

Caracterização de Alunos em Ambientes de Ensino Online: Estendendo o Uso da DAMICORE para Minerar Dados Educaçãoais.

Luis F. de S. Moro, Carla L. Rodrigues, Fernando R. H. Andrade, Alexandre C. B. Delbem, Seiji Isotani

Instituto de Ciências Matemáticas e de Computação – USP Avenida Trabalhador são-carlense, 400 – Centro - CEP: 13566-590 - São Carlos - SP

{luismoro, carlalrodriguez, fernando.heb, acbd, sisotani}@icmc.usp.br

Abstract. *With the popularization of the use of technological resources in education, a bigger amount of data related to the interactions between students and these resources has been stored. Analyzing this data in order to characterize these students is a very important task. In this work, partial results of a master degree's research, from educational data mining techniques are presented, identifying relevant behaviors of students who interacted with an educational web system. These results indicate that these techniques can help teachers to preview the performance of their students to facilitate a future development of a tool that assists this task for teachers.*

Resumo. *Com a popularização do uso de recursos tecnológicos na educação, uma significativa quantidade de dados com relação às interações entre alunos e esses recursos é armazenada. A análise destes dados, visando caracterizar tais alunos é tarefa muito importante. Nesse trabalho, são apresentados resultados parciais de uma pesquisa de mestrado que identificou, por meio de técnicas de mineração de dados educaçãoais, padrões relevantes de comportamentos de alunos que interagiram com um sistema educacional web. Os resultados apontam que essas técnicas podem auxiliar o professor a prever o desempenho dos seus alunos facilitando futuramente o desenvolvimento de uma ferramenta que facilite essa tarefa para os professores.*

1. Introdução

Atualmente é possível observar uma crescente inclusão de recursos tecnológicos na educação, que vão desde simples softwares educativos até os mais sofisticados Ambientes Virtuais de Aprendizagem (AVA), como, por exemplo, os Sistemas Tutores Inteligentes (ITS, do termo em inglês: *Intelligent Tutoring Systems*), que visam auxiliar professores e alunos nas tarefas de ensinar e aprender dentro e fora das salas de aula (Gotardo, Cereda e Junior, 2013).

Conseqüentemente, essa popularização faz com que uma enorme quantidade de dados seja armazenada, permitindo a realização de análises desses dados em busca de importantes informações e conhecimento sobre os alunos, como por exemplo, o motivo que leva determinados alunos a abandonarem um curso (Martins, Lopes e Raabe, 2012), a predição de um mau desempenho em uma determinada disciplina (Gotardo, Kaestner,

Noronha, 2012), qual o estilo cognitivo de um determinado aluno (Flores et al., 2013), entre outros. No entanto, essas análises não são triviais. De modo que para serem extraídas informações relevantes de um conjunto de dados armazenados, são necessários o desenvolvimento e a aplicação de ferramentas e técnicas específicas que auxiliam o processo de Mineração de Dados (MD).

Durante os últimos anos, os sistemas educacionais *web* vêm ganhando mais espaço na educação formal, o que gera um grande volume de dados relacionados às interações dos alunos com esses sistemas. No entanto, esse volume de dados cria novas dificuldades para os professores quando esses tentam acompanhar o progresso dos alunos ou detectar suas necessidades na realização das atividades propostas. Como resultado, a busca de informações sobre o comportamento e desempenho dos alunos é atualmente um campo interessante e crescente de investigação. Porém, é importante que esses comportamentos sejam identificados o quanto antes, para que alunos com dificuldades na aprendizagem de determinado conteúdo possam ser auxiliados e alunos que possuem facilidade na aprendizagem de determinado conteúdo possam receber novos desafios.

Para auxiliar os professores a identificar comportamentos dos alunos e tomar decisões de ensino que efetivamente os auxiliem no processo de aprendizagem, técnicas de Mineração de Dados Educacionais (EDM, do termo em inglês, *Educational Data Mining*) vêm sendo utilizadas (Kampff, 2009; Prata et al., 2009; Baker, 2010). Nesse sentido, a pesquisa aqui apresentada teve como objetivo analisar os dados provenientes do sistema educacional *web*, que disponibiliza exercícios específicos de matemática, o MathTutor. Essa análise foi realizada por meio de um processo de EDM que possibilitou o reconhecimento de padrões do comportamento dos alunos e gerou resultados que podem auxiliar os professores no processo de ensino e aprendizagem.

