

Wheel-Spinning : les élèves qui ne parviennent pas à maîtriser une compétence.

Joseph E. Beck et Yue Gong

Institut polytechnique de Worcester
{josephbeck, ygong}@wpi.edu

Résumé. Le concept d'apprentissage par la maîtrise est puissant : plutôt qu'un nombre fixe de pratiques, les élèves continuent à pratiquer une compétence jusqu'à ce qu'ils la maîtrisent. Cependant, une hypothèse implicite dans cette formulation est que les étudiants sont capables de maîtriser la compétence. Cette hypothèse est cruciale pour les tuteurs informatiques, car leur répertoire d'actions pédagogiques peut ne pas être aussi efficace qu'on le croit généralement. Que se passe-t-il si l'élève n'a pas les connaissances suffisantes pour résoudre les problèmes impliquant la compétence, et si le tuteur informatique n'est pas en mesure de fournir une instruction suffisante ? Cet article présente le concept de "Wheel-Spinning", c'est-à-dire les étudiants qui ne parviennent pas à maîtriser une compétence en temps voulu. Nous montrons que si un élève ne maîtrise pas rapidement une compétence dans ASSISTments ou dans le Tuteur cognitif, il est probable qu'il aura des difficultés et ne maîtrisera probablement jamais cette compétence. Nous discutons des liens entre un tel manque d'apprentissage et les comportements négatifs des étudiants, comme le jeu et le désengagement, et nous discutons des modifications à apporter à la conception des STI pour surmonter ce problème.

Mots clés : apprentissage par la maîtrise, modélisation de l'élève, Wheel-Spinning

1 Introduction

Les systèmes de tutorat intelligents (STI) sont des environnements d'apprentissage généralement efficaces pour la résolution de problèmes assistée par ordinateur. Les STI sont capables de fournir de l'aide aux élèves qui ont des difficultés à résoudre des problèmes, et il a été démontré qu'ils sont plus efficaces que les traditionnels devoirs papier-crayon pour aider les élèves à apprendre. Comparativement aux méthodes d'enseignement plus traditionnelles, les STI donnent généralement de bien meilleurs résultats selon les mesures définies par l'expérience [p. ex., 1], et un peu mieux selon les instruments normalisés. Bien qu'il soit tentant de supposer que tous les élèves bénéficient de l'utilisation d'un STI, cette hypothèse ne tient pas nécessairement, car un STI n'est pas un choix judicieux pour tous les apprenants.

Dans le cadre de l'apprentissage par la maîtrise [2], tel qu'il est mis en œuvre dans de nombreux STI, l'élève n'est pas confronté à un nombre fixe de problèmes, mais continue à résoudre des problèmes jusqu'à ce qu'il parvienne à maîtriser les compétences associées. En d'autres termes, une fois que l'élève a fini de résoudre un problème, éventuellement avec l'aide de l'ordinateur, s'il n'a pas encore maîtrisé la compétence associée, l'ordinateur lui présente un autre problème. Il existe une longue histoire de travaux sur l'apprentissage par la maîtrise dans le cadre de l'enseignement assisté par ordinateur [3], et ce modèle a un sens intuitif et répond certainement à la maxime "la pratique rend parfait", d'autant plus que la plupart des tuteurs fournissent une assistance à l'étudiant sous forme de conseils ou de décomposition du problème en étapes. Cependant, un peu de réflexion révèle quelques

les faiblesses du modèle. Si un élève a besoin d'aide pour résoudre les deux premiers problèmes, lui en présenter un troisième en espérant qu'il apprendra la compétence pourrait très bien être une stratégie judicieuse. Si l'élève a été incapable de résoudre vingt problèmes et a eu besoin d'une aide considérable pour chacun d'entre eux, il est probablement plutôt optimiste de croire que le vingt-et-unième problème lui permettra d'acquérir soudainement la compétence (dans l'ensemble de données que nous analysons, il n'y a que 1,4 % de chances qu'un tel élève maîtrise un jour la compétence, du moins dans les données recueillies pour cette étude).

L'hypothèse selon laquelle les élèves finiront par acquérir des compétences avec suffisamment de pratique ne fait pas seulement partie de la prise de décision du tutoriel, puisque le traçage des connaissances [4] suppose une probabilité constante d'apprendre la compétence à chaque tentative de résolution de problème. Cependant, tous les élèves ne sont pas capables d'acquérir des compétences dans un STI, et certains passent un temps considérable bloqué dans la boucle d'apprentissage de la maîtrise sans qu'aucun apprentissage ne se produise. En plus de faire perdre du temps à l'apprenant, une telle expérience est vraisemblablement frustrante, car on présente sans cesse aux apprenants des problèmes qu'ils sont clairement incapables de résoudre. Nous appelons ce phénomène le "Wheel-Spinning", en référence à une voiture enlisée dans la boue ou la neige ; ses roues tournent rapidement, mais elle ne va nulle part. De même, les élèves sont confrontés à de nombreux problèmes, mais ne progressent pas vers la maîtrise. Plus tard, nous discuterons des liens possibles avec d'autres comportements négatifs tels que le jeu.

