Agent d'apprentissage et la gestion du processus pédagogique

Abdellah BENNANE
CFIE, Rabat, Maroc
bennanea@yahoo.fr

RESUME. *Le but est de développer un système tuteur adaptatif, sachant que les échanges qui auront lieu lors de l'interaction entre le système et ses utilisateurs potentiels, se dérouleront avec un haut niveau d'interactivité et prendront en compte les caractéristiques individuelles de l'apprenant tels que le niveau d'étude. Pour atteindre cet objectif, généralement, on fait appel à des techniques issues de la machine automatique afin d'approprier le système aux conditions interne et externe de l'environnement, et de permettre à ce système de durer et d'interagir efficacement avec son utilisateur. L'objet de ce papier est d'automatiser la gestion des taches pédagogiques du système tuteur, en l'occurrence le choix des situations pédagogiques, par la création d'un agent d'apprentissage qui simplifie la logique manuelle et prend en charge le déroulement et la gestion du processus d'enseignement à travers des interactions naturelles.*

MOTS CLES : s*ystème adaptatif, Tuteur intelligent, module pédagogique, séquence pédagogique adaptative, apprentissage par renforcement, agent, environnement, interaction naturelle.*

SUMMARY. *The goal of this project is to develop an adaptive tutoring system. The exchanges that will take place at the time of the interaction between the system and its potential users, will be realized with a interactivity high-level and take into account the learner's individual features like study level. In this end, generally, one uses the artificial intelligence techniques in order to appropriate the system to the environment internal and external conditions, and allow this system to interact efficiently with its potentials user. The object of this paper is to automate and manage the pedagogical process of tutoring system, in particular the selection of the content and manner of pedagogic situations. We create a pedagogical learning agent that simplifies the manual logic and takes in charge the progress and the management of the teaching process (tutor-learner) through natural interactions.*

KEYWORDS*: adaptive system, tutoring system, agent, environment, natural interaction, reinforcement learning, pedagogic process.*

1. Problématique

La conception classique des tuteurs intelligents, est composée de quatre éléments, le modèle du domaine (connaissances), le modèle de l'élève, le module pédagogique, et le module de communication [Wenger87].
Le module pédagogique est le chef d'orchestre de l'action pédagogique entreprise dans un système tuteur. C'est à lui que revient le choix des situations que le système présente à l'élève. Traditionnellement, le module pédagogique est un ensemble de règles élaborées par l'auteur ou le concepteur du système. ***Ces règles sont définies et développées manuellement en se basant sur l'expérience de l'auteur du système tuteur. Cette logique « manuelle » est plus subjective qu'objective car elle est liée à l'auteur et ses expériences pédagogiques***. La tendance est de remplacer par des approches statistiques les approches traditionnelles basées sur des règles établies manuellement. Il faut développer des approches probabilistes et des modèles flexibles, qui permettent à la machine d'apprendre par elle-même à partir des données reçues [Aimeur2004]. La corrélation de la performance de l'apprentissage est basée sur le moteur de la déduction statistique qui rassemble de l'information au sujet du comportement de l'utilisateur pour chaque parcours de l'apprentissage individualisé et créant une distribution des probabilités pour l'ensemble entier du module de formation. [Sonwalkar2004]. L'apprentissage automatique s'est intéressé, pour une grande part, à développer les algorithmes qui peuvent étiqueter les nouvelles données, et non pas aux systèmes qui apprennent naturellement et en interaction avec les gens. Nous souhaitons contribuer à changer ce centre d'intérêt (focus) pour développer de nouveaux genres de systèmes qui apprennent avec les utilisateurs à travers une interaction naturelle [Picard-Papert2004].
Pour solutionner le problème posé, n'est-il pas possible de remplacer la logique « manuelle » par une logique « numérique » qui peut exploiter les parcours des apprenants ? Peut-on assurer l'adaptabilité de l'environnement d'enseignement aux apprenants et objectivement ? Ces questions expriment une hypothèse que nous voulons vérifier.
Plusieurs techniques issues de l'apprentissage automatique ont été utilisées pour résoudre ce genre de problèmes. L'automatisation permet de gagner de l'énergie et du temps et de vaincre les problèmes multiples de l'adaptabilité des systèmes tuteurs. Le module pédagogique peut être généré dynamiquement en se basant uniquement sur l'interaction de l'élève avec son environnement d'enseignement. Les efforts investis ont été dans deux directions, d'une part, la conception d'une base de données qui tient en compte les propriétés significatives de l'acte pédagogique, de l'autre, chercher et essayer parmi les techniques de l'apprentissage basées sur les calculs numériques, celle qui est plus appropriée au champ d'étude telles que les techniques de l'apprentissage par renforcement (RL).

