



Le moteur d'inference COTO. Application en analyse de données

Tu Bao Ho, Joël Quinqueton

► To cite this version:

Tu Bao Ho, Joël Quinqueton. Le moteur d'inference COTO. Application en analyse de données. RR-0555, INRIA. 1986. inria-00075999

HAL Id: inria-00075999

<https://inria.hal.science/inria-00075999>

Submitted on 24 May 2006

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.



CENTRE DE ROCQUENCOURT

Rapports de Recherche

N° 555

**LE MOTEUR
D'INFÉRENCE COTO
APPLICATION
EN ANALYSE DE DONNÉES**

**Tu Bao HO
Joël QUINQUETON**

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France
Tél. (1) 39 63 55 11

Août 1986

LE MOTEUR D'INFERENCE COTO APPLICATION EN ANALYSE DE DONNEES

HO Tu Bao
Université Paris 6

QUINQUETON Joël
I.N.R.I.A.

RESUME

Le but de ce rapport est donc de présenter les principes de construction et les aspects d'implémentation du moteur d'inférence COTO. La première application du COTO est en Analyse de Données comme une assistance intelligente du logiciel SICLA.

ABSTRACT

This report presents the design principles and implementation aspects of the inference engine COTO. The first application of COTO is in Data Analysis as an intelligent assistance for the software SICLA.



PAPIER RÉCUPÉRE ET RECYCLE

SOMMAIRE

I. Introduction

II. Moteur d'inférence COTO

II.1. Caractéristiques

II.2. Le schéma des systèmes experts avec COTO

II.3. Représentation de connaissances dans COTO

II.4. Le contrôle du système

II.5. La conversation et explication

III. Le logiciel d'analyse de données SICLA

III.1. Les phases principales d'une analyse de données

III.2. Présentation du logiciel SICLA

III.2.1. Les fonctionnalités du système

III.2.2. Architecture de SICLA

III.2.3. Les aides traditionnelles

IV. Le système expert CLAVETO

IV.1. Représentation de connaissances dans CLAVETO

IV.2. Les différents types de connaissances

IV.2.1. Les étapes préliminaires dans l'espace de représentation

IV.2.2. Choix des méthodes

IV.2.3. Aide à l'interprétation des résultats

IV.2.4. Règles concernant les stratégies

IV.2.5. Traitement des erreurs

V. Interface COTO-SICLA

VI. Conclusion et perspectives

VII. Exemple d'utilisation de CLAVETO

VIII. Références

I. INTRODUCTION

C'est avec l'arrivée des systèmes experts que l'Intelligence Artificielle s'est introduite dans une série de domaines d'applications. Les experts humains sont des gens capables de résoudre des problèmes, leur expertise étant basée essentiellement sur :

- . leur connaissances du domaine accumulées pendant des années d'expérience.
- . leur capacités de raisonner sur ces connaissances pour trouver la solution du problème.

En bref, le but des systèmes experts est de simuler ces deux possibilités pour reproduire sur machine les activités des experts humains. Un système expert se compose donc de deux parties principales : une *base de connaissances* et un *mécanisme de raisonnement*.

Ce rapport est composé de deux parties. Dans la première partie, nous présentons les principes et les aspects d'implémentation d'un système expert basé sur moteur d'inférence COTO en déterminant l'espace d'états, les opérateurs applicables et la stratégie du raisonnement. Dans la deuxième partie, l'application de COTO en Analyse de Données est présentée. L'intérêt de guider un logiciel d'analyse de données par un système expert est de choisir les méthodes les plus adaptées au problème de l'utilisateur et d'aider celui-ci à interpréter les résultats d'analyse.

II. MOTEUR D'INFERENCE COTO

II.1. Caractéristiques

COTO est implémenté en Fortran 77 de façon à permettre une insertion aisée dans une bibliothèque de programmes d'analyse de données, mais c'est un moteur d'inférence, donc un système "essentiel", indépendant du domaine d'application.

COTO admet des variables numériques et alphanumériques, qui sont obligatoirement instanciées au moment de l'inférence et ne sont pas des inconnues. Il est également capable d'effectuer les calculs numériques sur n'importe quel type d'expressions arithmétiques.

COTO permet l'utilisation de procédures externes (subroutines) et leur connexion aux règles.

COTO fonctionne en chaînage avant, chaînage arrière et en mixte selon le besoin de l'utilisateur.

COTO fonctionne en logique trivaluée : affirmation, négation, ignorance.

COTO dialogue avec l'utilisateur en langage quasi-naturel, c'est-à-dire compréhensible, en principe par un "non expert".

COTO peut expliquer son raisonnement en indiquant le chemin pour obtenir la solution, les règles ayant contribué à la progression à chaque moment, la raison de l'échec, ...

II.2. Le schéma des systèmes experts avec COTO

Nous nous limitons ici à une classe de systèmes experts basés sur des règles de production et faits de deux composantes :

1. **Une base de connaissances** : Il s'agit d'une base de règles **R** et d'une base de faits **F**. Les règles sont stockées dans un fichier de texte séparé du programme. Elles sont fournies par les experts d'un domaine précis qui utilisent des descriptions symboliques pour déclarer les situations et les actions correspondantes. Les faits sont des données relatives aux objets concrets à traiter, considérés comme les données du problème et entrés principalement par l'utilisateur du système.

2. Moteur d'inférence COTO : Il est le coeur du système et se compose de quatre modules:

(i) **Compilateur de règles :** Lecture de la base de règles stockée sous forme externe dans un fichier et construction de la représentation interne de ces règles sous forme d'un réseau d'inférence.

(ii) **Acquisition des informations:** Il s'agit d'un ensemble de procédures permettant l'acquisition et la modification des faits dans la base de faits.

(iii) **Inférence :** Il s'agit de deux opérateurs de base pour déclencher et désactiver les règles et de réaliser les deux modes de raisonnement en chaînage avant, en chaînage arrière et le contrôle du système.

(iv) **Dialogue :** Son rôle est l'interface homme-machine. Il peut poser des questions sur la situation considérée pour demander, par exemple les "symptômes" du problème et de répondre aux questions de l'utilisateur, par exemple pour expliquer le raisonnement.

II.3 Représentation de connaissances dans COTO

Le problème de la représentation de connaissances est de trouver une structure d'information interprétable par une machine.

Dans la forme externe, les connaissances sont représentées dans COTO comme les règles de production. Elles sont écrites en vrac, chacune étant sous la forme d'une succession de prémisses et de conclusions :

SI <premisses> ALORS <conclusions>

Les prémisses et les conclusions sont considérées comme les composantes de la règle, et elles ont la syntaxe suivante :

```

<composante> ::= <objet 1> <relation> <objet 2>
<objet 1>  ::= <mot> | (<mot>)
<objet 2>  ::= <mot> | (<mot>)
<relation> ::= > | >= | < | <= | = | <>  (comparaison des variables numériques)
<relation> ::= == | ><                    (comparaison des variables alphanumériques)
<relation> ::= := | <== | <--              (opérateurs d'affectation)
<mot>      ::= <caractère non blanc> | (<caractère non blanc>)

```

Les prémisses et les conclusions peuvent contenir des variables qui peuvent être de deux types bien distincts : numérique ou alphanumérique. Les variables numériques peuvent apparaître comme des inconnues dans des fonctions multivariées et prendre n'importe quelle valeur réelle. Les variables alphanumériques ne peuvent que prendre des valeurs exclusives dans un ensemble bien déterminé.

Concrètement, <objet 1> peut prendre l'une des formes suivantes:

- . Une **fonction** de plusieurs variables numériques représentée par le nom de la fonction et une liste des variables entre deux parenthèses. Dans ce cas, <objet2> peut être une autre fonction, une variable ou une constante.
- . Une **variable numérique**, dans ce cas <objet 2> peut être une autre variable numérique ou une constante.
- . Une **variable alphanumérique**, dans ce cas <objet 2> est une constante alphanumérique.
- . Une **proposition** qui prend une valeur vraie ou fausse.

Dans les prémisses, <relation> est l'un des opérateurs de comparaison. Dans les conclusions, c'est un opérateur d'affectation :

- := affectation pour une variable alphanumérique
- <-- affectation pour une variable numérique avec effacement
- <== affectation pour une variable numérique sans effacement

En conclusion, <objet 2> peut être un appel d'une fonction ou subroutine.

Dans la forme interne, les règles de COTO sont structurées comme un réseau d'inférence qui représente leurs relations. La description de chaque composante d'une règle dans ce réseau se compose de son étiquette, le nom d'objet dans le dictionnaire, son rôle et type, le sens et un pointeur qui pointe à l'adresse suivante de cet objet dans le réseau d'inférence.

Le rôle de la composante indique si l'objet est en prémisse ou en conclusion, si la relation est une comparaison ou une affectation, la nature de variable (numérique ou alphanumérique), le type de règle de production ...

Le sens de la composante indique que si c'est une affirmation ou une négation (pour une proposition), la valeur du seuil, la valeur numérique à affecter à une variable, l'adresse d'une variable alphanumérique ou une expression numérique, ...

Chaque objet peut apparaître dans plusieurs règles. Il est associé à un pointeur qui pointe à son adresse dans une autre règle concernant cet objet. Cela a pour but de propager des assertions dans le processus d'inférence.

Les faits sont enregistrés dans la base de faits sous forme des triplets <objet,valeur,type> qui présentent le nom de l'objet, la valeur affectuée, le type du fait (affirmé ou déduit).

Pour faciliter le travail de l'expert en représentant ses connaissances, certaines conceptions sont considérées comme intermédiaires et n'apparaient jamais devant les yeux de l'utilisateur. Elles sont représentées dans les règles de résumé qui ont la même forme que les règles de production mais le mot "Alors" est remplacé par "Ecrire".

II.4. Le contrôle du système

Dans ce paragraphe, nous décrivons la méthode de résolution de problème implémentée dans COTO en terme de l'espace d'états et d'opérateurs et la stratégie du contrôle du système.

La représentation de connaissances sous forme d'un réseau d'inférence décrit précédemment ne présente qu'une structure statique. On ajoute à COTO une mémoire du raisonnement pour stocker dynamiquement les informations sur le problème en cours de traitement.

