Un modèle d'élève par l'analyse statistique implicative. Prise en compte du contexte algébrique

Marie-Caroline Croset

Un modèle d'élève par l'analyse statistique implicative. Prise en compte du contexte algébrique.

Marie-Caroline Croset*

** Université Joseph Fourier*
Equipe MeTAH
Laboratoire LIG
46, Av F. Viallet
38031 Grenoble Cedex – France

Marie-Caroline.Croset@imag.fr
http://www.noe-kaleidoscope.org/people/croset/

Résumé. Le modèle d'élève que nous construisons cherche à donner une représentation fine des compétences cognitives des élèves dans le domaine de l'algèbre. Il se fonde sur les micro-actions enregistrées dans le micromonde Aplusix. Ces actions étant de niveau très fin, nous devons reconstruire les connaissances à un niveau plus abstrait : non pas une action isolée mais un ensemble d'actions « similaires » apparaissant dans un même contexte. Nous utilisons pour cela l'analyse statistique implicative qui permet de faire ressortir la stabilité des actions et de déterminer les contextes où elles ont eu lieu, en construisant des liaisons implicatives du contexte vers l'action.

1 Introduction

L'étude des connaissances des élèves établie à partir de productions papier étant fastidieuse, elle a souvent donné lieu à des recherches en intelligence artificielle afin de l'automatiser. Outre l'intérêt évident de pouvoir, ainsi, traiter de grandes bases de données et de rendre compte, tant à l'enseignant qu'au chercheur, de l'état d'apprentissage d'un groupe d'élèves, modéliser les connaissances peut aussi servir à orienter un tuteur artificiel dans le choix d'un exercice de remédiation adapté. Une remédiation abusive consisterait à réagir à chaque apparition d'erreur, sans tenir compte de « l'histoire » de l'élève. Ceci aurait pour conséquence que tout élève effectuant la même erreur recevrait en retour le même message, et ce, indépendamment de ses activités passées. Afin d'éviter cet écueil, certains auteurs suggèrent de rechercher dans l'erreur de l'élève à la fois la stabilité et un aspect conceptuel (Sleeman et al., 1989). (Brousseau, 2000), quant à lui, attribue le statut d'erreur à un écart significatif : reproductible, persistant et récurrent et ne pouvant, de ce fait, être expliqué par l'étourderie, le hasard ou une absence de contrôle.

Dans le but de construire un modèle des connaissances des élèves qui tienne compte des recommandations précédentes, il faut recueillir des informations sur les connaissances puis les organiser pour en donner une représentation ou un modèle qui justifiera les comportements dudit élève. A cette fin, nous avons utilisé la théorie de l'analyse implicative, introduite par (Gras, 1979). Le but de cette méthode est de trouver des relations significatives entre des attributs par un graphe orienté. L'« arc » entre deux attributs, a et b, sera représenté dans un graphe dit implicatif si l'intensité d'implication[1], qui mesure l'étonnement d'observer peu de contre-exemples à l'implication $a \Rightarrow b$, est supérieure à un seuil donné, eu égard à l'effectif des observations en jeu. Ce graphe implicatif est automatiquement généré par le logiciel CHIC. Nous faisons l'hypothèse que l'analyse statistique implicative permettra de :

– Discriminer les tâches qui présentent des difficultés à l'élève,
– Repérer parmi celles-ci les erreurs stables et d'en trouver des causes.

Après une description de la problématique en section 2, la section 3 expose les attributs choisis pour notre étude et les implications possibles entre ces attributs qui permettent de modéliser grossièrement un élève. Une modélisation plus fine, à nouveau basée sur l'analyse statistique implicative, est détaillée en section 4.

---

[1] Voir en particulier (Gras, 2004) pour une définition de l'intensité d'implication.

2 Problématique

2.1 Cadre de la recherche

Depuis 2003, notre équipe travaille sur la recherche d'invariants des connaissances dans le domaine transformationnel de l'algèbre[2] (Kieran, 2001). Pour cela, nous utilisons les protocoles d'élèves recueillis dans l'environnement informatique d'apprentissage pour l'algèbre, Aplusix (Bouhineau et al., 2001). Aplusix est un EIAH d'algèbre combinant des aspects micromonde et exerciseur. L'élève doit transformer des expressions algébriques selon le type d'activité : factorisation, développement, réduction d'expressions algébriques ou résolution d'équations. L'élève est libre de produire autant de pas de calcul qu'il le souhaite, un *pas de calcul* étant la transformation d'une expression E_i en une autre, E_{i+1}, cf. figure 1. Son pas de calcul est correct si l'équivalence des deux expressions est conservée et erroné si elle ne l'est pas. Le domaine de l'algèbre, et plus particulièrement le domaine transformationnel de l'algèbre, étant « formel », il s'est avéré judicieux de représenter les actions des élèves sous formes de règles correctes ou erronées. Par exemple, dans l'exemple illustré en figure 1, une règle erronée peut expliquer le premier pas de calcul : a – (b + c) → a – b + a – c.

