

BULLETIN

DE

L'AFIA



OCTOBRE 2000

N°43

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 reflètent le point de vue de leurs auteurs et n'engagent qu'eux-mêmes.

Pour contacter l'AFIA

Président :
Bertrand BRAUNSCHWEIG
Institut Français du Pétrole
Direction Informatique
Mathématiques Appliquées
B.P. 311
92506 Rueil Malmaison CEDEX
Mél. : Bertrand.Braunschweig@ifp.fr

Contributions au bulletin

Gérard SABAH
voir ci-contre

Serveur WEB
<http://www.afia.polytechnique.fr/>

Adhésions, Liens avec les adhérents
Marc AYEL
LIA-ESIGEC- Technolac
Université de Savoie
73376 Le Bourget du Lac CEDEX
Mél. : Marc.Ayel@univ-savoie.fr

Parrainage de manifestations
Sylvie PESTY
Laboratoire Leibniz, Institut IMAG
46, Avenue Félix Viallet
38031 Grenoble CEDEX
Mél. : Sylvie.Pesty@imag.fr

Membres d'honneur

**Jaques Pitrat, Jean-Paul Haton, Marie-Odile Cordier
Jean-Marc David, Daniel Kayser, Claude Vogel**

Personnes morales adhérentes à l'AFIA

ADIT, AI*IA, 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

Bertrand BRAUNSCHWEIG, président
Serge DUPUY, secrétaire
Marc AYEL, trésorier

**François ARLABOSSE, Jean-Paul BARTHÈS, Jérôme EUZENAT,
Patrick GALLINARI, Catherine GARBAY, Christine GOLBREICH,
Eunika MERCIER-LAURENT, Sylvie PESTY, Gérard SABAH,
Christian de SAINTE-MARIE, Jean-Paul SANSONNET,
Michèle SEBAG, Laurent SIKLÓSSY, Marc SCHOENAUER,
Fabien TORRE**

Comité de rédaction

Gérard SABAH
Rédacteur en chef
LIMSI — CNRS
B.P. 133, 91403 ORSAY Cedex
gs@limsi.fr

Jean CHARLET
Rubrique
« Conférences et comptes rendus »
INSERM U194 — DIAM
91 boulevard de l'Hôpital
75634 Paris Cedex 13
charlet@biomath.jussieu.fr

Patrick GALLINARI
Rubrique « Groupes de travail et collègues »
LIP6, Université Paris 6
8, rue du Capitaine Scott
75015 Paris
Patrick.Gallinari@lip6.fr

Brigitte GRAU
**Rubrique « Sommaires des revues »
et « petites annonces »**
LIMSI — CNRS
B.P. 133, 91403 ORSAY Cedex
grau@limsi.fr

**Rubriques « Présentation de laboratoires »
et « Présentation de sociétés »**
Gilles KASSEL
Université de Picardie Jules Verne — LARIA
5, rue du Moulin neuf
80000 — Amiens
kassel@laria.u-picardie.fr

et **Catherine BARRY-GRÉBOVAL**
PSI — INSA de Rouen
place Émile Blondel BP 08
76131 Mont Saint Aignan Cedex
Catherine.Greboval@insa-rouen.fr

Amedeo NAPOLI
Rubriques « thèses » et « livres »
LORIA - B.P. 239
F-54 506 Vandœuvre lès Nancy
Amedeo.Napoli@loria.fr

Serge STINCKWICH
Rubrique « IA et Internet »
GREYC — Université de Caen
Esplanade de la Paix
14032 CAEN CEDEX
Serge.Stinckwich@info.unicaen.fr

Jean-Daniel ZUCKER
Rubrique « Débats »
LIP6, Université Paris 6
4, Place Jussieu, 75232 Paris Cedex 05
Jean-Daniel.Zucker@lip6.fr



Réa. D.E.B. Copy – 01 46 31 64 53
ISSN 1273-1323
Dépot légal novembre 2000

Le retour du JEDA

Ce n'est pas un retour, en fait, plutôt une naissance, je l'avoue. Mais, franchement, il m'était difficile d'échapper à ce titre pour un éditorial qui vous annonce que nous nous préparons à un heureux événement, la naissance du JEDAI (prononcer « djèdaï », comme vous l'aurez deviné).

Le projet du Journal Électronique D'Intelligence Artificielle⁽¹⁾ a été conçu dans la douleur, douleur de voir la RIA s'enfoncer petit à petit dans les problèmes dont nous avons suffisamment parlé cet été. Je ne reviendrai donc pas sur ces problèmes, ni sur le *modus vivendi* convenu avec Hermès, qui les fera servir les abonnés 2000 comme ils l'entendent — sans doute par des numéros spéciaux.

Nous nous tournons aujourd'hui vers ce projet de bâtir un journal électronique d'IA francophone. D'autres ont pris avant nous de telles initiatives : les ETAI, Electronic Transactions on Artificial Intelligence, de l'ECCAI, existent depuis plusieurs années, et sont un réel succès notamment grâce à la persévérance de leur éditeur Erik Sandewall. Le nouveau magazine d'Elsevier sur les applications du « soft computing », dirigé par Raj Roy, sera 100 % électronique — à ma connaissance c'est la première fois qu'un grand éditeur comme Elsevier abandonne le papier pour une de ses revues. En complément de cette parution électronique, Elsevier publiera tous les ans un CD d'extraits du magazine, formule qui leur permet d'inclure des vidéos, des animations, des démonstrations de logiciels, toutes choses difficiles à rendre sur papier...

L'AFIA considère qu'une revue scientifique francophone d'IA est un outil de travail indispensable à notre communauté, en complément des publications anglophones existantes, et qu'il relève de sa responsabilité de développer une telle revue. Aujourd'hui, un tel outil doit nécessairement être pensé en liaison avec l'Internet. À côté du bulletin de l'association, qui n'a pas vocation à devenir une publication purement scientifique, et au sein de notre futur portail de l'IA francophone, le journal électronique sera idéalement placé pour servir de vecteur de diffusion privilégié des recherches et des applications de l'IA en France et dans la francophonie. Nous le voulons interactif, innovant, de bonne tenue scientifique, pouvant servir de référence, et surtout — point appréciable — gratuit et ouvert à tous, pas seulement aux adhérents de l'association.

Ce numéro contient un appel à candidature pour le poste de rédacteur en chef. Nous souhaitons mener cette opération très rapidement, afin d'être prêts à publier vos articles (que nous attendons nombreux) dès le début de l'année prochaine. Car, bien entendu, le contenu, c'est aussi et avant tout vous qui le ferez !

Que la Science soit avec vous.

Bertrand Braunschweig

(1) J'utilise JEDAI comme appellation provisoire en attendant que le futur comité de rédaction propose un nom définitif. Ne m'écrivez pas pour me signaler l'inversion du IA en AI, je suis au courant !

Journal Électronique d'Intelligence Artificielle ⁽¹⁾

Appel à candidature pour le poste de rédacteur en chef

L'AFIA recherche un rédacteur en chef pour son futur journal électronique d'intelligence artificielle, qui sera lancé début 2001. Ce journal électronique sera consultable librement sur le portail Web de l'association et deviendra un vecteur interactif privilégié de diffusion des recherches et applications développées par la communauté IA française et francophone. Nous envisageons d'accompagner le journal électronique d'une diffusion annuelle sous forme d'ouvrage et de CD-ROM, en liaison avec un éditeur.

L'appel concerne uniquement la prise en charge du contenu scientifique du journal. Le contenant, c'est-à-dire les moyens techniques de fonctionnement du journal, relève du groupe qui a la responsabilité du portail Web. Le rédacteur en chef sera chargé de constituer le comité de rédaction, de définir les processus de soumission, de revue et de publication, et de les mettre en œuvre. Plus généralement il devra également définir, en collaboration avec le bureau de l'AFIA, la politique de la revue électronique.

Les candidats sont invités à se déclarer, avant le 1er décembre 2000 au soir, en fournissant une lettre de motivation accompagnée d'un CV, par Mél à *Bertrand.Braunschweig@ifp.fr*.

La sélection du rédacteur en chef sera effectuée avant le 20 décembre 2000 par le comité dont la composition est indiquée ci-après.

Le comité " JEDAI "

François Arlabosse
Philippe Besnard
Bertrand Braunschweig
Raja Chatila
Michel Chein
Marie-Odile Cordier
Daniel Kayser
Jean-Paul Sansonnet

(1) Nom provisoire en attendant que le comité de rédaction propose un nom au journal.

Le Web est-il, deviendra-t-il une base de connaissances ?

C'est l'intitulé du débat suggéré par Jérôme Euzenat dans le dernier numéro. Deux débatteurs nous ont à ce jour répondu. Marie-Christine Rousset répond par l'affirmative à la première partie de la question en soulignant que le principal problème du Web, qui est déjà une base de connaissances de par sa construction, est sa **disparité**. Elle considère qu'une question constructive à poser est de savoir si le « *Web deviendra une base de connaissances utile, c'est-à-dire, réellement exploitable pour répondre de façon précise et fiable aux besoins des utilisateurs en recherche d'information, et si oui comment ?* ». Elle conclut sur le fait que les techniques d'Intelligence Artificielle, en particulier les techniques de représentation de connaissances, de fouille de données et de traitement des langues naturelles, ont des atouts majeurs pour apporter des solutions innovantes à ces problèmes. La contribution de Jean Charlet insiste sur la nécessité des méta-données pour exploiter ces connaissances.

Si vous avez envie de réagir, en quelques lignes ou quelques paragraphes, vos contributions sont toujours les bienvenues !

Le nouveau débat qui sera présenté dans le prochain numéro s'articulera autour de quelques questions clefs posées par Jean-Luc Soubie, Mario Borillo et Henri Prade. Celles-ci ont pour origine des dialogues qui ont eu lieu à la suite du spectacle de Jean-François Peyret « Histoire Naturelle de l'esprit », un spectacle original conçu à partir de la vie d'Alan Turing (cf. page suivante).

Participation de Marie-Christine Rousset.

Vaste question, question qu'il est légitime de poser, question à laquelle il est difficile de répondre sans être réducteur, question dont la réponse dépend fortement de la perception que l'on a de ce qu'est une *base de connaissances*.

Je vous épargnerai la définition que je donne à mes étudiants, que certains d'entre vous trouveraient trop stricte, et je prendrai, pour participer à ce débat, l'acceptation la plus large (et la plus floue) de la notion de base de connaissances. En particulier, je ne ferai pas de distinguo subtil entre « données », « connaissances » et « informations ».

On ne peut contester que la multitude des docu-

ments accessibles par le Web contient de la connaissance : celle des personnes ayant écrit ces documents avec la volonté de diffuser des éléments d'information sur tel ou tel sujet. De ce point de vue, le Web (plus précisément l'ensemble des documents accessibles par le Web) est donc bel et bien une immense base de connaissances constituée de connaissances distribuées et surtout très disparates : là est le problème !

Il me semble donc que la bonne question à poser est : « le Web est-il une base de connaissances utile, c'est-à-dire, réellement exploitable pour répondre de façon précise et fiable aux besoins des utilisateurs en recherche d'information ? »

Or, on ne peut que constater que, même avec l'aide des moteurs de recherche actuels les plus évolués, chercher une information précise sur le Web revient souvent à chercher une aiguille dans une meule de foin. Pour Umberto Eco : « *14 millions sites, it is like nothing since I am not able to choose* ». La réponse à la question : « le Web est-il une base de connaissances *utile* ? » me paraît aujourd'hui être non.

Ne nous décourageons pas et posons alors une nouvelle question : « **le Web deviendra-t-il une base de connaissances utile ? et, si oui, comment ?** »

C'est précisément cette question, avec les défis qu'elle engendre pour lui apporter une réponse positive, qui motive une partie de mes travaux récents et actuels.

La nécessité de structurer et d'organiser le vrac et la diversité des données accessibles par le Web s'impose : selon Lucien Sfez (un spécialiste en communication) : « *if we want to access any knowledge, we must learn to classify into hierarchies and structures* ».

Structurer et organiser les données du Web se heurte à deux difficultés majeures :

1. La quantité même des données accessibles par le Web, qui ne fait que croître, empêche une structuration manuelle, même guidée par des ontologies existantes ou que l'on construirait. Les solutions à la Yahoo (classification manuelle par les équipes de Yahoo des documents trouvés par leurs *crawlers* relativement à une hiérarchie de termes également conçue par eux-mêmes) sont des solutions *ad-hoc* qui montrent déjà leurs limites pour le Web d'aujourd-

d'hui et qui ne permettent pas de passer à l'échelle du Web de demain. Ma conviction est que pour structurer, regrouper, organiser les données à l'échelle du Web, il faut inventer des outils automatiques.

2. La diversité de la provenance des données, et le fait qu'elles sont véhiculées en majorité par des langues naturelles, entraîne une forte hétérogénéité sémantique qui complique un regroupement automatique des documents traitant de sujets similaires ou proches. Il faut faire face au multilinguisme, bien sûr, mais aussi aux problèmes de polysémie et de synonymie à l'intérieur d'une même langue. On peut penser que l'arrivée du format XML comme nouveau standard de description des données du Web va supprimer ce problème. L'intérêt de XML est certes d'offrir un format homogène de structuration des documents (et donc des données du Web) par un jeu de balises librement choisies par les auteurs des documents. Cette structuration pourra être exploitée par les outils automatiques d'interrogation de bases de documents XML. Il existe déjà plusieurs langages de requêtes XML proposés par la communauté « Bases de Données » (avec les moteurs de requêtes associés). Cependant, un format homogène des données ne supprime pas le problème de leur hétérogénéité sémantique.

On ne pourra empêcher des auteurs de documents XML de choisir des balises pour structurer leurs documents qui seront différentes du jeu de balises choisies par d'autres auteurs pour décrire des données pourtant relatives à des sujets similaires. Autant je pense que le format XML va s'imposer dans un avenir proche sur le Web, autant je doute que les auteurs de documents XML suivront des directives pour se conformer de façon stricte à des DTDs (c'est-à-dire à des jeux de balises) prédéfinis par tel ou tel groupe de normalisation.

Les techniques d'Intelligence Artificielle, en particulier les techniques de représentation de connaissances, de fouille de données et de traitement de langues naturelles, ont des atouts majeurs pour apporter des solutions innovantes à ces problèmes. Pour réussir, nous devons cependant prendre la mesure de la taille et du bruit des données que nous devons traiter en entrée des outils fondés sur ces techniques. Je pense réellement que la solution passe par une combinaison de techniques et de compétences variées associant des chercheurs en Intelligence Artificielle, en Bases de Données, et en Statistiques.

LE COIN DU TRÉSORIER

Les chiffres et lettres à droite de votre nom indiquent votre situation vis-à-vis de la trésorerie. Vous pouvez ainsi vérifier que vous êtes bien à jour de votre cotisation pour cette année.

<p>Dupont Jacques Laboratoire d'IA Villa les Systèmes Experts 79010 Mycin</p>	<div style="border: 1px solid black; width: 40px; height: 30px; margin: 0 auto;"></div> <div style="border: 1px solid black; border-radius: 50%; width: 100px; height: 20px; display: flex; align-items: center; justify-content: center; margin: 0 auto;"> m92/93/RIA96 </div>
---	--

- Le code RIA indique une adhésion couplée avec l'abonnement à RIA.
- La lettre "m" indique une adhésion en tant que personne morale
- Les chiffres sont les deux derniers chiffres de l'année.

Ne jetez pas l'enveloppe d'expédition du Bulletin sans regarder de plus près l'étiquette portant votre nom et adresse...

Autour du spectacle de Jean-François Peyret « Histoire naturelle de l'esprit »

Jean-François Peyret et sa troupe ont, au printemps 2000, donné à voir successivement à Rennes, Bobigny et Toulouse avec « Histoire Naturelle de l'Esprit (suite et fin) » un spectacle original conçu à partir de la vie d'Alan Turing, un des pères de l'informatique et un des précurseurs de l'intelligence artificielle.

Le spectacle théâtral de deux heures est conçu à partir de textes de scientifiques (biologistes, mathématiciens, etc.), de philosophes (en particulier Hannah Arendt) et d'éléments biographiques de la vie d'Alan Turing. Sa mise en scène originale fait appel à une technologie de pointe (traitement d'image et projection vidéo simultanée) et s'appuie sur un décor de cubes (clin d'œil au système informatique de dialogue en langage naturel SHRDLU de Terry Winograd ?) de sièges et de lampadaires à roulettes, renvoyant à une dimension ludique de l'enfance. Le corps, l'esprit et leur relation (si douloureusement vécue par Alan Turing) sont au cœur des interrogations portées par le spectacle, qui renvoie également à la problématique du théâtre et au jeu des acteurs.

Jean-François Peyret s'est lui-même exprimé sur ce sujet à l'occasion du spectacle donné au Théâtre National de Toulouse dans un entretien avec Sébastien Bournac, dont le texte est disponible au TNT.

« Au théâtre, on entre des textes dans les corps des comédiens (input) ; on ne sait pas si ce qui ressort est mécanique ou vivant (output) [...] »

C'est vrai que réfléchir sur le monde infernal de la machine, sur l'idée qu'une machine pourrait penser et dépasser l'homme, sur l'idée qu'il pourrait ne pas y avoir de différence entre un homme qui pense et une machine qui pense, nous a étrangement renvoyés au théâtre. La question essentielle étant : « est-ce que penser c'est manipuler des symboles ? », sachant que le théâtre lui-même ne fait pas autre chose que manipuler des symboles, jouer avec des codes, nous ne pouvions esquiver la question de savoir si la machine théâtrale pense ou ne pense pas. [...] Le comédien dit un texte, le traite sans avoir à le penser, et les spectateurs qui le reçoivent ne le pensent pas non plus. Mais n'en va-t-il pas toujours ainsi, même hors du théâtre, même sur la scène de la pensée professionnelle ? Qu'est-ce que comprendre ? qu'est-ce que penser ? En tout cas, cette question m'intéresse autant que celle de la psychologie des personnages [...] Que l'homme fabrique des machines " intelligentes " bouleverse nécessairement la définition qu'on se fait de l'esprit. Que l'homme n'ait peut-être plus le monopole de la pensée n'est pas sans

conséquence sur nos définitions de l'humain et de l'in-humain dès lors que l'esprit était l'apanage de l'homme. Il y a cette hypothèse, que je caresse volontiers, que si l'homme perd le privilège de la pensée, il ne lui restera plus que ses passions à traiter, [...] »

Le site de la compagnie permet d'accéder aux textes et à de nombreux éléments sur la gestation et la conception du spectacle (<http://www.tf2.asso.fr/total.htm>).

Les dialogues qui suivent ne visent pas à donner une idée et encore moins une analyse d'un spectacle très riche qui pendant près de deux heures a mis en scène, bien au-delà de l'anecdote biographique, une multiplicité de textes et de points de vue autour de questions telles que « Les machines pensent-elles ? » ou « Faut-il un corps pour penser ? » Ils se bornent à reprendre l'essentiel des propos tenus par trois chercheurs en intelligence artificielle de l'Institut de Recherche en Informatique de Toulouse (IRIT, CNRS, UPS, ENSEEIHT) lors d'un débat organisé à Toulouse le 12 avril 2000 à la suite d'une des représentations du spectacle réalisé par Jean-François Peyret au Théâtre National de Toulouse. Car cela n'est pas la moindre des vertus de ce spectacle que de faire réagir, s'exprimer et réfléchir le chercheur sur sa propre discipline et sur les questions souvent inédites qu'ouvrent ses travaux.

Qu'un metteur en scène utilise une matière scientifique difficile comme argument d'un spectacle, c'est assez inhabituel. Qu'il en joue d'une manière ludique, esthétique, et qui donne à réfléchir et à rêver, c'est encore plus remarquable. Que la science soit considérée comme partie intégrante de la culture contemporaine, qu'elle puisse contribuer à nourrir l'imaginaire de chacun, s'inscrire dans des recherches d'ordre esthétique, ne peut que faciliter sa vulgarisation, en l'aidant à sortir d'un isolement d'autant plus redoutable qu'elle est en train de bouleverser le monde.

Jean-Luc Soubie :

La question centrale autour de laquelle tourne la pièce de J.F. Peyret « Histoire naturelle de l'esprit » est : Les machines pensent-elles ?

À cette interrogation il nous est demandé d'apporter le point de vue du chercheur en informatique.

Mario Borillo :

Pour la recherche scientifique proprement dite, en particulier celle qui est associée aux « machines »,

c'est-à-dire à l'informatique précisément, la « pensée » en tant que telle, holistique, aux contours mal définis, territoire de la philosophie de l'esprit... n'est pas en soi un objet d'investigation. Mais nos travaux reposent implicitement sur des présupposés liant telle ou telle forme de pensée et de calcul.

Henri Prade :

C'est en effet une question que, paradoxalement, l'intelligence Artificielle n'aborde pas de front.

Jean-Luc Soubie :

Il n'existe évidemment pas de réponse péremptoire par oui ou non, fut-elle argumentée, à une telle question. Cependant, dans son activité quotidienne, le chercheur est conduit à s'interroger sur le rapport qu'entretiennent les résultats qu'il produit avec ce type d'interrogation de nature un peu fantasmatique.

Mario Borillo :

Ce que nous étudions en vue de leur reproduction par des machines sont des phénomènes plus spécifiques, par exemple des formes de raisonnement que nous mettons en jeu dans la vie quotidienne (le « sens commun » !), d'apprentissage de procédures, de reconnaissance de formes, de « compréhension » du langage... Ces phénomènes relèvent à coup sûr de la vie mentale, et ce sont des phénomènes cognitifs singuliers qui n'avaient pas fait l'objet jusqu'ici de ce type d'approche scientifique. Avec des résultats troublants, dont on peut parfois trouver l'écho dans la presse.

Songez à Deep Blue, grand maître échiquéen !

Jean-Luc Soubie :

Peut-être est-il insuffisant d'affirmer que penser est une activité humaine qui exclut que tout autre agent que l'humain puisse l'exercer. Cependant, toute définition de la pensée appelle la convocation d'une de ses caractéristiques troublantes, la réflexivité. En effet, la pensée ne se conçoit pas sans la conscience d'elle-même ; peut-on penser sans avoir conscience que l'on pense, alors que l'on peut agir sans avoir conscience que l'on a pensé pour agir. Ainsi de suite, jusqu'au vertige, si l'on pense qu'on pense, alors on pense qu'on pense qu'on pense...

Henri Prade :

Les machines peuvent être dotées de capacités de raisonnement et de prise de décision de plus en plus élaborées. La production de conclusions chez ces machines peut aller bien au-delà de la pure déduction logique, et permettre l'induction, le raisonnement par

analogie, le raisonnement par défaut, l'extrapolation, notamment. Les machines peuvent ainsi, par exemple, calculer, évaluer des situations, faire des diagnostics, planifier, répondre à des questions, (re) découvrir des lois mathématiques, mettre en évidence des régularités dans des données, percevoir l'environnement grâce à des capteurs, se mouvoir en évitant des obstacles, imiter des comportements humains, aider à la créativité, à la conception de nouveaux objets, voire faire des réalisations artistiques obéissant à des lois de composition et à de l'aléatoire.

Les machines effectuent leurs opérations souvent plus rapidement et de manière plus fiable que l'homme en exploitant efficacement de plus grandes masses d'informations. L'exemple récent de Deep Blue, vainqueur de Kasparov est effectivement là pour en attester.

Les machines commencent aussi à pouvoir réfléchir sur ce qu'elles sont en train de faire et ainsi à pouvoir faire évoluer leur comportement. Sans doute demain pourront-elles résumer des textes, appréhender des émotions, des intentions, voire faire des évaluations touchant à l'esthétique.

Jean-Luc Soubie :

Les informaticiens ont créé des mécanismes qui permettent de prendre en compte au sein d'une machine, au moyen de la récursivité, une accumulation, possiblement infinie de fonctions s'appliquant à elles-mêmes. L'exemple classique est le calcul de la factorielle, définie à chaque étape relativement à sa valeur à l'étape précédente $n! = n*(n-1)!$. L'un des problèmes est la définition de la condition d'arrêt de ces calculs qui, mal définie, peut conduire au blocage de la machine dans une boucle sans fin. L'humain sain d'esprit prend en compte le plus souvent, des éléments de son environnement pour adapter ces conditions d'arrêt, ce qu'il est pour longtemps impossible de faire faire à une machine hors d'un ensemble déterminé de critères définis *a priori*.

Mario Borillo :

Pour l'Intelligence Artificielle, il faut souligner que de telles performances ne sont possibles que dans la mesure où ses recherches ont pour premier objectif de construire des modèles formels des phénomènes évoqués précédemment. Ces modèles permettent d'écrire des programmes, qui, implantés sur des machines, font que celles-ci « se comportent » pour le phénomène modélisé — restriction essentielle — comme le ferait un sujet humain. On parlera alors, à juste titre, de simulation... Il est important de souligner que pour la première fois dans l'Histoire, des fonctions mentales importantes sont implantées dans de la matière étrangère.

re à notre corps et que les conséquences à terme de cette percée techno-scientifique ne peuvent manquer d'être extrêmement graves. Mais ceci est un tout autre problème, qui aura des conséquences non seulement économiques – on le voit déjà – mais aussi sur la civilisation. Pour revenir à l'objet du débat, comme l'a dit Jean-Luc, conscience, réflexivité, par exemple, leur sont étrangères. Il ne faut pas oublier qu'en réalité elles ouvrent aussi, sur le plan théorique, des horizons nouveaux à l'informatique.

Henri Prade :

Dans toutes les activités réalisés par la machine le paradigme est toujours le même : La machine ne fait que résoudre des problèmes... souvent difficiles... que l'homme lui pose et qui sont exprimables dans un cadre formel, éventuellement très général. Même si une seule machine avait toutes les capacités qu'on vient d'évoquer, on ne pourrait sans doute pas encore dire qu'elle pense.

En effet la machine ne pense pas au sens où elle ne peut pas ultimement décider de ses buts ni sans doute imaginer de nouvelles représentations du monde... même si la machine peut formuler des sous-problèmes, dans un espace prédéfini, par rapport à une tâche poursuivie qu'on lui a assignée... Une machine n'hésite jamais, ne doute pas ! Clairement au fur et à mesure que l'intelligence artificielle... et l'informatique... développent des machines capables d'effectuer des tâches réputées intelligentes, cela pose, par différence, la question de la nature, de la spécificité de la pensée humaine, qui est sans doute davantage que des capacités de raisonnement et de réflexion en relation avec la gestion de sensations et d'émotions.

Jean-Luc Soubie :

La question qui semble raisonnablement pouvoir être posée aujourd'hui serait plutôt : quel rôle peut-on faire jouer à la machine comme support des activités humaines ? Il suffit de regarder derrière nous pour constater la fantastique évolution, permise en premier lieu par la technologie, des registres d'intervention des machines dans l'accompagnement des activités humaines : de la calculatrice initiale, on est arrivé à l'outil de communication ou d'accès à des données produites dans le monde entier, voire à la prothèse professionnelle capable de résoudre des problèmes compliqués mais routiniers ou encore au partenaire de jeu. On voit bien que cette évolution peut s'inscrire sur un axe allant de l'objet matériel à l'homme et que le point de vue pris par les concepteurs des outils se rapproche de plus en plus de l'homme. Cependant, cette tendance

conduit-elle à concevoir des machines qui, grâce à leurs capacités issues d'une démarche anthropomorphique, auront vocation à se substituer à l'humain dans ce qu'il a encore de spécifique : la pensée ? Ou bien l'homme doit-il cultiver sa spécificité et utiliser la machine dans ce qu'elle sait le mieux faire, les tâches répétitives et lourdes, pouvant faire appel à des raisonnements complexes ?

Mario Borillo :

Les recherches que nous évoquons ici s'inscrivent dans un mouvement scientifique plus général qui est celui des sciences de la cognition, où elles retrouvent un spectre très large de disciplines comme l'ensemble des neurosciences cognitives, la psychologie, l'ergonomie, la linguistique, la philosophie de l'esprit... Ces recherches sont fondamentalement pluridisciplinaires, à la fois empiriques (expérimentales) pour produire des données observationnelles et théoriques dans la définition des modèles formels et computationnels. Elles participent in fine, sur un plan plus général, à concevoir des manières d'associer la complexité de l'univers mental aux structures et fonctions biologiques. Parvenus à ce point, une question cruciale ne peut manquer d'émerger : Quelle nouvelle conception de l'homme et de son « esprit » de ses rapports avec la « machine » biologique et non-biologique, se dégagera de ce mouvement et de ses résultats, aussi bien scientifiques que technologiques ? Avec quelles conséquences pour la civilisation... rien de moins !

