

Chapitre 12 : Guide d'utilisation des principales fonctionnalités du logiciel CHIC¹

Harrison Ratsimba-Rajohn

33 Chemin des Peyrères, 33610 CANEJAN, FRANCE

ratsimba@u-bordeaux3.fr

<http://mizara.ovh.org/recherche>

Résumé : Dans ce chapitre, nous avons choisi d'utiliser les données d'une recherche en cours, pour présenter et décrire les étapes et procédures à suivre dans un traitement par le logiciel CHIC dans le but, d'une part, d'obtenir des graphes implicatifs et des représentations des hiérarchies cohésitives et, d'autre part, d'en faciliter la lecture grâce aux diverses options et fonctionnalités de ce logiciel. Nous aborderons par étape la description des données recueillies, leur transformation, le choix des variables supplémentaires, la préparation du fichier-texte au format *.CSV requis par CHIC, les traitements pour obtenir un graphe implicatif en particulier avec la notion de conjonction de prémisses et de seuil d'originalité, le traitement pour obtenir une représentation de la hiérarchie cohésive, complété par les notions de niveaux significatifs, de contribution de groupes optimaux, de typicalité de variables supplémentaires ou d'individus. Dans un dernier nous aborderons par le traitement de variables sur intervalles.

1 Présentation

A partir d'un exemple de données, nous allons, dans cette partie, regarder les manières de manipuler CHIC afin d'obtenir des résultats d'analyses implicatives.

Les données collectées, sur lesquelles nous allons illustrer les manipulations sont issues d'une recherche² relative à l'utilisation des images en didactique de la biologie. Dans cette recherche on avait distingué deux types d'images : nous appellerons le premier "infographie" et le second "schéma". Brièvement, derrière "infographie" a été mise une représentation suffisamment proche de la "réalité" d'un processus cellulaire biologique et derrière "schéma" a été mise une représentation graphique suffisamment simplifiée et schématique de ce même processus. La¹ recherche qui a été menée consistait à étudier la manière dont des étudiants en

¹ Acronyme C.H.I.C pour désigner le logiciel : Classification Hiérarchique, Implicative, Cohésive

² Dahmani H.R., (2008-2009) *Recherche en cours dans le cadre de sa thèse*, ENS (Ecole Normale Supérieure), Laboratoire de Didactique de la Biologie, Kouba, 16000, Alger, Algérie), sous la direction de P. Schneeberger, Equipe du DAESL (Laboratoire de Didactique, d'Anthropologie, d'Epistémologie et des Sciences du Langage), Université Bordeaux II, 33000, Bordeaux, France (<http://pagesperso-orange.fr/daest/index.htm>), et en partenariat avec Dr I Kramer, chercheur à l'Institut Européen de Chimie et de Biologie de Bordeaux. (<http://www.cellbiol.net/home/index.php>)

SVT (Bac+2) "percevait" le même processus cellulaire selon qu'ils traitaient une infographie ou un schéma le représentant durant une évaluation écrite.

Nos propos n'entreront pas sur le fond de cette recherche en didactique de la biologie ni dans sa forme mais se concentreront sur la manière d'utiliser le logiciel CHIC. C'est pourquoi, d'une part, nous ne procéderons pas comme habituellement dans une phase normale d'une recherche où, avant de se lancer dans une analyse des données, on pose d'abord les différentes questions et hypothèses qui nous aident à traiter une problématique. D'autre part nous ne procéderons pas aussi aux interprétations qui en découlent. .

Les données recueillies sont les résultats de l'évaluation écrite soumise à 82 étudiants le 28 mars 2008. Les 82 étudiants reçoivent les mêmes questions mais 42 ont comme support les images sous forme d'infographie et 40 les ont sous forme de schéma. La répartition des 2 types d'image a été faite aléatoirement.

Sur chaque étudiant ont été observés 7 caractères :

1. L'étudiant a passé l'évaluation sous la forme schéma: OUI/NON (variable valorisée 1 ou 0)
2. L'étudiant a passé l'évaluation sous la forme infographie: OUI/NON (variable valorisée 1 ou 0)
3. La note moyenne obtenue par l'étudiant lors de 3 devoirs précédents. (note /20)
4. La note obtenue par l'étudiant lors de cette évaluation écrite (note /20)
5. La note obtenue par l'étudiant sur une question demandant l'intitulé d'une image (note /20)
6. La note obtenue par l'étudiant sur une question demandant de légènder l' image (note /20)
7. La note obtenue par l'étudiant sur une question demandant de décrire le processus représenté par l'image (note /20)

Nous avons désigné respectivement ces sept caractères par les codes de variables suivants :

Sch inf moy3 not3 titr LEG DES

Les données que nous allons traiter se présentent sous forme de tableau de 83 lignes et 9 colonnes.

Ainsi nous procéderons selon les étapes suivantes :

- Description des données recueillies et leur préparation;
- Mise en forme de ces données sur le tableur EXCEL, afin d'obtenir des données de type .CSV acceptées par le logiciel CHIC;
- Deux traitements avec 5 variables principales réelles et 2 variables supplémentaires binaires;
- Deux traitements avec des "variables sur intervalles" sans variables supplémentaires.

2 Description des données recueillies

Les données que nous allons traiter se présentent sous forme de tableau de 83 lignes et 9 colonnes. :

1. La première ligne contient les codes des variables et les autres lignes représentant les 82 individus qui ici sont les étudiants
2. La première colonne contient les codes des individus (ici ce sont des numéros de 1 à 86 car quelques numéros sont absents) puis les autres 7 colonnes contiennent les valeurs prises par les variables.

Ainsi dans le tableau ci-dessous l'individu 11 est absent, l'individu 12 a traité les questions avec les images sous forme de schéma, a obtenu 13/20 comme moyenne de 3 devoirs précédents, 16 comme note à l'évaluation; 0 comme note sur l'intitulé de l'image, 19 sur la légende et 18 sur la description ; par contre l'individu 8 a traité les questions ayant comme support l'infographie, il a obtenu 15/20 à la moyenne des 3 devoirs précédents, 17 à l'évaluation; 20 sur l'intitulé des images, 13 sur les légendes et 20 sur la description.

Notons que la première ligne de la première colonne est laissée vide.

