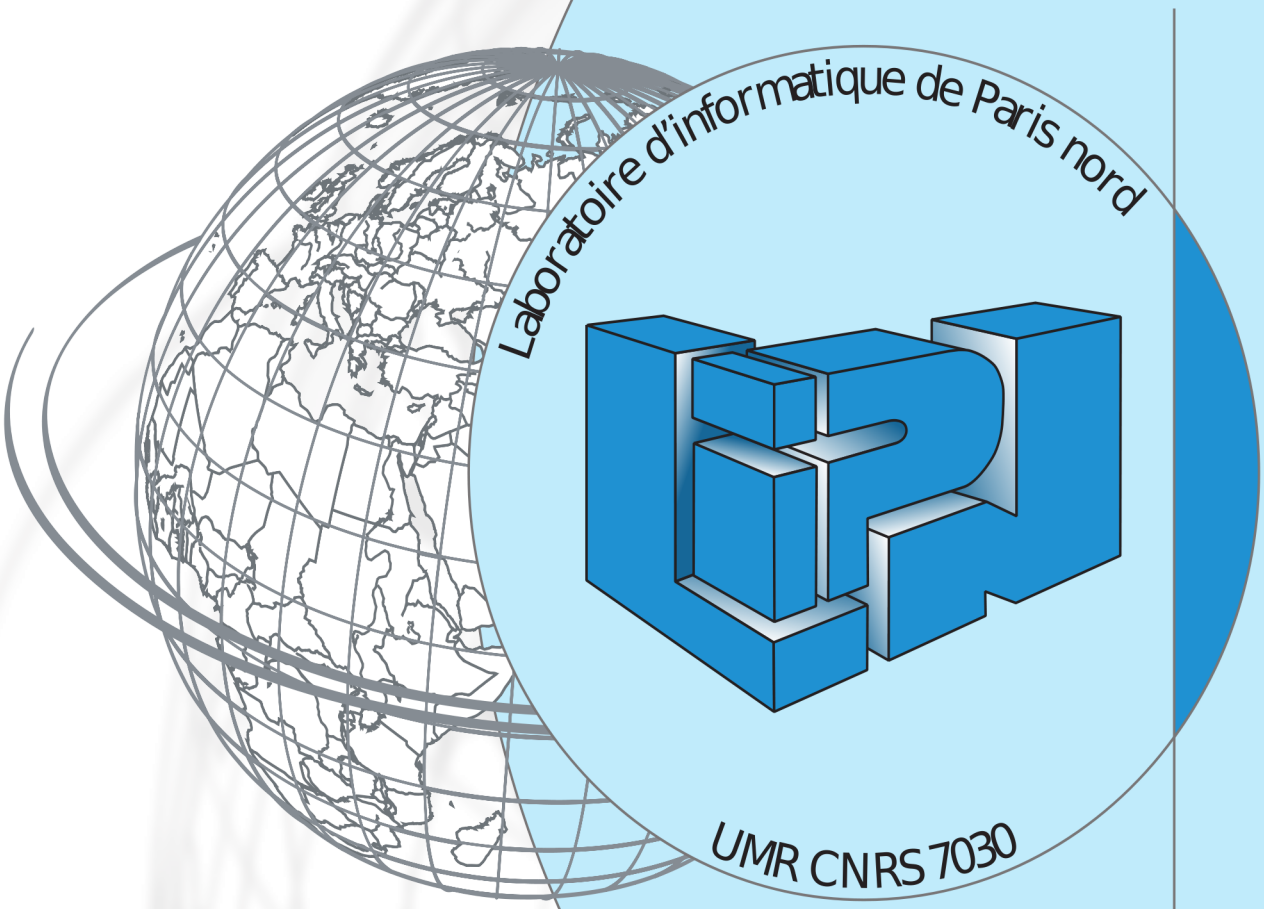
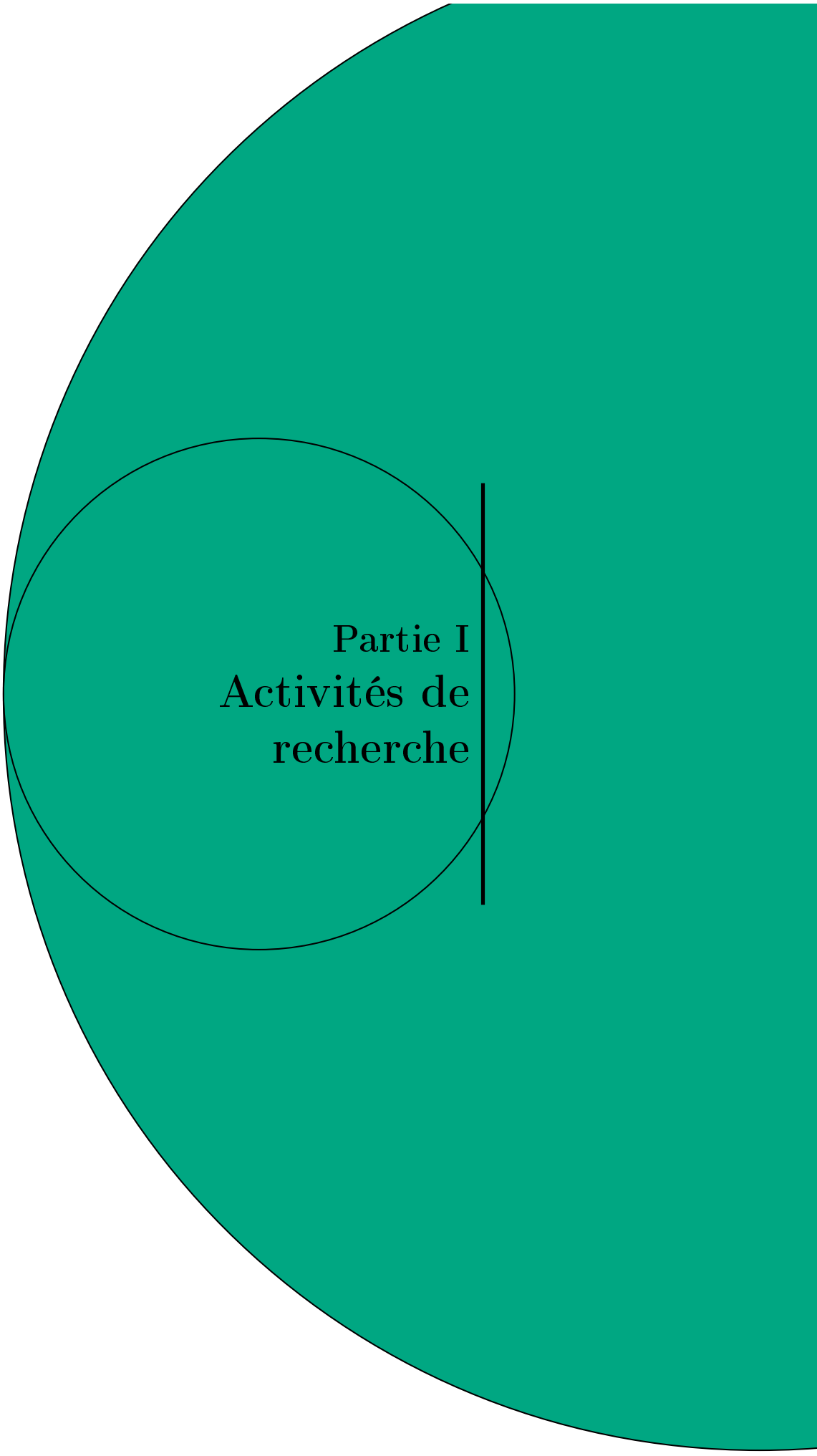


Rapport d'activité

2004-2008

Equipe A3



Partie I
Activités de recherche

Chapitre 1
A^3 (Apprentissage Artificiel et Applications)

1.1 Membres de l'équipe

Permanents

	Nom	*Prénom*	*Situation*	*Institution*
	Alphonse	Erick	MCF	IG
	Azzag	Hanane	MCF	IUTV
	Bennani	Younès	PU	IG
	Bouthinon	Dominique	MCF	IUTV
	Champesme	Marc	MCF	IG

1.1 Membres de l'équipe

	Gérard	Pierre	MCF	IUTV
	Guérif	Sébastien	MCF	IG
	Kanawati	Rushed	MCF	IUTV
	Osmani	Aomar	MCF	IUTV
	Rouveirol	Céline	PU	IG
	Soldano	Henry	MCF	IG
	Viennet	Emmanuel	MCF	IUTV

Doctorants

	Nom	*Prénom*	*Situation*	*Institution*
	Aseervatham	Sujeevan	ATER	IG
	Cabanes	Guénaël	AM	
	Dehni	Lahcène	ATER	IG
	Elati	Mohamed	ATER	IG
	Grozavu	Nistor	CDD	
	Karoui	Hager	ATER	IUTV
	Rogovschi	Nicoleta	CDD	

Membres associés

Nom	*Prénom*	*Situation*	*Institution*
Benabdeslem	Khalid	MCF	Université Lyon 1
Bossaert	Fabrice	MCF	IUT Bobigny
Krief	Francine	PU	Université Bordeaux 1
Lebbah	Mustapha	MCF	IUT Villetaneuse
Zucker	Jean-Daniel	DR	IRD Bobigny

personnels invités du 22/06/2005 au 1/9/2006

Prénom	*Nom*	*Origine*	*Financement*
2005			
Jose	Neuman de Souza	Federal University of Ceara (Brazil)	professseur invité
2006			
Jose	Neuman de Souza	Federal University of Ceara (Brazil)	professseur invité
2007			
Stefan	Kramer	TU Munich (Allemagne)	professseur invité
2008			
Attilio	Giordana	Université de Turin (Italie)	professseur invité

professeur invité : 1 mois d'invitation par l'université Paris 13

1.1.1 Evolution des membres de l'équipe

Départs

- A. El Fallah, MCF (thème SMA), promotion PU à Paris 10 (2002)
- A. Guerra-Hernandez, MCF, Univ. de La Cruz (Mexique), (Thèse soutenue en décembre 2003)
- K. Benabdeslem, MCF, Univ. Lyon 1, (Thèse soutenue en décembre 2003)
- F. Zehraoui, MCF, Univ. Evry, (Thèse soutenue en octobre 2004)
- H. Kadri-Dahmani, Post-doc, (Thèse soutenue en novembre 2005)
- F. Krief, MCF (thème diagnostic), promotion PU à Bordeaux 1 (2004)
- Ph. Dague, PU (thème diagnostic), mutation à Paris 11 (2005)
- A. Pauchet, (thème SMA), Post-doc chez France Télécom (2006)
- Z. Jrad, (thème SMA), Post-doc à l'INRIA (2006)
- H. Moungla, (thème SMA), Post-doc chez France Télécom (2006)

Arrivées

- P. Gérard, MCF IUTV (Apprentissage par renforcement), 2003 (thèse Paris 6)
- H. Karoui, doctorant, ATER (RaPC), octobre 2003
- S. Aseervatham, doctorant, AM (Apprentissage statistique), octobre 2004
- L. Dehni, doctorant, CDD ANR (Apprentissage connexionniste), octobre 2004
- E. Alphonse, MCF IG (Apprentissage symbolique), 2005 (thèse Paris 11)
- H. Azzag, MCF IUTV (Apprentissage bio-inspiré), 2006 (thèse Tours)
- C. Rouveirol, PU IG (Apprentissage symbolique), 2006 (auparavant MCF Paris 11, LRI Orsay)
- M. Elati, doctorant, AM (Apprentissage symbolique), octobre 2006
- N. Rogovschi, doctorante, CDD ANR (Apprentissage et réseaux sociaux), octobre 2006
- N. Grozavu, doctorant, CDD ANR (Apprentissage connexionniste), octobre 2006

1.1.2 Historique

L'équipe A³ (Apprentissage Artificiel & Applications) est issue d'une réorganisation des thèmes de recherche de l'équipe Adage (Apprentissage, Diagnostic et Agents) qui est devenue nécessaire après les départs respectifs d'Amal El Fallah (thème Agents) à Paris 10 et de Philippe Dague (thème Diagnostic) à Paris 11. Cette restructuration a été effectuée avec la volonté de tirer parti de la complémentarité de ses thèmes de recherche et de mettre en œuvre leur coopération afin de rendre plus visible cet axe de recherche au niveau national et international. En effet, le rayonnement du thème «Apprentissage Artificiel» du LIPN, se heurtait au manque d'une structure fédératrice et dynamique. La réorganisation des activités de recherche sur l'apprentissage artificiel à Paris-Nord (LIPN et ses proches voisins) selon une structure d'équipe de recherche fédératrice est un moyen prometteur pour renforcer la visibilité scientifique de ce thème de recherche. Cette structure a pour vocation de s'appuyer sur la qualité scientifique et la dynamique des chercheurs sur ce thème, les coopérations scientifiques sous forme de plusieurs axes de recherche inter-équipes du LIPN, des axes de recherche inter-laboratoires présents ou en cours d'élaboration. En effet, ces collaborations donnent naissance à une structure scientifique fédératrice des activités de recherche autour de l'apprentissage artificiel à Paris-Nord ayant un noyau au LIPN. Il s'agit d'un des premiers pôles sur cette thématique en Ile de France. Il faut souligner aussi que l'axe « Apprentissage Artificiel» est un des deux axes de recherche du LIPN présents dans le pôle de compétitivité Cap Digital-IMVN (Image, Multimédia et Vie Numérique).

Parallèlement à la prolongation des recherches au sein de chaque thème de l'équipe A³, plusieurs axes de recherche inter-équipes sont présents. En effet, la problématique de l'apprentissage est déjà largement partagée avec l'équipe RCLN, en particulier dans le cadre du projet InfoMagic (Pôle de compétivité Cap Digital), tout comme la classification hybride combinant des réseaux connexionnistes à un système de raisonnement à partir de cas. La résolution de problèmes de satisfaction de contraintes, utile notamment en raisonnement spatio-temporel, donne lieu à une coopération avec l'équipe OCAD sur le problème MAX-CSP. Enfin, des chercheurs en apprentissage symbolique de l'équipe participent à l'EPML IAPuces dirigée par Jean-Daniel Zucker, du Lim&Bio Bobigny de Paris-Nord. Soulignons enfin qu'une caractéristique essentielle du mode de fonctionnement de cette équipe est que les recherches y sont alimentées et évaluées grâce à de nombreuses applications, concernant la génomique, la parole, l'image, les télécommunications, Internet, le e-commerce et le e-learning, etc, donnant lieu pour la plupart à des contrats industriels.

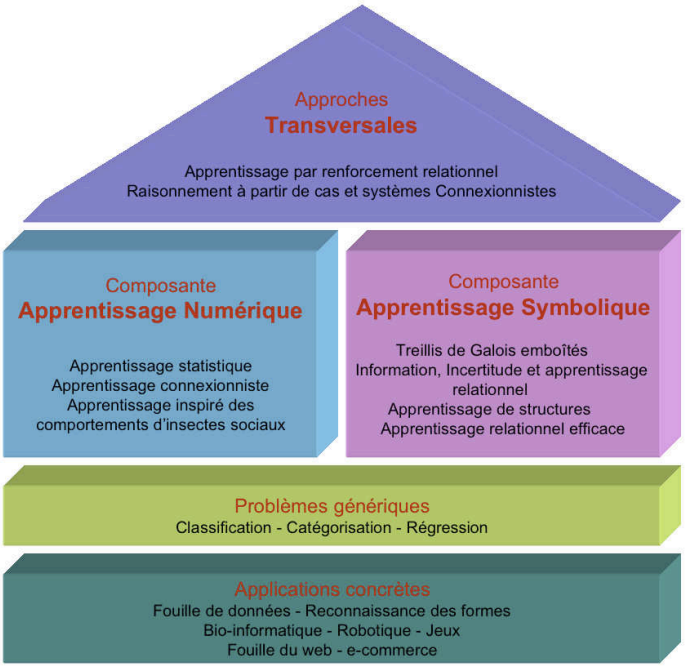
1.2 Fondements scientifiques et orientations de recherche

L'apprentissage artificiel est une discipline scientifique qui recouvre plusieurs aspects d'études mathématiques, statistiques et algorithmiques. Ses applications sont très stratégiques, notamment pour le traitement de données scientifiques, pour l'ingénierie, et pour les systèmes d'information. Les techniques d'apprentissage artificiel permettent de capturer des relations entre des événements ou des signaux caractéristiques d'un phénomène connu par des exemples (observations). Ces méthodes sont utilisées le plus souvent pour des problèmes pour lesquels on est capable de réaliser un grand nombre de mesures et dont on ne connaît pas la loi sous-jacente. Cela couvre une large gamme de problèmes pouvant aller de la modélisation de phénomènes naturels à l'identification et la commande de processus industriels en passant par la reconnaissance des formes et l'exploration des données. Les recherches dans ce domaine sont extrêmement actives depuis les années 80. Des laboratoires tant industriels qu'universitaires participent à cette recherche. Différentes voies ont déjà été explorées et les systèmes d'apprentissage sont devenus des outils incorporés dans différents logiciels commercialisés pour les statistiques, la fouille de données (Data Mining), le traitement du signal et l'identification de systèmes.

Les activités de recherche au sein de l'équipe A³ sont structurées autour de deux grandes composantes :

- Composante Apprentissage numérique,
- Composante Apprentissage symbolique,

Afin de tirer parti de la complémentarité de ses différents thèmes de recherche et de mettre en oeuvre leur coopération des approches transversales (symbolico-numériques) sont aussi explorées. Ces recherches sont alimentées, coordonnées et évaluées grâce à diverses applications dans le domaine de la reconnaissance des formes et de la fouille de données.



Orientations scientifiques de l'équipe A³

L'équipe A³ développe des recherches fondamentales et intensifie une politique de coopération avec les grands organismes et les industriels, politique qui se traduit par un accroissement de sa participation dans les actions ANR, RNRT et RNTL, les contrats avec les différents ministères, et une collaboration avec des entreprises dans le cadre soit de projets européens, soit de bourses CIFRE.

1.3 Composante Apprentissage Numérique

Participants : S. Aseervatham, H. Azzag, Y. Bennani, L. Dehni, N. Grozavu, S. Guérif, N. Rogovschi, E. Viennet

L'apprentissage numérique consiste à découvrir un concept sous une forme géométrique simple (hyperplan). Il se présente sous la forme d'optimisation d'un critère (fonction de coût) interprétable sur l'ensemble d'apprentissage. Nos activités de recherche pour ces méthodes sont centrées sur les apports des techniques d'apprentissage numériques (statistiques et bio-inspirées) à l'analyse, l'exploration et l'interprétation de données complexes et de taille importante. Notre travail de recherche s'oriente plus particulièrement vers l'extraction de connaissances à partir de données et l'aide à la décision. Nos objectifs et activités de recherche s'articulent autour de deux dimensions complémentaires : théorique et pratique. Les études théoriques portent sur la conception de nouveaux algorithmes d'apprentissage et le développement d'approches modulaires et hybrides faisant coopérer différentes techniques. La dimension pratique s'attache à valoriser les activités de recherche en intégrant les résultats dans des systèmes opérationnels. Cette intégration se fait dans le cadre d'applications réelles dans différents domaines : parole, image, biométrie, diagnostic, Data Mining, Web Mining, Marketing. Notre problème central est donc celui de l'analyse et l'exploration de données afin d'en comprendre le sens, de déceler des relations entre des événements, d'en déduire des modèles de comportement. Il s'agit de la transformation de données en connaissances qui est et sera de plus en plus touchée par la nécessité de développer, de maîtriser et de choisir de nouveaux outils d'analyse et d'exploration des données. Le principal objectif de ces outils est celui d'un processus automatisé qui prend ses sources dans les données et dont le but est l'extraction de connaissances pour l'aide à la décision.

Actuellement, nous développons les axes de l'apprentissage non-supervisé, l'apprentissage semi-supervisé et la visualisation de données structurées en séquences (temporelle, spatio-temporelles et évolutives).

Nous poursuivons aussi nos travaux sur la combinaison de modèles et la fusion de données et de décisions pour des problèmes de classement, de classification et de régression. Nous continuons également à travailler sur les techniques de sélection, d'extraction et de pondération de caractéristiques en apprentissage non-supervisé. Il nous paraît important de poursuivre l'étude en la généralisant à la modélisation de données dans de grandes bases de données issues des entrepôts de données. Nous envisageons de développer des activités sur l'exploration de données textuelles. Les objectifs visés concernent l'accès à l'information textuelle (recherche d'information, navigation dans les documents, filtrage et routage de documents), la gestion de données textuelles (structuration de documents et indexation, classification de textes), et l'analyse de contenu (extraction d'information, acquisition de terminologie spécialisée).

Ces recherches sont validées sur des problèmes de reconnaissance des formes et de fouille de données, en particulier sur des données temporelles, spatio-temporelles et évolutives (traces de navigations sur Internet, données comportementales, données de consommation, données de marketing, réseaux de capteurs, ...).

1.3.1 Approches statistiques

Apprentissage à base de noyaux pour le traitement de données textuelles

Cet axe est nouveau dans l'équipe, et concrétise la volonté affichée depuis plusieurs années par le laboratoire de développer des axes de recherche inter-équipes. Nous travaillons sur ce thème en relation avec plusieurs chercheurs de l'équipe RCLN, en

particulier dans le cadre du projet InfoM@gic et du projet Alvis. Une thèse est en cours sur les méthodes à noyaux appliquées au traitement de données structurées et semi-structurées.

Apprentissage semi-supervisé

Dans de nombreuses applications des techniques d'apprentissage, on dispose de grands volumes de données non étiquetées, mais seulement d'un petit échantillon de données étiquetées. Les algorithmes d'apprentissage séparent traditionnellement le traitement des données étiquetées (apprentissage *supervisé*) de celui des données non étiquetées (*non supervisé*). L'équipe A³ travaille depuis plusieurs années sur l'apprentissage non supervisé, en particulier via les cartes topologiques de Kohonen, et sur les classifieurs supervisés, en particulier les machines à vecteurs de support (SVM). Nous étudions actuellement comment aborder le cas semi-supervisé : modèles transductifs proposés par Vapnik, extraction de traits non supervisés, techniques de co-apprentissage. Nous espérons déboucher sur des approches applicables à des problèmes aussi différents que la fouille de textes ou la correction du biais lors de l'évaluation du risque crédit (traitement des refusés par les établissements bancaires).

Réseaux Sociaux et apprentissage

L'étude des réseaux sociaux est une discipline récente. Longtemps confiné aux mathématiciens (théorie des graphes aléatoires) et aux sociologues, ce domaine a connu un net regain d'intérêt à la fin des années 90, lorsqu'on a remarqué que la plupart des réseaux observés dans la nature ne se comportaient pas comme des graphes aléatoires simples, mais partageaient un certain nombre de propriétés fondamentales. Parmi celles-ci, citons l'invariance de la structure par changement d'échelle, qui se traduit par une distribution des degrés suivant une loi de puissance. L'étude de ces réseaux a bien sûr été fortement stimulée par l'essor du Web, qui offre de nombreuses occasions de créer des réseaux (entre pages Web, entre utilisateurs...) dont l'analyse est fort utile (recherche d information, animation de sites Web, systèmes de recommandations...) mais pose des problèmes techniques mal résolus.

Dans la cadre du projet ANR RNTL Septia (*Solution Automatique pour le Traitement des Interactions Clients*), nous travaillons sur l'apport des techniques d'analyse des réseaux sociaux pour l'apprentissage, dans le contexte de l'élaboration de systèmes de classification de messages et d'aide à la réponse (pour la gestion de centres d'appel). Nous essayons par exemple de comprendre comment définir la notion de communauté d'utilisateurs afin de maximiser les performances d'un système de classification d'utilisateurs (ou des messages émis par ces derniers). Cette problématique a des liens évidents avec les méthodes de clustering non supervisé étudiées depuis plusieurs années dans l'équipe A³, et donne lieu au travail de doctorat de Mlle N. Rogovschi.

1.3.2 Approches connexionnistes

Les réseaux de neurones artificiels (ou réseaux connexionnistes) sont des systèmes numériques permettant la modélisation de processus généraux par l'établissement de modèles fonctionnels [LI-3]. Ceux-ci sont identifiés à partir des entrées-sorties du processus par des algorithmes dits "d'apprentissage" qui s'apparentent à des techniques d'estimation statistiques. Nés en informatique dans le domaine de l'intelligence artificielle, ils ont connu depuis le début des années 80 un développement intensif dû au succès rencontré dans une très large gamme d'applications. Les réseaux connexionnistes offrent

une panoplie de techniques adaptatives pour de nombreux problèmes génériques : la classification, le classement, la modélisation, la prévision. Parmi les domaines où ils ont été utilisés, on peut citer le traitement du signal, de l'image, de la parole, le contrôle, la prévision, l'optimisation combinatoire, le diagnostic, la fouille de données, ... Il s'agit donc d'un outil général pouvant être utilisé dans des cadres différents avec une grande souplesse. Les modèles connexionnistes apparaissent comme une alternative ou plus souvent comme un complément aux techniques classiques. Les recherches dans le domaine sont extrêmement actives actuellement. Plusieurs laboratoires tant industriels qu'universitaires participent à cette recherche en Europe, aux États-Unis et au Japon. Différentes voies ont déjà été explorées et les réseaux connexionnistes sont devenus un outil incorporé dans différents logiciels commercialisés pour les statistiques, la fouille de données, le traitement du signal et l'identification de systèmes.

