Un indice d'interprétabilité des bases de règles floues

J.M. Alonso, Serge Guillaume, L. Magdalena

▶ **To cite this version:**

J.M. Alonso, Serge Guillaume, L. Magdalena. Un indice d'interprétabilité des bases de règles floues. Rencontres LFA francophones sur la Logique Floue et ses Applications (LFA), Oct 2006, Toulouse, France. p. 37 - p. 44. hal-00468036

HAL Id: hal-00468036
https://hal.science/hal-00468036

Submitted on 29 Mar 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Un indice d'interprétabilité des bases de règles floues

An interpretability index for fuzzy knowledge bases

José M. Alonso[1] Serge Guillaume[2] Luis Magdalena[3]

[1] ETSI Telecomunicacion, UPM

[2] Cemagref, Montpellier

[3] European Centre for Soft Computing

Ciudad Universitaria, 28040 Madrid, Spain, jmam@mat.upm.es

BP 5095, 34196 Montpellier Cedex 5, France, serge.guillaume@montpellier.cemagref.fr

Campus Universitario de Mieres, Mieres 33600 Spain, luis.magdalena@softcomputing.es

Résumé :

Ce travail décrit la proposition d'un système d'inférence floue pour évaluer l'interprétabilité de bases de règles floues. La structure est hiérarchique, les sorties intermédiaires ont du sens. Dans la version actuelle, seuls les problèmes de classification sont considérés. Les premiers résultats sont analysés et discutés.

Mots-clés :

Interprétabilité, règles linguistiques, classification

Abstract:

This work describes the proposal of a fuzzy inference system to assess the interpretability of a linguistic knowledge base. The intermediate outputs of the hierarchical structure are meaningful. Currently, the index only deals with classification problems. Preliminary results are discussed.

Keywords:

Interpretability, linguistic knowledge bases, classification

1 Introduction

L'interprétabilité est une question clé pour les systèmes d'inférence floue. Ce concept apparaît dans la littérature sous différents noms : transparence, intelligibilité, compréhensibilité,... Dès l'origine, les règles linguistiques dans les systèmes d'inférence floue ont été reconnues pour leurs avantages : conception simple et rapide d'un système, facilité de maintenance et d'adaptation. Leur utilisation s'inscrit dans la perspective de proposer des modèles utilisant un langage de description aussi proche que possible de celui employé par les experts humains. La logique floue constitue un cadre commode pour l'expression et la formalisation de la connaissance experte.

D'autre part, le formalisme flou est à la base de nombreuses méthodes d'apprentissage [7, 9]. Dans les cas où des techniques de modélisation plus classiques peuvent s'avérer performantes, on attend des systèmes d'inférence floue qu'ils apportent des éléments de connaissance qualitative sur le phénomène étudié, qu'ils soient interprétables.

Comment définir l'interprétabilité ? Il est difficile de trouver une définition formelle. L'encyclopédie libre Wikipedia[1] propose la suivante : *Soient T et S deux théories formelles. T est dite interprétable dans S si et seulement si le langage de T peut être traduit dans celui de S de telle manière que tous les théorèmes de T soient prouvés dans S.*

S'inspirant de cette définition, Mencar *et al.* [10] proposent un cadre formel pour les systèmes. Ils distinguent le langage *L* utilisé pour décrire le modèle, d'un langage *L'* qui permet d'expliquer le modèle à l'utilisateur, il s'agit en général du langage naturel. Si le système est interprétable alors la traduction de *L* vers *L'* est réalisée par l'utilisateur lui-même, de manière simple.

[1] http ://en.wikipedia.org

Cela signifie que l'utilisateur pourra comprendre le fonctionnement du système dans ses moindres détails et qu'il pourra se représenter son comportement. Lorsque le langage de modélisation utilise des règles linguistiques floues, la compréhension implique une lecture facile de chacune des règles et des termes linguistiques associés ainsi qu'une représentation du comportement de la base de règles prise dans son ensemble.

Une importante littérature a récemment été consacrée à l'interprétabilité des systèmes d'inférence floue[2]. Même si le formalisme flou favorise l'interprétabilité de la base de règles, il n'est pas suffisant pour la garantir. De nombreux travaux [5, 2, 3] visent à l'améliorer, notamment en imposant des contraintes sur le partitionnement des variables ou en s'efforçant de diminuer le nombre de règles.