Une prémisses ou conclusion d'une règle est dite **en négation** (contredite) si sa négation appartient à la base de faits. Trois états alternatifs d'une règle en cours du raisonnement sont définis, et à chacun d'eux correspond une valeur de la **fonction d'état s** :

- . **fermé** ($s = -1$) : si elle contient au moins une prémisses en négation.
- . **ouvert** ($s = 0$) : si elle ne contient aucune prémisses en négation et il existe des prémisses qui ne sont pas encore déclenchées.
- . **rempli** ($s = 1$) : si toutes les prémisses sont déclenchées et satisfaites.

Par cette définition, la base statique de règles est plongée dans un espace dynamique dans lequel on cherche la solution du problème.

L'espace d'états du problème est déterminé comme l'ensemble de toutes les configurations possibles de la base de règles. Chaque configuration est constituée par les états de toutes les règles en cours de raisonnement à un moment quelconque.

Au lancement du système, toutes les règles sont ouvertes et cet état est appelé l'état **initial**.

Un ou plusieurs états contenant au moins une règle remplie peuvent être acceptés comme la solution du problème et sont appelés **états finaux**. Quelques états intermédiaires doivent être atteints avant d'arriver aux états finaux et sont appelés **sous-états finaux**.

La fonction principale du moteur d'inférence COTO est de spécifier un ensemble d'actions convenables pour passer de l'état initial aux états finaux. Ce processus est une chaîne de changements d'états intermédiaires.

CHANGEMENT D'ETAT

{état de R} -----> {nouvel état de R}

Deux opérateurs essentiels opérant sur l'ensemble de règles sont définis pour réaliser ces changements d'états basés sur les faits dans la base de faits.

Notons la règle $R_m \in R: \wedge p_i \Rightarrow \wedge c_j, i \in I, j \in J$, où $\wedge p_i$ est la partie prémisses, $\wedge c_j$ est la partie conclusions.

Opérateur de déclenchement (déclen): déclencher par modus ponens et modus tollens toutes les règles ouvertes contenant en prémisse ou en conclusion un fait considéré. Si $s(R_m) = 0$ et soit $f \in F$ le dernier fait entré alors:

si $f \in \{p_i, i \in I\}$ et $\forall i \in I: p_i \in F$

alors $s(R_m) := 1$ et $F := F \cup \{c_j, j \in J\}$

si $f \in \{p_i, i \in I\}$ et $\exists i \in I: p_i \notin F$

alors $s(R_m) := 0$

si $\exists j \in J: f = \neg c_j$ et $\exists i \in I: p_i \notin F$ et $\exists k \in I: p_k \notin F$

alors $s(R_m) := 0$

si $\exists j \in J: f = \neg c_j$ et $\exists i \in I: p_i \notin F$

alors $s(R_m) := -1$ et $F := F \cup \{\neg p_i\}$

si $f \in \{p_i, i \in I\}$ et $\exists k \in I: p_k \notin F$ et $\exists j \in J: \neg c_j \in F$

alors $s(R_m) := -1$ et $F := F \cup \{\neg p_k\}$

Opérateur de désactivation (desact) : fermer toutes les règles ouvertes contenant une prémisses contredite par un fait.

Si $s(R_m) = 0$ et $\exists i : f = \neg p_i$ alors $s(R_m) := -1$

Le domaine d'application de ces deux opérateurs, bien sûr, est l'ensemble des règles ouvertes contenant le fait considéré (les règles candidates). Chaque fois qu'un fait est ajouté dans la base de faits, ces deux opérateurs sont appliqués. En conséquence, quelques règles parmi les règles candidates restent encore ouvertes, les autres deviennent remplies ou fermés. L'intérêt du réseau d'inférence décrit précédemment est les opérateurs de déclenchement et de désactivation sont facilement appliqués "à la fois" à toutes les règles candidates par propagation des nouvelles informations. Autrement dit, chaque nouveau fait affirmé ou déduit engendre un changement d'état de la base de règles déterminé par l'application de ces deux opérateurs sur les règles ouvertes concernées par ce fait : l'opérateur de déclenchement permet de déduire les nouveaux faits et l'opérateur de désactivation permet de réduire l'espace de recherche.

Le raisonnement de COTO est réalisé par une **stratégie mixte** de chaînage avant et de chaînage arrière dont l'intérêt est de gérer convenablement les changements d'état pour atteindre rapidement les états finaux.

En **chaînage avant**, on essaie de déduire tous les faits que l'on peut déduire jusqu'à ce qu'il n'y ait plus de règles ouvertes candidates. Ce mode de raisonnement est implémenté par la procédure appelée "fchain" suivante :

Procédure de fchain

Soit F la base de faits et l'ensemble de faits déjà traités $F' = \emptyset$ au départ.

Tant que $F' \neq F$ faire

debut

Soit f le premier fait de $F \setminus F'$ alors $F' := F' \cup \{f\}$,

O_1 , O_2 et O_3 les ensembles des règles ouvertes

contenant respectivement f et $\neg f$ en prémisses, et $\neg f$ en conclusion:

Appeler Declen pour les règles de O_1 et O_3

et Desact pour les règles de O_2 .

fin.

En **chainage arrière**, on se fixe un but et on essaie de le déduire. Si toutes les prémisses d'une règle contenant ce but en conclusion sont satisfaites, alors le but est atteint, car il sera ajouté à la base de faits par le chainage avant. Sinon le système considère chaque prémisses inconnue de la règle comme un nouveau but et recommence le processus.

Ce mode de raisonnement est implémenté par la procédure récursive appelée "bchain". Un arbre des buts est construit pas à pas. Chaque chemin de la racine (but final) à une feuille (sous but courant) est appelé **raisonnement**.

Un raisonnement est dit **cyclique** s'il contient deux fois le même sous but: cela veut dire que l'on cherche à démontrer un énoncé à partir de lui même!... Par exemple, un cycle est détecté si on cherche à démontrer le but d avec les règles suivantes :

$$\{ a \Rightarrow b; b \wedge c \Rightarrow d; e \Rightarrow a; b \Rightarrow e \}$$

Le raisonnement " $d \Leftarrow b \Leftarrow a \Leftarrow e \Leftarrow b$ " est cyclique : la dernière règle " $b \Rightarrow e$ " ne peut donc être appliquée dans ce cas.

Un raisonnement est dit **contradictoire** si les prémisses et les conclusions des règles utilisées sont contradictoires. Cela veut dire que le raisonnement ainsi construit déduira un fait de son contraire!... Par exemple, une contradiction est détectée si on cherche à démontrer le but d avec les règles suivantes :

$$\{ c \wedge b \Rightarrow d; a \Rightarrow b; e \wedge \neg c \Rightarrow a \}$$

Le raisonnement " $d \Leftarrow b \Leftarrow a \Leftarrow \neg c$ " est contradictoire car la branche " $d \Leftarrow b$ " utilise la règle " $c \wedge b \Rightarrow d$ " qui est contredite par le dernier sous but " $\neg c$ ". La règle " $e \wedge \neg c \Rightarrow a$ " ne peut donc être utilisée dans ce cas.

Supposons que g est le but ou sous but à atteindre et H le raisonnement en cours de construction.

Procédure de *bchain*(*g*, *H*)

Demander à l'utilisateur si *g* est vrai

Si $g \notin F$

Alors déterminer l'ensemble *G* des règles ouvertes contenant *g* en conclusion.

Si $G \neq \emptyset$ alors faire

debut

Choisir une règle *R* dans *G* telle que le raisonnement $H \cup \{R\}$ ne soit ni cyclique ni contradictoire.

Si il existe une telle règle *R*

alors, pour toutes les prémisses *p* de *R*, faire *bchain*(*p*, $H \cup \{R\}$)

sinon ECHEC

fin.

Quand un fait est demandé à l'utilisateur, les reponses peuvent être "oui", "non" ou encore "je ne sais pas" ("!"). Dans ce dernier cas, le fait est considéré comme probable et la procédure *bchain* précédente va essayer de le déduire.

Lors du lancement du processus, COTO demande à l'utilisateur l'ensemble des faits initiaux et un but à prouver.

L'application des deux opérateurs de déclenchement et de désactivation sur les faits affirmés initialement détermine certains chemins possibles dans l'espace d'états.

S'il y a un but demandé par l'utilisateur, COTO essaie de le prouver en descendant l'arbre de buts. S'il n'y a aucun but au début, COTO vérifie les règles ouvertes concernant les faits dans la base de faits, et il essaie de demander ou déduire les faits inconnus dans ces règles. S'il n'y a ni fait ni but donné par l'utilisateur au début, COTO essaie de poser les questions "les plus informantes" déterminées par les fréquences d'occurrence de chaque énoncé en partie prémisses des règles.

II.5. La conversation et explication

Elles jouent des rôles importants en interface dans le processus du travail de COTO.

Par conversation, COTO peut éventuellement fournir les conseils d'experts ou la solution du problème, les informations dont l'utilisateur a besoin.

Pour entrer les faits dans la base de faits, l'utilisateur tape des phrases qui décrivent l'objet. Malheureusement, elles ne coïncident pas toujours avec les mots de l'expert pour les mêmes objets. Dans ce cas, l'utilisateur a besoin de "dialoguer" avec COTO pour identifier ces objets. Plusieurs possibilités ont été développées dans COTO :

- . **Accès au dictionnaire** séquentiellement (ordre alphabétique), ou par des types différents de mots que COTO distingue par la nature de variable (proposition, numérique, alphanumérique, procédure,...)

- . **Consultation du dictionnaire** : l'objet peut être décrit par un ou plusieurs mots maîtres ou par un ensemble de synonymes. Dans ce cas, COTO liste tous les objets de son dictionnaire contenant ces mots maîtres ou ces synonymes. Alors l'utilisateur peut trouver facilement l'objet identifié, sinon le processus est répété ou s'arrête.

Par des réponses prédéfinies, l'utilisateur peut intervenir sur le système pour:

- . **Insertion** d'un ou plusieurs faits dans la base de faits.

- . **Annulation** d'un ou plusieurs faits dans la base de faits . Elle permet de modifier la base de faits ou reraisonner à partir d'un fait quelconque déjà affirmé.

- . **Les informations** : Guide de raisonnement par l'affichage des informations utiles sur la situation en cours d'analyse.

- . **Accès à la base de faits** par l'affichage des faits affirmés ou déduits.