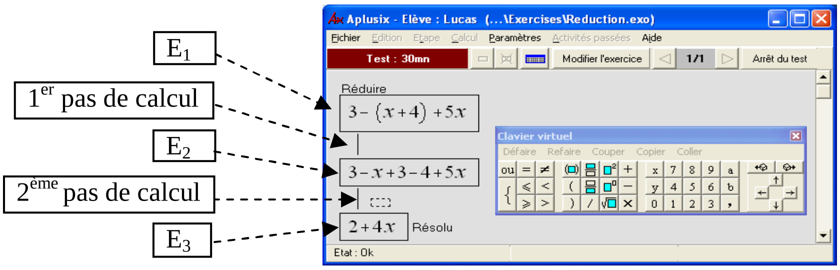

FIG 1. *Copie d'écran d'écran d'un élève ayant effectué deux pas de calculs Le premier comporte une transformation incorrecte et le second, trois transformations correctes : deux calculs numériques et une addition de monômes.*

Un pas de calcul peut comporter plusieurs transformations, applications de règles différentes, comme le montre le second pas de calcul de la figure 1. Nous découpons un pas de calcul en sous étapes élémentaires, chacune étant expliquée par l'application d'une règle[3], correcte ou erronée, comme le montre le tableau 1. Nous appellerons, dans la suite, ces sous étapes élémentaires, *transformations*. Chaque transformation est associée à une des quatre *tâches* de réduction, développement, factorisation ou de mouvement d'un membre dans une équation. Chaque tâche n'étant pas nécessairement la même que le thème de l'activité.

Pas de calcul	Activité	Transformations	Règle associée	Tâche
3(4 + 5x) = 2 → 15x = 6 – 12	Résolution d'équation	3(4 + 5x) → 12 + 15x	a(b + c)→ ab + ac	Développement
		12 + 15x = 6 → 15x = 6 – 12	a + b = c → a = c – b	Mouvement

TAB 1 – *Exemples de découpage d'un pas de calcul en transformations.*

2.2 Importance de la prise en compte du contexte algébrique

(Balacheff, 2000) explique que deux comportements, plus précisément dans notre cas l'utilisation de deux règles différentes, « peuvent sembler contradictoires, mais cette incohérence peut s'expliquer localement, soit par

2 Kieran appelle transformationnelle, l'approche qui consiste à changer la forme d'une expression/équation en conservant l'équivalence. Elle rappelle que des approches uniquement générationnelles, ou de modélisation, ne permettent pas de donner à l'élève tous les sens de l'algèbre : transformer une expression en une autre équivalente permet de relier une tâche algébrique à une réflexion conceptuelle.

3 Un diagnostic automatique a été construit, permettant d'associer à un pas de calcul une séquence de règles (Nicaud et al., 2005). Cette automatisation étant en cours d'amélioration, nous ne l'avons que partiellement utilisée pour les travaux exposés dans cet article.

le temps soit par la situation/contexte ». Précisons par un exemple : deux élèves, V et W, transforment l'expression $5x^4 + 3x^3$, en $8x^7$ en utilisant la règle erronée R : $ax^n + bx^p \rightarrow (a + b)x^{(n+p)}$. Supposons que, confrontés à l'expression $5x^2 + 3x$, l'élève V n'applique pas la règle R tandis que l'élève W l'applique en produisant $8x^3$. Devons-nous en déduire que le comportement de l'élève V est incohérent ? En fait, il se peut que V n'ait pas utilisé R dans le second cas parce que les exposants supérieurs à 3 n'ont pas le même statut, à ses yeux, que les exposants 1 et 2 : les *contextes algébriques* de $5x^4 + 3x^3$ et de $5x^2 + 3x$ sont différents. L'élève V a un comportement cohérent relativement à des contextes algébriques distincts.

Une des techniques classiques pour capturer la résistance d'un comportement à des situations différentes consiste à affecter à une règle un coefficient propre à chaque élève : le rapport du nombre d'utilisations de cette règle par le nombre d'opportunités de l'appliquer (Payne & Squibb, 1990). Seul le numérateur varie en fonction de l'élève. Plus ce coefficient est proche de 1, plus l'élève a une utilisation « stable » de cette règle. Cependant, que signifie le terme « opportunité d'application » ? Est-ce une occasion générique, ou, est-elle relative à l'élève ? Il est difficile de répondre de façon objective, car ce qui pourrait paraître être une opportunité d'application pour nous, ne le serait pas nécessairement pour l'élève. L'exemple étudié en 2.2 le montre : L'expression $5x^2 + 3x$ pourrait présenter, à première vue, une « opportunité d'appliquer » R. L'élève W la perçoit. Le nombre d'opportunités d'appliquer R devrait être crédité d'un point. Or, l'élève V n'applique pas R à cette expression et, comme nous l'avons expliqué, V n'en a pas pour autant une utilisation instable. Pour V, l'expression $5x^2 + 3x$ ne présente pas d'« opportunité d'appliquer » R. Pour cet élève, il ne faudrait pas diminuer le coefficient associé à R, en augmentant le dénominateur. Il y a dans cette technique, une anticipation forte des comportements des élèves.

