THÈSE

présentée à

L'Institut National Polytechnique de Grenoble - I.N.P.G.
Laboratoire LEIBNIZ - IMAG

pour obtenir le titre de

DOCTEUR

Spécialité:

Informatique
(Arrêté ministériel du 30 mars 1992)

INSS : UN SYSTEME HYBRIDE NEURO-SYMBOLIQUE POUR L'APPRENTISSAGE AUTOMATIQUE CONSTRUCTIF

par

Fernando Santos Osório

Soutenue le 3 février 1998 devant le jury composé de:

MM.	Philippe JORRAND	Directeur
	Alain GRUMBACH	Rapporteur
	Christian PELLEGRINI	Rapporteur, Président
	Bernard AMY	Examinateur (co-directeur)
	Frédéric ALEXANDRE	Examinateur
	Vincent RIALLE	Examinateur

A Denise et Felipe,

mes deux compagnons de toutes les heures

embarqués dans cette aventure avec moi.

Remerciements:

Je tiens tout d'abord à remercier Bernard Amy, qui m'a accepté dans son équipe, a consacré beaucoup de son temps à mon travail, et a fait de nombreuses remarques, toujours intéressantes, sur cette thèse. Bernard a toujours su diriger mes recherches dans la bonne direction; je dois beaucoup à lui pour ses nombreuses contributions à la réussite de cette thèse.

Je tiens à remercier aussi Alain Grumbach, Professeur à l'ENST de Paris, et à Christian Pellegrini, Professeur à l'Université de Genève, pour avoir accepté de juger ce travail.

Je tiens également à remercier Vincent Rialle, Maître de Conférences à l'Université Joseph Fourrier de Grenoble, et à Fréderic Alexandre, Chercheur à l'INRIA Lorraine, qui ont gentiment accepté de participer à mon jury de thèse.

Merci beaucoup à Philippe Jorrand, Directeur du Laboratoire Leibniz - IMAG/INPG, qui a accepté de m'accueillir dans ce laboratoire, et aussi, qui a accepté de passer toutes ces années à mon côté, comme mon directeur de thèse.

Je remercie à la CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior) et plus particulièrement au programme de coopération Franco-Brésilien CAPES-COFECUB, de m'avoir accordé un financement pour la réalisation de cette thèse. Le financement ainsi que le support administratif fourni par ces deux organismes a été une pièce fondamentale pour permettre mon séjour en France et la réalisation de cette thèse. Merci à Brigitte Plateau et Claudio Geyer, coordinateurs du projet de coopération CAPES-COFECUB entre l'UFRGS et l'INPG, de m'avoir permis de participer de son projet de coopération. Je tiens aussi à remercier particulièrement à Philippe Navaux de l'UFRGS.

Je remercie aussi à l'UNISINOS (Universidade do Vale do Rio dos Sinos, Brésil), pour m'avoir accordé un détachement de mes fonctions d'enseignant de 1993 à 1998, et principalement, pour son soutien constant qui m'a permis de mener mes travaux de doctorat jusqu'à la fin.

Je tiens aussi à remercier :

Les membres des laboratoires LIFIA et Leibniz qui m'ont beaucoup aidé dans la réalisation de cette thèse. Plus particulièrement, je remercie aux 'anciens de l'équipe' Arnaud Giacometti, Bruno Orsier, Abderrahim Labbi, Nicolas Szilas, Maria Malek, Françoise Fessant et Emmanuel Duriez pour leurs inestimables contributions sur le plan scientifique et aussi pour leur amitié. Je remercie aussi à tous ceux qui ont travaillé a mon côté dans le développement et l'utilisation du système INSS (Loïc Decloedt, Pierre-Yves Blond, Gerardo Reyes et Benjamin Deuker), ainsi qu'à Vincent Danel et Vincent Rialle pour leurs contributions dans l'application médicale. Je remercie particulièrement à Henri Saya, notre ingénieur système, à Jaime Sichman et à Remis Balaniuk, qui m'ont beaucoup aidé à "m'insérer" dans ce laboratoire d'un point de vue technique et aussi humain. Je remercie ~~beaucop~~ beaucoup ;^) aux *héros* qui ont fait la révision de ce document, les critiques sur son contenu, et la correction de mes innombrables fautes commises dans la langue de Molière: Nico, Eric Gauthier, Daniel Memmi, Sophie Rouzier et Bernard.

Tous les membres de l'équipe RESEAUX dirigée par Bernard Amy au laboratoire Leibniz pour leur accueil très chaleureux et leur disponibilité constante, avec qui j'ai eu des rapports aussi divers qu'enrichissants. *Une thèse est principalement le résultat d'un vrai travail d'équipe!*

Merci aussi à tous mes amis des autres groupes de recherche de Grenoble : l'équipe LAPLACE (en particulier à Olivier Lebeltel pour sa inestimable aide avec le petit Khepera), les équipes MAGMA et PRIMA, les amis des laboratoires TIMA, TIMC et LMC, aux membres de l'association InCognito, et tant d'autres compagnons de recherche et d'amitié.

Merci à Scott Fahlman, à Geoffrey Towell, à Olivier Michel, et à tous les autres chercheurs qui ont mis leurs outils, articles et travaux accessibles à travers l'Internet, et merci à tous les membres du projet Européen MIX. Vos contributions ont beaucoup apporté au développement de cette thèse.

