



**HAL**  
open science

# Amélioration de l'interprétabilité du diagnostic cognitif de l'apprenant par catégorisation des composantes de connaissance

Olivier Allègre, Amel Yessad, Vanda Luengo

► **To cite this version:**

Olivier Allègre, Amel Yessad, Vanda Luengo. Amélioration de l'interprétabilité du diagnostic cognitif de l'apprenant par catégorisation des composantes de connaissance. 10e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain, Marie Lefevre, Christine Michel, Jun 2021, Fribourg, Suisse. pp.34-45. hal-03292665

**HAL Id: hal-03292665**

**<https://hal.science/hal-03292665v1>**

Submitted on 23 Jul 2021

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

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

# Amélioration de l'interprétabilité du diagnostic cognitif de l'apprenant par catégorisation des composantes de connaissance

Olivier Allègre<sup>1,2</sup>, Amel Yessad<sup>1</sup>, and Vanda Luengo<sup>1</sup>

<sup>1</sup> MOCAH, LIP6, Sorbonne Université, Paris, France

<sup>2</sup> Kartable, [www.kartable.fr](http://www.kartable.fr)

**Abstract.** Le diagnostic cognitif permet de modéliser les connaissances d'un apprenant à partir de ses interactions avec un EIAH pour la recommandation adaptative de ressources ou la prédiction de l'état futur de l'apprenant. En étudiant les performances des algorithmes des modèles *knowledge tracing* et *factor analysis* sur des jeux de données restreints aux composantes de connaissance (CC) déclaratives ou procédurales, nous montrons que les modèles *factor analysis* sont plus performants sur des CC déclaratives, alors que les modèles *knowledge tracing* réagissent mieux sur les CC procédurales, même s'il est compliqué de comparer les deux modèles l'un à l'autre. Ces résultats suggèrent l'intérêt d'une catégorisation des CC par leur nature dans le choix des algorithmes de diagnostic cognitif, et que cette catégorisation permet l'amélioration de l'interprétabilité du modèle de l'apprenant.

**Keywords:** Apprentissage adaptatif · Diagnostic cognitif · Fouille de données éducationnelles · Modélisation de l'apprenant.

**Abstract.** Cognitive diagnosis is used to model the learner knowledge from the learners' traces on an ITS for adapted feedback or prediction of learner future state. By studying the performance of algorithms of *knowledge tracing* and *factor analysis* models on datasets restricted to either declarative or procedural knowledge components (KC), we show that *factor analysis* models outperform on declarative KCs while *knowledge tracing* ones outperform on procedural KCs, even if it is complicated to compare them. These results suggest the interest of categorizing the KCs by nature in cognitive diagnosis algorithms and that this categorization allows the enhancement of learner model interpretability.

**Keywords:** Adaptive learning · Cognitive diagnosis · Educative data mining · Learner modeling.

## 1 Introduction

Personnaliser l'enseignement est une des promesses de l'intelligence artificielle en éducation. Par l'intermédiaire de la modélisation de l'apprenant, il est possible de proposer aux élèves des ressources adaptées à leur niveau, dans un

but d'apprentissage ou d'évaluation. La visée du modèle de l'apprenant est de déterminer l'état dans lequel ce dernier se situe. Via la modélisation des capacités cognitives, de la motivation ou encore de la méta-cognition, l'état latent prédit de l'apprenant peut permettre d'adapter la rétroaction et la recommandation du EIAH.

Nous ne nous intéresserons dans ce papier qu'à la modélisation du savoir et des capacités de l'apprenant, très utilisée dans les systèmes éducatifs adaptatifs [11]. Afin de définir de tels modèles, l'usage le plus courant est de définir et de modéliser le contenu d'un domaine d'étude, par l'intermédiaire de composantes de connaissance (CC) qui correspondent aux constituants du savoir (e.g. "Connaître la table de multiplication de 3" ou "Accorder un adjectif régulier dans un groupe nominal minimal"). Les CC sont associées aux ressources pédagogiques consultables par l'apprenant ainsi qu'aux exercices permettant de l'évaluer. À l'aide d'algorithmes de diagnostic cognitif, il est enfin possible de déterminer des métriques de maîtrise de ces CC à partir des traces des apprenants sur les exercices. Beaucoup de caractéristiques de l'apprenant sont modélisables [2] et, même au sein de la modélisation de son savoir, il est possible d'adopter des stratégies de diagnostic cognitif extrêmement diverses. Le choix du modèle n'est pas simple : il est nécessaire d'avoir à la fois de bonnes performances sur la modélisation de l'apprenant et, spécifiquement dans le domaine de l'éducation, une interprétabilité du modèle par les enseignants [12] [10] et parfois même par les apprenants [4].

