L'Analyse Statistique Implicative (A.S.I.) en réponse à des problèmes fondateurs

Régis Gras, Pascale Kuntz

LINA– Ecole Polytechnique de l'Université de Nantes
La Chantrerie BP 60601 44306 Nantes cedex
regisgra@club-internet.fr, pascale.kuntz@polytech.univ-nantes.fr

Résumé: Partie de situations de didactique des mathématiques, la méthode d'analyse statistique implicative se développe au fil des problèmes rencontrés et des questions posées. Son objectif majeur vise la structuration de données croisant sujets et variables, l'extraction de règles inductives entre les variables et, à partir de la contingence de ces règles, l'explication et donc une certaine prévision dans divers domaines : psychologie, sociologie, biologie, etc.. C'est dans ce but que les concepts d'intensité d'implication, de cohésion de classes, d'implication-inclusion, de significativité de niveaux hiérarchiques, de contribution de variables supplémentaires, etc., sont fondés. De la même façon, au traitement de variables binaires (par exemple, des descripteurs) s'ajoutent progressivement ceux des variables modales, fréquentielles et, récemment, de variables-intervalles et de variables floues.

Mots-clés: implication statistique, induction, implication-inclusion, entropie, cohésion, graphe implicatif, arbre hiérarchique, niveau significatif, dialectique, variable binaire, variable modale, variable fréquentielle, variable-sur-intervalle, variable-intervalle, règle, méta-règle, règle généralisée, contribution, typicalité.

Préambule

Les connaissances opératoires de l´homme se constituent principalement selon deux composantes : celle des faits et celles des règles entre les faits ou entre règles elles-mêmes. Ce sont ses apprentissages qui, à travers sa culture et par ses expériences personnelles, lui permettent une élaboration progressive de ces formes de connaissances, en dépit des régressions, des remises en cause, des ruptures qui surgissent au détour d'infirmations décisives. Cependant, on sait que celles-ci contribuent dialectiquement à lui assurer un équilibre opératoire. Or les règles se forment inductivement de façon relativement stable dès lors que le nombre de succès, quant à leur qualité explicative ou anticipatrice, atteint un certain niveau (de confiance) à partir duquel elles seront susceptibles d´être mises en œuvre. En revanche, ce niveau (subjectif) non atteint, l´économie de l´individu le fera résister, dans un premier temps, à son abandon ou à sa critique. En effet, il est coûteux de substituer à la règle initiale une autre règle lors de l´apparition d´un faible nombre d'infirmations, dans la mesure où elle aurait été confortée par un nombre important de confirmations. Un accroissement de ce nombre d'instances négatives, fonction de la qualité de robustesse du niveau de confiance en la règle, conduira peut-être à un réajustement de celle-ci, voire à son abandon. Laurent Fleury (Fleury L. 1996), dans sa thèse, cite avec pertinence l´exemple -que je reprends- de la règle fort admissible : " toutes les Ferrari sont rouges". Cette règle, très robuste, ne sera pas abandonnée lors de l´observation d´un seul ou de deux contre-exemples. D´autant qu´elle ne manquerait pas d´être rapidement re-confortée.

Ainsi, à l´opposé de ce qui est légitime en mathématiques, où toute règle (théorème) ne souffre pas d´exception, où le déterminisme est total, les règles en sciences humaines, plus généralement en sciences dites "molles", sont acceptables et donc opératoires tant que le nombre de contre-exemples restera "supportable" eu égard à la fréquence de situations où elles seront positives et efficaces. Le problème, en analyse des données, est alors d´établir un critère numérique, relativement consensuel, pour définir la notion de niveau de confiance ajustable au niveau d´exigence de l´usager de la règle. Qu´il soit établi sur des bases statistiques a tout lieu de ne pas surprendre. Qu´il possède une propriété de résistance non linéaire au bruit (faiblesse du ou des premiers contre-exemples) peut également paraître naturel, conforme au sens "économique" évoqué plus haut. Qu´il s´effondre si les contre-exemples se répètent semble aussi devoir guider notre choix dans la modélisation du critère recherché. Ce texte présente le choix épistémologique que nous avons fait. En tant que tel il est donc réfutable mais le nombre de situations et d'applications où il s´est avéré pertinent et fécond nous conduit à en restituer ici la genèse.

1 Introduction

Différentes approches théoriques ont été adoptées pour modéliser l'extraction et la représentation de règles d'inférence imprécises (ou partielles) entre variables binaires (ou attributs ou caractères) décrivant une population d'individus (ou sujets ou objets). Mais les situations de départ et la nature des données ne modifient pas la problématique initiale. Il s'agit de découvrir des règles inductives non symétriques pour modéliser des relations du type *"si a alors presque b"*. C'est, par exemple, l'option des réseaux bayesiens (par ex. : Amarger S., Pearl J. 1988) ou des treillis de Galois (par ex. : Simon A. 2000). Mais le plus souvent, la corrélation et le test du χ^2, s'avérant inadaptés du fait de leur caractère symétrique, la probabilité conditionnelle (Loevinger 1947, Agrawal et al. 1993, Gras et al. 2004 a) reste le moteur de la définition de l'association, même quand l'indice de cette association retenu est de type multivarié (par ex. Bernard J.-M. 1999).

De plus et à notre connaissance, d'une part, le plus souvent les développements différents et intéressants se centrent sur des propositions d'un indice d'implication partielle pour des données binaires (cf. [Lerman et al. 2004] ou [Lallich et al 2005] dans ces actes ASI 05), d'autre part, cette notion n'est pas étendue à d'autres types de variables, à l'extraction et la représentation selon un graphe de règles ou selon une hiérarchie de méta-règles ; structures visant l'accès à la signification d'un tout non réduit à la somme de ses parties[1], c'est-à-dire fonctionnant comme un système complexe non linéaire. Par exemple, on sait fort bien, par l'usage, que la signification d'une phrase ne passe pas complètement par le sens de chacun des mots la composant.

Revenons à ce que nous croyons fertile dans la démarche que nous développons. Il semblerait que, dans la littérature, la notion d'indice d'implication ne soit pas non plus étendue à la recherche de sujets et de catégories de sujets responsables des associations. Ni que cette responsabilité soit quantifiée et conduise, de ce fait, à une structuration réciproque de l'ensemble des sujets, conditionnée par leurs relations aux variables.

Nous proposons justement ici ces prolongements après avoir rappelé le paradigme fondateur.

2 L'intensité d'implication dans le cas binaire

2.1 Situation fondamentale et fondatrice

Une population E d'objets ou de sujets est croisée avec des variables (caractères, critères, réussites, ...) que l'on interroge de la façon suivante : "*dans quelle mesure peut-on considérer qu'instancier la variable[2] a implique instancier la variable b ? Autrement dit, les sujets ont-ils tendance à être b si l'on sait qu'ils sont a ?"*. Dans les situations naturelles, humaines ou sciences de la vie, où les théorèmes (si a alors b) au sens déductif du terme ne peuvent être établis du fait des exceptions qui les entachent, il est important pour le chercheur et le praticien de "*fouiller dans ses données*" afin de dégager cependant des règles suffisamment fiables (des sortes de "théorèmes partiels", des inductions) pour pouvoir conjecturer[3] une possible relation causale, une genèse, pour décrire, structurer une population et faire l'hypothèse d'une certaine stabilité à des fins descriptives et, si possible, prédictives. Mais cette fouille exige la mise au point de méthodes pour la guider et pour la dégager du tâtonnement et de l'empirisme.

2.2 Mathématisation

Pour cela, à l'instar de la méthode de mesure de la similarité de I.C. Lerman (Lerman I.C. 1970, 1981), à l'instar de la démarche classique dans les tests non paramétriques (ex. Fischer, Wilcoxon, etc.), nous définissons (Gras R. 1979, Gras R et al.1996) la mesure de qualité confirmatoire de la relation implicative $a \Rightarrow b$ à partir de l'invraisemblance de l'apparition, dans les données, du nombre de cas qui l'infirment, c'est-à-dire pour lesquels a

1 C'est ce que souligne le philosophe L. Sève : « …dans le passage non additif, non linéaire des parties au tout, il y a apparition de propriétés qui ne sont d'aucune manière précontenues dans les parties et qui ne peuvent donc s'expliquer par elles » (« Emergence, complexité et dialectique », Odile Jacob, mai 2005). F.Wahl affirme de la même façon : « L'idée est que le système prime sur les parties, chaque partie se définissant par sa différence avec une autre et par son articulation avec toutes les autres. » (entretien recuilli par C.David, Nouvel Observateur n° 02228 des 16-22 août 2007).

2 Dans tout l'article, le mot « variable » désigne aussi bien une variable isolée en prémisse (exemple : « être blonde ») ou bien une conjonction de variables isolées (ex. : « être blonde et avoir moins de 30 ans et habiter Paris »)

3 « L'exception confirme la règle » nous dit l'adage populaire au sens où il n'y aurait pas d'exceptions s'il n'y avait pas de règle.

est vérifié sans que b ne le soit. Ceci revient à comparer l'écart entre le contingent et le théorique si seul le hasard intervenait[4]. Mais, dans le cadre de l'analyse de données, c'est cet écart qui est pris en compte et non pas l'énoncé d'un rejet ou de l'admissibilité d'hypothèse nulle.. Cette mesure est relativisée au nombre de données vérifiant respectivement a et non b, circonstance dans laquelle l'implication est précisément mise en défaut.. Elle quantifie "*l'étonnement*" de l'expert devant le nombre invraisemblablement petit de contre-exemples eu égard à une indépendance supposée entre les variables et aux effectifs en jeu.

Précisons. Un ensemble fini V de v variables est donné : a,b,c,...Dans la situation paradigmatique classique, il s'agit des performances (réussite-échec) à des items d'un questionnaire. A un ensemble fini E de n sujets x, on associe, par abus d'écriture, les fonctions du type : $x \rightarrow a(x)$ où $a(x) = 1$ (ou $a(x)$ = vrai si x satisfait ou possède le caractère a et 0 (ou $a(x)$ = faux) sinon. En intelligence artificielle, on dira que x est un exemple ou une instance pour a si $a(x) = 1$ et un contre-exemple dans le cas contraire.

La règle "$a \Rightarrow b$" est logiquement vraie si pour tout x de l'échantillon, b(x) n'est nul que dans le cas où a(x) l'est aussi ; autrement dit si l'ensemble A des x pour lesquels $a(x)=1$ est contenu dans l'ensemble B des x pour lesquels $b(x)=1$. Cependant, cette inclusion stricte n'est qu'exceptionnellement observée dans les expériences réelles. Dans le cas d'un questionnaire de connaissances, on pourrait en effet observer quelques rares élèves réussissant un item a et ne réussissant pas l'item b, sans que ne soit contestée la *tendance* à réussir b quand on a réussi a. Relativement aux cardinaux de E (soit n), mais aussi de A (soit n_a) et B (soit n_b), c'est donc le "poids" des contre-exemples (soit $n_{a\wedge\bar{b}}$) qu'il faudra prendre en compte pour accepter statistiquement de conserver ou non la **quasi-implication** ou la **quasi-règle** "$a \Rightarrow b$". Ainsi, c'est à partir de la dialectique exemples-contre-exemples que la règle apparaît comme le dépassement de la contradiction.

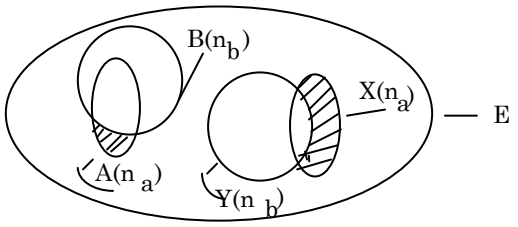
2.3 Formalisation

Pour formaliser cette quasi-règle, nous considérons, comme le fait I.C. Lerman pour la similarité, deux parties quelconques X et Y de E, choisies aléatoirement et indépendamment (absence de lien a priori entre ces deux parties) et de mêmes cardinaux respectifs que A et B. Soit $\bar{Y}$ et $\bar{B}$ les complémentaires respectifs de Y et de B dans E de même cardinal $n_{\bar{b}} = n-n_b$.

Nous dirons alors :

Définition 1 : $a \Rightarrow b$ est *admissible au niveau de confiance1-* α si et seulement si

$$Pr[\, Card(X \cap \bar{Y}) \leq card(A \cap \bar{B})\,] \leq \alpha$$



Les parties hachurées correspondent aux contre-exemples de l'implication a => b.

FIG 1

Il est établi (Lerman I.C. et al. 1981) que, pour un certain processus de tirage, la variable aléatoire $Card(X \cap \bar{Y})$ suit la loi de Poisson de paramètre $\frac{n_a n_{\bar{b}}}{n}$. Nous parvenons à ce même résultat en procédant différemment de la façon suivante :

4 « … [en accord avec Jung] si la fréquence des coïncidences n'excède pas de façon significative la probabilité qu'on peut leur calculer en les attribuant au seul hasard à l'exclusion de relations causales cachées, nous n'avons certes aucune raison de supposer l'existence de telles relations. », H.Atlan, « A tort et à raison. Intercritique de la science et du mythe », Seuil, 1986.

