

Um Panorama Acerca da Mineração de Dados Educacionais no Brasil

Pedro de Torres Maschio¹, Marcos Alves Vieira²,
Newarney Torrezão da Costa², Sara Luzia de Melo³, Cleon Xavier Pereira Junior²

¹IF Goiano - Campus Campos Belos – Campos Belos-GO – Brasil

²IF Goiano - Campus Iporá – Iporá-GO – Brasil

³Faculdade de Engenharia Elétrica - UFU – Uberlândia-MG – Brasil

pedromaschio.ptm@gmail.com, {marcos.vieira, newarney.costa}@ifgoiano.edu.br

sara.melo@yahoo.com.br, cleon.junior@ifgoiano.edu.br

Abstract. *Educational Data Mining (EDM) is an interdisciplinary field that uses data generated by educational environments in search of relevant informations to the education improvement. This work presents a Systematic Mapping Study (SMS) about the brazilian scenario, using the repository from Especial Comission of Informatics in Education, in relation to EDM. One of the mapping result indicates a high level of data exploration from undergraduate and a lack of basic education data exploration.*

Resumo. *A Mineração de Dados Educacionais (MDE) é uma área interdisciplinar que utiliza dados gerados por ambientes educacionais na busca de informações relevantes ao aprimoramento do ensino. Sendo assim, este trabalho apresenta um Mapeamento Sistemático da Literatura (MSL) sobre o cenário brasileiro, utilizando o repositório da Comissão Especial de Informática na Educação, com relação a MDE. Um dos resultados do mapeamento aponta um alto índice de exploração de dados oriundos da graduação em relação a análise dos dados da educação básica.*

1. Introdução

A educação, a exemplo de outras áreas, sofreu grandes transformações com o advento tecnológico. O ensino tem recebido cada vez mais o suporte de Ambientes Virtuais de Aprendizagem (AVAs) e recursos computacionais para aprimoramento. Além disso, há o crescente uso da modalidade de Ensino a Distância (EaD), que possui forte dependência de sistemas de gestão da aprendizagem (*i.e.*, Moodle¹). Todo este uso computacional na educação vem gerando uma grande quantidade de dados que podem servir como matéria-prima para pesquisa, visando uma constante melhoria na qualidade do ensino.

Uma área que tem como principal objetivo o desenvolvimento de métodos para explorar dados oriundos de ambientes educacionais é conhecida como Mineração de Dados Educacionais (MDE) [Baker et al. 2011]. A MDE explora estatística, aprendizado de máquina e algoritmos de mineração de dados aplicados a diferentes tipos de dados de ensino e permite descobrir novos conhecimentos, de modo a estabelecer bases para um processo

¹<https://moodle.org>

de aprendizagem mais eficaz [Romero and Ventura 2010]. Dentre as investigações possíveis de se realizar utilizando técnicas de MDE, podem ser citados trabalhos com objetivo de verificar potenciais riscos de evasões [Manhães et al. 2011], auxiliar na formação de grupos [Correia and Pimentel 2011] e analisar comportamento [Santos et al. 2015].

Diante do cenário de crescimento da área, este trabalho apresenta um Mapeamento Sistemático da Literatura (MSL) acerca da MDE em esfera nacional, tendo como base de dados os trabalhos publicados desde 2001 e disponíveis no acervo da Comissão Especial de Informática na Educação (CEIE). Assim, esta pesquisa tem como objetivo mapear técnicas, resultados e aplicações de MDE no contexto brasileiro.

2. Método da pesquisa

Para realizar o MSL nesta pesquisa, optou-se por seguir o processo apresentado por [Petersen et al. 2008], onde são estabelecidas cinco etapas, sendo elas: (i) definição das questões de pesquisa; (ii) condução da pesquisa para estudos primários; (iii) seleção de artigos relevantes através de critérios de inclusão e exclusão; (iv) classificação através dos resumos; (v) extração de dados e realização do mapeamento.

Seguindo o procedimento do MSL, foram estabelecidas quatro questões de pesquisa, sendo elas: **QP1:** Quais técnicas de MDE estão sendo utilizadas? **QP2:** Quais investigações têm sido realizadas com MDE? **QP3:** Qual nível de escolaridade (educação básica, superior, etc) tem sido explorado pela MDE? **QP4:** Quais tipos de dados estão sendo considerados relevantes para a pesquisa?

O próximo passo se deu em estabelecer uma *string* e um local para a busca. Esta pesquisa se limitou em explorar os seguintes veículos de publicação: Workshop de Informática na Educação (WIE); Simpósio Brasileiro de Informática na Educação (SBIE), Revista Brasileira de Informática na Educação (RBIE); Workshops do Congresso Brasileiro de Informática na Educação (WCBIE). Foram selecionados todos os trabalhos que continham o termo Mineração de Dados ou *Data Mining* a partir do ano de 2001.