L'interprétabilité ne doit pas nous faire oublier la précision : ces deux pôles sont essentiels pour la qualité des systèmes d'inférence floue. Que l'on parte d'un pôle [4] ou de l'autre [5], l'essentiel est de parvenir à un compromis.

Des indices sont nécessaires à la fois pour qualifier les systèmes et pour guider des processus d'apprentissage. Quantifier la précision est assez usuel. De nombreux indices, basés sur la différence entre la sortie attendue et la sortie calculée par le système, ont été étudiés. Il n'en est pas de même pour l'interprétabilité. Nauck [11] a, le premier, proposé un indice. Notre travail s'inscrit dans cette démarche.

L'article est organisé comme suit. La section 2 détaille les différents niveaux d'interprétabilité d'une base de règles. Dans la section 3 le système hiérarchique, avec ses bases de règles, est présenté en détail. Son comportement est illustré, dans la section 4, sur différents systèmes qui modélisent un problème de classification bien connu, les iris. Enfin, les principales conclusions sont rappelées dans la section 5.

[2]http ://sci2s.ugr.es/keel/interpretability.php, P. Villar. KEEL : Interpretability issues in fuzzy rule based systems.

2 Interprétabilité

L'interprétabilité du système, qui est la facilité d'analyse de son comportement, dépend à la fois des partitions des variables et des règles. Elle est liée à la complexité du modèle, nombre de paramètres, ainsi qu'aux mécanismes d'inférence.

Nous proposons d'examiner les niveaux suivants :

1. *Partitions des variables*
2. *Base de règles*
   - Analyse individuelle d'une règle ;
   - Interaction entre les règles.

La lisibilité des partitions floues est une condition nécessaire à l'interprétabilité des règles. Ce sujet a été largement étudié [6]. Rappelons les principales conditions :

- Distinguabilité : Chaque sous-ensemble flou représente un concept identifié, différent de celui de ses voisins.
- Un nombre d'étiquettes justifiables.
- Couverture : Chaque point de l'univers, x, doit appartenir de façon significative, $\mu(x) > \epsilon$, à au moins un sous-ensemble flou. ϵ est appelé le niveau de couverture.
- Normalisation : Les ensembles flous doivent être normaux, $\exists x$ tel que $\mu(x) = 1$.
- Recouvrement : Le niveau d'intersection entre deux voisins doit être significatif.

Les partitions floues fortes, comme celle de la figure 1, satisfont toutes ces conditions. Leur niveau de couverture est $\epsilon = 0.5$.

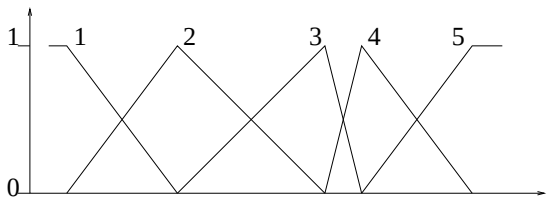

Figure 1 – Une partition floue forte

La seule lisibilité des partitions floues ne suffit pas à garantir l'interprétabilité de la règle. Dans cet exemple, $Si\ x\ est\ (A_1 \vee A_4) \wedge \neg A_3\ alors \ldots$, la combinaison des connecteurs logiques, qui rend l'ensemble défini par la prémisse non convexe, ne facilite pas l'interprétabilité. Pour la garantir, il faut imposer des conditions qui conduisent à des ensembles convexes, notamment en restreignant les combinaisons de sous-ensembles flous à des termes voisins [8].

Le nombre de variables utilisées dans la prémisse est très important : il est très difficile d'appréhender des interactions entre variables lorsque leur nombre dépasse trois. Mais l'exercice d'analyse se complique lorsque l'on essaie d'analyser le comportement de la base de règles prise dans son ensemble. En effet, la sortie calculée par le système résulte de l'agrégation de l'ensemble des règles activées par l'exemple. Le processus est donc d'autant plus facile que le nombre de règles activées par l'exemple est petit et que ces règles sont simples.

C'est en s'appuyant sur un raisonnement voisin que Nauck [11] a proposé un indice d'interprétabilité, I_{Nauck}. Sa valeur résulte du produit de 3 termes :

– Comp : Mesure de la complexité du classifieur, c'est le rapport du nombre de classes au nombre total de variables utilisées dans les règles (que nous appellerons *Nombre total de prémisses*).
– Part : C'est la moyenne, pour l'ensemble des variables d'entrée, de l'inverse du nombre d'étiquettes (moins un, car une partition doit comporter au moins 2 étiquettes) par variable.
– Cov : Caractérise la couverture du domaine de l'entrée par la partition floue. Ce paramètre vaut 1 pour une partition floue forte.