Classification connexionniste non supervisée et visualisation de données évolutives

Par opposition aux données temporelles, le recueil des données évolutives ne se fait pas avec une périodicité de la mesure : des mesures peuvent avoir été recueillies à différents moments. Cela prive les données de leur caractère temporel au sens statistique. L'évolution de ces données peut être vue sous forme de trajectoires temporelles (de taille variable et parfois multivariées). La visualisation de ces trajectoires dans l'espace de l'évolution devient difficile à interpréter dès que le nombre de trajectoires est élevé. L'un des objectifs principaux de l'analyse de ce type de données est d'obtenir une classification (ou typologie) évolutive et incrémentale de trajectoires en fonction de leurs positions et leurs évolutions. Cette typologie permettra de diminuer le nombre de ces trajectoires, en assimilant les trajectoires de la même typologie à la trajectoire prototype de la typologie.

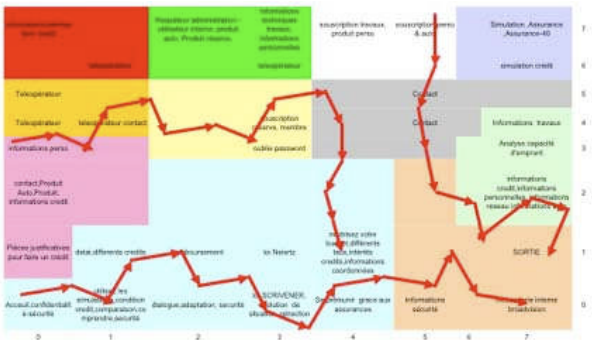


FIG. 1.1: Cartographie comportementale d'un site web et visualisation des trajectoires des Internautes

Nous avons proposé plusieurs approches d'apprentissage non supervisé pour l'exploration de ce type de données [CI-17, RI-4, CI-3]. La validation de ces approches a été effectuée en particulier sur les traces de navigation dans les sites web. La figure 1.1 montre un exemple de cartographie comportementale d'un site web avec la visualisation des navigations sous forme de trajectoires.

Nous travaillons sur d'autres approches à base d'algorithme d'apprentissage non-supervisé pour traiter ce type de données en quatre phases :

– *modélisation globale : recherche d'un modèle unique de représentation qui résume d'une manière globale l'ensemble des données (compromis).*

- *étude de l'interstructure : identification des proximités ou des différences entre les ensembles de données, sans description.*
- *étude de l'intrastructure : analyse fine des ressemblances ou des différences permettant d'expliquer les structures identifiées.*
- *étude de l'évolution des trajectoires : analyse et description de l'évolution du phénomène d'une manière globale ou étude des évolutions de chacun des éléments.*

Réduction des dimensions de l'espace de description

La taille des données peut être mesurée selon deux dimensions, le nombre de variables et le nombre d'exemples. Ces deux dimensions peuvent prendre des valeurs très élevées, ce qui peut poser un problème lors de l'exploration et l'analyse de ces données. Pour cela, il est fondamental de mettre en place des outils de traitement de données permettant une meilleure compréhension de la valeur des connaissances disponibles dans ces données. La réduction des dimensions est l'une des plus vieilles approches permettant d'apporter des éléments de réponse à ce problème. Son objectif est de sélectionner ou d'extraire un sous-ensemble optimal de caractéristiques pertinentes pour un critère fixé auparavant. La sélection de ce sous-ensemble de caractéristiques permet d'éliminer les informations non-pertinentes et redondantes selon le critère utilisé. Cette sélection/extraction permet donc de réduire la dimension de l'espace des exemples et rendre l'ensemble des données plus représentatif du problème. En effet, les principaux objectifs de la réduction de la dimension sont :

- *faciliter la visualisation et la compréhension des données,*
- *réduire l'espace de stockage nécessaire,*
- *réduire le temps d'apprentissage et d'utilisation,*
- *identifier les facteurs pertinents.*

Les algorithmes d'apprentissage artificiel requièrent typiquement peu de traits (features) ou de variables (attributs) très significatifs caractérisant le processus étudié. Dans le domaine de la reconnaissance des formes et de la fouille de données, il pourrait encore être bénéfique d'incorporer un module de réduction de la dimension dans le système global avec comme objectif d'enlever toute information inconséquente et redondante. Cela a un effet important sur la performance du système. En effet le nombre de caractéristiques utilisées est directement lié à l'erreur finale. L'importance de chaque caractéristique dépend de la taille de la base d'apprentissage (pour un échantillon de petite taille, l'élimination d'une caractéristique importante peut diminuer l'erreur). Il faut aussi noter que des caractéristiques individuellement peu pertinentes peuvent être très informatives si on les utilise conjointement.

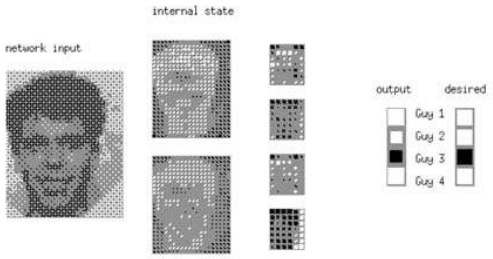


FIG. 1.2: Architecture d'un système connexionniste pour la reconnaissance faciale avant l'optimisation

La réduction de la dimension est un problème complexe qui permet de réduire le volume d'informations à traiter et faciliter le processus de l'apprentissage. Nous

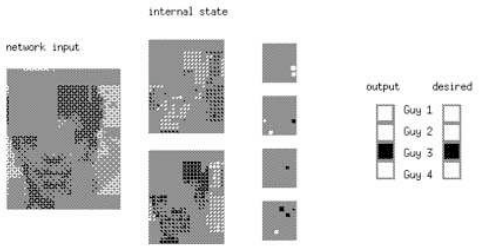


FIG. 1.3: Architecture d'un système connexionniste pour la reconnaissance faciale après optimisation par la mesure de pertinence HVS proposée par Y. Bennani et M. Yacoub en 2001. Les zones grises représentent les neurones supprimés.

pouvons classer toutes les techniques mathématiques de réduction des dimensions en deux grandes catégories :

- *la sélection de variables : qui consiste à choisir des caractéristiques dans l'espace de mesure,*
- *l'extraction de traits : qui vise à sélectionner des caractéristiques dans un espace transformé (dans un espace de projection)*

La première catégorie est appropriée quand l'acquisition de mesures des formes est coûteuse. Ainsi l'objectif principal de la sélection de caractéristiques dans ce cas est de réduire le nombre de mesures requises. Par contre, les techniques d'extraction de traits (deuxième catégorie) utilisent toute l'information contenue dans les formes pour la compresser et produire un vecteur de plus petite dimension. Ces techniques projettent un vecteur forme de l'espace de représentation dans un espace de dimension plus petite. Les systèmes d'apprentissage connexionniste sont un bon exemple de cette catégorie. En effet, les modèles connexionnistes connus pour une tâche de discrimination fournissent un système avec des aptitudes intéressantes pour l'analyse du processus. Les cellules cachées d'un Perceptron multi-couches apprennent comment extraire les caractéristiques significatives du signal d'entrée. Nous avons proposé dans cette direction plusieurs mesures de pertinence, en apprentissage supervisé et non supervisé [CO-21, CI-44].

Combinaison de modèles et fusion de décisions

Un de nos thèmes de recherche majeur est la conception et la mise en œuvre de systèmes qui font coopérer différents modèles entre eux et avec des techniques plus classiques dans les différents domaines d'application. Ces systèmes constituent une autre génération des modèles connexionnistes et apparaissent comme la seule solution pour la réalisation de tâches complexes et de grande taille. Les études théoriques portent sur les algorithmes d'apprentissage pour ce type de système, la décomposition de tâches, la conception de chaînes de traitement optimales. Nous travaillons sur la possibilité de coopération des algorithmes connexionnistes et d'autres méthodes plus classiques comme les Chaînes de Markov Cachées, afin de concevoir des systèmes modulaires hybrides plus robustes.

Classification topographique évolutive et stabilité

L'apprentissage non-supervisé évolutif est un problème assez récent et représente une voie de recherche très prometteuse ouvrant d'excellentes perspectives. En effet, l'apprentissage évolutif fait partie des sujets difficiles en apprentissage artificiel, essentiellement parce qu'il n'existe pas de théorie de l'apprentissage statistique réellement dévolue dans le cas où les exemples sont fournis en ligne et les concepts

de classes varient dans le temps. Ces systèmes permettront de découvrir un espace topologique (évolutif) d'un ensemble de données. Cet espace topographique préserve la notion de voisinage entre les exemples décrivant le processus à modéliser. Nous avons étudié la possibilité de rendre évolutive la découverte de cet espace topographique : i.e. le nombre de regroupement homogène (clusters) change dans le temps en fonction des informations complémentaires qui arrivent, la classification se poursuit dans le temps. Pour cet axe de recherche, nous travaillons aussi sur les critères d'évaluation et de stabilité [CI-60]. En effet, les algorithmes de classification sont aujourd'hui très nombreux, mais on n'a presque jamais de raison « théorique » de choisir l'un plutôt que l'autre. La stabilité est une propriété très souhaitable pour ce type d'algorithmes Ces recherches sont validées en particulier sur des données temporelles, spatio-temporelles, évolutives (traces de navigations sur Internet, données comportementales, données RFID, données de consommation, données de marketing, réseaux de capteurs, ...).

1.3.3 Approches inspirées des comportements d'insectes sociaux

Notre travail dans cette direction est consacré à l'étude de solutions originales de classification non supervisée. Ces solutions s'inspirent des propriétés du vivant pour proposer des constructions originales, dotées de capacités étonnantes et aptes à résoudre des problèmes difficiles vis-à-vis desquels les approches plus traditionnelles rencontrent des difficultés. Le principe des méthodes que nous développons se base sur des études menées en éthologie et consiste ainsi à extraire de ces études un modèle réalisable du point de vue informatique.

Ce nouveau domaine de recherche a pour objet de transformer la connaissance que les éthologues ont des capacités collectives de résolution de problèmes des insectes sociaux en techniques artificielles de résolution de problèmes en informatique. Comment un ensemble d'entités individuellement simples au fonctionnement décentralisé est-il capable collectivement de fournir une réponse complexe ? Ce principe de plusieurs entités simples en interaction qui émerge à un niveau collectif en une structure complexe (qualifiée d'intelligente) est présent dans différentes disciplines : biologie, physique, informatique, etc. Dans ces domaines, différents modèles ont vu le jour pour décrire et analyser de tels systèmes.

C'est dans ce vaste domaine de l'informatique biomimétique dans lequel nous tentons de contribuer en apportant de nouveaux modèles bio-inspirés pour résoudre des problèmes de classification non supervisée et hiérarchique sur des grandes masses de données. Les résultats obtenus par des simulations permettent ensuite d'évaluer la qualité du modèle relativement aux études biologiques. Les conséquences peuvent être utiles aux informaticiens afin d'améliorer le modèle lui même ou aux éthologues qui peuvent développer et tester leurs théories du comportement animal.

Simulation et visualisation 3D de la dynamique des blattes

Dans le cadre du projet *ANR Sillages* nous nous sommes intéressés à un processus d'auto-organisation qui est l'agrégation, et sa capacité à générer des organisations spatio-temporelles chez les espèces grégaires, telles que la blatte américaine *Periplaneta americana*. Cette dernière sera utilisée par les éthologues comme modèle expérimental.

Afin de faciliter le travail des éthologues, nous avons développé un outil qui simule le comportement des blattes *Periplaneta americana*. Notre but est que les chercheurs

puissent mieux cerner la dynamique de ces insectes et découvrent de nouvelles hypothèses les concernant.
Comme nous le montrerons dans la section suivante, ce simulateur nous a permis de découvrir que le comportement collectif des blattes peut construire de nouvelles méthodes de classification des données.
Notre simulateur modélise le dispositif expérimental présenté dans la figure 1.4.
Ce modèle utilise une chaîne de *Markov* qui simule le déplacement de chaque blatte suivant sa position courante : vers les abris ou vers l'environnement extérieur. Tous ces déplacements vont dépendre de probabilités liées à la qualité de l'abri, sa taille ou encore au nombre de blattes présentes sous l'abri à un instant t.

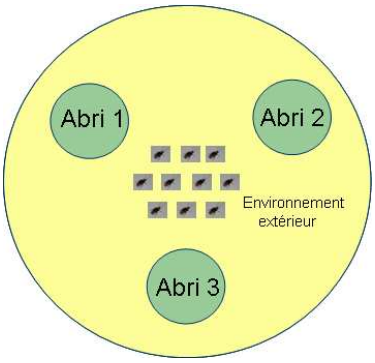


FIG. 1.4: Dispositif expérimental pour tester les blattes réel

Nous avons également greffer à notre simulateur un outil de visualisation 3D permettant de voir les déplacements de blattes de manière plus réaliste (voir figure 1.5).

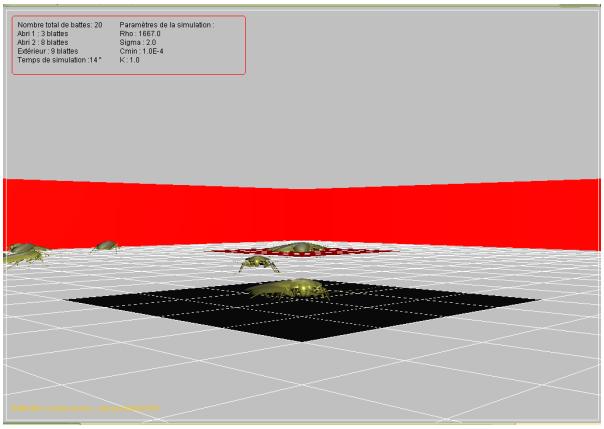


FIG. 1.5: Visualisation 3D de la dynamique du comportement des blattes

Classification non supervisée inspirée des comportements d'insectes sociaux

Cet axe a pour objectif de transformer la connaissance que les éthologues ont des capacités collectives de résolution de problèmes chez les blattes américaines, en techniques artificielles de Clustering. Nous avons ainsi imaginé un système dans lequel il s'agirait de modéliser la manière dont les blattes forment des agrégats

sous chaque abri, et d'utiliser ce comportement pour partitionner des données en différentes classes. Chaque blatte va ainsi représenter une donnée à classer et les abris les classes initiales.

Ces blattes (artificielles) vont ensuite construire de manière similaire un partitionnement en appliquant certaines règles comportementales. Les déplacements des blattes vers les différentes classes dépendent de la similarité entre les données.

Les résultats que nous avons obtenus sont très intéressants, nous avons pu les tester sur plusieurs types de données car notre approche ne dépend pas du type de données à classer : il suffit uniquement de définir une fonction de similarité pour ces nouvelles données.

Il reste beaucoup de problèmes à résoudre pour ce type d'approche et les résultats obtenus jusqu'ici sont plus qu'encourageants [CI-65]. Ce nouveau modèle se placera ainsi dans la lignée des travaux précédents sur les algorithmes de classification s'inspirant des insectes sociaux et plus généralement des systèmes biologiques.

Génération de graphes par des fourmis artificielles

En collaboration avec le laboratoire d'informatique de l'Université de Tours, nous avons développé un algorithme incrémental nommé AntGraph pour la construction de graphes de voisinage. Il s'inspire du comportement d'auto-assemblage observé chez une population de fourmis réelles où ces dernières se fixent progressivement à un support fixe puis successivement aux fourmis déjà fixées afin de créer une structure vivante.

Nous utilisons ainsi une approche à base de fourmis artificielles où chaque fourmi représente une donnée. Ce comportement a été précédemment utilisé pour construire de manière incrémentale un graphe à partir d'une mesure de similarité entre les données. Il s'agit en effet d'une généralisation de l'algorithme AntTree proposé par Hanane Azzag dans sa thèse.

Par cette méthode, le graphe obtenu est plus représentatif de la similarité existante entre les données avec en plus une complexité inférieure à celle observée dans les méthodes classiques (Voisins Relatifs, graphes de Gabriel, triangulation de Delaunay). Les résultats obtenus sont meilleurs, notamment en terme de temps de calcul [CI-57].

1.3.4 Domaines d'applications

Nous travaillons sur différents domaines d'applications, toutefois nous avons développé ces dernières années un thème principal qui est l'analyse et la visualisation de données structurées en séquences :

- *Exploration des traces de navigations sur Internet ;*
- *Analyse et visualisation des données comportementales : Marketing, RFID ;*
- *Diagnostic des systèmes complexes ;*
- *Identification multi-modale : vocale et visuelle.*

1.3.5 Résultats significatifs

Parmi les nombreux résultas de nos activités de recherche, nous pouvons mettre en avant les cinq contributions suivantes :

- *Mesures de pertinence pour la sélection de caractéristiques et l'optimisation des architectures connexionnistes ;*

- *Algorithmes de pondération des variables, durant l'apprentissage, en fonction de leur pouvoir discriminant ;*
- *Classification incrémentale de données évolutives, séquences ;*
- *Algorithmes de classification topographique à deux niveaux simultanés ;*
- *Sélection/pondération de variables en apprentissage non-supervisé à deux niveaux.*

1.3.6 Perspectives de recherche

Nous continuons actuellement nos travaux sur la combinaison de modèles et la fusion de données et de décisions pour des problèmes de classement, de classification et de régression. Nous continuons également à travailler sur les techniques de sélection et d'extraction de caractéristiques dans les réseaux connexionnistes. Il nous paraît important de poursuivre l'étude en la généralisant à la modélisation de données dans de grandes bases de données issues du web ou des entrepôts de données. Nous envisageons aussi de développer des activités sur l'exploration de données textuelles. Les objectifs visés concernent l'accès à l'information textuelle (recherche d'information, navigation dans les documents, filtrage et routage de documents), la gestion de données textuelles (structuration de documents et indexation, classification de textes), et l'analyse de contenu (extraction d'information, acquisition de terminologie spécialisée). La problématique de la classification topographique évolutive sera aussi abordée dans le cadre de la classification à base de noyaux, pour laquelle nous définirons un critère d'évaluation prenant en compte son double objectif (évolution et topographie). Nous proposerons aussi dans un contexte d'apprentissage non supervisé, des noyaux pour traiter des données structurées (séquences, arbres, graphes, ...). Enfin, nous souhaitons apporter quelques contributions permettant de faire naître un cadre théorique pour l'apprentissage non-supervisé évolutif.

1.4 Composante Apprentissage Symbolique

Participants : E. Alphonse, D. Bouthinon, M. Champesme, M. Elati, A. Osmani, C. Rouveirol, H. Soldano.

Cet axe a pour but d'appréhender des quantités de données importantes (cf. fouille de données) décrites de manière symbolique, i.e., telle que les exemples et les hypothèses sont représentés par des formules logiques. Dans ce cas, les exemples ainsi que les hypothèses candidates pour le concept cible sont organisés dans des espaces fortement structurés. Les techniques d'apprentissage symbolique, par opposition aux techniques d'apprentissage numérique, vont explorer l'espace des hypothèses en tirant parti de la structure de l'espace de recherche imposée par la relation structurant l'espace, afin d'identifier une hypothèse satisfaisant des critères de couverture de l'ensemble d'apprentissage.

L'équipe couvre un large spectre de techniques d'apprentissage (apprentissage supervisé, apprentissage descriptif), et conçoit des systèmes d'apprentissage pour des langages de concept allant de la logique attribut-valeur à des restrictions de la logique des prédicats comme Datalog non récursif avec négation. Orthogonalement, des travaux sont menés sur un nouveau cadre d'apprentissage permettant de formaliser l'apprentissage en présence d'exemples incomplets ou incertains.

Nous avons poursuivi des travaux dans le domaine de l'apprentissage attribut-valeur, sur le thème des treillis de Galois emboîtés et développé et étudié les performances d'un système d'apprentissage multi-agent. L'équipe est également active dans le domaine de l'apprentissage de structures (séquences, graphes) et en Programmation Logique Inductive, domaine dédié à l'Apprentissage Symbolique dans des langages de concepts restrictions de la logique du premier ordre. Ces langages de concept complexes sont requis dans le cadre d'applications en fouille de données issues du Web (link mining, graph mining), de développements en Bases de Données Inductives ou d'applications dans les domaines du Traitement Automatique de la Langue, de la Bio-chimie ou de la Bioinformatique.

Dans ces langages, les mécanismes associés à l'apprentissage sont coûteux (voire indécidables), la recherche heuristique joue donc un rôle primordial. Nous avons développé une nouvelle stratégie d'apprentissage relationnel efficace, fondées sur un changement de représentation, appelé propositionnalisation, qui projette un problème d'apprentissage en un problème attribut-valeur. Enfin, nous étudions les phénomènes de transition de phase liés à l'apprentissage relationnel et leurs effets sur l'efficacité de la recherche d'hypothèses, comme l'apparition d'importants phénomènes de plateaux au cours de la recherche d'une hypothèse candidate, rendant "aveugle" le système d'apprentissage.

1.4.1 Les axes de recherche

Treillis de Galois emboîtés - constructions de taxonomies

Notre travail utilise essentiellement des structures latticielles basées sur des correspondances de Galois définies entre les objets observés et leur description. L'inconvénient principal de ces structures est leur grande taille (en nombre de nœuds) qui en rend très difficile l'exploration directe par l'utilisateur, nous avons défini une notion de «projection» sur l'ensemble des objets que nous utilisons pour faire varier la granularité de cette structure en offrant à l'utilisateur la possibilité d'explorer les données selon un point de vue d'abord grossier (pour appréhender globalement l'ensemble des données) puis de plus en plus détaillé à mesure que l'intérêt de l'utilisateur se focalise.

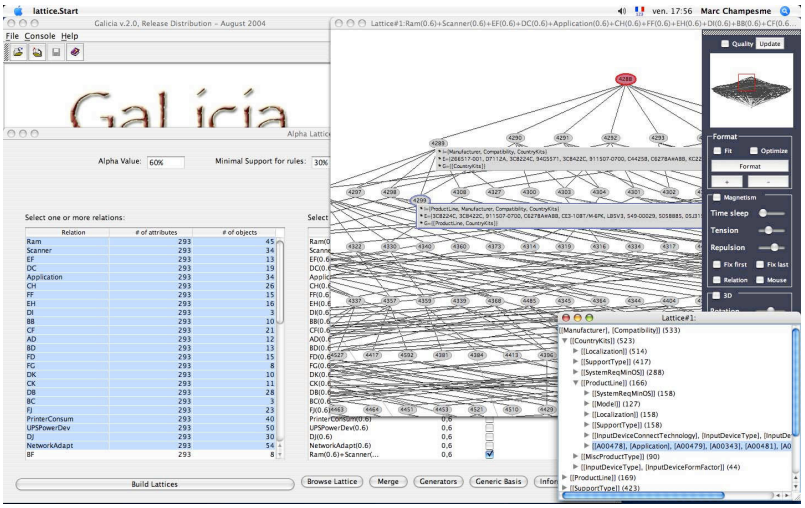


FIG. 1.6: La plate-forme Galicia étendue pour la prise en compte de la projection alpha (ici, treillis construit sur les données C/NET pour $\alpha = 60\%$)

La notion de projection "alpha" [RE-3] (cas particulier de projection que nous avons définie) permet en outre d'utiliser, quand elle existe, une partition des objets en classes, l'exploration des données étant alors guidée par cette partition. L'utilisation de telles partitions nous permet alors d'ignorer des propriétés marginales des objets, de manière analogue aux théories avec défauts exceptions, mais en restant dans le cadre théorique simple des correspondances de Galois "ordinaires" (i.e. domaine de l'analyse formelle de concept). Notre travail s'intéresse par ailleurs à l'extraction des règles induites par cette structuration.

Apprentissage dans un système multi-agents cognitif

Dans le cadre de la thèse d'A. Guerra Hernandez, nous nous sommes intéressés avec A. Elfallah-Segrouchni, à l'apprentissage symbolique multi-agents. Le SMA considéré est composé d'agents communicants de type BDI (Beliefs, Desires, Intentions). Au niveau collectif, un groupe d'agents BDI améliore son apprentissage de contextes de plans en échangeant des croyances avec d'autres agents [CO-13]. Nous nous intéressons actuellement à l'apprentissage incrémental collectif dans les Systèmes Multi-Agents dans le cas où les agents sont complètement autonomes.

Analyse de l'information mise en jeu au cours de l'apprentissage

Ce thème s'articule autour de 2 axes. D'une part, nous avons proposé une formalisation en logique du premier ordre de différents cadres d'apprentissage dont les données recèlent des formes variées d'incomplétude ou d'ambiguïté. En particulier, nous présentons un nouveau cadre d'apprentissage qui rend compte de situations jusqu'ici inexprimables du fait de la nature de l'incomplétude (incertitude sur les exemples positifs et négatifs + abstraction). Ce cadre permet de plus de représenter tout problème d'apprentissage multi-instances introduit par Dietterich ou d'apprentissage par satisfiabilité introduit par de Raedt.

D'autre part, nous menons une analyse de l'information mise en jeu au cours de l'apprentissage, qui vise à identifier et quantifier l'information fournie à l'apprenant. Le modèle d'information que nous proposons, fondé sur la théorie de l'information, permet de mesurer et comparer l'information apportée par différents types d'ex-

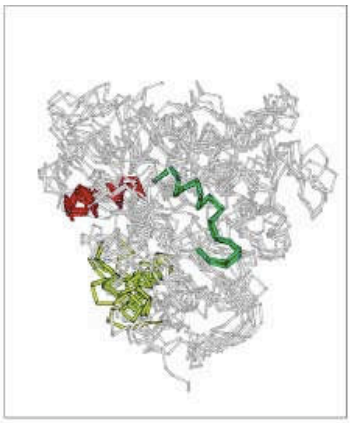
emples incomplets ou ambigus. En particulier, nous mettons en évidence un fort déséquilibre informationnel entre les exemples positifs et négatifs en apprentissage multi-instances. Un approfondissement de cet axe nous a permis d'établir des définitions et propriétés relatives à la corrélation existant entre éléments informationnels. Nous démontrons qu'un ensemble d'éléments informationnels apporte plus d'information que la somme des contributions de chacun d'eux. Enfin nous avons exhibé un critère informationnel permettant de guider des changements dynamiques de représentations au sein d'un espace de représentations possibles.

