

30 ans d'IA au LIFO

Christel Vrain
LIFO - Université d'Orléans
christel.vrain@univ-orleans.fr

11 mars 2019

Le LIFO (Laboratoire d'Informatique Fondamentale d'Orléans) et la fédération ICVL accueillent les 27, 28 et 29 mai à Orléans les journées plénières du GDR IA sur les aspects algorithmiques et formels de l'Intelligence Artificielle ¹. L'Intelligence Artificielle est devenue un domaine d'actualité avec les progrès remarquables accomplis ces dernières années. Derrière les réalisations les plus retentissantes, se cachent principalement des techniques issues de l'Apprentissage Automatique, plus connu sous le terme anglais de *Machine Learning*. Mais l'Intelligence Artificielle, c'est bien plus... et à titre d'exemple on peut citer les thématiques abordées dans le GDR IA : la représentation des connaissances et la modélisation des raisonnements, la gestion de l'incertitude avec en particulier les modèles graphiques comme les réseaux bayésiens, la modélisation de problèmes en Programmation par Contraintes ou en SAT, la planification, les systèmes multi-agents et la prise de décision collective, sans oublier l'apprentissage automatique. A ces travaux fondateurs s'ajoutent les domaines d'application phares tels que le traitement du langage naturel, la reconnaissance de la parole, le traitement d'images, la vision par ordinateur... avec ses applications actuelles comme les assistants vocaux, les voitures autonomes, la détection de fraudes...

Au LIFO, l'Intelligence Artificielle a toujours été présente. A la création du laboratoire d'informatique dans les années 80, sous l'impulsion de Bernard Lorho, les activités de recherche concernaient principalement la définition et l'implantation des langages de programmation. Cependant l'Intelligence Artificielle y avait déjà sa place avec trois équipes travaillant sur la Programmation Logique (Gérard Ferrand), la Démonstration Automatique de Théorèmes (Siva Anantharaman) et le Traitement des Langues Naturelles (Jean-Claude Bassano).

La Programmation Logique est un paradigme de programmation fondé sur une définition logique des relations, avec le même pouvoir de calcul que les langages de programmation classiques. Un programme logique s'écrit comme un ensemble de clauses : par exemple la clause $m\grave{e}re(X, Y) : \neg parent(X, Y), femme(X)$ se lit *X est la mère de Y si X est parent de Y et X est une femme*. Les travaux à l'époque portaient principalement sur la sémantique de tels langages (e.g. une fois défini un ensemble de clauses, que peut-on inférer de cet ensemble ? quels sont les faits vrais et les faits faux ?) et sur le diagnostic d'erreurs [14, 15]. Cela s'inscrivait bien dans les travaux du laboratoire axés principalement sur les langages de programmation. Les travaux sur cette thématique ont ensuite évolué vers la Programmation Logique avec Contraintes intégrant des contraintes, par exemple arithmétiques et permettant

1. <http://www.gdria.fr>

de représenter élégamment des relations numériques entre variables. Par exemple la formule $pair(X) : -X = Y + 2 \square pair(Y)$. exprime qu'un nombre X est pair si un nombre Y est pair, sous la contrainte $X = Y + 2$ [16].

La démonstration automatique de théorèmes comme son nom l'indique s'intéressait aux méthodes permettant d'automatiser le raisonnement. Les travaux étudiés au LIFO portaient sur la démonstration automatique dans les théories équationnelles, en général fondé sur des techniques de réécriture. L'idée était de concevoir et de réaliser des outils de preuve, afin d'aider les démarches formelles de validation et vérification. Cette thématique fut en particulier développée dans le cadre des programmes logiques : les programmes logiques avec négation et les programmes logiques contraints pouvaient être transformés en un programme de réécriture [1] et ainsi être étudiés sous cet angle.

L'équipe TAL (*Traitement automatique du Langage Naturel*) travaillait principalement sur la thématique de la recherche documentaire avec le développement des systèmes multi-experts DIALECT et DIALECT2 : les requêtes en langage naturel étaient alors transformées en triplets appelées *lexis*. Toujours dans le cadre de ce système, une architecture à base de réseaux de neurones avait été proposée pour une analyse morpho-syntaxique du français [2].