Henri Prade :

La compréhension de la nature de la pensée nous échappe encore largement et peut-être pour longtemps. L'homme a depuis plus de trois siècles volontiers comparé son corps et son cerveau à la machine la plus sophistiquée qu'il était parvenu à construire... les horloges au XI^e siècle, puis bientôt les automates, et aujourd'hui les ordinateurs et les robots..., pour se forger la représentation qu'il se faisait de lui-même en tant qu'individu ou que membre d'une société d'agents... développant une « intelligence collective ». Cette comparaison sert sans doute aussi son besoin de se confirmer dans sa supériorité sur la machine ! Inversement, les capacités de l'esprit humain constituent, sans doute encore pour longtemps, un ensemble de défis pour les chercheurs en intelligence artificielle dans leurs tentatives de les comprendre, de les formaliser et de les reproduire.

Jean-Luc Soubie :

Lorsqu'on constate la difficulté rencontrée par les

chercheurs pour définir les méthodes de constitution de connaissances de base quel que soit le domaine, sous forme d'ontologie formelle, indispensable pour envisager une future autonomie de la machine dans le domaine considéré, on est fondé à penser que longtemps encore les logiciels réalisés sur ces machines seront conçus dans un objectif bien déterminé. Cette démarche guidée par le but est seule garante de la faisabilité de systèmes informatiques de complexité maîtrisée. Ainsi à capacité de calcul égale, on pourra différencier deux machines par leur aptitude à s'adapter à l'activité de ceux qui les utilisent. On retrouve ici la problématique du travail coopératif, avec deux partenaires capables de réaliser ensemble une tâche complexe. Viser un comportement coopératif pour la machine, c'est lui incorporer et faire vivre en son sein une représentation valide de ce qu'elle est elle-même capable de réaliser, mais aussi une représentation de son partenaire en termes cognitifs et sociaux et enfin une représentation du contexte... physique, social et organisationnel... dans lequel est située cette activité coopérative. Les problèmes posés par la poursuite de cet objectif sont très difficiles à résoudre et font l'objet de recherches pluridisciplinaires associant psychologues, informaticiens et sociologues.

Mario Borillo :

Ce qui se dégage factuellement des recherches actuelles, leurs méthodologies et leurs objectifs originaux, leurs résultats surprenants, appellent aussi, et de manière urgente compte tenu de la gravité de leurs enjeux, à une réflexion d'une autre nature, enracinée dans les faits et connaissances disponibles, mais plus aventurée puisque soucieuse de leurs projections sur le plan sociétal, éthique... Ce qui exige de développer sans tarder une profonde et dynamique philosophie de la cognition et du calcul associant de manière critique le biologique, le mental, le computationnel, mais aussi l'anthropologique et l'éthique.

Jean-Luc Soubie :

On a aujourd'hui du mal à discerner les limites de cette association homme-machine dans leur nature et dans leurs effets possibles. C'est pourtant de cette confrontation que viendront les évolutions les plus importantes de la machine vers l'autonomie.

Henri Prade :

La machine reste un outil — surpuissant — en principe au service de l'homme. Cependant de nouvelles questions émergent, difficilement imaginables du temps de Turing : en quoi une possibilité d'accès plus

massive à l'information, à des capacités de résolution de problèmes toujours accrues, peut-elle influencer sur nos manières de penser, de se comporter ? En quoi le compagnonnage personne/machine peut-il transformer l'homme — et la société —, le mettre en position d'assisté, le submerger par la vitesse de la machine, la quantité d'informations brassées, le rendre « homme machinal » ? Jusqu'où peuvent nous mener les couplages entre cerveau et ordinateur ou entre biologie et informatique ?

Mario Borillo :

Pour conclure sur mes réactions de scientifique — et d'homme tout court — à un théâtre mettant en scène un matériau scientifique de cette nature, il s'agit pour moi d'un événement précurseur, extrêmement important. Qui offre l'occasion rare de débattre en public sur des questions encore étrangères à la culture que nous partageons, susceptibles pourtant de changer nos « visions du monde » et de nous-mêmes — un questionnement métaphysique. Dont nous savons aussi qu'elles vont avoir, à travers la technique, le travail, l'économie, la culture, la guerre... un impact considérable sur la civilisation mais encore largement imprévisible. Comment ne pas souhaiter une confrontation indispensable, analogiquement comparable par sa difficulté et ses enjeux à celle qui s'est déjà instaurée pour la biologie et les biotechnologies. Mais en beaucoup plus complexe, puisque comportant une dimension abstraite essentielle.

Jean-Luc Soubie :

Le spectacle de Jean François Peyret pose au travers d'éléments de biographie et de dialogues philosophiques un ensemble de questions fondamentales, qui n'appellent ni ne reçoivent de réponse, mais dont l'immense mérite est de provoquer la réflexion de celui qui restera longtemps le seul capable de la mener : L'humain.

Henri Prade :

La science est une dimension intégrante de la culture. Au-delà d'un « théâtre d'idée » — un genre que le spectacle de Jean-François Peyret renouvelle complètement — la science peut contribuer à nourrir l'imaginaire de chacun, s'inscrire dans des recherches d'ordre esthétique. Cela ne peut que contribuer à sa vulgarisation, à la sortir d'un certain « ghetto », et ainsi à entretenir des relations fructueuses entre art et science

(Propos recueillis par Catherine Gadon, responsable de la Commission Culture de l'Université Paul Sabatier, Toulouse III).

Présentation des équipes I & A et IASI du LRI

Yves Kodratoff et Marie-Christine Rousset

1. Équipe I&A (Responsable : Yves Kodratoff)

L'équipe compte actuellement une quinzaine de personnes :

- Huit enseignants-chercheurs : Jacques Ales Bianchetti, Isabelle Bournaud, Lola Canamero, Antoine Cornuejols, Marta Franova, Yves Kodratoff, Claire Nédellec, Céline Rouveirol,
- Six doctorants : Érick Alphonse, Sébastien Augier, Jérôme Azé, Antoine Ducoulombier, David Faure, Mohammed Ould Abdel Vetah

Collaborations très actives avec Gilles Bisson (Grenoble), Adeline Nazarenko (LIPN), Michèle Sebag (X Palaiseau)

L'équipe s'intéresse à l'apprentissage automatique de connaissances par ordinateur, domaine dans lequel elle a au cours des années acquis une solide réputation internationale, comme le montre sa participation à de nombreux contrats européens et au comité de programme de la plupart des conférences sur ce thème.

L'apprentissage automatique comporte une étape délicate (et discutée) de raisonnement inductif, par lequel on invente des hypothèses qui rendent compte au mieux des données. Dans notre équipe, cette recherche se fait selon plusieurs thèmes principaux : l'induction symbolique-numérique, l'induction symbolique en programmation logique inductive, et l'induction symbolique en preuves par récurrence.

L'utilisation de l'induction comme un outil scientifique a toujours soulevé une grande opposition d'une partie de la communauté scientifique. Cette opposition a été contredite récemment par l'ampleur des applications industrielles de l'Extraction de connaissances à partir des données (ECD, souvent appelée « Data Mining ») qui fait systématiquement usage du raisonnement inductif.

Dès la naissance de cette approche, l'équipe y a participé. Un exemple de découverte de tels « gemmes » (« *nuggets* »), constituant une connaissance nouvelle, a été donné dans la thèse de Gilles Venturini.

Dans une base de données juridique, décrivant les prévenus, et décrivant aussi la conclusion donnée à leur affaire, nous avons immédiatement découvert que la règle suivante était très probable :

```
SI      descripteur_nationalité
        = français
ET SI   descripteur_emploi
        = a_un_emploi
ALORS  conclusion_de_l'affaire
        = sans_suite.
```

Ceci est une règle effectivement cachée dans les données puisqu'aucun juge ne l'utilise, et elle est donc impossible à « déduire » d'une connaissance juridique particulière.

Dans tous les problèmes d'extraction de connaissances à partir de textes (ECT, souvent appelée « *text mining* ») un problème majeur vient de ce que le nombre de termes est tellement élevé que les algorithmes d'ECD sont incapables de découvrir des relations entre termes. La solution consiste à regrouper les termes en classes, ou concepts, et à rechercher des relations entre concepts. C'est pourquoi l'équipe consacre une part importante de ses efforts à l'aide à la construction automatique de taxonomies à partir de données textuelles. Ce travail a donné lieu à la création du logiciel Asium sous la direction de Claire Nédellec.

La tendance actuelle de l'équipe est de chercher à se développer surtout dans le domaine de l'ECT, tout en conservant une activité non négligeable sur deux thèmes plus traditionnels pour l'équipe, c'est-à-dire la programmation logique inductive (axe animé par Céline

Rouveirol) et le raisonnement analogique (axe animé par Antoine Cornuéjols).

1.1. Thème : Méthodes symbolique-numériques pour l'extraction des connaissances

Ce thème regroupe toutes les méthodologies développées dans l'équipe, ou adaptées d'autres équipes, afin d'effectuer de la découverte de connaissances dans les données (ECD, appelée KDD : « *Knowledge Discovery in DataBases* » par les Anglo-Saxons).

Les deux points forts de l'équipe sont les algorithmes génétiques et l'apprentissage coopératif.

- Algorithmes génétiques

Cet axe est essentiellement animé par Yves Kodratoff, assisté de Michèle Sebag. Après une pointe d'activité dans ce domaine vers 1995, notre effort se porte sur la recherche d'une expression relationnelle à variables représentant (apprise à partir d') un ensemble d'exemples, avec cette restriction que seul le quantificateur universel est autorisé. La programmation génétique nous permet de travailler avec un nombre de données de l'ordre du millier ce qui autorise à envisager des applications pratiques de ces techniques. L'implémentation d'un tel système d'apprentissage de relations est terminée et elle fait appel de façon systématique à des algorithmes parallèles.

- Apprentissage coopératif et révision des connaissances

L'apprentissage coopératif est nécessaire lorsque la prise en compte de nouvelles situations amène à réviser les connaissances acquises ou apprises précédemment. Les méthodes de révision s'attachent à compléter ou à corriger ces connaissances de manière incrémentale, sans repartir de zéro. Par exemple, un robot, qui savait ouvrir les portes à battants, doit pouvoir compléter ses

connaissances en apprenant à ouvrir les portes à glissières à partir de ce qu'il savait déjà. Ces méthodes nécessitent souvent la participation d'un utilisateur humain. Son rôle peut être double, il peut concerner l'apport de nouvelles connaissances (par exemple, le concept de portes à glissières) et le contrôle des méthodes de révision en l'absence d'exemples significatifs (par exemple, choisir de considérer la porte à glissière comme une porte ou comme un obstacle mobile).

Une application récente de cette approche consiste à automatiser la construction des « ontologies », c'est-à-dire des relations sémantiques entre mots du langage naturel à partir de textes. Quelle que soit la méthode d'analyse du langage naturel utilisée (par exemple, un analyseur syntaxique), elle fera certaines fautes, et les erreurs induites doivent être corrigées par un utilisateur, ce qui constitue la part interactive de la construction. La part automatique, implémentée à l'heure actuelle, se fonde sur une méthode de classification conceptuelle ascendante. La construction des ontologies se réduit alors à la correction des connaissances extraites par le système sur un corpus de textes, et non plus, comme c'est le cas actuellement, à la création laborieuse de ces ontologies « à la main ».

L'utilisation de méthodes non fondées sur la classification conceptuelle ascendante sont également en cours d'étude.

Cette recherche, ainsi que ses applications à la fouille de textes va devenir un des thèmes de recherche principaux de l'équipe dans les années à venir.

- Outils d'ECT de nature linguistique

Notre approche à l'ECT fait usage d'un analyseur linguistique (par exemple, un analyseur syntaxique, ou un analyseur statistique) permettant de mettre en évidence des relations entre les composants du discours, par exemple les relations de nature syntaxique. Le logiciel Asium permet de passer de ces relations à une taxonomie de concepts. En d'autres termes, on passe d'une connaissance syntaxique ou statistique à une connaissance sémantique.

Nous désirons maintenant « mesurer » l'apport de chaque structure syntaxique à la sémantique des concepts.

Par exemple, nous avons constaté que la relation « sujet-verbe » apporte très peu d'information sémantique, en général (mais pas toujours !). Par contre la relation « nom-préposition-nom » tend à en apporter beaucoup (mais pas toujours !).

Il faut définir plus exactement la nature et la qualité de ces apports.

L'approche développée dans Asium donne la possibilité d'apprentissage de schémas prédictifs à partir d'exemples dans les textes. Un effort important commence à prendre place sur ce thème de recherche.

- Outils d'ECT de nature statistique

La majorité des outils existants sont relatifs à la détection des associations entre descripteurs d'une base de données. Nous étudions l'intérêt d'autres outils : détection d'associations selon la théorie de l'intensité de l'implication, selon des réseaux de causalité bayésiens, ou encore selon une nouvelle définition des associations causales.

1.2. Thème : Programmation Logique Inductive

Dans ce thème, l'équipe s'attache à montrer l'importance des connaissances implicites (les « biais ») dans la conduite des inductions dans le cas d'induction symbolique en programmation logique inductive. Ces techniques sont nécessaires quand on veut des règles exprimant non plus des valeurs de descripteurs mais des relations entre descripteurs, c'est-à-dire apprises dans une restriction de la logique des prédicats.

Par exemple, considérons la règle qui dit que lorsqu'une voiture en suit une autre et que la voiture suivie freine, alors la voiture suiveuse freine aussi. Cela peut s'écrire comme :

```
SI          SUIT (voiture2, voiture1)
ET SI      FREINE (voiture1)
ALORS     FREINE (voiture2).
```

On voit que la règle s'applique à

n'importe quelles voitures qui se suivent et non plus à deux voitures particulières. C'est pourquoi ces règles sont beaucoup plus complexes que celles qui ne font qu'utiliser des propriétés des objets concernés, et elles demandent la mise au point de techniques très particulières.

Les applications les plus prometteuses de ces techniques se font en Biologie Moléculaire et en Fouille de Données dans les bases de données relationnelles.

Les points forts de l'équipe sont :

- la configuration de systèmes de PLI

Chaque système d'apprentissage traite de façon particulière le problème de la complexité calculatoire liée à l'utilisation d'une restriction de la logique des prédicats, c'est ce qu'on appelle les « biais » du système. Nous cherchons à trouver une représentation uniforme de ces biais.

Ces travaux ont permis de développer les systèmes de PLI générique Haiku, et plus récemment Haiku-DB. Ce dernier est dédié à un problème typiquement mal posé : la découverte dans les bases de données multi-relationnelles.

- la formalisation de la PLI

L'algorithme DiVS, hybride de l'algorithme de l'Espace des Versions et de l'algorithme de l'Étoile, permet d'apprendre l'ensemble des clauses maximale généralement couvrant au moins un exemple positif et rejetant l'ensemble des exemples négatifs d'un concept cible. Cet algorithme a été formalisé et étendu à l'apprentissage de clauses définies, puis à l'apprentissage de clauses contraintes.

Nous travaillons également à l'exploitation dynamique de contraintes sur le langage de concept. La tâche d'apprentissage est décrite à un niveau abstrait par la donnée d'un ensemble de propriétés attendues pour la définition du concept cible (exigences concernant aussi bien la couverture ou le rejet d'exemples, que la syntaxe de la définition recherchée ou encore des propriétés plus complexes, comme des schémas d'instanciation imposés). Nous avons défini des opérateurs optimaux vis-à-vis

de ces contraintes.

2. Équipe « Intelligence Artificielle et Systèmes d'Inférences »

(Responsable : Marie-Christine Rousset)

L'équipe regroupe actuellement une quinzaine de personnes :

- Huit enseignants-chercheurs titulaires : Philippe Chatalic, Christine Froidevaux, Hélène Gagliardi, Marie-Christine Rousset, Chantal Reynaud, Brigitte Safar, Daniel Schlienger, Véronique Ventos.
- Une post-doctorante : Nathalie Pernelle.
- Cinq doctorants : Alain Bidault, Jean-Paul Forest, François Goasdoué, Laurent Simon, Alexandre Termier.

L'équipe IASI travaille depuis plus de quinze ans sur la conception et la construction de Systèmes à Base de Connaissances (SBC). Les problèmes qui nous intéressent plus particulièrement portent sur la modélisation et la représentation déclarative de connaissances complexes ou nombreuses et sur la conception d'algorithmes d'inférence permettant de mécaniser des raisonnements sur ces connaissances.

Depuis trois ans, certains travaux de l'équipe se sont infléchis vers une combinaison de techniques d'Intelligence Artificielle et de Bases de données pour traiter des problèmes nouveaux suscités par l'explosion de l'accès à un grand nombre de sources d'information via le Web. Ces travaux s'articulent autour d'un thème fédérateur centré sur le problème de l'intégration de sources d'informations multiples et hétérogènes et se déclinent au travers de plusieurs projets effectués en collaboration avec différents partenaires académiques ou industriels. Ils sont le support d'un rapprochement avec le groupe Verso de l'INRIA-Rocquencourt.

Un autre axe de travail de l'équipe, initialisé par une collaboration avec EDF R&D (centre de Clamart), mené par Philippe Chatalic et Laurent Simon, concerne l'utilisation du langage de la logique propositionnelle comme support pour la modélisation de connaissances. Bien que d'un faible pouvoir

d'expression, la simplicité de ce formalisme autorise la mise en œuvre d'algorithmes de raisonnement efficaces. Initialement motivés par une problématique de validation de connaissances, nos travaux actuels se tournent vers le problème central de la satisfiabilité d'une formule booléenne. Nous nous intéressons tout particulièrement à la prise en compte de la structure des instances considérées, d'une part en utilisant des structures de données adaptées pour représenter de grands ensembles de clauses (tris, bdds) et d'autre part en étudiant les capacités des algorithmes à exploiter cette structure. Un important travail d'évaluation expérimentale et de comparaison des différents algorithmes existants, testés sur les grandes classes d'instances habituellement utilisées dans la communauté, est également en cours de réalisation. Ces travaux sont décrits dans [10], [11], [12], et [13].

Très récemment, certains membres de l'équipe (Christine Froidevaux et Jean-Paul Forest) ont commencé à travailler sur l'étude de techniques d'Intelligence Artificielle pour la détection automatique de gènes dans des séquences génomiques.

Cette présentation est focalisée sur les travaux relatifs au problème de l'intégration d'informations, qui est le thème fédérateur principal de l'équipe. Le lecteur intéressé par une présentation exhaustive et synthétique des thèmes de l'équipe est renvoyé à la page Web située à l'adresse : <http://www.lri.fr/Francais/Recherche/iasi.html>.

Le problème de l'intégration d'informations porte sur l'agrégation et la combinaison de réponses à des requêtes posées à différentes sources d'information (pouvant être réparties et hétérogènes) pour répondre à une requête globale de l'utilisateur. Ce problème se pose de façon cruciale face à l'explosion de l'accès à un grand nombre de sources rendant nécessaires des outils de médiation intelligents entre les utilisateurs et les sources d'informations existantes stockées ou accessibles par le Web. Les techniques de représentation de connaissances sont très utiles pour la

construction de tels outils. En particulier, certains langages de représentation de connaissances peuvent être utilisés pour définir de façon déclarative et flexible des vues abstraites au-dessus des données. Raisonner sur ces vues permet de comparer le contenu sémantique des différentes sources de données et ainsi de déterminer lesquelles sont pertinentes pour la construction du plan global de requêtes. Plus précisément, si les vues représentent les requêtes spécialisées exécutables sur les sources de données, la construction du plan global de requêtes est un problème de réécriture de la requête globale en fonction de vues.

Nous participons à plusieurs projets de recherche (**PICSEL**, **GAEL** et **Xyleme**) en collaboration avec différents partenaires du monde industriel et académique. Nous y abordons le problème de l'intégration d'information dans différents contextes et selon différents angles d'attaque.

PICSEL (<http://www.lri.fr/~picsel>) est le résultat d'une collaboration avec le centre de recherche de France Telecom qui s'est concrétisée par une convention de recherche de 3 ans (décembre 1997-décembre 2000) autour du projet PICSEL. L'objectif de PICSEL est de fournir un environnement souple et déclaratif de développement de serveurs d'information fondé sur la construction de médiateurs entre les utilisateurs et un ensemble de sources d'informations pré-existantes relatives à un même domaine d'application (e.g., le tourisme).

Un médiateur regroupe dans un cadre unificateur et homogène la description d'un domaine (appelée *ontologie*) et le contenu des différentes sources relatives à ce domaine. Ainsi, l'utilisateur n'a pas à connaître le format ni le contenu des sources disponibles relativement à son domaine d'intérêt et a l'illusion qu'il interroge un système centralisé et homogène. C'est la tâche du *moteur de requêtes* de prendre en entrée la requête de l'utilisateur et de déterminer l'ensemble des plans de requêtes spécialisés à exécuter sur les sources pour obtenir et

éventuellement combiner l'ensemble des réponses à la requête initiale compte tenu des sources disponibles et de leur contenu. Le domaine du tourisme a été choisi comme domaine d'expérimentation de PICSEL, en collaboration avec l'agence de voyage Degriftour.

L'intérêt de l'approche PICSEL est de séparer la construction de l'ontologie (relative à un domaine d'intérêt) de la mise en œuvre du moteur de requêtes, en se fondant rigoureusement sur une approche logique de la représentation des connaissances. Le moteur de requêtes est générique et peut être réutilisé pour le développement de serveurs d'informations dans des domaines variés. Un environnement convivial est offert au développeur de serveurs d'informations pour lui permettre de décrire le domaine d'application visé, sans avoir à coder dans un langage informatique. En particulier, l'ontologie du domaine peut être facilement mise à jour et la hiérarchie des concepts qui en découle automatiquement calculée et visualisée. L'approche et les algorithmes, fondés sur la modélisation du domaine, du contenu des sources et des requêtes de l'utilisateur en CARINALN, ont été décrits dans plusieurs articles de revues ou de conférences ([1], [2], [3], [4], [5], [6]).

L'environnement de description d'ontologies, les outils de classification automatique et de visualisation, ainsi que le moteur de requêtes ont été implémentés en Java et ont donné lieu au dépôt de deux logiciels ONTOCLASS et ONTOQUERY. Une ontologie du domaine du tourisme contenant près de 500 concepts et propriétés a été également construite.

Par ailleurs, nous avons dès le départ conçu PICSEL comme un environnement coopératif d'interrogation de sources d'informations. En particulier, nous avons étudié le problème de la relaxation de requêtes quand le médiateur détermine au niveau du calcul des plans de requêtes qu'il ne peut pas fournir de réponses à la requête de l'utilisateur. Ce travail a donné lieu à plusieurs publications ([7], [8], [9]).

GAEL est un projet effectué en colla-

boration avec le groupe VERSO de l'INRIA et une PME (Matchvision) spécialisée dans le commerce électronique. Ce projet est financé dans le cadre du RNRT (Réseau National de Recherche en Télécommunications). Il a démarré il y a moins d'un an. L'objectif général de GAEL est de concevoir un générateur de catalogues électroniques et de services d'interrogation associés, permettant à des non informaticiens de développer leurs propres catalogues électroniques de façon souple et flexible au dessus des bases de données stockant une description brute des produits qu'ils désirent mettre en ligne. Dans ce projet, nous prônons une approche déclarative fondée sur une spécification logique des classes de produits et des requêtes pour y accéder. Les techniques de représentation de connaissances sont alors au service du filtrage des données pour les structurer de façon adaptative en fonction des utilisateurs. Les données brutes sont cachées à l'utilisateur à qui on préfère donner une vision abstraite, condensée, qualitative, plus signifiante pour lui. Le problème est d'extraire *automatiquement* à partir des données brutes les connaissances qui traduisent ces données à un niveau d'abstraction plus pertinent pour un utilisateur humain ou pour des outils d'aide à la décision. Les données sont alors au service de la représentation de connaissances via des techniques de « *data mining* ».

Xyleme (<http://www-rocq.inria.fr/verso/LEVEL1/Xyleme.html>) est un projet très ambitieux, lancé par le groupe VERSO de l'INRIA en septembre 1999, et qui regroupe plusieurs équipes universitaires françaises et allemandes. L'objectif de Xyleme est de regrouper dans un immense entrepôt de données l'ensemble de tous les documents du Web de demain et de construire un moteur de recherche en rupture avec les moteurs de recherche actuels afin de dépasser leurs limites intrinsèques. Xyleme repose sur l'hypothèse que le langage XML va devenir dans un avenir proche le langage de définition et de description des données du Web. En levant les limitations du langage HTML, ce langage ouvre la voie à de

nouveaux services sur le Web. Il devient possible notamment de définir des langages de requêtes très riches qui dépassent de loin les requêtes par mots-clés que proposent les moteurs de recherche actuels.

Un document XML est dit semi-structuré car il consiste en un texte contenant un certain nombre de balises librement choisies pour leur signification et facilement repérables syntaxiquement dans le document. Les balises et leur structure d'imbrication sont décrites au travers de DTDs (Document Type Definition). Par exemple, un document décrivant une personne pourra se conformer à un DTD indiquant que le nom et le(s) prénom(s) de la personne se trouvent dans la partie du document délimitée par les balises < identité> et </identité>, et qu'à l'intérieur de cette zone, on trouvera le nom entre les balises < nom> et </nom> et les prénoms entre < prénom> et </prénom>. On pourra indiquer qu'on accepte qu'il n'y ait pas de prénoms ou plusieurs prénoms, mais qu'il doit y avoir un et un seul nom. Ensuite, on pourra trouver une zone de texte délimitée par les balises < adresse> et </adresse> pouvant elle-même être découpée en différents sous-éléments décrivant une adresse. Les DTDs et leurs balises fournissent une structure aux documents XML qui s'y conforment tout en permettant une certaine souplesse (certaines balises sont optionnelles, d'autres peuvent être répétées un nombre arbitraire de fois non précisé à l'avance). En même temps, un document XML contient beaucoup de texte entre ses balises et ce texte peut être analysé, indexé et interrogé par des techniques utilisées en recherche documentaire.

Notre contribution dans Xyleme est de mettre en œuvre des méthodes de regroupement automatique des documents présents dans l'entrepôt en classes de documents traitant du même sujet. Chaque classe représente un domaine qu'il s'agit de caractériser par un ensemble de descripteurs pertinents. Ces descripteurs, eux mêmes structurés, constituent les points d'entrée de l'interface sémantique d'interrogation de Xyleme.

PRÉSENTATION DE LABORATOIRES

Bibliographie (équipe IASI) :

- [1] Alon Levy and Marie-Christine Rousset, *Combining Horn Rules and Description Logics in CARIN*, Artificial Intelligence Journal, vol 104, September 1998.
- [2] François Goasdoué and Chantal Reynaud, *Modeling Information Sources for Information Integration*, 11th European Workshop on Knowledge Acquisition, Modeling and Management, EKAW'99, p. 121-138, Lecture Notes in AI 1621, Springer-Verlag, Dagstuhl Castle, Germany, May 26-29, 1999.
- [3] Marie-Christine Rousset, *Backward Reasoning in Aboxes for Query Answering*, Proceedings of the International Workshop on Description Logics (DL'99), Linköping, Sweden, July 99, Also in Proceedings of the 6th International Workshop on Knowledge Representation meets Databases (KRDB'99), Linköping, Sweden, July 99.
- [4] Marie-Christine Rousset, *Query Expansion in Description Logics and CARIN*, Proceedings of the AAAI Fall Symposium on Question Answering Systems, Cape Cod, November 99.
- [5] François Goasdoué and Marie-Christine Rousset, *Rewriting Conjunctive Queries using Views in Description Logics with Existential Restrictions*, Proceedings of the International Workshop on Description Logics (DL'00), Aachen, Germany, 2000.
- [6] François Goasdoué, Véronique Lattès and Marie-Christine Rousset, *The Use of CARIN Language and Algorithms for Information Integration : The PICSEL System*, International Journal of Cooperative Information Systems (IJCIS), à paraître.
- [7] Alain Bidault et Brigitte Safar, *Réparation de requêtes posées à un médiateur*, 12ème Congrès Francophone de reconnaissance des Formes et Intelligence Artificielle. RFIA 2000 (1) 225-235, Paris, février 2000.
- [8] Alain Bidault, Christine Froidevaux and Brigitte Safar, *Repairing Queries in a Mediator Approach*, 14th European Conference on Artificial Intelligence, ECAI 2000, Berlin, August 2000
- [9] Alain Bidault, Christine Froidevaux and Brigitte Safar, *Finding Successful Queries in a Mediator Context*, 4th International Conference on Flexible Query Answering Systems. FQAS 2000.
- [10] Ph. Chatalic et L. Simon, *Davis et Putnam 40 ans après : une première expérimentation*, Journées Nationales sur la Résolution Pratique de Problèmes NP-Complets, juin 1999.
- [11] Ph. Chatalic and L. Simon, *ZRes : the old DP meets ZBDDs*, in Proceedings of the 17th Conference of Automated Deduction (CADE), juin 2000
- [12] Ph. Chatalic and L. Simon, *Résolution sur des ensembles de clauses compressés : comment doper DP*, Journées Nationales sur la Résolution Pratique de Problèmes NP-Complets, juin 2000.
- [13] Ph. Chatalic and L. Simon, *Multi-Resolution on Compressed Sets of Clauses*, Twelfth International Conference on Tools with Artificial Intelligence, ICTAI-2000, Vancouver, 2000.