Apprentissage de structures

Nous nous intéressons à deux types de structures qui apparaissent dans un grand nombre d'applications, notamment en Bioinformatique : les séquences et les graphes. En ce qui concerne la recherche de structures séquentielles, nous avons travaillé sur la détection de *motifs relationnels répétés* dans les séquences, c'est-à-dire de motifs décrivant non seulement leurs composants mais aussi les relations qui les lient. L'algorithme KMRRelat et les conditions de son application à la recherche de motifs structuraux 3D sur les protéines ont fait l'objet de la thèse de N. El-Zant. Des recherches récentes nous ont conduits, avec N. Pisanti, en post-doc CNRS au LIPN, et M. Carpentier (thèse en sciences de la Vie à Paris VI, soutenue en décembre 2005) à une amélioration de cet algorithme permettant d'éviter la génération de (parfois très nombreux) candidats inutiles lors de la recherche de répétitions et aux représentations de ces motifs en vue de leur recherche dans les séquences de structures de protéines [CL-7]. Nous nous intéressons actuellement à la sélection et l'ordonnancement de ces motifs (de différentes longueurs) en vue d'obtenir des signatures structurales pour différentes familles de protéines.

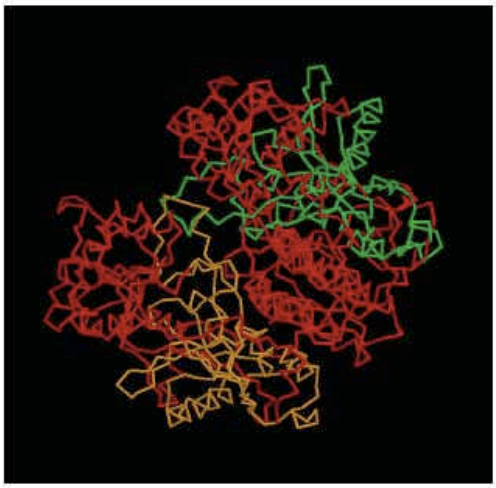
D'autre part, dans le cadre de la thèse de M. Elati, nous abordons l'apprentissage de structures de graphes sous-jacentes à des données propositionnelles. Ce travail est appliqué à l'extraction de réseaux de régulation à partir de données d'expression de gènes, en collaboration avec l'Institut Curie et l'Institut de Génétique Moléculaire de l'Université Paris 11. La détection de relations de régulation repose la plupart du temps sur l'hypothèse qu'un régulateur et son gène cible sont corrélés. Construire un graphe acyclique où chaque lien est une relation orientée de régulation et maximisant une mesure de corrélation donnée est un problème NP-complet, et a donné lieu aux développements de multiples approches heuristiques, dont bayésiennes. Cependant, ces méthodes sont peu adaptées à la détection de régulation coopératives, où plusieurs régulateurs se coordonnent pour affecter un gène cible et pour lesquelles l'expression d'un gène et d'un de ses régulateurs n'est pas nécessairement fortement corrélée. Nous avons développé une approche heuristique originale qui recherche pour chaque gène un réseau "local" représentant l'ensemble de ses régulateurs (activateurs et/ou inhibiteurs). La recherche du meilleur ensemble de régulateurs pour chaque gène cible est restreinte à l'ensemble des régulateurs fréquemment co-exprimés. Cette approche limite drastiquement le nombre d'ensemble de régulateurs candidats à explorer, tout en préservant les régulations coopératives. Elle a été appliquée avec succès à des données d'expression de levure et de patients atteints de tumeurs de vessie. Nous collaborons actuellement avec des statisticiens de l'Université d'Evry pour construire, à partir des réseaux locaux ainsi obtenus, un réseau global reliant les gènes et/ou les régulateurs entre eux, et qui s'appuie sur une information plus complexe que sur la détection de corrélations d'expression directes.

Programmation Logique Inductive



Motifs relationnels "structuraux" de longueur 18 pour les 5 cytochromes P450 4CP4 , 1ROM, 1CPT, 2HPD chain B et 3CPP. Les relations entre éléments (Cα des acides aminés) sont ici leurs distances internes.

Analyse intégrée d'une duplication directe proche interne (DDPI) dans la protéine 1a9x (carbamyl phosphate synthetase) et son gène. La duplication apparaît en vert et en orange dans la figure. La duplication se traduit par deux sous domaines structuraux similaires et symétriques. La duplication 3D est plus étendue que la duplication trouvée en acides aminés, cette dernière étant environ deux fois plus étendue que la duplication nucléique détectée.



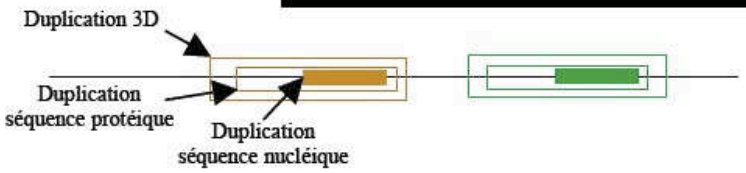


FIG. 1.7: Motifs relationnels structuraux et analyse intégrée d'une duplication 3D

L'équipe aborde dans ce thème les développements algorithmiques pour l'apprentissage dans des fragments de la logique du premier ordre, comme le Datalog non-recursif avec negation, selon deux axes privilégiés : l'utilisation de macro-opérateurs pour l'abstraction de l'espace de recherche, qui bénéficie de l'expertise acquise dans le cadre du thème «Treillis de Galois emboîtés - constructions de taxonomies», et la résolution algébrique du problème d'apprentissage dans des treillis distributifs. Ce deuxième axe s'intéresse à la formalisation et au développement de solveurs algébriques de problèmes d'apprentissage. Cela nous a notamment permis d'identifier des problèmes clés d'apprentissage de type MAX-CSP et CSP pondérés (*Weighted CSP*) qui sont des problèmes déjà étudiés dans l'équipe [CI-14]. Notamment sur ce dernier point, nous avons formalisé et développé une technique de résolution efficace dite «descendante dirigée par les données» dans des restrictions de la logique du premier ordre.

Dans un cadre transversal, nous étudions l'efficacité de ces techniques dans le cadre

empirique de la *transition de phase*. Ce cadre est une étude topologique de la complexité en moyenne de problèmes combinatoires NP-complets en fonction de paramètres d'ordre. Ce travail au sein de l'équipe fait suite aux nombreuses études et publications des différentes communautés d'Optimisation Combinatoire sur les problèmes de type SAT, CSP, TSP etc, avec la même finalité de développer un cadre d'analyse et de comparaison des approches algorithmiques proposées. Notamment, l'étude des transitions de phase permet le développement de générateurs de problèmes difficiles pour tout algorithme afin de faciliter l'analyse des performances des algorithmes en moyenne. Nous avons montré notamment que l'apparition d'une transition de phase (du test de couverture) en PLI était un cas pathologique pour la recherche heuristique des algorithmes d'apprentissage. Cependant, nos premiers résultats montrent que ce cas pathologique n'affecte systématiquement que les stratégies dites « générer-tester » et non pas la stratégie « descendante dirigée par les données ».

1.4.2 Résultats significatifs

- formalisation et développement d'une technique de résolution algébrique dite «descendante dirigée par les données» dans des restrictions de la logique du premier ordre [CI-48] ;
- étude de l'impact de la transition de phase du test de couverture sur la recherche heuristique en Programmation Logique Inductive [RI-11*]

1.4.3 Domaines d'applications

Les domaines visés par les méthodes d'apprentissage symbolique sont multiples, et en pleine extension, car de nombreux utilisateurs réclament des méthodes d'apprentissage dont les résultats sont exprimés sous forme de règles utilisant un vocabulaire explicite pour ces utilisateurs. Par exemple, nous avons travaillé sur des données semi-structurées décrivant des produits informatiques (données C/NET). Les deux autres axes applicatifs privilégiés de l'équipe sont :

- la Bioinformatique. Plus précisément l'équipe collabore avec le laboratoire MIG de l'INRA Jouy-en-Josas, avec l'Atelier de Bioinformatique, de l'Université Paris 6, ainsi qu'avec l'Institut Curie (Paris)
- le traitement de connaissances textuelles, avec le laboratoire MIG de l'INRA Jouy-en-Josas également

Ces domaines d'applications nous permettent d'établir des collaborations au sein de l'Université Paris 13 avec l'équipe RCLN et le Lim&Bio.

1.4.4 Perspectives de recherche

Les perspectives de l'apprentissage à partir d'exemples incomplets et de l'analyse de l'information mise en jeu au cours du processus d'apprentissage, sont à la fois appliquées et théoriques. Sur le plan des applications, la Bioinformatique est un terrain prometteur. En effet les changements de représentations en fonction de critères informationnels semblent intéressants pour guider la recherche de motifs répétés au sein de séquences biologiques : ainsi une séquence de nucléotides (ADN ou ARN) peut être représentée de façon plus abstraite comme une séquence d'acides aminés. Sur le plan théorique, nous nous intéressons à reproduire cette étude en présence d'informations révisables (non monotones).

Une part importante de notre activité est consacrée au développement d'un système hautement configurable d'apprentissage dans des langages de représentation au moins

aussi riches que le Datalog non-récursif. Nous allons valoriser à court terme ce développement en offrant à la fois aux spécialistes et aux praticiens de l'apprentissage une plate-forme pour l'expérimentation et pour la résolution de problèmes d'apprentissage. Cette mise à disposition va permettre de centraliser et de diffuser les différentes spécialisations de la plate-forme effectuées pour les applications ; elle servira également de plate-forme expérimentale pour le développement et la validation d'approches algorithmiques de l'apprentissage développées dans l'équipe A³ et notamment en Programmation Logique Inductive. Les problèmes ouverts de recherche dans le cadre relationnel que nous comptons aborder avec cette plate-forme sont la sélection/réduction du langage de concepts, en particulier à l'aide de macro-opérateurs, et le traitement de données incomplètes et bruitées.

Enfin, à plus long terme, nous souhaiterions développer un axe de recherche transversal (relationnel et numérique) dans l'équipe sur le thème de de l'*apprentissage relationnel probabiliste*, dont le but est d'effectuer du raisonnement de l'inférence probabiliste sur des représentations logiques, qui nous permettra de modéliser et d'apprendre des connaissances en utilisant la souplesse des représentations probabilistes et l'intelligibilité des représentations logiques.

1.5 Approches Transversales

Participants : P. Gérard, R. Kanawati, H. Karoui, C. Rouveirol

Cet axe rassemble les deux approches qui font collaborer des techniques d'apprentissage symbolique et numérique, l'apprentissage par renforcement relationnel d'une part, et le raisonnement à partir de cas. L'apprentissage par renforcement (AR) est une approche classiquement numérique. Un système automatique d'apprentissage par renforcement interagit avec son environnement et utilise l'information récoltée pour apprendre un comportement adéquat. Dans chaque situation qui peut se présenter, le système cherche à apprendre quelle action il vaut mieux entreprendre parmi celles qui sont à sa disposition. L'action la meilleure est celle qui devrait lui permettre de maximiser un cumul de récompenses scalaires obtenues sur le long terme. En effet, l'apprentissage se fonde sur une récompense scalaire sporadique, plus facile à définir par un concepteur que le détail des actions à entreprendre dans chaque situation possible. Le système cherche à maximiser le cumul de ces récompenses successives, sans connaître a priori son environnement, en effectuant un compromis entre récompenses immédiates et récompenses attendues dans le futur. Les systèmes d'AR peuvent être envisagés partout où un système autonome doit s'améliorer au fur et à mesure de son expérience. Les domaines d'application les plus courants sont la robotique, les jeux vidéo ou la fabrication industrielle.

Un problème central de l'AR est celui de la généralisation. En effet, dès que la taille des problèmes augmente, il est illusoire de vouloir apprendre quelle action entreprendre dans chacune des multiples situations possibles. Il est alors nécessaire d'utiliser des représentations propres des espaces d'états et d'actions de manière non-extensive. Depuis 2001, l'apprentissage par renforcement relationnel commence à explorer les représentations relationnelles pour remplacer les représentations attributs-valeurs, peu concises, peu aisées à généraliser, mais classiquement utilisées en apprentissage par renforcement. Les états et les actions sont alors représentés par des prédicats relationnels plutôt que par des vecteurs d'attributs. A ce jour, la plupart des travaux en AR relationnel portent sur les systèmes de régression relationnelle employés. Nous envisagerons plutôt l'adaptation des algorithmes classiques d'apprentissage par renforcement aux systèmes de PLI employés.

La deuxième approche transversale developpée dans l'équipe A³ concerne les travaux menés dans le domaine du raisonnement à partir de cas (RàPC). Le RàPC est une méthodologie de résolution de problèmes fondée sur l'adaptation des solutions éprouvées afin de résoudre de nouveaux problèmes similaires à ceux traités dans le passé. Naturellement, l'emploi de cette méthodologie ne peut être envisageable que dans des domaines où des problèmes similaires acceptent des solutions similaires. La figure 1.8 illustre le schéma classique d'une application RàPC : une base de cas dits cas sources regroupe les problèmes résolus et leurs solutions. La résolution d'un nouveau problème passe par cinq étapes successives : 1) l'élaboration qui permet de déterminer les attributs du nouveau problème à résoudre (appelé cas cible), 2) la rémémoration qui permet de trouver les cas sources les plus facilement adaptables par rapport au cas cible, 3) l'adaptation dont le rôle est de concevoir une solution au cas cible à partir des solutions appliquées aux cas sources retenues, 4) la révision de la solution proposée et 5) l'apprentissage qui permet d'ajouter le nouveau cas résolu si opportun.

Ainsi, la relation entre le domaine du RàPC et de celui de l'apprentissage automatique est double : d'une part, la cinquième phase du cycle RàPC assure une fonction d'apprentissage incrémental et paresseux, d'autre part des techniques d'apprentissage, numérique et symbolique, peuvent être employées pour mettre en œuvre les différentes phases du cycle RàPC. Nos travaux dans l'équipe A³ traitent les deux aspects de la relation entre les deux domaines.

Nous proposons une approche RàPC originale, que nous appelons RàPC guidé par les données. Selon cette approche, la mémoire d'un système RàPC est organisée en deux

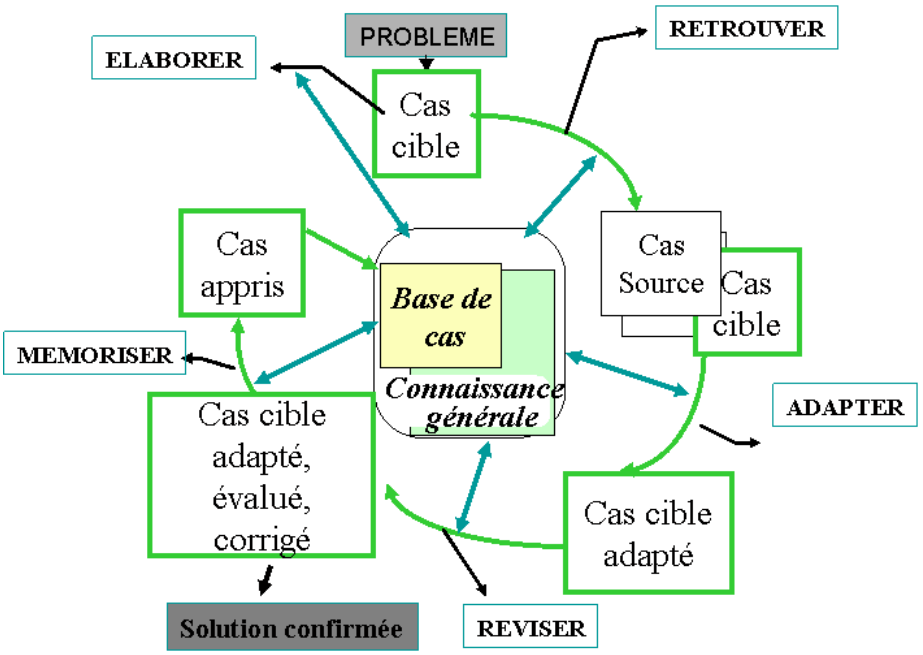


FIG. 1.8: *Le cycle RàPC*

niveaux : la base de cas source et une base de données brutes qui alimente la base de cas. Le cycle RàPC est modifié pour assurer une fonction de maintenance permanante du système (maintenance de la base de cas, des fonctions de similarité et des connsissance d'adaptation). Dans ce cadre, deux axes de recherche sont explorés :

- le dévelopement de systèmes RàPC hybrides combinant des approches symboliques et numériques,
- le dévelopement de systèmes RàPC coopératifs.

1.5.1 Les axes de recherche

Apprentissage par renforcement relationnel Le recherche en apprentissage par renforcement relationnel peut pourtant suivre deux voies complémentaires : d'une part, l'adaptation des algorithmes de programmation logique inductive (PLI) aux contraintes spécifiques des problèmes d'apprentissage par renforcement, largement étudiées à l'heure actuelle dans la communauté de PLI ; d'autre part, l'adaptation d'algorithmes de l'état de l'art d'apprentissage par renforcement aux systèmes de PLI employés, pour l'instant plus confidentielle.

Dans un premier temps, nous nous intéressons plus à l'exploitation de la nature relationnelle des données pour améliorer à la fois les techniques d'AR (indirect, hiérarchique etc) de manière à produire des algorithmes efficaces.

Raisonnement à partir de cas A la différence d'un système RàPC classique qui repose pour son fonctionnement sur la disponibilité préalable d'une base de cas sources, un système RàPC guidé par les données vise à alimenter une base de cas, initialement vide, au fur et à mesure de la présentation de nouveaux problèmes à résoudre. La mémoire d'un système RàPC guidée par les données est composée de deux parties : a) une base de cas source initialement vide et b) une base de données brutes contenant des observations sur le système cible. Par exemple, dans le

cadre d'une application de personnalisation automatique de sites Web, les données de trace de navigations (i.e. données de log web) constituent la base de données brutes, et la base de cas est composée d'un ensemble de motifs de navigation qui ont permis de prédire le comportement d'un utilisateur du site cible [CL-14].

Le cycle RàPC classique est modifié afin de prendre en compte la nature de la mémoire du système. D'une manière simplifiée, le nouveau cycle consiste à explorer la base de données brutes afin de proposer des cas sources dits potentiels lorsqu'une des phases de recherche, d'adaptation ou de révision du cycle RàPC classique, échoue. La phase d'apprentissage est aussi modifiée pour ne retenire dans la base de cas que les cas potentiels à partir desquels des bonnes solutions ont été derivées. Ce nouveau cycle permet de constituer une base de cas qui évolue avec la nature des problèmes à résoudre. Des techniques de maintenance asynchrones sont proposées afin d'éliminer de la base les cas à faible contribution [CI-20].

Nous appliquons cette approche dans deux domaines, la *classification et prédiction de séquences* et les *systèmes de recommandations coopératifs*. Le premier axe a fait le sujet de la thèse de F. Zehraoui où nous avons proposé un système hybride de classification et de prédiction de séqeunces de navigation dans un site web. Dans ce travail, les étapes de remémoration et d'adaptation des cas sont implantées sous forme d'une carte auto-organisatrice (la carte M-SOM-ART [CO-8]) conçue pour la classification des séquences et qui possède les deux propriétés de plasticité et de stabilité. Le deuxième axe fait le sujet de la thèse de H. Karoui. Une application pilote de partage et de recommendation de références bibliographiques est développée. Chaque utilisateur est assisté par un agent personel qui observe ses actions afin de détecter ses centres d'intérêt et qui coopère avec les autrs agents afin de recommander de nouvelles références pertinentes pour l'utilisateur associé. Chaque agent implante deux systèmes RàPC guidés par les données afin de 1) calculer les recommandations [CI-39] et 2) apprendre à associer une communauté d'agents à chaque centre d'intérêt détecté [CO-34].

1.5.2 Domaines d'applications

Un domaine d'applications priviligié pour nos travaux en RàPC concerne le dévelopement de système de recommandations [CL-16]. Deux champs d'activités sont particulièrement visés :

- le commerce électronique. Notre approche de traitement et de classification de séquences a été appliquée pour prédire le comportement d'utilisateurs d'un site de E-commerce afin de reconnaître leur intention (i.e. achat ou non achat). Le site peut alors être adapté dynamiquement en fonction de la classe prédite. Cette même approche peut servir aussi pour l'adaptation dynamique d'un site web afin de proposer des raccourcis de navigation.
- l'accès coopératif et intelligent à l'information distribuée. Notre objectif est de permettre à un groupe organisé d'utilisateurs (ex. équipe de recherche) de partager, d'une manière implicite et automatique, leurs expériences de recherche d'information. Trois applications différentes, construites selon la même architecture de type égal-à-égal (P2P), sont proposées : une application de partage de signets web [CL-16], une application de recommandation de références bibliographiques [RE-4] et un méta-moteur de recherche d'information sur le web orienté communautés [CI-1].

1.5.3 Résultats significatifs

Sur le plan théorique les résultats les plus signficatifs sont :

- la définition et la validation de l'approche RàPC guidée par les données [CI-20]
- la proposition et la validation d'un système hybride RàPC connexionniste pour la classification et la prédiction de séquences [CL-14]

Sur le plan applicatif, les réalisations les plus significatives sont :

- Le prototype COBRAS : un système égal à égal pour la recommandation de références bibliographiques. Ce système implante une approche RàPC pour la formation de comité dans les systèmes de type égal à égal [CI-39]
- Le prototype d'un moteur de recherche orienté communauté. Le principe est d'explorer les traces de l'utilisdation du moteur par une communauté d'utilisateurs afin d'améliorer le tri des résultats obtenus par des moteurs classiques de recherche sur le web [CI-10, CI-1].

1.5.4 Perspectives de recherche

L'axe d'apprentissage par renforcement relationnel est un axe qui débute tout juste, et qui est amené à se développer dans le futur. Il s'agit d'un travail novateur et exploratoire qui permet en outre d'initier une collaboration entre le LIPN et l'équipe de Luc de Raedt de l'Université Catholique de Leuven (Belgique).

En ce qui concerne l'axe RàPC, deux axes de recherche reliés sont en cours d'exploration :

- **L'apprentissage coopératif**. Deux problèmes nous semblent particulièrement intéressants : 1) l'identification de commaunautés d'experts dans un système RàPC coopératif et 2) l'étude des différents niveaux de coopération envisageables entre des systèmes RàPC autonomes. En ce qui concerne le premier problème, nous explorons une approche fondée sur l'utilisation des systèmes de réputations et d'analyse de communautés web. En ce qui concerne le deuxième problème, les travaux dans l'état de l'art se concentrent sur des stratégies d'échange de cas entre systèmes RàPC autonomes afin d'améliorer les performances globales du système. Nous voudrons explorer l'utilité d'autres types de coopération notamment la coopération pour la mise en place de chacune des phases du raisonnement, en particulier les phases de la recherche et de l'adaptation. Des algorithmes coopératifs pour le calcul de similarité de structures complexes (i.e. des séquences, des graphes) sont aussi à l'étude.
- **Techniques RàPC pour l'analyse des réseaux sociaux**. L'objectif est d'explorer l'application de la méthodologie RàPC guidé par les données dans des tâches d'identification des réseaux sociaux par l'exploration des traces d'interactions entre acteurs, mais aussi pour des tâches d'analyse des réseaux identifiés notamment pour la prédiction de l'évolution des réseaux et pour la prédiction de flux d'influences dans ce type de réseaux. Les systèmes de recommandation restent des applications cibles favorites, où les réseaux sociaux peuvent être employés à la fois pour enrichir les profils utilisateurs mais aussi pour cibler les utilisateurs de manière à augmenter l'impact des recommendations dans une communauté cible.

1.6 Animation de la recherche

1.6.1 Formation à la recherche

- **Habilitations à Diriger les Recherches**

Francine Krief : Thèse d'HDR en Informatique, Université Paris 6, Les réseaux haut débit : vers la gestion autonome et contextuelle.
Soutenue le 12 décembre 2003.
Coordinateur : G. Pujolle.
Jury : G. Pujolle, N. Agoulmine, D. Gaïti, A. Karmouch, Y. Bennani, P. Dague, O.C.M. Bandeira Durate, S. Zinaty.
Rapporteurs : N. Agoulmine, D. Gaïti, A. Karmouch.
Situation actuelle du diplômé : Professeur à Bordeaux.

- **Thèses soutenues**

Alejandro Guerra-Hernandez : Apprentissage d'Agents Rationnels BDI dans un univers Multi-Agents.
Financement de la thèse : Bourse Mexique
Soutenue le 6 Décembre 2003.
Directeurs : Amal Elfallah-Segrouchni et Henry Soldano.
Situation actuelle du diplômé : Maître de Conférences, Informatique Université de La Cruz (Mexique).

Khalid Benabdeslem : Approches Connexionnistes pour la classification et la visualisation des séquences évolutives : Application aux données issues d'usages de l'Internet.
Financement de la thèse : CDD sur Contrat de recherche.
Soutenue le 17 Décembre 2003.
Situation actuelle du diplômé : Maître de conférences à l'Université Lyon 1.