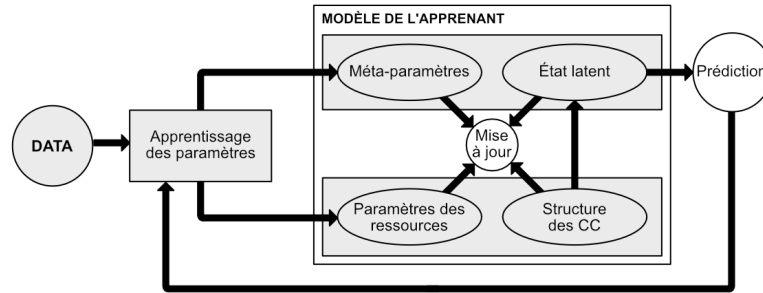
Cohen et Squire [5] introduisent en 1980 le terme de mémoire déclarative, mémoire du conscient incluant principalement des faits et des événements, qu'ils opposent à la mémoire procédurale, inconsciente cette fois, permettant d'acquérir capacités cognitives et motrices. En différenciant la nature déclarative ou procédurale du savoir de l'apprenant, nous pouvons donc émettre l'hypothèse que les CC, elles-mêmes, correspondent à l'une ou à l'autre. Nous nous interrogerons dans ce papier sur l'impact du choix des algorithmes de diagnostic cognitif sur le modèle de l'apprenant et ses constituantes procédurales et déclaratives. La catégorisation en nature des CC permet-elle une meilleure interprétabilité du diagnostic cognitif ?

Dans un premier temps, nous étudions la modélisation de l'apprenant et nous focalisons sur les algorithmes permettant la prédiction du savoir de l'apprenant. Nous détaillons ensuite les paramètres de ces algorithmes. Enfin, nous présentons l'étude expérimentale que nous avons menée sur des jeux de données de mathématiques et de français filtrés par la nature des CC et nous discutons les résultats obtenus.

## 2 Travaux associés

Le choix des éléments modélisés dans la modélisation de l'apprenant est large, mais nous ne nous intéressons qu'aux modèles basés sur les connaissances de l'apprenant, par le biais d'un modèle du domaine sous forme de décomposition

en CC. Cette structure de modèle de l'apprenant est représentée dans la Figure 1.



**Fig. 1.** Structure du modèle de l'apprenant, illustration inspirée de [11] : le modèle de l'apprenant comporte des équations de mise à jour de prédiction afin de déterminer l'état latent de l'apprenant à partir des paramètres de l'apprenant (méta-paramètres) et des ressources, eux même appris à partir des données issues des traces des apprenants.

La majorité des systèmes éducatifs adaptatifs présentés dans la littérature sont basés sur ce type de modèle [11] [1]. Chaque système s'appuie sur des équations de mise à jour et de prédictions spécifiques et des paramètres liés à l'apprenant (méta-paramètres) ou au domaine (paramètres des ressources), qui rendent possible l'interprétation du modèle de l'apprenant.

Nous nous focalisons sur les modèles *knowledge tracing* et *factor analysis* dans ce papier. On indexe les apprenants par  $i \in \llbracket 1, I \rrbracket$ , les exercices par  $j \in \llbracket 1, J \rrbracket$  et les CC par  $k \in \llbracket 1, K \rrbracket$ . On note  $Y_{i,j} \in \{0, 1\}$  la justesse de la réponse, binaire, de l'élève  $i$  à l'exercice  $j$ ,  $p_{ij}$  la probabilité que l'élève  $i$  réponde correctement à la question  $j$  et  $\sigma$  la fonction logistique.