Foram encontrados 98 trabalhos que continham termos da *string* de busca. Ao aplicar os critérios de seleção pelo título e resumo restaram 55 trabalhos. Após leitura na íntegra, 6 trabalhos foram descartados. Por fim, foram analisados 49 trabalhos, conforme gráfico da Figura 1. Os artigos analisados podem ser encontrados em: http://bit.ly/sbie_mde

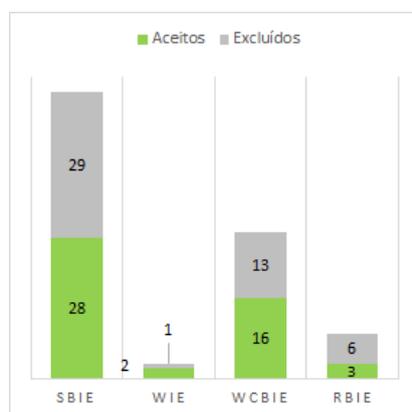


Figura 1. Trabalhos que contêm os termos da *string* de busca.

3. Resultados

3.1. QP1: Quais técnicas de MDE estão sendo utilizadas?

Foi possível observar que, no cenário brasileiro, a MDE utiliza os algoritmos clássicos de mineração de dados para obter resultados a partir dos dados explorados. A Figura 2 apresenta as técnicas utilizadas nos trabalhos analisados.



Figura 2. Técnicas de MDE utilizadas.

Por se tratar, em sua maioria, de classificação (seja previsão de desempenho, evasão, aprovação, etc), grande parte dos trabalhos apresentados fizeram uma análise dos dados utilizando algoritmos de aprendizagem de máquina. Além disso, algumas pesquisas buscaram traçar padrões de comportamento através dos dados educacionais utilizando regras de associação. Em menores proporções, também foram encontradas técnicas utilizando regressão linear, agentes inteligentes e algoritmos para mineração de texto. Em geral, houve um suporte da ferramenta Weka² para realizar o processo de mineração.

3.2. QP2: Quais investigações têm sido realizadas com MDE?

A Figura 3 apresenta um gráfico com as principais investigações dos trabalhos brasileiros em relação a MDE. No total foram encontrados 5 tipos distintos de investigação através da MDE.

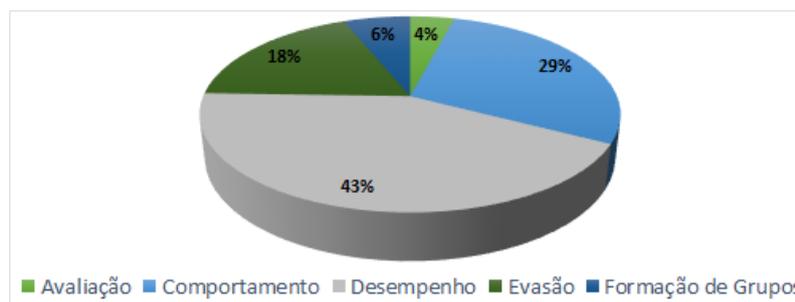


Figura 3. Objetivos com os quais realizou-se MDE.

Conforme observado, a maior parcela dos trabalhos se preocuparam em analisar o desempenho dos aprendizes. Neste caso, é importante ressaltar que houveram trabalhos que focaram somente em reprovação, outros em aprovação, no entanto a maioria dos trabalhos focaram no desempenho geral dos estudantes. Além disso, o comportamento

²Weka é um *software* que possui uma coleção de algoritmos de aprendizagem de máquina.

dos usuários durante a interação com AVAs aparecem com alta frequência de investigação, sendo que foi possível encontrar inclusive trabalhos que analisaram estados de ânimo [Santos et al. 2015]. Trabalhos de formação de grupos, avaliação e evasão também foram evidentes.

3.3. QP3: Qual nível de escolaridade (educação básica, superior, etc) tem sido explorado pela MDE?

Dos 49 trabalhos analisados, 25 apresentaram dados de sistemas educacionais e/ou AVAs presentes no nível de graduação. Este resultado confirma que, apesar de haver iniciativas também para o ensino fundamental e médio, o uso por sistemas computacionais que geram dados educacionais ainda é baixo. A Tabela 1 apresenta os resultados mapeados conforme o nível de escolaridade explorado.