Présentations de laboratoires dans le bulletin de l'AFIA

LIFIA, Grenoble	Bulletin n°1	IRISA, INRIA et Université de Rennes.....	Bulletin n°13
LRI, Orsay	Bulletin n°1	Société INGENIA	Bulletin n°14
Service Systèmes Experts, Renault.....	Bulletin n°1	LIPN, Université de Paris Nord.....	Bulletin n°14
CEDIAG, Bull.....	Bulletin n°2	Institut EURISCO	Bulletin n°15
CERT, ONERA, Toulouse	Bulletin n°2	LRDC, Université de Pittsburgh (USA)	Bulletin n°15
IRIT, Toulouse.....	Bulletin n°2	Société ISOFT.....	Bulletin n°16
LAAS, Toulouse.....	Bulletin n°2	Dépt. d'Info de l'Université d'Ottawa.....	Bulletin n°16
HEUDIASYC, UTC	Bulletin n°3	Equipe CHM, Université du Colorado (USA)	Bulletin n°17
IFP, Rueil Malmaison	Bulletin n°3	LIRMM, Montpellier	Bulletin n°19
DIAM, INSERM U194.....	Bulletin n°3	Institut autrichien de recherches en IA.....	Bulletin n°20
Lab. Math. Info., Fac Médecine de Marseille	Bulletin n°4	ENST Bretagne	Bulletin n°21
GMD, St. Augustin (RFA)	Bulletin n°4	LIA - Université de Savoie	Bulletin n°22
ONERA, Chatillon	Bulletin n°4	INRETS.....	Bulletin n°23
KSL, Université de Stanford (USA).....	Bulletin n°5	IRIN Nantes	Bulletin n°24
Dépt Applications de l'IA au CNET, Lannion	Bulletin n°5	CRIN - INRIA Lorraine.....	Bulletin n°25
LAFORIA, Univ. Pierre et Marie Curie	Bulletin n°6	DIRO - Université de Montréal	Bulletin n°26
L'institut FAW, ULM (RFA).....	Bulletin n°6	IRIT - Toulouse (1)	Bulletin n°28
Institut IIIA, Compiègne.....	Bulletin n°6	IRIT - Toulouse (2)	Bulletin n°29
LAIR, OHIO State University (USA)	Bulletin n°7	LAAS - Toulouse (1)	Bulletin n°30
ARAMIHS, Labo mixte MATRA-CNRS, ..	Bulletin n°7	Sony CSL	Bulletin n°31
CEA, Service SERMA, Saclay	Bulletin n°8	LAAS - Toulouse (2)	Bulletin n°32
Société ILOG	Bulletin n°8	LIMSI - Département CHM	Bulletin n°33
LAIAC, Université de Caen	Bulletin n°9	LAMSADE	Bulletin n°34
Institut Français du Pétrole	Bulletin n°10	Institut autrichien de recherches en IA.....	Bulletin n°36
DFKI (Centre allemand de recherches en IA)	Bulletin n°11	LIP6 (Univ. Pierre et Marie Curie)	Bulletins n°37 & 38
GRTC, Marseille	Bulletin n°11	GREYC (Univ. de Caen)	Bulletin n°40
Inst. d'Analyse des Systèmes, Ac. Russe	Bulletin n°12	LIFL (Univ. de Lille)	Bulletin n°41
Georges Mason Univ., Center for AI (USA)	Bulletin n°13	LRI (équipes IA et IASI)	Bulletin n°43

Dossier Connexionnisme, Apprentissage Numérique

Coordonné par Patrick Gallinari, LIP6 (Patrick.Gallinari@lip6.fr)

Introduction

En tant que méthodes d'apprentissage, les réseaux de neurones ont marqué les années quatre-vingt-dix en proposant de nouveaux paradigmes pour le calcul, de nouvelles méthodes de modélisation, et en réintroduisant les méthodes numériques au sein de l'IA. Le domaine s'est développé extrêmement vite en proposant à la fois des méthodes pratiques pour résoudre des problèmes de grande taille et en permettant l'essor de travaux théoriques sur les fondements de l'apprentissage et le développement de nouvelles classes de méthodes. Les travaux sur les réseaux de neurones ont ainsi servi de base pour de nouvelles directions et de nouveaux domaines en apprentissage. Aujourd'hui, la communauté des réseaux de neurones fait partie intégrante des communautés « apprentissage » et « fouille de données » tout en ayant conservé certaines de ses spécificités.

Un autre grand domaine de recherche est l'utilisation de ces modèles pour l'étude de phénomènes biologiques et cognitifs. Les réseaux de neurones y sont employés comme outils de modélisation pour étudier des phénomènes cognitifs, des mécanismes de traitement de l'information sensorielle, les repré-

sentations des différents types d'information mis en jeu dans ces mécanismes. Inversement, les modèles de calcul s'inspirent des avancées en neurosciences et en sciences cognitives pour proposer de nouvelles méthodes de traitement des données, de nouveaux systèmes de traitement de l'information.

Les réseaux de neurones se sont développés comme un domaine pluridisciplinaire avec dans le domaine de l'apprentissage des contributions provenant des mathématiques (optimisation, statistiques, systèmes non linéaires, théorie de la décision), de la physique statistique, de l'informatique (intelligence artificielle, reconnaissance des formes). En modélisation, il existe une interaction importante entre créateurs de modèles et chercheurs en neurosciences et sciences cognitives. Cette pluridisciplinarité est toujours présente et se reflète dans la présentation des équipes qui figurent dans ce dossier.

Les applications ont été centrales pour le développement du domaine et ont donné lieu à de nombreuses réalisations qui sont réellement opérationnelles. Des applications ont été réalisées dans la plupart des domaines traditionnels de l'ingénierie, pour le diagnostic, la prévision, la commande, la discrimination, la visualisation de données, etc.

Des simulations de grande ampleur de phénomènes physiques ont également été réalisées. Plus récemment, les applications liées à l'essor du Web, des grandes bases de données ont été explorées, et les réseaux de neurones sont largement utilisés pour la fouille de données, la recherche et l'extraction d'information textuelle ou multimédia, la visualisation, les télécommunications, etc.

Le bulletin de l'AFIA avait édité un dossier « connexionnisme » (bulletin n° 11) alors que le domaine en était encore à ses débuts en France. Ce dossier est l'occasion de faire le point sur l'évolution du domaine et l'avenir de ces outils, sur les applications qui ont été réalisées, sur les travaux théoriques dans le domaine de l'apprentissage avec en particulier les liens avec la communauté COLT, sur les méthodes issues des travaux sur les réseaux de neurones comme les machines à vecteurs supports, sur les applications dans les nouvelles technologies.

Pour ceux qui recherchent des pointeurs sur les travaux du domaine, les équipes, les logiciels, un bon point de départ est le site du réseau d'excellence NEuroNet spécialisé bien sûr sur les réseaux de neurones :

<http://www.kcl.ac.uk/neuronet/>

Liste des contributions

CEA Grenoble, Département de Recherche Fondamentale sur la Matière Condensée, SPSMS / Groupe Théorie
E3I, Laboratoire d'Informatique de l'Université de Tours
ERIC, pôle connexionnisme, Université Lyon 2
ESPCI - École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris - Laboratoire d'Électronique
France Télécom R&D, Groupe TNT
HEUDIASYC, UMR 6599, U.T. Compiègne, thème diagnostic
INRA - Unité de Biométrie et Intelligence Artificielle, Equipe Méthodes Mathématiques et Informatiques pour la Décision

INT, Département EPH, Equipe Réseaux de Neurones et Applications Télécom
IPSL - Neuratel - Equipe Modélisation et méthodes statistiques avancées, groupe de travail Neuratel de l'Institut Pierre Simon Laplace
ISC, Equipe « Connexionnisme et Modélisation Cognitive »
LEIBNIZ-IMAG - Equipe Réseaux d'Automates
LIP6 - Equipe Apprentissage numérique, Université Paris 6
LIPN - Equipe : ADAge : « Apprentissage, Diagnostic, Agents », Université Paris 13
LORIA/INRIA-Lorraine - Equipe CORTEX
PSI, INSA de Rouen, Thème Classification et apprentissage
SAMOS-MATISSE, UMR 8595, Université Paris 1

**CEA Grenoble
Département de Recherche
Fondamentale sur la Matière
Condensée, SPSMS/
Groupe Théorie**

17, rue des Martyrs
38054 Grenoble Cedex 9
Contact : Mirta B. Gordon
mirta.gordon@cea.fr
Tél. 04 76 88 43 30

Chercheurs concernés par le thème :
Mirta B. Gordon, Sebastian Risau-
Gusman, Christelle Godin.

Cadre général et objectifs, problématique

Nos travaux portent sur les propriétés d'apprentissage (supervisé et non supervisé) et de généralisation des systèmes adaptatifs, en particulier des réseaux de neurones mais aussi d'autres paradigmes, comme les algorithmes génétiques et les Machines à Exemples Support. Nous nous intéressons plus particulièrement aux propriétés typiques des grands réseaux, ou plus généralement, de systèmes avec un grand nombre de paramètres adaptables. Dans ces cas, on peut prédire les propriétés avec les outils théoriques de la Physique Statistique. Bien qu'essentiellement théoriques, nos études ont eu des retombées pratiques, car nous avons découvert un algorithme d'apprentissage très performant pour des neurones binaires. Une de ses caractéristiques est qu'il permet de trouver automatiquement la solution de marge maximale qui n'est autre que la solution recherchée avec les Machines à Exemples Supports, avec une méthode alternative. Ces Machines constituent une des voies prometteuses issues des recherches sur les réseaux de neurones.

Panorama des travaux effectués par l'équipe sur le thème

Nous avons étudié les propriétés de généralisation de réseaux de neurones entraînés utilisant la minimisation de différentes fonctions de coût. Nous avons montré que, dans le cas d'un perceptron apprenant un problème linéairement séparable, le classifieur optimal (bayésien) peut être obtenu par minimisation d'une fonction de coût. Nous

avons déterminé une équation différentielle qui permet de calculer cette fonction de coût, et nous avons corroboré les prédictions avec des simulations numériques. Nous avons aussi calculé la *capacité de réseaux à une couche cachée* obtenus par *apprentissage constructif*. Nous trouvons que, pourvu que l'apprentissage des neurones successivement incorporés au réseau minimise le nombre d'erreurs d'apprentissage, les algorithmes constructifs considérés introduisent dans la couche cachée le nombre de neurones *minimal* permettant d'apprendre les exemples. Actuellement, nous étudions les propriétés des Machines à Exemples Support avec les outils de la Physique Statistique.

Références

Mirta B. Gordon, D. R. Grempel, Learning with a temperature dependent algorithm, Europhys. Lett. 29 (1995) 257-262.
Bruno Raffin, Mirta B. Gordon, Learning and generalization with Minimerror, a temperature dependent learning algorithm, Neural Computation 7 (1995) 1206-1224.
Arnaud Buhot, Mirta B. Gordon, Cost function and patterns distribution of the bayesian perceptron, Physics Letters A 228 (1997) 73-78.
Mirta B. Gordon et Hélène Paugam-Moisy, Sciences cognitives — Diversité des approches, Collection : Interdisciplinarité et nouveaux outils. Hermès, 1997.
Arnaud Buhot, Juan Manuel Torres Moreno, Mirta B. Gordon, Finite size scaling of the bayesian perceptron, Phys. Rev. E 55 (1997) 7434-7440.
Juan-Manuel Torres Moreno, Mirta B. Gordon, Characterization of the Sonar Signals Benchmark, Neural Processing Letters 7 (1998) 1-4.
Juan Manuel Torres Moreno, Mirta B. Gordon, Efficient adaptive learning for classification tasks with binary units, Neural Computation 10 (1998) 1017-1040.
Arnaud Buhot, Mirta B. Gordon, Phase transitions in optimal unsupervised learning, Phys. Rev. E 57 (1998) 3326-3333.
Sebastian Risau-Gusman, Mirta B. Gordon, Understanding stepwise generalization of Support vector Machines : a toy model, in Advances in Neural Information Processing Systems 12, S. A. Solla, T. K. Leen and K.-R. Müller (eds.), MIT Press (2000) 321-327.
Arnaud Buhot, Mirta B. Gordon, Storage capacity of a constructive learning algorithm, J. Phys. A 33 (2000) 1713-1727.

**E3I, Laboratoire
d'Informatique
de l'Université de Tours**

École d'Ingénieurs en Informatique pour l'Industrie
64, avenue Jean Portalis
37200 TOURS

Contact : M. Crucianu
téléphone : 02 47 36 14 14
télécopie : 02 47 36 14 22
Mél : crucianu@univ-tours.fr, bone@univ-tours.fr, jean-pierre.asselin@auf-francophonie.org

Chercheurs concernés par le thème

Jean-Pierre Asselin de Beauville (professeur, en détachement auprès de l'AUF)

Michel CRUCIANU (maître de conférences, E3i – Université de Tours)
Romuald Boné (docteur)

Cadre général et objectifs, problématique

Nous nous sommes intéressés essentiellement à la mise au point de modèles de séries temporelles à partir de données, ainsi qu'à l'évaluation de ces modèles. Dans ce domaine d'utilisation, les réseaux de neurones étaient confrontés à quelques difficultés notables. Nous avons concentré nos efforts sur l'exploitation des connaissances a priori concernant les séries, sur la prise en compte des dépendances à long terme et sur l'évaluation de la qualité locale des modèles obtenus.

Panorama des travaux effectués par l'équipe sur le thème

Les avantages potentiels des réseaux de neurones récurrents sont difficiles à mettre en évidence à cause de l'incapacité des algorithmes de descente de gradient à prendre en compte les dépendances à moyen et long terme présentes dans les données. Les algorithmes constructifs que nous avons développés [1], [2] arrivent à surmonter cette difficulté grâce à l'ajout très ciblé d'un nombre réduit de connexions à retard. Les nombreuses évaluations expérimentales montrent que les paramètres supplémentaires de ces algorithmes peuvent être choisis selon des règles simples.

Dans toute tentative de modélisation

à partir de données, l'exploitation des connaissances a priori concernant le problème permet d'améliorer les performances grâce à une réduction éclairée de l'espace des modèles candidats. Nous avons proposé une technique générale qui permet, dans un cadre bayésien, de tirer profit des connaissances a priori [3].

Ce même cadre bayésien nous offre un moyen d'évaluer la qualité de chacune des prédictions individuelles d'un modèle, à travers des intervalles de confiance qui prennent en compte le bruit présent dans les données et la variabilité des modèles. Nous avons étendu l'apprentissage bayésien à l'utilisation des réseaux de neurones récurrents et non récurrents pour des problèmes temporels [3], [4].

Références

- [1] Boné, R., Crucianu, M., Asselin de Beauville, J.-P. (2000) Two constructive algorithms for improved time series prediction with recurrent neural networks, *IEEE International Workshop on Neural Networks for Signal Processing*, December 11-13, 2000, Sydney, Australia.
- [2] Boné, R., Crucianu, M., Asselin de Beauville, J.-P. (2000) An Algorithm for the Addition of Time-Delayed Connections to Recurrent Neural Networks, *Proceedings of ESANN 2000*, April 26-28, 2000, Bruges, Belgium, pp. 293-298.
- [3] Crucianu, M., Boné, R., Asselin de Beauville, J.-P. (2000) Bayesian learning for recurrent neural networks, *Neurocomputing Letters*, accepté pour publication.
- [4] Crucianu, M., Boné, R., Asselin de Beauville, J.-P. (1999) Bayesian learning for time series prediction with exogenous variables, *Proceedings of IJCNN'99*, Washington DC, July 10-16, 1999.

Laboratoire ERIC, Pôle Connexionnisme

Université Lumière Lyon 2
5 avenue Pierre Mendès-France,
69676 Bron Cedex

Contact : Hélène Paugam-Moisy (H. Paugam Moisy dirige également une équipe à l'Institut des Sciences Cognitives de Lyon - ISC -, présentée plus loin dans le dossier)
04 78 77 31 51 — hpaugam@univ-lyon2.fr

Chercheurs concernés : André ELISSEFF, Olivier TEYTAUD, Hélène PAUGAM-MOISY

Panorama des travaux effectués par l'équipe sur le thème

Connexionnisme et théorie de l'apprentissage :

Cette thématique est développée à ERIC depuis deux ans, en particulier dans le cadre du Working Group européen NeuroCOLT. Partant de la théorie statistique de l'apprentissage de Vapnik, nous nous intéressons à la complexité des réseaux de neurones et du modèle SVM. Nous avons établi des bornes d'erreur réalistes et efficaces en pratique, en utilisant des nombres de couverture. Nous nous sommes intéressés à la discrimination multi-classe et par une méthode d'étude directe, nous avons défini des SVM multi-classes. Nous avons obtenu des résultats théoriques sur le comportement asymptotique des SVM, en termes de nombre de vecteurs de supports, et nous avons comparé la qualité des séparations obtenues à la classification bayésienne. Nous nous sommes également penchés sur la notion de stabilité des algorithmes d'apprentissage.

Mise en œuvre et applications :

Nous avons implanté des algorithmes efficaces de SVM en appliquant ou redéfinissant diverses méthodes pour l'optimisation (méthode SMO, Chunking, algorithme de Franck et Wolfe). Nous appliquons ces méthodes à des problématiques industrielles et de recherche, notamment en fouille de données, en imagerie médicale et en Physique.

Références

- Elisseff, Paugam-Moisy, (1999), « JNN, a randomized algorithm for learning multilayer networks in polynomial time », *NeuroComputing*, 29:3-24.
- Guermeur, Paugam-Moisy, (1999), « Théorie de l'apprentissage de Vapnik et SVM, Support Vector Machines », dans *Apprentissage Automatique*, G. Venturini et M. Sebban édés., Hermès, 109-138.
- Droniou, Elisseff, Paugam-Moisy, Teytaud, (1999), « Contrôle de l'architecture et des

représentations internes dans les réseaux de neurones multicouches », conférence CAp'99, Palaiseau, 185-194.

Smola, Elisseff, Scholkopf, Williamson, (1999), « Entropy numbers for convex combinations and multilayer networks », in *Advances in Large Margin Classifiers*, MIT Press.

Guermeur, Elisseff, Paugam-Moisy, (2000), « A new multiclass SVM based on a uniform convergence result », *IJCNN'2000*, IV-183-188.

ESPCI École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris Laboratoire d'Électronique

10 rue Vauquelin, 75005 PARIS
Tél. : 01 40 79 45 41 ; fax : 01 47 07 13 93

Contact : Gérard Dreyfus
Gerard.Dreyfus@espci.fr
http://www.neurones.espci.fr

Chercheurs permanents concernés par le Thème : Ginette Bossavit, Directeur de Recherches au CNRS, Gérard Dreyfus, Professeur à l'ESPCI, Directeur du Laboratoire, Howard Gutowitz, Maître de Conférences à l'ESPCI, Claudine Masson, Directeur de Recherches au CNRS, Yacine Oussar, Maître de Conférences à l'ESPCI, Olivier Parodi Directeur de Recherches au CNRS, Brigitte Quenet, Maître de Conférences à l'ESPCI, Pierre Roussel, Maître de Conférences à l'ESPCI

Cadre général et objectifs, problématique

Deux axes de recherche :

Modélisation et apprentissage pour les Sciences de l'Ingénieur : nos recherches ont conduit à des développements théoriques nouveaux concernant la conception de modèles non linéaires (statiques ou dynamiques) par apprentissage, et à de nombreuses réalisations dans les milieux de l'industrie et de la finance

Neurosciences computationnelles : nos travaux actuels visent essentielle-

ment à une meilleure compréhension du codage des informations dans les systèmes nerveux.

Panorama des travaux effectués par l'équipe sur ce thème

1) Modélisation et apprentissage pour les Sciences de l'Ingénieur :

L'introduction et la formalisation de la technique de modélisation neuronale de connaissances ou modélisation neuronale semi-physique, constitue un de nos résultats majeurs des dernières années : elle permet d'incorporer les connaissances physiques (ou chimiques, biologiques, etc.), disponibles sous forme d'équations algébriques ou différentielles, dans la structure d'un réseau de neurones.

Parmi les autres résultats d'ordre méthodologique obtenus récemment au Laboratoire, citons

- des méthodes originales de sélection de modèles, notamment celle du *leave-one-out* virtuel ;
- la détermination automatique des variables pertinentes dans la modélisation d'un processus non linéaire (méthode du vecteur sonde) ;
- la définition et la mise en œuvre d'une forme canonique des modèles dynamiques non-linéaires ;
- des méthodes originales d'initialisation pour l'apprentissage de réseaux d'ondelettes ;

Ces progrès méthodologiques ont été reconnus par de nombreux industriels français et étrangers, avec lesquels nous avons mené à bien des projets appliqués conduisant à des réalisations opérationnelles :

- la modélisation neuronale de connaissances d'un procédé de séchage (collaboration avec 3M États-Unis) ;
- la prédiction de la qualité de soudures par points (collaboration avec SOL-LAC) ;
- la prédiction des propriétés mécaniques de caoutchoucs pour pneumatiques (collaboration avec MICHELIN) ; le modèle neuronal de connaissances mis au point dans le cadre de ces études est utilisé par plusieurs usines du groupe pour la mise au point de nouveaux matériaux ;
- le traitement automatique des dépêches de l'Agence France-Presse : création automatique de filtres d'infor-

mation en fonction de profils d'intérêt de l'utilisateur (application opérationnelle au sein du groupe de la Caisse des Dépôts et Consignations) ;

- la modélisation de réseaux d'hydrologie urbaine (collaboration avec la Direction de l'Eau et de l'Assainissement) ;
- la modélisation d'une colonne de distillation industrielle pour la détection précoce de dysfonctionnements (en collaboration avec ELF-ATOCHEM et NETRAL).

2) Neurosciences computationnelles :

appuyée sur plusieurs collaborations avec des laboratoires français et étrangers, cette activité a donné lieu à des travaux concernant notamment :

- la morphométrie, c'est-à-dire l'étude de la forme de l'arbre dendritique des neurones, forme dont l'impact sur la fonction des neurones constitue un important sujet de recherche dans la communauté des neurosciences ; il a été possible de montrer que, en dépit de la très grande complexité apparente de l'arbre dendritique, sa morphologie peut être décrite, de manière parcimonieuse, par un très petit nombre de paramètres.
- l'analyse de données électrophysiologiques par corrélation avec des données de génétique moléculaire (en collaboration avec l'équipe de Jean Rossier) ;
- l'étude des mécanismes de codage et de traitement de l'information sensorielle par les réseaux de neurones biologiques, qui constitue l'axe essentiel de nos recherches actuelles. Ces travaux ont été menés sur deux supports d'études : le système olfactif (notamment celui des invertébrés) et le système de communication des poissons électriques. L'intégration de données anatomiques, physiologiques et comportementales a permis notamment d'élaborer un modèle, entièrement traitable analytiquement, d'un étage de traitement de l'information du système olfactif : étant données des réponses électrophysiologiques à des stimuli olfactifs, nous pouvons proposer une famille de modèles susceptibles de produire ces réponses ; ces modèles peuvent servir d'hypothèses de travail pour d'autres expériences.

Références

- G. Dreyfus, Les réseaux de neurones, Encyclopaedia Universalis (janvier 2000).
- F. Wolinski, F. Vichot, M. Stricker, Using Learning-Based Filters to Detect Rule-based Filtering Obsolescence, RIAO'2000, Paris (2000).
- Y. Oussar, G. Dreyfus, How to be a gray box : Dynamic Semi-Physical Neural Modeling, Neural Networks, à paraître (article invité)
- G. Monari, G. Dreyfus, Withdrawing an example from the training set : an analytic estimation of its effect on a nonlinear parameterized model, Neurocomputing Letters, in press.
- Y. Oussar, G. Dreyfus, Initialization by Selection for Wavelet Network Training, Neurocomputing, in press
- B. Quenet, D. Horn, G. Dreyfus, Temporal Coding in an Olfactory Oscillatory Model, Computation and Neural Systems, Bruges (2000).
- B. Quenet, G. Dreyfus, D. Horn, The Codes Observed in Oscillating Neural Assemblies of the Olfactory Pathway May Be Reproduced by a Simple Neural Network, Neural Computing in Science and Technology, NCST'99, Tel Aviv (1999).
- G. Dreyfus, Y. Oussar, J.L. Ploix, Knowledge-based Dynamic Neural Modeling of Industrial Processes, Neural Computing in Science and Technology, NCST'99, Tel Aviv (1999), conférence invitée.
- A. Duprat, T. Huynh, G. Dreyfus, Towards a Principled Methodology for Neural Network Design and Performance Evaluation, in QSAR ; Application to the Prediction of LogP, J. Chem. Inf. Comp. Sci., 38, 586-594 (1998).
- G. Dreyfus, Y. Idan, The Canonical Form of Nonlinear Discrete-time Models, Neural Computation 10, 133-164 (1998).

France Télécom R & D, Groupe TNT

FTR & D, Groupe TNT, 22307 Lannion Cedex

Contact : Daniel Collobert

Tel : 02 96 05 21 20

Mél : daniel.collobert@francetelecom.fr

Chercheurs concernés par le Thème :

Ingénieurs de recherche : D. Bardouil, O. Bernier, F. Clérot, D. Collobert, M. Collobert, R. Féraud, J.E. Viallet
Techniciens : P. Gouzien, J. Guérin, B. Rolland.

Doctorants : S. Marcel, C. Leignel

Cadre général

Le groupe TNT (Techniques Neuromimétiques pour les Télécommunications) est chargé d'étudier, de proposer et de développer de nouvelles méthodes destinées aux services et aux réseaux de France Télécom, fondées principalement sur des machines à apprentissage statistique.

Ses domaines d'activité comportent en particulier le traitement de séquences d'images pour la communication de groupe, la prévision de séries temporelles pour le trafic et les réseaux de communications, le traitement de données et la fouille de données.

Panorama des travaux effectués

En ce qui concerne la communication de groupe, la problématique est d'analyser et de suivre en temps réel des objets dans des séquences d'images animées. La première réalisation est le système appelé LISTEN (Locating Individual Speaker and Tracking ENvironment), un système capable de localiser et suivre en temps réel un locuteur, par l'image, dans un cadre de terminal individuel. LISTEN est fondé sur des réseaux de neurones génératifs et contraints qui estiment la probabilité qu'une image donnée soit un visage. A l'heure actuelle le détecteur de visages utilisé est le plus performant au monde (mesures du taux de reconnaissance et du taux d'erreur).

Une extension concerne l'analyse des gestes, principalement la reconnaissance automatique d'un certain nombre de gestes de commande. Cette reconnaissance, couplée à une reconnaissance de la parole, devrait permettre la réalisation de commandes et d'interfaces naturelles pour les systèmes de visiophonie et de visioconférence, sans contrainte pour les utilisateurs (pas de fil à la patte...).

Pour ce qui est des applications liées aux réseaux de télécommunications, il est possible d'utiliser des réseaux de neurones pour réaliser des prédictions concernant des trafics fortement sporadiques caractéristiques des liaisons entre réseaux locaux. Ces prédictions sont suffisamment précises pour autoriser par exemple la renégociation en temps réel des contrats de trafic ATM contenant les descripteurs de trafic et les exigences de qualité de service.

Les activités de traitement de données et de fouille de données touchent un secteur sensible des entreprises. Nous travaillons, comme tous ceux qui sont concernés par cette thématique, sur l'analyse des modèles et la mise en forme de bases de données imparfaites, par exemple en utilisant, en sus des réseaux de neurones, des techniques classiques de régression et de classification.

Références

R. Féraud, F. Clérot, J.L. Simon and Serge Martin, « Kalman and Neural Network Approaches for the Control of a VP bandwidth in an ATM Network », IFIP'2000, Grenoble, 4 juin 2000.

R. Féraud, O. Bernier, J.-E. Viallet et M. Collobert, « A Fast and Accurate Face Detector for indexation of Faces Images », FG'2000 Conference on Automatic Face and Gesture Recognition, Grenoble, 28-30 mars 2000.

Sébastien Marcel, Olivier Bernier, Jean Emmanuel Viallet et Daniel Collobert « Hand Gesture recognition using Input-output Hidden Markov Models », FG'2000 Conference on Automatic Face and Gesture Recognition, Grenoble, 28-30 mars 2000.