	Sch	inf	moy3	Not3	titr	LEG	DES
1	1	0	19	19	10	20	18
2	1	0	12	19	10	20	20
5	1	0	13	17	10	20	16
6	1	0	14	17	20	19	20
7	0	1	16	17	20	16	18
8	0	1	15	17	20	13	20
9	1	0	10	17	20	19	18
10	1	0	18	17	10	19	20
12	1	0	13	16	0	19	18
						
84	0	1	5	2	0	2	3
86	1	0	9	2	10	3	2

TAB 1 – Données brutes : source : http://mizara.ovh.org/recherche/chic/donnees_brutes.xls

3 Transformation de données

Les données que nous allons traiter sont mises sur une feuille d'un classeur EXCEL. Mais CHIC ne peut pas traiter des données dont des valeurs de variables sont plus grandes que 1 (ni plus petites que 0), si ce ne sont pas des variables sur intervalles. Avec l'aide du tableur nous transformons ces valeurs comprises entre 0 et 20 en des valeurs proportionnelles comprises entre 0 et 1 pour obtenir des variables fréquentielles (Ici, il suffit de diviser par 20 chaque valeur) et nous obtenons le tableau :

	Sch	inf	moy3	not3	Titr	LEG	DES
1	1	0	0,95	0,95	0,50	1,00	0,90
2	1	0	0,60	0,95	0,50	1,00	1,00
5	1	0	0,65	0,85	0,50	1,00	0,80
6	1	0	0,70	0,85	1,00	0,95	1,00
7	0	1	0,80	0,85	1,00	0,80	0,90
8	0	1	0,75	0,85	1,00	0,65	1,00
9	1	0	0,50	0,85	1,00	0,95	0,90
10	1	0	0,90	0,85	0,50	0,95	1,00
12	1	0	0,65	0,80	0,00	0,95	0,90
			...				
84	0	1	0,25	0,10	0,00	0,10	0,15
86	1	0	0,45	0,10	0,50	0,15	0,10

TAB. 2 –: *Données en variables fréquentielles*

Selon la problématique étudiée on aurait pu choisir de diviser les valeurs d'une colonne par le maximum de la colonne. Ici nous avons fait le choix de diviser par la valeur maximale théorique. On aurait pu aussi obtenir pour une variable des valeurs binaires en demandant à Excel de mettre 0 si inférieur à tel seuil, sinon 1.

4 Choix des variables qui seront supplémentaires

En principe comme les variables à étudier sont les variables principales pour être prises en compte normalement dans CHIC pour les calculs des indices d'implication ou de similarité, nous pouvons alors considérer les deux premières variables Sch et inf comme pouvant être des variables supplémentaires. En effet les variables supplémentaires sont en général des descripteurs et sont ici des variables binaires (ou modales) et ils n'interviennent que dans le calcul des contributions des catégories. Par exemple, si on souhaite savoir si une implication est plutôt formée par des étudiants qui ont traité les schémas ou par ceux qui ont eu à traiter les infographies, on définit pour chaque individu le fait qu'il avait traité le schéma ou l'infographie. Ensuite, les variables Sch et inf ne sont prises en compte que lors de la recherche de la contribution ou la typicalité des catégories.. Pour indiquer à CHIC une variable supplémentaire, on ajoute au nom de la variable un espace et un "s". Ainsi les codes respectifs de ces deux variables supplémentaires sont "Sch s" et ".inf s". Nous obtenons alors le tableau suivant :

	Sch s	Inf s	moy3	not3	titr	LEG	DES
1	1	0	0,95	0,95	0,50	1,00	0,90
2	1	0	0,60	0,95	0,50	1,00	1,00
3	1	0	0,80	0,90	1,00	0,70	1,00
4	1	0	0,80	0,85	0,50	0,95	0,90
5	1	0	0,65	0,85	0,50	1,00	0,80
6	1	0	0,70	0,85	1,00	0,95	1,00
7	0	1	0,80	0,85	1,00	0,80	0,90
8	0	1	0,75	0,85	1,00	0,65	1,00
...

TAB. 3 – Données avec variables supplémentaires

5 Préparation du fichier-texte au format *.CSV requis par le logiciel CHIC

CHIC n'acceptant pas les données sous forme de tableau comme ci-dessus, les données doivent être sous la forme suivante (Deux valeurs adjacentes sont séparées par un "point-virgule") :

;	Sch s;	inf s;	moy3;	not3;	titr;	LEG;	DES
1;	1;	0;	0,95;	0,95;	0,5;	1;	0,9
2;	1;	0;	0,6;	0,95;	0,5;	1;	1
3;	1;	0;	0,8;	0,9;	1;	0,7;	1
4;	1;	0;	0,8;	0,85;	0,5;	0,95;	0,9
5;	1;	0;	0,65;	0,85;	0,5;	1;	0,8
6;	1;	0;	0,7;	0,85;	1;	0,95;	1
7;	0;	1;	0,8;	0,85;	1;	0,8;	0,9
8;	0;	1;	0,75;	0,85;	1;	0,65;	1
...							

TAB. 4 – Données sous format CSV

Ils seront contenus dans un fichier-texte de type (ou format) .CSV. On peut écrire par exemple ces données avec un traitement de texte simple (comme le bloc-note de Windows) en le sauvegardant sur un fichier-texte mais avec l'extension .CSV.

Toutefois, Il est bon de savoir que le tableur EXCEL peut enregistrer une feuille de calcul sous ce format .CSV. Il suffit de choisir dans le menu *Fichier* la rubrique *Enregistrer sous...*, de donner un nom sans tiret-bas, ni espace et enfin de choisir comme type de fichier : CSV (séparateur point-virgule) (*.csv). Signalons qu'avec le tableur Calc du logiciel libre Open Office la production de ce type de fichier .CSV est aussi faisable avec quelques précautions supplémentaires. A partir du traitement de texte WORD ou de celui WRITER d'Open Office, on peut également obtenir des valeurs séparées par des virgules en utilisant judicieusement la rubrique "convertir un tableau en texte" du Menu Tableau. Quelle que soit la manière d'obtenir le fichier-texte de type .CSV, nous l'appellerons "donnees01s.csv" en vue du

traitement par le logiciel CHIC. Il est essentiel, pour raison technique, que ce nom ne comporte ni espace, ni tiret-bas.

6 Traitement pour obtenir un graphe implicatif

Pour simplifier, mettons d'abord le fichier "donnees01s.csv" dans le même dossier que l'application Chic.exe.

6.1 Exécutons Chic.exe

A ce stade, il convient de lancer l'exécution du programme CHIC.exe.

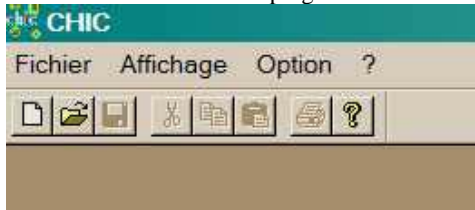


FIG. 1 – Écran d'accueil du logiciel CHIC

6.2 Choix des options

Choisissons les Options. Préparons les paramètres dans le menu Options. Dans un premier temps, pour une première approche avec un graphe implicatif, nous choisissons le Type d'implication selon la théorie classique avec le type de loi qui sera la loi binomiale. Historiquement ce sont les bases initiales des calculs. Par ailleurs, pour simplifier, nous garderons les couleurs des seuils proposés par défaut, puis dans les Options d'Apriori nous choisissons, pour commencer, 1 prémisses. Ceci signifie que, dans une autre analyse ultérieure, on peut paramétrer CHIC, à travers cette Option d'Apriori, pour représenter une implication dont la prémisse est formée par un ensemble de deux ou trois variables qui impliquerait d'autres variables. Dans un second temps, avec des éclaircissements théoriques et en fonction de la problématique traitée, on pourra choisir d'autres combinaisons d'options et de paramètres.