Et bien sûr, merci à toute la communauté brésilienne de Grenoble et à tous mes amis avec qui j'ai garde des souvenirs très agréables des bonnes moments vécus ensemble. Nos échanges de messages (et des "flames") à travers la "penduick-fm" resteront à jamais enregistrés dans la mémoire non-volatile des nos âmes... "SaluTCHE et Obrigado!".

Parmi la communauté brésilienne de Grenoble, je garderai un souvenir très particulier de tous ceux qui ont été toujours prêts à m'encourager dans ce long parcours : l'unique et inestimable João Paulo Kitajima, Silvio Nabeta, Ricardo Duarte et Silvana, Alvaro Guarda et Vera, Gilson Braviano et Teresinha, Paulo Fernandes et Marilia, José Celso et Néia, Gilberto Marchioro (Giba) et Camila, Alexandre Ribeiro et Helena.

Je remercie aussi à tous ceux et celles qui sont là-bas de l'autre côté du monde, au Brésil, et qui ont donné leur soutien à la réalisation de cette thèse. Merci particulièrement à mes beaux-parents Egydio et Eneida, à mes grands-parents Luiz Carlos et Maria, et à ma mère Beatriz, je n'aurais pas pu arriver jusqu'ici sans vous.

Finalement, merci, merci et encore merci à mon épouse qui a toujours su me soutenir (dans tous les sens de ce mot, et principalement dans le sans brésilien - "suportar, agüentar") et me donner les forces nécessaires pour arriver jusqu'à la fin de cette thèse. Je n'aurais jamais écrit chacune des 300 pages qui composent cette thèse sans son inestimable aide et amour. DeniSe, ma vraie DS à moi, tu es for me... formidable!

Merci à Denise et Felipe, sources de mon inspiration.

Merci à toute ma famille et plus particulièrement à mes parents Beatriz et Luiz Carlos.

Résumé:

Plusieurs méthodes ont été développées par l'Intelligence Artificielle pour reproduire certains aspects de l'intelligence humaine. Ces méthodes permettent de simuler les processus de raisonnement en s'appuyant sur les connaissances de base disponibles. Chaque méthode comporte des points forts, mais aussi des limitations. La réalisation de systèmes hybrides est une démarche courante qui permet de combiner les points forts de chaque approche, et d'obtenir ainsi des performances plus élevées ou un champ d'application plus large. Un autre aspect très important du développement des systèmes hybrides intelligents est leur capacité d'acquérir de nouvelles connaissances à partir de plusieurs sources différentes et de les faire évoluer.

Dans cette thèse, nous avons développée des recherches sur les systèmes hybrides neuro-symboliques, et en particulier sur l'acquisition incrémentale de connaissances à partir de connaissances théoriques (règles symboliques) et de connaissances empiriques (exemples). Un nouveau système hybride, nommé système INSS - *Incremental Neuro-Symbolic System,* a été étudié et réalisé. Ce système permet le transfert de connaissances déclaratives (règles symboliques) d'un module symbolique vers un module connexionniste (réseau de neurones artificiel - RNA) à travers un convertisseur de règles en réseau. Les connaissances du réseau ainsi obtenu sont affinées par un processus d'apprentissage à partir d'exemples. Ce raffinement se fait soit par ajout de nouvelles connaissances, soit par correction des incohérences, grâce à l'utilisation d'un réseau constructif de type Cascade-Correlation. Une méthode d'extraction incrémentale de règles a été intégrée au système INSS, ainsi que des algorithmes de validation des connaissances qui ont permis de mieux coupler les modules connexionniste et symbolique. Le système d'apprentissage automatique INSS a été conçu pour l'acquisition constructive (incrémentale) de connaissances. Le système a été testé sur plusieurs applications, en utilisant des problèmes académiques et des problèmes réels (diagnostic médical, modélisation cognitive et contrôle d'un robot autonome). Les résultats montrent que le système INSS a des performances supérieures et de nombreux avantages par rapport aux autres systèmes hybrides du même type.

Abstract:

Various Artificial Intelligence methods have been developed to reproduce intelligent human behaviour. These methods allow to reproduce some human reasoning process using the available knowledge. Each method has its advantages, but also some drawbacks. Hybrid systems combine different approaches in order to take advantage of their respective strengths. These hybrid intelligent systems also present the ability to acquire new knowledge from different sources and so to improve their application performance.

This thesis presents our research in the field of hybrid neuro-symbolic systems, and in particular the study of machine learning tools used for constructive knowledge acquisition. We are interested in the automatic acquisition of theoretical knowledge (rules) and empirical knowledge (examples). We present a new hybrid system we implemented: INSS - Incremental Neuro-Symbolic System. This system allows knowledge transfer from the symbolic module to the connectionist module (Artificial Neural Network - ANN), through symbolic rule compilation into an ANN. We can refine the initial ANN knowledge through neural learning using a set of examples. The incremental ANN learning method used, the Cascade-Correlation algorithm, allows us to change or to add new knowledge to the network. Then, the system can also extract modified (or new) symbolic rules from the ANN and validate them. INSS is a hybrid machine learning system that implements a constructive knowledge acquisition method. We conclude by showing the results we obtained with this system in different application domains: ANN artificial problems (The Monk's Problems), computer aided medical diagnosis (Toxic Comas), a cognitive modelling task (The Balance Scale Problem) and autonomous robot control. The results we obtained show the improved performance of INSS and its advantages over others hybrid neuro-symbolic systems.

TABLE DES MATIERES

TABLE DES FIGURES