White Paper

Approche xtractis®

pour la modélisation prédictive robuste
et l'optimisation multi-objectifs de processus complexes

Auteurs : Prof. Z. Zalila, J. Cuquemelle, C. Penet, A. Chikh, B. Lorentz, D. Deschamps, C. Assemat, S. Marbach, G. Gueydan, C. Leroux
xtractis@intellitech.fr

Première publication : 02/2008
Révision : v3.2 – 12/2013

INTELLITECH – 14 rue du Fonds Pernant – Technopolis III – 60200 Compiègne, FRANCE
Tel: +33 3 44 23 48 – Web: www.intellitech.ai – www.xtractis.ai
SARL au capital de 206 400 Euros – R.C.S. Compiègne B 419 510 128 – SIRET 419 510 128 00020 – Code APE 731Z

Avertissement

Zalila, Z., Cuquemelle, J., Penet, C., Chikh, A., Lorentz, B., Deschamps, D., Assemat, C., Marbach. S., Gueydan, G., Leroux, C. (2008-2013) *Approche x**tractis**® pour la modélisation prédictive robuste et l'optimisation multi-objectifs de processus complexes*, White Paper, v3.2, **intelli**tech, Compiègne, France, décembre 2013, 18 p, http://xtractis.ai/fr/approche-mathematique-algorithmique-unique/

Sommaire

1. INTRODUCTION

Les systèmes d'inférence floue sont des outils permettant de modéliser facilement et intuitivement un processus de prise de décision, permettant à un opérateur humain de formuler un avis ou une évaluation, de poser un diagnostic, de déterminer un classement ou une classification. La prise de décision s'effectue en mettant en relation des entrées – l'état de connaissance sur la situation – avec une sortie – la décision prise – grâce à des règles linguistiques, pour former un système de décision déterministe. Ce même concept de système flou peut aussi être mobilisé pour modéliser des phénomènes naturels, en reliant les variables d'entrées mesurées avec la variable de sortie étudiée.

La méthode classique de modélisation par système flou, héritée de l'Intelligence Artificielle Symbolique, consiste à expliciter linguistiquement le processus de décision d'un expert afin d'obtenir les règles nécessaires à la construction du système flou. Cependant, dans de nombreux cas (évaluation subjective, processus difficilement modélisables par des méthodes classiques, grande complexité du processus de décision, super-expert), l'expert est incapable d'expliciter *a priori* sa prise de décision sous la forme de règles linguistiques[1], du fait qu'il fait appel à des connaissances tacites ou implicites.

L'approche x**tractis**®, qui se présente comme une approche augmentée de la modélisation floue, et par là même de l'Intelligence Artificielle Symbolique, propose dans tous ces cas d'extraire automatiquement, par apprentissage, les règles de décision représentant le processus à modéliser. L'objectif, de prime abord paradoxal, consisterait à expliciter une connaissance implicite, qui échappe à l'expert qui la détient, ou encore à découvrir un modèle du comportement cognitif de l'expert.

Cet apprentissage s'effectue à partir d'une base de données d'exemples, comportant divers cas de prise de décision en fonction de différentes situations. Cette approche est semblable à l'entraînement d'un réseau de neurones à partir d'une base de données d'exemples, mais dispose toutefois des avantages du formalisme flou par rapport aux réseaux de neurones (*cf.* §2). x**tractis**® s'inscrit ainsi dans la branche du *Data Mining* (fouille de données) dénommée *Knowledge Discovery from Data* (Découverte de connaissances à partir de données) ou *Data-Driven* Modelling (modélisation dirigée par les données).

Les bases de données utilisées par x**tractis**® peuvent être de trois types : des données objectives (**O**) – résultats d'analyses ou de mesures physico-chimiques, données démographiques, financières … –, des données subjectives (**S**) – tests hédoniques –, et des données subjectives objectivées (**SO**) fournies par des capteurs humains, les panels

[1] La Psychologie Cognitive postulait que le processus de décision humaine était formalisable par des règles linguistiques du type « si Prémisse alors Conclusion ». Plus tard, il a été prouvé que les super-experts raisonnaient plus par reconnaissance de formes que par inférence de règles. Seuls les experts néophytes développeraient des raisonnements à base de règles, en faisant appel aux connaissances fraîchement apprises et explicitées par leurs pairs.

d'experts, qui représenteraient l'évaluation la plus objective possible de critères subjectifs et/ou sensoriels.

Les problèmes intrinsèques liés à tout processus d'apprentissage (*overfitting*[2], données bruitées ou incomplètes, faible nombre de données d'apprentissage) imposent de mettre en place des méthodes permettant d'analyser les bases de données utilisées, de superviser le déroulement du processus d'apprentissage, et également de valider la capacité de généralisation des modèles générés à des situations qui ne sont pas présentes dans la base d'apprentissage.

2. CARACTÉRISTIQUES ET PROPRIÉTÉS DU FORMALISME FLOU

2.1 Théorie du flou

La théorie du flou est un corpus mathématique permettant de représenter et manipuler des données floues, c'est-à-dire incertaines, imprécises ou subjectives. Les théories qui constituent ce corpus sont des généralisations de théories classiques :

théorie des ensembles → théorie des ensembles flous,

logique binaire → logique floue,

théorie des nombres et intervalles → arithmétique floue,

probabilités → théorie des possibilités.