2. Elaboration de la solution

2.1. Pédagogie différenciée et RL

La solution repose sur deux piliers, le premier

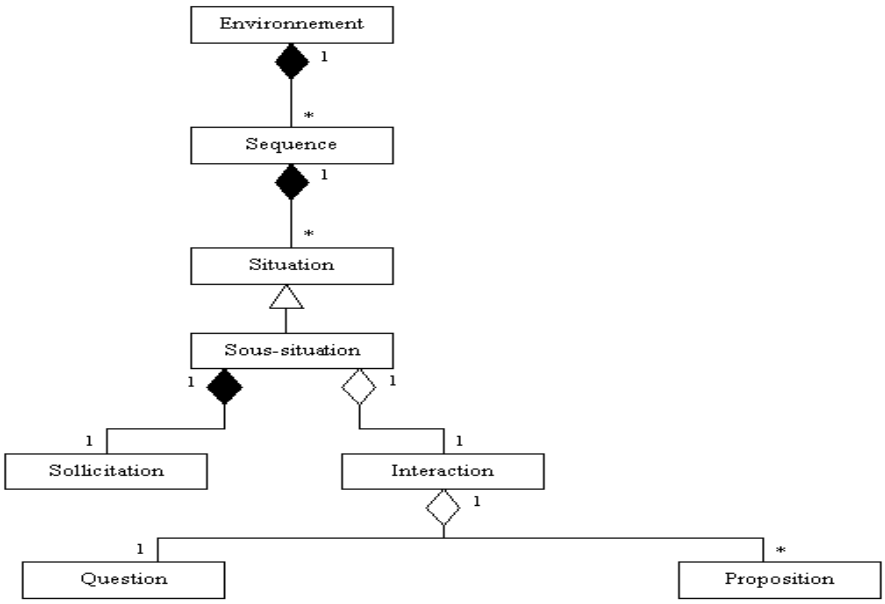

Figure 1 : Digramme de classes de l'« Environnement »

est l'utilisation de RL et le second est la manière avec laquelle nous structurons et organisons l'environnement d'enseignement.
Nous associons l'approche modulaire et la pédagogie différenciée [Przesmycki91] pour structurer et organiser l'environnement d'enseignement. Dans ce sens, une séquence pédagogique est une succession de situations. L'élaboration d'une situation repose sur deux paramètres importants, l'individualisation et la variété [Bennane2001]. Une pédagogie individualisée est une pédagogie qui reconnaît l'élève comme une personne ayant ses représentations propres de la situation de formation. Alors qu'une pédagogie variée est une pédagogie qui propose un éventail de démarches, s'opposant ainsi au mythe identitaire de l'uniformité, faussement démocratique, selon lequel tout le monde doit travailler au même rythme, dans la même durée, et par les mêmes itinéraires. Cette approche permet de contribuer à résoudre le problème de l'échec scolaire où la négligence du niveau d'étude des apprenants dans le processus d'évaluation fait défaut. En général, les enseignants et les formateurs quand ils préparent un module ou un cours, ils visent toute la classe et l'individu (apprenant) est ignoré par le fait que le cours (module) est