## 2.1 Knowledge tracing

Le modèle *knowledge tracing* repose sur la modélisation des processus d'apprentissage et d'oubli. Le *bayesian knowledge tracing* (BKT), un modèle caché de Markov, est le modèle *knowledge tracing* le plus répandu, où l'apprentissage est modélisé par une transition d'un état de non maîtrise d'une composante de connaissance à un état de maîtrise [6] à l'aide de quatre paramètres :  $P_i$  la probabilité a priori,  $P_l$  la probabilité de transition entre l'état de non-maîtrise à l'état de maîtrise,  $P_g$  la probabilité de deviner la bonne réponse à un exercice sans maîtriser la CC associée,  $P_s$  la probabilité de ne pas répondre correctement à un exercice malgré la maîtrise de la CC associée. L'apprentissage de tous

ces paramètres est généralement exécuté par un algorithme de maximisation de l'espérance. Nous considérons dans ce papier la version classique de BKT (que nous notons BKT) et trois de ses variantes :

- IDE [9], dans lequel les probabilités  $P_g$  et  $P_s$  sont dépendantes des exercices
- ILE [13], qui considère la probabilité  $P_l$  dépendante de l'exercice
- PPS [8], dans lequel la probabilité  $P_i$  dépend de la justesse de la première tentative de l'apprenant

## 2.2 Factor analysis

Le modèle *factor analysis* est fondé sur l'analyse de certaines observations par le biais d'un modèle logistique. Il permet de déterminer les facteurs communs parmi les interactions de l'ensemble des apprenants sur les ressources pédagogiques. Le choix des observations diffère entre les algorithmes du modèle *factor analysis* et mène à un ensemble de paramètres sur les apprenants et les ressources. Nous étudions ici quatre algorithmes de *factor analysis*, dont on précise les paramètres et l'équation de mise à jour.

- L'algorithme *item response theory* (IRT) modélise la capacité  $\theta_i$  de l'apprenant  $i$  ainsi que la difficulté  $d_j$  de la question  $j$ .

$$p_{ij} = \sigma(\theta_i \cdot d_j - \delta_j) \quad (1)$$

- L'algorithme *additive factor model* (AFM) modélise l'apprenant à l'aide de la difficulté  $\beta_k$  de la CC  $k$ , du biais  $\gamma_k$  de maîtriser la CC  $k$  pour chaque essai et  $N_{ik}$  le nombre d'essais de l'apprenant  $i$  sur un exercice associé à la CC  $k$

$$p_{ij} = \sigma \left( \sum_{k \in KC(j)} \beta_k + \gamma_k N_{ik} \right) \quad (2)$$

- L'algorithme *performance factor analysis* (PFA) modélise l'apprenant à l'aide du biais  $\beta_k$  de la CC  $k$ , des biais  $\gamma_k$  et  $\delta_k$  de maîtriser la CC  $k$  pour chaque essai réussi ou non, et du nombre de succès  $S_{i,k}$  et d'échecs  $F_{i,k}$  sur des exercices associés à la CC  $k$ .

$$p_{ij} = \sigma \left( \sum_{k \in KC(j)} \beta_k + \gamma_k S_{i,k} + \delta_k F_{i,k} \right) \quad (3)$$

- L'algorithme *learning factor analysis* (LFA) modélise l'apprenant à l'aide du biais  $\alpha_i$  de chaque apprenant, et comme pour l'additive factor model, la difficulté  $\beta_k$  de la CC  $k$ , des biais  $\gamma_k$  et  $\delta_k$  de maîtriser la CC  $k$  pour chaque essai réussi ou non, et du nombre de succès  $S_{i,k}$  et d'échecs  $F_{i,k}$  sur des exercices associés à la CC  $k$ .

$$p_{ij} = \sigma \left( \alpha_i + \sum_{k \in KC(j)} (\beta_k + \gamma_k S_{i,k} + \delta_k F_{i,k}) \right) \quad (4)$$

### 3 Méthodologie

#### 3.1 Données exploitées

Nous utilisons les données issues de traces d'apprenants de CE1 et CE2 sur des exercices interactifs sur l'ensemble du programme scolaire de français et de mathématiques de CE1 et CE2, issus de la plateforme web Kartable ([www.kartable.fr](http://www.kartable.fr)). La majorité des exercices sont des questions à choix multiples, où le nombre de réponses attendues est donné (environ 90% ; le reste correspond à des questions libres, des couples à relier ou des glisser-déposer). Les observations présentes dans ces jeux de données sont les suivantes : utilisateur, exercice réalisé, CC associées, nombre d'essais, nombre de succès, nombre d'échecs et date de l'interaction.