D'un point de vue formel, la théorie du flou peut être considérée comme une interface entre des données qualitatives ou des concepts symboliques et des valeurs quantitatives. Sa capacité intrinsèque à manipuler des données hétérogènes (quantitatives / qualitatives, précises / vagues) la rend plus apte à traiter des problèmes du Réel[3]. Aussi, l'inférence floue reproduirait-elle un processus de raisonnement approché, à l'image d'un raisonnement humain développé lors d'une prise de décision, qui exploiterait des concepts qualitatifs ou des données imprécises.

De plus, dans les cas où l'information disponible est incomplète, peu nombreuse ou de faible qualité, dans les situations fort courantes où la loi de Gauss n'est pas vérifiée, la théorie des possibilités offre une alternative plus efficiente à la théorie des probabilités pour représenter et traiter l'imprécision et l'incertitude.

[2] Voir §4.1

[3] Nous appelons indifféremment « Réel » ou « monde réel ».

2.2 Interprétabilité

À l'inverse d'une modélisation par réseaux de neurones[4], ou d'une modélisation sous forme d'équation mathématique qui peut être difficile à interpréter, un système flou est défini par une collection de règles linguistiques[5] dont l'interprétation est intuitive. Par exemple :

Règle 1 : « **SI** [C_citron vert] est **fort** **ET** [C_sucre] est **faible** **ALORS** [Goût acide] est **élevé** »

Règle 2 : « **SI** [C_citron vert] est **faible** **ET** [C_sucre] est **fort** **ALORS** [Goût acide] est **peu marqué** »

où [C_citron vert] est la variable prédictive « Concentration en citron vert », [C_sucre] est la variable prédictive « Concentration en sucre » et [Goût acide] est la variable à prédire

faible et **fort** sont des sous-ensembles flous qui caractérisent, sur une échelle quantitative, le fait qu'une concentration est considérée comme faible ou comme forte.

élevé et **peu marqué** sont des caractérisations, sur une échelle d'appréciation, de l'intensité du goût acide tel qu'évaluée par un paneliste.

Cette représentation sous forme de règles définit implicitement une fonction non linéaire de plusieurs variables, avec l'avantage d'être beaucoup plus facile à interpréter que sa forme fonctionnelle. Cette capacité à maintenir une interprétabilité correcte même pour des modèles multidimensionnels nous permet, dans la plupart des cas, de modéliser un processus à partir de l'ensemble de ses variables descriptives originelles, sans passer par une réduction subjective ou *a priori* de la dimensionnalité du problème[6]. Nous préservons ainsi au maximum la connexion entre le modèle et le processus réel.

2.3 Localité

Chaque règle floue définit un domaine d'expertise assimilable à un modèle local, pour lequel la décision à prendre est représentée par la conclusion de la règle. Étant défini par une compilation de règles, un système flou permet de combiner différents domaines d'expertise sur l'espace de décision. Ainsi, modifier un paramètre du système flou revient à modifier un comportement local du système, à l'inverse d'une modélisation par fonction

[4] Un réseau de neurones est une « boîte noire » dont il est très difficile de caractériser l'influence d'un paramètre donné, ou d'expliciter le cheminement d'une prise de décision.

[5] Selon le modèle « si *Prémisse* alors *Conclusion* » dérivé de la psychologie cognitive.

[6] L'ACP (Analyse en Composantes Principales) est souvent utilisée pour réduire la dimensionnalité d'un problème afin de simplifier sa modélisation. La complexité de la modélisation est ainsi plus faible, mais l'interprétabilité du modèle est perdue car il est difficile d'exprimer un lien entre une composante principale et le processus réel. De plus, même si la dimension est réduite, la connaissance de l'ensemble des variables originelles est nécessaire pour calculer les valeurs des composantes principales des points dont on veut prédire la sortie. D'autres raisons nous incitent à ne pas faire appel à l'ACP. En effet, l'ACP est un traitement linéaire donc peu adapté au traitement des processus complexes, tous non linéaires. En outre, nous montrons que l'ACP est peu robuste dans le cas d'une faible quantité de données, c'est-à-dire que la construction des axes d'inertie est sensible au choix des données de départ. La modélisation par PLS souffre des mêmes inconvénients.

polynomiale ou par réseau de neurones pour lesquelles la modification d'un paramètre peut avoir une influence globale sur la réponse du système.

Cette propriété de localité est très intéressante pour la stabilisation des processus d'apprentissage puisqu'elle évite des influences antagonistes de plusieurs points d'apprentissage très différents sur des paramètres identiques du modèle.

2.4 Traçabilité

Le résultat d'une inférence floue est traçable : toute prédiction peut être explicitée en observant les degrés de déclenchement des règles du système flou qui ont conclu à la réponse donnée.

2.5 Universalité

Un système flou est un approximeur universel de fonction non linéaire[7], ce qui permet de garantir l'existence d'un système flou *ad hoc* quel que soit le processus de décision à modéliser. La difficulté pour un modélisateur humain est alors de découvrir un tel système, lorsque le processus étudié est complexe.