Par conversation, COTO explique son raisonnement. En général, le système montre les pas dans le chemin pour arriver au but, les règles ayant contribué à la décision, la raison de la décision,... Pour la question "pourquoi" ou "pourquoi pas", COTO pointe sur les règles sélectionnées ou la raison d'échec quand le but à atteindre ne correspond pas aux faits acquis.

III. LE LOGICIEL D'ANALYSE DE DONNEES SICLA

L'Analyse de données, en bref, peut être considérée comme un ensemble de techniques de traitement de l'information qui a pour but à trouver une nouvelle représentation plus simple de l'information destinée à être soumise à un observateur, et à partir de celle-la, de donner une interprétation.

Depuis les années soixante ces méthodes ont connu des développements importants. Leur champ d'application s'est considérablement élargi et le nombre d'utilisateurs dans tous les domaines se multiplie.

Tandis que nombreux sont les logiciels puissants, leur utilisation demande des connaissances statistiques et expérimentales considérables. Face à un ensemble de méthodes, l'utilisateur de l'analyse de données ignore bien souvent lesquelles sont les plus adaptées à son problème et comment interpréter les résultats obtenus. L'acquisition d'une expertise en analyse de données n'est une tâche ni facile ni rapide et pourtant, tout praticien en biologie, géologie, physique, sciences sociales,... a besoin d'utiliser ces méthodes de façon permanente et efficace. C'est pour cette raison que les problèmes en analyse de données sont fréquemment mal résolus ou les résultats mal interprétés.

Depuis quelques années, beaucoup d'auteurs s'orientent vers les techniques d'intelligence artificielle, en particulier les systèmes experts pour résoudre ce type de problème. Un résumé de cette tendance peut être trouvé dans /DEQURA/, /GALE/.

Dans ce rapport nous présentons un autre système expert, CLAVETO, pour le logiciel d'analyse de données SICLA. C'est une application du moteur d'inférence COTO comme assistance intelligente en analyse de données. Implémenté en Fortran 77, ce moteur d'inférence est capable d'interagir avec SICLA de façon transparente pour l'utilisateur.

Dans le système expert CLAVETO, une base de connaissances en analyse de données est développée sur deux branches :

- . pour le **choix d'une méthode** la plus adaptée à l'information à analyser,
- . pour **l'aide de l'interprétation** des résultats obtenus après une analyse.

Les premiers résultats de ce système expert conduisent à un avantage considérable : on peut utiliser facilement le logiciel SICLA sans avoir une maîtrise parfaite de l'analyse de données.

III.1. Les phases principales d'une analyse de données

On peut classer les techniques de l'analyse de données en deux branches principales suivant la technique mathématique utilisée :

. **l'analyse factorielle** dont le but est d'obtenir une représentation de l'information dans un espace de dimension faible de telle manière qu'elle soit la moins déformée possible.

. **la classification automatique** qui groupe l'ensemble des méthodes et algorithmes consistants à découper une population d'objets en plusieurs classes en tenant compte des variables qui les caractérisent.

Si l'on définit une information comme un couple formé d'une représentation et de ses interprétations, on peut dire que ***l'objet de l'analyse de données est de donner une interprétation à une représentation de l'information*** /SIMON/.

Une représentation est le support matériel d'une information. Il se peut que nous ayons à considérer une représentation comme une variable. L'ensemble des variables est appelé **espace de représentation**. Le terme d'espace indique que des lois peuvent opérer sur les variables. Une analyse de données peut donc être considérée comme composée de trois phases essentielles :

- 1) Les étapes préliminaires dans l'espace de représentation
- 2) La réalisation d'une méthode de l'analyse de données
- 3) L'interprétation du résultat de l'analyse.

1. Les étapes préliminaires d'une analyse de données dans l'espace de représentation sont les suivantes :

a) **Détermination des individus et des variables.**: La méthode choisie pour l'analyse dépend tout d'abord de la nature de variables qui détermine la structure algébrique de l'espace de représentation.

b) **Extraction de tableau de données à analyser** : Les informations en analyse de données sont souvent représentées sous forme de tableau de données résultant de la définition de l'ensemble des individus et des variables. Selon l'espace de représentation et la structure algébrique qui lui est attachée, il y a différents types de tableau de données : tableau de mesures hétérogènes, tableau binaire, tableau de contingences, tableau de préférences, tableau des modalités, ...

c) **Descriptions élémentaires** : Soit par des mesures quantitatives (moyennes, variances, corrélations,...), soit à l'aide de description visuelles (histogrammes, diagrammes, hiérarchies,...). Elles ont pour but de donner une vue globale sur les données.

d) **Changement et codage de variables** : Création des nouvelles variables exploitables ou sur lesquelles on va effectivement travailler.

e) **Choix d'une mesure de ressemblance** (ou de dissemblance): Il s'agit de choisir dans l'ensemble des distances disponibles la plus adaptée à la méthode choisie et aux données du problème.

Les deux dernières phases sont les suivantes :

2. **Analyse de tableau des données** : Suivant la structure de données et l'objectif de recherche, les analyses sont choisies et effectuées.

3. **Interprétation des résultats de l'analyse** : C'est l'étape la plus délicate à mettre en oeuvre et sur laquelle il y a encore peu de recherche. L'interprétation des résultats est toujours l'une des plus difficiles pour l'utilisateur de l'analyse de données.

III.2. Présentation du logiciel SICLA

SICLA est un Système Interartif de CClassification Automatique permettant la gestion, la description élémentaire et le traitement des données par les méthodes d'analyse multidimensionnelle /LERAGO/.

Son objectif principal est la diffusion de méthodes issues de recherches récentes, principalement celles concernant les nuées dynamiques, mais il peut également accéder aux méthodes factorielles de la bibliothèque SPAD et aux méthodes de la bibliothèque MODULAD.

Le système SICLA propose un certain nombre de fonctionnalités couvrant les différentes phases présentées dans le paragraphe II., et son architecture est adaptée à la pratique de l'analyse de données.

III.2.1. Les fonctionnalités du système

SICLA comporte plusieurs modules réalisant les différentes fonctions suivantes :

- a) **Enregistrement des données** : saisie, contrôle, validation.
- b) **Description élémentaire** : calcul de divers indicateurs relatifs aux variables des données (min, max, moyennes, etc ...).
- c) **Analyse de données** : Ce sont les divers modules d'analyses statistiques regroupés sous les thèmes de classification de type nuées dynamiques , hiérarchie, méthodes factorielles, méthodes de discrimination.
- d) **Description des structures** : Les méthodes de l'analyse de données engendrant des structures (partition, hiérarchies, plans factoriels, etc ...) qui peuvent être analysées, mises en rapport par exemple avec des données différentes de celles qui ont servi à les calculer.

e) **Edition des données et des structures** : Ce sont les méthodes d'impression et de mise en page.

f) **Gestion des données** : Ce thème regroupe l'ensemble des modules permettant de sélectionner des individus ou variables , de créer de nouveaux tableaux, etc ...

L'ensemble de ces modules écrits en FORTRAN 77, permet l'analyse effective d'un tableau de données, son édition ou la mise en forme des résultats.

III.2.2. Architecture de SICLA

Le système SICLA est constitué d'un ensemble d'objets , représentations informatiques des structures mathématiques de l'analyse de données, et d'un ensemble de modules permettant d'effectuer les différentes tâches d'analyses, de gestion, d'édition, ...

L'orientation principale du système est le contrôle du "bon emploi" de l'information. Le système est régi par un ensemble de règles portant :

- . sur *la compatibilité des modules et des objets* (un module de gestion de tableau individus * variables ne peut opérer sur un tableau de distances par exemple).

- . sur *la compatibilité des méthodes statistiques et des tableaux de données* (une classification nuées dynamiques sur tableau de modalités ne peut s'exécuter sur un tableau de mesures par exemple)

- . sur *la conformité des données à la structure mathématique choisie* (le type tableau de contingence ne pourra être appliqué à un tableau comportant des valeurs négatives par exemple)

Cette dernière contrainte d'intégrité relèvera uniquement du logiciel SICLA, et ne pourra pas être traitée par un système expert car elle consiste en des vérifications de nature numérique portant sur les domaines de variation des variables.

III.2.3. Les aides traditionnelles

On offre généralement dans un système fonctionnant par un ensemble de commandes, une assistance on_line de la forme suivante :

HELP nom_commande

A la reception de cet ordre le système ira chercher et listera un fichier d'information lié à la commande.

Cette forme d'aide est utile pour un utilisateur averti, car elle permet de rappeler l'objectif ou les arguments liés à la procédure. Elle est insuffisante cependant pour un néophyte peu habitué à système. En effet il doit au préalable faire une recherche ou deviner la commande adaptée à son problème. Cela peut être particulièrement fastidieux si le système comporte un grand nombre de commandes (par exemple, SICLA comporte environ 50 commandes).

Aussi, pour une utilisation orientée vers des utilisateurs débutants, la solution généralement adoptée est celle du menu hiérarchique. Les procédures sont regroupées en divers sous-menus, l'utilisateur guidé par la hiérarchie accède plus facilement à la commande désirée.

Si le menu est bien organisé et les thèmes de sous-menu est bien choisis, cette structure permet une bonne description du système. Ainsi le menu associé à SICLA comporte quatre grandes rubriques **édition, gestion, description, analyse** qui reflètent les principaux types de traitement dans l'analyse d'un tableau de données.

Critique des schémas habituels

Les principaux reproches que l'on peut faire aux aides traditionnelles sont :

- . **Passivité** plus ou moins grande du système : c'est l'utilisateur qui doit prendre l'initiative, connaître ou deviner la commande qu'il cherche. Le système ne servant que d'aide_mémoire.

. **L'environnement** de l'utilisateur n'est pas pris en compte : il peut choisir une commande illicite sans que le système intervienne, ce dernier ne réagissant qu'à l'exécution de la procédure.

. **Le savoir de l'expert** du domaine, du concepteur du système ne sont pas pris en compte pour guider l'utilisateur dans le choix d'une commande et l'interprétation des résultats.

. **La nature statique** et figée d'un menu ne permet pas de suggérer des stratégies . Par un menu on aboutit toujours, et seulement, à une commande alors qu'une assistance par un système expert permettra de proposer des enchainements de méthodes dynamiques dépendants des données de l'utilisateur, des résultats obtenus et des objectifs poursuivis. Un ensemble cohérent de règles peut dans une certaine mesure permettre au système "d'accompagner" l'utilisateur dans sa progression dans l'étude d'un tableau de données.

