

Educational Data Mining versus Learning Analytics: estamos reinventando a roda? Um mapeamento sistemático

Barbara Moissa, Isabela Gasparini, Avanilde Kemczinski

PPGCA e PPGECCMT, Departamento de Ciência da Computação (DCC)
Universidade do Estado de Santa Catarina (UDESC) – Joinville, SC – Brasil

barbara.moissa@gmail.com, {isabela.gasparini,avanilde.kemczinski}@udesc.br

Abstract. *Virtual learning environments allow the gathering of several student data. Two research areas have emerged with the purpose of analyze these data and use the results to enhance learning and teaching processes: Educational Data Mining (EDM) and Learning Analytics (LA). However, since the appearance of the LA field, there is a discussion about this being a new research area or just a new name to the EDM field. This paper seeks to distinguish both areas through a systematic mapping of the literature. As results, differences were found in the use of algorithms, tools and techniques. Thus, it is clear that there are differences between both fields.*

Resumo. *Os ambientes virtuais de aprendizagem permitem que sejam coletados diversos dados sobre os alunos. Duas áreas de pesquisa surgiram com o propósito de analisar estes dados e utilizar os resultados para melhorar o processo de ensino-aprendizagem: Educational Data Mining (EDM) e Learning Analytics (LA). Entretanto, desde o surgimento da área de LA, há uma discussão sobre esta ser uma nova área de pesquisa ou apenas um novo nome para a área de EDM. Este artigo busca diferenciar ambas as áreas através de um mapeamento sistemático da literatura. Como resultados, diferenças foram encontradas no uso de algoritmos, instrumentos e técnicas. Assim, é possível afirmar que existem diferenças entre ambas as áreas.*

1. Introdução

Diversas ferramentas são utilizadas no contexto da educação (e.g., blogs, wikis, Moodle, Blackboard, entre outros), gerando registros cumulativos das atividades de seus usuários [Almazroui 2013]. Neste contexto, duas áreas de pesquisa emergiram: *Learning Analytics* (LA) — em português, analíticas de aprendizagem — e *Educational Data Mining* (EDM) — em português, Mineração de Dados Educacionais.

De acordo com Baker e Inventado (2014) a área de EDM teve seu primeiro trabalho em 1995, seu primeiro workshop relacionado no ano de 2000 e seu primeiro workshop oficial (utilizando o termo *Educational Data Mining*) no ano de 2005. EDM é uma área em crescimento, preocupada em desenvolver métodos para explorar os tipos de dados encontrados em ambientes educacionais e utilizá-los para melhor entender os alunos e o contexto em que estes aprendem [“International Educational Data Mining Society” 2015]. Já a área de LA teve sua primeira conferência realizada em 2010 [Baker and Inventado 2014]. De forma similar, LA é definida como a medida, coleta, análise e relato dos dados sobre os alunos e seus contextos com o objetivo de entender e otimizar o aprendizado e o ambiente em que este ocorre [“1st International Conference on Learning Analytics and Knowledge 2011 (LAK’11)” 2011].

Ambas as áreas possuem uma definição similar, que aborda o uso de dados educacionais para aprimorar o aprendizado do aluno e o ambiente em que ele está inserido e, desde o surgimento da área de LA, há questionamentos sobre haver ou não diferenças entre ambas as áreas. Mais especificamente, o que a área de LA possui de diferente de modo a justificar a sua existência frente à área de EDM? O objetivo deste artigo é diferenciar as áreas de LA e EDM através de um mapeamento sistemático da literatura. Para contextualizar as decisões tomadas durante o mapeamento, descrevemos as áreas na Seção 2. Em seguida, na Seção 3, descrevemos o processo adotado para a seleção de trabalhos e extração dos dados. Na Seção 4, mostramos os resultados obtidos e os discutimos, confrontando-os com afirmações encontradas na literatura. Por fim, realizamos as considerações finais na Seção 5.