[7] Cette propriété est également vraie pour les réseaux de neurones.

3. QUALITÉ ET REPRÉSENTATIVITÉ DES DONNÉES

La qualité des données de référence est primordiale dans tout processus d'apprentissage. En effet, ce qui va être modélisé lors de l'apprentissage est la connaissance implicitement contenue dans cette base de données. Elle doit donc être représentative du processus qu'on désire modéliser.

Il est de ce fait important, dans le cas où les données peuvent être soumises à une forte variabilité (données provenant de capteurs humains, **SO** et **S**, ou données provenant de capteurs instrumentaux très bruités), d'avoir plusieurs répétitions des mesures pour chaque point de la base de données. Ceci nous permet de réaliser une analyse et un filtrage des points aberrants[8].

Il est intéressant de noter ici une particularité intrinsèque aux systèmes flous, qui est leur capacité à réaliser une inférence même en présence d'entrées non renseignées. Le résultat de l'inférence est la valeur de sortie la plus possible[9] étant donné l'ignorance de la valeur d'une ou plusieurs entrées. Cette propriété nous permet de supprimer des points aberrants, ou de prendre en compte l'absence d'information éventuelle dans la base de données, sans avoir besoin de remplacer ces valeurs manquantes par des estimations[10].

Le fait de disposer de données non agrégées (avis et répétitions d'experts individuels, répétitions de mesures bruitées) permet de tenir compte de cette variabilité lors du processus de génération. Ainsi, l'apprentissage n'est pas réalisé sur des données agrégées, mais sur les données brutes, ce qui améliore la robustesse des modèles générés. En effet, ces modèles doivent obtenir une erreur de prédiction minimale sur l'ensemble des répétitions des points d'apprentissage et non pas seulement sur leur moyenne.

4. CAPACITÉ DE GÉNÉRALISATION DES MODÈLES GÉNÉRÉS

4.1 Overfitting

Tout processus d'apprentissage est soumis à un risque d'*overfitting* ou d'apprentissage par cœur. Il est en effet très aisé d'obtenir un modèle capable de prédire exactement les points de la base d'apprentissage, mais dont la capacité de généralisation à d'autres points est

[8] Dans un premier temps, l'analyse de l'intra-variabilité permet d'évaluer la cohérence ou la répétabilité d'un capteur. Puis, l'analyse de l'inter-variabilité permet d'évaluer le consensus du groupe de capteurs. L'objectif est de trouver un compromis pour maximiser la répétabilité et le consensus, tout en filtrant le moins de données possibles.

[9] Au sens de la théorie des possibilités.

[10] Dans les approches classiques de modélisation, on est souvent contraint de remplacer une valeur manquante sur une variable par une estimation, calculée à partir du domaine de définition de la variable ou des valeurs prises par les autres points d'apprentissage sur cette variable (méthode dite d'imputation). Nous considérons que cette approche revient à ajouter dans la base de données de l'information qui peut être fausse ou biaisée.

très faible voire nulle[11]. Ainsi, en créant un système flou comportant une règle de décision pour chaque cas d'apprentissage, on obtiendrait aisément des prédictions exactes (la conclusion de chaque règle vaut exactement la valeur de sortie pour le point considéré), mais les conclusions en dehors de ces cas d'apprentissage n'auraient aucun sens. On obtiendrait un résultat similaire en effectuant une régression polynomiale avec un degré élevé sur un faible nombre de points.

Il est donc nécessaire de mettre en place des méthodes pour restreindre les risques d'overfitting pendant la génération (méthodes de régulation), et pour contrôler la capacité de généralisation des modèles générés (méthodes de validation)[12].

4.2 Méthodes de régulation

Lors de la génération de modèle, nous injectons du bruit sur les données d'entrée afin de restreindre le risque de spécialiser une ou plusieurs règles sur des points particuliers de l'ensemble d'apprentissage. Ceci permet de limiter les fortes non-linéarités d'un modèle.

D'autre part, nous avons conçu et implémenté une approche holistique dite « du tout vers moins » permettant de déterminer la complexité du phénomène étudié : à chaque descente, x**tractis**® sélectionne une stratégie d'apprentissage parmi l'infinité de stratégies dont il dispose. En partant de l'intégralité des variables disponibles, il va diminuer progressivement le nombre de variables utilisées dans les prémisses des règles du système flou, en éliminant les variables qui ne participent pas à la qualité de la prédiction ; il minimise aussi le nombre de classes floues utilisées pour qualifier les variables. La complexité du phénomène est celle du modèle le plus performant et le plus compact découvert.

D'autre part, pour chaque descente, x**tractis**® construit plusieurs générations de systèmes flous en utilisant un nombre de règles plus ou moins important[13], et en effectuant une sélection du sous-ensemble de variables d'entrée le plus pertinent pour prédire la sortie. En choisissant les modèles ayant la structure la plus compacte possible (faible nombre de règles, de variables et de classes floues qualifiant les variables prédictives), c'est-à-dire la moins complexe possible, nous diminuons le nombre de paramètres nécessaires à sa description : nous limitons par conséquent les risques d'*overfitting* et augmentons dans le

[11] Malheureusement, force est de constater que la plupart des publications scientifiques de travaux de modélisation, toutes disciplines confondues, tombent dans ce piège, surtout lorsque le processus étudié est complexe. Aussi, les résultats présentés présentent-ils une faible pertinence : ils ne sont valides que pour les cas étudiés, ceux qui ont justement permis l'établissement du modèle !