2 Décrire le filage des roues

Nous définissons la roue folle comme un élève qui passe trop de temps à lutter pour apprendre un sujet sans le maîtriser. Certains élèves commencent à travailler avec un ITS en comprenant déjà la matière. D'autres maîtriseront la compétence relativement rapidement, peut-être avec l'aide de l'ITS. Aucun de ces groupes n'est problématique. Nous sommes préoccupés par les élèves qui passent trop de temps sans maîtriser la compétence. Cette définition comporte deux concepts qui doivent être opérationnalisés :

1. Qu'est-ce que cela signifie de *maîtriser* un sujet ?
2. Combien de temps est *trop long* ?

La réponse à ces deux questions varie quelque peu d'un système à l'autre, car la notion de maîtrise est un concept vague qui peut être concrétisé de diverses manières. Une des approches, proposée par Corbett et Anderson (1995), consiste à estimer les connaissances de l'élève, et lorsque la probabilité qu'un élève connaisse une compétence dépasse 0,95, on considère que l'élève a maîtrisé cette compétence. Une approche utilisée dans le système ASSIST-ments consiste à considérer qu'un étudiant a maîtrisé la compétence lorsqu'il obtient trois questions correctes d'affilée. En ce qui concerne le temps, la quantité idéale varie également selon le système. Un STI dont les problèmes demandent 10 minutes à résoudre devrait probablement exiger moins de problèmes pour la maîtrise qu'un système qui demande 20 secondes par problème.

Pour notre critère de maîtrise, nous avons décidé d'utiliser l'approche plus simple de trois réponses correctes consécutives. Le processus d'ajustement du modèle de traçage des connaissances est plutôt lent sur les grands ensembles de données, et sa capacité à désambiguïser les connaissances des élèves est mise en doute en raison de problèmes d'identifiabilité [5]. De plus, si d'autres personnes souhaitent reproduire notre modèle de traçage de la connaissance, elles ne pourront pas le faire.

Si les élèves travaillent sur d'autres ensembles de données, il sera plus facile de travailler avec un critère de maîtrise qui ne nécessite pas de souscrire à un cadre de modélisation particulier. Nous supposons également qu'une fois que les élèves maîtrisent une compétence, ils sont incapables de la démaîtriser. Pour être clair, nous pensons que l'oubli existe et qu'un système adaptatif du monde réel doit en tenir compte. Cependant, notre objectif ici est de comprendre comment les apprenants se comportent lors de la maîtrise initiale, et s'ils sont capables d'y parvenir dans un délai raisonnable. L'oubli de ce qui a été appris, bien que ce soit un sujet important, n'est pas central à cette question de recherche.

Pour savoir combien de temps est une quantité raisonnable à maîtriser, nous avons sélectionné 10 possibilités de pratique. Bien que ce seuil soit quelque peu arbitraire, nous verrons (voir Fig. 1) que les résultats ne sont pas très sensibles au seuil exact choisi. De plus, l'un des systèmes que nous analysons, ASSISTments, possède une fonction qui "bloque" les apprenants après qu'ils aient fait 10 tentatives pour une compétence dans une même journée, et leur demande de réessayer un jour ultérieur. Nous ne sommes pas sûrs de l'impact que cette fonction pourrait avoir sur les données, ni de ce que les élèves pourraient faire après avoir été bloqués (par exemple, demander de l'aide à quelqu'un). Par conséquent, nous avons utilisé 10 possibilités de pratique comme seuil de maîtrise dans un délai raisonnable. Voir la figure 1 pour une représentation visuelle du comportement de Wheel-Spinning dans le Cognitive Algebra Tutor (CAT) et dans ASSISTments.

Dans la figure 1, l'axe des x représente le nombre d'occasions de s'exercer qu'un élève a eues pour une compétence particulière, et l'axe des y représente la probabilité cumulative qu'un élève ait maîtrisé la compétence, c'est-à-dire qu'il ait réussi trois problèmes d'affilée. Par définition, aucun élève n'a maîtrisé une compétence lors des deux premières tentatives d'entraînement. Au troisième essai, environ 35 % des élèves des programmes CAT et ASSIST ont maîtrisé la compétence. Pour atteindre cette maîtrise aussi rapidement, ces élèves n'ont pas fait d'erreur dans les trois premiers problèmes ; ils n'ont donc pas bénéficié de l'encadrement disponible pour cette compétence. En d'autres termes, ces élèves ont répondu aux questions sans avoir besoin d'aide, et ont essentiellement utilisé le tuteur comme un simple devoir papier-crayon ; par conséquent, le STI ne devrait pas recevoir de crédit pour avoir aidé ces élèves.