Além desta seção introdutória, este trabalho está subdividido em seis seções. A seção 2 descreve os conceitos, técnicas e ferramentas utilizados de acordo com a metodologia adotada no decorrer da pesquisa, apresentada na seção 3. As seções 4 e 5, respectivamente, descrevem o experimento que gerou os dados utilizados na pesquisa e os resultados alcançados com a aplicação das técnicas de EDM. Por fim, a seção 6 traz as conclusões sobre tais resultados, bem como os trabalhos futuros a serem realizados para complementar essa pesquisa.

2. Conceitos Básicos e Trabalhos Relacionados

Atualmente, com a popularização e os avanços das tecnologias de informação e comunicação, quantidades enormes de dados são geradas pelos mais diversos sistemas computacionais. Um bom exemplo disso é a Wikipédia, que desde julho de 2001 até janeiro de 2013 teve 24.778 artigos inclusos em seu banco de dados (Wikipédia, 2013).

Para explorar esses dados de uma forma completa e sem a necessidade de retrabalhos, alguns passos devem ser seguidos. Esse conjunto de passos é denominado “Descobrimto de Conhecimento em Bases de Dados”, (do inglês *Knowledge Discovery in Data Bases - KDD*) e possui basicamente cinco fases, a “seleção”, o “pré-processamento”, a “transformação”, a “mineração de dados” e a “interpretação e avaliação dos dados” (Fayyad et al., 1996). A partir da aplicação do KDD em um conjunto de dados brutos selecionados no início do processo, é possível gerar

determinado conhecimento sobre esses dados. A aplicação desse processo em dados educacionais é atualmente conhecida como EDM.

De acordo com Baker et al. (2010), a EDM possui diferentes técnicas, tais como “predição”, “agrupamento”, “mineração de relações”, “destilação de dados para facilitar decisões humanas” e “descoberta com modelos”, as quais, atualmente, são amplamente estudadas e utilizadas por diversos pesquisadores, como Baker e Yacef (2009), Costa et al. (2012), Kampff (2009), Romero et al. (2008), entre outros.

De forma geral, as diversas técnicas de EDM são utilizadas para detectar e prever os comportamentos dos alunos frente à determinada tarefa/desafio ou situação educativa. Os comportamentos detectados podem indicar ao professor tendências e padrões que possam auxiliá-lo na compreensão das potencialidades e limites de seus alunos e, dessa forma, melhorar o processo de ensino e aprendizagem dentro e fora da sala de aula (Baker, Isotani e Carvalho, 2011).

Para desenvolver a fase de MD do trabalho aqui apresentado foram utilizadas técnicas de “agrupamento” que, como definido por Costa et al. (2012, p. 10,11), têm como objetivo “dividir o conjunto de dados em grupos, de forma que os objetos contidos nos dados fiquem agrupados naturalmente de acordo com a semelhança entre eles”. Aplicando essa definição a esta pesquisa, os alunos podem ser classificados de acordo com as características que possuem em comum.

Existem diversas pesquisas da área que investigaram os alunos e seus comportamentos. Kumar e Vijayalakshmi (2011) apresentaram resultados que mostram como árvores de decisão podem ser utilizadas para prever o desempenho acadêmico dos alunos. Nessa pesquisa foi aplicado o algoritmo de árvore de decisão C4.5 e assim obtiveram-se resultados capazes de prever alunos suscetíveis a falharem na disciplina. Para isso, foram utilizadas as notas já obtidas pelos alunos em avaliações anteriores e algumas outras características, como o sexo do aluno, sua situação financeira, etc. Em outra pesquisa Kumar e Vijayalakshmi (2012), visaram prever o desempenho dos alunos com base na pontuação obtida por eles em semestres anteriores. Foram avaliadas as notas dos alunos nos últimos quatro semestres e, baseados nesses dados e nos algoritmos *Decision Table* e *One R*, obtiveram-se resultados relevantes capazes de prever alunos que iriam reprovar em determinada disciplina. Em Quadri e Kalyanka (2010), objetivou-se, por meio da utilização de técnicas de árvores de decisão, descobrir quais alunos irão abandonar a faculdade, A característica identificada e considerada mais relevante nesse estudo foi a renda familiar do aluno. Além disso, outros dois critérios de avaliação foram as notas em semestres anteriores e a condição de o aluno ser ou não o primeiro filho da família.