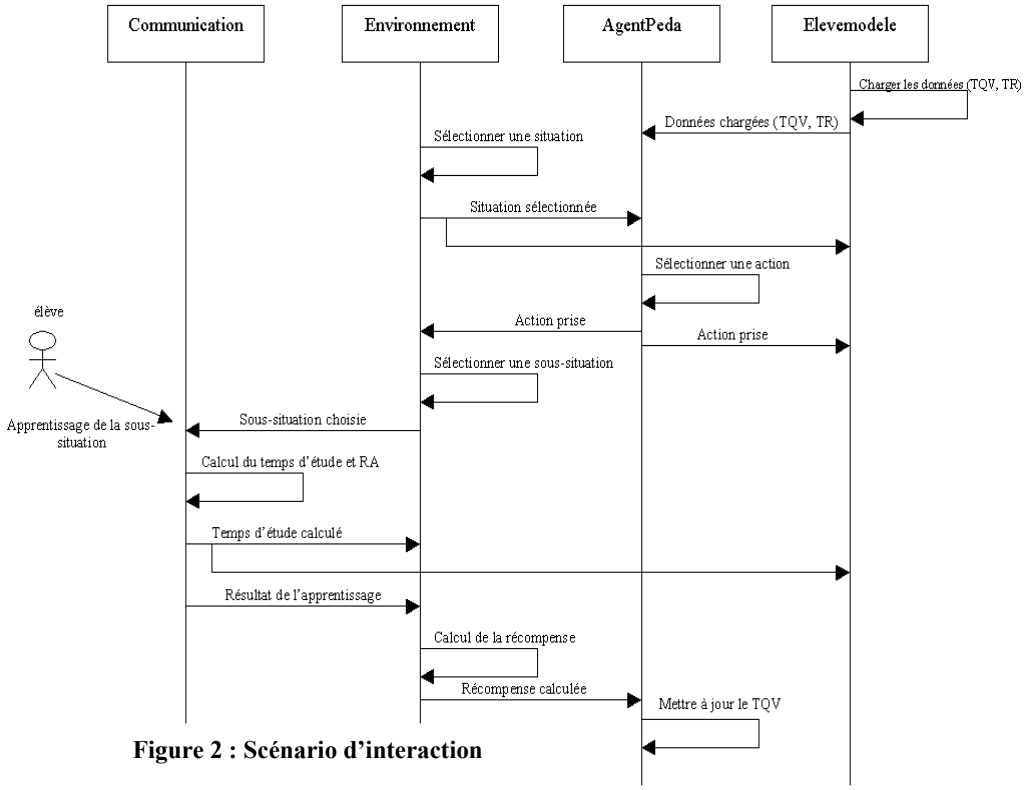

Figure 2 : Scénario d'interaction

collectif. Notre conception demande un effort supplémentaire de la part des enseignants et des formateurs afin que leur produit tient en compte les caractéristiques individuelles qui font la différence entre les apprenants et leur enseignement tels que le niveau d'étude par exemple. Par ce fait, une situation pédagogique sera un paquet de sous-situations.
Le choix de RL est dû à la nature de son modèle. Il est adapté au contrôle de l'apprentissage humain comme il a été fait depuis les travaux des pionniers tels que Watson, Pavlov, Skinner, Bellman, etc.. Notre choix part de l'idée suivante : dans un environnement d'enseignement automatique, nous trouvons deux agents, l'un externe au système (acteur), l'autre interne. Le premier agent est l'élève. C'est un agent naturel qui veut apprendre un cours, un module d'enseignement. Cet agent a besoin d'un tuteur afin qu'il puisse l'aider à apprendre en lui choisissant les situations adaptées à son niveau. Donc, on a besoin d'un agent pédagogique qui peut remplir cette fonction. Le second agent est l'agent pédagogique dont la fonction est d'aider les élèves à progresser en sélectionnant les situations les plus adaptées à leurs niveaux. L'agent pédagogique est un agent artificiel, interne car il fait partie d'un système de formation automatisée. L'agent pédagogique ne peut remplir cette fonction que s'il possède la faculté d'apprendre. La théorie et le modèle de RL peuvent assurer à l'agent pédagogique

d'apprendre par essai et erreur (par expérience). En d'autres termes, l'apprentissage de l'agent pédagogique ne peut se réaliser qu'à travers l'apprentissage des élèves.

