

Visite du LIF par les élèves de l'ENS Cachan

Propositions d'exposés, programme et plan d'accès

8 décembre 2009

1 BDAA

Bases de données et Apprentissage Automatique

Issue d'une double communauté, Base de Données et Apprentissage Automatique, l'équipe BDAA s'est scientifiquement fondée sur une thématique fédératrice : le data mining. Elle est en outre en charge de l'option Bases de Données Avancées des masters recherche (2IF) et professionnel (I2A) (voir [Master d'Informatique de Marseille](#)).

Depuis son apparition, le data mining a suscité de nombreuses contributions de la part de différentes communautés scientifiques et a élargi le spectre des types de connaissances extraites, des approches et techniques mise en oeuvre. De nouvelles problématiques et de nouveaux champs d'application sont apparus. Dans ce contexte général de l'extraction de nouvelles connaissances, nos travaux de recherche s'inscrivent dans les problématiques suivantes : fouille de bases de données multidimensionnelles ou d'entrepôts ; inférence grammaticale probabiliste ; apprentissage statistique (théorique et pratique).

Rémi Eyraud : *Apprentissage artificiel : des principes généraux à un algorithme particulier*

Le contenu serait divisé en deux parties : une première présentant les principes généraux de l'apprentissage, une seconde détaillant un algorithme particulier (algorithme RPNI d'apprentissage d'automates à états finis).

Liva Ralaivola : *Machines à vecteurs de support et ordonnancement d'instances*

Les machines à vecteurs de supports (SVM) constituent l'une des avancées majeures des dernières années en apprentissage statistique. Ces modèles, qui reposent sur des arguments statistiques fondés, nécessitent pour leur apprentissage la résolution d'un programme quadratique, pour lequel des algorithmes efficaces existent. Une des limitations des SVM est cependant qu'elles n'optimisent pas directement un critère de performance lié à une notion d'ordre, notion primordiale lorsque l'on s'intéresse à des tâches de recherche d'information (réponses d'un moteur de recherche à une requête, par exemple). Des formulations des SVM pour le problème d'ordonnancement existent mais la plupart nécessitent de construire/considérer un échantillon d'apprentissage dont la taille est de l'ordre du carré de l'échantillon d'apprentissage d'origine : le problème de l'ordonnancement est alors formulé sous la forme d'un problème de classification de paires. Une question intéressante est celle de savoir s'il est possible de s'affranchir de la construction de ces paires d'apprentissage. Une piste de recherche à explorer repose sur la formulation dite "2-norm" des SVMs, qui peut éventuellement faire le lien entre les SVMs classiques pour la classification et les SVMs pour l'ordonnancement.

2 ESCAPE

Systemes complexes, automates et pavages

L'objectif d'Escape est de proposer des contributions à l'étude des systèmes complexes en employant des méthodes issues de l'informatique théorique moderne.

La thématique des systèmes complexes est un sujet de recherche pluridisciplinaire en pleine expansion. Elle est motivée par un constat de confrontation de nombreuses disciplines à des objets de nature commune (des réseaux de très nombreuses entités connues en interaction) exhibant des comportements globaux complexes difficiles à expliquer et à prévoir.

L'étude des systèmes complexes vise à factoriser des outils, méthodes et théories issus de différentes disciplines dans le but d'expliquer, de prédire et de produire de tels phénomènes. Au sein des réseaux de recherche qui se mettent en place sur ces thèmes, l'équipe Escape participe à l'étude de modèles formels des systèmes complexes. En particulier, nous analysons les liens entre la notion de calcul et l'émergence de propriétés globales. La notion de calcul est comprise au sens large du déplacement de l'information, de son traitement, de son placement dans des structures géométriques.

Emmanuel Jeandel : Pavages du plan : théorie des puzzles

Dès notre plus jeune âge, nous apprenons à résoudre des puzzles, d'abord avec un petit nombre de pièces, puis de plus en plus. Dans cet exposé, nous étudierons les pavages, qu'on peut voir comme une modélisation mathématique très simple des puzzles. Nous verrons ainsi que résoudre un puzzle n'est pas si facile que ça.

