



MINERAÇÃO DE DADOS: A NOVA MANEIRA DE COMPREENDER O COMPORTAMENTO DO USUÁRIO DO ENSINO A DISTÂNCIA

Michele Bandeira
Centro Universitário Ritter dos Reis
michele.bandeira@gmail.com

Natália Braulio dos Santos
Centro Universitário Ritter dos Reis
natalia.braulio@gmail.com

Vinicius Gadis Ribeiro
Centro Universitário Ritter dos Reis
vinicius@uniritter.edu.br

Wilson Pires Gavião Neto
Centro Universitário Ritter dos Reis
wilson_gaviao@uniritter.edu.br

Resumo: Na literatura do design, é consenso que a combinação de mais de uma técnica de coleta de informações do usuário pode produzir melhores conclusões a respeito de seu comportamento. Contudo, a mineração de dados raramente tem sido envolvida na prática do Design Centrado no Usuário bem como considerada na prática do Design Instrucional (DI). O objetivo deste trabalho é apresentar um paralelo entre métodos tradicionais do Design Centrado no Usuário e as contribuições da área de Mineração de Dados no sentido de melhor compreender o comportamento de estudantes em Ambientes Virtuais de Aprendizagem (AVAs). Para isso, procede-se metodologicamente em 3 partes: (i) revisa-se práticas no âmbito da captura do feedback do estudante segundo a literatura do DI; (ii) compreende-se o funcionamento de técnicas de mineração de dados e o que elas se propõe a resolver e (iii) elucida-se como são ou podem ser empregadas no contexto do ensino à distância para melhor compreensão do comportamento do estudante. As principais conclusões são que a Mineração de Dados (MD) conecta práticas quantitativas e qualitativas empregadas no DI e oportuniza novas formas de análise da resposta dada pelo estudante aos estímulos gerados pelas atividades de ensino.

Palavras-chave: Design Centrado no Usuário, Mineração de Dados Educacionais, Design Instrucional, *User Experience*

Abstract: *In Design literature, there is consensus that the combination of more than one user information collection technique can produce better conclusions about their behavior. However, data mining has rarely been*

involved in the practice of User-Centered Design. The aim of this paper is to present a parallel between traditional methods of User-Centered Design and the contributions of the Data Mining area in order to better understand the behavior of students in Learning Management Systems (VLEs). Our methodology was divided in 3 parts: (i) revising practices in the student feedback capture according to DI 's literature; (ii) understanding data mining techniques and what they propose to solve and (iii) elucidating as are or may be used in the context of distance learning for better understanding of student behavior . The main conclusions are that the Data Mining (MD) connects quantitative and qualitative practices employed in DI and provides new forms of response analysis given by the student to the stimuli generated by educational activities.

Keywords: *User-centered Design, Educational Data Mining, Instructional Design, User Experience.*

1. INTRODUÇÃO

É constante a busca por mecanismos que aumentem o conhecimento sobre as demandas, as preferências e o comportamento de usuários. Existem modelos, processos e técnicas consagradas do Design que propõe centrar o projeto nas necessidades do usuário, através da observação de seu comportamento (CROSS, 2008; JONES, 1992; NORMAN, 2014), e também como uma forma de atingir a inovação (VON HIPPEL, 2009).

Tradicionalmente, a elicitación de necessidades de consumidores e/ou usuários é realizada através de diversas técnicas como entrevistas com usuários, consultas a especialistas, levantamento de experiências de sucesso/falha de produtos passados, prototipagem com análise de respostas de usuários, sessões de *brainstorming* ou mesmo um estudo de produtos concorrentes (OGLIARI et al., 2008). No entanto, desafios ainda persistem na tarefa de melhor compreender as necessidades de usuários/consumidores. No âmbito da educação, por exemplo, onde busca-se o *feedback* do estudante como forma de aprimorar-se experiências de ensino, a aplicação de formulários e entrevistas incorre no risco de registrar-se apenas o que o estudante percebe de seu comportamento em determinado momento, contando com sua memória da ação ocorrida, sendo que a memória pode ser falha, esquecida ou distorcida (ROMERO et al., 2011).

Por outro lado, com baixo custo e de forma abundante, a internet oferece riqueza de dados no sentido de compreender o comportamento de consumidores/usuários. Com a popularização da Web 2.0, muitos consumidores expressam suas experiências na internet através de blogs, fóruns e outros tipos de mídias sociais. Além disso, mesmo de forma não intencional, consumidores navegam através de *links* na web gerando um rastro de potenciais interesses que são armazenados pelos provedores do conteúdo visitado. A disponibilidade destes dados motivou o surgimento de áreas de estudo como a Mineração de Dados e/ ou *Big Data* (SIEGEL, 2013; RICCI et al.,2011), ambos focados em técnicas computacionais capazes de extrair conhecimento de grandes volumes de dados e estabelecer correlações entre usuários, suas preferências e características de produtos/serviços.