Notre équipe, quant à elle, a utilisé une autre approche dans le domaine de la résolution des équations, qui repose sur une construction a priori des théorèmes-en-acte[4] d'élèves. Cette construction manuelle anticipe les regroupements possibles des contextes. Elle est implantée dans un processus automatique afin de discriminer un corpus d'élèves en fonction des théorèmes-en-acte préconstruits (Chaachoua et al., 2006).

Dans le domaine de la factorisation, du développement et de la réduction d'expressions algébriques, nous avons décidé d'utiliser une troisième technique, à savoir l'analyse statistique implicative (Gras, 1999). L'avantage de cette méthode sur les précédentes est qu'il n'y a pas de construction a priori : aucun lien n'est décrit entre les comportements ni entre les contextes. De plus, le travail effectué dans le domaine des équations n'est pas nécessairement transposable dans un domaine plus complexe comme celui du développement ou factorisation d'expressions algébriques, où le nombre de règles en jeu est plus important.

3 Traitement des activités : repérer des erreurs

Un nombre important de données est nécessaire pour construire un modèle précis des connaissances de l'élève. Il est difficile de les obtenir pour des raisons pratiques. D'une part, les contraintes institutionnelles et temporelles empêchent d'avoir des données nombreuses sur un même élève ayant travaillé une heure sur chaque type d'activité. D'autre part, il y a un risque, non négligeable, que, sur une longue durée d'activités dans un domaine précis, les connaissances de l'élève évoluent, rendant la description de la stabilité inopérante. Cependant, nous l'avons vu dans la section 2, une même activité peut nécessiter différentes tâches. Au cours d'activités portant sur la résolution d'équation, il est possible de recueillir un ensemble d'informations sur la réduction assez conséquent pour être représentatif des connaissances de l'élève dans ce domaine. Il faut donc, dans un ensemble d'activités, repérer les tâches qui peuvent susciter une modélisation : celles qui, à la fois, sont en nombre suffisamment important et qui suscitent, de la part de l'élève, un comportement stable. Pour cela, nous effectuons un premier traitement sur les transformations d'un même élève, en leur associant des attributs que l'on dénommera, dans la suite, comme « *grossiers* ». Un second traitement sera effectué avec de nouveaux attributs dits « *fins* » sur une sous-partie de l'ensemble des transformations repérées par le premier traitement, cf. section 4.

---

[4] Au sens de (Vergnaud, 1991).

3.1 Les attributs

Après qu'un élève ait travaillé dans Aplusix, un premier fichier est créé[5], comportant des informations sur les activités de cet élève. En lignes, se trouvent les transformations, tandis qu'en colonnes, se trouvent deux types d'attributs :

– Les *actions* qui prennent deux modalités : la transformation est soit correcte soit erronée.

– Les *variables du contexte*. Nous définissons une variable du contexte comme étant une caractéristique de l'expression algébrique ; par exemple, le degré de l'expression. Nous faisons l'hypothèse que ces variables peuvent être significatives chez un élève, c'est-à-dire, peuvent avoir un impact sur son comportement. Elles sont au nombre de 15. Nous les avons regroupées en quatre catégories : le degré de l'expression initiale, son opérateur, la nature de ses coefficients (entiers, fractionnaires etc.) et la tâche associée à l'action.

Les valeurs des attributs sont, dans cette étude, binaires. Le fichier, traité par le logiciel CHIC, est une matrice constituée, en colonnes, par les 17 attributs et, en lignes, par les expressions algébriques. La valeur 1 à la ligne i, colonne j, signifie que la i-ème transformation peut être décrite par le j-ème attribut. Le tableau 2 en donne un exemple. Nous souhaitons trouver des liens entre les actions et les 15 variables de contexte.

Transformation	Erroné	Correct	Réduction	Développement	Opérateur+	Opérateur$\times$	Degré >3	Degré0
$5x^2 \times 7x^3 \rightarrow 21x$	1	0	1	0	0	1	1	0

TAB 2 –*Extrait d'une description grossière d'une transformation*

3.2 Etude de cas

Les analyses qui suivent ont été faites sur quelques productions d'élèves de classe de 3ème. Ils ont travaillé consciencieusement sur les expressions proposées pendant environ une heure, en mode « test », donc sans rétroaction de la part d'Aplusix, et sans intervention d'enseignant. Leurs réponses étaient automatiquement enregistrées par Aplusix. Les productions viennent de deux types d'expérimentations. La première a eu lieu dans le cadre du projet « Ecole et sciences cognitives » (Nicaud et al., 2005). Elle avait pour objet principal l'étude de la résolution d'équations. La seconde était constituée d'activités de factorisation et s'est déroulée en 2006. Nous avons testé ce que l'Analyse Statistique Implicative pouvait apporter comme informations sur l'ensemble des quatre tâches. Nous donnons en figure 2 les liens que le logiciel CHIC produit à partir des données sur un élève, l'élève Jean[6], ayant participé à la première expérimentation.