2. Educational Data Mining vs. Learning Analytics

O processo de EDM converte os dados brutos de sistemas educacionais em informações úteis que podem impactar na prática e na pesquisa educacional [Romero and Ventura 2010]. De acordo com Romero e Ventura (2010), este processo é similar ao processo de Data Mining: (1) pré-processamento; (2) análise; e (3) pós-processamento. Por sua vez, o processo de LA proposto por Clow (2012), consiste em ciclo no qual os dados são coletados do aluno, analisados e os resultados são utilizados para melhorar algum aspecto no processo de ensino-aprendizagem através de intervenções.

Já o processo de LA proposto Chatti et al. (2012) é similar ao processo de EDM e consiste em um ciclo composto de três etapas: (1) coleta de dados e pré-processamento; (2) análise e ação; (3) pós-processamento. De acordo com Chatti et al. (2012), neste processo há quatro elementos que devem ser considerados: (1) quais os dados que serão analisados (e.g., provenientes do ambiente virtual de aprendizagem ou de mídias sociais); (2) as pessoas envolvidas no processo (e.g., aluno, professor); (3) as técnicas que serão aplicadas para analisar os dados; e (4) os objetivos, que são categorizados como Monitoramento e Análise (MeA); Predição e Intervenção (PeI); Tutoria e Mentoria (TeM); Avaliação e Feedback (AeF); Adaptação (A); Personalização e Recomendação (PeI); e Reflexão (R).

Os processos de LA e EDM são diferentes. Enquanto o processo de LA é um ciclo, independentemente da referência adotada (e.g., Chatti et al. 2012; Clow 2012), o processo de EDM é sequencial e possui um fim. Além disso, no processo de LA a participação de pessoas (e.g., professores, alunos) e as intervenções são consideradas.

As duas áreas compartilham os objetivos de melhorar o processo de ensino-aprendizagem através do aperfeiçoamento dos processos de avaliação, da compreensão dos problemas da educação e do planejamento das intervenções [Siemens and Baker 2012]. Outro objetivo em comum é melhorar a qualidade das análises de dados em larga-escala para dar suporte à pesquisa e à prática [Siemens and Baker 2012].

De acordo com Siemens e Baker (2012), existem cinco diferenças entre as áreas: (1) a descoberta da informação que em EDM é automática e em LA aproveita-se do julgamento humano; (2) a abordagem que em EDM é reducionista (ênfase em reduzir os problemas em componentes e analisá-lo individualmente ou a sua relação com os demais) e em LA é holística (entender o sistema como um todo); (3) nas origens, pois LA originou-se na web semântica, nos currículos inteligentes, entre outros. e EDM originou-se no software educacional, na modelagem dos alunos, entre outros; (4) que

EDM foca na adaptação automática (sem participação humana) enquanto LA foca em informar e capacitar professores e alunos; e (5) nas técnicas e métodos, considerando as técnicas *Social Network Analysis*, *Sentiment Analysis*, *Influence Analysis*, entre outras, como técnicas e LA e as técnicas Classificação, Clusterização, Modelagem Bayesiana, Mineração de Relações, entre outras como técnicas de EDM.

3. Metodologia

Com o objetivo de diferenciar as áreas, um mapeamento sistemático foi realizado. Para isto, adotamos a metodologia proposta por Petersen et al. (2008), composta por quatro etapas que serão explicadas nas próximas subseções.

3.1. Definição das questões de pesquisa

Através deste mapeamento procuramos responder a pergunta “Há diferenças entre as áreas LA e EDM?”. Para isto, com base no exposto na Seção 2, definimos algumas questões de pesquisa que podem nos auxiliar a responder estas perguntas:

- Quais são os objetivos dos trabalhos? É possível classificar os trabalhos de acordo com a classificação proposta por Chatti et al. (2012) para a área de LA?
- Quais são as técnicas, algoritmos e instrumentos utilizados?
- Quais são os produtos dos trabalhos? Quais são as contribuições dos trabalhos?
- São utilizados dados reais? O trabalho considera/menciona aspectos éticos e de privacidade? O trabalho considera/menciona intervenções? Há a participação direta de outras pessoas (e.g. professores, alunos)?

Para encontrar trabalhos e responder a estas perguntas, conduzimos a busca conforme descrito na próxima subseção.

