. SETHI

COGNITIVE SCIENCE V30 N°5 2006

Harmony in Linguistic Cognition, P. SMOLENSKY
Early Child Grammars: Qualitative and Quantitative

SOMMAIRES DES REVUES

Analysis of Morphosyntactic Production, G. LEGENDRE
Phonotactics and Articulatory Coordination Interact in
Phonology: Evidence from Nonnative Production, L. DAVIDSON
Faithful Contrastive Features in Learning, B. TESAR
Dynamics of Phonological Cognition, A. I. GAFOS, S.
BENUS
Learning Phonology With Substantive Bias: An Experimental
and Computational Study of Velar Palatalization, C. WILSON

COGNITIVE SCIENCE V30 N°6 2006

The Import and Export of Cognitive Science, R. L. GOLD-
STONE, L. LEYDESDORFF
How Known Constructions Influence the Acquisition of Other
Constructions: The German Passive and Future Constructions,
K. ABBOT-SMITH, H. BEHRENS
Errors of Omission in English-Speaking Children's
Production of Plurals and the Past Tense: The Effects of
Frequency, Phonology, and Competition, D. E. MATTHEWS,
A. L. THEAKSTON
Pictures and Spoken Descriptions Elicit Similar Eye
Movements During Mental Imagery, Both in Light and in
Complete Darkness, R. JOHANSSON, J. HOLSANOVA,
K. HOLMQVIST
Decision Making and Confidence Given Uncertain Advice,
M. D. LEE, M. J. DRY
The Action-Sentence Compatibility Effect: It's All in the
Timing, K. L. BORREGGINE, M. P. KASCHAK
Phonological Abstraction in the Mental Lexicon, J. M.
MCQUEEN, A. CUTLER, D. NORRIS

APPLIED ARTIFICIAL INTELLIGENCE

<http://www.tandf.co.uk/journals/tfj08839514.html>

APPLIED AI V20 N°9 OCTOBER 2006

Artificial Intelligence and Graph Theory Tools for Describing
Switched Linear Control Systems, A. IBEAS, M. DE LA SEN
Metaheuristics for Handling Time Interval Coverage
Constraints in Nurse Scheduling, E. K. BURKE, P. de
CAUSMAECKER, S. PETROVIC, G. vanden BERGHE
Textual Article Clustering in Newspaper Pages, M. AIELLO,
A. PEGORETTI
Application of Hierarchical Neural Fuzzy Models to
Modeling and Control of a Bioprocess, L. A. C. MELEIRO,
R. J. G. B. CAMPELLO, R. MACIEL FILHO, W. C. AMARAL

APPLIED AI V20 N°10 NOV-DEC 2006

Self-organizing contextualized mobile workforce management
with collaborative art learning, S-T. YUAN, M. WU
Emotional and non-emotional persuasion, M. MICELI, F. DE
ROSIS, I. POGGI
An enhanced ant colony optimization metaheuristic for the
minimum dominating set problem, C. K. HO, Y. P. SINGH, H.
T. EWE

APPLIED INTELLIGENCE

<http://www.kluweronline.com/issn/0924-669X/>

APPLIED INTELLIGENCE V25 N°3 2006

A unified architecture for agent behaviors with selection of
evolved neural network modules, K-J. KIM, S-B. CHO
Granularity and disaggregation in compositional modelling
with applications to ecological systems, J. KEPPENS, Q.
SHEN
A classification method using a hybrid genetic algorithm com-
bined with an adaptive procedure for the pool of ellipsoids, K.
LEE, W. YOON, D. BAEK
Data-driven fuzzy rule generation and its application for
student academic performance evaluation, K. RASMANI,
Q. SHEN
Potential field method to navigate several mobile robots, S.
PRADHAN, D. PARHI, A. PANDA, R. BEHERA
Achieving dynamic, multi-commander, multi-mission
planning and execution, E. SANTOS, S. DELOACH, M. COX
Efficient data reduction in multimedia data, S. WANG, M.
DASH, L-T. CHIA, M. XU

COMPUTATIONAL INTELLIGENCE

<http://www.blackwellpublishing.com/journal.asp?ref=0824-7935&site=1>

COMPUTATIONAL INTELLIGENCE

V22 N°3-4 AUG-NOV 2006

Special issue

Introduction to the special issue on case-based reasoning in
the health sciences, I. BICHINDARITZ, C. MARLING
Memory organization as the missing link between case-based
reasoning and information retrieval in biomedicine, I.
BICHINDARITZ