[12] Nous appelons indifféremment « capacité de généralisation » ou « robustesse » d'un modèle prédictif.

[13] Une règle représente une connaissance partielle sur le processus à modéliser. Si le nombre de règles est faible, chaque règle matérialisera une connaissance générale sur le processus, alors qu'avec un nombre de règles plus important, la connaissance matérialisée par chaque règle sera plus spécifique. Généralement, plus le nombre de règles est élevé, plus le système sera précis mais avec une capacité de généralisation d'autant plus faible. Dans le cas contraire, la connaissance modélisée par chaque règle étant plus générique, les capacités de généralisation du modèle seront plus importantes, même si sa précision est moindre.

même temps l'interprétabilité du modèle. Habituellement, les meilleurs modèles sélectionnés comportent entre 2 et 6 règles. Toutefois, certains modèles de phénomènes complexes peuvent nécessiter jusqu'à une vingtaine de règles.

Si de telles méthodes de régulation permettent de diminuer le risque d'apprentissage par cœur, elles ne garantissent pas pour autant que les modèles générés soient effectivement robustes. Aussi est-il nécessaire de vérifier la capacité de généralisation des modèles générés à l'aide de méthodes de validation.

4.3 Méthodes de validation

DÉFINITION D'UN ENSEMBLE DE VALIDATION

La méthode de validation la plus utilisée consiste à séparer les données disponibles en un ensemble d'apprentissage et un ensemble de validation, habituellement dans des proportions de 2/3 pour l'apprentissage et de 1/3 pour la validation. Les modèles sont générés en utilisant uniquement l'ensemble d'apprentissage, puis on mesure leur performance sur l'ensemble de validation, ce qui constitue une estimation de leur erreur de généralisation. Cette méthode n'est en revanche efficace que lorsque l'ensemble de données disponibles est grand au regard de la dimensionnalité et de la complexité du phénomène à modéliser. Lorsque cette condition n'est pas respectée, l'estimation de l'erreur de généralisation est très fortement dépendante du découpage en bases d'apprentissage et de validation, ce qui remet en cause la validité même de cette estimation. Cette limitation est illustrée par la Figure 1.

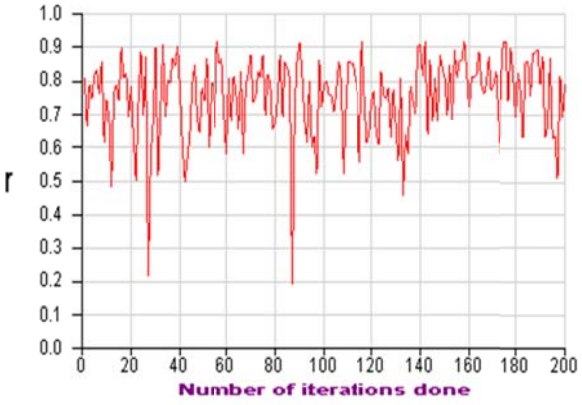

Figure 1 - Variabilité due au choix d'un ensemble de validation

Ce graphique représente le résultat de la même expérience répétée 200 fois sur le même jeu de données : tirage aléatoire d'un ensemble de validation de 30% des points

disponibles, génération d'un modèle avec les données restantes (en imposant la même stratégie d'apprentissage[14]) et prédiction par le modèle généré sur les points de validation. La Figure 1 représente la corrélation « valeur réelle/ valeur prédite » sur les points de validation de chaque expérience.

Utiliser un ensemble de validation pour valider un modèle reviendrait donc à choisir une expérience au hasard parmi les 200 représentées ici pour tirer des conclusions quant à la validité du modèle ; or, nous constatons que la corrélation en validation est très fluctuante : elle évolue entre 0.2 et 0.92 ! D'où le risque élevé d'exploiter un modèle non robuste en pensant qu'il était robuste, ou inversement de rejeter un modèle robuste en étant convaincu qu'il n'était pas robuste.

De plus, la définition d'un ensemble de validation unique interdit d'utiliser les données de validation lors de l'apprentissage, ce qui dans le cas d'un faible nombre de points disponible ne permet pas de générer un modèle représentatif du processus que l'on veut modéliser.

Dans la pratique, et particulièrement en conception de produit, il est très rare de disposer de suffisamment de points pour que cette méthode soit valide[15]. Nous avons donc recours à des méthodes dites de Validation Croisée, qui permettent de s'affranchir de la sensibilité à un partage de données particulier, et de générer des modèles à partir de l'ensemble des données disponibles tout en fournissant une estimation de leur erreur de généralisation. Cette estimation est calculée en réalisant plusieurs partitionnements différents de l'ensemble de points initial en un sous-ensemble d'apprentissage et un sous-ensemble de validation. Pour chaque partitionnement, un modèle de test[16] est généré à l'aide du sous-ensemble d'apprentissage, et ses prédictions calculées sur le sous-ensemble de validation sont confrontées aux valeurs réelles pour en déduire un indice de performance. En agrégeant ces indices de performances calculés à partir de plusieurs partitionnements, on obtient une estimation de la robustesse du modèle généré à partir de la base complète.

