

# Évolution du diagnostic cognitif de l'apprenant à l'aide d'une approche numérique et symbolique

Olivier Allègre

LIP6, Sorbonne Université  
olivier.allegre@lip6.fr  
1ère année de thèse\*

**Abstract.** L'objet de la recherche présentée ici est un dispositif de diagnostic cognitif hybride basé sur une association de la connaissance experte avec les données issues des traces des apprenants. Il est caractérisé par plusieurs couches de données numériques et symboliques. Ce diagnostic cognitif concentre les probabilités de maîtrise des composantes de connaissance du domaine construit par les experts, à l'aide des exercices réalisés par l'apprenant, et les relations existantes entre ces mêmes composantes de connaissance, dérivées des relations explicitées par les experts. L'objectif est d'adapter le modèle du domaine, correspondant à une vision consensuelle des experts sur les règles de structure régissant le domaine, à l'apprenant à l'aide des états de connaissance des apprenants partageant les mêmes caractéristiques d'apprentissage. L'apport relatif d'un tel système sera étudié à l'aide des résultats d'un ensemble d'apprenants sur la plateforme Kartable, puis comparé à d'autres systèmes d'apprentissage adaptatif.

**Keywords:** Apprentissage adaptatif · EIAH · diagnostic cognitif.

## 1 Introduction

L'utilisation de plateformes pédagogiques numériques apporte de nombreuses données issues des traces des apprenants, exploitées par des méthodes statistiques de fouille de données. Les experts ont, eux, une connaissance approfondie des domaines de connaissance et peuvent en produire des modèles basés sur les relations pouvant exister entre différentes composantes de connaissance. Ainsi, les systèmes de modélisation de l'apprenant se basent sur des méthodes d'extraction numériques ou sur des méthodes de représentation des connaissances symboliques. Si ces deux types sont très différents dans leur approches et leurs algorithmes, ils permettent l'élaboration d'un diagnostic cognitif de l'apprenant. Nous définissons le diagnostic cognitif par la prédiction de l'état des connaissances de l'apprenant (ce qu'il sait et ne sait pas) mais aussi de sa façon de raisonner. Le but de cette recherche est de prédire les connaissances de l'apprenant et les relations logiques existantes entre elles, tout en restant fidèle

---

\* Thèse financée par Kartable, [www.kartable.fr](http://www.kartable.fr), site web de soutien scolaire.

au modèle défini par les experts, à partir de la réalisation de plusieurs activités pédagogiques. Nous veillerons à concilier les règles de structure dictées par les experts avec les traces des apprenants. L'enjeu est d'adapter la vision experte du domaine à l'apprenant : la structure du domaine dictée par les experts correspond au mieux à un "apprenant moyen" mais en aucun cas à l'intégralité des apprenants. On souhaite déterminer les règles régissant le domaine associé à l'apprenant, en adaptant éventuellement les règles initiales données par les experts.

Dans un premier temps sera étudiée une structure de données permettant de stocker à la fois la représentation symbolique du domaine explicitée en amont par un groupe d'experts et la représentation numérique de l'apprenant regroupant ses résultats et ses caractéristiques. Ensuite, une stratégie sera mise en place pour créer un modèle de l'apprenant à partir de l'utilisation coordonnée du modèle du domaine déterminé par les experts et des données de l'apprenant récoltées lors de ses différentes actions, et d'en établir son diagnostic cognitif.

## 2 État de l'art

L'apprentissage adaptatif a débuté dès les années 1970, alimenté par l'intérêt grandissant pour l'intelligence artificielle [1]. Plusieurs outils numériques et symboliques existent permettant un diagnostic cognitif de l'apprenant. On s'attardera sur les techniques qui influencent la recherche présentée ici. On remarquera également que ces systèmes s'attardent soit sur la fabrication d'un modèle de l'apprenant sans prise en compte d'un modèle expert complet, soit, dans le cas où un modèle expert est initialement introduit, ce dernier n'est pas remis en cause. Notre recherche se place dans ce contexte, et cherche à rétablir une communication entre les deux modèles expert et apprenant.

Les outils numériques de diagnostic cognitif les plus représentatifs sont les modèles de calculs de maîtrise comme l'*item response theory* (IRT) et ses dérivés Learning Factor Analysis (LFA) [2] et Performance Factor Analysis (PFA) [14]. Des outils issus de l'*educational data mining* permettent d'extraire de l'information à partir des données éducatives. Les techniques de clustering [9] et de filtrage collaboratif [3] sont utilisées pour déduire des informations sur le niveau de maîtrise des apprenants. Le diagnostic cognitif y reste partiel, portant généralement sur le niveau global de l'apprenant ou sur ses caractéristiques. Ils ne caractérisent pas le raisonnement de l'apprenant, et restent assez limités en terme d'interprétabilité des résultats qu'ils peuvent produire. Ils ont néanmoins l'avantage de ne pas nécessiter d'intervention humaine lourde.