Dans les années 90, une nouvelle thématique a été développée au LIFO autour des Bases de Données Déductives (Dominique Laurent) et de l'Apprentissage Automatique (Christel Vrain). En Apprentissage, on distingue couramment deux problématiques : l'*apprentissage supervisé* où les concepts à apprendre sont connus (par exemple, la prédisposition à une maladie) et l'on dispose d'exemples positifs et négatifs illustrant ces concepts et l'*apprentissage non supervisé* où il faut découvrir les concepts sous-jacents aux observations (par exemple identifier des segments de marché à partir d'une population de clients). Les travaux se situaient en Programmation Logique Inductive (PLI), sous domaine de l'Apprentissage Symbolique Automatique dans lequel les connaissances ou modèles appris sont exprimés en logique (sous forme de programmes logiques) afin d'être plus facilement interprétables par un expert. Contrairement à la plupart des travaux menés en Machine Learning où les données sont en général décrites dans une unique table, par exemple une table de patients avec leurs symptômes et leur diagnostic, la Programmation Logique Inductive permet de prendre en compte des données relationnelles. Par exemple si on dispose d'informations sur des personnes, - leur rôle, leur liste de publications, leur relation,- on peut apprendre des connaissances de la forme $directeur(X, Y) : -professeur(X), \acute{e}tudiant(Y), m\grave{e}mePubli(X, Y)$, exprimant que si deux personnes publient ensemble, l'une étant professeur et l'autre étudiant alors la première dirige les travaux de la seconde. L'une des originalités des travaux de l'époque au LIFO était de se placer dans des sémantiques déclaratives (sémantique de Fitting, sémantique bien fondée) et de spécifier des méthodes d'apprentissage garantissant la correction des programmes appris par rapport à la sémantique attendue définie par l'expert [24]. Ces travaux s'enrichissaient des travaux menés en parallèle dans l'équipe Programmation Logique autour des sémantiques déclaratives et du diagnostic d'erreurs. Ils ont ensuite été étendus à l'apprentissage de contraintes dans le cadre de la Programmation Logique avec Contraintes [20].

Depuis le début des années 2000, les activités en Démonstration Automatique de Théorèmes se sont orientées vers la vérification et les travaux en Programmation Logique ont disparu au profit des Contraintes. Les recherches en Intelligence Artificielle ont été regroupées dans

l'équipe Contraintes et Apprentissage et portent principalement sur 3 axes : d'une part l'apprentissage automatique (Machine Learning) et la fouille de données (Data Mining), d'autre part le traitement de la langue et enfin les contraintes.

Concernant les contraintes, les travaux se sont déclinés autour de 4 thèmes : la résolution de contraintes dans des structures d'arbres [6, 12], l'apprentissage de solveurs de contraintes de domaines finis, la programmation par contraintes pour le traitement automatique des langues et la propagation par contraintes pour le Data Mining. Les travaux actuels se fondent principalement sur la Programmation par Contraintes, permettant de modéliser un problème par des variables et des contraintes sur ces variables. Son mécanisme de base est la propagation de contraintes qui consiste à supprimer des valeurs inconsistantes du domaine des variables de la contrainte, pour réduire l'espace de recherche. L'efficacité des solveurs dépend donc fortement de la capacité des propagateurs associés à chaque contrainte et nous avons proposé des méthodes de construction automatique de propagateurs pour des contraintes de table, représentées par leur ensemble de solutions, s'appuyant sur des méthodes d'Apprentissage Automatique [9, 19]. Récemment, un nouvel axe a été développé à la croisée entre Contraintes et Apprentissage et visant à modéliser des problèmes d'apprentissage sous contraintes utilisateur (principalement du clustering) en Programmation par Contraintes [7, 18, 8].

Les grammaires de propriétés (GP) constituent un formalisme permettant de décrire la langue naturelle en termes de contraintes locales, appelées propriétés. Ces propriétés peuvent être satisfaites ou violées indépendamment les unes des autres, ce qui rend possible la description d'énoncés agrammaticaux et également une notion de grammaticalité. Nous avons proposé une sémantique formelle des GP en théorie des modèles et nous avons développé un cadre en programmation par contraintes pour modéliser et résoudre le problème de l'analyse syntaxique en GP sous forme d'un problème d'optimisation sous contraintes. Ce cadre permet non seulement d'analyser des énoncés grammaticalement corrects mais aussi de trouver des structures syntaxiques les plus adaptées pour des énoncés agrammaticaux [13].