SOMMAIRES DES REVUES

Adaptation knowledge acquisition: a case study for case-based decision support in oncology, M. D'AQUIN, J. LIEBER, A. NAPOLI

Conceptual clustering and case generalization of two-dimensional forms, S. JÄNICHEN, P. PERNER

Visio-spatial case-based reasoning: a case study in prediction of protein structure, J. DAVIES, J. GLASGOW, T. KUO

Accounting for the temporal dimension in case-based retrieval: a framework for medical applications, S. MONTANI, L. PORTINALE

Predicting influenza waves with health insurance data, R. SCHMIDT, T. WALIGORA, L. GIERL

Case-based reasoning and knowledge discovery in medical applications with time series, P. FUNK, N. XIONG

gene-CBR: a case-based reasoning tool for cancer diagnosis using microarray data sets, F. DIAZ, F. FDEZ-RIVEROLA, J. M. CORCHADO

An evaluation of the usefulness of explanation in a case-based reasoning system for decision support in bronchiolitis treatment, D. DOYLE, P. CUNNINGHAM, P. WALSH

MINDS AND MACHINES

<http://www.springeronline.com/sgw/cda/frontpage/0,0,5-0-70-35534114-0,0.html>

MINDS AND MACHINES V16 N°2 2006

Neural networks discover a near-identity relation to distinguish simple syntactic forms, A. C. BALE, T. R. SHULTZ

In the hands of machines? The future of aged care, R. SPARROW, L. SPARROW

The logic of Searle's Chinese room argument, R. I. DAMPER

Quantity of experience: brain-duplication and degrees of consciousness, N. BOSTROM

The concept of strong and weak virtual reality, A. M. LISEWSKI

Book Review

M. Augier and J. G. March (eds): Models of a Man: Essays in Memory of H. Simon, P. ALLO

D. Gentner and S. Goldin-Meadow (eds): Language in Mind: Advances in the Study of Language and Thought., P. CHANDRA

M. T. Guasti, Language Acquisition: The Growth of Grammar., R. K. MISHRA

MINDS AND MACHINES V16 N°3 2006

Special Issue

Causality, Uncertainty and Ignorance, R. HAENNI, S. HARTMANN

Common causes and the direction of causation, B. WESLAKE

Dispositional versus epistemic causality, J. WILLIAMSON

Hitchcock's (2001) treatment of singular and general causation, C. JAKOB

The power of intervention, K. B. KORB, E. NYBERG

Homogeneity, selection, and the faithfulness condition, D. STEEL

Counterexamples to a likelihood theory of evidence, M. R. FORSTER

Scale construction from a decisional viewpoint, M. SMITHSON

How causal knowledge simplifies decision-making, R. GARCIA-RETAMERO, U. HOFFRAGE

ARTIFICIAL INTELLIGENCE REVIEW

<http://www.kluweronline.com/issn/0269-2821/contents>

INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES

<http://www.academicpress.com/ijhcs>

IJHCS V64 N°12 DECEMBER 2006

The visual characteristics of avatars in computer-mediated communication: Comparison of Internet Relay Chat and Instant Messenger as of 2003, H-S. KANG, H-D. YANG

Assessing facial beauty through proportion analysis by image processing and supervised learning, H. GUNES, M. PICCARDI

A cognitive approach to user perception of multimedia quality: An empirical investigation, S. Y. CHEN, G. GHINEA, R. D. MACREDIE

Knowledge engineering and psychology: Towards a closer relationship, N. MILTON, D. CLARKE, N. SHADBOLT

Matching users' diverse social scripts with resonating humanized features to create a polite interface, J-Y. TZENG

IJHCS V65 N°1 JANUARY 2007

Information security in the knowledge economy, Edited by A. S. VINZE, T.S. RAGHU

Information security in the knowledge economy, A. S. VINZE, T.S. RAGHU

SOMMAIRES DES REVUES

Worm damage minimization in enterprise networks, S. SANGUANPONG, U. KANLAYASIRI

The usability of passphrases for authentication: An empirical field study, M. KEITH, B. SHAO, P. JOHN STEINBART

Investigation of IS professionals' intention to practise secure development of applications, I. M.Y. WOON, A. KANKAN-HALLI

Mapping the contemporary terrorism research domain, E. F. REID, H. CHEN

Mining communities and their relationships in blogs: A study of online hate groups, M. CHAU, J. XU

Analyzing terror campaigns on the internet: Technical sophistication, content richness, and Web interactivity, J. QIN, Y. ZHOU, E. REID, G. LAI, H. CHEN