V. Lemaire, Olivier Bernier, Fabrice Clérot, Daniel Collobert, A new method to increase the margin of multilayer perceptrons, Neural Processing Letter (n469) 1999

Sébastien Marcel and Olivier Bernier, Hand Posture Recognition in Body-Face centered space, 3rd Gesture Workshop, Gif-sur-Yvette, Paris, 1999

Vincent Lemaire, Fabrice Clérot, Estimation of the blocking probabilities in a ATM Network node using Artificial Neural Networks for Connection Admission Control International Teletraffic Congress (ITC16) — 7 au 11 juin 1999 -Edinburgh.

Laboratoire Heudiasyc, UMR 6599, U.T.Compiègne, thème diagnostic

Université de Technologie de Compiègne,
B.P. 20 529
60205 Compiègne Cedex

Contact : Yves Grandvalet – 03 44 23 49 28 – Yves.Grandvalet@utc.fr

Chercheurs concernés par le thème :

Thierry Denœux, Pr. UTC, Christophe Ambroise, MdC UTC, Yves Grandvalet, CR CNRS, Mylène Masson, MdC Université d'Amiens

Cadre général et objectifs, problématique

Les travaux du thème se situent dans le cadre de l'apprentissage statistique et de la reconnaissance de formes. Nous privilégions les applications au diagnostic en général, et à la surveillance de l'environnement en particulier (capteurs logiciels pour la gestion de l'eau).

Panorama des travaux effectués par l'équipe sur le thème

Nous développons de nouveaux algorithmes de discrimination et de régression fondés sur la théorie des fonctions de croyance. Ce type d'approche permet d'une part de traiter de l'imprécision dans les données, et surtout rendre les sorties plus informatives dans un contexte de fusion de décisions ou d'aide à la décision.

Dans le cadre de l'apprentissage statistique, nous nous sommes concentrés sur les algorithmes de sélection de modèles. Nous avons développé des techniques permettant de mener conjointement la sélection de modèles et la sélection de variables, ou de pénaliser, de manière automatique, différents groupes de variables, ce qui représente une solution de remplacement intéressante aux techniques bayésiennes hiérarchiques.

Nos derniers travaux portent sur l'explication du succès des techniques d'ensemble.

Nos domaines d'application récents les plus importants sont le diagnostic des machines tournantes (Projet Européen Brite-Euram INQUAC avec Imperial College, Austrian Research Institute for Artificial Intelligence et les partenaires industriels Medav, Siemens, ZF, Automation Service et Moulinex) et la surveillance de la qualité de l'eau (Projet Européen Esprit EM2S, avec VKI Water Quality Institute, Danfoss, Hitec, Computas et Lyonnaise des Eaux).

Références

Denoeux, T. (2000), A Neural Network Classifier based on Dempster-Shafer Theory. IEEE Trans. on Systems, Man and Cybernetics A, 30 (2).
 Grandvalet, Y. (2000), Anisotropic Noise Injection for Input Variables Relevance Determination, IEEE Trans. on Neural Networks (à paraître).
 Valentin, N. et Denoeux, T. (2000), A neural network-based software sensor for coagulation control in a water treatment plant, Intelligent Data Analysis (à paraître).
 Masson, M.-H., Canu, S., Grandvalet, Y. et Lynggaard-Jensen, A. (1999), Software Sensor Design Based on Empirical Data, Ecological Modeling, vol 120.
 Denoeux, T. (1998), Application du Modèle des Croyances Transférables en Reconnaissance de Formes, Traitement du Signal, 14 (5).
 François, J., Grandvalet, Y., Denœux, T. et Roger, J.-M. (2000), Bagging belief structures in Dempster-Shafer K-NN rule, IPMU'2000.
 Grandvalet, Y. (2000), Bagging down-weights leverage points, IJCNN'2000, IEEE.
 Grandvalet, Y., Ambroise, C. et Canu, S. (1999), Local Learning by Sparse Radial Basis Functions, ICANN'99, IEE.
 Petit-Renaud, S. et Denoeux, T. (1999), Regression analysis using fuzzy evidence theory. FUZZ-IEEE'99, IEEE.
 Grandvalet, Y. et Canu, S. (1998), Outcomes of the equivalence of adaptive ridge with least absolute shrinkage, NIPS 11, MIT Press.

INRA — Unité de Biométrie et Intelligence Artificielle, Équipe Méthodes Mathématiques et Informatiques pour la Décision

Institut National de la Recherche Agronomique
 B.P. 27, 31326 Auzeville Cedex France

Responsable : Roger Martin-Clouaire
 Tel. + 33 5 61 28 52 86
 Mél : rmc@toulouse.inra.fr
Contact : Frédéric Garcia

Chercheurs concernés par le thème :
 Frédéric Garcia (CR)
 Tel. + 33 5 61 28 52 83

Mél : fgarcia@toulouse.inra.fr
 Régis Sabbadin (CR)
 Tel. + 33 5 61 28 54 76
 Mél : sabbadin@toulouse.inra.fr
 Laurent Peret (en thèse depuis septembre 2000)
 Tel. + 33 5 61 28 53 34
 Mél : peret@toulouse.inra.fr

Cadre général et objectifs, problématique

Les travaux méthodologiques de l'équipe concernent le développement de formalismes de représentation de stratégies décisionnelles adaptés à la simulation de problèmes de conduite de systèmes biophysiques, la modélisation de critères de décision non classiques et la conception de méthodes d'optimisation de stratégies fondées sur l'emploi de simulateurs. Ce dernier point est abordé principalement du point de vue de l'apprentissage par renforcement, où l'objectif est d'apprendre par l'emploi de simulations des comportements optimaux pour un critère considéré. Dans ce cadre, nos recherches portent sur le développement d'algorithmes efficaces d'apprentissage par renforcement, sur l'emploi de structures d'approximation permettant de traiter des problèmes de très grande dimension, et sur l'extension de l'apprentissage par renforcement à des modèles non probabilistes (qualitatifs).

Travaux effectués

Une adaptation des algorithmes d'apprentissage par renforcement aux problèmes de décision séquentielle en horizon fini a été proposée. Des résultats d'équivalence entre critères d'optimalité ont été établis. Des comparaisons théoriques et expérimentales ont été menées entre les algorithmes Qlearning et Rlearning [4,6,7].

Les algorithmes d'apprentissage par renforcement comme Qlearning ou TD peuvent être étudiés par la méthode de l'équation différentielle moyennée (ODE). Il est ainsi possible d'optimiser certains des paramètres de ces algorithmes, de manière à accélérer la convergence. Nous avons développé et appliqué cette technique aux algo-

rithmes Qlearning [6,8], TD [1] et Q [2].

Une théorie de la décision séquentielle qualitative fondée sur la théorie des possibilités a été développée. Des algorithmes de type « Programmation Dynamique » ont été proposés. Des travaux s'inspirant des techniques de l'apprentissage par renforcement sont en cours pour l'apprentissage automatique à partir de simulations de stratégies optimales au sens de ces critères « qualitatifs » [5].

Références

- [1] Frédéric Garcia and Florent Serre. Efficient Asymptotic Approximation in Temporal Difference Learning. Proceedings of the 14th European Conference on Artificial Intelligence (ECAI 2000), Berlin, Germany, 2000.
- [2] Frédéric Garcia et Florent Serre. AQ-learning : une approximation asymptotique optimisée de Q. Actes de la Seconde Conférence Apprentissage (Cap 2000), St-Etienne, 2000.
- [4] Frédéric Garcia. Use of Reinforcement Learning and Simulation to Optimize Wheat Crop Technical Management. Proceedings of the international congress on modelling and simulation (MODSIM 99), Hamilton, NZ, 1999.
- [5] Régis Sabbadin. Décision séquentielle qualitative possibiliste : algorithmes itératifs, apprentissage de stratégies. Actes de la Première Conférence Apprentissage (Cap 1999), Paris, 1999.
- [6] Frédéric Garcia et Seydina Ndiaye. Apprentissage par renforcement en horizon fini I : comparaison du Qlearning et du Rlearning. Apprentissage par renforcement en horizon fini II : analyse par la méthode de l'ODE. Chapitres 2 et 3 de l'ouvrage collectif Apprentissage automatique, sous la direction de Gilles Venturini et Marc Sebban, éditions Hermes, 1999.
- [7] Seydina Ndiaye. Apprentissage par renforcement en horizon fini : application à la génération de règles pour la conduite de culture. Thèse d'Informatique de l'Université Paul Sabatier de Toulouse, février 1999.
- [8] Frédéric Garcia and Seydina Ndiaye. A Learning Rate Analysis of Reinforcement Learning Algorithms in Finite-Horizon. Proceedings of International Conférence on Machine Learning (ICML'98), Madison, USA, 1998.

INT, Département EPH, Équipe Réseaux de Neurones et Applications Télécom

Institut National des Télécommunications,
9 rue Charles Fourier, 91011 EVRY
01 60 76 44 30

Contact : Bernadette Dorizzi
Mél : Bernadette. Dorizzi@int-evry.fr

Cadre général et objectifs, problématique

Le groupe s'est donné comme objectifs, depuis ses débuts de tester le bien fondé, l'intérêt et l'apport des techniques neuronales dans le traitement d'un certain nombre d'applications. Cette problématique nous a amenés à nous poser des questions sur les modèles eux-mêmes et à élaborer une recherche novatrice dans le domaine des réseaux de neurones.

Panorama des travaux effectués par l'équipe

Reconnaissance du manuscrit :

Reconnaissance de mots manuscrits en ligne, thèse de L. Duneau, 1994

Étude de modèles d'apprentissage automatique statistique, comme les modèles markoviens et de leur hybridation avec des modèles neuronaux, dans le cadre de la reconnaissance de mots manuscrits (collaboration avec le Lip6), thèse de Sonia Garcia-Salicetti, 1996, thèse de Z. Wimmer, 1998

Prédiction de séries temporelles :

Étude des réseaux récurrents (et plus particulièrement des réseaux récurrents simples) dans le cadre de la prédiction de séries temporelles : thèse de T. Czernichow, 1996

Elaboration d'algorithmes de sélection de variables dans le cadre prédictif, thèse de T. Czernichow, 1996

Fusion de données

Étude de l'apport des techniques neuronales pour la Modularité et la Coopération dans le cadre de la reconnaissance des chiffres manuscrits, thèse de P. Sébire, 1995

Fusion de réseaux neuronaux pour l'inversion de fonctions dans le cadre de la prédiction ionosphérique, thèse de F.

Jacquet, 1999

Fusion de décision à l'aide de modèles neuro-flous et experts, thèse d'E. Jouseau, 1999

Égalisation de canaux par approche de classification, thèse de J. Montalvao-Filho, 2000

Références

L. Duneau, B. Dorizzi, « On-line cursive script recognition : a user-adaptive system for word identification », *Pattern Recognition*, Déc.1996

T. Czernichow, A.Piras, K.Imhof, P.Caire, Y.Jaccard, B.Dorizzi, A. Germond, « Short Term Electrical Load Forecasting with Artificial Neural Networks », *Engineering Intelligent Systems*, vol4, n° 2, pp85-99, juin 1996

S. Garcia-Salicetti, B. Dorizzi, P. Gallinari, Z. Wimmer, « Adaptive Discrimination in an HMM-Based Neural Predictive System for On-Line Word Recognition », *ICPR 96*, 26-26 août 1996, Vienne, Autriche

B.Dorizzi, J.C.M.Mota, F.Albu, « A Step Towards Equalization for Radiomobile Channels : Neural Networks and Variable Selection », *Workshop on « Emerging Techniques for Communication Terminals »*, Toulouse, France, 07-09/07/97.

S. Garcia-Salicetti, B. Dorizzi, P. Gallinari, « A Neural Predictive System for On-Line Handwriting Recognition », soumis à *Internal Journal On Document Analysis, IJDAR (2000)*

J. R. Montalvao-F., J. C. M. Mota, B. Dorizzi, C.C.Cavalcante, « Bayesian Equalization Strategy : performance versus number of clusters » *Intern. Conf. on Signal Processing Application & Technology-ICSPT, 98*, Toronto, Canada, Septembre 1998

F. Jacquet, J. Caratori, B. Dorizzi, « Neural Networks models in the inversion of multidimensional ionograms », *IEEE-Eurasip, NSIP'99*, Antalya, Turkey, June 1999

E. Jouseau, B. Dorizzi, « Neural networks and fuzzy data fusion. Application to an on-line and real time vehicle detection system », *Pattern Recognition Letters 20 (1999) 97-107*

Z. Wimmer, B. Dorizzi, P. Gallinari, « Dictionary preselection in a Neuro-Markovian Word Recognition System », *ICDAR 1999*, Bangalore, India, septembre 1999

E. Jouseau, B. Dorizzi, « Predictive neural networks and fuzzy data fusion for on-line and real time vehicle detection, *Fusion'99* », *International Conference, San Francisco, July 1999.*

IPSL — Neuratel — Équipe Modélisation et méthodes statistiques avancées, animatrice du groupe de travail Neuratel à l'Institut Pierre Simon Laplace (IPSL) (<http://www.ipsl.jussieu.fr/>)

LODYC-Paris6, 4 place Jussieu Tour 26
Boite 100.75252 Paris Cedex 05

Contact : Sylvie Thiria
(thiria@lodyc.jussieu.fr
01-44 27 27 08)

Groupe de travail « Neuratel »

Sylvie Thiria (thiria@lodyc.jussieu.fr
01-44 27 27 08) Fouad Badran
(badran@cnam.fr 40 27 22 69) Carlos
Mejia (mejia@lodyc.jussieu.fr 01 44 27
27 09) Ludvine Gross (lgr@lodyc.jus-
sieu.fr) Meziane Yacoub (yacoub@
lodyc.jussieu.fr) Mustapha Lebbah
(Mustapha.Lebbah@renault.com)

Cadre général et objectifs, problématique

NEURATEL est un groupe de travail auquel participe l'ensemble des laboratoires de l'IPSL. Son objectif est de développer l'utilisation des méthodes neuronales en géophysique. Les sujets abordés jusqu'à maintenant concernent plus particulièrement la télédétection par satellite. La télédétection spatiale est confrontée à un double problème lié d'une part à la très grande quantité de données à traiter et, d'autre part, au fait que les fonctions de transfert qui conduisent aux variables géophysiques ne peuvent pas toujours s'écrire de manière rationnelle étant donnée la complexité des phénomènes mis en jeu. La stratégie de NEURATEL est de faire travailler ensemble à la fois les physiciens de l'IPSL et des mathématiciens spécialistes de traitements des données (CNAM, Paris 13, INT) afin de réaliser des chaînes complètes de traitement des données utilisant en même temps les résultats des recherches menées en physique et en statistiques.

L'approche neuronale est de plus tout à fait adaptée à l'intégration de données de natures physiques différentes dans un

même algorithme. Des actions sur l'utilisation de réseaux neuronaux ont ainsi été entreprises.

Panorama des travaux effectués par l'équipe

Les réseaux neuronaux se sont révélés particulièrement efficaces :

- au LODYC pour décoder les données des diffusiomètres de ERS1/2, NSCAT et QSCAT améliorer les algorithmes conduisant aux constituants biogéochimiques marins à partir des mesures multi-spectrales par satellites de couleurs de l'océan
- au LMD pour inverser les fonctions de transfert radiatif et pour classifier les nuages
- au SA pour identifier les gaz traces dans l'atmosphère
- au CETP pour retrouver le contenu en vapeur d'eau à partir de radiomètres micro-ondes.

Pour les prochaines années, le groupe NEURATEL a décidé de focaliser son action sur l'inversion des données satellitaires qui conduisent à la détermination de certains paramètres biogéochimiques océaniques à partir des capteurs multi-spectraux mesurant la couleur de l'océan (SeaWifs, MERIS sur ENVISAT), des variables atmosphériques telles que la température, l'eau liquide nuageuse, les précipitations (capteurs micro-ondes, IASI), le vecteur vent (QSCAT, SeaWinds sur ADEOS2) et de la concentration des constituants mineurs dans la troposphère comme l'ozone, le méthane et le monoxyde de carbone (IASI sur METOP et IMG sur ADEOS). Ces études sont généralement liées à un instrument particulier ou à une campagne de mesure particulière. Elles seront toujours réalisées d'une manière méthodologique afin que les solutions proposées dans une configuration particulière soient reproductibles à d'autres capteurs, à d'autres données, à des problèmes similaires.

Le groupe NEURATEL développe aussi des recherches théoriques qui portent sur les domaines suivants :

- détermination des barres d'erreurs et des matrices de co-variances des fonc-

- tions de transferts modélisées par des perceptrons multi-couches,
- amélioration des cartes topologiques de Kohonen par introduction de l'aspect probabiliste et application aux méthodes de classification,
- inversion et assimilation par méthode variationnelle mixte où certains éléments sont représentés par des réseaux de neurones ; des travaux sont entrepris sur ce sujet en assimilation de données de couleur de l'océan dans les modèles biogéochimiques,
- utilisation de méthodes mixte multi-experts — réseaux de neurones,
- fusion de données provenant de capteurs différents.

Les activités du groupe ont été soutenues par deux programmes Européens dont il est P.I. : NEUROSAT (1996-1999) et NAOC (200-2003). Par ailleurs, le groupe NEURATEL a entrepris une activité de formation dont les premières actions sont la mise en place en 1999 d'un DESS « Traitement des données : du capteur à l'exploitation des données » cohabilité par l'UVSQ et l'IN <http://carnot.physique.uvsq.fr/~jli-naires/DESS/index.html>.

Ce DESS forme des spécialistes capables d'assister les physiciens dans la réalisation de chaîne complexe de traitement des données. Il participe à l'organisation pour 2001 d'une école de printemps (CNES-CEA-ONERA) sur les réseaux de neurones et leurs applications à l'aéronautique, au spatial et à l'environnement <http://www-instn.cea.fr/Actualites/ecole-neurones.htm>

Références

- F. Anouar, F. Badran et S. Thiria (1998) « Probabilistic Self organizing map and Radial Basis Functions Networks », *Neurocomputing* 20 (1998) 83-96
- F. Badran, S. Thiria : (1997) « Neural Network smoothing in correlated time series context ». *Neural Networks*, vol 10 N° 8, pp. 1445-1453, 1997.
- F. Badran, Y. Stephan, N. Metoui, S. Thiria : « A general formulation of non-linear least square regression using multi-layered perceptrons ». *Soumis à IEEE transaction on neural networks* Décembre 1999) (40 p)
- F. Badran, S. Thiria : « Les perceptrons Multicouches : De la régression non-linéaire aux problèmes inverses » soumis à *Revue de*

Statistique Appliquée Septembre 2000

Gross L., Thiria S. Frouin R Mitchell B.G. (2000) Artificial neural networks for modeling transfer function between marine reflectance and phytoplankton pigment concentration *Journal of. Geophys. Res.* Vol 105, no. C2, pp3483-3949, february 15, 2000

Hadjji-Lazaro, J., C. Clerbaux, and S. Thiria, An inversion algorithm using neural networks to retrieve atmospheric CO total columns from high-resolution nadir radiances, *J. Geophys. Res.* Vol 104 NO 19 pp 23841-23854, 1999.

Mejia C, F. Badran, A. Bentamy, M. Crepon, S. Thiria, N. Tran : Determination of the Geophysical Model Function of NSCAT Scatterometer and its corresponding variance by the use of Neural Network, *J. Geophys. Res.* 104, 11539-11556 (NSCAT SPECIAL ISSUE).

Mejia C, S. Thiria, F. Badran, N. Tran and M. Crepon, : Determination of the Geophysical Model Function of ERS1 Scatterometer by the use of Neural Networks. *J. Geophys. Res.*, 103, 12853-12868. 1998

Richaume P., F. Badran, M. Crepon, C. Mejía, H. Roquet and S. Thiria, 1998 : Neural Network Wind Retrieval from ERS-1 Scatterometer Data, *J. Geophys. Res.*, vol 105, No. C4, Pages 8737-8751, 2000

M Yacoub, D Frayssinet, F Badran, S Thiria : 2000 « Classification based on Expert knowledge propagation using Probabilistic Self-Organizing Map : application to geophysics » in *Studies in Classification, Data Analysis, and Knowledge Organization*. Springer-Verlag (a paraître).

ISC, Équipe « Connexionnisme et Modélisation Cognitive »

Institut des Sciences Cognitives, UMR CNRS 5015

67 boulevard Pinel , 69675 Bron cedex

Contact : Hélène PAUGAM-MOISY

04 37 91 12 25

hpaugam@isc.cnrs.fr

Chercheurs concernés

Emanuelle Reynaud, Agnès Crépet, Didier Puzenat, Hélène Paugam-Moisy

Panorama des travaux effectués par l'équipe

Réseaux de neurones et parallélisme

Ayant constaté le relatif échec de la parallélisation des réseaux de neurones pour les ingénieurs (le parallélisme à

grain fin des réseaux de neurones est difficile à capturer sur les machines parallèles actuelles, efficaces pour des implantations à grain grossier), nous nous sommes tournés vers les réseaux modulaires (ex. modèle HME) dont la parallélisation est plus prometteuse. Notre nouvel objectif est d'exploiter les aspects asynchrones du parallélisme pour rendre compte des comportements du cerveau, dans le cadre de la modélisation de processus cognitifs.

Modélisation cognitive

Contrairement à de nombreuses équipes qui construisent des modèles biologiquement plausibles d'assemblées de neurones, nous nous plaçons davantage au niveau de la Psychologie Cognitive et nous nous intéressons à la modélisation d'architectures fonctionnelles pour mieux comprendre les phénomènes de mémorisation, de reconnaissance et d'identification. Nous élaborons un modèle de mémoire associative multi-modale en combinant plusieurs modèles connexionnistes, en les adaptant et en définissant leurs modes de coopération. Ces modèles, outre leur intérêt direct pour les sciences Cognitives, peuvent aussi être vus, a posteriori, comme de nouveaux paradigmes d'apprentissage adaptés à la classification d'objets multi-modaux et à la fusion de données.

Références

Bertolini, Paugam-Moisy, Puzenat, (1999), « Priming an artificial memory », in Proc. of IWANN'99, Int. Work. on Artificial Neural Networks, Lecture Notes in Computer Science, 1606:348-356.
Crépet, Paugam-Moisy, Reynaud, Puzenat, (2000), « A modular neural network for binding several modalities », in Proc. of ICAI'2000, Int. Conf. of Artificial Intelligence, Las Vegas, US.
Reynaud, Crépet, Paugam-Moisy, Puzenat, (2000), « A computational model for bidding sensory modalities », abstract to appear in Consciousness and Cognition, ASCC4, Assoc. for the Sc. Study of Consciousness.
Reynaud, Crépet, Paugam-Moisy, Puzenat, (2000), « A modular neural network model for a multi-modal associative memory », in Proc. of ICCNS'2000, Int. Conf. on Cognitive and Neural Systems, Boston, US.
Crépet, Paugam-Moisy, Reynaud, Puzenat, (2000), « Mémoire associative multimodale

et mécanismes de retour de l'information », C'Ap'2000, conférence sur l'apprentissage, éd. Colin de la Higuera, Hermès, 69-82.

LEIBNIZ-IMAG — Équipe Réseaux d'Automates

46 avenue Félix Viallet,
38000 Grenoble (France)
Tél. : (+ 33) 4 76 57 46 64
Fax : (+ 33) 4 76 57 46 02

Contact : Danniell Memmi
memmi@imag.fr

Web : www-leibniz.imag.fr/RESEAUX/

Chercheurs permanents :

Bernard Amy, Daniel Memmi

Cadre général et objectifs, problématique

L'équipe (une dizaine de personnes en moyenne) développe des modèles neuronaux pour traiter des tâches cognitives, et en particulier des problèmes de diagnostic et de contrôle, de raisonnement expert, et de recherche d'information.

Nous avons notamment étudié la conception de systèmes hybrides neuro-symboliques combinant étroitement connaissances explicites et apprentissage à partir des données. Il faut assurer la traduction entre les différents types de connaissance (compilation de règles à partir de réseaux entraînés) et la coordination entre les différents modules.

Nous nous intéressons aussi aux réseaux récurrents et aux systèmes dynamiques. Nous avons une longue expérience de l'utilisation de réseaux récurrents discrets pour la modélisation de processus séquentiels, dans des domaines très divers allant de la robotique à la linguistique.

Nous employons également les réseaux neuronaux pour la classification de documents et la recherche d'information. Des réseaux compétitifs à apprentissage non supervisé (type ART ou SOM) permettent de catégoriser des textes automatiquement pour faciliter l'accès à l'information pertinente. Nous cherchons en particulier à aider la recherche de pages Web.

L'apprentissage est naturellement un thème transversal commun à ces différents axes de recherche. Nous étudions des méthodes d'apprentissage pour divers types de réseaux de neurones et d'automates, en les confrontant aux algorithmes génétiques et aux systèmes multi-agents.

Panorama des travaux effectués par l'équipe

Notre équipe a mené à bien des applications industrielles de taille réelle, notamment dans le cadre de contrats européens.

Nous avons ainsi utilisé les réseaux neuronaux pour effectuer des tâches de diagnostic médical, et de diagnostic de pannes en robotique sous-marine. Nous avons développé pour cela plusieurs systèmes hybrides opérationnels combinant règles expertes et apprentissage à partir de cas pour un meilleur diagnostic.

La robotique mobile et autonome pose une gamme de problèmes allant de la perception à l'action en passant par la planification et le contrôle. Nous avons employé des réseaux récurrents pour la gestion de trajectoires sur un petit robot Khepera, mais aussi développé des systèmes hybrides pour le contrôle de véhicules terrestres et de robots sous-marins.

En partenariat avec une start-up, nous étudions actuellement la recherche collaborative d'informations sur le Web. Le travail en cours doit mener à la réalisation d'un produit commercialisable.

Références

B. Deuker, M. Perrier & B. Amy, Fault-diagnosis using neuro-symbolic hybrid systems, 9th International Workshop on Principles of Diagnosis (DX'98), Cape Cod, Mass., mai 1998.
B. Deuker, M. Perrier & B. Amy, Fault-diagnosis of subsea-robots using neuro-symbolic hybrid systems, OCEAN'98, Nice, sept. 1998.
F.S. Osorio & B. Amy, INNS : A hybrid system for constructive machine learning, Neurocomputing 28, p. 191-205, 1999.
I. Schadle, J.Y. Antoine & D. Memmi, Connectionist language models for speech understanding : the problem of word order variation, Eurospeech'99, Budapest, sept. 1999.

D. Memmi, Processing flexible word order languages with neural networks, Cahiers LEIBNIZ n° 03, mai 2000.

D. Memmi & J.-G. Meunier, Using competitive networks for text mining, NC'2000, Berlin, mai 2000.

LIP6 — ÉQUIPE APPRENTISSAGE NUMÉRIQUE

Université Paris 6
8 rue du Capitaine Scott
75015 Paris

Contact : Patrick Gallinari
Patrick.Gallinari@lip6.fr
<http://www-connex.lip6.fr/>
tel : 01 44 27 73 70

Chercheurs concernés par le thème :

Permanents : T. Artières, P. Gallinari, J.D. Kant, A. Lifchitz

En thèse : M. Amini, L. Haifeng, I. Kermia, A. Lopes, S. Marukatat, B. Piwowarski

Cadre général et objectifs, problématique

Les travaux de l'équipe portent sur le développement d'algorithmes d'apprentissage numérique pour les problématiques de l'accès à l'information. Les thèmes principaux aujourd'hui concernent le traitement de l'information textuelle et les interfaces pour objets mobiles, fondées sur les stylos.

Panorama des travaux effectués par l'équipe

Aspects formels des réseaux de neurones

Nous avons développé en particulier :

- des méthodes pratiques de contrôle de la complexité des réseaux fondées sur l'apprentissage bayésien.
- un nouvel algorithme de complexité polynomiale pour optimiser les combinaisons linéaires de classificateurs. Ces travaux ont fait l'objet d'une collaboration avec des chercheurs de l'Institut de Biologie et Chimie des Protéines de Lyon. Les logiciels développés sont utilisables pour faire de la prédiction « en ligne » de la structure secondaire des protéines sur le site Web à http://pbil.ibcp.fr/NPSA/npsa_mlr.html.

- des algorithmes de sélection de variables.

Traitement de séquences et Interfaces stylos

Après avoir travaillé pendant plusieurs années sur la modélisation de séquences pour différents domaines d'application, nos travaux sont principalement développés aujourd'hui dans le cadre de la reconnaissance de l'écriture en ligne. Un premier système a été réalisé fondé sur des modèles markoviens et des réseaux de neurones, et un prototype est disponible sur Web (www-poleia.lip6.fr/CONNEX/HWR). Nous travaillons actuellement sur la représentation des lexiques sous des formes compactes, sur des nouveaux modèles de séquences et sur des interfaces stylos pour les portables et les mobiles.

