Controle Inteligente de Tempo Livre em Tutoria Multissessão: Concepção, Implementação e Avaliação Empírica

Weber Martins[1], Viviane Margarida Gomes[2], Lauro Eugênio Guimarães Nalini[3]

[1]Escola de Engenharia Elétrica e de Computação – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

[2]Departamento de Tecnologia da Informação – Instituto Federal de Goiás (IFGO)
Inhumas – GO – Brasil

[3]Departamento de Psicologia – Pontifícia Universidade Católica de Goiás (PUC-Goiás)
Goiânia – GO – Brasil

weber@eee.ufg.br, viviane@inhumas.ifgo.edu.br, legn@ucg.br

Abstract. *This research proposes the intelligent control of free time (break intervals) in multi-session tutoring. The teaching strategy presents the content in modules with video, exercise, practical suggestion, free time, and revision exercise. Based on the student performance in exercises, the proposed system uses Reinforcement Learning to control pause durations. The experimental group (with intelligent control) has been compared to the control group (where decisions belong to the student). Results have shown significant and equivalent gains in knowledge retention.*

Resumo. *Esta pesquisa propõe o controle inteligente de tempo livre (pausas) em tutoria multissessão. A estratégia de ensino apresenta o conteúdo em módulos com vídeo-aula, exercício, sugestão prática, tempo livre e exercício de revisão. Baseado no desempenho do aluno nos exercícios, o sistema proposto utiliza Aprendizagem por Reforço para controlar a duração das pausas. O grupo experimental (com controle inteligente do tempo livre) foi comparado ao grupo controle (onde a decisão pertence ao próprio estudante). Resultados mostram ganhos significativos e equivalentes na retenção de conhecimento.*

1. Introdução

Sistemas Tutores Inteligentes (STI) são programas computacionais concebidos para prover instrução personalizada. A partir da interação homem-máquina, métodos de Inteligência Computacional adaptam dinamicamente o conteúdo e/ou o estilo de ensino às características do aprendiz [Murray 1999] [Martins et al. 2004].

Historicamente, os STI expõem o material em etapas consecutivas de tutoria, sem pausas [Martins et al. 2007]. Conforme ponto de vista comportamental em Psicologia, a pausa influencia o tempo dedicado ao estudo (sessão de tutoria), sendo a condição de liberdade apreciada pelo estudante [Osborne 1969] [Geiger 1996]. Portanto esta pesquisa introduz e busca controlar inteligentemente o tempo livre entre etapas da tutoria.

O presente artigo apresenta a fundamentação teórica, seguida pela descrição do sistema proposto, o trabalho de experimentação e os resultados obtidos e analisados, encerrando com a seção de conclusão.

2. Fundamentação Teórica

2.1. Aprendizagem por Reforço

Segundo [Kaelbling et al. 1996], Aprendizagem por Reforço é uma técnica adequada ao problema enfrentado por agentes que devem aprender em ambientes dinâmicos através de interações de tentativa e erro. A estrutura (do Inglês, *framework*) da Aprendizagem por Reforço fundamenta-se na relação entre estado-ação-recompensa. Estado, denotado por s, é o contexto de interação entre ambiente e agente a cada momento, onde o agente executa uma ação (a) e, por consequência, recebe uma recompensa (r) do ambiente.

Após explorar as ações, o agente aprende a selecionar boas ações (ou pares estado-ação) para atingir seu objetivo: maximizar o valor total das recompensas recebidas. A cada passo, o agente escolhe alguma ação a partir da política, denotada por π_t, ou seja, pela regra estocástica para seleção de ações como uma função de estados [Sutton and Barto 1998]. Métodos de seleção de ação definem $\pi_t(s, a)$ baseados no valor da ação $Q_t(s, a)$, tais como *Greedy*, ϵ-*Greedy* e *Softmax*.