Notons X (resp. Y) le sous-ensemble aléatoire de transactions binaires où a (resp. b) apparaîtrait, de façon indépendante, avec la fréquence $\frac{n_a}{n}$ (resp. $\frac{n_b}{n}$). Pour préciser le mode d'extraction, par exemple, des transactions spécifiées aux variables a et b, respectivement A et B, on énonce les hypothèses sémantiquement admissibles suivantes, relativement à l'observation de l'événement :[a=1 et b=0]. $A \cap \overline{B}$ [5] est le sous-ensemble des transactions, contre-exemples de l'implication $a \Rightarrow b$:

Hypothèses :

h1 : les temps d'attente d'un événement [a et non b] sont des variables aléatoires indépendantes ;

h2 : la loi du nombre d'événements arrivant dans l'intervalle de temps [t, t+T[ne dépend que de T ;

h3 : deux tels événements ne peuvent arriver simultanément

On démontre alors (par exemple dans (Saporta, 2006)) que le nombre d'événements se produisant pendant une période de durée n fixée suit une loi de Poisson de paramètre c.n où c est appelé cadence du processus d'apparitions pendant l'unité de temps.

Or pour chaque transaction supposée aléatoire, l'événement [a=1] a pour probabilité la fréquence $\frac{n_a}{n}$ l'événement [b=0] a pour probabilité la fréquence $\frac{n_{\overline{b}}}{n}$ donc l'événement conjoint [a=1 et b=0] a pour probabilité estimée par la fréquence $\frac{n_a}{n}.\frac{n_{\overline{b}}}{n}$ dans l'hypothèse d'absence de lien a priori entre a et b (indépendance)

On peut donc alors estimer la cadence c de cet événement par $\frac{n_a}{n}.\frac{n_{\overline{b}}}{n}$.

Ainsi pour une durée de temps n, les apparitions de l'événement [a et non b] suivent une loi de Poisson de paramètre :

$$\lambda = \frac{n_a . n_{\overline{b}}}{n}$$

Par suite, $\Pr[\mathrm{Card}(X \cap \overline{Y}) = s] = e^{-\lambda} \frac{\lambda^s}{s!}$

En conséquence, la probabilité pour que le hasard conduise, sous l'hypothèse d'absence de lien a priori entre a et b, a plus de contre-exemples que ceux qui ont été observés est :

$$\Pr\left[\mathrm{Card}(X \cap \overline{Y}) \leq \mathrm{Card}(A \cap \overline{B})\right] = \sum_{s=0}^{\mathrm{Card}(A \cap \overline{B})} \frac{\lambda^s}{s!} . e^{-\lambda}$$

Mais d'autres processus légitimes de tirage conduisent à une loi binomiale, voire une loi hypergéométrique (elle-même non sémantiquement adaptée à la situation en raison de sa symétrie). Dans des conditions de convergence convenables, ces deux lois se ramènent finalement à la loi de Poisson ci-dessus.

Dans le cas où $n_{\overline{b}} \neq 0$, nous réduisons et centrons cette variable de Poisson en la variable :

$$Q(a,\overline{b}) = \frac{\mathrm{Card}(X \cap \overline{Y}) - \frac{n_a n_{\overline{b}}}{n}}{\sqrt{\frac{n_a n_{\overline{b}}}{n}}}$$

Dans la réalisation expérimentale, la valeur observée de $Q(a,\overline{b})$ est $q(a,\overline{b})$. Elle estime un écart entre la contingence ($\mathrm{card}(A \cap \overline{B})$) et la valeur qu'elle aurait prise s'il y avait eu indépendance entre *a* et *b*.

---

5 Nous notons par la suite $\overline{v}$ la variable négation de v (ou non v) et $\overline{P}$ la partie complémentaire de la partie P de E.

Définition 2: $q(a,\bar{b}) = \dfrac{n_{a\wedge\bar{b}} - \dfrac{n_a n_{\bar{b}}}{n}}{\sqrt{\dfrac{n_a n_{\bar{b}}}{n}}}$

est appelé <u>indice d'implication,</u> nombre retenu comme indicateur de la non-implication de a sur b.

Dans les cas légitimant convenablement l'approximation (par exemple, $\frac{n_a n_{\bar{b}}}{n} \geq 4$), la variable $Q(a,\bar{b})$ suit approximativement la loi normale centrée réduite. <u>L'intensité d'implication</u>, <u>qualité de l'admissibilité</u> de $a \Rightarrow b$, pour $n_a \leq n_b$ et $n_b \neq n$, est alors définie à partir de l'indice $q(a,\bar{b})$ par :

Définition 3 :
L'<u>intensité d'implication</u> qui mesure la qualité inductive de a sur b est :

$\varphi(a,b) = 1 - \Pr[Q(a,\bar{b}) \leq q(a,\bar{b})] = \dfrac{1}{\sqrt{2\pi}} \int_{q(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt$ si $n_b \neq n$, $\varphi(a,b) = 0$ si $n_b = n$

Par suite, la définition de l'implication statistique devient :

Définition 4 :
L'implication $a \Rightarrow b$ est admissible au niveau de confiance 1-α si et seulement si : $\varphi(a,b) \geq 1 - \alpha$

Rappelons que cette modélisation de la quasi-implication mesure <u>l'étonnement de constater la petitesse des contre-exemples en regard du nombre surprenant des instances de l'implication</u>. C'est une mesure de la qualité <u>inductive</u> et <u>informative</u> de l'implication. Par conséquent, si la règle est triviale, comme dans le cas où B est très grand ou coïncide avec E, cet étonnement devient petit. Nous démontrons (Gras R., 1996) d'ailleurs que cette trivialité se traduit par une intensité d'implication très faible, voire nulle : *Si, n_a étant fixé et A étant inclus dans B, n_b tend vers n (B "croît" vers E), alors $\varphi(a,b)$ tend vers 0.* Nous définissons donc, par « continuité »: $\varphi(a,b) = 0$ si $n_b = n$. De même, si $A \subset B$, $\varphi(a,b)$ peut être inférieure à 1 dans le cas où la confiance inductive, mesurée par l'étonnement statistique, est insuffisante.

<u>Remarque 1</u> : La quasi-implication, d'indice $q(a,\bar{b})$ non symétrique, ne coïncide pas avec le coefficient de corrélation $\rho(a,b)$ qui est symétrique et qui rend compte de la liaison entre les variables a et b. En effet, nous démontrons que si $q(a,\bar{b}) \neq 0$ alors $\dfrac{\rho(a,b)}{q(a,\bar{b})} = -\sqrt{\dfrac{n}{n_b n_{\bar{a}}}}$

<u>Remarque 2</u> : Rappelons que nous considérons non seulement des conjonctions de variables du type "a et b" mais également des disjonctions comme "(a et b) ou c..." afin de modéliser les phénomènes qui relèvent de concepts comme il est fait en apprentissage ou en intelligence artificielle. Les calculs associés restent compatibles avec la logique des propositions reliées par des connecteurs.

<u>Remarque 3</u> : Contrairement à l'indice de Loevinger (Loevinger J. 1947) et à la probabilité conditionnelle (Pr[B/A])[6] et tous ses dérivés, l'intensité d'implication varie, non linéairement, avec la dilatation des ensembles E, A et B et s'affaiblit avec la trivialité (cf. Définition 3). De plus, elle résiste aux bruits, en particulier au voisinage de 0 pour $n_{a\wedge\bar{b}}$, ce qui ne peut que rendre statistiquement crédible la relation que nous voulons modéliser et établir. Enfin, on l'a vu, l'inclusion de A dans B n'assure pas une intensité maximale, la qualité inductive peut ne pas être forte, alors qu'au contraire Pr[B/A] est égale à 1 (Gras et al., 2004 a et cf. *Quality Measures in Data Mining*, Guillet F. and Hamilton H.-J., éditeurs). Dans le paragraphe qui suit, nous étudions de plus près le problème de la sensibilité et de la stabilité de l'indice d'implication en fonction de faibles variations des paramètres dans $q(a,\bar{b})$ en jeu par l'étude de sa différentielle.

---

[6] Comme nous l'avons dit, la probabilité conditionnelle représente la fonction la plus classique servant à la confirmation inductive d'une règle.

Remarque 4: D'autres modélisations, autres que celle de Poisson, sont possibles. Citons :

* *une modélisation binomiale* : considérant les variables duales card($A \cap \overline{Y}$) et card($X \cap \overline{B}$), où X et Y sont des parties choisies de façon indépendante dans E et respectant les propriétés cardinales respectives de A et B, tout élément de E, par exemple, a la probabilité $\frac{na}{n}\frac{n\overline{b}}{n}$ d'appartenir à $A \cap \overline{Y}$. Par suite :

$$\Pr\left[\text{card}(A \cap \overline{Y}) = k\right] = C_n^k \left(\frac{n_a n_{\overline{b}}}{n^2}\right)^k \left(1 - \frac{n_a n_{\overline{b}}}{n^2}\right)^{n-k} = \Pr\left[\text{card}(X \cap \overline{B}) = k\right]$$

* *une modélisation hypergéométrique* : on peut le voir rapidement en considérant encore les variables aléatoires card($A \cap \overline{Y}$) et card($X \cap \overline{B}$) où X et Y possèdent les mêmes propriétés cardinales respectives que A et B. On a, en effet :

$$\Pr[\text{card}(A \cap \overline{Y}) = k] = \frac{C_{na}^{k} C_{n-na}^{n-nb-k}}{C_n^{n-nb}} = \frac{n_a! n_{\overline{a}}! n_b! n_{\overline{b}}!}{k! n! (n_a - k)! (n_{\overline{b}} - k)! (n_{\overline{b}} - n_a - k)!}$$

$$= \frac{C_{n-n_b}^{k} C_{n_b}^{n_a - k}}{C_n^{n_a}} = \Pr[\text{card}(X \cap \overline{B}) = k]$$

Si la modélisation binomiale reste compatible avec la sémantique de l'implication, relation binaire non symétrique, il n'en est plus de même pour la modélisation hypergéométrique. Aussi, nous ne retiendrons que le modèle de Poisson et le modèle binomial.

2.4 Stabilité de l'indice d'implication

Etudier la stabilité de q, revient à examiner ses petites variations au voisinage des 4 valeurs entières observées (n, na, nb et $n_{a \wedge \overline{b}}$). Pour ce faire, il est possible d'effectuer différentes simulations en croisant ces 4 variables entières dont q dépend ((Fleury, 1996) ou (Gras et al, 2004a)). Mais, considérons ces variables comme nombres réels et q fonction continûment différentiable par rapport à ces variables contraintes à respecter les inégalités : na, nb et $n_{a \wedge \overline{b}} \leq \inf[na, nb])$ et $\sup[na, nb] \leq n$. Il suffit alors d'examiner la différentielle de q par rapport à ces variables et d'en conserver la restriction aux valeurs entières des paramètres de la relation $a \Rightarrow b$.

Différentielle de q

$$dq = \frac{\partial q}{\partial n} dn + \frac{\partial q}{\partial n_a} dn_a + \frac{\partial q}{\partial n_b} dn_b + \frac{\partial q}{\partial n_{a \wedge \overline{b}}} d\, n_{a \wedge \overline{b}}$$

Si l'on veut étudier comment varie q en fonction de $n_{\overline{b}}$, il suffit de remplacer n_b par $n\text{-}n_b$ et donc changer le signe de la dérivée de n_b dans la dérivée partielle.

En fait, l'intérêt de cette différentielle réside dans l'estimation de l'accroissement (positif ou négatif) de q, soit Δq, par rapport aux variations respectives Δn_a, $\Delta\, n_{\overline{b}}$ ou Δn_b et $\Delta\, n_{a \wedge \overline{b}}$. La sensibilité de l'indice q apparaît ainsi liée aux variations locales des 4 paramètres dont dépend la qualité de la règle.

Exemple : cas où seuls varient n_b et $n_{a \wedge \overline{b}}$ (dérivée partielle de n_a nulle)

$$\frac{\partial q}{\partial n_b} = 1/2\, n_{a \wedge \overline{b}} . \left(\frac{n_a}{n}\right)^{-1/2} (n\text{-}n_b)^{-3/2} + ½ . \left(\frac{n_a}{n}\right)^{1/2} (n\text{-}n_b)^{-1/2} > 0$$

$$\frac{\partial q}{\partial n_{a\wedge\bar{b}}} = \frac{1}{\sqrt{\frac{n_{a\wedge\bar{b}}}{n}}} > 0$$

Ainsi, si les accroissements Δn_b et $\Delta n_{a\wedge\bar{b}}$ .sont positifs, l'accroissement de $q(a,\bar{b})$ est également positif. Ceci s'interprète ainsi : si le nombre d'exemples de b et celui des contre-exemples de l'implication augmentent l'intensité d'implication diminue pour n et n_a constants. Autrement dit, cette intensité est maximum aux valeurs observées (n_b et $n_{a\wedge\bar{b}}$) et minimum aux valeurs $n_b + \Delta n_b$. et $n_{a\wedge\bar{b}} + \Delta n_{a\wedge\bar{b}}$.

Sensibilité de l'intensité d'implication

Considérons l'intensité d'implication φ comme fonction de $q(a,\bar{b})$:

$$\varphi(q) = \frac{1}{\sqrt{2\pi}} \int_q^{\infty} e^{-t^2/2} dt$$

On peut alors examiner comment $\varphi(q)$ varie lorsque q varie au voisinage d'une valeur donnée (a,b), sachant comment q varie lui-même en fonction des 4 paramètres qui le déterminent. Par dérivation de la borne d'intégration, on obtient :

$$\frac{d\varphi}{dq} = -\frac{1}{\sqrt{2\pi}} e^{-q^2/2} < 0$$

Ce qui confirme bien que l'intensité croît lorsque q décroît, mais la vitesse de croissance est précisée par la formule, ce qui permet d'étudier avec plus de précision les variations de $\varphi(q)$. La résistance de l'intensité d'implication aux bruits pourra donc être estimée à l'aide de ces différentes relations.