Après trois opportunités de pratique, le tuteur cognitif et les ASSISTments montrent tous deux une augmentation progressive du pourcentage d'élèves ayant maîtrisé la compétence. Après 6 occasions de pratiquer, 59% des élèves du tuteur cognitif et 55% des élèves des ASSISTments ont maîtrisé une compétence. Enfin, après 10 occasions de s'exercer, 69% des élèves du CAT et 62% des élèves des ASSISTments ont maîtrisé la compétence. Bien que nous ayons choisi 10 opportunités de pratique de manière quelque peu arbitraire, et sur la base d'un artefact possible dans l'ensemble de données ASSISTments, ce seuil se situe au-delà du "coude" des courbes de maîtrise pour les deux systèmes, et l'inspection de la figure 1 montre que la proportion d'étudiants ayant maîtrisé la compétence ne changerait pas sensiblement si le seuil était augmenté au-delà de 10. Par conséquent, nous sommes satisfaits de notre seuil pour le Wheel-Spinning, du moins pour une première analyse de ce problème.

Il est intéressant de comparer les courbes pour le CAT et les ASSISTments. Bien que, au départ, les deux tuteurs aient eu un nombre à peu près égal d'élèves qui connaissaient déjà la compétence, on peut en déduire que la TAO aide mieux les élèves à la maîtriser que les ASSISTments. De telles comparaisons doivent être faites avec prudence, car il existe une variété de facteurs qui pourraient influencer les différences dans les courbes entre les systèmes. Tout d'abord, la difficulté des problèmes peut varier. Un système avec des problèmes relativement

plus faciles montrerait que davantage d'élèves atteignent la maîtrise de la matière qu'un programme comportant des questions plus difficiles. Rendre les problèmes plus faciles n'est probablement pas un bon moyen de réduire le tournioement. Cependant, étant donné que dans les deux systèmes, environ 34 % des élèves maîtrisent immédiatement la compétence, il est peu probable que des problèmes plus faciles soient une explication. Deuxièmement, les modèles d'utilisation peuvent différer. Par exemple, si les élèves résolvent les problèmes d'une compétence particulière en grandes quantités dans le système ASSISTments, mais seulement quelques-uns par jour dans le système CAT, les élèves auront plus d'occasions d'apprendre en dehors du tuteur. Ainsi, nous ne pouvons pas conclure que la TAO est meilleure que les ASSISTments pour fournir une assistance qui empêche la rotation, mais nous pouvons conclure qu'une certaine combinaison de la TAO et de la manière dont elle est déployée semble fonctionner mieux que les ASSISTments. Cependant, le graphique permet de conclure qu'un nombre important d'étudiants ont des problèmes de rotation, et que ce problème n'est pas particulier à un seul STL, puisqu'il se produit dans deux tuteurs informatiques largement utilisés et ayant des approches pédagogiques différentes.