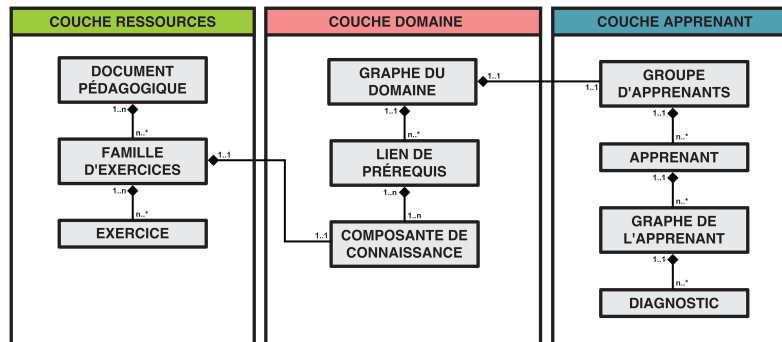
D'autres systèmes construisent des structures hiérarchiques de connaissances. Il peut s'agir de règles de production [13] ou de contraintes [11] basées sur des ensembles de déclarations d'experts. La décomposition du domaine d'apprentissage en composantes de connaissance dans la *théorie des espaces de connaissance* (KST) [8] et ses alternatives telles que *Competence-based Knowledge Space Theory* (CbKST) [10] utilisent l'explicitation de relations logiques entre compétences

par un ensemble d'experts. Ces modèles de diagnostic cognitif misent sur la *compréhension* de l'apprenant et sont donc plus facilement interprétables. Cependant, elles peuvent être coûteuses car elles nécessitent la disponibilité des experts qui doivent construire un modèle consensuel et partagé du domaine.

Des techniques mixtes existent également, à l'instar des réseaux bayésiens [4] qui associent concepts, compétences, misconceptions et erreurs des apprenants à l'aide de relations probabilistes. Le Bayesian Knowledge Tracing (BKT) [5] et ses variantes [6][16] utilisent des variables symboliques latentes et des données numériques afin de calculer des probabilités de maîtrise. Le *Partial Order Knowledge Structures* (POKS) [7] est basé sur la théorie KST, mais utilise tout de même les données d'utilisation pour inférer l'existence des relations entre les compétences. L'architecture de ces systèmes implique une explicabilité plus importante du diagnostic. Néanmoins, ces techniques considèrent leurs résultats sans nécessairement les comparer aux raisonnements experts pouvant les remettre en cause.

### 3 Architecture multi-couches et organisation des données

La stratégie mise en place dans le cadre de cette recherche débute par une organisation des données permettant l'interaction entre les données symboliques et les données numériques, dès leur acquisition, à l'aide de trois couches.



**Fig. 1.** Vue d'ensemble des trois couches. La couche ressources contient les documents pédagogiques et les exercices, regroupés en familles. La couche domaine contient les composantes de connaissances, associées à un graphe du domaine par des liens de prérequis. La couche apprenant contient les diagnostics des apprenants, regroupés en groupe de même caractéristiques d'apprentissage.

La **couche ressources** contient toutes les informations sur les ressources pédagogiques utilisables par l'apprenant. Elle contient également les liens hiérarchiques, donnés par le support guidant l'apprentissage (comme le découpage en chapitres et en thèmes) et les hyperliens existants entre les ressources. Pour chaque ressource

sont stockées ses différentes caractéristiques propres, et indépendantes de l'usage qui peut en être fait. On peut y ajouter des liens factuels, comme une étude sur la similitude des différents contenus. Les exercices de la couche ressources sont regroupés en familles d'exercices, associées à une composante de connaissance, définie ci-après, sur laquelle elle porte.