Les travaux actuels en TAL portent principalement sur la production automatique et semi-automatique d'outils et de ressources pour le TAL, notamment au niveau morphologique, syntaxique et sémantique et sur l'annotation de corpus oraux. Un effort particulier a été apporté sur la syntaxe de la langue : que ce soit dans sa modélisation, et pour l'étude de sa complexité (développement des méta-grammaires, et des grammaires de propriété), ou dans la visée d'applications reposant sur des techniques de Machine Learning et applicables immédiatement sur des données réelles (oral).

Les travaux en Apprentissage Automatique se sont diversifiés. Par exemple, une collaboration avec le BRGM a conduit l'équipe à s'intéresser à la recherche de motifs fréquents, en particulier dans des Systèmes d'Information Géographiques [23]. Les approches jusque là principalement symboliques ont du être adaptées pour traiter des données numériques [21, 22]. Le traitement des données incertaines a conduit à étudier l'Apprentissage Relationnel Statistique [11] permettant de coupler Programmation Logique Inductive et modélisation probabiliste. Les thématiques actuellement abordées en Machine Learning et Data Mining sont diverses : classification non supervisée [5, 7, 18, 8], classification supervisée [3, 17], apprentissage séquentiel/incrémental, apprentissage dans des représentations logiques (Programmation Logique Inductive), apprentissage de modèles de structuration (prétopologique) [4] et recherche de motifs fréquents [21, 22], contraintes discrètes pour le Data Mining [7, 18, 8]. Ils sont appliqués à la recherche d'informations, à la bio-informatique [10], aux réseaux sociaux et aux humanités numériques.

Pour conclure, notons que les travaux menés dans l'équipe Contraintes et Apprentissage, autour de l'Apprentissage Automatique, des Contraintes et du Traitement Automatique de la Langue, s'appuient (ou se sont appuyées) sur des collaborations locales (LLL, PRISME, ICOA, BRGM, INRA, ...), régionales (LIFAT, Université de Tours), nationales (GREYC, Université de Caen; ICube, Université de Strasbourg; AgroParisTech, ...) et internationales (Brésil, Californie, ...).

Sources et contributeurs : Ce panorama rapide des travaux menés au LIFO autour de l'Intelligence Artificielle s'appuie sur des rapports d'activités du LIFO, un historique du LIFO réalisé par G. Hains (directeur du LIFO de 2000 à 2005) et des contributions ou relectures de Sylvie Billot, Thi Bich Hanh Dao, Anaïs Lefeuvre-Halftermeyer et Marcilio Pereira de Souto.

Références

- [1] Siva Anantharaman and Gilles Richard. A rewrite mechanism for logic programs with negation. *Theor. Comput. Sci.*, 192(1) :77–106, 1998.
- [2] Dominique Archambault and Jean-Claude Bassano. A neural network for supervised learning of natural language grammar. In *Sixth International Conference on Tools with Artificial Intelligence, ICTAI '94, New Orleans, Louisiana, USA, November 6-9, 1994*, pages 267–273, 1994.
- [3] Victor H. Barella, Luís P. F. Garcia, Marcilio C. P. de Souto, Ana Carolina Lorena, and André C. P. L. F. de Carvalho. Data complexity measures for imbalanced classification tasks. In *2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–8, 2018.
- [4] Guillaume Cleuziou and Gaël Dias. Learning pretopological spaces for lexical taxonomy acquisition. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part II*, pages 493–508, 2015.
- [5] Guillaume Cleuziou and Jose G. Moreno. Kernel methods for point symmetry-based clustering. *Pattern Recognition*, 48(9) :2812–2830, 2015.
- [6] Alain Colmerauer and Thi-Bich-Hanh Dao. Expressiveness of full first-order constraints in the algebra of finite or infinite trees. *Constraints*, 8(3) :283–302, 2003.
- [7] Thi-Bich-Hanh Dao, Khanh-Chuong Duong, and Christel Vrain. Constrained clustering by constraint programming. *Artif. Intell.*, 244 :70–94, 2017.
- [8] Thi-Bich-Hanh Dao, Chia-Tung Kuo, S. S. Ravi, Christel Vrain, and Ian Davidson. Descriptive clustering : ILP and CP formulations with applications. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 1263–1269, 2018.
- [9] Thi-Bich-Hanh Dao, Arnaud Lallouet, Andrei Legtchenko, and Lionel Martin. Indexical-based solver learning. In *Principles and Practice of Constraint Programming - CP 2002, 8th International Conference, CP 2002, Ithaca, NY, USA, September 9-13, 2002, Proceedings*, pages 541–555, 2002.