Exemple d'un autre indice de qualité de règles : stabilité de l'indice « confiance »

Cet indice c est le plus connu et, historiquement, après celui de J. Lovinger, le plus utilisé grâce à la connaissance commune de la une publication anglo-saxonne (Agrawal et al. 1993). Il est à l'origine de plusieurs autres indices communément employés qui n'en sont que des variantes satisfaisant telle ou telle exigence sémantique.. De plus, il est simple et s'interprète aisément et immédiatement.

$$c(a,b) = \frac{n_{a\wedge b}}{n_a} \text{ ou } 1 - \frac{n_{a\wedge\bar{b}}}{n_a}$$

La première forme s'interprète ainsi comme une fréquence conditionnelle des exemples de b quand a est connu.

La sensibilité de cet indice aux variations des occurrences des contre-exemples se lit avec la dérivée partielle :

$$\frac{\partial c}{\partial n_{a\wedge\bar{b}}} = -\frac{1}{n_a}$$

Par conséquent, la confiance croît quand $n_{a\wedge\bar{b}}$ décroît ce qui est sémantiquement acceptable, mais la vitesse de variation est constante, indépendante de la vitesse de décroissance de ce nombre, des variations de n et de n_b. Le gradient de c ne s'exprime que par rapport à $n_{a\wedge\bar{b}}$ et à n_a. Ceci peut apparaître comme une restriction du rôle des paramètres dans l'expression de la sensibilité de l'indice.

3 Cas des variables modales et fréquentielles

3.1 Situation fondatrice

La recherche de Marc Bailleul (1991-1994) porte en particulier, sur la représentation que se font les enseignants de mathématiques de leur propre enseignement. Afin de la mettre en évidence, des mots significatifs leur sont proposés qu'ils doivent hiérarchiser. Leurs choix ne sont donc plus binaires, les mots retenus par un enseignant quelconque sont ordonnés du moins au plus représentatif. L'interrogation de M. Bailleul se centre alors sur des questions du type : « si je choisis tel mot avec telle importance alors je choisis tel autre mot avec une importance au moins égale ». Il a donc fallu étendre la notion d'implication statistique à des variables autres

que binaires. C'est le cas des variables modales qui sont associées à des phénomènes où les valeurs a(x) sont des nombres de l'intervalle [0,1] et qui décrivent des degrés d'appartenance ou de satisfaction comme le sont en logique floue, par exemple, les modificateurs linguistiques "peut-être", "un peu", "quelquefois", etc.. Cette problématique se retrouve également dans des situations où la fréquence d'une variable traduit un préordre sur les valeurs attribuées par les sujets aux variables qui leur sont présentées. Il s'agit de variables fréquentielles qui sont associées à des phénomènes où les valeurs de a(x) sont des réels positifs quelconques. On trouve une telle situation lorsque l'on considère le pourcentage de réussite d'un élève à une batterie de tests portant sur des domaines distincts.

3.2 Formalisation

J.B. Lagrange (Lagrange J.B. 1998) a démontré que, dans le cas modal,

- si a(x) et $\bar{b}(x)$ sont les valeurs prises en x par les variables modales a et $\bar{b}$, avec $\bar{b}(x)=1-b(x)$
- si s_a^2 et $s_{\bar{b}}^2$ sont les variances empiriques des variables a et $\bar{b}$ alors l'indice d'implication, qu'il dénomme *indice de propension*, devient :

Définition 5 : $q(a,\bar{b}) = \dfrac{\sum_{x\in E} a(x)\bar{b}(x) - \dfrac{n_a n_{\bar{b}}}{n}}{\sqrt{\dfrac{(n^2 s_a^2 + n_a^2)((n^2 s_{\bar{b}}^2 + n_{\bar{b}}^2)}{n^3}}}$ est l'indice de propension de variables modales

J.B. Lagrange prouve également que cet indice coïncide avec l'indice défini précédemment dans le cas binaire et pour une modélisation de Poisson (cf. 2.3). Le nombre de modalités de a et de b est justement 2, car dans ce cas :

$n^2 s_a^2 + n_a^2 = n n_a$, $n^2 s_{\bar{b}}^2 + n_{\bar{b}}^2 = n n_{\bar{b}}$ et $\sum_{x\in E} a(x)\bar{b}(x) = n_{a\wedge\bar{b}}$.

Cette solution apportée au cas modal est aussi applicable au cas des *variables fréquentielles*, voire *des variables numériques positives*, à condition d'avoir normalisé les valeurs observées sur les variables, telles que a et b, la normalisation dans [0,1] étant faite à partir du maximum de la valeur prise respectivement par a et b sur l'ensemble E.

Depuis, dans (Gras et Régnier, 2007), nous montrons que, pour une autre modélisation de l'indice d'implication (modélisation binomiale), l'indice de propension coïncide avec l'indice $q(a,\bar{b})$ du cas binomial.

Remarque

Dans [Régnier J.C. et Gras R. 2005.], nous considérons des variables de rang qui traduisent un ordre total entre des choix présentés à une population de juges. Chacun d'entre eux doit ordonner son choix préférentiel parmi un ensemble d'objets ou de propositions qui lui sont faites. Un indice permet de mesurer la qualité d'énoncé du type : « si l'objet a est rangé par les juges alors, généralement, l'objet b est rangé à un rang meilleur par les mêmes juges ». La proximité avec la problématique précédente conduit à un indice relativement voisin de l'indice de Lagrange, mais mieux adapté à la situation de variable-rang.

4 Cas des variables-sur-intervalles et variables-intervalles

4.1 Variables-sur-intervalles

4.1.1 Situation fondatrice

On recherche, par exemple, à extraire d'un ensemble de données biométriques la règle suivante, en estimant sa qualité : « si un individu pèse entre 65 et 70 kg alors en général il mesure entre 1.70 et 1.76 m ». Une situation comparable se présente dans la recherche de relation entre des intervalles de performances d'élèves dans deux disciplines différentes. La situation plus générale s'exprime alors ainsi : deux variables réelles a et b prennent un certain nombre de valeurs sur 2 intervalles finis $[a_1, a_2]$ et $[b_1, b_2]$. Soit A (resp. B) l'ensemble des

valeurs de a (resp. b) observées sur $[a_1, a_2]$ (resp. $[b_1, b_2]$. Par exemple ici, a représente les poids d'un ensemble de n sujets et b les tailles de ces mêmes sujets

Deux problèmes se posent :

1° peut-on définir des sous-intervalles adjacents de $[a_1, a_2]$ (resp. $[b_1, b_2]$.) afin que la partition la plus fine obtenue respecte au mieux la distribution des valeurs observées dans $[a_1, a_2]$ (resp. $[b_1, b_2]$.) ?

2° peut-on trouver les partitions respectives de $[a_1, a_2]$ et $[b_1, b_2]$ constituées de réunions des sous-intervalles adjacents précédents, partitions qui maximisent l'intensité d'implication moyenne des sous-intervalles de l'un sur des sous-intervalles sur l'autre appartenant à ces partitions ?

Nous répondons à ces deux questions dans le cadre de notre problématique en faisant choix des critères à optimiser pour satisfaire l'optimalité attendue dans chaque cas. A la première question, de nombreuses solutions ont été apportées dans d'autres cadres (par exemple, par Lahanier-Reuter D. 1998).

4.1.2 Premier problème

On va s'intéresser à l'intervalle $[a_1, a_2]$ en le supposant muni d'une partition initiale triviale de sous-intervalles de même longueur, mais pas nécessairement de même distribution des fréquences observées sur ces sous-intervalles.

Notons $P_0 = \{A_{01}, A_{02},..., A_{0p}\}$, cette partition en p sous-intervalles. On cherche à obtenir une partition de $[a_1, a_2]$ en p sous-intervalles $A_{q1}, A_{q2},..., A_{qp}$ de telle façon qu'au sein de chaque sous-intervalle on ait une bonne homogénéité statistique (faible inertie intra-classe) et que ces sous-intervalles présentent une bonne hétérogénéité mutuelle (forte inertie inter-classe). On sait que si l'un des critères est vérifié l'autre l'est nécessairement. (théorème de Koenig-Huyghens). Pour ce faire, on adoptera une méthode directement inspirée de la méthode des nuées dynamiques conçue par Edwin Diday (Diday E., 1972)(voir aussi Lebart L. et al. 2000) et adaptée à la situation présente. On obtient ainsi la partition optimale visée.

4.1.3 Deuxième problème

On suppose maintenant que les intervalles $[a_1, a_2]$ et $[b_1, b_2]$ sont munis de partitions optimales P et Q, respectivement, au sens des nuées dynamiques. Soit p et q les nombres respectifs de sous-intervalles composant P et Q. A partir de ces deux partitions, il est possible d'engendrer 2^{p-1} et 2^{q-1} partitions obtenues par réunions itérées de sous-intervalles adjacents respectivement de P et de Q [7].

On calcule les intensités d'implication respectives de chaque sous-intervalle réuni ou non à un autre de la première partition sur chaque sous-intervalle réuni ou non à un autre de la seconde, puis les valeurs des intensités des implications réciproques.

Il y a donc au total $2.2^{p-1}.2^{q-1}$ familles d'intensités d'implication, chacune d'entre elles nécessitant le calcul de tous les éléments d'une partition de $[a_1, a_2]$ sur tous les éléments d'une des partitions de $[b_1, b_2]$ et réciproquement.

On choisit comme *critère d'optimalité* la moyenne géométrique des intensités d'implication, moyenne associée à chaque couple de partitions d'éléments, réunis ou non, définies inductivement. On note les deux maxima obtenus (implication directe et sa réciproque) et on retient les deux partitions associées en déclarant que l'implication de la variable-sur-intervalle a sur la variable-sur-intervalle b est optimale lorsque l'intervalle $[a_1, a_2]$ admet la partition correspondant au premier maximum et que l'implication réciproque optimale est satisfaite pour la partition de $[b_1, b_2]$ correspondant au deuxième maximum.

---

7 Il suffit de considérer l'arborescence dont A1 est la racine, puis de le réunir ou non à A2 qui lui-même sera ou non réuni à A3, etc. Il y a donc 2p-1 branches dans cette arborescence.

4.2 Variables-intervalles

4.2.1 Situation fondatrice

On dispose des données fournies par une population de n individus (qui peuvent être chacun ou certains des ensembles d'individus, par ex. une classe d'élèves) selon p variables (par ex. notes sur une année en français, math, physique, …, mais aussi bien : poids, tailles, tour de poitrine, …). Les valeurs prises par ces variables selon chaque individu sont des intervalles de réels positifs. Par exemple, l'individu x donne la valeur [12 ; 15.50] à la variable note de math. E. Diday parlerait à ce sujet de p variables symboliques à valeurs intervalles définies sur la population.

On cherche à définir une implication d'intervalles, relatifs à une variable a, constitués eux-mêmes des intervalles observés, vers d'autres intervalles pareillement définis et relatifs à une autre variable b. Ceci permettra de mesurer l'association implicative, donc non symétrique, de certain(s) intervalle(s) de la variable a avec certain(s) intervalle(s) de la variable b, ainsi que l'association réciproque à partir de laquelle on retiendra la meilleure pour chaque couple de sous-intervalles en jeu, comme il vient d'être fait au § 4.1.

Par exemple, on dira que le sous-intervalle [2 ; 5,5] de notes de mathématiques implique généralement le sous-intervalle [4,25 ; 7,5] de notes de physique, ces deux sous-intervalles appartenant à une partition optimale au sens de la variance expliquée des intervalles respectifs de valeurs [1 ; 18] et [3 ; 20] prises dans la population. De même, on dira que [14,25 ; 17,80] en physique implique le plus souvent [16,40 ; 18] en mathématiques.

4.2.2 Algorithme

En suivant la problématique de E. Diday et ses collaborateurs, si les valeurs prises selon les sujets par les variables a et b sont de nature symbolique, en l'occurrence des intervalles de R^+, il est possible d'étendre les algorithmes ci-dessus [Gras R. 17 c, 2001]. Par exemple, à la variable a sont associés des intervalles de poids et à la variable b des intervalles de tailles, intervalles dus à une imprécision des mesures. Effectuant la réunion des intervalles I_X et J_X décrits par les sujets x de E selon respectivement chacune des variables a et b, on obtient deux intervalles I et J recouvrant toutes les valeurs possibles de a et de b. Sur chacun d'eux on peut définir une partition en un certain nombre d'intervalles respectant comme plus haut un certain critère d'optimalité. Pour cela, les intersections des intervalles tels que I_X et J_X avec ces partitions seront munies d'une distribution prenant en compte les étendues des parties communes. Cette distribution peut être uniforme ou d'un autre type discret ou continu. Mais ainsi, nous sommes ramenés à la recherche de règles entre deux ensembles de variables-sur-intervalles qui prennent comme précédemment dans le § 4.1, leurs valeurs sur [0,1] à partir desquelles on pourra chercher les implications optimales.

Remarque Quel que soit le type de variable considéré, se pose souvent le problème de surabondance des variables et donc de difficulté de représentation. C'est pour cette raison que nous avons défini une relation d'équivalence sur l'ensemble des variables qui nous permet de substituer à une classe d'équivalence une variable dite leader ([Gras et al. 02], [Couturier et al 04]).

5 L'implication-inclusion

5.1 Situation fondatrice et problématique

Deux raisons nous ont conduits à améliorer le modèle formalisé par l'intensité d'implication:

- lorsque la taille des échantillons traités, et en particulier celui de E, croît (de l'ordre du millier et plus), l'intensité $\varphi(a,b)$ a tendance à ne plus être suffisamment discriminante car ses valeurs peuvent être très voisines de 1, alors que l'inclusion dont elle cherche à modéliser la qualité, est loin d'être satisfaite (phénomène signalé dans (Bodin A. 1997) qui traite des grandes populations d'élèves à travers des enquêtes internationales) ;
- le modèle de la quasi-implication précédent retient essentiellement la mesure de la force de la règle $a \Rightarrow b$. Or, la prise en compte d'une concomitance de non b => non a (contraposée de l'implication) est indispensable pour renforcer l'affirmation d'une bonne qualité de la relation quasi-implicative, voire quasi-

causale, de a sur b [8]. En même temps, elle pourrait permettre de corriger la difficulté évoquée ci-dessus (si A et B sont petits par rapport à E, leurs complémentaires seront importants et réciproquement).