La Table 1 décrit le jeu de données en fonction de la matière des exercices desquels sont issues les traces des apprenants. Nous avons choisi de prendre deux matières (mathématiques et français) dont la structure pédagogique est très différente afin de vérifier si nos résultats dépendent de la matière. Par la suite, les mathématiques seront désignées par MAT et le français par FRA.

**Table 1.** Récapitulatif des ressources et des utilisateurs du jeu de données utilisé dans ce papier. Le terme "exercice" correspond aux questions présentées aux élèves tandis que le terme "interactions" correspond à la réponse d'un apprenant à cette question.

Matière	Nombre d'apprenants	Nombre d'exercices	Nombre d'interactions
MAT	6734	2314	1 393 525
FRA	7309	2880	2 093 946

#### 3.2 Catégorisation des composantes de connaissance

Les composantes de connaissance sont catégorisées en fonction de leur nature déclarative ou procédurale. Cette catégorisation des CC est effectuée à partir de la définition des mémoires déclarative et procédurale [3]. Les CC déclaratives portent sur un ensemble de connaissances factuelles ou explicites. Les CC procédurales, elles, portent sur des schémas de raisonnements implicites ou tacites.

La typologie des CC renseigne en partie la nature de ces dernières : l'usage d'un verbe d'action est d'usage dans la dénomination des CC et il donne une indication sur leur caractère déclaratif ou procédural. Une CC nommée "Savoir..." ou "Connaître..." aura tendance à être une CC déclarative, tandis qu'une CC nommée "Calculer..." ou "Déterminer..." sera plus facilement une CC procédurale. La catégorisation est automatiquement déterminée de cette façon, puis elle est comparée à une classification effectuée par deux enseignants sur un échantillon de 30 CC issues de l'ensemble des CC étudiées dans notre papier. L'accord entre la classification automatique et celle des enseignants a été mesuré par le calcul du kappa de Cohen  $\kappa = (p_a - p_e)/(1 - p_e)$ , où  $p_a$  est l'accord relatif

entre des enseignants avec la classification automatique et  $p_e$  la probabilité d'un accord aléatoire. Nous obtenons  $\kappa = 0.7$  en mathématiques et  $\kappa = 0.1$  en français. Malgré une volonté d'uniformisation des définitions, il semble complexe pour certains domaines (tel que le français) de définir la nature de toutes les CC. Nous considérons donc les mathématiques comme une matière où la catégorisation est plutôt correcte et le français comme une matière où la catégorisation est très bruitée, ce qui doit être pris en compte dans l'analyse des résultats. Dans la suite, les CC déclaratives seront désignés par DEC, les CC procédurales par PRO.

### 3.3 Mesure de la performance

Afin de mesurer la capacité de nos modèles à prédire la réponse d'un apprenant  $i$  à un exercice  $j$ , nous utilisons les métriques suivantes :

- AUC (pour *area under curve*), une métrique allant de 0 (lorsque les prédictions sont toutes erronées) à 1 (traduisant une classification parfaite). Ce score est calculé par l'intermédiaire de la courbe ROC.
- RMSE (pour *root-mean-square error*), mesurant la précision de la prédiction. Plus la corrélation des prédictions avec les données est forte, plus RMSE tend vers 0.

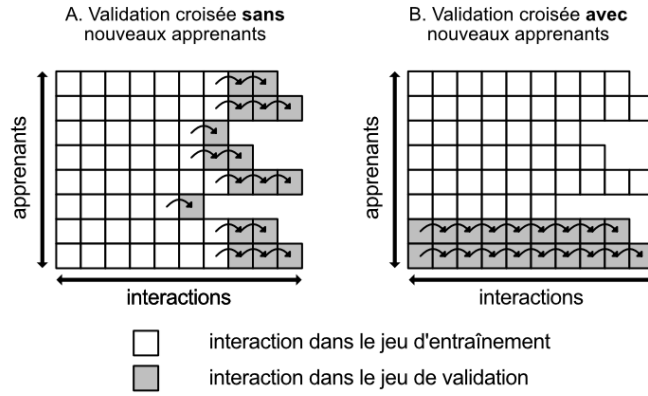
La performance des algorithmes étudiés est principalement mesurée à l'aide des valeurs d'AUC. La métrique RMSE n'est utilisée ici que pour départager deux algorithmes dont la métrique AUC serait égale : à performance de classification équivalente, on compare l'écart à la moyenne pour déterminer quel algorithme pourrait donner des résultats trop éloignés de la réalité.

