L'item choisi est celui pour lequel $|(1) - (2)|$ est maximum. Si la valeur absolue de cette différence est faible, cela signifie que le fait qu'un item ait été réussi ou manqué ne permet pas d'affirmer que l'objectif est maîtrisé.

Les éléments suivants sont nécessaires pour la mise en oeuvre d'un tel modèle:

- Le testeur doit définir un *seuil de réussite*, c'est à dire fixer la probabilité minimum acceptable qu'un élève maîtrise un objectif. Le test s'arrêtera par exemple lorsqu'une probabilité de 0.8 (quatre-vingt chances sur cent) que l'élève maîtrise les bases de l'arithmétique sera atteinte.
- Il est nécessaire de fournir un certain nombre de *probabilités de maîtrise à priori* pour que le réseau soit "calculable". C'est ici que réside la principale limitation de ce type de modèle. La détermination de ces probabilités à priori est malaisée et nécessite beaucoup de temps. Elle ne peut provenir que d'un expert du domaine (par exemple un enseignant qui sait où les élèves font des erreurs) ou d'une pré-enquête sur un grand nombre de sujets.
- En revanche, l'adjonction de *relations d' inclusion et de prérequis* entre les objectifs est relativement aisée. Les relations d'inclusion regroupent plusieurs objectifs élémentaires en un objectif plus général. Par exemple, l'ensemble constitué par l'addition, la soustraction, la multiplication et la division forme un nouvel objectif plus général que l'on pourrait nommer "arithmétique élémentaire". En structurant ainsi le domaine à tester, l'information qu'apporte la réussite ou l'échec d'un item élémentaire peut être propagée le long des arcs du graphe. Le système peut alors poser un diagnostic sur la maîtrise d'un objectif en utilisant un minimum d'items.

Référence

:

Collins, J. A., Greer, J. E., Huang, S. X. Adaptive Assessment using Granularity Hierarchies and Bayesian Nets. In Proceedings of the Third International Conference on Intelligent Tutoring Systems. Montreal, Canada. 1996.

P. Jermann & P. Dillenbourg, Université de Genève, FPSE, TECFA.

Patrick.Jermann@tecfa.unige.ch

pdillen@divsun.unige.ch

Stratégies de sélection des items dans les tests adaptatifs.

Les tests informatisés offrent certains avantages sur les tests papier-crayon et sur les interviews : précision des mesures, standardisation des conditions de passation et enregistrement automatique des données. Par contre, lorsque le test est conduit par une personne, celle-ci peut adapter les questions, passer des items inutiles ou poser certaines questions plus précises afin d'affiner le diagnostic. Un test adaptatif est capable de sélectionner dynamiquement des items en cours d'interaction.

Nous traiterons ici des stratégies de sélection des items. Celles-ci consistent à laisser de côté des items qui n'apporteront plus d'information pertinente par rapport à l'objectif du test. Ces stratégies permettent de réduire le temps de passation des tests et de sauvegarder la motivation du sujet. Quoi de plus frustrant, en effet, que de devoir continuer à répondre à des questions qu'on a aucune chance de réussir ? Ces stratégies permettent également d'économiser des items : dans le cas de mesures répétées, avant et après une séquence d'apprentissage par exemple, il peut être intéressant de garder un maximum d'items 'vierges'.

On distingue deux stratégies de sélection des items.

- Selon la *stratégie ascendante*, on ne présente pas les items supérieurs en difficulté à celui que le sujet n'a pas réussi, puisqu'il n'a en théorie aucune chance de réussir un item plus difficile (si l'épreuve est bien construite). Le test prend fin dès que le sujet ne répond pas correctement à un item. Comme certaines erreurs sont dues à l'inattention on peut fixer le critère d'arrêt de passation à plusieurs erreurs consécutives.
- La *stratégie descendante* consiste à commencer par les items les plus difficiles et à proposer des items plus faciles jusqu'au moment où le sujet réussit. On suppose dans ce cas qu'il n'échouera pas ensuite à des items plus faciles.

Le choix d'une stratégie montante ou descendante dépend de l'estimation à priori du niveau du sujet. Dans un pré-test, celui-ci n'est généralement pas supposé maîtriser les compétences mesurées. Aussi, la stratégie ascendante permettra un gain de temps maximum. Dans un post-test, par contre le sujet est censé maîtriser les compétences évaluées. La stratégie descendante sera alors préférable.

Les stratégies ascendante et descendante peuvent être mises en oeuvre à un plus haut niveau, celui des objectifs pédagogiques. Si le sujet échoue la partie de l'épreuve concernant l'objectif X et que l'objectif X est un pré-requis indispensable à l'objectif Y, il y a peu de chances qu'il maîtrise l'objectif Y. Inversement, si un élève maîtrise l'objectif Y, il est peu probable qu'il ne maîtrise pas l'objectif X. La relation de pré-requis entre deux objectifs est souvent plus facile à établir qu'un rapport de difficulté entre deux items.

Certaines stratégies de sélection appliquent les *méthodes de tri*. Considérons un test comportant 100 items classés par ordre croissant de difficulté. Le système sélectionne d'abord l'item central, de difficulté 50. Si le sujet échoue, le système ne considère plus que le segment 0-50 et recommence la procédure avec l'item 25. Si l'élève réussit, il considère le segment 25-50 et on répète la procédure jusqu'au moment où l'on obtient une précision de mesure considérée comme acceptable. Cette stratégie permet de minimiser le nombre moyen d'items nécessaires pour identifier le niveau du sujet. Elle exige cependant une mise au point minutieuse de la batterie d'items. La précision nécessaire quant à l'évaluation du niveau de difficulté des items ne peut en effet se faire qu'en expérimentant au préalable chaque item auprès de très nombreux sujets.

Certains chercheurs en intelligence artificielle tentent de développer des méthodes plus sophistiquées que celles que nous avons abordées jusqu’ici. Il s’agit plus dans ce cas de recherche fondamentale que de méthodes éprouvées et facilement applicables. Collins (1996) par exemple, présente un modèle basé sur les *réseaux probabilistes* (réseaux bayesiens ou réseaux de croyance) qui adresse le problème de la sélection des items dans les tests adaptatifs. Imaginons que les objectifs pédagogiques qu'un élève est censé maîtriser constituent un réseau. A chaque objectif est associé la probabilité que l'élève le maîtrise.

La règle de Bayes permet de répondre à des questions du type: "quelle est la probabilité que l'élève maîtrise l'objectif X étant donné qu'il a répondu juste/faux à l'item Q ?" (on parle dans ce cas de probabilités conditionnelles). Un réseau probabiliste applique cette règle de calcul pour établir la probabilité qu'un élément du réseau soit vérifié en utilisant les réponses de l’élève d’une part, et les relations de prérequis entre les éléments d’autre part.

Après chaque réponse de l'élève, le calcul de la probabilité associée à chaque objectif est réitéré et un nouvel item est sélectionné. Le choix se fait selon une fonction d'utilité qui mesure le pouvoir discriminant d'un item relativement à un ou plusieurs objectifs à tester. L'application de la fonction d'utilité revient à poser deux questions:

(1) quelle est la probabilité que l'objectif X soit maîtrisé si l'élève répond **juste** à l’item Q ?

(2) quelle est la probabilité que l'objectif X soit maîtrisé si l'élève répond **faux** à l’item Q ?