Alexander Shen : Tutorial

I'd be glad to provide some kind of tutorial (possible topics : probabilistic proofs and kolmogorov complexity, algorithmic randomness, interactive proofs, etc.) but unfortunately can do this only in English.

3 CRO

Combinatoire et Recherche Opérationnelle

Notre activité de recherche concerne principalement l'optimisation combinatoire (complexité, méthodes et algorithmes de résolution, programmation en nombres entiers, applications à la Recherche Opérationnelle), la théorie des graphes (structure, métrique et algorithmique des graphes et réseaux), la combinatoire (systèmes d'ensembles structurés), et la géométrie discrète et algorithmique (géométrie des distances, distances de chanfrein). De nombreux problèmes que nous considérons ont des applications pratiques directes dans les domaines des réseaux de communications, de la modélisation de systèmes complexes, de la planification de personnels, de l'analyse d'images et nous ont permis d'obtenir différents contrats avec France Telecom, Prologia, Air Liquide. Le groupe Combinatoire et Recherche Opérationnelle a également un financement ANR pour la période 2006-2010 sur le thème Optimisation Combinatoire : Théorie, Algorithmes et Applications.

Pierre Bonami : *Coupes disjointes pour la programmation mixte sous contraintes quadratiques*

Dans ces travaux nous proposons de construire des relaxations fortes de problèmes non-convexes sous contraintes quadratiques en utilisant des techniques de programmation disjointive et la méthodologie du lift-and-project. Nous présenterons plusieurs algorithmes ainsi que des expériences de calculs sur des problèmes tests de la littérature pour évaluer la force des relaxations obtenues.

(travaux avec Anureet Saxena, Axioma Inc., et Jon Lee, IBM T.J. Watson)

Victor Chepoi : *On isometric embedding of median graphs into Cartesian products of trees*

In this talk, we consider the problem of isometric embedding of a median graph into the Cartesian product of a minimum number of trees. After a brief excursion into the world of median structures, their properties, characterizations and connections with other discrete objects, we will define the respective embedding problem. Then we show that already for very particular median graphs this problem is as hard as coloring of arbitrary graphs. We continue with several characterizations of graphs isometrically embeddable into the Cartesian product of two trees. We show also that a special class of plane median graphs, called squaregraphs, and consisting of plane quadrangulations with inner vertices of degree > 3 , can be isometrically embedded into Cartesian products of at most five trees. We also formulate a couple of open questions. This is a joint work with Hans-Juergen Bandelt (Hamburg Universitaet) et David Eppstein (University of California, Irvine).

Yann Vaxès : *Diamètre, centre et arbres approximant les espaces géodésiques et les graphes delta-hyperboliques.*

La notion d'espace métrique δ -hyperbolique a été introduite par Gromov en 1987. Elle joue un rôle important dans le domaine de la théorie des groupes, de la géométrie des espaces à courbure non-négative. Certains travaux ont récemment montré l'intérêt de cette notion dans le domaine de l'algorithmique et des réseaux. Dans cet exposé, étant donné un ensemble S de points d'un espace δ -hyperbolique, nous présentons des algorithmes simples et efficaces pour approximer le diamètre de S avec une erreur 2δ et calculer de façon approximative le rayon et le centre d'une plus petite boule englobant S avec une erreur additive 3δ . Ces algorithmes s'exécutent en temps linéaire pour des espaces hyperboliques classiques et pour les graphes δ -hyperboliques. De plus, nous montrons que le centre des graphes δ -hyperboliques de degrés bornés peut être calculé en temps $O(|E|)$. Nous décrivons aussi une construction simple d'un arbre approximant les distances d'un graphe δ -hyperbolique $G = (V, E)$ sur n sommets avec une erreur additive $O(\delta \log n)$. Cette construction a une distorsion additive comparable à celle donnée par M. Gromov pour les espaces δ -hyperboliques mais elle peut-être implémentée en temps linéaire $O(|E|)$ (au lieu de $O(n^2)$). Il s'agit d'un travail en collaboration avec V. Chepoi (LIF), F. Dragan (Kent State University), B. Estellon (LIF), M. Habib (LIAFA).