On procède à une validation croisée à  $k = 5$  blocs. Pour les modèles de *knowledge tracing*, la validation croisée est basée sur les CC : le jeu de données de la phase de test porte sur des CC différentes du jeu de données de la phase d'apprentissage. La validation croisée pour les modèles logistiques est, elle, basée sur les apprenants : le jeu de données de la phase de test porte sur des apprenants différents du jeu de données de la phase d'apprentissage. La différence entre les deux types de validation croisée est représentée dans la Figure 2. Les deux types de modèles de prédiction ne seront donc pas comparés, car les métriques utilisées ne concordent pas, malgré le fait qu'elles portent sur le même jeu de données. La validation croisée basée sur les apprenants mène à des valeurs d'AUC plus élevées que celle basée sur les CC [7].

## 4 Résultats

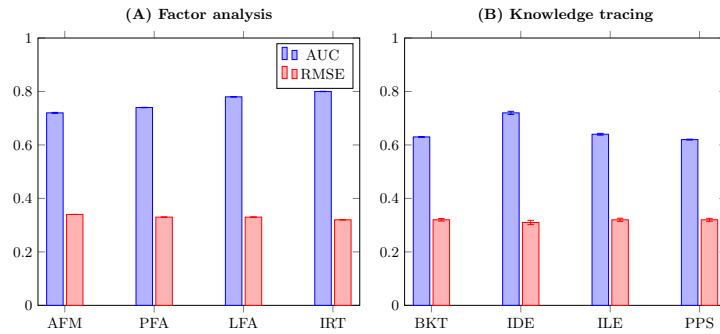
### 4.1 Comparaison des algorithmes de diagnostic cognitif

Les résultats détaillés des algorithmes que l'on a utilisé pour ce papier sont présentés dans la Fig. 3. On remarque que la variante KT-IDE (pour le modèle BKT) et IRT (pour le modèle *factor analysis*) ont de meilleures performances sur



**Fig. 2.** Comparaison de la méthode de validation entre (A) le modèle de *knowledge tracing* et (B) le modèle de *factor analysis*, illustration inspirée de [11].

notre jeu de données que les autres algorithmes et variantes étudiés. Les résultats présentés dans les parties 4.2 et 4.3 sont restreints à ces deux algorithmes. On note tout de même que des résultats de même tendance ont été observés pour les autres algorithmes de leur modèle respectif.

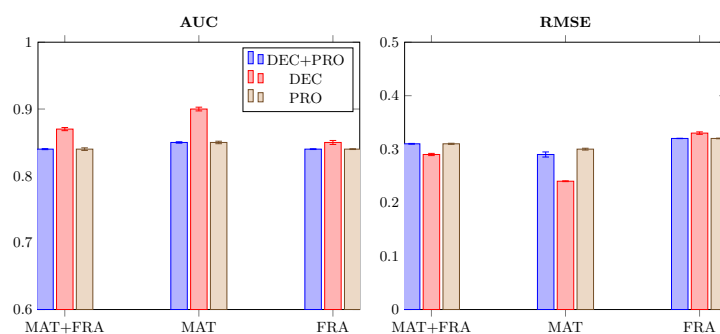


**Fig. 3.** Performance des différents algorithmes testés pour les modèles (A) *factor analysis* et (B) *knowledge tracing* : AUC et RMSE sont calculés pour chaque algorithme selon la méthodologie décrite dans la partie 3.3.



