LILI explique

1962 : «Checker»

Dans les années 60, Arthur Samuel, ingénieur chez IBM, réalisa un travail de pionnier en Intelligence Artificielle (IA) et en Apprentissage Machine : il développa le logiciel «Checker» capable d'apprendre à jouer aux dames anglaises en analysant des parties d'experts.

Ainsi, en 1962, Robert Nealy, champion de Stamford (Connecticut) perdit un match face au programme de Samuel. Pour ce domaine naissant qu'était l'IA, cet événement était d'importance mais sa portée réelle fût surestimée par les médias, Nealy n'étant pas non plus un joueur de premier plan. Cependant, suite à cette victoire, les dames furent classées, à tort, comme un problème résolu et les recherches se focalisèrent uniquement sur le jeu d'échecs.

Après une interruption de plus de 25 ans, la recherche repris, sous l'impulsion de Jonathan Schaeffer, et en 1992 son programme "Chinook" termina second du Championnat du Monde de Dames devant Marion Tinsley. Finalement, Chinook devint champion du monde en 1995.

Enfin en 2007, dans un article de Science, J. Schaeffer et son équipe du AICML (Alberta Ingenuity Centre for Machine Learning) annoncèrent, qu'après avoir analysé 5.10^{20} positions différentes, le jeu de dames anglaise était «résolu» : Chinook ne peut être vaincu, car il connaît à tout moment le meilleur coup à jouer. De surcroît, ce travail montre qu'avec deux joueurs «parfaits» une partie de dame se termine toujours par une nulle.

Aujourd'hui, il reste du travail à faire dans le domaine des autres jeux. Ainsi au Go les joueurs humains restent invaincus même si des résultats récents montrent que les choses pourraient évoluer assez rapidement.

Parole à ...

Bonjour Frédéric Prost (MCF UJF, équipe CAPP) ... Vous êtes un passionné d'Échecs, dans ce domaine, il n'y a pas eu de résultats aussi définitifs que ceux obtenus avec les dames anglaises, mais ou en est-on exactement ?

Effectivement, le jeu d'échec n'est pas «résolu» et ne risque pas de l'être, le nombre de positions à explorer étant de l'ordre du nombre estimé d'atomes dans l'univers ! Toutefois, depuis la défaite de Kasparov en 1997, les programmes ont beaucoup progressé et atteignent un niveau de «3300 ELO», soit 500 points de plus que les meilleurs Grands Maîtres. Autrement dit, face à un ordinateur, personne n'a plus la moindre chance de gagner un tournoi.

Cette progression a-t-elle eu un impact sur la pratique humaine du jeu ?

Enorme, oui … Maintenant, ces logiciels sont intégrés à la préparation des joueurs professionnels pour mettre au point des combinaisons «surprises». Cela a permis de faire évoluer le jeu vers un style plus ouvert : ainsi, certaines règles stratégiques classiques (contrôle du centre, …) sont battues en brèche, l'aspect tactique devenant très important au détriment parfois de l'esthétique des parties. Le niveau moyen des joueurs a également progressé. L'arrivée des logiciels a eu toutefois des aspects plus négatifs : l'apparition de cas de «triche» durant les tournois, certains joueurs essayant de se faire aider par des ordinateurs sur les coups cruciaux. On observe aussi une diminution de la fréquentation dans les clubs, chacun pouvant jouer de chez soi.

Y a-t-il encore des défis à relever au-delà de l'amélioration du niveau des programmes ?

Après le résultat de 1997, plusieurs chercheurs se sont demandé comment complexifier le jeu d'échec, sans en bouleverser les règles, afin que les êtres humains «reprennent» la main. Plusieurs variantes ont été proposées, je suis l'auteur d'une d'entre elles nommée «Twilight Chess» [1]. L'idée de base est simple : en plus des coups classiques l'utilisateur peut, à la place d'un déplacement retirer l'une de ses pièces du jeu, pour la remettre ultérieurement à la position qu'il souhaite. Ce faisant on passe de 40 coups jouables en moyenne à plus de 150 et les approches *forces brutes* qu'utilisent les ordinateurs sont alors totalement mises en échec.

Quel est l'impact de ces recherches pour l'informatique en général ?