Contribuindo para ampliar o conhecimento sobre o assunto, este trabalho visou prever o comportamento de alunos que utilizaram o software educacional *web MathTutor* no domínio de números decimais, com o diferencial de, no futuro, gerar uma ferramenta que possa ser utilizada por pessoas leigas em MD. Na próxima seção será apresentada a metodologia utilizada nessa pesquisa.

3. Metodologia

Nessa pesquisa, buscou-se por meio da utilização do KDD e de técnicas de EDM aplicadas aos dados resultantes de um experimento (Seção 4) no qual alunos interagiram

com um sistema educacional *web*, gerar resultados que auxiliem o professor a identificar o comportamento dos seus alunos frente aos conteúdos propostos.

Como pode ser visto na Figura 1, a primeira fase foi a “seleção”, na qual os dados que serão utilizados nas fases seguintes do KDD são escolhidos. Com os dados selecionados, foi realizada a fase de “pré-processamento”, na qual ocorreu a refinação dos dados. Posteriormente a isso, se fez a “mineração dos dados”, fase em que foram encontrados padrões dentro dos dados já trabalhados. Por fim esses dados foram avaliados em busca de algum conhecimento sobre os alunos que participaram do experimento.

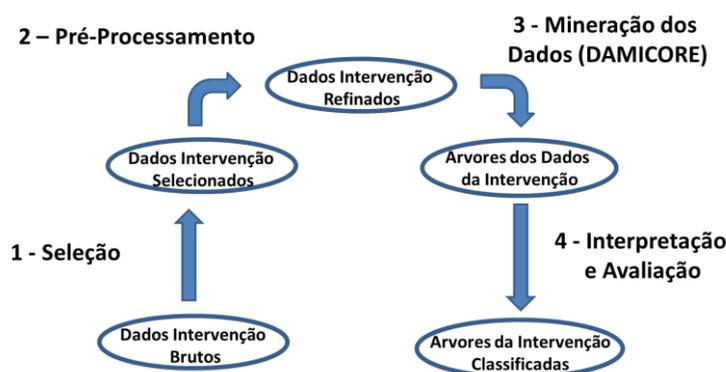


Figura 1 - Fases do KDD personalizadas.

Os dados utilizados na pesquisa são oriundos de um experimento realizado por McLaren et al. (2012), que envolveu três escolas dos Estados Unidos: *Hopewell*, *Huston* e *Steel Valey*. Esse experimento visava entender como os alunos se comportam ao aprender o assunto “números decimais”, do domínio de matemática, resolvendo exercícios com o apoio do sistema educacional *web* MathTutor. Na seção 4 são detalhadas as etapas desse experimento.

4. O Experimento

A realização do experimento consistiu em seis etapas: “pré-teste”, “questionário demográfico e de experiência matemática” (Q1), os problemas da “intervenção”, “questionário sobre a experiência do aluno com os materiais” (Q2), um “pós-teste imediato” e um “teste de retenção” (McLaren et al., 2012), todas elas realizadas com os alunos que interagiram com o sistema MathTutor.

Resumidamente, na etapa “pré-teste”, buscou-se classificar os alunos de acordo com o seu conhecimento prévio, de forma que não houve intervenção do professor na aplicação do teste. Na etapa “Q1”, os alunos foram indagados sobre suas características como “idade”, “sexo”, “conhecimento em matemática”, “conhecimento em números decimais”, “conhecimentos em informática”, etc. Na etapa “intervenção”, foi apresentado aos alunos um conjunto de exercícios guiados pelo sistema MathTutor. Na “Q2”, buscou-se por informações sobre a experiência do aluno com o sistema e com o conteúdo que foi apresentado. A etapa “pós-teste imediato” teve como objetivo descobrir quanto os alunos aprenderam com a “intervenção”. A etapa “teste de retenção” teve como objetivo verificar a quantidade de informações que os alunos conseguiram assimilar num longo período, após a “intervenção”.