Ce premier fichier comporte les 99 transformations (élémentaires) de l'élève. Comme l'indique le graphe implicatif, figure 2, au seuil 93, cet élève semble maîtriser l'action du mouvement dans les équations, notée « Mouvement ». Quand la tâche est une réduction d'une soustraction, « InOpMinus », avec des coefficients entiers, « WIntCoef », la transformation a tendance à être erronée. Ces informations sont importantes : elles semblent indiquer que, de manière stable, une tâche de soustraction provoque des erreurs chez cet élève. Au seuil 87, apparaît le fait que Jean a des difficultés particulièrement quand l'expression initiale est de degré 1, « InDeg1 ». On peut noter que rien n'est indiqué sur des transformations de développement ou de factorisation. Ceci s'explique soit par un comportement instable de l'élève soit par un manque de données sur ces tâches.

5 L'automatisation de ce processus est en cours.

6 Le nom des élèves a été modifié par respect pour leur anonymat.

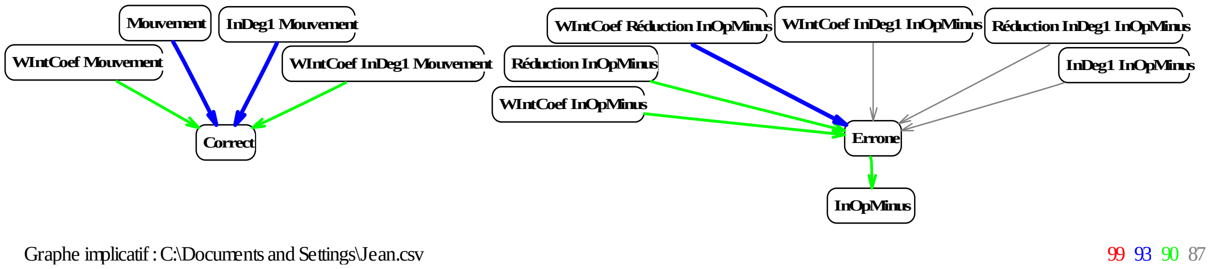

FIG 2. *L'analyse "grossière" a pour but de discriminer les tâches posant des difficultés à l'élève.*

Ce premier traitement répond à notre attente : l'Analyse Statistique Implicative permet de discriminer les variables de contexte suscitant des comportements stables. Dans ce but, les comportements étaient volontairement décrits « grossièrement ». Ceci demande, dans un second temps, à être affiné, ce que nous faisons dans la section suivante à l'aide de deux exemples.

4 Second traitement : traitement spécifique à une tâche

Le traitement effectué dans cette section porte à nouveau sur des transformations (élémentaires). Nous ne prenons, pour cette analyse, que les transformations repérées par le premier traitement. Ces transformations sont décrites dans cette section *finement* : sous la forme de règles.

4.1 Les attributs

A nouveau le fichier comprend en colonnes deux types d'attributs :

– Les *actions* vues comme des règles. Chaque transformation est décrite de manière *fine* en précisant la règle qui peut l'expliquer. Par exemple, dans la réduction de somme de deux termes, nous avons écrit 4 règles :
PlusCorrect : a + b → a + b,
PlusOp (PlusOpposé) : a + b → a – b,
OpPlusOp (OpposéPlusOpposé) : a + b → – a – b,
OpPlus (OpposéPlus) : a + b → – a + b.

– Les *variables du contexte*. Elles dépendent du type de règle. Ainsi, pour les quatre règles précédentes, nous avons retenu six variables : quatre pour les signes de a et de b et deux pour l'ordre de |a| et de |b|. Nous avons ajouté deux variables supplémentaires : la nature de a et de b (numérique versus monôme).

Transformation	PlusCorrect	PlusOp	(–,–)	(–,+)	(+,–)	(+,+)	|a|>|b|	|a|<|b|	Numérique S	Monôme S
– 4x – 3x → – x	0	1	1	0	0	0	1	0	0	1

TAB 3 – *Extrait d'une description fine d'une transformation*

4.2 Types d'implications

Notons, C, les variables de contexte et R, les règles (ou actions). Dans notre travail, quatre types de relation entre attributs peuvent apparaître :

– $C_i \Rightarrow C_j$; Cette quasi-implication est interprétable de deux manières. D'une part, elle peut être évidente mathématiquement. Par exemple, si l'expression est de degré 3, alors l'expression est de degré supérieur à 2. D'autre part, cette implication peut apparaître parce que les contextes n'ont pas été suffisamment combinés. Par exemple, si l'opérateur est « diviser », alors, l'expression est de « degré 0 ». Cette implication est liée aux choix faits lors de la constitution du corpus des expressions : il n'y avait pas de fractions rationnelles, une division était nécessairement numérique. En effet, certaines combinaisons de variables n'ont pas été envisagées en fonction du niveau des élèves.