Stochastic dynamics of music album lifecycle: An analysis of the new market landscape, S. BHATTACHARJEE, R. D. GOPAL, K. LERTWACHARA, J. R. MARSDEN

IJHCS V65 N°2 FEBRUARY 2007

Design and evaluation of visualization support to facilitate decision trees classification, Y. LIU, G. SALVENDY

The media equation and team formation: Further evidence for experience as a moderator, D. JOHNSON, J. GARDNER

CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses, R. MAZZA, V. DIMITROVA

Developing professional skills and social capital through computer supported collaborative learning in university contexts, D. FRANCESCATO, M. MEBANE, R. PORCELLI, C. ATTANASIO, M. PULINO

Knowledge sharing behavior in virtual communities: The relationship between trust, self-efficacy, and outcome expectations, M-H. HSU, T. L. JU, C-H. YEN, C-M. CHANG

A cyclic model of information seeking in hyperlinked environments: The role of goals, self-efficacy, and intrinsic motivation, P. DAVID, M. SONG, A. HAYES, E. S. FREDIN

COMPUTATIONAL LINGUISTICS

<http://mitpress.mit.edu/catalog/item/default.asp?sid=8563C099-9701-4DD2-85C8-8F3502E9C8AE&ttype=4&tid=10>

COMPUTATIONAL LINGUISTICS

V32 N°4 DECEMBER 2006

ACL Lifetime Achievement Award

Old Linguists Never Die, They Only Get Obligarily Deleted, E. HAJI_OVÁ

Automatic Evaluation of Information Ordering: Kendall's Tau, M. LAPATA

Unsupervised Multilingual Sentence Boundary Detection, T. KISS, J. STRUNK

N-gram-based Machine Translation, J. B. MARÍÑO, R. E. BANCHS, J. M. CREGO, A. DE GISPERT, P. LAMBERT, J. A. R. FONOLLOSA, M. R. COSTA-JUSSÀ

Book Reviews

Language, Cohesion and Form, M. Masterman (1910–1986) (ed. by Y. Wilks), Cambridge University Press (Studies in natural language processing, ed. by S. Bird and B. Boguraev), 2005, J. F. SOWA

One-on-One Tutoring by Humans and Computers, M. Evens and J. Michael, Lawrence Erlbaum Associates, 2006, P. JORDAN
Memory-Based Language Processing, W. Daelemans and A. van den Bosch, Cambridge: Cambridge University Press, 2005, S. KÜBLER

TREC: Experiment and Evaluation in Information Retrieval, E. M. Voorhees and D. K. Harman (eds.), Cambridge, MA: The MIT Press (Digital libraries and electronic publishing series, ed. by W. Y. Arms), 2005, N. STOKES

Ontology Learning from Text: Methods, Evaluation and Applications, P. Buitelaar, P. Cimiano and B. Magnini (eds.), Amsterdam: IOS Press (Frontiers in artificial intelligence and applications, ed. by J. Breuker et al., volume 123), 2005, C. BREWSTER

Last Words

Mark-up Barking Up the Wrong Tree, A. ZAENEN

Errata

Errata: Experiments on the Automatic Induction of German Semantic Verb Classes, S. SCHULTE IM WALDE

NATURAL LANGUAGE ENGINEERING

<http://journals.cambridge.org/action/displayJournal?jid=NLE>

NLE V12 N°4 DECEMBER 2006

Reading comprehension tests for computer-based understanding evaluation, B. WELLNER, L. HIRSCHMAN, W. GREIFF, L. FERRO
The importance of the lexicon in tagging biological text, L. H. SMITH, W. J. WILBUR, T. C. RINDFLESCHE

An algorithm for the unsupervised learning of morphology, J. GOLDSMITH

How much can part-of-speech tagging help parsing?, M. DALRYMPLE

USER MODELING AND USER-ADAPTED INTERACTION

<http://www.wkap.nl/jrnltoC.htm/0924-1868>

UMUAI V16 N°3-4 2006

Special issue

Preface to the special issue on user modeling to support groups, communities and collaboration, J. VASSILEVA, A. SOLLER, E. GAUDIOSO

Creating cognitive tutors for collaborative learning: steps toward realization, L. BOLLEN, B. M. MCLAREN, E. WALKER, J. SEWALL, A. HARRER

Modeling individual and collaborative problem-solving in medical problem-based learning, S. SUEBNUKARN, P. HADDAWY
Using shared representations to improve coordination and intent inference, R. ALTERMAN, J. INTRONE