Dans le cadre de SICLA, il existe déjà une assistance intelligente, CLAVECIN /DEQURA/, qui est réalisée à partir du moteur d'inférence OURCIN. Nous allons maintenant présenter un autre système expert pour guider l'utilisation de SICLA, basé sur le moteur d'inférence COTO, qui est incorporé dans SICLA avec les modules développés à partir de la maquette CLAVECIN.

IV. LE SYSTEME EXPERT CLAVETO

Dans ce paragraphe nous présentons un système expert guidant l'utilisation des programmes de l'analyse de données : CLAVETO . Ce système expert se compose d'une base de connaissance sur l'analyse de données construite par les auteurs du logiciel, et le moteur d'inférence COTO.

IV.1. Representation de connaissances dans CLAVETO

La connaissance d'expert est exprimée sous forme de règles de production opérant sur les déclarations de l'univers de l'analyse de données. Elles sont de la forme :

SI prémisses ALORS conclusions

Généralement, nous rappellerons "variables" certains mots du vocabulaire qui représentent les concepts utilisés en analyse de données et peuvent prendre plusieurs valeurs différentes. Ainsi, le "type de tableau de données", "l'objectif de la classification", "le type de la représentation", "la commande à exécuter", ... sont des variables.

Chaque variable est associée à un ensemble de valeurs qui la caractérise. Par exemple, pour la variable "type de tableau de données", on pourra préciser l'une des valeurs suivantes dans une situation concrète : {"mesure hétérogènes", "contingence", "binaire", "préférence", "modalités", "mesure homogènes", "hétérogène", "burt", "distances"}. Comme les cas particuliers, les variables peuvent prendre une seule valeur, ce sont des propositions, par exemple: "message d'erreur relatif à données hors bornes", ...

En correspondance avec les étapes principales d'une analyse de données et les thèmes du logiciel SICLA, les connaissances dans CLAVETO sont structurées sous forme arborescente des variables. Les variables définissent les classes de commandes à exécuter qui respectent la structure de SICLA.

Exemple :

```

--> TYPE DE TRAITEMENT == STRATEGIE
    --> TYPE DE TABLEAU DE DONNEES == MESURE HETEROGENES
        --> DEMARCHE PROPOSEE:
            --> DESCRIPTION UTILISER LES COMMANDES : DESQAN, CORREL, HISTO
            --> HOMOGENEISATION DU TABLEAU : CENTRAGE REDUCTION
            --> UTILISER : METRIQ
            --> PUIS POUR L'ANALYSE UTILISER LES COMMANDES :
                MNDQAN , WARD , ACOMP /SPAD/
        --> TYPE DE TABLEAU DE DONNEES == HETEROGENE
            --> DEMARCHE PROPOSEE:
                --> FAIRE DES DESCRIPTIONS ELEMENTAIRES
                    DES VARIABLES QUANTITATIVES ET QUALITATIVES
                --> HOMOGENEISER LE TABLEAU EN CODANT LES
                    VARIABLES QUANTITATIVES EN QUALITATIVES PAR CODQAN
                --> ANALYSER CE NOUVEAU TABLEAU
        --> TYPE DE TABLEAU DE DONNEES == CONTINGENCE
            --> DEMARCHE PROPOSEE:
                --> POUR L'ANALYSE UTILISER LES COMMANDES:
                --> ANALYSE DES CORRESPONDANCES CORBI /SPAD/
                --> CLASSIFICATION CROISEE : CROKI2
            .....

```

IV.2. Les différents types de connaissances

Nous présenterons maintenant les groupes de connaissances sur l'analyse de données. Elles sont classifiées par les auteurs selon les thèmes du traitement de SICLA. Mais bien entendu, comme dans tous les systèmes experts, les règles sont écrites en vrac et l'ordre d'écriture n'a aucune importance sur le raisonnement du système. Dans ces règles, les commentaires sont écrits en minuscules.

IV.2.1 Les étapes préliminaires dans l'espace de représentation.

I.V.2.1.1. Règles générales du système

Ce type de règles détermine la nature de variable selon le type de tableau des données et les conventions du système. Certaines règles intermédiaires définies par la volonté de l'expert qui représentent des faits invisibles à l'utilisateur (ils n'apparaîtront jamais devant ses yeux même lors d'explications) , sont écrites sous forme de règles de résumé.

Exemple :

REGLE 3

Si TYPE DE TABLEAU DE DONNEES == BINAIRE

Alors NATURE DE VARIABLES == QUANTITATIVES

REGLE 5

Si TYPE DE TABLEAU DE DONNEES == MODALITES

Alors NATURE DE VARIABLES == QUALITATIVES

et les règles intermédiaires, par exemple :

REGLE 41

Si TYPE DE TABLEAU DE DONNEES == HETEROGENE

et CHOIX DE VARIABLE == QUANTITATIVES

et TYPE DE VARIABLES == COMPTAGES

et ANALYSE DE DONNEES MULTIDIMENSIONNELLES

Ecrire CAS DE FIGURE <-- 3

REGLE 42

Si TYPE DE TABLEAU DE DONNEES == HETEROGENE

et CHOIX DE VARIABLE == QUANTITATIVES

et TYPE DE VARIABLES == RANG

et ANALYSE DE DONNEES MULTIDIMENSIONNELLES

Ecrire CAS DE FIGURE <-- 1

IV.2.1.2. Règles sur la description de données

C'est l'étape d'extraction d'un tableau de données et des descriptions élémentaires : calculs des moyennes, les min-max, calcul et édition des histogrammes, la matrice des corrélations, construction du tableau de forme forte, ...

Exemple :

REGLES 11

Si TYPE DE TRAITEMENT == DESCRIPTION DE DONNEES
 et NATURE DE VARIABLE == QUANTITATIVES
 et ETUDE DE LIAISON ENTRE VARIABLES
 Alors COMMANDE := CORREL
 et Calcul & impression de matrice de corrélations

REGLES 12

Si TYPE DE TRAITEMENT == DESCRIPTION DE DONNEES
 et NATURE DE VARIABLE == QUALITATIVES
 et ETUDE DE LIAISON ENTRE VARIABLES
 Alors COMMANDE := AKH12
 et Calcul & analyse de tableau de contingence

REGLES 36

Si TYPE DE TRAITEMENT == DESCRIPTION DE DONNEES
 et PRESENCE DE MULTIPARTITION
 et RECHERCHE DE FORME FORTE
 Alors COMMANDE := FF
 et Construction & analyse du tableau de forme forte

IV.1.2.3. Règles concernant la gestion des données

Celle-là aident à enregistrer des données, de manipuler des structures de données et de générer les variables. Ces types de règles sont importantes car les résultats des analyses

dependent de la qualité du tableau :

. Les règles sur l'enregistrement des données :

REGLE 14

Si TYPE DE TRAITEMENT == GESTION DE DONNEES
et ENREGISTREMENT DES DONNEES
Alors COMMANDE := ENRI

REGLE 15

Si TYPE DE TRAITEMENT == GESTION DE DONNEES
et ENREGISTREMENT DU DICTIONNAIRE DES VARIABLES
Alors COMMANDE := ENRV

. Règles sur la génération des variables

REGLE 19

Si TYPE DE TRAITEMENT == GESTION DE DONNEES
et CREATION DE DONNEES CENTRES DE GRAVITE
Alors COMMANDE := CRPACG

REGLE 20

Si TYPE DE TRAITEMENT == GESTION DE DONNEES
et CREATION DES NOUVELLES VARIABLES QUANTITATIVE
Alors COMMANDE := CREQAN
et Création de variables quantitatives par des
opérateurs arithmétiques & logiques

REGLE 21

Si TYPE DE TRAITEMENT == GESTION DE DONNEES
et CREATION DES NOUVELLES VARIABLES QUALITATIVE
Alors COMMANDE := CREQAL
et Création de variables qualitatives par des
opérateurs arithmétiques & logiques

Règles sur la manipulation des structures de données

REGLE 24

Si TYPE DE TRAITEMENT == GESTION DE DONNEES
 et STRUCTURE DE DONNEES A CHANGER == DONNEES COURANTES
 Alors COMMANDE := DONNEE
 et Changement de la structure de données courante

REGLE 25

Si TYPE DE TRAITEMENT == GESTION DE DONNEES
 et NATURE DE VARIABLES == QUALITATIVES
 et CREATION DE TABLEAU DE CONTINGENCE
 Alors COMMANDE := CRCONT
 et Creation de tableau de contingence

REGLE 27

Si TYPE DE TRAITEMENT == GESTION DE DONNEES
 et TYPE DE TABLEAU DE DONNEES == HETEROGENE
 et CREATION DE TABLEAU DE DISTANCES
 Alors SOIT LA COMMANDE : CRDSQN
 et Tableau de distances relatif aux quantitatives
 et SOIT LA COMMANDE : CRDSQL
 et Tableau de distances relatif aux qualitatives

IV.1.2.4. Règles concernant l'édition des données

Ce sont des règles sur l'édition (ecran) et l'impression (listing), par exemple:

REGLE 32

Si TYPE DE TRAITEMENT == EDITION DE DONNEES
 et PRESENCE DE HIERARCHIE
 Alors COMMANDE := EDHIAR
 et Edition d'un arbre hierarchique

REGLE 37

Si TYPE DE TRAITEMENT == EDITION DE DONNEES
et EDITION DE DONNEES DE L'ARCHIVE

Alors COMMANDE := EDAR

et Impression des constituants de l'archive

IV.1.2.5. Règles sur l'évaluation des données

Il n'est pas évidemment possible de vérifier la représentativité du tableau par les procédures automatiques mais des diagnostics préliminaires concernant les dimensions du tableau peuvent être fournis. Cela éviterait, par exemple, de voir des utilisateurs analyser par une régression des tableaux comportant 10 individus ou d'effectuer des analyses de types factorielles sur des ensembles comportant deux fois plus de variables que d'individus, etc ... (encore que ce genre de règle soit très controversé parmi les spécialistes en Analyse de Données).