5.2 Un indice d'inclusion

La solution[9] que nous apportons utilise à la fois l'intensité d'implication et un autre indice qui rend compte de la dissymétrie entre les situations S_1 = (a et b) et S'_1 (a et non b), ((resp. S_2 = non a et non b) et S'_2 = (a et non b)) en faveur de la première nommée. La faiblesse relative des instances qui contredisent la règle et sa contraposée est ainsi fondamentale. D'ailleurs, le nombre de contre-exemples $n_{a\wedge\bar{b}}$ à $a \Rightarrow b$ est celui à la contraposée. Pour rendre compte de l'incertitude liée à un éventuel pari de l'appartenance à une des deux situations (S_1 ou S'_1, (resp. S_2 ou S'_2)), c'est donc au concept d'entropie de Shannon (1949) que nous faisons référence :

$$H(b/a) = -\frac{n_{a\wedge b}}{n_a}\log_2\frac{n_{a\wedge b}}{n_a} - \frac{n_{a\wedge\bar{b}}}{n_a}\log_2\frac{n_{a\wedge\bar{b}}}{n_a},$$

est l'entropie conditionnelle relative aux cases (a et b) et (a et non b) lorsque a est réalisée

$$H(\bar{a}/\bar{b}) = -\frac{n_{a\wedge\bar{b}}}{n_{\bar{b}}}\log_2\frac{n_{a\wedge\bar{b}}}{n_{\bar{b}}} - \frac{n_{\bar{a}\wedge\bar{b}}}{n_{\bar{b}}}\log_2\frac{n_{\bar{a}\wedge\bar{b}}}{n_{\bar{b}}}$$

est l'entropie conditionnelle relative aux cases (non a et non b) et (a et non b) lorsque non b est réalisée.

Ces entropies, à valeurs dans [0,1], devraient donc être simultanément faibles et donc les dissymétries entre les situations S_1 et S'_1 (resp. S_2 et S'_2) devraient être simultanément fortes si l'on souhaite disposer d'un bon critère d'inclusion de A dans B. En effet, les entropies représentent l'incertitude moyenne des expériences qui consistent à observer si b est réalisé (resp. si non a est réalisé) lorsque l'on a observé a (resp. non b). Le complément à 1 de cette incertitude représente donc *l'information* moyenne recueillie par la réalisation de ces expériences. Plus cette information est importante, plus forte est la garantie de la qualité de l'implication et de sa contraposée. Nous devons maintenant adapter ce critère numérique entropique au modèle attendu dans les différentes situations cardinales.

Pour que le modèle ait la signification attendue, il doit satisfaire, selon nous, les contraintes épistémologiques suivantes :

1° il devra intégrer les valeurs de l'entropie et, pour les contraster, par exemple, intégrer ces valeurs au carré,

2° comme ce carré varie de 0 à 1, afin de dénoter le déséquilibre et donc l'inclusion, afin de s'opposer à l'entropie, la valeur retenue sera le complément à 1 de son carré tant que le nombre de contre-exemples sera inférieur à la moitié des observations de a (resp. de non b). Au delà de ces valeurs, les implications n'ayant plus de sens inclusif, on affectera au critère la valeur 0,

3° afin de prendre en compte les deux informations propres à $a \Rightarrow b$ et non $b \Rightarrow$ non a, le produit rendra compte de la qualité simultanée des valeurs retenues. Le produit a la propriété de s'annuler dès que l'un de ses termes s'annule, i.e. dès que cette qualité s'efface,

4° enfin, le produit ayant une dimension 4 par rapport à l'entropie, sa racine quatrième sera de la même dimension.

Posons $\alpha = \frac{n_a}{n}$, la fréquence de a et $\bar{\beta} = \frac{n_{\bar{b}}}{n}$, la fréquence de non b. Notons, en fonction de la fréquence t $= \frac{n_{a\wedge\bar{b}}}{n}$ de contre-exemples, les deux termes significatifs des qualités respectives de l'implication et sa contraposée :

$h_1(t) = H(b/a) = -(1-\frac{t}{\alpha})\log_2(1-\frac{t}{\alpha}) - \frac{t}{\alpha}\log_2\frac{t}{\alpha}$ si $t \in [0, \frac{\alpha}{2}[$ et $h_1(t) = 1$ si $t \in [\frac{\alpha}{2}, a]$

$h_2(t) = H(\bar{a}/\bar{b}) = -(1-\frac{t}{\bar{\beta}})\log_2(1-\frac{t}{\bar{\beta}}) - \frac{t}{\bar{\beta}}\log_2\frac{t}{\bar{\beta}}$ si $t \in [0, \frac{\bar{\beta}}{2}[$ et $h_2(t) = 1$ si $t \in [\frac{\bar{\beta}}{2}, \bar{\beta}]$

8 Ce phénomène est signalé par Y.Kodratoff dans son article publié dans les Actes du Colloque « Fouille dans les données par la méthode implicative », IUFM de Caen, juin 2000.

9 J.Blanchard apporte dans (Blanchard J. et al. 2005) une réponse à ce problème par une mesure de « l'écart à l'équilibre ».

D'où la définition permettant de déterminer le critère entropique :

Définition 6 : l'indice d'inclusion de A, support de a, dans B, support de b, est le nombre :

$i(a,b) = [(1-h_1^2(t))(1-h_2^2(t))]^{1/4}$

qui intègre l'information délivrée par la réalisation d'un faible nombre de contre-exemples, d'une part à la règle a => b et, d'autre part, à la règle non b $\Rightarrow$ non a

5.3 L'indice d'implication-inclusion

L'intensité d'implication-inclusion *(ou intensité entropique)*, nouvelle mesure de la qualité inductive, est le nombre :

$\psi(a,b) = [i(a,b).\ \varphi(a,b)]^{1/2}$

qui intègre à la fois l'étonnement statistique et la qualité inclusive.

La fonction ψ suivant la variable t admet une représentation qui a la forme indiquée par la figure 2, pour n_a et n_b fixés. On remarquera sur la figure 2 la différence de comportement de la fonction par rapport à la probabilité conditionnelle P(B/A), indice fondamental des autres modélisations de la mesure des règles, par exemple chez Agrawal et son école. Outre son caractère linéaire, donc peu nuancé, cette probabilité conduit à une mesure qui décroît trop vite dès les premiers contre-exemples et résiste ensuite trop longuement lorsque ceux-ci deviennent importants.

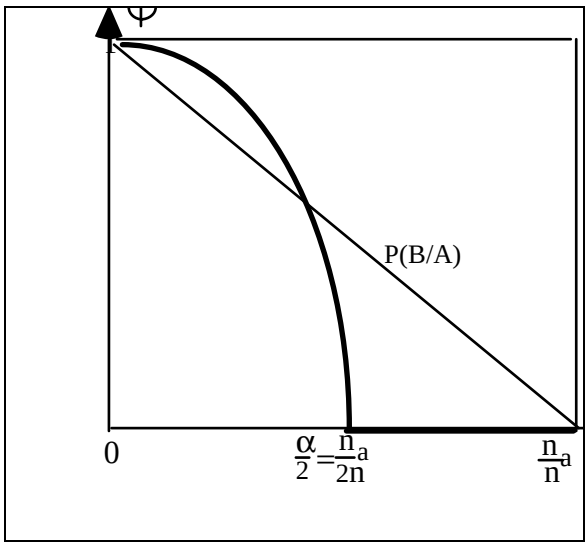


FIG 2

On constate que cette représentation de fonction continue de t traduit les propriétés attendues du critère d'inclusion :

* "réaction" lente aux premiers contre-exemples (résistance au bruit),

* "accélération" du rejet de l'inclusion au voisinage de l'équilibre soit $\frac{n_a}{2n}$,

* rejet au-delà de $\frac{n_a}{2n}$ ce que n'assurait pas l'intensité d'implication $\varphi(a,b)$.

Exemple 1

	b	$\bar{b}$	marge
a	200	400	600
$\bar{a}$	600	2800	3400
marge	800	3200	4000

L'intensité d'implication est $\varphi(a,b) = 0{,}9999$ (pour $q(a,\bar{b}) = -3{,}65$)

Les valeurs entropiques de l'expérience sont $h_1 = 0 = h_2$

La valeur du coefficient modérateur est donc : $i(a,b) = 0$

Par suite $\psi(a,b) = 0$ alors que $P(B/A) = 0{,}33333$

Ainsi, les fonctions "entropiques" "modérent" l'intensité d'implication dans ce cas où justement l'inclusion est médiocre.

Exemple 2

	b	$\bar{b}$	marge
a	400	200	600
$\bar{a}$	1000	2400	3400
marge	1400	2600	4000

L'intensité d'implication est 1 (pour $q(a, \bar{b})$ = - 8,43)
Les valeurs entropiques de l'expérience sont h_1 = 0,918 h_2 = 0,391
La valeur du coefficient modérateur est donc : i(a,b) = 0,6035
Par suite ψ (a,b) = 0,777 alors que P(B/A) = 0,66666

Remarque

La correspondance entre φ(a,b) et ψ(a,b) n'est pas monotone comme le montre le deuxième exemple suivant :

	b	$\bar{b}$	marge
a	40	20	60
$\bar{a}$	60	280	340
marge	100	300	400

L'intensité d'implication est inférieure à la précédente car: $q(a, \bar{b})$ = - 6,47
Les valeurs entropiques sont : h_1= 0,918, h_2 = 0,353
La valeur du coefficient modérateur est : i(a,b)= 0,608
Donc ψ (a,b) = 0,78 alors que P(B/A) = 0,66666

Ainsi, alors que φ(a,b) a décru du 1er au 2ème exemples, i(a,b) a crû de même que ψ (a,b). En revanche, la situation contraire est la plus fréquente. Notons que, dans les deux cas, la probabilité conditionnelle ne change pas.

Remarque

Nous renvoyons à (Lenca P. et al 2004) pour une étude comparative, très fouillée, des indices d'association pour des variables binaires. En particulier, les intensités d'implication classique et entropique (inclusion) présentées dans cet article sont confrontées à d'autres indices selon une entrée « utilisateur ».

6 Graphe d'implication

6.1 Problématique

A l'issue des calculs des intensités d'implication (classique ou entropique), nous disposons certes d'un tableau qui croise les variables, quelle que soit leur nature, et dont les éléments sont des nombres de [0 ;1]. Mais la structure sous-jacente entre ces variables est inapparente. L'utilisateur est aveugle face à un tableau carré qui peut présenter un nombre très grand de lignes et colonnes. Il ne peut embrasser simultanément l'enchaînement éventuel des règles qui sous-tendent la structure. Nous avons donc associé à ce tableau un graphe orienté, pondéré (par les intensités d'implication), sans cycle dont l'utilisateur peut contrôler la complexité de la représentation en fixant un seuil de prise en compte de la qualité implicative des règles.

6.2 Algorithme

La relation définie par l'implication statistique, si elle est réflexive et non symétrique, n'est pas **transitive** bien évidemment, comme l'induction et au contraire de la déduction. Or nous voulons qu'elle modélise la relation d'ordre partiel entre deux variables (les réussites dans notre exemple initial). Par convention, si a => b et si b => c, nous accepterons la fermeture transitive a => c seulement si ψ (a,c) $\geq$ 0,5, c'est-à-dire si la relation implicative de a sur c est meilleure que la neutralité en soulignant la dépendance entre a et c.

Par exemple, supposons qu'entre les 7 variables *a, b, c, d, e* et *f* existent, au seuil supérieur à 0,5, les règles suivantes : *e => c, a, f, b* ; *c => a, f* ;*b=> a, f* ; *g => d, f* ; *a=> f.*

On pourra alors traduire cet ensemble de relations par le graphe suivant[10] :

10 Les traitements automatiques des calculs et des graphiques sont exécutés à l'aide du logiciel C.H.I.C. (Classification Hiérarchique Implicative et Cohésitive) disponible sous Windows 95, 98, NT et XP. Ce logiciel, à partir d'une première version établie par R.Gras et H.Rostam, révisée sous Pascal par S.Ag Almouloud, [Ag Almouloud S. 1992], est maintenant développé par R.Couturier [Couturier et al. 2005] et constamment étendu par lui aux nouveaux concepts et nouveaux algorithmes.

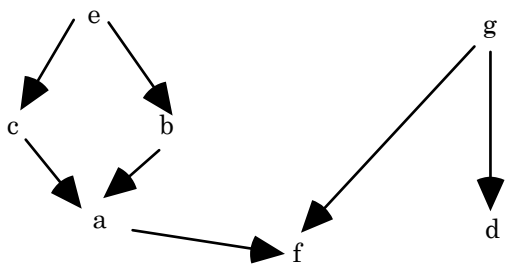


FIG. 3

Le graphe peut être réduit (resp. élargi) si nous élevons (resp. abaissons) le seuil de considération des règles. Notons que ce graphe n'est pas un treillis puisque, par exemple la variable a n'implique pas la variable (a ou non a) dont le support est E. A fortiori, ce ne peut être un treillis de Galois. Des options de CHIC permettent de supprimer à volontiers des variables, de déplacer leur image dans le graphe afin de décroiser les arcs ou de se centrer sur certaines variables sommets d'une sorte de « cône » dont les deux « nappes » sont constituées respectivement des variables « parents » et des variables « enfants ».