3.2. Condução da busca

A busca por trabalhos foi conduzida nos sete Mecanismos de Busca Acadêmica (MBAs) com melhor resultado na análise realizada por Buchinger et al. (2014): Web of Knowledge, Engineering Village, Scopus SciVerse, IEEE Xplore, ACM Digital Library, Science Direct e Springer Link. Para definir a expressão de busca diversos testes foram realizados. Utilizamos a seguinte expressão para procurar pelos trabalhos através do título, resumo e palavras-chave de artigos publicados entre 2010 e 2014: (“*educational data mining*” OR “*learning analytics*”) AND (*technique* OR *method* OR *approach* OR *application* OR *tool* OR *analysis*).

A próxima subseção aborda os critérios adotados para a seleção dos trabalhos encontrados durante a condução da busca.

3.3. Seleção dos trabalhos

A seleção dos trabalhos consiste em aplicar um conjunto de critérios de inclusão e exclusão de forma a serem incluídos ou excluídos do processo de análise. Incluímos os trabalhos que satisfizeram os seguintes critérios de inclusão (CIs):

- Artigo disponível para download na íntegra;
- Artigo científico de conferência ou revista;
- Estudo primário;

- Artigo completo (i.e., quatro ou mais páginas); e
- Escrito em inglês, português ou espanhol.

Após selecionar os artigos que satisfizeram os critérios de inclusão, foram excluídos todos os que satisfizeram a um ou mais dos critérios de exclusão (CEs):

- Artigo duplicado;
- Trabalho duplicado: diferentes artigos que tratam do mesmo trabalho científico, sendo considerado o trabalho mais detalhado;
- Trabalho sem relação com o foco do mapeamento proposto (e.g., LA e EDM citados apenas como resultados).

Ao aplicar os CIs nos 1.748 artigos encontrados, foram pré-selecionados 780 artigos. Destes, após aplicar os CEs restaram 280 artigos, cujos dados foram extraídos conforme explicado na próxima subseção.

3.4. Extração e mapeamento dos dados

Todos os artigos selecionados foram lidos para extrair os dados. Cada um dos artigos foi classificado como pertencente a LA ou EDM de acordo com o posicionamento dos autores. Foram extraídos os seguintes dados: objetivo; objetivo conforme a classificação proposta por Chatti et al. (2012); produtos; resultados; técnicas; algoritmos; instrumentos; consideração a aspectos éticos e de privacidade; consideração a intervenções; participação de terceiros e utilização de dados reais. Estes dados foram organizados em uma base de dados que nos permitiu analisá-los.

O processo descrito nesta seção foi realizado por três pesquisadores das áreas de Interação Humano-Computador, Informática na Educação e *Learning Analytics*. O processo foi realizado em 5 meses.

4. Resultados

Em relação ao posicionamento dos autores quanto à área de atuação, dos 280 artigos analisados, 66,43% (186 artigos) estão classificados na área de LA; 29,29% (82 artigos) como de EDM; e os outros 4,29% (12 artigos) não foram classificados.

Na Tabela 1, podemos ver os objetivos dos trabalhos de EDM (coluna à esquerda) e LA (coluna à direita) de acordo com o definido pelos autores dos trabalhos (devidamente categorizados), sendo que alguns trabalhos possuem mais de um objetivo. Em EDM foram encontrados 21 objetivos distintos. Os objetivos principais são utilizar técnicas com algum propósito específico, propor uma ferramenta, propor uma abordagem/método e entender o processo de ensino-aprendizagem. Ainda nesta tabela, percebemos que outros 5 objetivos distintos foram encontrados em dois trabalhos cada (totalizando 2,44% de relevância cada) e que outros 7 objetivos distintos foram encontrados uma única vez (cada um com uma relevância de 1,22%). A Tabela 1 também mostra os objetivos de LA, cujos principais objetivos são propor uma ferramenta, entender o processo de ensino-aprendizagem e propor uma abordagem/método. Além desses, nos trabalhos de LA foram encontrados outros 29 objetivos, totalizando 32 objetivos distintos encontrados.