O cenário acima aponta o surgimento de novos recursos de apoio ao projeto

de produtos/serviços inovadores, especificamente na etapa de melhor compreender demandas de consumidores/usuários (OHSAWA; KATSUTOSHI, 2009; BAE; KIM, 2011). Neste contexto, profissionais e autores da área do design emergem com denominações como *data-driven design*¹ ou *data-informed design*² para esta área que busca descobrir correlações dentre dados massivos e delega ao designer a tarefa de interpretar e apontar perspectivas sobre as relações identificadas.

Da mesma forma que consumidores produzem rastros digitais ao interagir com lojas virtuais, estudantes também podem ter seu comportamento analisado ao interagir com Ambientes Virtuais de Aprendizagem (AVAs) (PEÑA-AYALA, 2014; ROLIM et al., 2014). Esta recente área de estudo é conhecida como Mineração de Dados Educacionais (MDE) ou *Learning Analytics* (BAKER, ISOTANI E CARVALHO, 2011; ROMERO, 2011), e busca extrair conhecimento a partir de dados provenientes de interações em AVAs. Comparando-se com atividades realizadas por um usuário em um *e-commerce*, as quais são tipicamente associadas a simples acessos a conteúdos disponíveis na loja virtual, a diversidade de atividades comumente presente em AVAs (como testes, tutoriais e vídeos) permite uma modelagem potencialmente mais rica do comportamento do usuário/estudante (PAHL e DONNELLAN, 2002).

O objetivo deste trabalho é apresentar um paralelo entre métodos tradicionais do Design Centrado no Usuário e as contribuições da área de Mineração de Dados no sentido de melhor compreender o comportamento de estudantes em Ambientes Virtuais de Aprendizagem (AVAs). Para tal, é apresentada uma breve revisão de práticas de áreas como o Design Instrucional e a Mineração de Dados Educacionais ou *Learning Analytics*. A Figura 1 ilustra a confluência de áreas envolvidas neste tema ao mesmo tempo em que situa as contribuições deste trabalho dentro da recente literatura delimitada pela região central da figura.

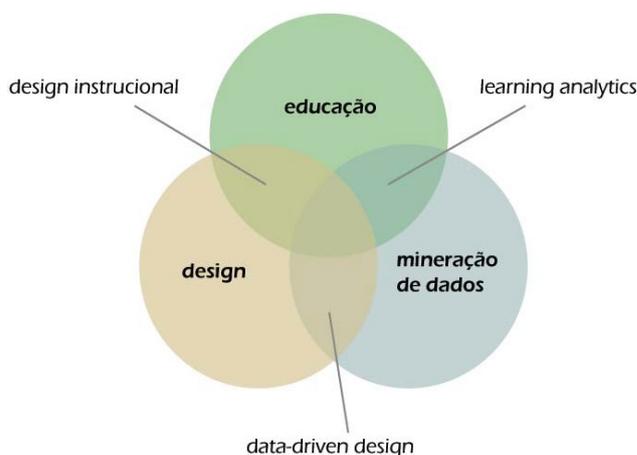


Figura 1 - Design, educação, mineração de dados e suas intersecções.

Fonte: elaborado pelos autores.

A organização deste artigo é feita da seguinte forma: a Seção 2 revisa métodos tradicionais do Design para conhecer o comportamento de usuários, sendo que a Seção 3 os contextualiza na prática do Design Instrucional. A Seção 4 apresenta as

¹ <https://uxmag.com/articles/six-myths-about-data-driven-design>

<http://datascopeanalytics.com/>

² <http://www.uxforthemasses.com/data-informed-design/>

ferramentas da Mineração de Dados exemplificadas no cenário do Ensino a Distância. A Seção 5 tem como foco esclarecer intersecções entre a MDE e o DI. Seção 6 apresenta as considerações finais deste trabalho.

2. OS INSTRUMENTOS PARA SE COMPREENDER O USUÁRIO

Design Centrado no Usuário é um conjunto de métodos que têm por base a projeção para os usuários e, de preferência, envolvendo-os no processo de desenvolvimento, visando seus comportamentos e necessidades (ABRAS; MALONEY-KRICHMAR; PREECE, 2004). O estudo do comportamento de usuários é consagrado no design e faz parte da literatura clássica da área (JONES, 1992; BOMFIM, 1995; MACCOMERK; VERGANTI; IANSITI, 2001; CROSS, 2008; VON HIPPEL, 2009; BONSIPE, 2012; NORMAN; VERGANTI, 2014).

Na prática, a coleta de dados ou a compreensão do comportamento de usuários pode não ser uma tarefa simples, uma vez que deve-se compatibilizar a complexidade da situação de uso com vantagens e desvantagens de diferentes instrumentos disponíveis para se realizar a tarefa de coleta de informações (GOODMAN; KUNIAVSKY; MOED, 2012; VREDENBURG et al., 2002; USABILITY, 2002). O Quadro 2 sumariza alguns instrumentos comumente mencionados na literatura de *User Research* (GOODMAN; KUNIAVSKY; MOED, 2012; USABILITY, 2002) bem como seus prós e contras quanto a qualidade dos dados gerados. Uma discussão abrangente sobre estes instrumentos está além do escopo deste trabalho, porém pode ser encontrada em (GOODMAN; KUNIAVSKY; MOED, 2012; ROHRER, 2014).