Esse experimento gerou uma planilha de dados com as informações sobre as interações dos alunos com o MathTutor (Planilha 1). Essa planilha possui diversos campos que armazenam cada uma das interações dos alunos com o sistema.

Utilizando os dados da Planilha 1, uma segunda planilha foi gerada (Planilha 2). Nessa nova planilha foi armazenado o desempenho de cada um dos alunos nos testes (“pré-teste”, “pós-teste imediato” e “teste de retenção”). Os dados armazenados em ambas as planilhas (Planilha 1 e Planilha 2) foram utilizados no desenvolvimento da pesquisa, conforme será apresentado na próxima seção.

5. Discussão

Conforme dito anteriormente, a pesquisa aqui apresentada teve como objetivo analisar os dados provenientes de um sistema educacional *web*, o MathTutor, para identificar padrões de comportamento e gerar resultados que possam auxiliar os professores no processo de ensino e aprendizagem. Para alcançar o objetivo proposto, utilizou-se de 4 fases do KDD sugeridas por Fayyad et al. (1996): “seleção”, “pré-processamento dos dados”, “mineração de dados” e “interpretação e avaliação” (Figura 1).

Na fase de “seleção”, foram escolhidos os atributos a serem utilizados para minerar os dados. Nessa fase optou-se por trabalhar com os dados referentes à etapa de “intervenção”, uma vez que são dados coletados no início do experimento, o que oferece ao professor mais tempo para interagir com os alunos e auxiliar na melhoria do seu rendimento. Além disso, foram selecionados os dados para interpretar e avaliar os resultados da fase de MD. Para esse contexto, foram utilizados dados da “intervenção”, do “pré-teste” e do “pós-teste”.

Após a fase de “seleção” foi realizado o “pré-processamento dos dados”, fase na qual a planilha com as informações sobre as interações dos alunos (Planilha 1) foi formatada. Com isso, os dados estavam prontos para serem inseridos na ferramenta de MD DAMICORE (Fase de MD), uma vez que não é necessário que a fase de transformação seja realizada, pois a DAMICORE não exige um tipo específico de arquivo para funcionar.

Na fase de MD, a DAMICORE foi aplicada aos dados resultantes das fases anteriores. A DAMICORE trabalha com a seguinte combinação de algoritmos: *Normalized Compression Distance* (NCD), da Teoria da Informação (Cilibrasi e Vitanyi, 2005); *Neighbor Joining* (NJ), da filogenética (Felsenstein, 2003) e *FastNewman* (FN), de Redes Complexas (McLaren et al., 2008). Assim, foi aplicada a técnica NCD para o cálculo da distância, para o agrupamento e para a compressão dos dados analisados. Na sequência a técnica NJ construiu uma rede filogenética e, por fim, a técnica FN identificou e formou os *clusters* dos dados de entrada (Sanches, Cardoso e Delbem, 2011).

Como resultado da fase de MD foram geradas “árvores” que agrupam os alunos de acordo com as suas características (Figura 2). A estrutura de cada “árvore” gerada é composta por “folhas” que representam cada um dos alunos identificados pelo seu código, e “ramos” (conjunto de folhas) que representam os agrupamentos dos alunos de acordo com o desempenho apresentado por eles na fase da “intervenção”.

Para realizar a última fase do KDD, fase de “interpretação e avaliação”, foi necessário analisar as “árvores” geradas pela DAMICORE, visando encontrar “ramos” que trouxessem informações relevantes sobre os alunos.