Tabela 1. Objetivos dos trabalhos de (a) EDM e (b) LA

(a) EDM			(b) LA		
Objetivo	n.	%	Objetivo	n.	%
Utilizar uma técnica	15	18,29%	Propor uma ferramenta	52	27,96%
Propor uma ferramenta	12	14,63%	Entender o processo de ensino-aprendizagem	40	21,51%
Propor uma abordagem/método	11	13,41%	Propor uma abordagem/método	27	14,52%
Entender o processo de ensino-aprendizagem	9	10,98%	Propor um framework	9	4,84%
Construir um modelo	6	7,32%	Utilizar uma técnica	9	4,84%
Propor um algoritmo	5	6,1%	Descobrir correlações entre elementos	7	3,76%
Descobrir preditores	4	4,88%	Avaliar uma ferramenta	7	3,76%
Propor uma técnica	3	3,66%	Construir um modelo	6	3,23%
Descobrir correlações entre elementos	3	3,66%	Descrever uma experiência	6	3,23%
5 objetivos	2	2,44%	Descobrir preditores	5	2,69%
7 objetivos	1	1,22%	Propor indicadores	3	1,61%
			5 objetivos	2	1,08%
			16 objetivos	1	0,54%

A Figura 1 mostra a classificação dos trabalhos de cada área conforme a classificação proposta por Chatti et al. (2012) para LA, explicada na seção 2. Conforme a Figura 1, a maioria destes trabalhos estão nas categorias MeA (44,64%), AeF (17,14%), PeI (24,64%). Por ser possível classificar os trabalhos de EDM de acordo com uma classificação proposta para LA, podemos perceber que temos uma grande sobreposição entre as áreas. Entretanto, verificamos também que EDM e LA possuem alguns objetivos que não se enquadram nestas categorias.

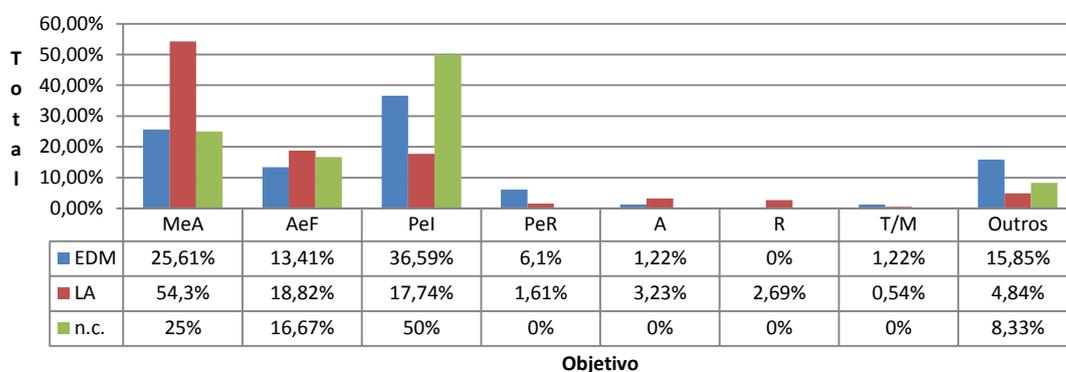


Figura 1. Distribuição dos artigos de acordo com o objetivo

A Tabela 2a e a Tabela 2b mostram as técnicas utilizadas, respectivamente, nos trabalhos de EDM e de LA. As principais técnicas de EDM são: Classificação, Visualização da Informação (InfoVis), Estatística Descritiva e Clusterização. Além destas, outras 24 técnicas distintas foram identificadas. Das técnicas que constam na Tabela 2, as que se destacam em EDM e não em LA são as Regras de Associação, *cross-validation* e *Sequential Pattern Mining*. Já em LA as principais técnicas são InfoVis e Estatística Descritiva. Foram identificadas outras 27 técnicas distintas. Das técnicas citadas nos artigos, as que se destacam em LA e não em EDM são os indicadores e *Social Network Analysis* (não encontrada nos trabalhos de EDM).