Notre proposition, décrite en détail dans la section suivante, s'en inspire.

3 Evaluer l'interprétabilité

Notre objectif est donc de proposer un dispositif, composé de plusieurs bases de règles floues, pour évaluer l'interprétabilité des systèmes d'inférence floue. La première étape de la conception est l'identification des variables influentes. Trois groupes peuvent être distingués :

1. Taille de la base de règles : Nombre de règles, de variables et d'étiquettes par variable.
2. Nature des règles :
   – Mamdani : La conclusion est un ensemble flou ;
   – Takagi-Sugeno : La conclusion est un polynôme, qui peut être réduit à une constante.
3. La structure de chacune des règles :
   – Nombre de variables utilisées dans la description de la règle ;
   – Diversité des opérateurs présents : Conjonction, disjonction, négation ;
   – Nature des termes linguistiques : Élémentaires, composés.

Dans la première ébauche que nous présentons tous ces paramètres ne sont pas pris en compte. Par exemple, seules les règles de type Mamdani sont considérées. L'extension aux règles de Sugeno dont la conclusion est une constante est immédiate.

La figure 2 donne une vue d'ensemble du système. Celui-ci est formé de six variables d'entrée, la sortie étant l'indice d'interprétabilité. Les valeurs de sortie, intermédiaires ou finale, résultent d'une défuzzification.

Les entrées sont groupées suivant la nature de l'information dont elles sont porteuses pour former une structure interne hiérarchique de quatre sous-systèmes dont les sorties intermédiaires ont un sens précis.

Le comportement du système est le suivant. Le premier sous-système, $BR1$, estime la dimension de la base de règles à partir du nombre total de règles et du nombre total de prémisses, c'est-à-dire le nombre de variables réellement utilisées dans l'ensemble des règles. Cette dimension sera d'autant plus petite qu'il

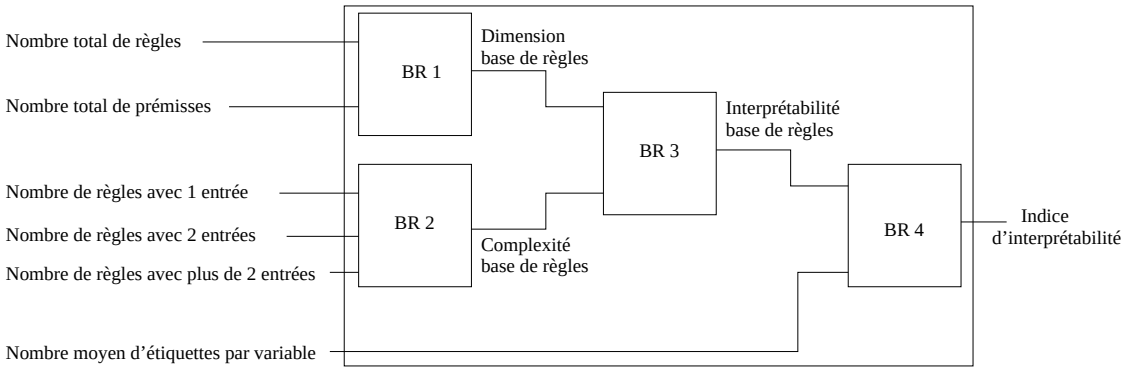

Figure 2 – Le système hiérarchique

y aura peu de règles et que celles-ci seront incomplètes, définies par un nombre réduit de variables d'entrée. Parallèlement, le deuxième sous-système, $BR2$, évalue la complexité de la base de règles, à partir du nombre de règles utilisant une, deux et au moins trois variables d'entrée.

A première vue les deux bases de règles peuvent apparaître comme redondantes car elles utilisent les mêmes informations : le nombre total de règles est la somme des règles avec 1, 2 et au moins 3 entrées ; le nombre total de prémisses se calcule à partir de ces 3 nombres. La présence de $BR2$ se justifie néanmoins car la manipulation de nombres globaux comme le fait $BR1$ peut cacher des disparités importantes. L'analyse d'une base de règles dépend du nombre de variables intervenant dans les règles. Il est facile de comparer des règles définies par une seule variable, cette analyse reste accessible dans le cas de deux variables, mais devient rapidement fastidieuse lorsque ce nombre augmente.