La question clé est la suivante : comment sélectionner les sous-situations les plus appropriées au niveau d'un apprenant donné dans le but d'atteindre l'objectif d'une séquence donnée ?. Afin de répondre à cette question, il faut d'abord présenter la fonction du modèle de l'élève [VanLehn88]. Le modèle de l'élève a deux fonctions. D'une part, c'est une mémoire qui stocke toutes les transactions qui passent d'un objet à l'autre du système et elle s'organise en utilisant le quadruplet (situation, action, situation suivante, temps d'étude) afin de collecter les données des transactions. De l'autre, c'est une ressource de données qu'on exploite pour extraire la fréquence de chaque transition et la moyenne des récompenses immédiates. Ces extractions nous permettent de construire deux tableaux, le tableau des fréquences des transitions et le tableau de la moyenne des récompenses immédiates. Ces deux tableaux sont envoyés au module pédagogique. Le module pédagogique est le moteur de l'adaptabilité du système tuteur et c'est à lui que revient le choix du parcours de l'apprentissage de l'apprenant. Au sens de l'apprentissage par renforcement, il a besoin d'un nombre de données et de fonctions afin de répondre aux exigences de la tâche. Son but est de déterminer une fonction optimale Q* [Sutton98] qui sera le critère sur lequel repose la sélection des sous-situations. Les valeurs de cette fonction sont définies par : $Q^*(s, a) = \sum_{s'} P^a_{ss'}.(R^a_{ss'} + \gamma.\max_{a'} Q^*(s', a'))$, $\forall$ (s, a) $\in$ SxA et γ un paramètre, $0 \leq \gamma \leq 1$, où $P^a_{ss'}$ et $R^a_{ss'}$ sont respectivement la probabilité de transition et la récompense. On fait appel à deux tableaux, le tableau des fréquences des transitions afin de calculer la distribution des probabilités de transition, et le tableau de la moyenne des récompenses immédiates dans le but de calculer les valeurs de la fonction de récompense. Donc, Déterminer Q* revient à déterminer une fonction de récompense (fr) et une distribution des probabilités de transition (p).

2.2. Distribution des probabilités et fonction de récompense

D'après l'organisation d'une séquence pédagogique comme nous l'avons définie, les successeurs d'une situation donnée (s) sont :

- soit elle même si l'action prise est accomplie par échec ;
- soit (s') si l'action prise est accomplie par succès.

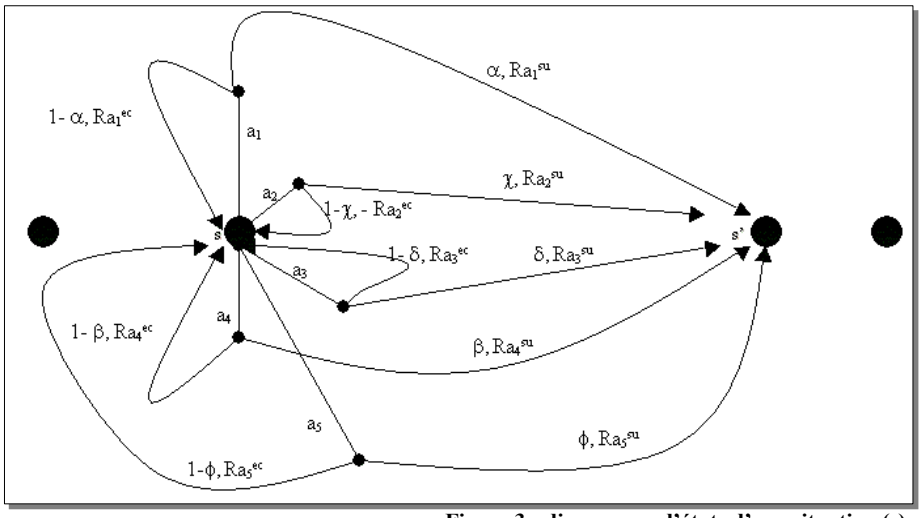

Figure 3 : diagramme d'états d'une situation (s) non terminal, à une situation (s')

La fonction de récompense est définie de l'ensemble SxAxS vers R. Elle est notée fr. Soulignons que les valeurs Ra_1^{su}, Ra_1^{ec} , Ra_2^{su}, Ra_2^{ec} , Ra_3^{su}, Ra_3^{ec} , Ra_4^{su}, Ra_4^{ec} , Ra_5^{su}, Ra_5^{ec} utilisées à titre indicatif dans la figure (3) sont inconnues.

Le choix d'une fonction de récompense n'est pas évident car dans plusieurs cas, la fonction de récompense proposée ne permet pas de satisfaire une hypothèse donnée [BeckWolf2000] dans le but de déterminer la fonction optimale. Nous soulignons qu'une simulation est indispensable afin de déterminer la fonction de récompense avant de passer à l'expérience réelle.

La distribution des probabilités de transition est définie de SxAxS vers [0, 1]. Elle est notée (p). Soulignons que les valeurs α, χ, δ, β, et ϕ

utilisées à titre indicatif dans la figure (3) sont inconnues.