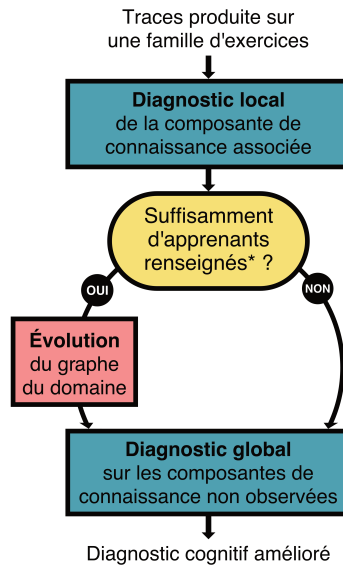
La **couche domaine** contient toute la connaissance experte sur la structure du domaine. Elle se compose principalement de composantes de connaissances, qui peuvent être déclaratives (portant alors sur de la connaissance pure) ou procédurales (auquel cas il s'agit de tâches techniques simples, telles les compétences définies dans la théorie CbKST). Nous faisons l'hypothèse de l'existence d'une bijection entre l'ensemble des familles d'exercices et l'ensemble des composantes de connaissance. La granularité de ces dernières définit donc le périmètre des exercices, afin qu'ils permettent le diagnostic de chacune des composantes de connaissance. La couche domaine contient également des **graphes du domaine**, inspirés des graphes de compétences de la théorie KST. Ils sont composés d'un ensemble de nœuds (les composantes de connaissances décrites précédemment) et de liens de prérequis augmentés numériquement à l'aide de tables de probabilités conditionnelles. L'existence et la pondération de ces relations sont définies par des experts.

Enfin, la **couche apprenant**, unique pour un apprenant donné, contient les traces de l'apprenant lors de ses interactions avec les différentes ressources pédagogiques qu'il a pu rencontrer au cours de ses sessions d'apprentissage, stockées dans un graphe directement dérivé du graphe du domaine. Expliqué plus précisément dans la partie 4, le **graphe de l'apprenant** permet le stockage de la maîtrise des composantes de connaissance du graphe du domaine correspondant à l'apprenant. Son association avec les relations de prérequis stockées dans le graphe du domaine donne lieu à un réseau bayésien, sur lequel sera mené un apprentissage des traces de l'apprenant, contraint par les règles et valeurs régissant le graphe du domaine associé à l'apprenant.

## 4 Évolution du modèle de diagnostic cognitif

On définit dans cette partie une méthode permettant une évolution du modèle de diagnostic cognitif. Ce dernier est décrit dans cette recherche par l'association des valeurs de maîtrise des composantes de connaissance stockées dans le graphe de l'apprenant avec les relations entre les composantes de connaissance stockées dans le graphe du domaine associé à l'apprenant. On procède en trois étapes.

À chaque fois qu'un apprenant produit des traces sur une famille d'exercices, on procède à un **diagnostic local** de la maîtrise de la composante de connaissance sur laquelle l'évaluation porte. Ce diagnostic dépend de la nature de la composante de connaissance. L'algorithme *Performance Factor Analysis ELO-Extended* [15], basé sur la théorie de réponse à l'item, est spécialement efficace pour tester les connaissances d'un apprenant [12]. Il est utilisé pour évaluer les



**Fig. 2.** Fonctionnement des trois phases de l'évolution du diagnostic cognitif. Les phases de diagnostics local et global ne sont pas soumises à condition. L'évolution a lieu dès que l'indicateur (\*) est enclenché, c'est-à-dire quand un nombre suffisant de diagnostics locaux est effectué par un nombre suffisant d'apprenants.

composantes de connaissances déclaratives. Les exercices concernant des composantes de connaissance procédurales sont évaluées par Bayesian Knowledge Tracing. Il permet, à l'aide notamment de toutes ses possibilités de complexification, une analyse plus effective de l'apprentissage de tâches techniques [12]. L'évaluation de la précision de cette étape sera la comparaison de la prédiction du diagnostic avec les résultats de l'apprenant sur la plateforme Kartable.

À l'aide des tables de probabilités conditionnelles stockées dans le graphe du domaine associé à l'apprenant, il est possible d'induire de proche en proche les probabilités de maîtrise des parents et enfants de la composante de connaissance diagnostiquée. On appelle **diagnostic global** cette phase de propagation dans le réseau bayésien (formé par l'association des graphes du domaine et de l'apprenant). La probabilité prédite de maîtrise des compétences non-évaluées sera comparée avec les résultats effectifs de l'apprenant afin d'évaluer la fiabilité du diagnostic global.

Une fois qu'un nombre suffisant de graphes d'apprenants, associés à un même graphe du domaine, est renseigné (c'est à dire que les graphes ont assez de diagnostics locaux), on procède à l'**évolution** de la structure de ce même graphe du domaine. Cet indicateur, sur le nombre de diagnostics locaux effectués et le nombre d'apprenants renseignés, reste à être déterminé. Cette étape permet une adaptation de la vision du domaine donnée par les experts à un groupe

d'apprenants partageant des caractéristiques d'apprentissage similaires, et pouvant diverger du stéréotype imaginé par les experts. Cette évolution est effectuée par apprentissage sur les réseaux bayésiens représentant les associations du graphe du domaine avec les graphes des apprenants étudiés. L'apprentissage est contraint par les valeurs initiales du graphe du domaine, censé refléter les règles de structure du domaine données par les experts. L'interprétabilité du graphe de l'apprenant est l'enjeu principal de cette phase. Il doit rester conforme au graphe du domaine initialement introduit par les experts : si un lien de prérequis s'avère inexistant, il faudra être capable de juger si son absence est un non-sens du point de vue des experts ou si cela répond à une logique en lien avec les caractéristiques d'apprentissage des apprenants. L'amélioration des prédictions de maîtrise des composantes de connaissance par le diagnostic global sera utiliser pour évaluer la fiabilité de l'évolution.