Exemples :

REGLE 97

Si TYPE DE TRAITEMENT == EVALUATION DE DONNEES

et NOMBRE DE VARIABLES > NOMBRE D'INDIVIDUS

Alors POUR ANALYSER UN TEL TYPE DE TABLEAU DE DONNEES

et CHOISISSEZ DES SOUS ENSEMBLE DE VARIABLES DE MANIERE QUE

et NOMBRE DE VARIABLES EST INFÉRIEUR A NOMBRE D'INDIVIDUS

REGLE 98

Si TYPE DE TRAITEMENT == EVALUATION DE DONNEES

et NOMBRE DE VARIABLES QUANTITATIVES > 100

Alors NOMBRE DE VARIABLES EST TROP IMPORTANT

et POUR UNE REDUCTION DE L'INFORMATION PERTINENTE

et CHOISIR DES SOUS ENSEMBLES DE VARIABLES

et COMPORTANT UNE DIZAINE DE VARIABLES

et FAIRE UNE DESCRIPTION DE DONNEES POUR SELECTIONNER

REGLE 99

Si TYPE DE TRAITEMENT == EVALUATION DE DONNEES
 et NOMBRE DE MODALITES TOTAL > 100
 Alors NOMBRE DE MODALITES EST TROP IMPORTANT
 et POUR UNE REDUCTION DE L'INFORMATION PERTINANTE
 et CHOISISSEZ DES SOUS ENSEMBLES DE QUESTIONS
 et RELATIF A UN MEME THEME DE L'ENQUETE PAR EXEMPLE

IV.1.2.6. Règles sur les codages

Exempls :

REGLE 89

Si TYPE DE TRAITEMENT == CODAGE
 et TYPE DE TABLEAU DE DONNEES == HETEROGENE
 Alors COMMANDE := CODQAN
 et Codage : Quantitatif ---> Qualitatif
 et Codage par effectifs égaux ou bornes imposées
 et VOIR AUSSI : CREQAN
 et Création de variables quantitatives par des
 et opérateurs arithmétiques & logiques

REGLE 90

Si TYPE DE TRAITEMENT == CODAGE
 et TYPE DE TABLEAU DE DONNEES == MODALITES
 Alors COMMANDE := CREQAL
 et Création de variables qualitatives par des
 et opérateurs arithmétiques & logiques

V.2.2. Choix des méthodes d'analyse

C'est la partie la plus importante de l'assistance fournie, les différents modules relatifs aux traitements possibles (gestion, analyse, interprétation,...) ont un contexte d'exécution

et des objectifs précis. Nous étudions pour chaque méthode d'analyse de données, les structures de la représentation de l'information sur laquelle elle peut s'appliquer, ses objectifs et nous les traduisons sous forme des règles. Un tableau qui figure les types de tableaux adéquats pour les méthodes factorielles peut trouver dans /DEQURA/.

Dans le système SICLA, ce sont les auteurs des programmes qui décident du champ d'application de la méthode, la base de connaissance ne reflétera que les différents choix des auteurs de programmes. Une fois la procédure choisie, l'utilisateur doit spécifier les paramètres de mise en oeuvre : les programmes proposent en général des options par défaut. Toutefois lorsque les paramètres dépendent de la nature des données (choix de métrique par exemple) le type du tableau étant un fait connu du système expert, il est possible de proposer avec la méthode choisie, la métrique usuellement utilisée pour le type de tableau de données.

Les règles concernant le choix d'une méthode de l'analyse sont divisées en trois sous groupes :

IV.2.2.1. Règles sur la classification automatique

REGLE 57

Si TYPE DE TRAITEMENT == CLASSIFICATION
 et CAS DE FIGURE = 1
 et OBJECTIF DE CLASSIFICATION == LES INDIVIDUS
 et TYPE DE REPRESENTATION == PARTITION
 Alors COMMANDE := MNDQAN
 et UTILISER LA COMMANDE INPAQN POUR INTERPRETER LE RESULTAT

REGLE 59

Si TYPE DE TRAITEMENT == CLASSIFICATION
 et CAS DE FIGURE = 1
 et OBJECTIF DE CLASSIFICATION == LES INDIVIDUS
 et TYPE DE REPRESENTATION == HIERARCHIE
 Alors SOIT COMMANDE := WARD
 et SOIT COMMANDE := SAUT
 et LA DISTANCE EUCLIDIENNE EST PROPOSEE

REGLE 61

Si TYPE DE TRAITEMENT == CLASSIFICATION
 et CAS DE FIGURE = 2
 et OBJECTIF DE CLASSIFICATION == LES INDIVIDUS
 et TYPE DE REPRESENTATION == HIERARCHIE
 Alors SOIT COMMANDE := WARD
 et Hiérarchie avec indice accroissement de l'inertie
 et SOIT COMMANDE := SAUT
 et Hiérarchie du saut minimum ou maximum
 et LE CENTRAGE REDUCTION EST FORTEMENT CONSEILLE
 et UTILISATION DE LA COMMANDE METRIQ
 et choix d'une métrique pour les variables quantitatives

REGLE 68

Si TYPE DE TRAITEMENT == CLASSIFICATION
 et CAS DE FIGURE = 3
 et OBJECTIF DE CLASSIFICATION == LES VARIABLES
 et TYPE DE REPRESENTATION == PARTITION
 Alors COMMANDE := CRSOVA
 et Transposition des données
 et PUIS LA COMMANDE : MNDQAN
 et Méthode des nuées dynamiques sur variables quantitatives
 et UTILISER COMMANDE INPAQN POUR INTERPRETER LE RESULTAT
 et LA DISTANCE DU KHI2 SUR LES PROFILS EST PROPOSEE
 et UTILISATION DE LA COMMANDE : METRIQ

REGLE 74

Si TYPE DE TRAITEMENT == CLASSIFICATION
 et CAS DE FIGURE = 4
 et OBJECTIF DE CLASSIFICATION == LES MODALITES
 et TYPE DE REPRESENTATION = PARTITION
 Alors COMMANDE := CROMUL
 et Classification croisée sur modalités

IV.2.2.2. Les règles sur l'analyse factorielle

Exemple :

REGLE 77

Si TYPE DE TRAITEMENT == FACTORIELLE
 et CAS DE FIGURE = 1
 Alors COMMANDE := SISPAD
 et Interface avec SPAD ou MODULAD
 et EXECUTER L'ETAPE : ACOMP DANS SPAD

REGLE 79

Si TYPE DE TRAITEMENT == FACTORIELLE
 et CAS DE FIGURE = 3
 Alors COMMANDE := SISPAD
 et Interface avec SPAD ou MODULAD
 et EXECUTER L'ETAPE : CORBI DANS SPAD

REGLE 80

Si TYPE DE TRAITEMENT == FACTORIELLE
 et CAS DE FIGURE = 4
 Alors COMMANDE := SISPAD
 et Interface avec SPAD ou MODULAD
 et EXECUTER L'ETAPE : MULTM DANS SPAD

IV.2.2.3. Règles sur l'analyse discriminante

Exemple :

REGLE 81

Si TYPE DE TRAITEMENT == DISCRIMINATION
 et NATURE DE VARIABLES == QUANTITATIVES
 et VARIABLES EXPLICATIVES == QUANTITATIVES
 et VARIABLE A EXPLIQUER == QUANTITATIVE
 Alors FAIRE UNE REGRESSION LINEAIRE

REGLE 82

Si TYPE DE TRAITEMENT == DISCRIMINATION
 et TYPE DE TABLEAU DE DONNEES == HETEROGEE
 et VARIABLES EXPLICATIVES == QUANTITATIVES
 et VARIABLE A EXPLIQUER == QUALITATIVE
 Alors COMMANDE := DISC
 et Discrimination linéaire
 et SOIT LA COMMANDE : DNP
 et Discrimination non paramétrique
 et SOIT LA COMMANDE : VOISIN
 et Méthode des K plus proches voisins

REGLE 84

Si TYPE DE TRAITEMENT == DISCRIMINATION
 et NATURE DE VARIABLES == QUALITATIVES
 et PRESENCE DE PARTITION
 et UNE PARTITION A DISCRIMINER
 et VARIABLES EXPLICATIVES == QUALITATIVES
 et NOMBRE DE CLASSES A DISCRIMINER > 2
 Alors FAIRE UNE CORRESPONDANCE MULTIPLE
 et PUIS UNE ANALYSE DISCRIMINANTE SUR LES FACTEURS

REGLE 85

Si TYPE DE TRAITEMENT == DISCRIMINATION
 et NATURE DE VARIABLES == QUALITATIVES
 et PRESENCE DE PARTITION
 et UNE PARTITION A DISCRIMINER
 et VARIABLES EXPLICATIVES == QUALITATIVES
 et NOMBRE DE CLASSES A DISCRIMINER = 2
 Alors COMMANDE := DISQL2
 et Discrimination qualitative

IV.2.3. Aide à l'interprétation des résultats

Les méthodes de l'analyse de données produisent des états de sorties : graphiques et tableaux d'indicateurs statistiques. Le dépouillement de ces résultats nécessite une bonne connaissance des méthodes et souvent l'assistance d'un statisticien pas toujours disponible. Pour faciliter cette phase, deux approches sont développées dans le cadre de SICLA :

- . Divers outils graphiques basés sur les normes GKS
- . Les règles d'interprétation des indicateurs statistiques.

L'interprétation des résultats obtenus après une analyse de données est encore peu étudiée. D'une part l'interprétation dépende des significations des indicateurs obtenus par les méthodes proposées par les experts de l'analyse de données. D'autre part elle dépende d'expériences d'analyseurs dans ses domaines de travail. L'intérêt d'avoir des règles d'interprétation guidant la plupart des utilisateurs de l'analyse de données nous conduit de développer ce point dans la base de connaissances et dans la construction du moteur d'inférence COTO. Nous commençons ici avec les règles sur l'interprétation d'une partition quantitative et qualitative. Avec la même technique, plusieurs règles d'interprétation des résultats pour les différentes méthodes pourront être développées.

Base sur les résultats récents, par exemple cf. /BOCCEL/, nous avons obtenus les règles suivantes : Interprétation globale de la partition, interprétation du rôle d'une variable dans la partition obtenue, interprétation des classes de la partition, ...