– $R_i \Rightarrow R_j$; Cette quasi-implication indique que, lorsque l'élève utilise R_i, il utilise *souvent* R_j. Dans notre expérimentation, il est impossible de l'obtenir. En effet, pour une expression donnée, l'élève que nous considérons, n'a qu'un seul comportement : l'utilisation de R_i ou de R_j est exclusive l'une de l'autre puisque une ligne correspond à une transformation élémentaire. Cependant, en termes de perspectives, cette relation pourrait être intéressante. Nous pourrions prendre, en ligne, non pas les transformations mais les pas de calcul, associer aux pas l'ensemble des règles les expliquant. Ces quasi-implications de l'utilisation d'une règle à une autre pourront alors ressortir. Elles permettraient d'expliciter quelles sont les règles dont la maîtrise est une condition suffisante à l'utilisation de telle autre règle.

– $R_i \Rightarrow C_j$; Cette quasi-implication signifie que le contexte C_j est *souvent* vérifié lorsque la règle R_i est utilisée. Ce qui est plus compréhensible par la contraposée : si l'expression n'est pas de contexte C_j, alors l'élève n'utilise presque jamais R_i.

– $C_j \Rightarrow R_i$; Cette quasi-implication exprime que lorsque l'expression est de contexte C_j, l'élève utilise *presque* toujours R_i.

Nous sommes intéressés par ces deux dernières implications. A cette fin, nous utilisons une option de CHIC qui permet de sélectionner uniquement certains attributs comme « sommets principaux ». Seuls les ancêtres et descendants des attributs choisis sont alors dessinés dans le graphe implicatif. Les parents et enfants directs sont en ligne pleine, tandis que les autres, comme les grands-parents, sont en pointillés. En choisissant les règles comme sommets centraux, cela nous permet de ne visionner que les deux derniers types d'implication.

4.3 Etudes de cas

Les attributs construits par une analyse a priori se veulent génériques : les mêmes attributs (pour une règle ou un ensemble de règles), quel que soit l'élève. L'analyse précise qui suit se limite à quelques productions d'élèves. Elle a pour but de valider le choix des attributs : sont-ils suffisamment précis pour modéliser les comportements de l'élève, pour faire ressortir la stabilité des actions ? Nous pouvons répondre par l'affirmative, grâce aux deux exemples présentés ci-après.

4.3.1 Un premier exemple

Nous avons souhaité affiner le premier « diagnostic » des erreurs repérées en section 3. Nous nous sommes concentrés sur les 21 transformations qui ont contribué à l'implication *(Réduction* $\wedge$ *InOpMinus* $\wedge$ *WIntCoef)* $\Rightarrow$ *Erroné*, ressorties lors du premier traitement. Nous avons ajouté les 24 autres transformations qui concernaient le contexte *(Réduction* $\wedge$ *WIntCoef)*.

Une analyse implicative sur une description fine du contexte et des actions permet de produire les comportements indiqués en figure 3. Ainsi, la position du signe moins et l'ordre de |a| et de |b| ont un impact sur le type d'action de l'élève. Expliquons ces implications :

– n°1 et n°2 : lorsque (a et b sont de signes positifs) ou que (a est de signe positif et b négatif et |a| est plus grand que |b|), la transformation est correcte (*PlusCorrect*) : somme de a et de b ; Exemple : 5–2→3 ou 5+2→7.

– n°3 : lorsque (a et b sont de signe négatif et |a| est plus petit que |b|), la transformation consiste à soustraire |a| à |b| (*OpPlus*) ; Exemple : – 2 – 5→ – 3.

– n°4 et n°5 : lorsque (a et b sont de signe négatif et |a| est plus grand que |b|) ou (a est de signe négatif et b positif), la transformation consiste à additionner a à l'opposé de b (*PlusOp*) ; Exemple : –5 – 2→ –3 ou –5 + 2→– 7 ou –2 + 5 → – 7.

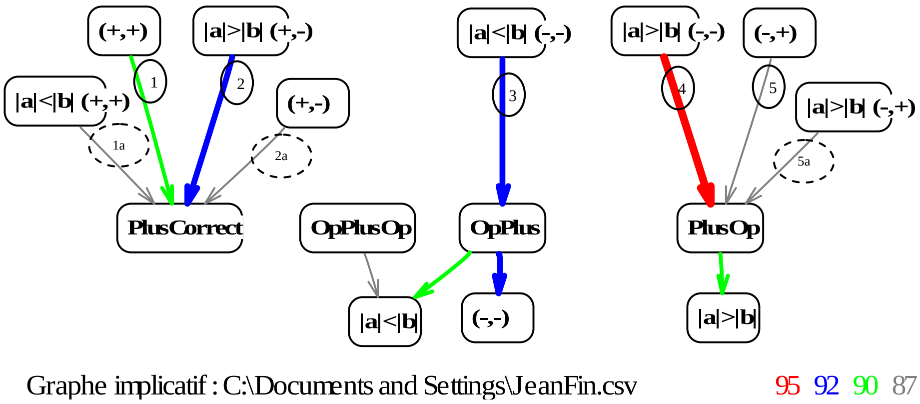

FIG 3. *Graphe implicatif de la réduction de a+b, où a et b sont de signe quelconque. Sur la première horizontale se situent les conditions principales, en seconde ligne, les conditions moins importantes (numéros en pointillé).*