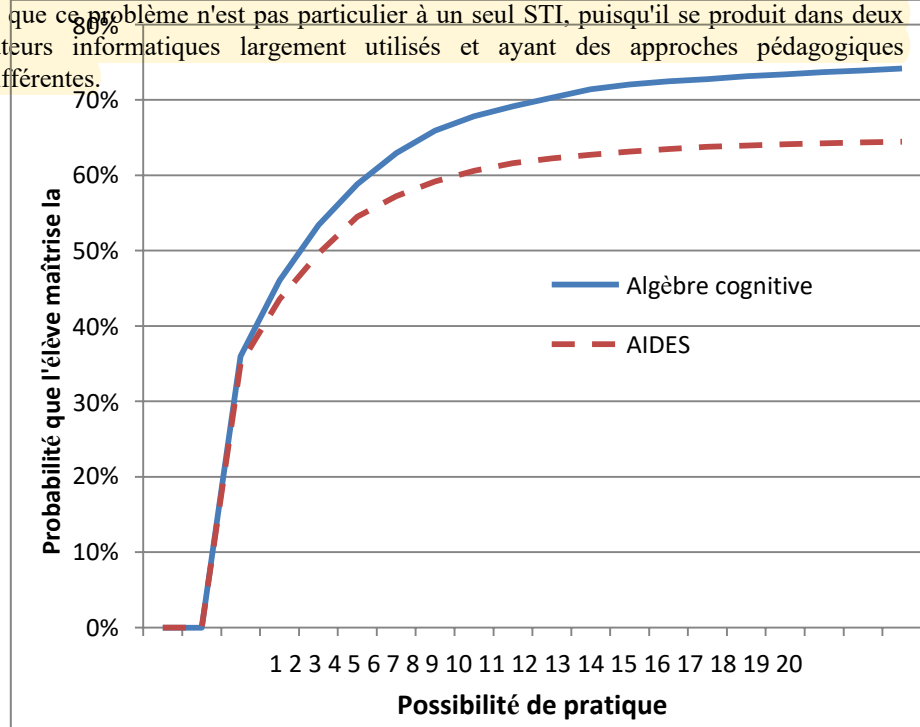


Fig. 1. Graphique de l'engrenage dans ASSISTments et Cognitive Algebra Tutor

3 Modélisation du filage des roues

Étant donné que la rotation des roues est au mieux non productive et éventuellement irritante pour l'élève, nous aimerions détecter ce comportement aussi rapidement que possible. Si nous pouvons prédire

que l'élève risque de s'essouffler, nous pouvons effectuer une autre action de tutorat, de nature plus pédagogique.

Notre approche consiste à considérer chaque paire élève-compétences et à examiner les cas où l'élève maîtrise la compétence ou, après avoir vu 10 problèmes, ne la maîtrise pas (Wheel-Spinning). Les données après la dixième rencontre ou après que l'élève ait maîtrisé la compétence sont ignorées. Cette définition présente une asymétrie, car un élève qui n'a vu que 7 problèmes mais n'a pas réussi à maîtriser la compétence est un cas de Wheel-Spinning indéterminé et n'est pas inclus dans cette analyse, alors qu'un élève qui l'a maîtrisée en 7 tentatives est inclus. Ainsi, cette approche sous-estime le Wheel-Spinning, et estime qu'il se produit dans 9,8% des données. Nous construisons un modèle de régression logistique pour prédire dans quelle catégorie l'élève va maîtriser la compétence ou faire la roue. Afin de déterminer la rapidité avec laquelle nous pouvons catégoriser les élèves, nous construisons un modèle distinct pour chaque nombre d'occasions de pratique que l'élève a eues sur la compétence actuelle. En d'autres termes, nous construisons un modèle pour le moment où l'élève commence à pratiquer la compétence (il n'a vu aucun problème), lorsqu'il a vu un problème, deux problèmes, etc. Nous adoptons cette approche pour deux raisons. Premièrement, nous voulons voir comment la précision évolue au fur et à mesure que nous accumulons des données sur l'élève. Deuxièmement, ce qui est important peut changer avec le temps. Demander un indice (la réponse au problème) sur le premier item peut ne pas poser de problème, mais demander une telle aide sur le quatrième problème peut être un indicateur négatif fort. Nous avons eu 258 990 problèmes résolus par 5997 étudiants. Après avoir supprimé les données indéterminées, notre ensemble de données se compose de 131 909 problèmes résolus par 5026 élèves. Cette analyse a utilisé des données collectées entre septembre 2010 et juillet 2011, avec des étudiants provenant principalement du nord-est des États-Unis. Nous ne disposons que des âges déclarés par les étudiants, et 75 % d'entre eux ont affirmé avoir entre 12 et 15 ans au 1er janvier 2011.

La variable dépendante est le fait qu'un élève fasse ou non une roue sur cette compétence. Les trois premières variables indépendantes suivent les performances de l'élève sur la compétence en question, et les trois suivantes examinent ses performances sur l'ensemble des compétences :

- Nombre antérieur de réponses correctes de l'élève sur cette compétence
- Temps de réponse à cette compétence. Nous avons d'abord transformé les temps de réponse de chaque item en un score Z pour cet item (pour tenir compte du fait que certains problèmes prennent plus de temps que d'autres). Nous avons ensuite pris la moyenne géométrique, $\gamma * \text{prior_average} + (1 - \gamma) * \text{new_observation}$, avec $\gamma = 0,7$. La moyenne géométrique est une méthode de synthèse des données séquentielles, mais elle donne moins de poids aux observations plus anciennes, car les observations antérieures sont diminuées de γ à chaque pas de temps.
- Combien de fois l'élève a atteint un indice de basculement sur cette compétence.
- La fréquence à laquelle l'élève devinait rapidement, calculée pour toutes les compétences, définie comme la soumission de réponses à moins de 2 secondes d'intervalle sur des items successifs. Nous avons pris la moyenne géométrique de la même manière que pour le temps de réponse.
- La fréquence à laquelle l'étudiant a donné une réponse rapide, calculée pour toutes les compétences, définie comme suit
répondant dans un laps de temps qui suggère un taux de lecture de plus de 400 mots par mi-nute. Nous avons pris la moyenne géométrique de cette caractéristique.
- La fréquence à laquelle l'élève a obtenu un indice sur le fond dans trois problèmes consécutifs, toutes compétences confondues ; un 1 indique que l'élève a demandé la réponse dans trois problèmes consécutifs. Nous avons pris la moyenne géométrique de cette caractéristique.