Exemple :

REGLE 119

Si INTERPRETATION DE LA PARTITION
 et INTERPRETATION GLOBALE DE LA PARTITION
 et INERTIE TOTALE EST FAIBLE
 et INERTIE INTER-CLASSES EST FORTE
 Alors QUALITE DE LA PARTITION N'EST PAS BONNE
 et ESSAYEZ DE REEXECUTER COMMANDE MNDQAN EN CHANGEANT
 et SOIT LE NOMBRE DE CLASSES SOIT LA DISTANCE CHOISIE

REGLE 122

Si INTERPRETATION DE LA PARTITION
 et INTERPRETATION DU ROLE DE LA VARIABLE DANS PARTITION
 et VARIABLE A INTERPRETER > 0
 Alors %_INERTIE DE LA VARIABLE CONSERVE PAR LA PARTITION
 <-- %_DISCRIMINANT(VARIABLE A INTERPRETER)
 et CONTRIBUTION DE LA VARIABLE A L'INERTIE EXTRAITE PAR
 PARTITION <-- CONTRIBUTION(VARIABLE A INTERPRETER)

REGLE 123

Si %_DISCRIMINANT(VARIABLE A INTERPRETER) > %_INERTIE
 EXPLIQUEE
 Alors CETTE VARIABLE EST PLUS DISCRIMINANTE QUE LA MOYENNE

REGLE 124

Si %_DISCRIMINANT(VARIABLE A INTERPRETER) < %_INERTIE
 EXPLIQUEE
 Alors CETTE VARIABLE EST MOINS DISCRIMINANTE QUE LA MOYENNE

REGLE 125

Si %_DISCRIMINANT(VARIABLE A INTERPRETER) = %_INERTIE
 EXPLIQUEE
 Alors CETTE VARIABLE EST AUSSI DISCRIMINANTE QUE LA MOYENNE

REGLE 126

Si %_INERTIE DE LA VARIABLE CONSERVE PAR PARTITION EST FORTE
 et CONTRIBUTION DE LA VARIABLE A L'INERTIE EXTRAITE EST FAIBLE
 Alors C'EST UN CAS EXCEPTIONNEL
 et CETTE VARIABLE EST TRES DISCRIMINANTE
 et MAIS LA CONTRIBUTION A L'INERTIE DE LA PARTITION EST FAIBLE

REGLE 130

Si %_DISCRIMINANT(VARIABLE A INTERPRETER) =
 MAX_DISCRIMINANT(VARIABLE A INTERPRETER)
 Alors CETTE VARIABLE EST LA PLUS DISCRIMINANTE

REGLE 131

Si $\%_DISCRIMINANT(VARIABLE \text{ A INTERPRETER}) =$
 $MIN_DISCRIMINANT(VARIABLE \text{ A INTERPRETER})$
 Alors CETTE VARIABLE EST LA MOINS DISCRIMINANTE

REGLE 132

Si INTERPRETATION DE PARTITION
 et INTERPRETATION DES CLASSES
 et CLASSE A INTERPRETER > 0
 Alors LA VARIABLE LA PLUS SIGNIFICATIVE DE CLASSE
 $\leftarrow MAX1_DISCRIMINANT(CLASSE \text{ A INTERPRETER})$
 et $\%_POURVOIR \text{ DISCRIMINANT DE CELLE-LA POUR LA CLASSE}$
 $\leftarrow DISCRIMINANT_MAX1(CLASSE \text{ A INTERPRETER})$
 et CONTRIBUTION DE CELLE-LA A L'INERTIE INTERCLASSE
 $\leftarrow CONTRIBUTION_MAX1(CLASSE \text{ A INTERPRETER})$

REGLE 133

Si INTERPRETATION DE PARTITION
 et INTERPRETATION DES CLASSES
 et CLASSE A INTERPRETER > 0
 Alors LA DEUXIEME VARIABLE SIGNIFICATIVE DE CLASSE
 $\leftarrow MAX2_DISCRIMINANT(CLASSE \text{ A INTERPRETER})$
 et $\%_POURVOIR \text{ DISCRIMINANT DE CELLE-LA POUR LA CLASSE}$
 $\leftarrow DISCRIMINANT_MAX2(CLASSE \text{ A INTERPRETER})$
 et CONTRIBUTION DE CELLE-LA A L'INERTIE INTERCLASSE
 $\leftarrow CONTRIBUTION_MAX2(CLASSE \text{ A INTERPRETER})$

REGLE 134

Si INTERPRETATION DE PARTITION
 et INTERPRETATION DES CLASSES
 et CLASSE A INTERPRETER > 0
 Alors LA VARIABLE LA MOINS SIGNIFICATIVE DE CLASSE
 $\leftarrow MIN1_DISCRIMINANT(CLASSE \text{ A INTERPRETER})$
 et $\%_POUVOIR \text{ DISCRIMINANT DE CELLE-LA POUR LA CLASSE}$
 $\leftarrow DISCRIMINANT_MAX2(CLASSE \text{ A INTERPRETER})$
 et CONTRIBUTION DE CELLE-LA A L'INERTIE INTERCLASSE
 $\leftarrow CONTRIBUTION_MIN2(CLASSE \text{ A INTERPRETER})$

REGLE 135

Si INTERPRETATION DE PARTITION
 et INTERPRETATION DES CLASSES
 et CLASSE A INTERPRETER ≥ 0
 Alors LA DEUXIEME VARIABLE MOINS SIGNIFICATIVE DE CLASSE
 $\leftarrow \text{MIN2_DISCRIMINANT}(\text{CLASSE A INTERPRETER})$
 et POUVOIR DISCRIMINANT DE CELLE-LA A L'INERTIE INTERCLASSE
 $\leftarrow \text{CONTRIBUTION_MIN2}(\text{CLASSE A INTERPRETER})$

Dans la première version de CLAVETO, une combinaison par multifenetrage de technique graphique et système expert est réalisée qui fournit l'utilisateur à la fois une vue visible et synthétique sur les données et des explications sur les indicateurs statistiques obtenus.

IV.2.4. Règles concernant les stratégies

La pratique de l'analyse de données conduit utilisateurs et chercheurs à voir des stratégies dans l'étude d'un problème donné. Elles sont en général limitées ou difficiles à mettre en oeuvre pour les bibliothèques usuelles d'analyse de données car la notion de structure de données n'existe pas en général pour les programmes et les enchainements nécessitent divers travaux informatiques pénibles. Par contre, le système SICLA permet des enchainements faciles de méthodes, de reprises de résultats, donc il est possible d'envigager des stratégies complexes, fonctions des données et des objectifs poursuivis. Un certain nombre de stratégies sont donc proposées à l'utilisateur par le système expert CLAVETO. Elles sont formulées sous forme de règles guidant l'utilisateur à chaque étape.

Exemple

REGLE 98

Si TYPE DE TABLEAU DE DONNEES == MESURES HOMOGENES
 et TYPE DE TRAITEMENT == STRATEGIE
 Alors DEMARCHE PROPOSEE :
 et DESCRIPTION UTILISER LES COMMANDES : DESCAL,CORREL,HISTO
 et ANALYSE UTILISER LES COMMANDES: MNDQAN,WARD,ACOMP /SPAD/

REGLE 100

Si TYPE DE TABLEAU DE DONNEES == MESURES HETEROGENES
 et TYPE DE TRAITEMENT == STRATEGIE
 Alors DEMARCHE PROPOSEE:
 et DESCRIPTION UTILISER LES COMMANDES: DESQAN,CORREL,HISTO
 et HOMOGENEISATION DU TABLEAU : CENTRAGE REDUCTION
 et UTILISER: METRIQ
 et POUR L'ANALYSE UTILISER : MNDQAN,WARD,ACOMP
 et UTILISER INPAQN POUR INTERPRETER DE RESULTAT DE MNDQAN
 et SOIT CREER 1 TABLEAU MODALITES PAR CODAGE: CODQ
 et ANALYSER DE NOUVEAU TABLEAU
 et POSSIBILITE AINSI MISE EN EVIDENCE

REGLE 102

Si TYPE DE TABLEAU DE DONNEES == BINAIRE
 et TYPE DE TRAITEMENT == STRATEGIE
 Alors DEMARCHE PROPOSEE:
 et POUR L'ANALYSE UTILISER : CROBIN,ACOMP /SPAD/

REGLE 103

Si TYPE DE TABLEAU DE DONNEES == CONTINGENCE
 et TYPE DE TRAITEMENT == STRATEGIE
 Alors DEMARCHE PROPOSEE:
 et POUR L'ANALYSE UTILISER LES COMMANDES:
 et ANALYSE DES CORRESPONDANCES CORBI /SPAD/
 et CLASSIFICATION: CROKI2

REGLE 106

Si TYPE DE TABLEAU DE DONNEES == MODALITES
 et TYPE DE TRAITEMENT == STRATEGIE
 Alors DEMARCHE PROPOSEE:
 et ETAPE 1:
 et CHOISIR LES ENSEMBLES DE VARIABLES ACTIVES & ILLUSTRATIVES
 et FAIRE UNE DESCRIPTION ELEMENTAIRE PAR : DESQAL,AKHI2
 et ELIMINER OU DECODER LES VARIABLES COMPORTANT LES
 MODALITES
 et FAIBLES EFFECTIFS PAR LES COMMANDES : ELIMOD,CREQAL

...

REGLE 117

Si TYPE DE TABLEAU DE DONNEES == MODALITES

et TYPE DE TRAITEMENT == STRATEGIE

et ETAPE 4 OK

Alors DEMARCHE PROPOSEE :

et ETAPE 5

et CREER UN TABLEAU CONTINGENCE MULTIPLE PAR CONTMU

et CROISANT LA PARTITION & LE VARIABLES ILLUSTRATIVES

et ANALYSE DE TABLEAU DE DONNEES PAR CORBI,WARD

IV.5. Traitement des erreurs

Lors d'une session SICLA, des erreurs peuvent se produire dues à des fausses manipulations de l'utilisateur. Un message alors est imprimé par le système mais l'action à entreprendre n'est pas toujours évidente. Un ensemble de règles explicitent les messages et indique les remèdes possibles.

Exemple :

REGLE 92

Si TYPE DE TRAITEMENT == ERREUR SICLA

et TYPE DE L'ERREUR == ERREUR LORS DE L'ENREGISTREMENT DE
DONNEES

et MESSAGE L'ERREUR RELATIF DONNEES HORS BORNES

et MAUVAISE DEFINITION DU DESCRIPTIF DES VARIABLES

Alors SOIT ERREUR SUR LA NATURE DES VARIABLES

et SOIT ERREUR SUR LE NOMBRE DE MODALITES

et COMMANDE := GENFIC

et CORRIGER LE DICTIONNAIRE A L'EDITEUR DE TEXTE

et RECOMMENCER L'ENREGISTREMENT PAR COMMANDE ENRFIC

V. INTERFACE CLAVETO

Quelques propriétés principales de l'interface CLAVETO sont les suivantes :

- . **Génération initiale la base de faits** concernant l'environnement de SICLA par la transmission de l'information sur les données enregistrées de SICLA à COTO.