O método *Softmax* determina a política como uma distribuição ordenada de probabilidades a partir das estimativas de ação-valor, como pela Distribuição de Gibbs (ver Equação 1), onde o parâmetro τ, temperatura, é responsável pela diferenciação das ações. Temperaturas altas tornam as ações equiprováveis, incentivando a exploração extensiva das ações (*exploration*). À medida que a temperatura diminui, as probabilidades das ações diferenciam-se até convergir (exploração intensiva ou *exploitation*).

$$\pi_t(a) = \frac{e^{\frac{Q_t(a)}{\tau}}}{\sum_{b=1}^{n} e^{\frac{Q_t(b)}{\tau}}} \tag{1}$$

No paradigma da Aprendizagem por Reforço, a caracterização do problema contribui diretamente na escolha do método de solução: Programação Dinâmica, Monte Carlo e Aprendizagem por Diferença Temporal. A modelagem revela o nível de conhecimento, completo ou incompleto, do ambiente dinâmico. Métodos de Programação Dinâmica exigem completo e acurado modelo do ambiente. Ao contrário, métodos de Monte Carlo e de Aprendizagem por Diferença Temporal aprendem da experiência, mesmo em ambiente totalmente desconhecido [Sutton and Barto 1998].

2.2. Análise Experimental do Comportamento

A Análise Experimental do Comportamento (AEC) é uma disciplina científica surgida em 1938, no contexto da Psicologia como ciência natural, com o objetivo geral de descrever e explicar as interações entre o organismo/indivíduo e o ambiente [Catania 1999]. Nos casos de comportamentos emitidos espontaneamente, as respostas agem ou "operam" sobre o ambiente e sofrem as consequências da ação. Em AEC, esse comportamento recebe o nome de comportamento operante [Skinner 2003].

Premack verificou que atividades com elevada frequência podem funcionar como reforçadores para atividades de baixa frequência. Tal constatação foi formulada como o princípio de Premack [Premack 1959], utilizado como fundamento para o sistema proposto desta pesquisa. O tempo livre, quando o aprendiz podia ouvir música ou jogar (atividades comumente mais frequentes), foi contingenciado como consequência às atividades de estudo no tutor (atividades comumente menos frequentes).

3. Sistema Proposto

O sistema proposto controla a duração das pausas baseado no desempenho do aluno. No contexto da tutoria, os estados referem-se aos módulos do curso, as ações são as escolhas do controle (conjunto de ações A, $\{diminuir, manter, aumentar\}$ a duração da pausa) e as recompensas resultam das notas ($N1$ e $N2$) nos exercícios.

Em cada estado s, o agente 1) seleciona uma ação a a partir da política $\pi_t(s, a)$, 2) determina r pela função-recompensa e 3) atualiza o valor da ação $Q(s, a)$, de forma incremental, conforme mostra o algoritmo da Figura 1 (onde o parâmetro α é igual a $\frac{1}{k}$, para k-ésima recompensa recebida para ação a). A política π fornece, pelo método *Softmax*, as probabilidades de escolha de cada ação (ver a Equação 1).

Inicialize $Q(s,a) \leftarrow (r_{min} + r_{max})/2$
Repita (para cada episódio):
  Inicialize s
  Repita (para cada passo do episódio):
  ❶ Escolha a de s usando política derivada de Q (Softmax)
  ❷ Tomada ação a, observe r, s'
  ❸ $Q(s,a) \leftarrow Q(s,a) + \alpha\,[r - Q(s,a)]$
    $s \leftarrow s'$;
  Até que s seja terminal

Figura 1. Algoritmo do sistema proposto.

A modelagem do problema da Aprendizagem por Reforço pressupõe perfis distintos de aluno: teórico, equilibrado e pragmático. Em simulações para ajuste da temperatura, o desempenho do aluno teórico melhora quando o tempo livre diminui, do aluno equilibrado, quando tempo livre mantém-se, e do pragmático, quando aumenta.