In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems, A. GATT, J. MASTHOFF
Design and evaluation of an adaptive incentive mechanism for sustained educational online communities, R. CHENG, J. VASSILEVA

Coalescing individual and collaborative learning to model user linguistic competences, T. READ, B. BARROS, J. PANCORBO, E. BÁRCENA

The impact of learning styles on student grouping for collaborative learning: a case study, A. ORTIGOSA, E. ALFONSECA, P. PAREDES, E. MARTÍN, R. M. CARRO

UMUAI V16 N°5 2006

Learned student models with item to item knowledge structures, M. C. DESMARAIS, P. MESHKINFAM, M. GAGNON
MASHA: A multi-agent system handling user and device adaptivity of Web sites, D. ROSACI, G. M. L. SARNÉ

COMPUTER SPEECH AND LANGUAGE

http://www.sciencedirect.com/science?_ob=JournalURL&_isn=08852308&_auth=y&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=be00614a6a8826664cf3538182118628

COMPUTER SPEECH & LANGUAGE V21

N°1 JANUARY 2007

A fast and memory-efficient N -gram language model lookup method for large vocabulary continuous speech recognition, X. LI, Y. ZHAO

Evaluating spoken dialogue systems according to de-facto standards: A case study, S. MÖLLER, P. SMEELE, H. BOLAND, J. KREBBER

‘Early recognition’ of polysyllabic words in continuous speech, O. SCHARENBERG, L. TEN BOSCH, L. BOVES

Speaker-adaptive learning of resonance targets in a hidden trajectory model of speech coarticulation, D. YU, L. DENG, A. ACERO

Backoff hierarchical class n -gram language models: effectiveness to model unseen events in speech recognition, I. ZITOUNI

Random forests and the data sparseness problem in language modelling, P. XU, F. JELINEK

Reformulating the HMM as a trajectory model by imposing explicit relationships between static and dynamic feature vector sequences, H. ZEN, K. TOKUDA, T. KITAMURA

Iterative reconstruction of speech from short-time Fourier transform phase and magnitude spectra, L. D. ALSTERIS, K. K. PALIWAL

Features based on filtering and spectral peaks in autocorrelation domain for robust speech recognition, G. FARAHANI, S.M. AHADI, M.M. HOMAYOUNPOUR

A two-pass approach for handling out-of-vocabulary words in a large vocabulary recognition task, O. SCHARENBERG, S. SENEFF, L. BOVES

Automatic pronunciation evaluation of foreign speakers using unknown text, N. MOUSTROUFAS, V. DIGALAKIS

COMPUTER SPEECH & LANGUAGE

V21 N°2 APRIL 2007

Environment adaptation for robust speaker verification by cascading maximum likelihood linear regression and reinforced learning, K.K. YIU, M.W. MAK, S.Y. KUNG

Prior knowledge guided maximum expected likelihood based model selection and adaptation for nonnative speech recognition, X. HE, Y. ZHAO

The common vector approach and its comparison with other subspace methods in case of sufficient data, M. B. GÜLMEZOGLU, V. DZHAFAROV, R. EDIZKAN, A. BARKANA

Modeling durations of syllables using neural networks, K. S. RAO, B. YEGNANARAYANA

SOMMAIRES DES REVUES

Accessing speech data using strategic fixation, S. WHITTAKER, J. HIRSCHBERG

A fuzzy decision tree-based duration model for Standard Yorùbá text-to-speech synthesis, O. A. ODEJOBÍ, S. H. S. WONG, A. J. BEAUMONT

Improving statistical machine translation using shallow linguistic knowledge, Y.-S. HWANG, A. FINCH, Y. SASAKI

Discriminative n-gram language modelling, B. ROARK, M. SARAFLAR, M. COLLINS

Partially observable Markov decision processes for spoken dialog systems, J. D. WILLIAMS, S. YOUNG

MACHINE LEARNING

<http://www.wkap.nl/jrnltoe.htm/0885-6125>

MACHINE LEARNING V65 N°2-3 NOV-DEC 2006

Guest editorial: Machine learning in and for music, G. WIDMER

Melodic analysis with segment classes, D. CONKLIN

Modeling, analyzing, and synthesizing expressive piano performance with graphical models, G. GRINDLAY, D. HELMBOLD

Aligning music audio with symbolic scores using a hybrid graphical model, C. RAPHAEL

A case based approach to expressivity-aware tempo transformation, M. GRACHTEN, J.-L. ARCOS, R. LÓPEZ MÁNTARAS