- . **Acquisition des faits** par conversation. L'utilisateur dialogue avec le système essentiellement sur l'objectif du problème.

- . **CLAVETO est incorporé dans SICLA.** Permettant l'utilisation de sous-routines externes, le moteur d'inférence COTO est connecté directement à SICLA. Cela permet à l'utilisateur de consulter sur son problème et d'exécuter les commandes indiquées sans quitter le système expert.

VI. CONCLUSION ET PERSPECTIVES

Le but de ce rapport est de présenter un premier travail concernant une application de l'intelligence artificielle à l'interprétation des résultats d'une analyse de données.

La base de connaissances présentée ici va être progressivement enrichie par l'expertise dont nous disposons au sein de l'équipe "Classification Automatique et Reconnaissances des Formes".

Un intérêt de ce travail, qui était déjà apparu dans CLAVECIN, est aussi de permettre une certaine diffusion de l'expérience acquise au sein de l'équipe lors des divers problèmes réels d'analyse de données qui nous ont été soumis.

VII. EXEMPLE D'UTILISATION DE CLAVETO

SICLA

SICLA (Systeme Interatif de Classification Automatique) permet la GESTION, la DESCRIPTION ELEMENTAIRE et le TRAITEMENT des donnees par les methodes d'analyse multidimensionnelle. Il permet egalement d'accéder aux methodes factorielles de SPAD et MODULAD.

SICLA est interface avec un SYSTEME EXPERT (COTO) qui a pour but d'aider les utilisateurs de SICLA comme une assistance intelligente. COTO souhaite que vous soyez content de ses services sur SICLA !

Voulez-vous relire la base de regles ?

? oui

Je lis actuellement la base de connaissances.

Je serai a vous dans quelques instants

Savez vous comment dialoguer avec le systeme ?

Si non regardez les conventions du systeme avant de l'utiliser

Repondez "oui" ou "non", S. V. P. !

? oui

En attendant la lecture de la base de connaissances,
jetez un coup d'oeil sur quelques conventions du systeme.
Pour repondre aux questions du systeme , vous tapez :

"o" ou "oui" pour l'affirmation

"n" ou "non" pour la negation

"!" comme JE NE SAIS PAS

"un chiffre" pour une valeur numerique

"un ordre " pour une valeur alphanumerique

Vous pouvez aussi demander le systeme n'importe quand :

"?" pourquoi ou pourquoi pas

"h" ou "help" demande des informations utiles

"t" ou "trace" affichage de la base de faits

"i" ou "insert" insertion d'un nouveau fait

"d" ou "delete" annulation d'un fait affirmé

"s" ou "stop" pour arreter le raisonnement
 J'ai lu 128 regles dans la base de connaissance
 J'ai donc enregistre 377 mots dans mon dictionnaire
 parmi ceux-ci 16 fonctions numeriques
 et 225 buts demandables

```
*****
* SICLA                            Ho                            Date : 17/06/86    *
* S.D : vote.sdo                   Archive : vote.sar                            *
* Listing : lis16_44                Mode interactif                            *
*****
* SICLA                            Commande INITIA                            Version 15/04/86    *
*****
* donnees : vote.sdo    individus : 96 * variables : 6                            *
* -----                            -----                            *
*                            quantitatives : 6    qualitatives : 0                            *
*****
=== Fin de INITIA ===
```

Veuillez entrer des faits ? (\$ pour FIN)

TABLEAU DES REELS

Vous me dites : TABLEAU DES REELS

Je ne reconnais pas ce fait.

Pourriez vous me donner un mot maitre ou un segment du fait ?

(mot maitre ... sinon "\$")

? TABLEAU

Celui que vous voulez affirmer est-il l'un des faits suivants ?

1. TYPE DE TABLEAU DE DONNEES
2. CREATION DE TABLEAU DE CONTINGENCE
3. CREATION DE TABLEAU DE DISTANCES
4. TYPE DE TABLEAU DE DONNEES HETEROGENE

Quel fait vou voulez entrer ?

Tapez son ordre ou "n" si vous ne le trouvez pas.

? 1

Vous voulez parler de : TYPE DE TABLEAU DE DONNEES

OK ! Elle peut prendre une valeur parmi les suivantes :

1. MESURES HETEROGENES
2. CONTINGENCE
3. BINAIRE
4. PREFERENCE
5. MODALITES
6. MESURES HOMOGENES
7. HETEROGENE
8. BURT
9. DISTANCES

Laquelle vous connaissez ? (tapez un ordre ou "n" si ca va pas)

? 1

OK ! je l'enregistre dans la base de faits. Le fait numero 4

Continuez de decrire des faits, S.V.P. ! (\$ pour FIN)

? \$

Avez vous un but a me soumettre ? (nom du but ou "non")

? n

Connaissez vous : TYPE DE TRAITEMENT

Est ce ?

1. DESCRIPTION DE DONNEES
2. GESTION DE DONNEES
3. EDITION DE DONNEES
4. CLASSIFICATION
5. FACTORIELLE
6. DISCRIMINATION
7. CODAGE
8. ERREUR SICLA
9. EVALUATION DES DONNEES
10. STRATEGIE

? 4

Vous me dites : TYPE DE TRAITEMENT == CLASSIFICATION

OK !

A partir de la structure de donnees enregistree :

Fait 4 : TYPE DE TABLEAU DE DONNEES == MESURES HETEROGENES

Au moyen de la REGLE 1

J'en arrive a la conclusion suivante :

1. NATURE DE VARIABLES := QUANTITATIVES

Pensez vous que ANALYSE DE DONNEES MULTIDIMENSIONNELLES ?

??

J'essaie de declencher toutes regles ouvertes concernees a :

ANALYSE DE DONNEES MULTIDIMENSIONNELLES

par exemple :

REGLE 46

SI ANALYSE DE DONNEES MULTIDIMENSIONNELLES

TYPE DE TABLEAU DE DONNEES == MESURES HETEROGENES (affirme)

ALORS CAS DE FIGURE <-- 2.0

Pensez vous que ANALYSE DE DONNEES MULTIDIMENSIONNELLES ?

?o

Avec les faits affirmes et deduits suivants :

Fait 7 : ANALYSE DE DONNEES MULTIDIMENSIONNELLES

Fait 4 : TYPE DE TABLEAU DE DONNEES == MESURES HETEROGENES

Au moyen de la REGLE 46

J'en arrive a la conclusion suivante :

2. CAS DE FIGURE <-- 2.0

Connaissez vous : OBJECTIF DE CLASSIFICATION

Est ce ?

1. LES INDIVIDUS

2. LES VARIABLES

3. LES MODALITES

?1

Vous me dites : OBJECTIF DE CLASSIFICATION == LES INDIVIDUS

OK !

Connaissez vous : TYPE DE REPRESENTATION

Est ce ?

1. PARTITION

2. HIERARCHIE

? 1

Vous me dites : TYPE DE REPRESENTATION == PARTITION

OK ! Avec les faits affirmes et deduits suivants :

Fait 5 : TYPE DE TRAITEMENT == CLASSIFICATION

Fait 8 : CAS DE FIGURE = 2.0

Fait 9 : OBJECTIF DE CLASSIFICATION == LES INDIVIDUS

Fait 10 : TYPE DE REPRESENTATION == PARTITION

Au moyen de la REGLE 55

J'en arrive a la conclusion suivante :

3. COMMANDE := MNDQAN

4. Methode des Nuees Dynamiques sur variables quantitatives

5. LE CENTRAGE REDUCTION EST FORTEMENT CONSEILLE

6. UTILISATION DE LA COMMANDE : METRIQ

7. Choix d'une metrique pour les variables quantitatives

8. UTILISER INPAQN POUR INTERPRETER LE RESULTAT

Voulez-vous executer une commande du SICLA ?

? oui

Tapez le nom de la commande , S.V.P. !

?MNDQAN

creation du fichier profil : sicla.profil

* SICLA Ho Date : 17/06/86 *

* S.D : Archive : *

* Listing : lis16_52 Mode interactif *

* version 1.2 : (faire help sicla) *

* SICLA Commande MNDQAN Version 1/10/85 *

entrez le nom de la structure de donnees choisie !

vote

-----> fichier : vote.sdo

```
*****
* donnees : vote.sdo      individus : 96      variables : 6      *
* -----                -----                -----      *
*                          quantitatives : 6  qualitatives : 0      *
*****
```

selection de variables ?

la\$

selection d individus ?

la\$

variables selectionnees :

pc ps udrp ere vert fn

L'ensemble des donnees a ete selectionne

nombre d essais ?

1

code conditions initiales ?

?

conditions initiales

1 = tirage au hasard des noyaux

2 = entree sur la console des noyaux de depart

(les noyaux peuvent etre de tailles differentes)

3 = la partition de depart est choisie au hasard

4 = la partition de depart est lue sur l archive

1

nombre de classes ?

4

code impression partition ?

1

essai numero : 1

convergence atteinte en 5 iterations

pourcentage d inertie expliquee : 57.48

entrez l abrege suivi du libelle a donner a la meilleure partition

vote

=== Fin de MNDQAN ===

STOP fin normale

Une autre execution du SICLA ?

? oui

Tapez le nom de la commande , S.V.P. !

INPAQN

* SICLA Ho Date : 17/06/86 *

* S.D : vote.sdo Archive : vote.sar *

```
* Listing : lis16 52      Mode interactif      *
```

* SICLA Commande INPAON Version 1/10/85 *

```
* donnees : vote.sdo      individus : 96 *      variables : 6      *
```

* quantitatives : 6 qualitatifs : 0 *

selection de variables ? 1a\$

selection d individus ? 1a\$

variables selectionnees :

pc ps udrp ere vert fn

L'ensemble des donnees a ete selectionne

```
* archive:  vote.sar                                     *
```

* nombre de constituants : 2 *

```
* rang * type * nom          * informations          *
```

* 1 * part * mm * 6 clas *

* 2 * part * vote * 4 clas *

entrez le rang de la partition ? Taper le numero du constituant a choisir entre (1 et 2)

2

=== Fin de INPAQN ===

STOP fin normale

Une autre execution du SICLA ?