Les valeurs des deux fonctions (fr) et (p) sont inconnues. Elles seront apprises par expérience afin de compléter notre modèle. Comment déterminer la distribution des probabilités? Nous élaborons une base de données dans laquelle on stocke les observations (s_t, a_t, s_{t+1}, td_t). Cette base de données sera exploitée pour déterminer les valeurs des deux fonctions que nous cherchons pour compléter le modèle, puis utiliser la programmation dynamique pour déterminer la fonction optimale Q*.

La distribution des probabilités de transition est définie de la façon suivante :

p : SxAxS $\rightarrow$ [0, 1] décrit les probabilités de transition de l'état courant (s) vers l'état suivant (s') quand l'action (a) est prise. Considérons une base de données de X observations (s_t, a_t, s_{t+1}, r_t) générées lors des expériences.

Mentionnons que les successeurs d'une situation donnée (s) sont, soit elle même si l'action prise est accomplie par échec ; soit (s') si l'action prise est accomplie par succès. Sur une estimation basée sur les fréquences relatives dans la base de données, (p) peut s'écrire de la façon suivante:

- p(s, a, s') =

$$\frac{|\{x \in LS / s_t(x) = s,\ a_t(x) = a,\ s_{t+1}(x) = s' \text{ et } s' \neq s\}|}{|\{x \in LS / s_t(x) = s,\ a_t(x) = a\}|}$$

si $s \neq s'$.

- $p(s, a, s') = \frac{|\{x \in LS / s_t(x) = s,\ a_t(x) = a,\ s_{t+1}(x) = s\}|}{|\{x \in LS / s_t(x) = s,\ a_t(x) = a\}|}$ si s'=s
- p(s, a, s') + p(s, a, s) =1

La fonction de récompense est définie de la façon suivante :

fr : SxAxS $\longrightarrow \Re$

(s, a, s') $\longrightarrow$ fr(s, a, s')

Pour déterminer la fonction de récompense, nous partons de l'hypothèse suivante : au cours de l'exploration, les sous situations les plus réussies seront considérées comme très faciles, les moins réussies seront considérées comme très difficiles et entre les deux se répartissent le reste des sous situations suivant les valeurs de la fonction optimale. Rappelons que la récompense est le moyen de communiquer à l'agent ce qu'on veut qu'il fasse et non pas comment il le fasse [Sutton, 98]. Sur une estimation basée sur les fréquences relatives dans la base de données, (fr) peut s'écrire de la façon suivante :

$$fr(s, a, s') \cong \frac{\sum_{o \in \{x \in LS / s_t(x)=s,\ a_t(x)=a,\ s_{t+1}(x)=s'\}} r_t(o)}{|\{x \in LS / s_t(x)=s,\ a_t(x)=a,\ s_{t+1}(x)=s'\}|}$$ où $r_t(o)$ est la récompense immédiate associée à l'observation o, à l'instant t. fr(s, a, s') est la moyenne des récompenses immédiates reçues le long des expériences antérieures [Jodogne003].

2.3. Algorithme de la fonction optimale

Rappelons que la fonction optimale Q*est définis par l'expression suivante :

$Q^*(s, a) = \sum_{s'} P^a_{ss'}.(R^a_{ss'} + \gamma \max_{a'} Q^*(s', a')) \ \forall \ (s, a) \in$ SxA et γ un paramètre, $0 \leq \gamma \leq 1$.

Les successeurs d'une situation (s) sont (s) et (s'). D'où Q*:

$Q^*(s, a) = p(s, a, s).[fr(s, a, s) + \gamma.\max_{a'} Q^*(s, a')] + p(s, a, s').[fr(s, a, s') + \gamma.\max_{a'} Q^*(s', a')]$

Les valeurs de Q* peuvent être déterminées par l'algorithme suivant :

Initialiser Q(s, a) à 0, pour tout* (s, a) ∈ SxA
Répéter
  Delta=0
  Pour (s, a) ∈ SxA
    Q=Q*(s, a)*
    $Q^*(s, a) = p(s, a, s).[fr(s, a, s) + \gamma.\max_{a'} Q^*(s, a')] + p(s, a, s').[fr(s, a, s') + \gamma.\max_{a'} Q^*(s', a')]$
    Delta=max(Delta, abs(Q- Q*(s, a))*
Jusqu'à (Delta < θ)

θ est un infiniment petit, et γ est un paramètre compris entre 0 et 1.