Pour l'instant, c'est assez théorique mais à terme, on peut imaginer, par exemple, donner naissance à de nouvelles méthodes de recherche arborescentes capables de mieux intégrer des notions fondamentales comme *l'initiative* (qui est le joueur actif) et la *quiescence* (qu'est ce qu'une position stable). Dans Twilight Chess, les pièces qui (ré)apparaissent sur l'échiquier peuvent être également examinées sous l'angle de la gestion d'évènements imprévus. Plus généralement, il faut voir qu'historiquement les travaux de recherche sur les échecs ont eu un impact dans toute l'informatique, les méthodes de recherche arborescente ou l'indexation des données bien sûr, mais on peut aussi rappeler que les premiers grands «réseaux sociaux» se sont constitués dès le début des années 90 pour gérer des communautés … de joueurs d'échecs !

[1] *Frédéric Prost: Twilight Chess - A Chess Variant Designed to Rehabilitate Human vs. Computer Challenge. ICGA Journal 32(2): 89-92 (2009).*

L'agenda de LILI

Keynote Speeches

Les Keynote Speeches du LIG invitent des chercheurs de renom international à venir présenter leur vision d'un domaine de l'informatique. Elles ont lieu tous les 1er jeudi du mois à 14h. Tous les renseignements et vidéos des événements sur : **http://www.liglab.fr/spip.php?page=keynotes**

Les prochains rendez-vous

- 5 janvier 2012 : Joseph Hellerstein (Google) – Integrating job management ...
- 2 février 2012 : Marie-Paule Cani (Grenoble INP) – *thème non-communiqué*
- 1 mars 2012 : Jean-Paul Haton (LORIA) – *thème non-communiqué*

Joseph L. Hellerstein – Google — 05/01/2012

Integrating Job Management and Machine Management in Google Data Centers

In cloud based services, many jobs run for months. This is a time frame that is on the order of machine lifetimes. As a result, scheduling the administration of machines (e.g., repairs, adding new machines, draining machines) must be integrated with scheduling jobs. At Google, we approach this problem by using an abstraction called «shapes» to describe both machines and jobs. I will discuss the shapes abstraction, and I will also describe optimization problems that we solve using shapes.

Holger Hermanns – Saarland University, Germany — 01/12/2011

From Concurrency Models to Numbers: Performance, Dependability, Energy

Prof. Holger Hermanns is chair for Dependable Systems and Software, and Dean of the Faculty of Mathematics and Computer Science at Saarland University, Germany. His research interests include modelling and verification of concurrent systems, resource-aware embedded systems, and compositional performance and dependability evaluation. He is a founding member and principal investigator of the German special research initiative AVACS and holder of several other national and European research grants.

Joseph Sifakis – Laboratoire VERIMAG — 03/11/2011

Une vision pour l'Informatique - La perspective système

Joseph Sifakis a fondé le laboratoire Verimag de Grenoble où il travaille encore à présent. Il est le lauréat, avec Edmund Clarke (Carnegie Mellon University) et Allen Emerson (Université du Texas à Austin) du Prix Turing 2007, et le premier Français à recevoir cette distinction. Il a également reçu en 2001 la médaille d'argent du CNRS. Il est l'un des créateurs de la méthode d'énumération et de vérification de modèles (model checking), pour laquelle il a reçu ce prix.

Dr. Wlodek Zadrozny – IBM — 06/10/2011

Building Watson: An overview of the Deep Question Answering project

Dr. Wlodek Zadrozny is a Research Staff Member in the IBM Thomas J. Watson Research Center. He joined IBM in 1985 and worked mostly on applications of natural language processing. In the Watson/DeepQA project he was responsible for acquiring and preparing textual resources. He also worked on applying this question answering technology to help desk support and medical diagnosis. Dr. Zadrozny is an author or co-author of about 40 patents and 50 technical papers.

Lettre d'Information du LIG

L'edito

N°1 – Janvier 2012

En ce début d'année 2012, ce premier numéro inaugure le lancement de la Lettre d'Information du LIG - dite LILI - minutieusement conçue et élaborée par notre très dynamique Mission Communication (MissCom).

Cette publication sera précieuse en interne pour diffuser toutes les informations ayant trait à la vie du laboratoire ; mais elle est destinée également à être un support de communication utile à l'attention de nos partenaires et nos tutelles.

LILI offre ainsi l'opportunité d'annoncer, à l'aide d'un support dédié, divers événements : les keynote speeches, les conférences organisées par des équipes du LIG, les événements à partager, les faits marquants à relever, les distinctions à honorer … LILI, donne aussi la parole aux personnels du laboratoire et aux équipes ; et à ce titre, LILI est aussi un formidable média pour notamment mieux connaître nos missions ou nos métiers respectifs ou encore pour découvrir ou redécouvrir les activités des différentes équipes composant notre grand laboratoire.