- Le nom de la compétence actuelle.

Nous avons ajusté ce modèle en utilisant la régression logistique dans SPSS. Les six premiers termes étaient des covariables, et le dernier terme était entré comme un effet fixe (c'est-à-dire un paramètre par compétence). Notez que nous n'avons pas pu avoir l'identité de l'utilisateur comme facteur dans ce modèle, car cela dépassait les capacités de SPSS ; par conséquent, la fiabilité statistique serait quelque peu surévaluée en raison de la non-indépendance des essais des étudiants [6], et nous ne faisons donc pas état de la fiabilité statistique. Le tableau 1 résume la précision du modèle. Chaque ligne indique la performance du modèle après avoir vu l'élève résoudre un nombre donné de problèmes sur la compétence. La deuxième colonne indique le pourcentage de paires élève-compétences qui ont abouti à une rotation de la roue. Lorsque les élèves commençaient à maîtriser une compétence, 9,8 % des données comprenaient du wheel spinning. Pour les élèves qui n'avaient pas maîtrisé un problème à la cinquième tentative, 38,5 % des données indiquaient une rotation de la roue. La troisième colonne est le R^2 , une mesure de l'ajustement modéré, qui va de 0 (impossibilité de prédire les données) à 1 (précision parfaite). Notez qu'avant même que l'élève ne commence à résoudre un problème sur la compétence, le modèle est capable de rendre compte de 13 % de la variation du filage de roue. Le modèle est capable d'utiliser les quatre dernières caractéristiques énumérées ci-dessus, le comportement de jeu de l'élève sur d'autres problèmes, et l'identité de la compétence actuelle, puisque ceux-ci ne dépendent pas de la performance de l'élève sur la compétence actuelle.

Tableau 1. Performance du modèle pour la prédiction du Wheel-Spinning

# problèmes sur cette compétence	Roue-rotation %	Nagelkerke R^2	Faux positifs%.	Faux négatifs%.
0	9.8%	0.13	0.3%	98.2%
1	9.8%	0.28	1.0%	88.6%
2	9.8%	0.39	1.6%	73.9%
3	20.5%	0.37	4.7%	60.5%
4	28.5%	0.41	9.2%	47.3%
5	38.5%	0.45	15.3%	33.7%
6	53.2%	0.44	27.0%	21.2%
7	67.5%	0.65	28.2%	10.2%
8	83.5%	0.85	3.9%	4.6%

La quatrième colonne indique les faux positifs, c'est-à-dire les cas où le modèle prévoit que l'élève fera une pirouette, mais où en fait l'élève maîtrise la compétence dans les 10 premières occasions de s'entraîner. Le modèle a un taux de faux positifs assez faible, principalement en raison du déséquilibre des classes car, au départ, le patinage est une petite minorité des données. Le taux de faux positifs continue d'augmenter au fur et à mesure que les élèves en roue libre constituent un pourcentage de plus en plus important de l'ensemble de données. Cependant, en général, si le modèle affirme qu'un étudiant va faire la roue, il a généralement raison.

La cinquième colonne indique les faux négatifs, c'est-à-dire le cas où le modèle prédit que l'élève maîtrisera la compétence, mais qu'au lieu de cela il fait la roue. Au départ, ce taux est extrêmement élevé, principalement parce que le modèle ne veut pas prédire le patinage en roue libre, la classe de demi-norité, sur la base de peu de données sur la connaissance de cette compétence par l'élève. Le modèle ne parvient pas à détecter la plupart des cas où l'élève va faire une pirouette, et il est un peu conservateur dans ses prédictions. Ainsi, en tant qu'alerte précoce

pour empêcher les élèves d'avoir des séquences de résolution de problèmes frustrantes sans maîtrise, le détecteur a encore besoin de travail supplémentaire. Cependant, une fois que les élèves ont résolu deux problèmes sur une compétence, le système détecte relativement bien la rotation des roues et peut potentiellement éviter aux élèves une certaine frustration.