Les valeurs de la fonction optimale Q* vont être utilisées dans la phase de l'exploitation pour sélectionner les actions.

3. Validation de la solution

Nous avons développé un didacticiel afin de vérifier la faisabilité de notre théorie. La base de données du didacticiel est développée et produite par un ensemble d'enseignants. Le didacticiel est doté d'un agent pédagogique qui gère le processus pédagogique entre les apprenants et l'environnement d'enseignement. Afin de mesurer l'évolution de l'apprentissage de l'agent pédagogique, nous avons utilisé trois critères d'évaluation. Le premier est la probabilité de succès des situations pédagogiques afin de mesurer son évolution d'une situation à une autre. Pour tous les parcours (itérations), nous calculons les fréquences de succès et d'échec de l'apprentissage des situations à partir de la base de données du système d'apprentissage, puis nous calculons la distribution des

probabilités de succès. Cette distribution nous permet de voir, d'une part, la tendance de cette fonction (croissante, décroissante, constante, en hausse, en baisse, etc), de l'autre, le niveau de l'amélioration de la probabilité de succès, en l'occurrence, entre les premières et les dernières situations de la séquence pédagogique. Nous pratiquons cette démarche pour les deux phases (exploration et exploitation) dans le but de comparer leurs résultats. Le second critère est le temps d'étude à investir pour apprendre la séquence pédagogique dans le but de mesurer le taux de réduction du temps d'étude. Pratiquement, nous calculons le temps d'étude pour chaque parcours (itération) et nous comparons les résultats des deux phases et nous déduisons le taux de réduction du temps d'étude. Le troisième critère est le gain relatif qui permet de mesurer la performance des élèves avant et après l'étude du didacticiel afin de déterminer l'impact du didacticiel sur les apprenants [Bennane2003].

Le but est d'apprendre une séquence optimale d'actions qui déplaceront le système d'un état arbitraire à un état objectif. La mesure de la qualité du choix des actions de l'agent pédagogique est l'objet de notre étude. Si l'élève accomplit et les situations avec une probabilité de succès intéressante et la séquence pédagogique avec un temps optimal, alors on peut dire que l'assistance de l'agent pédagogique est achevée avec succès.

L'essai du didacticiel et le calcul d'une fonction optimale (Q*) concerne l'essai du didacticiel « expression de but » dans les deux phases, l'**exploration** et l'**exploitation**. Les données collectées lors de la phase de l'exploration seront utilisées dans le but de déterminer les valeurs d'une fonction optimale, qui seront utilisées de leur part comme critère pour choisir les actions de l'agent pédagogique dans la phase de l'exploitation.

Pour atteindre ce but, nous avons choisi un échantillon de 60 élèves qui sont distribués en deux groupes. Le premier a essayé le didacticiel dans la phase de l'exploration, et le second a essayé le même didacticiel dans la phase de l'exploitation.

Nous avons essayé de vérifier trois choses, le niveau de réduction du temps d'apprentissage, le niveau de l'amélioration de la probabilité de succès des situations d'apprentissage et finalement l'impact de l'apprentissage du didacticiel sur les apprenants en utilisant le gain relatif. Les tables ci-dessous enregistrent les résultats de notre expérience pour les trois critères, le temps d'étude, la probabilité de succès et le gain relatif.

Entre les deux phases, nous avons enregistré une réduction du temps d'étude de 29,3% et une réduction de l'écart type de 72 %.

Temps d'étude	T_{min}	T_{max}	T_{moy}	σ
Exploration	883	2500	1331,45	352,27
Exploitation	768,9	1130,38	941,71	97,92
Synthèse			↓ 29,3 %	↓ 72 %

En ce qui concerne la distribution des probabilités de succès, nous avons enregistré une amélioration très nette de 53.5% entre les deux phases, sachant que la moyenne des probabilités est passée de 0.538 (dans l'exploration) à 0.826 (dans l'exploitation). Dans l'exploitation, la tendance de la courbe est généralement vers le haut sachant que la probabilité est passée de 0.508 (1ère situation) à 0.882 (dernière situation de la séquence). Cette variation représente une hausse (augmentation) de 73.6 %.