Tabela 2. Técnicas identificadas nos trabalhos de EDM e LA

(a) EDM			(b) LA		
Técnica	n.	%	Técnica	n.	%
Classificação	39	47,56%	Visualização da Informação	113	60,75%
Visualização da Informação	28	34,15%	Estatística descritiva	48	25,81%
Estatística descritiva	21	25,61%	Métricas	29	15,59%
Clusterização	19	23,17%	Indicadores	25	13,44%
Regras de Associação	15	18,29%	Social Network Analysis	22	11,83%
Métricas	13	15,85%	Clusterização	19	10,22%
Testes estatísticos	10	12,20%	Testes estatísticos	18	9,68%
Regressão	10	12,20%	Regressão	14	7,53%
Cross-validation	9	10,98%	Estatísticas	13	6,99%
Sequential Pattern Mining	6	7,32%	Classificação	12	6,45%
Descoberta de Conhecimento com Modelos	5	6,10%	Descoberta de Conhecimento com Modelos	6	3,23%
Indicadores	4	4,88%	2 técnicas	5	2,69%
2 técnicas	3	3,66%	2 técnicas	4	2,15%
Filtragem colaborativa	2	2,44%	Transformação de dados	3	1,61%
13 técnicas	1	1,22%	4 técnicas	2	1,08%
Trabalhos em que nenhuma técnica foi identificada	7	8,54%	9 técnicas	1	0,54%
			Trabalhos em que nenhuma técnica foi identificada	25	13,44%

A Tabela 3a mostra os algoritmos identificados nos trabalhos de EDM e a Tabela 3b mostra os algoritmos identificados nos trabalhos de LA. Em EDM os principais algoritmos são J48 (uma variação do algoritmo C4.5 implementada na ferramenta de mineração de dados Weka) e K-means. Além destes outros 102 algoritmos distintos foram identificados. Em LA o algoritmo que se destaca é o K-means, mas não em uma quantidade significativa (3,22%). Na área, outros 31 algoritmos também foram identificados.

Tabela 3. Algoritmos identificados nos trabalhos de EDM (a) e LA (b)

(a) EDM			(b) LA		
Algoritmo	n.	%	Algoritmo	n.	%
J48	13	15,85%	K-means	6	3,23%
K-means	11	13,41%	C4.5	3	1,61%
C4.5	8	9,76%	6 algoritmos	2	1,08%
Apriori	7	8,54%	24 algoritmos	1	0,54%
k-NN	5	6,1%	Trabalhos em que nenhum algoritmo foi identificado	152	81,72%
REPTree	5	6,1%			
Sequential Minimal Optimization	5	6,1%			
jRip	4	4,88%			
7 algoritmos	3	3,66%			
26 algoritmos	2	2,44%			
63 algoritmos	1	1,22%			
Trabalhos em que nenhum algoritmo foi identificado	35	42,68%			

As Tabelas 4a e 4b mostram os instrumentos encontrados (i.e., ferramentas utilizadas no trabalho para coletar ou analisar dados, auxiliar no desenvolvimento de outras ferramentas, padrões de dados, bancos de dados, entre outros), respectivamente, nas áreas de EDM e LA. Em EDM os principais instrumentos foram Weka e Moodle e RapidMiner. Além desses outros 48 instrumentos foram identificados. Se destacam em EDM, mas não em LA, os instrumentos Weka e RapidMiner. Em LA, os principais instrumentos utilizados são MS Excel, Moodle, R, ferramentas próprias e IBM SPSS.

Além destes, outros 196 instrumentos foram identificados. São destaques nesta área e não em EDM os instrumentos MS Excel, R, ferramentas próprias, ontologias e o Google Docs.