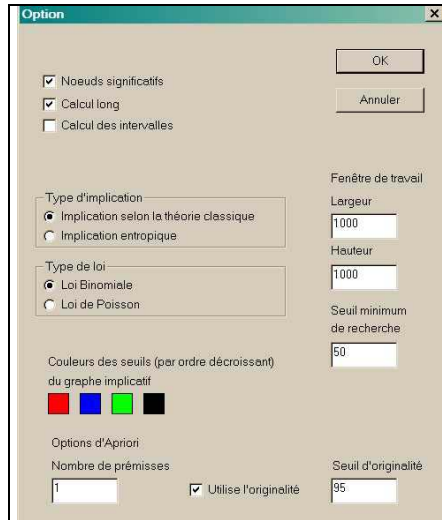


FIG. 2 – Menu options du logiciel CHIC

6.3 Traitement

Procédons au traitement en choisissant dans le Menu Fichier la rubrique : Nouveau traitement ... et en cherchant d'ouvrir le fichier "donnees01s.csv" que nous avons préparé auparavant.

6.4 Choix de l'analyse

CHIC nous propose de choisir l'analyse qui nous intéresse parmi les trois proposés : Arbre des similarités, Graphe implicatif, Arbre cohésitif.

6.5 Choix de Graphe implicatif

Ayant choisi Graphe implicatif, trois fenêtres s'ouvrent :

- Fenêtre 1 : Elle contient 5 tableaux d'information, de résultats de calculs intermédiaires (dont la matrice des corrélations) et de la matrice des indices d'implications (compris entre 0 et 100).
- Fenêtre 2 : Elle contient le tableau des valeurs traitées en plus des mêmes tableaux que la fenêtre 1 mais présentées d'une manière plus conviviale
- Fenêtre 3 : Elle présente le graphe implicatif calculé .

Le Menu Fenêtre permet de choisir la fenêtre à visualiser.

6.6 Paramétrage : choix des seuils d'indices d'implication

Si la fenêtre 3 n'affiche rien (comme c'est le cas avec nos données), il est nécessaire de vérifier à droite de la fenêtre la Barre des paramètres du graphe implicatif en y choisissant les seuils des indices d'implication qui doivent apparaître avec les couleurs correspondantes.

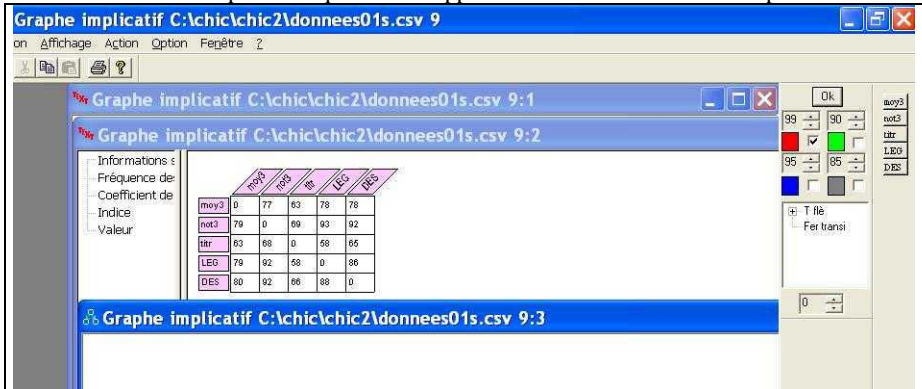


FIG. 3 – Matrice d'implication

Effectivement, sur l'image ci-dessus (FIG 3) seuls les indices au moins égaux à 99 ont été choisis alors que si nous consultons la matrice d'implication (Fenêtre 2) l'indice maximal obtenu sur nos données est : 93.

Ainsi nous pouvons choisir les indices à faire apparaître avec le paramétrage de ces seuils et des couleurs correspondantes

6.7 Paramétrage : différenciation des seuils

A partir de l'observation de la matrice d'implication, on peut décider de choisir les implications à faire apparaître en manipulant judicieusement les 4 seuils possibles. Dans la zone T flèche, après avoir développé le (+), on peut choisir l'épaisseur d'une flèche correspondant à un seuil. En cliquant sur Fer Transi et en choisissant 1 on peut obtenir le tracé de tous les arcs associés aux fermetures transitives aux seuils choisis, et avec respect des couleurs. Ces nouveaux arcs apparaissent en pointillés. Mais le choix de faire apparaître les fermetures transitives risque d'augmenter la complexité du graphique implicatif, même si on a la possibilité de choisir les variables qui apparaissent ou non soit en cliquant, soit en déclinquant sur un au moins des codes des variables vus sur la Barre des items placée à l'extrême droite de la fenêtre 3.

Après les différents choix effectués sur la Barre des paramètres du graphe implicatif et sur la Barre des items, on recalcule le graphique soit en mettant Ok, soit en faisant menu contextuel sur l'espace de travail des graphes (généralement sur PC : un clic droit), puis en choisissant la rubrique "Dessine le graphe de manière automatique". Cette action est aussi accessible dans le Menu Action parmi d'autres possibilités d'analyse. Si l'espace de travail pour dessiner le graphique est insuffisant, il suffit de modifier les dimensions en pixel dans le menu Option.

Remarquons que les deux variables supplémentaires n'apparaissent pas.

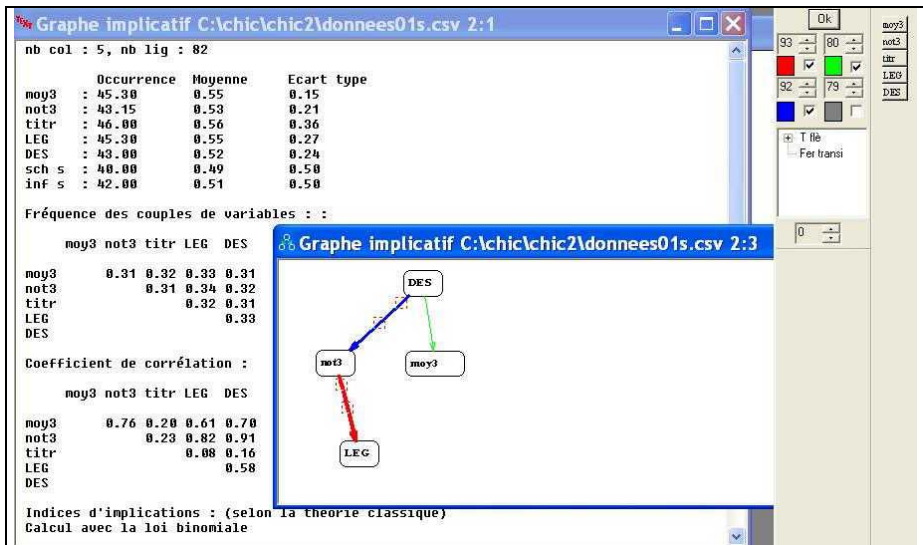


FIG. 4 – Occurrence, moyenne, écart-type, graphe implicatif

6.8 Utilisation de l'option : conjonction de variables

Si nous regardons de plus près, le graphe implicatif ci-dessus, il est assez dépouillé. La variable « titr » qui repère les étudiants ayant pu donner correctement ou non un titre aux images proposées n'entre même pas en jeu : elle n'apparaît pas dans la chaîne des implications au niveau que nous avons choisi (seuil d'intensité implicative 80). Cette variable semble ne pas avoir d'effet significatif. Mais le fait de la « coupler » avec d'autres variables nous apportera probablement plus d'information.