Accès à l'information textuelle

Nous avons commencé à développer des travaux sur ce thème il y a quatre ans et c'est devenu aujourd'hui une des activités fortes de l'équipe. Nous avons en particulier développé une famille de modèles qui couvre un ensemble de tâches allant de la recherche d'information (filtrage, routage...) à l'extraction d'information de surface (extraction des groupes de mots ou portions de textes pertinentes à un ensemble de requêtes). Ces systèmes sont fondés sur des modèles de Markov cachés ou des réseaux de neurones et permettent de traiter des séquences de termes. Nous travaillons actuellement sur la prise en considération de la structure des documents et sur le résumé.

Diagnostic des systèmes complexes

Nous avons abordé des problèmes de diagnostic temps réel pour des systèmes complexes du type réseaux de télécommunication et réseaux électriques avec le CNET et Schneider Electric. Nous avons en particulier travaillé sur la fusion spatiale de diagnostics locaux et proposé des modèles combinant réseaux Bayésiens et réseaux de neurones.

Références

Artières T., Marchand J.M., Gallinari P., Dorizzi B., 2000, multi modal segmental models for on line handwriting recognition, ICPR'2000.
Amini M.-R., Zaragoza H., Gallinari P., 2000, Learning for Sequence Extraction

Tasks, 6th Conference on « Content-Based Multimedia Information Access » (RIAO'2000), Collège de France, Paris.

Amini M.-R. Interactive Learning for Text Summarization, Proceedings of PKDD'2000 Workshop on Machine Learning and Textual Information Access

Piwowarski B. Learning in Information Retrieval : a Probabilistic Differential Approach. IRSG'2000

Zaragoza H., Gallinari P., Curtelin R., Leglaye F., 1998, Multiple multivariate regression and global sequence optimization, an application to large scale models of radiation intensity, Signal Processing. 64, 371-382.

Guermeur Y. Geourgeon C., Gallinari P. Deléage G., 1999, Improved Performance in Protein Secondary structure Prediction by Inhomogeneous Score Combination, Bioinformatics, Vol 15, n° 5, 413-421.

Gallinari P., Cibas T., 1999, Practical complexity control in multilayer perceptrons, Signal processing 74. Vol. 74, No. 1, April, 29-46.

Leray P. Gallinari P., 1999, Feature extraction with neural networks, Behaviormetrika, Vol 26, N°1, 145-166.

LIPN — Équipe : ADAGE : « Apprentissage, Diagnostic, Agents »

Groupe : « Apprentissage Connexionniste et Reconnaissance des Formes »
UNIVERSITE PARIS 13 — INSTITUT GALILEE — LIPN, CNRS
99, Avenue J-B. Clément
93430 Villetaneuse, FRANCE
Contact : Younes Bennani
Mél : younes@lipn.univ-paris13.fr
URL : <http://www-lipn.univ-paris13.fr>
Tél. : 01 49 40 36 07
Fax : 01 48 26 07 12

Chercheurs concernés par le thème :

- Permanents : Y. Bennani, E. Viennet.
- Doctorants : K. Benabdeslem, R. Hamiti, J. Kharroubi.
- Membres associés : F. Bossaert, V. Radevski, M. Yacoub.

Cadre général et objectifs, problématique

Nos activités sont centrées sur les apports des techniques connexionnistes à différents domaines de l'ingénierie et de l'informatique. Notre démarche de

recherche est fondée sur la confrontation de connaissances théoriques et de résultats pratiques. Les études théoriques portent sur la conception d'algorithmes d'apprentissage et le développement de systèmes modulaires et hybrides faisant coopérer des techniques connexionnistes et classiques. Les résultats de ces études sont validés sur des applications réelles dans différents domaines : parole, image, diagnostic, fouille de données.

Panorama des travaux effectués par l'équipe sur le thème

Nous nous sommes attachés au problème du contrôle de la capacité de généralisation des systèmes d'apprentissage connexionnistes. Nous avons montré qu'il est souvent plus simple d'utiliser des méthodes heuristiques pour déterminer l'optimum de la capacité d'un système, et choisir une architecture dont la capacité est adaptée à celle du problème. Nous avons proposé des méthodes structurelles fondées sur des mesures heuristiques permettant d'évaluer l'importance des unités dans un système connexionniste et ensuite de procéder à l'élagage des cellules les moins importantes. Une autre approche qui consiste à pondérer les variables en fonction de leurs participations à la discrimination a été également proposée.

Une grande partie de nos travaux sur les modèles connexionnistes a consisté à tester la possibilité de les adapter à différents domaines d'application. C'est dans cette voie que nous avons montré sur des problèmes réels (parole, image, diagnostic, « fouille du Web »), comment les données et les connaissances a priori du domaine peuvent être utilisées pour spécifier les architectures des systèmes connexionnistes dans les cas de la modélisation, de la classification et de la discrimination. Plus précisément, nous avons proposé des méthodologies de combinaison de classificateurs et de fusion de leurs décisions. Nous étudions aussi les propriétés de nouveaux algorithmes d'apprentissage, comme les Machines à Vecteurs de Support, ainsi que les modalités de leur mise en œuvre sur des problèmes réels (identification d'images, fouille de données).

Références

- Bennani Y. & Bossaert F. (2000), « Modular Connectionist Modelling and Classification Approaches for Local Diagnosis in Telecommunication Traffic Management », à paraître dans *International Journal of Computational Intelligence and Applications*.
 Yacoub M. & Bennani Y. (2000), « Features Selection and Architecture Optimization in Connectionist Systems », à paraître dans *International Journal of Neural Systems*.
 Radevski V. & Bennani Y. (2000), « Reliability Control in Committee Classifier Environment », *IJCNN'2000*.
 Fernandez R. & Viennet E. (1999), « Face Identification with Support Vector Machine », *ESANN'99*.
 Viennet E. & Fernandez R. (1999), « Machines à Vecteur de Support et Réseaux de Neurones : comparaisons expérimentales pour l'identification de visages », *CAP'99*.
 Nedjari T. & Bennani Y. (1998), « Symbolic-Connectionist Interaction », *ICCN'98*.
 Radevski V. & Bennani Y. (1997), « Combining Structural and Statistical Features for Handwritten Digit Recognition », *IJCNIS'97*.

LORIA/INRIA-Lorraine — Équipe CORTEX

B.P. 239
 F-54506 Vandoeuvre-les-Nancy
 Tel : (+ 33/0) 3 83 59 20 53
 Fax : (+ 33/0) 3 83 41 30 79
 Contact : Frédéric Alexandre
 Mél : falex@loria.fr

Chercheurs impliqués par le thème :

Frédéric Alexandre, DR INRIA, Hervé Frezza-Buet, Supélec, Bernard Girau, MdC Université de Nancy 2, Yann Guermeur, MdC Université de Nancy 1, Jean-Charles Lamirel, MdC Université de Strasbourg, Dominique Martinez, CR CNRS

Cadre général et objectifs, problématique

Le but de nos recherches est d'étudier les propriétés et les capacités d'un traitement automatique de l'information réalisé sur des bases distribuées, numériques et adaptatives. Plus précisément, nous cherchons à montrer qu'un tel type de traitement peut permettre la réalisation de systèmes « intelligents », c'est-

à-dire capables d'extraire de la connaissance à partir de données et de manipuler cette connaissance pour résoudre des problèmes. L'ensemble de ces capacités est obtenu par la mise au point de modèles connexionnistes neuromimétiques développés selon deux sources d'inspiration, les neurosciences (*computational neurosciences*) et l'apprentissage automatique (*machine learning*).

Panorama des travaux effectués par l'équipe sur le thème

Nos modèles inspirés des neurosciences s'intéressent à des mécanismes particuliers (e.g. traitement perceptif, mémoire), mais s'attachent également à réaliser des systèmes autonomes complets. Ainsi, des plates-formes sont réalisées, qui regroupent des capacités de perception, représentation interne, décision et action. Ces plates-formes sont implantées sur des simulateurs ou sur des robots réels. Elles regroupent un ensemble de modules fonctionnels, chaque module correspondant à la modélisation d'une structure neuronale particulière. Cette approche modulaire nous permet en particulier de nous intéresser à la définition des mécanismes numériques permettant l'émulation de différents types de mémoire : mémoire épisodique (hippocampe), mémoire de travail (cortex préfrontal), mémoire procédurale (cortex postérieur), ainsi que des possibles collaborations entre ces mémoires.

Nos travaux inspirés de l'apprentissage automatique proposent l'adaptation de modèles classiques du connexionnisme (MLP, SOM, SVM) et la définition de modèles originaux, tous ces modèles permettant d'étudier l'extraction de connaissances à partir de ces modèles ou leur utilisation dans un cadre expert, c'est-à-dire dans un cadre où des connaissances a priori permettent de spécifier l'utilisation de modèles ou leur combinaison. Des exemples de tels domaines sont la bio-informatique, la structuration de bases de connaissances, les télécommunications. Dans ce cadre, la mise au point d'algorithmes d'apprentissage doit être couplée à l'interprétation de leur fonctionnement et de leurs règles implicites d'extraction de

connaissances, afin de donner en retour à l'utilisateur des renseignements sur le travail effectivement réalisé.

Références

- F. Alexandre. Biological inspiration for multiple memories implementation and cooperation. In proc. International Symposium on Computational Intelligence, August 2000.
- L. Bougrain. Étude de la construction par réseaux neuromimétiques de représentations interprétables. Thèse de l'Université Henri Poincaré, Nancy 1, octobre 2000.
- H. Frezza-Buet. Un modèle de cortex pour le comportement motivé d'un agent neuromimétique autonome. Thèse de l'Université Henri Poincaré, Nancy 1, octobre 1999.
- H. Frezza-Buet, F. Alexandre and N. Rougier. A cerebral framework for the integration of biologically inspired temporal mechanisms for sequences processing. In : Neural, symbolic and reinforcement methods for sequence learning, R. Sun and L. Gilles eds, Springer, 2000.
- Y. Guermeur et D. Zelus Combining protein secondary structure prediction methods with a new multi-category SVM. ISMB'00, San Diego, 2000.
- N. Rougier. Modèles de mémoires pour la navigation autonome. Thèse de l'Université de Nancy 1, octobre 2000.

PSI, INSA de Rouen, Thème Classification et apprentissage

INSA de Rouen,
Technopôle du Madrillet,
76801 Saint Etienne du Rouvray
CEDEX

<http://psychaud.insa-rouen.fr/~scanu/>

Contact : Stéphane Canu
stephane.canu@insa-rouen.fr
tel (+ 33) 02 32 95 98 44
fax (+ 33) 02 32 95 98 46

Chercheurs concernés par le thème

Stéphane Canu, Professeur, Philippe Leray, Maître de conférence, Alain Rakotomamonjy, Maître de conférence, Xavier Mary, Doctorant, Patrick Vanoorenberge, Maître de conférence, Thierry Paquet, Maître de conférence, Laurent Heute, Maître de conférence

Cadre général et objectifs, problématique

- Théorie de l'apprentissage, régularisation, ondelettes, SVM
- Réseaux d'Yprels
- Représentation des incertitudes : Dempster-Shaffer
- Sélection des variables pertinentes en entrées d'un réseau de neurones
- Estimation bayésienne
- Applications : Géostatistique, prévision des pics d'ozones, capteurs logiciels, reconnaissance de l'écriture manuscrite.

Panorama des travaux effectués par l'équipe sur le thème

Nous avons montré qu'un modèle connexionniste original appelé réseaux d'Yprels était capable de classification distribuée et d'apprentissage incrémental et continu. Ce système auto-organisant permet de distribuer le problème en sous-tâches de moindre complexité dont la résolution est répartie sur plusieurs réseaux coopérants. Ce système permet en outre des reprises d'apprentissage sur de nouvelles données, la sélection des données nécessaires à l'apprentissage et la sélection des primitives nécessaires pour la résolution du problème.

Nous avons montré de plus qu'une étude préalable de la topologie des données d'apprentissage permet de distribuer le problème de classification de façon cohérente avec les données et d'évaluer a priori la pertinence des primitives. L'analyse de la topologie est réalisée par une hiérarchie indicée dont nous déduisons une hiérarchie de réseaux de neurones. Une approche multi-niveaux originale a été mise au point pour la détermination et l'extraction des clusters dans un jeu de données à partir de la hiérarchie indicée. Une méthode de remise à jour incrémentale de la hiérarchie indicée et des réseaux associés a été mise au point et validée sur des bases de données réelles.

Parallèlement, nous avons montré qu'il est possible de fiabiliser un problème de décision (optimisation du compromis erreur-rejet) en combinant les décisions prises indépendamment par plusieurs classifieurs, même dans le cas de sources d'information en cohérence

partielle. Le schéma de combinaison mis au point tient compte du comportement de chaque composant, en l'intégrant dans le schéma même de combinaison, ce qui permet de fusionner efficacement les solutions fournies par chaque composant. Cette solution a été validée sur des problèmes de reconnaissance de chiffres manuscrits.

Nous avons développé des systèmes de classifieurs robustes permettant de fiabiliser le compromis erreur-rejet en détection de mélanomes et en reconnaissance de caractères manuscrits. Dans ce dernier cas, nous avons intégré ces systèmes sur différentes plates-formes de lecture automatique de documents manuscrits (chèques bancaires français et formulaires). Nous avons proposé une stratégie d'interaction entre processus d'apprentissage et de reconnaissance dans le cadre de l'interprétation de textes manuscrits

Par ailleurs, des techniques de sélection de caractéristiques pertinentes ont été appliquées à la détection des mélanomes, en reconnaissance en ligne de caractères, en interprétation de la figure de REY, et à l'identification d'un capteur logiciel. De même, de nouveaux algorithmes ont été développés pour l'identification en géo-statistiques.

Références

- S. Canu, Ph Leray et A. Rakotomamonjy, « Une méthode de prévision à un pas de temps : application à la prévision de la qualité de l'air », exposé dans le cadre des premières journées automatique et environnement, le 9 mars 2000 à Nancy
- S. Canu, M Kanevski. « Pollution Mapping with Support Vector Regression » Presented at StatGIS99 conference in Klagenfurt (Austria), 20-21 September 1999. Proceedings will be published in the journal *Mathematische Geologie*.
- S. Canu and A Elisseff, « Why sigmoid-shaped function are good for learning ». Snowbird workshop, avril 1999.
- S. Canu, Y. Grandvalet and M. H. Masson, « Black-box Software Sensor Design for Environmental Monitoring », in International Conference on Artificial Neural Networks, Skovde, Sweden. Sep 2-4, 1998.

SAMOS-MATISSE, UMR 8595

Université Paris 1
Centre PMF
90, rue de Tolbiac
75634 PARIS CEDEX 13
TEL et FAX : 33 1 40 77 19 22
Contact : Marie Cottrell
cottrell@univ-paris1.fr

Chercheurs concernés :

Eric de Bodt (Lille 2), Marie Cottrell, Jean-Claude Fort (Nancy 1), Smail Ibbou, Serge Iovleff (Vannes), Ryad Kallel, Patrick Letremy, Fatine Maghrebi (Tunis 3), Morgan Mangeas (Inrets-Livic), Patrick Rousset, Joseph Rynkiewicz, Vincent Vigneron (Evry),

Cadre général et objectifs

Nos recherches portent sur les rapprochements entre les statistiques et les modèles connexionnistes, et en particulier sur les applications aux données économiques ou financières.

Panorama des travaux effectués par l'équipe sur le thème

Nous étudions la théorie de certains réseaux de neurones (algorithme d'auto-organisation de Kohonen, perceptrons multi-couches), l'utilisation des techniques neuronales en analyse de données, en prévision de séries temporelles. Nous étudions comment ces techniques se situent par rapport aux modèles linéaires et non-linéaires plus classiques, et étudions leurs propriétés en les comparant.

Nous développons des outils logiciels élaborés de visualisation de données de

grande dimension à partir de cartes de Kohonen, de prévision de séries temporelles au moyen de perceptrons gérés par des chaînes de Markov cachées, avec recherche automatique de la meilleure architecture.

Ces outils sont mis en œuvre dans des applications économiques et financières, en collaboration avec l'équipe MATISSE de Paris 1, et GERME de Lille 2.

Plus récemment, nous avons entrepris une étude à long terme sur la prévision de la pollution atmosphérique en région parisienne, à partir de méthodes neuronales combinant des perceptrons multi-couches et des chaînes de Markov cachées, en collaboration avec Alain Dutot du LISA (Créteil).

Actions de formation

Plusieurs chercheurs de l'équipe assurent la direction ou des cours pour les deux diplômes suivants :

DEA : Statistiques et Modèles aléatoires appliqués à l'économie et à la finance (Paris 1 — Paris 7)

DESS : Techniques de la décision en entreprise (en Économie à Paris 1)

Collaborations

Le SAMOS est membre du réseau Ile de France de sciences cognitives qui regroupe la plupart des centres de recherche en sciences cognitives de la région parisienne.

Depuis décembre 1994, les membres du SAMOS appartiennent au GDR, Mathématiques de la Perception artificielle et naturelle, coordonné par Laurent Younès (ENS Cachan).

Le SAMOS fait partie depuis mars 1998 du GDR « Méthodes Quantitatives

en Finance » créé en 1998 et coordonné par Nicole El Karoui.

Il entretient des collaborations prolongées avec Tunis 3, Malaga, Louvain la Neuve, La Havane, Turin, Barcelone.

Références

- J. Rynkiewicz, Hybrid HMM/MLP models for time series prediction, ESANN 1999, p. 455-462.
- F.Maghrebi, On a Hopfield net arising in the modelling and control of over-saturated signalized intersections, Neural Processing Letters, 10 (3), p. 161-169, 1999.
- M. Cottrell, P.Gaubert, P. Letremy, P. Rousset, Analysing and Representing multi-dimensional quantitative and qualitative data : Demographic study of the Rhône valley. The domestic consumption of the Canadian families chapitre du livre « Kohonen Maps », (Eds. E.Oja and S.Kaski), Elsevier, juin 1999, chap 1, p 1-14.
- P. Gaubert, M. Cottrell, A dynamic analysis of segmented labor market, ACSEG99, Fuzzy Economic Review, 1999, 4 (2), p.63-82, 1999.
- P.Rousset, Applications des algorithmes d'auto-organisation à la classification et à la prévision, Thèse de l'Université Paris 1, décembre 1999.
- E. de Bodt, M. Cottrell, Bootstrapping Self-Organizing Maps to Assess the Statistical Significance of Local Proximity, ESANN 2000.
- R. Kallel, M. Cottrell, V.Vigneron, Bootstrap for neural model selection, ESANN 2000.
- M. Cottrell, T.S. Turova, Use of an Hourglass model in neuronal Coding, à paraître dans Journal of Applied Probabilities, 2000.
- M. Cottrell, P. Gaubert, Classification of recurring unemployment workers and unemployment exits, ACSEG, Louvain 1998, EJESS, 2000, 14 (1), p.39-68.
- J. Rynkiewicz, Estimation de Modèles auto-régressifs à changements de régime markoviens. Thèse de l'Université Paris 1, soutenance prévue en décembre.



INGÉNIERIE DES CONNAISSANCES — ÉVOLUTIONS RÉCENTES ET NOUVEAUX DÉFIS

JEAN CHARLET, MANUEL ZACKLAD, GILLES KASSEL ET DIDIER BOURIGAULT

Éditeurs Eyrolles, (Collection technique et scientifique des télécommunications), 2000.

Notes de lecture d'Amedeo Napoli

Ce livre est une somme d'articles — il en compte 36 — et d'efforts, des auteurs mais aussi des éditeurs. Il contient une sélection d'articles présentés aux Journées « Acquisition des Connaissances » (JAC 95 et 96), puis « Ingénierie des connaissances » (IC 97 et IC 98). Il présente les nombreuses facettes actuelles de l'IC, qui prend de plus en plus d'importance à l'heure du Web, de la fouille de documents (de toute nature), de la mémoire d'entreprise, et de la gestion des connaissances. Ce livre constitue d'ores et déjà une référence de première importance en matière d'IC.

Il s'articule autour de sept parties principales, qui comptent (ce qui donnera une assez bonne idée du contenu du livre) : (1) modélisation des connaissances, (2) langages de modélisation et de représentation, (3) textes, terminologie et ontologies, (4) raisonnement à partir de cas, (5) hypertextes, hypermédia et Web, (6) gestion des connaissances, (7) systèmes d'information pour le travail coopératif et méthodes cognitives. Comme on le voit, l'éventail est large et recouvre bien l'ensemble des thèmes actuellement abordés (de près ou de loin) par l'IC. Les parties les plus représentées (les plus importantes ?) — c'est-à-dire les parties 1, 2, 3 et 7 — contiennent entre six et huit articles, très représen-

tatifs des recherches dans le domaine. Ainsi, le livre peut-il servir à la fois pour avoir les premières notions sur un thème donné, et pour appréhender de façon plus approfondie certaines directions d'investigation actuelles.

Il faut souligner la présentation extrêmement soignée de l'ouvrage, qui rend la lecture claire et très agréable. De plus, et c'est là une des particularités « réflexives » de l'ouvrage, l'index, qui est très complet, très fourni, et qui facilitera assurément la navigation dans les articles, a été construit à l'aide de LEXTER, un logiciel d'extraction de terminologies (dû à Didier Bourigault). La boucle est bouclée : les principes et les techniques de l'IC se voient là appliqués à l'IC elle-même, montrant par là l'intérêt et l'applicabilité de ces travaux de recherches.

Il ne faudrait pas terminer cette chronique sans saluer le travail des éditeurs (c'est leur tour puisque les auteurs ont déjà été applaudis lors des conférences...), qui ont fait là un excellent travail d'organisation et de supervision. Tout cet effort se voit récompenser par un ouvrage qui donne sa juste place à l'IC dans l'éventail des recherches en intelligence artificielle en France.

TECHNIQUES PRATIQUES POUR L'ÉTUDE DES ACTIVITÉS EXPERTES

PAR ANDRÉ BISSERET, SUZANNE SÉBILLOTTE, ET PIERRE FALZONEYROLLES

OCTAR'S Éditions, Toulouse, 1999.

Notes de lecture d'Amedeo Napoli

La première phrase de ce petit livre résume bien les objectifs des auteurs : « Provoquer des observables, pour mettre au jour des activités cognitives humaines, tel est l'objectif des techniques décrites dans ce livre ».

Ce livre traite de psychologie cognitive, d'ergonomie et d'observation des comportements : il propose une analyse des pratiques dans l'accomplissement d'une tâche, pour mieux la comprendre, l'appréhender,

et partant, améliorer l'accomplissement de cette tâche pour les exécutants. Pour le chercheur en IA, ce livre, qui se veut avant tout un outil pratique, est plutôt agréable à lire, et il contient de nombreux éléments en rapport avec la représentation des connaissances, la mémorisation, l'organisation des connaissances, la réutilisation, la recherche d'informations, l'analyse des tâches (voire l'acquisition de connaissances), et les traces de raisonnements. Il peut donc intéresser bon nombre de chercheurs et de praticiens de l'intelligence artificielle, comme les spécialistes de représentation des connaissances, des sciences de l'éducation, de raisonnement à partir de cas, entre autres domaines...

Le livre se divise en douze chapitres (mais où est donc passée la table des matières ?...), qui abordent autant de techniques d'interrogation permettant de mettre à jour les activités cognitives auxquelles s'intéressent les auteurs, techniques parmi lesquelles se trouvent le rappel, la catégorisation, le tri, la recherche de similarités, et l'élicitation.

Chaque chapitre est construit sur le même mode (ce qui permet de faire plus facilement le lien entre les

différentes techniques abordées), avec un exposé des principes, des objectifs et des prérequis, puis le cadre théorique, suivi d'une discussion sur les avantages et les inconvénients de la technique, des exemples d'applications et enfin une bibliographie.

Le livre traite les choses du point de vue des sciences cognitives, mais les chercheurs en IA y trouveront un bon nombre d'éléments et de façons de faire qui pourront leur permettre de revoir, ajuster ou adapter leur propre savoir faire, par exemple dans le cadre général de l'acquisition et de la représentation des connaissances.

Pour terminer, il faut mentionner qu'il est dommage que ce petit livre n'ait pas reçu une plus grande attention lors de sa publication, qu'une table des matières aurait été la bienvenue, sinon nécessaire, de même qu'un index, qui facilite et aiguille bien des recherches. Toutefois, il est possible de passer outre ces petits désagréments, et découvrir assez simplement, par la pratique, ce que peuvent apporter à l'intelligence artificielle les études en sciences cognitives et en ergonomie.

INTRODUCTION AU DATA MINING — ANALYSE INTELLIGENTE DES DONNÉES

PAR MICHEL JAMBU

Collection technique et scientifique des télécommunications, Eyrolles, Paris, 1999.

Notes de lecture d'Amedeo Napoli

Comme le suggère l'auteur lui-même, le livre aurait très bien pu s'appeler « Introduction à l'analyse intelligente des données », voire «... à la fouille de données », ce qui aurait été plus joli et ce qui aurait beaucoup mieux montré la relation étroite qui existe entre cet ouvrage et la réédition d'un classique du même auteur : « Méthodes de base de l'analyse des données » dont on trouvera la chronique ci-après.

Ce petit livre essaie de combler un certain vide dans la littérature francophone sur le sujet, car il existe peu d'ouvrages en français sur le sujet, destinés aux universitaires. Le livre se divise en six chapitres, qui présentent les objectifs et les méthodes de la fouille de données (qui est l'expression que nous utiliserons dans la suite). L'auteur introduit tout d'abord de façon très générale la gestion de l'information pour la prise de décision, avec une préoccupation plutôt liée à l'entre-

prise, et y montre la place de la fouille de données. Il analyse ensuite les types de données utilisés sur lesquels s'appliquent les processus de fouille de données, en faisant une très large part aux tableaux de données (de toutes sortes), ces tableaux étant les formats auxquels sont habitués les spécialistes d'analyse des données. L'auteur aborde ensuite les méthodes de fouille de données proprement dites, en mettant encore une fois l'accent sur les méthodes plus spécifiquement liées à l'analyse des données, comme les méthodes factorielles, de segmentation ou de régression. Le chapitre suivant propose une liste de critères permettant d'évaluer un système de fouille de données, selon les points de vue des différents utilisateurs (au sein d'une entreprise). Enfin, le sixième et dernier chapitre fournit une liste d'exemples d'applications de la fouille de données dans l'entreprise, en particulier pour améliorer la satisfaction

des clients, pour la gestion et le marketing, pour la détection de fraudes, etc.

Ce livre présente la fouille de données dans ses grandes lignes, parfois un peu rapidement, avec la vision que peut avoir un spécialiste de l'analyse des données. Le chercheur en intelligence artificielle apprendra assurément un certain nombre de choses dans cette optique, mais il risque d'être un peu déçu, car le point de vue symbolique qui lui est cher n'est pas du tout abordé (nous pensons à la classification conceptuelle, la classification par treillis et l'extraction de règles, les modèles et les systèmes de connaissances

pour la fouille de données...). De ce fait, le livre va surtout jouer le rôle d'ouvrage introductif pour les chercheurs ou étudiants qui désirent aborder le domaine ; ils y trouveront une bonne et très utile exposition des relations entre fouille de données et analyse des données classique, ce qui leur permettra de bien comprendre les origines du processus de la fouille de données et les enjeux sous-jacents. Pour le point de vue symbolique, il leur faudra aller consulter certains ouvrages de langue anglaise, ou encore certains articles du livre « Ingénierie des connaissances » dont on trouvera la chronique ci-dessus.

MÉTHODES DE BASE DE L'ANALYSE DES DONNÉES

PAR MICHEL JAMBU

Collection technique et scientifique des télécommunications, Eyrolles, Paris, 1999.

Notes de lecture d'Amedeo Napoli

Ce livre réactualise un ouvrage du même auteur publié il y a une dizaine d'années sur le même sujet. C'est en quelque sorte un « classique de l'analyse de données » comme il en existe un petit nombre dans la littérature francophone sur le sujet. Le livre propose un panorama très général de l'analyse des données et introduit les grandes méthodes de base de cette branche des mathématiques.