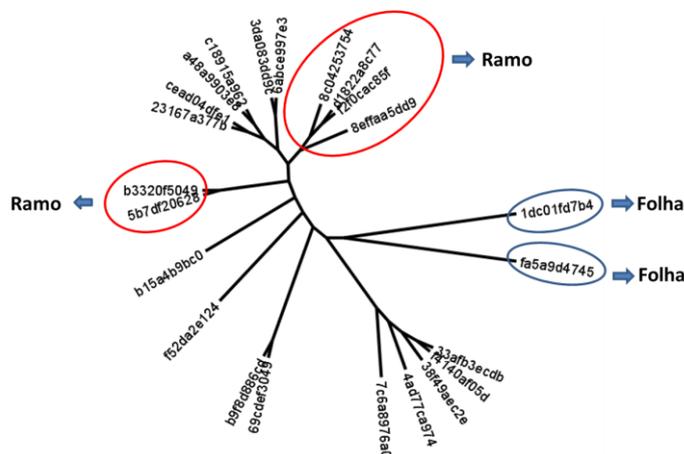


Figura 2 – “Árvore” gerada pela DAMICORE.

Assim, para analisar as “árvores”, recorreu-se às classificações dos alunos no “pós-teste”, contidas na Planilha 2. As “árvores” geradas na fase de MD (como a da Figura 2) foram coloridas de acordo com o desempenho dos alunos no “pós-teste”. As cores demonstram a classificação dos alunos de acordo com o desempenho obtido. Na Figura 3 pode-se observar a métrica definida para classificar os alunos bem como uma legenda na qual são descritas as cores que definem cada uma das faixas calculadas.

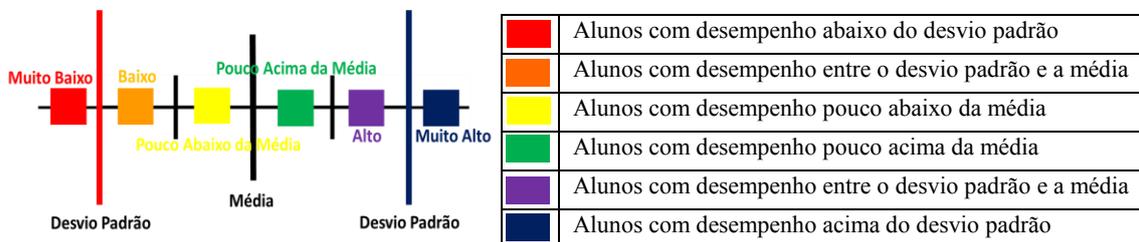


Figura 3 - Intervalos e cores utilizadas para classificação das “árvores”.

A “árvore” colorida resultante retrata o desempenho dos alunos na etapa da “intervenção” e as cores retratam o desempenho do aluno no “pós-teste” (Figura 4, árvore da esquerda). Portanto, se for possível verificar agrupamentos de cores semelhantes em alguns pontos desta “árvore”, pode-se afirmar que a DAMICORE conseguiu prever determinados comportamentos no “pós-teste” analisando os dados da “intervenção”.

Na Figura 4, a “árvore” da esquerda apresenta dois agrupamentos de alunos de acordo com o resultado que alcançaram no “pós-teste”, na qual o agrupamento 1 representa os alunos que obtiveram um alto desempenho e o agrupamento 2 os que obtiveram baixo desempenho. A “árvore” da direita exhibe dois agrupamentos de alunos de acordo com o tempo que dedicaram interagindo com o sistema, sendo os alunos do ramo 1 aqueles que interagiram por menos tempo e os alunos do ramo 2 aqueles que interagiram por mais tempo.