Une question se pose : quelle est la source de puissance du modèle ? En traçant les valeurs β estimées par la régression logistique, l'impact des caractéristiques est relativement stable d'un modèle à l'autre. L'importance du nombre de réponses correctes consécutives augmente à mesure que le nombre de problèmes vus augmente. Ce résultat a un sens intuitif ; par exemple, un élève qui a 0 réponse correcte à la suite du deuxième problème n'a pas autant de difficultés qu'un élève qui a 0 réponse correcte à la suite du sixième problème, car ce dernier a peu d'occasions d'obtenir 3 réponses correctes à la suite. Le temps de réponse normalisé de l'élève devient relativement moins important à mesure que le nombre de problèmes vus augmente. En fait, les élèves qui prennent relativement plus de temps que la moyenne pour résoudre un problème ont moins de chances de faire de la roue. Cette relation est un peu surprenante, mais pourrait peut-être être due au fait que les réponses rapides sont ambiguës et indiquent soit que l'élève est très habile (et qu'il est susceptible de maîtriser la compétence), soit qu'il entre une réponse aléatoire (et qu'il est susceptible de faire la roue). En général, les caractéristiques liées à la performance sur d'autres compétences deviennent nettement moins importantes au fur et à mesure que l'élève résout des problèmes sur cette compétence. Là encore, ce résultat est intuitif, car moins il y a de données disponibles sur cette compétence, plus les données sur les performances de l'élève dans d'autres compétences sont utiles.

Au-delà de la prédiction du Wheel-Spinning, nous avons également exploré sa relation avec le comportement négatif du gaming. Nous avons constaté que les 2 491 élèves qui n'ont jamais fait preuve de Wheel-Spinning ont obtenu un score moyen de 0,013 pour le jeu, tandis que les 366 élèves qui ont toujours fait preuve de Wheel-Spinning ont obtenu un score moyen de 0,163 pour le jeu. Ainsi, les élèves qui pratiquent le wheel-spinning sont également susceptibles de jouer. Mais cette relation se vérifie-t-elle chez un élève en particulier, c'est-à-dire que lorsqu'un élève fait la roue, est-il plus susceptible de jouer que lorsqu'il maîtrise une compétence ? Pour les 1207 élèves qui ont parfois fait la roue, leur score moyen de jeu était de 0,104 lorsqu'ils faisaient la roue, contre seulement 0,017 lorsqu'ils maîtrisaient une compétence en temps voulu. Ces chiffres sont similaires aux valeurs de jeu correspondantes pour les élèves qui faisaient toujours de la roue ou qui maîtrisaient toujours rapidement. Ce résultat suggère fortement que le jeu et le Wheel-Spinning sont liés. Cependant, il laisse en suspens la direction de la causalité.

4 Contributions

La principale contribution de cet article est d'identifier un nouveau problème de modélisation de l'étudiant qui peut être résolu par le module de prise de décision du tutorat d'un STI. La plupart des efforts de modélisation des étudiants (par exemple [5, 7, 8]) et la coupe KDD 2010 sur la modélisation des données éducatives se concentrent sur la prédiction du comportement des étudiants au niveau des réponses individuelles. Bien que cette approche valide clairement le modèle de l'étudiant, il est raisonnable de se demander pourquoi ce problème est intéressant en premier lieu, en particulier du point de vue de la création d'un STI efficace et adaptatif. Si nos modèles avaient la moitié du taux d'erreur de l'état actuel de la technique, cela améliorerait-il sensiblement la performance des tuteurs informatiques ? Il est difficile de savoir quelles décisions de tutorat seraient affectées par une telle amélioration.

Le modèle de l'apprentissage par la maîtrise ne permet que de légères améliorations du modèle de l'apprentissage par la maîtrise pour savoir quand considérer que l'élève a terminé une compétence. En revanche, un modèle solide de la rotation des roues a des implications claires sur la façon d'adapter l'enseignement à l'élève. Considérons la figure 2 comme un modèle possible d'intégration d'un détecteur de rotation dans un STI, en modifiant le cycle typique d'apprentissage par la maîtrise du STI pour ne pas toujours présenter un autre problème en cas d'erreur de l'élève. Si l'élève est susceptible de faire un tête-à-queue, il est inutile de lui proposer un autre problème, car il est peu probable qu'il maîtrise la compétence. Comme la résolution de problèmes est, statistiquement, un exercice futile pour cet apprenant à l'heure actuelle, faire autre chose que de la résolution de problèmes semble justifié. Il existe plusieurs méthodes d'enseignement possibles, notamment l'intervention de l'enseignant, le tutorat par les pairs ou l'intégration d'une instruction plus poussée dans le STI lui-même (p. ex. [9]). Nous ne disposons pas de données suffisantes pour prescrire une solution particulière au Wheel-Spinning, mais nous sommes prêts à conclure que la résolution de problèmes supplémentaires n'est pas une approche viable. En outre, le Wheel-Spinning peut être calculé à partir des fichiers journaux et ne nécessite pas de codeurs humains pour former un modèle, et il représente également un pourcentage modéré (10% à 35%) du comportement.

