Requisitos de Dados para uso de Técnicas de Análise Quantitativa de Dados Educacionais em AVAs

Juliano Cezar Teles Vaz¹, Raphael A. Dourado¹, Alex Sandro Gomes¹, Rodrigo Lins Rodrigues²

¹Centro de Informática - Universidade Federal de Pernambuco (UFPE) ²Departamento de Educação - Universidade Federal Rural de Pernambuco (UFRPE)

{jctv, rasd2, asg}@cin.ufpe.br, rlr@ded.ufrpe.br

Abstract. This paper aims to identify, from a literature review, a set of data requirements to be considered by any Learning Management Systems (LMS) that wishes to support Educational Data Mining (EDM) and Learning Analytics (LA) functionalities. In defining these requirements, we hope to help in conceiving architectures for new LMS as well the evolution of existing ones interested in incorporating EDM and LA features.

Resumo. Este trabalho propõe, a partir de uma revisão de literatura, um conjunto de requisitos de dados a ser considerado por qualquer Ambiente Virtual de Aprendizagem (AVA) que deseje suportar funcionalidades de Mineração de Dados Educacionais (MDE) e Learning Analytics (LA). Ao propor estes requisitos, espera-se auxiliar na concepção de arquiteturas de novos AVAs, bem como informar a evolução de plataformas existentes interessadas em incorporar funcionalidades de MDE e LA.

1. Introdução

A utilização de Ambientes Virtuais de Aprendizagem (AVAs) produz um grande volume de dados relativos às interações dos aprendizes com o ambiente. Estes dados, se corretamente coletados e armazenados, podem ser utilizados como insumo para diversos tipos de análise, capazes de extrair informações valiosas sobre os processos de aprendizagem, dificuldades e comportamento de cada estudante [Koedinger et al. 2009, Baker et al. 2011, Martins, Migueis & Fonseca 2018, Ray & Saeed, 2018].

Duas áreas de pesquisa emergentes que têm explorado estas possibilidades são a Mineração de Dados Educacionais (MDE) e Learning Analytics (LA), ou de forma mais geral, a *Ciência de Dados Educacionais* [Silva et al. 2017]. No entanto, muitas vezes as possibilidades de análise utilizando técnicas propostas por essas linhas de pesquisa são limitadas pela qualidade dos dados armazenados pelos AVAs. Dessa forma, um dos pré-requisitos para que as técnicas de MDE e LA consigam fornecer os melhores resultados possíveis é a capacidade do AVA de armazenar e disponibilizar um conjunto adequado de variáveis comportamentais, de forma bem estruturada e na granularidade correta [Papamitsiou & Economides 2014, Emmons, Light & Börner 2017].

Nesse sentido, este trabalho propõe, a partir de uma revisão de literatura, um conjunto de requisitos de dados necessários à incorporação de funcionalidades de MDE e LA aos AVAs. Tais requisitos podem auxiliar desenvolvedores e arquitetos a projetar plataformas capazes de utilizar todo o potencial da Ciência de Dados Educacionais como instrumento de apoio no processo de ensino-aprendizagem.

DOI: 10.5753/cbie.sbie.2018.1941

O restante deste artigo está organizado da seguinte forma: a Seção 2 descreve o método utilizado para extração dos requisitos; a Seção 3 apresenta os resultados obtidos; a Seção 4 discute as implicações destes resultados; e por fim, a Seção 5 apresenta as conclusões e possíveis trabalhos futuros.

2. Método

A identificação dos requisitos se deu por meio de uma revisão narrativa de literatura [Okoli & Schabram 2010]. Com base na opinião de especialistas, foi selecionado um conjunto inicial de trabalhos na área de MDE e LA, composto por revisões sistemáticas de literatura, teses de doutorado e pesquisas exemplares na área. A partir da leitura destes trabalhos e de pesquisas citadas pelos mesmos, buscou-se identificar ou inferir variáveis que eram comumente utilizadas nas pesquisas, suas granularidades e variações, e a partir disso compor um conjunto de variáveis que satisfizesse aos requisitos de dados da maioria dos trabalhos consultados.