## 5 Synthèse et perspectives

L'association du graphe de l'apprenant (valeurs de maîtrise des composantes de connaissance) et du graphe du domaine associé (contenant les tables de probabilités conditionnelles entre les composantes de connaissance) correspond au diagnostic cognitif défini dans l'introduction. L'état de connaissance de l'apprenant (au sens de Doignon et Falmagne [8]) a été évalué par diagnostics locaux effectués sur les traces de l'apprenant, puis par diagnostics globaux sur les composantes de connaissance non-observées. Les raisonnements de l'élève sont représentés dans notre modèle par les tables de probabilités conditionnelles du graphe du domaine associé à l'apprenant. Ce dernier est déterminé par apprentissage sur les états de connaissances des apprenants associé au même graphe du domaine, et contraint sur les règles de structure expertes initialement introduites. Cette association permet d'étudier les spécificités d'apprentissage de l'apprenant, avec une résistance aux comportements sporadiques non prédits par les experts.

Le système sera observé et testé sur la plateforme Kartable, permettant l'étude de la progression d'un nombre d'apprenants suffisant. Le domaine étudié est celui des mathématiques de niveau lycée. Ce système pourra également être comparé à d'autres modèles uniquement basés soit sur de la connaissance experte pure, soit sur des données d'utilisation d'apprenants seulement.

## References

1. Jaime R Carbonell. Ai in cai: An artificial-intelligence approach to computer-assisted instruction. *IEEE transactions on man-machine systems*, 11(4):190–202, 1970.
2. Hao Cen, Kenneth Koedinger, and Brian Junker. Learning factors analysis—a general method for cognitive model evaluation and improvement. In *International Conference on Intelligent Tutoring Systems*, pages 164–175. Springer, 2006.
3. Mohamed Amine Chatti, Anna Lea Dyckhoff, Ulrik Schroeder, and Hendrik Thüs. A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6):318–331, 2013.

4. Konstantina Chrysafiadi and Maria Virvou. Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications*, 40(11):4715–4729, 2013.
5. 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.
6. Ryan SJ d Baker, Albert T Corbett, and Vincent Alevén. More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *International conference on intelligent tutoring systems*, pages 406–415. Springer, 2008.
7. Michel C Desmarais, Peyman Meshkinfam, and Michel Gagnon. Learned student models with item to item knowledge structures. *User Modeling and User-Adapted Interaction*, 16(5):403–434, 2006.
8. Jean-Paul Doignon. Knowledge spaces and skill assignments. In *Contributions to mathematical psychology, psychometrics, and methodology*, pages 111–121. Springer, 1994.
9. Philippe Fournier-Viger, Jerry Chun-Wei Lin, Rage Uday Kiran, Yun Sing Koh, and Rincy Thomas. A survey of sequential pattern mining. *Data Science and Pattern Recognition*, 1(1):54–77, 2017.
10. Jürgen Heller, Christina Steiner, Cord Hockemeyer, and Dietrich Albert. Competence-based knowledge structures for personalised learning. *International Journal on E-learning*, 5(1):75–88, 2006.
11. Antonija Mitrovic. Fifteen years of constraint-based tutors: what we have achieved and where we are going. *User modeling and user-adapted interaction*, 22(1-2):39–72, 2012.
12. Jan Papousek, Radek Pelánek, and Vít Stanislav. Adaptive practice of facts in domains with varied prior knowledge. In *Educational Data Mining 2014*, 2014.
13. Gigliola Paviotti, Pier Giuseppe Rossi, and Dénes Zarka. Intelligent tutoring systems: an overview. *Pensa Multimedia*, 2012.
14. Philip I Pavlik Jr, Hao Cen, and Kenneth R Koedinger. Performance factors analysis—a new alternative to knowledge tracing. *Online Submission*, 2009.
15. 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.
16. Michael V Yudelson, Kenneth R Koedinger, and Geoffrey J Gordon. Individualized bayesian knowledge tracing models. In *International conference on artificial intelligence in education*, pages 171–180. Springer, 2013.