La dimension et la complexité sont ensuite combinées ($BR3$) pour caractériser l'interprétabilité de la base de règles. L'indice final prend également en compte le nombre moyen de sous-ensembles flous par variable ($BR4$).

Pour appréhender le fonctionnement du système, une analyse de chacune des bases de règles est nécessaire.

La première, illustrée sur la figure 3, estime une dimension de la base de règles. Lorsque le

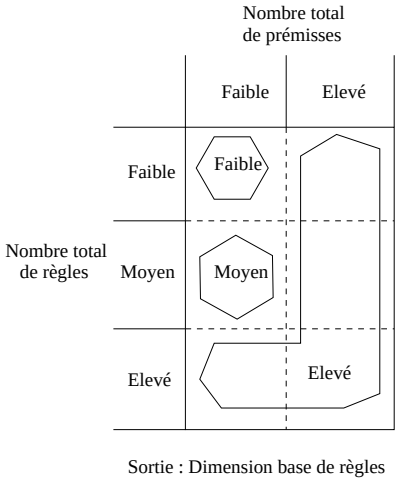

Figure 3 – La base de règles 1

Nombre total de prémisses est *Faible*, alors la sortie évolue comme le nombre de règles. Dans le cas ou le nombre de prémisses est élevé alors, quel que soit le nombre de règles, la sortie sera élevée. Ceci correspond à une base de règles de taille raisonnable dont toutes les règles sont complètes.

La figure 4 décrit le fonctionnement de la base $BR2$. Comme ce sous système comporte 3 entrées, il est représenté sous la forme de deux schémas. La partie supérieure de la figure illustre le cas où le *nombre de règles avec deux entrées ou plus* est *Faible*. La sortie, *complexité de la base de règles* est alors *Faible* sauf lorsque le *nombre de règles avec deux variables* est *Elevé* où elle devient *Moyenne*.

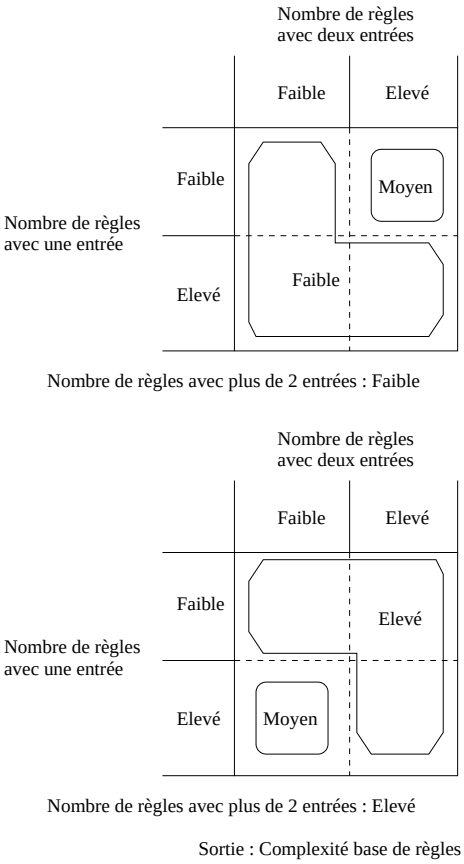

Figure 4 – La base de règles 2

De manière analogue, lorsque le *nombre de règles avec deux entrées ou plus* est *Elevé*, la sortie est généralement de type *Elevé*, sauf lorsque le *nombre de règles avec deux variables* est *Faible* où elle devient *Moyenne*.

L'*indice d'interprétabilité* de la base de règles est calculé par $BR3$, les règles de fonctionnement sont illustrées par la figure 5. Elles traduisent un principe simple : chacune des deux entrées, *dimension de la base de règles* et *complexité de la base de règles*, influe de manière proportionnelle sur la sortie. Ainsi lorsque l'une des deux est *Moyenne*, la sortie se comporte comme l'autre. Pour décrire les nuances de ce comportement, cinq termes linguistiques sont nécessaires.