LEAVE-ONE-OUT (LOO)

C'est la méthode de validation croisée la plus simple. Pour un ensemble initial de données de taille N, elle consiste à générer N partitionnements composés chacun d'un ensemble d'apprentissage de taille N-1, et d'un unique point de validation. Pour chaque partitionnement, un modèle est généré avec les N-1 points d'apprentissage et testé sur le point de validation restant. Le nuage de points obtenu représente, pour chaque point de la base initiale, la prédiction d'un modèle qui a été généré sans ce point. Cela permet

[14] Même stratégie d'apprentissage signifie que l'on utilise le même algorithme d'apprentissage avec les mêmes paramètres, tout en imposant d'extraire un modèle avec la même structure (mêmes variables d'entrée et même nombre de règles).

[15] Pour les problématiques de conception de produit, un point d'apprentissage représente un prototype ou un produit déjà commercialisé. Pour des raisons évidentes de coût de fabrication et de tests, leur nombre est forcément limité.

[16] Pour que l'estimation de la robustesse soit pertinente, les modèles générés pour chaque partitionnement doivent avoir la même structure (pour un modèle flou, cela signifie avoir les mêmes variables d'entrée, et le même nombre de règles) et les mêmes paramètres de génération que le modèle dont on cherche à estimer la robustesse.

d'estimer la capacité de généralisation du modèle sur l'ensemble des points d'apprentissage.

Dans le cas d'un algorithme d'apprentissage non déterministe (notamment avec l'injection de bruit), l'ensemble de l'opération peut être répété sur plusieurs cycles pour tenir compte de l'incertitude intrinsèque à la génération de modèle dans l'estimation de la capacité de généralisation.

MONTE-CARLO P-CROSS VALIDATION

Cette méthode permet d'explorer de manière plus complète l'influence de la distribution des points d'apprentissage disponible en effectuant des partitionnements comportant plus de points de validation. En effet, le LOO est un estimateur très peu biaisé car tous les points sauf un sont utilisés pour l'apprentissage de chaque modèle de test, mais il est fortement variant à cause du faible nombre de partitionnements différents disponibles. Afin de palier ces deux limitations, il est intéressant de créer des partitionnements de l'ensemble initial comportant plusieurs points de validation. L'estimateur idéal, dit *Exhaustive P-Cross Validation*, consiste à créer tous les partitionnements possibles dont l'ensemble de validation contient P points parmi les N points disponibles, de générer un modèle pour chaque partitionnement à l'aide des N-P points d'apprentissage restants, et de calculer l'erreur de généralisation avec les P points de validation. Il est dans la pratique très souvent impossible de réaliser tous les partitionnements comportant P points de validation (il y en a C_N^P), en revanche il est possible d'obtenir une approximation statistique de cet estimateur, ce que propose la P-validation croisée de Monte Carlo. Elle consiste à agréger les erreurs de généralisation de modèles générés sur des partitionnements construits par tirage aléatoire de P points parmi les N de l'ensemble complet. Ainsi, au bout d'un certain nombre de partitionnements (typiquement entre 50 et 400 selon la complexité des modèles et le nombre P de points tirés), l'agrégation de l'erreur sur les points de validation converge vers l'estimation de l'erreur de généralisation du modèle généré à l'aide la base complète.

Une particularité de notre approche dans l'estimation de l'erreur de généralisation à l'aide des méthodes de validation croisée consiste à ne pas calculer d'indice de performance sur chaque modèle de test dans le but de les agréger. Nous préférons conserver les prédictions de chaque modèle de test obtenues sur son sous-ensemble de validation, et calculer plusieurs indices de performance sur le nuage de points ainsi formé par les prédictions de tous les modèles de test. Cette agrégation la plus tardive possible permet de conserver le maximum d'information pour l'estimation de la robustesse. La Figure 2 montre la convergence de cette méthode : pour chaque itération, la corrélation affichée est celle du nuage de points composé des prédictions de toutes les itérations réalisées auparavant.

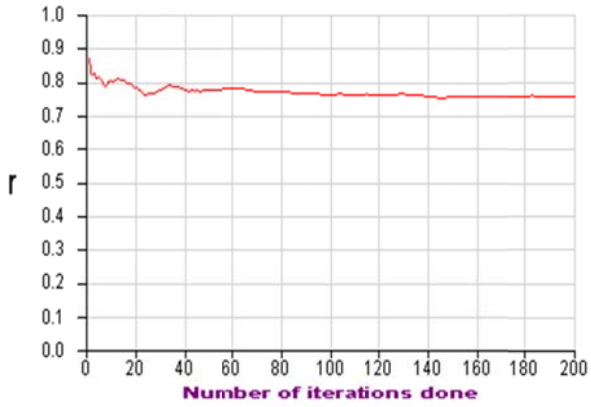

Figure 2 - Convergence de l'estimation Monte Carlo CV

Ces méthodes de validation croisée (de même que le choix du nombre P de points de validation utilisés pour une validation Monte Carlo) ont des caractéristiques différentes en termes de variance et de biais de l'estimation de l'erreur de généralisation. L'utilisation d'un seul de ces estimateurs de robustesse ne sera pas suffisante pour déterminer le modèle ayant la meilleure performance. Pour cette raison, nous utilisons systématiquement plusieurs méthodes de validation afin de sélectionner les modèles donnant satisfaction sur plusieurs estimateurs.