7 Implication entre règles et méta-règles ou R-règles

7.1 Situation fondatrice

En didactique des mathématiques, l'une des questions[11] que se pose le didacticien est de reconnaître, puis d'identifier la source des obstacles tant didactiques qu'épistémologiques, auxquels se heurte l'élève au cours de ses apprentissages. Or ces obstacles se fondent sur le terreau des conceptions que l'élève se construit. Elles sont elles-mêmes structurées, par des règles simples ou complexes, en un tout qui n'en soit pas la concaténation, mais qui soit une sorte de micro-théorie cognitive constituée, comme l'est une théorie mathématique d'axiomes, de théorèmes, de corollaires, de théorèmes de théorèmes, etc.. Cette structure, qui permet de dépasser le simple assemblage des parties afin d'atteindre un tout signifiant, n'est pas de nature hiérarchique classique symétrique où les classes de similarité entre variables se retrouvent emboîtées selon des partitions, des typologies de plus en plus grossières. Elle se présente plutôt selon trois schémas :

1. une règle de type $a \Rightarrow b$ (où a et b peuvent être elles-mêmes des règles) implique une variable c. c peut alors s'interpréter comme une conséquence de la règle,

2. une variable a implique une règle $b \Rightarrow c$ (où b et c peuvent être des règles). La règle $a \Rightarrow (b \Rightarrow c)$ peut se lire comme : de l'observation de a, on déduit la règle $b \Rightarrow c$. L'intuition peut être soutenue par le détour formel (a et b) $\Rightarrow$ c,

3. une règle $a \Rightarrow b$ implique une autre règle $c \Rightarrow d$. Cette situation est tout à fait comparable à la situation où un théorème a pour corollaire un autre théorème ou bien un lemme conduit à une proposition.

La structure qui paraît répondre à cette attente est donc hiérarchique, comme nous le montrerons plus loin, mais elle se doit d'être non symétrique, orientée. C'est cette structure que nous cherchons à formaliser maintenant. Elle doit permettre de passer au niveau épistémologique supérieur, celui des opérations sur les règles, sous forme de méta-règles ou règles généralisées (R-règles). On retrouve dans la structure visée des éléments en tout point comparables aux structures cognitives qui rendent compte de la genèse de la connaissance opératoire, au sens de Piaget où l'on passe d'un niveau à l'autre par abstraction réfléchissante : on s'élève d'une représentation des objets à celle des opérations sur les objets, puis aux opérations sur les opérations. C'est encore le cas en mathématiques où l'on passe des objets élémentaires mathématiques aux fonctions sur ces objets, puis à celle des fonctions sur les fonctions. Si par exemple en psychologie ou en sociologie, les variables en jeu sont des comportements, les règles généralisées (règles de règles) seront des profils comportementaux, des conduites générales. C'est donc un point de vue hiérarchique dynamique à l'opposé d'un point de vue statique, comme l'est une typologie. Plus encore, ce n'est pas dans la simple description individuelle des règles constituant la règle généralisée que s'exprime la propriété **émergente** du tout, propriété suradditive, presque étrangère à la « somme » de ses descripteurs. C'est de l'analyse de la structure de ce tout que s'interprète cette propriété,

11 Nous avons pour la première fois mathématisé cette question dans la thèse de (Larher A. 1991) pour structurer et hiérarchiser des raisonnements d'élèves en situation de preuve mathématique

qu'elle soit un profil, une conduite, une genèse ou un autre système dynamique non linéaire[12] (cf. pour un exemple (Gras R. et Kuntz P., 2005))

7.2 **Formalisation** (pour une formalisation détaillée voir (Gras R.et Kuntz P. 2005)) et (Kuntz P. actes ASI 05)

Les règles et règles généralisées à élaborer feront référence à des variables se structurant de façon ascendante en classes emboîtées et orientées. Mais une règle entre classes de variables ne prend véritablement son sens qu'à condition qu'à l'intérieur de chaque classe de variables dont on examine la relation avec d'autres, existe une certaine "**cohésion**" entre les variables qui la constituent ; ceci devant se faire en respectant l'ordre institué dans la classe. On souhaite ainsi que le "flux" implicatif d'une classe ***A*** sur une classe ***B*** soit nourri d'un "flux" interne à ***A*** et alimente un "flux" interne à ***B*** (ce mot *flux* est choisi pour sa connotation métaphorique hydraulique ou thermodynamique). Pour cela, le concept d'entropie H permettant de rendre compte du désordre entre des variables, nous définissons la cohésion entre deux variables par l'entropie de l'expérience à deux issues : implication ou non implication :

Définition 7 : La cohésion de la classe (a,b) est le nombre c(a,b) tel que :

. si $p = \psi(a,b)$ et $H = -p\log_2 p-(1-p)\log_2(1-p)$, alors $coh(a,b) = \sqrt{1-H^2}$

. si p = 1, alors coh(a,b) = 1

. si $p \leq 0{,}5$, alors coh(a,b) = 0

Intuitivement, la cohésion mesure le déséquilibre des occurrences des événements $\bar{a} \vee b$ et $a \wedge \bar{b}$ en faveur du premier.

Définition 8 : La cohésion de la classe ordonnée de variables ***A*** = (a_1,....,a_r) est alors définie par extension :

$$Coh(\underline{\boldsymbol{A}}) = \left(\prod_{\substack{i\in\{1,\ldots,r-1\} \\ j\in\{2,\ldots,r\}, j>i}} c(a_i, a_j) \right)^{\frac{2}{r(r-1)}}$$

C'est la moyenne géométrique des cohésions de couples qui, en tant que telle, s'annule dès que l'une des cohésions en jeu s'annule. A la cohésion correspond une distance ultramétrique qui permet d'indicer la hiérarchie et en justifie, a posteriori, la désignation. Cette cohésion permet en effet de l'indicer par une fonction positive. Il suffit de choisir pour une paire *(x,y)* réunie au premier niveau *d(x,y)=1-max [c(x,y), c(y,x)]*, puis *d(x,y)=[1-c($h_{x,y}$)]* où *c($h_{x,y}$)* est la cohésion de la plus petite classe $h_{x,y}$ contenant *x* et *y*. On a bien *d(x,x)=0* et la distance maximum est $\leq$ *1*. On démontre que *d* est ultramétrique (maximale inférieure), ce qui justifie le terme de hiérarchie (Gras et al. 2003 et Gras et al 2005). La construction automatique de cette hiérarchie est obtenue dans CHIC.

Enfin nous pouvons modéliser l'implication statistique d'une classe de variables sur une autre classe en exigeant du modèle qu'il intègre les informations suivantes :

- les cohésions respectives des 2 classes,
- une intensité d'implication extrémale des éléments d'une classe sur les éléments de l'autre,
- les cardinaux respectifs des 2 classes.

Chacune de ces informations crédite l'indice que nous retiendrons si :

- l'indice croît avec les cohésions de chaque classe et s'annule lorsque la cohésion de l'une d'entre elles est nulle,
- l'indice croît avec la liaison extrémale (minimale si l'on vise un degré d'exigence élevé, maximale si l'on recherche une souplesse réaliste),
- l'indice décroît avec les cardinaux des classes, eu égard à la prise en compte d'une liaison maximale.

Par suite, notant ***A*** et ***B*** des classes de variables d'éléments génériques a_i et b_j, puis Coh(***A***) et Coh(***B***) leurs cohésions respectives, l'intensité d'implication de ***A*** sur ***B*** est donnée par :

---

12 « C'est ainsi que sont privilégiés maintenant, comme dans nos réseaux d'automates, les processus dynamiques par rapport aux états, les procédures délocalisées, et en partie stochastiques, de créations de significations par rapport aux représentations ». H.Atlan, déjà cité

Définition 9 : L'intensité d'implication de ***A*** sur ***B*** est :

$$\psi(\underline{\boldsymbol{A}},\underline{\boldsymbol{B}}) = \left[\sup_{i\in\{1,\ldots,r\},\ j\in\{1,\ldots,s\}} \psi(a_i,b_j)\right]^{rs} .[\mathrm{Coh}(\underline{\boldsymbol{A}}).\mathrm{Coh}(\underline{\boldsymbol{B}})]^{1/2}$$

On pourra constater que cet indice satisfait les contraintes sémantiques déclarées ci-dessus.

Définissant une méthode de classification descendante classique par le critère de cohésion décroissante, , on obtiendra par exemple des arbres comme celui-ci :

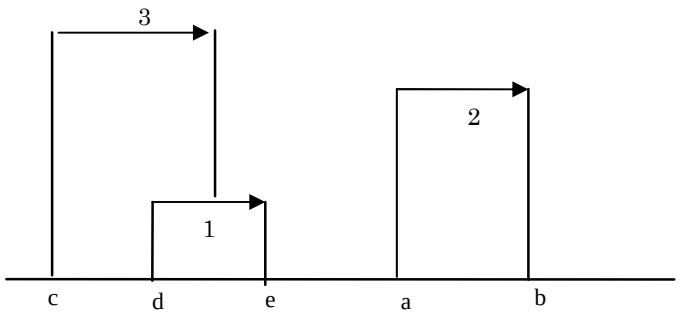


Fig. 4

Des interprétations de telles règles généralisées sont plus complexes, comme par exemple, illustrée par la figure 4, la règle $c \Rightarrow (d \Rightarrow e)$. Quelques règles généralisées sont réductibles à des assemblages aisément interprétables. Par exemple, $c \Rightarrow (d \Rightarrow e)$ se ramène logiquement à $c \wedge d \Rightarrow e$.

7.3 Application

Dans le cadre d'une enquête de l'Association des Professeurs de Mathématiques de l'Enseignement Public (APMEP) auprès de professeurs de mathématiques de classes terminales (séries scientifiques S et ES, littéraires LI et technologiques TE sont les variables supplémentaires), nous avons recueilli et analysé (Bodin et al., 1999) les réponses de 311 professeurs, à des classements (de 1 à 6) portant sur quinze objectifs qu'ils assignent à leur enseignement (A, B, C, …O)[13] et sur leurs opinions relatives à dix phrases susceptibles d'être communément énoncées (OP1, OP2,..OPX)[14]. La variable PER donne la possibilité d'énoncer les objectifs jugés non pertinents. Les 26 variables correspondantes ne sont pas binaires, sauf PER, mais ordinales (valeurs (1, 0.8, 0.6, 0.4, 0.2, 0.1, 0) pour les objectifs et (1, 0.5, 0) pour les opinions). Ainsi l'analyse intègre l'intensité des attitudes, d'un choix prioritaire d'un objectif à un choix plus secondaire, voire non retenu.

Les occurrences des 26 variables sont les suivantes :

A : 105.70 B : 8.80 C : 9.70 D : 140.00 E : 21.80 F : 138.70 G : 19.50 H : 44.80
I I : 83.10 J : 108.40 K : 77.60 L : 4.60 M : 90.20 N : 66.60 O : 33.20
OP1 : 81.50 OP2 : 147.50 OP3 : 242.50 OP4 : 229.00 OP5 : 190.00 OP6 : 240.00 OP7 : 200.00 OP8: 165.00 OP9 : 98.00 OPX : 207.00 PER : 254.

La hiérarchie orientée obtenue structure les 26 variables en plusieurs classes qui définissent des R-règles de longueur, d'interprétation et d'intérêt variés. Une aide à l'interprétation peut être apportée si l'on se souvient de la tautologie en logique formelle : a=>(b=>c) ⇔ (a et b)=>c. De plus, relativement à chaque classe maximale, le logiciel C.H.I.C. indique quelle variable supplémentaire contribue le plus à la formation de la classe. Cette information permet d'améliorer la compréhension et la signification de la classe.

Voici une partie de la hiérarchie où nous nous limitons dans un souci de clarté à deux classes maximales.

13 Par exemple, E symbolise l'objectif : « développement de l'imagination et de la créativité »

14 Par exemple, OP4 symbolise : « Pour corriger, j'aime bien un barème très détaillé sur les résultats à obtenir »

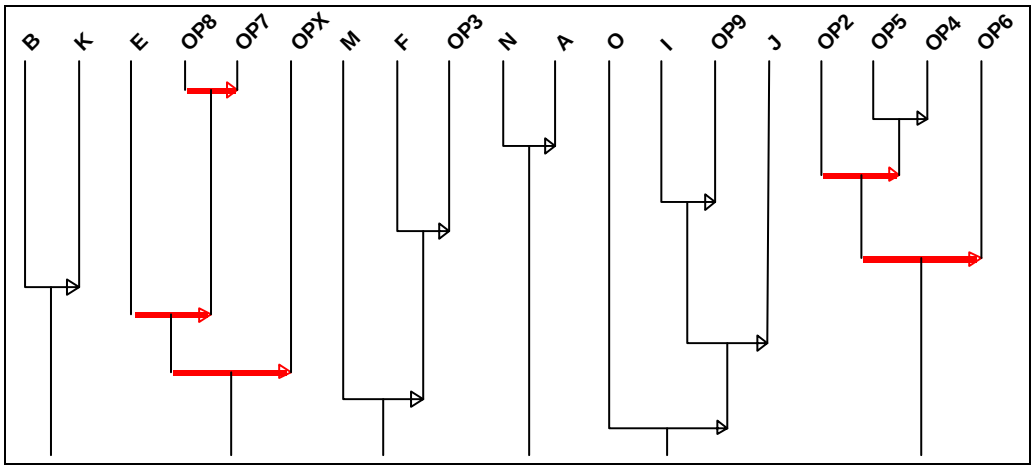


Fig. 5. Hiérarchie orientée pour le questionnaire

Des règles généralisées apparaissent sur ce graphe et sont significatives :

- **si E alors (si OP8 alors OP7))** qui peut s'interpréter ainsi : si un enseignant choisit E (imagination et créativité) alors en général il considère que pour que l'élève découvre un caractère de divisibilité par 4 (OP7), il suffit qu'il ait été entraîné à trouver lui-même exemple et contre-exemple (OP8). Cette règle est d'ailleurs constituée à un niveau significatif de la hiérarchie. Elle met l'accent sur la relation entre des comportements non dogmatiques de l'enseignant et, en conséquence, la volonté de placer l'élève en situation de recherche personnelle. Ainsi, nous pouvons interpréter cette R-règle comme l'indice d'une conception d'ouverture didactique ;

- **si (si OP2 alors (si OP5 alors OP4) alors OP6.** Explicitement, cette R-règle se lit de la façon suivante : si l'on considère un grand problème indispensable à l'examen (OP2), alors, considérant que la démonstration est le seule façon rigoureuse de faire des mathématiques (OP5), le barème de correction doit être précis. Mais alors, la prémisse de la R-règle étant satisfaite, l'enseignant a tendance à choisir OP6 (demande de programmes bien définis afin de savoir ce que l'on doit ou ne doit pas faire). On a là une conception d'une catégorie d'enseignants très soumise à l'institution et conservatrice dans ses choix pédagogiques. Ne soyons pas surpris, la démonstration en France est le fondement de l'activité mathématique (pays de Descartes), tout en étant difficile à évaluer ; le grand problème en est le critère d'évaluation. On retrouve ici une illustration plus synthétique des règles d'association qui sous-tendent, de façon homogène et cohérente, la R-règle, à savoir une conception de l'enseignement très classique qui exige un soutien institutionnel explicite et libérateur.