Le graphe implicatif ayant fait ressortir ces implications, cela dénote une stabilité : les actions sont certes différentes mais reliées à un contexte précis, elles présentent une cohérence. Une lecture des liaisons conjointes permet d'associer à ces implications ($C_j \Rightarrow R_i$), la règle abstraite[7] R_a, non construite a priori :

R_a : a+b → Signe(a) (a□b),
où □ est une addition si a et b sont de signe positif et une soustraction, sinon.

Cette règle est correcte dans les cas 1 et 2. Le graphe implicatif montre, cependant, deux phénomènes qui peuvent sembler incohérents, de prime abord, avec cette règle abstraite :

– l'implication 2a, ((+,–)⇒PlusCorrect), ressort et ce, quel que soit l'ordre de |a| et de |b|. Or, pour être cohérente avec R_a, la condition (+,–) n'est pas suffisante. Les variables supplémentaires peuvent expliquer cette incohérence. En effet, la variable supplémentaire la plus typique de 2a est Monôme (versus Numérique) avec un *risque* de 0.487 (respectivement 0.668). Ceci signifie que cette implication a tendance à avoir plus lieu lorsque a et b sont numériques que lorsque ce sont des monômes. Rappelons que l'action est correcte. Nous pouvons expliquer ce comportement par le fait que Jean est en phase d'apprentissage : il maîtrise mieux l'arithmétique que l'algèbre. Jean commence à perdre l'habitude d'utiliser R_a, dans le contexte d'un nombre positif à additionner à un nombre négatif. Cependant, confronté à de l'algèbre, son ancienne conception erronée ressurgit.

– lorsque (a et b sont de signe négatif et |a| est plus petit que |b|), ou lorsque (a est de signe positif et b négatif et |a| est plus petit que |b|), le graphe implicatif n'apporte pas d'informations. Or, nous avons vu que la variable supplémentaire Monôme avait un impact sur le comportement de Jean. En utilisant cette variable en variable principale, et avec 3 prémisses, ces informations ressortent au seuil 72 : ((+,–)∧|a|<|b|∧Monôme)⇒OpPlusOp. La transformation 4x-7x→3x, par exemple, contribue à cette implication.

4.3.2 Un second exemple

Le cas étudié précédemment est une précision d'une implication apparue lors du traitement « grossier ». Dans ce second exemple, nous allons affiner les descriptions des transformations sur lesquelles le premier traitement n'a apporté aucune information. Nous l'avons dit, ceci peut s'expliquer par une instabilité du comportement (transformation parfois correcte et parfois incorrecte dans une même tâche) ou par un manque de données. Le cas traité ici porte sur la tâche de factorisation dans des activités de factorisation, chez l'élève Jeanne. 41 transformations sont concernées. Cinq actions ont été repérées pour cette élève :

– l'expression n'est pas traitée, noté *NonRens (non renseigné)* ;

– l'expression est correctement factorisée, noté *RES (résolu)* ;

– l'expression est incorrectement factorisée et l'erreur porte sur le signe moins, noté *ER_SM (erreur signe moins)* ;

[7] Nous réfléchissons actuellement à un processus automatique qui permettrait de généraliser automatiquement les implications trouvées.

– l'expression est incorrectement factorisée et l'erreur est l'utilisation de la règle, notée *ER_Rien*, que nous explicitons sur un exemple : l'expression (8y – 5)(6y + 3) + (8y – 5) est transformée en (8y – 5)(6y + 3). L'interview qui a suivi l'activité a permis de justifier notre diagnostic a priori : Cette transformation n'est pas due à la perte d'un terme mais à l'application de la règle, ab + a $\rightarrow$ a(b), la mise en facteur de « a » dans le terme « a » laissant « rien » comme cofacteur ;

– l'expression est incorrectement factorisée et l'erreur n'est ni *ER_SM* ni *ER_Rien*. On la note *ER_Autre*.

D'autres règles erronées ont été écrites a priori, en amont de l'expérimentation, mais n'ont pas été utilisées par cette élève. Les variables de cette tâche sont au nombre de 29 (nature du facteur commun, des cofacteurs, leurs degrés, la visibilité du facteur commun, etc.). Avec une seule prémisse, au seuil de confiance 86 (Gras, 2004), différentes informations sur l'élève se dégagent du graphe implicatif, cf. figure 4 :

– l'expression n'est pas traitée, *NonRens*, lorsque le facteur commun ou un des cofacteurs sont des produits, respectivement *FProduit* et *Cof2Poly*, ou lorsque le facteur commun est une somme non adjacente, *FSum2Term_NoAdj*, (exemple : 2x+(x+2)*3+x²) ou opposée, *FSum2Term_Opp*, (exemple :–2x–7+(2x+7)(x+5)). Ces implications montrent que le simple fait de ne pas traiter un exercice peut être informatif sur la difficulté que perçoit l'élève.