4.4 Performances d'un modèle prédictif : Précision *versus* Robustesse

En apprentissage automatique, nous différencions deux mesures de la performance des modèles prédictifs : la Précision et la Robustesse. La précision est une mesure de la capacité descriptive du modèle. En ce sens, un modèle précis est capable de bien décrire les points d'apprentissage. La robustesse, ou capacité de généralisation, est une mesure de la capacité prédictive du modèle. En ce sens, un modèle robuste est capable de bien prédire sur des points qui lui sont inconnus. Un modèle robuste est nécessairement précis. Par contre, un modèle précis n'est pas nécessairement robuste[17].

Grâce à des techniques de modélisation linéaire, telles que la PLS, il est aisé d'extraire des modèles précis de phénomènes linéaires.

Par contre, toutes les techniques de modélisation non linéaire ne sont pas capables d'extraire des modèles précis de phénomènes non linéaires.

Un système flou étant un approximeur universel de fonctions non linéaires, x**tractis**® a la capacité de quasiment toujours découvrir un modèle capable d'approcher aussi finement

[17] Nous dirons que la précision est une condition nécessaire mais non suffisante de la robustesse d'un modèle. Inversement, la robustesse est une condition suffisante mais non nécessaire de la précision d'un modèle.

que souhaité un phénomène non linéaire : l'extraction de modèles non linéaires précis est aisée.

Toutefois, il est beaucoup plus difficile d'extraire des modèles à la fois précis **et** robustes. Dès que l'information contenue dans la base de données d'apprentissage permet l'extraction de modèles non linéaires (ou linéaires) précis et robustes, que nous dénommons « top-modèle », x**tractis**® déploie ses stratégies intelligentes d'apprentissage et finit par découvrir et extraire ces top-modèles. La durée nécessaire pour l'extraction automatique de tels modèles robustes n'est toutefois pas prédictible.

4.5 Problèmes mal posés

L'un des points forts de x**tractis**® est d'alerter le modélisateur lorsqu'il paraît peu vraisemblable de découvrir un modèle robuste. Un problème « mal posé »[18] peut avoir plusieurs solutions ; dans ce cas, il sera possible de découvrir plusieurs top-modèles ayant des structures différentes[19]. L'autre cas extrême est l'absence de solution au bout de plusieurs jours de calculs intensifs ; plusieurs raisons peuvent alors être avancées pour justifier l'impossibilité de trouver un modèle robuste :

La base d'apprentissage inclut des erreurs / du bruit soit dans les valeurs des prédicteurs potentiels, soit dans les valeurs des variables à prédire. Ces erreurs induisent des contradictions qui interdisent la découverte de modèles stables.

Les « bons » prédicteurs, qui pourraient expliquer la variable à prédire, n'ont pas été inclus dans la base d'apprentissage.

La base d'apprentissage ne contient pas suffisamment de points compte tenu de la complexité du processus.

La population de points d'apprentissage nécessite une segmentation : chaque segment sera régi par un modèle spécifique.

Le processus ou le phénomène à modéliser est purement aléatoire. Or un processus complexe ne peut être modélisé par x**tractis**® que s'il est un tant soit peu déterministe, même s'il est entaché d'aléa.

Dans les quatre premiers cas de figure, il suffit de compléter et/ou modifier et/ou corriger la base d'apprentissage avant de relancer x**tractis**® GENERATE pour extraire de nouveaux modèles.

---

[18] Un problème mathématique *mal posé* au sens d'Hadamard (1902) peut ne pas avoir de solutions ou avoir plusieurs solutions ou avoir des solutions instables par rapport aux données observées qui sont souvent bruitées. Un problème mathématique *bien posé* au sens d'Hadamard a toujours une solution qui est unique et qui dépend continûment des données, c'est-à-dire qu'une faible perturbation des données conduit à une faible perturbation de la solution.

[19] La structure d'un modèle x**tractis**® est définie par la liste des prédicteurs, l'ensemble des classes floues qualifiant chacun des prédicteurs et la liste des règles de décision utilisées par le modèle pour prédire la variable d'étude.

5. EXPLOITATION DES MODÈLES XTRACTIS®

5.1 Interprétation du modèle

Les modèles générés sont, comme nous l'avons mentionné, une compilation de règles linguistiques, induites à partir des exemples contenus dans la base d'apprentissage. Cela permet au modélisateur de confronter la connaissance automatiquement extraite à partir des données, avec l'expérience et les connaissances de l'expert métier sur les processus à modéliser, dans le but de valider ou de critiquer le modèle.

Notons que, grâce à notre approche, il est également possible de synthétiser de manière linguistique le comportement d'un processus totalement inconnu de l'expert et par conséquent de contribuer à la découverte de connaissances nouvelles sur le Réel, notamment en sciences. Elle se pose donc comme une alternative efficiente à l'approche modélisatrice classique *ab initio* jusqu'alors utilisée en sciences.