. En résumé, remarquons que ces deux R-règles correspondent à deux conceptions opposées de la mission enseignante, évidemment associées à deux profils typiques et inconciliables de professeurs.

Insistons sur l'accroissement, dans chacun de ces cas, de la richesse de l'analyse obtenue par l'association des règles d'association en des R-règles. Ce ne sont plus seulement des faits ou des comportements isolés qui sont extraits., mais plutôt des conduites générales, révélatrices elles-mêmes de phénomènes plus globaux, moins singuliers ou de représentations psychologiques profondes . Une typologie comme en fournissent les classifications traditionnelles (donc symétriques), ne pourrait pas rendre compte de la dynamique des faits ou comportements sous-jacents. C'est pourtant cette dynamique restituée par les règles généralisées qui, appuyée sur des nécessités (les prémisses des règles), conduit à des élucidations vives d'un fragment de théorie, éventuellement en voie de construction.

8 Niveaux significatifs d'une hiérarchie cohésitive

8.1 Situation fondamentale

Etant donné la multiplicité des niveaux de formation des classes, il est indispensable de dégager ceux qui sont les plus pertinents par rapport à l'intention classificatrice du chercheur et eu égard aux critères choisis. Ces niveaux semblent, dans des applications psycho-didactiques ou sociologiques, correspondre à des conceptions consistantes et stables, d'où leur intérêt pour l'expert. Leur mise en évidence conduit celui-ci à une étude plus approfondie et une interprétation plus assurée des classes généralisées obtenues. Ce problème s'est posé lors de

l'analyse d'un questionnaire présenté à des enseignants de mathématiques sur les objectifs de l'enseignement des mathématiques tels qu'ils les ressentaient (Bodin A. et Gras R. 1999). Les choix de base se sont structurés en attitudes de plus en plus nuancées au cours de l'ascendance de la hiérarchie.

Nous procédons alors de façon comparable à celle adoptée primitivement par I.C.Lerman (Lerman I.C. 1981) et relativement à la hiérarchie de similarité, mais en reconditionnant son approche. Depuis 2004, nous avons établi un autre mode de définition de niveaux significatifs (Gras R., Kuntz P. et Régnier J.C. 2004 b) et voir la conférence de P. Kuntz) basé sur un indice dit de cohérence. Mais nos préférons présenter ici le mode initial qui seul est programmé dans CHIC.

8.2 Préordre cohésitif

Considérons l'ensemble V des variables $\{a_1, a_2, ..., a_m\}$ et l'ensemble des couples (a,b) de VxV tels que $a \neq b$. Il existe m(m-1) tels couples auxquels on a associé leurs cohésions c(a,b) respectives.

Définition 10 : On appelle préordre initial et global cohésitif sur VxV (ou préordonnance), le préordre Ω induit par l'application cohésion c sur VxV.

Soit G(Ω) son graphe dans VxV. D'après les paragraphes précédents, il s'ensuit que:

* d'une part, la classe de préordre correspondant à c=0 contient tous les couples tels que $\psi(a,b) \leq 0{,}5$,

* d'autre part, si $n_a \leq n_b$ alors $c(b,a) \leq c(a,b)$.

Remarquons, par contre, que si $c(a,b) \leq c(c,d)$ on n'a pas nécessairement $c(b,a) \leq c(d,c)$ ou $c(b,a) \geq c(d,c)$.

8.3 Détermination des niveaux significatifs

Plaçons-nous à un niveau quelconque k de la hiérarchie. A ce niveau, se forme une classe de m_i variables ($2 \leq m_i \leq m$) dont la cohésion est moins bonne que celle des classes antérieurement formées, conformément à l'algorithme retenu, et meilleure que celles des classes à venir.

Soit $\prod_k$ la partition sur V définie à ce niveau constituée des classes qui y sont déjà formées et, éventuellement, des singletons non encore associés. $\prod_k$ est plus fine que $\prod_{k+1}$.

Soit $S\prod_k$ l'ensemble des couples séparés à ce niveau et $R\prod_k$ l'ensemble des couples qui y sont réunis pour la première fois, étant entendu que l'on dira que le couple (a,b) est réuni si a et b appartiennent à la même classe du type (...(...a,...))...b)...).

L'ensemble $G(\Omega) \cap [\, S\prod_k \text{ x } R\prod_k\,]$ est constitué des couples de couples qui au niveau k respectent le préordre initial. Par exemple, si l'on a c(e,f) < c(a,b) (donc ((e,f), (a,b)) $\in$ G(Ω)) et si au niveau k, e et f sont séparés alors que a et b se réunissent dans la classe qui se forme, le couple ((e,f),(a,b)) appartient à $G(\Omega) \cap [\, S\prod_k \text{x } R\prod_k]$.

Comme il a été fait dans le paragraphe **2** pour le cardinal de $A \cap \overline{B}$, nous associons au cardinal de $G(\Omega) \cap [\, S\prod_k \text{x } R\prod_k]$ l'indice aléatoire $card[G(\Omega^*) \cap [S\prod_k \text{x } R\prod_k]]$ où Ω^* est une préordonnance aléatoire dans l'ensemble, muni d'une probabilité uniforme, de toutes les préordonnances de même type cardinal que Ω. Cet indice a pour espérance 1/2 $card[\, S\prod_k \text{x } R\prod_k]$ et pour variance $card[\, S\prod_k \text{x } R\prod_k]$ $card[G(\Omega)]$.

Soit $s(\Omega, k)$ l'indice centré réduit obtenu :

$$\frac{(card[G(\Omega^*) \cap [S\Pi_k \text{x} R\Pi_k]] - 1/2\, card[S\Pi_k \text{x} R\Pi_k])}{(card[S\Pi_k \text{x} R\Pi_k]\, card[G(\Omega)])^{1/2}}.$$

Définition 11 : On appelle noeud significatif tout noeud correspondant à un maximum local de $s(\Omega, k)$ au cours de la constitution de la hiérarchie implicative. Nous dirons dans ce cas que la partition $\prod_k$ est en résonance partielle avec Ω.

Si, de plus, $G(\Omega) \cap [\, S\prod_k \text{ x } R\prod_k] = S\prod_k \text{x } R\prod_k$, nous dirons que la partition $\prod_k$ est en *résonance totale* avec Ω. Le logiciel d'analyse C.H.I.C. permet le traitement complet de données quantitatives, ainsi que la sortie du graphe d'implication et de la hiérarchie implicative en mentionnant les noeuds significatifs.

9 Typicalité et contribution des sujets et des variables supplémentaires

9.1 Situation fondatrice

Nous introduisons la notion de variable supplémentaire en analyse implicative à l'instar de la même notion définie en analyse factorielle, autrement dit, c'est une variable extrinsèque, un descripteur par exemple, n'intervenant pas directement dans les liaisons exprimées par la classification entre les variables dites principales de V, donc n'intervenant pas dans la structure de cet ensemble sous la forme graphe ou hiérarchie. Par exemple, une variable supplémentaire pourra représenter une catégorie de sujets (âge, sexe, catégorie socio-professionnelle, etc.). Cette notion doit permettre d'éclairer sur l'importance ou la superfluité de ces catégories dans la formation des règles ou de classes de règles. Savoir si ce sont plutôt les enseignants de classes littéraires qui adoptent telle attitude vis-à-vis des objectifs de l'enseignement des mathématiques est plus riche, sur le plan de l'information recherchée, que la seule identification de l'attitude.

Donc, il nous paraît très important au moment du choix des variables sur lesquelles portera l´analyse, en liaison avec l'expert, de bien distinguer les variables qui vont participer à la mise en évidence de règles ou de méta-règles, variables que nous appelons **principales**, des autres variables, dites **supplémentaires** qui sont descriptives des sujets et qui de ce fait sont objectives, non liées directement à leurs comportements de sujets. Les introduire dans l´analyse risquerait fort, comme nous l´avons observé, de créer des artefacts dans le graphe ou la hiérarchie, car elles y jouent alors un rôle attracteur trop puissant pour laisser les autres variables subjectives s´organiser entre elles. Comme on le sait, les processus de chaînage inhérents aux méthodes de classification, perturberaient les raisons véritables qui sont à la base des relations à extraire. En revanche, identifier les phénomènes objectifs qui jouent un rôle indirect dans les structures apparues est d´une autre importance.

Intéressons-nous pour l´instant à l´analyse cohésitive. A un niveau quelconque de la hiérarchie se forme une classe **C** de cohésion non nulle. Notre objectif, particulièrement dans le cas d'un noeud significatif, est de définir un critère permettant d'identifier un ou des sujets, puis la catégorie de sujets, ou tout autre variable supplémentaire, contribuant le plus à la constitution de cette classe. Le comportement de ces sujets sera ainsi en harmonie avec le comportement statistique à l'origine de la classe. Une approche comparable est faite conjointement pour étudier la typicalité et la contribution des sujets et des variables supplémentaires à la constitution d'un arc ou d'un chemin du graphe implicatif.

9.2 Puissance implicative de classe et de chemin

Plaçons-nous à un niveau k de la hiérarchie où viennent de se réunir, pour former **C**, deux classes $\underline{\boldsymbol{A}}$ et $\underline{\boldsymbol{B}}$ telles que $\underline{\boldsymbol{A}} \Rightarrow \underline{\boldsymbol{B}}$ au sens du paragraphe 8

Définition 12 : Le couple (a,b) tel que: $\forall i \in \underline{\boldsymbol{A}}$, $\forall j \in \underline{\boldsymbol{B}}$ $\psi(a,b) \geq \psi(i,j)$ est appelé couple générique de **C.** C'est ce couple, généralement unique, qui intervient par le sup. dans le calcul de l'implication de $\underline{\boldsymbol{A}}$ sur $\underline{\boldsymbol{B}}$**.** Le nombre $\psi(a,b)$ est appelé implication générique de **C**.

Mais, dans chaque sous-classe de **C**, existe également un couple générique. Précisément, si **C** est constituée de g ($g \leq k$) sous-classes (**C** comprise), il y a g couples génériques à l'origine de **C** et g intensités maximales d'intensités contingentes $\psi_1, \psi_2, \ldots, \psi_g$, qui leur correspondent.

Dans le cas d'un chemin **C** fermé transitivement (chaque arc de la fermeture admet une intensité d'implication au moins égale à 0.50), composé de g nœuds, **C** présente g(g-1)/2 arcs transitifs. A chacun de ces arcs, par exemple (a,b), on associe, comme pour une classe, l'intensité d'implication $\psi(a,b)$, que l'on dira encore générique.

Définition 13 : Le vecteur contingent générique ($\psi_1, \psi_2, \ldots, \psi_g$), élément de $[0,1]^g$, est appelé vecteur puissance implicative de **C.** Il traduit une force implicative interne à **C**.

9.3 Puissance implicative d'un sujet sur une classe ou un chemin et distance à cette classe ou ce chemin

Un sujet x quelconque respecte ou non l'implication du couple générique d'une classe ou d'un arc de chemin avec un ordre de qualité comparable. Associant logique formelle et considération sémantique, nous poserons, par exemple et en fonction des valeurs prises par a et b en x:

$\psi_x(a, \bar{b})$=1 si a=1 ou 0 et b=1; $\psi_x(a, \bar{b})$=0 si a=1 et b=0 ; $\psi_x(a, \bar{b})$=p si a=b=0 avec p∈]0,1].

Dans nos premières expériences, nous choisissions p=.5, valeur neutre. Dans le logiciel CHIC, le calcul des typicalités et des contributions se fait cependant en modulant ces valeurs afin de mieux prendre en compte la sémantique des valeurs attribuées par x à a et à b.

Ainsi, à x, nous pouvons associer g nombres $\psi_{x,1}$, $\psi_{x,2}$,..., $\psi_{x,g}$ correspondant aux valeurs prises en x par les g implications génériques de la classe ou du chemin **C**.

Définition 14 : Le vecteur ($\psi_{x,1}$, $\psi_{x,2}$,..., $\psi_{x,g}$), élément de $[0,1]^g$, est appelé vecteur contingent générique ou puissance implicative de x. Le sujet (ou peut-être les sujets) x_t, peut-être fictif, dont toutes les composantes du vecteur puissance sont (ψ_1, ψ_2,..., ψ_g), est appelé sujet typique optimal de **C**.