Le livre se divise en deux grandes parties, la première traitant de l'analyse élémentaire d'une variable, l'analyse conjointe de deux et de plusieurs variables, la seconde partie traitant de l'analyse approfondie des données avec les différentes méthodes d'analyse que sont les méthodes factorielles et les méthodes de classification par exemple. L'exposé écrit est complété par un CD-Rom qui contient une version du logiciel d'analyse des données STATlab, dû à l'auteur, et différents jeux de données pour expérimentation. Dans la première partie, l'auteur introduit les différents problèmes associés à l'analyse des données, les modes d'analyse pour une variable, l'analyse conjointe pour deux et plusieurs variables. L'exposé est clair et accompagné de nombreux exemples, ce qui rend le propos agréable à lire et à suivre. L'auteur prend soin de présenter les théories mathématiques et les difficultés sous-jacentes de façon très progressive. De plus, chaque chapitre est complété par des exercices dans lesquels les enseignants, tout comme les étudiants, pourront aller chercher des com-

pléments et des précisions. La deuxième partie introduit les méthodes classiques d'analyse de données, parmi lesquelles les analyses factorielles, en composantes principales, des correspondances binaires, des correspondances multiples, et enfin les méthodes de classification (pour ces dernières sont abordées partition, classification hiérarchique ascendante, segmentation et discrimination, mais rien sur les treillis). Le même souci pédagogique est présent, et l'auteur prend soin de bien imager son propos, ce qui devrait bien convenir au chercheur en intelligence artificielle, qui verra les nombreuses possibilités de traitement des données du monde réel relevant des méthodes d'analyse des données. À l'heure où la fouille de données (de toutes sortes) prend de plus en plus d'importance, il est nécessaire de bien comprendre la portée des méthodes classiques — et les méthodes d'analyse des données en font partie — pour mener ainsi des recherches dans le champ de l'intelligence artificielle sur des bases solides et cohérentes.

Pour résumer, par rapport aux méthodes de base de l'analyse des données, ce livre est complet, clair, bien documenté, et comporte beaucoup d'exemples. Il bénéficie, en outre, d'une très belle présentation. Il devrait de ce fait convenir à tous les chercheurs en intelligence artificielle qui sont intéressés par l'analyse des données et par la fouille de données numériques et symboliques.

IC'2000

11ES JOURNÉES INGÉNIERIE DES CONNAISSANCES TOULOUSE, CENTRE POUR L'UNESCO, 10 – 12 MAI 2000

<http://www.irit.fr/IC2000/>

Textes écrits, rassemblés et mis en forme

par P. Tchounikine, M.C. Jaulent, A. Mille, J. Charlet et N. Ausсенac-Gilles

Les journées Ingénierie des connaissances de l'an 2000, organisées par les laboratoires toulousains IRIT et ERSS, ont offert une cuvée réussie et dynamisante pour la communauté scientifique du domaine. Comme l'a rappelé P. Tchounikine (LIUM, Le Mans), président du comité de programme, dans sa rapide introduction à la conférence, l'appel à communication avait déjà été un succès puisque 58 soumissions ont été proposées (30 retenues). P. Tchounikine avait demandé au comité de programme de mettre en pratique la remarque de Jacques Pitrat dans le bulletin AFIA N°37, c'est-à-dire d'appliquer une méthode de relecture « disjonctive » (accepter un article qui, même s'il est insatisfaisant sur certains points, présente néanmoins au moins une idée réellement intéressante, par opposition à une méthode plus « conjonctive »), ce qui a conduit à l'acceptation de quelques articles atypiques mais intéressants (et, de fait, lors de la conférence, ces articles ont conduit à des échanges passionnants). L'accent a été ensuite mis sur la vitalité de l'Ingénierie des connaissances (IC), avec notamment l'apparition de 3 nouveaux thèmes (« Intranet, Internet et IC », « systèmes d'information et IC » et « ingénierie éducative »), mais également sur sa cohérence scientifique : l'analyse des articles présentés montre que l'IC n'est pas une juxtaposition de thèmes, mais bien une thématique scientifique cohérente. Enfin, illustrant la maturité de la discipline, 3 « conférences tutorielles » (affreux néologisme, qui a été reproché au président du comité de programme !, mais adopté faute de mieux — est-il appelé à subsister ?...) ont été proposées cette année, chacune proposant une vision synthétique d'une thématique phare. Ce principe de conférence tutorielle a été retenu comme une bonne idée, associant à la fois une approche pédagogique et une volonté de faire le point (avec une touche personnelle) sur un domaine de recherche. Une autre originalité de cette édition d'IC a été la part importante accordée aux réalisations, à travers une session de démonstrations (20 logiciels ont été montrés) et deux sessions parallèles sur les réalisations.

Journée du 10 mai 2000

C'est la « conférence tutorielle » de R. Teulier (CNRS, Cachan) qui a ouvert IC'2000. Intitulé

« L'ingénierie des connaissances aujourd'hui : quels enjeux pour l'organisation ? », l'exposé proposait une analyse historique et prospective de l'impact de l'IC sur les entreprises, en s'appuyant sur l'évolution parallèle des travaux sur les systèmes d'informations et sur les approches gestionnaires. Une série d'enjeux étaient clairement identifiés : continuer à construire une véritable ingénierie, réussir l'intégration dans les SI, conceptualiser les modes de travail et comprendre les impératifs organisationnels. La nécessité d'aborder ces enjeux en interaction avec les autres disciplines et ingénieries (assumer le carrefour pluridisciplinaire !) et l'importance des aspects collectifs étaient également soulignés. Cet exposé ne pouvait que titiller l'auditoire et le temps de discussion n'a pas permis de donner la parole à tous. Une série d'interventions a cependant permis d'insister sur le passage des problématiques individuelles aux problématiques collectives (P. Zaraté), sur le bouleversement introduit par le Web, qui s'impose à faible coût et dont il faut respecter les caractéristiques intrinsèques (P. Albert) ou encore sur le changement des usages, le problème de l'accès aux connaissances prenant le pas sur celui de la résolution de problèmes (P. Laublet).

Si le nombre d'articles portant sur les méthodes de résolution de problèmes s'est effectivement réduit par rapport aux années précédentes, la session qui y était consacrée n'en était pas moins intéressante. Le premier exposé, « Une modélisation au niveau connaissance du raisonnement à partir de cas » de B. Fuchs (IAE, Lyon 3) et A. Mille (LISI, Lyon 1) présentait une modélisation générique du raisonnement à partir de cas, fondée sur des formalismes de spécification et de décomposition de tâches, l'exemple de la tâche de remémoration étant approfondi. Le second, « Construire un Système à Base de Connaissances de type Tâche/Méthode à l'aide de graphes conceptuels », de F. Trichet, M. Leclère et C. Choquet (IRIN, Nantes), présentait l'apport des graphes conceptuels à la représentation des différents types de connaissances (domaine et raisonnement) utilisés lors d'une modélisation de type tâche-méthode, ce qui constitue une amélioration notable de travaux antérieurs (l'architecture DSTM, présentée à IC'98).

La pause repas ne se prolongea pas par une pause digestive puisque c'est la session « Épistémologie » qui

attendait les participants. L'exposé de B. Morand (GREYC, Caen), « Processus de représentation et raisonnement abductif — Une approche expérimentale du cas des diagrammes » proposait une contribution à une approche expérimentale des phénomènes de représentation, en l'occurrence centrée sur la conception de diagrammes. Les diagrammes successifs produits lors d'un processus de conception sont considérés comme des « faits de conception » observables, dont les états successifs permettent d'étudier les opérations permettant de passer de l'un à l'autre. B. Morand note que l'approche expérimentale qu'il propose permet d'envisager l'instrumentation de l'étude de phénomènes de cognition, qui n'est cependant selon lui envisageable que par la conception d'un système mixte (homme-machine). Non moins conceptuelle, l'intervention de N. Balacheff, « Les connaissances, pluralité de conceptions. Le cas des mathématiques » (LEIBNITZ, Grenoble) proposait une réflexion sur la notion de connaissance. L'auteur insistait sur la coexistence possible, chez un élève, de structures apparemment contradictoires mais cohérentes dans des contextes différents de mises en œuvre. Du point de vue de la conception de systèmes d'apprentissage (qui est le contexte de cette étude), la notion d'« erreur » peut ainsi être vue comme le symptôme d'une connaissance dont le fonctionnement sort de son domaine de validité. D'un point de vue plus général, l'exposé mettait en évidence le problème que constitue le simple emploi du terme « connaissance ». Finalement, si certains regrettaient l'absence des problématiques de l'« ingénierie », la plupart des auditeurs furent fortement interpellés par cette présentation atypique mais passionnante.

Intitulée « Ontologies et hypertextes », la session suivante proposait une première intervention de C. Desmoulin et M. Grandbastien (LORIA, Nancy) intitulée « Des ontologies pour indexer des documents techniques pour la formation professionnelle ». L'exposé présentait comment la conception d'ontologies (ontologie générique de documentation, ontologie du domaine de documentation, ontologies domaines et ontologie pédagogique) permettait aux formateurs d'exploiter les documentations techniques numérisées directement proposées par des constructeurs automobiles. L'exposé suivant de T. Dechilly et B. Bachimont (INA, Paris) « Une ontologie pour éditer des schémas de description audiovisuels, extension pour l'inférence sur les descriptions », proposait ensuite une contribution à l'instrumentation de l'indexation audiovisuelle. Celle-ci comprenait une méthode pour définir, à partir d'interprétations d'images et de la prise en compte du type d'informations qui y sont recherchées, une ontologie pertinente pour ensuite les indexer.

La dernière session de présentation de cette longue journée, consacrée à l'ingénierie éducative, regroupait des approches de conception d'hypermédia à vocation pédagogique. Les travaux proposés présentaient successivement « Une démarche de conception pour les hypermedias pédagogiques : l'enjeu d'une approche centrée sur l'information » (S. Crozat et P. Trigano – Heudiasyc, Compiègne), puis « La construction de cours hypermédia adaptés à l'ap-

prenant par agencement de briques élémentaires » (N. Delestre – PSI, Rouen). Ces 2 exposés (tout comme celui de C. Desmoulin et M. Grandbastien précédemment) mettaient bien en évidence les liens entre l'ingénierie éducative et l'ingénierie des connaissances.

Enfin, la session démonstration (organisée par B. Biebow) était alors proposée aux participants, fatigués mais rapidement ragaillardis par un buffet servi sur place. Si l'ergonomie de certains systèmes était mise à mal (difficile de tester un logiciel un verre à la main !), cette session (près de 20 démonstrations et 6 posters) illustrait cependant l'ancrage « ingénierie » de la conférence.

Le soir, une visite guidée de la ville permettait de contempler la place du Capitole, d'apprendre pourquoi beaucoup d'églises de la région n'avaient pas de transept ou encore l'histoire de la Basilique Saint-Sernin (qui illustre la couverture des actes).

Journée du 11 mai 2000

La journée a débuté par une conférence sur « Textes et ontologies » conjointement présentée par D. Bourigault (ERSS, Toulouse) et J. Charlet (AP-HP, Paris). Cette intervention avait pour objectif de refléter les points de vue et les réflexions menées depuis quelques années au sein du groupe Terminologie et Intelligence Artificielle (TIA). Le problème majeur de la discipline est de développer des méthodes pour construire des ontologies en sachant bien qu'il est impossible de délivrer des ontologies universelles. Dans le groupe TIA, une place particulière a été faite à la construction d'ontologies à partir de textes, plusieurs outils existant dans ce domaine. Les orateurs ont insisté sur la nécessité d'avoir des outils pour sélectionner, trier les termes pour que l'expert puisse les valider. Les orateurs ont présenté un thesaurus pour l'ingénierie des connaissances, le projet Th (IC)². Ce projet constitue une alternative, complémentaire, au projet (KA)² de la communauté internationale d'ingénierie des connaissances (*cf. <http://www.irit.fr/GRACQ/>*). L'objectif étant de construire un thesaurus en français décrivant la recherche francophone dans le domaine de l'ingénierie des connaissances à partir d'un corpus d'articles provenant des conférences d'IC.

La discussion suscitée par le tutoriel a tourné autour de la construction d'ontologies à partir d'autre chose que des textes et en particulier des aspects réutilisables. Une polémique existe entre ceux qui pensent que les ontologies peuvent être une partie réutilisable d'une conceptualisation et ceux qui sont irrités par ces idées de généricité et montrent, par exemple, que les outils syntaxiques ne sont pas réutilisables d'une langue à l'autre.

En prolongement, la session 5 était la première de deux sessions consacrées à la terminologie et l'acquisition de connaissances à partir de textes. Elle illustrait bien le caractère pluridisciplinaire de l'ingénierie des connaissances. Le premier article, présenté par B. Biebow, informaticienne (LIPN, Paris) présentait les étapes du processus de modélisation à partir de textes. Le second article, pré-

senté par J. Rebeyrolle, linguiste (ERSS, Toulouse), s'est intéressé à l'acquisition de connaissances à partir de textes lorsqu'aucun dictionnaire n'est disponible (utilisation de structures lexico-syntaxiques privilégiées comme les contextes définitoires et de marqueurs tels les marqueurs lexico-syntaxiques et les marqueurs typographiques et dispositionnels). Dans ces deux travaux, les résultats obtenus reposent sur une analyse linguistique menée sur un corpus constitué par le groupe TIA.

Deux aspects de l'évolution du capital « connaissances » d'une entreprise ont été évoqués au cours de la session 6. Le premier article présenté par T. Tounkara (LAMSADE, Paris) concernait l'acquisition d'informations de l'environnement lorsque l'entreprise est en état de veille technologique. Les auteurs ont proposé une extension de la méthode MKSM (cette démarche prolonge le problème de la capitalisation d'un patrimoine de connaissances vers celui de l'évolution de ce patrimoine). L'intérêt de ce travail était de montrer les lacunes des approches classiques (comme par exemple la défaillance des processus de veille quant à la représentation et la création de nouvelles connaissances). Le deuxième article présenté par S. Moisan (INRIA, Sophia Antipolis) s'attaquait à la difficulté particulière pour l'entreprise de gérer les connaissances contenues dans les codes informatiques. Les auteurs proposent deux nouveaux moyens pour la gestion opérationnelle des connaissances sur les codes (les livres de connaissances et le pilotage des codes).

Après un repas convivial et riche en discussions dans les locaux de l'Unesco, l'après-midi est consacrée à des aspects plus opérationnels de l'ingénierie des connaissances avec d'une part les langages formels et d'autre part plusieurs expériences pratiques dans le domaine.

La session 7 était consacrée à la modélisation et la représentation de connaissances à l'aide d'objets. Les contributions ont mis en évidence qu'une modélisation directe dans un formalisme donné est très difficile. Le premier article présenté par A. Napoli (LORIA, Nancy) était une proposition pour la représentation de structures dans un système de représentation de connaissances par objets. Il s'appuie sur des travaux effectués en reconnaissance des formes où les descriptions structurelles ont été étudiées pour l'appariement et la reconnaissance d'objets dans des scènes. Le deuxième article est présenté par J. Nobécourt (LIPN, Paris 13) sur Mdos : un langage semi-formel permettant une modélisation des connaissances sous forme de réseaux conceptuels, à la frontière entre l'extraction des primitives conceptuelles à partir des textes et le choix d'un formalisme. Le but est ici d'aider un ingénieur dans sa tâche de modélisation, appliquée dans le domaine de la supervision de réseaux de télécommunication. Le dernier article de la session est présenté par X. Cregut (ENSEEIH, Toulouse) sur la modélisation et la réutilisation de composants de procédés (intégration dans l'environnement Rhodes, un environnement d'assistance pour la description et l'exécution de processus de conception et gérant des procédés en s'appuyant sur la technologie des composants).

Un point méthodologique fort de l'ingénierie des connaissances qui s'impose dans les applications présentées en session 8 est le recueil et la capitalisation des connaissances à partir d'expériences concrètes. Ces expériences peuvent être des observations directes fournies par l'expert ou par des documents ou déjà des connaissances explicitées par l'expert. Un autre point méthodologique abordé a été la validation des modèles de connaissance. Les domaines d'applications concernés par ces expériences pratiques sont divers (domaine médical, agricole, accidents de la route, feux de forêts, contrôle en temps réel). On a pu remarquer qu'il s'agissait de domaines d'application pour lesquels la prise de décision était souvent peu formalisée, reposant sur l'expérience accumulée des experts.

À l'issue de cette session, nous avons été conviés par N. Aussenac-Gilles à nous rendre à la mairie de Toulouse où l'adjoint au maire, autour d'un cocktail, a fait une allocution remerciant les organisateurs et les participants de la conférence d'avoir choisi cette belle ville de Toulouse pour leurs échanges fructueux. La soirée s'est poursuivie à l'hôtel d'Assezat, joyau de la renaissance toulousaine, avec la visite du musée et une dégustation de vins régionaux.

Journée du 12 mai 2000

C'est la conférence tutorielle « Gestion des connaissances », présentée par J.-L. Ermine (CEA, Saclay) qui a ouvert la journée. L'exposé très clair, synthétique et pédagogique manquait peut-être un peu d'ouvertures vers des voies de recherche précises et prometteuses. Mais on ne saurait trop recommander à ceux qui démarrent dans le domaine de la gestion des connaissances de commencer leur lecture par les travaux présentés ici. Les questions ont permis d'aborder des concepts comme « l'intelligence émotionnelle », (intelligence de comportement ou de l'être) dans le partage et la transmission de connaissances, tandis que les différences d'approche entre américains et européens étaient distinguées – on met en place les outils et on attend de voir ce qui en ressort, versus on réfléchit sur les connaissances, les modes d'organisation et on instrumente avec des outils. L'expérience de l'intranet chez les professionnels a fait ressortir que les clés de réussite de mise en œuvre sont : 1/3 d'organisation, 1/3 de changement et 1/3 d'outils. Enfin, la façon de prendre en compte et de représenter le contexte a été discutée du point de vue de l'ingénierie des connaissances.

La fin de matinée a été consacrée à la session 9, « Gestion des connaissances » qui a commencé par un exposé dont le titre « Gérer les retours d'expérience pour maintenir une mémoire métier, étude chez PSA Peugeot Citroën » (A Faure – PSA, La Garenne Colombes – et G. Bisson – INRIA, Rhône-Alpes), résonnait particulièrement bien avec la conférence tutorielle. Il s'agissait en effet de gérer l'accès aux documents contenant les connaissances métier de façon à en améliorer la persistance et la réutilisation. La discussion a démontré que cette démarche,

pragmatique au départ, a nécessité de modéliser les connaissances sous forme d'un référentiel métier, présenté selon 4 points de vue inspirés de MKSM. Toutefois si les documents métiers ont été analysés par un groupe de travail constitué de chefs de service, il n'y a pas eu d'interviews des utilisateurs ou des auteurs des documents. Une façon d'assurer le passage réciproque d'une ontologie à une base documentaire par un « cycle vertueux » a été présentée dans l'exposé suivant, « Construction et exploitation d'une ontologie pour la gestion des connaissances d'une équipe de recherche » de G. Kassel (LaRIA Amiens), M.H. Abel, C. Barry, P. Boulitreau, C. Irastroza, S. Perpette. Il s'agissait en effet de pouvoir expliquer les documents par des connaissances et réciproquement illustrer des connaissances par des documents qui les présentaient. La session s'est terminée par l'exposé « Ingénierie de la coopération pour gérer les connaissances dans une entreprise étendue », par I. Boughzala, M. Zacklad et N. Matta (TechCico, Troyes), qui s'est attaché à montrer que le problème de la communication de connaissances et de la coopération entre partenaires impliqués dans une entreprise étendue était particulièrement important. La question du support de communication EDI-Web a peut-être été un peu rapidement évoquée. Les aspects organisationnels ont été majeurs dans cet exposé intéressant mais pas toujours accessible pour les auditeurs.

La session 10, « Système d'information et ingénierie des connaissances » a débuté par l'exposé « Ingénierie des besoins, ingénierie des connaissances : similarités et complémentarités des approches de modélisation » de F. Tort (ENS, Cachan), R. Teulier, G. Grosz et J. Charlet, qui comparait EKD (méthode développée à l'occasion du projet Elektra) à CommonKADS. EKD, piloté par les objectifs et les changements, a permis de modéliser des aspects importants de l'entreprise et des besoins qui s'y exprimaient. CommonKADS s'attache plus à mettre en évidence le triplet Contextes/ Concepts/Artefacts. Les deux approches sont capables de représenter l'amont, le contexte de l'entreprise et la réutilisation des connaissances.

Le second exposé de cette session « Extraction et intégration d'informations semi-structurées dans des pages Web, Projet Chimère », par M.S. Seget (LIRMM, Montpellier), P. Pompidor et D. Héryn, s'est intéressé à la récupération des pages Web issues de sites d'offres de services (train, hôtel). L'objectif final est de tester automatiquement les offres de service. Pour cela, le système (robot) induit une ontologie locale à chaque serveur conforme à une ontologie globale fabriquée. Cette ontologie doit permettre de remplir automatiquement le formu-

laire ainsi découvert. Le problème est très intéressant mais les difficultés semblent impressionnantes et la recherche menée pour les surmonter commence tout juste.

Une deuxième et dernière session consacrée à « Terminologie et acquisition de connaissances » est venue clore la conférence. L'exposé d'ouverture « Partage des termes, partage des connaissances ? Construire une modélisation unique de plusieurs corpus », proposé par les linguistes M.P. Jacques et A.M. Soubeille (ERSS, Toulouse), présentait une étude sur 4 institutions qui essayaient d'harmoniser leurs termes et leurs connaissances. Les termes et les relations lexicales sont issus, selon une analyse linguistique, du corpus de chaque institution et ont servi de départ à la construction d'une ontologie. Une tentative de fusion des 4 ontologies a été menée. Sur 700 concepts repérés, 4 seulement se sont révélés parfaitement identiques et une quarantaine proches. Il semble à l'évidence que l'effet de contexte joue très vite et que les efforts à faire pour rapprocher des usages de termes ont été vite impressionnants. Le second exposé de la session s'intitule « Relations morphologiques et structuration de terminologie », P. Zweigenbaum et N. Grabar (AP-HP, Paris), abordait l'apprentissage sur des couples de mots pour découvrir des relations morphologiques. Nous avons interprété les résultats principaux du travail comme capables d'améliorer sensiblement l'étiquetage des relations de référence (celles qui donnent un sens). Il s'agirait alors d'une amélioration importante de l'indexation dans les thésaurus. Le dernier exposé « Apport de la fouille de données textuelles pour l'analyse de l'information », Y. Toussaint, A. Simon et H. Cherfi (LORIA, Nancy) illustre une autre approche des textes, non plus fondée sur leur étude linguistique mais sur leur exploration selon des critères statistiques d'usage des termes. Les résultats de ces analyses (ici des classes au sens des treillis de Galois et des règles d'association) ont pu être interprétés par des experts pour faire de la veille scientifique sur de grosses quantités de documents techniques.

On pourra trouver, sur le site GRACQ (<http://www.irit.fr/GRACQ/>), les articles de la conférence in extenso (très prochainement), le présent compte rendu, ainsi que la description et les exposés des différentes journées co-organisées en 1999 et 2000 par le GRACQ et accompagnant la politique d'ouverture de la conférence IC. La conférence vous donne rendez-vous au sein de la plate-forme AFIA pour son édition 2001.

Ont également participé à la rédaction de ce compte rendu : S. Delaitre, B. Fuchs, Y. Prié, J.M. Héraud, P.A. Champin, L. Damase.

Journées francophones d'Extraction et de Gestion des Connaissances Appel à communications (Provisoire)

Nantes, le 18-19 Janvier 2001

<http://www.sciences.univ-nantes.fr/irin/EGC2001/> Mèl : cec2001@ireste.fr

Ecole polytechnique de l'université de Nantes, IRIN - Université de Nantes, ERIC - Université de Lyon 2

Les journées EGC'2001 sont organisées sous le parrainage scientifique de la conférence européenne PKDD (*Principles and Practice of Knowledge Discovery in Databases*), du PRC-GDR I3 (*Information, Interaction, Intelligence*), de l'association INFOR-SID (*Informatique des Organisations et Systèmes d'Information et de Décision*), et de la SFC (*Société Francophone de classification*).

Les actes seront édités par HERMES Science Publications.

Objectifs

Aujourd'hui, de grandes masses de données structurées ou semi-structurées sont accessibles dans les bases de données d'entreprises ainsi que sur la Toile. Aussi les entreprises ont-elles besoin de méthodes et d'outils capables d'extraire les connaissances pertinentes pour les décideurs, de les stocker, et de les diffuser aux différents acteurs de l'organisation. Pour répondre à cette attente du tissu industriel, de nombreux projets de recherche se développent autour de l'extraction de connaissances à partir de données (Knowledge Discovery in Data), ainsi que sur la gestion de connaissances (Knowledge Management).

L'objectif de ces journées consiste à rassembler, d'une part les chercheurs des disciplines connexes (apprentissage, statistiques et analyse de données, systèmes d'information et bases de données, ingénierie des connaissances...), et d'autre part les spécialistes d'entreprises qui déploient les méthodes et les outils adaptés à leurs besoins, afin de contribuer à la formation d'une communauté scientifique dans le monde francophone autour de cette double thématique de l'extraction et de la gestion de connaissances.

Ces journées seront le lieu privilégié d'échanges entre chercheurs et spécialistes industriels afin de présenter l'état de l'art de cette thématique centrale, d'en faire émerger les besoins et les enjeux, et d'en stimuler le développement.

Principaux thèmes (liste non limitative)

- Modèles de représentation des données et des connaissances.
- Méthodes de prétraitement. Données volumineuses.
- Méthodes et algorithmes d'extraction de connaissances à partir de sources diverses (Bases de données, textes, images, vidéos, données spatiales, données temporelles, données semi-structurées, Toile) : analyse de données (symboliques, numériques), réseaux neuromimétiques, réseaux bayésiens, programmation logique inductive, ensembles approximatifs, découverte de règles...
- Mesure de la qualité des connaissances.
- Visualisation des connaissances.
- Entrepôts de données.
- Mémoires d'entreprises.
- Synthèse de connaissances. Ontologies et méta-informations.
- Déploiement des connaissances (travail coopératif, workflow).
- Cycle de vie des connaissances.
- Méthodes de recueil, d'acquisition des connaissances
- Systèmes d'information et de gestion des connaissances.
- Applications : tout domaine d'application quel qu'en soit le secteur ou l'activité professionnelle est éligible : industrie, administration, défense, biomédical, télécommunications, gestion de la clientèle, ressources humaines, production et maintenance...

Information aux auteurs

L'ensemble des contributions retenues sera publié dans les actes de la conférence par HERMES Science Publications.

La langue officielle de la conférence est le français. Toutefois la conférence est ouverte aux contributions et exposés en anglais.

Deux formats de communication sont possibles:

- un format court : 5 pages maximum,
- un format long : 12 pages maximum.

Chaque exemplaire doit comporter en première page le titre de la communication, la liste des auteurs et leur affiliation, les coordonnées et le Mèl de l'auteur à contacter, ainsi qu'un résumé de 200 mots en français et en anglais. Les références bibliographiques sont à placer en fin du document. Les pages ne doivent pas comporter de numérotation. Les auteurs sont également invités à utiliser et à se conformer aux feuilles de style proposées par HERMES Science Publications. Ces feuilles de style sont téléchargeables à partir de l'URL : http://www.hermes-science.com/cnsi_act.htm

Les propositions de communication sont à envoyer sous la forme d'un fichier, au format PostScript uniquement, attaché à un courrier électronique (avant le 20 octobre 2000) adressé à : ecg2001@irin.univ-nantes.fr

Les fichiers PostScript peuvent être compressés avec un format de compression GZIP ou ZIP.

Un accusé de réception de l'article sera envoyé. Les auteurs doivent vérifier que le papier est bien parvenu au comité de programme.

Consulter le site web de la conférence pour les dernières recommandations.

Formes de participation

- COMMUNICATIONS DE RECHERCHE (format long). Sont concernés des travaux de recherche originaux, théoriques ou expérimentaux. Une grande rigueur devra être apportée à la rédaction des articles, tant sur le plan de la forme que sur celui du contenu scientifique.
- COMMUNICATIONS COURTES (format court). Sont concernés des travaux en cours ou au stade de résultats partiels de recherches ou appliqués. Ce type de soumission correspond à des travaux innovants de qualité, mais dont la maturité ne permet pas des élaborations complètes et des conclusions définitives.
- COMMUNICATIONS APPLIQUEES (format court). Sont concernés des applications industrielles et opérationnelles significatives, des études de cas, des prototypes avancés. Le contexte de l'application devra être clairement décrit et les auteurs mettront l'accent sur les problèmes rencontrés, les enseignements tirés et les retours d'expériences.

Dates importantes

Réception des propositions : **20 Octobre 2000**
Notification aux auteurs : 3 Novembre 2000
Réception des textes définitifs : 14 Novembre 2000

NB: Attention au délai très court entre la date de notification d'acceptation aux auteurs et celle de réception des textes définitifs.

Congrès/Colloques/Écoles d'été...