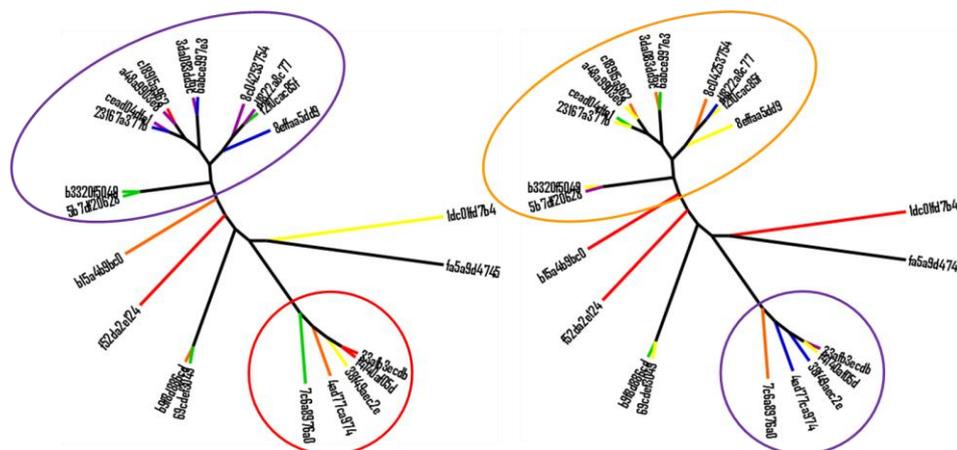


Figura 4 - Comparação entre “árvores” classificadas.

A fim de entender como a ferramenta de mineração de dados DAMICORE realizou esses agrupamentos de alunos, os atributos selecionados durante a fase de “seleção” foram utilizados para colorir as “árvores”, tornando possível analisar cada uma das “árvores” e compará-las com a “árvore” colorida do “pós-teste”. Dessa forma, foi possível verificar se cada um dos atributos analisados influenciou no desempenho dos alunos e de que forma isso ocorreu. Um exemplo pode ser analisado na mesma Figura 4, em que, na “árvore” da esquerda temos os alunos agrupados de acordo com o desempenho no “pós-teste” e, na “árvore” da direita, os alunos agrupados de acordo com o tempo que dedicaram à interação com o sistema. Pode-se notar que os alunos que obtiveram um desempenho acima da média (Figura 4, “árvore” esquerda, “ramo” 1) interagiram por menos tempo com o sistema (Figura 4, “árvore” direita, “ramo” 1) e os alunos com desempenho abaixo da média (Figura 4, “árvore” esquerda, “ramo” 2) interagiram por mais tempo com o sistema (Figura 4, “árvore” direita, “ramo” 2).

Após classificar e analisar todas as “árvores”, determinados padrões de comportamento apresentados pelos alunos foram identificados e, então, foram elaboradas as seguintes hipóteses:

- **Ótimo Desempenho (C1)** → Alunos que obtiveram bom desempenho no “pós-teste” apresentaram poucas interações e baixo tempo na “intervenção”, e poucos *missconceptions* no “pré-teste”.
- **Péssimo Desempenho (C2)** → Alunos que obtiveram mau desempenho no “pós-teste” apresentaram muitas interações e alto tempo na “intervenção”, e muitos *missconceptions* no “pré-teste”.
- **Gênero Dependente (C3)** → Alunos que obtiveram bom desempenho no “pós-teste” são do sexo feminino ou alunos que obtiveram mau desempenho são do sexo masculino.
- **Bom Desempenho (C4)** → Alunos que obtiveram bom desempenho no “pós-teste” apresentaram dois dos três seguintes comportamentos: poucas interações na “intervenção”, baixo tempo na “intervenção” e poucos *missconceptions* no “pré-teste”.
- **Mau Desempenho (C5)** → Alunos que obtiveram mau desempenho no “pós-teste” apresentaram dois dos três seguintes comportamentos: muitas interações

na “intervenção”, alto tempo na “intervenção” e muitos *missconceptions* no “pré-teste”.

As hipóteses elaboradas foram analisadas, verificando seus percentuais de acerto quando um determinado comportamento ocorreu no “pós-teste”. Por exemplo, para a hipótese C4, na totalidade das vezes em que os alunos possuíam um bom desempenho no “pós-teste”, foram registradas duas das três seguintes características: poucas interações e baixo tempo na “intervenção” ou poucos *missconceptions* no “pré-teste”. Essa relação foi analisada para as outras hipóteses e os resultados obtidos foram os seguintes: C1 = 61% de ocorrência, C2 = 44% de ocorrência, C3 = 41% de ocorrência, C4 = 100% de ocorrência e C5 = 67% de ocorrência.