4 LOMI

Logiques Non Classiques et Modèles d'Interaction

L'objectif de LOMI est l'étude de formalismes logiques pour la représentation des connaissances humaines. Ceci se traduit par le développement de concepts et de formalismes logiques visant à résoudre des problèmes de raisonnements de sens commun.

Profondément convaincus de la nécessité d'avancer à la fois sur les aspects conceptuels, sur les aspects formels et sur leur concrétisation algorithmique et logicielle, nous contribuons à ces différents niveaux en cherchant à formaliser des domaines intéressants, en développant de nouveaux formalismes, et en développant des démonstrateurs pour ceux-ci.

Camilla Schwind : Révision de théories et tableaux sémantiques

Cet exposé concerne la représentation des connaissances par des théories logiques et les changements de ces théories lors de l'arrivée d'une nouvelle information. Ces changements sont décrits par des fonctions de révision qui associent une nouvelle base de connaissances à une base donnée et une formule qui représente la nouvelle information.

Nous présentons des postulats qui décrivent quelles propriétés doit avoir une fonction de révision. Nous montrons qu'il y a une représentation très naturelle pour chaque fonction de révision en termes d'impliquants d'une formule.

5 MOVE

Modélisation et Vérification

L'équipe MOdélisation et VErification (MOVE) située au CMI et à Luminy effectue son activité de recherche sur des thèmes fondamentaux et appliqués entre logique et informatique en se concentrant sur les aspects de conception, analyse et synthèse de systèmes logiciels et matériel. L'équipe a la charge de l'option Fiabilité et Sécurité des Systèmes Informatiques du master informatique (voir [formations](#)).

Jérémy Chalopin : Rendez-vous d'agents mobiles

On dispose d'agents mobiles qui peuvent se déplacer sur un graphe, les agents étant initialement dispersés sur différents sommets du graphe. Chaque agent est autonome et tous les agents exécutent le même algorithme. On cherche à trouver des algorithmes permettant aux agents de tous se retrouver en un même sommet du graphe. Selon les hypothèses qu'on fait sur le graphe et les capacités des agents, le problème ne peut pas toujours être résolu. Dans cet exposé, j'essaierai de présenter des conditions nécessaires et/ou suffisantes permettant de résoudre le problème.

Luigi Santocanale : Variations sur la logique modale : la logique des jeux

Je vous présenterai d'abord la logique multimodale, aussi connue comme logique de Hennessy Milner, et ensuite sa sémantique sur les systèmes de transitions. Je discuterai les limites des systèmes de transition en tant que modèle des calculs interactifs, pour ensuite décrire un autre modèle – le modèle des jeux. Je présenterai enfin une variante de la logique multimodale, dite logique monotone, qui mieux s'adapte pour décrire des propriétés des systèmes interactifs. Je serai à votre disposition pour présenter des problèmes autour de cette logique, qui animent à présent ma recherche.

Séverine Fratani : *Contrôle de flot d'information*

Les mécanismes de sécurité standards ne procurent pas l'assurance que le comportement de bout en bout d'un système informatique satisfait des politiques de sécurité importantes telles que la confidentialité. Une politique de confidentialité de bout en bout affirme que les données d'entrée secrètes ne peuvent pas être inférées par un attaquant à travers l'observation des sorties du système ; cette politique contrôle le /flot d'informations./ Différentes approches directes sont utilisées pour spécifiées et exécuter des politiques de flot d'information, nous parlerons ici des systèmes de types de sécurité.