OCTOBRE 2000

18-20 octobre 2000, TICE'2000, Technologies de l'Information et de la Communication dans les Enseignements d'ingénieurs et dans l'industrie, Troyes.

Objectifs : Conférence « Modèles, Outils, Méthodes » consacrée aux nouvelles technologies de l'enseignement pour la formation technologique dans les contextes des Écoles d'ingénieurs et de l'industrie, ainsi qu'aux méthodes d'utilisation qui leur sont associées.

Date limite de réception des articles : 1^{er} février 2000

Notification aux auteurs : 1^{er} avril 2000

Réception des versions définitives : 15 juin 2000 :

Président : Claude Frasson (Université de Montréal)

Page d'accueil :

<http://www.univ-troyes.fr/tice>

DÉCEMBRE 2000

11-12 décembre 2000, JIOSC'2000, Journées Internationales d'Orsay sur les Sciences Cognitives.

Objectifs : L'Institut des Sciences Cognitives et de la Communication (ISCC) de l'Université Paris-Sud, structure qui fédère les chercheurs d'Orsay, Bures et Gif-sur-Yvette impliqués dans les Sciences Cognitives, propose l'apprentissage comme thème de son prochain colloque. À l'heure actuelle, les résultats les plus significatifs obtenus dans le domaine de l'informatique, de la neurobiologie, de la psychologie et de la didactique, sur l'apprentissage s'expriment souvent dans des cadres théoriques et méthodologiques propres à chaque discipline. Nous souhaitons que les JIOSC'2000 soient l'occasion de créer de nouveaux dialogues. Pour favoriser ces interactions, les contributions respectives des différents domaines de recherche seront confrontées au sein d'ateliers interdisciplinaires. Nous sollicitons des contributions qui, tout en présentant des résultats innovants et pointus dans une discipline donnée, aient le souci de poser des questions dans une perspective transdisciplinaire.

Date limite de réception des articles : 1^{er} mai 2000

Notification aux auteurs : 25 juin 2000

Réception des versions définitives :

1^{er} septembre 2000 :

Page d'accueil :

<http://www.limsi.fr/WkG/ISCC>

JANVIER 2001

18-19 janvier 2001, EGC'2001, journées

francophones d'extraction et de gestion des connaissances, Nantes.

Objectifs : Aujourd'hui, de grandes masses de données structurées ou semi-structurées sont accessibles dans les bases de données d'entreprises ainsi que sur la Toile. Aussi les entreprises ont-elles besoin de méthodes et d'outils capables d'extraire les connaissances pertinentes pour les décideurs, de les stocker, et de les diffuser aux différents acteurs de l'organisation. Pour répondre à cette attente du tissu industriel, de nombreux projets de recherche se développent autour de l'extraction de connaissances à partir de données (Knowledge Discovery in Data), ainsi que sur la gestion de connaissances (Knowledge Management).

L'objectif de ces journées consiste à rassembler, d'une part les chercheurs des disciplines connexes (apprentissage, statistiques et analyse et données, systèmes d'information et bases de données, ingénierie des connaissances), et d'autre part les spécialistes d'entreprises qui déploient les méthodes et les outils adaptés à leurs besoins, afin de contribuer à la formation d'une communauté scientifique dans le monde francophone autour de cette double thématique de l'extraction et de la gestion de connaissances. Ces journées seront ainsi le lieu privilégié d'échanges entre chercheurs et spécialistes industriels afin de présenter l'état de l'art de cette thématique centrale, d'en faire émerger les besoins et les enjeux, ainsi que d'en stimuler le développement.

Date limite de réception des propositions de communication sous format électronique : 20 octobre 2000

Notification aux auteurs :

3 novembre 2000

Réception des versions définitives :

14 novembre 2000

Page d'accueil :

<http://www.sciences.univ-nantes.fr/irin/EGC2001/>

29-31 janvier 2001, LMO'2001, Langages et modèles à objets, Le Croisic.

Objectifs : comme les années précédentes, cette conférence offrira un cadre pour comparer les notions d'objet en programmation, en représentation de connaissances, en bases de données et en système, afin d'étudier leurs spécificités, leurs points communs et leurs divergences, ainsi que les tendances futures. Thèmes : représentation par objets, programmation par objets, bases d'objets, génie des objets, approches multiparadigmes et ouverture à d'autres paradigmes, applications... Le comité de programme est particulièrement intéressé par les travaux qui font un pont

ou des comparaisons constructives entre les diverses approches des objets.

Date limite de réception des propositions de communication sous format électronique : 7 septembre 2000

Notification aux auteurs :

1^{er} novembre 2000

Réception des versions définitives :

27 novembre 2000

Page d'accueil :

<http://www.emn.fr/LMO2001>

MAI 2001

21-23 mai 2001, MFI'01, journées francophones, modèles formels de l'interaction, Toulouse.

Objectifs : L'interactivité est une tendance majeure des systèmes informatiques actuels et un important champ de recherche. Elle se décline sous plusieurs aspects : (a) interaction entre utilisateurs et systèmes informatiques, (b) interaction entre entités informatiques autonomes (agents) collaborant pour résoudre un problème, interconnectées sur un réseau local ou sur « la toile » et (c) intégration de ces deux aspects dans les divers « agents conversationnels », « agents de recherche », « assistants personnels », etc. Le bénéfice attendu concerne aussi bien la fiabilité et la reproductibilité des systèmes informatiques réalisés, ou l'élaboration de méthodologies de conception de ces systèmes, que la modélisation de l'interaction humaine elle-même dans des aspects cognitifs et sociaux dans une perspective de sciences humaines et sociales. Le but de ces « journées » est de rassembler des chercheurs de différentes communautés scientifiques (informatique, linguistique, sociologie, économie, sciences cognitives...) ayant en commun la volonté de formaliser tel ou tel aspect de l'interaction entre agents artificiels ou humains.

Date limite de réception des propositions de communication : 1^{er} décembre 2000

Notification aux auteurs : 9 février 2001

Réception des versions définitives :

16 Mars 2001

Page d'accueil :

<http://damas.ift.ulaval.ca/conferences/mfi01>

Nous annonçons dans cette rubrique les manifestations IA (i) se déroulant en France ou (ii) francophones et se déroulant hors de France. Pour inclure une annonce, envoyer un court texte (disquette Mac ou mail avec fichier attaché) à Jean Charlet (voir « l'ours » page 2 de ce bulletin).

SOMMAIRES DES REVUES

Un grand merci à Valérie Mangin (lip6) pour sa contribution à cette rubrique !

N'hésitez pas à envoyer un message à Brigitte Grau (grau@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.

N'hésitez pas à envoyer vos remarques à propos de l'ETAI.

REVUE D'INTELLIGENCE ARTIFICIELLE

http://www.editions-hermes.fr/periodiques/ia_cur.htm

ELECTRONIC TRANSACTIONS ON ARTIFICIAL INTELLIGENCE (ETAI)

Free publication and interactive reviewing on the internet

<http://www.ida.liu.se/ext/etai/>

ARTIFICIAL INTELLIGENCE

<http://www.elsevier.nl:80/inca/publications/store/5/0/5/6/0/1/>

ARTIFICIAL INTELLIGENCE V119 N°1-2 MAY 2000

- Semantical and computational aspects of Horn approximations, *MARCO CADOLI, FRANCESCO SCARCELLO*
- Qualitative rigid-body mechanics, *THOMAS F. STAHOVICH, RANDALL DAVIS, HOWARD SHROBE*
- Choosing social laws for multi-agent systems: Minimality and simplicity, *DAVID FITOUSSI, MOSHE TENNENHOLTZ*
- Semi-quantitative system identification, *HERBERT KAY, BERNHARD RINNER, BENJAMIN KUIPERS*
- The complexity of belief update, *PAOLO LIBERATORE*
- The Spatial Semantic Hierarchy, *BENJAMIN KUIPERS*
- Linear constraints for the interpretation of line drawings of curved objects, *MARTIN C. COOPER*
- Note about cardinality-based circumscription, *YVES MOINARD*
- Nature's way of optimizing, *STEFAN BOETTCHER, ALLON PERCUS*
- J.A. Anderson and E. Rosenfeld (Eds.), Talking Nets: An Oral History of Neural Networks, *NOEL E. SHARKEY*
- Gossiping Nets, *ROLF P. WÜRTZ*

ARTIFICIAL INTELLIGENCE V120 N°1 JUNE 2000

- Abductive reasoning through filtering, *CHITTA BARAL*
- Optimal auctions revisited, *DOV MONDERER, MOSHE TENNENHOLTZ*
- Bounding the cost of learned rules, *JIHIE KIM, PAUL S. ROSENBLUM*
- Backtracking algorithms for disjunctions of temporal constraints, *KOSTAS STERGIOU, MANOLIS KOUBARAKIS*
- A computational model of belief, *AARON N. KAPLAN, LENHART K. SCHUBERT*

ARTIFICIAL INTELLIGENCE V120 N°2 JULY 2000

- Internal reinforcement in a connectionist genetic programming approach, *ASTRO TELLER, MANUELA VELOSO*
- Credal networks, *FABIO G. COZMAN*
- Computer chess move-ordering schemes using move influence, *KIERAN GREER*
- Extending abstract argumentation systems theory, *P. BARONI, M. GIACOMIN, G. GUIDA*

AI MAGAZINE

<http://aaai.org/Magazine/>

COGNITIVE SCIENCE

http://www.elsevier.nl/gej-ng/29/18/show/Products/CITE/frames/frame_right.htm?area=Journals&factor=1

COGNITIVE SCIENCE V24 N°2 APRIL-JUNE 2000

- Getting to Best: Efficiency versus Optimality in Negotiation, *ELAINE B. HYDER, MICHAEL J. PRIETULA, LAURIE R. WEINGART*
- The nature and Processing of Errors in Interactive Behavior, *WAYNE D. GRAY*

- Solving Induction Reasoning Problems in Mathematics: Not-so Trivial Pursuit, *lisa a. haverty, R. KOEDINGER, DAVID KLAHR, MARTHA W. ALIBALI*
- Efficient Creativity: Constraint-Guided Conceptual Combination, *FIN-TAN J. COSTELLO, MARK T. KEANE*

APPLIED ARTIFICIAL INTELLIGENCE

<http://www.tandf.co.uk/JNLS/aai.htm>

APPLIED ARTIFICIAL INTELLIGENCE V14 N°4 APRIL 2000

- Guest editorial, *CHRISTIAN HOLZBAUR, THOM FRÜHWIRTH*
- University course timetabling using constraint handling rules, *SLIM ABDENNADHER, MICHAEL MARTE*
- A constraint-based framework for diagrammatic reasoning, *BERND MEYER*
- Symbolic constraints for meta-logic programming, *HENNING CHRISTIANSEN, DAVIDE MARTINENGI*
- A prolog constraint handling rules compiler and runtime system, *CHRISTIAN HOLZBAUR, THOM FRÜHWIRTH*
- On the incremental adaptation of CHR derivations, *ARMIN WOLF, THOMAS GRUENHAGEN, ULRICH GESKE*
- Selected recent project descriptions involving CHR

APPLIED ARTIFICIAL INTELLIGENCE V14 N°5 JUNE 2000

- Implementing a business process management system using adept: a real-world case study, *N. R. JENNINGS, P. FARATIN, T. J. NORMAN, P. O'BRIEN, B. ODGERS, J. L. ALTY*
- Cooperative learning strategies for intelligent tutoring systems, *ESMA AIMEUR, CLAUDE FRASSON, HUGO DUFORT*
- Compositional design and reuse of a generic agent model, *FRANCES M. T. BRAZIER, CATHOLIJN M. JONKER, JAN TREUR*

APPLIED ARTIFICIAL INTELLIGENCE V14 N°6 JULY 2000

- Fuzzy neural networks for learning

SOMMAIRES DES REVUES

fuzzy if-then rules, *R. J. KUO, P. C. WU, C. P. WANG*

• Ontological foundations for experimental science knowledge bases, *NATALYA FRIDMAN NOY, CAROLE D. HAFNER*

APPLIED ARTIFICIAL INTELLIGENCE V14 N°7 AUGUST 2000

Guest editorial, *PAOLO PETTA, JÖRG MÜLLER*

- On the sources of complexity in agent design, *MICHAEL WOOLDRIDGE*
- Vivid agents: theory, architecture, and applications, *MICHAEL SCHROEDER, GERD WAGNER*
- Information management at a bank using agents: theory and practice, *F. DIGNUM*
- Holonic transport scheduling with teletruck, *HANS-JÜRGEN BÜRCKERT, KLAUS FISCHER, GERO VIERKE*
- Proplant: multiagent system for production planning, *VLADIMIR MARIK, MICHAL PECHOUCEK, OLGA STEPANKOVA, JIRI LAZANSKY*

COMPUTATIONAL INTELLIGENCE
<http://www.blackwellpublishers.co.uk/asp/journal.asp?ref=08247935&src=cts>

COMPUTATIONAL INTELLIGENCE V16 N°3 AUGUST 2000

- Temporal Reasoning and Bayesian Networks, *AHMED Y. TAWFIK, ERIC M. NEUFELD*
- Parsing and Interpretation in the Minimalist Paradigm, *JAMES S. WILLIAMS, JUGAL K. KALITA*
- Choosing Rhetorical Structures To Plan Instructional Texts, *LEILA KOSSEIM, GUY LAPALME*
- Strategies in Human Nonmonotonic Reasoning, *MARILYN FORD, DAVID BILLINGTON*
- On Resolving Conflicts Between Arguments, *NICO ROOS*

MINDS AND MACHINES

JOURNAL FOR ARTIFICIAL

INTELLIGENCE PHILOSOPHY AND COGNITIVE SCIENCE

<http://www.wkap.nl/issuetoc.htm/0924-6495>

MINDS AND MACHINES V10 N°1 FEBRUARY 2000

- What the #*\$%! is a Subsymbol?!, *ISTVAN S. N. BERKELEY*
- Rethinking Autonomy, *RICHARD ALTERMAN*
- Reverse Psychologism, Cognition and Content, *TERRY DARTNALL*
- Limited Rationality in Action: Decision Support for Military Situation Assessment, *KATHRYN BLACKMOND LASKEY, BRUCE D'AMBROSIO, TOD S. LEVITT, SUZANNE MAHONEY*
- Satisficing Revisited, *MICHAEL A. GOODRICH, WYNN C. STIRLING, ERWIN R. BOER*
- Clarifying the Logic of Anti-Computationalism: Reply to Hauser, *SELMER BRINGSJORD*
- Ordinary Devices: Reply to Bringsjord's 'Clarifying the Logic of Anti-Computationalism: Reply to Hauser', *LARRY HAUSER*
- Universal Grammar or Common Syntax? A Critical Study of Jackendoff's Patterns in the Mind, *JAMES H. BUNN*
- Bringing Patterns into Focus: A Response to Bunn, *RAY JACKENDOFF*
- The Syntax of Galileo: Reply to Ray Jackendoff, *JAMES H. BUNN*
- Re-Entering the Chinese Room, *GRAHAM BUTTON, JEFF COULTER, JOHN R.E. LEE, WES SHARROCK*
- Colin Allen and Marc Bekoff, Species of Mind: The Philosophy and Biology of Cognitive Ethology, *LAWRENCE A. SHAPIRO*
- Gilles Fauconnier, Mappings in Thought and Language, *BARBARA ABBOTT*
- Sherman Wilcox, ed., Evolution of Communication, Vol. 1, No. 1, 1997, *JUSTIN LEIBER*
- Marc De Mey, The Cognitive Paradigm: An Integrated Understanding of Scientific Development, reprint, with a new introduction, *WILLIAM HERFEL*

MINDS AND MACHINES VOLUME 10, ISSUE 2, MAY 2000

- Using Extra Output Learning to Insert a Symbolic Theory into a Connectionist

Network, *M.R.W. DAWSON, D.A. MEDLER, D.B. MCCAUGHAN, L. WILLSON, M. CARBONARO*

- Bringing Knowing-When and Knowing-What Together: Periodically Tuned Categorization and Category-Based Timing Modeled with the Recurrent Oscillatory Self-Organizing Map (ROSOM), *MAURI KAIPAINEN, PASI KARHU*
- Deleting the Subject: A Feminist Reading of Epistemology in Artificial Intelligence, *ALISON ADAM*
- Computation and the Three Worlds, *MARIO RADOVAN*
- Is Design Relative or Real? Dennett on Intentional Relativism and Physical Realism, *REESE M. HEITNER*
- Janet W. Astington, Paul L. Harris and David R. Olson, eds., *Developing Theories of Mind*; Henry M. Wellman, *The Child's Theory of Mind*; Douglas Frye and Chris Moore, eds., *Children's Theories of Mind: Mental States and Social Understanding* Judith Felson Duchan, *JUDITH FELSON DUCHAN*
- M. Gams, M. Paprzycki and X. Wu, eds., *Mind Versus Computer: Were Dreyfus and Winograd Right?*, *Frontiers in Artificial Intelligence and Applications*, Vol. 43, Amsterdam: IOS Press, 1997, *CHARLES E. M. DUNLOP*
- Mark Crimmins, *Talk About Beliefs*, Cambridge, MA: MIT Press, 1992, *PHILIP L. PETERSON*
- Donald Gillies, *Artificial Intelligence and Scientific Method*, Oxford: Oxford University Press, 1996
- Ron McClamrock, *Existential Cognition: Computational Minds in the World*, Chicago: University of Chicago Press, 1995, *DIEGO MARCONI*
- James Trefil, *Are We Unique? A Scientist Explores the Unparalleled Intelligence of the Human Mind*, New York: John Wiley & Sons, 1997, *PAUL SCHWEIZER*
- Benny Shanon, *The Representational and the Presentational: An Essay on Cognition and the Study of the Mind*, Hemel Hempstead, Hertfordshire, U.K.: Harvester Wheatsheaf, 1993; Paramus, NJ: Prentice-Hall, 1994, *MARK H. BICKHARD*

SOMMAIRES DES REVUES

ARTIFICIAL INTELLIGENCE REVIEW

http://www.elsevier.nl/gej-ng/29/18/show/Products/CITE/frames/frame_right.htm?area=Journals&factor=1

ARTIFICIAL INTELLIGENCE REVIEW V14 N°1-2 APRIL 2000

- Editorial: Intelligent Help Systems for UNIX: Computational Models and Systems
- An Intelligent Human-Computer Interface for Provision of On-Line Help, *JENNIFER JERRAMS-SMITH*
- Virtues and Problems of an Active Help System for UNIX, *MARIA VIRVOU, JOHN JONES, MARK MILLINGTON*
- The Berkeley UNIX Consultant Project, *ROBERT WILENSKY, DAVID N. CHIN, MARC LURIA, JAMES MARTIN, JAMES MAYFIELD, DEKAI WU*
- The OSCON Operating System Consultant, *PAUL MC KEVITT*
- USCSH: An Active Intelligent Assistance System, *MANTON MATTHEWS, WALTER PHARR, GAUTAM BISWAS, HARISH NEELAKANDAN*

INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES

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

INTERNATIONAL JOURNAL OF HUMAN COMPUTER STUDIES V52 N°6 JUNE 2000

- Evaluating a domain-specialist-oriented knowledge management system, *TIMOTHY C. LETHBRIDGE*
- Creating an effective training environment for enhancing telework, *VISAWANATH VENKATESH, CHERI SPEIER*
- Using intentional models for the interface design of multi-level systems, *ALAN W. COLMAN, YING K. LEUNG*
- Calculators are needlessly bad, *HAROLD THIMBLEBY*
- Ontology-driven document enrichment: principles, tools and applications, *ENRICO MOTTA, SIMON BUCKINGHAM SHUM, JOHN DOMINGUE*
- WonderTools? A comparative study of ontological engineering tools, *A. J. DUINEVELD, R. STOTER, M. R. WEIDEN, B. KENEP, V. R. BENJAMINS*

INTERNATIONAL JOURNAL OF HUMAN COMPUTER STUDIES V53 N°1 JULY 2000

Editorial: Understanding work and designing artefacts, *ROBERT FIELDS, PETER WRIGHT*

- Activity theory, cognitive ergonomics and distributed cognition: three views of a transport company, *FRANÇOISE DECORTIS, SAMUEL NOIRFALISE, BERTHE SAUDELLI*
- Ethnography, theory and systems design: from intuition to insight, *CATRIONA MACAULAY, DAVID BENYON, ALISON CRERAR*
- Measurement in action: an activity-theoretical perspective on producer-user interaction, *MERVI HASU, YRJÖ ENGESTRÖM*
- Design guidelines for dealing with breakdowns and repairs in collaborative work settings, *MARK HARTSWOOD, ROB PROCTER*
- Building bridges: negotiating the gap between work practice and technology design, *TONI ROBERTSON*
- Putting ethnography to work: the case for a cognitive ethnography of design, *LINDEN J. BALL, THOMAS C. ORMEROD*
- Ethnographically informed analysis for software engineers, *STEPHEN VILLER, IAN SOMMERVILLE*
- The paradox of understanding work for design, *JOHN MCCARTHY*

INTERNATIONAL JOURNAL OF HUMAN COMPUTER STUDIES V53 N°2 AUGUST 2000

- Cognitive and gender factors influencing navigation in a virtual environment, *TIM R. H. CUTMORE, TREVOR J. HINE, KERRY J. MABERLY, NICOLE M. LANGFORD, GRANT HAWGOOD*
- Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics, *KATHERINE ISBISTER, CLIFFORD NASS*
- Structure, causality, visibility and interaction: propositions for evaluating engagement in narrative multimedia, *BRIDE MALLON, BRIAN WEBB*
- The coding principle and method for automatic recognition of Jia Gu Wen characters, *FENG LI, PENG-YUNG WOO*
- World Wide Web navigation aid,

MILENA HEAD, NORM ARCHER, YUFEI YUAN

COMPUTATIONAL LINGUISTICS

<http://mitpress.mit.edu/journal-contents.tcl?issn=08912017>

COMPUTATIONAL LINGUISTICS V26 N°2 JUNE 2000

- Integrating Text Planning and Linguistic Choice Without Abandoning Modularity: The IGEN Generator, *ROBERT RUBINOFF*
- A Model for Multimodal Reference Resolution, *LUIS ALBERTO PINEDA, GABRIELA GARZA*
- A Multi-Strategy Approach to Improving Pronunciation by Analogy, *YANNICK MARCHAND, ROBERT DAMPER*
- Models of Translational Equivalence among Words, *I. DAN MELAMED*
- Pipelines and Size Constraints, *EHUD REITER*
- **Book reviews**
- A Grammar Writer's Cookbook, Miriam Butt, Tracy Holloway King, Maria-Eugenia Nino, and Frederique Segond, by *MICHAEL MAXWELL*
- Local Constraints vs. Economy, David E. Johnson and Shalom Lappin, by *ANNIE ZAENEN*
- Predicative Forms in Natural Language and in Lexical Knowledge Bases, edited by Patrick Saint-Dizier, by *MANFRED STEDE*
- Lexical Semantics and Knowledge Representation in Multilingual Text Generation, Manfred Stede, by *BARBARA DI EUGENIO*
- The Mathematics of Syntactic Structure edited by Hans-Peter Kolb and Uwe Moennich, by *GERALD PENN*
- Foundations of Statistical Natural Language Processing, Christopher D. Manning and Hinrich Schuetze, by *LILLIAN LEE*
- Advances in Automatic Text Summarization, edited by Inderjeet Mani and Mark T. Maybury, by *MARK SANDERSON*
- Extended Finite State Models of Language, edited by Andras Kornai, by *ED KAISER*
- Optimality Theory, Rene Kager, by *JASON EISNER*

SOMMAIRES DES REVUES

- Systemic Functional Grammar in Natural Language Generation, Elke Teich, by *GRAHAM WILCOCK*

NATURAL LANGUAGE ENGINEERING

<http://www.cup.cam.ac.uk/journals/nle/nleETOC.HTM>

NATURAL LANGUAGE ENGINEERING V5 N°2

- Distinguishing systems and distinguishing senses: new evaluation methods for Word Sense Disambiguation, *PHILIP RESNIK, DAVID YAROWSKY*
- Matching the tagging to the task, *GEORGE A. MILLER, RANDEE TENGI, SHARI LANDES*
- A corpus-based bootstrapping algorithm for Semi-Automated semantic lexicon construction, *ELLEN RILOFF, JESSICA SHEPHERD*
- Verb sense disambiguation based on dual distributional similarity, *JEONGMI CHO, JUNGYUN SEO, GIL CHANG KIM*
- Semantic tagging of unknown proper nouns, *ALESSANDRO CUCCHIARELLI, DANILO LUZI, PAOLA VELARDI*
- Recognizing subjectivity: a case study in manual tagging, *REBECCA F. BRUCE, JANYCE M. WIEBE*
- SENSE: an analogy-based Word Sense Disambiguation system, *STEFANO FEDERICI, SIMONETTA MONTMAGNI, VITO PIRRELLI*

NATURAL LANGUAGE ENGINEERING V5 N°3

- An approach to program understanding by natural language understanding, *LETHA H. ETZKORN, LISA L. BOWEN, CARL G. DAVIS*
- Routing email automatically by purpose not topic, *HAMID KHOSRAVI, YORICK WILKS*
- A reestimation algorithm for probabilistic dependency grammars, *SEUNGMI LEE, KEY-SUN CHOI*
- Parsing with discontinuous phrases, *ALLAN RAMSAY*

Book reviews

Roger Garside, Geoffrey Leech, Anthony McEnery (eds). *Corpus Annotation: Linguistic Information from Computer Text Corpora.*

Longman. 1997, by *OLIVER MASON*, Dafydd Gibbon, Roger Moore, and Richard Winski (eds). *Handbook of Standards and Resources for Spoken Language Systems.* Mouton de Gruyter. 1997, by *GEOFFREY SAMPSON*

Ronnie W. Smith and D. Richard Hipp. *Spoken Natural Dialog Systems: A Practical Approach.* Oxford: Oxford University Press, 1994, by *DEBORAH DAHL*

NATURAL LANGUAGE ENGINEERING V5 N°4

- Context-sensitive spoken dialogue processing with the DOP model, *RENS BOD*
- MLDS: A translator-oriented MultiLingual dictionary system, *E. AGIRRE, X. ARREGI, X. ARTOLA, A. DIAZ DE ILARRAZA, K. SARASOLA, A. SOROA*
- Topic-based mixture language modeling, *YOSHIHIKO GOTOH, STEVE RENALS*
- Evaluating two methods for Treebank grammar compaction, *ALEXANDER KROTOV, MARK HEPPLE, ROBERT GAIZAUSKAS, YORICK WILKS*

Book reviews

Jennifer Pearson. *Terms in Context.* John Benjamins Publishing Co., Amsterdam. 1998, by *PAUL R. BOWDEN*

Nomi Erteschik-Shir, *The Dynamics of Focus Structure.* Cambridge: Cambridge University Press, 1997, by *MAYUMI MASUKO*

USER MODELING AND USER-ADAPTED INTERACTION

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

COMPUTER SPEECH AND LANGUAGE

<http://www.ideallibrary.com/links/toc/csla/latest>

COMPUTER SPEECH & LANGUAGE V14 N°3 JULY 2000

- ProSynth: an integrated prosodic approach to device-independent, natu-

ral-sounding speech synthesis, *RICHARD OGDEN, SARAH HAWKINS, JILL HOUSE, MARK HUCKVALE, JOHN LOCAL, PAUL CARTER, JANA DANKOVIOVA, SEBASTIAN HEID*

- Incorporating lip protrusion and larynx lowering into a time domain model for articulatory speech synthesis, *C. C. GOODYEAR*
- Tree-structured vector quantization for speech recognition, *M. BARSZCZ, W. CHEN, G. BOULIANNE, P. KENNY*
- Multi-resolution sub-band features and models for HMM-based phonetic modelling, *P. M. MCCOURT, S. V. VASEGHI, B. DOHERTY*
- Statistical language modeling with a class-based n-multigram model, *SABINE DELIGNE, YOSHINORI SAGISAKA*

MACHINE LEARNING

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

MACHINE LEARNING V40 N°3 SEPTEMBER 2000

- A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms, *TJEN-SIEN LIM, WEI-YIN LOH, YU-SHAN SHIH*
- Randomizing Outputs to Increase Prediction Accuracy, *LEO BREIMAN*
- Learning to Play Chess Using Temporal Differences, *JONATHAN BAXTER, ANDREW TRIDGELL, LEX WEAVER*
- A Study of Reinforcement Learning in the Continuous Case by the Means of Viscosity Solutions, *REMI MUNOS*