Classification-based melody transcription, D. P. W. ELLIS, G. E. POLINER

Bootstrap learning for accurate onset detection, N. HU, R. B. DANNENBERG

Aggregate features and ADABOOST for music classification, J. BERGSTRÄ, N. CASAGRANDE, D. ERHAN, D. ECK, B. KÉGL

Using duration models to reduce fragmentation in audio segmentation, S. ABDALLAH, M. SANDLER, C. RHODES, M. CASEY

MACHINE LEARNING V66 N°1 JANUARY 2007

Special issue

Guest editorial to the special issue on grammatical inference, G. PALIOURAS, Y. SAKAKIBARA

LARS: A learning algorithm for rewriting systems, R. EYRAUD, C. HIGUERA, J.-C. JANODET

Interactive learning of node selecting tree transducer, J.

CARME, R. GILLERON, A. LEMAY, J. NIEHREN

Learning finite-state models for machine translation, F. CASACUBERTA, E. VIDAL

Learning deterministic context free grammars: The Omphalos competition, A. CLARK

NEURAL NETWORKS

<http://www.elsevier.com/inca/publications/store/8/4/1/>

NEURAL NETWORKS V19 N°8 OCTOBER 2006

Special issue

Neurobiology of Decision Making, ed. by S. FUNAHASHI, D. LEE, M. RUSHWORTH

Stochastic models of decisions about motion direction: Behavior and physiology, J. DITTERICH

Rapid decision threshold modulation by reward rate in a neural network, P. SIMEN, J. D. COHEN, P. HOLMES

Extended LATER model can account for trial-by-trial variability of both pre- and post-processes, H. NAKAHARA, K. NAKAMURA, O. HIKOSAKA

Building bridges between neural models and complex decision making behaviour, J. R. BUSEMEYER, R. K. JESSUP, J. G. JOHNSON, J. T. TOWNSEND

The time course of saccadic decision making: Dynamic field theory, C. WILIMZIG, S. SCHNEIDER, G. SCHÖNER

Neural mechanism for stochastic behaviour during a competitive game, A. SOLTANI, D. LEE, X.-J. WANG

Computational algorithms and neuronal network models underlying decision processes, Y. SAKAI, H. OKAMOTO, T. FUKAI

Recurrent neural network architecture with pre-synaptic inhibition for incremental learning, H. OHTA, Y. PEGIO GUNJI

Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making, M. J. FRANK

The role of short-term depression in sustained neural activity in the prefrontal cortex: A simulation study, Y. IGARASHI, Y. SAKUMURA, S. ISHII

The misbehavior of value and the discipline of the will, P. DAYAN, Y. NIV, B. SEYMOUR, N. D. DAW

Neural voting machines, W. RICHARDS, H. S. SEUNG, G. PICKARD

Modeling of autonomous problem solving process by dynamic construction of task models in multiple tasks environment, Y. OHIGASHI, T. OMORI

Effects of reward expectancy on sequential eye movements in monkeys, J-W. SOHN, D. LEE

Both motivational and training factors affect response conflict choice performance in rats, J. E. HADDON, S. KILLCROSS

Contributions of prefrontal cue-, delay-, and response-period activity to the decision process of saccade direction in a free-choice ODR task, K. WATANABE, S. IGAKI, S. FUNAHASHI

Single trial-based prediction of a go/no-go decision in monkey superior colliculus, R. P. HASEGAWA, Y. T. HASEGAWA, M. A. SEGRAVES

Brain mechanism of reward prediction under predictable and unpredictable environmental dynamics, S. C. TANAKA, K. SAMEJIMA, G. OKADA, K. UEDA, Y. OKAMOTO, S. YAMAWAKI, K. DOYA

Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning, M. HARUNO, M. KAWATO

Behavioral inhibition and prefrontal cortex in decision-making, M. SAKAGAMI, X. PAN, B. UTTL

Neuropsychological correlates of decision-making in ambiguous and risky situations, M. BRAND, K. LABUDDA, H. J. MARKOWITSCH

Neural systems implicated in delayed and probabilistic reinforcement, R. N. CARDINAL

Weighing up the benefits of work: Behavioral and neural analyses of effort-related decision making, M.E. WALTON, S.W. KENNERLEY, D.M. BANNERMAN, P.E.M. PHILLIPS, M.F.S. RUSHWORTH

Goal-based action selection and utility-based action bias, K. MATSUMOTO, M. MATSUMOTO, H. ABE

NEURAL NETWORKS V19 N°9 NOVEMBER 2006

Special issue

Introduction to the special issue on "Brain & Attention", J.G. TAYLOR, C.A. NOBRE, K. SHAPIRO