Diante dos resultados obtidos é possível afirmar que, no caso do experimento estudado, a aplicação das técnicas da EDM e a utilização da ferramenta DAMICORE auxiliaram na obtenção de dados relevantes sobre o comportamento dos alunos. Além disso, é possível supor que a análise desses comportamentos pode favorecer a tomada de decisão do professor e trazer melhorias para o processo de ensino e aprendizagem, uma vez que antecipam situações nas quais os alunos se deparam com maiores dificuldades quanto à compreensão de determinado conteúdo ou têm necessidade de maiores desafios.

6. Conclusão e Trabalhos Futuros

O objetivo desse trabalho é apresentar os resultados parciais de uma pesquisa de mestrado que identificou, por meio de técnicas de EDM, padrões relevantes de comportamentos de alunos que interagiram com o sistema educacional *web*, que disponibiliza exercícios específicos de matemática, o MathTutor. Essa pesquisa, se enquadra em um projeto maior que visa encontrar soluções para desenvolver uma ferramenta de simples utilização para que leigos em MD possam obter conhecimento de dados educacionais. Os resultados obtidos e aqui apresentados foram satisfatórios e mostram que é possível identificar comportamentos dos alunos ao analisar dados das interações dos alunos com um sistema educacional *web*.

Apesar dos resultados positivos alcançados, ainda é necessário validar as hipóteses com um grupo maior de alunos, bem como verificar se os resultados alcançados conseguem realmente auxiliar os professores (usuários leigos aos conceitos de MD) a identificarem o comportamento de seus alunos de forma mais efetiva. Para conseguir essas validações, novos experimentos, com outros grupos de alunos e em outros contextos educativos, necessitam ser realizados. Dessa forma, está previsto como trabalho futuro o desenvolvimento de um projeto de pesquisa no qual o mesmo experimento será realizado em outro contexto educativo. O objetivo desse novo projeto será o de acompanhar o desempenho dos alunos e professores envolvidos para comparar se os comportamentos manifestados pelos alunos nessa pesquisa se repetem, validando com maior precisão, os resultados encontrados.

É previsto, por exemplo, que nesse novo experimento os professores participantes recebam as informações referentes aos comportamentos dos alunos a cada etapa, visando auxiliá-lo a identificar as necessidades manifestadas pelos alunos. Com essas informações à disposição, o professor poderá, durante o processo de ensino e

aprendizagem auxiliar aqueles alunos com maiores dificuldades e gerar novos desafios para aqueles com maiores facilidades no tema.

Em relação aos recursos de apoio ao professor, considera-se interessante o desenvolvimento de uma ferramenta que apresente a ele, de forma rápida e objetiva, resultados que o auxiliem a interagir com a sua turma. Para isso, deverão ser implementadas na ferramenta de EDM, técnicas de Interação Humano-Computador que aprimorem a forma de apresentação das funcionalidades para o professor tirar o máximo proveito de sua utilização.

Agradecimentos

Agradecimentos à CAPES e ao CNPQ pelo apoio financeiro. ICMC e Carnegie Mellon.

Referências

- Baker, R. S. J. D., Isotani, S. e Carvalho, A. M. J. B. Mineração de dados educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, v. 19(2), 2011.
- Baker, R. S. J. D. e Yacef, K. (2009) “The State of Educational Data Mining in 2009: A Review and Future Visions”, in *Journal of Educational Data Mining*, v. 1(1), p. 3-17.
- Baker, R.S.J.D., McGaw, B., Peterson, P. e Baker, E. (2010) “Data Mining For Education”, in *International Encyclopedia of Education* (3rd edition), B. McGaw, Peterson, P. e Baker Ed. Elsevier, Oxford, UK.
- Baker, R.S.J.d., Nkmabou, R., Mizoguchi, R., e Bourdeau, J. (2010) “Mining Data for Student Models”, in *Advances in Intelligent Tutoring Systems*, p. 323-338.
- Cetintas, S., Si, L., Xin, Y. P. e Tzur, R. (2013) “Probabilistic Latent Class Models for Predicting Student Performance” in *Conference on Information & Knowledge Management*, p. 1513-1516.
- Costa, E., Baker R. S. J. d., Amorim L., Magalhães J. e Marinho T. (2012) “Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações” em *Jornada de Atualização em Informática na Educação*, v. 1, p. 1-29.
- Fayyad, U. M., Piatetsky-Shapiro, G. e Smyth, P. (1996) *From Data Mining to Knowledge Discovery: An Overview*. *Advances in Knowledge Discovery and Data Mining*, p. 1-34.
- Flores, M. S., Frozza, R., Bagatini, D. e Cruz M. E. J. K. (2013) “Uso de Rede Neural Artificial para Determinar o Estilo Cognitivo do Aprendiz” em *Congresso Brasileiro de Informática na Educação*, p. 717-726.
- Gotardo, E., Kaestner, C. e Noronha, R. V. (2012) “Previsão de Desempenho de Estudantes em Cursos EAD Utilizando Mineração de Dados: uma Estratégia Baseada em Séries Temporais” em *Anais do Simpósio Brasileiro de Informática na Educação*.
- Gotardo, R., Cereda, P. R. M. e Junior, E. R. H. “Predição do Desempenho do Aluno usando Sistemas de Recomendação e Acoplamento de Classificadores” em *Congresso Brasileiro de Informática na Educação* p. 657-666.