En effet, CHIC permet de calculer des conjonctions entre les variables prémisses de l'implication et ceci de manière automatique.

Ainsi, si on choisit par exemple 3 prémisses, on obtient des règles de la forme $a \wedge b \wedge c \Rightarrow d$. Le choix du nombre de prémisses qui interviendront dans les règles se fait dans le menu « Option » au niveau des paramètres « Option d'Apriori ». Si vous choisissez 1, alors vous aurez seulement des règles de la forme $a \Rightarrow b$. Si vous choisissez 2, vous aurez des règles de la forme $a \wedge b \Rightarrow c$ et ainsi de suite. Le calcul des conjonctions peut rapidement engendrer un nombre considérable de règles. La figure ci-dessous donne une idée de la complexité de la représentation obtenue à partir de seulement 5 variables avec le choix de 2 prémisses et si on ne contrôle pas ce que nous appelons « le seuil d'originalité ».

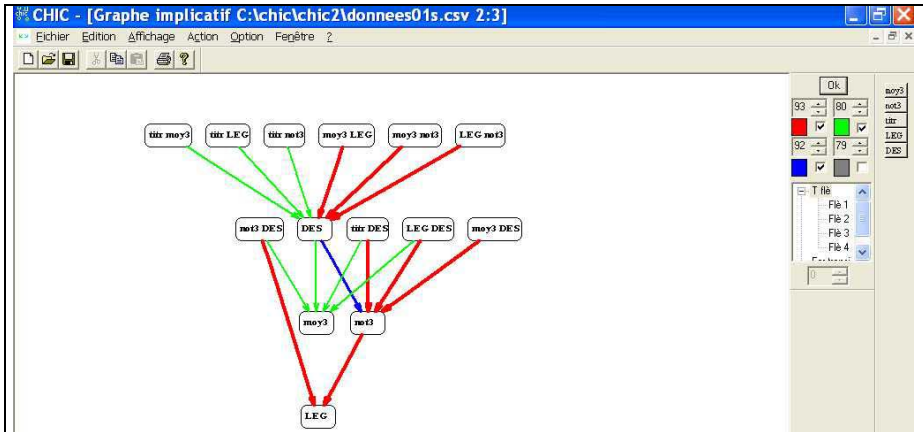


FIG. 5 – Conjonctions de variables, (2 prémisses, seuil de l'originalité 0)

Le graphe ci-dessus nous montre donc les 10 prémisses obtenues par les combinaisons de 5 éléments pris 2 à 2. Or, il est possible de retenir uniquement les règles ayant un caractère original, optimisant le nombre de conjonctions significatives pour le nombre de prémisses choisi. D'où l'option prévue à cet effet avec un seuil qui permet de retenir uniquement les règles supérieures à ce seuil. Le seuil d'originalité est un nombre compris entre 0 et 100. Pour approfondir cette notion, on peut consulter (Couturier R., 2004, 2008).

La figure ci-dessous, nous présente la conjonction qui serait la plus originale après qu'on ait choisi le seuil d'originalité 50 avec le seuil d'intensité implicative de 80 déjà préalablement sélectionné.. Dans notre exemple, c'est la conjonction de la variable « titr » et de la variable « moy3 » qui impliquerait la variable « DES ». Rappelons que la variable « titr » n'était même pas dans le graphe implicatif initial sans conjonction alors que sa moyenne est « la plus réussie ». Pourquoi apparaît-elle ici en conjonction avec la variable « moy3 » ?

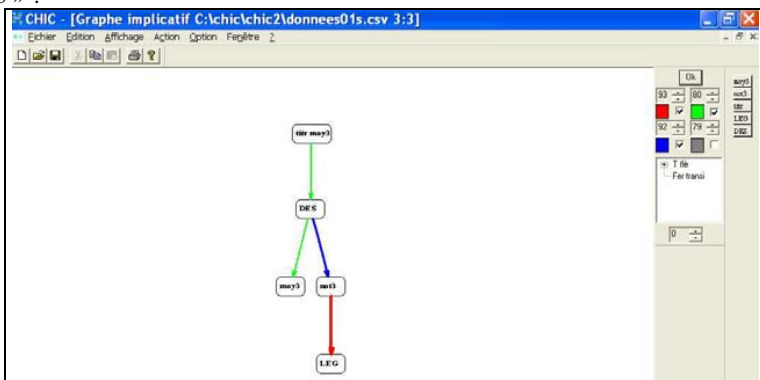


FIG. 6 – Conjonctions de variables, (2 prémisses et seuil de l'originalité 50)

Indéniablement, avec les conjonctions de variables le chercheur peut obtenir des informations complémentaires utiles au traitement de sa problématique.

7 Traitement pour obtenir un arbre cohésitif

7.1 Exécutons Chic.exe ou si CHIC est déjà ouvert, fermons le dernier travail

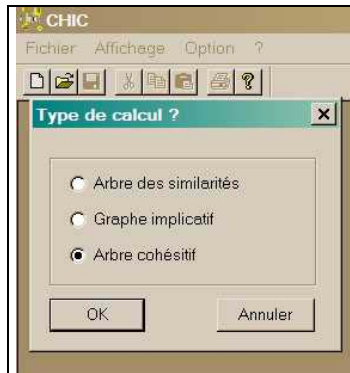


FIG. 7 – Choix du traitement : arbre cohésitif

7.2 Choix des options

Choisissons les Options. Préparons les paramétrages dans le menu Options. Dans un premier temps, pour une première approche avec un arbre implicatif, nous choisissons de faire apparaître les niveaux significatifs avec le Type d'implication selon la théorie classique et avec le type de loi qui sera la loi binomiale. Dans un second temps, avec des éclaircissements théoriques (cf. Partie 1) et en fonction de la problématique traitée on pourra choisir d'autres combinaisons d'options et de paramètres.

7.3 Début du traitement

Procédons au traitement en choisissant dans le Menu Fichier la rubrique : Nouveau traitement et en cherchant d'ouvrir le fichier "donnees01s.csv" que nous avons préparé auparavant.

7.4 Choix de l'analyse

CHIC nous propose de choisir l'analyse qui nous intéresse parmi les trois proposés : Arbre des similarités, Graphe implicatif, Arbre cohésitif.