Quadro 1: Prós e contras de instrumentos de coleta de dados e análise de comportamento de usuários comumente mencionados na literatura.

Instrumento/como funciona	Pros (+) e Contras (-)
Questionário: São feitas aos usuários um conjunto padrão de perguntas que pode ser em papel, pessoalmente, por telefone ou e-mail.	(+) pode coletar dados de muitos usuários rapidamente (-) pode ser um desafio desenvolver um instrumento confiável (-) o que o usuário fala pode não ser uma representação exata de seu comportamento
Entrevista: Entrevistador faz perguntas semi-estruturadas pessoalmente ou por telefone. Os entrevistados podem ser colaboradores, especialistas no conteúdo e os próprios usuários.	(+) baixo custo, forma direta para coletar dados (+) eficaz para identificar as necessidades e opiniões dos usuários (-) não irá revelar qualquer informação que os indivíduos não queiram ou que não saibam conscientemente (-) depende da memória do participante
Grupo Focal: Usuários participam, junto a um facilitador, de um grupo para discutir ideias e opiniões sobre o produto/serviço. São feitos encontros com grupos até que as respostas se tornem repetitivas.	(+) elicit objeções e inseguranças a respeito de um sistema ou a sua utilização que pode não ser descoberto através de outros meios (+) pode gerar grandes quantidades de dados num curto período de tempo (-) requer um facilitador experiente (-) sujeito a inconstâncias quanto ao que as pessoas pensam e o que dizem no grupo

<p>Estudo etnográfico de campo: Projetistas/ designers visitam usuários em seu local de trabalho para modelar seu comportamento através da análise de suas atividades e fatores do ambiente de trabalho.</p>	<p>(+) cria uma compreensão abrangente do problema que está sendo resolvido. (-) trabalho intensivo.</p>
<p>Análise de dados de log: As ações de usuários de um sistema são coletadas de um servidor de logs e examinadas para encontrar padrões de uso e áreas de problema em potencial.</p>	<p>(+) fornece um histórico de tráfego de uso de um sistema (+) rápido, fácil reunir dados sobre o comportamento dos usuários sem ter que realmente recrutá-los (-) arquivos de log não indicam por que os usuários fizeram ou deixaram de fazer certas ações</p>
<p>Análise de tarefa: Por meio de observação e/ou entrevistas com usuários experientes e novatos, os designers identificam os passos necessários para que os usuários alcancem os objetivos de uso.</p>	<p>(+) pode revelar informações exploráveis no design de softwares (por exemplo, atalhos que usuários experientes tomam) (-) pode ser demorado para realizar (-) caso não seja observado um usuário experiente, se pode inadvertidamente reproduzir uma maneira ineficiente de se completar uma tarefa (-) caso se observe somente usuários experientes, pode-se não descobrir problemas específicos dos principiantes.</p>

Fonte: elaborado pelos autores.

De especial importância para este trabalho é o instrumento *Análise de dados de logs*. Embora esta técnica não capture alguns porquês do comportamento do usuário, destaca-se que suas vantagens contornam quase que por completo as desvantagens dos demais instrumentos descritos no Quadro 1, ao mesmo tempo em que também partilha de parte das vantagens dos outros instrumentos.

3. DESIGN INSTRUCIONAL E A COLETA DA RESPOSTA/FEEDBACK DO ESTUDANTE

O Design Instrucional (DI) é parte importante no planejamento de um curso a distância. O DI deve ser modelado de forma que os perfis dos estudantes sejam levados em consideração, a metodologia de educação utilizada seja respeitada e a sequência de atividades esteja em conformidade com tais definições (NUNES; SCHIEL, 2013).

Segundo Branch (2010), Design Instrucional é um processo iterativo de (a) planejamento de objetivos de desempenho, (b) seleção de estratégias instrucionais, (c) seleção ou criação de materiais e (d) avaliação de resultados. Para tanto, existem diversos processos, sendo o chamado ADDIE um dos mais empregados. O processo ADDIE propõe 5 etapas (BRANCH, 2010), cada qual com seu objetivo: análise, projeto, desenvolvimento, implementação e avaliação. De maneira a manter um foco no estudante, as etapas de implementação e avaliação são de especial interesse por demandarem uma captura direta de suas reações, ou comportamento, e desempenho.

O Quadro 2 sumariza as fases de implementação e avaliação.

Quadro 2: Design instrucional: sumário das etapas de implementação e avaliação do processo ADDIE.

Fase	Objetivo	Como/questionamentos
Implementação	Preparo do ambiente de aprendizado e engajamento dos estudantes. Execução do curso propriamente dito.	Como está sendo a reação dos participantes e se há necessidade de adaptação/ajuste no conteúdo ou na forma como ele é disponibilizado. O <i>feedback</i> dos estudantes e dos professores é importante para que se possa fazer ajustes.
Avaliação	Comparação dos resultados esperados com os obtidos.	Os objetivos foram alcançados? O que foi aprendido? Precisa ser realizada alguma ação complementar? O <i>feedback</i> nesta fase torna possível reavaliar e iniciar o processo novamente, contribuindo para sua melhora/crescimento.