Após este levantamento, as variáveis foram agrupadas de acordo os contextos e características em comum. Na próxima seção é apresentado o conjunto final resultante deste estudo.

3. Resultados

A Tabela 1 descreve o conjunto de variáveis compiladas a partir da revisão de literatura. Para otimizar o uso do espaço nas referências bibliográficas, optou-se por citar os trabalhos encontrados por meio de revisões sistemáticas ou *surveys* utilizando "*apud*".

Tabela 1 - Conjunto de variáveis identificadas.

	Variável	Descrição e/ou exemplos	Utilizada ou referenciada em
1	Local de acesso	Quando disponível, a partir do IP (Internet Protocol) e/ou localização geográfica.	[Ramos 2016], [Rodrigues 2016], [Vieira, Parson & Byrd 2018], dentre outros.
2	Dispositivo utilizado	Smartphone, tablet, computador (PC).	[Siemens & Long 2011 apud [Vieira, Parson & Byrd 2018].
3	Recursos acessados (mídia)	Recursos como página web, vídeo, documento, apresentação (slides), fórum de discussão. Quais são os recursos mais utilizados, como estão sendo atualizados, onde surgem problemas de compreensão, etc.	[Ramos 2016], [Rodrigues 2016], [Scheffel et al. 2014], [Vieira, Parson & Byrd 2018], [You 2016], [Emmons, Light & Börner 2017], [Martins, Migueis & Fonseca 2018], [Ray & Saeed, 2018], dentre outros.
4	Atividades acessadas	Leitura textual, produção textual, quiz, jogo (gamificação), webquest (questionário), ou até mesmo atividades que requeiram submissão de arquivos de mídia (arquivo específico de imagem, áudio, codificação, vídeo, etc).	[Picciano 2012 apud Vieira, Parson & Byrd 2018], [Emmons, Light & Börner 2017], [You 2016], [Ray & Saeed, 2018], dentre outros.
5	Fluxo e tipo de interação	Identificação, por exemplo, de mensagens trocadas entre: o professor e estudante, os estudantes entre si, professores e tutores, o estudante e o ambiente, entre outros.	& Byrd 2018], [You 2016], [Emmons,
		Criação de nova thread (postagem), resposta	[Ramos 2016], [Rodrigues 2016],