Pierre-Alain Reynier : *Synthèse de contrôleurs : théorie et pratique au travers d'une étude de cas industrielle*

Les méthodes formelles visent à garantir la fiabilité des systèmes, en s'appuyant sur des modèles mathématiques. Je présenterai au travers d'une étude de cas réalisée dans le contexte d'un projet européen les outils théoriques mis en oeuvre (modèles, algorithmes), les outils logiciels utilisés et le type de résultats obtenus. L'étude de cas visait ici à synthétiser de façon automatique un contrôleur pour un système de pompe qui soit à la fois correct, tolérant aux perturbations, et optimal en terme d'énergie.

6 TALEP

Traitement Automatique du Langage Ecrit et Parlé

TALEP est l'acronyme de Traitement Automatique du Langage Ecrit et Parlé. L'équipe TALEP réalise des travaux linguistiques fondamentaux avec le souci d'une formalisation rigoureuse qui permet des traitements automatiques non triviaux de la langue.

Elle constitue et diffuse des ressources linguistiques électroniques (corpus, lexiques, grammaires, etc.) dont elle étend régulièrement le contenu. Elle conçoit des algorithmes et des programmes pour le traitement automatique des langues. Elle développe des logiciels génériques, comme CONTEXTES, ILLICO ou MACAON au moyen desquels des applications spécifiques sont réalisées.

Alexis Nasr : *Analyse syntaxique probabiliste*

L'objectif de l'analyse syntaxique probabiliste est de fournir l'analyse syntaxique la plus probable d'une phrase. Il s'agit d'une étape préalable à un certain nombre de traitements relevant du TAL (Traitement Automatique des Langues) tel que la traduction automatique ou la compréhension de texte.

Cet exposé commencera par décrire les principes généraux de l'analyse syntaxique probabiliste (construction de grammaires, modèles probabilistes, algorithmes d'analyse et de recherche des analyses les plus probables) et présentera deux projets en cours d'analyse syntaxique de l'anglais et du français.

Frédéric Béchet : *Analyse linguistique et traitement automatique de l'oral*

Cet exposé se focalisera sur l'analyse linguistique de messages oraux et plus particulièrement sur l'articulation des méthodes de Reconnaissance Automatique de la Parole (RAP) et de Traitement Automatique de la Langue (TAL) pour traiter le problème de la compréhension d'un message vocal.

Deux cadres applicatifs seront présentés : les systèmes de dialogue oral homme-machine et les systèmes d'indexation et d'accès à des corpus oraux de grandes tailles. Après une présentation assez générale des modèles utilisés en RAP pour transcrire un flux audio en flux textuel, nous nous intéresserons plus particulièrement aux spécificités de ces transcriptions automatiques. Nous justifierons notamment l'approche consistant à intégrer le plus possible le processus de transcription de parole et les processus d'analyse linguistique de ces transcriptions. Les méthodes employées sont principalement des méthodes basées sur de l'apprentissage automatique sur corpus et nous présenterons la combinaison de telles méthodes avec des modèles à base de connaissances linguistiques. Cet exposé sera illustré par des exemples et des démonstrations d'interfaces interactives de traitement de la parole.

Robert Pasero et Paul Sabatier : *Contraintes sur les expressions à analyser/produire*

D'un point de vue général, concevoir un système ayant pour tâche, dans le cadre de différentes applications, d'analyser et/ou de produire des expressions du langage naturel (mots, syntagmes, propositions, phrases, etc.) consiste à

- (1) définir différents types de représentations formelles au moyen de connaissances appropriées (connaissances lexicales, syntaxiques, conceptuelles, sémantiques, pragmatiques, ...),
- (2) d'exprimer des contraintes sur et entre ces représentations, et
- (3) d'exécuter/vérifier ces contraintes lors de l'analyse ou de la production de ces expressions.

Dans le cadre d'un modèle général de traitement automatique du langage naturel, nous nous intéresserons aux contraintes sur les expressions à analyser/produire. Nous illustrerons notre approche en présentant différentes applications que nous avons réalisées.

Lien : <http://www.lif.univ-mrs.fr/illico.html>