Fonte: elaborado pelos autores, adaptado de Branch (2010).

No contexto acima, Filatro (2008) observa que é possível avaliar a solução de ensino de duas formas: a primeira ocorre ao final de uma unidade do curso, e tem como objetivo identificar como os alunos reagiram a um programa em relação a conteúdo, tecnologia e metodologia. Também chamada de *avaliação da reação*, é feita por meio de questionários com questões objetivas e abertas. A segunda forma de avaliação, chamada *avalição formativa*, deve ocorrer durante todo o curso, desde antes do início até depois de seu fim, e proporciona uma análise mais completa, além de oferecer “subsídios para o aperfeiçoamento da solução proposta a partir de *feedbacks* de estudantes e educadores.” (FILATRO, 2008, p.32). Na prática, esta avaliação também é feita “por meio de listas de verificação que devem ser preenchidas pelo designer instrucional e por meio de testes de uso com estudantes potenciais” (FILATRO, 2008, p. 146).

Assim como sugerido por Filatro (2008) acima, quando o processo de DI demanda a coleta de dados do usuário (MATTAR, 2014), nesse caso o estudante, inúmeros outros trabalhos na área de DI também fazem uso dos instrumentos tradicionais de coleta mencionados na Seção 2, como entrevistas, questionários, grupos focais (SANTOS, 2011; SNELBECKER; MILLER; ZHENG, 2011) e observações contextuais ou etnográficas (DIRKSEN, 2012). Dessa maneira, é razoável concluir que as vantagens e limitações discutidas no Quadro 2 também estão presentes na rotina do profissional que projeta atividades instrucionais.

4. MINERAÇÃO DE DADOS: TAREFAS BÁSICAS

Mineração de dados (MD) é um processo orientado a extrair conhecimento implícito, previamente desconhecido, e potencialmente útil a partir de grandes e heterogêneos repositórios de dados (WITTEN; FRANK, 2005). Por esta razão a área de mineração de dados também é conhecida sob o nome de Descoberta de Conhecimento em Bases de Dados (do inglês *Knowledge Discovery in Databases* - KDD).

A MD tem sido usada em diversas áreas para identificar padrões de comportamento e encontrar *insights* que gerem melhorias em produtos e serviços. A compreensão de como isso ocorre inicia-se por estabelecer-se tarefas básicas de um sistema que tem por objetivo automatizar a descoberta de conhecimento sobre uma base dados. Neste sentido, pode-se distinguir dois tipos básicos de tarefas (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996): *predição* e *descrição*, que ainda dividem-se em tarefas de *classificação*, *regressão*, *agrupamento* ou *clustering*, *associação* e *sumarização*. Dentre estas, apresenta-se a seguir as tarefas mais empregadas no

contexto educacional (PEÑA-AYALA, 2014), bem como um detalhamento sobre dados de usuários comumente armazenados por AVAs e que têm servido de matéria-prima para técnicas de MD.

4.1 Ambientes virtuais e o rastro digital dos estudantes

Ao fazerem uso de AVAs, estudantes deixam rastros de suas interações: cada clique, juntamente com o instante em que ocorreu, fica gravado em um banco de dados. Desta forma, a análise desses dados (chamados *logs*) permite reconstruir o caminho percorrido por cada usuário. Nesse cenário, em que cada interação é armazenada, gera-se um volume considerável de dados, cuja análise manual torna-se exaustiva e frequentemente impraticável. Ainda que AVAs prestem suporte à construção de gráficos, destaca-se que a mineração de dados busca ir além de uma análise exploratória, conforme elucidado no início desta seção.

De acordo com o Manual do Moodle³ - um dos AVAs mais difundidos no mundo, e o mais utilizado no Brasil (ABED, 2015) - ao interagirem com o sistema, estudantes podem executar *atividades* e utilizarem-se de *recursos*. Atividades são ferramentas de comunicação, interação entre usuários e avaliação. Recursos permitem a educadores disponibilizar material de apoio às atividades, como arquivos, *links* para sites externos e vídeos. Desta forma, as interações de um estudante no sistema acabam por armazenar um rico conjunto de informações que reflete seu comportamento quanto ao acesso a materiais didáticos, participações em fóruns/*chats*, assiduidade, realização de tarefas e testes. A Tabela 1 ilustra parte de um relatório de *logs* extraído do AVA Moodle.

Tabela 1 - Relatório parcial de *logs* extraído do AVA Moodle: registro histórico de interações de estudantes. Coluna Ação ilustra atividades em fóruns, acesso a material de apoio e submissão de trabalhos.

Hora	Estudante	Ação
05/11/2015 19:27:04	João	<i>forum view</i>
05/11/2015 19:27:24	João	<i>forum view</i>
05/11/2015 19:29:31	João	<i>resource view</i>
05/11/2015 19:45:27	Ana	<i>resource view</i>
05/11/2015 20:15:31	Ana	<i>assign view</i>
05/11/2015 20:16:00	Ana	<i>assign submit</i>
⋮	⋮	⋮

Fonte: elaborado pelos autores.