- [10] Marcílio Carlos Pereira de Souto, Pablo A. Jaskowiak, and Ivan G. Costa. Impact of missing data imputation methods on gene expression clustering and classification. *BMC Bioinformatics*, 16 :64 :1–64 :9, 2015.
- [11] Quang-Thang Dinh, Matthieu Exbrayat, and Christel Vrain. Generative structure learning for markov logic networks based on graph of predicates. In *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 1249–1254, 2011.
- [12] Khalil Djelloul, Thi-Bich-Hanh Dao, and Thom W. Frühwirth. Theory of finite or infinite trees revisited. *TPLP*, 8(4) :431–489, 2008.
- [13] Denys Duchier, Thi-Bich-Hanh Dao, and Yannick Parmentier. Model-theory and implementation of property grammars with features. *Journal of Logic and Computation*, 24(2) :491–509, 2014.
- [14] Gérard Ferrand. Error diagnosis in logic programming, an adaption of e. y. shapiro’s method. *J. Log. Program.*, 4(3) :177–198, 1987.
- [15] Gérard Ferrand and Pierre Deransart. Proof method of partial correctness and weak completeness for normal logic programs. *J. Log. Program.*, 17(2/3&4) :265–278, 1993.
- [16] Gérard Ferrand and Arnaud Lallouet. A logic program characterization of domain reduction approximations in finite domain csps. In *Logic Programming, 18th International Conference, ICLP 2002, Copenhagen, Denmark, July 29 - August 1, 2002, Proceedings*, pages 478–479, 2002.
- [17] Luís P. F. Garcia, Ana Carolina Lorena, Marcilio C. P. de Souto, and Tin Kam Ho. Classifier recommendation using data complexity measures. In *24th International Conference on Pattern Recognition, ICPR 2018, Beijing, China, August 20-24, 2018*, pages 874–879, 2018.
- [18] Chia-Tung Kuo, S. S. Ravi, Thi-Bich-Hanh Dao, Christel Vrain, and Ian Davidson. A framework for minimal clustering modification via constraint programming. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 1389–1395, 2017.
- [19] Arnaud Lallouet, Thi-Bich-Hanh Dao, Andrei Legtchenko, and AbdelAli Ed-Dbali. Finite domain constraint solver learning. In *IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003*, pages 1379–1380, 2003.
- [20] Lionel Martin and Christel Vrain. Learning linear constraints in inductive logic programming. In *Machine Learning : ECML-97, 9th European Conference on Machine Learning, Prague, Czech Republic, April 23-25, 1997, Proceedings*, pages 162–169, 1997.
- [21] Ansaf Salleb-Aouissi, Christel Vrain, and Cyril Nortet. Quantminer : A genetic algorithm for mining quantitative association rules. In *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, pages 1035–1040, 2007.
- [22] Ansaf Salleb-Aouissi, Christel Vrain, Cyril Nortet, Xiangrong Kong, Vivek Rathod, and Daniel Cassard. Quantminer for mining quantitative association rules. *Journal of Machine Learning Research*, 14(1) :3153–3157, 2013.
- [23] Teddy Turmeaux, Ansaf Salleb, Christel Vrain, and Daniel Cassard. Learning characteristic rules relying on quantified paths. In *Knowledge Discovery in Databases : PKDD*

2003, *7th European Conference on Principles and Practice of Knowledge Discovery in Databases, Cavtat-Dubrovnik, Croatia, September 22-26, 2003, Proceedings*, pages 471–482, 2003.

- [24] Christel Vrain and Lionel Martin. Inductive learning of normal clauses. In *Machine Learning : ECML-94, European Conference on Machine Learning, Catania, Italy, April 6-8, 1994, Proceedings*, pages 435–438, 1994.