7.5 Traitement "Arbre cohésitif"

Ayant choisi "Arbre cohésitif", trois fenêtres s'ouvrent (cf. FIG.8) :

- Fenêtre 1 : Elle contient en plus des 4 tableaux d'information déjà présents lors du traitement pour le graphe implicatif, la matrice des indices de cohésion (valeurs de 0 à 1), la classification des niveaux de cohésions et l'indication des nœuds les plus significatifs, option que nous avons demandée.
- Fenêtre 2 : Elle contient le tableau des valeurs traitées en plus des mêmes tableaux que la fenêtre 1 mais présentées d'une manière plus conviviale
- Fenêtre 3 : Elle présente l'arbre cohésitif avec les nœuds significatifs mis en couleurs.

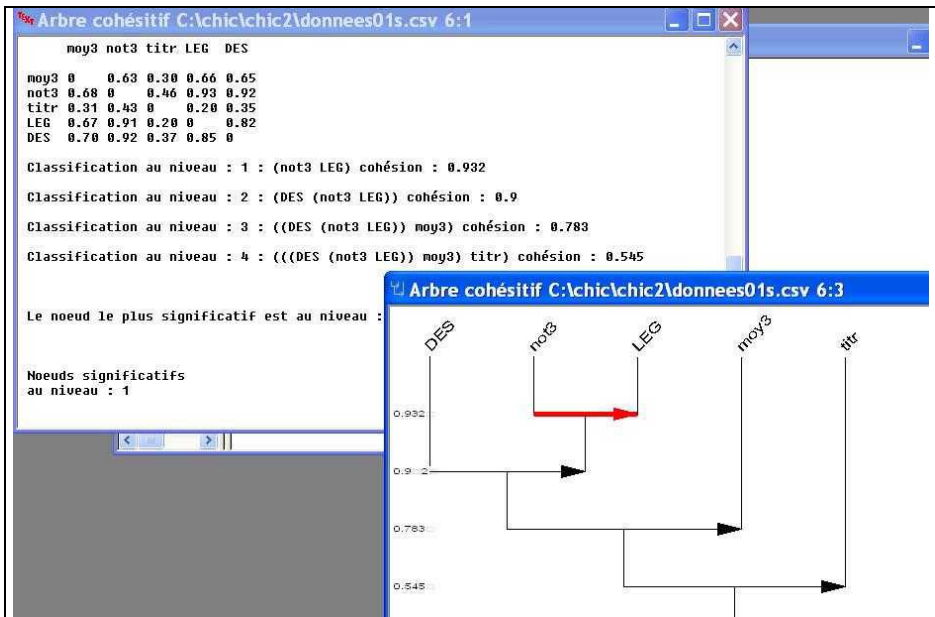


FIG. 8 – Les cohésions à chaque niveau

7.6 Compléments d'analyse

Dans le Menu "Action" nous pouvons choisir plus de possibilités d'analyse telle que « Calcul des contributions des variables supplémentaires ». Dès que ce complément d'analyse est demandé, dans la Fenêtre 1 apparaissent les informations correspondantes (cf. FIG 8):

7.7 Lecture des contributions des variables supplémentaires

La lecture de ces contributions nous révèle les risques d'erreur à déclarer que telle variable supplémentaire contribuerait à la formation de la classe de cohésion implicative

Exemple, avec nos données, pour la classe de cohésion implicative : DES, not3, LEG, moy3 (1,2,3) formée au niveau 3, on lit que "La variable supplémentaire sch contribue à cette classe avec un risque de : 0.00588 " et que "La variable supplémentaire inf contribue à cette classe avec un risque de : 0.996 ", autrement dit, normalement il y a 6 chances sur 1000 d'être dans l'erreur de déclarer que ce sont les étudiants ayant traité les schémas qui ont permis la construction de cette classe de cohésion implicative DES, not3, LEG, moy3 (1,2,3), alors que plus de 99 fois sur 100, on sera dans l'erreur de déclarer que ce sont les étudiants ayant traité l'infographie qui y ont contribué.

```

Contribution à la classe : not3,LEG ( 1 )
La variable sch contribue à cette classe avec un risque de : 2.62e-005
La variable inf contribue à cette classe avec un risque de : 1
La variable qui contribue le plus à cette classe est sch avec un risque de : 2.62e-005

Contribution à la classe : DES,not3,LEG ( 1,2 )
La variable sch contribue à cette classe avec un risque de : 0.000135
La variable inf contribue à cette classe avec un risque de : 1
La variable qui contribue le plus à cette classe est sch avec un risque de : 0.000135

Contribution à la classe : DES,not3,LEG,moy3 ( 1,2,3 )
La variable sch contribue à cette classe avec un risque de : 0.00588
La variable inf contribue à cette classe avec un risque de : 0.996
La variable qui contribue le plus à cette classe est sch avec un risque de : 0.00588

Contribution à la classe : DES,not3,LEG,moy3,titr ( 1,2,3,4 )
La variable sch contribue à cette classe avec un risque de : 0.418
La variable inf contribue à cette classe avec un risque de : 0.713
La variable qui contribue le plus à cette classe est sch avec un risque de : 0.418
  
```

FIG. 9 – Contributions des variables supplémentaires

7.8 Typicalité d'une variable supplémentaire

Dans le Menu Action de la Fenêtre 3, on peut demander aussi comme complément d'analyse la typicalité des variables supplémentaires. La Typicalité permet de connaître quelle variable supplémentaire (c'est-à-dire quel groupe d'individus) serait la plus typique d'une classe de cohésion implicative. La typicalité d'une variable supplémentaire sur une classe de cohésion implicative est aussi mesurée en risque d'erreur de la déclarer comme caractéristique de la classe. Ces typicalités s'affichent dans la Fenêtre 1.

7.9 Contributions des individus

Pour des analyses plus fines, avec CHIC et une rubrique de la Menu Action de la Fenêtre 3, on a la possibilité de connaître dans quelle mesure chaque individu a contribué à l'élaboration d'une classe de cohésion implicative. CHIC calcule en plus le sous-groupe d'individus qui a permis cette classe. C'est le **groupe optimal** qui a contribué le plus à la fabrication de cette classe. Ces informations sont affichées dans la Fenêtre 1

Exemple extrait de la Fenêtre 1 avec nos données sur la classe de cohésion implicative formée au niveau 3 :

- Contribution à la classe : DES, not3, LEG, moy3 (1, 2, 3)
- Contribution des individus, évaluée de 0 à 1