Proba. de succès	P_{min}	P_{max}	P_{moy}	Synthèse
Exploration	0,508	0,556	0,538	
Exploitation	0,508	0,968	0,826	↑ 73.6 %
Synthèse			↑ 53.5 %	

Pour le troisième critère, la moyenne globale du gain relatif est de l'ordre de 67%. Nous soulignons que la différence entre les moyennes du PRT et du PST est de 8,06. Cette différence reflète l'amélioration du niveau global des élèves de 102%. Ce sont de bons résultats qui donnent une idée nette sur le niveau très positif de l'impact du didacticiel sur l'ensemble de ses utilisateurs.

Notes	Min	Max	MOY	GR
PRT	06 / 20	10 / 20	07,87 / 20	
PST	14 / 20	18 / 20	15,93 / 20	
Synthèse			↑ 102 %	67%

En conclusion, l'agent d'apprentissage a permis une amélioration des probabilités de succès de 53.5 % entre les deux phases ; une amélioration des probabilités de succès de 73.6 % entre la 1ère et la dernière situation de la séquence ; une réduction du temps d'étude de 29 % entre les deux phases; et une amélioration du niveau global de 102% et un gain relatif de 67%.

4. Conclusion

Les chiffres enregistrés représentent de bons résultats. Cela est dû à l'intervention de l'agent pédagogique en choisissant les bonnes actions. Les résultats de nos expériences ont montré que l'automatisation des tâches pédagogiques du système tuteur est faisable et assurée. L'automatisation permet de gagner de l'énergie et du temps et de vaincre les problèmes multiples de l'adaptabilité des systèmes tuteurs. Notre approche voulait dépasser les démarches basées sur la logique manuelle pour rédiger l'ensemble des règles du module pédagogique. Le module pédagogique se génèrent dynamiquement en se basant uniquement sur l'interaction de l'élève avec son environnement d'enseignement.

Les efforts investis ont été dans deux directions, d'une part, la conception d'une base de connaissances qui tient en compte les propriétés significatives de l'acte pédagogique, de l'autre, chercher et essayer parmi les techniques de l'apprentissage basées sur les calculs numériques, celles qui sont plus appropriés au champ d'étude telles que les techniques de l'apprentissage par renforcement que nous avons essayées, etc. Nous soulignons que, nous avons utilisé une approche objective dans le but de générer dynamiquement une séquence pédagogique adaptative. L'objectivité se base uniquement sur des interactions naturelles entre le système tuteur avec ses utilisateurs potentiels.

Bibliographie

[Aimeur 2004]. Aimeur E. *L'intelligence artificielle : quel avenir ? L'Autre Forum, le journal des professeurs et des professeurs de l'université de Montréal. Volume 8, Numéro 2, Février 2004. Canada.* http://sgpum.umontreal.ca

[Picard-Papert 2004]. Picard R. W., Papert S., Bender W., Blumberg B., Breazeal C., Cavallo D., Machover T., Resnick M., Roy D. and Strohecker C. *Affective learning - a manifesto. BT Technology Journal • Vol 22 No 4 • October 2004.*

[Sonwalkar 2004]. Sonwalkar N. *Adaptive Learning : A Dynamic Methodology for Effective Online Learning.* http://idlsystems.com/media/brochures/ALS.pdf

[Bennane2003]. Bennane A., D'hondt T. *Tutoring and Adaptability: Case Study. MLMTA'03, p186-191, CSREA Press, USA. 2003, ISBN 1-932415-11-4.*

[Jodogne2003]. Jodogne S. ; *Introduction à l'apprentissage par renforcement;* http://www.montefiore.ulg.ac.be/~jodogne/cours/reinforcement-4.pdf.

[Bennane2001]. Bennane A., Manderick B., D'hondt T. *Generation of Training Situations and Adaptive Systems. ICCE2001, Korea. http://www.icce2001.org/pdf/p09/BE002.pdf*

[Beck2000]. Beck J., Beverly P.W & Beal C.R. *Advison: A machine learning architecture for intelligent tutor construction. AAAI2000.*

[Sutton98]. Sutton R., Barto A. G. *Reinforcement Learning, A Introduction; A Bradford Book, 1998.*

[PRZESMYCKI91]. Przesmycki H, *Pédagogie différenciée; Hachette 1991.*

[VanLehn88]. VanLehn K. 1988. *Student modeling. In Ploson and Richardson. Foundations of Intelligent Tutoring Systems. Lawrence Erlbaum Associates.*

[Wenger87]. Wenger E. (1987). *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge. Los Altos, CA: Morgan Kaufmann Publishers, Inc.*