– l'expression est correctement factorisée, *RES*, lorsque le facteur commun est numérique, *FNum*. Et plus précisément *FNum_1Fact*. L'expression –2x–14 contribue à l'implication « FNum $\Rightarrow$ RES ».

– La règle *ER_Rien* a comme condition nécessaire un cofacteur unitaire, noté *CofUnitVA* (exemple : (x – 4) + (x – 4)x $\rightarrow$ (x – 4)(x) ou y(5y – 1) – y $\rightarrow$ y(5y – 1)). Ce qui n'est pas surprenant et montre que le modèle a du sens. Cependant, avec deux prémisses au seuil 90, l'implication (CofUnitVA $\wedge$ FSum2Term) $\Rightarrow$ Er_Rien apparaît. L'élève Jeanne utilise particulièrement la règle Er_Rien lorsque le facteur commun est une somme évidente de deux termes (que l'on peut opposer à un facteur commun numérique, comme dans l'expression 5y+5).

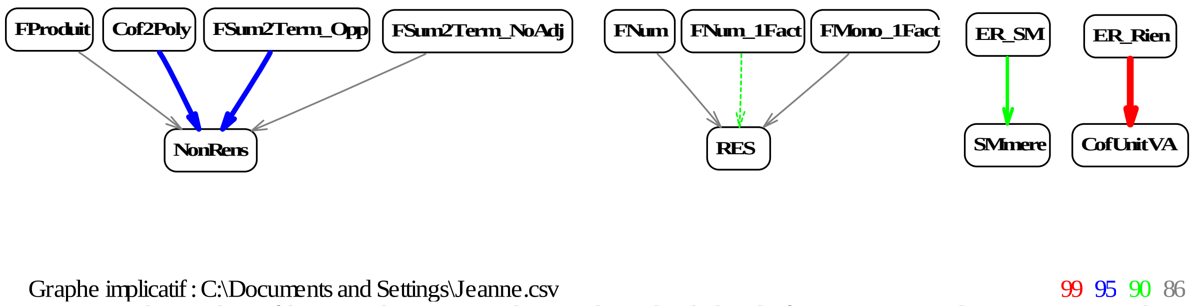

FIG 4. *Graphe implicatif à une prémisse entre les attributs de règles de factorisation et de contextes. Seuil 86.*

5 Conclusion

L'analyse statistique implicative a répondu à nos deux attentes. D'une part, elle a permis de discriminer des transformations sur des critères d'exactitude en les groupant par conditions portant sur les variables de contexte. D'autre part, le graphe implicatif a fait ressortir l'utilisation régulière de règles associées à des contextes algébriques précis. Ce résultat valide aussi nos choix d'attributs. Ce qui répond à notre désir d'obtenir un modèle d'élève qui ne capture pas la simple apparition intermittente d'un comportement, mais la persistance et la résistance de cette action à des contextes différents.

Nous avons, ainsi, pu obtenir une représentation des connaissances de l'élève. Le mot connaissance est vu, ici, comme un couple d'une règle et du contexte de son utilisation. Les résultats présentés dans cet article étant encourageants, nous prolongeons actuellement notre travail dans trois directions. En premier lieu, quoique l'utilisation de l'analyse statistique implicative soit satisfaisante, nous cherchons à comparer son apport aux autres techniques d'analyse de données, telle l'analyse des correspondances multiples. En parallèle, nous automatisons l'ensemble du processus : diagnostic des actions, remplissage des fichiers, reconstruction des règles abstraites. Le troisième travail porte sur la recherche de simplification de la lecture des résultats des graphes implicatifs. Elle nécessite, en effet, une certaine expertise. Nous souhaitons associer à une *séquence* (ou ensemble) de liens implicatifs, des règles abstraites sur lesquelles il sera possible de reconstruire des conceptions, sorte de réseau de règles. Il faut, pour ce faire, détecter quelles sont les séquences de liens implicatifs récurrentes chez les élèves, leur donner ensuite un sens et leur associer une règle abstraite. Il serait ainsi intéressant de s'intéresser non à un élève mais à un groupe d'élèves et de détecter, toujours pour une tâche donnée, quelles sont les séquences d'implications les plus significatives, en terme d'occurrence. Un graphe implicatif construit à partir d'une simple jonction des fichiers de chaque élève ne répondrait pas à notre attente. En effet, chaque ligne (associée à une transformation) deviendrait indépendante de l'élève ; il n'y aurait plus de distinction entre chaque élève. Les résultats de l'analyse perdraient alors leur sens cognitif car ce que l'on souhaite est bien les séquences d'implications les plus utilisées et non quelles sont les implications (contexte $\Rightarrow$ action) les plus fréquentes. Par exemple, si un groupe d'élèves utilise régulièrement les règles R_1 dans les contextes respectifs C_1 et C_2 et si un autre groupe utilise aussi régulièrement les règles R_2 et R_1 dans les contextes respectifs C_1 et C_2, les séquences que l'on souhaite faire ressortir sont les séquences d'implication $\{C_1 \Rightarrow R_1,\ C_2 \Rightarrow R_1\}$ *et* $\{C_1 \Rightarrow R_2,\ C_2 \Rightarrow R_1\}$. Une simple jonction des fichiers ferait ressortir une instabilité d'utilisation de règle dans le contexte C_1 et l'utilisation de R_1 dans le contexte C_2. Il faut donc qu'une *seule* ligne soit associée à *un* élève et qu'elle soit représentative des séquences que l'élève utilise. Nous nous proposons de construire de telles lignes associées non plus à des transformations mais à la séquence la plus représentative de chaque élève. Nous expliquons dans (Croset et al., 2007) comment nous construisons ce nouveau type de fichier et comment nous l'exploitons à l'aide de graphes de similarités pour faire ressortir des séquences d'implications fréquentes chez les élèves.