As ações recebem recompensas do ambiente, definidas pela função-recompensa $r = f(N1, N2)$, sendo $N1$ e $N2$ medidas do desempenho do aluno, de 0 a 1. Como os alunos só prosseguem após acertar o exercício, os erros consistem em perdas no valor máximo da nota, $Nota = 1 - (\sum_{i=1}^{ta} p_i)/p_R$, onde p_i e p_R são as perdas por tentativa e de referência e ta é a tentativa do acerto. Calcula-se a perda pela equação $p_i = H(x).(T - i + 1)^{-1}$, sendo *T* o total de alternativas (tentativas possíveis) e $H(x)$ a função que caracteriza cada tipo de resposta (Correta, Semelhante à Correta, Incorreta).

4. Experimento

Para testar o sistema proposto, o STI apresenta o conteúdo em sessões, como mostra a Figura 2, com duração do tempo livre controlada durante o curso. Os estados, $s_1, s_2, s_3, ..., s_{15}$, são caracterizados pelos módulos do curso "Introdução à Microinformática no Linux". Quanto às ações, o agente pode manter ou variar a duração da pausa em 25% (a menos ou a mais) do valor anterior. Portanto, numericamente, o conjunto de ações é: $A = \{0,75; 1; 1,25\}$. A função-recompensa é a média ponderada das notas $N1$ e $N2$ nos exercícios, conforme mostra a equação $r_t = (N1_t + 2N2_t)/3$.

Para cálculo de $N1$ e $N2$, a função $H(x)$ (*x = tipo de resposta*) foi definida, empiricamente, como *H(x)=0* para *x=Correta*, *H(x)=0,5* para *x=Semelhante à Correta*

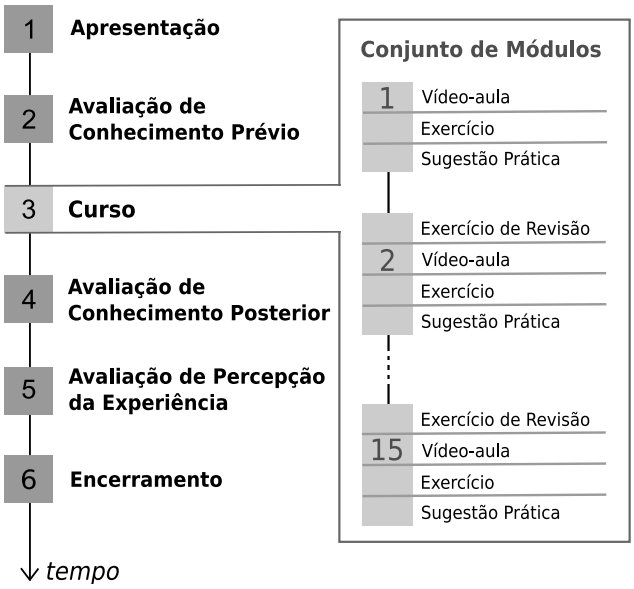

Figura 2. Etapas da tutoria.

e *H(x)=2* para *x=Incorreta*. As simulações ajustaram os valores inicial, intermediário e final da temperatura em 2, em $0,2$ (no Estado 8) e em $0,1$, respectivamente.

O STI, feito em Java, foi utilizado no Ubuntu/Linux. Dois grupos de 32 estudantes (entre 14 e 17 anos e iniciantes em Microinformática) compuseram a amostra, sendo o controle da pausa, inteligente no grupo experimental e no livre grupo controle.

5. Resultados

Na amostra geral e nos grupos, a nota final apresentou valores maiores que a inicial, com ganho na retenção de conhecimento igual a 45,2%, como mostra a Tabela 1. Sobre a percepção da experiência, as estatísticas sobre a etapa mais interessante para o grupo experimental e controle são, respectivamente: vídeo-aula - $53,1\%$ e $50,0\%$, exercício - $6,3\%$ e $37,5\%$, sugestão prática - $15,6\%$ e $3,1\%$, tempo livre - $9,4\%$ e $3,1\%$, e exercício de revisão - $15,6\%$ e $6,3\%$. A maioria dos alunos considerou o tempo livre suficiente (65,6%, no grupo experimental, e 78%, no grupo controle).