Lors du processus de génération, x**tractis**® effectue une sélection des variables les plus pertinentes pour prédire la variable d'étude, en calculant l'influence individuelle de chaque variable sur la qualité de prédiction du modèle, tout en tenant compte des synergies existantes dans certains groupes de variables[20].

5.2 Exploration topographique

Les modèles flous définissent implicitement des fonctions non linéaires multi-dimensionnelles. x**tractis**® fournit un ensemble d'outils permettant de représenter graphiquement en 2 ou 3 dimensions, et en temps réel, n'importe quelle coupe de la surface de décision analytique[21]. L'exploration de l'espace multidimensionnel de décision s'en trouve facilitée.

Une autre surface, nommée surface de *mapping* permet de représenter les zones d'influence de chaque règle du système flou, ainsi que des zones non modélisées : plus le degré de *mapping* est élevé, plus les points d'apprentissage de la zone considérée sont bien décrits par au moins l'une des règles du modèle. En général, des zones « non-mappées » ou non couvertes apparaissent subséquemment à un manque de données d'apprentissage dans ces zones combiné à des comportements non monotones au voisinage de ces mêmes zones. Le modélisateur peut ainsi cibler des zones de recherche potentiellement intéressantes pour une mise à jour ultérieure du modèle : la fourniture de

[20] Un groupe de variables en synergie peut expliquer un phénomène complexe, même si aucune de ces variables prises individuellement n'en serait capable (**théorie des signaux faibles**). En effet, un système complexe n'est pas réductible à la somme de ses parties : les interactions entre les parties sont tout aussi importantes que les parties composant le système complexe.

[21] Les modèles générés étant multidimensionnels, il est impossible de représenter graphiquement l'intégralité de leur surface de décision. Seules des coupes autour de valeurs fixées sont affichables.

nouveaux points d'apprentissage dans ces zones non mappées et la régénération du modèle à partir de la nouvelle base de données permettront de couvrir ces zones et ainsi d'étendre le domaine de prédiction du modèle.

5.3 Prédiction directe

Les modèles robustes générés peuvent être utilisés pour réaliser des prédictions de la variable étudiée (descripteurs d'un profil sensoriel, acceptation consommateur, risque de fuite, toxicité d'une molécule, diagnostic d'une pathologie,…), par exemple pour évaluer *a priori* l'impact d'une nouvelle formulation ou d'un nouveau dimensionnement de produit, ou de poser le diagnostic précoce pour un nouveau dossier. Nous parlons alors de *testing* virtuel ou de *screening* virtuel.

D'autre part, la rapidité du moteur d'inférence multithreads de x**tractis**® permet de réaliser des prédictions en temps réel, ou sur de grosses quantités de données[22] (contrôle en ligne, prédictions en C.R.M, en finance …).

Plus l'espace de décision sera couvert par les règles du modèle, plus l'utilisateur pourra utiliser son modèle dans un nombre élevé et varié de situations nouvelles : en effet, pour des raisons de confiance dans la prédiction, x**tractis**® interdit de prédire dans une zone non mappée. Cette alarme permet d'attirer l'attention de l'utilisateur sur les zones pauvres en données cohérentes pour permettre une modélisation fiable.

5.4 Optimisation par inversion de modèles

Cette fonctionnalité nécessite de disposer au préalable de modèles générés robustes. Elle permet de rechercher les valeurs d'entrées possibles qui permettent d'atteindre une sortie désirée, par exemple une préférence consommateur maximale, un profil sensoriel cible, une efficience maximale ou un risque de fuite minimal.

La sortie recherchée peut être décrite de manière nette (les valeurs cibles sont des singletons ou des intervalles), ou sous forme floue, ce qui permet de rendre les critères de recherche flexibles. Ceci permet de proposer des solutions satisfaisantes, dans le cas des problèmes fortement contraints où une solution optimale qui vérifierait des requêtes nettes ne pourrait être trouvée.

La recherche de solutions optimales peut également prendre en compte des contraintes (nettes ou floues) sur les entrées du système. Par exemple : « Quelle formulation maximise [Goût Acide] sachant que [C_Citron Vert] doit être inférieur à environ 8% ? ».

---

[22] Pour un modèle à 4 entrées, 4 règles et une sortie, le moteur d'inférence x**tractis**® PREDICT SERVER peut réaliser 860 000 prédictions / seconde sur un ordinateur quadricoeurs i7 920 2.67GHz fonctionnant avec un seul thread (1 cœur physique). Si l'on autorise le calcul parallèle avec 2 threads, on passe à 1 425 000 prédictions / seconde et à 2 410 000 prédictions / seconde avec 4 threads. Ces vitesses de calculs ne tiennent pas compte des temps d'accès disque éventuels pour le chargement et le stockage de données.

Il est aussi possible de rechercher des solutions optimales à une requête multi-objectifs en mixant simultanément l'inversion de modèles flous x**tractis**® et de modèles analytiques. Par exemple : « Quelle formulation maximise [Goût Acide] et dont l'évaluation du [Goût Sucré] est environ 4/10, sachant que [C_sucre] doit être égal à 2 x [C_citron vert] »

Les solutions optimales peuvent être représentées sur des coupes de la surface de décision, avec la possibilité d'afficher les zones comportant toutes les solutions optimales pour la coupe considérée.