Références

Balacheff N., Les connaissances, pluralité de conceptions. Le cas des mathématiques, Actes de la conférence Ingénierie de la connaissance (IC 2000), Toulouse, 2000.

Bouhineau D., Nicaud J.-F., Pavard X., Sander E., Un micromonde pour aider les élèves à apprendre l'algèbre, Sciences et Techniques Educatives, Numéro spécial : Environnements Interactifs d'Apprentissage avec Ordinateur EAIO'2001, Vol. 8, n°1-2, p 33-47, 2001.

Brousseau G., Les erreurs des élèves en mathématiques Petit X n°57, 2000.

Chaachoua H., Bittar M., Nicaud J.-F., Student's modelling with a lattice of conceptions in the domain of linear equations and inequations, Actes PME30, 2006.

Croset M.-C., Trgalova J., Nicaud J.-F., Student's Algebraic Knowledge Modelling: Algebraic Context as Cause of Student's Actions. Soumis pour publication dan**s** "The Statistical Implicative Analysis – Its theoretical foundations" Régis Gras, Einoshin Suzuki, Fabrice Guillet, Filippo Spagnolo (Ed.).

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

Gras R., Les fondements de l'analyse statistique implicative, Quaderni di Ricerca in Didattica, Vol. 9, 1999.

Gras R., L'analyse implicative : ses bases, ses développements, Revue Educaçao Matematica Pesquisa, Vol. 4, n°2, p 11-48, 2004.

Kieran C., The core of algebra: Reflections on its Main Activities, ICMI Algebra Conference, Melbourne, Australia, 2001.

Nicaud J.-F., Chaachoua H., Bouhineau D., Bittar M., Modélisation cognitive d'élèves en algèbre et construction de stratégies d'enseignement dans un contexte technologique, Project report of the "Ecole et sciences cognitives" research programme. Cahier du laboratoire Leibniz n°123, http://www-leibniz.imag.fr/, 2005.

Payne S.J., Squibb H.R., Algebra Mal-Rules and Cognitive Accounts of Error, Cognitive Science, Vol. 14, n°3, p 445-481, 1990.

Sleeman D.H., Kelly A.E., Martinak R., Ward R.D., Moore J.L., Studies of Diagnosis and Remediation with High School Algebra Students, Cognitive Science, Vol. 13, p 551-568, 1989.

Vergnaud G., La théorie des champs conceptuels, Recherches en Didactique des Mathématiques, Vol. 10/2.3, p 133-170, 1991.

Summary

We construct a student model in the algebra field. For gathering the data, we have used the Aplusix learning environment that allows students to make freely calculation steps and records all the students' actions. In a problem-solving environment, one way to build and update the student model is to precisely follow what the student is doing, by means of a detailed representation of cognitive skills.

The challenge is therefore to reconstruct high-level behaviors from this low-level data. We try to discover patterns of student behaviors. For that, we use the statistical implicative analysis which makes it possible to seek the stability of the actions and to determine the contexts where they appear. Implicative analysis allows us to build implicative connections between contexts towards actions.

Resumen

El modelo de alumno que construimos pretende dar una representación fina de las competencias cognoscitivas de los alumnos en el ámbito de la álgebra. Se base en las microacciones registradas en el software Aplusix. Estas acciones que son de nivel muy fino, tenemos que reconstruir los conocimientos a un nivel más abstracto: no una acción aislada sino un conjunto de acciones "similares" que aparecen en un mismo contexto. Utilizamos para eso el análisis estadístico implicative que permite hacer resultar la estabilidad de las acciones y de determinar los contextos donde tuvieron lugar, construyendo conexiones implicatives del contexto hacia la acción.