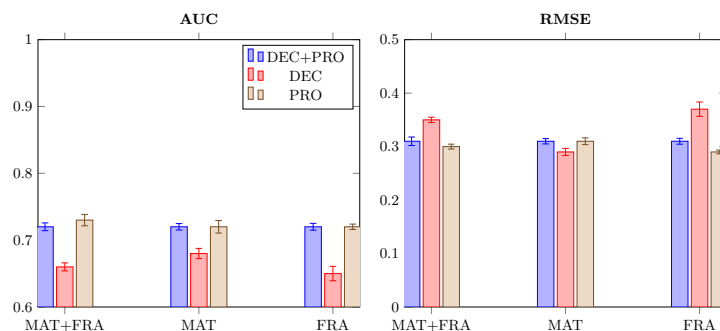
## 4.2 Filtrage par nature des composantes de connaissance et matière associée

Chaque interaction entre un apprenant et un exercice peut être filtrée en fonction de la nature des CC associées et de la matière concernée. La Figure 4 présente les métriques de performance de l'algorithme IRT sur les différents jeux de données filtrés. Les résultats sont groupés par matière concernée afin de pouvoir étudier la différenciation des données par la nature des CC sur les deux matières étudiées. Nous remarquons une meilleure performance d'IRT sur le jeu de données ne portant que sur les CC déclaratives (AUC +0,05 / RMSE -0,06 par rapport au jeu de données restreint aux CC procédurales). En comparant ces mêmes résultats en fonction de la matière associée aux CC par l'algorithme IRT, nous remarquons une meilleure performance globale de l'algorithme sur le jeu de données de mathématiques : l'écart de performance entre les mathématiques et le français est assez faible sur le jeu de données restreint aux CC procédurales (AUC +0,01 / RMSE -0,02), mais devient bien plus important sur le jeu de données restreint aux CC déclaratives (AUC +0,05 / RMSE -0,09).



**Fig. 4.** Performances de l'algorithme IRT sur les données filtrées par nature de CC et par matière associée. Les matières sont en abscisse pour comparer les performances en fonction du filtrage par nature de CC. À gauche, AUC. À droite, RMSE.

De la même manière, la Figure 5 présente les performances de l'algorithme KT-IDE en fonction de la nature des CC associées pour étudier l'impact de la différenciation des données par la nature des CC sur les deux matières étudiées. Cette fois, nous observons que ce sont sur les CC procédurales que l'algorithme de *knowledge tracing* réalise de meilleures prédictions (AUC +0,07 / RMSE -0,05 sur le jeu de données sur les deux matières). La matière concernée n'influe que très peu sur les résultats, nous ne remarquons qu'une légère amélioration sur les mathématiques (AUC = / RMSE -0,02 pour les mathématiques par rapport au français sur le jeu de données restreint aux CC procédurales).



**Fig. 5.** Performances de l'algorithme IDE sur les données filtrées par nature de CC et par matière associée. Les matières sont en abscisse pour comparer les performances en fonction du filtrage par nature de CC. À gauche, AUC. À droite, RMSE.

## 5 Discussion

### 5.1 Comparaison des algorithmes de diagnostic cognitif

Comme nous l'avons indiqué dans la partie 3.3, la façon dont sont calculées les métriques de performance des différents algorithmes rend impossible la comparaison inter-modèles. Néanmoins, il est possible de comparer les algorithmes appartenant à la même catégorie de modèle de l'apprenant.

Une nette domination d'IRT parmi les algorithmes de *factor analysis* sur toutes les métriques présentées est observée dans la partie 4.1. Du côté des algorithmes de *knowledge tracing*, les meilleurs résultats sont ceux de la variante KT-IDE. La nature des exercices sur lesquels sont récoltées les traces des apprenants peut être une des raisons de tels résultats. D'une part, les paramètres de *slip* et *guess* individualisés pour chaque exercice permettent une étude plus précise de tous les exercices sur lesquels sont récoltées les traces des apprenants. D'autre part, la théorie de réponse à l'item est plus performante sur des questions à choix multiples.