? n

Pensez vous que INTERPRETATION GLOBALE DE LA PARTITION ?

? n

Pensez vous que

INTERPRETATION DU ROLE DES VARIABLES DANS LA PARTITION ?

? oui

Pouvez vous me donner la valeur de : VARIABLE A INTERPRETER ?

? 1

Avec les faits affirmes et deduits suivants :

Fait 16 : UTILISER INPAQN POUR INTERPRETER LE RESULTAT

Fait 18 : INTERPRETATION DU ROLE DES VARIABLES DANS LA PARTITION

Fait 19 : VARIABLE A INTERPRETER = 1.0 > 0.0

Au moyen de la REGLE 116

J'en arrive a la conclusion suivante :

9. %_INERTIE DE LA VARIABLE CONSERVE PAR LA PARTITION

<-- %_DISCRIMINANT(VARIABLE A INTERPRETER) = 68.2

10. CONTRIBUTION DE LA VARIABLE A L'INERTIE EXTRAITE PAR PARTITION

<-- CONTRIBUTION(VARIABLE A INTERPRETER) = 30.7

Avec les faits affirmes et deduits suivants :

Fait 16 : UTILISER INPAQN POUR INTERPRETER LE RESULTAT

Fait 18 : INTERPRETATION DU ROLE DES VARIABLES DANS LA PARTITION

Fait 20 : %_DISCRIMINANT(VARIABLE A INTERPRETER) = 68.2

Fait 24 : %_INERTIE EXPLIQUEE = 57.48

Au moyen de la REGLE 117

J'en arrive a la conclusion suivante :

11. CETTE VARIABLE EST PLUS DISCRIMINANTE QUE LA MOYENNE

Avec les faits affirmes et deduits suivants :

Fait 16 : UTILISER INPAQN POUR INTERPRETER LE RESULTAT

Fait 18 : INTERPRETATION DU ROLE DES VARIABLES DANS LA PARTITION

Fait 20 : %_DISCRIMINANT(VARIABLE A INTERPRETER) = 68.2

Fait 26 : MAX_DISCRIMINANT(VARIABLE A INTERPRETER) = 30.7

Au moyen de la REGLE 122, J'en arrive a la conclusion suivante :

12. CETTE VARIABLE EST LA PLUS DISCRIMINANTE

Pensez vous que INTERPRETER UNE AUTRE VARIABLE ?

? oui

Pouvez vous me donner la valeur de : VARIABLE A INTERPRETER ?

? 4

Avec les faits affirmes et deduits suivants :

Fait 16 : UTILISER INPAQN POUR INTERPRETER LE RESULTAT

Fait 18 : INTERPRETATION DU ROLE DES VARIABLES DANS LA PARTITION

Fait 41 : VARIABLE A INTERPRETER = 4.0 > 0.0

Au moyen de la REGLE 116

J'en arrive a la conclusion suivante :

26. %_INERTIE DE LA VARIABLE CONSERVE PAR LA PARTITION

<-- %_DISCRIMINANT(VARIABLE A INTERPRETER) = 4.4

27. CONTRIBUTION DE LA VARIABLE A L'INERTIE EXTRAITE PAR PARTITION

<-- CONTRIBUTION(VARIABLE A INTERPRETER) = 0.2

Avec les faits affirmes et deduits suivants :

Fait 16 : UTILISER INPAQN POUR INTERPRETER LE RESULTAT

Fait 18 : INTERPRETATION DU ROLE DES VARIABLES DANS LA PARTITION

Fait 42 : %_DISCRIMINANT(VARIABLE A INTERPRETER) = 4.4

Fait 46 : %_INERTIE EXPLIQUEE = 57.48

Au moyen de la REGLE 118

J'en arrive a la conclusion suivante :

28. CETTE VARIABLE EST MOINS DISCRIMINANTE QUE LA MOYENNE

Avec les faits affirmes et deduits suivants :

Fait 16 : UTILISER INPAQN POUR INTERPRETER LE RESULTAT

Fait 18 : INTERPRETATION DU ROLE DE VARIABLE DANS LA PARTITION

Fait 42 : %_DISCRIMINANT(VARIABLE A INTERPRETER) = 4.4

Fait 49 : MIN_DISCRIMINANT(VARIABLE A INTERPRETER) = 0.2

Au moyen de la REGLE 123

J'en arrive a la conclusion suivante :

29. CETTE VARIABLE EST LA MOINS DISCRIMINANTE

Pensez vous que INTERPRETER UNE AUTRE VARIABLE ?

? non

Pensez vous que INTERPRETATION DES CLASSES ?

? oui

Pouvez vous me donner la valeur de : CLASSE A INTERPRETER ?

? 3

Avec les faits affirmes et deduits suivants :

Fait 16 : UTILISER INPAQN POUR INTERPRETER LE RESULTAT

Fait 19 : INTERPRETATION DES CLASSES

Fait 20 : CLASSE A INTERPRETER = 3.0 > 0.0

Au moyen de la REGLE 125

J'en arrive a la conclusion suivante :

14. LA VARIABLE LA PLUS SIGNIFICATIVE DE CLASSE

<-- MAX1_DISCRIMINANT(CLASSE A INTERPRETER) = 1.0

15. %_POUVOIR DISCRIMINANT DE CELLE-LA POUR CLASSE

<-- DISCRIMINANT_MAX1(CLASSE A INTERPRETER) = 32.8

16. CONTRIBUTION DE CELLE-LA A INERTIE INTERCLASEE

<-- CONTRIBUTION_MAX1(CLASSE A INTERPRETER) = 50.2

Avec les faits affirmes et deduits suivants :

Fait 16 : UTILISER INPAQN POUR INTERPRETER LE RESULTAT

Fait 19 : INTERPRETATION DES CLASSES

Fait 20 : CLASSE A INTERPRETER = 3.0 > 0.0

Au moyen de la REGLE 126

J'en arrive a la conclusion suivante :

17. LA DEUXIEME VARIABLE SIGNIFICATIVE DE CLASSE

<-- MAX2_DISCRIMINANT(CLASSE A INTERPRETER) = 4.0

18. %_POUVOIR DISCRIMINANT DE CELLE-LA POUR CLASSE

<-- DISCRIMINANT_MAX2(CLASSE A INTERPRETER) = 26.7

19. CONTRIBUTION DE CELLE-LA A L'INERTIE INTERCLASSE

<-- CONTRIBUTION_MAX2(CLASSE A INTERPRETER) = 48.3

Avec les faits affirmes et deduits suivants :

Fait 16 : UTILISER INPAQN POUR INTERPRETER LE RESULTAT

Fait 19 : INTERPRETATION DES CLASSES

Fait 20 : CLASSE A INTERPRETER = 3.0 > 0.0

Au moyen de la REGLE 127

J'en arrive a la conclusion suivante :

20. LA VARIABLE LA MOINS SIGNIFICATIVE DE CLASSE

<-- MIN1_DISCRIMINANT(CLASSE A INTERPRETER) = 5.0

21. % POUVOIR DISCRIMINANT DE CELLE-LA POUR CLASSE

<-- DISCRIMINANT_MIN1(CLASSE A INTERPRETER) = 0.2

22. CONTRIBUTION DE CELLE-LA A L'INERTIE INTERCLASSE

<-- CONTRIBUTION_MIN1(CLASSE A INTERPRETER) = 0.0

Avec les faits affirmes et deduits suivants :

Fait 16 : UTILISER INPAQN POUR INTERPRETER LE RESULTAT

Fait 19 : INTERPRETATION DES CLASSES

Fait 20 : CLASSE A INTERPRETER = 3.0 > 0.0

Au moyen de la REGLE 128

J'en arrive a la conclusion suivante :

23. LA DEUXIEME VARIABLE MOINS SIGNIFICATIVE DE CLASSE

<-- MIN2_DISCRIMINANT(CLASSE A INTERPRETER) = 4.0

24. % POUVOIR DISCRIMINANT DE CELLE-LA POUR CLASSE

<-- DISCRIMINANT_MIN2(CLASSE A INTERPRETER) = 2.6

25. CONTRIBUTION DE CELLE-LA A L'INERTIE INTERCLASSE

<-- CONTRIBUTION_MIN2(CLASSE A INTERPRETER) = 0.0

Pensez vous que INTERPRETER UNE AUTRE CLASSE ?

? non

Aucune inference ne peut etre encore effectuee !

Voulez vous reraisonner a partir d'un des faits affirmes ?

? non

GOOD BYE ! IT'S NICE TO SEE YOU AGAIN !

VIII. REFERENCES

- /BOCCEL/ BOCHI S., CELEUX G., " Quelques outils d'interprétation d'une partition", Rapport INRIA.
- /DEMQUI/ DEMONCHAUX E., QUINQUETON J., "OURCIN 2.1 : Manuel de reference", Rapport de Technique N°57, Septembre 1985, INRIA.
- /DEQURA/ DEMONCHAUX E., QUINQUETON J., RALAMBONDRAINY H., "CLAVECIN un système expert en analyse de données ", Rapports de recherche, N°431, Juillet 1985, INRIA.
- /GALE/ GALE W.A., "Knowledge representation in Data analysis", Quatrièmes journées internationals sur Analyse de données et Informatique, Versailles, Oct.1985.
- /HO/ HO T.B., "On the design and implementation of an expert system", Computer & Artificial Intelligence (à paraître fin 1986).
- /LAUREN/ LAURENT J.P., "La structure de controle dans les systèmes experts", Technique & Science Informatique, N°3, 1984.
- /LAURIE/ LAURIERE J.L., Intelligence Artificielle : Resolusion de problème par l'homme et la machine, Eyrolles, Fevrier 1986.
- /LERAGO/ LECHEVALIER Y.,RALAMBONDRAINY H., GOVAERT G., "Le système SICLA : Principes et Architecture, Rapport de recherche, N°500,Mars 1986 INRIA.
- /NILSON/ NILSON J., Principles of artificial intelligence, Tioga, 1980.
- /SIMON/ SIMON J.C.,La reconnaissance des formes par algorithmes, Masson 1984.
- /WEIKUL/ WEISS M.,KULIKOWSKI A.,Pratical guide of designing expert systems, Rowman & Allanheld Publishier,1984.

U

.

1

to

2

.

3

6

.

8