Tabela 4. Instrumentos identificados nos trabalhos de EDM e LA

(a) EDM			(b) LA		
Instrumento	n.	%	Instrumento	n.	%
Weka	23	28,05%	MS Excel	12	6,45%
Moodle	12	14,63%	Moodle	11	5,91%
Rapid Miner	8	9,76%	R	11	5,91%
MySQL	5	6,1%	Ferramenta própria	11	5,91%
IBM SPSS	5	6,1%	IBM SPSS	10	5,38%
SCORM	4	4,88%	Ontologias	9	4,84%
Próprio	3	3,66%	MySQL	8	4,3%
MS SQL Server	3	3,66%	Google Docs	6	3,23%
MS Excel	2	2,44%	6 instrumentos	4	2,15%
Experience API	2	2,44%	10 instrumentos	3	1,61%
41 instrumentos	1	1,22%	18 instrumentos	2	1,08%
Trabalhos em que nenhum instrumento foi identificado	20	24,39%	159 instrumentos	1	0,54%
			Trabalhos em que nenhum instrumento foi identificado	49	26,34%

As Tabelas 5a e 5b mostram, respectivamente, os produtos dos trabalhos analisados de EDM e LA. Como produto consideramos o que foi feito para que os resultados fossem obtidos. Em EDM os principais produtos são análises, ferramentas e abordagens/métodos. Em LA, os principais produtos são ferramentas e análises.

Tabela 5. Produtos identificados nos trabalhos de EDM e LA

(a) EDM			(b) LA		
Produto	n.	%	Produto	n.	%
Análise	22	26,83%	Ferramenta	77	41,4%
Ferramenta	20	24,39%	Análise	67	36,02%
Abordagem/Método	19	23,17%	Estudo de caso	32	17,2%
Avaliação	8	9,76%	Avaliação	27	14,52%
Estudo de caso	13	15,85%	Framework	17	9,14%
Comparação	12	14,63%	Abordagem/Método	15	8,06%
Experimento	7	8,54%	Modelo	11	5,91%
Modelo	7	8,54%	Algoritmo	5	2,69%
Algoritmo	6	7,32%	Comparação	4	2,15%
Aplicação	6	7,32%	Validação	3	1,61%
Validação	4	4,88%	Experimento	2	1,08%
Técnica	2	2,44%	Relato de experiência	2	1,08%
Framework	1	1,22%	Heurísticas	1	0,54%
			Indicadores	1	0,54%

A Tabela 6a mostra as contribuições dos trabalhos de EDM. Como contribuição consideramos a informação mais relevante do trabalho — e.g., uma análise (produto) que resultou em conhecimento sobre o processo de ensino-aprendizagem (contribuição); descrição uma ferramenta (produto e contribuição); descrição de uma ferramenta (produto) que foi avaliada (produto) e gerando conhecimento sobre como suas funcionalidades interferem no processo de ensino-aprendizagem (contribuição); uma comparação (produto) de ferramentas gerando conhecimento sobre estas (contribuição); entre outros. As principais contribuições são abordagens/métodos, ferramentas e conhecimento sobre o processo de ensino-aprendizagem. A Tabela 6b mostra as contribuições dos trabalhos de LA. As principais contribuições são ferramentas, conhecimento sobre o processo de ensino-aprendizagem e abordagens/métodos.

Tabela 6. Contribuições dos trabalhos de EDM

(a) EDM			(b) LA		
Contribuição	n.	%	Contribuição	n.	%
Abordagem/Método	19	23,17%	Ferramenta	44	23,66%
Ferramenta	13	15,85%	Conhecimento sobre o processo de ensino-aprendizagem	44	23,66%
Conhecimento sobre o processo de ensino-aprendizagem	11	13,41%	Abordagem/método	28	15,05%
Conhecimento sobre técnicas	8	9,76%	Framework	16	8,6%
Aplicação	8	9,76%	Modelo	12	6,45%
Algoritmo	6	7,32%	Aplicação	10	5,38%
Correlação entre elementos	4	4,88%	Conhecimento sobre ferramentas	8	4,3%
Modelo	4	4,88%	Conhecimento sobre técnicas	7	3,76%
Conhecimento sobre modelos	3	3,66%	Correlação entre elementos	6	3,23%
Preditores	2	2,44%	Preditores	5	2,69%
Técnica	2	2,44%	Indicadores	4	2,15%
6 contribuições	1	1,22%	Relato de experiência	4	2,15%
			Algoritmo	3	1,61%
			6 contribuições	1	0,54%

Dos trabalhos de EDM, apenas um (1,22%) considerou aspectos éticos e de privacidade. Os trabalhos de LA também tiveram uma baixa consideração (13 trabalhos, 6,99%) a estes aspectos. Em relação às intervenções, 17 trabalhos (20,73%) de EDM consideraram intervenções. Nos trabalhos de LA a consideração a intervenções foi um pouco mais alta (55 trabalhos, 29,57%). Em EDM os trabalhos utilizaram dados reais. Já em LA a maioria dos trabalhos utilizou dados reais, entretanto, 3 desses trabalhos (1,61%), além dos dados reais, utilizaram dados sintéticos (dados criados para realizar testes).