É razoável observar que os registros históricos de interações, como ilustrados na Tabela 1, refletem as respostas de estudantes aos estímulos gerados pelas atividades de ensino, e que diferentes aspectos do seu comportamento podem ser analisados não apenas a luz de abordagens consagradas da aprendizagem como o comportamentalismo (GRELLER; DRACHSLER, 2012), mas também como o construtivismo (KENT et al., 2016; BERLAND et al., 2014).

4.2 Tarefas preditivas

Uma tarefa preditiva caracteriza-se pela busca de um padrão de comportamento que seja capaz de prever o comportamento de uma futura entidade (FAYYAD; PIATETSKY-

³ Manual do Moodle disponível em https://docs.moodle.org/24/en/Course_overview_report.

SHAPIRO; SMYTH, 1996). Como exemplo no contexto educacional, pode-se citar uma tarefa de predição de insucesso acadêmico de estudantes em uma dada disciplina, em que busca-se estimar, sobre dados de edições passadas da disciplina, um padrão de comportamento de estudantes que tenha resultado em insucesso. Deste modo, uma vez que este padrão de comportamento seja identificado em um estudante que esteja cursando uma edição corrente da disciplina, pode-se inferir a predição de insucesso para ele, possibilitando que ações proativas sejam tomadas de forma a evitar a confirmação do insucesso.

De um ponto de vista prático, a Figura 2(a) ilustra uma transformação dos dados mostrados na Tabela 1, em que cada linha representa um estudante e seu comportamento em termos de totalizações de categorias de ações realizadas ao longo do período letivo. Assumindo-se que estes dados dizem respeito a uma disciplina já finalizada, em que já se conhece a situação final dos estudantes (Aprovado ou Reprovado), busca-se então descobrir um comportamento recorrente que esteja associado com o desempenho acadêmico de Aprovação ou Reprovação. Esta etapa denomina-se de *Treinamento*, e tem como resultado um *modelo preditivo* ou um simplesmente um *preditor*. Na figura 2(a), as colunas *forum*, *resource* e *assign* são denominadas de variáveis de entrada, e a coluna **Final** é denominada de variável de saída ou variável alvo, sendo que os valores que aparecem nesta variável de saída são denominados de *classes*. Isto é, no caso da variável **Final** mostrada na Figura 2(a), observa-se duas classes de estudantes: *Rep* (reprovados) e *Apr* (aprovados). Neste cenário, espera-se então que o preditor estimado seja resultado da descoberta de um padrão que relacione valores de variáveis de entrada (*forum*, *resource* e *assign*) com valores da variável de saída (**Final**).

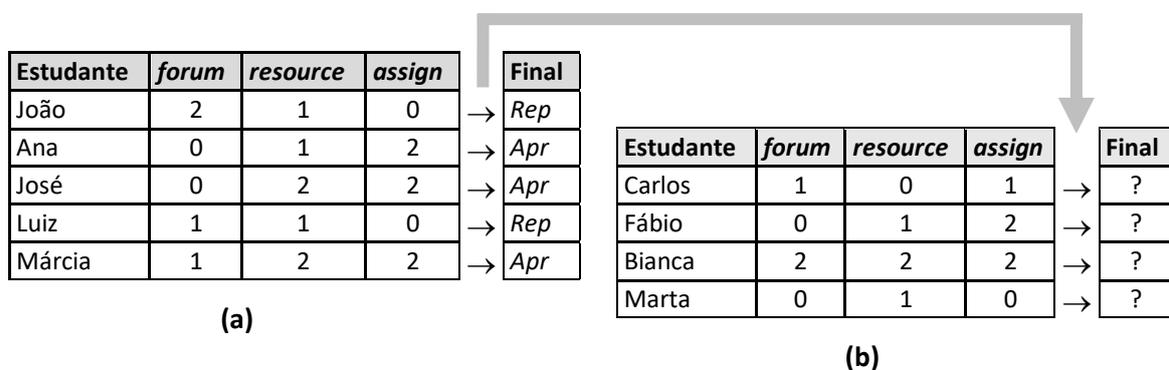


Figura 2 - (a) Etapa de treinamento de um modelo preditivo: associa quantificações históricas de interações (*fórum*, *resource* e *assign*) com o status final na disciplina (Aprovado ou Reprovado). (b) Etapa de predição: modelo preditivo infere um status final a partir de dados do período letivo corrente.
Fonte: elaborado pelos autores.

Finalmente, emprega-se o preditor na etapa denominada de predição, que caracteriza-se pela presença de valores de variáveis de entrada e pela ausência dos valores da variável de saída, conforme ilustrado na Figura 2(b). Neste cenário, a partir dos dados de comportamento correntes de cada estudante (isto é, totalizações de interações em *forum*, *resource* e *assign*), é papel do preditor inferir os valores (Rep ou Apr) da variável de saída para cada um deles.