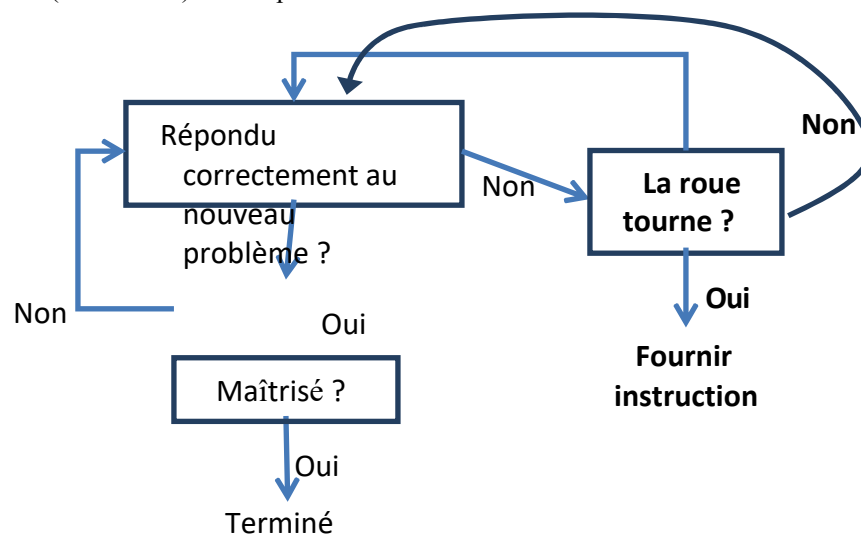


Fig. 2. Modèle de processus possible pour l'intégration de la rotation des roues dans un STI

Enfin, cet article propose une nouvelle approche pour évaluer l'impact d'un système de tutorat intelligent en utilisant la Fig. 1 comme visualisation. Nous pouvons séparer les étudiants en trois catégories. Tout d'abord, il y a les 35 % d'étudiants qui, à la fois dans les cours de TAO et d'ASSISTANCE, ont maîtrisé la matière sans faire d'erreurs ; comme ils connaissaient la compétence avant de commencer, ils n'ont pas bénéficié directement des composants de tutorat. Deuxièmement, il y a les 9,8% (limite inférieure) à 35% (limite supérieure) d'étudiants qui ont fait la roue, ils n'ont pas bénéficié du tuteur. Le troisième groupe d'étudiants est celui des étudiants qui ont *potentiellement* bénéficié du tuteur, un groupe comprenant 30 à 55% de la population étudiante. Nous ne pouvons pas déterminer si ces étudiants ont réellement bénéficié ou non du tuteur, car ils auraient peut-être maîtrisé la compétence avec une simple pratique du crayon et du papier. Cependant, la limite supérieure du pourcentage d'étudiants qui auraient pu être aidés par le tuteur existant est étonnamment faible.

5 Conclusions et travaux futurs

Ce travail est une première tentative de modélisation et de compréhension du Wheel-Spinning, et plusieurs questions restent sans réponse. Premièrement, à quoi ressemble le Wheel-Spinning dans d'autres systèmes ? Nous avons examiné deux tuteurs de mathématiques et avons trouvé des patrons de comportement largement similaires. Ce problème existe-t-il dans d'autres domaines de tutorat ? Une deuxième question est de savoir quelle est l'unité de mesure appropriée pour l'axe des x ? Ce travail a utilisé le nombre de problèmes, mais peut-être le temps total passé serait-il un meilleur indicateur ?

La deuxième dimension des travaux futurs consiste à comprendre la nature du wheel-spinning et à déterminer dans quelle mesure le détecteur est généralisable. Nous avons construit un ensemble de caractéristiques prédictives sur la base de nos croyances sur ce qui serait prédictif, mais il y a, nous l'espérons, des prédicteurs supplémentaires qui peuvent être mis à contribution pour ce problème. Une question connexe est de savoir si les prédicteurs sont cohérents entre les systèmes de tutorat ; dans quelle mesure un détecteur pour ASSISTments fonctionnerait-il sur CAT, et vice versa ? Un tel détecteur serait-il généralisable à un domaine non mathématique ? Une étape probablement importante dans ce processus est la normalisation appropriée des données, similaire à ce qui a été fait pour le temps de réponse.

Le troisième domaine de travail consiste à explorer la relation entre le Wheel-Spinning et les comportements négatifs tels que le jeu [10] et le comportement hors tâche [11]. Nous avons constaté que le jeu et le Wheel-Spinning sont des comportements corrélés, mais la direction de la causalité est une question, car il existe trois modèles plausibles :

- Le jeu entraîne la rotation de la roue. Les élèves ne prennent pas les problèmes au sérieux et demandent des conseils dont ils n'ont peut-être pas besoin. Les demandes d'aide étant considérées comme des réponses incorrectes, les élèves tournent en rond et ne parviennent pas à maîtriser le problème.
- La rotation de la roue entraîne le jeu. Les élèves qui ne comprennent pas la matière sont incapables de résoudre les problèmes et deviennent frustrés. Ces élèves n'ont d'autre choix que de demander de nombreux conseils, car de nombreux STI n'ont pas de composantes structurelles solides.
- Le jeu et la rotation des roues sont des symptômes qui sont affectés par une cause commune.