Dans ces conditions, on peut munir l'espace des puissances $[0,1]^g$ d'une métrique du type χ^2 afin d'accentuer les effets de fortes implications génériques.

Définition 15 : On appelle distance implicative d'un sujet x à la classe ou au chemin **C** le nombre:

$$d(x,\mathbf{C})= \left[\frac{1}{g} \sum_{i=1}^{i=g} \frac{[\psi_i - \psi_{x,i}]^2}{1-\psi_i} \right]^{\frac{1}{2}}$$

Ce nombre n'est autre que la distance dite du χ^2 entre les deux distributions $\{1-\psi_i\}_i$ et $\{1-\psi_{x,i}\}_i$ qui exprime les écarts entre les implications génériques empiriques et l'implication selon x. C'est pour cette raison que nous avons choisi le mot **typicalité** que nous allons définir plus bas**.** Elle permet de conférer à E une C-structure topologique discrète d'espace métrique**.** Si pour un i, $\psi_i = 1$, nous poserons, par convention, $\psi_{x,i} = 1$. Cette convention ne se fait pas contre nature puisque, dans ce cas, l'implication générique est maximale et significative d'une excellente liaison implicative entre ses deux termes, vérifiée par tous ou presque tous les sujets x de E. Ainsi, si le dénominateur s'annule, il en est de même du numérateur, sauf exception, et l'on pourra de toute façon attribuer la valeur 0 au quotient.

Remarque : Une classe C étant donnée, on peut définir une structure d'espace métrique sur E par la donnée de la distance indicée par C entre deux sujets quelconques de E, distance qui mesure la différence de comportement des sujets x et y à l'égard de C :

$$d_{\mathbf{C}}(x,y) = \left[\frac{1}{g} \sum_{i=1}^{i=g} \frac{[\psi_{x,i} - \psi_{y,i}]^2}{1-\psi_i} \right]^{\frac{1}{2}}$$

On voit alors que la distance de typicalité donnée plus haut n'est que la spécification de $d_{\mathbf{C}}$ aux sujets respectivement x et x_t. La distance $d_{\mathbf{C}}$ permet de conférer à E une **C**-structure topologique discrète. Cette topologie est équivalente à celle qui serait définie sur l'ensemble des vecteurs contingents ($\psi_{x,1}$, $\psi_{x,2}$,...,$\psi_{x,g}$) , sous-ensemble d'un espace vectoriel normé de dimension g et de norme : $\|\vec{x}-\vec{y}\| = d_{\mathbf{C}}(x,y)$. L'opérateur symétrique associé à la forme quadratique qui conduit à cette distance, a pour matrice la matrice diagonale d'éléments $[g(1-\psi_i]^{-1}$ pour i=1,…,g. Il est bien évident que la somme de deux tels vecteurs n'a qu'un sens théorique, c'est-à-dire hors du contexte dans lequel nous travaillons en A.S.I..

Une application intéressante peut consister à déterminer le ou les sujets appartenant à une boule de diamètre donné et de centre l'un des sujets pré-désignés, comme par exemple, l'individu optimal. En prolongement de cette approche métrique, le problème de complétion des données manquantes pourrait y puiser une solution originale.

9.4 Typicalité et contribution d'un sujet et d'une variable supplémentaire à une classe ou un chemin

9.4.1 Typicalité

Nous définirons la mesure de typicalité à partir du rapport entre la distance à C du sujet considéré et celle qui est la plus grande dans l'ensemble des sujets. Cette distance maximale est celle des sujets dont les $\psi_{x,i}$ sont tous nuls ou très faibles. Ces sujets sont donc les sujets les plus opposés aux règles génériques. La typicalité d'un sujet sera alors d'autant plus grande qu'il s'écartera de ces sujets, qu'il aura donc un comportement comparable à celui du sujet théorique optimal. La typicalité d'une catégorie de sujets ou d'une variable supplémentaire G s'en déduira :.

.

Définition 16 : La typicalité de x à **C** est : $\gamma(x, C)= 1 - \frac{\mathbf{d}(\mathbf{x},\mathbf{C})}{\max_{\mathbf{y}\in\mathbf{E}} \mathbf{d}(\mathbf{y},\mathbf{C})}$

et celle de G en est la moyenne dans G : $\gamma(G,\mathbf{C})=\frac{1}{\text{card}G}\sum_{x\in G}\gamma(x,G)$

La typicalité de x est un nombre de [0,1] qui vaut 1 si x est typique optimal et 0 si x est le plus en désaccord avec **C**. Afin de donner au chercheur le moyen de savoir ou de vérifier rapidement si telle catégorie de sujets qui l'intéresse est statistiquement déterminante dans la constitution d'une classe implicative ou d'un chemin transitif, un algorithme a été élaboré en s'appuyant sur les deux notions suivantes: groupe optimal et catégorie déterminante.

Définition 17 : Soit E la population étudiée. Un groupe optimal d'une classe implicative ou d'un chemin **C,** groupe noté GO(**C**), est le sous-ensemble de E qui accorde à **C** une typicalité plus grande que le complémentaire de GO(**C**) et qui forme avec celui-ci une partition en deux groupes maximisant la variance inter-classe de la série statistique des typicalités individuelles. Une telle partition est dite *significative*.

L'existence de ce groupe optimal est démontrée dans (Gras R. et Ratsimba-Rajohn H. 1997). Les propriétés utilisées sont aussi celles qui le sont pour établir l'algorithme sur lequel se basent les modules des programmes informatiques qui construisent, automatiquement dans C.H.I.C., chaque sous-groupe optimal.

Considérons une partition $\{G_i\}_i$ de E. Cette partition peut être définie par une variable supplémentaire correspondant par exemple à un descripteur de E. Soit X_i une partie aléatoire de E ayant le même cardinal que G_i, et Z_i la variable aléatoire Card ($X_i \cap GO(\mathbf{C})$). Z_i suit une loi binomiale de paramètres : card G_i et card GO(**C**) / card E.

Définition 18 : On appelle catégorie la plus typique de la classe implicative ou du chemin **C**, la catégorie qui minimise l'ensemble $\{p_i\}_i$ des probabilités p_i telles que:

$\forall$ i, p_i = Prob [card $G_i \cap$ GO(**C**)<Z_i]

Une catégorie G_0 est dite déterminante au seuil α si la probabilité associée p_0 est inférieure à α.

Ainsi, la signification d'une classe ou d'un chemin ayant été donnée par l'expert, il lui associera la sous-population la plus porteuse de ce sens. Cette approche est comparable à celle de I.-C. Lerman pour l'analyse des similarités, mais au moyen d'une modélisation et de concepts différents.

D'ailleurs, nous pouvons remarquer que nous pouvons associer au sous-groupe optimal une variable binaire correspondant à la fonction indicatrice de ce sous-ensemble de E. De la même façon, nous pouvons également associer à la catégorie G_i ou bien à la variable supplémentaire correspondante, une variable binaire dont l'indice de similarité $s=\frac{n_{a\wedge b}-\frac{n_a n_b}{n}}{\sqrt{\frac{n_a n_b}{n}}}$, au sens de I.C. Lerman, vérifie : p_i =Pr[S≥s], S étant la valeur aléatoire dont s est la réalisation. Ainsi, minimiser l'ensemble des probabilités $\{p_i\}_i$ revient à maximiser l'indice de similarité

entre les variables binaires, indicatrices de sous-ensembles, associées respectivement l'une au groupe optimal GO(**C**) et les autres aux différentes catégories $\{G_i\}_i$.

Cette remarque permet d'étendre efficacement la notion de variable supplémentaire la plus typique à des variables numériques, prenant leurs valeurs sur [0,1]. Il suffit dans ce cas d'extraire la plus forte des valeurs de similarité entre la variable binaire indicatrice définie par le groupe optimal et les différentes variables numériques placées en supplémentaire, l'indice étant calculé selon le principe retenu en analyse implicative pour les variables numériques (cf. définition 5). Nous savons que sa restriction au cas binaire coïncide avec sa valeur s dans le cas où les deux variables sont binaires.

Ainsi il est possible de dégager à la fois les individus et les groupes d'individus typiques d'une liaison ou d'un ensemble (classe ou chemin) de liaisons. Ce sont donc ceux qui sont le plus en accord avec la qualité de ces liaisons au sein de la population E considérée. Si par exemple la liaison entre les variables a et b est quantifiée par le nombre $\psi(a,b) = 0{,}92$, les individus x qui attribuent à cette liaison la valeur $\psi_x(a,b) = 0{,}90$ sont plus typiques que ceux qui lui attribuent la valeur 0,98. La nuance entre cette notion et celle de contribution que nous allons définir prend tout son sens dans l'étude des variables modales ou numériques.

9.4.2 Contribution

Cette notion se distingue de la précédente par l'examen de la responsabilité des individus, puis des variables supplémentaires, qui peuvent en être des descripteurs, *à l'existence* d'une liaison d'une règle ou d'une méta-règle entre variables principales. Supposons, en effet, que deux variables a et b (resp. plusieurs variables sur un chemin du graphe ou bien deux classes de la hiérarchie) soient réunies par un arc sur un graphe à un certain seuil (resp. en un chemin transitif **C** du graphe ou bien en une classe **C** dans une hiérarchie à un certain niveau). Connaissant la valeur $\psi_{x,i}$ attribuée par l'individu à la règle i : $a \Rightarrow b$ (resp. règle i du chemin **C** ou bien de la classe **C** constituée de g règles génériques) supposée admissible,

Définition 19 : On appelle *distance de x à (a,b) ou à **C*** :

$$d(x,\mathbf{C}) = \left[\frac{1}{g}\sum_{i=1}^{i=g}\left[1-\psi_{x,i}\right]^2\right] \text{ où g = 1 dans le cas de l'arc (a,b)}$$

On appelle *contribution de x à **C*** le nombre : $\gamma(x,\mathbf{C}) = 1\text{-}d(x,\mathbf{C})$

Cette contribution a pour maximum 1 dans le cas où l'individu x a donné la valeur 1 à toutes les règles i. Ceci permet de concilier sémantique et définition formelle. La suite des définitions et des algorithmes de calcul (contribution d'une catégorie ou d'une variable supplémentaire G, groupe optimal d'individus, catégorie ou variable supplémentaire la plus contributive) se transpose immédiatement à partir des principes de la typicalité. Mais dans les situations réelles, nous observons la nuance entre les deux concepts ce qui enrichit l'information exploitable par l'utilisateur. Cependant, <u>le concept de contribution est plus volontiers retenu pour l'interprétation.</u>

9.5 Application

Revenons à l'illustration évoquée dans le § 7.3 (Questionnaire présenté à des enseignants de mathématiques). Les occurrences des variables supplémentaires sont :

S(cientifique) : 155 ES(économique et sociale) : 68 LI(ttéraire) : 22 TE(chnologique) : 66. La hiérarchie cohésitive obtenue par CHIC à partir d'un nombre réduit des variables, afin de conserver les niveaux les plus significatifs, est donnée par la FIG. 3.

Considérons la classe **C**= $[E \Rightarrow (OP8 \Rightarrow OP7)] \Rightarrow OPX$. Son sens, analysé plus en détail dans (Bodin et al., 1999), est fortement marqué par l'importance accordée à l'imagination et à la recherche personnelle, par les enseignants d'accord avec ces objectifs et ces opinions, La variable la plus **typique** pour cette classe est S (série Scientifique) avec un risque de : 0.00393.

En effet, 116 des enseignants de S parmi les 155 de cette série qui ont répondu au sondage, figurent dans le groupe optimal **(GO)** de cardinal 201 relatif à **C**. Soit X une partie aléatoire de même cardinal (155) que S et Z la variable aléatoire égale au cardinal de l'intersection de X et du groupe optimal **GO**. Selon un modèle équiprobable de distribution des enseignants, Z suit la loi binomiale de paramètres 155 et 201/311 soit 0.656. La probabilité pour que Z soit plus grande que 116 est le risque annoncé, soit 0.00393**.** Mais pour **S**, c'est le couple (**S, (OP8, OP7**)) qui est mutuellement spécifique au seuil $\beta = 2.10^{-5}$. On retrouve une telle mutuelle spécificité

pour **TE** avec le couple **(TE, (B,K))** à un seuil 5.10^{-7} nous confirmant, sans surprise, que les enseignants des sections techniques **(TE)** considèrent que les mathématiques doivent être utiles à la vie professionnelle (**B**) et, en conséquence, aux autres disciplines (**K**) et y sont les plus attachés.

Les calculs de **contribution** à la classe **C** montrent que, cette fois, 111 enseignants sur les 311 sondés, participent au groupe optimal. Le nombre d'enseignants de **S** a diminué (il passe de 116 à 67) et, surtout, sa proportion est bien moindre que précédemment dans le **GO**. Ceci se ressent dans le seuil qui est 0.0251, soit un risque 6 fois plus élevé que pour la typicalité. Ce sont les enseignants sondés de **S** qui sont les plus typiques, c'est-à-dire « conformes » au comportement général de la population elle-même sondée. Mais ils sont moins contributeurs dans les relations strictes entre les 4 variables constituant **C**. Cette remarque nous montre les nuances apportées par les deux concepts : typicalité et contribution

Certaines liaisons apparues et commentées ci-dessus se retrouvent dans le graphe de la FIG. 4. Les contributions calculées dans CHIC montrent encore que les enseignants de la série S contribuent le plus au chemin : **E** $\Rightarrow$ **OP8** $\Rightarrow$ **OP7** $\Rightarrow$ **OPX** avec un risque d'erreur de 0.00746, la transitivité le long de ce chemin étant assurée au niveau 0.75.