O último aspecto analisado foi a participação de outras pessoas. Por participação direta entendemos que a pessoa participou na pesquisa, por exemplo, respondendo questionários ou realizando alguma atividade. Por participação indireta entendemos que as pessoas utilizaram algum sistema e tiveram seus dados coletados (de forma similar a uma ferramenta *web analytics*). Na maioria dos trabalhos de EDM (68,29%) e de LA (43,01%) houve a participação indireta de outras pessoas. A participação direta aconteceu em 24,39% dos trabalhos de EDM e em 30,11% dos trabalhos de LA.

5. Discussão

Primeiramente, é importante ressaltar que os resultados deste trabalho estão limitados aos trabalhos encontrados no processo descrito na Seção 3. Considerando que a área de EDM é mais antiga que a área de LA [Baker and Inventado 2014], é lógico pensar que a amostra analisada neste mapeamento deveria conter mais trabalhos de EDM. Entretanto, a quantidade de trabalhos da área de LA é significativamente maior do que a quantidade de trabalhos da área de EDM. Uma explicação pode estar relacionada com a expressão de busca adotada, visto que esta define quais artigos serão encontrados pelos MBAs. É possível que os termos definidos para buscar os artigos descrevam melhor a área de LA (que representa 66,43% dos artigos analisados neste mapeamento).

Desta forma, apesar de conseguirmos aplicar a classificação definida por Chatti et al. (2012) nos trabalhos de EDM analisados, em um contexto mais amplo da área isso pode não ser possível, mas ainda assim podemos perceber algumas diferenças (discutidas a seguir) que poderão guiar outras abordagens na diferenciação das áreas.

De acordo com Baker e Inventado (2014), as técnicas de *Domain Structure Discovery*, Mineração de Relações e Descoberta com Modelos são mais comuns nos trabalhos de EDM do que nos trabalhos de LA. Porém, conforme apresentado na Tabela 2a, técnicas de *Domain Structure Discovery* não foram identificadas nos trabalhos relacionados a área de EDM. Já as técnicas de Mineração de Relações (e.g., mineração de regras de associação, mineração de correlações, *Sequential Pattern Mining*, *Causal Data Mining*) foram mais encontradas nos trabalhos de EDM (21 trabalhos, 26%) do que nos trabalhos de LA (4 trabalhos, 2%). A técnica Descoberta de Conhecimento com Modelos obteve valores similares, com 5 trabalhos (6%) em EDM e 6 trabalhos (3%) em LA. Como a amostra de trabalhos de LA é maior, pode-se considerar que esta técnica é mais comum em EDM já que, proporcionalmente, consta no dobro de trabalhos.

Ainda de acordo com Baker e Inventado (2014), técnicas de análise de texto (e.g., *Text Mining*, *Discourse Analysis*, *Natural Language Processing*) se destacam em LA, sendo ocasionalmente descritas em EDM. Esta informação pode ser confirmada de forma sutil, pois em LA estas técnicas foram identificadas em 8 trabalhos (4%), enquanto que em EDM há 2 trabalhos (2%). Outra técnica que tem maior destaque em LA é SNA [Baker and Inventado 2014], sendo encontrada apenas nos trabalhos classificados como pertencentes a esta área.

Em LA poucos algoritmos foram identificados, mas foram descritos muitos instrumentos. Isto sugere que os pesquisadores da área estão preocupados mais com os resultados obtidos ao aplicar determinadas técnicas do que com os algoritmos propriamente ditos; indo ao encontro da afirmação de Siemens e Baker (2012) de que LA possui uma visão mais holística e EDM possui uma visão mais reducionista. Os dados expostos na Tabela 6 também contribuem com esta sugestão, visto que estes mostram que em EDM há mais contribuições com abordagens/métodos (23,17%) enquanto que em LA há mais contribuições com ferramentas (23,66%) e conhecimento sobre o processo de ensino-aprendizagem (23,66%).