Ces modèles ont des implications très différentes, puisque le premier modèle suggère que tenter d'agir directement sur l'humeur de l'élève est une approche viable. Le second modèle suggère que l'instruction est plus susceptible d'être bénéfique. En réalité, il y a probablement un mélange des deux comportements. Il est intéressant de noter que les premiers travaux sur la remédiation du jeu ont eu un effet positif [10], mais comprenaient des éléments qui tentaient à la fois de décourager le jeu, mais aussi d'ajouter un soutien pédagogique au-delà de ce qui était précédemment disponible dans le tuteur. Des études contrôlées qui fournissent des instructions aux étudiants qui sont susceptibles de jouer à la roulette seraient utiles pour déterminer laquelle des deux hypothèses candidates est la plus proche de la réalité.

Nous voyons deux conséquences claires de ce travail. Premièrement, il n'est peut-être pas judicieux d'utiliser toutes les données à des fins d'ajustement de modèle lors de la formation d'un modèle d'étudiant. Puisque la plupart des modèles d'étudiants supposent une probabilité fixe d'apprendre une compétence, de longues séquences de problèmes par des étudiants Wheel-Spinning sont susceptibles de sous-estimer le taux d'apprentissage de l'étudiant moyen. La distribution des taux d'apprentissage peut être considérée comme bimodale avec un groupe de paires de compétences d'étudiants Wheel-Spinning regroupées autour de 0. Ainsi, le travail sur la détection des facteurs contextuels qui affectent l'apprentissage [par exemple, 12] est un développement bienvenu.

La deuxième conséquence est que les concepteurs de STI doivent développer une solution de repli pour les échecs de l'apprentissage par la maîtrise. Le cycle simplifié de l'apprentissage par la maîtrise, qui consiste à " présenter les problèmes jusqu'à la maîtrise ", ne fonctionne pas pour de nombreux apprenants, même avec l'aide disponible dans deux tuteurs populaires. Une modification ou une intervention automatisée est justifiée si nous voulons éviter de frustrer les apprenants.

Remerciements

Ce travail a été soutenu par la National Science Foundation (subvention DRL-1109483) au Worcester Polytechnic Institute. Les opinions exprimées sont celles des auteurs et ne représentent pas nécessairement les vues de la Fondation.

Références

1. Koedinger, K.R., et al, Intelligent Tutoring Goes To School in the Big City. *International Journal of Artificial Intelligence in Education*, 1997. **8** : p. 30-43.
2. Bloom, B.S., *Caractéristiques humaines et apprentissage scolaire*. 1976, McGraw-Hill.
3. Frick, T.W., A comparison of three decision models for adapting the length of computer-based mastery tests. *Journal of Educational Computing Research*, 1990. **6**(4) : p. 479-513.
4. Corbett, A. et J.R. Anderson, Knowledge tracing : Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 1995. **4** : p. 253-278.
5. Beck, J.E. et K.-m. Chang. Identifiability : A Fundamental Problem of Student Modeling. in *International Conference on User Modeling*. 2007. Corfou, Grèce.
6. Menard, S., *Applied Logistic Regression Analysis. Applications quantitatives dans les sciences sociales*. 2001 : Sage Publications.
7. Pardos, Z., et al. Analyzing fine-grained skill models using bayesian and mixed effect methods. in *Thirteenth Conference on Artificial Intelligence in Education*. 2007 : IOS Press.
8. Chi, M., et al, Instructional Factors Analysis : A Cognitive Model For Multiple Instructional Interventions, in *Proceedings of Educational Data Mining*. 2011.
9. Koning, K.d., et al, Model-based reasoning about learner behaviour. *Artificial Intelligence*, 2000. **117** : p. 173–229.
10. Baker, R.S.J.d., et al. Adapting to When Students Game an Intelligent Tutoring System. in *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*. 2006..
11. Baker, R.S.J.d. Modeling and Understanding Students' Off-Task Behavior in Intelligent Tutoring Systems. in *Proceedings of ACM CHI 2007 : Computer-Human Interaction*. 2007.
12. Baker, R.S.J.d., A.B. Goldstein, et N.T. Heffernan, Detecting the Moment of Learning, in *Proceedings of the 10th Annual Conference on Intelligent Tutoring Systems*. 2010. p. 25-34.