A análise inferencial aponta ganhos significativos na retenção de conhecimento, conforme Teste t-*Student* pareado das notas inicial e final dos participantes. A hipótese nula (H_0) foi rejeitada ($\alpha_{efetivo} < 0,01$) na amostra geral e nos grupos. Para analisar o desempenho entre os grupos, aplicou-se o Teste t com amostras independentes. Com

Tabela 1. Estatística descritiva do desempenho.

Estatísticas	Média			Desvio Padrão		
Medida \ Amostra	Inteligente	Livre	Geral	Inteligente	Livre	Geral
Nota Inicial	4,13	3,88	4,00	1,55	1,52	1,53
Nota Final	6,71	6,56	6,64	1,95	1,75	1,84
Ganho Normalizado	45,2%	45,2%	45,2%	29,4%	23,4%	26,4%

$\alpha_{efetivo} > 0,05$, a prova de hipótese afirma equivalência entre os grupos, H_0 verdadeira.

No grupo experimental, os estudantes perceberam melhor o tempo livre como componente da estratégia de ensino. Quanto à técnica de Aprendizagem por Reforço, a instabilidade do ambiente (interação do aluno) penalizou a aprendizagem computacional.

6. Conclusão

Esta pesquisa introduz a inserção de pausas em STI, ressaltando seu forte potencial como componente na estratégia de ensino. Em trabalhos futuros, sugere-se a investigação sobre a eficácia do tempo livre na retenção de conhecimento, comparando STIs com e sem pausas. Para o controle inteligente da pausa, a Aprendizagem por Reforço poderia ser modelada com base em comportamento não-verbal, como ansiedade física e expressões faciais [Rodrigues and Carvalho 2005] [Sarrafzadeh et al. 2008].

Referências

Catania, A. C. (1999). *Aprendizagem: Comportamento, Linguagem e Cognição*. Ed. Artes Médicas Sul, Porto Alegre, 4a edição.

Geiger, B. (1996). A time to learn, a time to play: Premack's principle applied in the classroom. *American Secondary Education*, 25:2–6.

Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285.

Martins, W., Afonseca, U. R., Nalini, L. E. G., and Gomes, V. M. (2007). Tutoriais inteligentes baseados em aprendizado por reforço: Concepção, implementação e avaliação empírica. *Anais do XVIII Simpósio Brasileiro de Informática na Educação*, páginas 617–626.

Martins, W., Meireles, V., de Melo, F. R., and Nalini, L. E. G. (2004). A novel hybrid intelligent tutoring system and its use of psychological profiles and learning styles. *Lecture Notes on Computer Science*, 3220:830–832.

Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education*, 10:98–129.

Osborne, J. G. (1969). Free-time as a reinforcer in the management of classroom behavior. *Journal of Applied Behavior Analysis*, 2(2):113–118.

Premack, D. (1959). Toward empirical behavior laws: I. positive reinforcement. *Psychological Review*, 66:219–233.

Rodrigues, L. M. L. and Carvalho, M. (2005). STI-I: Sistemas tutoriais inteligentes que integram cognição, emoção e motivação. *Revista Brasileira de Informática na Educação*, 13(1):20–34.

Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C., and Bigdeli, A. (2008). "How do you know that I don't understand?" A look at the future of intelligent tutoring systems. *Computers in Human Behavior*, 24:1342–1363.

Skinner, B. F. (2003). *Ciência e Comportamento Humano*. Martins Fontes, São Paulo, 11a edição.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning*. Bradford Book/MIT Press, Cambridge, Massachusetts and London, England.