1 : 0.966	2 : 0.650	3 : 0.686	4 : 0.843	5 : 0.795	6 : 0.680
7 : 0.819	8 : 0.635	9 : 0.634	10 : 0.809	12 : 0.710	13 : 0.702
14 : 0.676	15 : 0.842	16 : 0.714	17 : 0.848	18 : 0.732	19 : 0.878
20 : 0.767	21 : 0.836	22 : 0.616	23 : 0.831	24 : 0.805	25 : 0.791
26 : 0.805	27 : 0.723	28 : 0.761	29 : 0.770	30 : 0.713	32 : 0.757
33 : 0.748	34 : 0.762	35 : 0.745	36 : 0.598	37 : 0.580	38 : 0.690
39 : 0.732	40 : 0.756	41 : 0.715	42 : 0.660	43 : 0.759	44 : 0.593
45 : 0.732	46 : 0.753	47 : 0.601	48 : 0.673	49 : 0.546	50 : 0.540
51 : 0.657	52 : 0.727	53 : 0.702	54 : 0.714	55 : 0.671	56 : 0.677
57 : 0.703	58 : 0.721	59 : 0.686	60 : 0.657	61 : 0.696	62 : 0.704
63 : 0.682	64 : 0.691	65 : 0.697	66 : 0.605	67 : 0.585	68 : 0.642
69 : 0.607	70 : 0.631	71 : 0.625	72 : 0.654	73 : 0.651	74 : 0.614
75 : 0.657	76 : 0.612	78 : 0.640	79 : 0.619	80 : 0.565	81 : 0.565
82 : 0.559	83 : 0.568	84 : 0.554	86 : 0.575		

TAB. 5 – Contribution de chaque individu à un niveau de classe de cohésions

Groupe optimal : card 34										
12	30	54	16	41	58	27	52	18	39	45
35	33	46	40	32	43	28	34	20	29	25
5	26	24	10	7	23	21	15	4	17	19
1										

TAB. 6 – Un groupe optimal d'individus contribuant à la formation d'un niveau de classe de cohésions

7.10 Typicalité de chaque individu

D'une manière analogue à la typicalité d'une variable supplémentaire, on peut connaître d'une manière plus fine la typicalité de chacun des individus et aussi, avec la liste "groupe optimal", le sous-groupe d'individus qui serait le plus typique d'une classe de cohésion implicative.

Exemple extrait de la Fenêtre 1 avec nos données sur la classe de cohésion implicative formée au niveau 3 :

- Typicalité à la classe : DES, not3, LEG, moy3 (1,2,3)

- Typicalité des individus évaluée de 0 à 1 :

1 : 0.797	2 : 0.566	3 : 0.402	4 : 0.887	5 : 0.842	6 : 0.562
7 : 0.798	8 : 0.301	9 : 0.547	10 : 0.679	12 : 0.641	13 : 0.281
14 : 0.590	15 : 0.845	16 : 0.571	17 : 0.841	18 : 0.410	19 : 0.846
20 : 0.760	21 : 0.813	22 : 0.082	23 : 0.775	24 : 0.745	25 : 0.698
26 : 0.725	27 : 0.637	28 : 0.525	29 : 0.664	30 : 0.354	32 : 0.611
33 : 0.520	34 : 0.630	35 : 0.592	36 : 0.150	37 : 0.013	38 : 0.366
39 : 0.558	40 : 0.583	41 : 0.415	42 : 0.249	43 : 0.584	44 : 0.088
45 : 0.545	46 : 0.572	47 : 0.189	48 : 0.263	49 : 0.020	50 : 0.009
51 : 0.216	52 : 0.486	53 : 0.476	54 : 0.501	55 : 0.294	56 : 0.307
57 : 0.432	58 : 0.481	59 : 0.412	60 : 0.253	61 : 0.432	62 : 0.431
63 : 0.414	64 : 0.400	65 : 0.442	66 : 0.279	67 : 0.006	68 : 0.201
69 : 0.112	70 : 0.265	71 : 0.229	72 : 0.246	73 : 0.246	74 : 0.131
75 : 0.319	76 : 0.142	78 : 0.283	79 : 0.168	80 : 0.000	81 : 0.000
82 : 0.010	83 : 0.043	84 : 0.018	86 : 0.070		

TAB. 7-: *Typicalité de chaque individu à un niveau de classe de cohésions*

Groupe optimal : card 40

62 61 57 65 53 58 52 54 33 28 45
 9 39 6 2 16 46 40 43 14 35 32
 34 27 12 29 10 25 26 24 20 23 1
 7 21 17 5 15 19 4

TAB. 8 – *Un groupe optimal d'individus typique d'un niveau de classe de cohésions*

7.11 Comparaison des individus avec le groupe optimal

La comparaison des individus associés à une variable supplémentaire typique d'une classe de cohésion implicative avec le groupe optimal des individus typiques de cette classe devrait apporter des informations intéressantes pour la problématique étudiée. Idem pour l'étude des contributions.

7.12 Utilité des groupes optimaux

La possibilité d'obtenir les groupes optimaux est intéressante si des variables supplémentaires ne sont pas au préalable définies.

8 Traitement de variables sur intervalles

Nous allons procéder à une analyse qui considère des **variables sur intervalles**.

Revenons sur la forme initiale de nos données mais sans les 2 variables supplémentaires Sch et Inf qui rappelons-le jouent le rôle de descripteurs. Au cours des traitements suivants nous allons regarder dans chacun des deux groupes comment se structurent les 5 variables principales du point de vue de la cohésion implicative. Nous avons choisi de procéder de

cette sorte puisque lors des analyses précédentes, nous avons observé que la variable supplémentaire Inf, contrairement à Sch, ne contribuerait nullement à la structuration des 5 variables principales sur les 82 individus.

Ainsi, à partir du tableau initial, nous créons deux tableaux. Le premier contenant les résultats du sous-groupe d'étudiants ayant 1 à la variable "Inf" (42 étudiants) et le second contenant ceux qui ont traité les images sous forme de schéma (40 étudiants). Les deux tableaux ont chacun les 5 variables principales.

1. La note moyenne obtenue par l'étudiant lors de 3 devoirs précédents. (note /20) (code : moy3)
2. La note obtenue par l'étudiant lors de cette évaluation écrite (note /20) (code : not3)
3. La note obtenue par l'étudiant sur une question demandant l'intitulé d'une image (note /20) (code : titr)
4. La note obtenue par l'étudiant sur une question demandant de légender l' image (note /20) (code : LEG)
5. La note obtenue par l'étudiant sur une question demandant de décrire le processus représenté par l'image (note /20) (code : DES)

Les 5 premières lignes du premier tableau "données-inf" contenant les notes des étudiants ayant traité les images sous forme "infographie" (la première colonne contient les codes des étudiants) se présentent ainsi :

	moy3 p	not3 p	titr p	LEG p	DES p
7	16	17	20	16	18
8	15	17	20	13	20
18	16	15	10	12	17
22	16	14	0	9	18

TAB. 9 – 1er groupe de données sous forme variables sur intervalles

Et les 6 premières lignes du second tableau "données-sch" contenant les notes des étudiants ayant traité les images sous forme "schéma" (la première colonne contient les codes des étudiants) se présentent ainsi :

	Moy3 p	not3 p	titr p	LEG p	DES p
1	19	19	10	20	18
2	12	19	10	20	20
3	16	18	20	14	20
4	16	17	10	19	18
5	13	17	10	20	16

TAB. 10 – 2ème groupe de données sous forme variables sur intervalles

Remarquons que la première ligne contient les codes respectifs des 5 variables principales, mais comme nous voudrions les traiter comme des variables sur intervalles, nous avons accolé à chaque code initial le caractère "espace" suivi du caractère "p". Par exemple, "moy3" est devenu "moy3 p".