Une telle fonctionnalité révolutionne le cycle classique de conception de produit, fondé sur le processus « essai-erreur » : elle réduit considérablement les coûts et le temps de convergence vers un produit optimal, en minimisant le nombre de prototypes à fabriquer et en réduisant le nombre de campagnes de tests à réaliser.

6. DOMAINES D'APPLICATION

6.1 Approche universelle

Cette approche de modélisation par extraction des connaissances implicitement contenues dans une base de données est universelle et peut s'appliquer à la modélisation de tout phénomène sur lequel on dispose d'une base de cas d'entrées/sorties. Elle ne nécessite aucune connaissance du domaine d'application de la part du modélisateur ou de l'expert métier, si ce n'est la présentation de l'ensemble des variables potentiellement explicatives[23] à partir duquel x**tractis**® pourra sélectionner seul les variables explicatives[24] du modèle qu'il créera. En effet, nous interdisons volontairement à x**tractis**® de créer de nouvelles variables combinées afin de maintenir l'interprétabilité du modèle extrait.

Il est important de noter qu'à partir d'une dizaine de dimensions, le problème devient inexorablement complexe et échappe totalement à l'entendement humain[25]. Aussi, nous parait-il nécessaire de faire appel à un automate intelligent qui, doté d'une puissance de calcul et de capacités de mémoire quasi-illimitées, pourrait proposer, en un temps relativement limité, des solutions efficientes aux problèmes complexes.

x**tractis**® aide déjà plusieurs grands groupes internationaux dans leur processus de découverte et de mise au point rapides de nouveaux produits (modélisation de profil sensoriel, de l'efficience d'une molécule, de nouvelles formulations), dans la gestion de la relation client (CRM) afin de prévoir et d'anticiper les comportements des clients

[23] Les variables potentiellement explicatives sont aussi appelées variables potentiellement prédictives ou prédicteurs potentiels du modèle (anglicisme tiré de *predictor*).

[24] Les variables explicatives sont aussi appelées variables prédictives ou prédicteurs du modèle.

[25] Les travaux de psychologie cognitive démontrent qu'un être humain est normalement capable de gérer les interactions simultanées entre 2, voire 3 variables au maximum, certainement du fait qu'il évolue dans un espace tridimensionnel ; tandis que les humains à très haut potentiel peuvent conceptualiser des interactions simultanées entre 7, 8 voire 9 variables au maximum.

(modélisation de la préférence de segments de consommateurs), dans leurs processus d'évaluation objectivo-subjective (modélisation du juste prix d'une transaction immobilière, de la valeur d'actifs mobiliers…), ainsi que dans leur démarche d'analyse de risque (modélisation du risque financier d'investissement, de l'écotoxicité d'une molécule, caractérisation de défauts dans des tubes d'acier, détection d'endommagement/rupture dans un matériau composite, diagnostic précoce de pathologies humaines, réceptivité à la chimiothérapie…). S'affirmant comme une solution technologique universelle, il est utilisé avec succès dans divers domaines : agro-alimentaire, cosmétique, sports & loisirs, produits d'hygiène, optique, automobile, santé, immobilier, finance comportementale, énergie, internet, sciences, espace, ressources humaines, marketing sensoriel et comportemental…

x**tractis**® peut aussi être utile dans tout domaine pour lequel les experts métier ont besoin d'aide pour exploiter de la connaissance enfouie dans des bases de données complexes. Aussi, considérons-nous notre robot comme un Assistant Virtuel de recherche ou comme une prothèse cognitive.

Conformément à sa politique d'excellence et d'avant-gardisme, **intelli**tech poursuit systématiquement ses actions de R&D pour découvrir et mettre au point de nouveaux algorithmes d'extraction automatique de connaissances, capables d'exploiter des clusters de stations GPU[26] ; de tels algorithmes seraient plus adaptés aux domaines d'application du *Big Data,* caractérisés par une très forte complexité, pouvant impliquer des millions de points d'apprentissage et/ou plusieurs milliers de dimensions.

6.2 Interface entre modélisateur et expert métier

L'interprétabilité des modèles générés permet au modélisateur de confronter ses résultats avec des experts métier du domaine d'application considéré. La représentation des modèles par règles floues et l'utilisation exclusive de variables originelles permet de rendre cette tâche beaucoup plus facile qu'avec d'autres types de modélisation. Ceci permet dans la plupart des cas de confirmer la pertinence d'un modèle, et parfois même de fournir des pistes d'étude pour faire évoluer l'état de connaissance du domaine d'application, que cela soit en sciences physiques, en sciences de l'Homme et de la Société, en sciences de la Vie et de la Terre ou en sciences pour l'Ingénieur.

[26] x**tractis**® exploite déjà le calcul parallèle CPU (processeurs multicœurs) et le calcul parallèle GPU (cartes graphiques CUDA) sur une station HPC (High Performance Computing). Une carte graphique de dernière génération propose une puissance de calculs équivalente à environ celle de 10 cœurs CPU.