Farida Zehraoui : Systèmes d'apprentissage connexionnistes et raisonnement à partir de cas pour la classification et le classement de séquences.
Financement de la thèse : CDD sur Contrat de recherche.
Soutenue le 2 octobre 2004 (Très Honorable+Félicitations).
Situation actuelle du diplômé : Maître de conférences à l'Université d'Evry.

Hakima Kadri-Dahmani : Mise à jour incrémentale des bases de données géographiques et maintien de leur cohérence.
Financement de la thèse : Institut Géographique National (IGN france) Soutenue le 22 novembre 2005.
Directeur : Aomar Osmani, Philippe Dague

Nahla El-Zant : Recherche de motifs relationnels dans les séquences : Application aux séquences biologiques.
Financement de la thèse : Allocataire.
Soutenue le 1er Décembre 2005.
Directeurs : Henry Soldano et Daniel Kayser.

Zeina El Jrad : Apport des techniques de l'intelligence artificielle dans la négociation dynamique de niveaux de service : Proposition d'une interface utilisateur pour l'Internet de nouvelle génération.
Financement de la thèse : CDD sur projet RNRT.
Soutenue le 15 mai 2006.
Directeur : Francine Krief.
Situation actuelle du diplômé : Post-doc, INRIA (Rocquencourt).

Alexandre Pauchet : Modélisation cognitive d'interactions humaines dans un cadre de planification multi-agents.
Financement de la thèse : Allocataire.
Soutenue le 22 Septembre 2006. (Très Honorable+Félicitations)
Directeurs : Amal El Fallah-Seghrouchni et Nathalie Chaignaud
Situation actuelle : Post-doc, France Télécom (Lannion).

Hassine Moungla : Evolution vers une Gestion Autonomique des Réseaux de Télécommunications : Extension de l'approche PBM.
Financement : Allocataire
Soutenue le 20 Novembre 2006.
Directeur : Francine Krief

Nejla Amara-Hachmi : Une architecture générique d'agents mobiles sensibles au contexte et auto-adaptables.
Financement : Bourse Tunisie
Soutenue le 6 Décembre 2006.
Directeur : Amal El Fallah
Situation actuelle : Post-doc, France Télécom.

Sébastien Guérif : Réduction de dimension en apprentissage non supervisé.
Financement de la thèse : Bourse CIFRE.
Soutenue le 11 Décembre 2006.
Directeur : Younès Bennani
Situation actuelle : MCF Université Paris 13.

- **Thèses en cours**

Sujeevan Aseervatham : Thèse Paris 13, depuis 10/2004.
Sujet : Apprentissage à noyaux et réduction de dimension.
Directeurs : Emmanuel Viennet et Younès Bennani
Financement : ATER

Guénaël Cabanes : Thèse Paris 13, depuis 10/2007.
Sujet : Classification topographique évolutive à deux niveaux simultanés.
Directeur : Younès Bennani
Financement : Allocataire

Lahcène Dehni : Thèse Paris 13, depuis 10/2004.
Sujet : Anlyse et visualisation des données évolutives.
Directeur : Younès Bennani

Financement : ATER

Mohamed Elati : Thèse Paris 13, depuis 10/2006.
Sujet : Apprentissage symbolique pour la bio-informatique.
Directeur : Céline Rouveirol
Financement : ATER

Hager Karoui : Thèse Paris 13, depuis 11/2003.
Sujet : Un *framework* multi-agents pour l'aide à l'accès coopératif et intelligent à l'information distribuée.
Directeurs : Rushed Kanawati et Laure Petrucci.
Financement : ATER

Nicoleta Rogovschi : Thèse Paris 13, depuis 10/2006.
Sujet : Apprentissage et réseaux sociaux.
Directeurs : Emmanuel Viennet et Younès Bennani
Financement : CDD sur projet ANR

Nistor Grozavu : Thèse Paris 13, depuis 10/2006.
Sujet : Apprentissage non supervisé et pondération local des variables.
Directeur : Younès Bennani
Financement : CDD sur projet InfoMagic

- **Encadrement de stages de recherche**
  - 2004 (Y. Bennani) : F. Ilo (DEA (IA-OC) Université P13), *Outils connexionnistes pour la prévision d'usures*
  - 2004 (Y. Bennani) : B. Laarabi (DEA Statistiques Université P7), *Sélection de variables en classification automatique*
  - 2004 (Y. Bennani et F. Krief) : L. Dehni (DEA (IA-OC) Université P13), *Apprentissage non supervisé et réseaux de capteurs*
  - 2006 (Y. Bennani et S. Guérif) : M. Belgacem (Master Informatique (MICR) Université P13), *Apprentissage connexionniste non supervisé avec pondération et sélection de variables*
  - 2006 (E. Viennet) : B. Berrabah (Master Informatique (MICR) Université P13), *Méthodes locales pour la réduction non-linéaire de dimension : Applications à la fouille de textes*
  - 2006 (D. Bouthinon et H.Soldano) : J. Boussaa (DESS Université CADI AYYAD, Marrakech (Maroc)), *Elaboration d'un système d'apprentissage d'ordre 1 à partir d'exemples ambigus*
  - 2006 (Y. Bennani et C. Recanati) : N. Rogovschi (Master Informatique (MICR) Université P13), *Systèmes d'apprentissage non supervisé connexionnistes et stochastiques pour la fouille de données structurées en séquences*
  - 2006 (R. Kanawati) : M. Lafer (Master Informatique Recherche, Université Paris Sud), *Identification de communautés d'experts dans un système P2P.*
  - 2007 (Y. Bennani) : M. Chelli (Master Informatique (MICR) Université P13), *Apprentissage non supervisé avec pondération locale des variables*
  - 2007 (Y. Bennani) : G. Cabanès (Master Informatique (MICR) Université P13), *Apprentissage non supervisé évolutif*
  - 2007 (P. Gérard) : Ch. Rodrigues (Master Informatique (MICR) Université P13), *Architectures d'Apprentissage par Renforcement Relationel*

- **Enseignements liés à la recherche**

Le travail de recherche d'une grande partie des membres de l'équipe a débouché sur la création de nouveaux enseignements au niveau du Master d'Informatique :

- 2003-2005 (Y. Bennani) : Systèmes d'apprentissage connexionnistes pour la ouille de données, DEA IA-OC Paris 8 - Paris 13
- 2003-2005 (H. Soldano et M. Champesme) : Apprentissage symbolique (option), master 2 Informatique, spécialité EID, Paris 13
- 2005-2006 (Y. Bennani) : Apprentissage numérique (option), master 2 Informatique, spécialité MICR, Paris 13
- 2005-2006 (Y. Bennani) : Apprentissage connexionniste (option), master 2 Informatique, spécialité EID, Paris 13
- 2005-2006 (Y. Bennani) : Reconnaissance statistique des formes (option), master 2 Informatique, spécialité EID, Paris 13
- 2005-2006 (E. Alphonse) : Apprentissage symbolique (option), master 2 Informatique, spécialité EID, Paris 13
- 2005-2007 (A. Osmani et R. Kanawati) : Apprentissage symbolique et RàPC (option), master 2 Informatique, spécialité MICR, Paris 13
- 2006-2007 (Y. Bennani et P. Gérard) : Apprentissage numérique (option), master 2 Informatique, spécialité MICR, Paris 13
- 2006-2007 (C. Rouveirol) : Fouille de données relationnelles (option), master 2 Informatique, spécialité MICR, Paris 13
- 2007-2008 (C. Rouveirol et H. Soldano) : Apprentissage symbolique, master 2 Informatique, spécialité MICR, Paris 13
- 2007-2008 (C. Rouveirol et J.D. Zucker) : Fouille de Données complexes en Bioinformatique, master 2 Informatique, spécialité MICR, Paris 13

1.6.2 Valorisation de la recherche

• Contrats de recherche

Type	Coordinateur	Resp. scient.	Nom	Financeur	Période	Durée (mois)	Budget global (Keuros)	Part LIPN (Keuros)
ANR		Y. Bennani	CADI	MENESR	2008-2010	24		200
ANR		E. Viennet	SEPTIA	MENESR	2007-2009	24	554	100
ANR		Y. Bennani	SILLAGES	MENESR	2005-2008	36	375	113
ARC	•	Y. Bennani	EUROC25	ONERA	2006-2006	6		3
CAP DIGITAL		A. Nazarenko	Infom@gic	DGE	2005-2006	12		48
BQR		Y. Bennani	Faces	Univ. Paris 13	2005-2006	12	30	5
CIFRE		Y. Bennani	AppMark	NUMSIGHT	2004-2006	24		10
ACI-IMPBio		H. Soldano	EVOLREP	MENESR	2003-2005	24		40
CIFRE		Y. Bennani	FouDon	GEOBS	2003-2004	12		8
RNRT		F. Krief	IP-SIG	MENESR	2003-2004	24		120
RNTL		Ph. Dague	RESEDA	MENESR	2001-2003	48		200
RNRT		Ph. Dague	MAGDA2	MENESR	2001-2003	48		180
R&D	•	Y. Bennani	WebMining	NUMSIGHT	2000-2003	48		151
5ième PCRDT	•	Ph. Dague	IDD	UE	2000-2003	36		145

Au niveau international

• Projet européen IDD

« Integrated Design process for on-board Diagnosis » (2000-2003)
Partenaires du projet :

- LIPN-CNRS : Ph. Dague
- Fiat, Daimler-Chrysler, PSA, Renault, Magneti-Marelli, OCC'M et les universités de Munich, de Turin.

Objectifs : L'objectif du projet IDD, « Integrated Design process for on-board Diagnosis », soumis dans le cadre du premier appel d'offres du programme Croissance compétitive et durable du 5ième PCRDT, était d'intégrer l'approche du raisonnement à base de modèles au processus de conception et de développement, afin de prendre en compte dès cette phase les nécessités du diagnostic. Nous sommes convaincus que le raisonnement à base de modèles ne pénétrera réellement l'industrie que s'il accompagne tout le cycle de vie d'un produit, de la conception à la maintenance, en passant par le développement, le test et le diagnostic, avec un partage des modèles dans ces différentes tâches et leur mise à jour en amont au niveau de la conception. Les principales tâches étudiées dans ce projet ont été la diagnosticabilité (détectabilité et discriminabilité) et la génération de systèmes de diagnostic à bord. Le LIPN a été en charge de la conception et la réalisation

d'interfaces entre les outils de CAO ou de simulation numérique (MATLAB/Simulink dans ce cas) et les systèmes de diagnostic à base de modèles (RAZ'R d'OCC'M dans ce cas), ce qui a nécessité des études sur la transformation automatisée de modèles (modèles numériques vers modèles qualitatifs). Il a réalisé le logiciel MOUSE (MOdel StrUcture Extraction) qui extrait les caractéristiques structurelles (composants et connexions) du modèle MATLAB/Simulink et discrétise le modèle numérique de simulation en un modèle qualitatif de diagnostic au niveau des composants, fournissant en sortie, au format xml, un modèle directement exploitable par RAZ' R. L'application retenue par PSA était le système de climatisation du véhicule et MOUSE a exploité le modèle Simulink de ce système.

- **Projet de coopération FS-KDDM**

Feature Selection for Knowledge Discovery and Data Mining. (1999-2003)
Partenaires du projet :

– LIPN-CNRS : Y. Bennani
– University of Skopje : D. Cakmakov

Objectifs : Ce projet porte sur l'étude de la fonction de base d'une méthode de sélection de variables (caractéristiques) qui est de choisir un sous ensemble de variables pertinentes à partir de la représentation vectorielle des formes observées. Généralement la phase de sélection et de décision (classement) s'effectuent l'une après l'autre dans un système classique de reconnaissance des formes, les méthodes connexionnistes sont très attrayantes car elles sont capables de mener les deux phases (sélection et classement) en parallèle : les variables sélectionnées optimisent le critère d'apprentissage. Y. Bennani et M. Yacoub ont proposé dans le cadre de ce projet quelques mesures de pertinence dont une dépendante des données et du système et une autre dépendante du système. Y. Bennani a aussi proposé une autre approche qui consiste à pondérer les variables en fonction de leurs participation à la discrimination. Ce projet a donné lieu à la publication d'un ouvrage sur le thème : Cakmakov D. & Bennani Y. (2002) , *Feature Selection for Pattern Recognition*, Informa Press, Ed. ISBN 9989-943-02-8.

- **Projet de coopération, COCPRA**

Combining and Optimizing Classifiers for Pattern Recognition Applications. (2004-2007)
Partenaires du projet :

– LIPN-CNRS : Y. Bennani
– University of Skopje : D. Cakmakov, B. Jaksimovski

Objectifs :

Ce projet concerne la conception et la mise en oeuvre de systèmes hybrides qui font coopérer différents rmodèles connexionnistes entre eux et avec des techniques plus classiques dans différents domaines d'application. Ces systèmes constituent une autre génération des modèles connexionnistes et apparaissent comme la seule solution pour la réalisation de tâches complexes et de grande taille. Les études théoriques portent sur les algorithmes d'apprentissage pour ce type de système, la décomposition de tâches, la conception de chaînes de traitement optimales. Y. Bennani a travaillé sur la possibilité de coopération des algorithmes connexionnistes et d'autres méthodes plus classiques comme les Chaînes de Markov Cachées, afin de concevoir des systèmes modulaires hybrides. Ces architectures présentent en théorie l'avantage de combiner les points forts de leurs composants et par exemple d'obtenir des systèmes qui soient à la fois discriminants et

incrémentaux. Y. Bennani et V. Radevski ont proposé aussi des approches pour la fusion de plusieurs types de données (structurelles et statistiques) et pour la fusion de décisions dans des systèmes multi-modulaires.

- **Projet de coopération, DCL&WSN**

Distributed Connectionist Learning for Wireless Sensor Networks " avec Federal University of Ceará, Brazil. (2005-2007)
Partenaires du projet :

– LIPN-CNRS : Y. Bennani
– Federal University of Ceará : J. Neuman de Souza, D. Gomes

Objectifs : Les estimations de la bande passante allouée au trafic vidéo IP sont réalisées par le biais de deux techniques de prédiction complémentaires. Une des contributions a consisté à intégrer un module connexionniste de prédiction dans une architecture de provisionning dynamique de la bande passante entre un fournisseur de VoD et un ISP (Internet Service Provider). Le fonctionnement de cette architecture est géré par un ensemble de politiques à l'instar de l'architecture PBM (Policy Based Management) de l'IETF (Internet Engineering Task Force) et réalisé à travers les technologies du Web. En effet, l'automatisation du provisionning est un concept nouveau qui n'est pas encore bien accepté au niveau des opérateurs qui souhaitent que toutes les procédures soient validées par des contrats papiers. Néanmoins, la complexité des changements associés à la dématérialisation de la société pousse vers des solutions d'automatisation complète de l'installation et de la configuration de nouveaux services. Pour réduire leurs frais opérationnels, les opérateurs vont recourir de plus en plus à l'automatisation des configurations de services et à leur contrôle. Ce domaine en pleine expansion s'appelle : *l'autonomic networking* ou encore les réseaux auto-gérables. L'objectif principal de cette étude consiste à développer des algorithmes d'apprentissage numériques adaptés à un environnement distribué. Ce type d'algorithmes est plus adaptés à notre domaine d'application.

Au niveau national

- **Projet Infom@gic**

[Dans le cadre du pôle de compétitivité Cap Digital (IMVN : Image, Multimédia et Vie Numérique)] (2006-2008)
Pilote industriel : Thales
Partenaires du projet :
BERTIN, EADS, EUROPLACE, INA, INTUILAB,ODILE JACOB, ONERA, PERTIMM , SPIKENET, TEMIS, THALES, VECSYS, XEROX , CEA, ENST, Univ. Paris 6, Univ. Paris 8, Univ. Paris 13, Univ. Paris Dauphine, Univ. Orsay, Univ. Marne la Vallée, CNRS

Objectifs : Ce projet consiste à mettre en place, sur une période de trois ans, un laboratoire industriel de sélection, de test, d'intégration et de validation sur des applications opérationnelles des meilleures technologies franciliennes dans le domaine de l'ingénierie des connaissances. Ce laboratoire s'appuie sur une plate-forme commune d'interopérabilité. Cette plate-forme doit couvrir les grands domaines fonctionnels et techniques de l'analyse d'information que sont la recherche et l'indexation, l'extraction de connaissances, et la fusion d'informations multimédias, sur tous les types de sources : texte, données, images, son. Elle incorpore des applications des technologies développées dans les domaines de la e-Education, et de la gestion des patrimoines culturels numériques.

Participation du LIPN-CNRS :

- Sous-Projet 2 : Extraction d'informations (ST 2.12), A. Nazarenko, T. Poibeau.
- Sous-Projet 3 : Analyse et fusion d'informations (ST 3.12 et ST 3.21), S. Aseervatham, Y. Bennani, N. Grozavu, S. Guérif, M. Lebbah, E. Viennet.

- **Projet (ANR) CADI**

Composants Avancés pour la DIstribution (2008-2010)

Partenaires du projet :

- LIPN-CNRS, Younès Bennani
- LIP6-CNRS, Massih-Reza Amini
- LITIS, Stéphane Canu
- KXEN, Erik Marcade
- Numsight, Eric Janvier
- SPACECODE, Georges Folcke
- SAMSE - La Boîte à Outils, Frédéric Dufau-Joël
- MONDOMIX, Marc Benaïche

Objectifs : Le projet CADI vise à produire un prototype de composants logiciels pour la Distribution, tant brick-and-mortar (magasins) que vente en ligne (site de e-commerce). Dans le nouveau modèle économique de la distribution, dit de la « Longue Traîne », on voit émerger d'un catalogue potentiellement illimité des multiplicités de marchés de micro-niches que le site commerçant peut espérer développer pour dériver des revenus importants. Pour cela, le site doit aider le client à trouver le produit susceptible de l'intéresser dans des micro-niches de produits adaptées à ses goûts et ses attentes et adapter ses pratiques marketing et commerciales traditionnelles pour prendre en compte les communautés qui se créent sur les blogs et les forums autour de ses produits. Visant un marché à l'intersection de quatre domaines en forte croissance (e-commerce, data mining, réseaux sociaux et RFID), représentant un marché potentiel de plusieurs milliards d'euros, le projet CADI produira, dans une approche Web services sur une architecture SOA, les prototypes des composants logiciels suivants :

- Analyse de comportement d'achat
- Moteur de recommandation
- Moteur de rating
- Analyse des communautés

- **Projet (ANR) SEPTIA**

Solution automatique pour le traitement des interactions clients (2007-2009)
Partenaires du projet :

- LIPN-CNRS : E. Viennet
- LIP6-CNRS : P. Gallinari
- KXEN : F. Fogelman
- TEMIS : Charles Huot
- EPTICA : Thierry Gandilhon

Objectifs : Le projet Septia vise à produire un logiciel avancé de traitement des interactions clients dans le cadre d'un centre de contacts. Axé de façon prédominante sur le traitement des emails, le projet étudiera également l'apport des données clients, des

données d'interactions téléphoniques et des données d'interactions inter-clients dans un portail collaboratif. Les caractéristiques principales :

- Extraction automatisée des thèmes des emails et constitution des ontologies associées, en utilisant de stechniques statistiques et sémantiques.
- Aide à la réponse (recherche automatique du "meilleur" interlocuteur.
- Supervision des activités (prévision de séries temporelles)
- Portail collaboratif : étendre les otils à un portail (forum, base client, chats, ...)

Le projet prévoit la construction d'un démonstrateur qui sera testé sur des données réelles de grande taille. Il sera l'occasion de développer des technologies avancées dans le domaine de la fouille de texte, de la fouille de données et de l'étude des "réseaux sociaux".

- **Projet (ANR) SILLAGES**

Suivi Individuel par Label éLectronique Au sein de Groupes complexES. (2006-2008) Partenaires du projet :

- LIPN-CNRS : Y. Bennani
- LEEC-CNRS : D. Fresneau
- ULB (Université Libre de Bruxelles) : J-L. Deneubourg
- SPACECODE : G. Folcke

Objectifs : Les structures sociales animales sont des systèmes dynamiques caractérisées par un grand nombre d'interactions entre individus. Malgré leur diversité, ces structures sont basées sur un nombre limité de règles génériques (amplifications, inhibitions, compétition pour les ressources). Les dynamiques du groupe et ses réponses résultent de la synergie entre ces interactions, des capacités individuelles à traiter l'information et de la diversité de ces réponses individuelles.

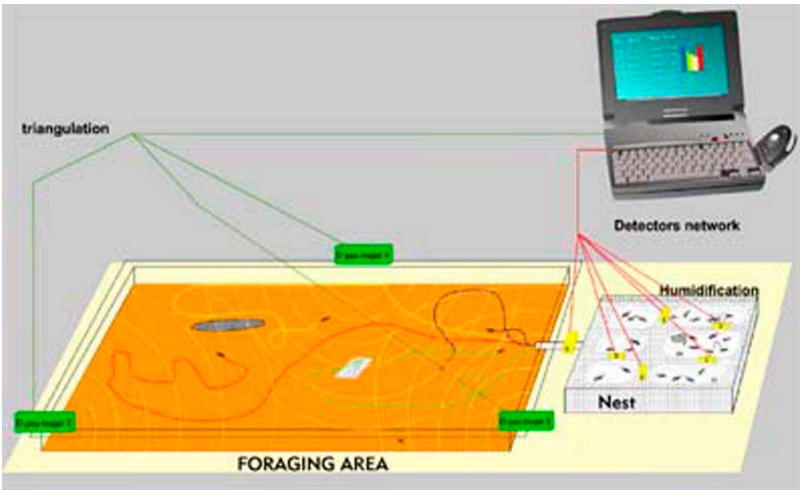


Fourmi avec le capteur RFID

Les objectifs de ce projet sont :

- le développement de nouveaux systèmes autonomes de suivi de l'activité spatio-temporelle des groupes, systèmes donnant accès à des données inaccessibles actuellement ;
- le développement d'outils d'analyse automatique de ces données et de quantification des comportements individuels ou collectifs ;
- l'intégration de ces quantifications aux outils de modélisation qui permettent de valider le lien entre les niveaux individuels et collectifs ;
- ces systèmes autonomes de suivi pourront contrôler des dispositifs et modifier l'environnement en fonction de l'identité des individus. Ceci permettra de tester notamment des hypothèses portant sur la plasticité comportementale.

Ces objectifs font de SILLAGES une action interdisciplinaire à l'intersection des sciences du comportement et du complexe et des sciences informatique et de l'ingénieur.



Dispositif expérimental RFID pour le suivi des fourmis

- **Projet EVOLREP, ACI-IMPBio**

Impact et évolution des répétitions dans les génomes : interactions entre innovation génétique et fonction biochimique (2003-2005).
Porteur : J. Pothier.
Atelier Bioinformatique, UFR 927, Université Pierre et Marie Curie, Paris.

Partenaires du projet :

- Unité Génétique des Génomes Bactériens, URA CNRS 2171, Institut Pasteur
- Equipe Structure et Dynamique du Génome, Institut Jacques Monod, UMR CNRS 7592
- Equipe Helix, INRIA Rhône-Alpes
- LIPN-CNRS : H. Soldano
- Equipe Contrôle de l'Expression Génique Laboratoire Adaptation et Pathogénie des Microorganismes, (CNRS UMR 5163), Université Joseph Fourier, Grenoble.

Objectifs : Une des énigmes majeures de l'évolution des génomes reste le rôle des phénomènes de duplication de l'ADN. Les évènements de duplication sont à l'origine de l'accroissement de la taille des génomes, des familles de gènes et de la création de la nouveauté fonctionnelle. Malgré cela, le rôle et les raisons de leur existence et les mécanismes qui contrôlent leur développement sont encore peu connus. Ce projet avait pour ambition l'étude intégrée des duplications intra-géniques dans les génomes des micro-organismes complètements séquencés (archeae, bactéries et levures) aux trois niveaux de description que sont les séquences nucléiques, protéiques et les structures 3D des protéines. Différentes méthodes de recherche (classiques ou développées dans le cadre du projet) ont été utilisées pour analyser les interactions que ces duplications intra-géniques peuvent avoir aux trois niveaux de description précités. En particulier le LIPN a travaillé sur l'algorthmique de la recherche de motifs structuraux répétés dans les structures 3D de protéines.

Dans le cadre de ce contrat, le CNRS (département STIC) a accordé une bourse de post-doc d'un an au LIPN, attribuée à Nadia Pisanti.

- **Bonus Qualité Recherche (BQR) [Face] (2005-2006)**

Projet de coopération LIPN-LEEC (BQR) : Détection et reconnaissance des mouvements faciaux liés au genre chez l'Homme. (2006-2007)
Partenaires du projet :

– LIPN-CNRS : Y. Bennani, S. Guérif
– LEEC-CNRS : C. Baudoin

Objectifs : Ce projet concerne l'étude de la signalisation et de la reconnaissance du genre à partir des mouvements faciaux. La prise en compte des mouvements faciaux est relativement récente et a été effective dans certains travaux portant sur l'expression des émotions et dans la communication parlée. Curieusement malgré des études comparatives, déjà anciennes sur les communications non verbales, peu de recherches récentes concernent le rôle des mouvements faciaux dans la signalisation d'appartenance à un groupe et dans la reconnaissance sociale chez l'Homme. Les objectifs du projet sont :

– de détecter les mouvements faciaux émis par des hommes et des femmes, appartenant à différents groupes culturels, dans des situations standardisées en situation contrôlée de laboratoire,
– de vérifier les résultats d'une étude pilote indiquant l'existence de mouvements faciaux latéralisés indicateurs du genre,
– de définir si ces mouvements sont perçus par des sujets « juges » des deux sexes, indépendamment d'indices morphologiques du visage,
– d'explorer ces mêmes informations par des réseaux connexionnistes (supervisés et non-supervisés),
– d'interpréter les phénomènes de causalité à partir des résultats de l'exploration de séquences spatio-temporelles des visages au moyen de modèles graphiques de type réseaux bayésiens.