Ces deux tableaux doivent donc être enregistrés dans deux fichiers-textes de type .CSV nommés respectivement "donnees-inf.csv" et "donnees-sch.csv"

Les 6 premières lignes du fichier "donnees-inf.csv" sont sous cette forme :

```
;moy3 p; not3 p; titr p; LEG p; DES p
7;16;17;20;16;18
8;15;17;20;13;20
13;15;16;20;11;16
18;16;15;10;12;17
22;16;14;0;9;18
```

TAB. 11 – données sous forme variables sur intervalles en format .CSV

Rappelons que lorsqu'une variable est proposée au logiciel CHIC comme **variable sur intervalles**, CHIC "découpe les valeurs de cette variable en un nombre fixe d'intervalles. Le nombre d'intervalles est choisi par l'utilisateur et ensuite l'algorithme des nuées dynamiques Diday (thèse d'Etat, 1972) constitue automatiquement les intervalles qui ont des limites distinctes. Cet algorithme a la particularité de construire des intervalles en minimisant l'inertie de chaque intervalle. Ensuite, un intervalle est représenté par une variable binaire et un individu a la valeur 1 s'il appartient à cet intervalle et 0 sinon. En utilisant une telle décomposition, un individu appartient à un seul intervalle". Durant leurs traitements, ces nouvelles présentations des données sont enregistrées par CHIC dans un fichier-texte de type .CSV dont le nom se termine par "_partition.csv"; par exemple pour nos données, nous obtenons "donnees-inf_partition.csv" ou "donnees-sch_partition.csv".

Les 6 premières lignes du fichier "donnees-inf_partition.csv" sont sous cette forme :

```
;moy3@1 i;moy3@2 i;moy3@3 i;not3@1 i;not3@2 i;not3@3 i;titr@1 i;titr@2
i;titr@3 i;LEG@1 i;LEG@2 i;LEG@3 i;DES@1 i;DES@2 i;DES@3 i
7;0;0;1;0;0;1;0;0;1;0;0;1;0;0;1
8;0;0;1;0;0;1;0;0;1;0;0;1;0;0;1
13;0;0;1;0;0;1;0;0;1;0;0;1;0;0;1
18;0;0;1;0;0;1;0;0;1;0;0;1;0;0;1
22;0;0;1;0;0;1;1;0;0;0;1;0;0;0;1
```

TAB. 12 – fichier .CSV de données binaires générées par chic à partir de variables sur intervalles

En effet, au moment où nous exécutons l'application CHIC, dans une première approche, nous paramétrons les options et en particulier nous ne cochons pas le paramètre "Calcul des intervalles"; En effet, si, en vue d'une analyse plus approfondie, nous le cochons, les traitements correspondants traiteront des **variables-intervalles** : c'est-à-dire "dans la suite du calcul, toutes les unions des intervalles d'une variable sont considérées", apportant par conséquent des résultats plus fins mais plus complexes, ainsi dans une première approche nous ne demandons pas cette option de "calcul des intervalles".

Après avoir paramétré les options, nous demandons un nouveau traitement sur le fichier "donnees-inf.csv" avec le type de calculs "Arbre cohésitif" et comme CHIC a repéré au moins une variable avec le code terminé par "une espace suivie de p", pour cette variable il pose à l'utilisateur une question sur le nombre de subdivisions que comportera la partition de

l'intervalle. Pour notre exemple, nous avons choisi la partition en 3 parties. Ainsi pour la variable "LEG p" nous allons obtenir les 3 nouvelles variables codées:

LEG@1 i LEG@2 i LEG@3 i

Dans la fenêtre 1, en plus des résultats intermédiaires habituels (5 tableaux d'information, de résultats de calculs intermédiaires (dont la matrice des corrélations) et de la matrice des indices de cohésions (compris entre 0 et 1)), nous obtenons des informations sur la partition obtenue par l'algorithme des nuées dynamiques, par exemple, le tableau suivant pour la variable sur intervalle LEG p:

Partitions optimales :
LEG1 de 2 à 4
LEG2 de 6 à 9
LEG3 de 11 à 16

TAB. 13 – Exemple de partition en 3 parties obtenue par l'algorithme des nuées dynamiques

Ce tableau nous informe qu'un étudiant ayant eu entre 2 et 4 inclus sera classé dans la variable LEG1, que celui qui a eu entre 6 et 9 inclus sera dans LEG2, et que dans LEG3 seront classés ceux ayant une note comprise entre 11 à 16. En gros LEG1 sont ceux qui ne savent pas légender et LEG3 ceux qui ont pu reconnaître une majorité de légendes et LEG2 ceux qui sont juste moyens. Pour information, la répartition des étudiants sur ces subdivisions est LEG1 : 8, LEG2 : 25, LEG3 : 9.

La fenêtre 3 nous donne l'arbre cohésitif suivant des variables sur intervalles : (cf FIG 9)

Brièvement, nous pouvons lire sur ce graphe qu'il y a 3 niveaux significatifs et qu'au premier niveau significatif le fait de ne pas savoir décrire entraînerait le fait d'avoir une mauvaise note à l'évaluation ; autrement dit en appliquant la contraposée "le fait de ne pas avoir une mauvaise note à l'évaluation" impliquerait que "l'étudiant ne soit pas mauvais sur la description". Et qu'au second niveau significatif de la cohésion (LEG1, (DES1,not31)) impliquerait MOY31, c'est-à-dire, en gros, "ne rien connaître sur le fait de savoir légender" ou "ayant une mauvaise note à l'évaluation même si la description n'est pas mauvaise" impliquerait "une mauvaise moyenne au 3 devoirs précédents"

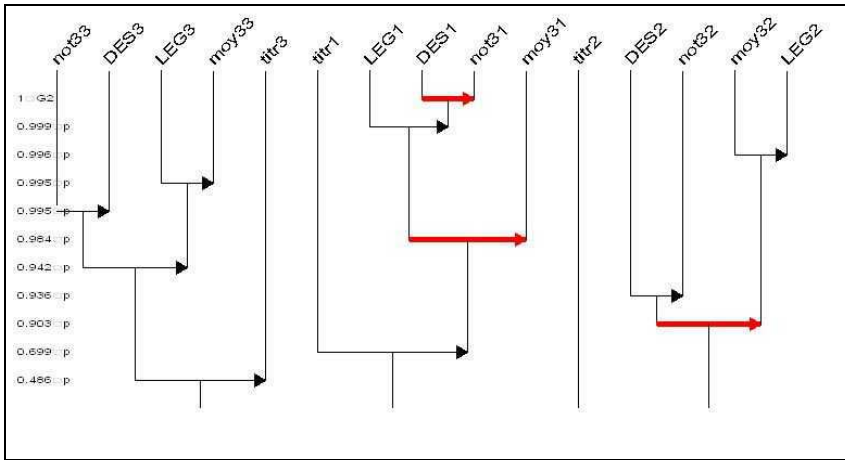


FIG. 10 – arbre cohésitif suivant des variables sur intervalles (1er sous groupe de données)

Or, nous ne retrouvons pas ces structures de cohésions implicatives si nous menons l'analyse sur le sous-groupe d'étudiants ayant traité les images sous forme de schéma.