La dernière base de règles permet de prendre en compte le *nombre moyen d'étiquettes par variables*. Lorsque celui-ci est *Faible*, il agit

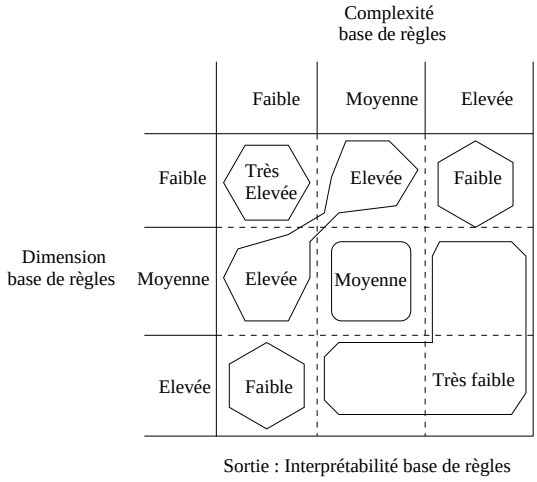

Figure 5 – La base de règles 3

comme un élément neutre : l'*indice d'interprétabilité* final se comporte comme celui de la complexité de la base de règles, la sortie du sous-système 3. S'il est *Elevé* alors la sortie est proportionnellement réduite.

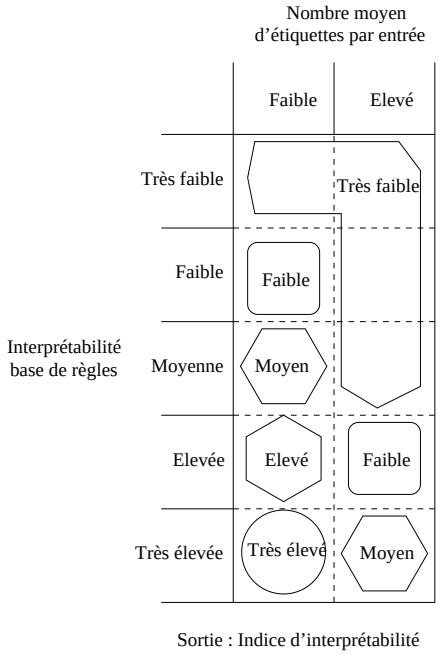

Figure 6 – La base de règles 4

La correspondance entre les valeurs numériques et les étiquettes linguistiques dépend de la nature du problème à traiter. Le nombre de paramètres d'un modèle est directement relié à

sa complexité. Dans cette version préliminaire, nous adressons seulement les problèmes de classification, et dans ce cas la complexité peut être caractérisée par le nombre de classes, C.

Nous avons également introduit un paramètre de réglage, P. La sensibilité du système à la valeur de ce paramètre est étudiée dans la section 4.

Pour chacune des variables, d'entrée ou de sortie intermédiaire, une gamme de variation a été définie. Au sein de cet intervalle, une partition floue forte régulière a été construite. Les bornes du domaine de variation, ainsi que le nombre de termes de la partition, sont indiqués dans le tableau 1.

Tableau 1 – Description des variables

Variable	Domaine	Labels
Nombre total de règles	[*C*, *P***C*]	3
Nombre total de prémisses	[*C*, (2**P*)**C*]	2
Dimension de la base de règles	[0, 1]	3
Nombres de règles avec 1 entrée	[*C*, *P***C*]	2
Nombres de règles avec deux entrées	$[0, \left(\frac{P}{2}\right)*C]$	2
Nombres de règles avec plus de 2 entrées	$[0, \left(\frac{P}{4}\right)*C]$	2
Complexité de la base de règles	[0, 1]	3
Interprétabilité de la base de règles	[0, 1]	5
Nombre moyen d'étiquettes par entrée	[2, 9]	2
Indice d'interprétabilité	[0, 1]	5

La limite supérieure du nombre de prémisses est fixée à deux fois celle du nombre de règles, qui ne dépend que de C et de P. La limite du nombre de règles avec une seule variable d'entrée est aussi celle du nombre de règles. Le nombre moyen d'étiquettes par variable est calculé sur les seules variables d'entrée. Pour les problèmes de régression, la sortie devrait également être prise en compte, en comptabilisant le nombre de sous-ensembles flous ou bien le nombre de conclusions distinctes dans les règles lorsqu'il s'agit de scalaires.

4 Premiers résultats

Le système hiérarchique a été testé sur les *Iris*. Ce jeu de données bien connu comprend 150 exemples, décrits par 4 variables, la longueur et la largeur des sépales et des pétales, et correspondant à 3 variétés d'iris, setosa, versicolor et virginica.