- **Convention Industrielle de Formation par la Recherche, Numsight**

CIFRE N° 654/2002 : Extraction de connaissances à partir de données évolutives : application aux données comportementales. (2003-2006)

Partenaires du projet :

– LIPN-CNRS : Y. Bennani, S. Guérif
– NumSight : E. Janvier

Objectifs : Ce projet de recherche possède deux aspects : un aspect fondamental qui consiste à développer de nouveaux algorithmes d'apprentissage non supervisé pour le traitement et l'exploration de données comportementales, et un aspect applicatif qui rentre dans le cadre des activités de Numsight Consulting France en ingénierie de Marketing. Dans le cadre de cette convention CIFRE, nous avons proposé d'étudier un ensemble de problématiques des modèles connexionnistes et leur mise en oeuvre pour lesquels nous pensons qu'ils offrent des avantages et des spécificités par rapport aux techniques classiques d'analyse des données.

– L'étude est structurée en quatre phases :
– Modélisation globale : cherche un modèle unique de représentation qui résume d'une manière globale l'ensemble des données.
– Étude de l'interstructure : consiste à identifier des proximités ou des différences entre les ensembles de données, sans description.
– Étude de l'intrastructure : permet grâce à une analyse fine des ressemblances ou des différences d'expliquer la structure identifiée.

– Étude de l'évolution des informations : analyse et décrit l'évolution du phénomène d'une manière globale ou étudier les évolutions de chacun des éléments.

- **ARC-ONERA**

Action de Recherche Coopérative avec l'ONERA « Algorithmes d'apprentissage artificiel pour le traitement des profils atmosphériques ». (2005-2006).

Partenaires du projet :

– LIPN-CNRS : Y. Bennani, S. Guérif
– ONERA : T. Huet

Objectifs : Cette étude consiste à définir une méthode de classification automatique permettant une réduction notable des profils atmosphériques constituant les scènes EUROC25 issus de Météo-France. Le domaine EUROC25 comprend 129 points Ouest-Est par 105 points Nord-Sud. Chaque point de grille est caractérisé par 21 niveaux comprenant l'altitude, la pression, la température et l'humidité relative. La situation du profil par rapport à la mer est également précisée dans le fichier. Sous-jacente à ces données, une information fréquentielle est également présente, sous la forme d'un coefficient d'absorption calculé à partir des données précédemment citées. L'étude s'articule autour de l'exploration et l'extraction de connaissances à partir de données EUROC25. L'évolution des données au cours du temps et leur position spatiale nécessitent d'explorer simultanément plusieurs ensembles de données multi-variables. En statistique, des méthodes d'analyse factorielle multiple et de classification ont été développées ces dernières années, mais elles sous-estiment le plus souvent le caractère spatial des données. Par contre, dans le domaine de l'apprentissage artificiel, et en particulier connexionniste, le traitement de ce type de données est un problème qui représente une voie de recherche très prometteuse ouvrant d'excellentes perspectives. Dans le cadre de cette prestation entre l'ONERA et le Laboratoire d'Informatique de Paris Nord (LIPN, UMR 7030 du CNRS), Y. Bennani et F. Carton (étudiant en Master Pro EID) ont proposé d'étudier un ensemble de problématiques des systèmes d'apprentissage connexionnistes et leur mise en oeuvre pour lesquels ils pensent qu'ils offrent des avantages et des spécificités par rapport aux autres techniques. Cela les conduira à réaliser soit des études ponctuelles sur quelques points très ciblés qui doivent pouvoir s'insérer dans d'autres études déjà menées à l'ONERA, soit à développer des nouveaux outils connexionnistes de classification automatique.

- **Projet RESEDA [RNTL]**

RESEDA (Réalisation d'un Environnement Système Evolutif pour le Diagnostic Automobile). (2001-2005)

Partenaires du projet :

– LIPN-CNRS : S. Aseervatham, Y. Bennani, P. Dague, A. Osmani
– Renault : P. Dubois
– Silicomp : V. Joloboff, A. Boutin
– Trialog : P. Robin

Objectifs : Ce projet se propose 3 objectifs :

– Définir une architecture de composants logiciels évolutifs de diagnostic en utilisant des techniques d'objets distribués, et la technologie de code mobile Java.

- Construire un démonstrateur qui permettra de valider l'infrastructure en utilisant l'état de l'art des méthodes de diagnostic chez les constructeurs.
- Elaborer, dans une démarche plus exploratoire, les méthodes et outils qui permettront de bénéficier du retour d'expérience sur le parc de véhicule en circulation pour faire évoluer les composants de diagnostic en tenant compte des problèmes constatés sur le terrain.

- **Projet MAGDA 2 [RNRT]**

MAGDA 2 : Modélisation et Apprentissage pour une Gestion Distribuée des Alarmes De bout En boUt . (2001-2003)

Partenaires du projet :

- LIPN-CNRS : P. Dague, F. Krief, A. Osmani, E. Viennet
- Alcatel : G. Desoblin, H. Papini, A. Aghasaryan, E. Marilly, B. Lapraye
- ILOG : B. Berstel, P. Feillet, E. Mazeran, C. de Sainte Marie
- France Telecom : D.Collobert, C. Dousson
- IRISA : A. Benveniste, Eric Fabre, C. Jard, M-O. Cordier

Objectifs : L'objectif du projet MAGDA 2 est de trouver des solutions avancées aux problématiques de la gestion des réseaux hétérogènes et de la prise en compte d'événements faisant intervenir l'interaction entre la couche réseau et la couche service . A cette fin, nous envisageons d'une part l'étude et le développement d' algorithmes robustes de corrélation et de diagnostic pour les réseaux de télécommunication ainsi que pour l'interaction entre la couche réseau et la couche services. Par ailleurs, nous développerons des techniques d'automatisation (partielle) de l' acquisition des connaissances et modèles nécessaires. Le projet MAGDA2 s'appuie de manière essentielle sur le succès du projet MAGDA. Il vise à étendre les résultats aussi bien techniques qu'industriels de MAGDA en profitant des connaissances communes acquises et de la dynamique que les partenaires ont su créer au sein de ce projet. Plus, Il vise à lever des limitations identifiées par MAGDA et à résoudre les nouveaux problèmes qu'engendre la mise en oeuvre industrielle des techniques développées dans ce projet.

- **Projet IP-SIG [RNRT]**

IP-SIG : Signalisation générique du monde IP. (2003-2004)
Partenaires du projet :

- LIPN-CNRS : F. Krief
- ENST, ALCATEL, UTT, ISEP, LIP6, THALES

Objectifs : Le monde IP a longtemps lutté pour se passer d'une signalisation grâce à l'adresse IP qui est routée et qui se situe dans tous les paquets IP. Cependant, depuis déjà une dizaine d'années, le monde IP se pose la question d'une signalisation permettant de gérer le réseau différemment des solutions existantes. Les évolutions MPLS et GMPLS, la voix sur IP ainsi que les techniques de contrôle par politique ont abouti à de nouvelles signalisations complètement incompatibles. L'IETF a décidé en 2002 de lancer un groupe de travail pour essayer d'unifier toutes ces solutions ou tout au moins de les faire coexister. Le but de ce projet est de participer aux travaux de l'IETF en proposant une norme de signalisation universelle capable de répondre au moins aux besoins de sécurité, de QoS et de mobilité. Notre idée dans ce projet est d'explorer la voie du protocole COPS comme protocole fédérateur. La raison en est simple, COPS est la solution poussée par

de nombreux industriels pour arriver à faire de la qualité de service, COPS est flexible et peut parfaitement s'adapter à la gestion de la sécurité et de la mobilité.
Les objectifs de ce projet sont :

- La définition d'une architecture de signalisation universelle qui est capable de mettre en place une connexion sécurisée tout en tenant compte des besoins de QoS et de mobilité.
- La définition d'un SLS standard.
- Une étude prospective sur la dynamisation du SLS.

- **ARC-THALWEG**

Projet ARC-THALWEG : Action de Recherche Coopérative avec le projet AIR de l'INRIA, "Modèles de traitement d'image pour les données turbulentes". (2002-2004) Partenaires du projet :

- LIPN-CNRS : Y. Bennani
- LMD-Ecole Polytechnique : M. Desbois, R. Roca, A. Szantai
- LPS-ENS-Ulm : J.P. Nadal
- INRIA-Rocquencourt : H. Yahia, I. Herlin, J. Grazzini, A. Turiel

Objectifs : Ces travaux sont basés sur l'élaboration de nouvelles méthodes de traitement d'image dans lesquelles on s'intéresse à la caractérisation de la turbulence dans un signal. Il est donc fait appel à des résultats récents utilisant les domaines de la physique et du traitement du signal, de la théorie des ondelèttes, des multifractales, du traitement d'image et de l'apprentissage artificiel. Ces méthodes sont actuellement appliquées à la détermination de certaines zones caractéristiques dans des images météorologiques infrarouges Meteosat. On cherche à segmenter les zones correspondant à une activité pluviométrique dans un nuage, en tenant compte des propriétés de turbulence dans le signal image. Cela conduit à la détermination de zones particulières appartenant à la "MSM ("Most Singular Manifold" ou "variété la plus singulière") contenant l'information des zones de transition maximale dans le signal. Ces ensembles sont de nature multifractale. Dans une première phase, on a donc tenté de définir des attributs reliés à la MSM et conduisant à des opérateurs de traitement d'image permettant la détermination des zones pluviométriques.

- **Numsight [Web mining]**

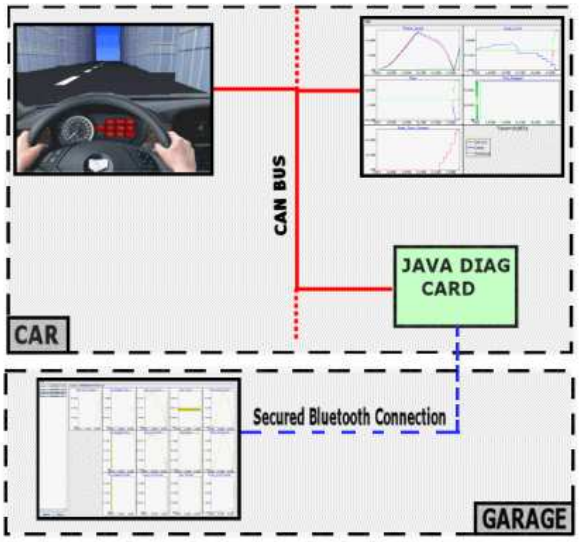
Apprentissage numérique pour la modélisation des Internautes. (2000-2003) Partenaires du projet :

- LIPN-CNRS : K. Benabdeslem, Y. Bennani, R. Kanawati, S. Salotti, F. Zehraoui
- NumSight : E. Janvier, F. Martin

Objectifs : La croissance accélérée du nombre d'utilisateurs d'Internet, pour la plupart novices, motive fortement le développement des approches d'aide adaptative pour faciliter l'accès à l'information.L'objectif de ce projet est le développement de systèmes d'apprentissage connexionnistes pour réaliser d'une part des tâches de classification visant à détecter des profils de comportement et d'autre part de reconnaissance dynamique basée sur l'évolution à court terme des traces de navigations des internautes afin d'anticiper le profil du navigateur ou de détecter l'émergence de comportements caractéristiques d'un profil déjà connu par le système.

- **Réalisations informatiques (logiciels) et Brevets**

Prototype Diagnostic embarqué (RESEDA) : Le démonstrateur de diagnostic embarqué a été réalisé dans le but d'évaluer, en temps réel, les méthodes de diagnostics mises au point par le LIPN dans le cadre du projet RESEDA. Le démonstrateur est constitué d'une unité centrale de diagnostic intégrant un microprocesseur Java et d'un module de communication sans fils "Bluetooth". La carte se connecte au système à diagnostiquer par un bus CAN.



Architecture du système de diagnostic embarqué

Pour la simulation, la carte peut se connecter à un PC, par une liaison CAN, et surveiller le comportement d'un simulateur développé sous Matlab/Simulink ou Java. Les informations de diagnostic peuvent, ensuite, être transmises à un système distant par une liaison Bluetooth sécurisée.

Auteur : S. Aseervatham

http ://www-lipn.univ-paris13.fr/ aseervatham/projects/cardiag.html

Plateforme Galicia : Les travaux sur les treillis de Galois emboîtés sont actuellement implémentés dans le cadre de la «plateforme Galicia»

Auteurs : M. Champesme et H. Soldano

(http ://www.iro.umontreal.ca/ galicia/)

RNA : Un programme de recherche de la structure secondaire commune à un ensemble de séquences d'ARN non alignées.

Auteurs : D. Bouthinon et H. Soldano

http ://www-lipn.univ-paris13.fr/ bouthinon/rnaHomedb/index.html

MACS (Modular Anticipatory Classifier System) : MACS est un Sytème de Classeur qui utilise l'apprentissage de modèles dans une architecture de type «Dyna», tout en offrant des capacités de généralisation.

Auteur : P. Gérard

http ://animatlab.lip6.fr/RechercheProjetXBotMACS

Method for processing Internet Request files : Brevet International, [N° de publication : WO 02/059781, date de publication : 01/08/2002]

Auteurs : K. Benabdeslem, Y. Bennani, E. Janvier

Un moteur de recherche sur le Web orienté communauté : (Système développé en projet de Master). Logiciel développé en Java, MySQL.

Auteur R. Kanawati

Un premier prototype du système CORBAS : Système de gestion et de recommandation de références bibliographiques. Système réalisé dans la cadre de la thèse de H.Karoui. Système dévelopé sur la plate-orme multi-agent Jade.

Auteur R. Kanawati

1.6.3 Diffusion de la recherche

La vie scientifique de l'équipe A^3 est organisée autour de plusieurs manifestations (trimestriel en alternance : chaque mois une manifestation) visant à introduire une dynamique permanente entre les différents chercheurs sur la thématique de l'apprentissage. Citons quelques exemples :

- **Séminaire «Apprentissage de Paris-Nord» (APN)**

Responsables : P. Gérard et M. Lebbah (LIMBio).
La fréquence des interventions est tous les mois, les thèmes abordés portent essentiellement sur l'apprentissage et la fouille de données (http ://www-lipn.univ-paris13.fr/A3/Seminaire/).

- **Tutoriels A^3**

Responsable : A. Osmani
En alternance avec le séminaire, ces exposés sont d'intérêt général (sous forme de cours) et durent de 1h à 1h30 et sont suivis d'une discussion scientifique.

- **Atelier des doctorants A^3**

Responsables : D. Bouthinon et H. Moungla
En complément au « séminaire junior », cet atelier est organisé sous forme d'une table ronde par les doctorants qui s'intéressent de près ou de loin aux techniques d'apprentissage. Le but est de leur permettre de présenter leurs travaux de recherche, les problèmes liés à leur sujet de thèse, ou de partager avec les autres doctorants et les permanents leurs expériences.

- **Journées thématiques «Apprentissage Artificiel et Fouille de Données» (AAFD)**

Responsables : Y. Bennani et E. Viennet
Tous les deux ans, l'équipe A^3 organise les journées (AAFD). L'objectif de ces journées est de faire le point sur l'état de l'art des techniques et des applications utilisant l'apprentissage pour la fouille de données. Ces journées se présentent sous forme d'un forum de rencontres entre chercheurs et industriels oeuvrant dans le domaine de la fouille de données (http ://www-lipn.univ-paris13.fr/A3/AAFD06/).

- **Actions de vulgarisation**

S. Aseervatham, Y. Bennani, L. Dehni, et S. Guérif, ont présenté des démonstrations de plusieurs prototypes, lors de visites du LIPN par des universitaires, des membres du CNRS ou des industriels ou lors de manifestations telles que les journées « Science en

fête - Savante Banlieue » : systèmes de reconnaissance vocale, de reconnaissance visuelle, de détection et d'identification de perturbations sur le réseau téléphonique commuté, de diagnostic autonome et évolutif, économie d'énergie dans les réseaux de capteurs sans fil, systèmes de diagnostics pour automobiles.

1.6.4 Evaluation de la recherche

- **Organisation d'événements scientifiques, Comité d'organisation, de programme et de lecture**

  - Comité de programme de la conférence francophone Extraction et Gestion des Connaissances (EGC'08), Sophia Antipolis, 29 janvier- 1 février 2008 ; Y. Bennani
  - Comité de programme de EPNACS'07 - Emergent Properties in Natural and Artificial Complex Systems, Dresden, Germany October 4-5, 2007 ; Y. Bennani
  - Comité de programme de International Symposium on Signal Processing and its Applications (ISSPA) in conjunction with the International Conference on Information Sciences, Signal Processing and its Applications (ICISSPA) 12 - 15 February 2007, Sharjah, United Arab Emirates (U.A.E.) ; Y. Bennani
  - Comité de programme de Intenational Symposium on Programming and Systems, May 7-9, 2007, Algeria ; Y. Bennani
  - Comité de programme de European Conference on Machine Learning (ECML) 2007, Warsaw, Poland ; C. Rouveirol
  - Comité de programme de ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) 2007, San José, California (U.S.A) ; C. Rouveirol
  - Comité de programme de IEEE International Conference on Data Mining series (ICDM) 2007, Omaha NE (USA) ; C. Rouveirol
  - Comité de programme de International Conference on Inductive Logic Programming (ILP) 2007, Corvalis, OR (U.S.A.) ; C. Rouveirol
  - Comité de programme du 6th Workshop on Multi-Relational Data Mining (MRDM) 2007, workshop attaché à ECML/PKDD 2007, Warsaw, Poland ; C. Rouveirol
  - Comité de programme du workshop 'Machine Learning in Systems Biology', 2007, Evry ; C. Rouveirol
  - Comité de programme de la conférence d'Apprentissage (CAp) 2007, plateforme AFIA 2007, Grenoble ; Y. Bennani, C. Rouveirol
  - Comité de programme de l'atelier Bases de Données Inductives, plateforme AFIA 2007, Grenoble ; C. Rouveirol
  - Comité de programme de l'atelier 'Réseaux d'Interactions, Analyse, Modélisation, Simulation' (RIAMS) 2007, attaché à la conférence 'Integrative Post Genomics', Lyon ; C. Rouveirol
  - Comité de programme de European Conference on Machine Learning and European Conference on Principles and Practice of Knowledge Discovery (co-location) (ECML/PKDD 2006) ; C. Rouveirol, E. Alphonse
  - Comité de programme de The Second International Workshop on Mining Complex Data - MCD'06 - in Conjunction with IEEE ICDM'06, Hong Kong, 18 - 22 December 2006, The IEEE International Conference on Data Mining (ICDM) ; Y. Bennani
  - Comité de programme de la 7ème édition de la conférence francophone Extraction et Gestion des Connaissances (EGC'07), Namur, Belgique, 23-26 janvier 2007 ; Y. Bennani
  - Comité de programme de l'atelier «Fouille de Données et Algorithmes Biomimétiques» dans le cadre de la conférence Extraction et Gestion de Connaissances (EGC'07) ; Y. Bennani

- Comité de programme de l'atelier «Visualisation de données» dans le cadre de la conférence Extraction et Gestion de Connaissances (EGC'07) ; Y. Bennani
- Comité de programme de SIGMAP is part of ICETE - The International Joint Conference on e-Business and Telecommunications Conference, 7-10 August 2006 Setúbal, Portugal ; Y. Bennani
- Comité de programme de la Conférence francophone sur l'apprentissage automatique CAp 2006, 22-24 mai 2006 - Trégastel ; Y. Bennani
- Comité de programme de FW-EGC06 « Fouille du Web » ; Y. Bennani
- Comité de programme de la conférence internationale WSOM'05, Paris ; Y. Bennani
- Comité de programme The Seventh International Symposium On Programming and Systems, ISPS'05, May 9-11, 2005 Algiers ; Y. Bennani
- Comité de programme de ALAMAS'2005 et ALAMAS'2006 ; H. Soldano
- Comité de programme de GRES 2006 ; Y. Bennani
- Comité de programme de CLA'06 ; H. Soldano
- Comité de programme : PSACE'06, ateliers française sur le RàPC (04-06) ; R. Kanawati
- Comité de programme de CoSSoM'06, Complex Systems and Self-organization Modelling, A workshop for ESM'2006, October 23-25, 2006, LAAS, Toulouse, France ; Y. Bennani et H. Kadri
- Comité de programme des 2es Journées thématiques «Apprentissage Artificiel et Fouille de Données» (AAFD'06), 27-28 avril 2006, Université Paris 13, Institut Galilée ; E. Alphonse, Y. Bennani, D. Bouthinon, M. Champesme, P. Gérard, R. Kanawati A. Osmani, H. Soldano, E. Viennet
- Comité d'organisation CBRSEQ : International Workshop on Applying CBR to Time series prediction). Prmière édition à Trondheim, 2003 et deuxième édition à Madrid Septembre 2004 ; R. Kanawati avec S. Salotti et M. Malek
- Comité d'organisation de l'atelier française sur le raisonnement à partir de cas. Villetaneuse, Mars 2004 ; R. Kanawati avec S. Salotti, F. Zehroui
- Comité d'organisation de l'atelier «Fouille de Données et Algorithmes Biomimétiques» dans le cadre de la conférence Extraction et Gestion de Connaissances (EGC'07) ; H. Azzag
- Comité d'organisation de GRES 2006 ; L. Dehni, Z. El Jrad, H. Moungla
- Comité d'organisation des 2es Journées thématiques «Apprentissage Artificiel et Fouille de Données» (AAFD'06), 27-28 avril 2006, Université Paris 13, Institut Galilée ; Y. Bennani et E. Viennet
- Comité de lecture du Numéro spécial de la Revue des Nouvelles Technologies de l'Information « Fouille du Web » ; Y. Bennani
- Comité de lecture du Numéro spécial de la Revue des Nouvelles Technologies de l'Information (AAFD'06) ; Y. Bennani et E. Viennet

- **Expertises**

  - Expertise sur projets ANR (2005-2007) ; Y. Bennani
  - Expertise sur projets ACI Masses de Données (2004) ; Y. Bennani
  - Expertise sur projets ANR Blanc, Jeunes Chercheurs et Masses de Données (2007) ; C. Rouveirol
  - Expertise sur dossiers masters (2007) ; Y. Bennani

- **Jurys de thèses et d'HDR depuis 2003 (hors Paris 13)**

  - Jurys de thèse : Y. Bennani (25 dont 13 comme rapporteur), H. Soldano (1), R. Kanawati (2 à l'étranger dont 1 comme rapporteur), C. Rouveirol (2 dont 1 comme

rapporteur).
- Jury d'Habilitation à diriger des Recherches : Y. Bennani (4 dont 2 comme rapporteur).

1.6.5 Collaborations

- **Collaborations internationales**

  - University of Skopje (D. Cakmakov), Feature Selection for Knowledge Discovery and Data Mining ; Y. Bennani.
  - University of Skopje (D. Cakmakov, B. Jaksimovski), Combining and Optimizing Classifiers for Pattern Recognition Applications ; Y. Bennani.
  - Federal University of Ceará (J. Neuman de Souza, D. Gomes), Distributed Connectionist Learning for Wireless Sensor Networks ; Y. Bennani.
  - University of Torino, Machine Learning group, (Attilio Giordana) ; Y. Bennani, H. Azzag.
  - University of Torino, Machine Learning group, (Lorenza Saitta) ; Y. Bennani, C. Rouveirol.
  - Université Libre de Bruxelles, Centre Interdisciplinaire de Phénomènes Non-linéaires et de Systèmes Complexes, (Jean-Louis Deneubourg) ; H. Azzag, Y. Bennani.
  - Université Catholique de Leuven (Kurt Driessens) ; P. Gérard, Céline Rouveirol.
  - TU München - Fakultät für Informatik (Stefan Kramer) ; E. Alphonse et C. Rouveirol.

- **Collaborations industrielles (hors contrat)**

  - Convention Cifre avec la société Géobs (2003-2004, thèse de S. Guérif)
  - Convention Cifre avec la société NumSight (2004-2006, thèse de S. Guérif)
  - Société KXEN (E. Viennet, comité scientifique)

1.7 Publications principales

Livres

2006 [:2006fx] Younès Bennani, éditeur. *Apprentissage Connexionniste*. Hermès, 2006, ISBN 2-7462-1337-0.

Chapitres de livres

2007 [RK07] Rushed Kanawati. *Agents RàPC coopératifs pour la recherche d'information*, Jean Renaud, Brigitte Chebel Morello, Béatrice Fuchs et Jean Lieber, éditeurs, *Raisonnement à partir de cas 2, surveillance, diagnostic et maintenance*, volume 2 de *Informatique et système d'information*, Chapitre 6, pages 177–195. Hermès, 2007.

[zehraoui07] Farida Zehraoui, Rushed Kanawati et Sylvie Salotti. *CASEP2 : Système Hybride pour le traitement de séquences de navigation sur le Web*, Jean Renaud, Brigitte Cheble-Morello, Bétarice Fuchs et Jean Lieber, éditeurs, *Raisonnement à Partir de Cas 2 : surveillance, diagnostic et maintenance*, volume 2 de *Informatique et système d'information*, pages 151–176. Hermès, 2007.