En effet, le traitement des données contenues dans le fichier "données-sch.csv" où les 5 variables principales sont des variables sur intervalles (données des étudiants ayant traité l'évaluation avec les images sous forme de schémas).; nous a produit l'arbre cohésitif suivant

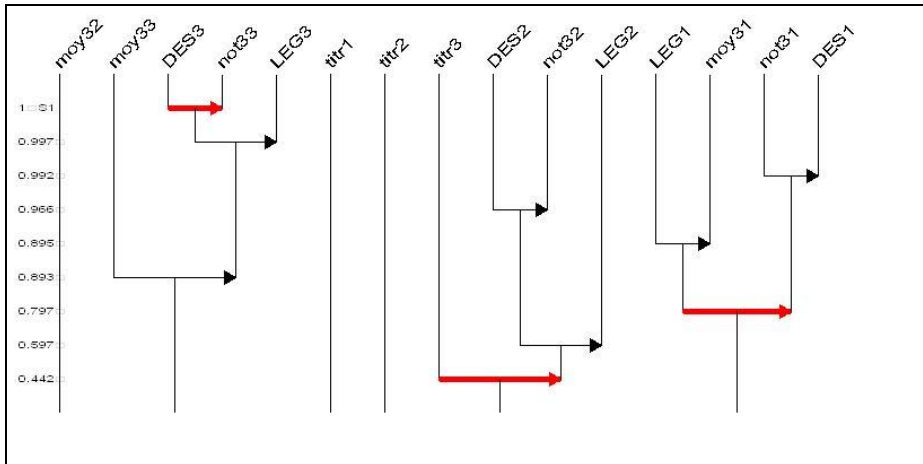


FIG. 11 – arbre cohésitif suivant des variables sur intervalles (2ème sous groupe de données)

Pour ce dernier traitement, nous avons aussi choisi 3 subdivisions pour chaque variable sur intervalle. Sur ce résultat on peut lire que pour ce groupe d'étudiants, au premier niveau significatif c'est le fait de savoir décrire qui impliquerait probablement le fait d'avoir une bonne note sur l'évaluation.; C'est-à-dire qu'il y aurait certains étudiants qui ont eu de bonnes

notes sur l'évaluation sans même savoir décrire. Ce qui correspond bien au tableau de contingence suivant entre les variables sur intervalles not33 et Des3 (codés respectivement dans le fichier "donnees-inf_partition.csv" **not3@3 i** et **DES@3 i**)

Sch	not3@3 i		
DES@3 i	0	1	Total
0	26	2	28
1	0	12	12
Total	26	14	40

TAB. 14 – *Tableau de contingence*

En comparant les deux structures de classification des cohésions implicatives (les deux arbres cohésitifs) on peut remarquer des différences qui doivent amener à se poser des questions sur le fond de la problématique.

En effet, d'une part, dans le groupe ayant traité l'infographie on aurait donc significativement au premier niveau DES1=>not31 alors que pour l'autre groupe on aurait not31=>DES1 (au 3^{ème} niveau de classification). Et d'autre part, on peut remarquer entre les deux sous-groupes les différences de positions des performances des étudiants lors de l'évaluation (reflétées par MOY31, MOY32, MOY33) par rapport à l'ensemble des autres variables principales dans les classifications des cohésions implicatives. Par ailleurs la structure globale de l'arbre produite par le fichier donnees-inf.csv (étudiant ayant traité par infographie) semble plus "cohérent" que celle produite par donnees-inf.csv (étudiant ayant traité les schémas).

En outre, notons que sur notre exemple de données les informations apportées par les arbres cohésitifs ne sont pas accessibles à partir des graphes implicatifs si on les avait demandés. Et aussi, si nous avions demandé les arbres des similarités, nous n'aurions pas vu les rôles dissymétriques que jouent les variables not31 et DES1 selon que l'on soit dans tel ou tel groupe.

9 Conclusion

En conclusion, afin de montrer les principales procédures d'utilisation du logiciel CHIC nous avons procédé dans cette partie, à deux types d'analyse que permet cette application : celle qui permet d'obtenir un graphe implicatif avec des variables supplémentaires et celle qui aboutit à des arbres cohésitifs. Dans les procédures, il y a la préparation d'une part des fichier-textes au format *.CSV sur lesquels les données sont organisées dans le format exigé par CHIC, après les avoir saisies et préparées à partir d'un tableur, d'autre part, les choix des options et des paramètres spécifiques aux calculs de graphe implicatif ou à ceux de l'arbre cohésitif. Pour des analyses plus approfondies, nous avons vu comment on obtient les contributions respectives des variables supplémentaires dans les calculs et même s'il n'y a pas de variables supplémentaires, CHIC nous a donné la possibilité d'obtenir les groupes optimaux qui ont contribué aux calculs d'une classification ou ceux qui en sont typiques. Pour une analyse plus fine, nous avons demandé pour chacun des deux groupes d'individus un arbre cohésitif sur les 5 variables principales. Cela nous a permis de voir des différences

de structures et les dissymétries qui devraient aider à poser les bonnes questions concernant la problématique à approfondir.

Références

- Couturier, R., (2008). *CHIC : utilisation et fonctionnalités* Laboratoire d'Informatique de l'université de Franche-Comté de Belfort-Montbéliard, BP 527, 90016 Belfort, raphael.couturier@iut-bm.univ-fcomte.fr
- Couturier, R, R. Gras et F. Guillet (2004). Reducing the number of variables using implicative analysis In International Federation of Classification Societies, IFCS 2004, Springer Verlag: Classification, Clustering, and Data Mining Applications, p. 277--285, ISBN 3-540-22014-3, Chicago, USA, July 2004.
- Couturier, R. (2008). *CHIC, Cohesive Hierarchical Implicative Classification*, in R. Gras, E. Suzuki, F. Guillet and F. Spagnolo (Eds) *Statistical Implicative Analysis* Berlin : Springer-Verlag, p. 41-54

Ressources bibliographiques en ligne

http://www.ardm.asso.fr/chic_web/bibliographie_chic/bibliographie_chic_base.xml

Summary

In this article, within the framework of an implicative statistic analysis, we made the choice to use the datas of a current research, in order to present and describe the steps and procedures to be follow with the CHIC software. The aim is, on the one hand, to get an implication graph and an oriented hierarchical tree, and on the other hand, to make easier the reading of the results thanks to the various possibilities and CHIC functionalities. Step by step, we will deal with the description of collected datas, the transformation of these datas, the choice of the variables which will be supplementary variables, the preparation of the text file under .CSV format which will be compatible with CHIC, the processing to get an implicative graphic, particularly with the conjunction of premises and the threshold of originality concepts, the processing to get an oriented hierarchical tree completed by the significant levels, contribution of an optimal set of individuals and computation of typicality and contribution concepts, and we will finish with processing on interval variables.