Ce premier numéro de LILI présente ainsi une nouvelle équipe du LIG arrivée durant l'année passée, AMA, dirigée par Eric Gaussier, dont la thématique de recherche porte sur les modèles et les algorithmes pour l'apprentissage automatique ; Cette nouvelle équipe est hébergée sur le campus de Grenoble Saint-Martin d'Hères au centre «équation» en attendant la construction du nouveau bâtiment qui accueillera la plupart des équipes du LIG, le laboratoire Verimag et une partie du laboratoire Jean Kuntzmann.

Au-delà de l'événementiel s'inscrivant dans une actualité, je suis convaincu que LILI saura inscrire la trace - autrement intangible - d'une histoire de laboratoire. Meilleurs vœux à toutes et tous pour cette nouvelle année, bonne lecture et merci à Gilles Bisson et à toute son équipe pour leur remarquable travail !

Hervé Martin
Directeur du LIG

Sommaire

- Edito et Sommaire ..... page 1
- Focus sur l'apprentissage : l'équipe AMA ..... page 2/3
  - *Acquérir automatiquement des connaissances*
- Focus sur le LIG ..... page 4
  - *Faits marquants, le LIG en chiffres*
  - *Le pôle d'expérimentation scientifique*
  - *Les HDR de l'année 2011*
- LILI explique ..... page 5
  - *Checker : le premier programme de jeu «apprenant»*
  - *Questions/Réponses : parole à Frédéric Prost (CAPP)*
- L'agenda de LILI ..... page 6
  - *Keynote Speeches*

La «Lettre d'Information du LIG» est une publication du Laboratoire d'Informatique de Grenoble (LIG) - 2012 - Directeur de la publication : Hervé Martin. Rédaction en chef : Gilles Bisson. Ont collaboré à ce numéro : Béatrice Buccio, Nadine Mandran, Céline Coutrix et Pierre Laforgue. Conception et maquette du numéro : Mission Communication (MissCom). Contact de la publication : <misscom.lig@imag.fr>. ISBN xxx-x-xxxx-xxxx-x

Focus sur ... l'équipe AMA

Focus sur le LIG

Présentation

Nom	Equipe AMA (Apprentissage : Modèles et Algorithmes)		
Responsable	Eric Gaussier		
Domaine	Apprentissage Automatique		
Site	**http://www.liglab.fr/spip.php?article917**		
Membres	Permanents	(post)-doctorants	Ingénieurs
	8 (CNRS, UJF, UPMF)	17	1
Mots-clefs	Apprentissage automatique, fouille de textes, modélisation de l'information, données complexes et temporelles, perception, modèle sociaux.		

Les recherches de l'équipe AMA s'inscrivent dans le cadre général de l'apprentissage automatique et de la modélisation de l'information. Notre objectif est de développer des modèles théoriques et des algorithmes sur deux axes qui répondent à des besoins distincts :

(1) modéliser, à des fins d'accès et de prédiction, les grands volumes d'informations complexes et dynamiques,

(2) analyser et simuler des systèmes sociaux complexes.

Ces recherches sont menées par des chercheurs d'horizons variés et de compétences complémentaires : mathématiques appliquées et statistiques, apprentissage automatique, recherche d'information, perception, intelligence artificielle et sciences de la cognition.

L'apprentissage ?

La problématique de l'*apprentissage automatique* est apparue dans la fin des années 50, conjointement avec l'essor de *l'intelligence artificielle*. Le but de l'apprentissage est d'étudier l'adaptation des programmes informatique à leur environnement par l'acquisition, de façon autonome, de connaissances. Dès les origines, deux axes de recherche principaux sont apparus :

- l'apprentissage visant à faire de la modélisation et de la résolution de problèmes (le «Machine Learning » anglo-saxon), dont la finalité est voisine de celle de *l'analyse des données*,
- l'apprentissage en tant qu'étude et modélisation de mécanismes cognitifs ; axe entretient des liens très étroits avec la recherche en *sciences cognitives*.

Durant ces dernières années, en réponse à l'augmentation en nombre et en complexité des données et connaissances à traiter, l'intérêt de disposer de processus d'analyse automatiques et adaptatifs s'est considérablement accru. Ainsi les domaines d'application utilisant les outils de l'apprentissage ne cessent d'augmenter. Parmi les plus actifs en 2012 on peut citer l'analyse de réseaux sociaux, l'étude des usages du Web, la classification et catégorisation de textes en langage naturel, l'apprentissage d'ontologies, la bio-informatique et la chemo- informatique, la surveillance de données sensibles, l'aide à la décision et à la gouvernance, la robotique, la reconnaissance de formes, etc.