2006 [Pisanti :2006li] Nadia Pisanti, Henry Soldano, Mathilde Carpentier et Joël Pothier. Implicit and explicit representation of approximated motifs. C. S. Iliopoulos, K. Park et K. Steinhoefel, éditeurs, *Algorithms in Bioinformatics*, pages 1–15. College Press, 2006.

Articles dans des revues internationales avec comité de lecture sélectif

2007 [RI-2*] Erick Alphonse et Aomar Osmani. On the connection between the phase transition of the covering test and the learning success rate in ilp. *Machine Learning Journal*, 2007.

2005 [Zehraoui :2005wy] Farida Zehraoui et Younès Bennani. New self-organising maps for multivariate sequences processing. *International Journal of Computational Intelligence and Applications*, 5(4) :439–456, 2005.

Articles dans des revues nationales ou internationales

2006 [HKP06LP-SGAI-06] H. Karoui, R. Kanawati et L. Petrucci. An intelligent peer-to-peer multiagent system for collaborative management of bibliographic databases. *SGAI's expert update magazine*, 8(3) :22–31, printemps 2006.

2005 [Ventos :2005vp] Véronique Ventos et Henry Soldano. Treillis de Galois alpha. *Revue d'Intelligence Artificielle RSTI série RIA*, 19(4-5) :799–227, 2005.

Communications dans des conférences internationales avec comité de lecture

2007 [Azzag2007j] Hanane Azzag, Younès Bennani et Mustapha Lebbah. A stochastic algorithm for clustering inspired from cockroaches behavior. *V International Conference on Microelectronics and Computer Science*, 2007.

[cabanes :2007] Guénaël Cabanes et Younès Bennani. A simultaneous two-level clustering algorithm for automatic model selection. *(ICMLA'07), International Conference on Machine Learning and Applications*, Cincinnati, Ohio, USA, décembre 2007.

[Azzag2007a] Julien Lavergne, Hanane Azzag, Christiane Guinot et Gilles Venturini. Construction incrémentale et visualisation de graphes de voisinage par des fourmis artificielles. Monique Noirhomme-Fraiture et Gilles Venturini, éditeurs, *Extraction et gestion des connaissances (EGC'2007)*, volume I de *Revue des Nouvelles Technologies de l'Information (RNTI-E-9)*, pages 135–146, Namur, Belgique, Janvier 2007. Cépadués Éditions.

2006 [Alphonse2006b] Erick Alphonse et Céline Rouveirol. Extension of the top-down data-driven strategy to ILP. *16th International Conference on Inductive Logic Programming (ILP 2006)*, /LNAI 4455/, pages /49–63/, 2006.

[Guerif :2006xz] Sébastien Guérif et Younès Bennani. Selection of clusters number and features subset during a two-levels clustering task. *Proceeding of the 10th International Conference Artificial intelligence and Soft Computing 2006*, pages 28–33, Palma de Mallorca, Spain, août 2006.

[karoui06b] Hager Karoui, Rushed Kanawati et Laure Petrucci. COBRAS : Cooperative CBR System for Bibliographical Reference Recommendation. *8th European Conference on Case-Based Reasonning (ECCBR'06)*, volume 4106 de *LNCS*, pages 76–90, Fethiye, Turkey, 2006. springer.

2004 [MMaRK05] Maria Malek et Rushed Kanawati. A Data Driven CBR Approach : A Contineous Maintenance Strategy. *International Conference on Advances in Intelligent Systems : Theory and Applications*, page 6. IEEE, Novembre 2004.

[Zehraoui :2004qb] Farida Zehraoui et Younès Bennani. M-SOM-ART : Growing self organizing map for sequence clustering and classification. *ECAI'04, European Conference on Artificial Intelligence*, 2004.

2003 [Bennaceur :2003mw] Hachemi Bennaceur et Aomar Osmani. Computing lower bound for MAX-CSP problems. *International Conference on Industrial, Engineering Applications of Artificial Intelligence and Expert Systems*, Lecture Notes in Artificial Intelligence, pages 614–624. Springer, 2003.

[kanawati03b] Rushed Kanawati. A Multi-agent System for Improving Result Ranking Service of Web Search Engines. *10th ISPE International conference on concurrent engineering (CE 2003)*, volume 1, pages 91–99, Funchal, Madeira, June 2003. Balkema.

[Zeboulon :2003kv] Arnaud Zeboulon, Younès Bennani et Khalid Benabdeslem. Hybrid connectionist approach for knowledge discovery from web navigation patterns. *ACS/IEEE International Conference on Computer Systems and Applications*, juillet 2003.

[kanawati07b] Rushed Kanawati et Maria malek. Computing Social Networks for Information Sharing : A Case-based Approach, volume = 4564, year = 2007. Schuler Douglas, éditeur, *Second international conference on Online Communities and Social Computing*, LNCS, pages 86–95, Beijing, China, July. Springer.

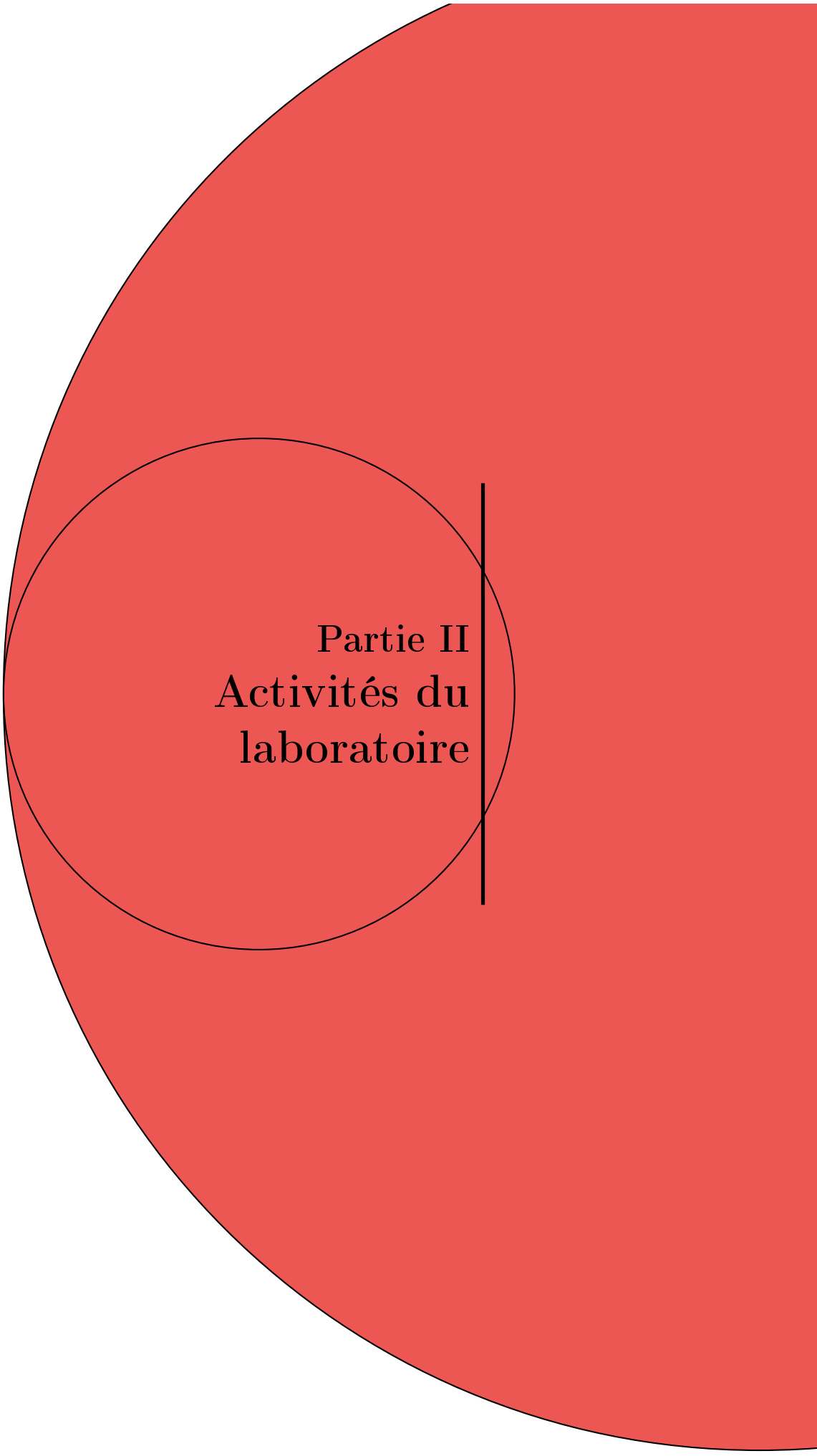
Workshops et autres conférences avec comité de lecture

2006 [karoui06d] Hager Karoui, Rushed Kanawati et Laure Petrucci. Cooperative CBR system for peer agent committee formation. *Proc. 5th Workshop on Agents and Peer-to-Peer Computing (AP2PC'06)*, pages 76–90, Hakodate, Japan, mai 2006.

2005 [Guerif :2005ya] Sébastien Guérif, Younès Bennani et Eric Janvier. μ-SOM : Weighting features during clustering. *Proceedings of the 5th Workshop On Self-Organizing Maps (WSOM'05)*, pages 397–404, Paris, France, septembre 2005.

2004 [Guerra-Hernandez :2004xi] Alejandro Guerra-Hernandez, Amal El Fallah-Seghrouchni et Henry Soldano. Learning in BDI multi-agent systems. *Fourth International Workshop on Computational Logic in Multi-Agent Systems (CLIMA IV)*, volume 3259 de *Lecture Notes in Computer Science*, pages 218–233. Springer, 2004.

[Zehraoui :2004xc] Farida Zehraoui et Younès Bennani. SOM-ART : Incorporation des propriétés de plasticité et de stabilité dans une carte auto-organisatrice. *Atelier FDC : Fouille de Données Complexes dans un processus d'extraction de connaissances (EGC 2004)*, pages 169–180, janvier 2004.



Partie II
Activités du laboratoire

Chapitre 1
Formation par la recherche

1.1 Master Informatique

Le Master (Sciences, Technologie, Santé) **mention Informatique** offre deux spécialités de formation qui s'appuient notamment sur des axes forts et des compétences bien développées au LIPN :

une spécialité à finalité professionnelle : Exploration Informatique des Données (EID)

une spécialité à finalité recherche : Modélisation Informatique des Connaissances et du Raisonnement (MICR)

La spécialité EID a pour objet l'acquisition de connaissances approfondies dans des domaines particuliers complémentaires : l'informatique et l'extraction de connaissances à partir de données. Cette spécialité est fondée sur l'acquisition de techniques destinées à favoriser l'exercice d'un métier déterminé dans le domaine de l'exploration des données.

La spécialité MICR a pour objectif de fournir aux étudiants les éléments fondamentaux leur permettant d'entamer une recherche tant en intelligence artificielle qu'en traite-

ment automatique du langage, en spécifications de systèmes complexes ou en théorie de la programmation. Dans cette spécialité, plusieurs parcours-types sont possibles, on peut cependant en préfigurer quelques uns :

- Parcours « Traitement Automatique des Langues Naturelles » : Traitement automatique des langues, Ingénierie des connaissances textuelles, Recherche et extraction d 'information, Ressources lexicales ;
- Parcours « Spécifications et Programmation » : Systèmes parallèles asynchrones, Systèmes infinis, temporisés et hybrides, Programmes et preuves, Unité d'approfondissement au choix ;
- Parcours « Logique et Programmation » : Calculs de processus, Programmes et preuves, Concurrence, Unité d'approfondissement au choix ;
- Parcours « Apprentissage Artificiel et Fouille de Données » : Apprentissage numérique, Apprentissage symbolique et RaPC, Apprentissage statistique, Fouille de données relationnelles.

Ce master est ouvert à la pluridisciplinarité qui est un des atouts de l'université Paris 13, avec par exemple une UE de linguistique sur les ressources lexicales et une UE sur la bio-informatique.

Formation par la recherche

1.2 Master Mathématique Informatique

Le Master (Sciences, Technologie, Santé) **mention Mathématique et Informatique** a pour objectif de fournir aux étudiants les éléments fondamentaux leur permettant d'acquérir une véritable double compétence en Informatique et en Mathématiques et leurs applications. Elle offre l'accès à des cours très complémentaires dans les domaines de l'informatique, des mathématiques discrètes et continues, des phénomènes aléatoires et de la combinatoire. Elle permet, outre les aspects théoriques, de toucher aux problèmes applicatifs sources de modèles complexes. La formation s'appuie à la fois sur les laboratoire LIPN et LAGA de l'université Paris 13.

Outre les thèses possibles dans les laboratoires, en particulier au LAGA ou au LIPN, les étudiants peuvent se tourner vers les activités de recherche et développement en entreprise. Cette seconde voie n'est pas à négliger, les entreprises devenant un endroit où les compétences solides en mathématiques et informatique appliquées, attestées par une formation par la recherche, sont de plus en plus recherchées. Le master mathématiques et informatique vise à donner au travers de la formation le bagage indispensable d'un bon Ingénieur en Recherche et Développement ou d'un Consultant auprès d'entreprises de divers secteurs de pointe.

La première année est consacrée aux apprentissages des outils de base en mathématiques appliquées et en informatique. En fin d'année, un travail personnel d'étude et de recherche mené en coordination avec l'équipe pédagogique doit permettre à l'étudiant de mettre par lui-même en application les connaissances acquises pendant l'année. La seconde année est l'occasion d'approfondissement des connaissances en choisissant parmi de nombreux sujets : résolution numérique d'équations aux dérivées partielles, recherche opérationnelle, combinatoire, mathématiques financières.

1.3 Thèses et Habilitations

Thèses en cours

Nom Prénom	1ère insc.	Situation
Titre	*Directeur*	*Co-encadrant*

BENCHETTARA Hamimed	12/01/07	sans
Prédiction de liens dans les réseaux sociaux	*C. Rouveirol*	*R. Kanawati*
RODRIGUES BENTO Christophe	10/01/07	sans
Apprentissage par renforcement relationnel	*C. Rouveirol*	*P. Gérard*
CABANES Guénaël	10/01/07	Alloc (07-10)
Apprentissage non supervisé à base de noyaux pour la classification topographique évolutive	*Y. Bennani*	
GROZAVU Nistor	11/01/06	contrat Cap Digital
Réduction de dimensions en apprentissage non supervisé	*Y. Bennani*	
ROGOVSCHI Nicoleta	11/01/06	contrat Septia
Apprentissage numérique et réseaux sociaux	*Y. Bennani*	*E. Viennet*
DEHNI Lahcène	10/01/04	(salarié m-temps) ; ANR
Apprentissage connexionniste non supervisé sous contraintes et réseaux de capteurs	*Y. Bennani*	

Thèses soutenues (depuis 2004)

Nom Prénom	Situation ultérieure	Soutenance	Mention
Titre		*Directeur*	*Co-directeur*
Jury : Jury			
ELATI Mohamed		12/18/07	
Apprentissage de réseaux de régulation génétique à partir de données d'expression		*C. Rouveirol*	
Jury : Jean-François Boulicaut (rapp), Jean-Daniel Zucker (rapp), Florence d'Alché-Buc, Younès Bennani, Amedio Napoli, François Radvanyi (invité)			
ASEERVATHAM Sujeevan		12/12/07	THF
Apprentissage à base de Noyaux Sémantiques pour le Traitement de données textuelles		*Y. Bennani*	*E. Viennet*
Jury : Massih-Reza Amini, Christophe Fouqueré (pdt), Cyril Goutte, Jean-François Marcotorchino, Alain Rakotomamonjy, Michèle Sebag			
KAROUI Hager		12/11/07	TH
Système coopératif de type égal-à-égal pour la recommandation : Application à la gestion et la recommandation de références bibliographiques		*L. Petrucci*	*R. Kanawati*
Jury : Jean-Paul Barthès, Catherine Berrut, Sylvie Després, Alain Mille, Enric Plaza, Laurence Vignollet			
GUERIF Sébastien	MCF P13	12/11/06	
Approches connexionnistes pour l'exploration et l'extraction de connaissances à partir de données spatio-temporelles		*Y. Bennani*	
Jury : C. Bertelle (rapp), G. Venturini (rapp), P. Kuntz, M. Magnusson, JD Zucker, E. Ecosse (invité), E. Janvier (invité)			
AMARA - HACHMI Nejla	post-doc FT	12/06/06	

GAMA : une architecture générique pour le développement d'agents mobiles sensibles au contexte et auto-adaptatifs — *A. El Fallah*

Jury : Olivier Boissier, Salima Hassas, Adeline Nazarenko, Henry Soldano

MOUNGLA Hassine — Post-doc INRIA — 11/20/06

Supervision distribuée de la QoS dans les réseaux IP contrôlés par politique des opérateurs de telecommunications — *F. Krief*

Jury : A. Karmouch, N Agoulmine, M. Salaün, Y Bennani, K Chen

PAUCHET Alexandre — post-doc FT R&D Lannion ; MDC Rouen — 09/22/06 — TH

Modélisation cognitive d'interactions humaines dans un cadre de planification multi-agents — *A. El Fallah* — *Jean-Paul Sansonnet (Rapp)*

Jury : Sylvie Pesty (Rapp), D. Kayser, H. Soldano, N. Chaignaud

EL FERKH - JRAD Zeina — post-doc Inria (Axis) — 05/15/06 — TH

Apport des techniques de l'Intelligence Artificielle dans la négociation dynamique des niveaux de service : proposition d'une interface utilisateur pour l'Internet de nouvelle génération — *F. Krief*

Jury :

EL ZANT - KADHI Nahla — 12/01/05 — TH

Apprentissage de motifs relationnels dans des objets partiellement ordonnés et application à la caractérisation de motifs dans l'ADN — *D. Kayser* — *H. Soldano*

Jury :

KADRI-DAHMANI Hakima — post-doc IGN Belgique — 11/22/05 — TH

Méthodes de modélisation pour la gestion de la notion de cycle de vie dans les systèmes d'information géographique — *P. Dague* — *A. Osmani*

Jury : Younès Bennani , Cyrille Bertelle, Philippe De Mayer, Gérard Duchamp, Anne Ruas, Jan Beyen

VAUVERT Guillaume — Post-doc INRIA-LIP6 — 01/15/04 — TH

Atteinte de consensus basé sur l'échange de préférences entre agents rationnels et autonomes : application à la formation d'alliances — *A. El Fallah-Seghrouchni*

Jury : Philippe Dague, Richard Emilion (rapp), Suzanne Pinson (rapp), Henry Soldano (invité), Patrick Taillibert, Jean-Daniel Zucker

ZEHRAOUI Farida — MCF Evry (LAMI) — 10/12/04 — THF

Systèmes d'apprentissage connexionniste et raisonnement à partir de cas pour la classification et le classement des séquences — *Y. Bennani*

Jury : Philippe Dague, Maurice Milgram, Alain Mille (rapp), Nicolas Nicoloyannis (rapp), Sylvie Salotti, Rushed Kanawati (invité), Eric Janvier (invité), Françoise Fessant (invité)

Habilitations à Diriger des Recherches soutenues (depuis 2004)

Nom Prénom — Soutenance

Titre

Jury :

Viennet Emmanuel 12/18/07

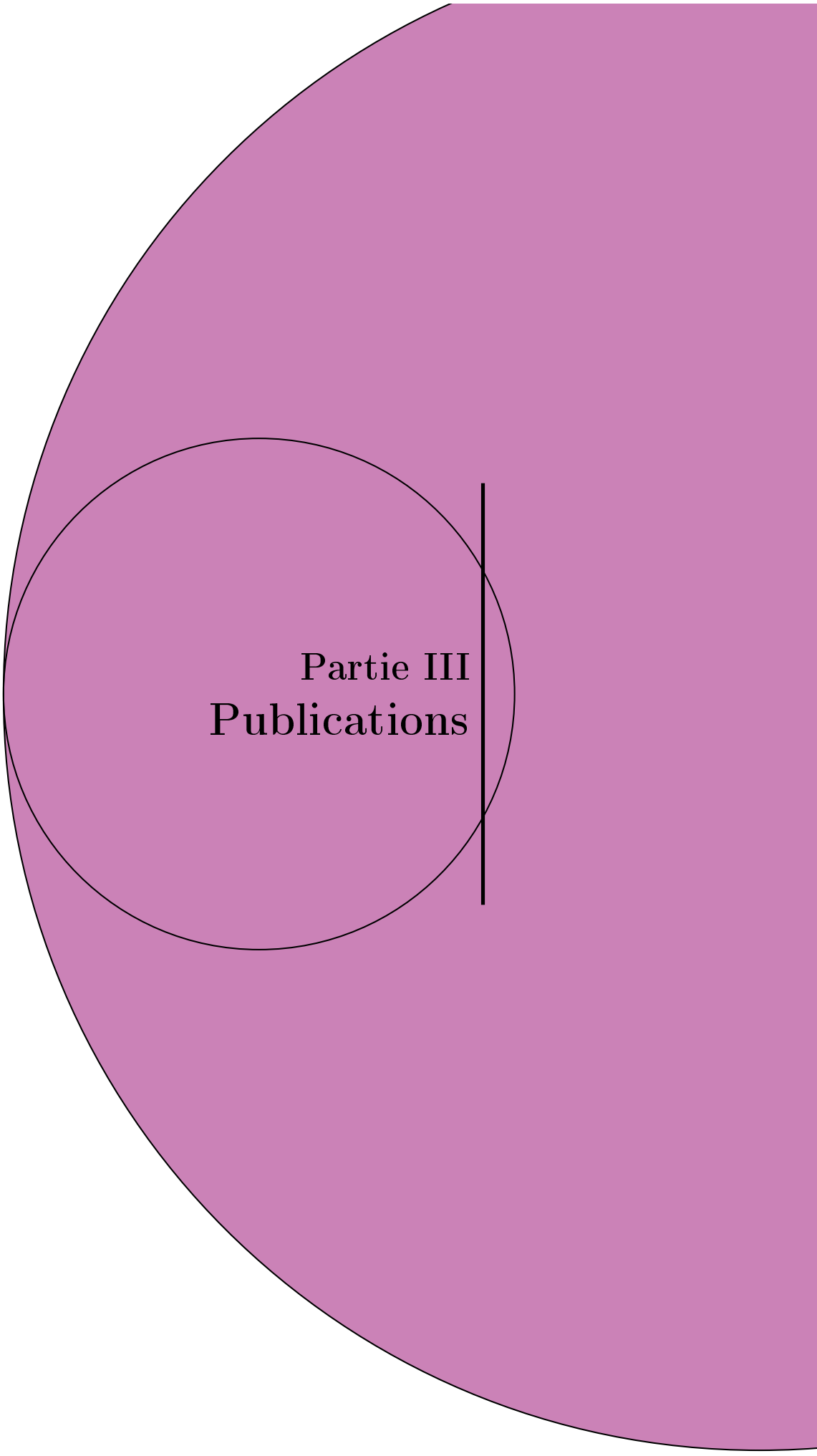
Contributions aux méthodes d'apprentissage numérique pour la fouille de données structurées

Jury : Stéphane Canu (rapp), Pascale Kuntz (rapp), Gilbert Ritschard (rapp), Younès Bennani, Françoise-Fogelman-Soulié, Christophe Fouqueré, Gilbert Saporta

Formation par la recherche

Chapitre 2
Organisation de conférences

Le LIPN a soutenu ou organisé les conférences et manifestations suivantes :
- *Implicit Computational Complexity and Logic*, septembre 2004, université Paris 13 (ACI GEOCAL), 30 participants, dont 8 étrangers.
- *Workshop on Implicit Computational Complexity*, 13-17/2/2006, (GEOCAL, *Geometry of Computation*, 2006), Marseille. 20 exposés et 50 participants (de 7 pays).
- *2es Journées thématiques "Apprentissage Artificiel & Fouille de Données"*, 27-28/4/2006, http://www-lipn.univ-paris13.fr/A3/AAFD06/
- *Journées ALEA*, 2005, (CIRM, SMF-CNRS, Luminy)
- *Marches aléatoires discrètes : théorie et applications en combinatoire, biologie algorithmique, informatique théorique, probabilités, physique statistique* (DRW-2003), Institut Henri Poincaré, 2003



Partie III
Publications

Les références comportant « * » sont acceptées et à paraître.

Articles dans des revues avec comité de lecture (ACL)

Articles dans des revues internationales avec comité de lecture sélectif

2007 [RI-11*] Erick Alphonse et Aomar Osmani. On the connection between the phase transition of the covering test and the learning success rate in ilp. *Machine Learning Journal*, 2007.

[RI-10] Danielo Gomes, Nazim Agoulmine, Younès Bennani et Neuman De Souza. Predictive connectionist approach for vod bandwidth management. *International Journal for Computer Communications*, 30(10) :2236–2247, 2007.

[RI-9*] Hakima Kadri-Dahmani, Cyrille Bertelle, Gérard Duchamp et Aomar Osmani. Consistent updating of geographical database as emergent property over influence system. *International Journal of Modelling Identification and Control, special Issue Modelling Complex Systems*, 2007.

[RI-8*] Fabien Picarougne, Hanane Azzag, Gilles Venturini et Christiane Guinot. A new approach of data clustering using a flock of agents. *Evolutionary Computation*, 2007.