Selective attention through phase relationship of excitatory and inhibitory input synchrony in a model cortical neuron, J. MISHRA, J-M. FELLOUS, T. J. SEJNOWSKI

The micro-structure of attention, N. R. TAYLOR, M. HARTLEY, J. G. TAYLOR

V4 receptive field dynamics as predicted by a systems-level model of visual attention using feedback from the frontal eye field, F. H. HAMKER, M. ZIRNSAK

Attention in natural scenes: Neurophysiological and computational bases, E. T. ROLLS, G. DECO

Modeling attention to salient proto-objects, D. WALTHER, C. KOCH

Attention as a controller, N. J.H. KORSTEN, N. FRAGOPANAGOS, M. HARTLEY, N. TAYLOR, J. G. TAYLOR

Analyzing and shaping human attentional networks, M. I. POSNER, B. E. SHEESE, Y. ODLUDAS, Y. TANG

Cognitive conflict without explicit conflict monitoring in a dynamical agent, R. WARD, R. WARD

NEURAL NETWORKS V19 N°10 DECEMBER 2006

Dynamic switching of neural codes in networks with gap junctions, Y. KATORI, N. MASUDA, K. AIHARA

Psychology and Cognitive Science

A neural network implementation of a saliency map model, M. DE BRECHT, J. SAIKI

A competitive neural model of small number detection, M. C. CASEY, K. AHMAD

Neuroscience and Neuropsychology

A theoretical network model to analyse neurogenesis and synaptogenesis in the dentate gyrus, M. BUTZ, K. LEHMANN, I. E. DAMMASCH, G. TEUCHERT-NOODT

Mathematical and Computational Analysis

Trainable fusion rules. I. Large sample size case, S. RAUDYS

Trainable fusion rules. II. Small sample-size effects, S. RAUDYS

Global exponential stability of recurrent neural networks with time-varying delays in the presence of strong external stimuli, Z. ZENG, J. WANG

Global attraction and stability for Cohen-Grossberg neural networks with delays, K. LU, D. XU, Z. YANG

Invariance priors for Bayesian feed-forward neural networks, U. V. TOUSSAINT, S. GORI, V. DOSE

Differential learning algorithms for decorrelation and independent component analysis, S. CHOI

Growing Hierarchical Tree SOM: An unsupervised neural network with dynamic topology, A. FORTI, G. L. FORESTI

Global exponential stability of impulsive high-order BAM neural networks with time-varying delays, D. W.C. HO, J. LIANG, J. LAM

TWNFI - a transductive neuro-fuzzy inference system with weighted data normalization for personalized modeling, Q. SONG, N. KASABOV

Terminated Ramp-Support Vector Machines: A nonparametric data dependent kernel, S. MERLER, G. JURMAN

Support vector classifiers via gradient systems with discontinuous righthand sides, L. V. FERREIRA, E. KASZKUREWICZ, A. BHAYA

Missing data imputation through GTM as a mixture of -distributions, A. VELLIDO

Engineering and Design

Fast computation of a gated dipole field, G. MENGOV, K. GEORGIEV, S. PULOV, T. TRIFONOV, K. ATANASSOV

A single network adaptive critic (SNAC) architecture for optimal control synthesis for a class of nonlinear systems, R. PADHI, N. UNNIKRISHNAN, X. WANG, S.N. BALAKRISHNAN

NEURAL NETWORKS V20 N°1 JANUARY 2007

Editorial for 2007: Another year of exciting special issues!, S. GROSSBERG, M. KAWATO, J. G. TAYLOR

Psychology and Cognitive Science

Topological relationships between brain and social networks, S. SAKATA, T. YAMAMORI

Perception through visuomotor anticipation in a mobile robot, H. HOFFMANN

The interaction of implicit learning, explicit hypothesis testing learning and implicit-to-explicit knowledge extraction, R. SUN, X. ZHANG, P. SLUSARZ, R. MATHEWS

Neuroscience and Neuropsychology

Towards cortex sized artificial neural systems, C. JOHANSSON, A. LANSNER

Mathematical and Computational Analysis

Stable concurrent synchronization in dynamic system networks, Q-C. PHAM, J-J. SLOTINE

Neural network explanation using inversion, E. W. SAAD, D. C. WUNSCH II

New necessary and sufficient conditions for absolute stability of neural networks, T. CHU, C. ZHANG