6. Considerações finais

Através do estudo das áreas de EDM e LA e do mapeamento sistemático realizado, podemos verificar que as duas áreas possuem definições e objetivos similares. Ambas possuem como objetivo analisar dados educacionais para entender o processo de ensino-aprendizagem e melhorá-lo. Os resultados do mapeamento apresentado neste artigo corroboram que os objetivos de ambas as áreas são similares. Em ambas as áreas análises são conduzidas de forma similar, utilizando as mesmas técnicas e métodos. Poucas diferenças são encontradas. Em EDM predomina o uso de técnicas típicas de Data Mining (e.g., clusterização, classificação) e de técnicas estatísticas. Em LA são utilizadas as mesmas técnicas, focando no uso de técnicas de Visualização da Informação (que são consideravelmente mais utilizadas). Nos trabalhos de LA analisados, encontramos uma técnica não utilizada nos trabalhos de EDM: *Social Network Analysis*.

Outras diferenças encontradas entre as áreas estão no uso dos algoritmos (maior em EDM) e instrumentos (maior em LA). Em relação aos dados (i.e., reais ou não), às intervenções e aos aspectos éticos e de privacidade nenhuma diferença significativa foi

encontrada. Os produtos de cada área também são similares, visto que ambas as áreas abordam problemas através de ferramentas e análises.

Com base neste mapeamento, concluímos que há diferenças sutis entre as áreas. A principal delas está relacionada ao foco de pesquisa. LA busca principalmente entender o processo de ensino-aprendizagem como um todo ou desenvolver ferramentas para auxiliar neste processo, em uma abordagem mais voltada para os fatores humanos e utiliza as técnicas de EDM como um instrumento. Já EDM procura aplicar determinadas técnicas em novos contextos e desenvolver métodos para realizar as análises. Entretanto, ambas as áreas podem colaborar entre si para atingir o seu objetivo comum: melhorar o processo de ensino-aprendizagem através do uso da tecnologia.

Com os dados deste mapeamento é possível realizar outras análises (e.g., analisar a triangulação dos dados de produtos e contribuições, verificar o uso de algoritmos com o uso de instrumentos) para descobrir diferenças ainda não identificadas entre as áreas. Este mapeamento também pode contribuir com outras abordagens para diferenciar as áreas e auxiliar novos pesquisadores a identificarem diferenças entre as áreas para melhor categorizar seus trabalhos.

Referências

- 1st International Conference on Learning Analytics and Knowledge 2011 (LAK'11) (2011). <https://tekri.athabascau.ca/analytics/>, [accessed on Sep 28].
- Almazroui, Y. A. (2013). A survey of Data mining in the context of E-learning. *International Journal of Information Technology & Computer Science (IJITCS)*, v. 7, n. 3, p. 8–18.
- Baker, R. S. J. De and Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In: Larusson, J. A.; White, B.[Eds.]. *Learning Analytics: From Research to Practice*. Springer New York. p. 61–75.
- Buchinger, D., Cavalcanti, G. A. de S. and Hounsell, M. da S. (28 apr 2014). Mecanismos de busca acadêmica: uma análise quantitativa. *Revista Brasileira de Computação Aplicada*, v. 6, n. 1, p. 108–120.
- Chatti, M. A., Dyckhoff, A. L., Schroeder, U. and Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, v. 4, n. 5/6, p. 318–331.
- Clow, D. (2012). The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. ACM.
- International Educational Data Mining Society (2015). <http://www.educationaldatamining.org/>, [accessed on Sep 28].
- Petersen, K., Feldt, R., Mujtaba, S. and Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering. In *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE'08)*. British Computer Society.
- Romero, C. and Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 40, n. 6, p. 601–618.
- Siemens, G. and Baker, R. S. J. d. (2012). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. ACM Press.