[RI-7] Arthur Tenenhaus, Alain Giron, Emmanuel Viennet, Michel Béra, Gilbert Saporta et Bernard Fertil. Kernel logistic PLS : a tool for supervised nonlinear dimensionality reduction and binary classification. *Computational Statistics & Data Analysis*, 51(9) :4083–4100, mai 2007.

2006 [RI-6] Khalid Benabdeslem et Younès Bennani. Dendogram-based svm for multi-class classification. *International Journal of Computing and Information Technology*, 14(4) :288–291, 2006.

2005 [RI-5] Alejandro Guerra-Hernández, Amal El Fallah-Seghrouchni et Henry Soldano. On learning intentionality. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, 9(25) :9–18, 2005. Première version publiée dans Proceedings of IberAgents 04 Workshop.

[RI-4] Farida Zehraoui et Younès Bennani. New self-organising maps for multivariate sequences processing. *International Journal of Computational Intelligence and Applications*, 5(4) :439–456, 2005.

2004 [RI-3] Nahla El-Zant et Henry Soldano. Finding repeated flexible relational words in sequences. *Journal of Systemics, Cybernetics and Informatics*, 2(4), 2004.

[RI-2] Francine Krief. Self-aware management of IP networks with QoS guarantees. *International Journal on Network Management*, 2004.

2003 [RI-1] Louise Travé-Massuyès, L. Ironi et Philippe Dague. Mathematical foundations of qualitative reasoning. *AI Magazine*, 24(4) :91–106, 2003.

Articles dans des revues nationales ou internationales

2007 [RE-11*] Sujeevan Aseervatham et Emmanuel Viennet. Méthodes à noyaux appliquées aux textes structurés. *Revue des Nouvelles Technologies de l'Information* [ED-6], pages 185–205. 230 pages, publication en cours.

[RE-10*] Hanane Azzag, David Da Costa, Christiane Guinot et Gilles Venturini. Un aperçu de la fouille visuelle de données. *Revue des Nouvelles technologies de l'Information (RNTI)*, 2007. Journées AAFD-2006 (Apprentissage Artificiel et Fouille de données).

[RE-9] Khalid Benabdeslem et Younès Bennani. Approche connexionniste pour l'analyse des données issues d'usage d'internet : Classification et visualisation. *Revue des Nouvelles Technologies de l'Information*, 2007.

[RE-8*] Younès Bennani, Sébastien Guérif et Emmanuel Viennet. Réduction des dimensions de l'espace de description. *Revue des Nouvelles Technologies de l'Information*, 2007.

[RE-7*] Emmanuel Viennet et Françoise Fogelman Soulié. Le traitement des refusés dans le risque crédit. *Revue des Nouvelles Technologies de l'Information*, 2007.

2006 [RE-6] Jérôme Azé, Mathieu Roche, Erik Alphonse, Ahmed Amrani, Thomas Heitz et Amar-Djallil Mezaour. Défi Fouille de Textes : reconnaissance automatique des auteurs de discours - Campagne DEFT'05 (TALN'05). *Revue des Nouvelles technologies de l'Information (RNTI)*, 2006.

[RE-5] Rushed Kanawati. Refléxions sur l'apport de l'exploration des trace d'usage pour améliorer le tri de résultats des moteurs de recherche sur le web. pages 150–154, Janvier 2006.

[RE-4] H. Karoui, R. Kanawati et L. Petrucci. An intelligent peer-to-peer multi-agent system for collaborative management of bibliographic databases. *SGAI's expert update magazine*, 8(3) :22–31, printemps 2006.

2005 [RE-3] Véronique Ventos et Henry Soldano. Treillis de Galois alpha. *Revue d'Intelligence Artificielle RSTI série RIA*, 19(4-5) :799–227, 2005.

2004 [RE-2] Philippe Dague et Louise Travé-Massuyès. Raisonnement causal en physique qualitative. *Numéro spécial Raisonnement Causal, Intellectica*, 1(38), novembre 2004.

[RE-1] Olivier Sigaud et Pierre Gérard. Apprentissage par renforcement indirect dans les systèmes de classeurs. *JEDAI*, 2004.

Autres communications sur invitation

2007 [AI-1] M. Elati et C. Rouveirol. Génération d'associations candidates sous contrainte de support adaptative. *BDI'2007 : atelier bases de données inductives, Plate-forme AFIA*, 2007.

Communications avec actes (ACT)

Communications dans des conférences internationales avec comité de lecture

[CI-68*] Erick Alphonse et Aomar Osmani. Phase transition and heuristic search in relational learning. *Sixth Int. Conference on Machine Learning and Applications (ICMLA 2007)*. IEEE press, 2007.

[CI-67] Sujeevan Aseervatham, Emmanuel Viennet et Younès Bennani. A semantic kernel for semi-structured documents. *ICDM 2007*, octobre 2007. International Conference on Data Mining (ICDM'07).

[CI-66] Jérôme Azé, Lucie Gentils, Claire Toffano-Nioche, Valentin Loux, Jean-François Gibrat, Philippe Bessières, Céline Rouveirol, Anne Poupon et Christine Froidevaux. Towards a semi-automatic functional annotation tool based on decision tree techniques. *International Workshop on Machine Learning in Systems Biology (MLSB 2007)*, 2007.

[CI-65] Hanane Azzag, Younès Bennani et Mustapha Lebbah. A stochastic algorithm for clustering inspired from cockroaches behavior. *V International Conference on Microelectronics and Computer Science*, 2007.

[CI-64] Hanane Azzag, Julien Lavergne, Christiane Guinot et Gilles Venturini. On building graphs of documents with artificial ants. *16th International World Wide Web Conference (WWW2007)*, pages 1299–1301, Banff, Alberta, Canada, May, 8-12th 2007.

[CI-63] Etienne Birmlé, Mohamed Elati, Céline Rouveirol et Christophe Ambroise. Identification of functional modules based on transcriptional regulation structure. *International Workshop on Machine Learning in Systems Biology (MLSB 2007)*, 2007.

[CI-62] Khellaf Bouandas et Aomar Osmani. Mining association rules in temporal sequences. *IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2007, part of the IEEE Symposium Series on Computational Intelligence 2007*, pages 610–615, 2007.

[CI-61] Gauvain Bourgne, Henry Soldano et Amal ElFallah-Seghrouchni. Smile : Sound multi-agent incremental learning. *International Conference on Autonomous Agents and Multiagent Systems*, pages 164–171, Honolulu, Hawai, 5 2007. ACM.

[CI-60] Guénaël Cabanes et Younès Bennani. A simultaneous two-level clustering algorithm for automatic model selection. *(ICMLA'07), International Conference on Machine Learning and Applications*, Cincinnati, Ohio, USA, décembre 2007.

[CI-59] Nistor Grozavu, Younès Bennani et Sébastien Guérif. Towards a dimensionality reduction through unsupervised learning and local variable weighting. *ICMCS'07*, Chisinau, Moldavie., September 19-21, 2007 2007.

[CI-58] Sébastien Guérif et Younès Bennani. Dimensionality Reduction Through Unsupervised Features Selection. *Proceedings of the 10th International Conference on Engineering Applications of Neural Networks (EANN 2007)*, pages 98–106, Thessaloniki, Greece, aug 2007. Publishing Centre Alexander T.E.I. of Thessaloniki.

[CI-57] Julien Lavergne, Hanane Azzag, Christiane Guinot et Gilles Venturini. Construction incrémentale et visualisation de graphes de voisinage par des fourmis artificielles. Monique Noirhomme-Fraiture et Gilles Venturini, éditeurs, *Extraction et gestion des connaissances (EGC'2007)*, volume I de *Revue des Nouvelles Technologies de l'Information (RNTI-E-9)*, pages 135–146, Namur, Belgique, Janvier 2007. Cépadués Éditions.

[CI-56*] Julien Lavergne, Hanane Azzag, Christiane Guinot et Gilles Venturini. Incremental construction of neighborhood gra-vphs using the ants self-assembly behavior. *the annual IEEE International Conference on Tools with Artificial Intelligence (ICTAI'2007)*, Patras, Greece, October 29-31 2007. the Conference and Cultural Center of the University of Patras, accepted full paper.

[CI-55] Julien Lavergne, Hanane Azzag, Christiane Guinot et Gilles Venturini. Une approche incrémentale d'une méthode de classification non supervisée par nuages d'insectes volants. *XIVe Rencontre de la Société Francophone de Classification (SFC'2007)*, Paris, France, 5-7 septembre 2007. ENST.

[CI-54] Mustapha Lebbah, Nicoleta Rogovschi et Younés Bennani. BeSOM : Bernoulli on self organizing map. *International Joint Conferences on Neural Networks. IJCNN 2007, Orlando, Florida, August 12-17*, 2007.

[CI-53] Mustapha Lebbah, Nicoleta Rogovschi et Younés Bennani. BeSOM : Bernoulli on self organizing map. *Cap'2007 : conférence francophone sur l'apprentissage automatique, Plate-forme AFIA*, 2007.

[CI-52*] Catherine Recanati, Nicoleta Rogovschi et Younès Bennani. Sequencing of verbs - a study on tense and aspect using unsupervised learning. Borovets, Bulgary, Sep. 2007. Institute for Parallel Processing, Bulgarian Academy of Sciences.

[CI-51*] Catherine Recanati, Nicoleta Rogovschi et Younès Bennani. The structure of verbal sequences analyzed with unsupervised learning techniques. Poznań, Poland, Oct. 2007.

[CI-50] Nicoleta Rogovschi et Emmanuel Viennet. Finding community structure in social networks using statistical approaches. *V International Conference on Microelectronics and Computer Science*, septembre 2007.

2006

[CI-49] Erick Alphonse et Aomar Osmani. On the connection between the phase transition of the covering test and the learning success rate. *16th International Conference on Inductive Logic Programming (ILP 2006)*, pages 27–29, 2006.

[CI-48] Erick Alphonse et Céline Rouveirol. Extension of the top-down data-driven strategy to ILP. *16th International Conference on Inductive Logic Programming (ILP 2006)*, /LNAI 4455/, pages /49–63/, 2006.

[CI-47] Sujeevan Aseervatham, Aomar Osmani et Emmanuel Viennet. BitSPADE : A lattice-based sequential pattern mining algorithm using bitmap representation. *Proceedings of the 2006 IEEE International Conference on Data Mining.*, pages 792–797, décembre 2006.

[CI-46] Khalid Benabdeslem et Younès Bennani. Dendogram based SVM for multi-class classification. *28th International Conference Information Technology Interfaces*, pages 173–178, Cavtat / Dubrovnik, Croatia., juin 2006.

[CI-45] Hakima Kadri Dahmani, Gérard Duchamp, Rawan Ghnemat, Hatem Hadj Kacem et Cyril Bertelle. Emerging decision support systems for geographical informations systems. *ICELM2*, 2006.

[CI-44] Sébastien Guérif et Younès Bennani. Selection of clusters number and features subset during a two-levels clustering task. *Proceeding of the 10th International Conference Artificial intelligence and Soft Computing 2006*, pages 28–33, Palma de Mallorca, Spain, août 2006.

[CI-43] Sébastien Guérif, Younès Bennani et Claude Baudoin. Connectionist and ethological approaches for discovering salient facial movement features in human gender recognition. *Proceeding of the 28th International Conference Information Technology Interfaces*, pages 189–194, Cavtat, Croatia, juin 2006.

[CI-42] Nejla Amara Hachmi. An ontology-based model for mobile agents adaptation in pervasive environments. *4th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA'06)*, pages 1106–1109, mars 2006.

[CI-41] Zeina Jrad, Francine Krief, Lahcène Dehni et Younès Bennani. Artificial Intelligence techniques in the dynamic negotiation of QoS : A user interface for the Internet New Generation. *International Federation for Information Processing (IFIP) TC6 Conference Smartnet in Autonomic Networking*, septembre 2006.

[CI-40] Zeina Jrad, Francine Krief et Rim Hammi. Video applications quality improvement in wireless systems : QoS negotiation and rate control. *International Conference on Digital Telecommunications ICDT'06*, France, septembre 2006.

[CI-39] Hager Karoui, Rushed Kanawati et Laure Petrucci. COBRAS : Cooperative CBR System for Bibliographical Reference Recommendation. *8th European Conference on Case-Based Reasonning (ECCBR'06)*, volume 4106 de *LNCS*, pages 76–90, Fethiye, Turkey, 2006. springer.

[CI-38] Maria Malek et Rushed Kanawati. CBR-based sequence analysis : An overview. *Case-Based Reasoning in Knowledge Discovery and Data Mining*, juin 2006.

[CI-37] Adeline Nazarenko, Erick Alphonse, Julien Derivière, Thierry Hamon, Guillaume Vauvert et Davy Weissenbacher. The ALVIS format for linguistically annotated documents. *Proceedings of the 5th international conference on Language Resources and Evaluation, LREC 2006*, pages 1782–1786. ELDA, 2006.

2005 [CI-36] Hakima Kadri Dahmani, Cyril Bertelle, Gérard Duchamp et Aomar Osmani. Evolutive complex scheduling in interaction networks for quality improvement in geographical data base updating. *MCS 2005 - Modeling, Computation and Systems - in IEEE-ICECS*, 2005.

[CI-35] Lahcène Dehni, Francine Krief et Younès Bennani. Power control and clustering in wireless sensor networks. *IFIP Med-Hoc-Net 2005 : Mediterranean Ad Hoc Networking Conference*, pages 31–40, juin 2005.

[CI-34] Zeina Jrad, Badr Benmammar, Joeseph Correa, Francine Krief et Nadek Mbarek. A user assistant for QoS negotiation in a dynamic environment using agent technology. *Second IFIP International Conference on Wireless and Optical Communications Networks WOCN*, mars 2005.

[CI-33] Maria Malek et Rushed Kanawati. A data-driven CBR approach : A contineous maintenance strategy. *Proceedings of the first International conference on Advances in Intelligent systems : Theory and applications (AISTA'05)*, Luxembourg, décembre 2005.

[CI-32] Hassine Moungla et Francine Krief. Conflict resolution in policy based management. *The 16th Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, septembre 2005.

[CI-31] Alexandre Pauchet, Nathalie Chaignaud et Amal El Fallah-Seghrouchni. A cognitive model of interaction for software agents. *IEEE IAT - International Conference on Intelligent Agent Technology*, pages 281–284, septembre 2005.

[CI-30] Nadia Pisanti, Henry Soldano et Mathilde Carpentier. Incremental Inference of Relational Motifs with a Degenerate Alphabet. *16th Annual Symposium on Combinatorial Pattern Matching (CPM)*, volume 3537, pages 229–240, 2005.

[CI-29] Véronique Ventos et Henry Soldano. Alpha Galois lattices : An overview. *Proceedings of ICFCA'05, Lecture Notes on Computer Science*, 3403 :298–313, 2005.

2004 [CI-28] Nejla Amara-Hachmi. A generic component-based methodology for extensible mobile agents design. *IEEE International Conference on Advances in Intelligent Systems : Theory and Applications AISTA'04*, novembre 2004.

[CI-27] Nejla Amara-Hachmi, Gilles Klein et Amal El Fallah-Seghrouchni. Towards MAS's design regarding communications and mobility. *Turkish Symposium on Artificial Intelligence and Neural Networks TAINN'04*, juin 2004.

[CI-26] Khalid Benabdeslem et Younès Bennani. An incremental SOM for web navigation patterns clustering. *26th International Conference Information Technology Interfaces ITI 2004*, juin 2004.

[CI-25] Badr Benmammar et Francine Krief. Agents for wireless environments. *International Conference on Telecommunication Systems, Modeling and Analysis (ICTSM 2004. IFIP WG 7.3)*, juillet 2004.

[CI-24] Hakima Kadri Dahmani et Aomar Osmani. How to update databases in a consistent way ? *UDMS 2004 : 24th Urban Data Management Symposium*, 2004.

[CI-23] Alejandro Guerra-Hernández, Amal El Fallah-Seghrouchni et Henry Soldano. Distributed learning in intentional BDI multi-agent systems. *5th Mexican International Conference on Computer Science (ENC 2004)*, pages 225–232. IEEE Computer Society, 2004.

[CI-22] Zeina Jrad et Francine Krief. An intelligent interface for the dynamic negotiation of QoS. *ARCADE. ACS/IEEE International Conference on Pervasive Services. ICPS'2004*, juillet 2004.

[CI-21] Francine Krief et Dominique Bouthinon. A learning and intentional local policy decision point for dynamic QoS provisioning. *IFIP TC6, Third International Conference on Network Control and Engineering for QoS, Security and Mobility, NetCon 2004*, pages 277–288, 2004.

[CI-20] Maria Malek et Rushed Kanawati. A Data Driven CBR Approach : A Contineous Maintenance Strategy. *International Conference on Advances in Intelligent Systems : Theory and Applications*, page 6. IEEE, Novembre 2004.

[CI-19] Hassine Moungla et Francine Krief. Service differentiation over GMPLS. *IEEE conference/ ICT'2004. Lecture Notes on Computer Science*, août 2004.

[CI-18] Véronique Ventos, Henry Soldano et Thibaut Lamadon. Alpha Galois lattices. *Int. Conf. on Data Mining (ICDM'04)*, pages 555–558. IEEE Press, 2004.

[CI-17] Farida Zehraoui et Younès Bennani. M-SOM-ART : Growing self organizing map for sequence clustering and classification. *ECAI'04, European Conference on Artificial Intelligence*, 2004.

[CI-16] Farida Zehraoui et Younès Bennani. M-SOM : Matricial self organizing map for sequences clustering and classification. *IJCNN'04, International Joint Conference on Neural Networks*, juillet 2004.

[CI-15] Farida Zehraoui, Rushed Kanawati et Sylvie Salotti. CASEP2 : Hybrid Case-based Reasoning System for Sequence Processing. *Advances in Case-Based Reasoning, 7th European Conference (ECCBR 2004)*, pages 449–463, 2004.

2003

[CI-14] Hachemi Bennaceur et Aomar Osmani. Computing lower bound for MAX-CSP problems. *International Conference on Industrial, Engineering Applications of Artificial Intelligence and Expert Systems*, pages 614–624, 2003.

[CI-13] Khellaf Bouandas et Aomar Osmani. Optimal algorithm for temporal patterns discovery. *FLAIRS-2003*, pages 455–460. AAAI press, 2003.

[CI-12] Faouzi Boufares, Hacemi Bennaceur et Aomar Osmani. On the consistency of cardinality constraints in uml modelling. *10th ISPE International Conference on Concurrent Engineering*, volume 1, pages 287–292. A. A. Balkema Publishers, 2003.

[CI-11] Zeina Jrad, Francine Krief et Badr Benmammar. A intelligent user interface for the dynamic negotiation of QoS. *International Conference on Telecommunications, ICT'2003*, février 2003.

[CI-10] Rushed Kanawati. A Multi-agent System for Improving Result Ranking Service of Web Search Engines. *10th ISPE International conference on concurrent engineering (CE 2003)*, volume 1, pages 91–99, Funchal, Madeira, June 2003. Balkema.

[CI-9] Larbi Ait Mohamed et Aomar Osmani. A model based on domain theory for the supervision of the telecommunications networks. *The 10th International Conference on telecommunications ICT-2003*, pages 31–40. IEEE Springer, 2003.

[CI-8] Hassine Moungla et Francine Krief. For an intelligent policy-based telecommunication network management. *International Conference on Telecommunication Systems, Modelling and Analysis. ICTSM'2003. IFIP-IEE. WG 7.3*, octobre 2003.

[CI-7] Hassine Moungla et Francine Krief. Vers une gestion intelligente des réseaux d'opérateurs offrant des services IP. *International Congress On Networks And Services Mangement, GRES'2003*, février 2003.

[CI-6] Aomar Osmani. Qualitative point sequential patterns. *7th International Conference on Knowledge Based Intelligent Information KES2003*, 2003.

[CI-5] Aomar Osmani. STCSP : A representation model for sequential patterns. *Foundations and Applications of Spatio-Temporal Reasoning. AAAI Sumposium*, 2003.

[CI-4] Alexandre Pauchet, Nathalie Chaignaud et Amal El Fallah-Segrouchni. BDIGGY : a cognitive architecture for cooperative problem solving. *IEEE IAT - International Conference on Intelligent Agent Technology*, pages 523–526, octobre 2003.

[CI-3] Arnaud Zeboulon, Younès Bennani et Khalid Benabdeslem. Hybrid connectionist approach for knowledge discovery from web navigation patterns. *ACS/IEEE International Conference on Computer Systems and Applications*, juillet 2003.

[CI-2] Farida Zehraoui, Rushed Kanawati et Sylvie Salotti. Case base maintenance for improving prediction quality. *The 6th International Conference on Case-Based Reasoning (ICCBR'03)*, volume 2689 de *Lecture Notes in Computer Science*, pages 261–275. springer, June 2003.

[CI-1] Rushed Kanawati et Maria malek. Computing Social Networks for Information Sharing : A Case-based Approach, volume = 4564, year = 2007. Schuler Douglas, éditeur, *Second international conference on Online Communities and Social Computing*, LNCS, pages 86–95, Beijing, China, July. Springer.

Workshops et autres conférences avec comité de lecture

2007 [CO-47] Lucile Chalet, Hanane Azzag, Christiane Guinot et Gilles Venturini. Un algorithme préliminaire pour la structuration visuelle de données inspiré des modéles de construction de nids chez les insectes sociaux. *Atelier Fouille du Données et Algorithmes Biomimétiques aux 7ème EGC'07 Extraction et Gestion des Connaissances*, pages 13–24, jan 2007.

[CO-46] M. Elati, P. Neuvial, C. Rouveirol, F. Radvanyi et E. Barillot. A methodology for discovering cooperative regulatory pattern from ternary gene expression data. *Cap'2007 : conférence francophone sur l'apprentissage automatique, Plate-forme AFIA*, 2007.

[CO-45] Sébastien Guérif et Younès Bennani. Sélection de Variables en Apprentissage Numérique Non Supervisé. *Conférence francophone sur l'Apprentissage automatique (CAp'2007)*, pages 221–236, Grenoble, France, July 2007.

[CO-44] Rushed Kanawati. A community aware web search engine. *In proceedings HCI International 2007*, Beijing, china, july 2007. poster.

[CO-43] Hager Karoui. Systèmes de recommandation de références bibliographiques et de formation de comités. *Actes des 3èmes Rencontre Inter-Associations (RIAs 07) : Recherche et Extraction d'Information*, Toulouse - IRIT, 2007.

[CO-42] Julien Lavergne, Hanane Azzag, Christiane Guinot et Gilles Venturini. Construction et exploration interactive d'un graphe de documents. *3émes Rencontres Inter-Associations (RIA's 2007)*, Toulouse, France, 12-13 mars 2007. IRIT.

[CO-41] Catherine Recanati et Nicoleta Rogovschi. Enchaînements verbaux - étude sur le temps et l'aspect utilisant des techniques d'apprentissage non supervisé. pages 379–388, Toulouse, France, Juin 2007.

[CO-40] Nicoleta Rogovschi, Younès Bennani et Catherine Recanati. Apprentissage neuro-markovien pour la classification non supervisée de données structurées en séquences. *Actes de Extraction et Gestion des Connaissances (EGC'07)*, Atelier « Fouille de données et algorithmes biomimétiques », pages 49–60, Namur, Belgique, Jan. 2007. Cépaduès Editions.

[CO-39] Nicoleta Rogovschi et Emmanuel Viennet. Finding community structure in large social networks. Poster at NATO Advanced Study Institute on Mining Massive Data Sets for Security, septembre 2007.

2006 [CO-38] Nejla Amara-Hachmi et Amal El Fallah-Seghrouchni. Modélisation d'information contextuelle pour des agents mobiles sensibles au contexte. *Nouvelles technologies de la répartition (NOTERE'06)*, Toulouse, juin 2006.

[CO-37] Khalid Benabdeslem et Younès Bennani. Classification et visualisation des données d'usages d'Internet. *Atelier " Fouille du Web " 6ème EGC'06 (Extraction et Gestion des Connaissances*, pages 29–40, janvier 2006.

[CO-36] Philippe Bessières, Robert Bossy, Alain-Pierre Manine, Erick Alphonse et Claire Nédellec. Getting the unknown from the known in bacteria, and the role of text mining. *Int. ECML'06 Workshop on Data and Text Mining for Integrative Biology*, 2006.

[CO-35] Jérome Galy, Emmanuel Viennet et Maria Ruiz. Création des départements Télécommunications et Réseaux dans les IUT au Vénézuela. *Vème congrès France-Amérique latine, Coopération technologique et professionnelle*, novembre 2006.

[CO-34] Hager Karoui, Rushed Kanawati et Laure Petrucci. Cooperative CBR system for peer agent committee formation. *Proc. 5th Workshop on Agents and Peer-to-Peer Computing (AP2PC'06)*, pages 76–90, Hakodate, Japan, mai 2006.

[CO-33] Hager Karoui, Rushed Kanawati et Laure Petrucci. Une approche ràpc pour la formation de comité d'agents dans un système de recommandation égal à égal. *14ème Atelier de Raisonnement à Partir de Cas*, pages 63–74, Besançon, France, Mars 2006.

[CO-32] Adeline Nazarenko, Claire Nédellec, Erick Alphonse, Sophie Aubin, Thierry Hamon et Alain-Pierre Manine. Semantic annotation in the Alvis project. *Int. Workshop on Intelligent Information Access (IIIA 2006)*, 2006.

[CO-31] Aomar Osmani et Khellaf Bouandas. DATTES : An algorithm for discovering patterns in temporal sequence. *Proceedings of ECAI temporal and Spatial reasoning Workshop*, 2006.