6	Tipos de mensagem / postagem	a uma <i>thread</i> de postagens (pode esta ter tipos como, postagem em fórum, aula, dúvida, central de ajuda, etc).	
7	Tempo de acesso	Por atividade, por curso, por página, por videoaula, podendo ser agrupado também por horas do dia, turnos, horas ativas e/ou inativas no AVA.	Byrd 2018], [You 2016], [Borner &
8	Finalização de partes específicas de um curso do AVA	Como seção, aula, disciplina, módulo, curso, atividade, entre outros.	[Ahn et al. 2013; Jacovina et al. 2015; Jugo, Kovačić, & Slavuj 2015; Minović et al. 2015; Buckingham et al. 2012 apud Vieira, Parson & Byrd 2018], [You 2016], dentre outros.
9	Novos usuários	Controle de quantidade e frequência de novos usuários podendo ser eles tipificados por: estudante, professor, tutor, visitante, entre outros.	
10	Avaliação nas postagens	Deve-se obter <i>feedback</i> acerca das postagens de todas as instâncias da plataforma, como um fórum de discussão, facilitando a compreensão de como cada de usuário avalia um determinado recurso.	[Papamitsiou & Economides 2014], [Scheffel et al. 2014], [You 2016], [Emmons, Light & Börner 2017], [Martins, Migueis & Fonseca 2018], [Picciano 2012; Verbert, Manouselis, Drachsler & Duval 2012 apud Vieira, Parson & Byrd 2018], [Ray & Saeed, 2018], dentre outros.
11	Clicks em links e botões	Ocorrências em determinado intervalo de tempo, podendo ser inclusive em mídia (como vídeo) dentro do AVA, como forma de capturar a movimentação a nível de sequenciamento, repetição	[Papamitsiou & Economides 2014], [Chen, et al. 2016; Freeman 2016 apud Vieira, Parson & Byrd 2018], [You 2016], dentre outros.
12	Quantidade de acessos	Aumento e/ou queda de acesso do usuário individualmente e em classe (turma, ano, instituição, série, disciplina) em diversos períodos (dia, semana, mês, ano, todo período).	[Ramos 2016], [Rodrigues 2016], [You 2016], [Emmons, Light & Börner 2017], [Davis et al. 2016; apud Vieira, Parson & Byrd 2018], dentre outros.
13	Notas / Pontuação	Capturadas nas atividades práticas (sejam elas automatizadas como webquest, ou não, como produção textual). As informações desse tipo são importantes para analisar a taxa de sucesso global e parcial dos estudantes, o nível de domínio dos conceitos e competências, bem como os erros mais frequentemente diagnosticados.	[Scheffel et al. 2014], [You 2016], [Emmons, Light & Börner 2017], [M. Tiwari et al. apud Emmons, Light & Börner 2017], [Martins, Migueis & Fonseca 2018], [Minović et al. 2015 apud Vieira, Parson & Byrd 2018], [Ray & Saeed, 2018], dentre outros.
14	Informações demográficas	Compreensão acerca dos perfis de usuários do AVA, com quesitos como: gênero, idade, escolaridade, localização, língua, entre outros, além de aspectos culturais como: fatores motivacionais e objetivos de carreira que podem favorecer também para análise qualitativa.	Halawa 2015 apud Emmons, Light & Börner 2017], [Martins, Migueis & Fonseca 2018], [Vieira, Parson &

4. Discussão

Ao observar o conjunto de requisitos apresentados na Seção 3, nota-se que alguns deles podem requerer que o próprio desenho da plataforma seja repensado, como por exemplo o registro de *feedback* dos usuários a itens do AVA (variável 10). Desse modo, o ideal é que a arquitetura dos AVAs sejam pensadas já com tais requisitos de dados em mente. No caso de plataformas já estabelecidas, o esforço para adequação tende a ser significativamente maior.

Além da captura e armazenamento das variáveis descritas na Tabela 1, é importante também que, na medida do possível, elas possam ser facilmente extraídas da base de dados do AVA. Como apontado por Romero et al. (2010), boa parte do tempo gasto em estudos de MDE e LA é destinado à extração e limpeza dos dados, os quais muitas vezes precisam ser extraídos por meio de complexos scripts SQL que requerem o entendimento de todo o modelo de dados da plataforma. Uma plataforma que ofereça relatórios pré-computados — como, por exemplo, a tabela *mdl_logstore_standard_log* do Moodle — facilita muito o trabalho dos analistas.

Outro ponto importante notado durante o levantamento, foi a necessidade de ser possível reconstruir temporalmente as sequências de ações dos usuários (variável 11). Para tal, um subsistema de *logging* que forneça tais dados de forma pré-computada pode tornar mais ágil a análise dos dados. Trabalhos recentes na literatura têm chamado a atenção para a carência de estudos focados em processos de aprendizagem, os quais necessitam primordialmente de informações sobre as sequências de ações dos estudantes [Bogarín et al. 2018, Vieira et al. 2018].

Por fim, uma necessidade constante nas variáveis identificadas é o armazenamento dos dados em um baixo nível de granularidade. Isto permite que análises mais ricas e detalhadas sejam realizadas e que os analistas agreguem os dados da forma que lhes for mais adequado (*drill up*), ao em vez de serem forçados a usar as agregações padrão fornecidas pela plataforma.