Faits marquants

Personnes distinguées

- Jim Crowley (PRIMA) : Membre senior IUF
- Marie-Christine Rousset (HADAS) : Membre senior IUF, Chevalier de l'Ordre national du Mérite
- Yves Demazeau (MAGMA) : Président de l'AFIA (*Association Française pour l'Intelligence Artificielle*), ECCAI Fellow 2010
- Hubert Garavel (VASY) : Prix Gay Lussac-Humboldt
- Derrick Kondo (MESCAL) : Google Research Award
- Denis Trystram (MOAIS) : Lauréat du partenariat entre Google et le CNRS, IUF 2010
- Christine Collet (HADAS) : Chevalier de l'Ordre national du Mérite
- Brigitte Plateau (MESCAL) : Chevalier de la Légion d'honneur

Équipes distinguées

- ADELE est partenaire de «OpenTheBox» qui est l'un des 6 projets retenus par le ministère pour l'appel national Programme d'Investissements d'Avenir.
- STEAMER, l'INED et Géographie-Cités ont conçu et développé une application Web proposant des cartes interactives de la population mondiale.
- IIHM a produit une application iPhone/iPad pour l'affichage 3D.

Faits marquants du laboratoire

Création d'un Laboratoire International Associé (LIA) du CNRS : *Laboratoire International en Calcul Intensif et Informatique Ambiante (LICIA)* : **http://www.licia-lab.org**

le LIG à l'international – les unités partenaires –

- IPAL (Image, Perception, Access and Language) à Singapour, dont Jean-Pierre Chevallet, membre du LIG a été co-directeur pendant 5 ans.
- LAFMIA (Laboratoire Franco-Mexicain d'Informatique et d'Automatique) au Mexique, dont Genoveva Vargas-Solar, membre du LIG, est directrice adjointe.
- MICA (Multimédia, Informations, Communications et Applications) à Hanoï, dont Eric Castelli, ex-membre du LIG, est co-directeur.

LIG en chiffres

Le laboratoire LIG compte 600 personnes (chiffre de mai 2011), se répartissant en :

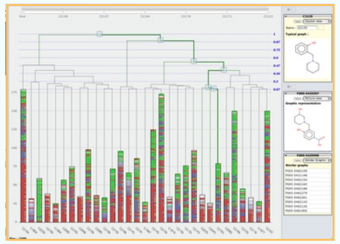
- 188 enseignants-chercheurs/chercheurs
- 45 ITA/IATOS
- 159 doctorants répartis dans 23 équipes
- 122 non permanents (visiteurs, ...)
- 86 contractuels

Travaux de recherche

T1 : Modèles théoriques pour le traitement de données complexes

Les recherches que nous menons dans cette thématique se situent dans le cadre de l'apprentissage automatique, de la modélisation de l'information et de l'analyse des données. Nous nous intéressons dans ce cadre à la définition de propriétés formelles que les systèmes d'accès à l'information doivent satisfaire et à la formalisation de la notion de «pertinence» ainsi qu'à la définition de mécanismes de sélection d'information appropriés.

Nous travaillons également sur la définition et l'apprentissage d'indices de proximité permettant de quantifier les ressemblances



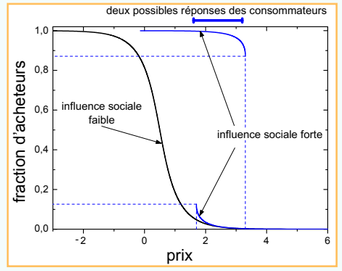
entre données, et plus particulièrement les données relationnelles (arbres, graphes et hyper-graphes notamment) et les séquences temporelles. Ce travail, dont l'un des objectifs est de définir un formalisme unifié des mesures existantes en apprentissage et en analyse des données, nous a déjà conduit à proposer de nouveaux algorithmes d'apprentissage pour différentes distances et similarités : cosinus généralisé, similarités en co-classification, distances entre séries temporelles.

Une autre problématique importante est liée à l'apprentissage en très grandes dimensions. Les systèmes d'apprentissage actuels doivent être aptes à manipuler de grandes masses de données : certains systèmes de catégories, comme DMOZ, possèdent plusieurs centaines de milliers de catégories, organisées hiérarchiquement. Afin de sélectionner la meilleure stratégie de catégorisation dans de tels systèmes, il importe de revoir les résultats existants (par exemple, les bornes sur l'erreur de généralisation) dans un cadre où plusieurs systèmes de catégories sont en concurrence.