[CO-30] Emmanuel Viennet, Françoise Fogelman Soulié et Benoît Rognier. évaluation de techniques de traitement des refusés pour l'octroi de crédit. Société Française de Statistique, éditeur, *38èmes Journées de Statistique*, mai 2006.

2005

[CO-29] Erick Alphonse, Sophie Aubin, Philippe Bessières, Gilles Bisson, Eric Gicquello, Thierry Hamon, Alain-Pierre Manine, Adeline Nazarenko et Claire Nédellec. Information extraction from MedLine on bacillus subtilis regulations. *BACELL 2005 European Workshop*, 2005.

[CO-28] Nejla Amara-Hachmi. A framework for building adaptive mobile agents. *Doctoral Mentoring Program at the 4th International Joint Conference on Autonomous agents and Multi-Agent Systems AAMAS 2005*, page 1369. ACM Press, juillet 2005.

[CO-27] Nejla Amara-Hachmi et Amal ElFallah-Seghrouchni. A framework for context-aware mobile agents. *4th Workshop on Ambient Intelligence - Agents for Ubiquitous Environments AmbiAgents05*, juillet 2005.

[CO-26] Nejla Amara-Hachmi et Amal ElFallah-Seghrouchni. GAMA : Architecture générique d'agents mobiles adaptables. *Plate-forme de l'Association Française pour l'Intelligence Artificielle - AFIA05, Conférence : Rencontres Jeunes Chercheurs en Intelligence Artificielle RJCIA'05*, juin 2005.

[CO-25] Nejla Amara-Hachmi et Amal ElFallah-Seghrouchni. Towards a generic architecture for self-adaptive mobile agents. *5th European Workshop on Adaptive Agents and MultiAgent Systems*, mars 2005.

[CO-24] Sujeevan Aseervatham et Aomar Osmani. Mining short sequential patterns for hepatitis type detection. *Discovery Challenge (ECML/PKDD)*, 2005.

[CO-23] Lahcène Dehni, Younès Bennani et Francine Krief. Low energy adaptive connectionist clustering for wireless sensor networks. *IEEE/IFIP MATA 2005 : 2nd International Workshop on Mobility Aware Technologies and Applications*, volume 3744, pages 405–415, octobre 2005.

[CO-22] Lahcène Dehni, Younès Bennani et Francine Krief. Une approche neuronale adaptative de routage minimisant la consommation d'énergie dans les réseaux de capteurs. *colloque GRES 2005 : Gestion de REseaux et de Services*, pages 115–127, février 2005.

[CO-21] Sébastien Guérif, Younès Bennani et Eric Janvier. μ-SOM : Weighting features during clustering. *Proceedings of the 5th Workshop On Self-Organizing Maps (WSOM'05)*, pages 397–404, Paris, France, septembre 2005.

[CO-20] Zeina Jrad. Spécification d'une interface intelligente pour la prochaine génération de réseaux IP. *Journées Scientifiques DNAC 2005. Les réseaux IP : synthèse et perspective, controle, gestion, mobilité et sécurité.*, avril 2005.

[CO-19] Zeina Jrad, Francine Krief et Sébastien Guérif. Une interface utilisateur pour faciliter l'accès à l'internet nouvelle génération. *Journées doctorales Informatique et Réseau (JDIR 2005)*, Troyes, France, décembre 2005.

[CO-18] Rushed Kanawati. Un système coopératif pour le tri des résultats de moteurs de recherche. *13ième atelier sur le raisonnement à partir de cas - Plate-forme AFIA 2005*, mai 2005.

[CO-17] Hager Karoui. Agents RàPC pour la gestion coopérative de bases bibliographiques personelles. *Atelier raisonnement à partir de cas, plate-forme AFIA*, pages 86–97, Nice, mai 2005.

[CO-16] Dorra Maghri et Nejla Amara-Hachmi. Intégration d'un agent mobile dans un système multi-agents hierarchique. *Manifestation des Jeunes Chercheurs francophones dans les domaines des STIC MajecSTIC*, Rennes, novembre 2005.

[CO-15] Hassine Moungla et Francine Krief. A probabilistic heuristic for conflict detection in policy based management of Diffserv networks. *Proceedings of IEEE/IFIP MATA 2005 : 2nd International Workshop on Mobility Aware Technologies and Applications*, octobre 2005.

[CO-14] Alexandre Pauchet, Nathalie Chaignaud et Amal El Fallah-Segrouchni. Un modèle cognitif de l'interaction pour agents logiciels. *3èmes Journées Francophones des Modèles Formels de l'Interaction (MFI'05)*, mai 2005.

2004 [CO-13] Alejandro Guerra-Hernandez, Amal El Fallah-Seghrouchni et Henry Soldano. Learning in BDI multi-agent systems. *Fourth International Workshop on Computational Logic in Multi-Agent Systems (CLIMA IV)*, volume 3259, pages 218–233, 2004.

[CO-12] H. Karoui, R. Kanawati et L. Petrucci. An intelligent peer-to-peer multi-agent system for collaborative management of bibliographic databases. *Proc. Workshop on Case Based Reasoning, Cambridge, England*, décembre 2004.

[CO-11] Hassine Moungla. La différentiation de service dans les réseaux MPLS/GMPLS. *DNAC 2004*, février 2004.

[CO-10] Aomar Osmani. Cyclic and linear intervals for reasoning about repetitions in time. *Proceedings of AAAI Workshop*, pages 1–6, 2004.

[CO-9] Véronique Ventos, Henry Soldano et Thibaut Lamadon. Treillis de Galois alpha. *Conférence d'Apprentissage (CAP'04)*, pages 175–190, Montpellier, 2004. Presses Universitaires de Grenoble.

[CO-8] Farida Zehraoui et Younès Bennani. SOM-ART : Incorporation des propriétés de plasticité et de stabilité dans une carte auto-organisatrice. *Atelier FDC : Fouille de Données Complexes dans un processus d'extraction de connaissances (EGC 2004)*, pages 169–180, janvier 2004.

[CO-7] Farida Zehraoui, Rushed Kanawati et Sylvie Salotti. Case Representation, Retrieval and Reuse in the Hybrid System for Sequence Processing CASEP2. Rushed Kanawati et Sylvie Salotti, éditeurs, *Second International workshop on applying case-base reasoning on time series prediction*, pages 339–350, September 2004.

2003 [CO-6] Badr Benmammar et Francine Krief. La technologie agent et les réseaux sans fil. *Congrès Des Nouvelles Architectures pour les Communications. DNAC 2003*, octobre 2003.

[CO-5] Gilles Klein et Francine Krief. Mobile agents for dynamic sla negotiation. *International Workshop on Mobile Agents for Telecommunication Applications. MATA 2003*, octobre 2003.

[CO-4] Francine Krief et Zeina Jrad. An intelligent environment for dynamic negotiation, provisioning and control of QoS in IP networks. *Colloque Gestion de Réseau et de Service. GRES 2003*, février 2003.

[CO-3] Alexandre Pauchet, Nathalie Chaignaud et Amal El Fallah-Seghrouchni. Résolution coopérative de problèmes intégrant planification et interaction. *Journées Francophones sur les Systèmes Multi-Agents*, pages 309–313, 2003.

[CO-2] Alexandre Pauchet, Nathalie Chaignaud et Amal El Fallah-Segrouchni. Modelling planning and interaction in the framework of human cooperative problem solving. *AAMAS Workshop on Human and Multi-Agent Systems*, pages 31–38, juillet 2003.

[CO-1] Olivier Sigaud et Pierre Gérard. Apprentissage par renforcement indirect dans les sytèmes de classeurs. *Actes des journèes PDMIA*, 2003.

Direction d'ouvrages (DO)

Livres

2006 [LI-3] Younès Bennani, éditeur. *Apprentissage Connexionniste*. 2006, ISBN 2-7462-1337-0.

2005 [LI-2] Benoît Charroux, Aomar Osmani et Yann Thierry-Mieg. *UML 2*. Pearson, juillet 2005, ISBN 2-7440-7124-2.

2003 [LI-1] Louise Travé-Massuyès et Philippe Dague, éditeurs. *Modèles et raisonnements qualitatifs*. IC2. octobre 2003, ISBN 2-7462-0744-3.

Chapitres de livre

2007 [CL-16] Rushed Kanawati. *Agents RàPC coopératifs pour la recherche d'information*, Jean Renaud, Brigitte Chebel Morello, Béatrice Fuchs et Jean Lieber, éditeurs, *Raisonnement à partir de cas 2, surveillance, diagnostic et maintenance*, volume 2 de *Informatique et système d'information*, Chapitre 6, pages 177–195. Hermès, 2007.

[CL-15] Maria Malek et Rushed Kanawati. *Case-based Reasoning in Knowledge Discovery and Data Mining*, Sankar K. Pal, David W. Aha et Kalyan M. Gupta, éditeurs, *Case-Based Reasoning in Knowledge Discovery and Data Mining*, Chapitre CBR-based sequence analysis : An overview. Wiely, 2007.

[CL-14] Farida Zehraoui, Rushed Kanawati et Sylvie Salotti. *CASEP2 : Système Hybride pour le traitement de séquences de navigation sur le Web*, Jean Renaud, Brigitte Cheble-Morello, Bétarice Fuchs et Jean Lieber, éditeurs, *Raisonnement à Partir de Cas 2 : surveillance, diagnostic et maintenance*, volume 2 de *Informatique et système d'information*, pages 151–176. Hermès, 2007.

2006 [CL-13] Younès Bennani. *Outils de simulation des systèmes connexionnistes*, Chapitre 11, pages 343–358. Volume 1 de ApprentissageConnexionniste [LI-3], 2006, ISBN 2-7462-1337-0.

[CL-12] Younès Bennani. *Réseaux à dictionnaires : Learning Vector Quantization*, Chapitre 5, pages 123–140. Volume 1 de ApprentissageConnexionniste [LI-3], 2006, ISBN 2-7462-1337-0.

[CL-11] Younès Bennani. Réseaux de neurones artificiels. J. Akoka et I. Wattiau, éditeurs, *Encyclopédie de l'informatique et systèmes d'information*, volume Section Intelligence Artificielle, pages 1217–1231. Editions Vuibert, 2006.

[CL-10] Younès Bennani. *Séparateurs connexionnistes linéaires*, Chapitre 1, pages 21–46. Volume 1 de ApprentissageConnexionniste [LI-3], 2006, ISBN 2-7462-1337-0.

[CL-9] Rushed Kanawati. Raisonnement à partir de cas : principes, méthodes et applications industrielles. Jean Renaud, Béatrice Fuchs et Jean Lieber, éditeurs, *Raisonnement à Partir de Cas : principes, méthodes et applications industrielles*, Chapitre Agents RàPC coopératifs pour la recherche d'information sur le Web. 2006.

[CL-8] Hassine Moungla. *L'Auto-configuration dans la gestion autonome, l'Autonomic communication*. 2006.

[CL-7] Nadia Pisanti, Henry Soldano, Mathilde Carpentier et Joël Pothier. Implicit and explicit representation of approximated motifs. C. S. Iliopoulos, K. Park et K. Steinhoefel, éditeurs, *Algorithms in Bioinformatics*, pages 1–15. College Press, 2006.

[CL-6] Emmanuel Viennet. *Réseaux à Fonctions de Base Radiales*, Chapitre 4, pages 105–122. Volume 1 de Bennani [LI-3], 2006, ISBN 2-7462-1337-0.

[CL-5] Farida Zehraoui, Rushed Kanawati et Sylvie Salotti. Raisonnement à partir de cas : principes, méthodes et applications industrielles. Jean Renaud, Béatrice Fuchs et Jean Lieber, éditeurs, *Raisonnement à Partir de Cas : principes, méthodes et applications industrielles*, Chapitre CASEP2 : Système hybride pour le traitement des séquences. 2006.

2005 [CL-4] Zeina Jrad. *Les Plates-formes Multi-agents, Intelligence artificielle et réseaux*. 2005.

[CL-3] Hassine Moungla. *L'Intelligence artificielle et la supervision des réseaux de télécommunication, Intelligence artificielle et réseaux*. 2005.

2004 [CL-2] Francine Krief. *Application des Agents dans la Gestion par Politique, Intelligence artificielle et réseaux*. 2004.

2003 [CL-1] Hakima Kadri-Dahmani et Aomar Osmani. *Updating Data in GIS : How to maintain Database Consistency, Enterprise Information Systems IV book*, pages 163–169. Kluwer, 2003, ISBN 1-4020-1086-9.

Ouvrages scientifiques (OS)

Édition d'ouvrages collectifs

2007 [ED-7] Hanane Azzag et Gilles Venturini. Fouille du Données et algorithmes biomimétiques. *Atelier Fouille du Données et Algorithmes Biomimétiques aux 7ème EGC'07 Extraction et Gestion des Connaissances*, page 60, jan 2007.

[ED-6] Y. Bennani et E. Viennet, éditeurs. *Apprentissage artificiel et fouille de données, numéro spécial de la Revue des Nouvelles Technologies de l'Information*. Cepadues, 2007. 230 pages, publication en cours.

2004 [ED-5] Rushed Kanawati et Sylvie Salotti, éditeurs. *Proceedings of the second international workshop on applying case-based reasoning for sequence processing*, Madrid, September 2004.

[ED-4] Rushed Kanawati et Sylvie Salotti, éditeurs. *Second International Workshop on Applying CBR to Time Series Prediction*, Madrid, Spain, septembre 2004. In conjunction with ECCBR'04.

[ED-3] Rushed Kanawati, Sylvie Salotti et Farida Zehraoui, éditeurs. *12ème Atelier de Raisonnement à partir de cas (RàPC 2004)*, Villetaneuse, mars 2004. LIPN.

2003 [ED-2] Rushed Kanawati, Maria Malek et Sylvie Salotti, éditeurs. *First International Workshop on Applying CBR to Time Series Prediction*, Trondheim, Norway, juin 2003. NTNU, departement of information Science. In conjunction with ICCBR'03.

[ED-1] Rushed Kanawati, Maria Malek et Sylvie Salotti, éditeurs. *Proceedings of the first international workshop on applying Case-based reasoning for sequence processing*, Trondheim, june 2003.

Rapports de contrats

2007 [RC-21] Younès Bennani et Emmanuel Viennet. Réduction des dimensions de l'espace de description (Rapport d'avancement 1, ST3.21, Projet Infom@gic, Pôle Cap Digital). Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, juillet 2007.

2006 [RC-20] Sujeevan Aseervatham et Emmanuel Viennet. Méthodes à Noyau appliquées aux textes structurés (Rapport d'avancement 1, ST3.12, Projet Infom@gic, Pôle Cap Digital). Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, juillet 2006.

[RC-19] Younès Bennani et Lahcène Dehni. Apprentissage non-supervisé pour la visualisation et la classification de données spatio-temporelles (Rapport d'avancement 1, ST3.21, Projet ANR SILLAGES Suivi Individuel par Label éLectronique Au sein de Groupes complexES). Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, juillet 2006.

[RC-18] Dominique Bouthinon et Henri Soldano. Concept learning from pieces of information. Rapport Technique 2006-03, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, 2006.

[RC-17] Julien Derivière et Adeline Nazarenko. Deliverable D5.2 : Report on theory and software of normalization options for IR (platform conception). Technical report, ALVIS IST-1-002068 STREP project, 2006.

2005 [RC-16] Armen Aghasaryan, Yves Cochet, Francine Krief, Eric Moreau, Hassine Moungla et Sophie Piekarec. State of the art on service contract expression languages, policy and network resource models. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, janvier 2005. Rapport 3.1.1- Self Aware management project (SWAN- RNRT project).

[RC-15] Sébastien Guérif et Younès Bennani. Apprentissage numérique non supervisé partir de données évolutives. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, Villetaneuse, France, juin 2005. Deuxième rapport de recherche ANRT (CIFRE 654-2002).

[RC-14] Zeina Jrad, Badr Benmammar, Francine Krief et Nader Mbarek. Dynamique de l'environnement : Scénarios, simulations et maquette. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, juillet 2005. Livrable de projet RNRT : IP-SIG (IP-Signaling).

[RC-13] Hassine Moungla. Specification of conflict resolution module. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, mars 2005. Rapport 3.4.2- Self Aware management project (SWAN- RNRT project).

[RC-12] Hassine Moungla et Francine Krief. Policy based management specification. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, mars 2005. Rapport 3.4.1- Self Aware management project (SWAN- RNRT project).

2004 [RC-11] Yamina Elmrabet, Sujeevan Aseervatham, Philippe Dague et Aomar Osmani. Réalisation du système de diagnostic pour le système EMB du véhicule. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, 2004.

[RC-10] Thierry Gaudré, Vania Joloboff, Yamina Elmrabet, Sujeevan Aseervatham, Philippe Robin, Philippe Dague et Aomar Osmani. Réalisation d'un environnement système évolutif pour le diagnostic automobile,. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, 2004.

[RC-9] Sébastien Guérif et Younès Bennani. Approches connexionnistes pour l'exploration et l'extraction de connaissances partir de données spatio-temporelles : application aux données géographiques. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, Villetaneuse, France, avril 2004. Premier rapport de recherche ANRT (CIFRE 654-2002).

[RC-8] Zeina Jrad, Badr Benmammar, Francine Krief et Nader Mbarek. Dynamique de l'environnement. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, juillet 2004. Rapport de projet RNRT : IP-SIG (IP-Signaling).

[RC-7] Hassine Moungla et Francine Krief. Spécification d'un modèle GMPLS dérivé du modèle générique. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, février 2004. Modélisation et Apprentissage pour une Gestion Distribuée des Alarmes De bout En boUt (projet RNRT-MAGDA2).

2003 [RC-6] Younès Bennani et Khalid Benabdeslem. Codage et compression de données issues d'internet. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, janvier 2003. RT 1, Contrat de recherche Maxeem-NumSight.

[RC-5] Younès Bennani et Khalid Benabdeslem. Visualisation des traces de navigation par apprentissage non-supervisé. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, mars 2003. RT 2, Contrat de recherche Maxeem-NumSight.

[RC-4] Zeina Jrad, Khaldoun Al Agha, Walid Dabbous, Ghamri Y. Doudane, Francine Krief, Frédéric Le Garrec, Sandrine Masson, Pascale Minet, Thi-Mai-Trang Nguyen, Guy Pujolle, Rafael Rizo et Rares Serban. ARCADE (ARchitecture de Contrôle Adaptative Des Environnements IP). Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, janvier 2003. Rapport final projet RNRT.

[RC-3] Larbi Ait Mohamed et Aomar Osmani. Sélection des MIB standards et des MIB CISCO. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, janvier 2003.

[RC-2] Emmanuel Viennet. Estimation de la QoS de bout en bout par apprentissage numérique. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, décembre 2003. Modélisation et Apprentissage pour une Gestion Distribuée des Alarmes De bout En boUt (projet RNRT-MAGDA2).

[RC-1] Arnaud Zeboulon, Younès Bennani et Khalid Benabdeslem. Classification et reconnaissance de séquences de navigations par SOM et modèle de mélanges. Rapport Technique, Laboratoire d'Informatique de Paris Nord - UMR CNRS 7030 - Institut Galilée - Université Paris 13, juin 2003. RT 3, Contrat de recherche Maxeem-NumSight.

Rapports

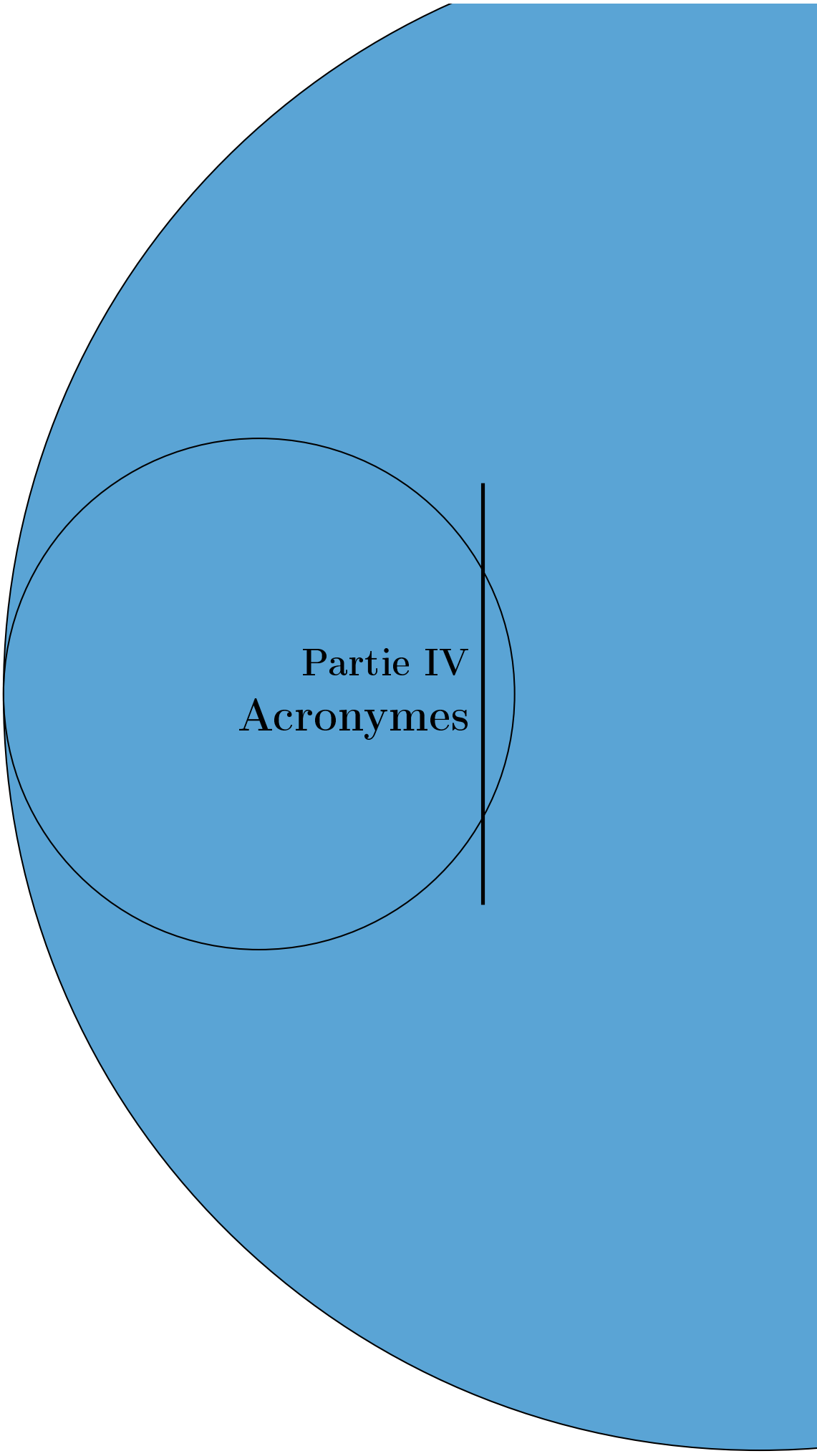
2006 [Ra-1] Catherine Recanati, Younès Bennani et Nicoleta Rogovschi. Analyse de séquences verbales - une segmentation numérique des temps et catégories aspectuelles dans des récits d'accidents. 19 pages, 2006.

Travaux en cours de soumission

2007 [So-3] Sujeevan Aseervatham, Emmanuel Viennet et Younès Bennani. Semi-structured document categorization with a semantic kernel. *International Journal of Computational Intelligence and Applications*, 2007. Soumis en août 2007.

[So-2] Hanane Azzag, Fabien Picarougne, Pascale Kuntz, Christiane Guinot et Gilles Venturini. A survey of biomimetic algorithms for data clustering. *Journal of Soft Computing, Special Issue on Evolutionary and Metaheuristic-based Data Mining*, 2007.

[So-1] Mustapha Lebbah et Younès Bennani. A probabilistic self-organizing map for binary data topographic clustering. *International Journal of Computational Intelligence and Applications (IJCIA)*, 2007.



Partie IV
Acronymes

A³ Apprentissage Artificiel et Applications

ACI Action Concertée Incitative

AM allocataire moniteur

ANR Agence National de la Recherche

ARC Australian Research Council

ATER attaché temporaire d'enseignement et de recherche

BQR Bonus Qualité-Recherche (université)

CDD Contrat à Durée Déterminée

CIFRE Conventions Industrielles de Formation par la Recherche

CNRS Centre National de la Recherche Scientifique

DR directeur de recherche

DEA Diplôme d'Etudes Approfondies

DESS Diplôme d'Etudes Supérieures Spécialisées

DGE Direction Générale de l'Equipement (ministère)

EA Equipe d'Accueil

IG Institut Galilée

IUTV IUT de Villetaneuse

LAGA Laboratoire Analyse, Géométrie et Applications, université Paris 13 - UMR 7539

Lim&Bio Laboratoire d'Informatique Médicale et Bio-Informatique, université Paris 13 - EA

LIPN Laboratoire d'Informatique de Paris-Nord - UMR 7030

LRI Laboratoire de Recherches en Informatique, université Paris 11 - UMR 8623

MCF maître de conférences

MENESR Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche

OCAD Optimisation Combinatoire et Algorithmique Distribuée

PCRDT Programme Cadre de Recherche et Développement Technologique

PU professeur des universités

RCLN Représentation des Connaissances et Langage Naturel

RNRT Réseau National de Recherche en Télécommunication

RNTL Réseau National de recherche et d'innovation en Technologies Logicielles

SMF Société Mathématique de France

UE Union européenne

UMR Unité Mixte de Recherche