Chacune des variables d'entrée a été partitionnée par une grille régulière de 5 sous-ensembles flous. Puis, plusieurs systèmes ont été conçus par induction automatique de règles. Parmi les nombreuses méthodes d'apprentissage disponibles [7, 9], nous avons utilisé celles implémentées dans le logiciel Fispro[3] : la méthode de *Wang et Mendel* (WM)[4], l'algorithme de prototypage rapide, *Fast Prototyping Algorithm* (FPA), et les *Arbres de décision* (FDT). Ces trois algorithmes sont très différents : WM est une méthode représentative de l'approche prototypes tandis que FPA représente l'approche régions. Ces deux méthodes génèrent des règles complètes, définies par l'ensemble des variables. Ce n'est pas le cas pour les règles induites par les arbres de décision. Quelle que soit la méthode d'induction, le système résultant peut être simplifié. Nous avons appliqué la procédure décrite en détail dans [1] et implémentée dans le logiciel KBCT[5]. Le processus itératif s'appuie sur deux opérations de base, fusion et élimination, appliquées aux partitions et aux règles. Les étiquettes qui sont utilisées de manière conjointe, comme les règles qui ont une conclusion identique, peuvent être fusionnées sous certaines conditions. Ensuite les étiquettes non utilisées, et les règles redondantes, peuvent être éliminées. L'objectif de la simplification est d'obtenir des bases de règles plus com-

[3] http ://www.inra.fr/bia/M/fispro

[4] Pour en savoir plus sur les algorithmes d'induction et leur implémentation, se référer à la documentation du logiciel.

[5] http ://www.mat.upm.es/projects/advocate/kbct.htm

pactes, cette procédure est donc susceptible d'améliorer l'indice d'interprétabilité.

Huit systèmes ont ainsi été construits : Quatre correspondant aux algorithmes originaux, WM, FPA, FDT, auquel s'ajoute une procédure d'élagage pour les arbres de décision, FDT-P ; et les quatre résultant de leur simplification. Le tableau 2 résume les caractéristiques des différents systèmes.

Tableau 2 – Résultats sur les iris

Méthode	#R	#P	#R1	#R2	#R+2	#E
WM	32	128	0	0	32	5
WM-S	7	28	0	0	7	5
FPA	37	148	0	0	37	5
FPA-S	5	20	0	0	5	4,75
FDT	18	45	3	6	9	5
FDT-S	9	21	2	3	4	3,25
FDT-P	7	11	3	4	0	5
FDT-P-S	4	6	2	2	0	3

Chacune des colonnes du tableau correspond à une entrée du système hiérarchique : *#R* est le nombre total de règles, *#Rx* le nombre de règles avec *x* variables d'entrée, *#P* le nombre total de prémisses et *#E* le nombre moyen d'étiquettes par variable.

Les effets de l'élagage, pour les arbres de décision, et de la simplification, pour l'ensemble des configurations, sont évidents. La simplification agit à la fois sur le nombre de règles et le nombre d'étiquettes par variables, en combinant par exemple deux étiquettes qui sont toujours utilisées ensemble en une nouvelle qui est leur enveloppe convexe.

L'indice d'interprétabilité a été calculé pour des systèmes hiérarchiques construits avec différentes valeurs du paramètre de réglage, P. Les résultats sont résumés dans le tableau 3. La première colonne indique la valeur de P. Pour comparaison, l'indice proposé par Nauck est calculé dans la dernière ligne.

Les premières lignes du tableau soulignent l'effet du paramètre P. Lorsqu'il est égal à deux seul le système le plus compact, construit par l'arbre de décision avec une procédure d'élagage puis de simplification, présente un indice d'interprétabilité non nul. Lorsque P augmente, c'est à dire lorsqu'on repousse les limite de ce qui est admissible, de plus en plus de configurations ont un indice non nuls. Une valeur de 7 ou 8 pour P représente un bon compromis : la dynamique est suffisamment importante sans que tous les systèmes soient considérés comme interprétables.

L'indice de Nauck, de conception plus simple, est également moins riche. Il est non nul pour tous les systèmes testés et son domaine de variation est très faible. Seule la configuration la plus compacte présente une valeur supérieure à 0.1. Le nombre d'étiquettes par variable semble avoir une importance exagérée, ce qui conduit à évaluer la configuration FDT-S plus interprétable que FDT-P, alors que cette dernière compte moins de règles et un nombre plus réduit de prémisses.