T2 : Apprentissage et perception

Beaucoup de systèmes intelligents évoluent dans un environnement complexe et les tâches qu'ils doivent résoudre nécessitent de modéliser cet environnement. Dans certains cas, lorsque l'environnement est fortement dynamique, la résolution du problème ne peut être pensée dans l'a priori et fait collaborer des processus hétérogènes et autonomes (e.g. acteurs humains et artificiels). Une autre spécificité des applications considérées est le manque de corpus annotés. Nous nous appuyons dans ce cas sur des modèles de collaboration entre agents qui permettent d'obtenir à la volée des éléments de solution, afin de faire progresser la résolution ou l'apprentissage sur d'autres parties des données.

Une autre partie de notre activité sur la modélisation de l'environnement porte sur l'utilisation de techniques de classification à partir de données brutes fournies par un capteur (typiquement des nuages de points 3D fournis par un capteur laser) ainsi que sur la reconnaissance d'objets mobiles avec des



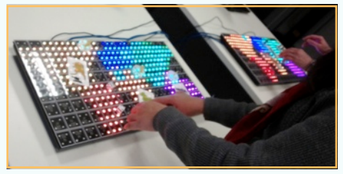
capteurs télémétriques et des caméras, en distinguant la reconnaissance de catégories d'objets génériques et la reconnaissance d'instances d'objets particuliers.

Enfin, nous nous intéressons dans cet axe à la visualisation de données structurées. En effet, la classification ascendante apporte une aide efficace dans l'analyse et la modélisation de données. Toutefois, lorsque l'on travaille avec des arbres contenant plus de quelques centaines de feuilles, la représentation et l'exploration visuelle de la classification deviennent difficiles. Nous avons introduit ces dernières années un nouveau modèle de visualisation, les «Stacked Trees», qui permettent de représenter simultanément à l'écran les informations contenues dans plusieurs dizaines de milliers d'objets.

T3 : Apprentissage pour la modélisation de systèmes sociaux

Le comportement macroscopique ou global des systèmes sociaux dépend des caractéristiques des individus (leurs préférences idiosyncratiques, leurs capacités d'apprentissage et leurs types d'interactions) d'une manière qui n'est pas prédictible à partir des données individuelles. Dans nos travaux théoriques, il s'agit de dépasser les modèles avec des interactions positives et symétriques (où les individus ont tendance à adopter un comportement mimétique ou de conformité avec les autres), pour traiter des cas comme celui des biens publics où les interactions ne sont pas symétriques. Nous arrivons ainsi à prédire des cycles dans l'évolution du système ou des effets d'hystérésis en fonction des politiques publiques. Nous travaillons également sur des données empiriques, par exemple l'évolution de la délinquance et sa modélisation.

Enfin, nous nous intéressons aussi dans ce



cadre à l'étude et la modélisation des réseaux sociaux sur le web, et plus particulièrement à la diffusion d'information dans ces réseaux et à la modélisation de leur dynamique. Nous avons dans ce cadre proposé un nouveau modèle de diffusion de l'information qui repose sur plusieurs facteurs : contenu de l'information diffusée, topologie du réseau, propension de chaque utilisateur à diffuser.

Retrouvez la liste complète des publications de l'équipe sur Pistou : **http://pistou.imag.fr/teams/show/32**

Pôle d'expérimentations

Le pôle d'expérimentation scientifique du LIG, ou pôle d'ingénierie multidisciplinaire, permet aux ingénieurs chargés du soutien à la recherche de partager leurs expertises spécifiques dans 6 grands domaines :

- Intelligence ambiante, réseaux de capteurs, robotique, interaction.
- Ingénierie des logiciels, ingénierie linguistique.
- Calcul haute performance, grilles.
- Usages et médiation, bâtiment intelligent, ingénierie RFI.
- Analyse sonore et multimodale.
- Informatique en socio-économie.

Le pôle comprend également une cellule d'appui transversale, Marvelig, en charge de :

- La virtualisation, les intergiciels, la pérennisation des prototypes.
- L'expérimentation sur panels d'utilisateurs, les méthodes statistiques et qualitatives.

HDR 2011

- Alan Schmitt (SARDES) – Analyse statique pour manipulations de données structurées hiérarchiquement.
- Michel Vacher (GETALP) – Analyse sonore et multimodale dans le domaine de l'assistance à domicile.
- Gwen Salaün (VASY) – Process Calculi : Bridges and Application to Distributed Systems.
- Agnès Front (SIGMA) – Réutilisation et socialisation dans l'ingénierie des méthodes.
- Sophie Dupuy-Chessa (IIHM/SIGMA) – Modélisation en Interaction Homme-Machine et Système d'Information : à la croisée des chemins.