Awaking and sleeping of a complex network, R. LÓPEZ-RUIZ, Y. MORENO, A.F. PACHECO, S. BOCCALETTI, D.-U. HWANG

Engineering and Design

Pipelining of Fuzzy ARTMAP without matchtracking: Correctness, performance bound, and Beowulf evaluation, J. CASTRO, J. SECRETAN, M. GEORGIPOULOS, R. DEMARA, G. ANAGNOSTOPOULOS, A. GONZALEZ

Robust Bayesian clustering, C. ARCHAMBEAU, M. VERLEYSSEN
A learning algorithm for adaptive canonical correlation analysis of several data sets, J. VÍA, I. SANTAMARÍA, J. PÉREZ

ARTIFICIAL INTELLIGENCE IN MEDICINE

<http://www.elsevier.com/inca/publications/store/5/0/5/6/2/7/>

AI IN MEDICINE V38 N°2 OCTOBER 2006

Special issue

Temporal Representation and Reasoning in Medicine, ed. by C. COMBI, G. POZZI

Temporal representation and reasoning in medicine: Research directions and challenges, K-P. ADLASSNIG, C. COMBI, A. K. DAS, A. T. KERAVNOU, G. POZZI

Distributed, intelligent, interactive visualization and exploration of time-oriented clinical data and their abstractions, Y. SHAHAR, D. GOREN-BAR, D. BOAZ, G. TAHAN
Extending the event calculus for tracking epidemic spread, H. CHAUDET

Spatiotemporal reasoning about epidemiological data, P. REVESZ, S. WU

Towards a comprehensive treatment of repetitions, periodicity and temporal constraints in clinical guidelines, L. ANSELMA, P. TEREZIANI, S. MONTANI, A. BOTTRIGHI

Fuzzy theory approach for temporal model-based diagnosis: An application to medical domains, J. PALMA, J. M. JUAREZ, M. CAMPOS, R. MARIN

AI IN MEDICINE V38 N°3 NOVEMBER 2006

Knowledge discovery in traditional Chinese medicine: State of the art and perspectives, Y. FENG, Z. WU, X. ZHOU, Z. ZHOU, W. FAN

From vagueness in medical thought to the foundations of fuzzy reasoning in medical diagnosis, R. SEISING

A closed-loop hybrid physiological model relating to subjects under physical stress, E. EL-SAMAHY, M. MAHFOUF, D. A. LINKENS

Using classification trees to assess low birth weight outcomes, P. KITSANTAS, M. HOLLANDER, L. LI

Cell-nuclear data reduction and prognostic model selection in bladder tumor recurrence, D. K. TASOULIS, P. SPYRIDONOS, N. G. PAVLIDIS, V. P. PLAGIANAKOS, P. RAVAZOULA, G. NIKIFORIDIS, M. N. VRAHATIS

Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room, M. GREEN, J. BJÖRK, J. FORBERG, U. EKELUND, L. EDENBRANDT, M. OHLSSON

Modeling in biomedical informatics: Pastime or necessity?: A dialogue on lessons from the past and directions for the future, STUDENT & TEACHER

AI IN MEDICINE V39 N°1 JANUARY 2007

Temporal abstraction in intelligent clinical data analysis: A survey, M. STACEY, C. MCGREGOR

Learning recurrent behaviors from heterogeneous multivariate time-series, F. DUCHÊNE, C. GARBAY, V. RIALLE

Side chain placement using estimation of distribution

SOMMAIRES DES REVUES

algorithms, R. SANTANA, P. LARRAÑAGA, J. A. LOZANO
Advanced fuzzy cellular neural network: Application to CT
liver images, S. WANG, D. FU, M. XU, D. HU
Dynamic knowledge validation and verification for CBR tele-
dermatology system, M. H. OU, G. A.W. WEST, M. LAZA-
RESCU, C. CLAY

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

<http://www.elsevier.com/inca/publications/store/5/0/5/7/8/7/>

IJAR V43 N°3 DECEMBER 2006

Bookmaking over infinite-valued events, D. MUNDICI
On the calculation of the bounds of probability of events using
infinite random sets, D. A. ALVAREZ
Extensions of the multicriteria analysis with pairwise
comparison under a fuzzy environment, M-S. KUO, G-S.
LIANG, W-C. HUANG

IJAR V44 N°1 JANUARY 2007

Special issue

Genetic Fuzzy Systems and the Interpretability–Accuracy
Trade-off, ed. by J. CASILLAS, F. HERRERA, R. PÉREZ,
M.J. DEL JESUS, P. VILLAR
Analysis of interpretability-accuracy tradeoff of fuzzy sys-
tems by multiobjective fuzzy genetics-based machine
learning, H. ISHIBUCHI, Y. NOJIMA