### 5.2 Influence de la nature des composantes de connaissance

Nos résultats sur les performances des deux modèles de l'apprenant deviennent bien plus éloquentes lorsque l'on s'intéresse à la comparaison des performances des algorithmes sur les jeux de données filtrés par la nature des CC. D'une part, pour le modèle *factor analysis*, le jeu de données limité aux CC déclaratives donne lieu à de meilleures performances que sur les jeux de données limités aux CC procédurales. Les algorithmes de *factor analysis* sont plus adaptés à la prédiction de réussite d'un apprenant sur des exercices portant sur des CC déclaratives. Les modèles logistiques semblent à même de modéliser la mémoire

(à savoir la mémoire déclarative de Cohen [5]). D'autre part, le modèle *knowledge tracing* est plus performant sur le jeu de données limité aux CC procédurales. Là encore, la structure de ce modèle semble plus adaptée pour étudier le phénomène de compréhension, propre aux tâches techniques. Il existe un gain de performance à limiter les algorithmes de *knowledge tracing* aux seules CC procédurales.

Nous comprenons ici l'importance de l'interprétabilité des paramètres utilisés dans le modèle de l'apprenant. L'augmentation des performances par la catégorisation des CC pour l'un et l'autre des modèles utilisés révèle un choix pertinent des observations par rapport à l'objet à modéliser. Les paramètres, résultant du choix des observations, sont ainsi plus significatifs dès lors que leur détermination dépend de la nature des CC sur lesquels ils sont calculés. Le modèle *knowledge tracing* repose sur des méta-paramètres particulièrement adaptés aux CC procédurales, tandis que le modèle *factor analysis*, lui, considère des paramètres associés aux CC déclaratives. En restreignant l'usage de BKT aux CC procédurales, le paramètre lié à l'apprentissage (noté  $P_l$  dans la partie 2), au-delà de permettre une bonne prédiction des résultats de l'apprenant, récupère tout le sens qu'il est censé avoir : il mesure la capacité de l'apprenant de passer d'un état de non-maîtrise d'une CC à un état de maîtrise, et modélise le moment du déclic, propre aux tâches techniques associées aux CC procédurales. Nous comprenons qu'il n'y aurait que très peu de sens à un tel paramètre pour une CC déclarative. De la même manière, un paramètre tel que le niveau de l'apprenant peut avoir un champ d'interprétation extrêmement large. Cependant, une fois limité aux seules CC déclaratives, il ne s'agit plus d'un niveau global censé mesurer la capacité d'un élève à comprendre, appliquer, etc, mais bien de la seule capacité d'apprendre et de retenir des faits. Comme le périmètre d'observation est plus restreint (car seule une nature de CC est considérée lors de l'apprentissage), les paramètres sont plus spécialisés et sont destinés à mesurer une métrique propre à la fonction cognitive sollicitée. Ils deviennent donc plus facilement interprétables.

Dans l'optique d'un système éducatif adaptatif où l'on chercherait l'interprétabilité du modèle de l'apprenant, le fait d'avoir des paramètres compréhensibles (voir partie 2) et adaptés en fonction des ressources étudiées devient très intéressant. De telles métriques permettent d'accéder à une analyse sur les différentes fonctions cognitives de l'apprenant. De plus, la catégorisation des CC permet en effet une amélioration des prédictions du modèle de l'apprenant.

### 5.3 Influence de la matière

Nous avons pu également observer des différences entre les deux matières observées. L'écart de performance entre *factor analysis* et *knowledge tracing* en fonction de la nature des CC est également visible lorsque l'on restreint les jeux de données aux mathématiques ou au français. Néanmoins, le contraste est plus marqué pour les mathématiques que pour le français. Nous rappelons que les deux matières ont été sujettes à une catégorisation de qualité différentes ( $\kappa = 0.7$  pour les mathématiques,  $\kappa = 0.1$  pour le français). Nous expliquons cette

différence entre les deux matières par les divergences remarquées dans la partie 3.3 sur la catégorisation des CC. D'une part, il apparaît que la catégorisation des CC procédurales est plus fiable que celle des CC déclaratives dans les deux matières, ce qui explique les similitudes des résultats entre mathématiques et français sur le jeu de données restreint aux CC procédurales. D'autre part, il existe d'après notre étude auprès des experts plus de cas de CC considérées comme procédurales mais déclarées déclaratives en français qu'en mathématiques. La performance obtenue sur le jeu de données restreint aux CC déclaratives est donc plus contrastée, avec de meilleurs résultats sur les mathématiques que sur le français.