5. Conclusões

Neste artigo propomos, a partir de uma revisão de literatura, a identificação de um conjunto de requisitos de dados necessários à utilização de técnicas de MDE e LA em AVAs. Dentre os trabalhos futuros suscitados por esta pesquisa, estão a utilização de um protocolo de revisão sistemática para garantir maior rigor na seleção dos trabalhos utilizados para compilar os requisitos, bem como a verificação e validação dos resultados por meio de entrevistas com especialistas em LA e MDE e desenvolvedores de AVAs.

Agradecimentos

Os autores agradecem ao apoio financeiro da FACEPE (processo BIC-1303-1.03/17) e do CNPq (processos 140973/2017-6 e 475634/2013-6).

Referências

Baker, R. S. J. D., Isotani, S. & Carvalho, A. (2011). "Mineração de Dados Educacionais: Oportunidades para o Brasil. In Revista Brasileira de Informática na Educação", p. 1-11.

- Bergdahl, N., Fors, U., Hernwall, P. & Knutsson, O. (2018). "The Use of Learning Technologies and Student Engagement in Learning Activities". In Nordic Journal of Digital Literacy, p. 113–130.
- Bogarín, A., Cerezo, R., Romero, C. (2018). "A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery", v. 8, n. 1, p. e1230.
- Emmons, S. R., Light, R. P. & Börner, K. (2017). "MOOC visual analytics: Empowering students, teachers, researchers, and platform developers of massively open online courses". In Journal of the Association for Information Science and Technology, p. 2350–2363.
- Koedinger, K. R., Baker, R. S., de Carvalho, A. M. J. A., Raspat, J., Aleven, V. & Corbett, A. T. (2009). "Educational software features that encourage and discourage 'gaming the system'". In Proceedings of the International Conference on Artificial Intelligence in Education, p. 475-482.
- Martins, M. P. G., Migueis, V. L & Fonseca, D. S. B. (2018). "Educational data mining: A literature review. In 13th CISTI, 2018, p. 1-6.
- Papamitsiou, Z. & Economides, A. (2014). "Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence". In Educational Technology & Society, p. 49–64.
- Okoli, C. & Schabram, K. (2010). "A Guide to Conducting a Systematic Literature Review of Information Systems Research," . Sprouts: Working Papers on Information Systems, 10(26).
- Ramos, J. L. C. (2016) "Uma abordagem preditiva da evasão na educação a distância a partir dos construtos da distância transacional". Tese de Doutorado. Recife: CIN/UFPE. Disponível: https://repositorio.ufpe.br/handle/123456789/21052.
- Ray, S. & Saeed, M. (2018). "Applications of Educational Data Mining and Learning Analytics Tools in Handling Big Data in Higher Education". In Applications of Big Data Analytics, p. 135–160.
- Rodrigues, R. L. (2016). "Uma abordagem de mineração de dados educacionais para previsão de desempenho a partir de padrões comportamentais de autorregulação da aprendizagem". Tese de Doutorado. Recife: CIN/UFPE. Disponível: https://repositorio.ufpe.br/handle/123456789/21132.
- Romero, C., Ventura, S., Pechenizkiy, M. & Baker R. S. J. D. (2010). "Handbook of educational data mining", CRC Press.
- Scheffel, M., Drachsler, H., Stoyanov, S. & Specht, M. (2014). "Quality Indicators for Learning Analytics". Educational Technology & Society, p. 117–132.
- Silva, L. A., Silveira, I. F., Silva, L., Rodrigues, R., & Ramos, J. L. C. (2017). "Ciência de Dados Educacionais: definições e convergências entre as áreas de pesquisa". In Workshops do Congresso Brasileiro de Informática na Educação. p. 764-773.
- Vieira, C., Parsons, P. & Byrd, V. (2018). "Visual learning analytics of educational data: A systematic literature review and research agenda". In Computers & Education, p. 119–135.
- You, J. W. (2016). "Identifying significant indicators using LMS data to predict course achievement in online learning". In The Internet and Higher Education, p. 23–30.