Improving the accuracy while preserving the interpretability
of fuzzy function approximators by means of multi-
objective evolutionary algorithms, J. GONZÁLEZ, I. ROJAS,
H. POMARES, L. J. HERRERA, A. GUILLÉN, J. M.
PALOMARES, F. ROJAS

Genetic learning of accurate and compact fuzzy rule based
systems based on the 2-tuples linguistic representation,
R. ALCALÁ, J. ALCALÁ-FDEZ, F. HERRERA, J. OTERO
Interpretability-preserving genetic optimization of linguistic
terms in fuzzy models for fuzzy ordered classification:
An ecological case study, E. VAN BROEKHOVEN, V.
ADRIAENSSENS, B. DE BAETS

IJAR V44 N°2 FEBRUARY 2007

Special issue

Fuzzy Decision-Making Applications, ed. by C. KAHRAMAN
A fuzzy approach to R&D project portfolio selection, C.
CARLSSON, R. FULLÉR, M. HEIKKILÄ, P. MAJLENDER
Decision tree search methods in fuzzy modeling and classifi-
cation, L.F. MENDONÇA, S.M. VIEIRA, J.M.C. SOUSA
Prioritization of organizational capital measurement indicators
using fuzzy AHP, F. TUNÇ BOZBURA, A. BESKESE
Sensory evaluation based on linguistic decision analysis, L.
MARTÍNEZ
A fuzzy-set approach to treat determinacy and consistency of
linguistic terms in multi-criteria decision making, J. MA, D.
RUAN, Y. XU, G. ZHANG
A fuzzy approach to value and culture assessment and an
application, A. SOYER, Ö. KABAK, U. ASAN

LES DOSSIERS DU BULLETIN

IA dans l'industrie	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 RNNTL.....	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
IA et Terminologie.....	n°32		
Décision et IA.....	n°31		
IA et Image	n°30		

Vos recensions dans le Bulletin

Vous avez récemment lu un livre sur l'I.A. et vous l'avez aimé ?

Faites-nous part de ce coup de coeur !

Nous la publierons, pour en faire bénéficier le plus grand nombre.

Adhésion individuelle et abonnement		<input type="checkbox"/> Demande	<input type="checkbox"/> Renouvellement
Nom :		Prénom :	
Affiliation :			
Adresse postale :			
N° de téléphone :		N° de télécopie :	
Adresse électronique :			
Activités (à titre professionnel / à titre privé (<i>rayez la mention inutile</i>)) :			
	Consultation du bulletin sur WEB (pour une personne)	Envoi du bulletin papier + un accès pour consultation du bulletin sur WEB	
<input type="checkbox"/> Adhésion simple :	30 Euros	60 Euros	
<input type="checkbox"/> Adhésion étudiant (sur justificatif) :	15 Euros	30 Euros	
<input type="checkbox"/> Adhésion de soutien	Sans objet	90 Euros	
<input type="checkbox"/> Abonnement au bulletin sans adhésion	Sans objet	55 Euros	
<input type="checkbox"/> Adhésion au collège <i>IAD-SMA</i> : ajouter 7,5 Euros pour les étudiants, 15 Euros pour les autres. <input type="checkbox"/> Adhésion au collège <i>Cafe (Apprentissage)</i> : gratuit.			

Adhésion Personne morale		<input type="checkbox"/> Demande	<input type="checkbox"/> 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 :	
<input type="checkbox"/> Laboratoires universitaires	100 Euros	30 Euros	
<input type="checkbox"/> Personnes morales non universitaires	300 Euros	30 Euros	
<input type="checkbox"/> Adhésion de soutien	600 Euros	Sans objet	
<input type="checkbox"/> j'accepte que les renseignements ci-dessus apparaissent dans l'annuaire de l'AFIA. <input type="checkbox"/> 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 TVA non applicable, article 293B du CGI			

SOMMAIRE DU BULLETIN N° 63

Editorial	3
Dossier	4
Présentation de Laboratoires	27
Conférences	28
Resumé de thèses et d'HDR	31
Offres adhérents	41
Sommaires des revues	42
Les dossiers du bulletin de l'Afia	53

CALENDRIER DE PARUTION DU BULLETIN DE L'AFIA

<i>Hiver</i>	<i>Eté</i>
Réception des contributions: 15 décembre	Réception des contributions: 15 juin
Sortie le 31 janvier	Sortie 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