10 Conclusion

Ce panorama du développement de l'analyse statistique implicative montre, s'il est nécessaire, comment une théoris de traitement de données se construit pas à pas en réponse à des problématiques présentées par les experts de domaines variés. Elle apparaît donc autrement que comme vue de l'esprit puisque directement applicable aux situations qui conduisent à sa genèse. Les extensions apportées aux types de données traitées, aux modes de représentation de leurs structures, aux relations entre les sujets, leurs descripteurs et les variables sont bien les fruits des interrogations gourmandes des experts. Ses fonctions respectives de révélateur et d'analyseur semblent opérer avec bonheur dans de multiples domaines applicatifs.

On aura remarqué que la base théorique est simple ce qui pourrait être la raison de sa fécondité. Même si les remises en question de choix théoriques primitifs ne sont pas apparentes ici, cette genèse ne s'est pas faite sans conflits entre les réponses attendues, la facilité de leur accès et donc ces réponses ont été sources de restaurations voire de refontes. Quoi qu'il en soit cette méthode d'analyse de données aura permis et permettra encore, je l'espère, la mise en lumière de structures vivantes grâce à l'approche non symétrique qui en est le fondement.

Parmi les travaux ultérieurs proposés à notre équipe, l'un portera sur une extension de l'A.S.I. à des variables vectorielles en réponse à des problèmes posés en protéomique. Un autre portera plus extensivement sur la relation entre l'A.S.I et le traitement des ensembles flous. La fonction de l'opérateur "implication" en logique floue sera précisée et illustrée par des applications. A travers un autre sujet, nous tenterons de construire une méthode qui permettrait de résoudre par l'A.S.I. le problème des vacuités d'un tableau de données. Citons également le travail en cours sur la recherche d'un indice de qualité d'un graphe implicatif et d'une hiérarchie de R-règles. Enfin, il est bien évident que ces travaux seront conduits interactivement avec des applications et, en particulier, celle de l'apport de l'ASI à la règle de classification dans les feuilles des arbres de classification. (cf. Texte de G. Ritschard dans cet ouvrage).

Références

Ag Amouloud S., (1992), L'ordinateur, outil d'aide à l'apprentissage de la démonstration et de traitement de données didactiques, thèse de l'Université de Rennes 1, 2 novembre 1992

Agrawal R., Imielinsky T. et Swami A,(1993), Mining association rules between sets of items in large databases, Proc. of the ACM SIGMOD'93, (1993)

Amarger S., Dubois D. et Prade H., (1991, Imprecise quantifiers and conditional probabilities" - in Symbolic and quantitative approaches to uncertainty (R. KRUSE, P. SIEGEL), Springer-Verlag, 1991, 33-37.

Aze J. et Kodratoff Y. (2001), "Evaluation de la résistance au bruit de quelques mesures d'extraction de règles d'association », Extraction des connaissances et apprentissage, Hermès, Vol 1, n° 4, 2001, p. 143-154

Bailleul M., (1994), Analyse statistique implicative: variables modales et contribution des sujets. Application à la modélisation de l'enseignant dans le système didactique, Thèse de l'Université de Rennes 1, juin 1994.

Bailleul M. et Gras R., (1995), L'implication statistique entre variables modales, Mathématique, Informatique et Sciences Humaines, E.H.E.S.S. Paris, n°128, (1995), 41-57

Bernard J.-M. et Poitrenaud S, (1999) L'analyse implicative bayesienne d'un questionnaire binaire : quasi-implications et treillis de Galois simplifié", Mathématiques, Informatique et Sciences Humaines, n° 147, 1999, 25-46

Blanchard J., Kuntz P., Guillet F. et Gras R.,(2004), Mesure de la qualité des régles d'association par l'intensité d'implication entropique, Mesures de qualité pour la fouille de données, RNTI-E-1, p. 33-44, 2004.

Blanchard J., Guillet F., Briand H. et Gras R. (2005) Ipee : Indice probabiliste d'écart à l'équilibre pour l'évaluation de la qualité des règles, actes Atelier Qualité des Données et des Connaissances, pp. 26-34, 2005.

Bodin A. (1997), Modèles sous-jacents à l'analyse implicative et outils complémentaires. Prépublication IRMAR. n°97-32, (1997)

Bodin A.et Gras R., Analyse du préquestionnaire enseignants avant EVAPM-Terminales, Bulletin n°425 de l'Association des Professeurs de Mathématiques de l'Enseignement Public, 772-786, Paris

Couturier R. (2001), Traitement de l'analyse statistique implicative dans CHIC, Actes des Journées sur la Fouille dans les données par la méthode d'analyse implicative, IUFM Caen, p. 33-50.

Couturier R, Gras R., et Guillet F.. [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, July 2004.

Couturier R, Gras R. [2005] : CHIC : Traitement de données avec l'analyse implicative, Extraction et Gestion des Connaissances, Volume I1, RNTI, Cépaduès, Paris, p.679-684, , ISBN 2.85428.683.9

David J, Guillet F., Gras R. and Briand H. [2006] : Conceptual hierarchies matching : an approach based on discovery of implication rules between concepts, *In Proc. ECAI 2006, 17th European Conference on Artificial Intelligence, Riva del Garda, Italy, IOS Press*

Diday E, (1972), Nouvelles méthodes et nouveaux concepts en classification automatique et reconnaissance des formes, Thèse d'Etat, Université de Paris VI, 1972

Fleury L. (1996), Extraction de connaissances dasn une base de données pour la gestion de ressources humaines, Thèse d'Université, Université de Nantes, 22 novembre 1996

Ganascia J.G., (1987), AGAPE et CHARADE : deux techniques d'apprentissage symbolique appliquées à la construction de bases de connaissances, Thèse d'Etat, Université de Paris Sud, 1987

Goodman R.M. et Smyth P. (1989), "The induction of probabilistic rule set. The ITRULE algorithm", Proceedings of sixth international conference on machine learning, 1989, p. 129-132

Gras R., (1979), Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques, Thèse d'Etat, Université de Rennes I, 1979

Gras R ; et Larher A., (1992), L'implication statistique, une nouvelle méthode d'analyse de données, Mathématique, Informatique et Sciences Humaines, E.H.E.S.S. Paris, n°120 1992, 5-31

Gras R. et Ratsimaba-Rajohn H., Analyse non symétrique de données par l'implication statistique. RAIRO-Recherche Opérationnelle, 30-3, AFCET, Paris, 1996, 217-232

Gras R., Briand H. et Peter P., (1996), Structuration sets with implication intensity", Proceedings of the International Conference on Ordinal and Symbolic Data Analysis - OSDA 95, E.Diday, Y.Chevallier, Otto Opitz, Eds., Springer, Paris 1996, 147-156

Gras R., Ag Almouloud S., Bailleul M., Larher A., Polo M., Ratsimba-Rajohn et Totohasina A (1996), L'implication Statistique, Collection Associée à "Recherches en Didactique des Mathématiques", La Pensée Sauvage, Grenoble, 1996

Gras R., Briand H., Peter P. et Philippé J., (1997), Implicative statistical analysis, Proceedings of International Congress I.F.C.S., 96, Kobé, Springer-Verlag, Tokyo , 1997, 412-419

Gras R., Kuntz P., Couturier R. et Guillet F., (2001), Une version entropique de l'intensité d'implication pour les corpus volumineux,.Extraction des Connaissances et Apprentissage (ECA), vol. 1, n° 1-2, 69-80. Hermès Science Publication, 2001

Gras R., Kuntz P. et Briand H., (2001), Les fondements de l'analyse statistique implicative et quelques prolongements pour la fouille de données, Mathématiques et Sciences Humaines, n° 154-155, p 9-29, ISSN 0987 6936, 2001

Gras R., Diday E., Kuntz P et Couturier R. (2001), Variables sur intervalles et variables-intervalles en analyse statistique implicative, Actes du 8ème Congrès de la Société Francophone de Classification, Université des Antilles-Guyane, Pointe-à-Pitre, 17-21 décembre 2001, pp 166-173

Gras R(égis)., Guillet F., Gras R(obin). et Philippé J., (2002) Réduction des colonnes d'un tableau de données par quasi-équivalence entre variables, Extraction des connaissances et apprentissage, Hermès, Volume 1, n°4/2001, p 197-202, ISBN 2-7462-0406-1, 2002

Gras R., Kuntz P. et Briand H. (2003), Hiérarchie orientée de règles généralisées en analyse implicative, Extraction des Connaissances et apprentissage, Hermès, p 145-157, ISSN 0992-499X, ISBN 2-7462-0631-5, 2003

Gras R., Couturier R.., Blanchard J., Briand H., Kuntz P., Peter P., (2004 a) : Quelques critères pour une mesure de qualité de règles d'association. Un exemple : l'implication statistique, Mesures de qualité pour la fouille de données, RNTI-E-1, Cépaduès –Editions, p 3-32, I.S.B.N. 2.85428.646.4

Gras R., Kuntz P. et Régnier J.C., (2004 b), Significativité des niveaux d'une hiérarchie orientée en analyse statistique implicative, Classification et fouille de données, Ed M. Chavent et M.Langlais, RNTI-C-1,

Gras R. et Kuntz P., (2005), Discovering R-rules with a directed hierarchy, Soft Computing, A Fusion of Foundations, Methodologies and Applications, Volume 1, p. 46-58, ISSN 1432-7643, Springer Verlag, 2005

Gras R. et Régnier J.C. (2007), Différents modèles d'indices d'implication en Analyse Statistique Implicative, soumis à Revue de Statistiques Appliquée

Gras R. et Kuntz P. [2007]. Reduction of Redundant Rules in Statistical Implicative Analysis. *Selected Contributions in data Analysis and Classification*, P. Brito, P. Bertrand, G. Cucumel, E. de Caravlho, eds, Springer, p.367-376, ISBN : 1431-8814

Kuntz P., Gras R., Blanchard J. (2002), Discovering Extended Rules with Implicative Hierarchies, Conference on the new frontiers of statistical data mining and knowledge discovery, juin 2002, Knoxville, Tennesee

Lagrange J.B., (1998), Analyse implicative d'un ensemble de variables numériques ; application au traitement d'un questionnaire à réponses modales ordonnées", Revue de Statistique Appliquée., Institut Henri Poincaré, Paris, 1998, 71-93

Lahanier-Reuter D. (1998),.,Etude de conceptions du hasard : approche épistémologique, didactique et expérimentale en milieu universitaire, Thèse de l'Université de Rennes 1, 1998

Lallich S ;, Lenca P. et Vaillant B., Variations autour de l'intensité d'implication, (2005), Actes ASI 05, Université de Palerme

Larher A. (1991), Implication statistique et applications à l'analyse de démarches de preuve mathématique, Thèse de l'Université de Rennes 1, 4 Février 1991.

Lebart L., Morineau A. et Piron M., Statistique exploratoire multidimensionnelle, Dunod, 2000

Lenca P., Meyer P., Vaillant P., Picouet P. et Lallich S., (2004), Evaluation et analyse multi-critères de qualité des règles d'association, Mesures de qualité pour la fouille de données, RNTI-E-1,Cépaduès, p. 219-246, 2004

Lerman I.-C., (1981)., Classification et analyse ordinale des données, Dunod, Paris, 1981.

Lerman I.-C., Gras R. et Rostam H., (1981), Elaboration et évaluation d'un indice d'implication pour des données binaires, I et II, Mathématiques et Sciences Humaines , 1981, n° 74,, 5-35 et n° 75, 5-47

Lerman I.C. et Azé J., (2004), Indice probabiliste discriminant (de vraisemblance du lien) d'une règle d'association en cas de « très grosses » données, Mesures de qualité pour la fouille de données, RNTI-E-1, p. 69-94, 2004

Loevinger J., (1947), A systematic approach to the construction and evaluation of tests of abilities", Psychological Monographs, 61, n° 4, 1947

Pearl J., (1988), Probabilistic Reasoning in intelligent systems, San Mateo, CA, Morgan Kaufmann., 1988

Polo-Capra M., (1996), Le repère cartésien dans les systèmes scolaires français et italien : étude didactique et application de méthodes d'analyse statistiques multidimensionnelles, Thèse de l'Université de Rennes 1, 9 Décembre 1996.

Régnier J.C. et Gras R., Statistique de rangs et analyse statistique implicative, Revue de Statistique Appliquée, 2005 ;, LIII, p. 5-38

Saporta G. (2006), Probabilités, Analyse de Données et statistique, Ed. Technip, Paris

Sebag M. et Schoenauer (1991), « Un réseau de règles d'apprentissage », Induction symbolique-numérique à partir de données, Cépaduès Editions, 1991

Shannon C.E. et Weaver (1949)., The mathematical theory of communication, Univ. of Illinois Press, 1949.

Simon A., (2001), [29] SIMON A., Outils classificatoires par objets pour l'extraction de connaissances dans des bases de données, Thèse de l'Université de Nancy 1, 2000

Zighed, D. A. et R. Rakotomalala (2000). Graphes d'induction: apprentissage et data mining. Paris: Hermes Science Publications.

Summary

Initiated for didactical mathematic situations, the implicative statistical analysis method has been improved with lessons learned. It mainly consists in giving structure to data linking subjects to variables and providing inductive data mining between variables. Moreover, based on the contingency of rules, it allows to explain and then to make forecasts in some investigation fields: psychology, sociology, biology, ... Those objectives entailed to define the notion of implicative intensity, classes cohesion, inclusive implication, significativity in hierarchical levels, contribution of supplementary variables, ... In the same way, in addition to binary variables (for example descriptors), modal variables, frequency ones, and recently, interval variables and fuzzy ones, have been studied.