Il est nécessaire ici de prendre en compte le niveau correspondant aux exercices étudiés. En classes de CE1 et CE2, la différence entre exercices portant sur des composantes de connaissance déclarative et procédurale semble bien plus marqué en mathématiques qu'en français. En mathématiques, les CC déclaratives portent sur de la restitution pure et les CC procédurales sur des tâches mécaniques demandant une rigueur méthodologique. En français, par contre, la nature des composantes de connaissance est plus floue : comment déterminer le type de mémoire sollicité par un exercice demandant le pluriel d'un nom ? Les kappa témoignent de la difficulté pour des experts de discriminer entre des CC déclaratives et procédurales. Les différences de performance entre les matières, elles, mettent en lumière l'importance de la classification par leur nature des CC pour un gain réel en interprétabilité.

## 6 Conclusion et perspectives

Nous avons mis en évidence dans ce papier une différence de performance des algorithmes modélisant l'apprenant en fonction de la nature des composantes de connaissance sur lesquelles sont récoltées les données issues des traces d'élèves. Cette différence a été étudiée dans deux matières différentes (les mathématiques et le français), ce qui permet d'améliorer la vraisemblance de notre hypothèse.

Nous comprenons alors l'importance de l'interprétabilité des paramètres utilisés dans ces algorithmes qui, au delà d'obtenir une précision suffisante, peuvent et doivent permettre d'accéder à des informations sur l'apprenant. Il devient possible de déterminer l'aisance d'un apprenant vis-à-vis de sa mémoire déclarative ou procédurale, tout en ne modélisant que son savoir, ce qui ouvre le champ à de nouveaux diagnostics cognitifs.

Le corollaire de ces résultats est l'intérêt pour un EIAH adaptatif de combiner des modèles permettant de modéliser les deux types de mémoire, déclarative et procédurale, d'un apprenant. Sous la réserve que les modèles de *knowledge tracing* puisse avoir des performances équivalentes aux modèles d'additive factors, la combinaison des deux laisse entrevoir un gain significatif pour le diagnostic. La détermination de la performance d'un tel modèle de diagnostic hybride devrait faire l'objet de futures recherches.

## References

1. Vincent Aleven and Kenneth R Koedinger. Knowledge component (kc) approaches to learner modeling. *Design Recommendations for Intelligent Tutoring Systems*, 1:165–182, 2013.
2. Vincent Aleven, Elizabeth A McLaughlin, R Amos Glenn, and Kenneth R Koedinger. Instruction based on adaptive learning technologies. *Handbook of research on learning and instruction*, pages 522–560, 2016.
3. John R Anderson. Act: A simple theory of complex cognition. *American psychologist*, 51(4):355, 1996.
4. Susan Bull and Judy Kay. Open learner models. In *Advances in intelligent tutoring systems*, pages 301–322. Springer, 2010.
5. Neal J Cohen and Larry R Squire. Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, 210(4466):207–210, 1980.
6. Albert T Corbett and John R Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.
7. Mohammad Khajah, Robert V. Lindsey, and Michael Mozer. How deep is knowledge tracing? In *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016, Raleigh, North Carolina, USA, June 29 - July 2, 2016*. International Educational Data Mining Society (IEDMS), 2016.
8. Zachary A Pardos and Neil T Heffernan. Using hmms and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset. *Journal of Machine Learning Research W & CP*, 40, 2010.
9. Zachary A Pardos and Neil T Heffernan. Kt-idem: Introducing item difficulty to the knowledge tracing model. In *International conference on user modeling, adaptation, and personalization*, pages 243–254. Springer, 2011.
10. Zachary A Pardos, Qing Yang Wang, and Shubhendu Trivedi. The real world significance of performance prediction. *International Educational Data Mining Society*, 2012.
11. Radek Pelánek. Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. *User Modeling and User-Adapted Interaction*, 27(3-5):313–350, 2017.
12. Radek Pelánek, Jirí Rihák, and Jan Papoušek. Impact of data collection on interpretation and evaluation of student models. In *Proceedings of the sixth international conference on learning analytics & knowledge*, pages 40–47, 2016.
13. Sarah Schultz and Trenton Tabor. Revisiting and extending the item difficulty effect model. In *In Proceedings of the 1st workshop on massive open online courses at the 16th annual conference on artificial intelligence in education*, pages 33–40. Citeseer, 2013.