Tarefas de mineração de dados de cunho preditivo ainda se distinguem em **classificação** e **regressão** (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). O tipo do dado que consta na variável de saída é o determinante. No caso de uma tarefa de

classificação, a variável de saída apresenta dados categóricos ou discretos, como exemplificado pela variável **Final** na Figura

2(a), em que os estudantes são categorizados em *Apr* ou *Rep*. Sendo assim, a tarefa preditiva ilustrada na Figura 2 trata de uma tarefa de classificação. Por outro lado, se a variável **Final** fosse representada em termos de valores numéricos e contínuos, como uma nota numérica e real entre 0 e 10, a tarefa de predição seria tratada como uma tarefa de regressão.

4.2.1 Seleção de variáveis

Como detalhado acima, tarefas preditivas buscam estabelecer uma relação entre variáveis de entrada (**forum**, **resource** e **assign**) e uma variável de saída (**Final**: Aprovado ou Reprovado). Na prática, é bastante comum algumas variáveis de entrada se mostrarem mais importantes que outras na predição dos valores da variável de saída. Neste contexto, destaca-se a grande utilidade de técnicas de seleção de variáveis relevantes (GUYON; ELISSEFF, 2003), vista que podem prestar suporte a decisões no sentido de se melhor compreender fatores/ variáveis determinantes com respeito a variável de saída.

Como exemplo no contexto educacional, pode-se citar conclusões a respeito da relevância de atividades em fóruns de discussão para propósitos de predição do conceito final do estudante em uma dada disciplina (MACFADYEN; DAWSON, 2010).

4.3 Tarefas descritivas

Em tarefas de descrição, o objetivo é explorar ou descrever um conjunto de dados (CARVALHO et al., 2011). As tarefas descritivas são genericamente divididas como descrito a seguir.

Agrupamento ou clustering: em que (dados de) entidades são agrupadas de acordo com sua similaridade. Com base nesta comparação por similaridade, busca-se identificar um conjunto finito de grupos que melhor caracterizem um conjunto de entidades.

Como exemplo no contexto educacional, sobre os dados ilustrados na Tabela 1, pode-se citar uma tarefa que visa identificar possíveis agrupamentos de estudantes a partir de suas interações em um AVA. Como possível resultado, está a constatação de grupos de (comportamento de) estudantes caracterizados por uma maior interação social através das ferramentas de chat e fóruns de discussão.

Associação: consiste em encontrar padrões frequentes, ou regras frequentes, de associação entre características de entidades. A construção das regras se dá sobre um conjunto de itens, ditos frequentes, e categorizados como *Antecedentes* e *Consequentes*, na forma de *Antecedente* \Rightarrow *Consequente*.

Como exemplo de conjunto de itens, pode-se definir as ações/interações ilustradas na Tabela 1, isto é, { *forum view*, *resource view*, *assign view*, ...}. Neste contexto, uma tarefa descritiva de mineração de regras de associação sobre um histórico de interações de estudante em AVAs, poderia resultar em regras como:

assign view \Rightarrow *forum post*: indicando que é frequente um comportamento de estudantes que, após tomarem conhecimento sobre uma atividade de entrega de trabalho (*assign view*), recorrem ao fórum de discussão, ou mesmo

assign view \Rightarrow *resource view*: indicando que também é frequente um outro comportamento de estudantes que, após tomarem conhecimento sobre a atividade

de entrega (*assign view*), recorrem a um material disponibilizado pelo educador (*resource view*) para prestar suporte ao aprendizado.

5. MINERAÇÃO DE DADOS COMO FERRAMENTA PARA O DESIGN INSTRUCIONAL

Embora inúmeros trabalhos da área de Mineração de Dados Educacionais (MDE) tenham direcionado resultados úteis ao planejamento de experiências de aprendizado, poucos são os trabalhos que envolvem o processo prático do DI (Seção 3) na discussão. No contexto de uso de MD, Educação e Design, a aplicação da mineração de dados no planejamento do design instrucional é feito por Santos, Nunes e Schiel (2014), porém sobre um curso fictício. Por esta razão, o foco desta seção, e deste artigo, está no sentido de elucidar intersecções entre a MDE e o DI, partindo-se de práticas comumente adotadas no DI, na tarefa de compreender a resposta do estudante, e chegando-se nas potenciais contribuições da MDE para tal propósito.

Instrumentos de coleta de informações e compreensão do comportamento de usuários mencionados no Quadro 1, como questionários e entrevistas, são muito usados na prática (BARGAS-AVILA; HORNBAEK, 2011), e consistem em uma abordagem que essencialmente lança hipóteses e busca validá-las (BAE; KIM, 2011). Neste contexto, como discutido na Seção 4, a MD pode contribuir ao ter por base uma abordagem diferente: em lugar de se predefinir hipóteses, busca-se descobri-las. Como exemplo, destaca-se a utilidade de técnicas de seleção (ou descoberta) de variáveis relevantes, conforme discutido na Seção 3.2.1, no sentido de apoiar o projetista da instrução sobre os aspectos da experiência de aprendizado que demandam maior atenção.