Tabela 1. Escolaridade dos dados avaliados

Nível de Ensino	Quantidade
Fundamental	2
Médio	7
Superior	25
Técnico	1
Não informado	14

Em alguns casos não foi possível verificar o nível de escolaridade, pois são trabalhos que apresentam análise a partir de dados coletados do uso de alguma ferramenta específica que não limita o nível dos usuários. Além disso, observou-se um trabalho que avalia o dados de um curso técnico na modalidade a distância. Apesar da presença em todos os estados brasileiros de Institutos Federais que promovem ensino médio integrado a cursos técnicos, estes não foram encontrados como objetos de avaliação.

3.4. QP4: Quais tipos de dados estão sendo considerados relevantes para a pesquisa?

Para responder a QP4, houve uma busca geral nos trabalhos selecionados visando encontrar quais dados foram levados em consideração para o processo de MDE, conforme Figura 4. Como a maioria dos trabalhos tratam de AVAs na modalidade EaD, foi possível encontrar uma grande investigação sobre interações em fóruns, *chats*, dentre outros, através de avaliações quantitativas. Ainda no mesmo cenário de número de interações, incluíram-se trabalhos que avaliaram a quantidade de acessos aos AVAs, o tempo de permanência e de interação com o ambiente.

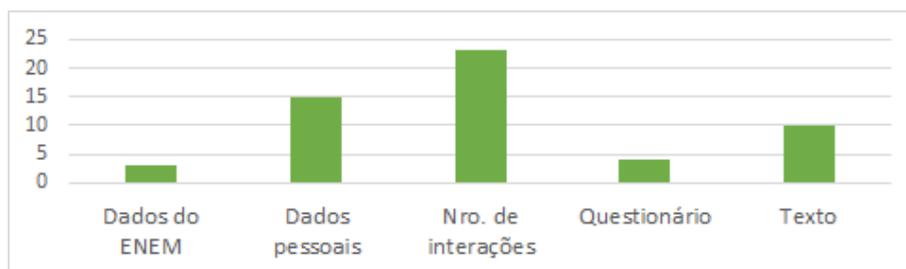


Figura 4. Tipos de dados usados para MDE.

De forma isolada ou em conjunto com os dados pessoais, análises através do número de interações foram bem predominantes nas investigações. Em relação aos dados pessoais, foram incluídos o sexo, a idade, o desempenho em avaliações e em disciplinas, dentre outros. Algumas pesquisas exploraram os dados disponibilizados pelo Governo Federal para investigar a educação básica brasileira. Além do número de interações e dados pessoais, tiveram trabalhos que fizeram mineração sobre textos em fóruns e *chats*. Neste caso, foi possível encontrar trabalhos que visavam analisar comportamentos através das postagens.

4. Considerações finais

Este trabalho apresentou um Mapeamento Sistemático da Literatura acerca do cenário brasileiro com relação à Mineração de Dados Educacionais. Como fonte de busca foram utilizados os artigos disponibilizados na base da CEIE. O MSL contou com 49 trabalhos que apresentavam de forma completa o uso da MDE, ou seja, os trabalhos tiveram que levantar uma investigação, apontar os dados necessários para avaliação, apresentar as técnicas utilizadas para minerar esses dados e, por fim, demonstrar os resultados obtidos a partir da mineração.

Pôde-se concluir que o Brasil tem explorado principalmente dados de ferramentas gerados por meio de interações com estudantes do ensino superior através da modalidade EaD e, apesar desta informação não aparecer de forma gráfica, fica evidente nos trabalhos analisados. Ainda há uma baixa exploração de investigações com relação a educação básica em geral. Além disso, não foi possível encontrar nenhum trabalho com objetivo de investigar dados oriundos de cursos técnicos integrados ao ensino médio. Sendo assim, é possível avaliar algumas oportunidades já relacionadas à realidade atual.

Referências

- Baker, R., Isotani, S., and Carvalho, A. (2011). Mineração de dados educacionais: Oportunidades para o Brasil. *Brazilian Journal of Computers in Education*, 19(02):03.
- Correia, C. F. and Pimentel, E. P. (2011). Mineração de dados na formação de turmas para a recuperação paralela na educação básica. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 1.
- Manhães, L. M. B., Da Cruz, S. M. S., Costa, R. J. M., Zavaleta, J., and Zimbrão, G. (2011). Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 1.
- Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. (2008). Systematic mapping studies in software engineering. In *EASE*, volume 8, pages 68–77.
- Romero, C. and Ventura, S. (2010). Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(6):601–618.
- Santos, F. D., Bercht, M., and Wives, L. (2015). Classificação de alunos desanimados em um avela: uma proposta a partir da mineração de dados educacionais. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 26, page 1052.