- Kampff, A. J. C. (2009) “Mineração de Dados Educacionais para Geração de Alertas em Ambientes Virtuais de Aprendizagem como Apoio a Prática Docente”. Tese de Doutorado apresentada à Universidade Federal do Rio Grande do Sul.
- Kumar, S. A. e Vijayalakshmi, M.N. (2011) “Efficiency of Decision Trees in Predicting Student’s Academic Performance” in International Conference on Computer Science, Engineering and Applications, p. 335–343.
- Kumar, S.A. e Vijayalakshmi, M.N. (2012) “Mining Of Student Academic Evaluation Records in Higher Education” in Recent Advances in Computing and Software Systems, p. 67-701.
- Martins, L. C., Lopes, D. A. e Raabe, A. (2012) “Um Assistente de Predição de Evasão aplicado a uma disciplina Introdutória do curso de Ciência da Computação” em anais do Simpósio Brasileiro de Informática na Educação.
- McLaren, B.M., Adams, D., Durkin, K., Gogvadze, G., Mayer, R.E., Rittle-Johnson, B., Sosnovsky, S., Isotani, S. e van Velsen, M. (2012) “To err is human, to explain and correct is divine: A study of interactive erroneous examples with middle school math students” in: Ravenscroft, EC-TEL, v. 7563, p. 222–235.
- Prata, D.N., Baker, R.S.J.d., Costa, E., Rosé, C.P., Cui, Y. e de Carvalho, A.M.J.B. (2009) “Detecting and Understanding the Impact of Cognitive and Interpersonal Conflict in Computer Supported Collaborative Learning Environments” in Proceedings of the International Conference on Educational Data Mining, p. 131-140.
- Quadri, M. N. e Kalyankar, N.V. (2010) “Drop Out Feature of Student Data for Academic Performance Using Decision Tree Techniques” in Global Journal of Computer Science and Technology, v. 10 i. 2, p. 2.
- Romero, C., Ventura S., Espejo, P. G., e Hervás C. (2008) “Data mining algorithms to classify students” in Proceedings of Educational Data Mining. p. 20-21.
- Sanches, A. Cardoso, J. M. P. e Delbem, A. C. B. (2011) “Identifying Merge-Beneficial Software Kernels for Hardware Implementation” in International Conference on Reconfigurable Computing and FPGAs, p. 74-79.
- Werner, L., McDowell, C. e Denner J. (2013) “Middle School Students Using Alice: What Can We Learn from Logging Data?” in ACM technical symposium on Computer science education, p. 507-512.
- Cilibrasi, A. R. e Vitanyi A. P. "Clustering by compression" IEEE Trans. Information Theory. V. 51(4), p. 1523-1545, 2005.
- Felsenstein J.. Inferring phylogenies. Sinauer Associates. 2003.
- McLaren, B.M., Lim, S.J. e Koedinger, K.R.. When and how often should worked examples be given to students? New results and a summary of the current state of research. Proceedings of the 30th Annual Conference of the Cognitive Science Society. P. 2176–2181, 2008.