Ambientes educacionais representam um cenário de dados propício à utilização de técnicas de MD, uma vez que a adoção de AVAs é bastante difundida e, com isso, tem-se à disposição as vantagens da Análise de Dados de *Logs* como um poderoso instrumento de coleta de dados do usuário/estudante, como descrito na Seção 2.

Embora vantajosos, dados de *logs* retratam o comportamento do usuário/estudante de uma maneira quantitativa (ROHRER, 2014) e, conforme o Quadro 1 podem não carregar desejáveis informações qualitativas, como por exemplo, a situação final do estudante (Aprovado ou Reprovado) que gerou os dados de *logs*. Como discutido na Seção 3.2, dados qualitativos são necessários a variáveis de saída em tarefas preditivas de mineração de dados, sobretudo em tarefas de classificação. Contudo, é da rotina de instituições de ensino produzir/armazenar dados qualitativos a respeito de seus estudantes, como desempenho acadêmico e características de disciplinas cursadas. Este rico contexto de dados, qualitativo e quantitativo (PEÑA-AYALA, 2013), pressupõe uma combinação de resultados de instrumentos complementares de coleta de dados de usuários/estudantes, envolvendo questionários, entrevistas e análise de *logs*, cabendo a mineração de dados o papel de descobrir conhecimento latente sobre esta diversa e massiva quantidade de dados.

Com relação a compreensão do comportamento de estudantes, a área da educação pode beneficiar-se com o uso da mineração de dados de diversas maneiras. Referências como (PEÑA-AYALA, 2013 e 2014) e (ROMERO et al., 2011) são abrangentes com relação a potencialidades e diversidade de resultados já alcançados. Tendo por base as abordagens descritas na Seção 3, a mineração de dados educacionais (MDE) tem evoluído nos seguintes objetivos gerais (PEÑA-AYALA, 2013):

Modelagem do estudante: busca caracterizar estudantes quanto a conhecimento, habilidades, motivação, satisfação, cognição, atitudes, experiências e progressos de aprendizado, bem como quanto a problemas que afetam negativamente o aprendizado: cometer muitos erros, chutar respostas, não explorar recursos disponíveis.

Predição de desempenho de estudantes: busca prever o conceito final ou outros tipos de resultados de aprendizado como retenção/evasão.

Gerar recomendações: busca identificar o contexto educacional corrente do estudante de maneira a recomendar conteúdos personalizados.

Em resumo, a MD oferece ao DI a oportunidade de aprimorar suas práticas pela possível descoberta de conhecimento sobre dados massivos de estudantes, ao mesmo tempo em que possibilita a criação de modelos que podem ser aplicáveis e testados em diferentes contextos educacionais.

6. CONCLUSÃO

Este artigo analisou as práticas do design instrucional na atividade de capturar o *feedback* do estudante para uso no aprimoramento das experiências instrucionais. Como forma de contribuir com tais práticas, apresentou-se a mineração de dados, suas tarefas e como tem sido aplicada em contextos educacionais, tendo em vista a abundância de dados produzida atualmente por estudantes ao interagir com ambientes virtuais de aprendizagem.

Ao possibilitar o estabelecimento de relações entre grandes volumes de dados quantitativos, como logs de estudantes em AVAs, e dados qualitativos, como sucesso/insucesso acadêmico, a contribuição da mineração de dados pode ser entendida como um catalisador de práticas qualitativas e quantitativas empregadas tradicionalmente no Design Instrucional. Tendo a mineração de dados à disposição, o designer de projetos de Ead pode analisar grandes volumes de *feedbacks* de usuários/estudantes, agrupá-los por certas características, prever seu comportamento e, sobretudo, descobrir hipóteses em lugar de predeterminá-las.

REFERÊNCIAS

- ABRAS, Chadia; MALONEY-KRICHMAR, Diane; PREECE, Jenny. **User-centered design. Bainbridge, W. Encyclopedia of Human-Computer Interaction.** Thousand Oaks: Sage Publications, v. 37, n. 4, p. 445-456. Thousand Oaks, 2004.
- BAE, J. K.; KIM, J. Product development with data mining techniques: A case on design of digital camera. **Expert Systems with Applications**, 38(8):9274 – 9280, 2011.
- BAKER, R.; ISOTANI, S.; CARVALHO, A. **Mineração de Dados Educacionais: Oportunidades para o Brasil.** Revista Brasileira de Informática na Educação, v. 19, n. 02, p. 03, 2011.
- BARGAS-AVILA, J. A.; HORNBAEK, K. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.** ACM, p. 2689-2698, 2011.
- BEHAR, Patricia Alejandra. **Competências em educação à distância.** Penso Editora. Porto Alegre, 2013. BERLAND, Matthew et al. **Educational data mining and learning analytics: Applications to constructionist research.** Technology, Knowledge and

Learning, 19(1-2), pp.205-220, 2014.

BONSIEPE, Gui. **Design como prática de projeto**. Ed. Blucher, São Paulo, SP, 2012.

BOMFIM, G. A. **Metodologia para desenvolvimento de projetos**. Editora Universitária / UFPB. João Pessoa, 1995.