MACHINE LEARNING V41 N°1 OCTOBER 2000

- Technical Note: Naive Bayes for Regression, *EIBE FRANK, LEONARD TRIGG, GEOFFREY HOLMES, IAN H. WITTEN*
- Selecting Examples for Partial Memory Learning, *MARCUS A. MALOOF, RYSZARD S. MICHALSKI*
- Lazy Learning of Bayesian Rules, *ZIJIAN ZHENG, GEOFFREY I. WEBB*
- A Cognitive Bias Approach to Feature Selection and Weighting for Case-Based Learners, *CLAIRE CARDIE*

MACHINE LEARNING V41 N°2 NOVEMBER 2000

- Maximizing Theory Accuracy Through Selective Reinterpretation, *SHLOMO ARGAMON-ENGELSON, MOSHE KOPPEL, HILLEL WALTERS*
- Learning Changing Concepts by Exploiting the Structure of Change, *PETER L. BARTLETT, SHAI BEN-DAVID, SANJEEV R. KULKARNI*
- A Formalism for Relevance and Its Application in Feature Subset Selection, *DAVID A. BELL, HUI WANG*
- Adaptive Versus Nonadaptive Attribute-Efficient Learning, *PETER DAMASCHKE*
- Phase Transitions in Relational Learning, *ATTILIO GIORDANA, LORENZA SAITTA*

NEURAL NETWORKS

http://www.elsevier.nl/cite/54/show/jrnl_index.htm

NEURAL NETWORKS V13 N°3 APRIL 2000

- Self-organized hierarchical structure in a plastic network of chaotic units, *J. ITO, K. KANEKO*
- Improving local minima of Hopfield networks with augmented Lagrange multipliers for large scale TSPs, *M. MARTIN-VALDIVIA, A. RUIZ-SEPULVEDA, F. TRIGUERO-RUIZ*
- Learning non-stationary conditional probability distributions, *D. HUSMEIER*
- The patchwork engine: image segmentation from shape symmetries, *G. J. VAN TONDER, Y. EJIMA*
- Position invariant recognition in the visual system with cluttered environments, *S. M. STRINGER, E. T. ROLLS*
- Local minima and plateaus in hierarchical structures of multilayer perceptrons, *K. FUKUMIZU, S. AMARI*
- Learning in higher order Boltzmann machines using linear response, *M. A. R. LEISINK, H. J. KAPPEN*
- A recurrent neural network for solving linear projection equations, *YOU-SHEN XIA, JUN WANG*
- Efficient perceptron learning using constrained steepest descent, *S. J. PERANTONIS, V. VIRVILIS*
- Information complexity of neural networks, *M. A. KON, L. PLASKOTA*

- A connectionist model for convex-hull of a planar set, *A. DATTA, S. PAL, N. R. PAL*
- Multilayer neural networks for solving a class of partial differential equations, *S. HE, K. REIF, R. UNBEHAUEN*
- A fruitful blend, or a trinket-box? A book review of The MIT Encyclopedia of the Cognitive Sciences. R.A. Wilson, F.C. Keil (Eds.); MIT Press, Cambridge, MA, 1999, *RAJEEV D. S. RAIZADA*

NEURAL NETWORKS V13 N°4-5 JUNE 2000

- Bias reduction in skewed binary classification with Bayesian neural networks, *P.J.G. LISBOA, A. VELLIDO, H. WONG*
- Independent component analysis: algorithms and applications, *A. HYVÄRINEN, E. OJA*
- Evolutionary robots with on-line self-organization and behavioral fitness, *D. FLOREANO, J. URZELAI*
- A generalized Hebbian rule for activity-dependent synaptic modifications, *T. KITAJIMA, K. HARA*
- Mutual information of sparsely coded associative memory with self-control and ternary neurons, *D. BOLLE, D.R.C. DOMINGUEZ, S. AMARI*
- Construction of confidence intervals for neural networks based on least squares estimation, *I. RIVALS, L. PERSONNAZ*
- A new algorithm for learning in piecewise-linear neural networks, *E.F. GAD, A.F. ATIYA, S. SHAHEEN, A. EL-DESSOUKI*
- Evolution and generalization of a single neurone. III. Primitive, regularized, standard, robust and minimax regressions, *S. RAUDYS*
- Determining the number of centroids for CMLP network, *M. LEHTOKANGAS*
- Defining a neural network controller structure for a rubbertuator robot, *M. OZKAN, K. INOUE, K. NEGISHI, T. YAMANAKA*
- An efficient learning algorithm for improving generalization performance of radial basis function neural networks, *ZHENG-OU WANG, TAO ZHU*
- Adaptive Resonance Theory Microchips: Circuit Design Techniques - T. Serrano-Gotarredona, B. Linares-Barranco, A.G. Andreou; Kluwer Academic Publishers, Boston, 1998, hardbound, ISBN: 0-792-38231-5, US\$ 120.00, *A.E. HUBBARD, T.A. HINCK*

Academic Publishers, Boston, 1998, hardbound, ISBN: 0-792-38231-5, US\$ 120.00, *A.E. HUBBARD, T.A. HINCK*

NEURAL NETWORKS V13 N°6 JULY 2000

- Neural networks with a continuous squashing function in the output are universal approximators, *J.L. CASTRO, C.J. MANTAS, J.M. BENITEZ*
- Implications of physical symmetries in adaptive image classifiers, *T. SAMS, J.L. HANSEN*
- Visual cortical mechanisms of perceptual grouping: interacting layers, networks, columns, and maps, *W.D. ROSS, S. GROSSBERG, E. MINGOLLA*
- Blind source separation in the presence of weak sources, *J.-P. NADAL, E. KORUTCHEVA, F. AIRES*
- Blind signal processing by the adaptive activation function neurons, *S. FIORI*
- Results concerning the absolute stability of delayed neural networks, *M. JOY*
- Dynamic topology representing networks, *J. SI, S. LIN, M.-A. VUONG*
- Consistency of posterior distributions for neural networks, *H.K.H. LEE*
- On the exact solution of the Parity-N problem using ordered neural networks, *E. LAVRETSKY*
- Approximation of continuous and discontinuous mappings by a growing neural RBF-based algorithm, *A. ESPOSITO, M. MARINARO, D. ORICCHIO, S. SCARPETTA*
- Stable behavior in a recurrent neural network for a finite state machine, *K. ARAI, R. NAKANO*
- A neural network for 500 word vocabulary word spotting using non-uniform units, *H.-J. YU, Y.-H. OH*
- Ionic current and metabolism for brain scanners (a three state-model of modular activation), *J.F. GOMEZ-M*

ARTIFICIAL INTELLIGENCE IN MEDICINE

http://www.elsevier.nl/gej-ng/29/18/show/Products/CITE/frames/frame_right.htm?area=Journals&factor=1

ARTIFICIAL INTELLIGENCE IN MEDICINE V19 N°3 JULY 2000

- Knowledge-based information management in intensive care and anesthesia, *MICHEL DOJAT, SILVIA MIKSCHE, JIM HUNTER*

- Multiple signal integration by decision tree induction to detect artifacts in the neonatal intensive care unit, *CHRISTINE L. TSIEN, ISAAC S. KOHANE, NEIL MCINTOSH*
- Towards symbolization using data-driven extraction of local trends for ICU monitoring, *D. CALVELO, M.-C. CHAMBRIN, D. POMORSKI, P. RAVAUX*
- Knowledge discovery and knowledge validation in intensive care, *KATHARINA MORIK, MICHAEL IMBOFF, PETER BROCKHAUSEN, THORSTEN JOACHIMS, URSULA GATHER*
- A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU, *PETER J.F. LUCAS, NICOLETTE C. DE BRUIJN, KARIN SCHURINK, ANDY HOEPELMAN*

ARTIFICIAL INTELLIGENCE IN MEDICINE V20 N°1 SEPTEMBER 2000

- Artificial intelligence in medicine and medical decision making Europe, *WERNER HORN*
- Guideline-based careflow systems, *S. QUAGLINI ET AL.*
- Internet-based decision-support server for acute abdominal pain, *HANS-PETER EICH, CHRISTIAN OHMANN*
- Intelligent analysis of clinical time series: an application in the diabetes mellitus domain, *R. BELLAZZI, C. LARIZZA, P. MAGNI, S. MONTANI, M. STEFANELLI*
- Machine learning for survival analysis: a case study on recurrence of prostate cancer, *BLA ZUPAN, JANEZ DEMAR, MICHAEL W. KATTAN, J. ROBERT BECK, I. BRATKO*
- A cooperative framework for segmentation of MRI brain scans, *LAURENCE GERMOND, MICHEL DOJAT, C. TAYLOR, C. GARBAY*

ARTIFICIAL INTELLIGENCE IN MEDICINE V20 N°2 OCTOBER 2000

- AI planning and scheduling in the medical hospital environment, *CONSTANTINE D. SPYROPOULOS*
- Toward interactive scheduling systems for managing medical resources, *ANGELO ODDI, AMEDEO CESTA*
- Continual planning and scheduling

for managing patient tests in hospital laboratories, *CATHERINE C. MARINAGI, CONSTANTINE D. SPYROPOULOS, CHRISTOS PAPTAEODOROU, STAVROS KOKKOTOS*

- Hybrid optimization techniques for the workshift and rest assignment of nursing personnel, *C. VALOIXIS, E. HOUSOS*
- Erratum to Theme Issue on « Evolutionary Computation in Medicine » - [Artificial Intelligence in Medicine 19 (1) (2000)]
Erratum to « Knowledge discovery and knowledge validation in intensive care » - [Artificial intelligence in Medicine 19 (2000) 225-249]

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

http://www.elsevier.nl/gej-ng/29/18/show/Products/CITE/frames/frame_right.htm?area=Journals&factor=1

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING V25 N°1 JULY 2000

- MV-algebras, multiple bets and subjective states, *BRUNELLA GERLA*
- Approximate reasoning and conflict resolution, *RONALD R. YAGER*
- Monetary unit sampling: a belief-function implementation for audit and accounting applications, *PETER R. GILLET*

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING V25 N°2 OCTOBER 2000

- Nonlinear nonnegative multiregressions based on Choquet integrals, *ZHENYUAN WANG, KWONG-SAK LEUNG, MAN-LEUNG WONG, JIAN FANG, KEBIN XU*
- A neural model for fuzzy Dempster-Shafer classifiers, *ELISABETTA BINAGHI, IGNAZIO GALLO, PAOLO MADELLA*
- Minimal approximate hitting sets and rule templates, *STAAL VINTERBO, ALEKSANDER ØHRN*
- Design and rule base reduction of a fuzzy filter for the estimation of motor currents, *DAN SIMON*

THE JOURNAL OF LOGIC PROGRAMMING

<http://www.elsevier.nl/inca/publications/store/5/0/5/7/3/1/index.htm?menu=cont>

THE JOURNAL OF LOGIC PROGRAMMING V44 N°1-3

- Special issue: abductive logic programming, *MARC DENECKERANTONIS KAKAS*
- Abducing through negation as failure: stable models within the independent choice logic, *DAVID POOLE*
- Conditional reasoning in logic programming, *D. GABBAY, L. GIORDANO, A. MARTELLI, N. OLIVETTI, M. L. SAPINO*
- Abductive logic programming and disjunctive logic programming: their relationship and transferability, *CHIAKI SAKAMA, KATSUMI INOUE*
- An abductive approach to disjunctive logic programming, *JIA-HUAI YOU, LI YAN YUAN, RANDY GOEBEL*
- ACLP: Abductive Constraint Logic Programming, *A. C. KAKAS, A. MICHAEL, C. MOURLAS*
- Computing abduction by using TMS with top-down expectation, *NOBORU IWAYAMA, KEN SATOH*
- An abductive event calculus planner, *MURRAY SHANAHAN*

JOURNAL OF LOGIC PROGRAMMING V45 N°1-3

- An abstract machine for efficiently computing queries to well-founded models, *KONSTANTINOS SAGONAS, TERRANCE SWIFT, DAVID S. WARREN*
- Dynamic updates of non-monotonic knowledge bases, *J.J. ALFERES, J.A. LEITE, L.M. PEREIRA, H. PRZYMUSINSKA, T.C. PRZYMUSINSKI*
- Introduction to the constraint language NCL, *JIANYANG ZHOU*
- Argumentation-based abduction in disjunctive logic programming, *KEWEN WANG*
- A simple polynomial groundness analysis for logic programs, *ANDY HEATON, MUHAMED ABO-ZAED, MICHAEL CODISH, ANDY KING*

DATA MINING

<http://www.research.microsoft.com/research/datamine/Issues.htm>

PÉPITE : un système d'assistance au diagnostic de compétences

par Stéphanie Jean

Thèse de l'Université du Maine, soutenue le 21 janvier 2000,
à l'Université du Maine.

Résumé

Notre travail de recherche a pour objectif de modéliser les connaissances des élèves en algèbre élémentaire en construisant leur profil cognitif. À cette problématique de modélisation de l'apprenant, s'ajoute une forte préoccupation d'usage : nous souhaitons concevoir un système utilisable dans l'enseignement. Pour lever en partie les difficultés posées par la modélisation de l'apprenant, nous proposons une façon nouvelle d'envisager le diagnostic. D'une part, le point de départ de notre travail est une analyse didactique multidimensionnelle validée dont le résultat est un outil papier-crayon qui aboutit à la création de profils cognitifs des élèves. C'est l'établissement de ce profil que nous avons informatisé. D'autre part, nous considérons le logiciel élève, permettant le recueil des observables, comme faisant partie intégrante du diagnostic. De la qualité des observables dépend en effet en partie la qualité du modèle de l'apprenant construit. Enfin, avec PÉPITE, nous nous adressons aux enseignants comme utilisateurs finals d'un logiciel et nous leur proposons non pas un système de diagnostic automatique, mais un système d'assistance au diagnostic. L'élaboration du profil de l'élève se fait en trois étapes auxquelles correspondent trois modules dans PÉPITE : le recueil des observables avec le logiciel élève, le diagnostic, puis la présentation des profils à l'enseignant. Chaque étape pose des problèmes particuliers. La réalisation de ce système a conduit à plusieurs résultats. Nous avons tout d'abord pu recueillir des observables fiables et représentatifs du fonctionnement des élèves. Nous avons également montré, en réalisant un prototype du module de diagnostic, que l'analyse multidimensionnelle automatique est possible pour des réponses d'élèves à différents types de questions, incluant des questions ouvertes. Nous avons de plus proposé un logiciel enseignant facilitant l'appropriation des profils par l'enseignant. Enfin, le travail fortement interdisciplinaire qui caractérise notre recherche nous a conduit à concevoir une méthode de conception, la méthode de conception différenciée, permettant aux concepteurs informaticiens de s'adresser de façon adaptée aux différents membres de l'équipe de conception interdisciplinaire.

Mots clés : EIAO, assistance au diagnostic, pluridisciplinarité.

Contact : Stéphanie Jean, IC2 — LIUM, 72085 Le MANS
CEDEX 9 (FRANCE).

Mél : Stephanie.Jean@univ-lemans.fr — <http://www-ic2.univ-lemans.fr/~jean/>

Tel : 33/0 2.43.83.38.75 — Fax : 33/0 2.43.83.38.68

Adresse où le document peut être récupéré : http://www-ic2.univ-lemans.fr/~jean/these_HTML/

Systèmes classificatoires par objets pour l'extraction de connaissances dans des bases de données

Arnaud Simon

Thèse de l'Université Henri Poincaré à Nancy 1 présentée le
23 septembre 2000 à l'Université Henri Poincaré à Nancy.

Résumé.

Cette dernière décennie a vu évoluer la problématique relative au traitement des données d'une part par un facteur d'échelle — les données ainsi que la représentation qui leur est associée sont de plus en plus nombreuses — et de disponibilité (notamment par le développement du WEB) de l'autre. L'étude des techniques permettant d'exploiter ces masses de données maintenant disponibles fait l'objet du courant de recherche appelé extraction de connaissances dans des bases de données ou ECBD. L'approche de l'ECBD que nous avons adoptée est fondée sur l'étude d'outils classificatoires par objets. Les techniques classiques d'arbres de décision et principalement de treillis de Galois sont adaptées — ces techniques relèvent des problématiques issues des modèles fondés sur des « tableaux de données » — au cas où les données relèvent d'un système de représentation des connaissances par objets, ou système de RCO. Un système de RCO a pour fonction de stocker et d'organiser les connaissances autour de la notion de hiérarchie d'objets et de fournir des services inférentiels et d'accès à l'information fondés sur des méthodes classificatoires. Par ailleurs, un système de RCO permet de manipuler des données « complexes » car leurs attributs ne se réduisent pas à des types de base atomique mais peuvent être d'autres objets (relations) ; aussi bien mono que multi-valués. Les fonctionnalités de représentation et d'inférence des systèmes de RCO sont mises à profit pour l'ECBD pour elles-mêmes mais aussi pour la construction de structures classificatoires d'arbres de décision et de treillis de Galois. Les structures ainsi construites sont des hiérarchies d'objets qui relèvent du modèle de RCO. Une telle approche a pour avantages de prendre en considération les connaissances disponibles et de produire des structures directement interprétables et réutilisables, pour mener à bien des inférences ou des constructions de nouvelles structures. Les structures de treillis de Galois sont également mises à profit pour la construction de toutes les règles d'association partielles. Nous montrons que notre approche possède l'avantage de construire moins de règles, car des règles redondantes ne sont pas produites, de permettre une mise à jour des règles pour tenir compte de nouveaux individus ou de nouvelles propriétés et d'améliorer l'expressivité des règles.

Mots clés : extraction de connaissances dans les bases de données (fouille de données), systèmes de représentation de connaissances par objets, treillis de Galois, arbres de décision.

Contact : Arnaud Simon, Department of Computing Science,
University of Newcastle upon Tyne, Newcastle upon Tyne,
NE1 7RU, England.

Mél : arnaud.simon@ncl.ac.uk

Adresse où le document peut être récupéré :
<http://www.loria.fr/~simona>

Classification de séquences audiovisuelles

Jean Carrive

Thèse de l'Université Pierre et Marie Curie (Paris 6), soutenue le 12 septembre à l'Institut National de l'Audiovisuel, Centre Pierre Sabbagh (83/85 rue de Patay — 75013 Paris).

Résumé.

Dans le contexte de l'indexation automatique de documents audiovisuels, nous nous intéressons aux collections d'émissions de télévision respectant une structure générale commune, telles que certains journaux télévisés ou certains magazines d'actualité ou de variétés. Nous cherchons à représenter et à reconnaître dans de telles émissions des séquences caractéristiques pouvant être annotées comme un tout par un(e) documentaliste, comme une chanson dans un magazine de variétés ou un reportage de journal télévisé. Nous proposons un langage de représentation de ces classes de séquences appelées canevas. La définition d'un canevas repose d'une part sur un vocabulaire d'événements primitifs directement observables, tels que les logos ou les jingles, et d'autre part sur un ensemble

de relations temporelles prescrites entre ces événements. Nous proposons de représenter les événements primitifs dans un formalisme de logiques de descriptions et nous étendons l'algèbre d'intervalles temporels de Allen avec un opérateur d'itération noté « * » désignant des séquences d'événements temporellement adjacents ainsi qu'avec des contraintes spécifiques faisant intervenir une négation. Nous présentons le problème de la reconnaissance des canevas comme un problème de satisfaction de contraintes et nous proposons une méthode de résolution efficace ainsi qu'une implémentation de cette méthode dans un système de satisfaction de contraintes existant. Nous exposons enfin une série d'expérimentations conduites à l'Institut National de l'Audiovisuel (INA) sur un corpus de journaux télévisés.

Mots clefs : indexation audiovisuelle, collection d'émissions, logiques de descriptions, raisonnement temporel, satisfaction de contraintes

Contact : Jean Carrive, DRP — INA, 4, avenue de l'Europe, 94366 Bry-sur-Marne.

Mél : jcarrive@ina.fr

Tél. : 01 49 83 29 69 — Fax : 01 49 83 25 82

Pour l'envoi d'un résumé de thèse ou d'habilitation :

Envoyer à Amedeo Napoli (adresse page 2 de ce Bulletin) un fichier texte ou un document word attaché contenant les informations suivantes :

Titre de la thèse	Mots clefs (5 maximum)
Prénom et nom du candidat	
Type de la thèse	Adresse postale du candidat
(Thèse de l'Université de xxx, ou Habilitation à diriger	Mél
des recherches de l'Université de xxx)	Adresse Web
Date et lieu de soutenance de la thèse	Téléphone et télécopie (facultatif)

Résumé (une demi-page)

Adresse où le document peut être récupéré (le cas échéant)

L'ensemble de ces informations doit tenir en 350 mots environ.

Dossier Documents et Intelligence Artificielle

Date de parution prévue : janvier 2001

(bulletin numéro 44)

Coordinatrice : Claudie FAURE, ENST, département TSI, 46 rue Barrault, 75634 Paris cedex 13
cfaure@tsi.enst.fr

Objectifs

Depuis un peu plus de dix ans, l'analyse du document s'affirme comme un domaine de recherche en constituant des associations (Groupe de Recherche en Communication Écrite en France), des comités techniques (TC 10 et TC 11 au sein de l'International Association of Pattern Recognition), en organisant de nombreuses conférences dont le Colloque National sur l'Écrit et le Document qui est maintenant le Colloque International Francophone sur ce thème, la conférence IEEE (International Conference on Document Analysis and Recognition), le Colloque International sur le Document Électronique. Des numéros spéciaux de revues scientifiques sont consacrés à ce domaine qui dispose maintenant de sa propre revue : International Journal on Document Analysis and Recognition.

Les objectifs de l'analyse automatique du document s'inscrivent dans la continuité des fonctions qui lui sont principalement attachées : la communication et la mémorisation. La plupart des documents sont sous forme papier, le traitement de leur image numérique permet d'automatiser des tâches (comme le tri postal ou la gestion des chèques) mais aussi de construire des versions numériques structurées de ces documents pour sauver les savoirs qui meurent sur des supports fragiles, rendre l'information largement accessible, faciliter sa mise à jour et son usage. Les documents peuvent aussi être produits ou modifiés directement sous forme numérique à l'aide de dispositifs de saisie des gestes graphiques. La reconnaissance de l'écriture et de signes, la segmentation et catégorisation des composantes physiques, la structuration logique, voire une lecture complète dans certaines applications, seront les tâches essentielles de l'analyse du document.

Le choc produit ces dernières années par l'évolution rapide des technologies de la communication a un effet considérable sur le domaine de l'analyse du document. En conséquence, on assiste à un rapprochement avec la

communauté scientifique des spécialistes de la gestion, la représentation, la diffusion des documents électroniques. On constate aussi une forte convergence avec le domaine de l'interaction homme-machine sur des questions relatives à l'interaction par le geste, la lecture active, la conception et la visualisation des documents.

L'objectif de ce dossier est de faire connaître qui fait quoi dans le domaine du document, de mesurer les effets de l'évolution récente des technologies de communication en termes d'orientation des thèmes d'études, et de signaler des coopérations effectives entre des équipes à compétences complémentaires.

Merci d'envoyer, dès que possible, votre intention de participer à ce dossier, à l'adresse électronique suivante : cfaure@tsi.enst.fr

Présentation des contributions

Les contributions expliqueront la problématique visée et présenteront les concepts, méthodes, techniques, expériences mises en œuvre. Les équipes du secteur public ou privé se présenteront par :

- nom et adresse de l'équipe, du laboratoire ou de l'entreprise
- chercheurs et ingénieurs concernés
- titre de la (ou des) études menée(s)
- types de méthodes principalement utilisées
- panorama des travaux effectués ou en cours (en précisant les coopérations avec d'autres équipes et l'évolution des thèmes et des objectifs des études sur le document dans cette l'équipe)
- bibliographie (principales références sur les travaux décrits)

Le volume souhaité des contributions est d'environ une demi à une page, à envoyer

***** Avant le 15 novembre 2000 *****

à Claudie Faure, par courrier électronique, au format Word ou RTF dépourvu de toute mise en forme (ni tableau ni tabulation) hormis le gras et l'italique.

PRÉVISIONS DE DOSSIERS DANS LES PROCHAINS BULLETINS

IA et fouille de données - avril 2001 (bulletin numéro 45)

IA et médecine - juillet 2001 (bulletin numéro 46)

Adhésion et abonnement **Demande** **Renouvellement**

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

	Consultation du bulletin sur WEB (un accès)	Envoi du bulletin papier + un accès pour consultation du bulletin sur WEB
<input type="checkbox"/> Adhésion simple :	200 francs	400 francs
<input type="checkbox"/> Adhésion étudiant (sur justificatif) :	100 francs	200 francs
<input type="checkbox"/> Adhésion de soutien	400 francs	800 francs
<input type="checkbox"/> Abonnement au bulletin sans adhésion	/	350 francs

- Adhésion au collège IADSMA : ajouter100 francs
 Adhésion au proto-collège *Apprentissage* :*exceptionnellement gratuit pour l'année 2000*
 Adhésion + abonnement à la Revue de l'Intelligence Artificielle : ajouter819,95 francs (réglés par chèque personnel ET envoi de la revue à l'adresse personnelle)
 Adhésion + abonnement à la Revue de l'Intelligence Artificielle : ajouter1311,91 francs (réglés par bon de commande OU envoi de la revue à l'adresse professionnelle)

Veillez trouver un règlement (à l'ordre de l'AFIA) de francs

Adhésion Personne morale **Demande** **Renouvellement****Organisme :**

Nom et prénom du représentant : cachet de l'organisme :
 Fonction :
 Adresse postale :
 N° de téléphone :
 N° de télécopie :
 Adresse électronique :

	Envoi du bulletin papier + 5 accès pour consultation du bulletin sur WEB
<input type="checkbox"/> Laboratoires universitaires	1500 francs
<input type="checkbox"/> Personnes morales non universitaires	3000 francs
<input type="checkbox"/> Adhésion de soutien	4500 francs

Je joins un bon de commande pour une somme de francs

- j'accepte que les renseignements ci-dessus apparaissent dans l'annuaire de l'AFIA
 j'accepte que les renseignements ci-dessus soient transmis à l'ECCAI pour constituer un fichier européen.
 je m'oppose à toute diffusion des renseignements ci-dessus.

Ce bulletin d'adhésion doit être retourné à :

René QUINIOU : INRIA / IRISA, Campus de Beaulieu 35042 Rennes Cedex
 Société Générale, Résidence du Val de Seine, 78430 LOUVECIENNES
 code banque30003, code guichet 01902, numéro de compte 00037283856 clef RIB 39

TVA non applicable, article 293B du CGI

Vous pouvez également adhérer par Internet, à l'URL : <http://www.afia.polytechnique.fr/accueil/adhesion.html>

Les dossiers du *Bulletin de l'AFIA*

IA et connexionnisme	Bulletin n°43	I.A. et musique	Bulletin
IA et Vie Artificielle	Bulletin n°42	n°23 Apprentissage	Bulletin n°22
IA et CHM	Bulletin n°41	Les explications dans les SBC	Bulletin n°20
IA et EIAH	Bulletin n°40	Pétrole-Chimie	Bulletin n°19
Plates-formes multi-agents	Bulletin n°39	Le raisonnement à partir de cas	Bulletin n°18
IA et WEB	Bulletin n°38	I.A. et temps-réel	Bulletin n°17
Mémoires d'entreprises	Bulletin n°36	Planification et action	Bulletin n°16
IA et logique	Bulletin n°35	Traitement automatique des langues	Bulletin n°15
Ingénierie des connaissances	Bulletin n°34	I.A. et médecine	Bulletin n°14
IA et Télécommunications	Bulletin n°33	Diagnostic à base de modèles	Bulletin n°13
IA et Terminologie	Bulletin n°32	Validation des SBC	Bulletin n°12
Décision et IA	Bulletin n°31	Le connexionnisme	Bulletin n°11
Raisonnement IA et Image	Bulletin n°30	I.A. et jeux	Bulletin n°10
Raisonnement temporel et spatial	Bulletin n°29	E.I.A.O.	Bulletin n°9
Systèmes Multi-agents	Bulletin n°28	I.A. et gestion	Bulletin n°8
IA et robotique	Bulletin n°27	Conception et I.A.	Bulletin n°7
I.A. et biologie moléculaire	Bulletin n°26	Intelligence artificielle distribuée	Bulletin n°6
I.A. et droit	Bulletin n°25	Acquisition des Connaissances	Bulletin n°5
I.A. et fusion de données	Bulletin n°24	IA et ordonnancement	Bulletin n°4

S O M M A I R E D U B U L L E T I N N ° 4 2

La vie de l'A.F.I.A.	4
Débats	5
Présentation de Laboratoires	11
Dossier IA et Connexionnisme	16
Livres	29
Conférences	32
Calendrier	36
Sommaire des revues	38
Résumés habilitations et thèses	44
Appels à dossiers	46

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: 30 septembre
Sortie le 30 avril	Sortie le 31 octobre