5 Conclusion

L'interprétabilité représente une des motivations essentielles d'utilisation des systèmes d'inférence floue. Toutefois, bien qu'une littérature importante ait été consacrée à la conception de systèmes interprétables, les travaux dédiés à la mesure de l'interprétabilité sont peu nombreux.

La structure proposée est hiérarchique, cela permet de combiner les 6 entrées du système de manière nuancée tout en maintenant des bases de règles intermédiaires interprétables. L'ensemble est paramétré par le nombre de classes à discriminer, qui rend compte de la complexité du problème étudié, ainsi que par un paramètre de réglage dont l'effet est analysé.

Cette version préliminaire, qui ne permet de qualifier que les systèmes de classification, est implémentée au sein du logiciel libre KBCT. Elle peut être étendue pour prendre en compte les problèmes de régression ainsi que d'autres caractéristiques importantes de la base de règles.

Tableau 3 – Indice d'interprétabilité : variations avec P et comparaison avec celui de Nauck

P	WM	WM-S	FPA	FPA-S	FDT	FDT-S	FDT-P	FDT-P-S
2	0	0	0	0	0	0	0	0.456
3	0	0	0	0	0	0	0.105	0.651
4	0	0	0	0	0	0	0.244	0.726
5	0	0	0	0	0	0	0.372	0.769
6	0	0	0	0	0	0.030	0.452	0.797
7	0	0	0	0.029	0	0.103	0.500	0.817
8	0	0	0	0.120	0	0.194	0.535	0.831
9	0	0	0	0.191	0	0.286	0.563	0.843
10	0	0.035	0	0.248	0	0.358	0.585	0.851
I_{Nauck}	0.006	0.027	0.005	0.040	0.016	0.077	0.068	0.333

Ainsi, si le nombre de règles permet de caractériser le fonctionnement global du système, son fonctionnement local dépend du nombre maximum de règles déclenchables simultanément et des opérateurs d'agrégation et de défuzzification. Les deux objectifs sont contradictoires : pour minimiser le nombre de règles, et analyser l'influence des variables, le nombre d'étiquettes par variable doit être petit. Ceci favorise le déclenchement simultané d'un plus grand nombre de règles. A l'inverse augmenter le nombre d'attributs conduit à des règles très spécifiques qui risquent de n'apporter qu'une information très locale.

Références

[1] José M. Alonso, L. Magdalena, and S. Guillaume. A simplification process of linguistic knowledge bases. In *Joint EUSFLAT-LFA 2005,* pages 1045–1050, Barcelona, Spain, September, 7-9, 2005.

[2] P. Carmona, J. L. Castro, and J. M. Zurita. Learning maximal structure fuzzy rules with exceptions. *Fuzzy Sets and Systems*, 146(1) :63–77, August 2004.

[3] J. Casillas, O. Cordón, M. J. del Jesús, and F. Herrera. Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. *IEEE T. on Fuzzy Systems,* 2005.

[4] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena. *Accuracy improvements in linguistic fuzzy modeling,* volume 129. Studies in Fuzziness and Soft Computing, Springer-Verlag, Heidelberg, 2003.

[5] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena. *Interpretability issues in fuzzy modeling*, volume 128. Studies in Fuzziness and Soft Computing, Springer-Verlag, Heidelberg, 2003.

[6] P.-Y. Glorennec. Algorithmes d′apprentissage pour systèmes d′inférence floue. *Editions Hermès*, Paris, 1999.

[7] S. Guillaume. Designing fuzzy inference systems from data : an interpretability-oriented review. *IEEE Transactions on Fuzzy Systems*, 9(3) :426–443, June 2001.

[8] S. Guillaume and L. Magdalena. An or and not implementation that improves linguistic rule interpretability. In *IFSA World Congress*, volume 2809 of LNCS, pages 302–313, Beijing, China, July 2005.

[9] Eyke Hüllermeier. Fuzzy methods in machine learning and data mining : Status and prospects. *Fuzzy Sets and Systems*, 156 :387–406, 2005.

[10] C. Mencar, G. Castellano, and Anna M. Fanelli. Some fundamental interpretability issues in fuzzy modeling. In *Joint EUSFLAT-LFA 2005*, pages 100–105, Barcelona, Spain, Sept, 2005.

[11] D. D. Nauck. Measuring interpretability in rule-based classification systems. In *FUZZ-IEEE 2003*, volume 1, pages 196–201, St. Louis, USA, May, 25-28, 2003.