BRANCH, R. M. **Instructional design: The ADDIE approach**. Athens: Springer Science+Business Media, 2010.

CARVALHO, André et al. **Inteligência Artificial - Uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2011.

CROSS, Nigel. **Engineering design methods: strategies for product design**. John Wiley & Sons, 2008.

DIRKSEN, Julie. **Design for how people learn**. New Riders, 2012.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. **From data mining to knowledge discovery in databases**. AI magazine, 17(3), p.37, 1996.

GOODMAN, Elizabeth; KUNIAVSKY, Mike; MOED, Andrea. **Observing the User Experience: A Practitioner's Guide to User Research** (2nd ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2012.

GRELLER, Wolfgang; DRACHSLER, Hendrik. **Translating Learning into Numbers: A Generic Framework for Learning Analytics**. Educational technology & society 15(3), p.42-57, 2012.

GUYON, Isabelle; ELISSEFF, André. **An introduction to variable and feature selection**. The Journal of Machine Learning Research v. 1, n. 3 , p. 1157-1182, 2003.

JONES, John Christopher. **Design methods**. (2nd ed.) John Wiley & Sons. Nova Iorque, 1992.

KENT, Carmel et al. **Interactivity in online discussions and learning outcomes**. Computers & Education, 97, pp.116-128, 2016.

MACFADYEN, L. P.; DAWSON, S. Mining LMS data to develop an “early warning system” for educators: A proof of concept. **Computers and Education**, v. 54, n. 2, p. 588–599, 2010.

MATTAR, João. **Design educacional: educação a distância na prática**. São Paulo: Artesanato Educacional, 2014.

NORMAN, Donald A.; VERGANTI, Roberto. **Incremental and radical innovation: Design research vs. technology and meaning change**. Design Issues, v. 30, n. 1, p. 78-96, 2014.

NUNES, I. D.; SCHIEL, U. **Utilização de Grafo de Alcançabilidade para a verificação de Design Instrucional bem formado**. XXXIII CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. Anais... . p.1396–1405, 2013. Maceió, Brazil.

OGLIARI, Andre et al. **Projeto integrado de produtos: planejamento, Concepção e Modelagem**. Manole, São Paulo 2008.

OHSAWA, Yukio; KATSUTOSHI, Yada. **Data mining for design and marketing**. CRC Press, 2009.

PAHL, C; DONNELLAN, D. **Data mining technology for the evaluation of web-based teaching and learning systems**. Int. Conference on E-Learning in Business, Government and Higher Education, 2002.

PALLOFF, Rena M.; PRATT, Keith. **O Aluno Virtual-um guia para trabalhar com estudantes on-line**. Penso Editora. Porto Alegre, 2004.

PEÑA-AYALA, A. **Educational data mining**: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, v. 41, n. 4 PART 1, p. 1432–1462, 2014.

PEÑA-AYALA, A. *Educational Data Mining: Applications and Trends*. Vol. 524. Springer, 2013.

RICCI, Francesco; ROKACH, Lior; SHAPIRA, Bracha. **Introduction to recommender systems handbook**. Springer US, 2011.

ROHRER, Christian. **When to use which user experience research methods**. Jakob Nielsen's alertbox, October, 2014. Disponível na internet por http em: <<https://www.nngroup.com/articles/which-ux-research-methods/>>. Acesso em 09/06/2016.

ROLIM, Vitor B.; CORDEIRO, Filipe Rolim; FERREIRA, Rafael. **Reconhecimento de Padrões Aplicados a Comentários de Fóruns Educacionais**. Centro de Informática, UFPE ,João Pessoa. 2014.

ROMERO, C.; VENTURA, S.; PECHENIZKIY, M.; BAKER, R. S. J. d. **Handbook of Educational Data Mining**. Boca Raton: Chapman & Hall/CRC Press, 2011.

SANTOS, Antonio. **Bringing Reality into the Classroom**. *Instructional Design: Concepts, Methodologies, Tools and Applications*. IGI Global, 219-237, 2011.

SIEGEL, E. **Predictive analytics**. Hoboken: John Wiley & Sons, Inc., 2013.

SNELBECKER, G. E.; MILLER, S. M.; ZHENG, R. Z. **Functional Relevance and Online Instructional Design**. *Instructional Design: Concepts, Methodologies, Tools and Applications*. IGI Global, 342-358, 2011.

USABILITY Consulting Services. **User Centered Design Methods**. Indiana University, 2002. Disponível na internet por http em: <http://www.indiana.edu/~usable/presentations/ucd_methods.pdf>. Acesso em 09/06/2016.

VON HIPPEL, Eric. **Democratizing innovation: the evolving phenomenon of user innovation**. *International Journal of Innovation Science*, v. 1, n. 1, p. 29-40. Cambridge, 2009.

VREDENBURG, Karel, et al. A survey of user-centered design practice. **Proceedings of the SIGCHI conference on Human factors in computing systems**. ACM, 2002.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical machine learning tools and techniques**. Morgan Kaufmann, 2005.